



Projekt SKALA

Abschlussbericht

Zitierung

Landeskriminalamt NRW (2018): Abschlussbericht Projekt SKALA. Düsseldorf.

Das Wichtigste in Kürze

Projekt SKALA – Was wurde getan?

- Ziel des Projektes SKALA war es, die Möglichkeiten und Grenzen der Prognose von Kriminalitätsbrennpunkten sowie die Effizienz und Effektivität darauf aufbauender polizeilicher Interventionen im Rahmen eines Pilotversuchs zu prüfen.
- Das Landeskriminalamt Nordrhein-Westfalen hat hierzu ein eigenes, unabhängiges System zur Kriminalitätsprognose entwickelt.
- Die Modell- und Prognoseerstellung wurde für Wohnungseinbruchdiebstahl, Einbruchdiebstahl aus Gewerbeobjekten und Kraftfahrzeugdelikte durchgeführt.
- Zur Prognose der Kriminalitätsrisiken (Predictive Policing) wurden statistische Wahrscheinlichkeiten der ausgewählten Delikte auf Basis raum- und zeitbezogener Daten berechnet.
- Die Berechnungsmodelle sind transparent und im Detail nachvollziehbar.
- Das im Projekt entwickelte Visualisierungs- und Auswertetool SKALA | MAP steht zukünftig allen Polizeibehörden des Landes Nordrhein-Westfalen zur Darstellung und Auswertung von Kriminalität – auch unabhängig von Kriminalitätsprognosen – zur Verfügung.

Projekt SKALA – Was zeigen die Ergebnisse?

- Die Einbruchwahrscheinlichkeit ist in ausgewählten Prognosegebieten häufig drei- bis vierfach höher als in anderen Wohnquartieren eines Polizeibezirks.
- Der Einfluss ausgewählter Variablen im Prognosemodell ist saisonal und je nach Polizeibezirk unterschiedlich.
- Die Prognosegüte hängt entscheidend von der Qualität und zeitlichen Verfügbarkeit der Daten ab.
- SKALA kann die Einsatzplanung von zentraler Stelle unterstützen und bietet neue Ansatzpunkte für die Kriminalprävention.

Inhaltsverzeichnis

| | | | |
|--|------------|--|-----------|
| Glossar | | | |
| 1 Einführung | 1 | | |
| 1.1 Hintergrund | 1 | | |
| 1.2 Projekt SKALA | 2 | | |
| 1.3 Projektvorgehen | 2 | | |
| 2 Projektorganisation | 5 | | |
| 2.1 Auftraggeber | 5 | | |
| 2.2 Projektleitung und Geschäftsführung | 5 | | |
| 2.3 Datenschutz | 5 | | |
| 2.4 Öffentlichkeitsarbeit | 6 | | |
| 2.5 Teilprojekte | 6 | | |
| 2.6 Zusammenarbeit | 7 | | |
| 2.7 Schnittstellen | 7 | | |
| 2.8 Berichtswesen und Controlling | 7 | | |
| 2.9 Kommunikation | 7 | | |
| 2.10 Kosten und Ressourcen | 7 | | |
| 3 Anlage der Untersuchung | 9 | | |
| 3.1 Ansprüche und Grundsätze | 9 | | |
| 3.2 Hypothesengestütztes Vorgehen | 10 | | |
| 3.2.1 Rational-Choice-Theorien | 11 | | |
| 3.2.2 Routine-Activity- und Lifestyle-Theorie | 12 | | |
| 3.2.3 Theory of Target Search/Crime-Pattern-Theory | 14 | | |
| 3.2.4 Desorganisationstheorie | 15 | | |
| 3.2.5 Broken-Windows-Theorie | 16 | | |
| 3.2.6 Defensible-Space-Ansatz | 17 | | |
| 3.2.7 Near-Repeat-Ansätze | 19 | | |
| 3.2.8 Professionelle Akteure/Expertenannahmen | 21 | | |
| 3.3 Operationalisierung und Indikatorenbildung | 22 | | |
| 3.4 IT-Hard- und Software | 22 | | |
| 3.5 Datenschutz | 23 | | |
| 3.6 Öffentlichkeitsarbeit | 25 | | |
| 3.7 Predictive Policing als Prozess | 26 | | |
| 3.8 Prognoseflächen als räumliche Bezugsgröße | 27 | | |
| 3.9 Prognosezeit als temporale Bezugsgröße | 31 | | |
| 3.10 Datenqualität im Predictive-Policing-Prozess | 31 | | |
| 3.11 Qualitätsmetriken | 31 | | |
| | | 3.11.1 Statistisch seltene Ereignisse | 32 |
| | | 3.11.2 Die „Trefferrate“ (Hit Rate) | 32 |
| | | 3.11.3 Der Predictive Accuracy Index | 33 |
| | | 3.11.4 Der Standardized Accuracy Efficiency Index | 33 |
| | | 3.11.5 Die Konfusionsmatrix | 34 |
| | | 3.11.6 Die Fallzahldifferenz | 34 |
| | | 3.12 Variabilität und Validität der Qualitätsmetriken | 34 |
| | | 3.12.1 Einflussdimension: Prognose-Delikt | 35 |
| | | 3.12.2 Einflussdimension: Prognose-Dauer | 35 |
| | | 3.12.3 Einflussdimension: Prognose-Raum | 36 |
| | | 3.12.4 Bewertung | 37 |
| | 4 | Methodische Umsetzung der Untersuchung | 39 |
| | 4.1 | Einordnung des methodischen Vorgehens | 39 |
| | 4.2 | Darstellung und Interpretation von Prognosen | 39 |
| | 4.3 | Modellerstellung | 41 |
| | 4.3.1 | 1. Schritt: Datenaufbereitung | 42 |
| | 4.3.2 | 2. Schritt: Georeferenzierung der Daten | 43 |
| | 4.3.3 | 3. Schritt: Modell- und Prognoseerstellung | 44 |
| | 4.3.4 | 4. Schritt: Visualisierung | 47 |
| | 4.3.5 | Analytisches Modell | 50 |
| | 4.4 | Testung von Prognosemodellen | 51 |
| | 4.4.1 | Entscheidungsbaummodelle | 51 |
| | 4.4.2 | Random Forest | 54 |
| | 4.4.3 | Weitere Modellansätze | 54 |
| | 4.4.4 | Modellvalidierung | 54 |
| | 4.5 | Testung von Variableneinflüssen im Modell | 55 |
| | 4.5.1 | Analyse soziostruktureller Daten | 56 |
| | 4.5.2 | Analyse der Modellperformance | 61 |
| | 4.6 | RZC-Analyse und Scoreberechnung | 63 |
| | 4.6.1 | Durchführung | 64 |
| | 4.6.2 | Bewertung | 66 |
| | 4.7 | Evaluierung der RZC-Scores | 69 |
| | 4.7.1 | Methodik | 69 |
| | 4.7.2 | Bewertung | 71 |
| | 4.8 | Prognose anderer Delikte | 75 |
| | 4.9 | Austausch und Kooperationen | 77 |

| | | |
|----------|---|-----------|
| 5 | Praktische Umsetzung der Prognosen | 79 |
| 5.1 | Beschreibung der Organisation | 79 |
| 5.2 | Vorbereitung der Umsetzung | 79 |
| 5.3 | Umsetzung | 79 |
| 5.4 | Bewertung der Umsetzung | 80 |
| 6 | Fazit und Ausblick | 83 |
| 6.1 | Bewertung der Ergebnisse | 83 |
| 6.2 | Ausblick | 84 |
| | Abbildungsverzeichnis | 85 |
| | Tabellenverzeichnis | 87 |
| | Literatur | 89 |

Glossar

Ausreißer: In der Statistik spricht man von einem Ausreißer, wenn ein Beobachtungswert scheinbar nicht in eine erhobene Messreihe passt, also den Erwartungen widerspricht. Im Allgemeinen handelt es sich dabei um besonders große oder kleine Messwerte.

Binäres Merkmal: Siehe *Dichotomie*.

Clusteranalyse: Bei der Clusteranalyse werden die Elemente einer Stichprobe zu Gruppen zusammengefasst. Zur Einteilung der Gruppen (Cluster) können verschiedenste Merkmale wie beispielsweise soziostrukturelle Daten verwendet werden. Ziel ist es, dass die Elemente jedes Cluster möglichst ähnliche Eigenschaften haben und gleichzeitig zwischen den Clustern der maximale Unterschied besteht.

Data Mining: Beim Data Mining wird mittels mathematischer/statistischer Verfahren versucht, bislang unbekannte Strukturen oder Zusammenhänge in großen Datenbeständen zu entdecken.

Datenaggregation: Datenaggregation meint die Hochskalierung oder Zusammenfassung von kleinräumigen Daten (z. B. auf Gebäude- oder Straßenabschnittsebene) zu übergeordneten Einheiten (z. B. Wohnquartierebene). Bei rational skalierten Daten wird in der Regel der Mittelwert oder die Summe aller Punktdaten im jeweiligen Gebiet berechnet.

Dichotomie: Dichotomie beschreibt eine Aufteilung eines Merkmals in genau zwei Ausprägungen (z. B. das Merkmal „Einbruch“ mit den Ausprägungen „ja“ oder „nein“).

Entscheidungsbaum: Als Entscheidungsbaum wird in der Statistik ein hierarchisches Modell bezeichnet, bei welchem der Datensatz mittels sogenannter „Entscheidungsknoten“ in mehrere sich gegenseitig ausschließende Untergruppen aufgeteilt wird.

Entscheidungsknoten: Ein Entscheidungsknoten ist ein Element (Blatt) eines Entscheidungsbaummodells, das die Daten anhand einer Variable entsprechend ihres Wertebereiches aufspaltet.

Exploration: Die Exploration bezeichnet in der Statistik die Annäherung an ein Phänomen ohne Vorwissen bezüglich bestehender Zusammenhänge. Sie dient zumeist der Bildung von Hypothesen.

Gamification: Gamification umfasst die Anwendung spielerischer Elemente in einem spielfremden Zusammenhang

mit dem Ziel der Motivationssteigerung. Die Aufgabenbewältigung wird vor allem durch Spielelemente unterstützt, die eine individuelle oder kollaborative Problemlösung fördern.

Geoinformationssystem (GIS): Bei einem GIS im engeren Sinne handelt es sich um eine Software, die komplexe Geoverarbeitungs- und -analyseprozesse ermöglicht. Im Vordergrund steht das Organisieren und Analysieren räumlicher Daten.

Georeferenzierung: Unter Georeferenzierung versteht man die Zuweisung raumbezogener Informationen zu einem Datensatz.

Gini-Index: Der Gini-Index bzw. Gini-Koeffizient ist ein gängiges statistisches Maß, welches Ungleichverteilungen beschreibt. Ungleichverteilungsmaße zeigen an, zu welchem Grad zwei Variablen von einer Gleichverteilung abweichen. Der Gini-Index wird aus der Lorenz-Kurve abgeleitet. Berechnungsgrundlage sind aufsteigend sortierte Listen der zwei untersuchten Variablen. Der Gini-Koeffizient reicht von 0 (völlige Gleichverteilung) bis 1 (maximale Ungleichverteilung).

Heatmap: Eine Heatmap ist eine zweidimensionale Darstellung von Daten anhand verschiedener Farben. Die geografische Dichte der Daten wird dabei, in Anlehnung an ein Wärmebild, farblich abgestuft visualisiert.

Hellfeld: Das Hellfeld bezeichnet die bekannt gewordene bzw. registrierte Kriminalität. Amtliche Statistiken wie die Polizeiliche Kriminalstatistik dokumentieren ausschließlich dieses so genannte Hellfeld.

Hotspot-Analyse: Bei einer Hotspot-Analyse werden, unter Rückgriff auf historische Kriminalitätsdaten, Orte identifiziert, die sich in einem bestimmten zeitlichen Kontext durch eine hohe Kriminalitätsbelastung auszeichnen. Dabei liegt der Fokus auf der Unterscheidung von Gebieten mit hohem und niedrigem Kriminalitätsaufkommen.

Indikator: Grundsätzlich ist ein Indikator ein mess- und/oder berechenbarer Index, welcher eine Variable beschreibt, deren Ausprägungen selbst nicht unmittelbar gemessen werden können.

Koinzidenz: Koinzidenz bedeutet das Zusammentreffen von mehreren Merkmalen oder Ereignissen.

Konfusionsmatrix: Die Konfusionsmatrix ist ein klassisches Schema für die Berechnung von Qualitätsmetriken in den Sozialwissenschaften.

Korrelationsmatrix: Die Korrelationsmatrix ist ein Schema von Korrelationskoeffizienten für mehrere Variablen, wobei Korrelationskoeffizienten den Zusammenhang zweier Variablen darstellen.

Messniveau (Skalenniveau): Das Messniveau oder Skalenniveau beschreibt die Eigenschaft von Merkmalen hinsichtlich ihrer Ausprägungen. Dabei unterscheidet man nominal, ordinal und metrisch skalierte Merkmale. Nominal skalierte Daten besitzen verschiedene Ausprägungen, die aber in keine Reihenfolge gebracht werden können. Ordinal skalierte Daten sind dadurch gekennzeichnet, dass sie in eine sinnvolle Reihenfolge gebracht, aber nicht miteinander verrechnet werden können (natürliche Rangfolge). Metrisch skalierte Daten können in eine sinnvolle Reihenfolge gebracht und miteinander verrechnet werden.

Mittelwert (arithmetischer): Der arithmetische Mittelwert (umgangssprachlich auch als Mittelwert oder Durchschnitt bezeichnet) stellt ein Lagemaß für Verteilungen dar. Er wird berechnet, indem man alle auftretenden Ausprägungen einer Variable aufsummiert und die Summe anschließend durch die Anzahl der Ausprägungen dividiert. Einzelne, extrem hohe oder niedrige Werte können den arithmetischen Mittelwert besonders im Falle geringer Fallzahlen stark beeinflussen.

Modalwert: Der Modalwert oder Modus stellt ein Lagemaß für Verteilungen dar und bezeichnet die am häufigsten vorkommende Ausprägung einer Variable.

Modellgüte: Die Modellgüte ist die berechnete Vorhersagekraft (*performance*) eines Modells anhand statistischer Indizes. Hierbei werden berechnete (modellerte/prognostizierte) und tatsächliche Werte gegenübergestellt.

Modellperformance: siehe *Modellgüte*.

Normalisierung: Bei der Normalisierung werden Variablen mit eventuell unterschiedlichen Einheiten in vergleichbare Werte desselben Skalenbereichs (z. B. 0 bis 1) transformiert, um diese anschließend miteinander verrechnen zu können.

Operationalisierung: Operationalisierung meint die Übersetzung von sozialwissenschaftlichen Hypothesen in konkrete, empirisch beobachtbare Indikatoren.

Overfitting (Überanpassung): *Overfitting* ist der Zustand, bei welchem ein Modell zu viele Eingabevariablen und/oder Parameter enthält. Dadurch besteht die Gefahr, dass ein Prognosemodell, das anhand eines bestimmten Datensatzes trainiert wurde, nicht optimal auf andere Daten übertragbar ist und somit nicht die bestmöglichen Analyseergebnisse erzielt.

Polygon: Polygone sind abgeschlossene geometrische Flächen, die durch die geografischen Koordinaten einzelner Punkte und der Verbindungslinien zwischen diesen Punkten begrenzt werden.

Prädiktor: Als Prädiktoren werden Variablen bezeichnet, die genutzt werden, um Werte einer anderen Variable (Zielvariable) vorherzusagen.

Precision: Die Precision zeigt hier an, in wie viel Prozent der Prognosegebiete tatsächlich in der Folgeweche eingebrochen wurde und ist demnach ein Index für die Wahrscheinlichkeit, dass in ein prognostiziertes Gebiet tatsächlich eingebrochen wird.

Predictive Policing: Predictive Policing setzt sich aus den englischen Wörtern *predict* (vorhersagen) und *policing* (Polizeiarbeit) zusammen und bedeutet vorausschauende Polizeiarbeit.

Prognosegebiet: Die Gebiete (hier Wohnquartiere), für die die höchsten Kriminalitätswahrscheinlichkeiten im Vergleich zu den anderen Gebieten des gesamten Prognosebereichs berechnet wurden, werden hier als Prognosegebiete definiert. Dabei wurde ihr Anteil auf etwa 1,5 Prozent der Gesamtquartieranzahl jedes Polizeibezirkes beschränkt

Prognosegüte: Die Prognosegüte meint hier den Grad an „Korrektheit“ einer erstellten Prognose, welche mithilfe statistischer Qualitätsmetriken berechnet werden kann.

Pruning: *Pruning* ist das „Zurückschneiden“ eines erstellten Entscheidungsbaumes.

Qualitätsmetrik: Eine Qualitätsmetrik ist ein statistischer Index zur Berechnung der Modell- oder Prognosegüte.

Rangkorrelationskoeffizient (Spearman): Der Rangkorrelationskoeffizient beschreibt den Zusammenhang zweier Variablen und bildet diesen auf einem Wertebereich von „-1“ bis „+1“ ab.

Recall: Der Recall zeigt hier den relativen Anteil der Wohnquartiere an, die anteilig an den tatsächlichen WED der Folgeweche durchschnittlich in der Prognose(-auswahl) enthalten waren. Er ist somit ein Indikator für die Wahrscheinlichkeit, dass ein WED-Gebiet prognostiziert wird.

Regressionsanalyse: Bei der Regressionsanalyse handelt es sich um ein statistisches Analyseverfahren, das die Beziehungen einer abhängigen zu einer oder mehreren unabhängigen Variablen zu modellieren versucht.

Robustheit (Modelle): Die Robustheit von Modellen bezieht sich auf die Übertragbarkeit eines (Prognose)modells auf neue Datensätze (Werte der Vorhersagevariablen) und

eine damit verbundene hohe Modellgüte auch bei Eingangsdaten, die nicht bei der Modellerstellung verwendet wurden.

Score: Der Score ist eine statistische Maßzahl (Index) und ergibt sich aus der (gewichteten) Summe von Merkmalen.

Shapefiles: Shapefiles sind ein Dateiformat für Geodaten, das ursprünglich für die Software ArcView der Firma ESRI entwickelt wurde.

Signifikanz: Die statistische Signifikanz gibt an, wie hoch die Wahrscheinlichkeit ist, dass ein in einer Stichprobe gefundenes Ergebnis nicht zufällig ist, sondern die Grundgesamtheit abbildet.

Variable: Eine Variable bezeichnet ein Merkmal, das empirisch erfasst werden kann und unterschiedliche Ausprägungen annimmt. Beispielsweise hat die Variable Geschlecht die Ausprägungen „männlich“ und „weiblich“.

Validität: Validität bezieht sich hier auf die Verlässlichkeit und Korrektheit von Qualitätsmetriken zur Berechnung einer Prognosegüte.

1 Einführung

Zur Einführung wird im Folgenden das Projekt „System zur Kriminalitätsauswertung und Lageantizipation“ (SKALA) vorgestellt. Zunächst werden der Hintergrund des Projektes sowie der Projektanlass dargelegt (Kapitel 1.1). Nachfolgend werden das Projekt und seine Ziele beschrieben (Kapitel 1.2). Das Kapitel schließt mit einem Überblick zum Projektvorgehen (Kapitel 1.3).

1.1 Hintergrund

Mit Erlass vom 05.02.2015 – 422-62.02.02 Projekt Predictive Policing – beauftragte das Ministerium für Inneres und Kommunales des Landes Nordrhein-Westfalen das Landeskriminalamt Nordrhein-Westfalen (LKA NRW) mit der Durchführung eines Forschungsprojektes zu Predictive Policing. Bei Predictive Policing¹ handelt es sich um vorausschauende Polizeiarbeit. Auf der Basis verschiedener, kriminalitätsrelevanter Datenquellen wird anhand komplexer mathematischer Verfahren prognostiziert, wo in Zukunft die Entwicklung von Brennpunkten im räumlichen Kriminalitätsgeschehen wahrscheinlich ist. Derartige Verfahren werden in einzelnen Ländern bereits angewendet.

Die Funktionsweise solcher Verfahren besteht darin, unter Rückgriff auf komplexe Datenverarbeitungsprogramme, verschiedene Datenquellen miteinander in Beziehung zu setzen. Zentral ist die raum- und zeitbezogene Zusammenführung der Datenquellen, die es ermöglichen soll, Strukturen und Zusammenhänge zwischen potenziellen Einflussfaktoren und dem Kriminalitätsgeschehen zu bestimmten Zeiten und an gewissen Orten zu erkennen und so Kriminalitätsentwicklungen besser einschätzen zu können. Primärer Zweck ist die Unterstützung einer strategischen und zielgerichteten Polizeiarbeit, die auf Basis bekannter, kriminalitätsrelevanter Faktoren mögliche Brennpunkte frühzeitig identifiziert. Dadurch soll ein ressourceneffizienter Einsatz der Polizeikräfte², und im Idealfall eine Senkung der Kriminalitätshäufigkeit erreicht werden.

Die Aussagen einiger Softwareanbieter, mit ihren Predictive Policing-Umsetzungen künftige Kriminalitätsbrennpunkte vorhersagen und damit zur Verhinderung von Kriminalität beitragen zu können, sind in Medien, Politik und Polizei viel diskutiert worden. Aus wissenschaftlicher Sicht war ein derartiges Unterfangen aber zunächst grundsätzlich kritisch zu sehen: Beispielsweise hat Wolfgang Streeck (2009: 11), der Leiter des Max-Planck-Instituts für Gesellschaftsforschung,

die Unmöglichkeit der Vorhersage von sozialen Phänomenen wie folgt begründet: „Jede Betrachtung gesellschaftlicher Prozesse hat es mit Fallzahlen zu tun, die niedriger sind als die Zahl der Faktoren, die als Erklärung infrage kommen“. Das gilt uneingeschränkt auch für Fragen der Kriminalitätsentwicklung. Damit mussten die von den Softwareanbietern in Aussicht gestellten, vielversprechenden Möglichkeiten zur gesicherten Prognose des Kriminalitätsgeschehens aus wissenschaftlicher Sicht zunächst mit Vorsicht behandelt werden. Systematische Evaluationsstudien, die den eindeutigen, kausalen Erfolg derartiger Interventionen belegen, liegen bislang nicht ausreichend vor.³

Darüber hinaus zeichnen sich die besagten Prognoseprogramme in der Regel (i. d. R.) durch einen ausgeprägten Datenempirismus aus: Sie basieren üblicherweise auf Techniken des Data Mining⁴, zu dem komplexe Software benötigt wird. Hierbei handelt es sich um eine so genannte explorative Vorgehensweise, die üblicherweise frei von inhaltlichen oder theoretischen Vorannahmen ist. Diese bietet zwar den Vorteil, dass auch unerwartete Zusammenhänge aufgedeckt werden können, die im Vorhinein möglicherweise nicht antizipiert und erst durch die neuartigen technischen Möglichkeiten erkannt werden. Die Vorgehensweise birgt jedoch auch Gefahren: Kausale Wirkungszusammenhänge, die möglicherweise wissenschaftlich bereits beforscht sind, können allein anhand solcher mathematisch-statistischer Verfahren nicht von Zufallsergebnissen und inhaltlich bedeutungslosen Koinzidenzen mehrerer Merkmale unterschieden werden. Eine derartig theorieleose Analyse von Daten ist aus wissenschaftlicher Perspektive ausschließlich zur Exploration eines bislang unbekanntes Phänomens zu akzeptieren, nicht aber um Auffälligkeiten und Muster in einem bereits beforschten inhaltlichen Bereich wie der raum-zeitlichen Verteilung von Kriminalität in Großstädten zu erklären. Da-

¹ „Der Begriff Predictive Policing setzt sich aus dem englischen „predict“ (vorhersagen) und „policing“ (Polizeiarbeit) zusammen, kurz: vorhersagende Polizeiarbeit“ (Pollich/Bode 2017: 3).

² Aus Gründen der leichteren Lesbarkeit wird in dem vorliegenden Abschlussbericht die gewohnte männliche Sprachform bei personenbezogenen Substantiven und Pronomen verwendet. Dies impliziert keine Benachteiligung des weiblichen Geschlechts, sondern soll im Sinne der sprachlichen Vereinfachung als geschlechtsneutral zu verstehen sein.

³ Eine Ausnahme für den deutschsprachigen Raum bildet die kürzlich veröffentlichte Evaluation des Max-Planck-Instituts zur Software Precobs, die sich allerdings nur auf einen Evaluationszeitraum von sechs Monaten bezieht und auch keinen eindeutigen, kausalen Wirkungszusammenhang feststellen kann (vgl. Gerstner 2017).

⁴ Beim Data Mining wird mittels mathematischer/statistischer Verfahren versucht, bislang unbekanntes Strukturen oder Zusammenhänge in großen Datenbeständen zu entdecken. Zur Vertiefung wird auf Ester/Sander 2000 verwiesen.

neben war auch der Nutzen einer statistischen Kriminalitätsprognose für die praktische polizeiliche Aufgabenwahrnehmung bisher nicht geprüft.

Diese Diskrepanz zwischen den Aussagen der Softwareindustrie und der kritischen Haltung der Wissenschaft war Anlass, ein wissenschaftlich begleitetes Praxisprojekt zur Methodik von Predictive Policing zu initiieren. In diesem Rahmen sollte mit mehreren Pilotbehörden in Nordrhein-Westfalen (NRW) ein entsprechendes Verfahren der vorausschauenden Polizeiarbeit eingesetzt, wissenschaftlich begleitet und dessen Nutzen und Praktikabilität geprüft werden. Die elementaren Bereiche

- eines hypothesengestützten Vorgehens
- eigener softwaretechnischer Lösungen und
- kriminalfachlicher Expertise

sollten hierbei gezielt vereint werden, um raum- und zeitbezogene Prognosen zu erstellen und für NRW ein eigenständiges System zur Kriminalitätsprognose zu entwickeln und zu testen.

1.2 Projekt SKALA

Das Projekt startete am 25.02.2015 unter Federführung der Abteilung 3 des LKA NRW zunächst unter der Projektbezeichnung „Predictive Policing“. Im Rahmen des als Pilotversuch gestalteten Projektes wurden die Kreispolizeibehörden (KPB) Duisburg und Köln als Pilotbehörden ausgewählt. Die Wahl fiel auf diese Behörden, weil es sich um große, strukturell ähnliche Bezirke mit einer erhöhten Kriminalitätsbelastung, einer heterogenen Stadtviertelstruktur sowie speziellen Kriminalitätsbrennpunkten handelt. Die existierenden räumlichen und strukturellen Unterschiede zwischen den Bezirken ermöglichten es außerdem, erkannte Muster zu vergleichen und hinsichtlich konkreter räumlicher oder sonstiger Gegebenheiten vor Ort besser einordnen zu können.

Am 01.01.2017 wurde das Projekt, verbunden mit einer Erweiterung auf die Polizeibezirke Düsseldorf, Essen und Gelsenkirchen sowie Bonn (ab dem 01.06.2017), in das Projekt SKALA (System zur Kriminalitätsauswertung und Lageantizipation) überführt.⁵ Ziel dieser Erweiterung während des laufenden Projektes war es, mit den bis dahin gewonnenen Erkenntnissen der wissenschaftlichen Begleitung, Analyse und Evaluation, anhand einer breiteren Datenbasis, Vergleichsmöglichkeiten zwischen verschiedenen Polizeibezirken zu schaffen. Dies ermöglichte den erweiterten Abgleich von Prognosemodellen unter verschiedenen räumlichen Bedingungen.

Die Hauptziele des Projektes SKALA bestanden darin, (1) die Möglichkeiten und Grenzen der Prognose von Kriminalitätsbrennpunkten sowie (2) die Effizienz und Effektivität darauf aufbauender polizeilicher Interventionen im Rahmen eines Pilotversuchs zu prüfen. Im Fokus des vorliegenden Berichts steht das erstgenannte Ziel. Zum zweiten Ziel wird ein gesonderter Evaluationsbericht vorgelegt.

Aus Gründen der Komplexitätsreduktion wurde im Projekt SKALA zunächst eine Festlegung auf ein einzelnes Delikt, den Wohnungseinbruchdiebstahl (WED), vorgenommen. Einerseits ist dieses Delikt (aus theoretisch-wissenschaftlicher Sicht) durch Interventionen (polizeilich) beeinflussbar und raum-zeitlich variabel, andererseits bot es sich aufgrund des aktuell hohen Fallaufkommens und des sich daraus ergebenden polizeilichen Handlungsdrucks für die Untersuchung an. Das Verfahren sollte zudem auf seine Geeignetheit für andere Deliktfelder geprüft werden.

1.3 Projektvorgehen

Nachfolgend werden die im Projekt getätigten Arbeitsschritte in Form von Teilzielen dargelegt. Im Verlauf des Berichtes werden diese Schritte umfassender erläutert.

Teilziel 1: Wissenschaftlich-theoriegeleitete Hypothesen sind erstellt.

Nach einer Aufarbeitung des relevanten Forschungsstandes wurde eine wissenschaftlich-theoriegeleitete Hypothesenbildung bezüglich der potenziell erklärenden Merkmale von Kriminalitätsbrennpunkten des WED vorgenommen. Die Ableitung geeigneter Hypothesen aus dem Forschungsstand fand unter temporärer Hinzuziehung externer Experten aus den Bereichen Kriminalgeografie und Stadtsoziologie statt (siehe Kapitel 3.2).

Teilziel 2: Daten zur Prüfung der Hypothesen sind bereitgestellt.

Aufbauend auf den entwickelten Hypothesen erfolgte die gezielte Auswahl relevanter Datenquellen, die zu einer empirischen Prüfung geeignet sind. Die erforderlichen Daten ergaben sich aus den in den Hypothesen identifizierten potenziellen Vorhersagevariablen für die raum-zeitliche Verbreitung des WED.

Neben polizeilichen Daten (aus dem Vorgangsbearbeitungssystem der Polizei NRW), die Informationen zum Kriminalitätsaufkommen in bestimmten Stadtbezirken liefern, gehörten auch Daten zur Bevölkerungsstruktur der verschiedenen Stadtbezirke (beispielsweise quartierbezogene Altersstrukturen, Haushaltsstrukturen, Zu- und Wegzüge) zum Kerndatenbestand. Je nach Inhalt der aufgestellten Hypothesen kam die Verwendung weiterer Datenquellen in Betracht.

⁵ Im Nachfolgenden wird sprachlich ausschließlich auf das Projekt SKALA eingegangen, wenngleich hierbei auch Phasen des Projektes Predictive Policing betroffen sind.

Hierzu gehörten beispielsweise Daten über Verkehrsinfrastrukturen oder meteorologische Daten. Darüber hinaus wurden zur Exploration möglicher unerwarteter Zusammenhänge zusätzliche Datenquellen hinzugezogen (siehe Kapitel 3.3).

Teilziel 3: Eine deutschlandweite Ausschreibung der Software ist erfolgreich abgeschlossen.

Nach einer Erhebung der Anforderungen an die zu beschaffende Software zur Auswertung der Daten (fachliche und technische Anforderungen, Support) wurde die Ausschreibung auf dem Vergabemarktplatz NRW veröffentlicht. Als Ergebnis des Vergabeverfahrens wurde eine geeignete Software identifiziert und beschafft (siehe Kapitel 3.4).

Teilziel 4: Die Hypothesen sind geprüft, unerwartete Zusammenhänge wissenschaftlich bewertet und erste Prognosen sind erstellt.

Nach der Beschaffung der Software erfolgte die Prüfung der wissenschaftlich generierten Hypothesen. Zur Prüfung der Aussagekraft der ausgewählten Merkmale wurde die Kriminalitätsentwicklung vergangener Jahre auf die Prognostizierbarkeit von WED durch verschiedene (ebenfalls in der Vergangenheit erhobene) Merkmale aus den Datenbeständen hin untersucht. Nur durch das Zugrundelegen theoretisch und inhaltlich begründeter Annahmen, die zudem einer empirischen Prüfung auf Basis retrograder Daten standgehalten haben, konnten verlässliche Erklärungsansätze des künftigen Kriminalitätsaufkommens generiert werden.

Um allerdings die potenziellen Vorteile des Data Mining nicht ungenutzt zu lassen, wurden neben der theorie- und hypothesengestützten Untersuchung auch explorative Analysen durchgeführt. Hierdurch entdeckte unerwartete Effekte wurden dann mit wissenschaftlichen Methoden weiter analysiert. Dies war beispielsweise durch Hinzuziehen weiterer theoretischer Erklärungsmodelle oder anderer Datenquellen realisierbar. Nachdem die wissenschaftlich generierten Hypothesen auf Basis retrograder Daten geprüft und relevante Vorhersagevariablen von Brennpunkten des WED identifiziert worden waren, wurde die Modellierung der Prognose begonnen. In dieser Projektphase spielten auch Testläufe mit verschiedenen Modellvarianten (beispielsweise hinsichtlich des Gültigkeitszeitraumes) eine bedeutende Rolle, die je nach Komplexität einen hohen Zeiteinsatz und Expertenwissen erforderten (siehe Kapitel 4.3).

Teilziel 5: Die Prognosen sind in einem geeigneten Format an die Pilotbehörden übermittelt.

In Absprache mit den Pilotbehörden wurde ein geeignetes Format für die Praxis erarbeitet und entsprechend umgesetzt (siehe Kapitel 5.3).

Teilziel 6: Interventionen auf Basis der Prognoseergebnisse sind in den Pilotbehörden umgesetzt.

Auf Basis der erstellten Kriminalitätsprognosen haben die Pilotbehörden präventive und repressive Maßnahmen getroffen. Die Maßnahmen wurden in eigener Zuständigkeit entwickelt und umgesetzt (siehe gesonderten Evaluationsbericht).

Teilziel 7: Die Bedeutungen von weiteren geografischen Räumen sind auf ihre Vergleichsmöglichkeiten untersucht.

Mit der Erweiterung des Projektes auf die Polizeibezirke Düsseldorf, Essen, Gelsenkirchen (ab 01.01.2017) und Bonn (ab 01.06.2017) wurde es ermöglicht, die Relevanz von zusätzlichen raumbezogenen Besonderheiten für die Entstehungszusammenhänge des WED und dessen Prognose zu beurteilen. Die sich daraus erschließenden Erkenntnisse flossen in die weitere Modell- und Prognoseerstellung ein und erlaubten die Einschätzung der Übertragbarkeit auf bislang noch nicht einbezogene Prognoseflächen (siehe Kapitel 4.5).

Teilziel 8: Weitere technisch-methodische Analysemodelle sind geprüft.

Die Erkenntnisse im Projekt dokumentierten, dass ergänzende Analysemodelle zur Erreichung der Projektziele geprüft werden müssen. Denkbar waren beispielsweise weitere methodisch-statistische Ansätze sowie stärker geografisch ausgerichtete Analysemodelle beziehungsweise (bzw.) Visualisierungsmethoden. Im Projekt SKALA wurde eine solche tiefgehende Prüfung durchgeführt (siehe Kapitel 4.3.5).

Teilziel 9: Andere Kriminalitätsphänomene sind auf ihre Relevanz für Kriminalitätsprognosen hin geprüft.

Die Analysen konzentrierten sich zunächst auf den WED. Im weiteren Projektverlauf wurden auch andere Deliktfelder (Einbruchdiebstahl aus Gewerbeobjekten und Kraftfahrzeugdelikte) auf ihre Prognostizierbarkeit hin geprüft (siehe Kapitel 4.8).

Die Laufzeit des Projektes betrug knapp drei Jahre. Der Start des Projektes war Februar 2015, die operative Feldphase endete im Dezember 2017. Die einzelnen Arbeitsschritte lassen sich mit einer Zeitleiste visualisieren (siehe Abbildung 1).

Abbildung 1: Zeitleiste

Februar 2015**Vorbereitungsphase**

- Rekrutierung/Einarbeitung Personal
- Beschaffung Hard-/Software
- Akquise von Daten
- Erstellung von Prognosemodellen (WED)

Oktober 2015**Dezember 2017****Umsetzungsphase**

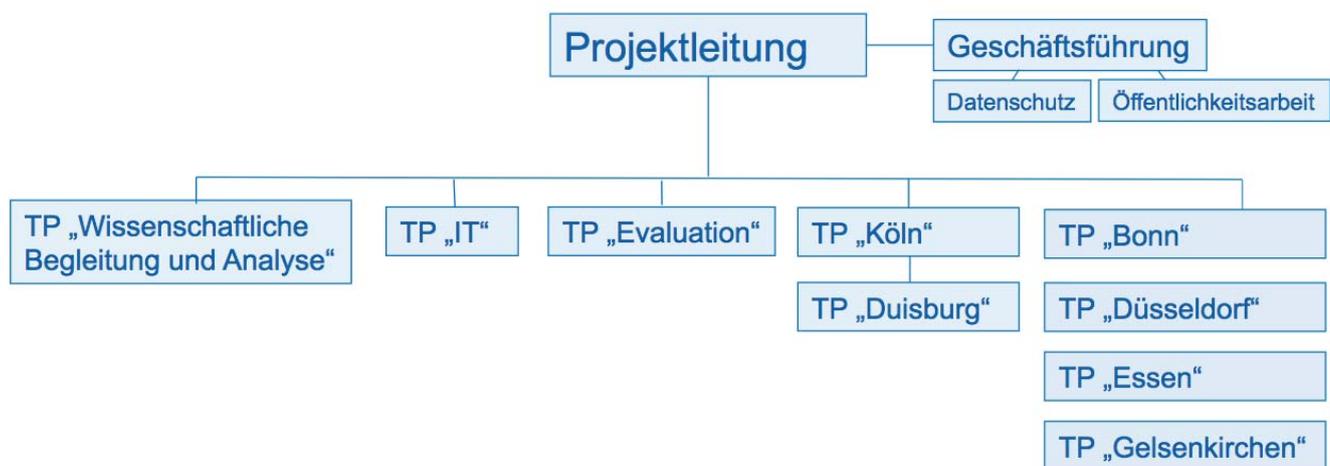
- Erstellung von Kriminalitätsprognosen
- Umsetzung von Maßnahmen
- Ständige Weiterentwicklung der Modelle
- Prüfung von weiteren Prognosemodellen (Geschäfts-ED, Kfz-Diebstahl,...)



2 Projektorganisation

Nachfolgend wird die Projektorganisation dargestellt, beginnend mit der Benennung bzw. Beschreibung des Auftraggebers (Kapitel 2.1), der Projektleitung mit Geschäftsführung (Kapitel 2.2), des Datenschutzes (Kapitel 2.3), der Öffentlichkeitsarbeit (Kapitel 2.4) sowie der einzelnen Teilprojekte (Kapitel 2.5). Anschließend werden kurz die Zusammenarbeit innerhalb des Projektes (Kapitel 2.6) sowie Schnittstellen des Projektes (Kapitel 2.7) dargelegt. Abschließend wird auf die Berichtslegung und das Controlling (Kapitel 2.8), die Kommunikation innerhalb des Projektes (Kapitel 2.9) sowie auf Kosten und Ressourcen (Kapitel 2.10) eingegangen. Zur übersichtlichen Darstellung der Projektorganisation wird nachfolgend zunächst ein entsprechendes Schaubild visualisiert (siehe Abbildung 2).

Abbildung 2: Projektorganisation



2.1 Auftraggeber

Auftraggeber war das Ministerium für Inneres und Kommunales des Landes Nordrhein-Westfalens.

2.2 Projektleitung und Geschäftsführung

Auftragnehmer war das LKA NRW. Mit der Leitung des Projektes wurde der Leiter der Abteilung „Strategische Kriminalitätsbekämpfung“ des LKA NRW, Herr Joachim Eschmann, beauftragt. Stellvertretende Projektleiterin und zugleich Geschäftsführerin war Frau Daniela Robach. Der Projektleiter vertrat das Projekt nach außen. Seiner Geschäftsführung wurden zudem die Bereiche Öffentlichkeitsarbeit wie auch Datenschutz unmittelbar zugeordnet. Der Projektleiter trug zudem die Verantwortung für die technische, organisatorische und inhaltliche Umsetzung, insbesondere für

- die Erreichung der Projektziele
- die Festlegung, Koordination und Überwachung der Vorgehensweise im Projekt
- das Übertragen von Aufgaben und Verantwortung unter Berücksichtigung der individuellen Fähigkeiten der Mitarbeiterinnen und Mitarbeiter

- die Motivation der Mitarbeiterinnen und Mitarbeiter, Konfliktlösungen und ausgewogene Arbeitsverteilung und
- die Erstellung der Projektberichte.

Die Geschäftsführerin trug insbesondere die Verantwortung für folgende Aufgaben der Projektorganisation:

- Erarbeiten und Aktualisieren des Planes für den Einsatz der Mitarbeiterinnen und Mitarbeiter entsprechend dem Projektablauf und Festlegen von Aufwand und Terminen
- Sachliche und termingerechte Projektabwicklung
- Projektdokumentation

2.3 Datenschutz

Von Projektbeginn an wurde das Vorhaben datenschutzrechtlich begleitet, um eventuelle Probleme hinsichtlich der Verwendung von Datenquellen oder des Zusammenführens verschiedener Datensätze frühzeitig zu erkennen. Personenbezogene Daten wurden nicht verwendet. Der Datenschutzbeauftragte des LKA NRW stand in regelmäßigem Austausch mit der Landesbeauftragten für Datenschutz und Informationsfreiheit NRW (siehe Kapitel 3.5).

2.4 Öffentlichkeitsarbeit

Der Bereich Öffentlichkeitsarbeit (ÖA) konzentrierte sich auf zwei Schwerpunkte: Zur Herstellung von Offenheit und Akzeptanz innerhalb der Polizei wurde bereits frühzeitig ein Konzept zur internen Öffentlichkeitsarbeit entworfen und umgesetzt. Nicht zuletzt aufgrund des hohen medialen Interesses an Predictive Policing wurde ebenfalls ein stringentes Konzept zur externen Öffentlichkeitsarbeit entwickelt und realisiert (siehe Kapitel 3.6).

2.5 Teilprojekte

Das Projekt SKALA wurde aufgrund seiner Komplexität in mehrere Teilprojekte gegliedert:

Teilprojekt Wissenschaftliche Begleitung und Analyse (TP WBA)

Das TP WBA setzte sich aus sieben festen Mitarbeiterinnen und Mitarbeitern zusammen. Die Leitung erfolgte anfänglich durch Frau Prof. Dr. Daniela Pollich und später durch Herrn Dr. Felix Bode. Vier Mitarbeiterinnen und Mitarbeiter verfügten über kriminologisches Hintergrundwissen (Studium oder Promotion im Bereich der Kriminologie oder Soziologie) und zwei über wissenschaftliche Expertise im Bereich Data Science (Promotion im Bereich Geografie bzw. Physik).

Im Rahmen dieses Teilprojektes erfolgte zunächst die Erstellung der Hypothesen auf Basis wissenschaftlicher Theorien und bisheriger Forschungsergebnisse. Die Hypothesen wurden anhand ausgewählter Daten zunächst retrograd überprüft. Darauf aufbauend wurden erste Kriminalitätsprognosen modelliert (siehe Kapitel 4.3). Weiterhin oblag es dem Teilprojekt, unerwartete Ergebnisse auf deren inhaltliche Aussagekraft zu überprüfen. Hierfür wurden im Rahmen des Teilprojektes geeignete Interventionen der wissenschaftlichen Prüfung in Reaktion auf den jeweiligen Befund entwickelt und umgesetzt (siehe z. B. Kapitel 4.3.5).

Im Projektverlauf hat es sich darüber hinaus als unerlässlich erwiesen, auf externe Expertise, beispielsweise von Universitäten, zurückzugreifen. Das TP WBA suchte, unterhielt und pflegte die entsprechenden Kontakte, identifizierte Beratungsbedarfe und koordinierte die Durchführung von Analysen, die ein entsprechendes Expertenwissen einbeziehen. Insbesondere die Zusammenarbeit mit dem Lehrstuhl für Datenanalyse und Visualisierung der Universität Konstanz war von großem Wert für die Projektdurchführung.

Teilprojekt IT (TP IT)

Das TP IT wurde durch Herrn Oliver Heinze geleitet. Das Teilprojekt war für die Durchführung des Vergabeverfahrens zur Auswahl der erforderlichen Soft- und Hardware sowie die entsprechenden Schulungen und den Support verantwortlich. Eine Bereitstellung der Daten, die sich zur Prüfung der generierten Hypothesen als notwendig erwiesen hat, oblag ebenfalls dem TP IT. Weiterhin waren die Mitarbeiter

des Teilprojektes mit der Pflege und Betreuung des Auswertesystems (Hard- und Software) und der technischen Unterstützung des TP WBA betraut (siehe Kapitel 3.4).

Das TP IT fertigte zudem in regelmäßigen Abständen Sicherheitskopien der Daten und ein Systembackup an und stellte die technische Lauffähigkeit des regelmäßigen Prognoseverfahrens sicher.

Teilprojekt Evaluation (TP Evaluation)

Die Evaluation wurde in einem eigenen Teilprojekt organisiert. Gewählt wurde eine Mischform aus externer und interner Evaluation. Das Evaluationsteam zur gemeinsamen Durchführung bestand aus Herrn Dr. Christian Erzberger (Gesellschaft für innovative Sozialforschung und Sozialplanung e. V./GISS-EV aus Bremen) und der Zentralstelle Evaluation (ZEVA) des LKA NRW. Als eigenständiges Sachgebiet vereint die ZEVA seit 14 Jahren methodische Kompetenz, Institutionsverständnis und polizeiliche Feldkenntnisse. Die Leitung erfolgte durch Herrn Ingo Dungs. Die Verantwortung für die Erstellung des Evaluationsberichtes lag bei Herrn Dr. Christian Erzberger.

Das TP Evaluation hatte die Aufgabe, die Einführung von SKALA im Zuge einer Prozessevaluation zu begleiten und – im Zuge einer Wirkungsevaluation – zu ermitteln, welchen Mehrwert SKALA für den Bereich der Organisation, aber auch hinsichtlich der Effektivität polizeilicher Interventionen, die auf Kriminalitätsprognosen aufbauen, erbringen kann. Die Evaluation sollte im Schwerpunkt evidenzbasierte Informationen zur Entscheidung über die Einführung von SKALA liefern und im Falle einer weiteren Umsetzung den Rückgriff auf robuste Erkenntnisse zulassen (Evaluationszweck). Ergebnisse, Schlussfolgerungen und Empfehlungen beruhen dabei auf systematisch empirisch ermittelten Daten.

Das Design der Evaluation wurde summativ in drei Komplexe untergliedert. Während in der Analysephase Fragen hinsichtlich der Praktikabilität, der *usability* und der Qualität der Kriminalitätsprognosen im Vordergrund standen, zielte eine Untersuchung in der Umsetzungsphase auf besondere Bedingungen für das Gelingen, Umsetzungshürden und Konzepttreue in den Pilotbehörden ab. In einem dritten Komplex sollte den Wirkungen und Effekten unter dem Gesichtspunkt nachgegangen werden, ob diese auf die Kriminalitätsprognosen und die sich anschließenden Interventionen kausal zurückgeführt werden können. Näheres ist dem gesonderten Evaluationsbericht zu entnehmen.

Teilprojekte Umsetzung Bonn (BN), Duisburg (DU), Düsseldorf (D), Köln (K), Essen (E) und Gelsenkirchen (GE)

Für die Pilotbehörden wurden jeweils gesonderte Teilprojekte geschaffen. Die Zuständigkeit der Pilotbehörden bestand in erster Linie darin, auf Basis der Kriminalitätsprog-

nosen geeignete präventive und/oder repressive Maßnahmen zu initiieren und sich an der Identifizierung der relevanten (polizeilichen) Daten zu beteiligen sowie sich bereits frühzeitig in die Umsetzung des Projektes einzubinden. Die konkrete Ausgestaltung dieser Maßnahmen oblag den Pilotbehörden. Sie standen hierbei im engen und ständigen Austausch mit dem TP WBA und dem TP Evaluation.

2.6 Zusammenarbeit

Alle Projektmitglieder gewährleisteten während der Projektarbeit die bestmögliche Abstimmung mit allen Beteiligten. Die jeweils verantwortlichen Projektmitglieder stimmten sich selbstständig untereinander und mit den anderen Projektmitgliedern ab, wenn Schnittstellen zu anderen Arbeitspaketen erkannt wurden. Sie sorgten für die fortgesetzte Informationsweitergabe der Arbeitsergebnisse an die Geschäftsführerin bzw. den Projektleiter.

2.7 Schnittstellen

Die Schnittstellen und Abhängigkeiten waren innerhalb des Projektes wie folgt feststellbar:

- Forschungsprojekt WED der Kriminalistisch-Kriminologischen Forschungsstelle (KKF) - LKA NRW
- Dezernat 31 - LKA NRW
- Abteilung 4 - LKA NRW
- Landesamt für Zentrale Polizeiliche Dienste (LZPD) NRW

2.8 Berichtswesen und Controlling

Der Projektleiter legte dem Auftraggeber ab dem 01.04.2015 insgesamt neun Quartalsberichte vor. Das Projekt endet mit Abgabe dieses Abschlussberichtes.

2.9 Kommunikation

Der Projektleiter oder die Geschäftsführerin luden zu Besprechungen und Workshops des Projektes ein. Zur Informationssteuerung wurde ein Funktionspostfach (SKALA@polizei.nrw.de) eingerichtet, das von der Geschäftsführung betreut wurde. Für die Dauer des Projektes wurde zudem die SharePoint-Technologie des LZPD NRW als zentrale Kommunikationsplattform genutzt. Teilprojektleiter-Sitzungen fanden alle zwei Wochen im LKA NRW in Düsseldorf statt. In den Teilprojektleiter-Sitzungen erfolgte ein intensiver Austausch über die aktuelle Entwicklung des Projektes.

2.10 Kosten und Ressourcen

Die veranschlagten Kosten für das Projekt betragen 639.000,00 Euro. Im Projektzeitraum wurden tatsächlich 511.742,38 Euro verausgabt (Stand 18.12.2017). Die Aufgaben der Projektmitglieder konnten grundsätzlich in der Regelarbeitszeit wahrgenommen werden. Für manche Projektmitglieder ergaben sich phasenbedingt Belastungen, die über die reguläre wöchentliche Arbeitszeit hinausgingen.

3 Anlage der Untersuchung

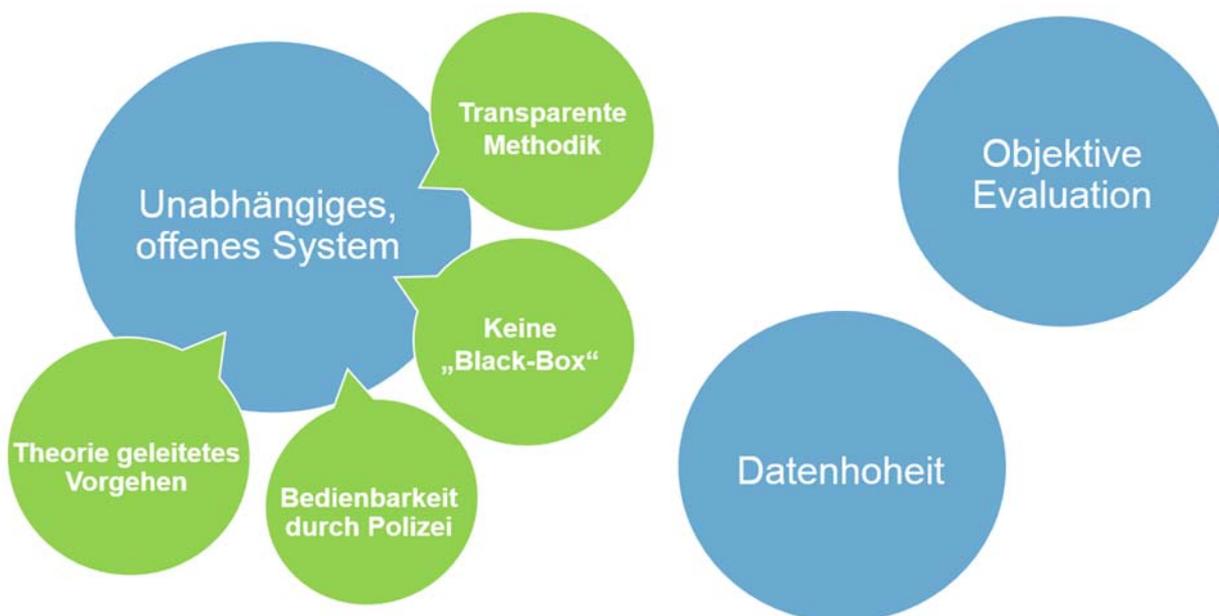
Das dritte Kapitel widmet sich der Anlage der Untersuchung. Im Fokus stehen die Rahmenbedingungen des Projektes. Hierfür werden zunächst die Ansprüche und Grundsätze im Projekt dargestellt (Kapitel 3.1). Anschließend wird das hypothesengestützte Vorgehen erläutert (Kapitel 3.2) sowie die damit in Verbindung stehende Operationalisierung und Indikatorenbildung (Kapitel 3.3). Neben einer Beschreibung der genutzten IT-Hard- und Software (Kapitel 3.4), dem Umgang mit Datenschutz (Kapitel 3.5) und der Öffentlichkeitsarbeit im Projekt (Kapitel 3.6) wird auch auf den Predictive-Policing-Prozess eingegangen (Kapitel 3.7). Hierbei werden Wohnquartiere als räumliche Bezugsgröße (Kapitel 3.8) und die Prognosezeit als temporale Bezugsgröße (Kapitel 3.9) thematisiert, ebenso wie die Datenqualität im Predictive-Policing-Prozess (Kapitel 3.10). Das Kapitel schließt mit einer Darstellung von gängigen Qualitätsmetriken im Bereich von Predictive Policing (Kapitel 3.11) sowie den damit einhergehenden Problemen von Variabilität und Validität (Kapitel 3.12).

3.1 Ansprüche und Grundsätze

Mit Beginn des Projektes bestand der Anspruch, das Projekt zu evaluieren, die Datenhoheit zu behalten und die Prüfung der Methodik von Predictive Policing unabhängig und mit einem offenen System (Software) durchzuführen. Durch den Anspruch an ein offenes und unabhängiges System sollte der Forderung einer transparenten, nachvollziehbaren Methodik Rechnung getragen werden. Darüber hinaus bestand der Anspruch, das System durch Polizeibedienstete bedienbar zu halten. Abschließend sollte die Prüfung der Methodik

von Predictive Policing theoriegeleitet verlaufen. Das bedeutet, dass sich methodisch nicht nur auf den bereits in der kriminologischen und polizeiwissenschaftlichen Literatur empirisch bewährten Near-Repeat-Ansatz fokussiert wurde, sondern auch weitere Kriminalitätstheorien und -ansätze Eingang in die Prüfung fanden. Abbildung 3 illustriert den Anspruch und die Philosophie im Projekt.

Abbildung 3: Ansprüche und Grundsätze



3.2 Hypothesengestütztes Vorgehen

Zunächst wurden exemplarisch raum-zeitliche Prognosen für den WED modelliert. Sowohl zur Phänomenologie und Erklärung dieses speziellen Deliktes, als auch zur Verteilung von Kriminalitätsphänomenen in Raum und Zeit im Allgemeinen liegen bereits zahlreiche theoretische und empirische wissenschaftliche Arbeiten vor (siehe Kapitel 3.2.1-3.2.5). Diese bildeten einen wesentlichen Grundstein der Modell- und Prognoseerstellung.

Die Aufarbeitung und Nutzbarmachung wissenschaftlicher Arbeiten erfolgte dabei im TP WBA, das auf Basis von theoretischen Überlegungen und empirischen Ergebnissen aus der Forschungsliteratur wissenschaftlich-theoriegeleitete Hypothesen erstellte. Denn die Auswahl der aus den Hypothesen zu identifizierenden Datenquellen ergibt sich nahezu automatisch durch eine stringente Operationalisierung und zeigt, dass die wissenschaftliche Fundierung einer zielgerichteten Umsetzung in die Polizeipraxis nicht widersprechen muss, sondern dieser erst die notwendige Struktur verleiht (vgl. Pollich/Bode 2017: 2).

Im vorliegenden Bericht werden zunächst die relevanten theoretischen Grundlagen aus der Kriminalsoziologie und der Kriminologie komprimiert dargestellt, denn eine eigene Theorie zu Predictive Policing gibt es nicht (vgl. Pollich/Bode 2017:3). Hierbei wird Augenmerk auf diejenigen Ansätze gelegt, die in der Forschungsliteratur besonders prominent sind und die sich auch im vorliegenden Forschungsprojekt zur Operationalisierung eignen. Das bedeutet, es wurden konkrete projektespezifische Hypothesen aus den dargestellten Ansätzen abgeleitet. Außerdem wurden in diesem Rahmen die zentralen, in den Hypothesen genannten Konstrukte – spezifisch für jeden Ansatz – in konkrete, empirisch beobachtbare Indikatoren übersetzt (Operationalisierung). Im Anschluss wurden die zahlreichen aus den Hypothesen abgeleiteten Indikatoren zusammengefasst und systematisiert. So konnten letztendlich die potenziell erforderlichen Datenquellen identifiziert werden, die zur Messung der einzelnen Indikatoren notwendig waren und zu Zwecken der Hypothesenprüfung und Modell- und Prognoseerstellung beschafft werden mussten.⁶

Die hier dargestellten Theorien und Ansätze – und damit auch die daraus abgeleiteten Hypothesen – stellen lediglich eine Auswahl aus einer Vielzahl dar, die grundsätzlich geeignet ist, um sich der raum-zeitlichen Verteilung des WED theoretisch anzunähern. Die Auswahl wurde einerseits aufgrund der nachgewiesenen Relevanz und Verbreitung

der theoretischen Ansätze getroffen. Andererseits wurde bereits an dieser Stelle auf eine spätere Nutzbarkeit der Ansätze im Projekt Wert gelegt. Da im Projekt ausschließlich bereits vorliegende Daten verwendet werden sollten, die keinen konkreten Personenbezug aufweisen, sondern lediglich Informationen über räumliche Einheiten beinhalten, waren nicht alle Ansätze für das vorliegende Vorhaben gleichermaßen geeignet. Es erfolgte insoweit eine Konzentration auf Ansätze, deren zentrale Konstrukte ohne Hinzuziehung von Daten auf individueller Ebene empirisch erfasst werden konnten.

Ergänzend muss darauf hingewiesen werden, dass die erstellten Hypothesen oftmals nicht universell gültig sind. Beim WED war beispielsweise davon auszugehen, dass verschiedene Typen von Tätern aktiv sind, für die möglicherweise verschiedene „Rationalitäten“ gelten. Forschungsarbeiten und -überlegungen (vgl. beispielsweise Bernasco/Nieuwebeerta 2005: 308; Fox/Farrington 2012: 1593f.; Kersting/Willing 2014: 721), insbesondere aber auch die praktische polizeiliche Erfahrung zeigen, dass WED nicht von einer homogenen Tätergruppe begangen werden. Vielmehr erscheint eine Unterteilung sinnvoll. Denkbare Tätertypen sind beispielsweise professionelle Einbrecherbanden, Betäubungsmittelkonsumenten oder auch Täter, die das Opfer bereits im Vorfeld der Tat kannten (vgl. beispielsweise LKA NRW 2017). Da Art und Anzahl der Tätertypen jedoch nicht abschließend festgelegt werden können, werden im Weiteren keine bestimmten Typen als gegeben angenommen. Dennoch ist die Annahme unterschiedlicher Tätertypen für die Hypothesenbildung elementar: Die Bedeutung einiger einbezogener Variablen (z. B. Tageszeit) kann für verschiedene Tätergruppen unterschiedlich sein. Das kann im Extremfall dazu führen, dass Hypothesen für verschiedene Gruppen von Tätern genau entgegengesetzt formuliert werden müssen. Als Beispiel hierfür soll der sozioökonomische Status dienen: Möglicherweise wählen professionelle Täter ihre Ziele danach aus, wo sie den höchsten Profit erwarten (womit die Beuteerwartung eine prominente Stelle in der Rationalitätsüberlegung einnimmt). Damit werden sie überwiegend in wohlhabenderen Gebieten nach Tatobjekten suchen, die eine hohe Beute erwarten lassen, möglicherweise aber auch gut gesichert sind. Im Falle professionell agierender Täter ist davon auszugehen, dass diese mit dem entsprechenden Werkzeug und den technischen Fähigkeiten ausgestattet sind, diese Sicherungsmaßnahmen zu überwinden und so die Eintrittswahrscheinlichkeit zu erhöhen. Im Falle von WED zum Ziele der Finanzierung einer Drogensucht kann sich der sozioökonomische Status eines

⁶ Eine wissenschaftliche Auswahl der Daten bewahrt zudem davor, explorativ und mit großem Aufwand eine Vielzahl an Daten auf ihre Aussagekraft im Zusammenhang mit Kriminalitätsprognosen zu untersuchen (vgl. Pollich/Bode 2017: 10).

Gebietes womöglich anders auswirken. Diese gehen unter Umständen weniger geleitet von der Erwartung hoher Beute, sondern von einer hohen Wahrscheinlichkeit, Beute jedweder Art zu erzielen, vor. Damit ist davon auszugehen, dass diese Täter nach günstigen Tatgelegenheiten, wie offenen stehenden oder schlecht gesicherten Türen oder Fenstern, suchen. Was für einen Tätertypus im Augenblick der Tat „rational“ erscheint, könnte beispielsweise auch auf ähnliche Weise durch ein hohes Maß an Routine, einen hohen zeitlichen oder emotionalen Druck beeinflusst werden. Zahlreiche in den folgenden Hypothesen angenommene Ursachen für den WED können je nach Tätertypus somit in ihrer Art und Weise variieren.⁷

3.2.1 Rational-Choice-Theorien

Theorien rationalen Handelns werden etwa seit den 1980er Jahren verstärkt in den Sozialwissenschaften eingesetzt, um menschliches Handeln zu erklären. Im Wesentlichen geht deren Adaption aus den Wirtschaftswissenschaften auf James Coleman (1986, 1990) zurück, im deutschsprachigen Raum wurden derartige Ansätze unter anderem (u. a.) von Hartmut Esser (1999, 2007) fortgeschrieben.

Den Grundstein für die Anwendung sozialwissenschaftlicher Theorien rationalen Handelns auf delinquentes Handeln legten u. a. Cornish und Clarke (1986) mit ihrem Sammelband „The Reasoning Criminal“. In Deutschland wurden derartige Ansätze im Rahmen eines umfassenderen Erklärungsmodells, beispielsweise von Eifler (2002: 80ff.) oder von Lüdemann und Ohlemacher (2002: 51ff.), in die Kriminalsoziologie eingebracht. Lüdemann und Ohlemacher (2002: 54) fassen die zentralen Annahmen der Theorie folgendermaßen zusammen: „Je größer die subjektive Wahrscheinlichkeit ist, mit der eine Person angenehme Handlungsfolgen bei Ausführung einer bestimmten Handlungsalternative erwartet und je positiver sie diese Handlungsfolgen bewertet und je geringer die subjektive Wahrscheinlichkeit ist, mit der die Person unangenehme Handlungsfolgen bei Ausführung dieser Handlungsalternative erwartet, desto eher bzw. häufiger wird sie diese Handlungsalternative auch ausführen.“

Im Kern beschreiben die Theorien der rationalen Wahl die Maxime „Strebe nach Dingen, die möglich und zuträglich sind; und meide ein Handeln, das undurchführbar und/oder schädlich ist“ (Esser 1999: 257). Diese Rationalitätsunterstellung ist jedoch zunächst dadurch eingeschränkt, dass das, was als möglich, nützlich oder auch schädlich eingeschätzt wird, auch durch Straftäter stets nur aus einer subjektiven Sicht heraus geschehen kann (vgl. Esser 1999: 340f.; Diekmann/Voss 2004: 20). Außerdem stehen einem Individuum niemals alle nötigen Informationen zur Verfü-

gung, um in einem ökonomischen Sinne umfassend „rational“ zu entscheiden. Entscheidungsprozesse müssen zudem stets durch die Individuen vereinfacht werden, um sich nicht in ihrer Komplexität zu verlieren (vgl. Simon 1957: 198ff.; Esser 1997: 70f.). Bezogen auf den WED besteht die Einschätzung des Nutzens beispielsweise in einer gewissen Beuteerwartung. Der Täter schätzt ein, mit welcher subjektiven Wahrscheinlichkeit er die erwartete Beute tatsächlich erhalten kann, z. B. bewertet er seine technischen Fähigkeiten, in ein Haus einzudringen oder die Wahrscheinlichkeit, dass sich tatsächlich Beute in der erhofften Höhe in einem Objekt befindet. Unangenehme Handlungsfolgen sind beim WED für den Täter negative, üblicherweise strafrechtliche, Sanktionen. Die subjektive Wahrscheinlichkeit, dass diese eintreten, bewertet der Täter beispielsweise durch seine persönliche Einschätzung, bei einem Einbruch entdeckt zu werden. Die Aufrechnung von subjektiv erwarteten Kosten und Nutzen wird formal dadurch realisiert, dass der erwartete Nutzen einer Handlungsoption (z. B. eines Einbruchs oder des Absehens von einem Einbruch), also die Beuteerwartung, mit dessen Eintrittswahrscheinlichkeit multipliziert wird. Selbiges gilt für die subjektiv erwarteten Kosten, das heißt die Bestrafung. Erwartete Kosten werden dann vom erwarteten Nutzen abgezogen und so der so genannte „Nettonutzen“ (Lüdemann/Ohlemacher 2002: 54) ermittelt. Die Option mit dem höchsten „Nettonutzen“ wird dann umgesetzt. Es ist aber darauf hinzuweisen, dass diese Annahme modellhaft ist, das heißt dass nicht im engen Sinne davon ausgegangen wird, dass Personen diese Kosten-Nutzen-Abwägung kognitiv tatsächlich in Form von „Rechenschritten“ abarbeiten.

Ausgehend von dem berechtigten Einwand, menschliches Handeln, beispielsweise auch das von Kriminellen, sei bei weitem nicht immer „rational“ im engeren Sinne, wurden die Rational-Choice-Theorien weiterentwickelt. Es wird in der Folge von einer „variablen Rationalität“ (Kroneberg 2005: 344) gesprochen. Diese Erweiterung der Theorie kann auch gewinnbringend für die Erklärung kriminellen Handelns sein. Vereinfacht gesagt besteht die Erweiterung darin, dass die „Kalkulation“ der vielversprechendsten Handlungsalternative nicht bei jedem Menschen und in jeder Situation gleich abläuft, sondern dass das, was in die Ermittlung der günstigsten Handlungsalternative einfließt, situationsabhängig verschieden sein kann. Die Handlungsauswahl verläuft dann in mehreren Schritten. Zunächst wird ein sogenannter *frame* definiert, das heißt die (Ausgangs-)Situation wird durch den Handelnden, z. B. den Straftäter, bestimmt (vgl. Esser 2001: 259ff.; Kroneberg 2005: 350f.). Solche Ausgangssituationen können beim WED z. B. die gezielte Suche nach hochwertigen Objekten durch professionell organisierte Banden sein

⁷ Möglicherweise zeigen sich stärkere Zusammenhänge zwischen dem Tätertypus und der Tageszeit der Tatbegehung als zwischen dem Tätertypus und der räumlichen Verteilung der Taten. Dies bleibt empirisch zu prüfen.

oder die eher ungeplante Suche eines Betäubungsmittelkonsumenten nach Gelegenheitsbeute (Beschaffungskriminalität). Daneben sind zahlreiche andere *frames* denkbar.

Beginnend bei der situativen Rahmung kommt es beispielsweise darauf an, ob ein Täter hierfür ein „Handlungsrezept“ (ein so genanntes Skript) eingeübt hat. Hat ein Einbrecher eine bestimmte Objektauswahl bereits vielfach durchgeführt, wird er eher seinem eingeübten Handlungsschema folgen als ein Einbrecher, der Entscheidungen für oder gegen einen WED noch nicht oder erst selten durchgeführt hat (vgl. Esser 2001: 261ff.; Kroneberg 2005: 351f.). Je nachdem, ob ein Handlungsrezept vorliegt, das einfach und routiniert befolgt werden kann, oder nicht, folgt letztendlich die Entscheidung für eine Handlungsalternative mehr oder weniger „rational“. Wenn eine Situation in einer bestimmten Art und Weise neuartig oder komplex für den Handelnden ist, wird dieser seine Entscheidung eher nach der oben skizzierten rationalen „Kalkulation“ treffen (vgl. Esser 2001: 291ff.; Kroneberg 2005: 352f.). Auch andere Faktoren, wie eine emotionale Belastung, ein psychischer oder körperlicher Ausnahmezustand, eine besondere Wertorientierung und hoher Zeitdruck können die „Rationalität“ von (delinquenten) Personen beeinträchtigen und variabel werden lassen (vgl. beispielsweise Kroneberg 2007: 220ff.; Eifler 2009: 193f.; Pollich 2010: 91ff.).

Rational-Choice-Theorien beruhen demnach auf der Annahme, dass ein Täter eine Kosten-Nutzen-Abwägung durchführt, bevor er beispielsweise einen Einbruch begeht. Dessen Wahrscheinlichkeit der Durchführung steigt, wenn die Erwartung positiver Handlungsfolgen die der negativen überlagert (vgl. Lüdemann/Ohlemacher 2002: 54). Hieraus lassen sich z. B. folgende Hypothesen ableiten (Auswahl):

Je höher der **Wert des erwarteten Diebesguts** in einem Gebiet ist, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je geringer der **Einbruchaufwand** in einem Gebiet ist, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je niedriger die **Entdeckungswahrscheinlichkeit** in einem Gebiet ist, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

3.2.2 Routine-Activity- und Lifestyle-Theorie

Die so genannte Routine-Activity- und die Lifestyle-Theorie gehen von ähnlichen Annahmen aus (vgl. Garofalo 1987: 24ff.). Deshalb werden sie hier in einem Kapitel zusammengefasst. Was beide im Vergleich zu den Rational-Choice-Theorien auszeichnet, ist eine Erweiterung des Blickwinkels über die Tätermotivation hinaus (vgl. Garofalo 1987: 41; Cohen/Felson 1979: 590).

Die Routine-Activity-Theorie geht auf Cohen und Felson (1979) zurück und hatte ursprünglich das Ziel, den Kriminalitätsanstieg in den USA zwischen 1960 und 1975, insbesondere den immensen Anstieg der Einbruchkriminalität im Hellfeld um 200 Prozent, zu erklären. Die Autoren führten den besagten Anstieg in ihrer Theorie auf einen Wandel der alltäglichen Routinehandlungen (Routine Activities) der Menschen zurück. Die Struktur dieser Routineaktivitäten beeinflusst den Anstieg (vgl. Cohen/Felson 1979: 588f.; 593) zufolge die kriminellen Gelegenheiten. Routine Activities werden dabei verstanden als wiederkehrende bzw. regelmäßige und verbreitete Aktivitäten, die Menschen zum Zwecke ihrer Existenzsicherung ausführen (Arbeitsweg, Berufstätigkeit, Einkauf etc.).

Drei Elemente strukturieren dem Ansatz zufolge die Verteilungsmuster der Kriminalität (vgl. Cohen/Felson 1979: 589):

- **Motivierte Täter:**
Motivierte Täter, die auch (z. B. technisch oder körperlich) in der Lage sind, die kriminellen Handlungen umzusetzen (vgl. Cohen/Felson 1979: 590).
- **Geeignete Ziele:**
Die Eignung resultiert beispielsweise aus dem Wert, der Sichtbarkeit, der Zugänglichkeit oder der Transportierbarkeit (vgl. Cohen/Felson 1979: 591).
- **Abwesenheit fähiger Bewacher:**
Fähige Bewacher sind beispielsweise die Polizei, aber auch soziale Kontrollinstanzen wie beispielsweise Anwohner (vgl. Cohen/Felson 1979: 590).

Das Fehlen schon eines dieser Elemente verhindert den Autoren zufolge das Auftreten eines Kriminalitätseignisses (vgl. Cohen/Felson 1979: 589). Kriminelle Handlungen werden in diesem Ansatz als *events* gesehen, die in Raum und Zeit stattfinden (vgl. ebd.). Delinquente Neigungen von (potenziellen) Tätern werden als gegeben angenommen, ohne dass ihre Ursachen gesondert untersucht werden (dennoch ist deren Bedeutung den Autoren zufolge nicht zu unterschätzen; vgl. Cohen/Felson 1979: 605). Dagegen ist es das Erklärungsinteresse, wie die raum-zeitliche Organisation sozialer Aktivitäten Täter dabei unterstützt, ihre delinquenten Neigungen in Handlungen umzusetzen. Da illegale Handlungen auf Routinehandlungen im Allgemeinen basieren, spielt die räumliche und zeitliche Struktur legaler Routine Activities eine wichtige Rolle bei der Festlegung des Ortes, der Art und der Quantität illegaler Handlungen (vgl. Cohen/Felson 1979: 589).

Die Lifestyle-Theorie wurde von Hindelang, Gottfredson und Garofalo entwickelt, um Unterschiede zwischen verschiedenen sozialen Gruppen hinsichtlich ihres Viktimisierungsrisikos zu erklären (vgl. Miethe/Meier 1979: 36). Sie geht, ähnlich wie die Routine-Activity-Theorie (jedoch unter Verwendung anderer Begrifflichkeiten) davon aus, dass der Lebensstil einer Person diese stärker oder weniger stark risikanten bzw. kriminogenen Situationen aussetzt und somit

deren Viktimisierungsrisiko beeinflusst. Die wesentlichen Merkmale, wodurch sich Personen in ihrem Lebensstil unterscheiden, sind dabei demografische Unterschiede (vgl. genauer Hindelang et al. 1978: 241ff., zusammenfassend Miethe/Meier 1979: 32ff.). Damit sind die Lifestyles weitgehend deckungsgleich mit den Routine Activities nach Cohen und Felson (1979: 593f.) zu verstehen. Aus Gründen der Einfachheit wird daher im Folgenden – obwohl durchaus einige Unterschiede zwischen den Theorien bestehen – hauptsächlich die Routine-Activity-Theorie zur Hypothesenbildung herangezogen.

Die Routine-Activity- und die Lifestyle-Theorie eignen sich in erster Linie zur Erklärung von Opferwerdung. Darauf aufbauend generieren Miethe und Meier (1979) ein umfassenderes heuristisches Modell zur Integration von Theorien der Opferwerdung sowie der Täterschaft und des sozialen Kontexts, in dem Täter und Opfer zusammentreffen (vgl. Miethe/Meier 1979: 63ff.). Die Eckpunkte dieses Modells sind:

- **Sources of offender motivation:**
Die Motivation von Tätern resultiert, den Autoren zufolge, aus ökonomischer Benachteiligung, schwachen sozialen Bindungen, kriminalitätsbegünstigenden Einstellungen, psychologischen oder biologischen Merkmalen, allgemeinen Bedürfnissen sowie der Zugänglichkeit zu nichtkriminellen Handlungsalternativen (Miethe/Meier 1979: 64).
- **Victim Characteristics that provide opportunities:**
Hierunter verstehen Miethe und Meier (1979: 66) die Nähe zu Tätern, die Ausgesetztheit gegenüber kriminalitätsbelasteten Situationen (dies dürfte wohl eher für Körperverletzungsdelikte gelten), die Attraktivität als Ziel und die Abwesenheit fähiger Bewacher. Diese Merkmale helfen einerseits, einen kriminogenen sozialen Kontext zu definieren, andererseits wirken sie auch einzeln als Risikofaktoren für Kriminalität.
- **Social context:**
In diesem Zusammenhang sind der Ort (physischer Ort, Dunkelheit, Geschwindigkeit/Rhythmus), zwischenmenschliche Beziehungen und der soziale Handlungskontext (Zuhause, Schule, Arbeit, Freizeit) relevant.

Die Routine-Activity-Theorie basiert demzufolge u. a. auf den Annahmen der Rational-Choice-Theorien. Kriminalität entsteht nach diesem Ansatz vor allem im Zusammenhang mit alltäglichen, menschlichen Routinehandlungen. Die Struktur der Routine Activities beeinflusst hierbei die kriminellen Gelegenheiten. Routine Activities werden als wiederkehrende bzw. regelmäßige und verbreitete Aktivitäten, die Menschen zum Zwecke ihrer Existenzsicherung ausführen (Arbeitsweg, Berufstätigkeit, Einkauf etc.), verstanden (vgl. Cohen/Felson 1979: 588f.; 593). Hieraus lassen sich z. B. folgende Hypothesen ableiten (Auswahl):

Je höher die **Aktivität der Bewohner** in einem Gebiet ist, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je höher die **soziale Aktivität im Allgemeinen** in einem Gebiet ist, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Sofern Routine Activities die kriminellen Gegebenheiten entsprechend beeinflussen, strukturieren dem Ansatz zufolge drei Verteilungsmuster (Motivierte Täter, geeignete Ziele, Abwesenheit fähiger Bewacher) die Kriminalität. Hieraus lassen sich folgende Hypothesen ableiten (Auswahl):

Je **geeigneter die Ziele** in einem Gebiet sind, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je weniger **fähige Bewacher** in einem Gebiet ihre Routineaktivitäten ausführen, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je mehr **günstige Gelegenheiten** in einer Wohnumgebung (z. B. offen stehende Fenster) vorhanden sind, desto mehr Wohnungseinbrüche werden verübt.

Aus den Eckpunkten des Modells von Miethe und Meier (1979: 63ff.) können folgende Hypothesen abgeleitet werden (Auswahl):⁸

Je stärker die **ökonomische Benachteiligung** von Personen, desto höher ihre Motivation, Wohnungseinbrüche zu begehen.

⁸ Für den Zweck von Predictive Policing, zu dem hier keine gesonderten Daten erhoben wurden, eignen sich ausschließlich allgemein zugängliche Daten auf der Makroebene. Aus diesem Grund werden nicht zu sämtlichen Einflüssen auf die Tätermotivation (*sources of offender motivation*), die zuvor dargestellt wurden, Hypothesen formuliert. Beispielsweise können psychologische Merkmale nicht ohne Personenbezug und nicht ohne eine eigene Datenerhebung erfasst werden. Für die weiteren Eckpunkte (*victim characteristics that provide opportunities/social context*) werden an dieser Stelle ebenfalls keine gesonderten Hypothesen formuliert, da sie im Rahmen anderer Hypothesen benannt und operationalisiert werden.

Je geringer die **soziale Einbindung** von Personen, desto höher ihre Motivation, Wohnungseinbrüche zu begehen.

Je geringer die **Zugangsmöglichkeiten zu nichtkriminellen Handlungsalternativen** von Personen, desto höher ihre Motivation, Wohnungseinbrüche zu begehen.

3.2.3 Theory of Target Search/Crime-Pattern-Theory

Die Theory of Target Search geht auf Forschungen von Paul und Patricia Brantingham (2013) zurück, baut auf der Crime-Pattern-Theory auf und beschreibt einen theoretischen Ansatz zu Auswahlmechanismen von Tätern im Hinblick auf geeignete Angriffsobjekte. Der Fokus dieser Theorie liegt auf dem kriminellen Ereignis. Dabei werden für das Verständnis Aspekte der Situation (Ort und Zeit), der Gesetzgebung, des Täters und des Opfers betrachtet (vgl. Brantingham/Brantingham 2013: 535). Effektive Kriminalitätsvoraussagen erscheinen anhand dieser Perspektive wahrscheinlicher, denn Kriminalität ist ein Produkt sich in einem bestimmten Muster wiederholender menschlicher Aktivitäten (vgl. Brantingham 2011: 200). Das Konzept der Routineaktivitäten wird vom Ehepaar Brantingham somit angewendet, um Bewegungsbilder von Personen in Raum und Zeit nachzubilden (vgl. Pollich/Bode 2017: 6). Bei einer entsprechenden Kriminalitätsprognose sollte der Fokus folglich auf dem täglichen Leben und den Routine Activities liegen. Bevor es zu einem kriminellen Ereignis kommt, muss der potenzielle Täter jedoch erst auf ein geeignetes Angriffsobjekt bzw. Ziel treffen (vgl. Brantingham/Brantingham 2013: 536). Dieses Ziel muss einfach, sicher und profitabel sein (vgl. Brantingham/Brantingham 1995: 5). Der Prozess der Suche nach diesem Ziel wird als Target Search bezeichnet. Er bildet die theoretische Ausgangsbasis für die Crime-Pattern-Theory, die folgende Annahmen beinhaltet (vgl. Brantingham/Brantingham 2013: 536):

- Kriminalität ist komplex.
- Kriminalität passiert nicht zufällig.
- Täter und Opfer sind nicht ungewöhnlich in der Strukturierung von Raum und Zeit.
- Kriminelle Gelegenheiten und Tätigkeiten sind strukturiert.

Die Crime-Pattern-Theory geht davon aus, dass *target searchers* in einer bestimmten Art und Weise voraussagbar vorgehen: Täter, die beispielsweise eine Bahn benutzen, begehen ihre Taten i. d. R. auch in der Nähe der Bahnstationen (vgl. Brantingham/Brantingham 2013: 547f.). Ein kriminelles Ereignis findet der Theorie zufolge nur statt, wenn sich einem potenziellen Täter eine Kriminalitätsgelegenheit in Raum und Zeit bietet. Hierbei ist eine Vielzahl an Einflüssen maßgeblich (vgl. Brantingham/Brantingham 2013: 537ff.):

- Strukturierende und variierende Routine Activities (z. B. von Tag zu Tag)
- Struktur von Zeit und Entfernung (z. B. müssen Menschen zu bestimmten Zeiten an ihrer Arbeitsstelle sein)

- Struktur von menschlichen Bewegungsmustern
- Aktivitäts-/Vergnügungsmöglichkeiten
- Soziale Netzwerke
- Umgebungsstruktur (z. B. kulturelle Zusammensetzung eines Viertels)
- Weg des potenziellen Täters (z. B. auf dem Weg zur Arbeit)
- Wissen (Insider-/Outsider-Wissen) des potenziellen Täters
- Günstige Gelegenheit für den potenziellen Täter
- Urbane Gebiete (*crime generators*, *crime attractors* und *crime neutral areas*)

Den urbanen Gebieten kam im Projekt eine besondere Bedeutung zu, da es sich bei den Bezirken der Pilotbehörden im Wesentlichen um nordrhein-westfälische Großstädte handelt. *Crime generators* sind Gebiete, die viele Menschen kriminalitätsunabhängig anziehen, z. B. Einkaufszentren oder Bahnhöfe. Sie generieren Kriminalität durch ihr alleiniges Vorhandensein, da sie bei potenziellen Tätern leicht Begehrlichkeiten wecken können, Tatgelegenheiten eröffnen und Fluchtmöglichkeiten (An-/Abreise) bieten (vgl. Brantingham/Brantingham 2013: 537ff.). *Crime attractors* sind definiert als Gebiete, die potenzielle Täter besonders attraktiv finden (vgl. Brantingham/Brantingham 2013: 537ff.). Im Kontext des WED handelt es sich folglich insbesondere um Gebiete, die von den Tätern gezielt angefahren werden, z. B. wohlhabendere Viertel oder schlecht gesicherte Viertel. Die Übergänge zwischen *crime generators* und *crime attractors* sind zum Teil fließend. Es sind Indikatoren denkbar, die, mit leicht veränderter Konnotation, beidem gleichermaßen zuzuordnen wären (z. B. Haltestellen im öffentlichen Personennahverkehr [ÖPNV]). *Crime neutral areas* sind Gebiete, die weder viele Menschen anziehen noch besonders attraktiv für potenzielle Täter sind (vgl. Brantingham/Brantingham 1995: 7ff.).

Außerdem sind bauliche Merkmale des Raumes von Bedeutung für die Verteilung der Kriminalität (vgl. genauer Brantingham/Brantingham 1995: 10ff.). Hierbei handelt es sich um zentrale Plätze (*nodes*), Verkehrswege (*paths*), geografische Teilungen (*edges*) und Raumnutzungen (*land uses*).

Darüber hinaus unterscheiden die Autoren im Wesentlichen vier Formen von kriminellen Ereignissen (vgl. Brantingham/Brantingham 2013: 543ff.):

- Ereignisse basierend auf persönlichen Beziehungen, z. B. Tötungsdelikte
- Ereignisse basierend auf (günstigen) Gelegenheiten, z. B. Diebstahl im Einkaufszentrum
- Ereignisse basierend auf Aufklärung (Planung der Tat), z. B. gezielter Diebstahl aus einem gesicherten Objekt
- Ereignisse basierend auf Zweckmäßigkeit, das heißt Bereitschaft für eine entsprechende Tat

ist vorhanden, aber abhängig von den oben genannten Einflüssen

Die ersten drei Arten von kriminellen Ereignissen sind den Brantinghams (Brantingham/Brantingham 2013: 547f.) zufolge wenig geeignet, um Vorhersagen durchzuführen. Die vierte Art von kriminellen Ereignissen, basierend auf der Zweckmäßigkeit, eröffnet nach ihnen allerdings Potenzial für Vorhersagezwecke (vgl. ebd.).

Die Theory of Target Search beschreibt demnach einen theoretischen Ansatz über Mechanismen zur Auswahl geeigneter Angriffsobjekte (Brantingham/Brantingham 2013). Der Prozess der Suche des potenziellen Täters nach einem geeigneten Ziel wird als Target Search bezeichnet. Hinter dieser Annahme bildet die Crime-Pattern-Theory die theoretische Ausgangsbasis.⁹ Daraus lassen sich zunächst z. B. folgende Hypothesen ableiten (Auswahl):

Je **wohlhabender die Einwohner eines Gebietes** sind, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je **schlechter die Gebäude eines Gebietes gesichert** sind, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.¹⁰

Bei den Hypothesen zu den *crime neutral areas* wird davon ausgegangen, dass diese Gebiete komplementäre Merkmale zu den *crime generators* und *crime attractors* aufweisen. Deshalb wird auf die Formulierung von Hypothesen verzichtet. Im Hinblick auf Merkmale des urbanen Raumes ergeben sich beispielsweise folgende Hypothesen (Auswahl):

Je besser die **Verkehrsanbindung** eines Gebiets ist, desto mehr Wohnungseinbrüche werden verübt.

Je weniger **räumliche Teilungen**¹¹ in einem Gebiet vorhanden sind, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je mehr **Treffpunkte für disorderly people** in einem Gebiet vorhanden sind, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

3.2.4 Desorganisationstheorie

In der ersten Hälfte des 20. Jahrhunderts entstand in Chicago eine stark empirisch basierte Forschungstradition, die sich mit der Verteilung von Kriminalität in Städten befasste.

Bedeutende Vertreter dieser Tradition sind Shaw und McKay (1969[1942]), auf die im Wesentlichen die so genannte Theorie sozialer Desorganisation zurückgeht. Diese dient der Erforschung einer verstärkten Konzentration von Kriminalität in bestimmten Stadtgebieten und befasst sich demnach mit der räumlichen Verteilung von Kriminalität und ist in erster Linie makrosoziologisch ausgerichtet (vgl. Kubrin et al. 2009: 82, 89; Eifler 2002: 24).

Die Autoren stellten zunächst fest, dass die Kriminalitätsraten in der Stadt Chicago um das Zentrum herum am höchsten waren und zum Stadtrand hin wieder eine Abnahme zu beobachten war.¹² Allerdings erwies sich diese konzentrische Verteilung der Kriminalität als zeitlich nicht stabil (vgl. Shaw/McKay 1969[1942]: 50ff.). Die hoch belasteten Zonen zeichnen sich den Autoren zufolge dadurch aus, dass sie vom Verfall betroffen sind und dort eine niedrige soziale Kontrolle herrscht (vgl. Shaw/McKay 1969[1942]: 140ff.; Kubrin et al. 2009: 82ff.; Kunadt 2011: 24ff.).

In diesen sozial benachteiligten Gebieten herrscht den Autoren zufolge ein hohes Maß an sozialer Desorganisation, die als „Unfähigkeit lokaler Gemeinschaften, allgemein gültige Werte ihrer Bewohner umzusetzen und gemeinschaftliche Probleme zu lösen“ (Kunadt 2011: 28) definiert werden kann (vgl. Shaw/McKay 1969[1942]: 185f.). Empirische Studien der Autoren haben gezeigt, dass sich soziale Desorganisation verstärkt dann manifestiert, wenn folgende Faktoren gleichzeitig auftreten (vgl. Shaw/McKay 1969[1942]: 185ff.):

- **niedriger sozioökonomischer Status:**
Mit sinkendem sozioökonomischen Status ist von einer höheren Kriminalitätsrate auszugehen, weil konventionelle Ziele durch die ökonomischen Einschränkungen nicht auf konventionellen Wegen erreicht werden können, das Maß an sozialer Kontrolle niedriger ist und Gefühle einer relativen Deprivation vorherrschen. Weiterhin existiert ein geringes Maß an Bindung und Engagement gegenüber gesellschaftlichen Einrichtungen und Behörden und ein hohes Maß an Frustration durch das Leben in einer sozial und ökonomisch bedrückenden Umgebung (vgl. Shaw/McKay 1969[1942]: 185ff. und zusammenfassend beispielsweise Miethe/Meier 1979: 26f.; Kunadt 2011: 29). Allerdings kann auch ein höherer sozioökonomischer Status die Attraktivität für Kriminalität in einem Gebiet steigern, beispielsweise weil dieses lukrativer für Eigentumskriminalität erscheint. Wenn beide Ef-

⁹ An dieser Stelle wird nicht auf alle von Brantingham/Brantingham (2013: 537ff.) genannten Einflüsse eingegangen, sondern lediglich auf die bislang noch nicht näher beleuchteten Aspekte. Hierbei steht das raum-zeitliche Zusammentreffen eines potenziellen Täters mit einem geeigneten Angriffsobjekt im Mittelpunkt.

¹⁰ Es kann dabei von einem Interaktionseffekt zwischen dem Wohlstand und dem Ausmaß der Sicherung ausgegangen werden.

¹¹ Z. B. Flüsse oder Autobahnen.

¹² Die Annahme eines konzentrischen Aufbaus US-amerikanischer Metropolen geht auf Park et al. (1925: 50ff.) zurück.

fekte sich überlagern, kann der Effekt des sozio-ökonomischen Status statistisch möglicherweise verschwinden (vgl. Miethe/Meier 1979: 26).

- hohe ethnische Heterogenität:
Ethnische Heterogenität steht den Autoren zufolge für Werteunterschiede und damit für das Fehlen gemeinsamer Werte. In Gebieten, in denen gemeinsame Werte geteilt werden, kann Heterogenität dazu führen, dass diese effektiv umgesetzt werden. Heterogenität kann außerdem für objektiv messbare Gegebenheiten wie soziale Ungleichheit und wahrgenommene Ungerechtigkeit stehen, die wiederum die Motivation zu kriminellem Handeln fördern (vgl. Shaw/McKay 1969[1942]: 185ff.; Miethe/Meier 1979: 27; Kunadt 2011: 29f.).
- hohe Mobilität:
Die starke Fluktuation von Bevölkerung geht einher mit gesteigerten Möglichkeiten, sich kriminelle Normen anzueignen. Zudem sinkt die informelle soziale Kontrolle, es herrscht eine höhere Anonymität und die Anzahl der Tatgelegenheiten vergrößert sich (vgl. ebd.).

Dabei ist zu beachten, dass nicht diese drei Faktoren selbst Kriminalität verursachen, sondern dass sie auf soziale Bindungen und Kontrolle wirken, die wiederum das Kriminalitätsaufkommen beeinflussen (vgl. Kubrin et al. 2009: 89). In den betroffenen Gemeinschaften werden außerdem durch spezifische Lernprozesse delinquente Normen und Werte erlernt (vgl. Shaw/McKay 1969[1942]: 175f.; Eifler 2002: 25).

Sowohl diese Lernprozesse als auch neuere Erweiterungen der Desorganisationsansätze (beispielsweise Bursik/Grasmick 1995; Kornhauser 1978 zusammenfassend beispielsweise Eifler 2002: 26; Kunadt 2011: 34ff.) sind zwar für die Sozialwissenschaften als vielversprechend anzusehen, wurden jedoch im Rahmen des Projektes nicht weiter verfolgt, da sie überwiegend die Erhebung von selbstberichteten Individualdaten – i. d. R. erhoben durch Befragungen – erfordern. Solche Daten wurden im Projekt nicht verwendet. Zudem eignen sich diese Ansätze nur eingeschränkt zur Erklärung des Deliktes WED. Zwar gilt dies auch für die Desorganisationstheorie von Shaw und McKay, jedoch nimmt diese in der wissenschaftlichen Diskussion um die Verteilung von Kriminalität im Raum eine grundlegende Rolle ein und wurde deshalb auch im Projekt verstärkt beleuchtet.

Nach Shaw und McKay (1969[1942]: 185) geht folglich das erhöhte Aufkommen von Kriminalität in bestimmten Stadtvierteln auf das gleichzeitige Vorliegen eines niedrigen ökonomischen Status, ethnischer Heterogenität und einer erhöhten Mobilität zurück. Da dieser Ansatz sich zu einem Klassiker zur Erklärung der Verteilung von Kriminalität im Raum entwickelt hat (vgl. Kubrin et al. 2009: 82), wurde er an dieser Stelle auch auf den WED übertragen. Im Rahmen

einer allgemeinen Diskussion von Theorien zur Verteilung von Kriminalität im sozialen Raum formulieren Miethe und Meier (1979: 26f.) auch Hypothesen zur Prüfung der sozialen Desorganisationstheorie, an die die folgenden Hypothesen angelehnt sind (Auswahl):

Je niedriger der **sozioökonomische Status** in einem Gebiet ist, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je größer die **Mobilität** der Bevölkerung in einem Gebiet ist, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je größer die **ethnische Heterogenität** in einem Gebiet ist, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Jedoch werden diese Merkmale der Wohnumgebung durch soziale Bindungen und soziale Kontrolle vermittelt, bevor sie auf das Kriminalitätsaufkommen in einem Gebiet wirken. Da Hypothesen zur sozialen Einbindung und zur informellen sozialen Kontrolle im Rahmen anderer theoretischer Ansätze in ähnlicher Form formuliert werden, wird an dieser Stelle darauf verzichtet.

3.2.5 Broken-Windows-Theorie

Ein Experiment des Psychologen Philip Zimbardo (1973) kann als Basis der Broken-Windows-Theorie verstanden werden (vgl. Wilson/Kelling 1982: 31). Ziel des Experimentes war es, anhand zerstörter Auto-Wracks, Vandalisierungsverläufe in US-amerikanischen Innenstädten systematisch zu untersuchen. Dazu wurden jeweils ein Auto mit abgeschraubten Nummernschildern und geöffneter Motor- und Heckklappe (als notwendige „Auslösersignale“) abgestellt

- a) am Bronx Campus der New York University, wo innerhalb von zehn Minuten erste Gebrauchsteile abgebaut wurden. Nach 26 Stunden waren alle brauchbaren Teile abgebaut und nach weiteren neun Stunden begann die wahllose Zerstörung (vgl. Zimbardo 1973: 86f.)
- b) am Stanford University Campus, Palo Alto, CA wo die auslösenden Reize keine Wirkung zeigten. Das Auto stand über eine Woche unberührt. Daraufhin wurde das Auto zunächst durch die Versuchsleitung gezielt stark zerstört. Es folgte ein nächtlicher vandalistischer Übergriff auf das Autowrack durch Passanten (vgl. Zimbardo 1973: 88f.).

Die Bedingung für einen derartigen Vandalismus scheint nach Zimbardo ein Gefühl der Anonymität zu sein, das in New York eher vorherrscht. Zusammen mit minimalen Auslöser-Reizen reicht das Zimbardo zufolge für eine völlige Zerstörung des Autowracks aus. Ist diese gesellschaftliche Anonymität nicht gegeben (Palo Alto), sind extremere Reize und Modelle von Zerstörung und Aggression sowie ein noch

höheres Maß an Anonymität (Dunkelheit, Menschenmenge) vonnöten. Zimbardo schreibt der beobachteten Zerstörung von Autos eine gesellschaftliche Dimension zu (vgl. Zimbardo 1973: 89f.): „What is being destroyed here is not simply a car, but the basic fabric of social norms which must regulate all communal life“ (Zimbardo 1973: 90).

Auf dem Experiment von Zimbardo basiert die genannte Broken-Windows-Theorie von Wilson und Kelling (1982). Den Autoren zufolge sind Fußstreifen durch die Polizei empfehlenswert, da sie zwar nicht das Kriminalitätsaufkommen verringern, sich die Bewohner in den bestreift Gebiete jedoch sicherer fühlten, da ein Signal für die Wahrung der öffentlichen Ordnung gesetzt wurde. Eine Ursache für das Unsicherheitsgefühl der Bürger sehen die Autoren u. a. in einer empfundenen „Belästigung“ durch Obdachlose, Betrunkene, Prostituierte, Jugendliche (*disorderly people*) und Verwahrlosungserscheinungen im öffentlichen Raum (*disorder*; vgl. Wilson/Kelling 1982: 29f.).

Aufbauend auf ihren Überlegungen zur Entstehung des Unsicherheitsgefühls der Bevölkerung konstatieren die Autoren Zusammenhänge zwischen *disorder* und tatsächlicher Kriminalität. Zunächst kann das Ausmaß von „realer“ Kriminalität und *disorder* von Außenstehenden nicht adäquat unterschieden werden. *Disorder* und Kriminalität sind außerdem untrennbar durch eine Entwicklung verbunden. Wenn ein Fenster eines Gebäudes zerbrochen ist und nicht repariert wird, werden die übrigen Fenster bald auch beschädigt sein. Das nicht reparierte Fenster wirkt wie ein Signal, dass durch das Beschädigen weitere Fenster keine negativen Konsequenzen zu erwarten sind. Sogar gesetzestreue Bürger würden durch offensichtlich vernachlässigten Besitz zu Vandalismus verleitet (vgl. Wilson/Kelling 1982: 30f.). Die Vermutung von Wilson und Kelling (1982: 31) ist, dass derartige Phänomene am Ende zu einem Zusammenbruch der gemeinschaftlichen Kontrollmechanismen und der sozialen Kontrolle führen. Steigt die empfundene „Bedrohung“ durch *disorderly people* an, so werden die Bewohner annehmen, dass auch die Kriminalität ansteigt und sich dementsprechend verhalten. Sie gehen weniger auf die Straße und verhalten sich vorsichtiger. Ein betroffenes Gebiet ist sodann sehr anfällig für einen tatsächlichen Anstieg der Kriminalität.

Wilson und Kelling (1982: 5) führen weiter aus, dass die polizeiliche Arbeit sich in den 1960er Jahren von der Aufrechterhaltung der öffentlichen Ordnung (*order-maintenance*) auf die Kriminalitätsbekämpfung verlagert hätte. Dabei würde

den Autoren zufolge der Zusammenhang zwischen Ordnungserhalt und Kriminalprävention jedoch übersehen. Die Entkriminalisierung von lediglich „störendem“ Verhalten ist den Autoren zufolge ein Fehler. Einen Obdachlosen festzunehmen mag ihnen zufolge ungerecht erscheinen, aber das Versäumnis, etwas gegen viele störende Personen zu tun, könne eine Gemeinschaft zerstören. Dies entspricht der zuvor skizzierten Verbindung zwischen einem zerbrochenen und mehreren zerbrochenen Fenstern.¹³

Das Auto-Experiment von Zimbardo (1973) bezieht sich demzufolge weniger auf delinquentes Handeln im Allgemeinen, als vielmehr gezielt auf Vandalismus und damit *disorder*-Phänomene. Daher ist es schwierig, die Theorie auf den WED zu übertragen. Für den Bereich des WED werden nachfolgend beispielhafte Hypothesen formuliert, die eine direkte Verbindung zwischen den Ursachen von *disorder* und WED postulieren (Auswahl):

Je höher die **Anonymität in der Nachbarschaft** in einem Gebiet ist, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je mehr ***disorderly people*** in einem Gebiet anwesend sind, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je mehr ***disorder*-Phänomene** in einem Gebiet auftreten, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Da sich nach Wilson und Kelling (1982: 31f.) vor allem das Fehlen gemeinschaftlicher Kontrollmechanismen und sozialer Kontrolle auf Kriminalität auswirken, kann zudem festgehalten werden (Beispiel):

Je weniger **soziale Kontrolle** in einem Gebiet vorherrscht, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

3.2.6 Defensible-Space-Ansatz

Der Defensible-Space-Ansatz stammt vom Architekten und Stadtplaner Oscar Newman (1973; 1996) und hat seine Wurzeln in der sozialen Desorganisationstheorie. Den Schwerpunkt des Ansatzes bilden kriminalpräventive Überlegungen (vgl. Eifler 2002: 25), jedoch eignet er sich gleichwohl zur Erklärung von Kriminalität im Zusammenhang mit baulichen Merkmalen eines Wohngebiets. Ausgangspunkt

¹³ Auf den Ideen der Broken-Windows-Theorie bauen auch die viel diskutierten Zero-Tolerance-Ansätze der New Yorker Polizei auf (vgl. Lüdemann/Ohlemacher 2002: 142ff.; Schwind 2011: 331). Die Verbindung zwischen der Heranziehung dieses Ansatzes als einem von zahlreichen Erklärungsansätzen von Kriminalität im Raum und bestimmten polizeilichen Interventionen hatte aber keinerlei Implikationen für die im vorliegenden Projekt geplanten Interventionen. Die Ableitung einer Zero-Tolerance-Strategie aus den obigen Ausführungen war für die Pilotbehörden nie vorgesehen.

ist die Frage, warum in einer großen Hochhaussiedlung, die von Menschen mit niedrigem sozioökonomischen Status bewohnt wird, besonders an den zahlreichen *interior* und *exterior common public areas* Verfallserscheinungen und Vandalismus auftreten, während in einer vergleichbaren Siedlung mit Mittelschicht-Bewohnern bzw. solchen, in denen weniger dieser öffentlichen Orte existieren, derartige Phänomene nicht auftreten (vgl. Newman 1996: 10ff.).

Zentrales Element des Defensible-Space-Ansatzes ist die Kontrolle über Räume (Newman unterscheidet zwischen *private*, *semi-private*, *semi-public* und *public spaces* im Kontext von Wohngebäuden), genauer jene, die außerhalb von *dwelling units* (das Innere von Wohnungen, Häusern [privater Raum]) liegen. Newman untersucht speziell die Gruppierung dieser Wohneinheiten in verschiedenen Arten von Gebäuden, die gleichzeitig innen- und außenliegende *nonunit*-Räume (das heißt nicht-private Räume) mit verschiedenem Charakter produziert. Er klassifiziert (vgl. Newman 1996: 14ff.):

- a) *Single-family houses*: Einfamilienhäuser, die untergliedert in freistehende Häuser, semi-freistehende Häuser (Doppelhaushälften) und Reihenhäuser sind. Innerhalb der vier Wände ist immer der private Raum einer Familie. Innenliegenden Räume sind nicht öffentlich zugänglich.
- b) *Walkups*: Mehrfamilienhäuser ohne Fahrstuhl. Es gibt geteilte, nicht-private Räume innerhalb des Gebäudes. Auch die freien Flächen vor und hinter den Gebäuden sind meistens semi-öffentlich.
- c) *Highrises*: Hochhäuser mit Fahrstuhl, die dadurch gekennzeichnet sind, dass sich zahlreiche Familien die Innenbereiche teilen. Daher sind diese Bereiche als semi-öffentlich oder gar öffentlich zu beurteilen. Auch die zugehörigen Außenflächen sind als öffentlich zu bezeichnen.

Der „Anspruch“ einer Familie auf ein bestimmtes Territorium bzw. die Kontrolle hierüber reduziert sich proportional zum Anstieg der Anzahl an Familien, die sich dieses Territorium teilen: Je größer die Anzahl an Personen, die sich ein Territorium teilen, desto weniger Rechte an diesem können die Personen, Newman (1996: 17) zufolge, (subjektiv) beanspruchen. Je weniger Familien sich ein Territorium teilen, desto einfacher ist es hingegen, sich informell über eine adäquate Nutzung zu verständigen. Wenn die Anzahl der Nutzer steigt, kann es dazu kommen, dass lediglich die Nutzung als Durchgang noch möglich erscheint (Räume bleiben ungenutzt) und jegliche andere Nutzung zulässig ist (Müll abstellen, Vandalismus). Außenstehende können dann leichter, weil unbemerkter, eindringen.

Die drei Arten von Gebäuden unterscheiden sich nach Newman (1996: 18ff.) auch hinsichtlich der Kontrolle, die auf den angrenzenden Straßen ausgeübt werden kann, beispielsweise durch die Lage von Vorgärten, Hauseingängen,

Parkplätzen und Außenflächen. Hierbei ist insbesondere die „Anbindung“ der Straßen/Gehwege an möglichst privaten und bestimmten Wohneinheiten zugeordneten Außenflächen von Bedeutung. Diese erhöhen die Kontrolle bzw. den Einfluss der Bewohner und auch das Interesse an der dortigen Sicherheit.

Weiterhin hat sich in einer Studie von Newman und Franck (1980: 2, 206f.) u. a. gezeigt, dass, je mehr Personen sich einen Hauseingang sowie gemeinschaftlich genutzte Flächen teilen, es umso schwieriger ist, individuelle Ansprüche auf diese zu erheben. Denn dies erschwert zwischen Bewohnern und Fremden zu unterscheiden und sich mit anderen Bewohnern über Instandhaltung und Kontrolle dieser Bereiche zu verständigen. Eine steigende Gebäudegröße führt daher nicht nur zu mehr Kriminalität, sondern auch zu einer geringeren Nutzung (semi-)öffentlicher Räume, weniger sozialen Kontakten zwischen den Nachbarn, einem geringeren Gefühl der Kontrolle über innen- und außenliegende öffentliche Räume, mehr Kriminalitätsfurcht und einer weniger stabilen Gemeinschaft.

Außerdem bezieht Newman (1996: 23ff.) soziale Faktoren in seinen Ansatz ein, die mit der baulichen Umgebung interagieren. Als empirisch bedeutsame soziale Faktoren (die auch untereinander stark korrelieren) erweisen sich der Anteil der Bevölkerung, der wohlfahrtsstaatliche Leistungen bezieht, der Anteil der alleinerziehenden Eltern, die Leistungen beziehen und das Pro-Kopf-Einkommen.

Der Ansatz von Newman (1973, 1996) befasst sich mit dem Einfluss der architektonischen Gestaltung von Wohngebieten auf das Kriminalitätsaufkommen. Insbesondere die Menge an Personen, die sich Wohngebäude teilen und die damit verbundene Kontrolle über den Raum spielen bei dem Ansatz eine bedeutende Rolle. Aus dem Defensible-Space-Ansatz lassen sich konsequenterweise folgende Hypothesen ableiten (Auswahl):

Je mehr **Personen in den Wohngebäuden** eines Gebiets leben, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je **größer die Wohnkomplexe** in einem Gebiet sind, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Weiterhin bezieht Newman (1996: 23ff.) auch soziale Faktoren ein, die mit der baulichen Umgebung interagieren (Auswahl):

Je mehr **Empfänger wohlfahrtsstaatlicher Leistungen** in einem Gebiet leben, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je stärker das Ausmaß der **Anonymität der Wohnumgebung** in einem Gebiet ist, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je stärker das Ausmaß der **sozialen Isolation** der Bewohner eines Gebiets ist, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

3.2.7 Near-Repeat-Ansätze

Aufbauend auf dem sogenannten „Kirkholt Burglary Prevention Project“ aus dem Jahr 1988, das sich mit Präventionsmaßnahmen für WED befasste (vgl. Forrester et al. 1988), entwickelte sich das Konzept der Repeat Victimisation und seine verschiedenen Derivate, insbesondere der so genannte Near-Repeat-Ansatz.

Das erste Konstrukt, die Repeat Victimisation (im Folgenden RV), bezeichnet das Wiederauftreten von Kriminalität an denselben Orten und/oder gegen dieselben Personen (vgl. Pease 1998: 1). Im Gegensatz dazu bezeichnet das besser bekannte und allgemeinere Konzept des Near Repeat das Wiederauftreten von Kriminalität an den gleichen oder benachbarten Orten (vgl. Bernasco 2008: 414). Die Annahme des Near Repeat ist empirisch gut bewährt (vgl. beispielsweise Polvi et al. 1991; Farrell/Pease 1993; Johnson et al. 2007: 210). Die Erklärungsleistung der reinen RV- oder Near Repeat ist allerdings als eingeschränkt zu beurteilen. Der Ansatz besagt nicht „... what it is about the predicted hot spots that makes them hot...“ (Groff/La Vigne 2002: 37). Gründe für die wiederholte Auswahl derselben oder benachbarter Objekte oder Opfer bleiben bei diesen Ansätzen allenfalls implizit. Im Kontext der Near-Repeat-Ansätze werden lediglich rudimentär theoretisch basierte Versuche der Erklärung unternommen, weshalb eine erneute Viktimisierung i. d. R. zeitlich in engem Zusammenhang mit der ersten Tat auftritt. Dies kann darauf zurückzuführen sein, dass

- das Objekt für alle potenziellen Täter interessant ist (*flag risk*) oder
- die erste Tat die Wahrscheinlichkeit weiterer Taten verstärkt (*boot risk*).

Beim *flag risk* handelt es sich damit um einen beständigen Risikofaktor (z. B. eine für Täter „attraktive“ Wohngegend). Unter dem *boot risk* ist dagegen kein beständiger Risikofaktor zu verstehen. Dieses Risiko ist abhängig davon, welche Erfahrungen der Täter bei seiner ersten Tat im Objekt gemacht hat (vgl. Pease 1998: 8f.).

Inzwischen existiert eine Vielzahl empirischer Überprüfungen der Near-Repeat-Ansätze. Allen Ansätzen ist gemeinsam, dass Viktimisierung als der beste Prädiktor für zukünftige Viktimisierung zu sehen ist: Viele Täter nutzen auch im Kontext des WED vermutlich Erkenntnisse, die sie bei der ersten Tat erlangt haben (vgl. Pease 1998: 3).¹⁴ Der Täter hat beim Zurückkehren zum gleichen Objekt den Vorteil, dass er das Gebiet wie auch den Aufbau des Objektes bereits kennt (vgl. Bernasco 2008: 414). Konzeptionell aber weitgehend offen bleibt bei beiden Ansätzen, inwiefern tatsächlich davon ausgegangen werden kann, dass bei einem wiederholten Einbruch innerhalb kurzer Zeit der gleiche Täter hierfür verantwortlich ist. Bernasco (2008) geht, am Beispiel von Den Haag (NL), dieser Frage empirisch nach.¹⁵ Zusammenfassend wird durch die Studie bestätigt, dass tatsächlich häufig dieselben Täter bei RV und Near Repeat festzustellen sind (vgl. Bernasco 2008: 423ff.).

Weitere Studien, beispielsweise die von Bernasco und Nieuwbeerta (2005), fokussieren im Kontext der Diskussionen zu Repeat-Ansätzen auch auf die Motivation der Täter für die Auswahl bestimmter Objekte. Dabei gehen sie über die Annahmen der RV- und Near-Repeat-Ansätze im engen Sinne hinaus und untersuchen den Einfluss verschiedener Merkmale der räumlichen und sozialen Umgebung von Objekten, die die Funktionsweise des Near-Repeat-Phänomens erhellen könnten (vgl. Bernasco/Nieuwbeerta 2005: 298ff.). Auch Farrell et al. (1995: 395) haben zuvor bereits Erklärungsansätze diskutiert, die sich im Wesentlichen an die Rational-Choice- und Routine-Activity-Theorie anlehnen. RV-Raten sind ihnen zufolge ein Produkt aus Umgebungs- und Individualfaktoren.

Eine Studie von Bowers und Johnson (2005: 67) stellte zudem fest, dass direkt benachbarte Häuser von Einbruchobjekten ein erhöhtes Einbruchrisiko haben, genau wie Häuser, die auf der gleichen Straßenseite liegen und solche, die eine ähnliche Optik haben. Townsley et al. (2003) zeigen auf, dass Near Repeats vornehmlich in den Vororten mit einheitlicher Bauweise stattfindet. Johnson et al. (2007: 210) haben die Ubiquität des Near-Repeat-Phänomens anhand von zehn Gebieten in fünf Staaten näher untersucht. Sie kommen zu dem Schluss, dass das Near-Repeat-Phänomen tatsächlich ubiquitär ist. Europa und die USA unterscheiden sich aber in diesem Zusammenhang, da in Europa das Near-Repeat-Phänomen über längere Distanzen feststellbar ist, in den USA ist es dagegen stärker lokal ausgerichtet (vgl. Johnson et al. 2007: 212). Darüber hinaus ließ

¹⁴ Beispielsweise entfallen nach dem British Crime Survey 41 Prozent der Vermögensdelikte auf zwei Prozent der Personen, die Opfer von Eigentumskriminalität geworden sind (vgl. Pease 1998: 3). Im Bereich von gewerblichen Einbrüchen entfallen 59 Prozent der Taten auf drei Prozent der Geschäfte, im produzierenden Sektor entfallen 63 Prozent auf acht Prozent der Firmen (vgl. Pease 1998: 4).

¹⁵ Der Autor macht vor allem auf die Problematik aufmerksam, dass eine solche Messung schwierig ist, da beispielsweise manche Objekte anhand ihrer Merkmale viele unabhängig voneinander agierende Täter anlocken (vgl. Bernasco 2008: 413).

sich von ihnen feststellen, dass ein Near Repeat häufig zeitlich nah nach dem ersten Einbruch zu registrieren ist.¹⁶ Der zeitliche Aspekt, dass die Wahrscheinlichkeit von RV mit zunehmender Zeit geringer wird, wurde bereits von Polvi et al. (1991) festgestellt.

Trotz ihrer empirischen Bewährung muss einschränkend darauf hingewiesen werden, dass RV-Ansätze nicht für alle Arten von Kriminalität relevant sind, insbesondere nicht für solche, die mit dem Tod enden (z. B. Mord) oder solche, die kein unmittelbares Opfer haben (z. B. Drogendelikte) (vgl. Pease 1998: 21f.). Auch ist bislang konzeptionell noch unklar, ob ein Repeat zwingend dieselben Personen oder Orte voraussetzt, ob ein Repeat in eine gewerbliche Filiale, aber an einem anderen Ort, als Repeat gewertet wird, ab wann ein Versuch schon als Repeat bewertet wird, innerhalb welchen Zeitraumes ein Repeat noch als solcher gilt, über welche Straftatenpalette ein Repeat als solcher zu sehen ist und weitere Aspekte (vgl. Pease 1998: 28ff.). Diese offenen Fragen sind sicherlich auch das Produkt der nur ansatzweisen theoretischen Durchdringung des Near-Repeat-Ansatzes.

Vorausgegangene Kriminalität bleibt Pease (1998: 5) zufolge jedoch der beste Prädiktor für die Polizei (und ihre Datenbestände), wenn andere Datenquellen nicht vorhanden sind oder benutzt werden können. Beim Fehlen anderer Datenquellen kann es demnach für die Polizei als effizienteste Lösung angesehen werden, sich auf die Objekte zu konzentrieren, in die bereits eingebrochen wurde (vgl. Pease 1998: 16). Hierdurch begründet sich auch der große Vorteil dieser Ansätze, und zwar der des relativ überschaubaren Datenbedarfes und der damit einhergehenden Datenbeschaffung (vgl. Pollich/Bode 2017: 4).

Die Repeat-Ansätze beruhen demzufolge auf der Grundannahme, dass sich Straftaten, im Wesentlichen Einbruchdelikte, in Raum und Zeit häufen und es oft zu Mehrfachviktimisierungen kommt. Diese Mehrfachviktimisierungen können sowohl ein bestimmtes Opfer wie auch ein bestimmtes Gebiet betreffen. Handelt es sich um das gleiche Opfer oder um denselben Ort, wird dies als RV bezeichnet (vgl. Pease 1998: 1). Handelt es sich dagegen um benachbarte Orte/unmittelbar in der Nachbarschaft gelegene Objekte, wird von Near Repeat gesprochen (vgl. Bernasco 2008: 414). Beim Near-Repeat-Ansatz handelt es sich nicht um eine klassische Theorie, sondern um ein Bündel von Ansätzen, die auf empirischen Beobachtungen basieren.¹⁷ Deshalb ist es an

dieser Stelle nicht möglich, Hypothesen direkt theoretisch abzuleiten. Die erstellten Hypothesen basieren auf empirischen Erkenntnissen verschiedenster Arbeiten aus der Forschungsliteratur, die oftmals nicht im klassischen Sinne hypothestetend durchgeführt wurden, sondern nach Regelmäßigkeiten zur gezielten Ableitung kriminalpräventiver Maßnahmen gesucht haben. Die hier erstellten Hypothesen umfassen deshalb keine Annahmen über kausale Zusammenhänge, sondern bewegen sich stark auf einer deskriptiven Ebene (Auswahl):

Wenn in einem Objekt ein Einbruch stattgefunden hat, dann finden in diesem Objekt oder **in räumlicher Nähe** weitere Einbrüche statt.

Spezifizierung: Je **näher** andere Objekte an dem betroffenen Objekt liegen, desto höher ist die Wahrscheinlichkeit, dass dort ebenfalls eingebrochen wird.

Wenn in einem Objekt ein Einbruch stattgefunden hat, dann finden in diesem oder benachbarten Objekt(en) **in zeitlicher Nähe** weitere Einbrüche statt.

Spezifizierung: Je **kürzer** der Einbruch in das betroffene Objekt zurückliegt, desto höher ist die Wahrscheinlichkeit, dass in dieses oder benachbarte Objekte ebenfalls eingebrochen wird.

Aus den beschriebenen empirischen Befunden lassen sich folgende Hypothesen ableiten, in denen die Annahme räumlicher und zeitlicher Nähe jeweils in einer Hypothese zusammengefasst wird (Auswahl):¹⁸

Wenn in einem Objekt ein Einbruch stattgefunden hat, dann finden in Objekten auf der **gleichen Straßenseite** in räumlicher und zeitlicher Nähe weitere Einbrüche statt (vgl. Bowers/Johnson 2005: 67).

Wenn in einem Objekt ein Einbruch stattgefunden hat, dann finden in Objekten **gleicher Bauart** in räumlicher und zeitlicher Nähe weitere Einbrüche statt (vgl. Bowers/Johnson 2005: 67; Bernasco/Nieuwebeerta 2005: 298ff.).

Über den Near Repeat im engeren Sinne hinaus lassen sich auch Annahmen ableiten, die sich mit den verschiedenen Formen der (Allgemein)Kriminalität beschäftigen (Auswahl):

¹⁶ Dies ist vermutlich darauf zurückzuführen, dass der Täter noch in Erinnerung hat, wie ein Objekt oder eine Wohngegend aussieht (vgl. Johnson et al. 2007: 215). Weiterhin beeinflusst der ÖPNV der verschiedenen Städte die feststellbare Reichweite beim Near-Repeat-Phänomen (vgl. Johnson et al. 2007: 216).

¹⁷ Pollich und Bode (2017: 4) beschreiben den Near-Repeat-Ansatz als eine Art von Heuristik und weisen darauf hin, dass es keine sozialwissenschaftliche Theorie im erkenntnistheoretischen Sinne ist, da wenig theoretische Erklärungsversuche unternommen werden.

¹⁸ Schweer, ein deutscher Vertreter des Near-Repeat-Ansatzes, geht zudem davon aus, dass so genannte „Triggerkriterien“ „anhand ihrer Tatmerkmale eine überdurchschnittliche Wahrscheinlichkeit aufweisen, dass sie in Near Repeats auftreten“ (beide Schweer 2015: 14) und damit eine prognostische Funktion erfüllen. Um welche Kriterien es sich hierbei allerdings für den WED handelt, geht aus den Ausführungen nicht ausreichend hervor. Aufgrund dieser fehlenden empirischen und theoretischen Fundierung wird hier auf die Erstellung von Hypothesen zu „Triggerdelikten“ bzw. Deliktmerkmalen verzichtet.

Je mehr **Eigentumskriminalität** im Allgemeinen in einem Gebiet registriert wird, desto mehr Wohnungseinbrüche werden in räumlicher und zeitlicher Nähe in diesem Gebiet verübt (angelehnt an Pease 1998: 3f.).

Je mehr **Allgemeinkriminalität** in einem Gebiet registriert wird, desto mehr Wohnungseinbrüche werden in räumlicher und zeitlicher Nähe in diesem Gebiet verübt.

3.2.8 Professionelle Akteure/Expertenannahmen

Um der polizeilichen Praxis und dem daraus resultierenden Erfahrungswissen ausreichend Raum zu geben, wurden über die wissenschaftlich abgeleiteten Hypothesen hinaus auch Einschätzungen professioneller polizeilicher Akteure¹⁹ integriert. Solche Einschätzungen sind wichtig, um eine Rückkoppelung der wissenschaftlich hergeleiteten Hypothesen mit der polizeilichen Praxis abzugleichen und bei Bedarf entsprechend anzupassen (vgl. Pollich/Bode 2017: 7). Hierbei wurden nur solche Annahmen berücksichtigt, die möglicherweise geeignet sind, das Risiko von WED an bestimmten Orten und zu bestimmten Zeiten vorherzusagen.²⁰ Während die weiter oben formulierten Hypothesen häufig bereits Gegenstand empirischer Forschungsarbeiten waren, sind die im Folgenden dargestellten Annahmen nicht aus theoretischen Ansätzen abgeleitet und bislang zum Teil empirisch ungeprüft.

Obwohl die Licht- und Witterungsverhältnisse bereits im Zuge der Operationalisierung mehrerer Konstrukte (z. B. Entdeckungswahrscheinlichkeit) thematisch angerissen wurden, wurde bislang ihre eigenständige Bedeutung als Vorhersagevariable für WED nicht ausreichend deutlich. Da sie insbesondere für die Ausrichtung polizeilicher Handlungen von hoher Relevanz sind, werden an dieser Stelle nochmals gesondert Hypothesen zur Wirkung von tageszeitbedingten Licht- und Witterungsverhältnissen formuliert (Auswahl):²¹

Je früher die **Dämmerung eintritt (dunkle Jahreszeit)**, desto mehr Wohnungseinbrüche werden verübt (vgl. hierzu auch Landeskriminalamt NRW 2012: 1).

Wenn der **Mond** die nächtliche **Sicht** in einem Gebiet begünstigt, dann werden mehr Wohnungseinbrüche in diesem Gebiet verübt.

Wenn das **Wetter** die **Fluchtmöglichkeiten** in einem Gebiet einschränkt (Schnee, Eis), dann werden weniger Wohnungseinbrüche in diesem Gebiet verübt.

Wenn das **Wetter** die **Hinterlassung von Spuren** in einem Gebiet fördert (Schnee, Matsch), dann werden weniger Wohnungseinbrüche in diesem Gebiet verübt.

Wenn das **Wetter** die **Sicht** in einem Gebiet **einschränkt** (Nebel, Regen), dann werden mehr Wohnungseinbrüche in diesem Gebiet verübt.

Neben äußeren Einflüssen könnte die Bedeutung polizeiinterner Aspekte ebenfalls eine Auswirkung auf das Aufkommen des WED haben. In diesem Zusammenhang ließen sich folgende Hypothesen spezifizieren (Auswahl):

Je größer die aktuelle **polizeiliche Auslastung** in einem Gebiet durch Einsätze ist, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je geringer die **raumbezogene polizeiliche Dichte** (Wegstrecken) in einem Gebiet ist, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.²²

Je mehr **verdächtige Wahrnehmungen** in einem Gebiet aufkommen, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je mehr **Beuteabsatzmöglichkeiten** in einem Gebiet vorhanden sind, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Aber auch im Zusammenhang mit dem Vorhandensein von bestimmten Gebäudetypen lassen sich folgende, aus dem Erfahrungswissen der Polizei abgeleitete, Hypothesen bilden (Auswahl):

Je näher **Objekte an einem Bahnhof** liegen, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Je mehr **gemeinsame Wohn- und Geschäftsräume** in einem Gebiet vorhanden sind, desto mehr Wohnungseinbrüche werden in diesem Gebiet verübt.

Abschließend ist anzumerken, dass sich die zuvor dokumentierte Auswahl an Hypothesen auf die Lesart des jeweiligen theoretischen Ansatzes bezieht. Dieser Umstand bedingt, dass in Teilen auch gegensätzliche Hypothesen (z. B. im Zusammenhang mit dem Einfluss des sozioökonomischen Status) möglich sind. Die Prüfung der für das Projekt SKALA relevanten Bedeutsamkeit einzelner Vorhersagevariablen war Gegenstand der weiteren Untersuchung (siehe Kapitel 4.5).

¹⁹ Zur Begriffsbestimmung vgl. Bode 2013: 134.

²⁰ Ermittlungsunterstützende Annahmen aus der Praxis (beispielsweise zum Zusammenhang zwischen Tätergruppierung und Modus Operandi) werden hier nicht berücksichtigt, da diese nicht ausreichend zur Vorhersage von Kriminalitätsbrennpunkten geeignet sind.

²¹ Die tatsächlichen tageszeitbedingten Licht- und Witterungsverhältnisse müssen miteinbezogen werden, beispielsweise, ob die Sicht an einem Tag tatsächlich wetterbedingt begünstigt war oder aber entgegen der Wetterprognose Sicht behindernd (etwa durch Bewölkung).

²² Die raumbezogene polizeiliche Dichte muss die Einwohnerdichte in einem Gebiet mitberücksichtigen.

3.3 Operationalisierung und Indikatorenbildung

Entsprechend der für die Hypothesen relevanten Indikatoren wurden die erforderlichen Daten, die nicht durch die polizeilichen Vorgangsdaten abgedeckt werden konnten, kommerziell beschafft. Auf eine Auflistung aller im Projektverlauf getesteten polizeilichen Vorgangsdaten (Polizeidaten) und zusätzlich erworbenen Daten wird hier aus Gründen der Komplexität verzichtet. Die zusätzlich beschafften Daten werden im Folgenden unter dem Begriff „soziostrukturelle Daten“

zusammengefasst, beinhalten aber z. B. auch soziodemografische oder gebäudespezifische Daten. Darüber hinaus wird angemerkt, dass bei weitem nicht alle durch die Hypothesen identifizierten und potenziell bedeutsamen Daten erworben werden konnten. Dies ist im Wesentlichen darauf zurückzuführen, dass viele relevante Indikatoren sich nicht in entsprechenden Datenquellen widerspiegeln lassen bzw. diese beschafft werden konnten. Abbildung 4 zeigt jedoch, um welche Art von Daten es sich gehandelt hat.

Abbildung 4: Datenbasis



Neben den Polizeidaten, bei denen es sich um Vorgangsdaten wie Tatort, Tatzeit, Beute oder Modus Operandi handelte, flossen bei der Modell- und Prognoseerstellung folgende weitere Daten ein:

- **Einwohnerstruktur**
(z. B. Haushalte nach Altersgruppen, Haushalte mit Kindern, DINKS²³ oder Bildungsschichten)
- **Gebäudestruktur**
(z. B. Baujahr, Bauart [Einfamilienhaus, Reihenhaushaus, Mehrparteienhaus, ...], Bauweise [marode bis exklusiv], Wohnlage [schlecht bis gut] oder Gartengröße)
- **Verkehrs-anbindung**
(z. B. Entfernungen zu Autobahnen, Bundesstraßen, Haltestellen oder Bahnhöfen)
- **Kaufkraft**
(z. B. Haushalte nach Einkommen)
- **Mobilität**
(z. B. Zulassung von PKW-Typen oder Krädern)

Im Zusammenhang mit der Gewährleistung und Einhaltung des Datenschutzes wird auf Kapitel 3.5 verwiesen. Das häufig im Zusammenhang mit Predictive Policing diskutierte Near-Repeat-Phänomen floss ebenfalls als „Einflussgröße“

in die Modell- und Prognoseerstellung ein. Für eine detaillierte Beschreibung wird in diesem Kontext auf Kapitel 4.6 verwiesen.

3.4 IT-Hard- und Software

Die Prüfung der Methode von Predictive Policing wurde technisch im LKA NRW mit SKALA realisiert. Aufbauend auf dem zuvor dargestellten hypothesengestützten Vorgehen war demnach der Einsatz einer Software erforderlich, die dem Anspruch und der Philosophie des Projektes Rechnung trug und gleichzeitig eine entsprechende Offenheit und Flexibilität gewährleisten konnte. Konkret lautete der Ausschreibungstext für die Leistungsbeschreibung:

„Das zu beschaffende Analyse- und Prognosesystem muss von mindestens drei Anwendern an zentraler Stelle hinsichtlich der Modellierung und Erzeugung von Analysen und Prognosen vollwertig genutzt werden können. Alle Anwender besitzen die Möglichkeit, auf bereits erstellte Modelle und denselben zu analysierenden Datenbestand zuzugreifen, sowie unabhängig voneinander Berechnungen zu initiieren. Das zu beschaffende System ist in der Lage, die Ergebnisse je nach Sinngehalt numerisch, in Form von Tabel-

²³ Doppelverdiener ohne Kinder (deutsch)/Double Income No Kids (englisch).

len, Grafiken und mittels eines vom Auftraggeber beigestellten Geoinformationssystems (wünschenswert ArcGIS) kartographisch darzustellen. (...)

Es beinhaltet in der gelieferten Version bereits alle gängigen, aber auch aktuelle Algorithmen und ermöglicht die Einbindung und Nutzung externer Algorithmen, ohne dass dafür ein bieterseitiger Aufwand nötig ist. Ebenso soll das System die Verwendung datenbankinterner Algorithmen zulassen bzw. unterstützen. Das System soll die Open-Source-Integration mit R (R Foundation) ermöglichen. Ergebnisse können so exportiert werden, dass sie auch von Produkten anderer Hersteller verwendet werden können.

Das zu beschaffende System muss es ermöglichen, Daten zu importieren und in einem eigenen Speicher zu verwalten. Es soll flexible Möglichkeiten der Datenhaltung bieten, die einfache Integration neuer Daten ermöglichen und unterstützen sowie ungewollte Datenmanipulationen verhindern. Das System muss den Datenimport aus relationalen Datenbanken ermöglichen. Zusätzlich soll der Import webbasierter Quellen möglich sein und eine SSL-Verbindungsverschlüsselung unterstützen. Das Erfassen, Transformieren, Bereinigen und Validieren von Daten wird vom System durch leistungsfähige, automatisierte Datenaufbereitungsfunktionen unterstützt, aber auch eine systeminterne, manuelle Datenaufbereitung ist möglich.

Das zu beschaffende System muss in der Lage sein, Analysen durchzuführen und Prognosen zu erzeugen, sowie eine Verarbeitung und Analyse von Ortsdaten unterstützen. Dafür muss es neben den etablierten Algorithmen zur Ausreißerererkennung, Klassifikation, Clusteranalyse, Assoziationsanalyse und Regressionsanalyse auch neue, aktuelle Algorithmen anbieten. Das System soll eigenständig in der Lage sein, den Algorithmus mit der präzisesten Prognose zu ermitteln und dem Anwender anzubieten. Und es soll Ergebnisse von uni-, bi- und multivariaten Analysen in Tabellenform und in Form von Grafiken darstellen können.

Das System ermöglicht neben der Analyse strukturierter Daten auch die Analyse unstrukturierter Daten. Es ermöglicht die Darstellung erkannter Muster in Zusammenhang mit nicht in die Analyse einbezogener strukturierter Daten. Aufgrund denkbarer Veränderungen im Täterverhalten während des Analysezeitraums soll das System Möglichkeiten des Überwachten, des unüberwachten und des bestärkenden Lernens beinhalten.“

Als Ergebnis eines öffentlichen Vergabeverfahrens wurde die Data- und Text-Mining-Software SPSS Modeler von IBM beschafft. Mit dieser komplexen Software ist es möglich, verschiedene Datenquellen miteinander in Beziehung zu setzen, zu analysieren und darauf aufbauend Prognosemodelle zu erstellen. Zentral ist hierbei die raum- und zeitbezogene Zusammenführung und Analyse.

Der technische Support wurde durch die System Vertrieb Alexander (SVA) GmbH geleistet. Die SVA GmbH führte in ihren Räumlichkeiten eine mehrtätige Schulung aller Mitarbeiter, die mit dem IBM SPSS Modeler arbeiteten, durch.

Das TP IT stellte in einem separaten Auswerteraum die erforderliche Hard- und Software bereit. Der IBM SPSS Modeler wurde in einem geschlossenen Netzwerk (Server im Serverraum des LKA NRW, Client-PCs im Auswerteraum) betrieben. Das Netzwerk hatte keine Anbindung an das Internet wie auch keine Anbindung an das Polizeinetz (CN-Pol). Sämtliche Datentransfers wurden im Projektzeitraum über das Netzwerklaufwerk und über USB-Datenträger (mittels „Datenschleuse“ zwecks Virenprüfung) realisiert. Ein Sicherheitskonzept und Verfahrensverzeichnis wurden in Abstimmung mit dem Datenschutz- und IT-Sicherheitsbeauftragten erstellt.

Das LZPD NRW stellte die, mit dem Datenschutz abgestimmten, polizeilichen Daten (Vorgangsdaten) auf einem Netzlaufwerk im CN-Pol bereit. Für die Pilotbehörden wurden, zur Übermittlung der Analyseergebnisse (Kriminalitätsprognosen), Netzwerklaufwerke mit den erforderlichen Zugriffsberechtigungen eingerichtet. Über einen Discoverer-Zugang wurde seitens des LZPD NRW zudem gewährleistet, dass eigenständige Erhebungen von Vorgangsdaten durch das TP WBA vorgenommen werden konnten.

Anfängliche Performanceprobleme, die sich durch längere Berechnungszeiten äußerten, wurden mit der Installation einer SQL-Pushback-Datenbank und der Verlagerung von Rechenoperationen in die Datenbank behoben. Das Erstellen von Prognosemodellen lief nach der Umstellung deutlich schneller. Mit der Erweiterung des Projektes auf die Polizeibezirke Bonn, Düsseldorf, Essen und Gelsenkirchen erfolgte darüber hinaus ein Ausbau von Hardwarekomponenten zur Performancesteigerung des Systems. Dies betraf sowohl die Server- wie auch die Client-Hard- und -Software.

Zur Verbesserung der kollaborativen Arbeit bei der Modell- und Prognoseerstellung wurde im Projektverlauf ein Microsoft Surface Hub beschafft und eingesetzt.

Für die Georeferenzierung der Daten, geostatistische Analysen und die kartografische Aufbereitung der Kriminalitätsprognosen wurden zu Beginn des Projektes mehrere Lizenzen für das Geoinformationssystem ArcGIS von ESRI beschafft. Ebenfalls einhergehend mit der Erweiterung des Projektes wurden diese Lizenzen ergänzt, u. a. auf eine Spatial-Analyst-Lizenz zur Erstellung von Heatmaps.

3.5 Datenschutz

Das Projekt SKALA wurde schon im Vorfeld des Projektstarts datenschutzrechtlich eng begleitet. Der Datenschutzbeauftragte (DSB) des LKA NRW war organisatorisch direkt an die Projektleitung und Geschäftsführung angebunden.

Die enge und kontinuierliche Einbindung des DSB in das Projekt gewährleistete, dass datenschutzrechtliche Fragen zur Verwendung spezifischer Daten stets zeitnah geklärt werden konnten.

Das verwendete Datenmaterial bestand aus polizeilichen Daten, soziostrukturellen Daten sowie öffentlich zugänglichen Kalender- und meteorologischen Daten. Bei den polizeilichen Daten handelte es sich ausschließlich um bereits erhobene Daten aus dem Vorgangsbearbeitungssystem der Polizei NRW, dem Integrationsverfahren Polizei (IGVP). Dies beinhaltete Fallinformationen wie Tatzeit, Modus Operandi oder Beutehöhe. Die Daten waren zwingend zur Generierung der Kriminalitätsprognosen erforderlich. Hierunter fiel auch der (georeferenzierte) Tatort eines Deliktes. Bestimmte personenbezogene Daten von Tätern, Opfern oder anderen Personen wurden nicht genutzt. Die Einzelangaben sind zwar im Einzelfall (Beispiel: Tatort war ein Einfamilienhaus mit ein bis zwei Hausbewohnern) grundsätzlich geeignet, eine Person bestimmbar zu machen. Es ist jedoch anzumerken, dass die Datenfelder als reine Fallinformationen im Zusammenspiel mit anderen Parametern genutzt wurden, um auf Basis kriminologischer Erkenntnisse computerunterstützt und in Verbindung mit soziostrukturellen Daten Berechnungsgrundlagen für Prognosen räumlicher und zeitlicher Kriminalitätsrisiken zu erhalten. Es war nicht Ziel des Projektes, aufgrund einer möglichen Einzelfallkonstellation von Fallinformationen eine theoretisch mögliche Bestimmbarkeit einer Person anzustreben, da das für die Analyse sowie die Prognosemodellierung und -darstellung unerheblich war. Zusammenfassend zielen die Prognosen somit ausschließlich auf die räumliche und zeitliche Entwicklung von Kriminalitätsrisiken ab und niemals auf Täter oder Opfer. Deshalb wurden auch keine Täter- oder Opferdaten in die Modell- und Prognoseerstellung einbezogen. Es fand ausschließlich eine Verwendung von (Hintergrund-)Informationen zu Delikten in Raum (auf einem bestimmten Aggregationsniveau wie auf beispielsweise auf Straßenabschnittebene) und Zeit statt. Der grundsätzliche Sinn des Data Mining besteht darin, in großen Datenmengen Muster zu erkennen und daraus Regeln abzuleiten, also gerade nicht in der Ermittlung eines einzelnen bestimmten Datenfeldes. Darüber hinaus war eine theoretische Bestimmbarkeit konkreter Personen nicht ohne die Zuhilfenahme zusätzlicher polizeilicher Verfahren möglich und wurde auch vom Verfahren SKALA bewusst verhindert. Entsprechende Daten sind im Übrigen für die Modell- und Prognoseerstellung auch irrelevant.

Die zur Prüfung der Hypothesen erforderlichen soziostrukturellen Daten wurden im Wege eines öffentlichen Vergabeverfahrens von der NEXIGA GmbH erworben. Grundlegende Anforderungen an den Anbieter waren, dass die Erhebung sämtlicher Informationen, die in den angebotenen

Daten enthalten sind, unter Beachtung des Bundesdatenschutzgesetzes erfolgte und die Daten vollständig anonymisiert zur Verfügung gestellt werden.

Sämtliche Daten, die gegebenenfalls Informationen zu Personen beinhalten konnten, wurden – je nach Verfügbarkeit – auf der Ebene von Wohnquartieren, Straßenabschnitten oder höher aggregiert. Der Begriff Wohnquartier definiert sich gemäß der NEXIGA GmbH hierbei als Untereinheit aus den ursprünglich stammenden Stimmbezirken bei Wahlen und somit einer Gebietseinheit von durchschnittlich 400 Haushalten (vgl. Nexiga 2017). Lediglich die Informationen zu baulichen Merkmalen von Gebäuden (beispielsweise Gebäudehöhe, Art des Gebäudes) wurden objektbezogen erfasst und verarbeitet. Ein Rückschluss auf einzelne Personen war damit nicht möglich und ohnehin zur Erreichung der Projektziele irrelevant.

Weiterhin erfasst und für die Nutzung bei der Kriminalitätsprognose vorgesehen waren öffentlich zugängliche Kalenderdaten, wie beispielsweise Feiertage, Schulferien oder Zeiten des Sonnenauf- bzw. -untergangs. Eine Routine zum Einlesen der öffentlich zugänglichen meteorologischen Daten des Deutschen Wetterdienstes, die jeweils für die Einzugsbereiche bestimmter Wetterstationen gelten, wurde umgesetzt. Auch hierbei handelte es sich um Daten ohne Personenbezug.

Die polizeilichen Daten wurden durch das LZPD NRW über das gesicherte Polizeinetz zur Verfügung gestellt. Eine Nutzung dieser Daten fand ausschließlich durch die Mitglieder des TP WBA statt. Die Daten der NEXIGA GmbH wurden mittels eines verschlüsselten Downloads von einem FTP-Server, zu dem ausschließlich die Mitglieder des TP WBA einen passwortgeschützten Zugang erhielten, übermittelt.

Für die Datenanalyse wurde ein eigener Auswerteraum im LKA NRW mit passwortgeschützten Clients eingerichtet, zu dem ausschließlich ebenfalls die Mitglieder des TP WBA Zugang hatten. Alle externen Personen (beispielsweise Bedienstete von Unternehmen im Falle von Wartungsarbeiten am System oder fachliche Berater bei der Modell- und Prognoseerstellung) wurden zur Wahrung des Datengeheimnisses gemäß dem Gesetz über die förmliche Verpflichtung nichtbeamteter Personen (Verpflichtungsgesetz) verpflichtet, bevor sie im Rahmen des Projektes tätig wurden. Zudem fand die Arbeit externer Personen am System zu keiner Zeit unbegleitet statt. Die Auswertungen erfolgten in einem Sondernetz auf einem Server in einem gesicherten Technikraum, der weder mit dem sonstigen polizeilichen Datennetz noch mit dem Internet verbunden war. Sämtliche in das Sondernetz transferierten Daten wurden im Interesse der IT-Sicherheit über einen „Schleusenrechner“ eingespeist. Auf diese Weise erfolgte eine Prüfung sämtlicher Daten auf Viren und sonstige Schadsoftware.

Aus vorgenannten Gründen wurden ein Verzeichnismäßig § 8 DSGVO NRW und ein IT-Sicherheitskonzept zur Prüfung der technisch-organisatorischen Maßnahmen gemäß § 10 DSGVO NRW entwickelt. Die datenschutzrechtlichen Vorgaben wurden hierbei streng nach der kriminalfachlichen Erforderlichkeit erfüllt. Zudem wurden für die Anwendung SKALA | MAP (siehe Kapitel 4.3.4) ein Musterverfahrensverzeichnis und IT-Sicherheitskonzept erstellt, die den Pilotbehörden für eine notwendige, örtlich spezifizierte Anpassung und Vorabkontrolle durch den dortigen behördlichen DSB bereitgestellt wurden.

Der Landesdatenschutzbeauftragte bzw. die Landesdatenschutzbeauftragte (LDI) in NRW wurden mehrfach mit Berichten über den Stand der Fortentwicklung informiert und entsprechende Stellungnahmen gefertigt (zuletzt am 16.06.2016). Darüber hinaus gab es mehrfach Einladungen zur vertiefenden Erläuterung sowie einer Führung durch den Auswerterraum mit Darstellung der Hard- und Softwarekomponenten. Letztere Einladungen wurden allerdings seitens der LDI NRW nicht wahrgenommen.

Im Projektzeitraum wurden zwei Anfragen gemäß § 4 Abs. 1 IFG NRW in Abstimmung mit dem Datenschutzbeauftragten des LKA NRW beschieden. In einem Fall wurde eine Teilbeauskunftung durchgeführt, ansonsten lagen die Ablehnungsvoraussetzungen gemäß § 6 lit. a) IFG NRW vor.

3.6 Öffentlichkeitsarbeit

Die Öffentlichkeitsarbeit war im Projekt SKALA direkt an die Projektleitung und Geschäftsführung angegliedert. Sowohl für die interne wie auch die externe Öffentlichkeitsarbeit wurden spezifische Konzepte, mit festen Ansprechpartnern aus den Teilprojekten, entwickelt und realisiert.

Das Interesse am Projekt SKALA war während der gesamten Projektlaufzeit groß. Es gab insgesamt 50 schriftliche und 22 mündliche Presseanfragen. Die Presseanfragen kamen u. a. von der Frankfurter Allgemeinen Zeitung, der Rheinischen Post, der BILD-Zeitung, dem c't-Magazin, RTL, N24 und dem Westdeutschen Rundfunk. Viele der Anfragen führten zu persönlichen oder telefonischen Interviews oder es wurden Filmbeiträge, beispielsweise für WDR-Westpol oder das RTL-Nachtjournal, gedreht. Die Verwertung der vom LKA NRW gelieferten Beiträge fand außerordentlich sachlich statt. Beiträge, die das Projekt SKALA durchweg kritisch oder negativ beleuchteten, gab es nicht. Im März 2017 informierte sich der Bundesvorsitzende des Bundes Deutscher Kriminalbeamter vor Ort über den Sachstand (Ansatz, Fortschreibung und erste Erfahrungswerte) des Projektes. Im Mai 2017 stellte der damalige Innenminister Ralf Jäger das Projekt SKALA auf einer Landespressekonferenz im LKA NRW erstmalig der Öffentlichkeit vor.

Im Rahmen wissenschaftlicher Auseinandersetzungen hat Predictive Policing als Themenfeld an enormer Prominenz

gewonnen. Darüber hinaus wurden zahlreiche wissenschaftliche Artikel durch Mitglieder des TP WBA schon während der Projektlaufzeit veröffentlicht, wie der beispielhaften Auswahl zu entnehmen ist:

- Seidensticker, Kai (2017):
Kriminalität in Mikrosegmenten
In: forum kriminalprävention 04/2017
- Stoffel et al. (2017):
Qualitätsmetriken im Bereich von Predictive Policing. Die Variabilität und Validität von Treferraten.
In: Polizei & Wissenschaft 04/2017
- Pollich/Bode (2017):
Predictive Policing: Zur Notwendigkeit eines (sozial-) wissenschaftlich basierten Vorgehens
In: Polizei & Wissenschaft 03/2017
- Bode et al. (2017):
Variabilität und Validität von Qualitätsmetriken im Bereich von Predictive Policing
In: Konstanzer Online-Publikations-System (KOPS)

Auf polizeiinternen Seminaren und Fachtagungen wie auch auf externen Veranstaltungen mit Bezug zur Thematik Predictive Policing wurden zahlreiche Vorträge gehalten. Eine beispielhafte Auswahl ist der nachfolgenden Übersicht zu entnehmen:

- Fachtagungen K, V, GE und Leiter Leitungsstäbe
- Seminare beim Landesamt für Ausbildung, Fortbildung und Personalangelegenheiten der Polizei NRW (LAFP NRW)
- Seminare der Deutschen Hochschule der Polizei (DHPol)
- Deutscher Städte- und Gemeindebund
- Europäischer Polizeikongress

Ferner wurden Informationsblätter (Fact Sheets) in Deutsch und Englisch gestaltet, um über die wesentlichen Fakten des Projektes (Ziele und Projektbeteiligte) zu informieren.

Neben der breiten internen wie externen Öffentlichkeitsarbeit wurde zudem während der Projektlaufzeit auch ein eigenes Logo für SKALA entwickelt:

Abbildung 5: SKALA-Logo

3.7 Predictive Policing als Prozess

In der Praxis umfasst Predictive Policing mehrere Arbeitsschritte und Prozesse, die aufeinander aufbauen, angefangen mit der Erfassung und Aufbereitung von Daten, die für die Kriminalitätsprognosen erforderlich sind. Die mit Predictive Policing verbundenen Prozesse werden in der Literatur meist mit dem von Perry et al. (2013: 12) entwickelten Prediction-Led Policing Business Process visualisiert:

Abbildung 6: Prediction-Led Policing Business Process

Quelle: Perry et al. 2013: 12.

Der Prediction-Led Policing Business Process dokumentiert den idealtypischen Kreislauf von Arbeitsschritten und die Wechselwirkungen, die mit Predictive-Policing-Verfahren verbunden sein können. Die Ausgestaltung des methodischen Prozesses der Kriminalitätsprognose wird jedoch in dem von Perry et al. visualisierten Business Process nicht thematisiert. Eine Möglichkeit der Illustration (siehe Abbildung 7) eines solchen Prozesses findet sich bei Bode et al. (2017: 2). Die Prozessabbildung ermöglicht eine Einsicht in

die einzelnen Schritte zur Umsetzung von Predictive Policing aus polizeilicher Sicht, wie sie auch im Projekt SKALA stattgefunden hat. Methodische Abweichungen von dieser Prozessabbildung sind denkbar, dürften aber zumindest ähnliche Ausgestaltungen haben, sofern auf Techniken des maschinellen Lernens zurückgegriffen wird.²⁴

Schritt 1: Daten

Im ersten Schritt wird mit der Sichtung, Auswahl und Aufbereitung von Daten begonnen. Mit der entsprechenden Software ist es möglich, verschiedene Datenquellen raum- und zeitbezogen zusammenzuführen. In der weiteren Umsetzung (Schritt 2 und 3) existieren eine Vielzahl unterschiedlicher Methoden, angefangen bei rein auf den Near-Repeat-Ansatz fokussierten Lösungen (vgl. Balogh 2016) bis hin zu Ansätzen, die sich stärker auf verschiedenste wissenschaftliche Theorien stützen (vgl. Pollich/Bode 2017). Im Projekt SKALA wurde sich bewusst, wie zuvor in Kapitel 3.2 beschrieben, auf ein hypothesengestütztes Vorgehen konzentriert und damit ein statistischer Ansatz gewählt.

Die Modellierung und die Prognoseberechnung lassen sich begrifflich unter „Statistik“ und „Maschinelles Lernen“ fassen und beinhalten die methodische Ausgestaltung.

Schritt 2: Modellierung

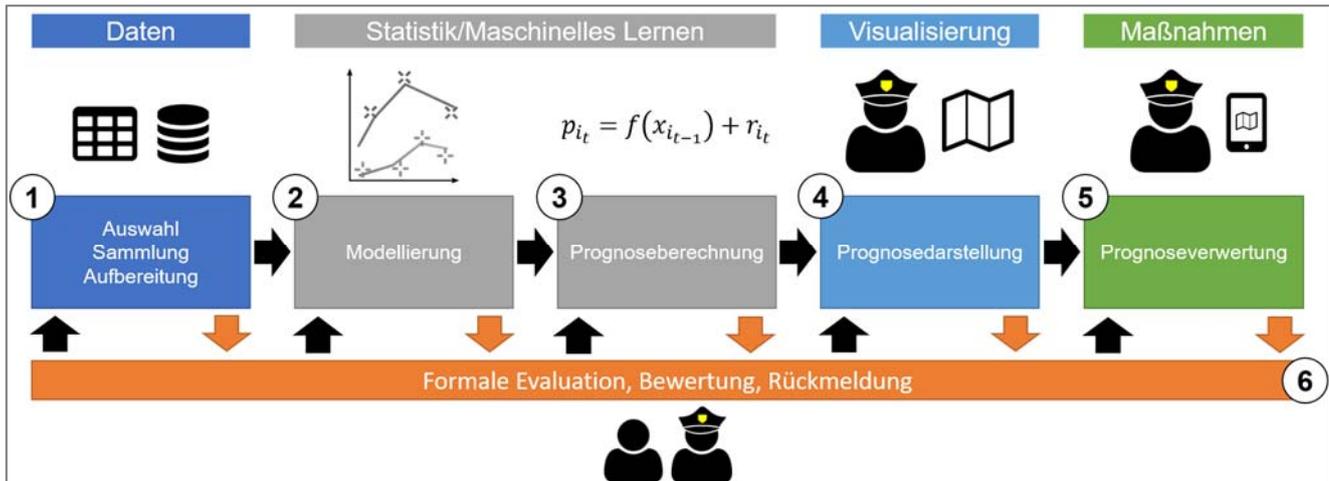
Schritt 2 beginnt i. d. R. mit der Verwendung eines konkreten Modells, unter Nutzung von historischen Daten, um die Kriminalitätslage möglichst angemessen in allen gewünschten Facetten abzubilden. Ziel ist es, hierauf aufbauend eine Prognoseberechnung durchzuführen. In diesem Zusammenhang sind z. B. Modellierungen mittels Regressionen (vgl. Box et al. 2015: 305 ff.), Entscheidungsbäumen (z. B. Chi-square Automated Interaction Detection [CHAID], vgl. Kass 1980) oder künstlicher neuronaler Netze möglich (vgl. Zhang/Qi 2005). Im Projekt SKALA wurde im Wesentlichen mit Entscheidungsbäumen und einem CHAID-Algorithmus gearbeitet

Schritt 3: Prognoseberechnung

Mit dem zuvor erstellten Modell erfolgt eine Anwendung auf aktuelle oder künftige Daten, um die Wahrscheinlichkeit eines bestimmten Deliktes in einem Gebiet zu ermitteln. Hierbei handelt es sich um den eigentlichen Schritt der Prognoseberechnung, der den wesentlichen Bestandteil im Predictive-Policing-Prozess darstellt. Das Prognoseergebnis im Projekt SKALA war eine Auswahl an Wohnquartieren, die ein höheres Kriminalitätsrisiko aufweisen als andere Wohnquartiere in derselben Stadt und im gleichen Gültigkeitszeitraum. Die Prognosen wurden wöchentlich berechnet und an die Pilotbehörden weitergegeben (siehe Kapitel 4.3).

²⁴ In der nachfolgenden Beschreibung des Predictive-Policing-Prozesses wird sich eng an die Formulierungen und den Aufbau des Artikels von Bode et al. (2017) bzw. Stoffel et al. (2017) angelehnt.

Abbildung 7: Predictive-Policing-Prozess



Quelle: Bode et al. 2017: 2.

Schritt 4/5: Prognosedarstellung und -verwertung

Im vierten Schritt muss die adäquate Darstellung der Kriminalitätsprognosen vorgenommen werden, um sie anschließend bei operativen Polizeikräften einsetzen zu können (Schritt 5). Mit zunehmender Digitalisierung sind neben Karten im Papierformat zwischenzeitlich auch Prognosevisualisierungen auf Tablet-PCs oder Mobilfunkgeräten möglich, was dazu führte, dass im Projekt SKALA Kriminalitätsprognosen auch an das Projekt „Mobil auf Streife“ übermittelt wurden, sodass die Kriminalitätsprognosen auch auf Tablet-PC in den entsprechend ausgestatteten Pilotbehörden visualisiert werden konnten. Darüber hinaus wurde in Zusammenarbeit mit dem Lehrstuhl für Datenanalyse und Visualisierung der Universität Konstanz das Visualisierungs- und Auswertetool SKALA | MAP entwickelt. Gerade das Forschungsfeld der Visual Analytics (vgl. Keim et al. 2010) konnte hier besondere Flexibilität und Aufgabenangemessenheit garantieren.

Schritt 6: Evaluation

Die Evaluation, bezogen auf den gesamten methodischen Predictive-Policing-Prozess, ist in Schritt 6 enthalten. Im Regelfall handelt es sich um die formale, statistische Beschreibung von beobachteten Effekten. Da durchgehende Plausibilitätsprüfungen und ad hoc Verifikationen von Zwischenergebnissen stets erforderlich wären, z. B. um die Eignung der gewählten Methode in den Schritten 2 und 3 oder der gewählten Visualisierungstechnik in Schritt 4 kontinuierlich sicherzustellen, wird dies von Bode et al. (2017: 2) mit durchgehenden Ein- (orangene Pfeile) und Ausgaben (schwarze Pfeile) visualisiert. Für das Projekt SKALA und die Ergebnisse der Evaluation wird an dieser Stelle auf den gesonderten Evaluationsbericht verwiesen.

3.8 Prognoseflächen als räumliche Bezugsgröße

Die räumliche Dimension von Kriminalität wurde bereits seit Mitte des 20. Jahrhunderts zum Gegenstand kriminologischer und kriminalgeografischer Forschung, insbesondere

durch die sogenannte Chicago-Schule (vgl. Park et al. 1925). Das im Projekt SKALA gewählte methodische Vorgehen lässt sich der georäumlichen Modellierung zuordnen, da Kriminalitätsprognosen für räumlich abgegrenzte Gebiete generiert wurden. Dabei bedingt grundsätzlich der gewählte methodische Ansatz die Entscheidung, ob Prognosen für ganze Städte oder nur für einzelne Teilbereiche einer Stadt generiert werden sollen, wobei die letzte Option zusätzlich einer nachvollziehbaren Entscheidung darüber bedarf, welche Gebiete einer Stadt in die Kriminalitätsprognose einbezogen werden sollen. Aus diesem Grund kam auch der Wahl einer geeigneten Flächeneinheit für die Prognosen im Projekt SKALA ein hoher Stellenwert zu. Der Begriff „räumliche Bezugsgröße“ bezieht sich auf Gebiete, in welche der gesamte Prognoseraum (z. B. eine Stadt) für die Modell- und Prognoseerstellung unterteilt wird. Hierzu ist die Anwendung eines Rasters, das auf den Prognoseraum gelegt wird, ein sehr gängiges Prinzip. Im Falle von SKALA wurde als räumliche Bezugsgröße jedoch die Wohnquartierebene gewählt, bei der die Wohnquartiergrenzen durch konstruierte Polygone definiert werden. Polygone sind abgeschlossene geometrische Flächen, die durch geografische Koordinaten einzelner Punkte und der Verbindungslinien zwischen diesen Punkten begrenzt werden.

Die Gebiete (hier Wohnquartiere), für die die höchsten Kriminalitätswahrscheinlichkeiten im Vergleich zu anderen Gebieten des gesamten Prognoseraumes berechnet wurden, werden hier als Prognosegebiete definiert. Dabei wurde ihr Anteil auf etwa 1,5 Prozent der Gesamtquartieranzahl jedes Polizeibezirkes beschränkt (siehe Kapitel 4.2).

Die Wahl einer geeigneten räumlichen Bezugsgröße kann als Grundsatzfrage bei der Umsetzung von Predictive Policing verstanden werden und ist stark vom gewählten theoretischen Ansatz der Umsetzung abhängig. Denn je nach gewählter räumlicher Dimension können sich unterschiedliche

Wirkungen in den einzelnen Prozessschritten – von der Modellierung bis zur Umsetzung polizeilicher Interventionen – einstellen. Diese Auswirkungen sollen nachfolgend skizziert werden, insbesondere mit Blick auf die methodischen Gründe, aufgrund derer im Projekt SKALA Wohnquartiere bewusst als räumliche Bezugsgröße ausgewählt wurden.

Die Festlegung auf eine bestimmte Art von Prognosefläche als räumliche Bezugsgröße bedingt, dass durch die konstruierten Polygone räumliche Grenzen innerhalb eines Gebietes erzeugt werden, wodurch die reale und natürliche Struktur des Raumes verzerrt werden kann. Die Folge kann sein, dass Zusammenhänge zwischen Kriminalitätsereignissen, die eigentlich nur eine geringe räumliche Distanz zueinander aufweisen, im Prozess der Modellierung und Prognoseberechnung nicht erkannt werden, da diese durch die künstlich erzeugten Polygongrenzen räumlich voneinander getrennt wurden.

Darüber hinaus wird durch die gewählte räumliche Bezugsgröße regelmäßig die Aggregation von Daten erforderlich, da sich diese meist auf unterschiedliche räumliche Bezugsgrößen beziehen, wie beispielsweise einzelne Punkte oder auch große Flächen/Gebiete. Die gewählte räumliche Bezugsgröße definiert somit das „Aggregationsniveau“ bei der Datenaufbereitung und der Modell- und Prognoseerstellung. So mussten im Projekt SKALA beispielsweise Daten, die sich auf die Gebäude- oder Straßenabschnittsebene beziehen, auf Wohnquartierebene aggregiert werden. Eine solche Aggregation von Daten, z. B. durch die Bildung von Mittel- oder Modalwerten, führt zwangsläufig dazu, dass Detailinformationen verloren gehen. Dieser Verlust nimmt grundsätzlich mit der Ausdehnung der Raumeinheiten zu. Zu große Prognoseflächen beherbergen dementsprechend häufig nur einen geringen Detailgrad und die Aussagekraft der darin enthaltenen Informationen ist niedrig. Wird der Prognoseraum allerdings zu klein gewählt, kann dies zu einer Vielzahl von unüberschaubaren und kaum mehr handhabbaren Details führen, was wiederum für die Modellierung von Kriminalitätsprognosen nicht zuträglich ist. Die räumliche Heterogenität der Daten ist in solchen Fällen oft sehr groß, was die Berechnung von statistischen Zusammenhängen zwischen möglichen Einflussfaktoren und dem Kriminalitätsgeschehen erschwert. Die Problematik der Aggregation von Daten auf unterschiedliche räumliche Ebenen bzw. die Abhängigkeit der Aussagekraft der Daten von der Definition der Raumeinheit ist in der Literatur als *Modifiable Areal Unit Problem* (MAUP) hinreichend bekannt (vgl. insbesondere Openshaw 1983).

Grundsätzlich besteht die Möglichkeit, einzelne Delikte und deren Verortung im Raum als eigene Prognosefläche zu betrachten und darauf aufbauend Kriminalitätsprognosen zu erstellen. Dies bringt aber keinen großen Erkenntnisgewinn, da der Eintritt eines kriminellen Ereignisses (z. B. ein WED an einer konkreten Anschrift) im Verhältnis zu allen anderen

in Frage kommenden Räumen (beispielsweise alle anderen potenziellen Anschriften in einer Stadt, an denen ein WED auftreten könnte) zu selten ist. Anders formuliert tendiert die statistische Wahrscheinlichkeit, dass am nächsten Tag in Haus X eingebrochen wird, gegen Null. Dagegen ist die statistische Wahrscheinlichkeit, dass im Stadtviertel Y, in dem Haus X liegt, am nächsten Tag eingebrochen wird, schon deutlich höher. Kriminalitätsprognosen müssen sich folglich immer auf einen ausreichend großen Raum beziehen. Auf diese räumliche Bezugsgröße werden alle in der Vergangenheit dokumentierten, darin liegenden Kriminalitätsereignisse aggregiert.

Die Betrachtung von Predictive-Policing-Verfahren kommerzieller Anbieter zeigt, dass in diesem Zusammenhang die unterschiedlichsten Raumdefinitionen für Prognosegebiete als räumliche Bezugsgrößen verwendet werden, um Prognose relevante Flächen auf einer Karte darzustellen. Neben der Flächenform variieren auch die Größen der einzelnen Flächen sowie die räumliche Abdeckung der Gesamtfläche. Zumeist werden Gitterzellen (Raster) bestimmter Größen, sogenannte Boxen oder aber Kreise, als Prognoseflächen benutzt. Die Praxis, anhand von Boxen (siehe Abbildung 8) Predictive Policing methodisch umzusetzen, stellt hierbei die wohl prominenteste Variante dar. Die Definition von Boxen erfolgt dabei über ein Gitternetz mit einer bestimmten Rastergröße, das über eine Stadt gelegt wird und diese so in einzelne Boxen teilt. Die Rastergrößen reichen von 150m x 150m bis zu 400m x 400m pro Box und werden zumeist normativ festgelegt, wobei auch Methoden der Punktanalyse zur Definition eines passenden Rasters angewendet werden können (beispielsweise Nearest-Neighbour-Analysis; vgl. Bailey/Gatrell 1995).

Je nach gewählter Ausdehnung der entsprechenden Boxen ermöglicht diese Methode eine kleinräumige Abgrenzung von Räumen und somit auch die Fokussierung auf detaillierte Raumphänomene. Darüber hinaus ist eine gitternetzartige Einteilung des Raumes in nur kurzer Zeit umsetzbar und erfordert eine geringe Implementierungsdauer. Allerdings gehen mit dieser Art der Raumdefinition auch beträchtliche Risiken und eine große Unschärfe einher. Bloße Rastereinteilungen orientieren sich – insbesondere in Regionen organisch gewachsener Städte – weder an geografischen Barrieren noch an individuellen Merkmalen der Stadtarchitektur. So läuft eine Einteilung in kleinräumige Raster Gefahr, räumlich und auch sozial homogene Gebiete wahllos zu zerteilen und in der Folge als jeweils abgeschlossene Einheiten zu behandeln. Dies führt dazu, dass der Raum in seiner Funktion lediglich auf eine im weitesten Sinne willkürliche Projektionsfläche von Kriminalität reduziert wird und dieser in der Folge aus kriminalgeografischer Sicht zur Beschreibung von Kriminalität nur eingeschränkt tauglich ist.

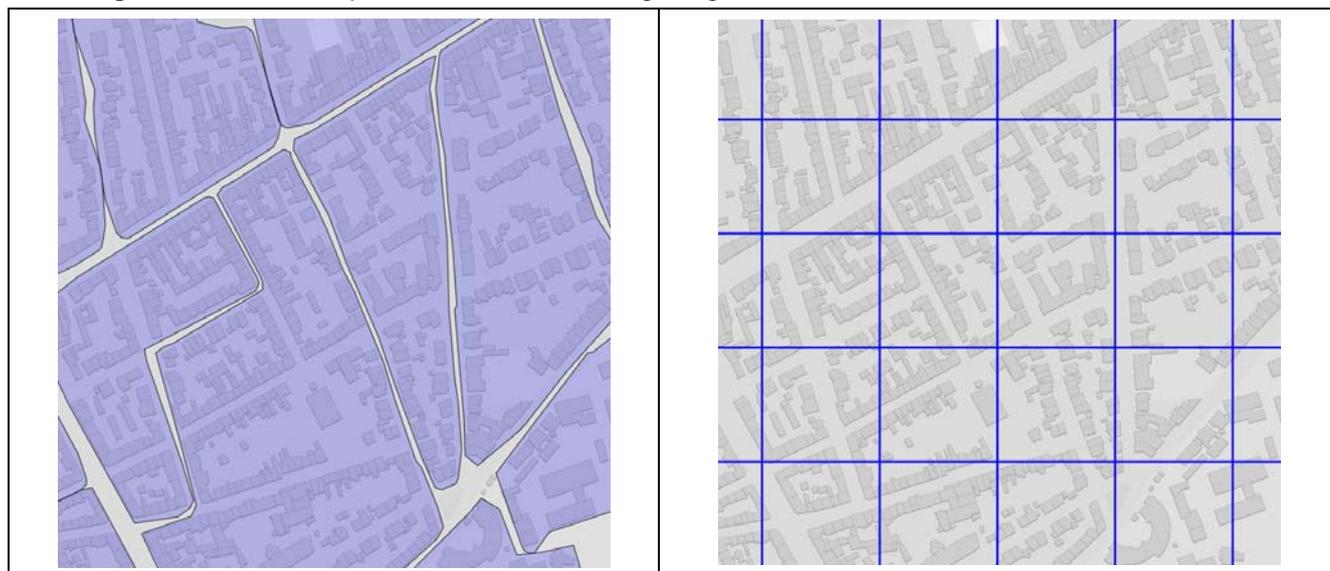
Ähnliche Probleme ergeben sich bei der räumlichen Definition durch Kreise. Im Gegensatz zu Zellen, die als räumliche

Bezugsgröße bei der Modellierung und Prognose verwendet werden, dienen Kreise i. d. R. primär der Visualisierung von Kriminalitätshotspots oder Risikogebieten, nicht aber als Prognoseebene. Beispielsweise werden um Mittelpunkte von mehreren Vorgangskorridoren Kreise mit einem definierten Radius gezogen, um darzustellen, dass es in diesen Bereichen in der letzten Zeit vermehrt zu Einbrüchen kam. Die Kreise orientieren sich daher in ihrer Ausdehnung an einem zuvor festgelegten Radius und teilen ebenfalls ein homogenes Gebiet ausschließlich aufgrund der Entfernung zum Zentrum des Kreises.

Um eine möglichst hohe Passung der Prognosemodellierung und -umsetzung mit den örtlichen Gegebenheiten der Stadtstruktur zu erzielen, erhielten im Projekt SKALA Wohnquartiere (siehe Abbildung 8) als räumliche Bezugsgröße den Vorzug vor den oben angegebenen Lösungen. Diese orientieren sich an sozialen Räumen und/oder geografischen Barrieren, was dazu führt, dass anhand unterschiedli-

cher Merkmale (beispielsweise anhand der Anzahl der Einwohner) möglichst homogene Einheiten abgebildet werden. Zudem hat sich das Wohnquartier in der Wissenschaft und insbesondere in der städtebaulichen Kriminalprävention als kleinräumige Analyseeinheit etabliert und gilt als gute Betrachtungsgröße (vgl. Abt/Schröder 2017: 42). Eine kleinräumige Betrachtung, die in diesem Kontext die lokal unterschiedlichen Strukturen einbezieht, ist auch nach Fotheringham et al. (2002: 1 f.) einer globalen Betrachtung vorzuziehen, die Homogenität ausschließlich anhand von gleichförmigen Quadraten erzeugt und suggeriert. Je stärker sich die datenbasierte Abbildung an realen räumlichen Gegebenheiten orientiert, desto genauere Aussagen können mit Kriminalitätsprognosen getroffen werden. Dieses Vorgehen erfordert allerdings, im Gegensatz zur Rastererstellung, zusätzliches Datenmaterial zur Bestimmung der jeweiligen Quartiergrenzen und ist im Verhältnis zur Erstellung von Boxen zeitintensiver in der Berechnung. Exemplarisch wird der Unterschied zwischen Wohnquartieren und Boxen, insbesondere im Hinblick auf die Stadtstruktur, in Abbildung 8 dargestellt.

Abbildung 8: Illustration Wohnquartiere und Boxen als Prognosegebiete



Quelle: Bode et al. 2017: 5.

Die im Projekt SKALA gewählten Wohnquartiere orientieren sich in ihrem Ursprung an ehemaligen Stimmbezirken, die durchschnittlich 400 Haushalte mit größtmöglicher Homogenität zusammenfassen (vgl. Nexiga 2017). Dementsprechend kann die Fläche der Wohnquartiere erheblich variieren. Eher kleine Wohnquartiere finden sich vermehrt im Stadt(teil)zentrum, wohingegen die Quartiere am Stadtrand bzw. in eher ländlich geprägten Gebieten sehr groß ausfallen können. Sie zeichnen aber ein Bild der Stadtstruktur, das sich sehr nah an der tatsächlich wahrnehmbaren Struktur orientiert. Entsprechend positive Rückmeldungen von operativen Polizeikräften aus den Pilotbehörden im Projekt SKALA runden diese theoretische Berechnung praxistauglich ab.

Wohnquartiere bilden beim WED die diesem Deliktfeld inhärente Tatgelegenheitsstruktur ab, also die in einem Quartier existenten Haushalte. Gleiches gilt für die später ebenfalls berechneten Prognosen für Kraftfahrzeugdelikte. Im Rahmen der Prüfung der Prognose-tauglichkeit von Einbruch-diebstahl aus Gewerbeobjekten stellten sich Wohnquartiere dagegen als weniger geeignet heraus, da sich auch die Tatgelegenheitsstruktur an anderen Faktoren orientiert. Hier wird deutlich, dass die Raumdefinition entscheidend von der Wahl des Deliktes bzw. der zuvor durchgeführten Deliktanalyse abhängt. Somit wurden für den Einbruchdiebstahl aus Gewerbeobjekten nachträglich und ergänzend zu den Wohnquartieren eigene Prognoseflächen als räumliche Bezugsgröße erstellt, die sich an der Verteilung der Gewerbe-

objekte bzw. der Flächendefinition innerhalb der Städte orientieren. Flächen, die als Gewerbefläche oder Flächen gemischter Nutzung ausgewiesen wurden, wurden zu handhabbaren Einheiten zusammengefügt, sodass sinnvolle und für die Fachdienststellen und operativen Polizeikräfte nachvollziehbare Flächen entstanden.

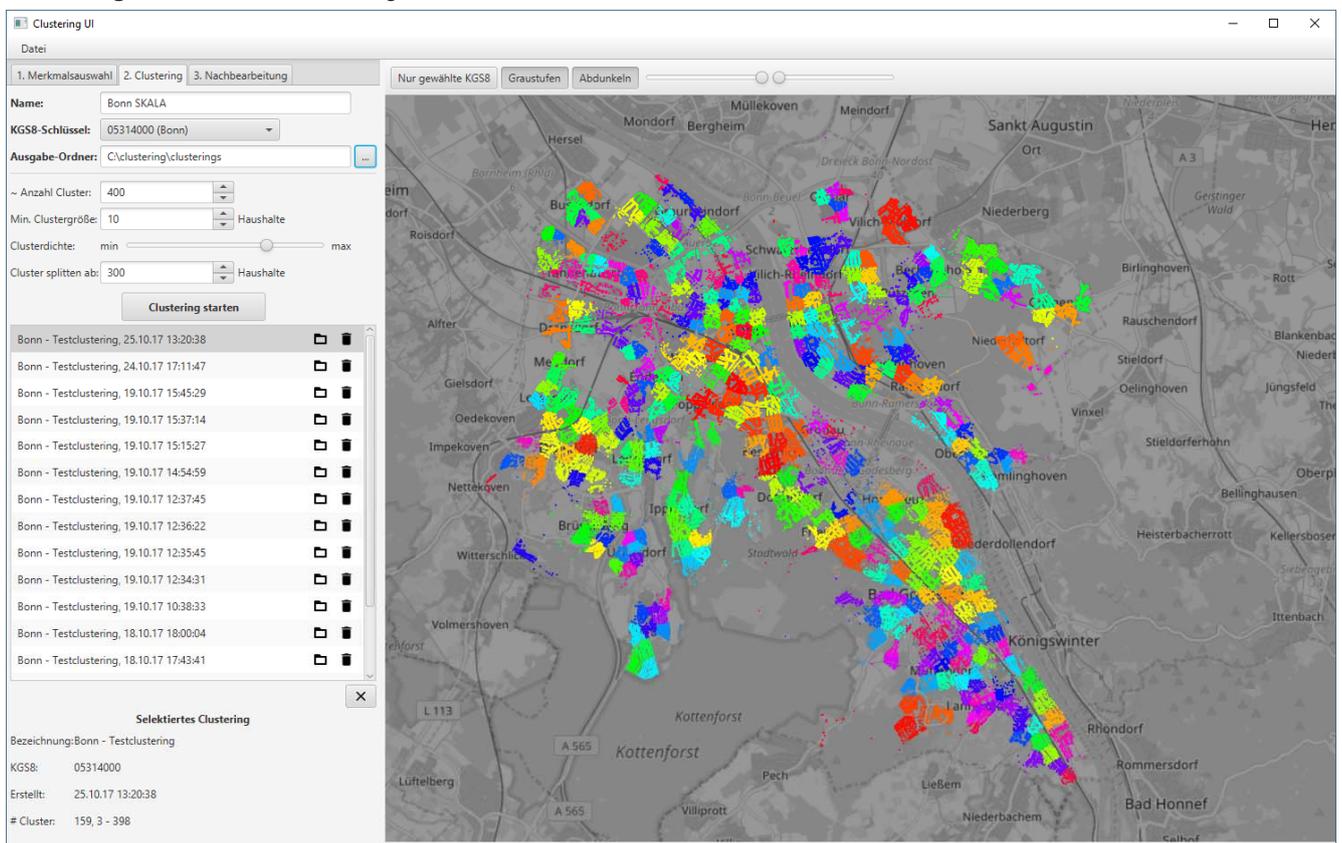
Alternativ zur Umsetzung mit Wohnquartieren als Prognoseflächen wurden auch schon existente Raumdefinitionen der Polizei auf ihre Eignung geprüft. Beispielsweise orientiert sich die polizeiliche Tätigkeit, insbesondere durch die Steuerung über vorhandene Einsatzleitsysteme, an bekannten Raumdefinitionen. Polizeiinspektionen (PI), Wachbezirke, Fahndungsräume oder andere Einheiten können demnach, unter Beachtung der bereits angesprochenen Problemfelder, als Prognoseflächen genutzt werden. Im Ergebnis war die Nutzung solcher Gebiete allerdings nicht zu realisieren, da die Raumdefinitionen einer PI oder eines Wachbereichs grundsätzlich ein zu großes Gebiet umfassen und hierauf erstellte Kriminalitätsprognosen deshalb keinen Mehrwert für operative Polizeikräfte gebracht hätten. Fahndungsräume stellten in diesem Kontext zwar eine interessante Alternative dar, konnten aber aufgrund von nur punktuell vorhandenem digitalen Datenmaterials nicht in die Analyse einbezogen werden.

Um noch mehr Homogenität bei den Wohnquartieren im Hinblick auf Einwohneranzahl, Bebauung und Sozialstruktur zu gewinnen, wurden diese in Zusammenarbeit mit dem

Lehrstuhl für Datenanalyse und Visualisierung der Universität Konstanz überarbeitet. Denn es stellte sich heraus, dass die tatsächliche Anzahl an Haushalten in den Wohnquartieren zum Teil stark schwankte, da die Gebiete auf sehr alten Wahlbezirksgrenzen beruhen. Ferner waren deswegen auch noch teilweise Gewerbe- und Grünflächen in den Wohnquartieren enthalten. Die Wohnquartierstruktur wurde folglich nicht immer optimal abgebildet. Mit Mitarbeitern der Universität wurde aus diesen Gründen ein eigenständiges Clustering-Tool entwickelt und programmiert, das den Raum anhand von soziostrukturellen und soziodemographischen Merkmalen homogen aufteilen kann. Dies ermöglichte ein auf das zu prognostizierende Delikt ausgelegtes geografisches Clustering anhand von selbst zu bestimmenden Merkmalen. Für diesen Zweck wurde der Clusteralgorithmus DBSCAN (Density-Based Spatial Clustering of Applications with Noise; vgl. Ester et al. 1996) genutzt. Die hierdurch neu berechneten Wohnquartiere weisen eine noch stärkere Homogenität auf als dies bei den vorhandenen Wohnquartieren bereits der Fall war und ermöglichten eine höhere Güte des Prognosemodells (größere Wahrscheinlichkeiten).

Exemplarisch zeigt Abbildung 9 einen Screenshot des Clustering-Tools. Mit dem Tool ist es perspektivisch möglich für andere Gebiete, in denen SKALA Anwendung finden soll, homogene Wohnquartiere als Basis für Kriminalitätsprognosen, zu erstellen. Das Clustering-Tool ist interaktiv und bedarf keiner speziellen IT-Fachkenntnisse.

Abbildung 9: Screenshot Clustering-Tool



3.9 Prognosezeit als temporale Bezugsgröße

Neben der räumlichen Bezugsgröße ist auch die Prognosezeit als temporale Bezugsgröße von Bedeutung. Denn die gewählte Prognosezeit wirkt sich erheblich auf den Predictive-Policing-Prozess aus. So bestimmt die Prognosezeit Entscheidungen zur retrograden Zeitspanne der einzubeziehenden Daten(-sätze) zwecks Modellierung, deren jeweilige Gewichtung sowie die zeitliche Gültigkeit einer erstellten Kriminalitätsprognose.

Die Gültigkeit von Prognosen kann sich auf unterschiedliche Zeiträume beziehen und ist abhängig von dem gewählten Delikt, dem Prozess der Prognoseübermittlung und den beabsichtigten polizeilichen Reaktionen auf eine Prognose. Kurze Zeiträume bieten sich insbesondere an, wenn das Delikt über eine hohe Fallzahl verfügt und eine entsprechende Dynamik zu erwarten ist (z. B. WED), wohingegen geringe Fallzahlen und wenig Dynamik für längere Gültigkeitszeiträume sprechen (z. B. Einbruchdiebstahl aus Gewerbeobjekten). Der Prozess der Prognoseübermittlung setzt sich aus der technischen Umsetzung der Prognose(-visualisierung) und den Bewertungsschnittstellen zusammen. Je mehr Schnittstellen im Prozess vorhanden sind, desto länger muss der Gültigkeitszeitraum gewählt werden, um eine Umsetzung von Interventionen am Ende des Prozesses innerhalb dieses Zeitraumes noch zu ermöglichen. Täglich erstellte Kriminalitätsprognosen mit einer Gültigkeit von ebenfalls einem Tag dürfen demzufolge keine bzw. nur wenige Schnittstellen durchlaufen, bis diese operativ bemaßnahmen werden. Eine so erzeugte operative Dynamik ist in Anbetracht der beabsichtigten Interventionen und verfolgten Ziele kritisch zu reflektieren. Je größer der Gültigkeitszeitraum, desto eher verschieben sich Intention und Wirkung einer Kriminalitätsprognose und verändern ihren Charakter hin zu einem strategischen Werkzeug. Da die fachliche Bewertung der Prognosen einen wichtigen Aspekt in der Umsetzung dieses Projektes darstellt, wurde ein Gültigkeitszeitraum von sieben Tagen für den WED gewählt. Dieser Zeitraum wurde von allen Projektbeteiligten, insbesondere von den operativen Polizeikräften, als gutes Maß empfunden. Für die Prognose von Kraftfahrzeugdelikten galt dasselbe. Für die Umsetzung der Prüfung der Prognosefähigkeit von Einbruchdiebstahl aus Gewerbeobjekten wurde hingegen, aufgrund der niedrigeren Fallzahlen und geringeren Dynamik, ein Gültigkeitszeitraum von 14 Tagen gewählt. Auch dieser Zeitraum wurde mit den operativen Polizeikräften abgestimmt.

3.10 Datenqualität im Predictive-Policing-Prozess

Predictive Policing besteht, wie zuvor dargestellt, aus mehreren aufeinander aufbauenden Schritten. Alle Prozessbestandteile hängen von den zu verarbeitenden Daten und deren Aufbereitung ab. Kommen bereits Datenfehler im ersten Schritt zustande, wirken sich diese auf den gesamten Predictive-Policing-Prozess aus. Die Verlässlichkeit des gesamten Prozesses und damit der Ergebnisse kann somit durch

fehlerhafte Datenauswahl oder -aufbereitung in Frage gestellt werden. Es gibt unterschiedlichste Arten von Fehlerquellen und Fehlertypen in Datensätzen, die bei der Datenaufbereitung im Rahmen von Predictive Policing berücksichtigt werden müssen und welche auch im Rahmen des Projektes SKALA sehr akribisch behandelt wurden, um spätere Modellfehler zu vermeiden. Einige Datenfehler haben einen eher technischen Ursprung und beruhen auf der Tatsache, dass Daten falsch oder unzureichend in der Datenbank eingetragen wurden. Dazu zählen beispielsweise Gebäudekoordinaten, die sich auf einen falschen Punkt beziehen oder fehlende Angaben beim Beuteschema. Zum anderen liegen häufig schlicht keine besseren oder genaueren Daten vor. Ein Beispiel ist der Tatzeitpunkt, der sich aufgrund fehlender Information in einigen Fällen über mehrere Wochen erstreckt, beispielsweise wenn das Opfer erst nach einer Urlaubsabwesenheit merkt, dass eingebrochen wurde.

Neben diesen konkreten, datenbezogenen Problemen nimmt bei der Bewertung von Predictive Policing ebenso die sogenannte Datenunsicherheit eine bedeutende Rolle ein (vgl. Morgan et al. 1990; Fritsch et al. 1998; Kinkeldey et al. 2014). Hier ist nicht bekannt, in welchem Umfang Fehler in den verwendeten Daten enthalten sind. Mit Bezug zu Kriminalitätsprognosen für den WED kann so bei der Erfassung ein schwankender Präzisionsgrad vorhanden sein. So ist es z. B. nicht ungewöhnlich, wenn Einbrüche von den Opfern verspätet zur Anzeige gebracht werden, sodass der exakte Tatzeitraum nur schwer eingegrenzt werden kann. Hierdurch offenbart sich aber ein grundsätzliches, methodisches Problem von Predictive Policing, wenn mit automatischen Datenanalysemethoden operiert wird. So besteht die grundsätzliche Annahme, dass das Delikt mit den vorliegenden Daten hinreichend genau in Raum und Zeit beschrieben ist. In der Praxis ist dies aber häufig nicht gegeben und sollte deshalb stets bei der Bewertung von angewandten Methoden, insbesondere im Hinblick auf ihre Qualität, beachtet werden.

3.11 Qualitätsmetriken

Polizeiliche wie mediale Diskussionen zu den unterschiedlichen Arten softwarebasierter Umsetzungsmöglichkeiten von Predictive Policing greifen immer wieder die Frage nach der Wirksamkeit auf, insbesondere im Zusammenhang mit sogenannten „Trefferraten“. Mit der Berechnung von „Trefferraten“ soll auf unterschiedliche Art und Weise geprüft werden, ob in einem Gebiet tatsächlich im Anschluss an eine Kriminalitätsprognose ein entsprechendes Delikt geschehen ist. Die Berechnung solcher Qualitätsmetriken (zur Begriffsbestimmung vgl. Stoffel et al. 2017) sind jedoch mit einer Vielzahl an methodischen Problemstellungen verbunden, sodass valide Aussagen zur Prognosequalität, insbesondere im Hinblick auf einen Vergleich von „Trefferraten“ unterschiedlicher Softwareumsetzungen, nicht möglich sind.

Ein grundsätzliches Problem bei der Berechnung von „Trefferraten“ im Rahmen von Predictive Policing ist die Tatsache, dass in Prognosegebieten Taten durch gezielte präventive Maßnahmen vermieden werden sollen. Den Erfolg von Predictive Policing an „Trefferraten“ zu messen ist daher generell mehr als fraglich.

Die Güte eines konkreten Predictive-Policing-Modells kann in diesem Kontext, wie in Kapitel 4.7 dargestellt, nur dann grundsätzlich abgeschätzt werden, wenn retrograde Kriminalitätsprognosen für einen sehr langen Zeitraum berechnet werden, in dem keine gezielten präventiven Maßnahmen im Zusammenhang mit Predictive Policing stattfanden. Hierfür werden die entsprechenden Qualitätsmetriken angewendet und für den (langen) Gesamtzeitraum statistisch gemittelt. Dies ist hinsichtlich einer internen Weiterentwicklung zur Verbesserung eines Prognosemodells von Bedeutung. Da die Qualitätsmetriken aber maßgeblich vom Prognosedelikt, der räumlichen und temporalen Bezugsgröße abhängen, sind Modellvergleiche unterschiedlicher Predictive-Policing-Verfahren mittels „Trefferraten“ nicht möglich, denn die Voraussetzungen identischer Analysesetups und Rahmenbedingungen zur Berechnung von Qualitätsmetriken sind nicht gegeben. Im Projekt SKALA wurden deshalb keine „Trefferraten“ berechnet. Dennoch sind bei Predictive-Policing-Verfahren Anwender wie Kritiker stets auf der Suche nach einem Qualitätsmaß, das solche Anforderungen objektiv, valide und zuverlässig erfüllt. Im Folgenden sollen deshalb die wesentlichen Probleme, die der Berechnung solcher Qualitätsmetriken stets inhärent sind, dokumentiert werden:²⁵

3.11.1 Statistisch seltene Ereignisse

Im Zusammenhang mit statistisch basierten Predictive-Policing-Verfahren werden Merkmale aus einem historischen Datensatz auf mögliche Korrelationen mit dem zu prognostizierenden Delikt, unter Beachtung der jeweiligen Stärke des Zusammenhangs, untersucht. Dies dient der Identifizierung von typischen Vorhersagevariablen. Liegen diese in einem bestimmten Raum vor, kann dort von einer erhöhten Kriminalitätswahrscheinlichkeit ausgegangen werden. Beispiele für solche Merkmale sind bei polizeilichen Vorgangsdaten die Tageszeit, der Wochentag oder der Modus Operandi. Kommt das zu prognostizierende Delikt jedoch selten vor, äußert sich dies in einer generell geringen Eintrittswahrscheinlichkeit: Es wird von „statistisch seltenen Ereignissen“ gesprochen.

Beim Vorliegen von statistisch bedeutsamen Vorhersagevariablen und damit der Erhöhung der Eintrittswahrscheinlichkeit eines Deliktes in einem bestimmten Gebiet, bleibt die Wahrscheinlichkeit bei grundsätzlich seltenen Ereignissen

dennoch gering. Sind beispielsweise im Schnitt ca. fünf Prozent aller Wohngebiete tatsächlich von Einbrüchen innerhalb einer Woche betroffen, könnte die vom Prognosemodell zu errechnende Wahrscheinlichkeit einzelner Wohngebiete 25 Prozent betragen. Das würde einer Verfünffachung des Grundrisikos entsprechen, wenngleich die Wahrscheinlichkeit, dass in diesem Gebiet kein Einbruch geschehen wird (zu 75 Prozent), überwiegt. So erklärt sich auch die Tatsache, dass in einem prognostizierten Gebiet im Gültigkeitszeitraum häufig keine entsprechende Tat eintritt. Werden Qualitätsmetriken angewandt, wird aber versucht genau diese, insgesamt unwahrscheinlichen Ereignisse zu erfassen. Darüber hinaus besteht die Besonderheit, dass die Polizei stets daran arbeitet, dass das prognostizierte Ereignis nicht eintritt. Dies muss bei Anwendung und Interpretation entsprechender Qualitätsmetriken ebenfalls stets berücksichtigt werden.

Diese zwei grundlegenden Besonderheiten sind bei nachfolgenden Qualitätsmetriken bei der Interpretation zu beachten. Darüber hinaus handelt es sich nur um eine Auswahl der gängigen, in der kriminologischen und polizeiwissenschaftlichen Literatur immer aufkommenden, Berechnungsmodelle im Rahmen von Predictive Policing. Die bereits angesprochene Kritik, die die Anwendung solcher Metriken bei Predictive-Policing-Verfahren mit sich bringt, erfolgt nach den zunächst objektiven Darstellungen der jeweiligen Qualitätsmetriken.

3.11.2 Die „Trefferrate“ (Hit Rate)

Die „Trefferrate“, auch als Hit Rate (HR) bekannt, ist die in der medialen Landschaft wohl am meisten aufgegriffene Qualitätsmetrik bei Predictive-Policing-Verfahren (vgl. beispielsweise Berk 2008; Chainey et al. 2008; Hunt et al. 2014; Mohler et al. 2015). Ihre Berechnung bezieht sich immer auf einen bestimmten Zeitraum (t), z. B. eine Woche, und lautet wie folgt:

$$HR[\%] = \frac{\text{Anzahl Delikte in den Prognosegebieten}}{\text{Anzahl aller Delikte}} \cdot 100$$

Beispiel-Sachverhalt: Haben sich in einem Gültigkeitszeitraum von einer Woche in einer Stadt 100 WED ereignet und lagen 15 WED in vorher als relevant berechneten Prognosegebieten, so ergibt sich mit der zuvor beschriebenen Berechnung eine „Trefferrate“ von 15 Prozent. Es wurden demnach 15 Prozent aller in der vorausgegangenen Woche eingetretenen WED „richtig“ vorhergesagt. Die Berechnungsformel wie auch das Ergebnis lassen sich einfach nachvollziehen.

Die Bezugsgröße, zuvor die Anzahl aller Delikte, kann aber auch angepasst werden, beispielsweise auf die Anzahl der

²⁵ In der Darstellung der verschiedenen Berechnungsmodelle wird sich erneut eng an die Formulierungen und den Aufbau des Artikels von Bode et al. (2017) bzw. Stoffel et al. (2017) angelehnt.

(als relevant berechneten) Prognosegebiete (HR_P). Die Berechnung lautet in dem Fall:

$$HR_P[\%] = \frac{\text{Anzahl getroffener Prognosegebiete}}{\text{Anzahl Prognosegebiete}} \cdot 100$$

Beispiel-Sachverhalt: Mit den zuvor beispielhaft angenommenen und eingetretenen 15 WED in den Prognosegebieten würde sich, wenn insgesamt 20 Prognosegebiete zuvor als relevant berechnet worden wären, eine „Trefferrate“ von 75 Prozent ergeben (konsequenterweise verbietet es sich, „Mehrfachtreffer“ zu zählen). Das würde bedeuten, dass in 75 Prozent aller Prognosegebiete tatsächlich mindestens ein WED stattgefunden hat.

3.11.3 Der Predictive Accuracy Index

Chainey et al. (2008) haben, um die „Trefferrate“ in Relation zur Größe der Prognosegebiete und der Gesamtfläche der Stadt setzen zu können, den Predictive Accuracy Index (PAI) entwickelt.²⁶ Der PAI berechnet sich, indem die zuvor berechnete „Trefferrate“ durch den Anteil der Prognosegebiete an der Gesamtfläche der Stadt geteilt wird, konkret:

$$PAI = \frac{\frac{n}{a} \cdot 100}{\frac{N}{A} \cdot 100}$$

- n: Anzahl Delikte in Prognosegebieten
- N: Anzahl aller Delikte
- a: Gesamtfläche der Prognosegebiete
- A: Gesamtfläche des betrachteten Gebiets

(vgl. Chainey et al. 2008: 14).

Beispiel-Sachverhalt: In der beispielhaften Annahme, dass die 20 Prognosegebiete insgesamt eine Fläche von 10 km² haben und die Gesamtfläche der Stadt 120 km² beträgt, berechnet sich der zuvor dokumentierte PAI wie folgt:

$$\frac{HR[\%]}{\text{Anteil der Prognosegebieten an der Gesamtfläche}} = \frac{15}{10/120 \cdot 100} = \frac{15}{8,3} = 1,81$$

Die Messung mithilfe des PAI ist im angloamerikanischen und angelsächsischen Raum sehr populär (vgl. z. B. Drawve 2014; Levine 2008, Mohler et al. 2015; Wang et al. 2012), wird aber von Van Patten et al. (2009) sowie von Hart und Zandbergen (2012) nur unter zusätzlicher Nutzung des Recapture Rate Index (RRI)²⁷ empfohlen.²⁸

²⁶ Der PAI wurde ursprünglich für Hotspot Mapping (Brennpunktkartierung) entwickelt. Brennpunktkarten basieren grundsätzlich auf Techniken der Kerndichteschätzung (vgl. Silverman 1986), bei denen Interpolationsverfahren zur Darstellung von Hotspots angewendet werden (vgl. Braga 2005; Chainey et al. 2008).

²⁷ Der RRI vergleicht die Anzahl der Delikte des aktuellen Gültigkeitszeitraumes mit dem vorherigen Gültigkeitszeitraum.

3.11.4 Der Standardized Accuracy Efficiency Index

Public Engines²⁹ hat darüber hinaus, um auch die Güte bzw. Effektivität der Prognosegebiete zu berücksichtigen, den Standardized Accuracy Efficiency Index (SAEI) gebildet (vgl. Public Engines 2014; Motorola Solutions 2015). Die Berechnung erfolgt aus der Achievable Efficiency, dem Anteil aller Delikte an allen möglichen Prognosegebieten, der Observed Efficiency, dem Anteil der Delikte in Prognosegebieten an allen möglichen Prognosegebieten und der Accuracy, die der „Trefferrate“ entspricht. Konkret werden die Metriken wie folgt berechnet:

$$SAEI = \frac{(\text{Achievable Efficiency} - (\text{Observed Efficiency} \cdot \text{Accuracy}))}{\text{Achievable Efficiency}}$$

$$\text{Achievable Efficiency} = \frac{\text{Anzahl aller Delikte}}{\text{Anzahl aller möglichen Prognosegebiete}}$$

$$\text{Observed Efficiency} = \frac{\text{Anzahl Delikte in den Prognosegebieten}}{\text{Anzahl aller möglichen Prognosegebiete}}$$

$$\text{Accuracy} = \frac{\text{Anzahl Delikte in den Prognosegebieten}}{\text{Anzahl aller Delikte}} \triangleq HR$$

(vgl. Public Engines 2014: 7)

Beispiel-Sachverhalt: Exemplarisch dargestellt am zuvor genutzten WED-Beispiel und annehmend, dass in einer Stadt 500 Prognosegebiete vorhanden sind, muss noch die Achievable und Observed Efficiency berechnet werden. Die Accuracy (in diesem Fall die „Trefferrate“) wurde bereits zuvor mit dem Wert 0,15/15 Prozent berechnet (siehe Kapitel 3.11.2). Bei den angenommenen 500 Prognoseflächen wird für die Achievable Efficiency demnach die Anzahl aller Delikte im Gültigkeitszeitraum durch die Anzahl aller möglichen Prognoseflächen geteilt, also: 100 / 500 = 0,2. Für die Observed Efficiency wird die Anzahl der Delikte in den Prognosegebieten im Gültigkeitszeitraum durch die Anzahl aller möglichen Prognoseflächen geteilt, konkret: 15 / 500 = 0,03. Das bedeutet, die Berechnung gestaltet sich folgt:

$$\frac{(0,2 - (0,03 \cdot 0,15))}{0,2} = \frac{0,1955}{0,2} = 0,9775$$

Trotz fehlender Angabe eines Wertebereichs für die Interpretation, kann der SAEI als Maßzahl gedeutet werden, die einen Mittelweg zwischen der maximal erreichbaren „Trefferrate“ (Accuracy) und der Effizienz (Achievable und Observed Efficiency) bilden soll. Höhere Werte beschreiben

²⁸ Hotspot Mapping unterscheidet sich methodisch und inhaltlich von Predictive Policing, weshalb auf Qualitätsmetriken für Hotspot Mapping (wie dem RRI) hier nicht weiter eingegangen wird.

²⁹ Public Engines ist ein ehemaliger, kommerzieller Anbieter für Datenanalysen und wurde im Jahr 2015 von Motorola Solutions übernommen.

danach Kriminalitätsprognosen, die genauer und effizienter sind. Der exakte Zusammenhang bleibt jedoch beim SAEI unklar, sodass seriöse Vergleiche zwischen SAEI-Werten verschiedener Prognosemodelle nicht möglich sind.

3.11.5 Die Konfusionsmatrix

Die Konfusionsmatrix ist ein klassisches Schema für die Berechnung von Qualitätsmetriken in den Sozialwissenschaften und kann auch in Predictive-Policing-Verfahren angewandt werden. Durch die Erstellung einer Vierfelder-Tabelle lassen sich Falsch Positive, Falsch Negative, Richtig Positive und Richtig Negative übersichtlich darstellen.

Beispiel-Sachverhalt: Mit der vorigen beispielhaften Annahme würde dies, wenn sich in 100 von den 500 möglichen Prognosegebieten ein WED ereignet hat, wie folgt aussehen:

Tabelle 1: Illustration Konfusionsmatrix

| | | Delikt aufgetreten | |
|----------|------|--------------------|---------------|
| | | Ja (Richtig) | Nein (Falsch) |
| Prognose | Pos. | 15 | 5 |
| | Neg. | 85 | 395 |
| Σ | | 100 | 400 |

Richtig Positive: 15 von 20. Es handelt sich um Gebiete, die als Prognosegebiete vorausgesagt wurden und in denen sich ein WED ereignet hat.

Falsch Positive: 5 von 20. Es handelt sich um Gebiete, die als Prognosegebiete vorausgesagt wurden und in denen sich kein WED ereignet hat.

Richtig Negative: 85 von 480. Es handelt sich um Gebiete, die nicht als Prognosegebiete vorausgesagt wurden und in denen sich ein WED ereignet hat.

Falsch Negative: 395 von 480. Es handelt sich um Gebiete, die nicht als Prognosegebiete vorausgesagt wurden und in denen sich kein WED ereignet hat. Mit der Bildung von entsprechenden Metriken, beispielsweise einer Falsch-Positiv-Rate, kann das jeweilige Modell sodann näher analysiert und bewertet werden.

3.11.6 Die Fallzahldifferenz

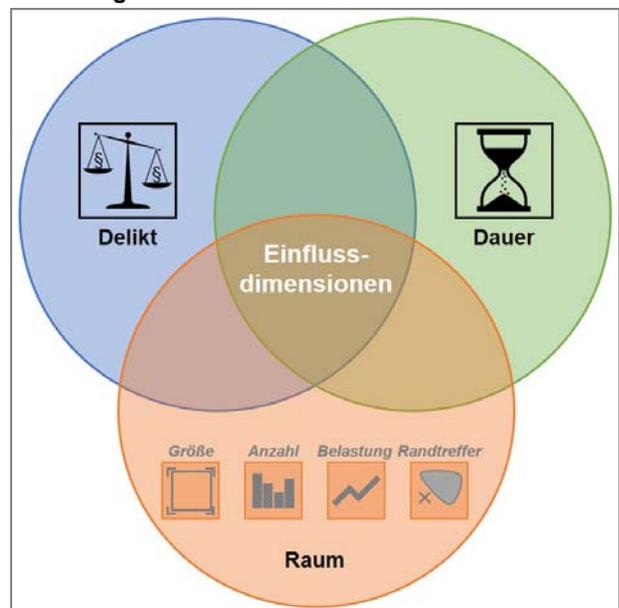
Es wird häufig diskutiert, vor allem im Rahmen von Evaluation, die Prognosegüte über die Fallzahldifferenz zu ermitteln, also durch einen Vergleich von Fallzahlen vor und nach der Implementierung von Predictive Policing. Die Berechnung erfolgt durch einfache Subtraktion. Mit der Fallzahldifferenz wird folglich nicht auf einzelne Kriminalitätsprognosen oder Prognosemodelle fokussiert, sondern auf Predictive Policing als Ganzes. Diese Art der methodischen Überprüfung ist aber nicht valide, denn Korrelation bedeutet nicht zwingend Kausalität. Kausalität lässt sich im Bereich von Predictive Policing nicht allein mit der Fallzahldifferenz feststellen, denn es bleibt stets offen, ob ein möglicher Fallzah-

lenrückgang auch tatsächlich auf die Umsetzung von Predictive Policing zurückzuführen ist. Kriminalität als soziales Phänomen ist immer multikausal bedingt. Darüber hinaus besteht mit dem Einsatz von Predictive Policing die Besonderheit, dass polizeiliches Einsatzverhalten auch unabhängig von der Erstellung einer Kriminalitätsprognose stattfindet. Konkret: Es wird z. B. auch in Reaktion auf Notrufe oder Notlagen agiert. In diesem Zusammenhang bleibt aber stets offen, ob ein gewisser Fallzahlenrückgang auf bewusstes Verhalten der Polizei auf Kriminalitätsprognosen zurückzuführen ist oder ob andere Ereignisse hierfür maßgeblich gewesen sind. Aussagen zur Wirksamkeit von Predictive Policing, ausschließlich reduziert auf einen Fallzahlenrückgang im Beobachtungszeitraum, sind folglich ohne inhaltliche Substanz.

3.12 Variabilität und Validität der Qualitätsmetriken

Die Einordnung und der Vergleich von „Trefferraten“ erscheinen zunächst leicht, insbesondere mit dem Wissen über die zuvor im Kapitel 3.11 dokumentierten Qualitätsmetriken. Nicht selten wird medial zudem von hohen „Trefferraten“ einzelner Predictive-Policing-Verfahren berichtet, wie beispielsweise von 80 prozentigen „Trefferraten“ der von der Polizei Zürich für Predictive Policing eingesetzten Software (vgl. Golem 2014; Focus 2015). Alle zuvor dokumentierten Qualitätsmetriken werden jedoch von drei wesentlichen Einflussdimensionen beeinflusst (siehe Abbildung 10).

Abbildung 10: Einflussdimensionen



Quelle: Bode et al. 2017: 9.

Der Einbezug dieser Dimensionen in die Qualitätsmetriken wird nachfolgend erläutert und zeigt auf, dass sich die jeweiligen Ergebnisse dadurch sehr verschieden berechnen las-

sen und sich damit eine Variabilität in den Metriken manifestiert. Dies hat in der Konsequenz wiederum Auswirkungen auf ihre Validität.³⁰

3.12.1 Einflussdimension: Prognose-Delikt

Umsetzungen von Predictive Policing auf statistischer Ebene erfordern immer eine sogenannte Zielvariable, die das Ziel der raum- und zeitbezogenen Vorhersage festlegt. Es bieten sich Deliktfelder an, die nicht statisch sind (sonst wären es Hotspot-Analysen) oder der Kontrollkriminalität zuzurechnen sind (z. B. Betäubungsmittelkriminalität). Im Projekt SKALA war die Zielvariable für WED der Wohnungseinbruch als Ereignis, sowie im weiteren Projektverlauf der Einbruchdiebstahl aus Gewerbeobjekten oder Kraftfahrzeugen. Das Ereignis der jeweiligen Straftat wurde stets als dichotome Zielvariable behandelt.

Allgemein ist bei der Auswahl und Festlegung auf ein bestimmtes Prognose-Delikt keine Variabilität in den Ergebnissen bei der Berechnung der Qualitätsmetriken zu erwarten. Es kommt jedoch zu einer systematischen Verzerrung, wenn Delikte, die nicht prognostiziert wurden, anschließend als volle oder halbe Treffer gewertet werden, wie z. B. ein Keller-Einbruch bei einer WED-Prognose. Es darf aber nur zwingend das Delikt in das Qualitätsmaß einbezogen werden, das zuvor auch prognostiziert wurde. Ansonsten werden Maßzahlen verzerrt und stellen fehlerhafte und in keinem Fall verwertbare Zahlen dar. Bestrebungen für die Definition einheitlicher Validierungsindizes sind aber derzeit noch keine gängige Praxis im Bereich von Predictive Policing, wie sich auf zahlreichen Diskussions- und Vernetzungsworkshops im In- und Ausland während der Projektlaufzeit herausstellte. Diese inhärente Subjektivität bei der definitorischen Festlegung auf eine bestimmte Auslegung lässt damit keinen Vergleich zu anderen Umsetzungen von Predictive Policing zu. Die Ergebnisse sind aufgrund unterschiedlicher Berechnungsweisen variabel und somit nicht valide.

3.12.2 Einflussdimension: Prognose-Dauer

Kriminalitätsprognosen werden immer für einen bestimmten Zeitraum generiert (Gültigkeitszeitraum). Dieser Zeitraum richtet sich nach dem Deliktfeld sowie dem polizeilichen Bedarf bzw. Nutzen und kann sehr unterschiedlich sein. Denkbar sind beispielsweise Tages- oder Wochenprognosen.

Das Ergebnis der angewandten Qualitätsmetrik folgt hierbei einem einfachen Grundprinzip: Tendenziell wird das Ergebnis bestimmter Qualitätsmaße schlechter, je kürzer der Gültigkeitszeitraum ist. Denn je länger der Zeitraum, auf den sich die Kriminalitätsprognose bezieht (Gültigkeit), desto

größer ist die Wahrscheinlichkeit, dass entsprechende Delikte in dem vorhergesagten Prognosegebiet auftreten können.

Abbildung 11 zeigt beispielhaft für einen Stadtteil von Köln eine WED-Tagesprognose, eine WED-Wochenprognose und eine WED-Zweiwochen-Prognose. Erwartungsgemäß ist die „Trefferrate“ für den Gültigkeitszeitraum eines Tages (siehe Abbildung 11 a) am schlechtesten. Die Wahrscheinlichkeit, dass ein Delikt in den Prognosegebieten auftritt, steigt mit der Anzahl der Delikte über die Zeit, sodass sie in der Abbildung 11 bei einer Zweiwochen-Prognose auf 50 Prozent steigt (siehe Abbildung 11 c). Es ist demnach wichtig zu wissen, wenn „Trefferraten“ zu Predictive Policing angegeben werden, für welchen Gültigkeitszeitraum diese erstellt wurden. Auch hier gilt, dass immer derselbe Gültigkeitszeitraum bei entsprechenden Berechnungen gewählt werden muss, da sonst keine interpretierbaren Maßzahlen vorliegen.

³⁰ Die jeweiligen Einflussdimensionen werden einzeln erläutert, wirken aber stets gemeinsam auf die zu berechnende Qualitätsmetrik.

Abbildung 11: Illustration der Auswirkung variabler Gültigkeitszeiträume auf die „Trefferrate“
Prognosegebiete sind blau dargestellt, Gebiete mit Treffern rot, WED gelb.



a) Gültigkeitszeitraum: Ein Tag.
„Trefferrate“: 0 %

b) Gültigkeitszeitraum: Eine Woche.
„Trefferrate“: 33,33 %

c) Gültigkeitszeitraum: Zwei Wochen.
„Trefferrate“: 50,00 %

Quelle: Bode et al. 2017: 10.

3.12.3 Einflussdimension: Prognose-Raum

Im Zusammenhang mit der Berechnung von Qualitätsmetriken ist die räumliche Bezugsgröße die komplexeste, da sie sich in vier sogenannte Sub-Dimensionen aufteilt (siehe Abbildung 10: Einflussdimensionen):

1. Größe der Prognosegebiete

Mit zunehmender Größe der Prognosegebiete steigt die „Trefferrate“ oder der Indexwert, da mit wachsender Größe der Prognosefläche die Chance, dass im prognostizierten Gebiet das erwartete Ereignis eintritt, z. B. ein WED, erhöht wird. Das bedeutet zudem im Umkehrschluss: Je kleiner das Prognosegebiet, desto unwahrscheinlicher wird es, dass das erwartete Ereignis im prognostizierten Gebiet eintritt.

2. Anzahl der Prognosegebiete

Die Anzahl der Prognosegebiete beeinflusst auch die zuvor dargestellten Qualitätsmetriken in ihrer Variabilität und Validität. So gilt grundsätzlich: Je mehr Prognosegebiete als relevant berechnet werden, desto größer ist die Gesamtfläche der Prognosegebiete und folglich umso höher ist die „Trefferrate“. Denn auch hier steigt die Trefferwahrscheinlichkeit, wenn mehr Prognosegebiete und damit potenzielle „Trefferflächen“ vorhanden sind. Im Hinblick auf eine praktikable Umsetzung muss aus den zuvor berechneten Wahrscheinlichkeiten immer eine bestimmte Auswahl getroffen werden, die den operativen Polizeikräften zur Verfügung gestellt wird. Dies ist aber mit methodischen Schwierigkeiten verbunden: Zum einen kann die Auswahl anhand einer festgelegten Anzahl von Prognosegebieten erfolgen oder aber anhand eines Schwellenwertes. Die Festlegung auf eine feste Anzahl birgt das Risiko, dass in bestimmten Gebieten keine polizeilichen Interventionen stattfinden, obwohl diese unter

Umständen jedoch die gleichen oder sehr ähnliche Wahrscheinlichkeiten wie die Prognosegebiete aufweisen. Mit der Festlegung auf einen immer wieder neu zu bestimmenden Schwellenwert wird dieses Problem der Festlegung umgangen, was aber zur Folge hat, dass von Prognose zu Prognose eine unterschiedliche Anzahl an Prognosegebieten generiert wird und eine vergleichende Anwendung der genannten Qualitätsmetriken nur schwer möglich macht.

3. Belastung der Prognosegebiete

Einfluss auf die „Trefferrate“ hat auch die kriminelle Grundbelastung bestimmter Orte einer Stadt, sogenannte Kriminalitätsbrennpunkte (Hotspots). Werden Kriminalitätsprognosen auf solche Gebiete einer Stadt gelegt, ist die Wahrscheinlichkeit, dass sich dort weitere Delikte der gleichen Art ereignen, erwartungsgemäß sehr hoch. Die Folge ist, dass die „Trefferrate“ steigt. Solche Kriminalitätsprognosen sind allerdings von geringer Aussagekraft und der Erkenntniswert ist niedrig, denn sie bilden das ab, was polizeilich bereits bekannt ist. Streng genommen handelt es sich in diesen Fällen methodisch um Hotspot-Analysen und nicht um Predictive Policing.

4. Randtreffer an Prognosegebieten

Randtreffer sind sogenannte „Beinahe-Treffer“. Sie können einen erheblichen Einfluss auf die Qualitätsmetrik haben, wenn sie in die Bewertung des Berechnungsmodells einfließen. Abbildung 12 zeigt beispielhaft, wie bei einer WED-Prognose Randtreffer mit quartierbasierten Gebieten aussehen können.

Abbildung 12: Illustration Randtreffer

Schraffiert ist das Prognosegebiet, gelbe Kreuze markieren WED.



Quelle: Bode et al. 2017: 11.

So liegen die drei WED direkt neben dem Prognosegebiet. Mit dem Wissen, dass Straftäter sich nicht an die künstlichen Grenzen eines Prognosegebietes halten, liegt es nahe, dass der ein oder andere Analyst oder Softwarehersteller solche Randtreffer ebenfalls in die Berechnung der „Trefferrate“ einbeziehen möchte (vgl. Public Engines 2014: 7). Hierdurch wird eine Vergleichbarkeit zu anderen Modellen unmöglich, da regelmäßig nicht bekannt ist, ob und welche Delikte als Randtreffer gewertet werden. Auch die Frage, wie Randtreffer gewertet werden (viertel, halb oder ganz) und was noch als Randtreffer gilt (Entfernung zum Prognosegebiet), erschwert in dieser Hinsicht die Vergleichbarkeit von unterschiedlich berechneten Qualitätsmetriken.

3.12.4 Bewertung

Zusammenfassend wird festgestellt, dass jegliche Form von Variabilität in der Berechnung der Qualitätsmetriken die Vergleichbarkeit von Prognosemodellen bei Predictive Policing in Frage stellt. Dadurch ist auch eine sichere Bewertung der Prognosegüte, anhand der zuvor dokumentierten Qualitätsmetriken, nicht ohne weiteres möglich. Weiterhin besteht bei der Berechnung von Qualitätsmetriken im Bereich von Predictive-Policing-Verfahren ein häufig anzutreffendes Paradoxon: Es wird versucht etwas zu messen, was eigentlich verhindert werden soll (und durch verstärkten Polizeieinsatz wahrscheinlich auch tatsächlich in den einigen Gebieten verhindert wird).

Aus den vorgenannten Gründen wurden im Projekt SKALA keine „Trefferraten“ für Kriminalitätsprognosen berechnet. Grundsätzlich ist es jedoch sinnvoll, für die interne Modellweiterentwicklung und -verbesserung gewisse Qualitätsindizes zur Abschätzung der Prognosegüte zu ermitteln. Ziel ist

es in dem Fall, die Indizes nicht mit anderen Softwareprodukten zu vergleichen, sondern mit unterschiedlichen Prognose setups desselben (eigenen) Modells. Dazu kann intern die Festlegung einheitlicher und vergleichbarer Kriterien aller zuvor beschriebenen Punkte, die bei der Berechnung der Qualitätsindizes zu beachten sind, sichergestellt werden. Ein adäquates Vorgehen solch einer internen Modellgüteabschätzung ist in Kapitel 4.7 dargestellt.

4 Methodische Umsetzung der Untersuchung

In diesem Abschnitt wird die methodische und technische Umsetzung der Generierung und Visualisierung der Kriminalitätsprognosen im Projekt SKALA beschrieben. Beginnend mit einer kurzen wiederholenden Erläuterung, warum ein hypothesengestütztes Vorgehen bei der methodischen Umsetzung wichtig war (Kapitel 4.1), widmet sich dieses Kapitel der Dokumentation getesteter und verwendeter Analyse- und Prognoseverfahren (Kapitel 4.2). Hierfür wird die technische Umsetzung umfassend erläutert (Kapitel 4.3), einschließlich der Testung von verschiedenen Prognosemodellen (Kapitel 4.4) und Variableneinflüssen (Kapitel 4.5). In diesem Kontext wird auch beschrieben, wie Raum-Zeit-Clusteranalysen und Scoreberechnungen durchgeführt (Kapitel 4.6) und evaluiert wurden (Kapitel 4.7). Ferner wird betrachtet, inwiefern auch die Prognose von anderen Delikten möglich ist (Kapitel 4.8). Der Abschnitt schließt mit einer Darstellung der Austausch- und Kooperationserfahrungen im Projektzeitraum (Kapitel 4.9).

4.1 Einordnung des methodischen Vorgehens

SKALA basiert auf einem hypothesengestützten Vorgehen und greift auf kriminologische und sozialwissenschaftliche Theorien zurück (siehe Kapitel 3.2). Ein solches Vorgehen ermöglicht, komplexe Wirkzusammenhänge zu erkennen, und geht über die Möglichkeiten der bloßen Anwendung des Near-Repeat-Ansatzes hinaus. Für eine umfassende Prüfung der grundsätzlichen Methodik von Predictive-Policing-Verfahren war dieses methodische Vorgehen unerlässlich. Die wissenschaftlichen Annahmen determinieren hierbei die weiteren Entscheidungsschritte, ohne einer zielgerichteten praktischen Umsetzung zu widersprechen. Darüber hinaus verleiht diese Art der methodischen Umsetzung die erforderliche Struktur (vgl. Pollich/Bode 2017: 2). Ein hypothesengestütztes Vorgehen kann zudem, neben der Möglichkeit verbesserter polizeilicher Interventionen, zu einem Erkenntnisgewinn für die Polizei über betreffende Gebiete und Delikte führen, da zielgerichtet Ursachen eines bestimmten Kriminalitätsfeldes in seinem raum-zeitlichen Kontext untersucht werden können.

Im Fokus der (technisch) methodischen Umsetzung von SKALA standen Aussagen über raum- und zeitbezogene Wahrscheinlichkeiten zu Kriminalitätsrisiken. Durch die gewählte statistische Betrachtungsweise war sichergestellt, dass Kriminalitätsprognosen für das gesamte Stadtgebiet, und nicht nur für Teilbereiche, erstellt wurden. Die Modellerstellung wurde mit dem IBM SPSS Modeler umgesetzt. Mit der komplexen Data-Mining-Software war es möglich, verschiedene Datenquellen miteinander in Beziehung zu setzen, zu analysieren und darauf aufbauend Prognosemodelle zu erstellen. Zur kartografischen Aufbereitung der erstellten Prognosen wurden ArcGIS-Produkte von ESRI sowie ein in Zusammenarbeit mit dem Lehrstuhl für Datenanalyse und Visualisierung der Universität Konstanz entwickeltes Visualisierungs- und Auswertetool (SKALA | MAP) eingesetzt.

4.2 Darstellung und Interpretation von Prognosen

Kriminalitätsprognosen werden i. d. R. visuell auf geografischen, meist digitalen Karten visualisiert. Eine solche Visualisierung ist jedoch immer eine Abstraktion und vereinfachte Darstellung der Realität und hängt von der gewählten Darstellungsweise des Kartographen ab. Dadurch kann die Interpretation desselben Inhalts bei der Betrachtung einer Karte in Abhängigkeit von ihrer Darstellungsweise unterschiedlich ausfallen (vgl. Belina 2009: 192). Durch eine geografische Karte wird somit eine spezifische sozialräumliche Wirklichkeit vermittelt (vgl. Frers et al. 2013: 168). Da schon die Produktion einer Karte durch die Wahrnehmung und Bewertung des Kartographen beeinflusst wird und im weiteren Verlauf auch die Wahrnehmung der betrachtenden Personen beeinflusst, muss die Art der Darstellung fortlaufend überprüft und kritisch reflektiert werden. Insbesondere gilt es, die Erwartungshaltung und Interpretationsweisen der jeweiligen am Predictive-Policing-Prozess beteiligten Zielgruppen zu erheben und Wandlungsprozesse aufzugreifen.

Innerhalb des Projektes SKALA wurde die visuelle Darstellung der Kriminalitätsprognosen stets an die Bedürfnisse der jeweiligen Zielgruppe, insbesondere die Wünsche der operativen Polizeikräfte, angepasst. Das TP WBA hat parallel und stetig die Visualisierung wie auch die Anpassungen kritisch reflektiert. Es zeigte sich beispielsweise, dass die farbliche Darstellung von Prognosegebieten erheblichen Einfluss auf die Bewertung durch die jeweiligen Fachdienststellen der Pilotbehörden hatte. So erfolgte anfangs eine Unterscheidung von WED-Prognosegebieten anhand unterschiedlicher Wahrscheinlichkeiten oder anhand verschiedener Modelle zur Prognoseerstellung, indem das jeweilige Gebiet gelb oder rot eingefärbt wurde. In diesem Zusammenhang fiel auf, dass insbesondere die roten Gebiete häufig als die wichtigen Gebiete bewertet und dementspre-

chend eher an die operativen Polizeikräfte übermittelt wurden als die gelben. Im weiteren Verlauf wurden sodann unterschiedliche Modelle der Prognoseerstellung mit unterschiedlichen Ziffern (Modell 1, Modell 2) dargestellt, was aber auch dazu führte, dass manche die Ziffern als Entscheidungskriterium nutzten, da hierin eine vermeintliche Rangfolge erkannt wurde. Eine solche Rangfolge innerhalb der Modelle gab es allerdings nicht. Im weiteren Projektverlauf wurde deshalb auf Modellbezeichnungen mit Ziffern verzichtet.

Neben dieser visuell-inhaltlichen Bewertung von Farben und Ziffern ergab sich zudem eine Herausforderung in der Darstellung von Wohnquartieren, in denen sich Grünflächen oder Flüsse befanden. Rückmeldungen aus den Pilotbehörden zeigten, dass es zu Irritationen führte, wenn Grünflächen oder Flüsse im WED-Prognosegebiet enthalten waren, sich dort aber tatsächlich kein Einbruch hätte ereignen können. Dieser Umstand führte zu individuellen Anpassungen der Wohnquartiere, indem Flüsse oder Grünflächen bei der grafischen Aufbereitung der Prognosegebiete bewusst „ausgeschnitten“ wurden.

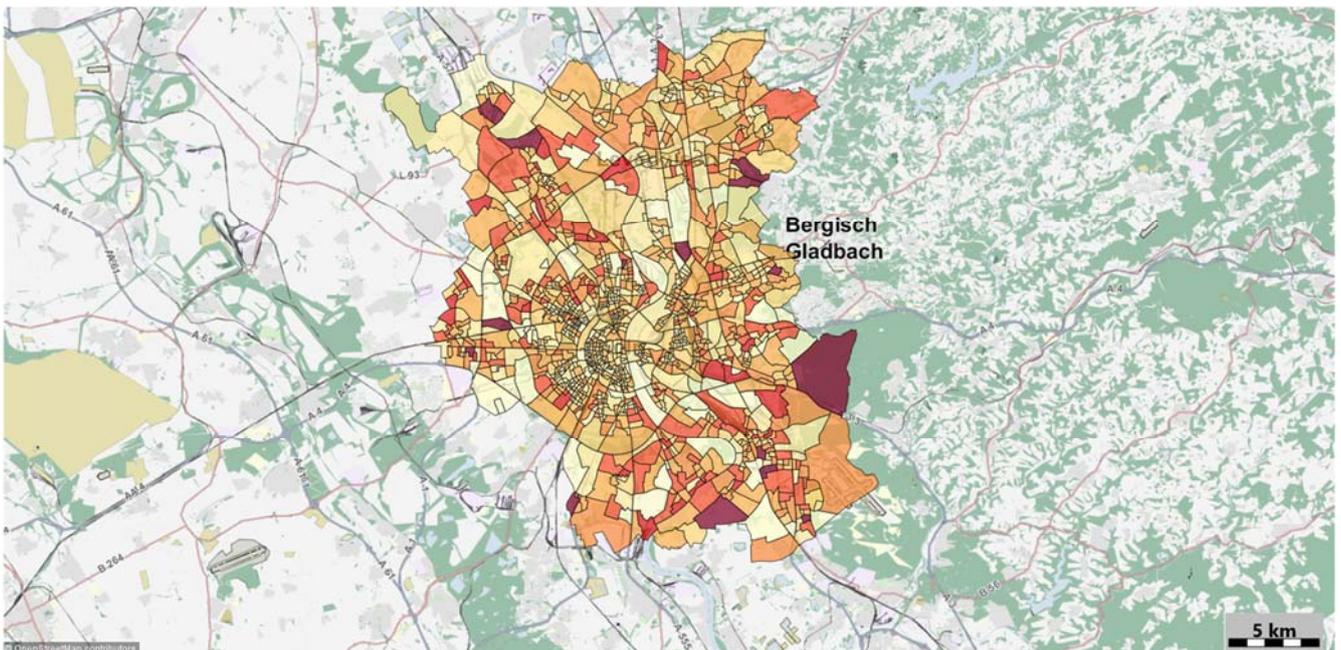
Bezogen auf die Anzahl der zu übermittelnden Prognosegebiete an die Pilotbehörden wurden zunächst nur die mit einer hohen Wahrscheinlichkeit belegten Gebiete übermittelt. Um den Pilotbehörden eine handhabbare Anzahl an Prognosegebieten zur Verfügung zu stellen, wurden jeder Behörde etwa 1,5 Prozent aus der Gesamtquartieranzahl aller möglichen Wohnquartiere übermittelt (siehe Tabelle 2). Die Rangfolge ergab sich dabei aus der zuvor erfolgten Priorisierung nach den größten Wahrscheinlichkeitswerten.

Tabelle 2: Quartierauswahl

| Behörde | Wohnquartiere | Davon übermittelt (ca. 1,5 %) |
|---------------|---------------|-------------------------------|
| Köln | 1360 | 20 |
| Duisburg | 481 | 8 |
| Düsseldorf | 593 | 9 |
| Essen | 777 | 12 |
| Gelsenkirchen | 277 | 5 |
| Bonn | 503 | 8 |

Dies hatte zur Folge, dass die Pilotbehörden mit unterschiedlich vielen Prognosegebieten beliefert wurden. Da eine solche Einschränkung auf nur wenige Wohnquartiere die Vergleichbarkeit der einzelnen Gebiete untereinander bzw. die Sicht auf die gesamte Lage im Zuständigkeitsbereich erheblich erschwert und eine Bewertung der Kriminalitätsprognosen aufgrund fehlender Informationen zum Umfeld nur eingeschränkt möglich war, wurde ergänzend zu dieser Auswahl eine separate Datei mit allen Wahrscheinlichkeitswerten der Wohnquartiere für den WED übermittelt. Das im Projekt SKALA entwickelte Visualisierungs- und Auswertetool SKALA | MAP ermöglicht, neben weiteren Funktionen, die Visualisierung solcher WED-Wahrscheinlichkeiten und stand allen Pilotbehörden ab April 2017 zur Verfügung (siehe Kapitel 4.3.4). Abbildung 13 zeigt eine Beispielparte für WED-Wahrscheinlichkeiten des Polizeibezirks Köln. Hierbei gilt: Je dunkler die Einfärbung, desto höher die Wahrscheinlichkeit eines WED. Die Darstellung von allen Wahrscheinlichkeitswerten der Wohnquartiere mit SKALA | MAP stellt eine Besonderheit im Projekt dar und hebt sich von anderen Predictive-Policing-Verfahren ab.

Abbildung 13: Beispielparte WED-Wahrscheinlichkeiten des Polizeibezirks Köln³¹



³¹ Kartenmaterial aller Abbildungen im Bericht © OpenStreetMap-Mitwirkende, Lizenziert unter CC BY-SA: www.openstreetmap.org/copyright.

4.3 Modellerstellung

Zur Modell- und Prognoseerstellung wurde der IBM SPSS Modeler genutzt (siehe 3.4). Mit dieser Anwendung war es möglich, die externen soziostrukturellen Daten mit den polizeilichen Vorgangsdaten in Beziehung zu setzen, zu analysieren und darauf aufbauend Prognosemodelle zu erstellen. Zentral war in diesem Kontext die raum- und zeitbezogene Zusammenführung und Analyse. Die (technisch) methodische Umsetzung der Modell- und Prognoseerstellung, insbesondere mit Bezug zum WED, wird nachfolgend dokumentiert.

Die polizeilichen Vorgangsdaten wurden, i. d. R. wöchentlich (montags), über eine automatisierte Abfragemaske aus dem polizeilichen Vorgangsbearbeitungssystem IGVP („Integrationsverfahren Polizei“) der Polizei NRW bezogen. Für diese Abfragen wurde der Oracle Discoverer genutzt (zur Auswahl der Daten siehe Tabelle 3). Die Daten unterliegen noch bis zur Abverfügung an die Staatsanwaltschaft einem gewissen Wandel. Das bedeutet, Datensätze können sich im Laufe der Zeit wieder ändern, z. B. kann das Delikt spezifischer eingegrenzt, die Beutesumme ergänzt oder der Modus Operandi korrigiert werden. Für das Delikt WED werden beispielsweise häufig detaillierte Beute- und

Schadensauflistungen erst nach geraumer Zeit nachgemeldet und im Vorgang ergänzt. Bei der Datengrundlage lagen folglich, im Hinblick auf die konkreten, tatsächlichen Tatumstände, stets gewisse Unsicherheiten vor. Neben der wöchentlichen Aktualisierung des Datenbestandes (Erhebung eines Delta-Datensatzes) wurde deshalb auch in regelmäßigen Abständen (je nach Delikt monats- oder quartalsweise) ein neuer Gesamtdatensatz aus IGVP bezogen, um eine stets gute Datenqualität für die Modell- und Prognoseerstellung zu gewährleisten. Neben den Nachteilen in Bezug auf die Datenqualität, lag der Vorteil einer wöchentlichen Aktualisierung und Ergänzung des Gesamtdatensatzes aber darin, Fälle unmittelbar nach der Erfassung zur Analyse nutzen und so angemessen schnell auf die aktuelle Kriminalitätslage reagieren zu können.

Tabelle 3: Polizeiliche Vorgangsdaten aus IGVP

| Merkmale | Erläuterung |
|-----------------------------------|---|
| Vorgangsprimärschlüssel | Eindeutiger Schlüssel des Vorgangs |
| Erfassungsdatum | Datum der Vorgangserfassung |
| Delikt 1 Serial | Deliktschlüssel des ersterfassten Deliktes |
| Delikt 1 Klartext | Klartext des ersterfassten Deliktes |
| Delikt 2 Serial | Deliktschlüssel des zweiterfassten Deliktes |
| Delikt 2 Klartext | Klartext des zweiterfassten Deliktes |
| Delikt 3 Serial | Deliktschlüssel des dritterfassten Deliktes |
| Delikt 3 Klartext | Klartext des dritterfassten Deliktes |
| X Koordinate | Georeferenz (X) des Ereignisortes |
| Y Koordinate | Georeferenz (Y) des Ereignisortes |
| Tatzeitdatum Von | Datum vom Beginn der Tatzeit |
| Tatzeituhrzeit Von | Uhrzeit vom Beginn der Tatzeit |
| Tatzeitdatum Bis | Datum vom Ende der Tatzeit |
| Tatzeituhrzeit Bis | Uhrzeit vom Ende der Tatzeit |
| Tatort Gemeinde Klartext | Klartextbezeichnung der Tatortgemeinde |
| Tatort Strasse1 | Name der ersten Straße mit Tatortbezug |
| Tatort Hausnummer1 | Hausnummer zur ersten Straße mit Tatortbezug |
| Begehungsweise 1 | Katalogwert der ersterfassten Tatbegehungsweise |
| Begehungsweise 2 | Katalogwert der zweiterfassten Tatbegehungsweise |
| Begehungsweise 3 | Katalogwert der dritterfassten Tatbegehungsweise |
| Gesamtschaden | Gesamtwert des entwendeten Gutes zzgl. verursachten Sachschäden |
| Versuch | Dichotome Information zum Tatstadium (Versuch/Vollendung) |
| Geo Qualitätsmarker | Index für die Genauigkeit der Vorgangs-Georeferenzierung |
| Lageschlagwort 1 | Katalogwert zur Kennzeichnung temporärer Phänomene und Serien |
| Lageschlagwort 2 | Katalogwert zur Kennzeichnung temporärer Phänomene und Serien |
| Sache Art | Oberbegriff (Katalogwert) der angegangenen Sache |
| Sache Ggstandsart Klartext | Bezeichnung der Gegenstandsart |
| Sache Erlangtes Gut | Dichotome Information, ob Gut erlangt worden ist |
| Schlagwort Klart Liste | Liste aller gewählten Schlagwörter |
| Schlagwort Lokal Liste | Liste aller gewählten behördenspezifischen Schlagwörter |
| Tatörtlichkeit 1 Serial | Schlüssel der Tatörtlichkeit |
| Tatörtlichkeit 1 Klar | Katalogwert der Tatörtlichkeit |
| Sache Wert SUM | Summe der Sachwerte eines Vorgangs |

Die Abfrage- und Transferrountinen waren im Rahmen des Projektes zeitlich überschaubar. Durch das Arbeiten in einem Sondernetz mussten die Daten aus dem Polizeinetz auf einen externen Datenträger (USB-Stick) transferiert werden, um sie anschließend wieder in das Sondernetz einzuspeisen. Dies gestaltete sich, insbesondere bei einer Aktualisierung des Gesamtdatenbestandes, als sehr zeitaufwendig. Bei einer etwaigen Einführung von SKALA in NRW sind an dieser Stelle Optimierungen zwingend erforderlich, vor allem im Hinblick auf die Automatisierung der entsprechenden Prozesse und die direkte Anbindung an das Polizeinetz.

In Bezug auf die Modellerstellung ermöglichte die Nutzung des IBM SPSS Modeler die komplexe, aber transparente Aufbereitung der Datensätze wie auch die anschließende Modell- und Prognoseberechnung. Mithilfe einer anwenderfreundlichen Benutzeroberfläche ermöglichte die Software Data-Mining-Schritte auf verhältnismäßig einfache Weise zu realisieren. Muster und Trends ließen sich übersichtlich in den eigenen Daten aufdecken. Somit waren ohne umfangreiche Programmierkenntnisse unterschiedlichste Datensatzoperationen sowie Modell- und Prognoseerstellung möglich.³² Die erstellten Modelle und Kriminalitätsprognosen waren in der Folge transparent und Fehlerquellen ließen sich immer schnell detektieren und beseitigen. Die meisten Prozessschritte wurden daher im Rahmen des Projektes SKALA im IBM SPSS Modeler erarbeitet und umgesetzt.

Grundsätzlich wird im IBM SPSS Modeler mit sogenannten Knoten operiert. Jeder dieser Knoten ermöglicht eine spezifische Operation (z. B. die Umbenennung von Feldern oder das Löschen von Duplikaten). Durch die Verbindung der Knoten wird der eingelesene Datensatz weitergereicht und Schritt für Schritt aufbereitet. Die individuelle und auf die eigenen Bedürfnisse zugeschnittene Zusammenstellung der Knoten wird als Stream bezeichnet. Der erstellte SKALA-Stream zur Generierung von WED-Prognosen gliederte sich in vier Schritte:

1. **Datenaufbereitung**
2. **Georeferenzierung der Daten**
3. **Modell- und Prognoseerstellung**
4. **Visualisierung**

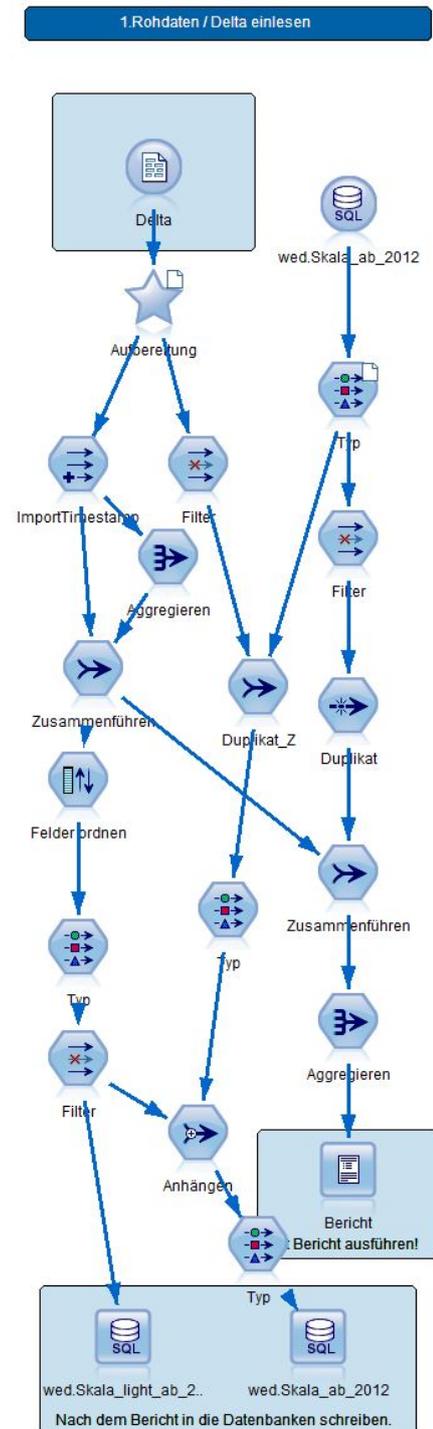
Diese Schritte werden nachfolgend, mit beispielhaften Grafiken aus dem IBM SPSS Modeler erläutert.

4.3.1 1. Schritt: Datenaufbereitung

Der erste Teil dieses Streams fokussierte sich auf die Datenaufbereitung. Diese stellte eine große Herausforderung

dar, da die polizeilichen Vorgangsdaten stets (mit jeder Aktualisierung des Gesamtdatensatzes wie auch mit der Einspeisung eines Delta-Datensatzes) in ein einheitliches Format gebracht werden mussten. Eine beispielhafte Illustration für den Aufbau des Streams zur Datenaufbereitung ist Abbildung 14 zu entnehmen.

Abbildung 14: Beispiel-Stream: Datenaufbereitung



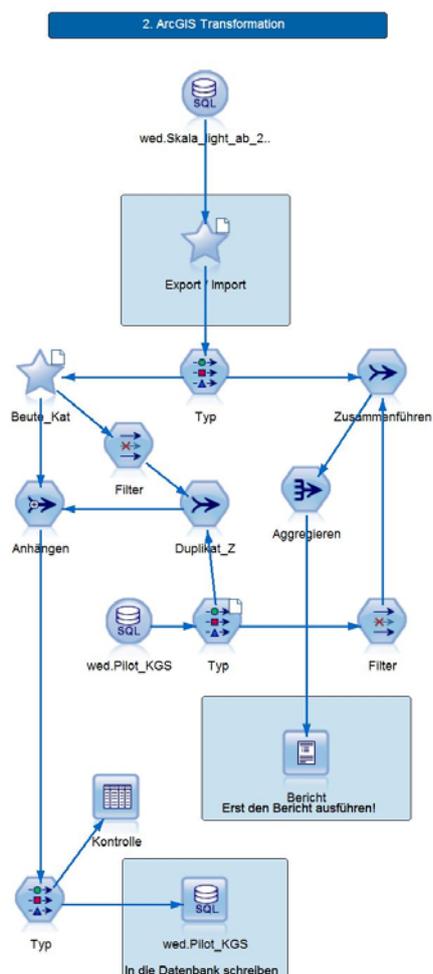
³² Während der Projektlaufzeit zeigte sich allerdings auch immer wieder, dass bestimmte Datenverarbeitungsprozesse, z. B. die Erstellung eines Kreuzproduktes, aus Performancegründen vom IBM SPSS Modeler abgetrennt werden sollten. Hier boten eigens programmierte Skripte schnellere Umsetzungsmöglichkeiten.

Im ersten Schritt der Datenaufbereitung stand zunächst der Import der zu integrierenden Vorgangsdaten für den Abfragezeitraum. Dieser Schritt musste bei jedem Prognosezyklus wie auch bei einer Aktualisierung des Gesamtdatensatzes durchgeführt werden. Anschließend erfolgte in mehreren Prozessschritten die Aufbereitung des Datensatzes. Dies umfasste im Wesentlichen die Festsetzung der richtigen Merkmalsbezeichnungen, der Speichertypen sowie der Messniveaus (Skalenniveaus). Darüber hinaus wurden leere Werte (*missings*) ersetzt oder einheitlich definiert und offensichtliche Datenfehler sowie Testvorgänge/Duplikate gelöscht. Im Zuge der Datenaufbereitung mussten unterschiedliche Datensätze zusammengeführt, aggregiert und gefiltert werden. Ziel war es, einen einheitlich nutzbaren Datensatz für die weitere Verarbeitung zu generieren.

4.3.2 2. Schritt: Georeferenzierung der Daten

Der zweite Teil des Streams befasst sich insbesondere mit der Georeferenzierung der Daten (siehe Abbildung 15).

Abbildung 15: Beispiel-Stream: Georeferenzierung



Die Georeferenzierung stellt im Predictive-Policing-Prozess grundsätzlich das zentrale Element dar, um das Kriminalitätsgeschehen adäquat im Raum abbilden zu können und darauf aufbauend entsprechende Modelle und Prognosen zu erstellen. Hierzu werden die einzelnen Datensätze geografisch zusammengeführt. Die Georeferenzierung im Projekt SKALA bestand im Wesentlichen aus der räumlichen Zusammenführung der Datensätze durch die mitgelieferten Koordinaten von polizeilichen Vorgangsdaten und den externen soziostrukturellen Daten. Das bedeutet, die Vorgangsdaten für die Modell- und Prognoseerstellung mussten spezifischen Gebäuden, Straßenabschnitten und Wohnquartierpolygonen zugeordnet werden. Im Projekt SKALA und mit Bezug zum WED wurden alle Vorgänge seit dem 01.01.2012 georeferenziert in das für die spätere Verwendung erforderliche Koordinatensystem ETRS-1989³³ transformiert und auf die im dritten Schritt erfolgende Zusammenführung mit den soziostrukturellen Daten vorbereitet (konkrete Zuordnung von WED zu Wohnquartieren und deren soziostruktureller Merkmale). Für diesen Zweck wurde auf das Geoinformationssystem ArcGIS zurückgegriffen. Im SKALA-Stream mussten deswegen die Vorgänge zunächst aus dem Stream exportiert werden, um sie anschließend in ArcGIS einlesen zu können. Nach der erfolgten Georeferenzierung in ArcGIS wurden die überarbeiteten Vorgänge wieder zurück in den SKALA-Stream importiert.

Die konkrete räumliche Zuordnung der Vorgänge und der soziostrukturellen Daten, die auf unterschiedlichen Aggregationsniveaus (Gemeindeebene, Wohnquartierebene, Straßenabschnittebene und Gebäudeebene) vorlagen (siehe Kapitel 4.5), gestaltete sich dabei wie folgt: Über die georäumliche Lage der WED zu den ebenso im Raum verorteten soziostrukturellen Daten wurde die Verbindung der verschiedenen Datensätze hergestellt, indem jeder Tat die Referenz (Gebäudeschlüssel) des am nächst gelegenen Gebäudes zugeordnet wurde. Die Gebäudedaten wurden in diesem Zusammenhang bereits von der Anbieterfirma der soziostrukturellen Daten mit einem eindeutigen Schlüssel übermittelt. Dieser Schlüssel ermöglichte es, die Daten der anderen Aggregationsniveaus (z. B. Wohnquartierebene) zuzuordnen. Darüber hinaus mussten diese bei der Aufarbeitung der Datensätze in ArcGIS in ein einheitliches Koordinatensystem transformiert werden.

Vorgangs-Duplikate³⁴ wurden im Rahmen der Georeferenzierung entfernt, um identische Tatorte nicht doppelt zu referenzieren. Ansonsten wurden vor Löschung des Duplikats wichtige Informationen aus dem Datensatz extrahiert. Duplikate wurden aus diesen Gründen zunächst dahingehend

³³ Die Polizeidaten liegen in einem Gauß-Krüger-Koordinatensystem vor, während das Gros an externen Daten in einem ETRS-1989-Koordinatensystem vorliegt.

³⁴ Fallduplikate können vorhanden sein, wenn zu dem zugrundeliegenden Vorgang mehrere Sachen erfasst wurden (z. B. entwendetes Gut). Die Anzahl der im Datensatz vorhandenen identischen Vorgangsschlüssel hängt somit von der Anzahl der erfassten Sachen ab.

aufbereitet, dass die Einträge beim entwendeten Gut mit dem georeferenzierten Hauptvorgang verkettet wurden. Dadurch wurde gewährleistet, dass bei der Duplikatslöschung keine wichtigen Informationen bezüglich der Beute verloren gingen. Zusätzlich wurden aus der Vielzahl an Katalogwerten für „Gegenstandsarten“ sogenannte Beutekategorien gebildet, um die im WED erlangte Beute vorab entsprechend gruppieren zu können.

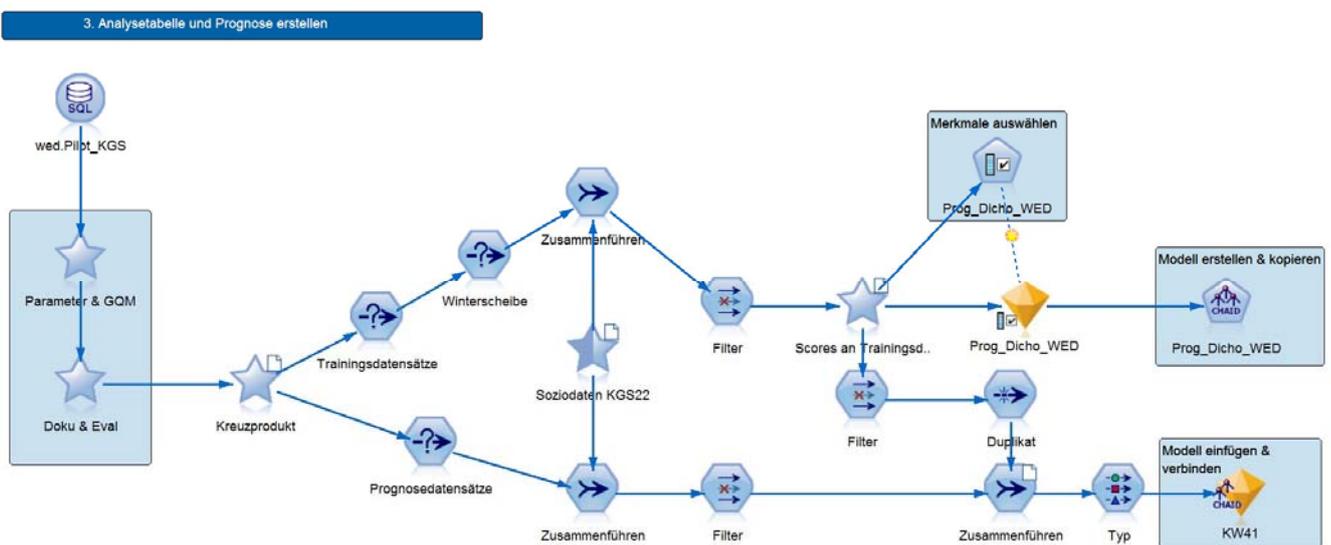
Neben der Georeferenzierung der Daten erfolgte auch eine Datenaufbereitung der Datums- und Zeitvariablen, da in diesem Bereich häufig fehlende oder fehlerhafte Werte zu verzeichnen waren. So wurden beispielsweise für den WED aus den Datumsfeldern, die die Wochentage als Klartext aufwiesen, numerische Werte abgeleitet. Außerdem wurde die Differenz der Tatzeiten (Beginn und Ende der er-

fassten Tatzeit) sowohl in Stunden, wie auch in Tagen berechnet. Taten, die eine größere Differenz als drei Tage aufwiesen, wurden im späteren Verlauf der Analyse ausgeschlossen, da solche Vorgänge zu statistischen Ausreißern geführt hätten. Taten mit einer Zeitdifferenz von bis zu drei Tagen wiesen aus statistischer Sicht eine überschaubare Datenunsicherheit auf und waren auch fachlich (z. B. ein WED, der sich am Wochenende ereignet hat) begründbar. Im Anschluss an die Georeferenzierung und die weitere Aufarbeitung des Datensatzes erfolgte die Modell- und Prognoseerstellung.

4.3.3 3. Schritt: Modell- und Prognoseerstellung

Im dritten Schritt fand die eigentliche Modell- und Prognoseerstellung statt. Abbildung 16 zeigt den ersten Teil dieser Prognoseerstellung.

Abbildung 16: Beispiel-Stream: Prognoseerstellung, 1. Teil



Vorbereiten der Daten für die Prognoseerstellung

Im ersten Abschnitt der Modell- und Prognoseerstellung erfolgte zunächst eine interne Dokumentation des Datensatzes, um stets nachvollziehen zu können, wie viele Datensätze zu welchem Zeitpunkt in die Modell- und Prognoseerstellung eingeflossen sind. Dies diente der Nachvollziehbarkeit des eigenen methodischen Vorgehens und gewährleistete die schnellere Detektion von Datenfehlern. Darüber hinaus wurden nunmehr die Vorgänge mit schlechten Geoqualitätsmarkern (GQM)³⁵ aus dem Datensatz entfernt.

Im Anschluss an die Löschung von schlechten GQM und die interne Dokumentation erfolgte die Bildung eines Kreuzproduktes. Im Kreuzprodukt wurde für jedes Wohnquartier der

Polizeibezirke (zum Ende der Projektlaufzeit 3.991 Quartiere) für jeden Tag seit dem 01.01.2012 erfasst, ob sich ein WED in dem spezifischen Wohnquartier ereignet hat. Dies ermöglichte die anschließende Aussage, ob z. B. in Wohnquartier X am Datum Y ein WED stattgefunden hat und unter welchen polizeilichen und soziostrukturellen Bedingungen sich dieser WED ereignet hat.

Konkret wurden hierfür die WED-Vorgänge zunächst über die jeweilige KP, den Wohnquartierschlüssel und das Datum aggregiert, sodass für jeden Tag im Datensatz, der mit mindestens einem Einbruch belastet war, die entsprechende Tatortbehörde, das Wohnquartier, in dem der Tatort lag, die summarische Anzahl der Einbrüche und die Scha-

³⁵ GQM verweisen auf die Güte der Georeferenzierung von polizeilichen Vorgangsdaten. Ist als Tatort beispielsweise lediglich eine Stadt bekannt und keine konkrete Adresse (z. B. Diebstahl während einer S-Bahn-Fahrt), so wird die Tat geografisch in die Stadtmitte verortet. Der GQM wird daraufhin als schlecht klassifiziert.

denssumme vorlagen. Um für die Erstellung des Kreuzproduktes auch die Tage in den Wohnquartieren ohne Einbruch einbeziehen zu können, wurde über einen *full outer join* für jedes Wohnquartier jeder einzelne Tag im Analysezeitraum über einen tabellarischen Kalender angefügt. Tage mit einem Einbruch wurden mit „1“ klassifiziert, Tage ohne Einbruch mit „0“. Diese Berechnung ermöglichte es, dass für jeden Tag weitere Daten, wie Informationen zu Ferienzeiten, Feiertagen, Mondphasen oder örtlichen Wetterlagen an den Datensatz gefügt werden konnten. Anschließend wurde der Datensatz nach Wohnquartierschlüsseln und Datum sortiert, sodass mittels einer Syntax jedem Quartier für jeden Tag der Gesamtzeitreihe die Anzahl der tatsächlichen WED innerhalb der Folgeweche (theoretischer, retrograder Gültigkeitszeitraum) zugeordnet werden konnte. Die dadurch erstellte (Ziel-)Variable stellte das Ergebnis dichotomisiert dar ($x \geq 1$, dann „wahr“/„1“, andernfalls „falsch“/„0“). Dadurch wurde das Modell flexibler und es fixierte sich nicht ausschließlich auf hoch frequentierte WED-Brennpunkte. Retrograd ließ sich somit für jedes Wohnquartier und jeden einzelnen Tag identifizieren, ob ein WED in diesem Quartier innerhalb des festgesetzten Gültigkeitszeitraumes erfolgte. Bei der Modellerstellung wird diese Zielvariable dann mittels der Eingabevariablen retrograd für jedes Wohnquartier und jede Woche der Gesamtzeitreihe berechnet. In Anlehnung an das Near-Repeat-Phänomen wurde zusätzlich für jeden Tag im Quartier die Anzahl der Tage seit dem letzten Einbruch fortlaufend berechnet. Aufgrund der zuvor beschriebenen umfangreichen Berechnungen mit großen Datenmengen nahm die Erstellung des Kreuzproduktes einen Großteil der Gesamtberechnungszeit in Anspruch. Der so erstellte Datensatz bot aber anschließend die Grundlage für die weiteren Berechnungsschritte, insbesondere die der konkreten Prognoseerstellung.

Nach der Kreuzprodukterstellung wurde der Datensatz in einen Trainings- und einen Prognosedatensatz aufgeteilt. Für den Trainingsdatensatz flossen alle Daten zwischen dem 01.01.2012 und dem Tag des letzten Datensatzes, abzüglich der Tagesanzahl des Gültigkeitszeitraumes (z. B. sieben Tage) ein. Zusätzlich wurde der Trainingsdatensatz mit Zeitscheiben an die jeweiligen Jahreszeiten angepasst, indem mit einer Jahreszeitscheibe und einer Winterzeitscheibe gerechnet wurde. Beispielsweise wurden für WED-Prognosen in den Wintermonaten auch nur Datensätze zur Analyse herangezogen, die in der Vergangenheit in Wintermonaten lagen. So ließen sich die strukturellen, saisonalen Unterschiede zwischen den WED adäquat abbilden und für die Modell- und Prognoseerstellung heranziehen. Darüber hinaus entsprach dieses Vorgehen den Erkenntnissen der hypothesengestützten Strukturanalyse zum WED (vgl. LKA NRW 2012).

Den Analyseteil unmittelbar vor der eigentlichen Prognoseerstellung bildete das Anfügen der verbleibenden soziostrukturellen Daten an den Datensatz sowie das Anfügen

von eigens für die Modellerstellung generierten Scores, die ebenfalls als potenzielle Vorhersagevariablen dienten. Zur Bildung der Scores wurden Berechnungen unter Verwendung von automatisiert laufenden Entscheidungsbäumen und einem CHAID-Algorithmus durchgeführt. Sie bildeten unabhängig voneinander die strukturellen Unterschiede von WED in den Sommer- und Wintermonaten, anhand der polizeilichen und soziostrukturellen Vorgangsdaten, ab. Konkret: Es wurde berechnet, wie sich die WED in den Sommer- und Wintermonaten strukturell voneinander unterscheiden. Aufbauend auf der Feststellung dieser Unterschiede wurden sogenannte „Profi-Scores“ gebildet, mit der Annahme und in Anlehnung an die Studie des LKA NRW aus dem Jahr 2012, dass professionelle Täter eher in den Wintermonaten agieren. Ebenfalls wurde im Laufe der Projektzeit der sogenannte PIAV-Merker als weitere Vorhersagevariable aufgenommen. PIAV bedeutet „Polizeilicher Informations- und Analyseverbund“. Der Merker wird von Sachbearbeitern regelmäßig gesetzt, wenn Hinweise auf professionelle Tatstrukturen mit überörtlichen Tätern vorliegen. Unter gewissen Relevanzkriterien, z. B. bei Serienstraftaten, bandenmäßiger Tatbegehung oder Bezügen in ein anderes Land bzw. das Ausland, werden die Daten bundesweit polizeilich bereitgestellt. Da jedoch die WED-Vorgänge, die mit dem Merker PIAV gekennzeichnet sind, nur einen Anteil von ca. sechs Prozent am Gesamtdatenbestand haben, waren die erzielten Ergebnisse statistisch und im Zusammenhang mit der Modell- und Prognoseerstellung von geringerer Bedeutung und wurden nicht genutzt. Nach den zuvor erläuterten, vorbereitenden Schritten erfolgte die eigentliche Prognoseerstellung.

Prognoseerstellung

Mit Fokus auf den Anspruch, ein transparentes und nachvollziehbares Prognosemodell zu generieren und das Prüfen der Methodik zu gewährleisten, wurde nach längerem explorativen Testen verschiedener Modellierungsoptionen zunächst das CHAID-Entscheidungsbaummodell gewählt. Entscheidungsbaummodelle bieten ein hohes Maß an Transparenz und Anpassungsflexibilität. Darüber hinaus gewährleisteten sie im Projekt SKALA die stetige Weiterentwicklung und ließen das hypothesengestützte Vorgehen und Testen der unterschiedlichen Vorhersagevariablen auf ihre Relevanz und Bedeutung zu. So ließen sich beispielsweise Entscheidungsbaummodelle automatisch nach Priorisierung einzelner, ausgewählter Vorhersagevariablen aufbauen oder aber die Aufspaltung (*split*) der Entscheidungsbäume erfolgte manuell gezielt im Hinblick auf bestimmte Vorhersagevariablen. Für die Vorhersage wurde relativ schnell Abstand von vollständig automatisiert aufbauenden Entscheidungsbäumen unter Einbezug aller infrage kommenden Vorhersagevariablen genommen. Sie ließen zwar häufig gute Zusammenhänge erkennen und ein Prognosemodell ließ sich hierauf aufbauen, aber der Nutzen war oft nicht gegeben, da keine inhaltliche Kausalität vorhanden war oder aber die berechneten WED-Prognosen zu statisch wurden.

Beispielsweise ergibt ein Zusammenschluss aus Merkmalen, die vor allem die Einbruchhäufigkeit abbilden, i. d. R. sehr statische Kriminalitätsprognosen.

Mit Hilfe der Entscheidungsbaummodelle zeigte sich relativ schnell, dass vor allem die polizeilichen Vorgangsdaten den höchsten Erklärwert in Bezug auf Kriminalitätsprognosen bieten. Vor allem bei den in Anlehnung an das Near-Repeat-Phänomen erstellten Variablen, beispielsweise die Information zum letzten Einbruch im jeweiligen Wohnquartier, konnten starke Einflüsse auf die Zielvariable festgestellt werden. Auch die grundsätzliche Häufigkeit von WED in einem Quartier erzielte sehr hohe Zusammenhangswerte. Allerdings wurde diese Variable aus inhaltlichen Gründen nur bedingt einbezogen, da sonst die WED-Prognosen zu statisch wurden und polizeilich bereits bekannte Brennpunkte fokussierten. Wurden bei den Entscheidungsbaummodellen hingegen die soziostrukturellen Daten, wie die Wohnlage, als erste Vorhersagevariable ausgewählt, ergab dies ebenfalls eine eher statische WED-Prognose. Dies lässt sich damit erklären, dass sich diese Merkmale, im Gegensatz zum Kriminalitätsgeschehen, im Wochenverlauf nicht dynamisch verändern (jährliche Aktualisierung). Die Aufspaltung der Entscheidungsbäume erfolgte demnach i. d. R. zunächst anhand der Variable „Tage seit dem letzten Einbruch im Quartier“. Je kürzer der Abstand hier zu verzeichnen war, desto stärker war der Zusammenhang. Anschließend wurde der Baum weiter aufgespaltet. Die weiteren Variablen unterschieden sich hierbei stark nach Jahreszeit und aktuellem Tatgeschehen. Häufig wurde auf die zuvor erwähnten Scores zurückgegriffen oder aber es wurden soziostrukturelle Daten, z. B. Daten zur Kaufkraft, zur Einwohnerstruktur (Haushalte im Haus, DINKS-Haushalte, ...), zur Passantenfrequenz oder zur Entfernung zur Autobahn zum vertiefenden Aufbau des Entscheidungsbaum genutzt. Variablen, die während des gesamten Projektzeitraumes keinen Zusammenhang zum WED aufwiesen, wurden bei der Modell- und Prognoseerstellung nicht weiter berücksichtigt.

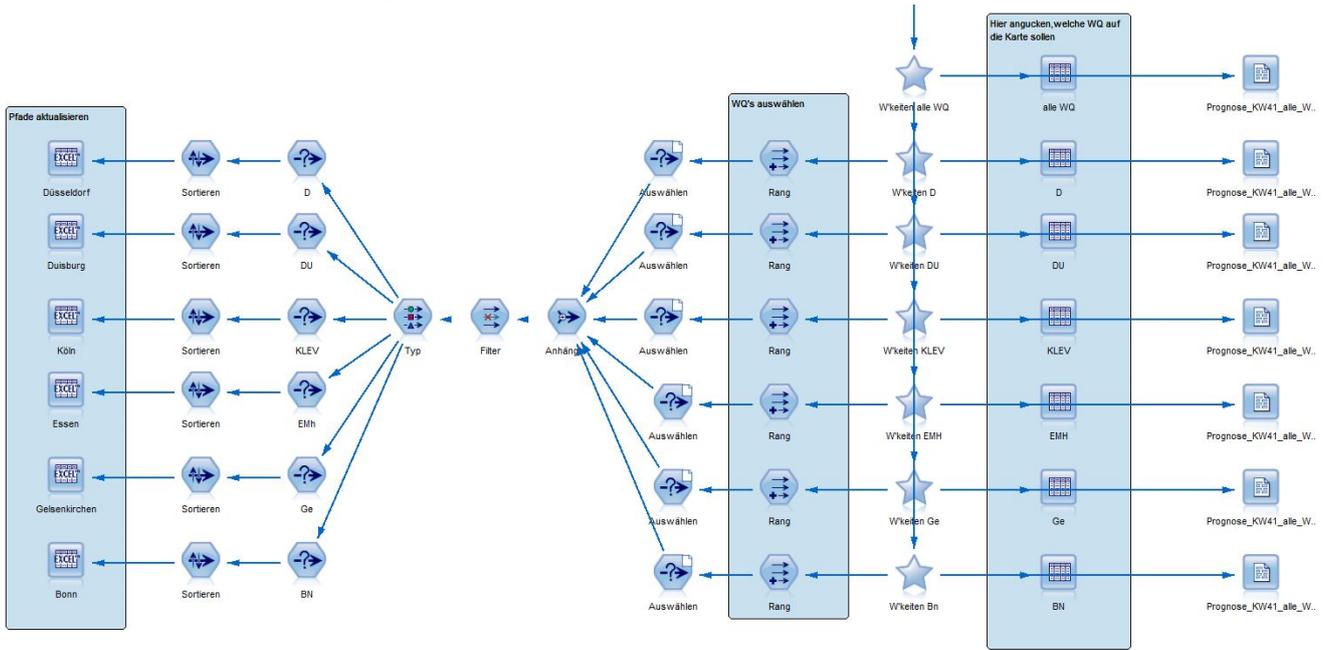
Zu Beginn des Projektes SKALA wurde ein relativ starker Zusammenhang zwischen WED und dem Wetter erwartet. Über den Deutschen Wetterdienst wurden deshalb entsprechende Datensätze bezogen und als Vorhersagevariablen in das Entscheidungsbaummodell eingebettet. Der erwartete starke Zusammenhang ließ sich allerdings während des gesamten Projektzeitraums nicht feststellen. Zumindest für die allesamt urban strukturierten Polizeibezirke war kein nennenswerter Zusammenhang (z. B. weniger Einbruch bei schlechterem Wetter) erkennbar. Das bedeutet, dass in Großstädten das Wetter keinen bedeutsamen Einfluss auf das Einbruchgeschehen hat. Ferner ist zu beachten, dass Informationen zum Wetter über Wetterstationen bezogen wurden und für eine Vielzahl an (geografisch unterschiedlichen) Wohnquartieren gelten. Je nach geografischer Lage war das tatsächliche Wetter demnach im Wohnquartier

durchaus anders als vom Deutschen Wetterdienst prognostiziert bzw. gemessen. Eine kleinräumigere Abbildung und Differenzierung ließ sich über die Daten allerdings nicht realisieren, sodass diesen ein gewisses Maß an Datenunsicherheit inhärent war.

Die auf Basis der selektierten Vorhersagevariablen mit dem CHAID-Algorithmus erstellten und getesteten Entscheidungsbaummodelle wurden anschließend auf den Prognosedatensatz angewendet. Jedem Wohnquartier wurde dadurch eine eigene statistische Einbruchwahrscheinlichkeit zugewiesen. Im Verhältnis zur Grundwahrscheinlichkeit (durchschnittliche Wahrscheinlichkeit eines WED im Wohnquartier für den Gültigkeitszeitraum) ließ sich regelmäßig eine zwei- bis dreifache Steigerung der Einbruchwahrscheinlichkeit berechnen. In Wintermonaten war die Wahrscheinlichkeit teilweise auch um den Faktor vier erhöht. Wenn beispielsweise berechnet wurde, dass für einen Gültigkeitszeitraum von einer Woche durchschnittlich ca. fünf Prozent aller Wohnquartiere von Einbrüchen betroffen sind und entsprechend die durchschnittliche Einbruchwahrscheinlichkeit je Quartier fünf Prozent war, ergab eine Steigerung der Wahrscheinlichkeit auf 20 Prozent in bestimmten Quartieren eine Vervierfachung des Einbruchrisikos bzw. des Ausgangsrisikos. Die mit dem höchsten Risiko behafteten Wohnquartiere wurden als WED-Prognosegebiete definiert.

Nach der Berechnung der Prognosegebiete, also der Wohnquartiere, die ein statistisch höheres Risiko für einen Einbruch im Gültigkeitszeitraum aufwiesen, erfolgte die konkrete Auswahl und Übermittlung der Quartiere an die Pilotbehörden. Die Ausgabe der Prognose erfolgte zunächst als Excel-Datei mit Verweisen zu den ausgewählten Wohnquartieren. Abbildung 17 zeigt diesen Auswahlprozess.

Abbildung 17: Beispiel-Stream: Prognoseerstellung, 2. Teil



4.3.4 4. Schritt: Visualisierung

Die Bereitstellung der Kriminalitätsprognosen erfolgte, neben einer für die Pilotbehörde editierbaren Datengrundlage, auch visuell für die operativen Polizeikräfte. Die Visualisierung der Prognosegebiete war in diesem Kontext der letzte Schritt in der Modell- und Prognoseerstellung.

Zu Beginn des Projektes wurden mit ArcGIS klassische Druckkarten (im PDF- oder PNG-Format) erstellt. Dies umfasste eine wöchentliche Übersichtskarte aller Prognosegebiete sowie entsprechende Detailkarten. Abbildung 18 und 19 zeigen dies beispielhaft für den Polizeibezirk Köln.

Abbildung 18: Beispiel-Übersichtskarte Köln

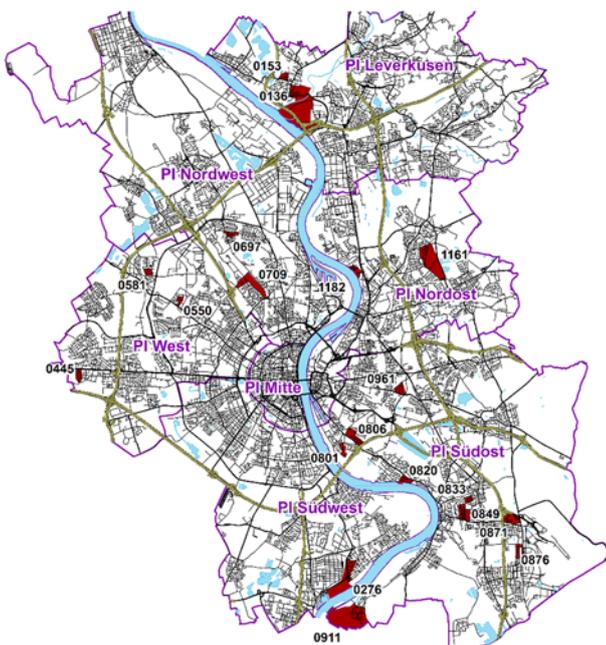


Abbildung 19: Beispiel-Detailkarte Köln



Die Bereitstellung der Karten und des Datensatzes erfolgten über ein SharePoint-Laufwerk im Polizeinetz. Jeden Montag wurden hier die neuen Prognosegebiete für den WED digital abgelegt. Ausgewiesene Bedienstete, i. d. R. aus den Auswertedienststellen der Pilotbehörden, nahmen anschließend eine Bewertung vor und entschieden unter Einbeziehung ihrer kriminalfachlichen Expertise, welche der vom LKA NRW generierten Gebiete mit höheren Wahrscheinlichkeiten eines WED den operativen Polizeikräften mitgeteilt wurden.

SKALA | MAP

Im Zusammenhang mit den Anforderungen aus den Pilotbehörden an die Visualisierung von Prognosegebieten wie auch dem Anspruch im Projekt offen und transparent vorzugehen, wurde in Zusammenarbeit mit dem Lehrstuhl für Datenanalyse und Visualisierung der Universität Konstanz das Visualisierungs- und Auswertetool SKALA | MAP entwickelt. Die Entwicklungszeit betrug drei Monate.

SKALA | MAP dient folgenden Zwecken:

1. Visualisierung der vom LKA NRW erstellten Kriminalitätsprognosen mit Möglichkeiten der einfachen und flexiblen Anpassung eigener Darstellungsweisen (Farben, Transparenz etc.)
2. Bereitstellung der grafischen Darstellung von Kriminalitätsprognosen auf Basis von Quartieren für die Organisationseinheiten der eigenen KPB
3. Bereitstellung der grafischen Darstellung von Delikten (als Punkte oder Heatmap) aufgrund georeferenzierter Daten aus IGVP (später ViVA) für die Organisationseinheiten der eigenen KPB
4. Auswertung visualisierter Kriminalitätsprognosen und Delikte sowie Einbezug von ergänzenden Informationen, basierend auf geografischen Daten (z. B. soziostrukturelle Daten)

SKALA | MAP unterstützt damit den Prozess der vorbeugenden Bekämpfung von Kriminalität unter Zuhilfenahme wissenschaftlicher Prognosetechniken. Es ermöglicht den KPB eine komfortable Auswahl und Darstellung der Prognosegebiete. Die Möglichkeit der grafischen Darstellung anderer Delikte aufgrund georeferenzierter Daten unterstützt sie bei dieser Auswahl. Abbildung 20, 21 und 22 zeigen beispielhaft SKALA | MAP Ausschnitte für den Polizeibezirk Köln.

Abbildung 20: Beispiel-Übersichtskarte KPB Köln aus SKALA | MAP (Prognose mit Heatmap)

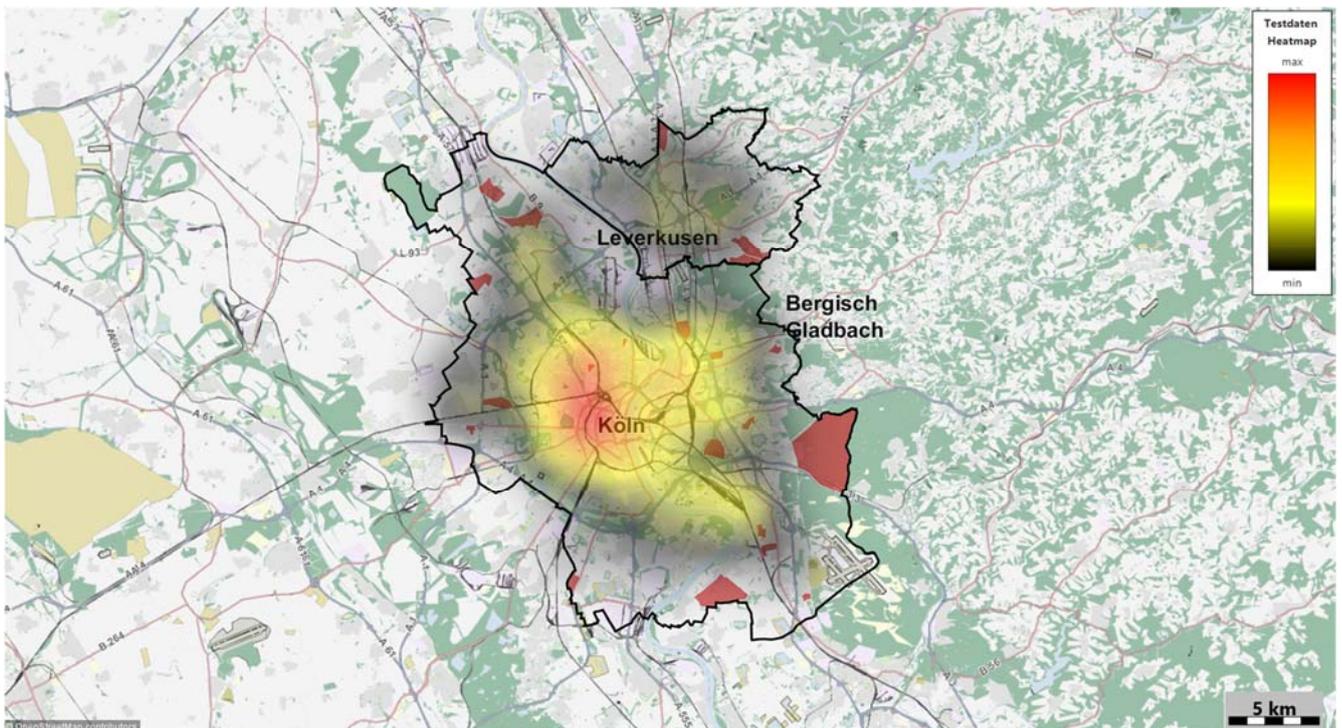


Abbildung 21: Beispiel-Übersichtskarte KPB Köln aus SKALA | MAP (Prognose mit Heatmap im Detail)

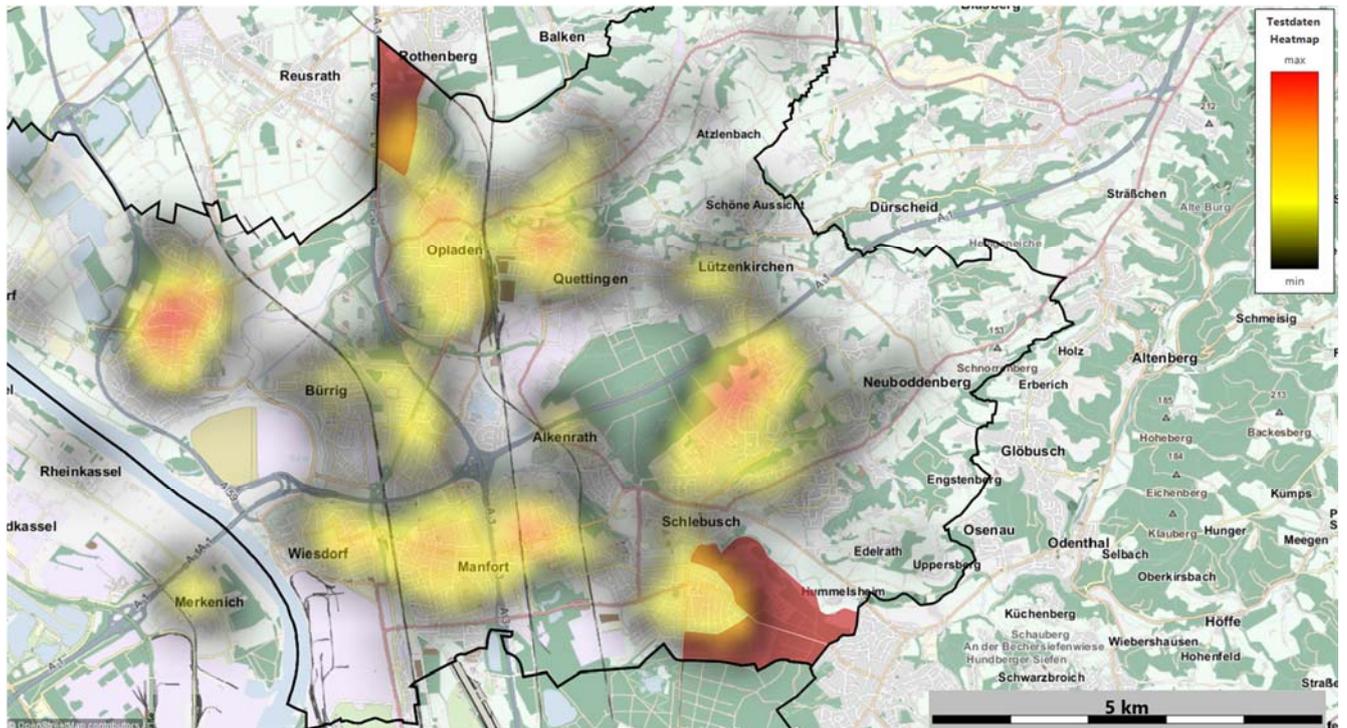
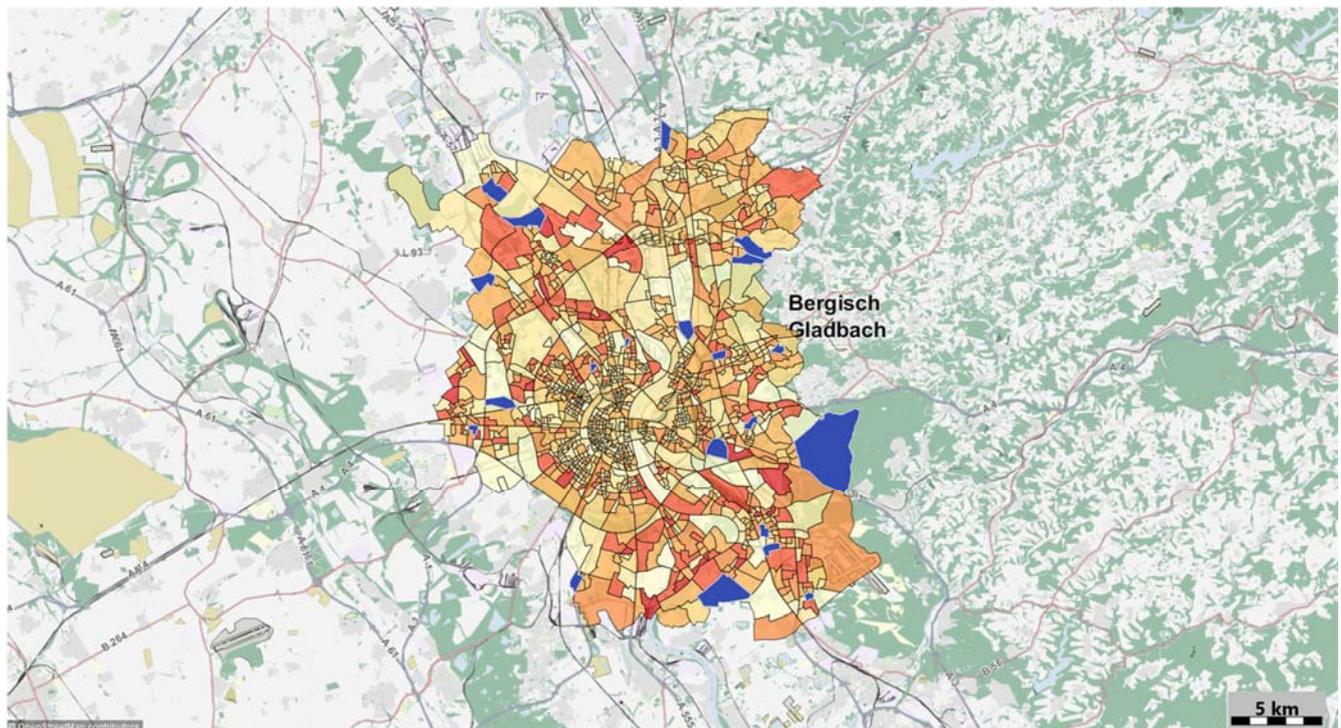


Abbildung 22: Beispiel-Übersichtskarte KPB Köln aus SKALA | MAP (Prognose mit allen Wahrscheinlichkeiten)



Jeder Pilotbehörde wurde SKALA | MAP zur Verfügung gestellt. Hiermit war es möglich, individuell und anwenderfreundlich eigene Karten zu erstellen und mit Zusatzinformationen zu verknüpfen (z. B. Heatmaps). Die Resonanzen aus den Pilotbehörden zu SKALA | MAP waren sehr positiv. Darüber hinaus erhielten zahlreiche weitere KPB auf konkrete Nachfrage Zugriff auf das Visualisierungs- und Auswertetool, um so die örtliche Kriminalität unter den verschiedenen Aspekten der Visualisierung adäquat

analysieren und darstellen zu können oder aber Lagebilder visuell zu verbessern.

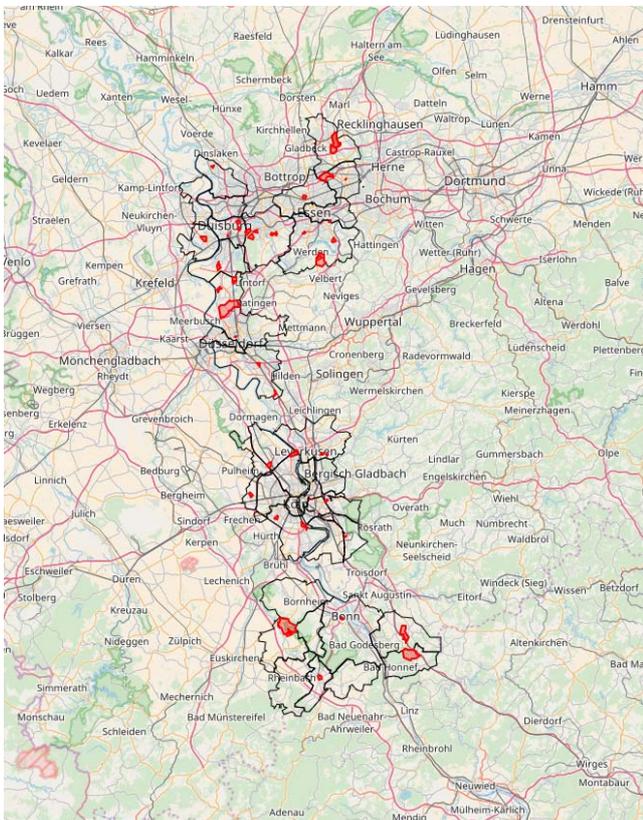
Damit unterstützt das im Projekt entstandene Visualisierungs- und Auswertetool SKALA | MAP die Aufgabebewältigung der Polizei nicht nur in Bezug auf die Visualisierung von Prognosegebieten, sondern verbessert durch einfache und flexible Darstellungsmöglichkeiten der räumlichen Verteilungen von Kriminalität das Erkennen spezifischer Muster.

SKALA | Web

Zusätzlich zu SKALA | MAP wurde auch eine web-basierte Visualisierung realisiert. Die web-basierte Darstellung der Prognosegebiete wurde wöchentlich aktualisiert und einen Tag nach Herausgabe der Prognose im Intranet veröffentlicht. Zuvor hatten alle Pilotbehörden rückgemeldet, in welchen der vom LKA NRW übermittelten Prognosegebiete hoheitliche Maßnahmen getroffen werden sollen.

Die Basiskarte der Intranetseite stammt, ebenso wie bei dem Visualisierungs- und Auswertetool SKALA | MAP von OpenStreetMap (<https://www.openstreetmap.de/index.html>). Die Intranetseite zur web-basierten Darstellung der Prognosegebiete ist unter dem Link <http://33-skala.polizei.nrw.de/skala/> verfügbar (Stand: Januar 2018). Abbildung 23 zeigt beispielhaft die Darstellung der Prognosegebiete in SKALA | Web.

Abbildung 23: Beispiel-Prognose, web-basiert



WED-Prognosegebiete werden rot dargestellt. Die Bezirksgrenzen der PI sind schwarz visualisiert. Sobald der Nutzer seinen *cursor* über eines der dargestellten Prognosegebiete zieht, werden die letzten vier Kennziffern des Wohnquartierschlüssels angezeigt. Dies erleichterte die interne Zuordnung der Gebiete. Gleiches gilt für die Anzeige der jeweiligen Polizeiinspektion. Abbildung 24 zeigt dies exemplarisch für die KPB Köln.

Abbildung 24: Beispiel-Prognose (Detail), web-basiert



Für die wöchentliche Aktualisierung von SKALA | Web mussten die *Shapefiles* der Prognosegebiete in entsprechende Polygone im json-Format umgewandelt werden, so dass die darzustellenden Daten von dem hinterlegten html-Skript gelesen werden konnten. Diese Umwandlung erfolgt unter Verwendung des Geoinformationssystems ArcGIS und der Open-Source-Software ORG2GUI (<http://ogr2gui.ca/>). Die json-Dateien enthielten nur die Polygongrenzen sowie Namen der Prognose- und PI-Bezirke.

Tablet-PC

Perspektivisch ist auch eine Prognoseübermittlung und -visualisierung anhand von Tablet-PC realisierbar. In diesem Fall bestünde die Möglichkeit einer dynamischen Prognose-darstellung. Das bedeutet, dass die jeweiligen Kriminalitätsprognosen beispielsweise anhand der aktuellen Uhrzeit und des jeweiligen Wochentags angepasst und nur für diese Zeiträume eingeblendet werden. Zudem wäre die Umsetzung in Verbindung mit einer GPS-Komponente möglich, die die bereits bestreiften Prognosegebiete automatisch aus-, und nach einer bestimmten Zeit, wieder einblendet (Stichwort: *gamification*).

4.3.5 Analytisches Modell

Die zuvor beschriebene methodische Umsetzung der Modell- und Prognoseerstellung fokussiert vor allem auf eine langfristige statistische Betrachtung. Im Laufe des Projektes SKALA kam vonseiten der Pilotbehörden immer wieder die Kritik auf, dass sich angehende Tatserien, die beispielsweise einen speziellen Modus Operandi aufwiesen, nicht in den WED-Prognosen widerspiegelten. Analysen anhand bekannter Tatserien aus den KPB zeigten hierbei, dass eine Abgrenzung zu anderen Serien oder Einzeltaten anhand des verfügbaren Datenmaterials nicht trennscharf möglich war, obwohl die Homogenität der Merkmale innerhalb der jeweiligen Serie stets hoch war. Als Ergänzung zum statistischen und entscheidungsbaumbasierten Ansatz der Modell- und Prognoseerstellung wurde deshalb ein ergänzendes Prognosemodell erstellt, das den Fokus auf mögliche Tatserien legt. Das analytische Modell ist unabhängig vom wesentlich umfangreicheren statistischen Modell und ergänzt dieses je nach Datenlage. Das analytische Modell verfolgt

den Gedanken einer detaillierteren Analyse einzelner WED auf Grundlage ihrer individuellen Merkmale und unter Berücksichtigung des raum-zeitlichen Tatgeschehens.

In der methodischen Durchführung wurde zunächst auf die zuvor bereinigten, georeferenzierten und in der Datenbank hinterlegten Datensätze zurückgegriffen. Diese wurden, wie in Kapitel 4.6 beschrieben, auf räumliche und zeitliche Zusammenhänge untersucht und bei einem möglichen Zusammenhang mit einer Serien-ID gekennzeichnet. Ein Zusammenhang zwischen zwei WED wurde angenommen, wenn die Taten maximal 14 Tage auseinander lagen und räumlich nicht mehr als 300 Meter voneinander entfernt stattgefunden haben. Diese Definition wurde im Hinblick auf das bereits empirisch gut bewährte Near-Repeat-Phänomen festgelegt und spiegelte damit auch die gängigen Befunde der Forschung (vgl. Bowers et al. 2005) aus diesem Bereich wider. Ein so hergestellter theoretischer Zusammenhang zwischen mehreren WED wurde im Anschluss durch weitere Informationen angereichert. Das heißt, es wurden für jede Tat Informationen bezüglich der deliktbezogenen Belastung (WED) sowie der Belastung durch Delikte der allgemeinen Straßenkriminalität in diesem Bereich ermittelt. Die Grundbelastung wurde jahresspezifisch unterschiedlich gewichtet, sodass die Bedeutung der Belastung mit größerer zeitlicher Distanz zur betreffenden Tat kontinuierlich abnahm. Aus diesen Berechnungen ergab sich ein Hotspot Index, der, bezogen auf die einzelnen WED, Aussagen darüber zuließ, inwieweit das jeweilige Wohnquartier grundsätzlich mit Kriminalität belastet war. Zusätzlich erfolgte eine Berechnung der Tage seit dem letzten Einbruch im Wohnquartier bzw. Straßenabschnitt, um eine kurzfristige Entwicklung der Delikte in diesem Bereich darstellen zu können.

Im weiteren Verlauf der Modell- und Prognoseerstellung wurden in Anlehnung an den Ansatz der *homogeneous areas* und des *contagion process of victimisation* (vgl. Townsley et al. 2003) sowie auf Grundlage der jeweils vorhandenen soziostrukturellen Daten zusätzliche Informationen zu den prognostizierten Auswirkungen des stattgefundenen Deliktes auf eventuelle Folgedelikte generiert. Hierzu wurden die soziostrukturellen Daten zur sozialen Schicht, der Wohnlage und der vorhandenen Gebäudetypen innerhalb eines Wohnquartiers genutzt, um so den Grad der Homogenität abbilden zu können.

Der Gefährdungsgrad des jeweiligen Wohnquartiers wurde anschließend anhand einer „Übertragungsquote“ (vgl. Townsley et al. 2003) bestimmt. Diese gab an, wie gut das jeweilige Wohnquartier an Bundesstraßen oder Autobahnen angebunden war bzw. wie weit entfernt von einer Bundesstraße oder Autobahn der letzte WED stattgefunden hat. In diesem Zusammenhang wurde grundsätzlich davon ausgegangen, dass eine gute Verkehrsanbindung des Tatortes zu einer höheren Gefahr der Viktimisierung der umliegenden Bereiche führte. Mit Blick auf die WED-Studie des LKA

NRW (2017) wurde dabei der Einfluss der Entfernung zu einer Bundesstraße höher gewichtet als die Entfernung zu einer Autobahn.

Aufgrund der in der Datenbank hinterlegten Falldaten wurde sodann die Möglichkeit der Auswertung des Modus Operandi sowie der Beute genutzt, um die Bedeutsamkeit eines WED besser beurteilen zu können. Hierzu mussten aufgrund der Vielzahl der möglichen Ausprägungen in den polizeilichen Vorgangsdaten Kategorien für die Beute und den Modus Operandi gebildet werden, die sich in der Art der Ausführung bzw. der unterstellten Professionalität des Täters voneinander unterschieden. Neben dieser Gewichtung für die Beurteilung des WED wurde zudem die letzte Art der Tatbegehung innerhalb des Wohnquartiers miteingefasst, sodass ein Vergleich zwischen aktueller und vergangener Tat, im Hinblick auf den Modus Operandi, möglich wurde. Bei der Beute war es darüber hinaus notwendig, den Fokus darauf zu richten, welche Beute professionelle Täter nicht entwerfen würden. Denn lediglich ein hoher Beutewert, beispielsweise Schmuck als entwendetes Gut, lässt keinen automatischen Rückschluss auf professionelle Täter zu.

Die angesprochenen Indikatoren wurden am Ende des Modells in die Variable „Profi_norm“ überführt. Diese ermöglichte es, mit einem Wertebereich zwischen 0 und 1, eine Aussage darüber zu treffen, ob es sich bei der betrachteten Tat möglicherweise um eine für die Kriminalitätsprognose bedeutsame Tat handeln könnte. Registrierte Taten, die einen PIAV-Merker aufwiesen, wurden ebenfalls in die Auswahl aufgenommen und am Ende des Modells anhand der weiteren Indikatoren bewertet. Das analytische Modell schloss somit die festgestellte Lücke und ermöglichte den detaillierten Blick zur Erkennung möglicher Tatserien.

4.4 Testung von Prognosemodellen

In dem vorherigen Kapitel (4.3) wurde umfassend dokumentiert, mit welchen Schritten ein Prognosemodell im Projekt SKALA realisiert wurde. Für das statistische Prognosemodell wurde mit dem CHAID-Algorithmus gearbeitet. Es wurden jedoch auch andere Modellansätze getestet. Die Wahl eines Entscheidungsbaummodells wurde aus wissenschaftlichen und inhaltlichen Gründen getroffen. Diese werden nachfolgend, insbesondere mit Blick auf andere anwendbare und angewendete Modelle, tiefergehend erläutert.

4.4.1 Entscheidungsbaummodelle

Entscheidungsbaummodelle sind Machine-Learning-Algorithmen (vgl. Quinlan 1986), die es ermöglichen, Daten zu klassifizieren und Datenlücken, basierend auf den vorhandenen Daten, zu schließen. Außerdem können sie verwendet werden, um für weitere Analysen von vielen zur Verfügung stehenden Variablen genau diejenigen auszuwählen, die die Zielvariable (z. B. WED als Prognosedelikt) am meisten beeinflussen.

Ein sehr großes Anwendungsgebiet von Entscheidungsbaummodellen ist die Vorhersage basierend auf Daten der Vergangenheit (vgl. Song/Lu 2015: 130). Sie werden grundsätzlich in vielen unterschiedlichen Bereichen angewendet, u. a. im Finanzwesen, Marketing, Ingenieurwesen und der Medizin (vgl. Maimon/Rokach 2015: 1). In der Medizin werden sie beispielsweise genutzt, um bestimmte Krankheiten auf Basis unterschiedlicher Symptome der Patienten zu diagnostizieren (vgl. Song/Lu 2015: 130).

Ein Entscheidungsbaum ist ein hierarchisches Modell, bei welchem der Datensatz mittels sogenannter „Entscheidungsknoten“ in mehrere sich gegenseitig ausschließende Untergruppen aufgeteilt wird. Von einem Entscheidungsknoten gehen zwei oder mehrere Verzweigungen oder sogenannte Äste ab. Das Kriterium für die Aufspaltung hängt von den Merkmalsausprägungen und dem Wertebereich des jeweiligen Attributes ab. Ein Entscheidungsknoten bezieht sich immer auf eines der verwendeten Attribute. Der erste Entscheidungsknoten an der Wurzel des Baumes³⁶ ist das einflussreichste Attribut für die Vorhersage der Zielvariablen. In einem Entscheidungsbaum können einzelne Attribute mehrfach in unterschiedlichen Entscheidungsknoten vorkommen (vgl. Maimon/Rokach 2015: 12f.). Ziel der Entscheidungsbaumanalyse ist es, einen Baum mit einer optimalen Unterteilung des Datensatzes zu erstellen, sodass mit den Vorhersagevariablen die Zielvariable in möglichst vielen Fällen korrekt bestimmt wird. Somit kann ein Entscheidungsbaummodell nicht nur zur Klassifizierung, sondern auch zur Vorhersage der Zielvariable dienen.

Grundsätzlich unterscheidet man Klassifizierungs- und Regressionsbäume. Bei der Nutzung von Entscheidungsbaummodellen als Klassifizierungsmodell wird der Terminus Klassifizierungsbaum benutzt, bei der Nutzung des Entscheidungsbaumes als Regressionsmodell wird der Begriff Regressionsbaum verwendet. Ob ein Entscheidungsbaum als Klassifizierungs- oder Regressionsmodell genutzt wird, hängt von der Aufgabenstellung und der Art der eingehenden Daten ab. Liegen vornehmlich numerische Daten vor und soll der Wert einer (numerischen) Zielvariable abgeschätzt werden, bietet es sich an, ein Regressionsmodell zu verwenden. Numerische Daten werden hierbei in unterschiedliche Kategorien entsprechend ihrer Werteskala aufgeteilt und somit die Untergruppen durch Wertebereiche definiert. Im Projekt SKALA wurde der Entscheidungsbaum in Form eines Klassifizierungsbaumes genutzt, da u. a. ordinal (kategorische) und nominal skalierte Daten verwendet wurden. Ordinale Daten sind dadurch gekennzeichnet, dass sie in eine sinnvolle Reihenfolge gebracht aber nicht miteinander verrechnet werden können (z. B. Qualität der Wohnlage). Nominal skalierte Daten besitzen verschiedene Ausprägungen, die

aber in keine Reihenfolge gebracht werden können (z. B. PKW-Marke). Bei ordinal und nominal skalierten Daten wird der Entscheidungsbaum entsprechend der jeweiligen Ausprägung bzw. den Kategorien eines Attributes aufgespalten.

Nachfolgend werden, in Anlehnung an Maimon/Rokach (2015: 10f.), die wesentlichen Vor- und Nachteile von Entscheidungsbaummodellen zusammengefasst.

Ein wesentlicher Vorteil von Entscheidungsbaummodellen ist, dass sie durch ihre Anlage und Methodik sehr transparent und damit auch für Außenstehende leicht zu verstehen sind. Entscheidungsbäume sind selbsterklärend und können i. d. R. einfach nachvollzogen werden, sodass mit ihnen dem in Kapitel 3.13.1 beschriebene Anspruch einer transparenten (keine „Black Box“) Methodik Rechnung getragen werden kann. Darüber hinaus können sowohl nominale als auch numerische Eingabevariablen berücksichtigt werden. Außerdem kann ein Entscheidungsbaum auch mit Datensätzen umgehen, die Fehler oder Datenlücken enthalten. Gerade solche fanden sich relativ häufig in den polizeilichen Vorgangsdaten, da beispielsweise beim WED Tatzeiten nicht genau eingegrenzt werden konnten oder aber die Beutehöhe zum Zeitpunkt der Datenerhebung noch nicht im Vorgangsbearbeitungssystem der Polizei erfasst war. Da die Methode aber nicht parametrisch ist, beinhaltet sie auch keine Annahmen über die Verteilungsfunktion der verwendeten Variablen. Ein Nachteil des Entscheidungsbaumansatzes ist, dass die Einflussstärke der Attribute jeweils isoliert getestet wird und somit eine möglicherweise sinnvolle Kombination von Attributen nicht berücksichtigt wird. Darüber hinaus ist ein wesentlicher Nachteil bei Entscheidungsbaummodellen, dass sie sehr sensitiv sind, was die Verwendung vieler (auch wenig einflussreicher) Attribute oder Daten mit einem starken „Rauschen“ (*random error* oder *noise*) betrifft. Das bedeutet, dass bei unbedachter Anwendung dieses Modellansatzes die Gefahr hoch ist, dass das erstellte Entscheidungsbaummodell instabil und somit schlecht übertragbar und für Kriminalitätsprognosen ungeeignet ist. Wird die Komplexität eines Entscheidungsbaums deshalb nicht gezielt durch die Modellparametrisierung eingeschränkt und die Variablen sinnvoll gewählt, können schon geringfügige Änderungen des Trainingsdatensatzes die Struktur eines Baumes komplett verändern und der Algorithmus würde unter Umständen nicht die optimale Auswahl von Attributen finden. Insofern sind Komplexität und Robustheit konkurrierende Modelleigenschaften, die bei Erstellung eines statistischen Modells berücksichtigt werden müssen (vgl. Song/Lu 2015: 132). Je komplexer ein Modell ist, desto höher ist die Gefahr des *overfittings* (vgl. Berthold et al. 2010: 102ff.), das häufig mit einer schlechten Übertragbarkeit einhergeht. Um dies zu vermeiden, können vor

³⁶ Bei Darstellungen von Entscheidungsbaummodellen befindet sich die Wurzel grundsätzlich „oben“.

der Modellerstellung Stoppregeln anhand von Modellparametern definiert werden, die den maximalen Umfang eines Entscheidungsbaumes eingrenzen. Dies sind u. a.

- die minimale Anzahl von Datensätzen in einem Blatt,
- die minimale Anzahl von Datensätzen eines Entscheidungsknotens, sowie
- die Tiefe (maximale Anzahl der Verzweigungen).

Neben Stoppparametern kann nach der Baumerstellung auch eine „Beschneidung“ des Baumes (*pruning*) durchgeführt werden. Dabei werden einzelne Verzweigungen und Äste eines Baumes entfernt, die nicht zu einer hohen Modellgüte beitragen. Je nach Ansatz werden hierfür z. B. Chi-Quadrat-Tests verwendet.

Im Hinblick auf die verwendeten Algorithmen gibt es verschiedene Arten. Nachfolgend werden zwei der gängigsten Algorithmen dokumentiert und beschrieben, da sie auch im Rahmen des Projektes SKALA getestet und angewendet wurden:

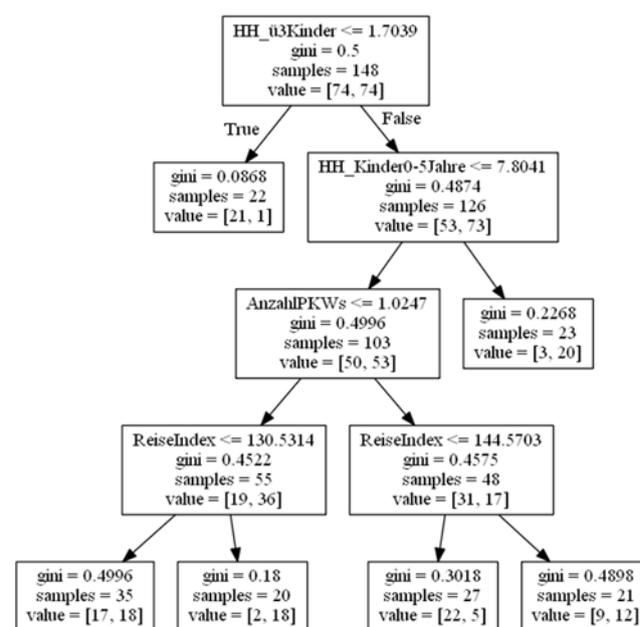
CART-Algorithmus

CART steht für *classification and regression trees*. Mit dem CART-Algorithmus können sowohl Klassifizierungs- als auch Regressionsbäume (C&R-Bäume) erstellt werden. Der CART-Algorithmus ist dadurch gekennzeichnet, dass er nur binäre Verzweigungen zulässt. Von jedem Entscheidungsknoten gehen genau zwei Äste ab. In dem Beispiel des Klassifizierungsbaumes, der in Abbildung 25 dargestellt ist, beinhaltet die Untergruppe der linken Abzweigung die Daten, für die das Kriterium des Entscheidungsknotens mit „wahr“ beantwortet wird, der rechte Knoten die Daten, für die das jeweilige Kriterium mit „falsch“ beantwortet wird.

Zur Erläuterung des CART-Klassifizierungsbaum-Beispiels: Die Wurzel des Baumes und dementsprechend das einflussreichste Attribut ist hier die Variable „HH_ü3Kinder“ (relative Anzahl der Haushalte mit drei Kindern oder mehr). Der Algorithmus hat als Entscheidungskriterium HH_ü3Kinder mit $\leq 1,7039$ berechnet. Der Datensatz umfasst eine Größe von 148 (*samples*). Mit *value* sind die jeweiligen Ausprägungen der Zielvariablen gemeint. In diesem Beispiel gibt es zwei Ausprägungen: „wenige Einbrüche“ und „viele Einbrüche“. Dies wurde zuvor anhand eines Grenzwertes definiert und bedeutet, dass der verwendete Trainingsdatensatz genau 74 Daten von Kategorie „1“ (wenig Einbrüche) und 74 Daten der Kategorie „2“ (viele Einbrüche) hatte. Der

CART-Algorithmus definiert das Spaltungskriterium im Entscheidungsknoten basierend auf dem Grad der Homogenität innerhalb der Klassen, die durch die Datenaufspaltung erzeugt werden. Der statistische Index, der im Falle des Klassifikationsbaumes über die Auswahl des Entscheidungsknotens und des Aufspaltungskriteriums entscheidet, ist der Gini-Index³⁷. Dieser ist auch im Beispiel (siehe Abbildung 25) dargestellt.

Abbildung 25: Beispiel-CART-Klassifizierungsbaum



CHAID-Algorithmus

Im Gegensatz zum CART-Algorithmus kann der CHAID-Algorithmus nicht-binäre Bäume generieren, was bedeutet, dass ein Knoten auch mehr als zwei Verzweigungen haben kann. Diese werden kategorial in Abhängigkeit vom Wertebereich der Trainingsdaten gruppiert. Bei jeder Aufspaltung sucht der Algorithmus nach einer Vorhersagevariable, die die Kategorie der Zielvariable am stärksten erklärt. Bei der Entscheidung, ob eine Aufspaltung mittels einer bestimmten Variable erfolgt, testet der Algorithmus die Hypothese in Bezug zur Abhängigkeit zwischen der aufgespaltenen Einflussvariable und der kategorialen Zielvariable. Dies geschieht unter Verwendung des Chi-Quadrat-Tests zur Unabhängigkeit zwischen den Daten mittels eines vordefinierten Signifikanzniveaus. Sobald der Chi-Quadrat Test ergibt, dass die aufgespaltene Einflussvariable und die Zielvariable unabhängig voneinander sind, wird das Wachstum des Knotens gestoppt. Andernfalls erfolgt die Aufspaltung des aktu-

³⁷ Der Gini-Index bzw. Gini-Koeffizient ist ein gängiges statistisches Maß, welches Ungleichverteilungen beschreibt. Ungleichverteilungsmaße zeigen an, zu welchem Grad zwei Variablen von einer Gleichverteilung abweichen. Der Gini-Index wird aus der Lorenz-Kurve abgeleitet. Berechnungsgrundlage sind aufsteigend sortierte Listen der zwei untersuchten Variablen. Der Gini-Koeffizient reicht von 0 (völlige Gleichverteilung) bis 1 (maximale Ungleichverteilung).

ellen Blattes und eine weitere optimale Aufspaltungsmöglichkeit wird gesucht. Durch die Festlegung eines geeigneten Signifikanzniveaus kann in diesem Zusammenhang *overfitting* methodisch vermieden werden (vgl. Maimon/Rokach 2015: 57).

4.4.2 Random Forest

Der Random-Forest-Algorithmus ist ein Ensemble mehrerer kombinierter Entscheidungsbäume. Für jeden einzelnen Baum wird aus der gesamten Anzahl der Einflussvariablen des Trainingsdatensatzes eine kleinere Variablenzahl N zufällig gesampelt. Es wird darauf aufbauend jeweils die beste Aufspaltung, unter dieser Variablenauswahl, berechnet. Hierbei handelt es sich nicht um die beste Aufspaltung unter Nutzung aller Einflussvariablen. Am Ende werden die Ergebnisse des Ensembles gemittelt. Durch die Zufallskomponente bei den Aufspaltungen an den Entscheidungsknoten ist jeder einzelne Baum des Ensembles i. d. R. etwas weniger genau als der Baum mit einer exakten Aufspaltung. Allerdings führt die Kombination mehrerer „ungefährer“ Bäume in Form eines Modell-Ensembles zu einer Abnahme der Varianz und regelmäßig zu einer verbesserten Genauigkeit im Vergleich zu einfachen Entscheidungsmodellen mit exakten Aufspaltungen. Wie alle Ensemble-Methoden ist das Random-Forest-Modell somit grundsätzlich robuster und verlässlicher als ein einfaches Entscheidungsbaummodell (Maimon/Rokach 2015: 106f.).

4.4.3 Weitere Modellansätze

Abgesehen von den entscheidungsbaum-basierten Modellansätzen wurden im Projekt SKALA weitere Modelle getestet (siehe Kapitel 4.5.2). Hierfür wurde primär mit der Standardparametrisierung gearbeitet, die von der jeweils verwendeten Software (IBM SPSS Modeler bzw. Python-Pakete) vordefiniert war. Aus diesem Grund werden die nachfolgenden alternativen Modellansätze nicht ausführlich erläutert, sondern lediglich aufgezählt. Die alternativen Modellansätze, die getestet wurden und ebenfalls für prädiktive Analysen bei der Umsetzung von Predictive Policing verwendet werden können, sind:

- **Logistische Regression**
Bei der logistischen Regression handelt es sich um eine Regressionsanalyse auf Basis binärer Variablen. Diskrete Variablen mit mehr als zwei Ausprägungen werden entsprechend in Serien binärer Dummy-Variablen transformiert (vgl. Friedman et al. 2001: 119ff.).
- **Naive Bayes**
Hierbei handelt es sich um einen einfachen Klassifikator, der auf dem Bayes-Theorem basiert und insbesondere bei einer sehr hohen Anzahl von Variablen, das heißt sehr vielen Dimensionen, geeignet ist (vgl. Friedman et al. 2001: 210ff.).
- **AdaBoost / Adaptive Boosting**
Hierbei handelt es sich um einen Maschine-Learn-

ing-Algorithmus, der in den 1990er Jahren entwickelt wurde und bei dem die sogenannte Exponential-Loss-Funktion minimiert wird (vgl. Friedman et al. 2001: 343ff.).

- **k-Nearest Neighbors**
Bei k-Nearest Neighbors handelt es sich um eine nicht-parametrische Klassifizierungsmethode, bei der im mehrdimensionalen Raum nahe gelegene Elemente/Datenpunkte innerhalb eines vorher definierten Radius zusammengefasst werden (vgl. Sammut/Webb 2011: 714; Friedman et al. 2001: 463ff.).
- **Künstliche Neuronale Netze (Neural Net) oder auch ANN (Artificial Neural Network)**
Das Prinzip neuronaler Netze basiert auf dem Vorbild des Nervensystems, genauer der Verknüpfung, dem Informationsaustausch und dem Lernen von Nervenzellen (Neuronen). Entsprechend bestehen neuronale Netze aus miteinander verknüpften Datenknoten. Die Struktur des Netzwerks erstellt der lernende Algorithmus selbst auf Basis von gewichteten Verbindungen zwischen Einflussvariablen und der Zielvariable (vgl. Sammut und Webb 2011: 716).

Insbesondere im Hinblick auf die Transparenz, die leichte Nachvollziehbarkeit und die weiteren oben beschriebenen Vorteile wurde sich im Rahmen des Projektes SKALA bei der (technisch) methodischen Umsetzung für die Verwendung von Entscheidungsbäumen entschieden. Ein Modellvergleich zeigte zudem die verhältnismäßig gute Modellperformance (siehe Kapitel 4.5.2).

4.4.4 Modellvalidierung

Innerhalb des Projektes war es erforderlich, die erstellten Modelle bzw. den gewählten Algorithmus auf seine Güte zu testen. Dies ermöglichte eine interne Modellweiterentwicklung und -verbesserung. Im Kapitel 3.11 wurde bereits auf das Problem der Messbarkeit und Vergleichbarkeit von Prognosemodellen im Rahmen von Predictive Policing umfassend eingegangen. Da kein Vergleich mit anderen Predictive-Policing-Verfahren angestrebt wurde, kam in diesem Fall eine Modellvalidierung mit zu berechnenden Indexwerten aus einer Konfusionsmatrix in Betracht. Bei der Validierung von Modellen handelt es sich um die Quantifizierung der Vorhersagegenauigkeit eines erstellten Modells. Für die Validierung wird grundsätzlich ein separater Datensatz verwendet, der nicht an der Modellerstellung beteiligt war. Anschließend wird überprüft, wie viele der prognostizierten Daten mit den tatsächlichen Daten übereinstimmen. Eine sehr gängige Methode für die Validierung von Modellen zur Vorhersage kategorialer Zielvariablen, wie sie in diesem Projekt i. d. R. vorlagen, ist die Berechnung von Evaluierungsindizes, deren Berechnung mit Hilfe einer sogenannten Konfusionsmatrix (siehe Tabelle 4) veranschaulicht werden kann.

Tabelle 4: Konfusionsmatrix zur Modellvalidierung

| | | Testdaten | |
|------------|---|---------------------|---------------------|
| | | 1 | 0 |
| Vorhersage | 1 | TruePositive | FalsePositive |
| | 0 | FalseNegative | TrueNegative |

Die Testdaten sind in diesem Fall die tatsächlich gemessenen Daten der Zielvariable (hier WED). Vorhersagen sind die entsprechenden Daten, die vom Modell prognostiziert wurden. In Tabelle 4 stünde der Wert „1“ z. B. für „WED-anfälliges Gebiet“ und „0“ für „kein WED-anfälliges Gebiet“. Ist für ein Wohnquartier beispielsweise sowohl der prognostizierte als auch der tatsächliche Wert eine „1“, würde es unter die Kategorie TruePositive fallen, das heißt ein Wohnquartier wurde richtig prognostiziert. Ein Wohnquartier würde falsch prognostiziert, wenn im Gültigkeitszeitraum tatsächlich wenige WED stattfanden (FalsePositive) oder von WED betroffene Wohnquartiere nicht in der Prognose enthalten waren (FalsePositive) oder tatsächlich anfällige Wohnquartiere nicht als solche berechnet (FalseNegative) wurden. Die Werte für TruePositive, TrueNegative, FalsePositive und FalseNegative sind die Anzahl der berücksichtigten Daten, die in die jeweilige Kategorie fallen.

Auf Basis der Elemente aus der Konfusionsmatrix lassen sich unterschiedliche Evaluierungsindizes berechnen, die nachfolgend – da hierauf in den weiteren Kapiteln immer wieder Bezug genommen wird – beschrieben werden:

Der Recall [%] wurde wie folgt berechnet:

$$\text{Recall} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}} \cdot 100$$

Der Recall ist der Anteil von richtig prognostizierten Wohnquartieren mit dem Wert „1“, anteilig an der Gesamtanzahl von tatsächlich vorliegenden Wohnquartieren mit dem Wert „1“.

Die Precision [%] wurde wie folgt berechnet:

$$\text{Precision} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}} \cdot 100$$

Die Precision ist der Anteil von richtig prognostizierten Wohnquartieren mit dem Wert „1“, anteilig an der Gesamtanzahl von Prognosegebieten.

Sowohl beim Recall als auch bei der Precision liegt der Fokus der Indizes in diesem Fall bei den Gebieten, die als WED-anfällig prognostiziert wurden. Die Verwendung dieser Indizes ist demnach im Rahmen von Predictive Policing

sinnvoll, um zu überprüfen, wie verlässlich Prognosen sind, die höhere Einbruchwahrscheinlichkeiten berechnen.

Um grundsätzlich die Güte (Modellperformance) von Klassifizierungsmodellen zu testen, wird auch häufig der Accuracy Index verwendet:

$$\text{Accuracy} = \frac{(\text{TruePositive} + \text{TrueNegative})}{\Sigma \text{Val}} \cdot 100$$

Die Anzahl von insgesamt für die Validierung verwendeten Wohnquartieren, das heißt die Summe von TruePositive, TrueNegative, FalsePositive und FalseNegative, wird hier mit ΣVal abgekürzt.

Es ist gängige Praxis, eine sogenannte Resampling-Methode zu nutzen, um robustere Validierungsergebnisse zu erhalten. Eine verbreitete Methode, die auch hier angewendet wurde, ist die n -fache Kreuzvalidierung. Bei der n -fachen Kreuzvalidierung wird der Datensatz zufällig aufgeteilt in n sich gegenseitig ausschließende Untergruppen von ungefähr gleicher Größe. Entsprechend wird das Modell n Mal generiert (trainiert) und getestet (evaluiert). Jedes Mal wird es anhand einer Untergruppe trainiert und anhand aller übrigen $n-1$ Untergruppen mittels eines vordefinierten Indizes (hier: Accuracy) getestet (vgl. Maimon/Rokach 2015: 51f.). Der finale Validierungsindex (hier Acv) ist dann der Mittelwert der n berechneten Indizes.

4.5 Testung von Variableneinflüssen im Modell

Die Ausrichtung des Projektes SKALA und das damit verbundene hypothesengestützte Vorgehen erforderten die vorhandenen polizeilichen Datenbestände mit soziostrukturellen Daten anzureichern. In der konkreten Modell- und Prognoseerstellung flossen diese dann als Einflussvariablen in das Vorhersagemodell ein (siehe Kapitel 4.3). Für die Überprüfung, ob auch nachweislich ein Zusammenhang zwischen diesen Daten und den polizeilichen (WED-)Vorgangsdaten bestand und ob sich die Einflussstärken je nach Stadt oder Jahreszeit ändern, wurden sowohl die soziostrukturellen Daten als auch die polizeilichen Vorgangsdaten auf ihre entsprechenden Einflüsse als Variablen im Modell überprüft. Ein wichtiges Ziel dieser Analyse war, aus der Vielzahl von vorhandenen Attributen eine kleinere Auswahl herauszufiltern, der im jeweiligen Modell der größte Einfluss unterstellt werden konnte. Entsprechend wurde die Einflussstärke der verwendeten Attribute ermittelt und verglichen. Zusätzlich wurden unterschiedliche Modelle und deren Güte getestet und verglichen. Diese Überprüfung ermöglichte im Prozess die kontinuierliche Verbesserung der Modelle, da diese dadurch feiner modelliert werden konnten. Im Folgenden wird auf die Testung der Variableneinflüsse konkreter eingegangen.

4.5.1 Analyse soziostruktureller Daten

Die nachfolgende Zusammenfassung der Analyse von soziostrukturellen Daten als Vorhersagevariablen sowie deren Zusammenhang mit WED-Vorgängen innerhalb von Wohnquartieren der Polizeibezirke umfasste folgende Analyse-schritte:

1. Überprüfung der Zusammenhangsstärke der Variablen untereinander und dem Einbruchgeschehen in Wohnquartieren mittels Korrelationsmatrizen
2. Test, Vergleich und Optimierung unterschiedlicher Modellansätze, insbesondere des CART-Entscheidungsbaummodells und Random-Forest-Modells
3. Identifizierung von Variablen, die die höchste Rangordnung und somit größte Bedeutung (*importance*) im Entscheidungsbaummodell haben
4. Validierung generierter Modelle
5. Separate Modellgenerierung und Validierung für den gesamten Datensatz sowie für einzelne Bezirke und Jahreszeiten (Sommer/Winter) zwecks Vergleichs

Die soziostrukturellen Daten lagen auf vier unterschiedlichen räumlichen Aggregationsniveaus (Kreis-Gemeinde-Schlüsselnummern; KGS) vor:

1. **KGS8:** Gemeindeebene
(z. B. PKW-Neuzulassungen)
2. **KGS22:** Wohnquartierebene
(z. B. Kaufkraft [Mio €], Arbeitslosenquote oder die Anzahl der Haushalte mit Kindern)
3. **KGS36:** Straßenabschnittsebene
(z. B. Entfernung zur Bundesstraße in Metern oder die Affinität für Urlaubsreisen)
4. **KGS44:** Gebäudeebene
(z. B. Affinität für Wohneigentum oder Größe/Exklusivität der Wohnungen des Hauses)

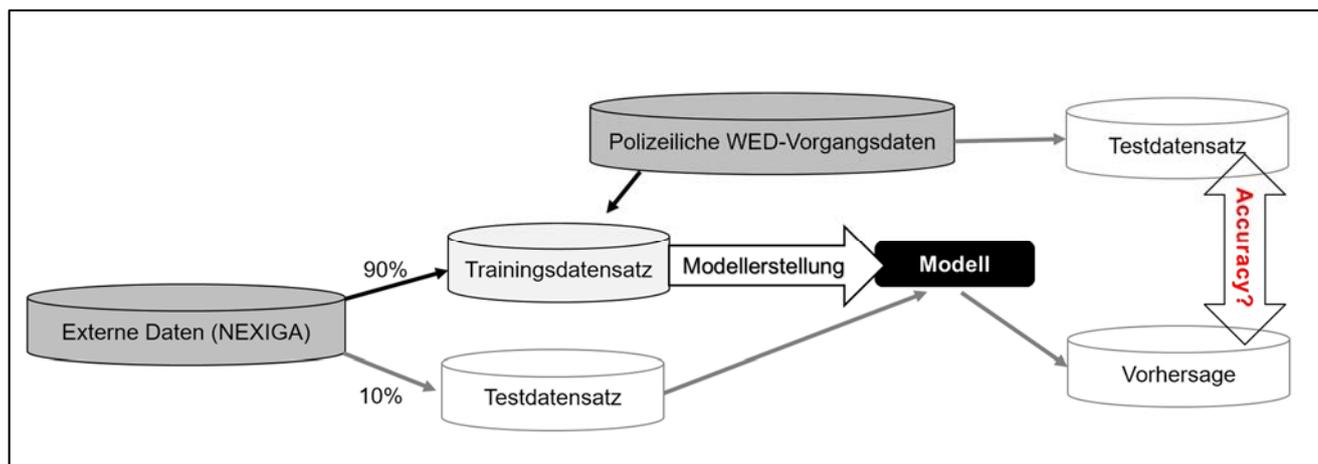
Jeder Gemeinde, jedem Wohnquartier, jedem Straßenabschnitt und jedem Haushalt sind jeweils mehrstellige Kennziffern (Schlüssel) zugeordnet. Sie sind dadurch eindeutig identifizierbar. Die Schlüssel auf Straßenabschnittsebene sind jeweils einer Hausnummer mit Koordinaten zugeordnet, die den kleinstmöglichen Abstand zu Gemeinde, Postleitzahl, Wohnquartier, Bezirk und Ortsteil hat. Die KGS22-Daten umfassen durchschnittlich 400 Haushalte mit hoher Homogenität. Die Wohnquartiergrenzen beziehen sich in diesem Zusammenhang auf frühere amtliche Stimmbezirke (vgl. Nexiga 2017). In der aktuell (Stand 2017) zur Verfügung stehenden Datenbasis variierte die Anzahl der Haushalte in einem Wohnquartier jedoch in Teilen recht stark (zwischen 1 und 997 Haushalten), sodass ein Teil der soziostrukturellen Daten in Bezug auf die Gesamtanzahl von Haushalten je Wohnquartier normalisiert werden musste. Für die Analyse wurden die vorliegenden externen soziostrukturellen Daten aus dem Jahr 2016 verwendet. Diese

bilden allerdings, aufgrund der Datenerfassungs-, Aufbereitungs- und Bereitstellungsmodalitäten und einer damit einhergehenden zeitlichen Verzögerung, das Jahr 2014 ab.

Für die durchgeführte Analyse war es zunächst wichtig, alle Daten auf dasselbe Aggregationsniveau zu skalieren. Als Analyseniveau wurde die Wohnquartierebene gewählt, da diese mit der aktuellen Modellierungsebene der im Projekt SKALA genutzten räumlichen Bezugsgrößen für Kriminalitätsprognosen übereinstimmte. Grundsätzlich bestehen je nach Datentyp (nominal, ordinal oder rational/metrisch) unterschiedliche Möglichkeiten der Datenaggregation. Datenaggregation bedeutet in diesem Fall die Hochskalierung oder Zusammenfassung von kleinräumigen Daten (z. B. von Gebäude- oder Straßenabschnittsebene) auf Wohnquartierebene. Bei rational skalierten Daten wird i. d. R. der Mittelwert oder die Summe aller Punktdaten im jeweiligen Gebiet berechnet. Hingegen ist die Mittelwertbildung nicht möglich bzw. sinnvoll, wenn es sich um nominal oder ordinal skalierte Daten handelt. Daher wurden bei ordinal skalierten Daten, z. B. der Variable „Affinität zu Wohneigentum“ von eins (sehr niedrig) bis neun (sehr hoch) oder bei nominal skalierten Daten (wie z. B. der „Familienstruktur“) bestimmte Kategorien des jeweiligen Attributes ausgewählt (z. B. „Familien mit Kindern“). Anschließend wurde die relative Anzahl dieser Ausprägungen im jeweiligen Wohnquartier ermittelt.

Abbildung 26 fasst die wesentlichen Schritte der methodischen Herangehensweise zusammen. Zentral ist die Aufteilung der polizeilichen Vorgangsdaten und der externen, soziostrukturellen Daten in einen Trainings- und einen Testdatensatz. Mit dem Trainingsdatensatz wird das Modell erstellt. Mit dem Testdatensatz wird das Modell validiert. Die Validierung und damit die Überprüfung der Modell-Performance erfolgt mittels des Accuracy Index, dessen Berechnungsgrundlage bereits zuvor erläutert wurde (siehe Kapitel 4.4.4). Bei der Validierung wird geprüft, inwieweit das anhand des Trainingsdatensatzes erzeugte Modell für die Prognose auf Basis anderer Eingangsdaten (hier: Testdatensatz der soziostrukturelle Daten) verwendet werden kann. Bei der Vorhersage handelt es sich, genau wie bei dem Testdatensatz der polizeilichen WED-Vorgangsdaten, um die Einbruchhäufigkeiten je Wohnquartier. Der hier verwendete Testdatensatz (sowohl soziostrukturelle Daten als auch polizeiliche WED-Vorgangsdaten) umfasste zehn Prozent des Gesamtdatensatz in Bezug auf die Gesamtanzahl vorhandener Wohnquartiere. Bei dem Modell handelt es sich um das jeweils getestete Modell, wie beispielsweise das Entscheidungsbaummodell.

Abbildung 26: Datenaufteilung für Modellerstellung und Validierung mittels Accuracy Index



Die gesamte Analyse wurde automatisiert auf Basis von Python durchgeführt. Dies geschah unter Verwendung gängiger Zusatzpakete, einschließlich Pandas (<http://pandas.pydata.org/>) und Scikit-learn (<http://scikit-learn.org/>).

Der statistische Zusammenhang der Daten wurde mittels des Spearman Rangkorrelationskoeffizienten (r_{SP}) untersucht. Starke Korrelationen der soziostrukturellen Daten als Vorhersagevariablen untereinander können hierbei bei der Modellierung und der Ergebnisinterpretation von Nachteil sein. Zum einen erschweren untereinander korrelierende Vorhersagevariablen die Interpretation des Entscheidungsbaummodells und der Analyseergebnisse, zum anderen besteht bei einer sehr großen Anzahl von Vorhersagevariablen, insbesondere wenn diese untereinander korrelieren, die Gefahr des *overfittings* im Modell. Bei der Modellierung wird daher stets empfohlen, eine möglichst geringe Anzahl an Eingabevariablen zu verwenden, die einen möglichst starken Einfluss auf die Zielvariable haben. Bei diversen Testläufen zeigte sich auch in dieser Analyse, dass die Verwendung einer maximalen Variablenanzahl einschließlich wenig einflussreicher Attribute nicht zu einer Verbesserung der Vorhersagekraft des Entscheidungsbaumes führte. In der Folge wurden in einem mehrstufigen Prozess (Korrelationsanalysen, Variablenauswahl und Modellentwicklung) schrittweise wenig einflussreiche und stark miteinander korrelierende Variablen gezielt von der Analyse ausgeschlossen. Dies führte zu einer Reduktion der Merkmale, für die, unter Verwendung zugrundeliegender und wissenschaftlicher Erkenntnisse (siehe Kapitel 3.2), ein Zusammenhang mit polizeilichen WED-Vorgängen als möglich und plausibel eingestuft wurde. Die Anzahl reduzierte sich konkret auf elf Variablen, die für die weitere Verwendung in Entscheidungsbaummodellen zur Prognose von Einbruchwahrscheinlichkeiten für die Pilotbehörden als sinnvoll erachtet wurden (siehe Tabelle 5). Es sei explizit darauf hingewiesen, dass diese Auswahl weder fest noch allgemeingültig ist, sondern stark von der Anzahl, der Auswahl und der Aufbereitung der

verwendeten Daten abhängt. Daher ist nicht gänzlich auszuschließen, dass andere der insgesamt zur Verfügung stehenden Variablen ebenfalls einen Einfluss auf die Ergebnisse des Entscheidungsbaumes haben können.

Einschränkend muss an dieser Stelle zudem angemerkt werden, dass für die Analyse soziostrukturelle Daten verwendet wurden, die für das Jahr 2016 bereitgestellt wurden. Auch wenn die Ergebnisse keine gravierenden Unterschiede zu Daten aus 2015 oder 2017 erwarten lassen, wären die Analyseergebnisse bei einem anderen Datensatz sicherlich nicht identisch. Ebenso kann sich die Art und Weise, wie Datenlücken behandelt werden, auf die Ergebnisse auswirken. Für solche Problemstellungen wäre es deshalb möglich, zunächst Klassifizierungsmodelle zu nutzen, um Datenlücken zu schließen, bevor die eigentliche Analyse durchgeführt wird. Die Anzahl der Datenlücken wurde allerdings im Verhältnis zu der insgesamt großen Datenmenge als gering eingestuft, sodass nicht von einem großen Einfluss auf die Analyseergebnisse ausgegangen wurde, der diesen zeitlichen Mehraufwand gerechtfertigt hätte. Den größten Einfluss auf die Analyseergebnisse hat vermutlich die Art und Weise, wie die Gebäude- und Straßenabschnittsspezifischen Daten auf die Wohnquartierebene skaliert werden (hier insbesondere die Auswahl an Kategorien ordinal skalierten Attribute für die Berechnung des wohnquartierspezifischen Attributwertes in Form der relativen Anzahl entsprechender Merkmalsausprägungen).

Tabelle 5: Ausgewählte Attribute soziostruktureller Daten als Prädikatoren für eine Modellierung

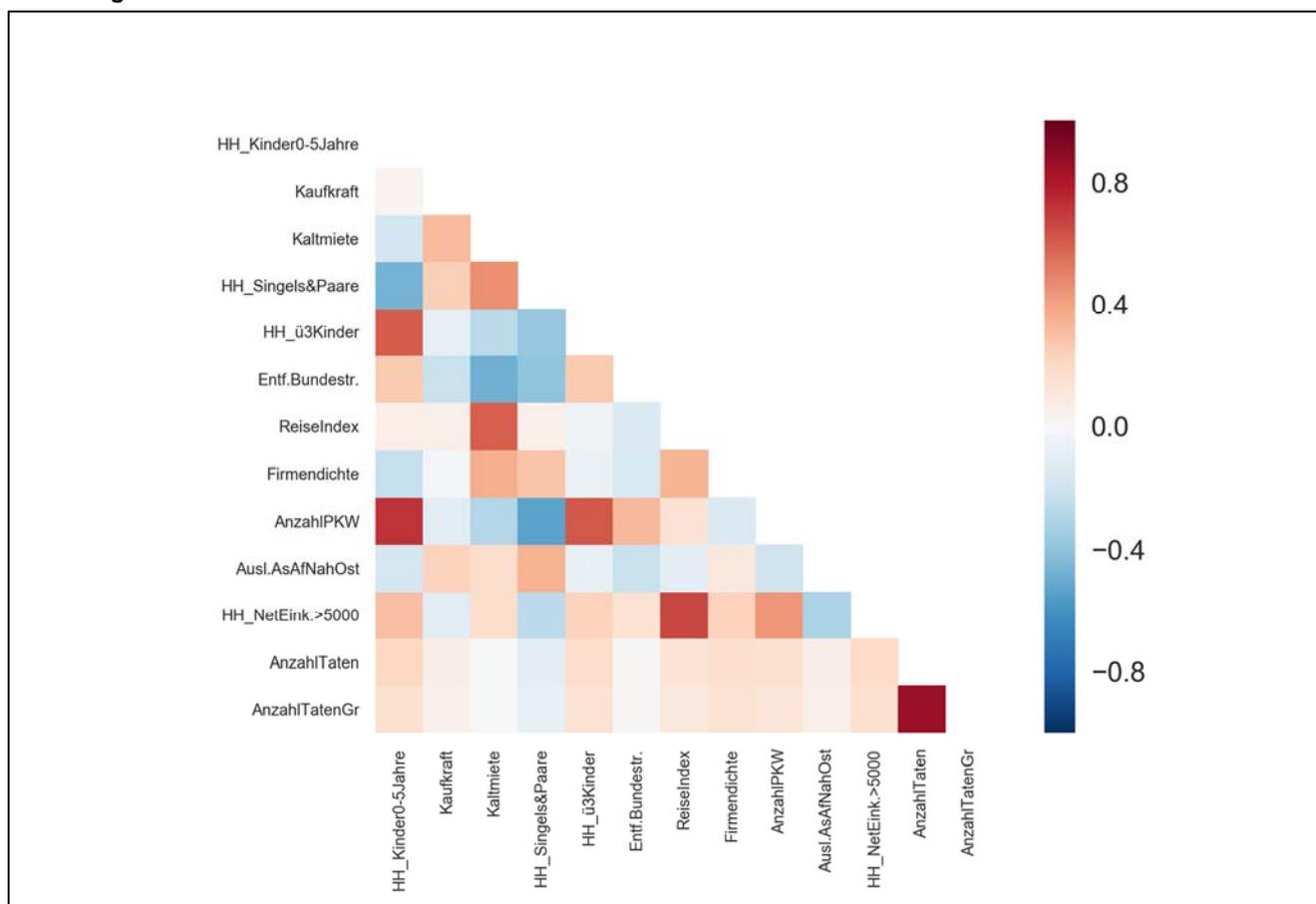
| Variablenkürzel [Einheit] | Erläuterung |
|--|--|
| HH_Kinder0-5Jahre [Anzahl/HH _{wq}] | Anzahl der Haushalte (HH) mit Kindern im Alter von 0-5 Jahren (anteilig an der Gesamtanzahl von Haushalten im Wohnquartier [HH _{wq}]) |
| Kaufkraft [Mio. €] | Durchschnittliche Kaufkraft der Straßenabschnitte im Wohnquartier in Millionen € |
| Kaltmiete [€] | Durchschnittliche Nettokaltmiete im Wohnquartier in € |
| HH_Singels&Paare [Anzahl/HH _{wq}] | Familienstruktur der Bewohner des Hauses (KGS44): Anzahl der Haushalte (HH) der Kategorien 1 (Singles), 2 (vorwiegend Singles), 3 (Paare) und 4 (vorwiegend Paare) pro Gesamtanzahl von HH im Wohnquartier (HH _{wq}) |
| HH_ü3Kinder [Anzahl/HH _{wq}] | Anzahl der HH mit drei oder mehr Kindern (anteilig an der Gesamtanzahl von HH im Wohnquartier) |
| Entf.Bundestr [m] | Durchschnittliche Entfernung zur Bundesstraße in Metern |
| ReiselIndex | Affinität für Urlaubsreisen als Index (Berechnungsgrundlage unbekannt) |
| Firmendichte [Anzahl/km ²] | Firmendichte (Anzahl Firmen pro km ²) |
| AnzahlPKW [Anzahl/ HH _{wq}] | Anzahl (gemeldeter) PKW anteilig pro HH _{wq} |
| Ausl.AsAfNahOst [Anzahl/ HH _{wq}] | Ausländer nach Herkunft (Asien, Afrika, Naher Osten) pro HH _{wq} |
| HH_NetEink.>5000 [€] | Anzahl der HH mit Nettoeinkommen \geq 5.000€ |

Im Zusammenhang mit der Vorhersagbarkeit von Einbruchhäufigkeiten auf Basis der soziostrukturellen Daten wurden, unter Verwendung der ausgewählten Attribute, unterschiedliche Modelle und spezifische Modelleinstellungen getestet. Hierfür wurde die rational skalierte Zielvariable „AnzahlTaten“ (Gesamtanzahl der polizeilichen WED-Vorgänge im Zeitraum 2012 bis 2016) in die Variable AnzahlTatenGr mit zwei Gruppen („0“ \rightarrow „wenig Vorgänge“ und „1“ \rightarrow „viele Vorgänge“) kategorisiert. Der jeweilige Grenzwert zur Einteilung der Wohnquartiere in diese Gruppen war abhängig von dem jeweils eingehenden Datensatz. Für die sommer- und winterspezifische Analyse wurden Grenzwerte von neun bzw. 15 WED-Vorgängen gewählt. Der Grenzwert war im Sommer niedriger als im Winter, da zu dieser Jahreszeit deutlich weniger WED-Vorgänge im Datensatz vorhanden waren. Für den Gesamtdatensatz (Sommer und Winter) wurden alle Wohnquartiere in Gruppe „0“ (wenig Taten) eingestuft, wenn im Gesamtzeitraum weniger als 23 WED-Vorgänge vorkamen. Die Einstufung in Gruppe „1“ (viele Taten) erfolgte dementsprechend, wenn 23 oder mehr WED-Vorgänge vorkamen. Diese als geeignet erachtete Definition der Grenzwerte wurde anhand mehrerer Testläufe über den Zeitraum der Jahre 2012 bis 2016 festgelegt. Ausschlaggebend war u. a., dass die Grenzwerte sowohl für die Analyse des Gesamtdatensatzes (alle Bezirke zusammen) als auch für die stadtspezifische Analyse infrage kamen. Der Sommerdatensatz beinhaltete Daten der Monate April bis September, der Winterdatensatz Daten die Monate Oktober bis März.

Als Entscheidungsbaum wurde der CART-Algorithmus als Klassifizierungsmodell verwendet (siehe Kapitel 4.4.1). Die

Validierung erfolgte mittels einer sechsfachen Kreuzvalidierung unter Verwendung des Validierungsindex Acv , der sich auf die Accuracy bezieht (siehe Kapitel 4.4.4).

Abbildung 27 verdeutlicht in diesem Kontext die Zusammenhangsstärke ausgewählter Attribute der soziostrukturellen Daten (siehe Tabelle 5) untereinander sowie mit der Anzahl der WED-Vorgänge je Wohnquartier für den Gesamtdatensatz (alle Pilotbezirke; Sommer und Winter).

Abbildung 27: Korrelationsmatrix: Soziostrukturelle Daten und Anzahl WED im Zeitraum 2012-2016

Erläuterung: Korrelationsmatrix mit Spearman Rangkorrelationskoeffizient (r_{Sp}) für die Auswahl soziostruktureller Daten und Anzahl der WED-Vorgänge (AnzahlTaten) im Zeitraum 2012-2016 pro Wohnquartier für den Gesamtdatensatz aller Pilotbezirke.

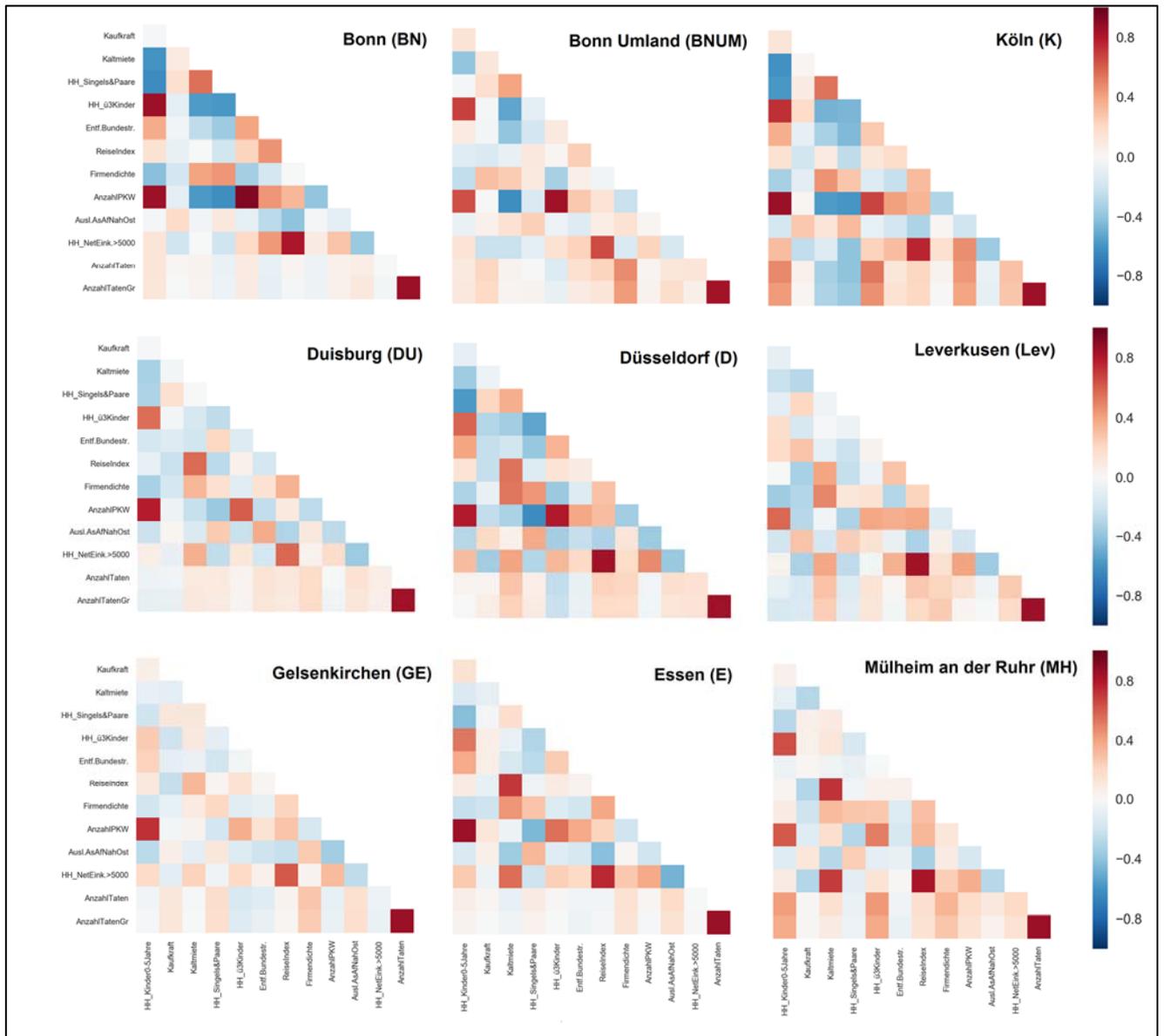
Ein Rangkorrelationskoeffizient r_{Sp} von +1 deutet auf eine perfekt positive Korrelation hin (je höher Variable 1, desto höher Variable 2). Ein r_{Sp} -Wert von -1 wäre eine perfekt negative Korrelation (je höher Variable a, desto niedriger Variable b). Ein Koeffizient von 0 (weißer Bereich der Werteskala) bedeutet, dass kein linearer Zusammenhang des jeweiligen Variablenpaars besteht.³⁸

Im Ergebnis zeigt Abbildung 27 eine starke Korrelation einiger Variablen untereinander. Beispielsweise korreliert die durchschnittliche Gesamtanzahl an PKW in einem Wohnquartier (AnzahlPKW) sehr stark mit der relativen Anzahl an Haushalten mit Kindern im Alter von null bis fünf Jahren im selben Wohnquartier (HH_Kinder_0-5Jahre). Gleichzeitig korreliert die durchschnittliche Gesamtanzahl von Haushalten mit einem Haushaltsnettoeinkommen ≥ 5.000 Euro (HH_NetEink.>5000) recht stark mit dem „ReiseIndex“. Bei der Interpretation dieser Ergebnisse gilt es zu beachten,

dass aus hohen Zusammenhangesstärken keine kausalen Zusammenhänge zwischen Variablen abgeleitet werden können. Wie aus Abbildung 27 hervorgeht, korrelieren „AnzahlTaten“ und „AnzahlTatenGr“ wie erwartet sehr stark, da es sich hier um dieselbe Zielvariable handelt (im Falle von „AnzahlTatenGr“ um die kategorisierte/gruppierte Variante von „AnzahlTaten“).

Insgesamt zeigt sich zunächst eine geringe Zusammenhangsstärke zwischen den soziostrukturellen Variablen und dem WED (Zielvariable). Legt man den Fokus aber differenziert auf die Polizeibezirke, betrachtet also die einzelnen Bezirke unabhängig voneinander, zeichnet sich in den Korrelationsmatrizen ein differentes Bild ab. Die Korrelationsmatrizen der Zusammenhangsanalyse für die einzelnen Bezirke (siehe Abbildung 28) unterscheiden sich danach teilweise deutlich von der in Abbildung 27 dargestellten Matrix.

³⁸ Da die Berechnung der Signifikanz direkt von der Anzahl der Daten abhängt, sind in vielen Fällen allein aufgrund der großen Datenmengen entsprechende Korrelationen möglich.

Abbildung 28: Korrelationsmatrix: Soziostrukturelle Daten und Anzahl WED im Zeitraum 2012-2016 nach Bezirken

Erläuterung: Korrelationsmatrix mit r_{Sp} für die Auswahl soziostruktureller Daten und Anzahl der WED-Vorgänge (Anzahl-Taten) im Zeitraum 2012-2016 pro Wohnquartier nach Bezirken.

So ist die Korrelation zwischen einigen soziostrukturellen Daten und WED (Zielvariable: AnzahlTatenGr) in einigen Bezirken stärker als dies in Abbildung 27 deutlich wird (Köln, Mülheim an der Ruhr und das Bonner Umland). Bei anderen Bezirken fällt diese Korrelation dagegen ähnlich oder geringer aus. Darüber hinaus unterscheiden sich die Korrelationen der Variablen auch untereinander, wenn nach den Bezirken entsprechend separiert analysiert wird. So ist beispielsweise die Korrelation zwischen „HH_Kinder_0-5Jahre“ und „HH_ü3Kinder“ in Gelsenkirchen und Leverkusen deutlich geringer als bei allen anderen Bezirken. Die separate Betrachtung der Zusammenhänge spricht somit dafür, perspektivisch die Modellierung von Kriminalitätsprognosen separat für einzelne Bezirke bzw. zusammengefasst für ähnlich strukturierte Bezirke durchzuführen. Weiterhin

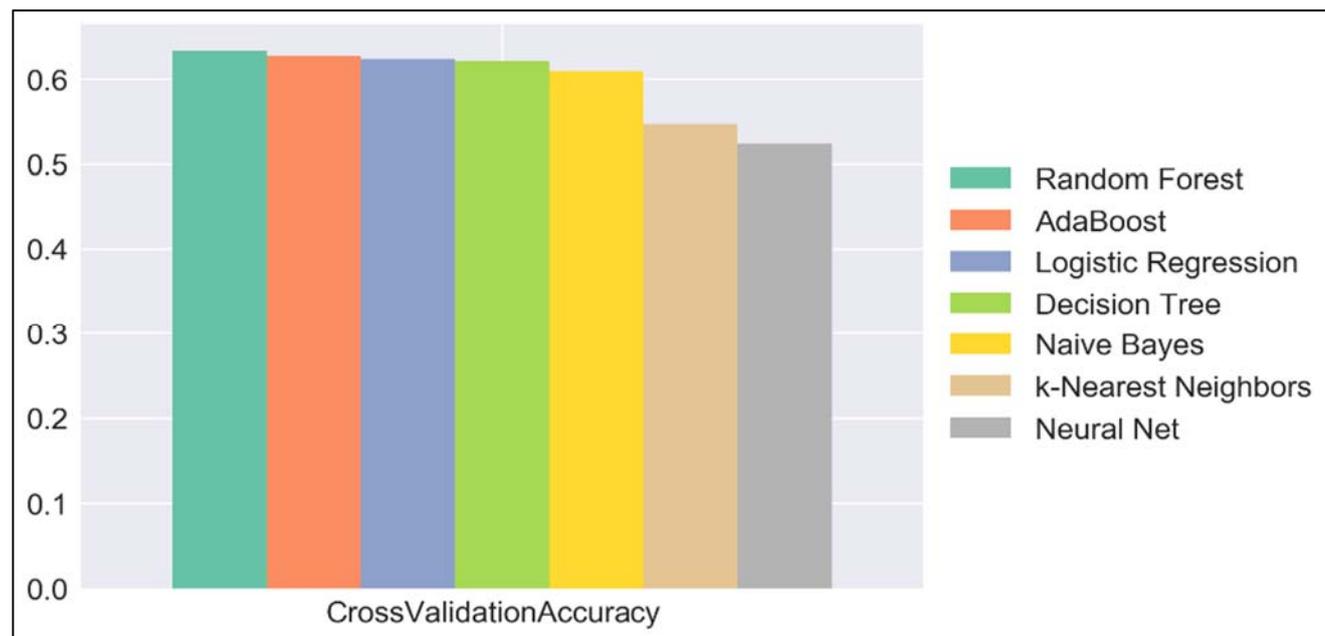
wurde bei der Analyse deutlich, dass die Zusammenhangsstärke einzelner Variablen deutlich davon abhängig war, ob der Datensatz für das ganze Jahr oder nur für die Sommer- oder Wintermonate verwendet wurde. Wurden z. B. nur Daten der Wintermonate einbezogen, so zeigte sich ein deutlich höherer r_{Sp} für einen Großteil der Attribute als dies im Sommer der Fall war. Dieser fiel für die Wintermonate sogar geringfügig höher aus als für den Gesamtdatensatz. Bei einigen Pilotbehörden variierte der r_{Sp} sehr stark in Abhängigkeit vom jeweiligen Attribut. Grundsätzlich war der r_{Sp} jedoch im Winter höher als im Sommer. Diese Erkenntnis deckt sich mit der bisherigen kriminologischen Forschung zum WED und der saisonalen Einbruchphänomenologie. Das Prognosemodell wurde deshalb entsprechend angepasst, wobei eine Modell- und Prognoseerstellung mittels Zeitscheiben jahreszeitlich angepasst durchgeführt wurde.

4.5.2 Analyse der Modellperformance

Im Zusammenhang mit der Güte des Prognosemodells bzw. der Modellperformance zeigt Abbildung 29 anhand der Cross Validation Accuracy (A_{cv}) die Performance der unterschiedlichen getesteten Modelle. Die Modellperformance ähnelt sich danach sehr stark. Im Falle des Gesamtdatensatzes (alle Pilotbehörden und das gesamte Jahr) erreichte das Random-Forest-Modell die höchste Accuracy (0,63), die jedoch nur geringfügig höher war als bei einem

Großteil der anderen getesteten Modelle, u. a. dem einfachen CART-Algorithmus (*decision tree*) mit einer Accuracy von 0,62. Da die Performance des Random-Forest-Modells in den meisten Fällen geringfügig höher war als bei dem einfachen Entscheidungsbaummodell und das Modell insgesamt stabiler ist als ein einfacher Entscheidungsbaum (vgl. Maimon/Rokach 2015: 99ff.), beziehen sich die im Folgenden zusammengefassten Ergebnisse immer auf das Random-Forest-Modell.

Abbildung 29: Vergleich Modellperformance anhand der Cross Validation Accuracy



Ähnlich wie bei der Zusammenhanganalyse variierte die Modellperformance geringfügig zwischen den Bezirken und in Abhängigkeit davon, welche Jahreszeit berücksichtigt wurde. Im Falle des Datensatzes aller Bezirke war die Modell-Accuracy im Winter ($A_{cv} = 0,66$) geringfügig höher als im Sommer ($A_{cv} = 0,61$). Für das gesamte Jahr war die Accuracy bei den meisten Bezirken geringfügig höher (Düsseldorf: 0,67; Duisburg: 0,68; Essen: 0,64; Köln: 0,73). Dies spricht erneut dafür perspektivisch die Modellierung von Kriminalitätsprognosen separat für einzelne Bezirke bzw. zusammengefasst für ähnlich strukturierte Bezirke durchzuführen. Bei Mülheim an der Ruhr ($A_{cv} = 0,60$), Bonn ($A_{cv} = 0,56$), Leverkusen ($A_{cv} = 0,57$) und Gelsenkirchen ($A_{cv} = 0,55$) war die Modellperformance niedriger als bei der Modellevaluierung für den Gesamtdatensatz. Dies hing mit großer Wahrscheinlichkeit damit zusammen, dass die Datenmenge dieser Bezirke geringer war als bei den anderen (größeren) Bezirken. Wurde dagegen das für alle Bezirke erstellte Modell einzeln für diese angewendet und evaluiert, war die A_{cv} für Mülheim an der Ruhr ($A_{cv} = 0,64$), Bonn ($A_{cv} = 0,57$) und Leverkusen ($A_{cv} = 0,61$) geringfügig höher als bei Anwendung der bezirkspezifisch erstellten Modelle. Diese Erkenntnis spricht dafür, den Trainingsdatensatz für

kleinere Bezirke bei der Modellerstellung zusammenzufassen, entweder in Form der bereits bestehenden polizeilich-organisatorischen Zuordnung (Köln und Leverkusen, Essen und Mülheim, Bonn mit Umland) oder aber zusammengefasst mit allen Bezirken.

Abbildung 30 stellt die relative Einflussstärke (*importance*), kurz Imp, der ausgewählten Attribute beim Random-Forest-Modell für den Gesamtdatensatz aller Bezirke dar. Der Imp-Wert ist unmittelbar abhängig von der Anzahl eingehender Attribute, da die Imp-Summe aller Attribute 1 ergibt. Entsprechend ist nur das relative Verhältnis der Imp-Werte ausschlaggebend, nicht der absolute Wert. Je höher der Imp-Wert eines Attributes im Vergleich zu den anderen Attributen ist, desto höher dessen Einflussstärke. Im Falle von Entscheidungsbaummodellen, einschließlich Random Forest, bedeutet eine hohe Einflussstärke, dass das jeweilige Attribut relativ weit oben in der Hierarchie des Entscheidungsbaumes (das heißt näher an dessen Wurzel) als Entscheidungsknoten ausgewählt wird. Das jeweilige Attribut spielt dann eine sehr große Rolle in Bezug auf die Klassifizierung der Zielvariablen. Abbildung 30 verdeutlicht, dass die Einflussstärke der einzelnen Variablen, in Abhängigkeit vom je-

weils für die Modellerstellung gewählten Datensatz (Gesamtdatensatz aller Pilotbehörden oder nur Daten der Sommer- oder Wintermonate), sehr stark variiert. In allen Fällen hatten die Attribute „Firmendichte“, „HH_Kinder0-5Jahre“, „HH_NetEink.>5000“, „HH_ü3Kinder“ und der „ReiseIndex“ eine sehr große Einflussstärke. Die einflussreichste Variable

war für das Gesamtjahr „HH_ü3Kinder“, im Sommer „Ausl.AsAfNahOst“ und im Winter „HH_NetEink.>5000“.

Abbildung 30: Einflussstärke der soziostrukturellen Daten beim Random-Forest-Modell (Gesamtdatensatz)

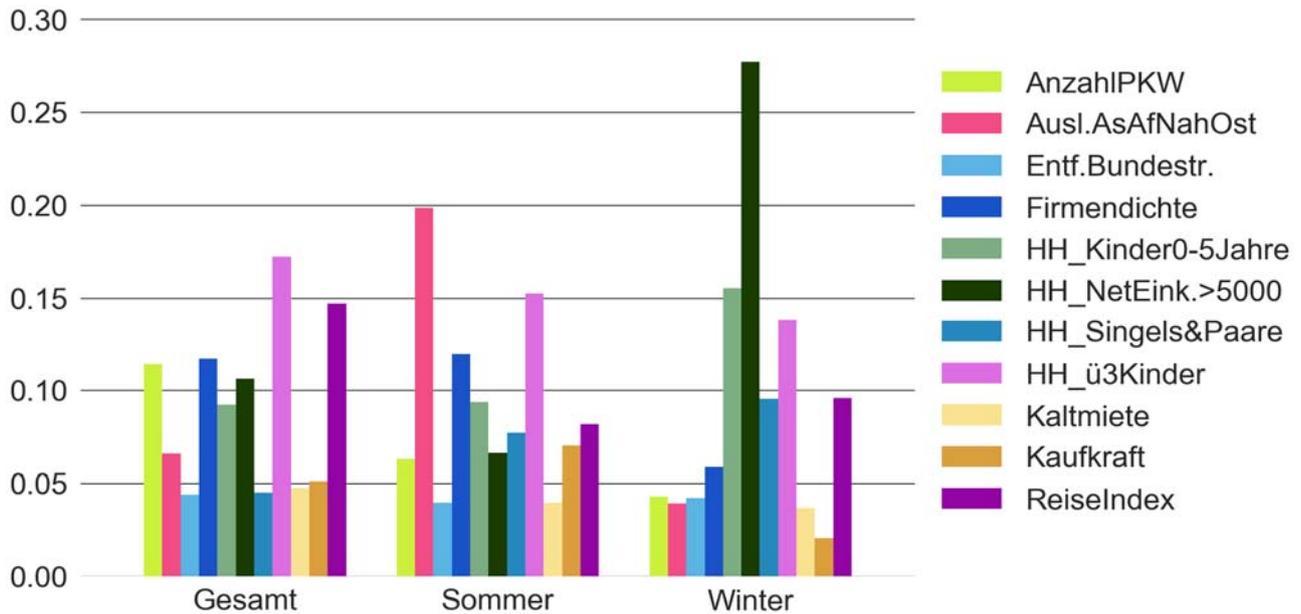
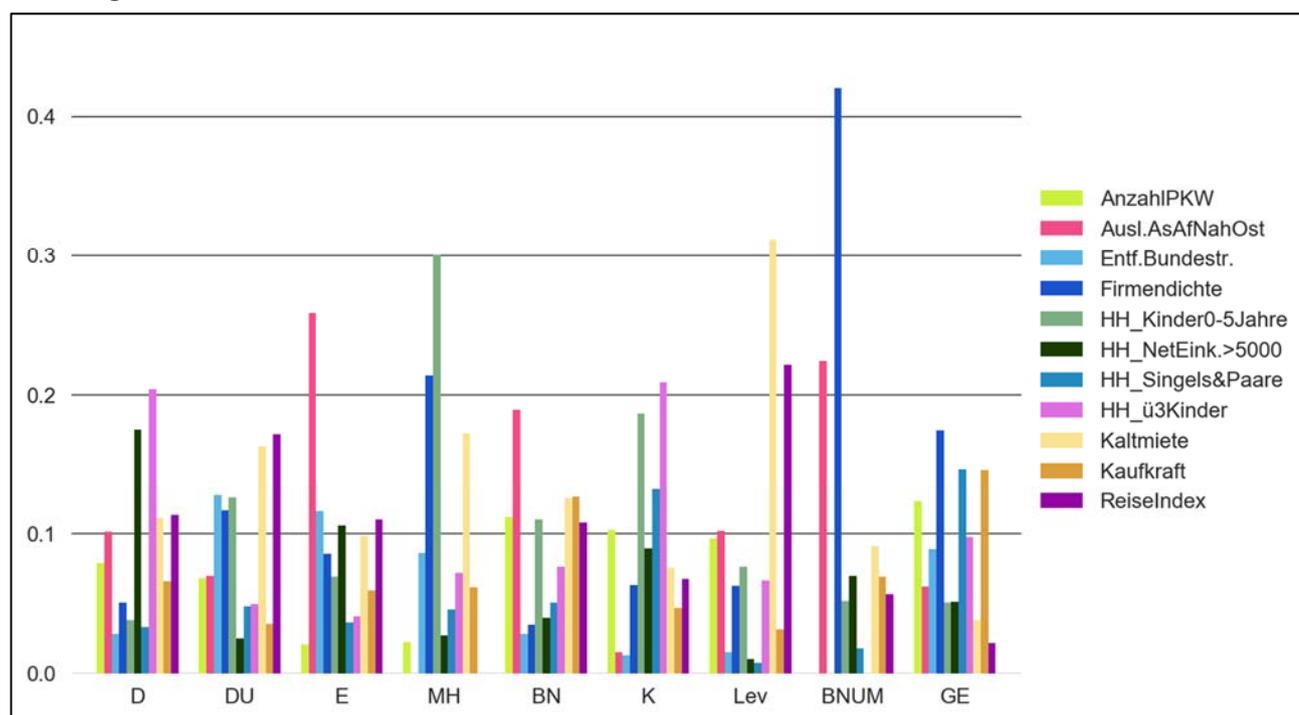


Abbildung 31 zeigt dagegen im Kontrast die Einflussstärke der ausgewählten Attribute für die einzelnen Bezirke. Auch hier variiert das Verhältnis der Imp-Werte sehr stark von Bezirk zu Bezirk. Das bedeutet, dass die ausschlaggebendsten Attribute bei der Modellerstellung für jeden Bezirk sehr unterschiedlich sind. Dies gilt nicht nur für die Ergebnisse des hier verwendeten Random-Forest-Modells, sondern auch für den einfachen (CART)-Entscheidungsbaumalgorithmus.³⁹ Diese Differenzen in den Einflussstärken einzelner Attribute zeigen sich beispielsweise deutlich im Attribut der „Kaltmiete“ beim Gesamtdatensatz (siehe Abbildung 30), in der es eine eher untergeordnete Rolle einnimmt, während es bei Duisburg und Leverkusen (siehe Abbildung 31) eine der einflussreichsten Variablen ist. Das Attribut der „Firmendichte“ hat zudem beim Datensatz für Mülheim (MH), Leverkusen (Lev), das Bonner Umland (BNUM) und Gelsenkirchen (GE) den stärksten Einfluss. In Essen (E) hat darüber hinaus das Attribut „Ausl.AsAfNahOst“ den stärksten Einfluss, während das in allen anderen Bezirken (außer Düsseldorf) eine eher untergeordnete Rolle spielt.

Auch Bezirke, die zunächst intuitiv als strukturell ähnlich im Projekt SKALA eingestuft wurden (z. B. Gelsenkirchen und Duisburg) oder bei Bezirken einer KP (z. B. Essen und Mülheim), zeigt Abbildung 31 starke Unterschiede in Bezug auf die Einflussstärke der einzelnen Attribute. Auffallend ist in diesem Zusammenhang der starke Unterschied der einzelnen Einflussstärken zwischen Bonn Stadt (BN) und dem Bonner Umland (BNUM). Das heißt, dass die Analyseergebnisse zwischen BN und BNUM stark unterschiedlich sind. Der Grund ist wahrscheinlich, dass es sich beim Bonner Umland um Kleinstädte und Kommunen mit größtenteils ländlichem Charakter handelt und im Gegensatz dazu Bonn (Stadt) mit über 320.000 Einwohnern eine andere Stadt- und Infrastruktur aufweist. Perspektivisch ist es demnach sinnvoll, Polizeibezirke wie Bonn bei der Modellerstellung als auch bei der Kriminalitätsprognose entsprechend separat zu behandeln (Unterscheidung zwischen Stadt und Umland).

³⁹ Der CHAID-Algorithmus wurde im Rahmen dieser Analyse nicht getestet. Die Ergebnisse des CART-Algorithmus lassen sich hierbei aber mit großer Sicherheit entsprechend übertragen.

Abbildung 31: Einflussstärke der soziostrukturellen Daten beim Random-Forest-Modell nach Bezirken



Die Ergebnisse zeigen zusammenfassend, dass sich die Einflussstärken der Attribute je nach Jahreszeit und Polizeibezirk stark unterscheiden. Dementsprechend sind die Ergebnisse nicht automatisch übertragbar auf andere Bezirke oder Zeiträume.

Aufbauend auf dieser Erkenntnislage sollte bei einer zukünftig Modell- und Prognoseverfeinerung auf eine Trennung (bzw. inhaltlich sinnvolle Zusammenfassung) von Bezirken und Regionen methodisch fokussiert werden. Ferner ist es sinnvoll, die Modell- und Prognoseerstellung mit der gesamten Auswahl, der im Rahmen dieser Analyse als relevant berechneten Attribute, durchzuführen und anschließend dem Algorithmus die Entscheidung zu überlassen, in welcher Art und Weise die Attribute als Entscheidungsknoten ausgewählt werden.

4.6 RZC-Analyse und Scoreberechnung

Im Rahmen der ersten Modell- und Prognoseerstellung stellte sich bereits anfangs im Projekt SKALA heraus, dass die „Tage seit dem letzten Einbruch“ im Quartier eine anscheinend wichtige Einflussvariable darstellen. Konkret: Je geringer der zeitliche Abstand zwischen zwei Einbrüchen war, desto höher die Wahrscheinlichkeit weiterer Einbrüche im entsprechenden Wohnquartier. Da die Berechnung auf der Basis von Quartieren erfolgte und folglich solche Einbrüche bei der Berechnung nicht berücksichtigt werden konnten, die außerhalb dieser Wohnquartiergrenzen lagen, wurde eine Raum-Zeit-Cluster(RZC)-Analyse konzipiert und durchgeführt. Diese ermöglichte die Identifizierung von Clustern räumlich und zeitlich naher WED-Vorgänge unabhängig von Quartiergrenzen. Ziel war es entsprechende

Scores zu bilden, die später in das Entscheidungsbaummodell als zusätzliche Einflussvariablen einfließen konnten.

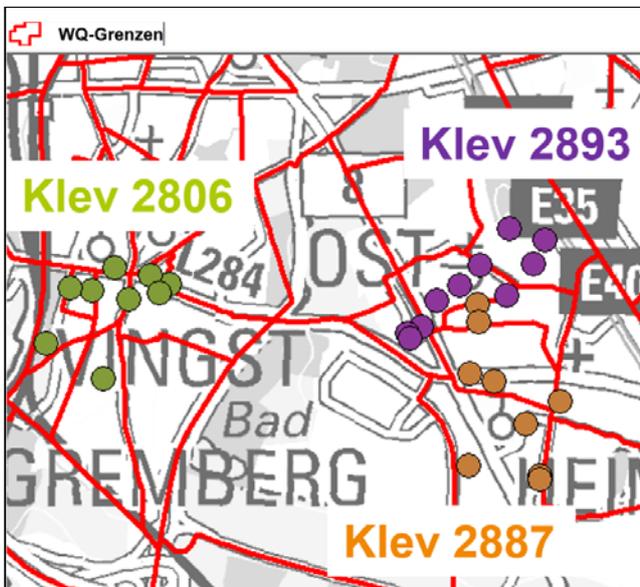
Im Kontext der Festlegung auf eine Definition von räumlich-zeitlicher Nähe ermöglichten Pretests, dass eine maximale räumliche Distanz von 300 Metern und ein maximaler zeitlichen Abstand von 14 Tagen eine sinnvolle definitorische Festlegung sind. Andere räumliche und zeitliche Abstände wurden ebenfalls getestet (z. B. 7 Tage und 500 Meter, 3 Monate und 500 Meter). Der maximale zeitliche und räumliche Abstand für die Definition eines Near-Repeat-Phänomens kann nicht genau festgelegt werden. Dies wird auch bei Gluba et al. (2015) deutlich, die Near Repeats für räumliche Abstände von 0-400 Metern und zeitliche Abstände von 0-27 Tagen analysieren. Allerdings zeigte sich sowohl bei Gluba et al. (2015) als auch bei Analysen im Rahmen des Projektes SKALA, dass es sich bei einem Großteil der Near Repeats um Vorgänge mit sehr kurzen räumlichen Distanzen (<300 Meter) und sehr kurzen Zeitabständen (<7 Tage) handelt. Zudem liegen Vorgänge mit räumlichen Distanzen von >300 Metern häufig nicht innerhalb derselben Wohnblocks, sodass ein Zusammenhang dieser Taten als weniger wahrscheinlich eingestuft wird. Daher wurde die Wahl von 300 Meter zur Abgrenzung eines Near Repeats befürwortet. Hingegen wird die Wahl des zeitlichen Abstands als flexibler eingestuft und davon ausgegangen, dass die Definition von 7 statt 14 Tagen auch geeignet wäre. Es ist zu berücksichtigen, dass mit Zunahme des gewählten Zeitabstandes die Größe der Raum-Zeit-Cluster (RZ-Cluster) zunimmt. Durch die Wahl eines größeren Zeitabstandes würden RZ-Cluster berechnet, die sich über verhältnismäßig längere Zeiträume erstrecken, wobei Taten zu Beginn der

Clusterserie womöglich nicht in Zusammenhang mit Vorgängen am Ende des Clusters stehen. Hingegen wird bei der Wahl eines kürzeren Zeitabstandes von maximal 14 Tagen eine höhere Wahrscheinlichkeit vermutet, dass die Vorgänge innerhalb eines Clusters teilweise in Zusammenhang stehen.

Die RZC-Analyse und Score-Berechnung wurde vollautomatisiert mittels eigens verfasster Microsoft SQL-Server Abfragen (Transact-SQL) und R-Skripte durchgeführt. Durch die Automatisierung der Clusteranalyse war es schnell möglich, neue raum-zeitliche Abstände zu definieren.

Die Clusteranalyse ermöglicht es weiterhin, über die Wohnquartiergrenzen hinaus, raum-zeitlich zusammenhängende WED-Vorgänge zu identifizieren und sodann entsprechend zu gruppieren. Dies wird exemplarisch anhand von Abbildung 32 deutlich, in dem sich die drei dargestellten RZ-Cluster über mehrere Wohnquartiere erstrecken.

Abbildung 32: Illustration Raum-Zeit-Cluster



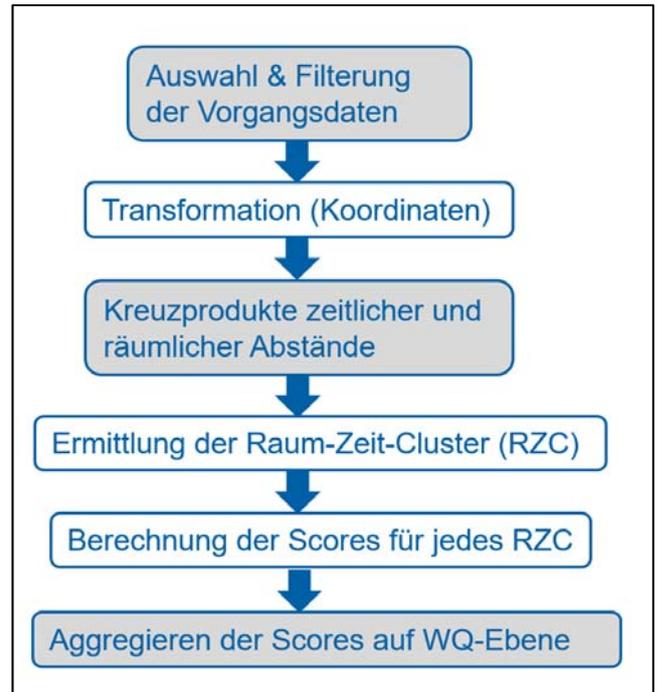
Finalisiert wurde die Analyse, indem jedem WED-Vorgang eine RZC-ID zugeordnet wurde. Jedes raum-zeitlich abgegrenzte Cluster ist durch unterschiedliche Merkmale gekennzeichnet, wie etwa die Anzahl der WED-Vorgänge des jeweiligen RZ-Cluster, der Dauer (Tage) des RZ-Cluster oder der Aktualität (letztes Tatzeitdatum im RZ-Cluster).

Die so berechneten RZC-Scores wurden jede Woche neu berechnet, um die Aktualität des Einbruchgeschehens stets adäquat abbilden zu können und flossen anschließend unmittelbar in die Modell- und Prognoseerstellung ein. Der mögliche Mehrwert der RZC-Scores wurde, wie in Kapitel 4.7 beschrieben, evaluiert.

4.6.1 Durchführung

Die einzelnen Analyseschritte zur Ermittlung der RZC-Scores werden in Abbildung 33 dargestellt.

Abbildung 33: Analyseschritte Ermittlung RZC-Scores



In die Berechnung der RZC-Scores wurden im ersten Schritt nur Daten einbezogen, die einen Geoqualitätsmarker (GQM) von „1“, „2“ oder „9“ im polizeilichen Vorgangsbearbeitungssystem IGVP aufwiesen. Das bedeutet, die räumliche Information in Form von Geokoordinaten wurde vom System als verlässlich eingestuft. Des Weiteren wurden nur Vorgangsdaten in der Analyse verwendet, bei welchen der Zeitabstand zwischen „Tatzeit-Beginn“ und „Tatzeit-Ende“ maximal 72 Stunden betrug, um statistische Ausreißer zu negieren. Hierunter fielen z. B. Einbrüche, deren Tatzeitraum wegen Urlaub oder sonstiger längerer Abwesenheit der Bewohner, nicht genau eingegrenzt werden konnte.

Im zweiten Schritt wurden die in der Vorgangsdatenbank enthaltenen Koordinaten vom Projektionsformat „ETRS 1989 UTM Zone 32N (EPSG: 25832)“ in das Format „WGS 1984 (EPSG: 4326)“ umgewandelt, das für räumliche Berechnungen mittels SQL-Server und R benötigt wurde.

Daraufhin (dritter Schritt) wurden Kreuzprodukte der zeitlichen und räumlichen Abstände jedes einzelnen Vorgangs mit jedem anderen Vorgang berechnet und anschließend alle Vorgangspaare gefiltert, die eine maximale räumliche Distanz von 300 Metern und einen maximalen zeitlichen Abstand von 14 Tagen aufwiesen. Alle WED-Vorgänge, bei denen nicht mindestens ein weiterer WED-Vorgang in zeitlicher und räumlicher Nähe vorhanden war, waren kein Teil eines RZ-Clusters und wurden somit bei der Analyse nicht weiter berücksichtigt. Um den zeitlichen Abstand von zwei

Vorgängen zu berechnen, wurde der „Tatzeit-Ende“-Zeitpunkt gewählt. Um die Berechnung zu beschleunigen, wurde das Kreuzprodukt separat für jede Pilotbehörde berechnet (Verkürzung der Berechnungszeit von ca. zwölf Minuten auf ein bis zwei Minuten). Die Ergebnistabelle des Kreuzproduktes wurde auf der SKALA-Datenbank gespeichert. Sie enthielt alle WED-Vorgangspaare mit einer maximalen räumlichen Distanz von 300 Metern und einem maximalen zeitlichen Abstand von 14 Tagen, einschließlich der zugehörigen Daten, die für die RZC-Analyse benötigt wurden. Diese Tabelle wird vom R-Analyseskript geladen, das die folgenden Schritte durchführt:

Zur Ermittlung der RZ-Cluster (vierter Schritt) wurden räumlich und zeitlich abgeschlossene RZ-Cluster durch mehrere Schleifen iterativ mit einer individuellen RZC-ID versehen, wobei zusammengehörenden Vorgängen dieselbe RZC-ID zugeordnet wurde. Das R-Script enthielt am Ende eine Korrekturschleife und eine Kontrollfunktion, um sicherzustellen, dass jede RZC-ID nur einmal vergeben wurde und dass sich kein RZ-Cluster zeitlich und räumlich mit einem anderen RZ-Cluster überschneidet.

Konkret wurde jedes RZ-Cluster anhand der folgenden Merkmale berechnet:

- **DauerT: Dauer**, wie lange ein RZ-Cluster besteht (Zeitlicher Abstand zwischen dem ersten und letzten Vorgang im RZ-Cluster [Tage])
- **NR_{VinRZC}: Anzahl** der Vorgänge eines RZ-Cluster
- **Freq: Frequenz** von WED-Vorgängen während der Dauer eines RZ-Cluster (NR_{VinRZC} dividiert durch DauerT)
- **RZC-Ende: Datum des letzten Vorgangs** im RZ-Cluster
- **ZIV: Zeitraum [Tage] zwischen RZC-Ende** und aktuellem Analysedatum

Zur Berechnung der RZC-Scores (fünfter Schritt) wurden für jedes RZ-Cluster normalisierte Faktoren für die Dauer, Frequenz und Aktualität eines RZ-Cluster berechnet, die jeweils Werte zwischen 0 und 1 annehmen. Der normalisierte Faktor für die Dauer eines RZ-Cluster wurde wie folgt berechnet:

$$\text{Dauer}_{\text{RZC}} = \frac{DT_i - DT_{\min}}{DT_{\max} - DT_{\min}}$$

DT_i ist die Dauer des jeweils betrachteten RZC; DT_{min} und DT_{max} sind der minimale und maximale DT-Wert aller RZ-Cluster eines Bezirks zum Zeitpunkt der Analyse.

Der normalisierte Faktor für die Frequenz an Vorgängen während des Zeitraums eines RZ-Cluster wurde, wie anhand der folgenden Berechnung, durchgeführt:

$$\text{Freq}_{\text{RZC}} = \frac{\text{Freq}_i - \text{Freq}_{\min}}{\text{Freq}_{\max} - \text{Freq}_{\min}}$$

mit

$$\text{Freq}_i = \frac{\text{NR}_{\text{VinRZC}_i}}{DT_i}$$

Der normalisierte Faktor für Aktualität eines RZ-Cluster ergab sich rechnerisch aus:

$$\text{Akt365}_{\text{RZC}} = 1 - \left(\frac{ZIV_i - ZIV_{\min}}{ZIV_{\max} - ZIV_{\min}} \right)$$

Je höher der jeweilige Faktor, desto länger dauerte ein Cluster an (Dauer_{RZC}), desto höher war die Frequenz von Vorgängen im Cluster (Freq_{RZC}) oder desto höher war die Aktualität eines RZ-Cluster (Akt365_{RZC}).

Anfänglich wurde der gesamte Datensatz (Beginn im Jahr 2012) für die Berechnung des Aktualitätsfaktors einbezogen. Hierbei war der Aktualitätsscore zunächst auch für diejenigen RZ-Cluster recht hoch (>0,8), in denen der letzte WED-Vorgang teilweise ein halbes Jahr zurücklag. Da dies der definitorischen und inhaltlichen Auslegung der Abbildung eines aktuellen Einbruchgeschehens widersprach, wurden anschließend für die Berechnung von Akt365_{RZC} schließlich nur RZC-Vorgänge einbezogen, die weniger als 365 Tage vom aktuellen Zeitpunkt (Durchführung RZC-Analyse) zurück lagen. Dadurch wurde realisiert, dass Akt365_{RZC} >0,8 nur noch auf die Cluster zutrifft, die weniger als ca. 2,5 Monate zurücklagen, ca. 0,5 bis 0,8 für die Cluster, die ca. 2,5 Monate bis 6 Monate zurücklagen, <0,5 für die Cluster, die 6-12 Monate zurücklagen und 0 für alle Cluster, bei welchen das letzte Einbruchdatum mehr als ein Jahr zurücklag.

Der RZC-Score RZCSg gewichtete die Dauer, Vorgangsfrequenz und Aktualität des RZ-Clusters gleich und ergab sich durch Addition der drei Faktoren:

$$\text{RZCSg} = \text{Dauer}_{\text{RZC}} + \text{Freq}_{\text{RZC}} + \text{Akt365}_{\text{RZC}}$$

RZCSg kann demnach Werte zwischen 0 und 3 annehmen. Ein hoher Wert bedeutet in dem Kontext, dass das RZ-Cluster von hoher Bedeutung für die Kriminalitätsprognosen ist, da sich das Cluster (1.) über viele Tage erstreckt hat und/oder (2.) die Frequenz der Vorgänge im RZ-Cluster sehr hoch ist und/oder (3.) die Aktualität des RZ-Clusters hoch ist.

Um zusätzlich solche RZ-Cluster sehr hoch zu priorisieren, in denen der letzte WED-Vorgang nur wenige Tage zurück liegt, also die Fokussierung auf eine noch stärkere Betonung der Aktualität zu legen, wurde zusätzlich der Faktor Akt30_{RZC} berechnet. Die Berechnung dieses Faktors war ähnlich wie die Berechnung von Akt365_{RZC}. Allerdings wurden nur RZ-Cluster mit einem Einbruchdatum einbezogen, das maximal 30 Tage vor dem Analysezeitpunkt lag. Für alle anderen RZ-Cluster wurde Akt30_{RZC} auf 0 gesetzt. Au-

ßerdem wurde $Akt30_{RZC}$ nach Normalisierung mit dem Faktor zwei multipliziert, sodass RZ-Cluster, die sehr aktuell waren, besonders hoch gewichtet wurden. Somit konnten bei der WED-Prognose gezielt Gebiete ausgegeben werden, in denen sich sehr aktuell Cluster von zeitlich und räumlich nahen WED ereignet haben und in denen entsprechend die Wahrscheinlichkeit von Folgetaten als sehr hoch eingestuft wurde.

Der Faktor $Akt30_{RZC}$ ging in die Berechnung des zweiten Scores $RZCSa$ ein, der den Zeitabstand zum letzten WED-Vorgang im RZ-Cluster zusätzlich gewichtete:

$$RZCSa = RZCSg + 2 \cdot Akt30_{RZC}$$

Schließlich (Schritt sechs) wurden die berechneten Faktoren ($Dauer_{RZC}$, $Frez_{RZC}$, $Akt365_{RZC}$, $Akt30_{RZC}$) und Scores ($RZCSg$, $RZCSa$) auf Wohnquartierebene aggregiert, um sie im Rahmen der SKALA-Prognosen entsprechend anwenden zu können. Dazu wurde zunächst für jedes Wohnquartier und jedes RZ-Cluster ermittelt, wie viele WED-Vorgänge eines Clusters anteilig in jedem durch dieses räumlich tangierte Wohnquartier vorkamen. Angenommen ein RZ-Cluster mit insgesamt zehn Vorgängen und einem Score von $RZCSa = 3,5$ erstreckt sich über zwei Wohnquartiere (WQ1 und WQ2), wobei zwei Vorgänge in WQ1 und acht Vorgänge in WQ2 liegen, so würde dem WQ1 für dieses RZ-Cluster ein gewichteter $RZCSa$ -Wert von $0,7 (= 0,2 \cdot 3,5)$ und dem WQ2 ein $RZCSa$ -Wert von $2,8 (= 0,8 \cdot 3,5)$ zugeordnet. Die Gewichtungsfaktoren (λ_w) wären dementsprechend $0,2$ und $0,8$. Schließlich wurden jeweils alle gewichteten Scores separat für jedes Wohnquartier aufsummiert, um den Gesamtscore für jedes Wohnquartier zu erhalten. $RZCSg^{WQ}$, der auf Wohnquartierebene aggregierte Score $RZCSa$, wurde hierbei wie folgt berechnet:

$$RZCSg^{WQ} = \sum_1^{n_{RZC}} \lambda_w \cdot RZCSg$$

n_{RZC} ist die Anzahl der RZ-Cluster, die im Gesamtzeitraum der Berechnung (seit 01.01.2012) in diesem Wohnquartier lagen. Nach dieser Berechnung ist $RZCSa^{WQ}$ der auf Wohnquartierebene aggregierte $RZCSa$ -Score.

Eine Besonderheit der Berechnung von $RZCSa^{WQ}$ ist, im Gegensatz zu $RZCSg^{WQ}$, dass auch $RZCSa^{WQ}$ auf 0 gesetzt wird, wenn $Akt365_{RZC}$ gleich 0 ist. Dadurch wurde vermieden, dass Wohnquartiere einen besonders hohen $RZCSa^{WQ}$ erhielten, in denen es seit mehr als einem Jahr zu keinem Einbruch kam, alleine weil die Anzahl der WED-Vorgänge je Wohnquartier insgesamt hoch war und sich dadurch bei Aufsummierung der $RZCSa^{WQ}$ -Werte ein hoher $RZCSa^{WQ}$ -Wert ergab. Denn $RZCSa^{WQ}$ soll primär Wohnquartiere im Bereich sehr aktueller RZ-Cluster hoch gewichteten.

$RZCSg^{WQ}$ zeigt im Wesentlichen die Wohnquartiere an, die im Gesamtzeitraum seit Anfang 2012 stark zur Bildung von RZ-Clustern neigten. Im Gegensatz dazu ist $RZCSa^{WQ}$ dann, in Anlehnung an das Near-Repeat-Phänomen, besonders hoch, wenn es in jüngster Zeit zu einer RZC-Bildung kam.

Folgende Ergebnistabellen der RZC-Analyse wurden abschließend in der Datenbank gespeichert und jeden Montag für die Kriminalitätsprognosen neu erstellt:

1. raumzeit.WED_RaumZeitClusterIDs_extent14Tage300m_ScoresWQs
2. raumzeit.WED_RaumZeitClusterIDs_extent14Tage300m_ScoresWQsSUM

Erstere beinhaltet alle Wohnquartiere und deren zugehörigen RZ-Cluster mit allen errechneten Eigenschaften und Scores. Letztere beinhaltet nur jeweils eine Zeile pro Wohnquartier mit den auf Wohnquartierebenen summierten Scores über alle RZ-Cluster.

4.6.2 Bewertung

In Tabelle 6 sind beispielhaft RZ-Cluster mit den zugehörigen Merkmalen und Scores aufgelistet, die am 23.10.2017 berechnet wurden. Jedem RZ-Cluster ist eine einmalige ID zugeordnet, bestehend aus dem Behördenkürzel (z. B. „D“ für Düsseldorf) und der Nummer. Im RZ-Cluster „D 2295“ gab es im Zeitraum zwischen dem 05.11.2016 und dem 04.01.2017 insgesamt zwölf raum-zeitlich nahe WED-Vorgänge, wobei die RZC-Dauer (DT) von 60 Tagen überdurchschnittlich hoch ist. Somit trägt $Dauer_{RZC}$ maßgeblich zu dem finalen Score-Wert von $0,99$ bei. $RZCSg$ und $RZCSa$ sind in diesem Fall identisch, da RZC-Ende mehr als drei Monate zurück liegt. Im Falle von „B 114“ ist $RZCSa = 0$, da der letzte Einbruch in diesem Cluster am 29.03.2012 stattfand und somit deutlich mehr als 365 Tage zurück liegt. Im Gegensatz dazu ist $RZCSa$ bei RZ-Cluster „EM 2722“ mehr als doppelt so hoch wie $RZCSg$. Dies ist darauf zurückzuführen, dass der letzte RZC-Vorgang am 21.10.2017 nur wenige Tage vor dem Analysedatum stattfand. Dadurch ist $Akt30_{RZC}$ mit $0,82 \cdot 2$ (hier nicht aufgeführt) sehr hoch und entsprechend auch $RZCSa$.

Tabelle 6: Beispielauswahl: Merkmale von RZ-Clustern

| RZC-ID | NR _{VinRZC} | RZC-Beginn | RZC-Ende | DT [Tage] | Freq _{RZC} | Dauer _{RZC} | Akt365 _{RZC} | RZCSg | RZCSa |
|-----------|----------------------|------------|------------|-----------|---------------------|----------------------|-----------------------|-------|-------|
| D 2295 | 12 | 05.11.2016 | 04.01.2017 | 60 | 0,0 | 0,72 | 0,26 | 0,99 | 0,99 |
| EM 2722 | 5 | 10.10.2017 | 21.10.2017 | 11 | 0,1 | 0,15 | 1,00 | 1,21 | 3,13 |
| BN 114 | 9 | 28.03.2012 | 29.03.2012 | 1 | 0,9 | 0,01 | 0,00 | 0,91 | 0,00 |
| D 331 | 7 | 15.09.2012 | 11.11.2012 | 57 | 0,0 | 0,69 | 0,00 | 0,69 | 0,00 |
| KLev 4940 | 6 | 23.09.2017 | 07.10.2017 | 14 | 0,1 | 0,12 | 0,96 | 1,14 | 2,07 |

Abbildung 34 zeigt alle RZCSg- und RZCSa-Werte für RZ-Cluster, bei welchen RZCSa > RZCSg ist. Dies sind zum Auswertzeitpunkt (Kalenderwoche 41, 2017) 80 RZ-Cluster von insgesamt 14.524 RZ-Clustern. In diesem Zusammenhang wird deutlich, dass die Höhe des RZCSa-Wertes tendenziell mit der Höhe RZCSg-Wertes zunimmt. Dies zeigt sich auch am Beispiel von „EM 2722“ und „KLev 4940“ (siehe Tabelle 6) und hängt damit zusammen, dass Akt365_{RZC} ein wesentlicher Faktor bei der Berechnung von RZCSg ist. Insgesamt ist für 12.787 der 14.524 berechneten RZ-Cluster der RZCSa-Wert 0, da diese RZ-Cluster seit mehr als einem Jahr abgeschlossen sind. Für 1.457 von 12.787 ist der Wert für RZCSa und RZCSg identisch. In diesem Fall sind die RZ-Cluster mehr als 30 Tage, jedoch maximal 365 Tage alt.

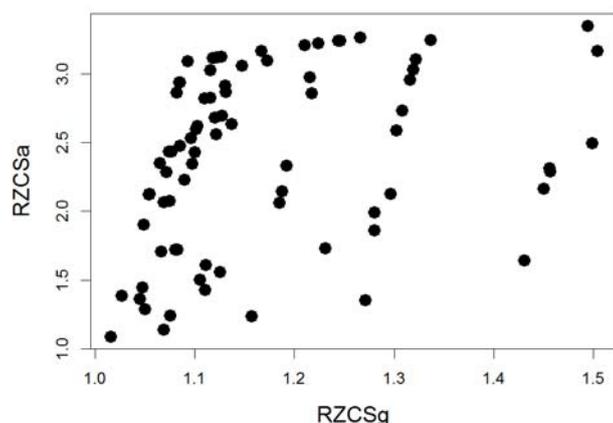
Abbildung 34: RZC-Scores (RZCSa > RZCSg)

Abbildung 35 illustriert beispielhaft die für die Kalenderwoche 43 (2017) berechneten Scores RZCSa für Wohnquartiere in Düsseldorf. In der Illustration wird deutlich, dass die aggregierten Werte RZCSa^{WQ} in den Wohnquartieren höher sind, in denen viele WED-Vorgänge als Teil unterschiedlicher RZ-Clustern lagen und wo einige RZ-Cluster hohe RZCSg-Werte aufwiesen. In dem rot markierten Wohnquartier ist RZCSa^{WQ} verhältnismäßig groß, da es hier aktuell zu einer Bildung von RZ-Clustern kam. Tabelle 7 dokumentiert ergänzend die RZ-Cluster, die in diesem Wohnquartier (KGS22 = 05111000100382) liegen und deren Scores (RZCSa) in die Berechnung von RZCSa^{WQ} eingegangen sind. Insgesamt lagen in diesem Wohnquartier im Gesamtzeitraum (seit dem 01.01.2012 bis zum Analysedatum vom 23.10.2017) 27 RZ-Cluster. Allerdings sind nur vier dieser

RZ-Cluster aktuell genug, um in die Berücksichtigung der RZCSa^{WQ} einzugehen. In Tabelle 7 sind nicht alle 27 RZ-Cluster aufgeführt. Da die Anzahl der RZ-Cluster und die Aktualität der jüngsten RZ-Cluster in diesem Wohnquartier sehr hoch sind, ist auch der RZCSg^{WQ} für dieses Quartier verhältnismäßig hoch. Wie Abbildung 35 ebenfalls zeigt, besitzen jedoch auch andere Wohnquartiere einen hohen RZCSg^{WQ}-Wert. Dies sind Quartiere, in welchen der letzte Einbruch schon mehrere Monate zurück lag, da die Aktualität hier nicht so hoch gewichtet wird wie bei RZCSa.

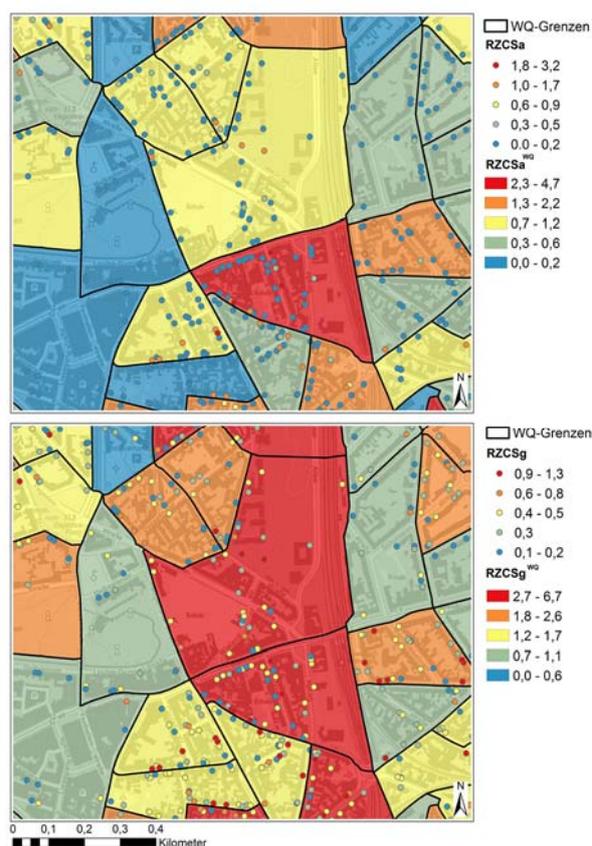
Abbildung 35: Illustration KPB Düsseldorf auf Wohnquartierebene

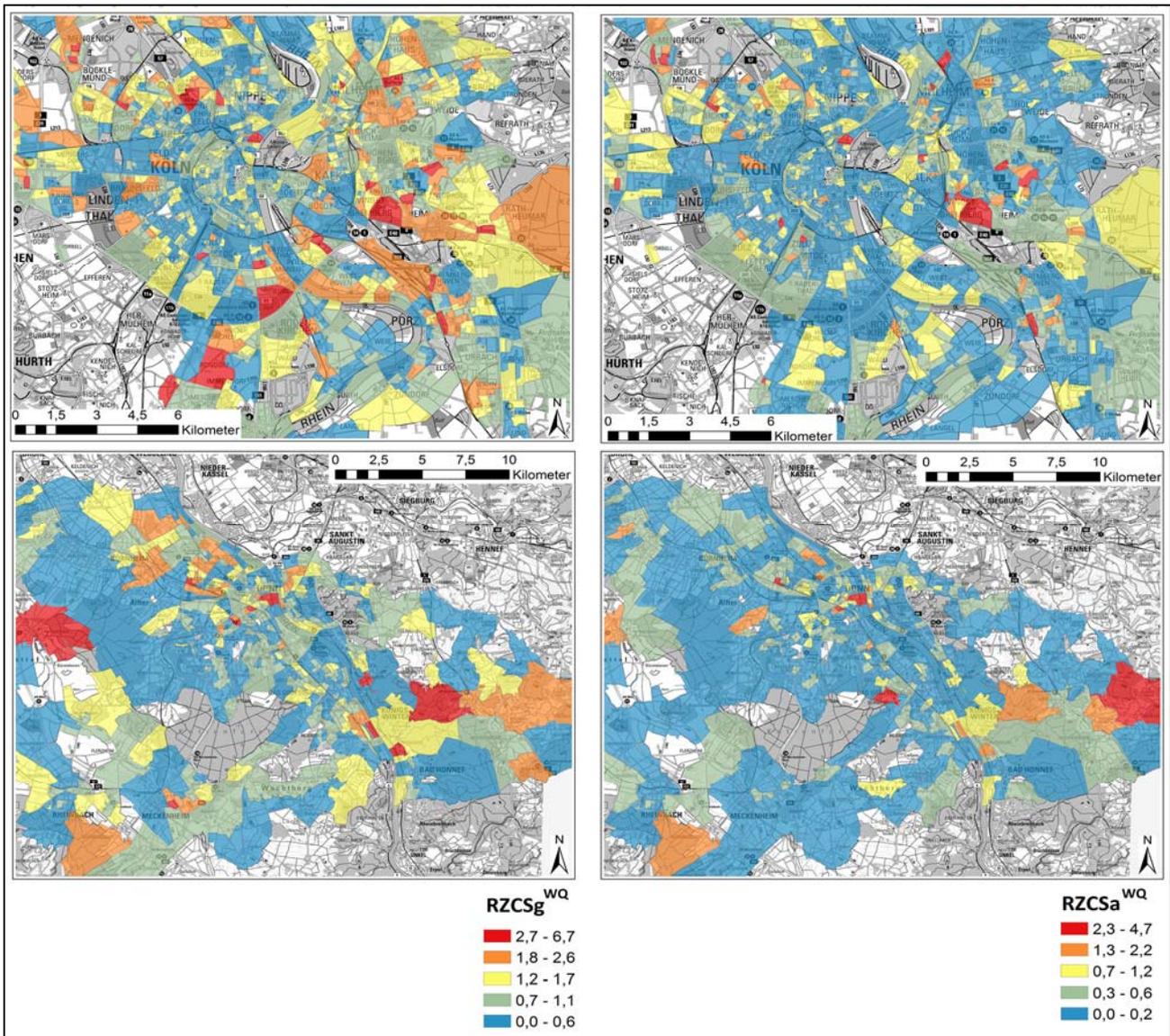
Tabelle 7: RZCS für WQ = 05111000100382

| RZC-ID | RZC-Ende | RZCSg | RZCSa |
|--------|------------|-------|-------|
| D 2592 | 22.10.2017 | 1,17 | 3,17 |
| D 2384 | 12.02.2017 | 0,90 | 0,90 |
| D 2313 | 18.11.2016 | 0,42 | 0,42 |
| D 2282 | 02.11.2016 | 0,32 | 0,32 |
| D 1025 | 02.03.2014 | 0,10 | 0,00 |
| D 1594 | 08.05.2015 | 0,24 | 0,00 |
| D 201 | 29.05.2012 | 0,22 | 0,00 |
| D 944 | 31.12.2013 | 0,15 | 0,00 |
| ... | ... | ... | ... |

In Abbildung 36 sind die für Kalenderwoche 43 (2017) berechneten Verteilungen RZCSa^{WQ} und RZCSg^{WQ} für Wohnquartiere diverser Polizeibezirke dargestellt. Die absoluten Werte sind zwischen RZCSa^{WQ} und RZCSg^{WQ} nicht ver-

gleichbar, da sich diese jeweils auf eine andere Berechnungsgrundlage beziehen. Allerdings ist zu erkennen, dass im Falle von RZCSg^{WQ} eine größere Anzahl an Wohnquartieren einen verhältnismäßig hohen RZCSg^{WQ}-Wert besitzt. Dies liegt darin begründet, dass in die Berechnung von RZCSg^{WQ} auch RZCSg-Werte von RZ-Clustern eingehen, bei denen der letzte RZC-Vorgang mehr als ein Jahr zurückliegen kann. Im Gegensatz dazu fließen in die Berechnung von RZCSa^{WQ} nur RZCSa-Werte von aktuelleren RZ-Clustern ein, bei denen der letzte Einbruch maximal ein Jahr zurück liegt. Daher ist der Anteil an Wohnquartieren mit verhältnismäßig hohen RZCSa^{WQ}-Werten geringer als bei RZCSg^{WQ}. Zudem lässt sich erkennen, dass die Wohnquartiere, die im Gesamtzeitraum stark zu RZC-Bildung neigten, meist nicht mit den Gebieten im Stadtzentrum übereinstimmen, in denen grundsätzlich häufig ein hohes Deliktaufkommen beim WED zu beobachten war. Der Umstand ist vor allem auf die Berechnung der Scores bzw. der Berücksichtigung der Aktualität in den RZC-Scores zurückzuführen.

Abbildung 36: RZCSg^{WQ} (links) und RZCSa^{WQ} (rechts) für Wohnquartiere in Köln (oben) und Bonner Raum (unten)



4.7 Evaluierung der RZC-Scores

Nach der Konzipierung und Durchführung der RZC-Analyse und der Bildung entsprechend darauf aufbauender Scores für das genutzte Entscheidungsbaummodell, wurde weiterhin geprüft, inwieweit diese Scores tatsächlich zur Modellverbesserung beitragen. Da sich das bisherige Vorgehen bei der Modellierung lediglich auf die Wohnquartierebene bezog und somit nicht das Erkennen von Zusammenhängen über die Wohnquartiergrenzen ermöglichte, wurde ein entsprechender Mehrwert der Scores erwartet.

Die hier nachfolgend zusammengefasste Beurteilung der Qualität prognostizierter Einbruchwahrscheinlichkeiten auf Basis der RZC-Scores, im Vergleich zu vorher verwendeten Variablen im Modell, erfolgte durch eine statistische Gegenüberstellung mit tatsächlich gemeldeten Vorgängen und deren raum-zeitlichen Verteilungen. Ergänzend dazu sollte quantifiziert werden, inwiefern die auf der Annahme des Near-Repeat-Phänomens retrograd berechneten wöchentlichen Einbruchwahrscheinlichkeiten mit dem tatsächlichen Einbruchgeschehen der Folgewoche übereinstimmen. Diese Abschätzung der Modellvalidität sowie die Vorhersagekraft der Scores war eine wichtige Voraussetzung für die interne Weiterentwicklung des Modells und darüber hinaus für die Beurteilung der Verlässlichkeit berechneter Einbruchwahrscheinlichkeiten im Rahmen des Projektes SKALA.

Die retrograden Berechnungen für diese Analyse erfolgten für einen Zeitraum von drei Jahren (2014 bis 2016) und für alle Pilotbehörden. In diesem Zeitraum wurden SKALA-Prognosegebiete noch nicht oder kaum gezielt im Rahmen des Projektes SKALA polizeilich bemaßnahmt. Aus diesem Grund und auch aufgrund der methodischen Herangehensweise dieser Analyse konnte eine Verfälschung der Evaluierungsergebnisse durch präventiv-polizeiliche Interventionen im Zusammenhang mit dem Projekt SKALA weitgehend ausgeschlossen werden. Nichtsdestotrotz haben auch in diesem Zeitraum sicherlich allgemeine polizeiliche Aktivitäten in den Wohnquartieren wie auch gezielte Einbruchpräventionsmaßnahmen außerhalb des Projektes SKALA stattgefunden. Durch die Wahl eines entsprechend langen Zeitraumes von drei Jahren und durch die Bildung darauf abgestimmter Evaluierungsindizes konnten solche Einflüsse aber außer Acht gelassen werden, sodass sich diese Herangehensweise für eine interne Abschätzung der Modellvalidität und Vorhersagekraft der Scores anbot. Die Abgrenzung zu Qualitätsmetriken wie einer „Trefferrate“ ist wichtig. „Trefferraten“ wie sie häufig im Rahmen von Predictive Policing verwendet werden, beziehen sich i. d. R. nur auf die aktuelle Kriminalitätsprognose oder einen sehr kurzen Zeitraum. Es ging bei der Evaluierung der RZC-Scores aber um eine interne Abschätzung der Vorhersagekraft.

Abgesehen von den vorhandenen Wohnquartiergrenzen wurden in Zusammenarbeit mit dem Lehrstuhl für Datenanalyse und Visualisierung der Universität Konstanz neue

Wohnquartiergrenzen ermittelt (siehe Kapitel 3.8), um eine höhere Homogenität in den Quartieren zu erreichen. Die hier beschriebene Analyse wurde sowohl für die sogenannten neuen Wohnquartiere als auch für die ursprünglichen Wohnquartiere getestet und verglichen.

Die Untersuchung der Dynamik von berechneten Prognosegebieten wurde ebenfalls durchgeführt. Das bedeutet, es wurde gezielt evaluiert, inwieweit sich die auf Basis der RZC-Scores ermittelten Prognosegebiete durchschnittlich von Woche zu Woche unterscheiden und ob sie – wie erwartet – durch die stärkere Fokussierung auf das aktuelle Einbruchgeschehen dynamischer in Raum und Zeit werden. Denn es wäre wenig zielführend, jede Woche beinahe dieselben Gebiete (Einbruch-Hotspots) als Kriminalitätsprognose herauszugeben.

4.7.1 Methodik

Die komplette Analyse erfolgte vollautomatisiert mittels eines selbst erstellten R-Skriptes und wurde separat für jede der in Tabelle 8 aufgeführten Variablen durchgeführt. Anschließend wurden die Analyseergebnisse, insbesondere im Hinblick auf die Vorhersagekraft einzelner Variablen, verglichen. Die ermittelten Einbruchwahrscheinlichkeiten ergeben sich aus der Höhe der jeweils berechneten Variablen.

Tabelle 8: Variablen RZC-Scores-Evaluierung

| Variablenkürzel | Variablenbeschreibung |
|-------------------------|--|
| Σ WED | Gesamtanzahl der Vorgänge im Quartier |
| WED _{lastDate} | Tatzeitdatum des letzten WED-Vorgangs im Quartier |
| Dauer _{RZC} | RZC-Score für die Länge des Zeitraums, die ein RZ-Cluster besteht. Je länger der Zeitraum, desto höher RZCS-Dauer |
| Freq _{RZC} | RZC-Score für die Frequenz, mit welcher Vorgänge im jeweiligen Zeitraum des Clusters (der Dauer) auftreten. Je höher die Frequenz, desto höher RZCS-Frequenz |
| Akt365 _{RZC} | RZC-Score für die Aktualität. Je geringer der zeitliche Abstand zwischen dem letzten Vorgang im Cluster und dem jeweiligen Beginn einer Kalenderwoche, desto höher RZCS-Aktualität |
| RZCSg | RZC-Score, der durch Addition der normalisierten Faktoren RZCS-Dauer, RZCS-Frequenz und RZCS-Aktualität zusammenfasst |
| RZCSa | RZC-Score, der RZCS-Final mit einem zusätzlichen Aktualitätsfaktor multipliziert, der nur Taten berücksichtigt, die maximal 30 Tage zurückliegen. Somit wird die Aktualität eines Clusters stärker gewichtet als die Frequenz und die Dauer eines Clusters |

Die hier evaluierten RZCS (siehe Tabelle 8) wurden durch die im Kapitel 4.6 beschriebene RZC-Analyse ermittelt, die nur Vorgänge berücksichtigt, die sich in zeitlicher und räumlicher Nähe (maximal 14 Tage und 300 Meter) zu mindestens einem anderen WED ereigneten.

WED_{lastDate} und Akt365_{RZC} haben eine ähnliche Aussagekraft, da sie Gebiete höher gewichten, in denen sich vor kurzer Zeit ein WED ereignet hat. Allerdings bezieht WED_{lastDate} alle Vorgänge ein, wohingegen Akt365_{RZC} nur diejenigen berücksichtigt, die Teil eines RZ-Clusters sind. Ein weiterer wesentlicher Unterschied ist, dass sich der WED bei Akt365_{RZC} auch außerhalb des jeweiligen Wohnquartiers befinden kann, sofern es Teil eines Clusters ist, das sich – aber nicht allein – über dieses Wohnquartier erstreckt. Der dritte Unterschied ist, dass Akt365_{RZC} keine Vorgänge berücksichtigt, die länger als ein Jahr vor dem jeweiligen Analysedatum liegen. Für alle Wohnquartiere, die seit mehr als einem Jahr nicht Teil eines RZ-Clusters waren, wurde Akt365_{RZC} auf 0 gesetzt. Die Variable Σ WED zeigt die Anzahl an WED-Vorgängen für jedes Wohnquartier im gesamten Zeitraum vom 01.01.2014 bis zu Beginn der jeweiligen Kalenderwoche an und ist somit ein Indikator für die grundsätzliche WED-Anfälligkeit bzw. -verletzlichkeit eines Wohnquartiers. Wohnquartiere mit hohen Σ WED-Werten sind meist solche, die sich in zentralen Lagen der Bezirke befinden oder in Wohngebieten, die bei der Polizei i. d. R. ohnehin als WED-Brennpunkte bekannt sind.

Für die Evaluierung wurden zunächst die in Tabelle 8 aufgeführten Variablen retrograd für jede der 157 Kalenderwochen im Zeitraum von drei Jahren (2014 bis 2016) für jedes Wohnquartier einzeln berechnet. Anschließend wurde der Datensatz separat für jede Kalenderwoche und separat für jede Variable absteigend nach der Variablengröße sortiert, sodass die Wohnquartiere mit den höchsten Variablenwerten oben in der Tabelle gelistet wurden. Die Auswahl an Wohnquartieren mit den höchsten Variablenwerten wurde

als Prognose für die Folgewoche angenommen, die entsprechend mit der tatsächlichen Verteilung der Taten in der Folgewoche iterativ für 156 Kalenderwochen verglichen wurde. Dieser Vergleich und damit die eigentliche Evaluierung erfolgten mittels der statistischen Indizes Recall und Precision (siehe Kapitel 4.4.4). Die Elemente der Konfusionsmatrix (siehe Tabelle 4, Kapitel 4.4.4), als Basis für die hier berechneten Indizes, sind:

- **TruePositive:** Anzahl der Wohnquartiere, die in der Auswahl/Prognose enthalten sind und in denen auch eingebrochen wurde.
- **TrueNegative:** Anzahl der Wohnquartiere, die nicht in der Auswahl/Prognose enthalten sind und in denen auch nicht eingebrochen wurde.
- **FalsePositive:** Anzahl der Wohnquartiere, die in der Auswahl/Prognose enthalten sind, in den aber nicht eingebrochen wurde.
- **FalseNegative:** Anzahl der Wohnquartiere, die nicht in der Auswahl/Prognose enthalten sind, in denen aber eingebrochen wurde.

Der Recall zeigt demnach den relativen Anteil der Wohnquartiere an, die anteilig an den tatsächlichen WED der Folgewoche durchschnittlich in der Prognose(-auswahl) enthalten waren. Er ist somit ein Indikator für die Wahrscheinlichkeit, dass ein WED-Gebiet prognostiziert wird. Die Precision zeigt an, in wie viel Prozent der Prognosegebiete tatsächlich in der Folgewoche eingebrochen wurde und ist demnach ein Index für die Wahrscheinlichkeit, dass in ein prognostiziertes Gebiet tatsächlich eingebrochen wird.

Recall und Precision wurden separat für jede Kalenderwoche und jede Pilotbehörde berechnet und anschließend über alle Kalenderwochen gemittelt. Werte für Recall- und Precision variieren in Abhängigkeit von der Anzahl der als Prognosegebiete definierten Wohnquartiere. Um einen Maßstab für eine sinnvolle Größe der Gebietsauswahl zu er-

halten, wurde vor der eigentlichen Evaluierung zunächst berechnet, in wie viel Prozent der Wohnquartiere des jeweiligen Polizeibezirks in einer Kalenderwoche im Durchschnitt eingebrochen wurde. Dieser Wert betrug für alle Pilotbehörden etwa 6,3 Prozent und schwankte zwischen 5,8 Prozent (Duisburg) und 6,7 Prozent (Düsseldorf). Es wurde somit zunächst eine Auswahl von 6 Prozent der Wohnquartiere mit den jeweils höchsten Variablenwerten festgelegt. Zusätzlich wurden die Indizes für eine Auswahl der höchsten 2 Prozent und 4 Prozent berechnet, da Rückmeldungen aus den Pilotbehörden im Projektverlauf ergaben, dass eine kleinere Auswahl wesentlich praxistauglicher ist. Entsprechend wurden die Indizes separat für eine (Top-)Auswahl von 2, 4 und 6 Prozent der Wohnquartiere eines Polizeibezirks mit den höchsten Variablenwerten berechnet und verglichen. Die Anzahl an Prognosegebieten, die sich aus der jeweiligen Top-Auswahl [%] ergibt, ist in Tabelle 9 dokumentiert. Aufgrund der unterschiedlichen Anzahl an Wohnquartieren in den jeweiligen Polizeibezirken variiert die Anzahl der ausgewählten Gebiete stark. Für den Polizeibezirk Gelsenkirchen bedeuten beispielsweise 6 Prozent der Wohnquartiere eine Auswahl von 16 Gebieten, für Köln ergeben sich bei 6 Prozent der Quartiere 80 Gebiete.

Tabelle 9: Top-Auswahl der Prognosegebiete

| | Gesamt | Top2 % | Top4 % | Top6 % |
|---------|--------|--------|--------|--------|
| D | 593 | 11 | 23 | 35 |
| DU | 481 | 9 | 19 | 28 |
| EM | 780 | 15 | 31 | 46 |
| GE | 275 | 5 | 11 | 16 |
| KLev | 1338 | 26 | 53 | 80 |
| BN | 499 | 9 | 19 | 29 |
| alleBez | 3966 | 75 | 156 | 234 |

Erläuterung: Anzahl an Prognosegebieten der Polizeibezirke für eine unterschiedlich große Top-Auswahl [%]. Die Berechnung erfolgte anhand der höchsten Score-Werte bei der Prognoseerstellung.

Durch die separate Berechnung von Recall und Precision für jede Kalenderwoche und jeden Polizeibezirk sowie die anschließende Mittelung der Werte über alle 156 Kalenderwochen, wurden Zufallsfehler herausgemittelt. Die Indizes erreichen in ihrer Aussagekraft folglich eine entsprechende Robustheit. Die retrograden Berechnungen über drei Jahre ermöglichen daher eine zuverlässige Abschätzung der Vorhersagekraft der untersuchten Variablen in Bezug auf die Einbruchwahrscheinlichkeit der Folgeweche.

Als Referenz wurde separat für jeden Bezirk die gleiche Anzahl an Gebieten, wie sie für die Berechnung der Indizes verwendet wurde, zufällig aus der Gesamtheit der Wohnquartiere eines Bezirkes gezogen. Von einer verlässlichen Abschätzung einer erhöhten Einbruchwahrscheinlichkeit für die Folgeweche mittels der untersuchten Variablen kann dann ausgegangen werden, wenn der jeweils berechnete

Indexwert für die Auswahl auf Basis der höchsten Variablenwerte deutlich höher ist als der Indexwert, der mit der Zufallsstichprobe berechnet wurde.

Um die zeitliche Dynamik der mittels der aufgeführten Variablen berechneten Prognosegebiete zu untersuchen, wurde für jede einzelne Variable ermittelt, wie stark sich die Prognosegebiete von Woche zu Woche im Durchschnitt unterscheiden. Dazu wurde separat für jeden Bezirk iterativ geprüft, wie viele Prognosegebiete im Durchschnitt mit der Auswahl an Prognosegebieten der Folgeweche und darauf folgender Wochen übereinstimmen. Es wurde der Index Var_{Dyn} ermittelt, der anzeigt, wie viel Prozent der Gebiete sich im Durchschnitt zwischen zwei Kalenderwochen unterscheiden. Je größer Var_{Dyn} , desto mehr unterscheiden sich die Prognosegebiete zweier Wochen. Je geringer Var_{Dyn} , desto statischer wäre das Prognosemodell auf Basis der jeweiligen Variable. Var_{Dyn} wurde separat für die Variablen $Akt365_{RZC}$, $RZCSg$ und $RZCSa$ für zeitliche Abstände von 1-79 Wochen zweier Kalenderwochen berechnet und verglichen. Im Gegensatz zu der Evaluierung mittels der Indizes Recall und Precision wurden bei Berechnung von Var_{Dyn} unabhängig von der Bezirksauswahl jeweils die 20 Wohnquartiere mit den höchsten Variablenwerten als Prognosegebiete definiert und diese bezüglich ihrer zeitlichen Dynamik analysiert.

4.7.2 Bewertung

Zur Bewertung der Vorhersagekraft der berechneten RZC-Scores in Bezug auf die Einbruchwahrscheinlichkeiten in Wohnquartieren zeigt Abbildung 37 den durchschnittlichen und über alle Bezirke gemittelten Precision- und Recall Index [%] für die retrograden WED-Prognosen (Top2%-Auswahl) mittels der getesteten Variablen (siehe Tabelle 8). Für alle Variablen waren sowohl Precision als auch Recall deutlich höher als der aus der Zufallsstichprobe berechnete Referenzwert. Demnach ermöglichen es alle Vorhersagevariablen eindeutig, Wohngebiete mit höheren Einbruchwahrscheinlichkeiten für die Folgeweche zu prognostizieren. Sowohl Precision als auch Recall waren für $RZCSa^{WQ}$ am höchsten (19,5 Prozent und 5,9 Prozent), gefolgt von $RZCSg^{WQ}$, $Akt365_{RZC}$ und $\sum WED$. Der Referenzwert betrug jeweils etwa zwei Prozent. Durch den $RZCSa^{WQ}$ waren somit die jeweiligen Wahrscheinlichkeiten des WED in den ausgewählten Wohnquartieren durchschnittlich etwa zehnfach (Precision) und dreifach (Recall) höher als im Vergleich zu einer zufälligen Gebietsauswahl. In allen Pilotbehörden lag die Vorhersagekraft von $\sum WED$ deutlich über der Vorhersagekraft von $WED_{lastDate}$. Demzufolge lag die berechnete Einbruchwahrscheinlichkeit höher in Wohnquartieren, die insgesamt sehr stark von WED betroffen sind, als in Wohnquartieren, in denen der Zeitabstand zum letzten Einbruch im Wohnquartier am höchsten war. Die Vorhersagekraft von $RZCSa^{WQ}$ war für alle Pilotbehörden deutlich größer als die von $RZCSg^{WQ}$. Das lässt darauf schließen, dass eine stärkere Gewichtung der Aktualität bei der Berechnung der RZC-

Scores durchaus zu erhöhten Wahrscheinlichkeiten und demnach einer stärkeren Vorhersagekraft der Variablen bzw. Scores führt. Die Vorhersagekraft von $RZCSa^{WQ}$ war im Vergleich zu den anderen getesteten Variablen am höchsten, auch im Falle der Top4 % und Top6 %-Auswahl. Allerdings nahm der relative Abstand der Indexwerte zwischen $RZCSa^{WQ}$ und den anderen Variablen bei einer höheren Top-Auswahl geringfügig ab. Hervorzuheben ist außerdem, dass

sowohl Recall als auch Precision bei $Akt365_{RZC}$ deutlich höher waren als bei $WED_{lastDate}$. $Akt365_{RZC}$ wurde im Gegensatz zu $WED_{lastDate}$ für die RZ-Cluster unabhängig von den Wohnquartiergrenzen bzw. darüber hinausgehend berechnet. Dies verdeutlicht, dass es durchaus sinnvoll ist, den im Hinblick auf das Near-Repeat-Phänomen ermittelten Zeitabstand zum jüngsten Einbruch auf Basis der RZ-Cluster zu ermitteln und in die Modell- und Prognoseerstellung einzubeziehen. Die Prognosegüte kann so gesteigert werden.

Abbildung 37: RZC-Score: Vergleich (Precision und Recall) für die Top2 % der Prognosegebiete

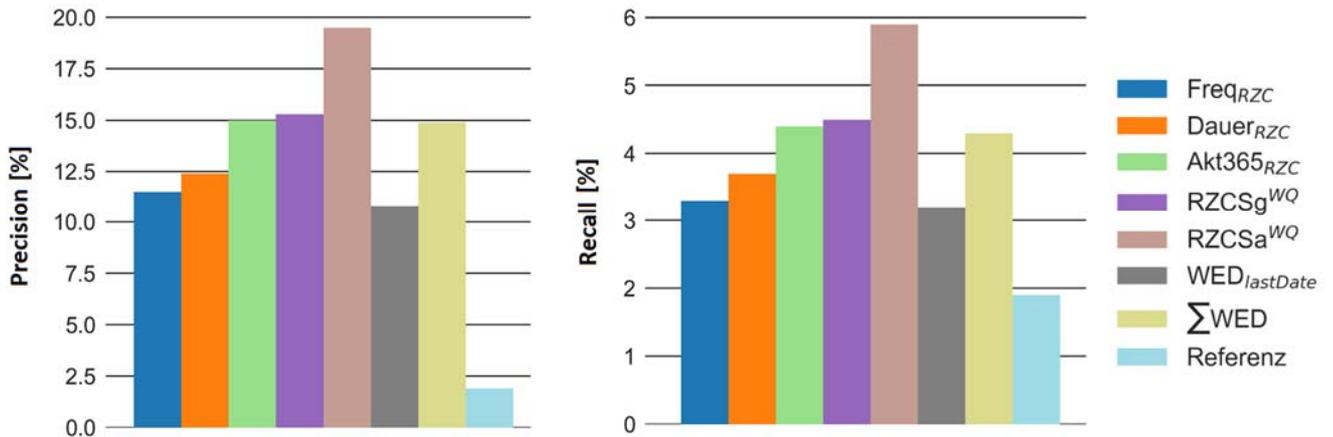


Abbildung 38: RZC-Score: Vergleich (Precision und Recall) für die Top2 % der Prognosegebiete nach Pilotbehörden

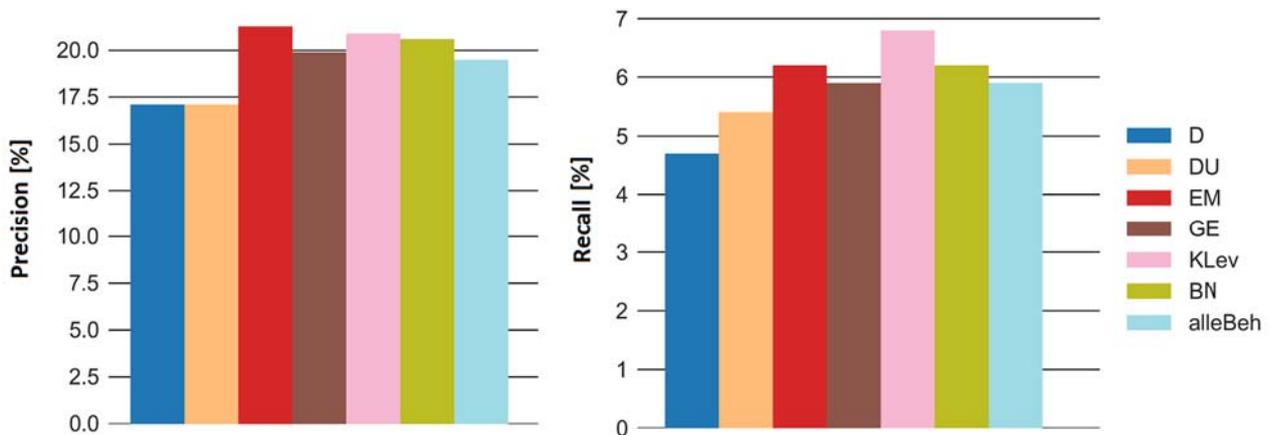


Abbildung 38 verdeutlicht, dass Recall und Precision zwischen den einzelnen Bezirken leicht schwanken. Für Köln/Leverkusen (KLev) und Essen/Mülheim (EM) war die Vorhersagekraft $RZCSa^{WQ}$ verhältnismäßig am höchsten.

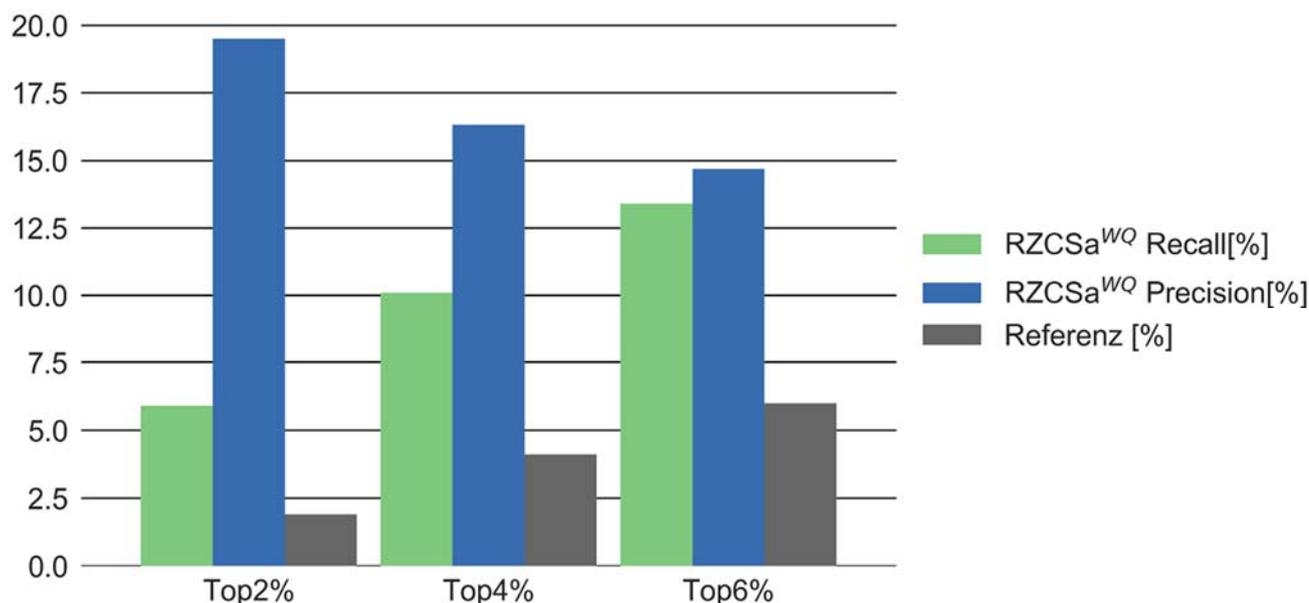
Abbildung 39 verdeutlicht, wie stark der Recall- und Precision-Wert von der prozentuellen Anzahl der als „Prognosegebiet“ definierten Wohnquartiere (Top-Auswahl [%]) auf Basis der höchsten $RZCSa^{WQ}$ -Werte abhängt. Bekanntermaßen ist die Zu- oder Abnahme von Recall und Precision gegensätzlich (vgl. Maimon/Rokach 2015: 35). Im Kapitel 3.12.3 zur Einflussdimension des Raumes wurde dies ebenfalls bereits umfassend beschrieben. Mit prozentueller Zunahme der ausgewählten Prognosegebiete nimmt demnach

der Recall zu. Dies ist plausibel, denn mit einer höheren Anzahl an Prognosegebieten ist auch die Wahrscheinlichkeit höher, dass diese eine größere Anzahl der tatsächlichen WED in der Folgewoche abdecken. Im Gegensatz dazu ist die Precision durchschnittlich höher bei einer kleineren Anzahl an ausgewählten Prognosegebieten. Dies ist wie folgt zu erklären: Es wird davon ausgegangen, dass je höher der Variablenwert (hier $RZCSa^{WQ}$) eines Wohnquartiers ist, desto höher ist auch die Wahrscheinlichkeit eines WED in der Folgewoche für dieses Wohnquartier. Entsprechend werden die Wohnquartiere mit den höchsten Variablenwerten als

Prognosegebiete definiert. Wenn eine kleine Anzahl an Wohnquartieren mit den höchsten Score-Werten ausge-

wählt wird, so ist erwartungsgemäß auch die Wahrscheinlichkeit höher, dass in diesen Gebieten tatsächlich ein WED-Vorgang in der Folgewoche verzeichnet wird.

Abbildung 39: RZC-Score: Vergleich (Precision und Recall) für die Top% der Prognosegebiete

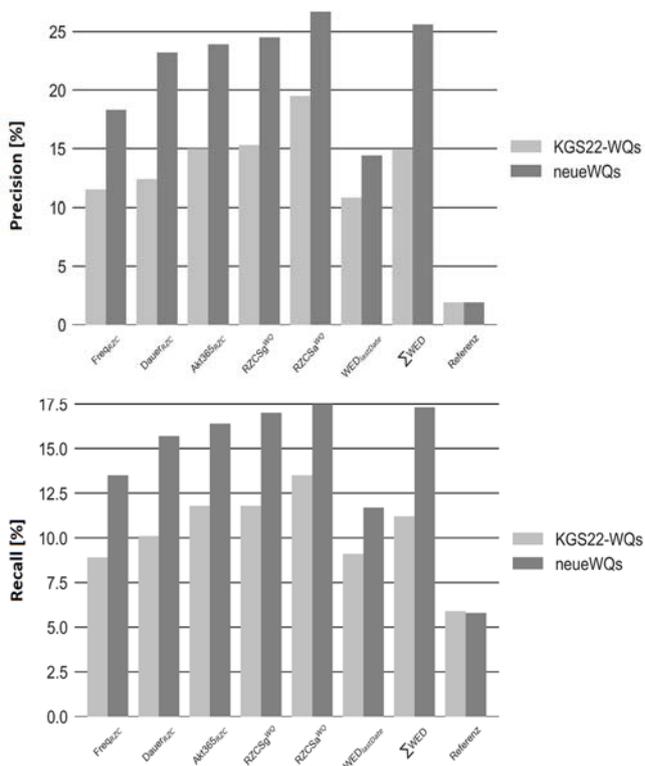


Beim Vergleich der durchschnittlichen Precision- und Recall-Indizes für die bisher verwendeten Wohnquartiere (KGS22) und die neu berechneten Wohnquartiere fällt auf (siehe Abbildung 40), dass die Vorhersagkraft aller Variablen bei Verwendung der neu berechneten Wohnquartiere deutlich höher ist. Dies deckt sich mit der anfänglichen Annahme, dass die neu berechneten Wohnquartiere im Hinblick auf die Einwohnerzahl, Bebauung und Sozialstruktur homogener sind und so bessere Zusammenhänge erkannt werden können. Die relative Zunahme der richtig prognostizierten Prognosegebiete betrug ca. 3 Prozent ($WED_{lastDate}$) bis ca. 12 Prozent (ΣWED). Somit dienen die externen Daten nicht nur direkt zur Vorhersage von WED anfälligeren Gebieten, sondern können durch Clusterbildung von homogeneren Raumeinheiten in Form der neu berechneten Wohnquartiere auch die Vorhersagekraft der hier analysierten, auf Wohnquartierebene aggregierten Variablen, deutlich verbessern.

Weiterhin fällt auf (siehe Abbildung 40), dass die Vorhersagekraft von RZCSa^{WQ} bei Verwendung der neu berechneten Wohnquartiergrenzen insgesamt um etwa sieben Prozent zugenommen hat und im Vergleich zu den anderen hier getesteten Variablen immer noch am höchsten ist (Precision = 26 Prozent, Recall = 20 Prozent). Die relative Höhe der RZCSa^{WQ}-Indizes im Verhältnis zu der Indexhöhe der anderen Variablen hat jedoch abgenommen. Während bei den KGS22-WQ die Indexwerte von RZCSa^{WQ} etwa zwei bis vier

Prozent höher waren als die von RZCSg^{WQ}, waren sie bei den neu berechneten Wohnquartieren nur noch etwa 0,7 Prozent höher.

Abbildung 40: RZC-Score: Vergleich für Wohnquartiere

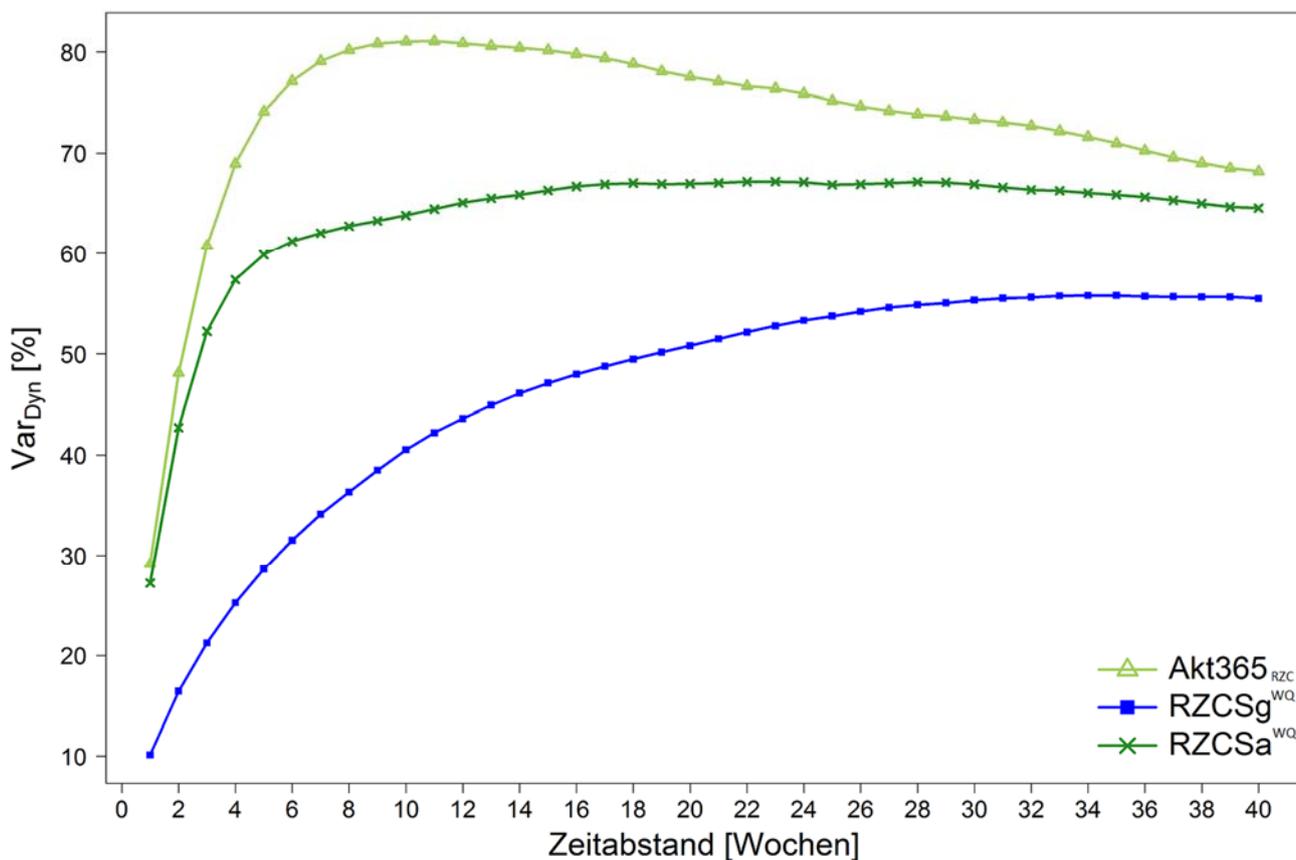


Besonders stark hat sich durch die Verwendung der neu berechneten Wohnquartiergrenzen auch die Vorhersagekraft von ΣWED , im Vergleich zu $RZCSa^{WQ}$, erhöht. Im Falle der Precision (Top2%-Auswahl) war beispielsweise $RZCSa^{WQ}$ bei Verwendung der KGS22-Wohnquartiergrenzen durchschnittlich 4,6 Prozent höher als ΣWED , mit den neu berechneten Wohnquartiergrenzen sind es 1,1 Prozent. Anzumerken ist allerdings, dass sich die Prognosegebiete sehr unterschiedlich in den Bezirken der KPБ verteilt haben, je nachdem ob ΣWED oder $RZCSa^{WQ}$ als Vorhersagevariable gewählt wurde. Im Falle von ΣWED sind es Wohnquartiere,

die grundsätzlich häufig von WED betroffen sind und deswegen eine höhere Grundwahrscheinlichkeit aufweisen. Im Falle von $RZCSa^{WQ}$ werden Wohnquartiere prognostiziert, in denen sehr aktuell Cluster von räumlich und zeitlich nahen WED-Vorgängen lagen.

Im Zusammenhang mit der Untersuchung zur zeitlichen Dynamik der WED-Prognosen veranschaulicht Abbildung 41 den durchschnittlichen Anteil Var_{Dyn} [%] nicht übereinstimmender Prognosegebiete (hier: KGS22-Wohnquartiere) zweier Kalenderwochen, basierend auf $Akt365_{RZC}$, $RZCSg^{WQ}$ und $RZCSa^{WQ}$ und für unterschiedliche Wochenabstände.

Abbildung 41: RZC-Score: Vergleich Zeitliche Dynamik der WED-Prognosen in Wochenabständen



Je höher Var_{Dyn} ist, desto mehr unterscheiden sich die Prognosegebiete zweier Kalenderwochen und desto dynamischer waren im Durchschnitt die mittels der jeweiligen Scores retrograd ermittelten Prognosegebiete im Zeitverlauf. Je geringer Var_{Dyn} , desto mehr Prognosegebiete sind identisch und desto statischer sind die Kriminalitätsprognosen. Hierbei verdeutlicht Abbildung 41, dass $Akt365_{RZC}$ und $RZCSa^{WQ}$, die ausschließlich ($Akt365_{RZC}$) oder primär ($RZCSa^{WQ}$) durch das letzte Tatzeitdatum der RZ-Cluster definiert werden, deutlich dynamischere Prognosen als $RZCSg^{WQ}$ erzeugten. Sowohl für $Akt365_{RZC}$ als auch für $RZCSa^{WQ}$ betrug Var_{Dyn} für den Abstand einer Kalenderwoche etwa 30 Prozent. Das Maximum von Var_{Dyn} wurde für $Akt365_{RZC}$ bei etwa zehn Wochen mit ca. 80 Prozent erreicht und für $RZCSa^{WQ}$ bei etwa 18 Wochen mit ca. 65 Prozent.

Bei Verwendung von $RZCSg^{WQ}$ betrug Var_{Dyn} für den Abstand einer Kalenderwoche ca. 10 Prozent. Das bedeutet, dass $RZCSg^{WQ}$ allein recht statische Kriminalitätsprognosen lieferte, die sich im Durchschnitt nur etwa zehn Prozent von der Prognose der Vorwoche unterscheiden. Obwohl $RZCSg^{WQ}$ auch die Aktualität der RZ-Cluster als einen Faktor berücksichtigt, fallen die Dauer und die Frequenz der Clustervorgänge bei $RZCSg^{WQ}$ stark ins Gewicht. Da sich die $RZCSg$ -Werte der einzelnen RZ-Cluster im Gesamtzeitraum bei der Datenaggregation auf Wohnquartierebene aufsummieren, dauert es länger, bis ein Wohnquartier mit einem sehr hohen $RZCSg^{WQ}$ -Wert in der Rangfolge deutlich sinkt, auch wenn andere Wohnquartiere womöglich kurz vor dem Prognosezeitpunkt sehr viel stärker von WED betroffen waren. Wie Abbildung 41 ebenfalls zeigt, nahm Var_{Dyn} für

RZCSg^{WQ} mit Zunahme des Wochenabstandes stetig zu. Nach etwa 35 Kalenderwochen, und damit deutlich später als bei Akt365_{RZC} und RZCSa^{WQ}, wurde das Maximum von Var_{Dyn} mit ca. 55 Prozent erreicht. Das heißt, dass sich die retrograd ermittelten Prognosegebiete auf Basis von RZCSg^{WQ} durchschnittlich am stärksten unterschieden, wenn sie etwa 35 Kalenderwochen auseinander lagen.

Var_{Dyn} war ab der jeweiligen Anzahl an Kalenderwochen, in dem das Maximum von Var_{Dyn} erreicht ist, zunächst etwa gleichbleibend und/oder nahm dann mit weiterer Zunahme des zeitlichen Abstandes zweier Kalenderwochen stetig ab. Abbildung 41 zeigt dies sehr deutlich. Die Anzahl identischer Prognosegebiete zweier Kalenderwochen nimmt demnach nicht linear mit Zunahme des Zeitabstandes ab. Der Verlauf von Var_{Dyn} ist vermutlich darauf zurückzuführen, dass die WED-Anfälligkeit bestimmter Gebiete in Abhängigkeit von ihrer Lage, ihrer Gebäude- und soziostrukturellen Merkmale jahreszeitlich schwankt. So ist z. B. erwartbar, dass sowohl Tages- und Jahreszeiten die zeitlichen und räumlichen Bewegungsmuster von Anwohnern und Tätern beeinflussen. Damit verbunden würde auch die Attraktivität bestimmter Wohngebiete für Täter jahreszeitlichen Schwankungen unterliegen. Diese Annahme wird auch von der hypothesengestützten Strukturanalyse zum WED getragen (vgl. LKA NRW 2012), spiegelt sich in den Ausführungen zum hypothesengestützten Vorgehen im Projekt SKALA (siehe Kapitel 3.2) wider und zeigte sich auch bei den Modellierungstests (siehe Kapitel 4.5.2). Letztere hatten, im Vergleich zur Analyse für das gesamte Jahr, eine verbesserte Vorhersagegenauigkeit von WED auf Basis soziostruktureller Daten ergeben, wenn die Analyse separat für Sommer und Winter (Zeitscheiben) durchgeführt wurde.

Es ist zu erwarten, dass die Dynamik der Prognosegebiete auf Basis von \sum WED noch geringer ist als für RZCSg^{WQ}, da \sum WED die Aktualität des letzten Einbruchs im Quartier gar nicht berücksichtigt. Aufgrund der geringen Dynamik der Prognosegebiete wurden regelmäßige WED-Prognosen auf Basis der Variablen \sum WED und RZCSg^{WQ}, trotz ihrer relativ hohen Vorhersagekraft, als wenig sinnvoll angesehen. Da RZCSa^{WQ} auch durchgehend höhere Wahrscheinlichkeiten im Vergleich zu allen anderen Variablen lieferte und eine höhere Dynamik aufwies als RZCSg^{WQ}, wurde RZCSa^{WQ} zunächst als Prognosevariable bevorzugt.

Zusammenfassend kann festgehalten werden, dass die hier untersuchten Variablen eine Vorhersage von Einbruchwahrscheinlichkeit ermöglichen, die für die ausgewählten Prognosegebiete deutlich über der Grundwahrscheinlichkeit liegen. Beispielsweise waren im Falle von RZCSa^{WQ} die jewei-

gen Wahrscheinlichkeiten des WED in ausgewählten Wohnquartieren durchschnittlich etwa 10-fach (Precision) und 3-fach (Recall) höher als im Vergleich zu einer zufälligen Gebietsauswahl. Die Analyse hat eindeutig gezeigt, dass von allen getesteten Variablen die prognostizierten Einbruchwahrscheinlichkeiten und damit die Vorhersagekraft für den Score RZCSa^{WQ} am höchsten war, der insbesondere die Aktualität der im Wohnquartier liegenden RZ-Cluster, aber auch die Dauer und die Frequenz der RZ-Cluster berücksichtigt. Das bedeutet, dass von allen vorgangsdatenbezogenen Variablen RZCSa^{WQ} am bedeutendsten für die wöchentliche Berechnung von Einbruchwahrscheinlichkeiten für die Folgewoche ist. Somit hat die Berechnung der RZ-Cluster und RZC-Scores (siehe Kapitel 4.6) einen eindeutigen Mehrwert im Hinblick auf die WED-Prognoseerstellung. Bei einem Vergleich der Recall- und Precision-Werte für die unterschiedliche Anzahl an Prognosegebieten wurde deutlich, dass die Precision mit Abnahme der Gebietsauswahl [%] zunimmt, während der Recall geringer ist, je weniger Gebiete ausgewählt werden. Die bisherigen Erfahrungen im Projekt SKALA zeigten hierbei, dass die Pilotbehörden mit einer kleineren Auswahl an Prognosegebieten (1,5-2 Prozent aller Wohnquartiere) gute Erfahrungen gemacht haben, insbesondere im Zusammenhang mit der Bewertung der WED-Prognosen und anschließender Planung von darauf aufbauenden polizeilichen Interventionen.

4.8 Prognose anderer Delikte

Neben der Betrachtung des WED wurde auch der Frage nachgegangen, inwiefern andere Deliktfelder möglicherweise mit den gleichen oder ähnlichen Modellen prognostiziert werden können. Für diese Prüfung kamen insbesondere solche Delikte in Betracht, die nicht statisch sind (sonst wären es Hotspot-Analysen) oder der Kontrollkriminalität zuzurechnen sind (z. B. Betäubungsmittelkriminalität). Um dem Anspruch einer umfassenden Prüfung gerecht zu werden, wurden deshalb im Projektverlauf Delikte, die in einem möglichen Tat- und Täterzusammenhang zum WED stehen, auf ihre Bedeutung bei der Modell- und Prognoseerstellung geprüft. Hierbei handelte es sich um Einbruchdiebstahl aus Gewerbeobjekten und Kraftfahrzeugdelikte.

Herausforderungen bei der Modell- und Prognoseerstellung zeigten sich vor allem in der Definition bzw. Eingrenzung des Deliktfeldes. Denn ein grundlegender Prozessbaustein ist die exakte Ab- und Eingrenzung des zu prognostizierenden Deliktes im Rahmen von Predictive-Policing-Verfahren.⁴⁰ Die nicht genaue Eingrenzung bedingt jedoch das Vermengen verschiedener Tat- und Täterstrukturen und erschwert die analytische Betrachtung. Aussagekräftige und belastbare Zusammenhänge auf Basis historischer Krimina-

⁴⁰ Das LAPD fasst beispielsweise WED und Einbruchdiebstahl aus Gewerbeobjekten als *burglary* zusammen und differenziert nicht weiter, wie sich bei einem Besuch im Dezember 2016 herausstellte.

litätsdaten können auf diese Art und Weise verfälscht werden. Die verschiedenen Modi Operandi oder Tatgelegenheitsstrukturen lassen bei einer nicht differenzierten Betrachtung keinen allgemeingültigen Schluss zu. Neben den vermengten Strukturunterschieden lässt sich auch eine auf das Prognoseergebnis ausgerichtete Intervention der Polizei nur schwer zielgerichtet steuern, denn verschiedene Vorgehensweisen der Täter oder delikt spezifisch gefährdete Orte erfordern unterschiedliche polizeiliche Interventionen. Deshalb ist in vielerlei Hinsicht die Definition bzw. Eingrenzung des Deliktfeldes bei der Umsetzung von elementarer Bedeutung.

Die Schwierigkeiten, die im Gegensatz zum relativ gut abgrenzbaren WED auftreten können, werden nachfolgend am Beispiel des Deliktfeldes Einbruchdiebstahl aus Gewerbeobjekten beschrieben:

Eine verbindliche Definition zum Einbruchdiebstahl aus Gewerbeobjekten ist nicht vorhanden. Auch die internationale kriminologische wie polizeiwissenschaftliche Forschung lässt dies bislang offen. Auf internationaler Ebene wird häufig nur zwischen *residential*, *non-residential*, *domestic* und *commercial burglaries* unterschieden (vgl. Prenzler 2009: 4f; Butler 2005: 29). Welche konkreten Gesetzesverstöße hierzu zählen bleibt offen. Sofern aber eine juristische Definition wie beim WED nicht vorliegt, muss das zu prognostizierende Delikt, hier der Einbruchdiebstahl aus Gewerbeobjekten, vorab zwingend definiert und beschrieben werden. Die Idee, zunächst eine Negativabgrenzung zum WED definitorisch vorzunehmen, zeigte jedoch im Projekt SKALA, dass ein solcher Datensatz eine Vielzahl von Delikten und Tatorten mit jeweils unterschiedlichen Spezifikationen vermengt und eine Zusammenhangsanalyse stark verzerrte bzw. unmöglich machte. So würde z. B. der Einbruch in einen metallverarbeitenden Betrieb mit dem Einbruch in eine Gartenlaube definitorisch gleichgesetzt. Um eine Ab- und Eingrenzung des zu prognostizierenden Deliktes zu realisieren wurde deshalb eine Ergänzung der juristischen Definition von §243 Abs. 1 Nr. 1 des Strafgesetzbuches (StGB)⁴¹ vorgenommen. Diese Ergänzung wurde mit der Polizeilichen Kriminalstatistik (PKS) durchgeführt, indem definitionsnahe Delikte anhand der PKS-Schlüsselnummern gezielt zusammengefasst wurden. So konnte einerseits die Nähe zur juristischen Deliktbeschreibung und andererseits der tatobjektbezogene Gewerbecharakter abgebildet werden. Für den Einbruchdiebstahl aus Gewerbeobjekten wurden die folgenden PKS-Schlüsselnummern verwendet:

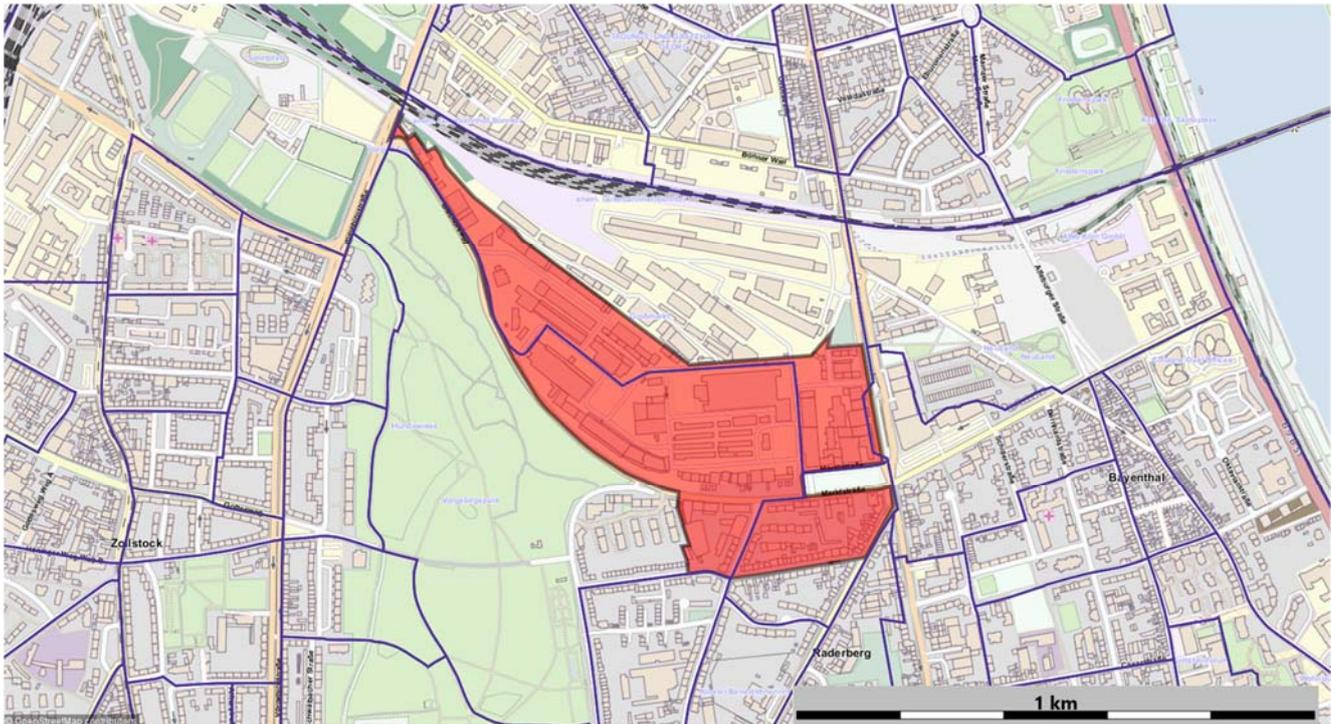
- 4000... Sonstiger „schwerer“ Diebstahl §§ 243 – 244a StGB
- 405... „Schwerer“ Diebstahl in/aus Banken, Sparkassen, Postfilialen und -agenturen
- 411... „Schwerer“ Diebstahl in/aus Diensträumen
- 412... „Schwerer“ Diebstahl in/aus Büroräumen
- 413... „Schwerer“ Diebstahl in/aus Werkstätten
- 414... „Schwerer“ Diebstahl in/aus Fabrikations- und Lagerräumen
- 418... „Schwerer“ Diebstahl in/aus Gaststätten und Kantinen
- 425... „Schwerer“ Diebstahl in/aus Kiosken, Warenhäusern, Verkaufsräumen, Selbstbedienungsläden, Schaufenstern, Schaukästen und Vitrinen (ohne Ladendiebstahl)
- 445... „Schwerer“ Diebstahl in/aus überwiegend unbezogenen Neu- und Rohbauten, Baubuden und Baustellen

Die Oberkategorie „Sonstiger schwerer Diebstahl §§ 243 – 244a StGB“ wurde bewusst inkludiert, da bei einer Erfassung der Straftaten innerhalb der Polizei erfahrungsgemäß in vielen Fällen zunächst keine Spezifizierung des Deliktes stattfindet. Anschließend mussten jedoch hier Prüfungen auf Fehlerfassungen oder unzutreffende Deliktqualifizierungen erfolgen, z. B. die Ausfilterung von nicht gewerblichen Tatornten wie Grünanlagen oder Gartenhäuser. Weitere Feinjustierungen in der Definition erfolgten mit den Pilotbehörden. Nach der definitorischen Festlegung mussten die Prognoseflächen angepasst werden. Für die Prognose von WED wurden, wie zuvor erläutert, Wohnquartiere verwendet. Der Einbruchdiebstahl aus Gewerbeobjekten findet jedoch i. d. R. (außer bei Mischbebauung) in Gewerbegebieten statt, sodass eine manuelle Anpassung der Prognoseflächen mit ArcGIS erfolgen musste. Abbildung 42 zeigt dies beispielhaft. Mit dieser Art und Weise der Definition von Einbruchdiebstahl aus Gewerbeobjekten lag abschließend eine entsprechende Ab- und Eingrenzung des zu prognostizierenden Deliktes vor.

Neben dem Einbruchdiebstahl aus Gewerbeobjekten wurden auch Kraftfahrzeugdelikte auf ihre Prognoseeignetheit getestet. Unter dem Begriff der Kraftfahrzeugdelikte wurden definitorisch Straftaten gefasst, die sich vorzugsweise auf Eigentumsdelikte im Bereich der Kfz-Kriminalität beziehen, wie der Diebstahl von, an und aus Kraftfahrzeugen, einschließlich der entsprechenden Qualifizierungstatbestände. Als Prognosefläche wurden Wohnquartiere gewählt.

⁴¹ „In besonders schweren Fällen wird der Diebstahl mit Freiheitsstrafe von drei Monaten bis zu zehn Jahren bestraft. Ein besonders schwerer Fall liegt in der Regel vor, wenn der Täter zur Ausführung der Tat in ein Gebäude, einen Dienst- oder Geschäftsraum oder in einen anderen umschlossenen Raum einbricht, einsteigt, mit einem falschen Schlüssel oder einem anderen nicht zur ordnungsgemäßen Öffnung bestimmten Werkzeug eindringt oder sich in dem Raum verborgen hält“.

Abbildung 42: Illustration Gewerbegebiet als Prognosegebiet



Zusammenfassend lassen sich der Einbruchdiebstahl aus Gewerbeobjekten und Kraftfahrzeugdelikte ebenfalls prognostizieren. Das bedeutet, es könnten Prognosen für Gewerbe- bzw. Wohnquartiere erstellt werden, die höhere Wahrscheinlichkeiten für einen Eintritt des jeweiligen Deliktes erwarten lassen. Die erhöhten Wahrscheinlichkeiten unterscheiden sich wie beim WED je nach Jahreszeit und Gültigkeitszeitraum. Für den Einbruch in gewerbliche Objekte wurden mit 14-tägigen Prognosen bis zu vierfach höhere Wahrscheinlichkeitswerte ermittelt. Für Kraftfahrzeugdelikte wurden bis zu dreifach höhere Wahrscheinlichkeitswerte, im Verhältnis zur durchschnittlichen Grundwahrscheinlichkeit, erreicht.

In diesem Zusammenhang ist anzumerken, dass die Prüfung der weiteren Delikte nicht derart ausführlich wie für den WED erfolgen konnte. Darüber hinaus wurden die soziostrukturellen Daten primär für die WED-Prognosen beschafft und sind daher nur eingeschränkt für Kriminalitätsprognosen in anderen Deliktfeldern nutzbar. Eine Modellverfeinerung und -verbesserung erfordert deshalb zukünftig auch den Einbezug und die Implementation von weiteren deliktsspezifischen Daten, z. B. zur Struktur von Gewerbegebieten.

4.9 Austausch und Kooperationen

Im Hinblick auf die unterschiedlichen Umsetzungsmöglichkeiten von Predictive Policing wurden über die gesamte Projektlaufzeit internationale wie nationale Treffen und Workshops besucht. Mit Beginn des Projektes wurde sich zunächst im Frühjahr 2015 über die Möglichkeiten der Umsetzung von Predictive Policing in Amsterdam (Niederlande) erkundigt. Ein entsprechender Gegenbesuch im LKA NRW konnte im Projektverlauf ebenfalls realisiert werden.

Durch die im November 2015 geschlossene Kooperationsvereinbarung mit dem Lehrstuhl für Datenanalyse und Visualisierung der Universität Konstanz⁴² wurde die kontinuierliche, fachliche Unterstützung des Projektes aus wissenschaftlicher Sicht gewährleistet. Mitarbeiter der Universität Konstanz unterstützten mehrfach während der Projektlaufzeit bei der Modell- und Prognoseerstellung. Darüber hinaus erfolgte die Programmierung des Visualisierungs- und Auswertetools SKALA | MAP durch Mitarbeiter der Universität Konstanz. Für die Gewährleistung eines kontinuierlichen Austausches wurden zudem regelmäßig gegenseitige Workshops und Treffen in Konstanz und Düsseldorf organisiert.

Vertreter der Geschäftsführung, des TP WBA und der KPB Duisburg besuchten im Dezember 2016 das Los Angeles Police Department (LAPD), um sich über methodische wie auch praktische Herausforderungen im Zusammenhang mit Predictive Policing auszutauschen. Das LAPD arbeitet seit

⁴² Rahmenvertrag mit der Steinbeis Transfer GmbH, vertreten durch Prof. Dr. Keim, Leiter des Steinbeis-Kompetenzzentrums Interaktive Datenanalyse und Visualisierung.

2011 mit der Software PREDPOL und prognostiziert mit einem Near-Repeat-Modell für die Deliktfelder Einbruch, Diebstahl von/aus Kraftfahrzeugen, Straßenraub, Vandalismus und Mord/Totschlag mit Waffen im öffentlichen Raum. PREDPOL bezieht keine soziostrukturellen Daten in die Modell- und Prognoseerstellung ein. Diese wird ausschließlich auf Basis der polizeilichen Daten zum Delikt, Tatort und zur Tatzeit erstellt. Als Prognoseflächen werden, im Gegensatz zu den Wohnquartieren im Projekt SKALA, 150 x 150m große Boxen gewählt. Abbildung 43 zeigt dies beispielhaft.

Abbildung 43: PREDPOL Boxen als Prognosegebiete



Bei Gesprächen mit operativen Polizeikräften des LAPD wurde deutlich, dass PREDPOL als gutes Hilfsmittel der alltäglichen Polizeiarbeit angesehen wird. Im Rahmen ihrer eigenen Evaluation habe sich zudem gezeigt, dass in den Gebieten, in denen PREDPOL Anwendung fand, eine deutliche Reduzierung der Gesamtkriminalität festzustellen war. Im Zusammenhang mit der Bekanntgabe der Prognosegebiete an operative Polizeikräfte, werden beim LAPD Karten im PDF-Format übermittelt. Eine dynamische Karte als IT-Anwendung, wie in NRW mit SKALA | MAP realisiert, ist nicht verfügbar und nicht geplant. Eine Live-Vorstellung von SKALA | MAP führte jedoch zu großem Interesse auf Seiten des LAPD.

Das TP WBA nahm darüber hinaus an der „4. Europäischen Konferenz für räumliche Kriminalitätsanalyse“ in München teil. Im Dezember 2016 erfolgte außerdem fernmündlich ein Erfahrungsaustausch mit einer Mitarbeiterin des Projektes CriPA („Crime Predictive Analytics“)⁴³ aus Österreich.

Weiterhin besuchte das TP WBA im März 2017 einen Diskussions- und Vernetzungsworkshop „Prognosetechnologien“ in Freiburg. An diesem Workshop nahm u. a. auch das MPICC (Max-Planck-Institut für ausländisches und internationales Strafrecht Abteilung Kriminologie), vertreten durch Herrn Prof. Dr. Oberwittler, teil. Herr Dominik Gerstner, ebenfalls vom MPICC, führte verantwortlich die Evaluationsstudie (vgl. Gerstner 2017) zur Prognosesoftware PRECOBS in Baden-Württemberg durch.

Darüber hinaus wurden vonseiten des Projektes zahlreiche Vorträge und Workshops bei der DHPol, dem LAFP NRW und in interessierten Polizeibehörden gehalten. Auf dem Thementag der Polizeilichen Euregio Rhein-Maas-Nord in Kleve (November 2017) wurde durch das TP WBA ein eintägiger Workshop zu Interventionsstrategien im Zusammenhang mit Predictive-Policing-Verfahren organisiert. In Lyon (Frankreich) wurde das Projekt SKALA auf einer internationalen Konferenz zum Thema Predictive Policing umfassend vorgestellt (Januar 2018).

Im medialen Kontext wurde häufig auf die besonders umfassende Prüfung der Methode von Predictive Policing durch NRW hingewiesen, sodass auch auf den letzten großen internationalen kriminologischen Tagungen, wie dem Stockholm Criminology Symposium (Juni 2017) und der Conference of the European Society of Criminology in Cardiff (September 2017), das Interesse am Projekt SKALA groß war. Die auf den jüngsten kriminologischen Kongressen geführten Diskussionen mit Mitarbeitern des Projektes SKALA verdeutlichten dessen hohe Reputation im wissenschaftlichen Umfeld.

Die zunehmende Prominenz des in NRW gewählten methodischen Vorgehens führte zudem dazu, dass sich der stellvertretende Polizeipräsident der Polizei Nanjing (China) im Rahmen eines Besuchs beim LKA NRW auch das Projekt SKALA vorstellen ließ. Weiterhin kamen auch Anfragen aus der Wirtschaft, wie beispielsweise von IBM, das Projekt SKALA bei der weltweiten Eröffnung des Watson IOT-Centers in München im Herbst 2017 vorzustellen. Durch diese Vorstellung wurde auch ein anschließender, vertiefender methodischer Austausch mit dem Präsidenten von ZITIS (Zentrale Stelle für Informationstechnik im Sicherheitsbereich) möglich.

⁴³ Das Institut für Wirtschafts- und Innovationsforschung der JOANNEUM RESEARCH Forschungsgesellschaft mbH erforscht im Rahmen des Projektes prädiktive Analysemethoden zur Bekämpfung von Kriminalität.

5 Praktische Umsetzung der Prognosen

Im fünften Abschnitt wird die praktische Umsetzung der Kriminalitätsprognosen in den Pilotbehörden näher beschrieben. Hierfür wird zunächst erläutert, welcher organisatorische Aufwand in den Pilotbehörden für die Implementierung von SKALA erforderlich war (Kapitel 5.1). Anschließend wird darauf eingegangen, welche vorbereitenden Maßnahmen für die praktische Umsetzung notwendig waren (Kapitel 5.2), wie sich die konkrete Umsetzung im Praxisbetrieb gestaltete (Kapitel 5.3) und welche Implementierungsempfehlungen vonseiten der beteiligten Pilotbehörden gegeben werden können (Kapitel 5.4).

5.1 Beschreibung der Organisation

Die Organisation der Teilprojekte in den Pilotbehörden erfolgte individuell und mit Fokus auf die jeweiligen besonderen Bedingungen. Nachfolgend werden aber die wesentlichen organisatorischen Strukturen beschrieben, die grundsätzlich bei allen Pilotbehörden im Projekt SKALA vorherrschend waren.

Die Teilprojekte wurden alle durch einen Beamten des höheren Polizeivollzugsdienstes der Direktion Kriminalität geleitet. Die Anzahl der involvierten Mitarbeiterinnen und Mitarbeiter in den unterschiedlichen Organisationseinheiten belief sich durchschnittlich auf etwa fünf.

Die im Rahmen des Maßnahmenpaketes der Landesregierung NRW einzelnen KPB zugeteilten Stellen für Regierungsbeschäftigte (Befristung des Arbeitsvertrages auf zwei Jahre) konnten mit qualifizierten Bediensteten, überwiegend aus den Fachrichtungen der Informatik und Geografie, besetzt werden. Die organisatorische Anbindung erfolgte überwiegend in den jeweiligen Analysestellen.

Im Projektzusammenhang standen der Teilprojektleitung und der Geschäftsführung in allen Pilotbehörden namentlich benannte Ansprechpartner in den Direktionen Gefahrenabwehr/Einsatz, Verkehr und dem Leitungsstab zur Verfügung. Dies gewährleistete den direktionsübergreifenden Informationsaustausch und erleichterte die gemeinsame Abstimmung bei Kräfte- und Einsatzkoordination.

5.2 Vorbereitung der Umsetzung

Der in der Vorbereitungsphase zu leistende personelle und technische Aufwand war generell von den Faktoren

- Größe der Behörde,
- technische Infrastruktur der Behörde und
- Auswerte- und Analysekompetenz in der Behörde

abhängig.

Im Zusammenhang mit der Größe der jeweiligen Behörde und der damit verbundenen Anzahl an Entscheidungsebenen waren viele organisatorische Entscheidungen im Vorfeld erforderlich, um das Projekt SKALA sachgerecht platzieren und koordiniert umsetzen zu können. Durch das

besondere mediale Interesse an Predictive-Policing-Verfahren und die neue Art und Weise, präventiv polizeilich tätig zu werden, nahm die interne Öffentlichkeitsarbeit einen herausragenden Stellenwert ein. Zu diesem Zweck wurde ein Kommunikationsworkshop der Pressestellen zur Formulierung von Kernbotschaften für SKALA veranstaltet. Die interne Öffentlichkeitsarbeit war während des gesamten Projektverlaufes von besonderer Bedeutung. So war es erforderlich, regelmäßig Transparenzveranstaltungen und Workshops innerhalb der Pilotbehörden auf allen Ebenen durchzuführen. Projektbezogene Informationen, vor allem zum Hintergrund, den Zielen, der Methodik und den polizeilichen Interventionen, wurden ergänzend auf den lokalen Intranetseiten der jeweiligen Pilotbehörden publiziert. Vereinzelt wurden auch Dienstanweisungen zur konkret geplanten Umsetzung erlassen, in denen z. B. Abläufe erläutert, Hinweise zur operativen Umsetzung gegeben und Ansprechpartner benannt wurden.

Für die interne Dokumentation mussten die Wohnquartiere als räumliche Bezugsgrößen in dem Einsatzleitsystem eCebius hinterlegt werden. Dies ermöglichte allen Pilotbehörden die Zuordnung von polizeilichen Einsätzen und Interventionen zu diesen Gebieten. Die Recherche von Einsätzen in den Prognosegebieten erforderte zudem eine Implementierung eigener Schlagworte zum SKALA (z. B. SKALA-WED) in eCebius durch das LZPD NRW.

5.3 Umsetzung

Die durch das TP WBA übermittelten Kriminalitätsprognosen wurden in den Pilotbehörden zunächst einer kriminalfachlichen Bewertung unterzogen, um das polizeiliche Erfahrungswissen stets einzubinden. Da die Prozessschritte im Einzelnen sehr different sind, werden diese im Folgenden so beschrieben, wie sie sich grundsätzlich bei allen beteiligten Pilotbehörden darstellten. Es wird darauf verzichtet lokale Besonderheiten der konkreten Umsetzung zu erörtern:

1. Schritt: Übermittlung

Die Kriminalitätsprognosen wurden nach der Übermittlung zunächst gesichtet, um erste mögliche Schwerpunkte zu erkennen. Diese erste Sichtung konnte z. B. anhand einer kartografischen Darstellung erfolgen.

2. Schritt: Bewertung und Auswahl

Im zweiten Schritt erfolgte die detaillierte kriminalfachliche Bewertung und Auswahl der Prognosen. Die Art und Weise erfolgte zwischen den Pilotbehörden sehr heterogen, wurde aber grundsätzlich von Bediensteten aus entsprechenden Fachkommissariaten oder Auswertedienststellen durchgeführt. In den meisten Fällen wurde zur Unterstützung der Bewertung die retrograde Kriminalitätslage hinzugezogen.

Nach einer Einschätzung zur Kriminalitätslage und den übermittelten Prognosen wurden diese auf polizeiliche Interventionsmöglichkeiten hin geprüft. Diese Prüfung umfasste in einigen Pilotbehörden die regelmäßige Diskussion in entsprechenden (Brennpunkt-)Konferenzen. In diesen wurde regelmäßig ein Maßnahmenpaket mit allen beteiligten Direktionen der KPB abgestimmt, dass sich u. a. an zeitlichen, taktischen und personalplanerischen Bedingungen orientierte. Konkrete inhaltliche Kriterien der Auswahl bzw. Abgrenzung waren z. B.:

- Dichtekarten/Heatmaps
- Fallentwicklungen in bestimmten Gebieten
- Aktuelle Fallentwicklung in der Prognosewoche
- Erkenntnisse aus Ermittlungskommissionen
- Beobachtungs- und Feststellungsberichte
- Großeinsätze/Veranstaltungen

In vielen Fällen wurde zur Erkennung von Tatserien jeder angezeigte WED im Hinblick auf Tatort, Tatzeit und Modus Operandi ausgewertet. Bei Übereinstimmungen wurden die Taten intern als Serie definiert und bei der fachlichen Bewertung und Auswahl der Kriminalitätsprognosen entsprechend berücksichtigt.

3. Schritt: Rückmeldung

Die von den Pilotbehörden für offene polizeiliche Interventionen ausgewählten Prognosegebiete wurden im Anschluss an das LKA NRW zurückübermittelt, um eine entsprechende Web-Visualisierung zu ermöglichen (siehe Kapitel 4.3.4).

4. Schritt: Operative Umsetzung

Die ausgewählten Prognosegebiete wurden im Anschluss an die Bewertung und Rückmeldung (Schritt 2 und 3) den operativen Polizeikräften visuell übermittelt. Die Visualisierung erfolgte in den Pilotbehörden auf unterschiedliche Art und Weise und konnte mit eigens erstellten Übersichts- und Detailkarten, SKALA | MAP oder der Web-Visualisierung realisiert werden. In den Pilotbehörden wurde ein internes Controlling geführt. Dieses ermöglichte in Einzelfällen, dass in einer Prognosewoche die noch nicht mit polizeilichen Interventionen belegten Prognosegebiete gezielt ermittelt und

im Anschluss noch ausstehende Maßnahmen getroffen werden konnten.

Im Wesentlichen wurden von den KPB folgende polizeiliche Interventionen geplant und umgesetzt:

- Verdeckte Aufklärungsmaßnahmen
- Verkehrskontrollen an Knotenpunkten
- Verkehrskontrollen an Autobahn-Anschlussstellen
- Präsenz in Prognosegebieten
- Präsenz im ÖPNV
- Präventionsberatung.

Für eine detaillierte Auflistung der Polizeiaktivitäten wird auf den gesonderten Evaluationsbericht verwiesen.

Darüber hinaus wurden im Projektzeitraum mehrere landesweite Aktionstage zur Bekämpfung des WED durchgeführt. In diesem Rahmen wurden die flächendeckenden Großeinsätze unter Berücksichtigung der Prognosen konzipiert und ausgerichtet.

5.4 Bewertung der Umsetzung

In diesem Abschnitt wird auf die Bewertung der Umsetzung und die damit verbundenen Prozesse in den Pilotbehörden sowie Herausforderungen und Schwierigkeiten eingegangen. Es erfolgt keine Bewertung der Methode Predictive Policing bzw. des Gesamtprojektes. Im Zusammenhang mit einer möglichen Einführung von SKALA in der Polizei NRW wurde vonseiten der beteiligten Pilotbehörden zudem eine Vielzahl an Erfahrungen und Empfehlungen diskutiert, die nachfolgend in wesentlichen Teilen dargestellt werden.

Erfahrungen

- Die operative Umsetzung erforderte in den ersten Wochen nach Projektstart einen erhöhten personellen Ressourceneinsatz.
- Die im Rahmen des Maßnahmenpaketes der Landesregierung NRW den KPB zugeteilten zusätzlichen Personalressourcen konnten mit qualifiziertem Personal, insbesondere aus dem Bereich Informatik/Geovisualisierung, erfolgreich besetzt werden. Dieses zusätzliche Personal war für den Einsatz von SKALA nicht erforderlich, steigerte aber die Analysekompetenz der KPB. Sollten den KPB zukünftig weitere Regierungsbeschäftigtenstellen zur Verfügung gestellt werden, bietet es sich an, die Auswahl des Fachpersonals entsprechend vorzunehmen und so eine Stärkung der Auswert- und Analysekompetenz herbeizuführen.
- Im Zusammenhang mit der Wahl des Wohnquartiers als räumliche Bezugsgröße wurde signalisiert, dass der Einsatz in kleinen, genau definierten Einsatzräumen insgesamt positiv sei und einer Rasterlösung grundsätzlich vorzuziehen ist.
- Die für das Projektcontrolling erforderliche Einbindung der Wohnquartiere in die bestehenden Struk-

turen des Einsatzleitsystems sowie die Auswertungen von getroffenen polizeilichen Interventionen anhand der neu eingerichteten Schlagworte wurden als sehr zeitintensiv beschrieben, da notwendige Schnittstellen, aufgrund der in Teilen veralteten IT-Strukturen dieser Systeme, nicht vorhanden oder nicht kompatibel waren.

Empfehlungen

- Jeder Polizeibezirk weist unterschiedliche Strukturen auf, die es bei einer entsprechenden Einführung von SKALA zu beachten gilt.
- Eine direktionsübergreifende Umsetzung ist essentiell.
- Die sachgerechte Bewertung und Auswahl der vom LKA NRW übermittelten Prognosen wie auch die Planung und Durchführung darauf aufbauender Interventionen kann nur gelingen, wenn sich alle beteiligten Direktionen entsprechend einbringen.
- Eine Beteiligung von Fachpersonal, z. B. aus Fachkommissariaten oder Auswertedienststellen, ist zwingend erforderlich.
- Die begleitende interne Öffentlichkeitsarbeit ist von entscheidender Bedeutung. Das Interesse und die Motivation der am Projekt beteiligten Bediensteten im operativen Dienst waren sehr hoch und angebotene Transparenzveranstaltungen wurden stark frequentiert. Die im Vorfeld durchgeführten Veranstaltungen, Workshops und Intranet-Publikationen gestalteten sich zeitaufwendig, waren aber nach den Rückmeldungen der Pilotbehörden zwingend erforderlich, um das notwendige Maß an Transparenz und Verständnis für das Projekt zu gewährleisten.
- Über diese Interventionen hinaus wird eine zentrale Informationsveranstaltung zur Einführung als hilfreich erachtet. Ergänzend könnten Multiplikatoren in den Behörden Schulungen (Wissenstransfer) vornehmen.

6 Fazit und Ausblick

Im letzten Abschnitt des Berichtes werden zunächst die Ergebnisse bewertet (Kapitel 6.1). Anschließend wird ein Ausblick gegeben (Kapitel 6.2).

6.1 Bewertung der Ergebnisse

Ziele des Projektes waren (1) die Prüfung von Möglichkeiten und Grenzen der Prognose von Kriminalitätsbrennpunkten sowie (2) die Prüfung der Effizienz und Effektivität darauf aufbauender polizeilicher Interventionen. In diesem Bericht wurde das erste Ziel fokussiert. Für die Ergebnisse zum zweiten Ziel wird auf den gesonderten Evaluationsbericht verwiesen.

Im Projekt SKALA wurden Wahrscheinlichkeiten von WED sowie anderer Deliktfelder auf Basis raumbezogener Daten für Wohnquartiere in ausgewählten Polizeibezirken erfolgreich berechnet. Diese Kriminalitätsprognosen wurden auf Grundlage eines hypothesengestützten Verfahrens erstellt. Diese Verfahrensweise gewährleistete, dass die Modell- und Prognoseerstellung auf belastbaren wissenschaftlichen Theorien und Forschungsbefunden basiert. Hierdurch grenzt sich die Vorgehensweise von vielen anderen Predictive-Policing-Verfahren ab, die häufig nur auf dem Near-Repeat-Ansatz beruhen. Darüber hinaus wurden systematisch polizeiliche und soziostrukturelle Daten auf ihre Aussagekraft im Zusammenhang mit Prognosen geprüft. Es wurden Kriminalitätsprognosen für WED, Einbruchdiebstahl aus Gewerbeobjekten und Kraftfahrzeugdelikte erstellt. Für alle drei Deliktfelder konnten häufig drei- bis vierfach erhöhte Kriminalitätswahrscheinlichkeiten im Vergleich zur Grundwahrscheinlichkeit berechnet werden. Hierdurch hat sich gezeigt, dass Kriminalitätsprognosen für unterschiedliche Deliktfelder erstellt werden können.

SKALA wurde ferner erfolgreich in unterschiedlichen KPB angewandt. Hierbei handelte es sich im Wesentlichen um städtisch geprägte Regionen. Weiter wurde geprüft, inwiefern die Prognosemodelle auch für ländlich geprägte Regionen geeignet sind. Als Muster wurde hierfür u. a. die KPB Bonn ausgewählt, da sie sowohl sehr urbane wie auch ländliche Regionen aufweist. In diesem Kontext zeigte sich, dass in ländlicheren Regionen keine automatische Übertragbarkeit der bereits erstellten Prognosemodelle möglich war. Zudem zeigte sich, dass auf Grund eines relativ geringeren Fallzahlenaufkommens die Erstellung wöchentlicher Kriminalitätsprognosen in solchen Bezirken nicht zielführend ist. Durch Analysen, die auf die strukturellen Unterschiede zwischen städtisch und ländlich geprägten Regionen fokussieren, konnten allerdings Ansätze gefunden werden, die die Ermittlung von Einbruchwahrscheinlichkeiten in ländlicheren Regionen ermöglichen könnten. Eine abschließende Bewertung erfordert jedoch eine tiefergehende Prüfung.

Die Berechnung der Kriminalitätswahrscheinlichkeiten bezog sich jeweils auf die Gesamtfläche der KPB. Mit dieser Vorgehensweise wurde gewährleistet, dass für jedes Wohnquartier in den ausgewählten Polizeibezirken eine eigene individuelle Anfälligkeit für WED in der Prognosewoche ermittelt wurde. Viele andere Predictive-Policing-Verfahren beziehen sich lediglich auf Teilbereiche von Städten oder Regionen. Die Wohnquartiergrenzen im Projekt SKALA orientierten sich anfänglich an früheren Stimmbezirken mit einer ungefähren Gebietseinheit von 400 Haushalten. Im Projektverlauf sind in Zusammenarbeit mit dem Lehrstuhl für Datenanalyse und Visualisierung der Universität Konstanz die Wohnquartiergrenzen überarbeitet worden, um eine höhere Homogenität in den Quartieren zu erreichen. Dies ermöglichte eine höhere Modellgüte.

Die Modell- und Prognoseerstellung erfolgte unter Verwendung von Entscheidungsbaummodellen. Darüber hinaus wurden weitere technisch-methodische Analysemodelle geprüft. Entscheidungsbaummodelle hatten eine vergleichsweise gute Performance. Sie sind zudem transparent und nachvollziehbar, sodass sie im Rahmen des Projektes favorisiert wurden. Darüber hinaus wurden auch die Einflussstärken der soziostrukturellen Daten auf WED geprüft. Hierbei zeigen die Ergebnisse zusammenfassend, dass sich die Einflussstärken der Variablen je nach Jahreszeit und Bezirk stark unterscheiden. Dementsprechend sind die Ergebnisse nicht automatisch auf andere Bezirke oder Zeiträume übertragbar.

Die Modell- und Prognoseerstellung erfolgte im Wesentlichen mit dem IBM SPSS Modeler. Der Vorteil dieser Anwendung ist die verhältnismäßig leichte Bedienbarkeit und die Möglichkeit, erste Datenanalyseaufgaben, einschließlich einer Modell- und Prognoseerstellung, innerhalb kurzer Zeit durchführen zu können. Nachteile liegen in den nicht ausreichend vorhandenen Schnittstellen zu anderen Softwarelösungen, wie beispielsweise ArcGIS, Python oder R. Darüber hinaus stehen nur eingeschränkte Möglichkeiten zur Verarbeitung und Analyse räumlicher Daten (z. B. Transformation von Koordinaten) zur Verfügung. Somit erforderten beispielsweise Analysen, wie die Berechnungen der RZ-Cluster, die Programmierung von selbst konzipierten R-Skripten in Kombination mit komplexen SQL-Abfragen.

Im Projektzeitraum wurde über einen Discoverer-Zugang seitens des LZPD NRW gewährleistet, dass eigenständige Erhebungen von polizeilichen Vorgangsdaten durch das TP

WBA jederzeit vorgenommen werden konnten. Diese Abfrage- und Transferrountinen waren im Rahmen des Projektes zeitlich überschaubar. Durch das Arbeiten in einem Sondernetz mussten die Daten jedoch aus dem Polizeinetz auf einen externen Datenträger (USB-Stick) transferiert werden, um sie anschließend wieder in das Sondernetz einzuspeisen. Dies gestaltete sich zeitaufwendig.

Zur Prognosevisualisierung für die Pilotbehörden wurde in Zusammenarbeit mit dem Lehrstuhl für Datenanalyse und Visualisierung der Universität Konstanz das Visualisierungs- und Auswertetool SKALA | MAP entwickelt. SKALA | MAP ermöglicht es, auf der Grundlage digitaler Karten, die durch das LKA NRW erstellten Prognosen darzustellen, anzupassen und den Organisationseinheiten für polizeiliche Interventionen zur Verfügung zu stellen. Es bietet den KPB die Möglichkeit der räumlichen Darstellung von Delikten auch unter Nutzung der Visualisierung von Heatmaps. SKALA | MAP unterstützt folglich den Prozess der vorbeugenden Bekämpfung von Kriminalität unter Zuhilfenahme wissenschaftlicher Prognosetechniken und ermöglicht den KPB eine komfortable Auswahl und Darstellung der Prognosegebiete. Positive Rückmeldungen aus den Pilotbehörden zu SKALA | MAP führten zu einer Zertifizierung der Software durch das LZPD NRW, sodass die Anwendung künftig durch alle Polizeibehörden zur Darstellung und Auswertung von Kriminalität genutzt werden kann.

6.2 Ausblick

Im Zusammenhang mit der Prüfung der Möglichkeiten und Grenzen der Prognose von Kriminalitätsbrennpunkten wurden verschiedene Forschungsbedarfe festgestellt, die es im Rahmen potenziell weiterer Forschungstätigkeiten in diesem Feld zu berücksichtigen gilt:

Ein hypothesengestütztes Vorgehen hat sich im Projekt SKALA bewährt und sollte bei Kriminalitätsprognosen anderer Deliktfelder berücksichtigt werden. Im Projekt SKALA wurden drei Deliktfelder betrachtet (WED, Einbruchdiebstahl aus Gewerbeobjekten und Kraftfahrzeugdelikte). Eine Erweiterung auf andere Deliktfelder, beispielsweise Raub, ist anzustreben. Dabei ist zu berücksichtigen, dass die bisher im Projekt verwendeten Wohnquartiere nicht ohne weiteres für andere Deliktfelder geeignet sind, sondern jeweils angepasst werden müssen.

Im Kontext von räumlichen Bezugsgrößen (z. B. Wohnquartieren) rückt perspektivisch auch eine Analyse von Kriminalität in Mikrosegmenten in den Fokus. David Weisburd, weltweit führender Vertreter der *place-based criminology*, konnte in seinen Studien zeigen, dass sich in verschiedenen Städten die Hälfte der Gesamtkriminalität in nur fünf bis sieben Prozent der Straßenabschnitte (Mikrosegmente) ereignet (vgl. Steenbeck/Weisburd 2016; Weisburd et al. 2012;

Weisburd/Amram 2014). Da sich im Projekt SKALA ebenfalls gezeigt hat, dass eine große Anzahl von Tatorten im Zeitverlauf wiederholt von Einbrüchen betroffen war, wurde eine erste Analyse dieser Art für die Städte Essen und Mülheim an der Ruhr bereits durchgeführt (vgl. Seidensticker 2017). Die Ergebnisse sind vielversprechend, sodass dieser Ansatz zukünftig stärker fokussiert werden sollte. Eine präzisere Datenaggregation, z. B. auf Straßenabschnittsebene, würde es ermöglichen, präventive Bemühungen kleinräumiger zu konzentrieren. Darüber hinaus könnte eine Übertragung dieser Mikrosegment-Erkenntnisse zur Verfeinerung der Prognosemodelle genutzt werden.

SKALA wurde erfolgreich auf unterschiedliche städtisch geprägte Polizeibezirke angewandt. Es hat sich allerdings gezeigt, dass eine Übertragung des Modells auf ländliche Regionen nicht ohne weiteres möglich ist. Die Erstellung von Kriminalitätsprognosen für ländlich geprägte Polizeibezirke setzt voraus, dass die Modell- und Prognoseerstellung inhaltlich sinnvoll auf die jeweiligen soziostrukturellen Gegebenheiten angepasst wird.

Für die Modell- und Prognoseerstellung wurden Entscheidungsbaummodelle in Form von Einbaum-Modellen genutzt. Mit Bezug auf die konkreten Algorithmen und eine Modellweiterentwicklung wäre zukünftig ein Random-Forest-Modell zu bevorzugen, da es im Projekt SKALA, im Vergleich zu den Einbaum-Modellen, nachweislich robustere Analyseergebnisse lieferte. Die Modell- und Prognoseerstellung basiert auf polizeilichen und soziostrukturellen Daten. Die soziostrukturellen Daten müssen in regelmäßigen Abständen aktualisiert werden. Darüber hinaus muss stetig geprüft werden, inwiefern neue Datenquellen im Rahmen der geltenden Datenschutzrichtlinien zur Verbesserung der Prognosemodelle beitragen können. Weiterhin zeigte sich in diesem Zusammenhang, dass die Modellgüte entscheidend von der Qualität und zeitlichen Verfügbarkeit der Daten abhängt. Erste Schritte für die Automatisierung einzelner Analyseverfahren wurden bereits im Projektverlauf mit den Programmiersprachen R und Python realisiert.

Die Nutzung des Discoverer-Zuganges des LZPD NRW erwies sich im Projekt SKALA als zeitaufwendig. Eine komfortablere und schnellere IT-Lösung ist anzustreben. An dieser Stelle wären Netzwerklaufrwerke denkbar. Im Zusammenhang mit der Datenqualität würde eine bessere und schnellere Verfügbarkeit der Daten die Analysequalität steigern und so möglicherweise eine bessere Modellgüte ermöglichen. Der Umstand, dass bestimmte Informationen zum Tatgeschehen erst relativ spät im Laufe des Ermittlungsverfahrens Eingang finden, z. B. Informationen zur Beuteart und -höhe, erschweren jedoch diesen Anspruch. Ergänzend wäre es für unstrukturierte Datensätze wie Kurzsachverhalte denkbar, Text-Mining-Methoden anzuwenden.

Abbildungsverzeichnis

| | |
|--|-------|
| Abbildung 1: Zeitleiste | S. 4 |
| Abbildung 2: Projektorganisation | S. 5 |
| Abbildung 3: Ansprüche und Grundsätze | S. 9 |
| Abbildung 4: Datenbasis | S. 22 |
| Abbildung 5: SKALA-Logo | S. 26 |
| Abbildung 6: Prediction-Led Policing Business Process | S. 26 |
| Abbildung 7: Predictive-Policing-Prozess | S. 27 |
| Abbildung 8: Illustration Wohnquartiere und Boxen als Prognosegebiete | S. 29 |
| Abbildung 9: Screenshot Clustering-Tool | S. 30 |
| Abbildung 10: Einflussdimensionen | S. 34 |
| Abbildung 11: Illustration der Auswirkung variabler Gültigkeitszeiträume auf die „Trefferate“ | S. 36 |
| Abbildung 12: Illustration Randtreffer | S. 37 |
| Abbildung 13: Illustration WED-Wahrscheinlichkeiten des Polizeibezirks Köln | S. 40 |
| Abbildung 14: Beispiel-Stream: Datenaufbereitung | S. 42 |
| Abbildung 15: Beispiel-Stream: Georeferenzierung | S. 43 |
| Abbildung 16: Beispiel-Stream: Prognoseerstellung, 1. Teil | S. 44 |
| Abbildung 17: Beispiel-Stream: Prognoseerstellung, 2. Teil | S. 47 |
| Abbildung 18: Beispiel-Übersichtskarte Köln | S. 47 |
| Abbildung 19: Beispiel-Detailkarte Köln | S. 47 |
| Abbildung 20: Beispiel-Übersichtskarten KPB Köln aus SKALA MAP (Prognose mit Heatmap) | S. 48 |
| Abbildung 21: Beispiel-Übersichtskarten KPB Köln aus SKALA MAP (Prognose mit Heatmap im Detail) | S. 49 |
| Abbildung 22: Beispiel-Übersichtskarten KPB Köln aus SKALA MAP (Prognose mit allen Wahrscheinlichkeiten) | S. 49 |
| Abbildung 23: Beispiel-Prognose, web-basiert | S. 50 |
| Abbildung 24: Beispiel-Prognose (Detail), web-basiert | S. 50 |
| Abbildung 25: Beispiel-CART-Klassifizierungsbaum | S. 53 |
| Abbildung 26: Datenaufteilung für die Modellerstellung und Validierung mittels Accuracy Index | S. 57 |
| Abbildung 27: Korrelationsmatrix: Soziostrukturelle Daten und Anzahl WED im Zeitraum 2012-2016 | S. 59 |

| | |
|---|-------|
| Abbildung 28: Korrelationsmatrix: Soziostrukturelle Daten und Anzahl WED im Zeitraum 2012-2016 nach Bezirken | S. 60 |
| Abbildung 29: Vergleich Modellperformance anhand der Cross Validation Accuracy | S. 61 |
| Abbildung 30: Einflussstärke der soziostrukturellen Daten beim Random-Forest-Modell (Gesamtdatensatz) | S. 62 |
| Abbildung 31: Einflussstärke der soziostrukturellen Daten beim Random-Forest-Modell nach Bezirken | S. 63 |
| Abbildung 32: Illustration Raum-Zeit-Cluster | S. 64 |
| Abbildung 33: Analyseschritte Ermittlung RZC-Scores | S. 64 |
| Abbildung 34: RZC-Scores (RZCSa > RZCSg) | S. 67 |
| Abbildung 35: Illustration RZC-Scores KPB Düsseldorf auf Wohnquartierbenbe | S. 67 |
| Abbildung 36: RZCSg ^{WQ} (links) und RZCSa ^{WQ} (rechts) für Wohnquartiere in Köln (oben) und Bonner Raum (unten) | S. 68 |
| Abbildung 37: RZC-Score: Vergleich (Precision und Recall) für die Top2% der Prognosegebiete | S. 72 |
| Abbildung 38: RZC-Score: Vergleich (Precision und Recall) für die Top2% der Prognosegebiete nach Pilotbehörden | S. 72 |
| Abbildung 39: RZC-Score: Vergleich (Precision und Recall) für die Top% der Prognosegebiete | S. 73 |
| Abbildung 40: RZC-Score: Vergleich für Wohnquartiere | S. 73 |
| Abbildung 41: RZC-Score: Vergleich Zeitliche Dynamik der WED-Prognosen in Wochenabständen | S. 74 |
| Abbildung 42: Illustration Gewerbegebiet als Prognosegebiet | S. 77 |
| Abbildung 43: PREDPOL Boxen als Prognosegebiete | S. 78 |

Tabellenverzeichnis

| | |
|--|-------|
| Tabelle 1: Illustration Konfusionsmatrix | S. 34 |
| Tabelle 2: Quartierauswahl | S. 40 |
| Tabelle 3: Polizeiliche Vorgangsdaten aus IGVP | S. 41 |
| Tabelle 4: Konfusionsmatrix zur Modellvalidierung | S. 55 |
| Tabelle 5: Ausgewählte Attribute soziostruktureller Daten als Prädikatoren für eine Modellierung | S. 58 |
| Tabelle 6: Beispielauswahl: Merkmale von RZ-Clustern | S. 67 |
| Tabelle 7: RZCS für WQ = 05111000100382 | S. 68 |
| Tabelle 8: Variablen RZC-Scores-Evaluierung | S. 70 |
| Tabelle 9: Top-Auswahl der Prognosegebiete | S. 71 |

Literatur

- Abt, J./Schröder, A. (2017): Städtebauliche Kriminalprävention. In *Bauwelt* 6/2017, S. 40-47.
- Bailey, T./Gatrell, A. (1995): *Interactive Spatial Data Analysis*. Harlow.
- Balogh, D. (2016): Near Repeat-Prediction mit PRECOBS bei der Stadtpolizei Zürich. In *Kriminalistik*, 5/2016, S. 335-341.
- Belina, B. (2009): Kriminalitätskartierung - Produkt und Mittel neoliberalen Regierens, oder: Wenn falsche Abstraktionen durch die Macht der Karte praktisch wahr gemacht werden. In *Geographische Zeitschrift* 9 (4), S. 192-212.
- Berk, R./Sherman, L./Barnes, G./Kurtz, E./Ahlman, L. (2009): Forecasting murder within a population of probationers and parolees: a high stakes application of statistical learning. In *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, H.1, Jg. 172, S. 191-211.
- Berk, R. (2008): Forecasting Methods in Crime and Justice. In *Annual Review of Law and Social Science*, H.1, Jg. 4, S. 219-238.
- Bernasco, W. (2008): Them Again? Same-Offender Involvement in Repeat and Near Repeat Burglaries. In *European Journal of Criminology* 5, S. 411-431.
- Bernasco, W./Nieuwebeerta, P. (2005): How do Residential Burglars Select Target Areas? A New Approach to the Analysis of Criminal Location Choice. In *British Journal of Criminology* 44, S. 296-315.
- Berthold, M./Borgelt, C./Höppner, F./Klawonn, F. (2010): *Guide to intelligent data analysis: how to intelligently make sense of real data*. London.
- Bowers, K./Johnson, S. (2005): Domestic Burglary Repeats and Space-Time Clusters. The Dimensions of Risk. In *European Journal of Criminology* 2, S. 67-92.
- Box, G./Jenkins, G./Reinsel, G./Ljung, G. (2015): *Time series analysis. Forecasting and Control*, Fifth Edition. New Jersey.
- Bode, F. (2013): *Beurteilung und Wertung richterlicher Entscheidungen durch Polizeibeamte. Eine empirische Untersuchung am Beispiel jugendlicher Mehrfach- und Intensivtäterkriminalität*. Hamburg.
- Bode, F./Stoffel, F./Keim, D. (2017): Variabilität und Validität von Qualitätsmetriken im Bereich von Predictive Policing. In Konstanzer Online-Publikations-System (KOPS), URL: <http://nbn-resolving.de/urn:nbn:de:bsz:352-0-402496>, zuletzt aufgerufen im Januar 2018.
- Bowers, K./Johnson, S. (2005): Domestic Burglary Repeats and Space-Time Clusters. The Dimensions of Risk. In *European Journal of Criminology* 2, S. 67-92.
- Braga, A. (2005): Hot spots policing and crime prevention: A systematic review of randomized controlled trials. In *Journal of Experimental Criminology*, Vol. 1, S. 317-342.
- Brantingham, P. (2011): Crime and Place. Rapidly Evolving Research Methods in the 21st Century. In *Cityscape* 13(3), S. 199-203.
- Brantingham, P./Brantingham, P. (2013): The Theory of Target Search. In Cullin, F./Wilcox, P. (Hg.): *The Oxford Handbook of Criminological Theory*, S. 535-553.
- Brantingham, P./Brantingham, P. (1995): Criminality of Place. Crime Generators and Crime Attractors. In *European Journal on Criminal Policy and Research* 3(3), S. 1-26.
- Bursik, R./Grasmick, H. (1995): *Neighborhoods and Crime. The Dimensions of Effective Community Control*. New York.
- Buttler, G. (2005): Commercial burglary: what offenders say. In Gill, M. (Hg.): *Crime at work: studies in security and crime prevention*, S. 29-41. Basingstoke.
- Chainey, S./Tompson, L./Uhlir, S. (2008): The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. In *Security Journal*, 21(1), S. 4-28.
- Cohen, L./Felson, M. (1979): Social Change and Crime Rate Trends: A Routine Activity Approach. In *American Sociological Review* 44, S. 588-608.
- Coleman, J. (1986): Social Theory, Social research, and a Theory of Action. In *American Journal of Sociology*, 91(6), S. 1309-1335.
- Coleman, J. (1990): *Foundations of Social Theory*. Cambridge.
- Cornish, D./Clarke, R. (1986): *The Reasoning Criminal. Rational Choice Perspectives on Offending*. New York.
- Diekmann, A./Voss, T. (2004): Die Theorie rationalen Handelns. Stand und Perspektiven. In A. Diekmann/T. Voss

- (Hg.): Rational-Choice-Theorie in den Sozialwissenschaften, S. 13-29.
- Drawve, G. (2014): A Metric Comparison of Predictive Hot Spot Techniques and RTM. In *Justice Quarterly*, H. 3, Jg. 33, S. 369-397.
- Eifler, S. (2009): *Kriminalität im Alltag. Eine handlungstheoretische Analyse von Gelegenheiten*. Wiesbaden.
- Eifler, S. (2002): *Kriminalsoziologie*. Bielefeld.
- Esser, H. (2007): Der Handlungsbegriff in der modernen Soziologie. In *Altmeyden, K./Hanitzsch, T./Schlüter, C. (Hg.): Journalismustheorie: Next Generation. Soziologische Grundlegung und theoretische Innovation*, S. 27-46.
- Esser, H. (2001): *Soziologie. Spezielle Grundlagen. Band 6: Sinn und Kultur*. Frankfurt.
- Esser, H. (1997): Die „Definition“ der Situation und die Rationalität des Handelns. In *Meleghy, H./Niedenzu, J./Preglau, M./Traxler, F./Schmeikal, B. (Hg.): Soziologie im Konzert der Wissenschaften. Zur Identität einer Disziplin*, S. 69-90.
- Esser, H. (1999): *Soziologie. Spezielle Grundlagen. Band 1: Situationslogik und Handeln*. Frankfurt.
- Ester, M./Sander J. (2000): *Knowledge Discovery in Databases*. Berlin.
- Ester, M./Kriegel, H./Sander, J./Xu, X. (1996): A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, AAAI Press, 1996, S. 226-231.
- Farrell, G./Pease, K. (1995): Like Taking Candy. Why does Repeat Victimization occur? In *British Journal of Criminology* 35, S. 384-399.
- Farrell, G./Pease, K. (1993): *Once Bitten, Twice Bitten: Repeat Victimization and its Implications for Crime Prevention*. Crime Prevention Unit Series Paper46. London.
- Focus (2015): *Vorhersage per Smartphone, Polizei-App sagt Ihnen wann der Dieb ins Haus kommt. Trefferquote bei über 80 Prozent*. URL: http://www.focus.de/digital/handy/warte-auf-den-einbrecher-vorhersage-per-smartphone-polizei-app-sagt-euch-wann-der-dieb-ins-haus-kommt_id_4725470.html, zuletzt aufgerufen Januar 2018.
- Forrester, D./Chatterton, M./Pease, K. (1988): *The Kirkholt Burglary Prevention Project*, Rochdale. Crime Prevention Unit Paper 13. London.
- Fotheringham, A./Brunsdon, C./Charlton, M. (2002): *Geographically Weighted Regression. The Analysis of Spatially Varying Relationships*. West Sussex.
- Fox, B./Farrington, D. (2012): *Creating Burglary Profiles Using Latent Class Analysis: A New Approach to Offender Profiling*. In *Criminal Justice and Behavior* 39(12), S. 1582-1611.
- Frers, L./Krasmann, S./Wehrheim, J. (2013): *Geopolicing und Kriminalkartierungen. Wie Polizeien sich ein Bild machen*. In *Kriminologisches Journal*, 45. Jg., H. 3, S. 166-179.
- Friedman, J./Hastie, T./Tibshirani, R. (2001): *The elements of statistical learning*. Bd. 1. New York.
- Fritsch, D./Glemser, M./Klein, U./Sester, M./Strunz, G. (1998): *Zur Integration von Unsicherheit bei Vektor- und Rasterdaten*. In *Geo-Informationssysteme - Zeitschrift für raumbezogene Information und Entscheidungen*, Vol. 11, Heft 4, S. 26-35.
- Garofalo, J. (1987): *Reassessing the Lifestyle Model of Criminal Victimization*. In *Gottfredson, M./Hirschi, T. (Hg.): Positive Criminology*, S. 23-42.
- Gerstner, D. (2017): *Predictive Policing als Instrument zur Prävention von Wohnungseinbruchdiebstahl. Evaluationsergebnisse zum Baden-Württembergischen Pilotprojekt P4*. Freiburg im Breisgau.
- Golem (2014): *Polizei nutzt Predictive-Policing Software gegen Einbrecher* URL: <https://www.golem.de/news/bayern-polizei-nutzt-predictive-policing-software-gegen-einbrecher-1408-108388.html>, zuletzt aufgerufen im Januar 2018.
- Gluba, A./Heitmann, S./Hermes, N. (2015): *Reviktimisierungen bei Wohnungseinbrüchen*. In *Kriminalistik*, 6/2015, S. 368-375.
- Groff, E./La Vigne, N. (2002): *Forecasting the Future of Predictive Crime Mapping*. In *Crime Prevention Studies* 13, S. 29-57.
- Hart, T./Zandbergen, P. (2012): *Effects of Data Quality on Predictive Hotspot Mapping - Final Technical Report*. Washington. URL: <https://www.ncjrs.gov/pdffiles1/nij/grants/239861.pdf>, zuletzt aufgerufen im Januar 2018.
- Hindelang, M./Gottfredson, M./Garofalo, J. (1978): *Victims of Personal Crime*. Cambridge.
- Hunt, P./Saunders, J./Hollywood, J. (2014): *Evaluation of the Shreveport Predictive Policing Experiment*. Santa Monica.

- Johnson, S./Bernasco, W./Bowers, K./Elffers, H./Ratcliffe, J./Rengert, G./Townsend, M. (2007): Space-Time Patterns of Risk: A Cross National Assessment of Residential Burglary Victimization. In *Journal of Quantitative Criminology* 23, S. 201-219.
- Kass, G. (1980): An Exploratory Technique for Investigating Large Quantities of Categorical Data. In *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, Vol. 29, No. 2, S. 119-127.
- Keim, D./Kohlhammer, J./Ellis, G./Mansmann, F. (2010): Mastering the Information Age – Solving Problems with Visual Analytics. Goslar.
- Kersting, S./Willing, S. (2014): Was sagen uns Tatmerkmale über den Wohnungseinbruchdiebstahl? Vorstellung eines kriminalistisch-kriminologischen Forschungsprojekts. In *Kriminalistik*, 12/2014, S. 720-723.
- Kinkeldey, C./MacEachren, A./Schiewe, J. (2014): How to Assess Visual Communication of Uncertainty? A Systematic Review of Geospatial Uncertainty Visualisation User Studies. In *The Cartographic Journal*, H. 4, Jg. 51, S. 372-386.
- Kornhauser, R. (1978): *Social Sources of Delinquency*. Chicago.
- Kroneberg, C. (2005): Die Definition der Situation und die variable Rationalität der Akteure. In *Zeitschrift für Soziologie* 34(5), S. 344-363.
- Kroneberg, C. (2007): Wertrationalität und das Modell der Frame-Selektion. In *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59(2), S. 215-239.
- Kubrin, C./Stucky, T./Krohn, M. (2009): *Researching Theories of Crime and Deviance*. Oxford.
- Kunadt, S. (2011): *Sozialer Raum und Jugendkriminalität. Zum Einfluss der Wohnumgebung auf delinquentes Handeln*. Münster.
- Landeskriminalamt NRW (2017): *Forschungsbericht Wohnungseinbruchdiebstahl. Basisbericht*. Düsseldorf.
- Landeskriminalamt NRW (2012): *Wohnungseinbruchdiebstahl. Eine hypothesenprüfende Strukturanalyse*. Düsseldorf.
- Levine, N. (2008): The "Hottest" Part of a Hotspot: Comments on "The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime". In *Security Journal*, H. 4, Jg. 21, S. 295-302.
- Lüdemann, C./Ohlemacher, T. (2002): *Soziologie der Kriminalität. Theoretische und empirische Perspektiven*. Weinheim.
- Maimon, O./Rokach, L. (2015): *Data Mining with Decision Trees. Theory and Applications*. Singapore. New York.
- Miethe, T./Meier, R. (1979): *Crime and its Social Context. Toward an Integrated Theory of Offenders, Victims, and Situations*. Albany.
- Mohler, G./Short, M./Malinowski, S./Johnson, M./Tita, G./Bertozzi, A./Brantingham, J. (2015): Randomized Controlled Field Trials of Predictive Policing. In *Journal of the American Statistical Association*, H. 512, Jg. 110, S. 1399-1411.
- Morgan, M./Henrion, M./Small, M. (1990): *Uncertainty*. Cambridge.
- Motorola Solutions (2015): *Predictive Analytics vs. Hotspotting. A Study of Crime Prevention, Accuracy and Efficiency*. URL: <https://www.motorolasolutions.com/content/dam/msi/docs/products/smart-public-safety-solutions/ilps/Predictive-Analytics-vs-Hotspotting-White-Paper.pdf>, zuletzt aufgerufen im Januar 2018.
- Newman, O. (1973): *Defensible Space. Crime Prevention Through Urban Design*. New York.
- Newman, O. (1996): *Creating Defensible Space*. U.S. Department of Housing and Urban Development, Office of Policy Development and Research.
- Newman, O./Franck, K. (1980): *Factors Influencing Crime and Instability in Urban Housing Developments*. Washington.
- Nexiga (2017): *Geodaten auf höchster Ebene*, URL: <http://www.nexiga.com/geodaten-auf-hoehster-ebene>, zuletzt aufgerufen im Januar 2018.
- Openshaw, S. (1984): The modifiable areal unit problem. In *Concepts and Techniques in Modern Geography*. Norwich.
- Park, R./Burgess, E./McKenzie, R. (1925): *The City*. Chicago.
- Pease, K. (1998): Repeat Victimization. Taking Stock. In *Police Research Group: Crime Detection and Prevention Series Paper 90*.
- Perry, W./McInnis, B./Price C./ Smith, S./Hollywood, J. (2013): *Predictive Policing. The Role of Crime Forecasting in Law Enforcement Operations*. Santa Monica.

- Pollich, D. (2010): Problembelastung und Gewalt. Eine soziologische Analyse des Handelns jugendlicher Intensivtäter. Münster.
- Pollich, D./Bode, F. (2017): Predictive Policing: Zur Notwendigkeit eines (sozial)wissenschaftlich basierten Vorgehens. In *Polizei & Wissenschaft*, 03/2017. S. 2-12.
- Polvi, N./Looman, T./Humphries, C./Pease, K. (1991): The Time Course of Repeat Burglary Victimization. In *British Journal of Criminology* 31, S. 411-414.
- Prenzler, T. (2009): Preventing Burglary in Commercial and Institutional Settings: A Place Management and Partnerships Approach. CRISP Report. Alexandria.
- Public Engines (2014): Predictive Analytics vs. Hot Spotting. A Study of Crime Prevention, Accuracy and Efficiency. URL: <https://www.motorolasolutions.com/content/dam/msi/docs/products/smart-public-safety-solutions/ilps/Predictive-Analytics-vs-Hotspotting-White-Paper.pdf>, zuletzt aufgerufen im Januar 2018.
- Quinlan, J. (1986): Induction of decision trees. In *Machine learning* 1 (1), S. 81-106.
- Sammut, C./Webb, G. (2011): Encyclopedia of machine learning. New York.
- Schweer, T. (2015): „Vor dem Täter am Tatort“ – Musterbasierte Tatortvorhersagen am Beispiel des Wohnungsgebrauchs. In *Die Kriminalpolizei*, 1/2015, S. 13-16.
- Schwind, H. (2011): Kriminologie. Eine praxisorientierte Einführung mit Beispielen. Heidelberg.
- Seidensticker, K. (2017): Kriminalität in Mikrosegmenten. Ergebnisse einer Auswertung für Essen und Mülheim an der Ruhr. In *forum kriminalprävention*, 04/2017, S. 27-31.
- Shaw, C./McKay, H. (1969[1942]): Juvenile Delinquency in Urban Areas. Chicago.
- Simon, H. (1957): Models of Man. Social and Rational. New York.
- Silverman, B. (1986): Density Estimation for Statistics and Data Analysis. London.
- Song, Y./Lu, Y. (2015): „Decision tree methods: applications for classification and prediction“. In *Shanghai Archives of Psychiatry* 27 (2): S. 130-135.
- Steenbeck, W./Weisburd, D. (2016): Where the action is in crime? An examination of variability of crime across different spatial units in The Hague. *Journal of Quantitative Criminology*, 32(3), S. 449-469.
- Stoffel, F./Bode, F./Keim, D. (2017): Qualitätsmetriken im Bereich Predictive Policing. Die Variabilität und Validität von Trefferraten. In *Polizei & Wissenschaft*, 04/2017. S. 2-15.
- Streeck, W. (2009): Man weiß es nicht genau: Vom Nutzen der Sozialwissenschaften für die Politik. In *MPIfG Working Paper* 09/11.
- Townsend, M./Homel, R./Chaseling, J. (2003): Infectious Burglaries. A Test of the Near Repeat Hypothesis. In *British Journal of Criminology* 43, S. 615-633.
- Van Patten, I./McKeldin-Coner, J./Cox, Deana (2009): A Microspatial Analysis of Robbery: Prospective Hot Spotting in a Small City. In *Crime Mapping: A Journal of Research and Practice*, 1(1), S. 7-32.
- Wang, D./Ding, W./Lo, H./Stepinski, T./Salazar, J./Morabito, M. (2012): Crime hotspot mapping using the crime related factors – a spatial data mining approach. In *Applied Intelligence*, H. 4, Jg. 39, S. 772-781.
- Weisburd, D./Amram, S. (2014): The law of concentrations of crime at place: the case of Tel Aviv-Jaffa. In *Police Practice and Research* (15), S. 101-114.
- Weisburd, D./Groff, E./Yang, S. (2012): The Criminology of Place: Street Segments and Our Understanding of the Crime Problem. Oxford.
- Wilson, J./Kelling, G. (1982): The Police and Neighborhood Safety. Broken Windows. In *The Atlantic Monthly* 249(3), S. 29-38.
- Zhang, P./Qi, M. (2005): Neural network forecasting for seasonal and trend time series. In *European Journal of Operational Research*, H. 2, Jg. 160, S. 501-514.
- Zimbardo, P. (1973): A Field Experiment in Auto Shaping. In *Ward, C. (Hg.): Vandalism*, S. 85-90.

Herausgeber

Landeskriminalamt Nordrhein-Westfalen
Abteilung 3
Völklinger Straße 49
40221 Düsseldorf

Redaktion:

Dr. Felix Bode

Dr. Hanna Post

Lukas Schulte, B.A.

Kai Seidensticker, M.A.

SKALA@polizei.nrw.de

poststelle.lka@polizei.nrw.de

www.lka.polizei.nrw

Stand: 21.08.2018

