



LANDESKRIMINALAMT
NIEDERSACHSEN

März 2021

PreMAP – Prüfung einer Erweiterung und Modifizierung des Prognoseansatzes

Beitrag des LKA Niedersachsen /
Fachstab „Kriminologische Forschung und Statistik“
zum Abschlussbericht der erweiterten Pilotierung



Niedersachsen

Inhalt

1. Auftrag des Teilprojekts Wissenschaft	3
2. Vorbereitung.....	4
2.1 Identifikation für Prognosen geeigneter Deliktsbereiche.....	4
2.2 Operationalisierung der potentiellen Erweiterungsdelikte.....	5
2.3 Erwerb kleinräumiger Daten.....	6
2.4 (Neu-)Kategorisierung von NIVADIS-Merkmalen.....	7
3. Vorgehen	9
3.1 Prognoseebenen und Prognoseeinheiten.....	9
3.2 Datengrundlage.....	10
3.3 Untersuchte Prognoseansätze	10
3.3.1 Tatbezogene Prognoseansätze: „Near Repeat: Ja / Nein“ für ein einzelnes Delikt.....	11
3.3.2 Raumbezogene Prognoseansätze: Delikthäufigkeits-Klasse eines Wohnquartiers in der Folgewoche.....	13
4. Ergebnisse.....	16
4.1 Tatbezogene Prognoseansätze: Evaluation der Prognosequalität.....	17
4.1.1 Erläuterungen zu den Evaluationsblättern in Anhang B.....	17
4.1.2 Delikttyp 1 „WED / TWE“	20
4.1.3 Delikttyp 2 „Schwerer Diebstahl aus Nichtwohnungen (ohne Ladendiebstahl)“	21
4.1.4 Delikttyp 3 „Diebstahl in / aus KFZ“	22
4.1.5 Delikttyp 4 „Diebstahl von KFZ“	23
4.1.6 Fazit zu den tatbezogenen Prognosen.....	24
4.2 Raumbezogene Prognoseansätze: Evaluation der Prognosequalität	25
4.2.1 Erläuterungen zu den Tabellen und Abbildungen in diesem Abschnitt.....	25
4.2.2 Delikttyp 1 „WED / TWE“ (Deliktschlüssel 435 und 436).....	26
4.2.3 Delikttyp 2 „Schwerer Diebstahl aus Nichtwohnungen (ohne Ladendiebstahl)“ (Deliktschlüssel 410, 415, 425).....	35
4.2.4 Delikttyp 3 „Diebstahl in / aus KFZ“ (Deliktschlüssel 350 und 450).....	41
4.2.5 Delikttyp 4 „Diebstahl von KFZ“ (Deliktschlüssel 300, 345, 400, 410 und 445)	48
4.2.6 Fazit zu den raumbezogenen Prognosen.....	54
5. Zusammenfassung und Handlungsempfehlungen.....	55
Anhang A: In die Prognoseansätze einbezogene Variablen.....	57
A.1 Tatbezogene Prognoseansätze	57
A.2 Raumbezogene Prognoseansätze.....	60
Anhang B: Evaluationsblätter zu den tatbezogenen Prognoseansätzen.....	63

1. Auftrag des Teilprojekts Wissenschaft

Das Teilprojekt Wissenschaft hatte im Rahmen der erweiterten Pilotierung von PreMAP gemäß Erlass MI 23.35-12337/14-9 den Auftrag, den Predictive Policing-Ansatz im Bereich des Wohnungseinbruchdiebstahls inhaltlich fortzuentwickeln und zu erproben. Zur Optimierung der Ausweisung von Risikogebieten sollten weitere geeignete Parameter für Prognosemuster geprüft und bezogen auf die Bekämpfung des WED sukzessive ab dem 1. November 2018 verfügbar gemacht werden.

Mit der Spezifizierung des Ziels des Teilprojekts Wissenschaft in Abstimmung mit der Projektleitung wurde deutlich gemacht, dass eine Einführung neuer Datengrundlagen oder Prognoseansätze nur dann stattfinden sollte, wenn diese sich im Rahmen einer Prüfung bewähren würden. Im Zwischenbericht der erweiterten Pilotierung vom 1. Juli 2019 hieß es entsprechend auf Seite 5, dass Prognoseansätze für den WED und ggf. weitere Phänomenbereiche aus dem In- und Ausland recherchiert, mit eigenen Ideen für Kriminalitätsprognosen abgeglichen und geprüft werden sollten. Als tauglich eingeschätzte Ansätze sollten unter Verwendung geeigneter Datenquellen weiterentwickelt werden und bei festgestellter Eignung in PreMAP integriert werden.

2. Vorbereitung

2.1 Identifikation für Prognosen geeigneter Deliktsbereiche

Da gemäß Auftrag die Bandbreite der Prognosen – bislang war PreMAP nur auf Wohnungseinbrüche ausgerichtet – ausgeweitet werden sollte, wurde in einem ersten Schritt eruiert, welche weiteren Deliktsbereiche sich für Prognosen anbieten. Geeignete Delikte mussten dabei einige Anforderungen erfüllen. So mussten sie in einem Umfange vorkommen, dass datengestützte Prognosen möglich waren. Das Dunkelfeld durfte nicht allzu groß sein, damit davon ausgegangen werden konnte, dass die registrierten Fälle auch das Gros der Grundgesamtheit abbilden. Ferner sollten die Delikte, die in die Prüfung einbezogen werden sollten, im öffentlichen Raum wahrnehmbar sein, um eine Wirksamkeit polizeilicher Reaktionen überhaupt potentiell zu ermöglichen. Die Auswirkungen der Delikte sollten im Vergleich zu anderen hoch sein; hierbei konnten psychische, physische oder materielle Auswirkungen auf der Individualebene oder auch solche auf der gesellschaftlichen Ebene umfasst sein. Schließlich sollte es keine anderen geeigneten Möglichkeiten der Prognose zukünftiger Risiken in Bezug auf die benannten Delikte geben.

Im Dezember 2018 wurde eine Abfrage bei den Pilotdienststellen der erweiterten Pilotierung initiiert, in der diese um Benennung geeigneter Deliktsbereiche unter den oben genannten Rahmenbedingungen gebeten wurden. 10 von 15 Inspektionen gaben entsprechende Rückmeldungen:

- Für viele Dienststellen waren vor allem KFZ-bezogene Delikte für Prognosen geeignet. Für Harburg, Lüneburg-Lüchow-Uelzen, Wolfsburg-Helmstedt, Osnabrück und Hildesheim hatten **Komplettentwendung hochwertiger Fahrzeuge, Kfz-Beschädigungen und Diebstähle aus KFZ** eine große Relevanz.
- Weiterhin als geeignet für die Erweiterung wurden der **Schwere Diebstahl in / aus gewerblich genutzten Räumen (Produktion, Gastronomie, Handel)** angesehen. Hier befürchteten insbesondere die Inspektionen Wilhelmshaven / Friesland, Delmenhorst / Oldenburg-Land / Wesermarsch und Hannover-Ost, dass es in diesem Deliktsbereich weiterhin verstärkt zu Serientaten durch Tätergruppierungen kommen könnte.
- Zudem wurden **Branddelikte** durch die Inspektionen Delmenhorst / Oldenburg-Land / Wesermarsch und Wolfsburg-Helmstedt aufgrund lokaler Tatserien für eine intensive prospektive Analyse vorgeschlagen.
- Das zusätzliche **Phänomen falscher Polizeibeamten** gewann den Meldungen zufolge ebenso an Relevanz. Beispielsweise kam es in Delmenhorst zu mehrfachen Vorfällen dieser Art. Dieses Phänomen, so berichteten Dienststellen, trat in einem räumlich begrenzten Bereich auf, sodass neben der relevanten Vorhersehbarkeit die Umsetzung in PreMap gewünscht wurde.
- Außerdem wurde auf den Bereich **Körperverletzungsdelikte** durch die PI Wolfsburg-Helmstedt aufmerksam gemacht.

Die genannten Delikte wurden in der Folge einer Prüfung unterzogen, bei für die **Branddelikte** eine Prüfung im Rahmen der deliktischen Erweiterung von PreMAP aufgrund des zu geringen Zahlengerüsts ausgeschlossen wurde. Die Häufigkeitszahl bewegte sich 2009–2018 bei ca. 16 und ist damit eindeutig zu gering. Auch der **Schwere Diebstahl in / aus Geschäften, Gaststätten, Kantinen, Hotels und Pensionen** wies in den vergangenen Jahren nur eine durchschnittliche Häufigkeitszahl von 30 auf; er wurde zwar nicht ausgeschlossen, sollte aber gegebenenfalls im Rahmen der datengestützten Prognoserechnungen mit anderen schweren Diebstählen gemeinsam behandelt werden. Das **Phänomen falsche Polizeibeamte** ist in NIVADIS schwerer fass- und selektierbar, sodass es auch ausgeschlossen wurde; wahrscheinlich wäre das

quantitative Niveau aber ohnehin zu gering gewesen. **Körperverletzungen** weisen zwar ausreichende Fallzahlen aus, es gibt aber für das Vorkommen dieser Taten im Hellfeld durchaus andere und einfache geeignete Erklärungsmöglichkeiten, wie örtliche Nähe zum Nachtleben und

Auf Sinnhaftigkeit / Optimierbarkeit geprüfte Delikttypen

- **Wohnungseinbruch (inklusive Tageswohnungseinbruch) – Delikttyp 1**
- **Schwerer Diebstahl aus Nichtwohnungen (ohne Ladendiebstahl) – Delikttyp 2**

(gemeinsame Betrachtung der folgenden Deliktschlüssel:)

- Schwerer Diebstahl in / aus Dienst-, Büro-, Fabrikations-, Werkstatt- und Lagerräumen
 - Schwerer Diebstahl in / aus Gaststätten, Kantinen, Hotels und Pensionen
 - Schwerer Diebstahl in / aus Kiosken, Warenhäusern, Verkaufsräumen, Selbstbedienungsläden, Schaufenstern und Schaukästen (ohne Ladendiebstahl)
- **Diebstahl in / aus KFZ – Delikttyp 3**
 - **Diebstahl von KFZ – Delikttyp 4**

zu Großveranstaltungen; ein Tool wie PreMAP wäre zum Erkennen solcher Zusammenhänge nicht notwendig gewesen. **Sachbeschädigungen an KFZ** schließlich wurden ausgeschlossen, da diese weitgehend im Dunkelfeld verbleiben.

2.2 Operationalisierung der potentiellen Erweiterungsdelikte

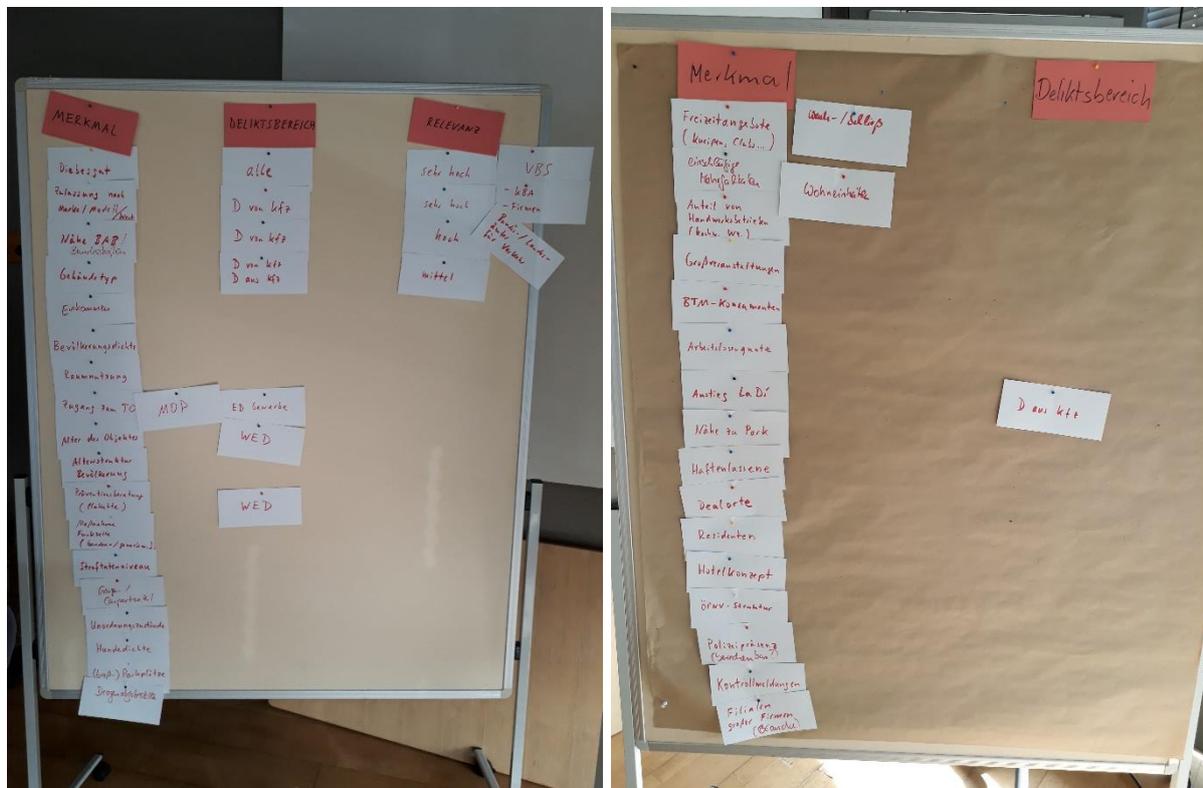
Insgesamt elf Vertreterinnen und Vertreter der PreMAP-Pilotdienststellen wurden daraufhin vom Teilprojekt Wissenschaft für den 7. März 2019 zu einem Workshop ins LKA geladen, in dessen Rahmen gemeinsam überlegt werden sollte, wie die potentiellen Erweiterungsdelikte für PreMAP operationalisiert werden könnten. Folgende Fragen wurden dabei adressiert:

- Welche Merkmale könnten einen Beitrag dazu leisten, Risiken für Wiederholungstaten einzuschätzen?
- Welche Merkmale sprechen für eine beginnende Tatserie?
- Wie können diese Merkmale erhoben werden? Liegen sie bereits polizeilich vor?
- Wie können die Merkmale in eine Risikoprognose integriert werden?
- Welche Indikatoren können herangezogen werden, um die Qualität einer möglichen Prognose einschätzen zu können?
- Welche Maßnahmen können ergriffen werden, um Taten im konkreten Deliktsfeld zu verhindern?

Um alle Teilnehmenden auf denselben Stand zu bringen, wurde eine 20minütige Einführung zum aktuellen Stand des Projekts PreMAP gegeben und das Vorgehen für den Workshop auseinandergesetzt. Danach wurde in der ersten Workshop-Phase in einer moderierten Gruppendiskussion erörtert, welche Einflussfaktoren für die verschiedenen Delikte maßgeblich seien. Hierbei ging es primär um ein Brainstorming; Gedanken wurden also ohne tiefgehende Diskussion aufgenommen und thematisch auf Metaplanwänden festgehalten (siehe untenstehende Bilder). Es wurden Einflussfaktoren identifiziert, die deliktsübergreifend Relevanz entfalten, und solche, die sehr spezifisch nur für einzelne Delikte Bedeutung haben. Sofern die Teilnehmenden Hinweise hatten, über welche Quellen empirischer Zugang zu den entsprechenden Einflussfaktoren gefunden werden könnte, wurden diese ebenfalls aufgenommen. Eine Herausforderung bestand darin, die polizeilich sozialisierten Teilnehmenden immer wieder zu bestärken, sich bei der Suche nach geeigneten Faktoren von prägenden selbst erlebten Fällen und Ereignissen zu lösen. Dies

gelang schlussendlich recht gut, sodass im Ergebnis eine Vielzahl von einflussnehmenden Faktoren auf die potentiellen Erweiterungsdelikte generiert werden konnte.

ABBILDUNG 1: DOKUMENTATION DER GEDANKEN IM RAHMEN DES BRAINSTORMINGS



Da die erste Phase des Workshops sich zeitlich deutlich umfangreicher gestaltete als geplant, unterblieb die Adressierung der Frage nach geeigneten polizeilichen Maßnahmen zur Reduzierung erkannter Kriminalitätsrisiken in Form der in Rede stehenden Delikte in einer zweiten Phase. Dies war aber insoweit unschädlich, als für die weitere Arbeit im Teilprojekt Wissenschaft die Ergebnisse der ersten Phase maßgeblich waren.

Ergänzend zum benannten Workshop wurde eine umfangreiche Literaturrecherche durchgeführt, um aufgrund kriminologischer und kriminalistischer Erkenntnisse weitere Einflussfaktoren zu gewinnen.

Bis zum 15. März 2019 wurden alle Ergebnisse zusammengetragen.

2.3 Erwerb kleinräumiger Daten

Das Gros der einflussnehmenden Variablen bestand in Merkmalen aus dem Vorgangsbearbeitungssystem NIVADIS; hierunter fielen vor allem die tatbezogenen Merkmale wie Tatzeit, Delikt, Tatmittel, Modus Operandi und Stehlgut. Darüber hinaus stand in der Liste der Einflussvariablen aber auch eine Vielzahl soziodemographischer und sozioökonomischer Daten. Beispielsweise wurde die durchschnittliche Kaufkraft als Faktor erkannt, der Anreize für Diebstähle schafft. Die Arbeitslosenquote und auch der Anteil kleiner Kinder im Haushalt indiziert dagegen, dass Menschen in der Regel häufiger daheim sind und damit eine größere soziale Kontrolle herrscht. Derartige Daten liegen in den polizeilichen Systemen nicht vor. Daher wurde geprüft, wo solche Informationen bezogen werden konnten.

Das Statistische Landesamt konnte zwar viele der Einflussfaktoren grundsätzlich abbilden, leider war dies dort nicht in der notwendigen Kleinräumigkeit möglich; die Angaben endeten im besten Falle auf Ebene einer Kommune, was für kleinräumige Prognosen ungeeignet ist.

Bestimmte Firmen dagegen bieten Daten, wie sie benötigt wurden, kleinräumig auf Ebene von Nachbarschaften, Straßenabschnitten oder gar Hausebene gegen Bezahlung an. Nachdem eine Recherche der einschlägigen Firmen vorgenommen wurde, wurden mit dem Dezernat 41.1 der Zentralen Polizeidirektion (ZPD) die Rahmenbedingungen einer möglichen Ausschreibung besprochen. Im Anschluss wurden drei professionelle Anbieter (Gesellschaft für Konsumforschung, infas360 und Nexiga) mit der Bitte um Erstellung einer Preisinformation angeschrieben. Zu diesem Zweck wurde den Firmen eine Liste der benötigten Daten übermittelt, damit diese angeben konnten, welche Daten konkret geliefert werden konnten und welche nicht. Da offen war, ob die Hinzuziehung von externen Daten einen bedeutenden Mehrwert versprach, sollte sich die Kalkulation der Preisinformation aus Kostengründen zunächst nur auf die beiden Städte Hannover und Braunschweig beziehen. Die aufgerufenen Preise lagen zwischen 7.500 und 20.000 Euro plus Mehrwertsteuer. Alle Bieter konnten die Vorgaben gewünschter Daten nicht komplett erfüllen, aber sehr wesentlich. Bei den Möglichkeiten gab es nur kleine Unterschiede. Aufgrund des Preisniveaus wurde durch die ZPD entschieden, dass eine Ausschreibung nicht notwendig war, eine Vergabe an den Bieter mit dem günstigsten Angebot wurde dadurch freihändig möglich. Den Zuschlag bekam die Firma Nexiga, deren Daten auch die Kolleginnen und Kollegen der Polizei Nordrhein-Westfalen für ihr Predictive Policing-Projekt SKALA nutzen. Insofern konnte angenommen werden, dass der günstigste Anbieter zudem einen Zugang zu polizeilichen Bedürfnissen und Anforderungen hat.

Dank der sehr guten Vorbereitung und Durchführung seitens der ZPD war es möglich, die kleinräumigen Daten zum 1. August 2019 für ein Jahr zur Nutzung in der KFS des Landeskriminalamtes zur Verfügung zu haben.

2.4 (Neu-)Kategorisierung von NIVADIS-Merkmalen

Für die ersten Pilotierungsphasen von PreMAP wurden die enthaltenen Merkmale zu Stehlgut, Tatmittel und Modus Operandi (unterteilt in die angegangene Stelle und das konkrete Vorgehen) zu Kategorien zusammengefasst, um die Komplexität zu reduzieren und eine Verdichtung zu erreichen. So wurden bezogen auf die NIVADIS-Katalogebene 4 aus insgesamt 2.065 Merkmalen 178 Kategorien gebildet. Diese Kategorien wurden seinerzeit aber vor dem Hintergrund gebildet, dass sie das Delikt „Wohnungseinbruchdiebstahl“ abbilden. Entsprechend fanden sich viele NIVADIS-Merkmale in Auffangkategorien wie „Sonstiges“, wenn sie keinerlei offensichtlichen Bezug zum Wohnungseinbruch aufwiesen. Die Bezüge können sich aber von Delikt zu Delikt unterscheiden, sodass durch das Teilprojekt Wissenschaft eine neue Kategorisierung vorgenommen wurde, die den Anspruch hatte, für ein breiteres Spektrum an Delikten anwendbar zu sein, zumindest aber für die in Rede stehenden. Folgende Verdichtungen fanden im Ergebnis statt:

- Beim Stehlgut wurden, von der NIVADIS-Katalogebene 2 ausgehend, 52 Merkmale beibehalten, ein Merkmal wurde anhand der Katalogebene 3 in vier Unterkategorien aufgeteilt, und die verbleibenden 19 Merkmale zu sieben Merkmalskategorien verdichtet. (Implizit wurden so 1.132 Merkmale der NIVADIS-Katalogebene 4 zu 63 Merkmalskategorien verdichtet.)
- Bei den Tatmitteln wurden, von der NIVADIS-Katalogebene 1 ausgehend, 10 Merkmale beibehalten, zwei zu einer Merkmalskategorie zusammengefasst, und vier anhand der Katalogebene 2 in insgesamt 21 Unterkategorien aufgeteilt. (Implizit wurden so 547 Merkmale der NIVADIS-Katalogebene 4 zu 32 Merkmalskategorien verdichtet.)

- Bei den Modi Operandi wurden 386 Merkmalen der NIVADIS-Katalogebene 4 zum einen zu 32 Merkmalskategorien für die angegangene Stelle verdichtet, zum anderen zu 51 Merkmalskategorien für das konkrete Vorgehen.

Die Kategorisierungsarbeiten waren am 15. Dezember 2019 abgeschlossen.

3. Vorgehen

3.1 Prognoseebenen und Prognoseeinheiten

Bislang verfolgen die datengestützten Prognosen bei PreMAP den Ansatz, ausgehend von einer Ausgangstat und deren Tatmerkmalen das Risiko für Folgetaten im Sinne des Near Repeat-Ansatzes zu bestimmen. Es handelt sich also um einen tatbezogenen Ansatz.

Im Teilprojekt Wissenschaft wurde auf der einen Seite versucht, diesen Ansatz zu übertreffen, und zwar durch alternative Prognoseansätze und / oder durch die probeweise Einbeziehung zusätzlicher kleinräumiger sozioökonomischer Daten. Dazu wurden tatbezogene Prognoseansätze mit dem bekannten Ziel untersucht, für eine einzelne Tat als Prognoseeinheit das Merkmal „Wird ein Near Repeat haben: Ja /Nein“ zu prognostizieren, idealerweise zusammen mit einer Wahrscheinlichkeitsangabe für „Ja“. Dabei weist die Prognoseeinheit selbst Merkmale (Stehlgüter, Charakteristika des Modus Operandi, Tatmittel) auf und liegt zusätzlich in einem Gebiet, das ebenfalls durch bestimmte Merkmale gekennzeichnet ist. Wenn letztere in den Prognoseansatz einbezogen werden, werden die sozioökonomischen Merkmale des Gebiets also als zusätzliche Merkmale der Tat interpretiert.

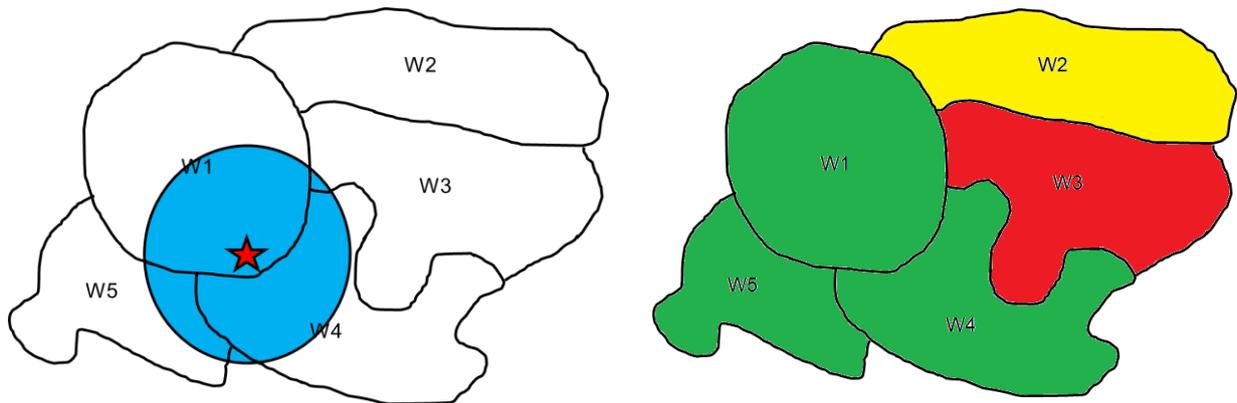
Auf der anderen Seite wurden im Teilprojekt Wissenschaft auch Ansätze geprüft, die als raumbezogene Prognoseansätze und damit als grundsätzliche Innovation gegenüber dem bisherigen PreMAP-Ansatz bezeichnet werden können. Hier sollte für ein bestimmtes Gebiet („Wohnquartier“ in der Abgrenzung des Datenlieferanten Nexiga) mit seinen Merkmalen die Delikthäufigkeit in einem vorgegebenen Zeitraum (Kalenderwoche) prognostiziert werden. Auf Seiten der Delikte ging nicht nur eine Tat, sondern die Gesamtheit aller Taten des jeweiligen Delikttyps in dem Gebiet in einer Woche in diese Berechnung ein. Die Prognoseeinheit war entsprechend das jeweilige Gebiet im zu prognostizierenden Zeitraum, also die Wohnquartier-Woche-Kombination (WWK). Hier werden die Merkmale der aufgetretenen Delikte im Gebiet in geeigneter Aggregation (z. B. Gesamtzahl oder Summe) als zusätzliche Merkmale des Gebietes interpretiert.

Prognosen grundsätzlich auf zwei Ebenen

- **Tatbezogene Prognosen:** Prognose von Near Repeats auf Basis von Tatmerkmalen einer Ausgangstat
- **Raumbezogene Prognosen:** Prognose von Tathäufigkeiten in einem Gebiet (0, 1, >1 Taten im Prognosezeitraum) auf Basis entsprechender räumlicher Merkmale (ggf. einschließlich der Merkmale der in dem Gebiet erfassten Taten)

Abbildung 2 zeigt die Unterschiede beispielhaft. Das linke Bild stellt mit dem Stern eine Ausgangstat dar, die im Gebiet W1 vorgefallen ist. Basierend auf den Merkmalen der Tat und des Gebiets W1 (interpretiert als weitere Merkmale der Tat) wird das Auftreten einer Folgetat als Ja/Nein-Variable prognostiziert, der Risikobereich bei einer Ja-Prognose ist ein Kreis um die Ausgangstat herum. Anders im rechten Bild: Die Merkmale des Gebiets W1, zu denen auch die in geeigneter Weise aggregierten Merkmale der dort in einem Bezugszeitraum (hier: Kalenderwoche) verzeichneten Taten gehören, gehen in die datengestützte Prognose der Delikthäufigkeit im gleichen Gebiet im Folgezeitraum ein.

ABBILDUNG 2: BEISPIELHAFTES DARSTELLUNG VON PROGNOSEEBENE (RISIKOBEREICHE) BEI TAT- (LINKS) UND RAUMBEZOGENEN (RECHTS) DATENGESTÜTZTEN PROGNOSEN



3.2 Datengrundlage

Die Datengrundlage für die angestellten Prognosen waren zunächst wie beim aktuellen PreMAP-Ansatz die NIVADIS-Merkmale Tatmittel, Stehlgut und Modus Operandi (angegangene Stelle, Vorgehen), allerdings mit den oben beschriebenen Neukategorisierungen. Zusätzlich gingen bei einigen untersuchten Ansätzen die eingekauften sozioökonomischen Daten der Firma Nexiga in die Berechnungen ein. Diese lagen je nach betrachteter Variable auf in der Regel auf Wohnquartierebene oder feingranularer vor. Die Daten aus dem Vorgangsbearbeitungssystem NIVADIS und von Nexiga mussten zunächst aufeinander bezogen werden. Bei den tatbezogenen Ansätzen müssen dem Einzeldelikt das korrekte Wohnquartier und nach Möglichkeit der korrekte Straßenabschnitt zugeordnet werden, bei den raumbezogenen Ansätzen dem einzelnen Wohnquartier die in einer bestimmten Woche dort begangenen Delikte. Bei letzterem wurden außerdem einige der vorliegenden Daten unterhalb der Wohnquartierebene durch Bildung des Medians für das Wohnquartier zusammengefasst.

Einige sozioökonomische Kennzahlen für die Wohnquartiere wurden durch eigene Berechnungen aus den Nexiga-Daten ermittelt: Einwohner je Flächeneinheit, Haushalte je Flächeneinheit, Haushalte in selbstgenutztem Wohneigentum je Flächeneinheit, Zuzüge je Einwohner, Fortzüge je Einwohner, Arbeitslose je Einwohner, PKW je Haushalt, PKW je Flächeneinheit, gewerbliche PKW je Haushalt, gewerbliche PKW je Flächeneinheit, Garagen je Haushalt, Garagen je Flächeneinheit, Garagen je PKW.

Die Nexiga-Daten beziehen sich auf den Erhebungstichtag 01.01.2018. Sie werden für den gesamten betrachteten Zeitraum von 2017 bis 2019 verwendet, womit implizit unterstellt wird, dass eventuelle Veränderungen in diesem Zeitraum klein genug waren, um die Prognosequalität der untersuchten Ansätze nicht zu beeinträchtigen.

In Anhang A ist aufgelistet, welche Variable in welchem datengestützten Prognoseansatz verwendet wurde.

3.3 Untersuchte Prognoseansätze

Die hier dokumentierten Untersuchungen zu unterschiedlichen Delikttypen und Prognoseansätzen wurden mit der Datenanalyse-Software KNIME in der Version 4.0 durchgeführt. Die Boxplots zu den Kennzahlen zur Prognosequalität wurden mit der Statistik-Software-Umgebung R in der Version 3.6.1 erstellt.

Bei den als Alternativen zum aktuellen PreMAP-Ansatz untersuchten Prognoseansätzen war es grundsätzlich empfehlenswert, die Input-Variablen zu standardisieren, d. h. durch Abziehen des

Mittelwertes und anschließendes Teilen durch die Standardabweichung dahingehend vergleichbar zu machen, dass jetzt alle Input-Variablen um den Wert 0 schwanken und gewissermaßen in Standardabweichungseinheiten gemessen werden. Auf diese Weise wird die unterschiedliche Streuung der Input-Variablen berücksichtigt und zudem die Maßeinheit herausgerechnet, also beispielsweise, ob eine Bevölkerungsangabe in der Einheit Personen oder der Einheit „Tausend Personen“ in den Daten enthalten ist.

Bei Prognoseansätzen ist es grundsätzlich ungünstig, wenn es starke lineare Zusammenhänge, im einfachsten Fall hohe paarweise Korrelationen, zwischen den Input-Variablen gibt, denn hierdurch werden die Ergebnisse – also die Prognosen – grundsätzlich unsicherer. Als einfache Gegenmaßnahme wurden alle paarweisen Korrelationen berechnet (und zwar jeweils neu für jeden Delikttyp und jede regionale Eingrenzung der Trainingsmengen,¹ aber nicht für jede einzelne Trainingsmenge), und in den Fällen, wo die Korrelation zweier Variablen über dem Schwellenwert von 0,8 lag, wurde eine der beiden Variablen aus der Analyse ausgeschlossen.² Das ist einer der Gründe für die Abweichungen zwischen den Spalten der Tabellen in Anhang A.

Bei Untersuchungen von datengestützten Prognoseansätzen ist es üblich, die vorhandene Datenmenge in zwei Teilmengen aufzuteilen. Mit dem größeren Teil als Trainingsmenge wird der Ansatz trainiert, im kleineren Teil als Testmenge³ wird mit geeigneten Kennzahlen geprüft, wie gut die Prognosen für die Output-Variable zu den dort bereits bekannten tatsächlichen Werten der Output-Variable passen.

Um sich nicht der Gefahr auszusetzen, aufgrund relativ „untypischer“ Trainingsdaten mit einem Ansatz nur ein einzelnes Modell mit schlechter Prognosequalität zu trainieren, werden zu jedem untersuchten Prognoseansatz jeweils 100 verschiedene 90-%-Trainingsmengen zufällig gezogen und die jeweils verbleibenden 10-%-Testmengen für die Kennzahlberechnung zur Prognosequalität verwendet. Man erhält dabei für einen Prognoseansatz 100 Modelle und damit zu jeder Kennzahl 100 Werte, die in geeigneter Weise gemeinsam betrachtet werden müssen. Für die hier untersuchten Prognoseansätze werden die „mittlere“ Höhe und die Streuung der 100 Werte – typischerweise bestimmte Anteilswerte, die unten näher erläutert werden – jeweils durch Boxplots visualisiert (vgl. die Erläuterungen in Unterabschnitt 4.1.1).

3.3.1 Tatbezogene Prognoseansätze: „Near Repeat: Ja / Nein“ für ein einzelnes Delikt

3.3.1.1 Aktueller PreMAP-Ansatz: *Near Repeat-Einflüsse (NRE) als merkmalspezifische Score-Beiträge*

Zur Prognose von zeitlich und räumlich nahen Folgetaten („Near Repeats“, NR) bei Wohnungseinbrüchen – nur diese waren bislang Gegenstand von PreMAP –, wird im aktuellen PreMAP-Ansatz mit einem Score als approximativem Risikomaß⁴ gearbeitet, der ausschließlich auf den in NIVADIS verzeichneten Charakteristika der Tat beruht, die allerdings noch zu speziellen Prognosemerkmalen verdichtet werden, die alle binäre, also 1-0- bzw. Ja-Nein-Variablen sein müssen. Ein Near Repeat wird prognostiziert, falls der Score-Wert einen festgelegten Schwellenwert übersteigt; im Umkreis der betreffenden Tat wird dann ein Risikogebiet ausgewiesen, um geeignete polizeiliche Maßnahmen zur Verhinderung der Near Repeat-Tat zu ermöglichen.

¹ In den Teildaten für Braunschweig oder Hannover allein treten eventuell andere Korrelationen auf als in den Gesamtdaten, daher können zum Beispiel bei den Braunschweig-Modellen für einen Ansatz andere Variablen herausfallen als bei den Gesamtmodellen für den gleichen Ansatz.

² Für die Prognosequalität spielt es keine Rolle, welche von beiden man ausschließt.

³ Man spricht auch von „Validierungsmenge“.

⁴ Die Score-Werte sind nicht auf das Intervall [0, 1] beschränkt, können also nicht als Risikoangabe im Sinne einer mathematischen Wahrscheinlichkeit betrachtet werden.

Zur Berechnung des Score-Wertes für eine bestimmte Tat werden zuvor aus Daten für einen Basiszeitraum merkmalspezifische Score-Beiträge („Near Repeat-Einflüsse“, NRE) ermittelt, und zwar wie folgt: Die Daten für alle Delikte werden nach der Eigenschaft „Hat Near Repeat: Ja / Nein“ in die Deliktgruppen ohne und mit Near Repeat (Folgetat nach maximal 72 Stunden in maximal 500 m Entfernung) aufgeteilt. In beiden Gruppen werden für jedes Merkmal die Anteile der Delikte mit der Ausprägung „Ja“ an allen Delikten ermittelt, also die relativen Häufigkeiten der „Ja“-Ausprägungen. Der NRE für ein Merkmal ergibt sich als Differenz „Anteil in der Deliktgruppe mit NR“ minus „Anteil in der Deliktgruppe ohne NR“, kann also positiv sein (bei höherem Anteil in der Gruppe mit NR) oder negativ (bei höherem Anteil in der anderen Gruppe).⁵ Bei einem positiven NRE ist das Merkmal ein Indikator zugunsten eines Near Repeats, bei einem negativen NRE ist es ein Indikator gegen ein Near Repeat. Der Score-Wert eines einzelnen Deliktes wird berechnet als Summe der NRE-Werte für alle seine Merkmale, die die Ausprägung „Ja“ aufweisen.

In den hier untersuchten Varianten des Ansatzes wurden die untersuchten Deliktdaten für die Jahre 2017 bis 2019 auch als Basisdaten für die Ermittlung der NRE-Werte verwendet. Dadurch könnte die Prognosequalität der untersuchten PreMAP-Ansätze etwas besser ausfallen als in einem Echtbetrieb, in dem ja immer Daten aus einem zurückliegenden Basiszeitraum für den Prognosezeitraum verwendet werden.

3.3.1.2 Aktueller PreMAP-Ansatz mit standardisierten Input-Variablen

Aufgrund der weitgehend automatisierten Auswertungsläufe ergab sich aus der oben beschriebenen Standardisierung der Input-Variablen eine Variante des aktuellen PreMAP-Ansatzes als Zufallsentdeckung: Der aktuelle PreMAP-Ansatz wurde ebenfalls mit den standardisierten Input-Variablen durchgerechnet,⁶ und überraschenderweise ergaben sich bessere Prognoseergebnisse als ohne die Standardisierung. Daraufhin wurde der modifizierte Ansatz für alle untersuchten Delikttypen geprüft und lieferte zumindest bei einem Teil eine bessere Prognosequalität als der Ansatz ohne Standardisierung.

3.3.1.3 Bayesianische Logistische Regression (BLR)

Ein etablierter Ansatz zur datengestützten Ermittlung von Zusammenhängen zwischen verschiedenen Variablen ist die Lineare Regression. Sie ermöglicht es, den Einfluss von ausgewählten Input-Variablen auf eine Output-Variable zu quantifizieren und – unter der Voraussetzung, dass die Werte für die Input-Variablen bekannt oder plausibel erwartbar sind – für Prognosen zu nutzen. Für den Fall, dass es sich bei der Output-Variable wie hier um eine binäre Variable handelt („Delikt wird ein Near Repeat haben: Ja / Nein“), wird häufig die Logistische Regression als geeignete Weiterentwicklung der Linearen Regression angewendet. Für den auch hier vorliegenden Fall, dass auch in den Input-Variablen zahlreiche binäre Variablen enthalten sind (z. B. „Tatmittel Bohrmaschine: ja / nein“), existiert als zusätzliche Weiterentwicklung der Ansatz der

⁵ Aus wahrscheinlichkeitstheoretischer Sicht kann man eine relative Häufigkeit immer als empirischen Näherungswert für eine theoretische Wahrscheinlichkeit ansehen. Die hier betrachteten relativen Häufigkeiten entsprechen also den beiden Wahrscheinlichkeiten, mit denen das betrachtete Merkmal bei den Delikten mit NR bzw. bei denen ohne NR die Ausprägung „Ja“ aufweist. Der NRE ist also als merkmalspezifische Wahrscheinlichkeitsdifferenz zwischen den beiden Deliktgruppen zu interpretieren.

⁶ Dass der Durchlauf ohne Fehlermeldung möglich war, liegt daran, wie der aktuelle PreMAP-Ansatz in KNIME programmiert wurde: Die relativen Häufigkeiten wurden nämlich nicht durch Auszählen der 1-Werten (für „Ja“) und Teilen durch die Gesamtzahl ermittelt, sondern mathematisch äquivalent als Mittelwert der jeweiligen Input-Variable mit den möglichen Zahlenwerten 0 und 1. Diese Berechnung funktionierte ohne Fehlermeldung auch für die standardisierten Input-Werte. Für die Verwendung des Mittelwertes, der jetzt nicht mehr einer relativen Häufigkeit entspricht, kann möglicherweise eine wahrscheinlichkeitstheoretische Begründung gefunden werden. Hier reicht erst einmal die rein empirische Begründung aus, dass der Ansatz mit standardisierten Input-Variablen teilweise die Prognosequalität verbessert.

Bayesianischen Logistischen Regression, der ursprünglich für die automatisierte Textanalyse entwickelt wurde. Wir haben diesen Ansatz in verschiedenen Varianten für alle untersuchten Delikttypen geprüft, und zwar unter Verwendung des KNIME-Funktionsmoduls („KNIME-Knoten“) „BayesianLogisticRegression (3.7)“.⁷

3.3.1.4 Tree Ensemble

Ein ebenfalls sehr verbreiteter Ansatz zur Prognose von binären Variablen ist der Entscheidungsbaum („Decision Tree“). Dabei wird datengestützt ein Baumdiagramm ermittelt, in dem man ablesen kann, in welcher Gewichtung die einbezogenen Input-Variablen auf Verzweigungen hin zu einer oder zur anderen Ausprägung der Output-Variable einwirken.⁸ Um mögliche Verzerrungen in den Daten herauszumitteln, kann man anstelle eines einzelnen Entscheidungsbaums mehrere Bäume mit zufälligen Variationen der einbezogenen Daten ermitteln und die Ergebnisse aller Bäume in geeigneter Weise zusammenfassen – man spricht von Zufallswäldern („Random ForestsTM“) oder Baum-Ensembles.⁹ Eine Variante dieser Vorgehensweise, die sich in vorangegangenen Untersuchungen als besonders vielversprechend erwiesen hat, ist in dem Funktionsmodul „Tree Ensemble Learner“ in KNIME implementiert und wurde für drei der hier untersuchten Delikttypen verwendet.

3.3.2 Raumbezogene Prognoseansätze: Delikthäufigkeits-Klasse eines Wohnquartiers in der Prognosewoche

Bei den raumbezogenen Prognoseansätzen wurde hier das Ziel verfolgt, für jedes Wohnquartier (eine räumliche Einheit mit je ca. 400 Haushalten) in Braunschweig und Hannover für die kommende Woche die Zugehörigkeit zu einer von drei Delikthäufigkeits-Klassen (Klasse 0: 0 Delikte, Klasse 1: 1 Delikt, Klasse 2: 2 und mehr Delikte) zu prognostizieren. Das würde eine Ampeldarstellung mit den Farben grün-gelb-rot für die zunehmende Delikthäufigkeit ermöglichen. Weiterhin ergibt sich implizit eine gröbere Klassenzuordnung, nämlich die Klasse-0-Prognosen (0 Delikte) und die Nicht-Klasse-0-Prognosen (1 oder mehr Delikte, im Folgenden kürzer „Nicht-0-Prognosen“ oder „Alarme“).¹⁰

⁷ Erläuterungen zu den KNIME-Funktionsmodulen, darunter auch Literaturangaben zu den methodischen Grundlagen, finden sich unter www.nodepit.com (zuletzt abgerufen am 29.03.2021). Dort kann rechts oben die verwendete KNIME-Version und damit die passende Erläuterungs-Version ausgewählt werden (hier: „v4.0“).

⁸ Ein bekanntes Beispiel ist die Modellierung der Output-Variable „Wird den Titanic-Untergang überleben: Ja / Nein“ mit den Informationen aus der Passagierliste als Input-Variablen. Dabei zeigt sich, dass die Variable Geschlecht den größten Einfluss hat (weibliche Personen überleben tendenziell, männliche nicht), gefolgt von anderen Input-Variablen (z. B. Alter, gebuchte Klasse, Preis des Tickets), die möglicherweise bei Frauen und Mädchen doch noch die Überlebenschance reduzieren oder sie bei Männern und Jungen erhöhen.

⁹ In gewisser Hinsicht wird hier der Gedanke, dass man durch mehrfache Anwendung eines Ansatzes mit zufälligen Variationen der einbezogenen Daten eine bessere Prognosequalität erreichen kann, doppelt angewendet: innerhalb des Tree-Ensemble-Ansatzes mit 100 Entscheidungsbäumen, und dann noch einmal in der Form, dass wir den Tree-Ensemble-Ansatz mit 100 Trainingsmengen geprüft haben. Die Boxplots auf den Evaluationsblättern in Anhang B zeigen aber, dass die Variation in einem Ensemble nicht dazu führt, dass es auf der Ebene der 100 Trainingsmengen keine Variation der Prognosequalitätskennzahlen mehr gibt, denn dann müssten sich sehr flache Boxen und sehr kurze Whiskers oder im Extremfall eine waagerechte Linie ergeben.

¹⁰ Prognoseansätze, die von vorneherein nur eine binäre Klassenzugehörigkeit „Delikte: Ja / Nein“ liefern würden, könnten hinsichtlich der Prognosequalität andere Ergebnisse aufweisen, wurden aber hier nicht untersucht.

3.3.2.1 Benchmark-Ansatz: Prognose der Output-Variablen in der Prognosewoche durch ihren Wert in der Vorwoche

Ein sehr einfacher Prognoseansatz besteht darin, die Klassenzugehörigkeit eines Wohnquartiers in der Prognosewoche allein durch seine Klassenzugehörigkeit in der Vorwoche zu bestimmen. Dieser Ansatz kommt ohne das Trainieren eines Modells und ohne jegliche Input-Daten aus, er benötigt nur die tatsächlichen Ausprägungen der interessierenden Output-Variablen (Delikthäufigkeits-Klasse 0, 1 oder 2). Dieser „datenignorierende“ Minimalansatz wurde für alle Delikttypen als zentraler Vergleichsmaßstab (Benchmark-Ansatz) mit untersucht.

Um weiteren Variablen in die Prognose für die Prognosewoche t einzubeziehen, bestehen grundsätzlich zwei Möglichkeiten:

1. Man trainiert das Prognosemodell für den Zusammenhang zwischen der Output-Variablen (Delikthäufigkeits-Klasse) in der Prognosewoche (Woche t) und den Input-Variablen in der Vorwoche (Woche $t-1$).¹¹
2. Man trainiert das Prognosemodell für den Zusammenhang zwischen der Output-Variablen (Delikthäufigkeits-Klasse) in der Woche $t-1$ und den Input-Variablen in Woche $t-1$. Dann übernimmt man den modellbestimmten Wert der Output-Variablen – nicht den tatsächlich beobachteten wie im Benchmark-Ansatz – als Prognose für die Prognosewoche t .¹²

Die zweite Herangehensweise hat sich in unseren Untersuchungen als grundsätzlich erfolgreicher erwiesen, daher wird im Folgenden auf die Ergebnisdokumentation für die erstgenannte Variante verzichtet.

3.3.2.2 Multinomiale Logistische Regression (MLR)

Zur oben beim BLR-Ansatz erwähnten „einfachen“ Logistischen Regression für binäre Output-Variablen existiert auch eine Weiterentwicklung für nominale Output-Variablen mit mehr als zwei Ausprägungen wie den hier betrachteten drei Klassenzugehörigkeiten, nämlich die Multinomiale Logistische Regression (MLR). Die entsprechenden Berechnungen wurden hier mit dem KNIME-Funktionsmodul „Logistic Regression Learner“ durchgeführt, mit dem auch MLR-Modelle trainiert werden können.

3.2.2.3 Multinomiale Logistische Ridge-Regression (MLRR)

Bei jedem Regressionsansatz wird die Stärke des Zusammenhangs zwischen der einzelnen Input-Variablen und der Output-Variablen durch einen Koeffizienten ausgedrückt, der beim Trainieren des Modells¹³ ermittelt wird. Im Allgemeinen ermittelt man diese Koeffizienten ohne weitere Vorgaben aus den in den Daten vorgefundenen Input-Output-Kombinationen durch geeignete

¹¹ Dabei kann man auch die Delikthäufigkeits-Klasse in Woche $t-1$ als Input-Variable einbeziehen. So wird gewissermaßen der Benchmark-Ansatz um zusätzliche Input-Variablen neben der Delikthäufigkeits-Klasse erweitert. Aufgrund langer Laufzeiten und der zeitlich begrenzten Datenverfügbarkeit wurden solche Prognoseansätze, bei denen die Häufigkeitsklasse als Input- (Woche $t-1$) und als Output-Variable (Woche t) einbezogen wird, nicht geprüft.

¹² Hier wird gewissermaßen der Benchmark-Ansatz dahingehend abgewandelt, dass anstelle der beobachteten Häufigkeitsklasse in Woche $t-1$ die modellbestimmte Häufigkeitsklasse in Woche $t-1$ als Prognose für die Häufigkeitsklasse in Woche t verwendet wird. Dabei wird unterstellt, dass die modellbestimmte Häufigkeitsklasse in Woche $t-1$ aussagekräftiger ist als die tatsächliche und dass sie daher auch eine bessere Prognose für Woche t ermöglicht. Die Gültigkeit dieser Unterstellung wird im Folgenden durch den Vergleich der Prognosequalitäten des Benchmark-Ansatzes und der Ansätze mit der hier skizzierten Vorgehensweise empirisch überprüft.

¹³ Bei Regressionsansätzen spricht man auch vom „Schätzen“ des Modells, es handelt sich dennoch um ein exaktes mathematisches Verfahren.

Rechenschritte. Es gibt aber auch Ansätze, die den Koeffizienten beim Trainieren bestimmte Restriktionen auferlegen, um bestimmte Werte oder eine bestimmte Wahrscheinlichkeitsverteilung der Koeffizienten zu begünstigen.

Ein solcher Ansatz ist die Multinomiale Logistische Ridge-Regression (MLRR), die sich ganz allgemein als geeignet erwiesen hat, um die ungünstigen Auswirkungen von Multikollinearität abzuschwächen. Multikollinearität liegt vor, wenn drei oder mehr Input-Variablen einen engen linearen Zusammenhang zueinander aufweisen. Sie kann als Mehr-Variablen-Verallgemeinerung der oben besprochenen hohen Korrelation zwischen zwei Variablen angesehen werden und wird durch die oben beschriebene Maßnahme, bei zwei hoch korrelierten Variablen eine davon aus dem Prognoseansatz herauszulassen, nicht ausgeschlossen.

Der MLRR-Ansatz wurde ebenfalls mit dem KNIME-Funktionsmodul „Logistic Regression Learner“ – mit einer entsprechend angepassten Einstellung – umgesetzt.

3.2.2.4 Multinomial Functional Tree (MFT)

Der Functional-Tree-Ansatz ist eine Kombination des Entscheidungsbaumansatzes mit dem Ansatz der Logistischen Regression, die in geeigneter Form an den Verzweigungen bzw. am Ende der Zweige zum Einsatz kommt, und existiert auch in einer Variante für multinomiale Output-Variablen (MFT-Ansatz), die hier zur Prognose der drei Delikthäufigkeits-Klassen verwendet wurde. Die Berechnungen wurden mit dem KNIME-Funktionsmodul „FT (3.7)“ (für „Functional Tree“) durchgeführt.

Untersuchte Prognoseansätze:

- **Tatbezogene Prognosen („Tat wird Near Repeat haben: Ja / Nein“):**
 - **Aktueller PreMAP-Ansatz:** Prognose / Risikobewertung anhand eines Scores mit Score-Beiträgen je Tatmerkmal
 - **Aktueller PreMAP-Ansatz mit standardisierten Daten:** Prognose / Risikobewertung anhand eines Scores mit Score-Beiträgen je Tatmerkmal (nach mathematischer Umskalierung)
 - **Bayesianische Logistische Regression (BLR):** Form der Regression, die besonders für binäre Input- und Output-Variablen geeignet ist
 - **Tree Ensemble:** Prognosen auf Grundlage von mehreren Entscheidungsbäumen
- **Raumbezogene Prognosen: („Gebiet wird in der Prognosewoche zu folgender Häufigkeitsklasse gehören: 0 Delikte / 1 Delikt / mehr als 1 Delikt“)**
 - **Benchmark-Ansatz:** Häufigkeitsklasse in Woche t-1 als Prognose für die Woche t (Prognosewoche), „datenignorierender Ansatz“
 - **Multinomiale Logistische Regression (MLR):** Form der Regression, die zur Prognose von nominalen Output-Variablen mit mehr als zwei möglichen Ausprägungen verwendet wird (hier: Zugehörigkeit zu einer von drei Klassen)
 - **Multinomiale Logistische Ridge-Regression (MLRR):** Eine Weiterentwicklung der MLR, die besonders robust gegen miteinander korrelierende Input-Variablen ist
 - **Multinomial Functional Tree (MFT):** Prognose durch Kombination von Entscheidungsbäumen und logistischen Regressionen

4. Ergebnisse

Im Folgenden wird die Prognosequalität der verschiedenen tat- bzw. raumbezogene Prognoseansätze evaluiert. Dabei werden nicht alle im Rahmen der erweiterten Pilotierung gerechneten Ansätze präsentiert, da dies den Rahmen sprengen würde. Stattdessen wurde eine Auswahl vorgenommen, die besonders relevante Ansätze umfasst. Die Charakteristika und Qualitätskennzahlen der hier präsentierten tatbezogenen Ansätze sind in Anhang B jeweils einzeln auf einem Evaluationsblatt dokumentiert und kommentiert, im Unterabschnitt 4.1.1 findet sich eine umfangreiche Erläuterung zu den Evaluationsblättern als Lesehilfe. Diese ansatzspezifischen Evaluationsergebnisse zu den vier untersuchten Delikttypen werden in den nachfolgenden Unterabschnitten weiter verdichtet. Für die raumbezogenen Prognoseansätze sind ähnliche Charakterisierungen durch Tabellen und Abbildungen direkt in den kommentierenden Text im übernächsten Abschnitt eingebettet, um durch synoptische Darstellungen eine bessere Vergleichbarkeit – bei insgesamt weniger augenfälligen Unterschieden als im Fall der tatbezogenen Ansätze – zu ermöglichen. Dort, unter 4.2.1, finden sich ebenfalls ausführliche Erläuterungen als Lesehilfe.

Relevante Kennzahlen zur Bewertung der Prognosequalität:

- **TP = True Positives:** Anzahl der Delikte, für die zutreffend ein Near Repeat prognostiziert wurde („zutreffende Alarme“)
- **FP = False Positives:** Anzahl der Delikte, für die fälschlich ein Near Repeat prognostiziert wurde („falsche Alarme“)
- **TPR = True Positive Rate:** Anteil der Delikte, für die zutreffend ein Near Repeat prognostiziert wurde („zutreffende Alarme“) an denjenigen Delikten, die auch tatsächlich ein Near Repeat aufwiesen.
- **FPR = False Positive Rate:** Anteil der Delikte, für die fälschlich ein Near Repeat prognostiziert wurde („falsche Alarme“), an denjenigen Delikten, die in Wahrheit kein Near Repeat aufwiesen
- **PPV = Positive Predictive Value = $TP / (TP + FP)$:** Anteil der zutreffenden an allen Alarmen, auch: positiver Vorhersagewert, Genauigkeit (Precision)
- **Kriterien für hohe Qualität eines Prognoseansatzes:** hoher TPR-Wert, niedriger FPR-Wert (damit auch PPV möglichst hoch)

4.1 Tatbezogene Prognoseansätze: Evaluation der Prognosequalität

4.1.1 Erläuterungen zu den Evaluationsblättern in Anhang B

TB 1.3.G Delikttyp 1: WED / TWE (Deliktschlüssel 435 und 436)

1

Prognoseansatz: Bayesianische Logistische Regression (BLR)

2

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	831	60	891
Hannover, Landeshauptstadt	3.359	419	3.778
Gesamt	4.190	479	4.669

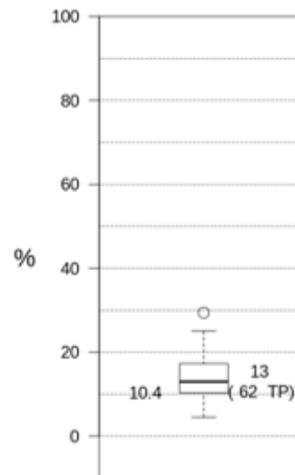
3

Trainings- und Testdaten: H u. BS gesamt
TPR, FPR berechnet für: H u. BS gesamt

4

Gesamtmodell, Gesamtauswertung

Nexiga-Daten einbezogen: Nein
Korrelationsschwelle: 0,8
Wahrscheinlichkeitsschwelle: 0,95

True-Positive-Rate
(Testdaten)False-Positive-Rate
(Testdaten)**5**

Charakteristische Beobachtungen:

6

- ➔ Mediane: eindeutig schlechter als bei TB 1.2.G (TPR-Median niedriger, FPR-Median höher); Vergleich mit TB 1.1.G uneindeutig (beide niedriger als dort)
- ➔ Aber: TPR-Boxunterkante sowie unteres TPR-Whisker-Ende höher und FPR-Boxoberkante niedriger als bei TB 1.2.G, FPR-Streuung insgesamt ähnlich wie dort
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 12,4 \%$ (etwas schlechter als bei TB 1.2.G)

Besonders relevante Modellvergleiche:

7

- TB 1.n.G für n = 1, 2 (alternative Ansätze, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- TB 1.4.G (gleicher Ansatz mit Einbeziehung der Nexiga-Daten, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Bes. rel. Modellvergleiche“.
- TB 1.7.G (alternativer Ansatz mit Einbeziehung der Nexiga-Daten, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Charakteristische Beobachtungen“.

- 1) Durch das Kürzel „TB“ wird gekennzeichnet, dass es sich bei dem präsentierten Ergebnis um einen tatbezogenen Ansatz handelt, der für ein gegebenes Delikt die Eigenschaft „Wird ein Near Repeat haben: Ja / Nein“ prognostiziert. Die erste Ziffer steht für den Delikttyp (auch ausführlich mit PKS-Schlüsseln in der Überschrift), die zweite für den betrachteten Prognoseansatz und seine konkrete Umsetzung (einbezogene Input-Daten – Variablenauswahl und/oder Regionalauswahl). Die Kürzel „G“, „H“ und „BS“ geben an, ob die Evaluation der Prognosequalität für Hannover und Braunschweig zusammen („Gesamtregion“) oder für Hannover oder Braunschweig alleine durchgeführt wurde (s. u.: „TPR / FPR berechnet für ...“)
- 2) Hier wird auf den verwendeten Prognoseansatz verwiesen. In diesem Fall handelt es sich um die Bayesianische Logistische Regression. Details und Beschreibungen zu allen untersuchten Ansätzen finden sich im Abschnitt 3.3 „Untersuchte Prognoseansätze“.
- 3) Diese Tabelle, die sich auf jedem Evaluationsblatt findet, bietet Orientierung die Datengrundlage betreffend. Sie stellt für die Städte Braunschweig und Hannover sowie die Summe beider dar, wie viele Fälle des in Rede stehenden Delikts in dem für Modelltraining und -evaluation betrachteten Zeitraum (2017, 2018, 2019) vorliegen, wie viele davon tatsächlich ein Near Repeat nach spätestens drei Tagen im Umkreis von 500 Metern aufwies bzw. dies nicht taten. Im aufgeführten Beispiel umfasste das verwendete Datenmaterial für H und BS zusammen 4.669 Delikte, davon hatten 479 tatsächlich ein Near Repeat (dritte Zeile). Die entsprechenden Angaben für H und BS finden sich in den Zeilen darüber.
- 4) In diesem Block sind die wesentlichen Charakteristika des Prognoseansatzes dargestellt:
 - a. Trainings- und Testdaten: Trainingsdaten werden verwendet, um ein Modell (eine konkrete Ausprägung des Prognoseansatzes) zu trainieren. Eine kleine Teilmenge (hier: 10 %) der Gesamtdaten wird üblicherweise als Testdaten zurückgehalten, um die Prognosequalität des Modells anhand von Daten, mit denen es nicht trainiert wurde, zu evaluieren. Aus diesem Eintrag geht hervor, ob die 100 Trainingsmengen jeweils als Zufallsstichprobe aus den Gesamtdaten (Hannover und Braunschweig) oder aus den Teildaten für Hannover oder Braunschweig gezogen wurden. In diesem Falle erfolgte dies für die Gesamtdaten (Hannover und Braunschweig).
 - b. Nexiga-Daten einbezogen: Hier ist angegeben, ob zusätzlich zu den NIVADIS-Daten die zu erprobenden sozioökonomischen Daten in den Ansatz eingeflossen sind. Bei „nein“ wurden ausschließlich Daten aus NIVADIS verwendet. In Anhang A.1 ist für jeden hier dargestellten Ansatz angegeben, welche Input-Variablen genau einbezogen wurden.
 - c. Korrelationsschwelle: Bei Prognoseansätzen ist es grundsätzlich ungünstig, wenn es starke lineare Zusammenhänge, im einfachsten Fall hohe paarweise Korrelationen, zwischen den Input-Variablen gibt. In den Fällen, wo die Korrelation zweier Variablen über dem hier angegebenen Schwellenwert lag, wurde eine der beiden Variablen aus der Analyse ausgeschlossen.
 - d. Wahrscheinlichkeitsschwelle: Einige der untersuchten Prognoseansätze liefern neben der bloßen Ja-Nein-Prognose für das Merkmal „Delikt wird ein Near Repeat haben“ einen Schätzwert für die Wahrscheinlichkeit für ein Near Repeat. Die entsprechende Schwelle ist an dieser Stelle angegeben. Beim aktuellen PreMAP-Ansatz sowie bei seiner Variante mit standardisierten Input-Variablen ist stattdessen die Score-Schwelle angegeben, oberhalb derer ein Near Repeat prognostiziert wird.

- 5) In diesem Bereich wird die Prognosequalität des jeweiligen Ansatzes durch Boxplots – ausführlicher: „Box-and-Whiskers-Plots“ – der True Positive Rates (TPR) und der False Positive Rates (FPR) für die 100 Modelle dargestellt, die der Ansatz für 100 unterschiedliche Trainingsmengen geliefert hat. Der linke Boxplot fasst die 100 TPR-Werte graphisch zusammen. Die eingezeichnete Box hebt die mittleren 50 % der Werte aus den 100 Durchläufen hervor. Oberhalb der Box liegen 25 % der aufgetretenen Werte, in der Box 50 % und unterhalb der Box wiederum 25 %. Im Beispiel liegt die Box-Unterkante bei einer TPR von 15,1 %, also liegt die TPR für drei Viertel der Durchläufe über diesem Wert. Der Median – der als „mittlerer Wert“ genau in der Mitte der aufsteigend sortierten TPR-Werte aller Durchläufe zu finden ist – ist als dicker waagerechter Strich im Boxinneren eingetragen. Er beträgt im Beispiel 19 %, was 91 True-Positive-Fällen, also „zutreffenden Alarmen“ entspricht. Der höchste Wert oberhalb der Box, der nicht weiter als die eineinhalbfache Boxhöhe von der Box entfernt liegt, wird am Ende der gestrichelten Linie, die oben aus der Box herausragt („oberer Whisker“), durch einen waagerechten Strich markiert (hier: ca. 35 %, ist auch der höchste auftretende Wert). Sollte es noch höhere Werte geben, werden diese „Ausreißer“ einzeln durch Kringel markiert. Entsprechendes gilt für den unteren Whisker (hier: ca. 5 %, ist auch der niedrigste auftretende Wert) und eventuelle Ausreißer nach unten. Die entsprechende Darstellung für die 100 aufgetretenen FPR-Werte ist im rechten Boxplot dargestellt. Hier ist anstelle der Box-Unterkante die Box-Oberkante beschriftet (hier 16 %, drei Viertel der auftretenden Werte liegen darunter). Weiterhin findet sich wieder der Median der FPR über alle Durchläufe (14,7 %, entspricht 615 False-Positive-Fällen, also „falschen Alarmen“). Im Beispiel liegen – nicht zuletzt aufgrund der geringen Boxhöhe – Ausreißer oberhalb bzw. unterhalb der Whiskers vor.
- 6) An dieser Stelle werden die wesentlichen Aussagen zum präsentierten Prognoseansatz dargestellt.
- 7) Die Prognosequalität eines Ansatzes lässt sich besser einschätzen, wenn sie mit der anderer Ansätze verglichen wird. Hier sind jene angegeben, die für einen Vergleich besonders relevant sind.

4.1.2 Delikttyp 1 „WED / TWE“

Beim Delikttyp 1 „WED / TWE“ liefert der aktuelle PreMAP-Ansatz¹⁴ TB 1.1 mit Daten aus rein polizeilicher Quelle (VBS NIVADIS) in der untersuchten Form für beide Städte (Hannover und Braunschweig zusammen) eine mäßige Prognosequalität mit einem PPV von 12,2 %. Der PPV kann auf 13,6 % erhöht werden, indem man die gleichen Input-Daten in standardisierter Form verwendet (TB 1.2) Damit einher geht eine TPR-Verteilung¹⁵, die nun günstigerweise relativ zur FPR-Verteilung weiter oben liegt (ablesbar an den Boxplots). In der Einzelstadtbetrachtung für Hannover liefert dieser Ansatz nicht so gute Ergebnisse (PPV: 13,1 %) wie für beide Städte zusammen. Für Braunschweig alleine liefert der Ansatz TB 1.2 sogar noch einen ungünstigeren PPV (12,5 %), obwohl die TPR- und die FPR-Verteilung günstiger sind als für die Gesamtregion.¹⁶

Ein alternativer Prognoseansatz (Bayesianische Logistische Regression, BLR) liefert zunächst in der Variante mit den reinen Polizeidaten (TB 1.3) für beide Städte zusammen eine minimale PPV-Verbesserung (auf 12,4 %) gegenüber dem aktuellen PreMAP-Ansatz, aber nicht in der gleichen Höhe wie der modifizierte PreMAP-Ansatz mit standardisierten Input-Daten (TB 1.2). Dies gelingt erst durch die Einbeziehung der externen sozioökonomischen Nexiga-Daten in den BLR-Ansatz (TB 1.4). Hier ergibt sich bei günstigerem TPR- und FPR-Median als in TB 1.2 (und damit auch als in TB 1.1) ein PPV von 14,4 %. In der Einzelstadtbetrachtung für Hannover liefert das BLR-Gesamtmodell mit Nexiga-Daten keine Verbesserung gegenüber TB 1.2 (PPV: 13,0 %), aber eine deutliche in der Braunschweig-Betrachtung¹⁷ (PPV: 20,0 %). Die Einzelstadt-Prognosequalität des BLR-Ansatzes mit Nexiga-Daten konnte weder für Hannover noch für Braunschweig durch Trainieren von jeweiligen Einzelstadt-Modellen (TB 1.5.H bzw. TB 1.6.BS) verbessert werden. Das bedeutet, die Daten in der einen Stadt liefern dem Ansatz nützliche Informationen für die Prognosen in der jeweils anderen Stadt.

Ein weiterer alternativer Prognoseansatz (Tree Ensemble, TB 1.7.G, hier ohne Einbeziehung der Nexiga-Daten) wurde nur in einer Variante untersucht, die deutlich höhere TPR- und FPR-Werte lieferte als die bisher beschriebenen Ansätze, was die Vergleichbarkeit einschränkt. Dennoch lassen sich der TPR- unterhalb des FPR-Medians, die TPR-Boxunterkante von 0 %, die sehr hohe FPR-Boxoberkante und der sehr niedrige PPV (8,6 %) so interpretieren, dass dieser Ansatz die bisher festgestellten Prognosequalitäten nicht erreicht.¹⁸

¹⁴ Der hier untersuchte Ansatz entspricht im Hinblick auf das rechnerische Prognoseverfahren dem aktuell angewendeten PreMAP-Ansatz, aber nicht im Hinblick auf die Verdichtung der polizeilichen Daten. Letztere stimmt mit der für die übrigen hier betrachteten Ansätze überein.

¹⁵ Zur Erinnerung: Der jeweilige Ansatz wurde mit 100 verschiedenen Trainingsmengen geprüft, so dass sich auch 100 TPR- und FPR-Werte ergaben (die nicht alle verschieden sein müssen). Bei der zusammenfassenden Betrachtung aller 100 Werte spricht man auch von der empirischen Verteilung oder kurz von der Verteilung der Werte, also hier beispielsweise von der Verteilung der TPR-Werte oder kürzer von der TPR-Verteilung.

¹⁶ Auf den ersten Blick erscheint es verwunderlich, dass der Ansatz TB 2.1 jeweils für Hannover *und* Braunschweig einen schlechteren PPV liefert als für die Gesamtregion, für die man eine Art von „mittlerem Ergebnis“ der beiden Teilregionen erwarten könnte, zumal bei TB 2.1.H/BS die Prognosequalität mit denselben 100 Testmengen für dieselben 100 trainierten Modelle evaluiert wird wie bei TB 2.1.G. Die Erklärung, warum der Ansatz für beide Teilregionen schlechter sein kann als für die Gesamtregion, ist die, dass eben für TB 2.1.H und TB 2.1.BS nur die jeweilige Teil-Testmenge betrachtet wird und sich dadurch im Zähler *und* im Nenner der TPR- und FPR-Berechnung jetzt H- und BS-spezifische Anzahlen anstelle der Anzahlen für die Gesamtregion finden. (Weitere Abweichungen in Zähler und Nenner ergeben sich in den einzelnen Durchläufen durch die Zufallsziehung der Trainingsmengen. Dies ist auch einer der Gründe, warum überhaupt Quoten – engl. „Rates“ – und nicht absolute TP- und FP-Anzahlen betrachtet werden.)

¹⁷ Der hohe PPV beruht allerdings auf einer deutlichen Verbesserung der FPR-Verteilung gegenüber TB 1.2.BS. Eine denkbare gleichzeitige Verbesserung der TPR-Verteilung wurde nicht geprüft.

¹⁸ Auch ein Tree-Ensemble-Ansatz mit Einbeziehung der Nexiga-Daten, der allerdings nicht mit der neu-entwickelten Kategorisierung der NIVADIS-Daten geprüft wurde, erreichte im Hinblick auf die relative

Damit hat sich beim Delikttyp 1 der BLR-Ansatz mit Einbeziehung der Nexiga-Daten als der beste Ansatz für H und BS zusammen erwiesen (und potenziell auch für Braunschweig alleine), aber eine Variante des aktuellen PreMAP-Ansatzes mit zusätzlicher Standardisierung der polizeilichen Input-Daten liefert die besten Prognosen für Hannover.

4.1.3 Delikttyp 2 „Schwerer Diebstahl aus Nichtwohnungen (ohne Ladendiebstahl)“

Beim Delikttyp 2 „Schwerer Diebstahl aus Nichtwohnungen (ohne Ladendiebstahl)“ muss darauf hingewiesen werden, dass die später zu prognostizierende Output-Variable „Delikt hat ein Near Repeat: Ja / Nein“ in den als Trainings- und Testmengen verwendeten Daten jeweils „schlüsselscharf“ innerhalb eines PKS-Dreisteller-Schlüssels ermittelt wurde.¹⁹ Damit kann z. B. ein Einbruch in einen Verkaufsraum in den Daten kein „tatsächliches“ Near Repeat-Delikt eines Einbruchs in eine Gaststätte sein. Dennoch soll der Prognoseansatz sich auf die Gesamtheit der drei Schlüssel beziehen und einen solchen Near Repeat-Zusammenhang prognostizieren können. Damit entsprechen die „tatsächlichen“ Near Repeats in den Trainings- und Testmengen jeweils nur einem Teil der Near Repeats, die der Prognoseansatz liefern soll und mit denen er im günstigsten Fall trainiert werden sollte. Ohne die „schlüsselscharfe“ Abgrenzung wären also besser trainierte Prognosemodelle zu erwarten, allerdings gäbe es in den Testmengen auch mehr „Positives“ und damit tendenziell niedrigere TPR-Werte (höherer Nenner) sowie tendenziell höhere FPR-Werte (niedrigerer Nenner).

Trotz dieser Einschränkung besitzt bei diesem Delikttyp (in der gemeinsamen Betrachtung der drei Schlüssel) der aktuelle PreMAP-Ansatz mit den neu kategorisierten NIVADIS-Daten (TB 2.1) eine deutlich bessere Prognosequalität für die Gesamtregion als beim Delikttyp 1 (PPV: 32,0 %, TPR-Box deutlich oberhalb der FPR-Box). Für Hannover alleine ist die Qualität noch besser (PPV: 35,0 %), für Braunschweig dagegen in der Medianbetrachtung maximal schlecht (TPR-Median von 0,0 % führt zu einem PPV von 0,0 %). Durch die Standardisierung der Input-Variablen (TB 2.2) kann bei diesem Delikttyp die Prognosequalität für beide Städte zusammen nicht verbessert werden (PPV: 31,3 %), allerdings verbessert sich in dieser Variante die Einzelstadt-Prognosequalität für Braunschweig (PPV: 8,0 %, aber bei einer weiterhin problematischen TPR-Boxunterkante von 0,0 %).

Der BLR-Ansatz ohne Nexiga-Daten (TB 2.3) liefert für beide Städte zusammen eine sehr ähnliche, aber minimal geringere Prognosequalität (PPV: 31,5 %). Durch die Einbeziehung der Nexiga-Daten (TB 2.4) kann diese deutlich gesteigert werden (PPV: 41,2 %). In der Einzelstadtbeurteilung fällt sie für Hannover alleine sogar noch besser aus (PPV: 42,7 %) – und damit auch deutlich besser als bei TB 2.1.H. Für Braunschweig alleine ist die Prognosequalität deutlich schlechter als für beide Städte zusammen, aber immerhin deutlich besser als bei TB 2.2.BS (PPV: 19,4 %, die TPR-Boxunterkante ist allerdings immer noch 0). Ein Trainieren des gleichen Ansatzes mit auf die jeweilige Stadt eingeschränkten Trainingsdaten bringt weder für Hannover (TB 2.5.H) noch für Braunschweig (TB 2.6.BS) eine weitere Verbesserung. Auch hier liefern also die Daten in der einen Stadt dem Ansatz nützliche Informationen für die Prognosen in der jeweils anderen Stadt.

Lage der TPR- zur FPR-Box sowie auf den PPV nicht die Prognosequalität der übrigen Ansätze, vgl. Evaluationsblatt TB 1.7.G im Anhang B.

¹⁹ Dieser Delikttyp wurde zusammengefasst aus den drei Schlüssel-Dreistellern 410 „Schwerer Diebstahl in / aus Dienst-, Büro-, Fabrikations-, Werkstatt- und Lagerräumen“, 415 „Schwerer Diebstahl in / aus Gaststätten, Kantinen, Hotels und Pensionen“ und 425 „Schwerer Diebstahl in / aus Kiosken, Warenhäusern, Verkaufsräumen, Selbstbedienungsläden, Schaufenstern und Schaukästen (ohne Ladendiebstahl)“.

Der Tree-Ensemble-Ansatz liefert in der hier untersuchten Variante ohne Nexiga-Daten für beide Städte zusammen trotz einem FPR-Median unterhalb des TPR-Medians, dafür aber bei einer ungünstig hohen FPR-Boxoberkante eine sehr viel schlechtere Prognosequalität als alle anderen Ansätze zu diesem Delikttyp (PPV: 19,7 %).

Zusammenfassend ist festzuhalten, dass beim Delikttyp 2 der BLR-Ansatz mit Einbeziehung der Nexiga-Daten die höchste Prognosequalität aufweist, und zwar sowohl für beide Städte zusammen als auch in Form eines mit den Gesamtdaten trainierten Modells für jede in der Einzelstadtbetrachtung.

4.1.4 Delikttyp 3 „Diebstahl in / aus KFZ“

Beim Delikttyp 3 „Diebstahl in / aus KFZ“ weist der aktuelle PreMAP-Ansatz mit den neu kategorisierten NIVADIS-Daten (TB 3.1) noch einmal eine deutlich bessere Prognosequalität für beide Städte zusammen auf (PPV: 43,3 %) als bereits beim Delikttyp 2, in der Einzelbetrachtung für Hannover wieder noch besser (PPV: 45,0 %) und für Braunschweig wieder deutlich schlechter (PPV: 23,7 %, hier aber immerhin ohne eine TPR-Boxunterkante von 0,0 %). Die Verwendung standardisierter Input-Daten (TB 3.2) bringt in der Gesamtbetrachtung keine Verbesserung, sondern eine sichtbare Verschlechterung (PPV: 40,7 %).

Auch der BLR-Ansatz ausschließlich mit den polizeilichen Daten (TB 3.3) liefert für beide Städte zusammen eine Prognosequalität unterhalb derer des aktuellen PreMAP-Ansatzes (PPV: 41,0 %). Erst durch die Einbeziehung der Nexiga-Daten (TB 3.4) lässt sich hier in der Gesamtbetrachtung eine sehr deutliche Verbesserung erzielen (PPV: 57,4 %), die in der Hannover-Betrachtung noch übertroffen wird (PPV: 58,3 %), während sie sich in der Braunschweig-Betrachtung in eine Verschlechterung verwandelt (PPV: 22,5 %). Die insgesamt sehr niedrigen TPR- und FPR-Werte sowie die außerdem wieder auftretende ungünstige TPR-Boxunterkante von 0,0 % wurden hier durch eine Anpassung der Wahrscheinlichkeitsschwelle behoben (TB 3.4a.BS). Daraus resultierte nun doch eine bessere Prognosequalität des Ansatzes in der Braunschweig-Betrachtung (PPV: 24,5 %) als bei TB 3.1.BS.

Das Trainieren des Ansatzes TB 3.4 mit Einzelstadt-Trainingsdaten liefert für Hannover eine minimale Verschlechterung der Prognosequalität (PPV: 58,1 %), für Braunschweig aber noch einmal eine sichtbare Verbesserung (PPV: 28,3 %) gegenüber der Variante TB 3.4a.BS. Mit anderen Worten, die Daten in Braunschweig liefern nützliche Informationen für die Prognosen in Hannover, aber die Daten in Hannover tragen nicht zur Prognosequalität in Braunschweig bei.

Der untersuchte Tree-Ensemble-Ansatz (TB 3.7), ebenfalls wieder ohne Einbeziehung der Nexiga-Daten, liefert für beide Städte zusammen eine schlechte Prognosequalität, denn hier führt eine TPR-Verteilung, die deutlich unterhalb der FPR-Verteilung liegt, zu einem für diesen Delikttyp sehr geringen PPV (15,9 %).²⁰

Insgesamt weist auch beim Delikttyp 3 der BLR-Ansatz mit Einbeziehung der Nexiga-Daten die höchste Prognosequalität auf, und zwar sowohl in der Gesamtbetrachtung für beide Städte zusammen als auch als Gesamtmodell in der Einzelstadtbetrachtung für Hannover. Für Braunschweig lässt sich bei diesem Delikttyp erstmals die Prognosequalität dadurch verbessern, dass in den Ansatz nur Einzelstadt-Trainingsdaten, also nur solche aus Braunschweig selbst, einbezogen werden („Braunschweig-Modell“).

²⁰ In einem Tree-Ensemble-Ansatz mit Einbeziehung der Nexiga-Daten, der allerdings nicht mit der neuentwickelten Kategorisierung der NIVADIS-Daten geprüft wurde, konnte der Abstand zwischen TPR- und FPR-Box verringert werden, der PPV blieb aber mit 25,5 % immer noch deutlich unter den Werten der anderen Ansätze, vgl. Evaluationsblatt TB 3.7.G im Anhang B.

4.1.5 Delikttyp 4 „Diebstahl von KFZ“

Beim Delikttyp 4 „Diebstahl von KFZ“ liefert der aktuelle PreMAP-Ansatz mit den neu kategorisierten NIVADIS-Daten (TB 4.1) in der untersuchten Form für beide Städte zusammen eine sehr mäßige Prognosequalität (PPV: 9,0 %). Dabei ergibt sich erstmals für diesen Ansatz die ungünstige Kombination mit einem TPR-Median unterhalb des FPR-Medians, außerdem erstmals in der Gesamtbetrachtung die sonst vereinzelt für die Braunschweig-Betrachtung zu bemängelnde TPR-Boxunterkante von 0,0 %. Die beiden letztgenannten Mängel können durch die Verwendung derselben Input-Daten in standardisierter Form (TB 4.2) behoben werden, dadurch erhöht sich auch der PPV etwas auf 11,0 %, was aber immer noch eine mäßige Prognosequalität anzeigt.

Der Wechsel zum BLR-Ansatz – zunächst ohne Einbeziehung der Nexiga-Daten (TB 4.3) liefert für beide Städte zusammen eine weitere Verbesserung (PPV: 13,5 %), und eine weitere – diesmal auffallende – Verbesserung resultiert aus der Einbeziehung der Nexiga-Daten (TB 4.4), die einen PPV von 18,0 % mit sich bringt. In der Hannover-Auswertung sind die entsprechenden Gesamtmodelle (TB 4.4.H) allerdings etwas schlechter (PPV: 17,1 %), und bei der Braunschweig-Auswertung (TB 4.4.BS) ergibt sich erstmals das Problem, dass die ganze Box des TPR-Boxplots auf den Wert 0,0 % zusammengeschrumpft ist, dass also mindestens drei Viertel aller TPR-Werte für die 100 Testmengen 0,0 % betragen.

Schon die Tatsache, dass damit auch ein TPR-Median von 0,0 % vorliegt, hat zur Folge, dass auch der Braunschweig-PPV für diesen Ansatz den Wert 0,0 % besitzt.²¹ Das Problem der TPR-Boxoberkante auf der Nulllinie kann dadurch behoben werden, dass der gleiche Ansatz nur mit Braunschweig-Daten (einschließlich der Nexiga-Daten) trainiert wird (TB 4.6.BS), dabei bleibt aber der TPR-Median und damit der PPV bei 0,0 %. Auch für Hannover lässt sich durch eine Beschränkung auf Hannover-Trainingsdaten (TB 4.5.H) eine Verbesserung gegenüber dem mit den Gesamtdaten trainierten Ansatz erreichen (PPV: 19,0 %). Also leisten die Daten in Braunschweig keinen Beitrag zur Prognosequalität in Hannover, und umgekehrt ist das auch nicht der Fall.

Beim hier untersuchten Tree-Ensemble-Ansatz ohne Nexiga-Daten zeigt sich erstmals ein TPR-Median deutlich oberhalb des FPR-Medians und damit ein PPV in ähnlicher Größenordnung wie bei den anderen Ansätzen. Aufgrund der sehr großen Streuung bei TPR- und FPR-Werten, die sich in Boxen niederschlägt, die vollständig oder fast vollständig von 0 bis 100 % reichen, ist dieser Ansatz dennoch kein guter Kandidat für die beabsichtigten Prognosen.

Bei Delikttyp 4 kann also festgestellt werden, dass der BLR-Ansatz mit Einbeziehung der Nexiga-Daten die besten Prognosen liefert, und zwar am besten, indem separate Prognosemodelle für Hannover und für Braunschweig trainiert werden.

²¹ Da hier anders als bei TB 2.1.BS neben dem TPR-Median auch die TPR-Boxoberkante auf der Null-Linie liegt, ist der PPV von 0,0 % hier in gewissem Sinne „noch schlimmer“ als der gleiche Wert dort.

4.1.6 Fazit zu den tatbezogenen Prognosen

Als Zusammenfassung für alle vier Delikttypen kann festgestellt werden, dass der BLR-Ansatz mit Einbeziehung der Nexiga-Daten grundsätzlich ein aussichtsreicher Prognoseansatz ist. Seine PPV-Ergebnisse für die Gesamtmodelle steigen an in der folgenden Reihenfolge der Delikttypen: „WED / TWE“, „Diebstahl von KFZ“, „Schwerer Diebstahl aus Nichtwohnungen (ohne Ladendiebstahl)“ und „Diebstahl in / aus KFZ“. Die Prognosequalität der Gesamtmodelle ist in der Braunschweig-Betrachtung oft sichtbar bis deutlich schlechter als in der Hannover- und in der Gesamtbetrachtung. Vereinzelt lässt sich die Prognosequalität durch die Einschränkung der Trainingsdaten auf die jeweilige Stadt verbessern (Einzelstadt-Modelle). Ohne die Nexiga-Daten liefert der BLR-Ansatz meist keine besseren Ergebnisse als der aktuelle PreMAP-Ansatz in der hier untersuchten Form. Schließlich bringt die aufgrund eines Zufalls erprobte Standardisierung der rein polizeilichen Input-Daten für den aktuellen PreMAP-Ansatz eine leichte Verbesserung der Prognosequalität bei den Delikttypen „WED / TWE“ und „Diebstahl von KFZ“, aber nicht bei den beiden anderen untersuchten Delikttypen.

4.2 Raumbezogene Prognoseansätze: Evaluation der Prognosequalität

4.2.1 Erläuterungen zu den Tabellen und Abbildungen in diesem Abschnitt

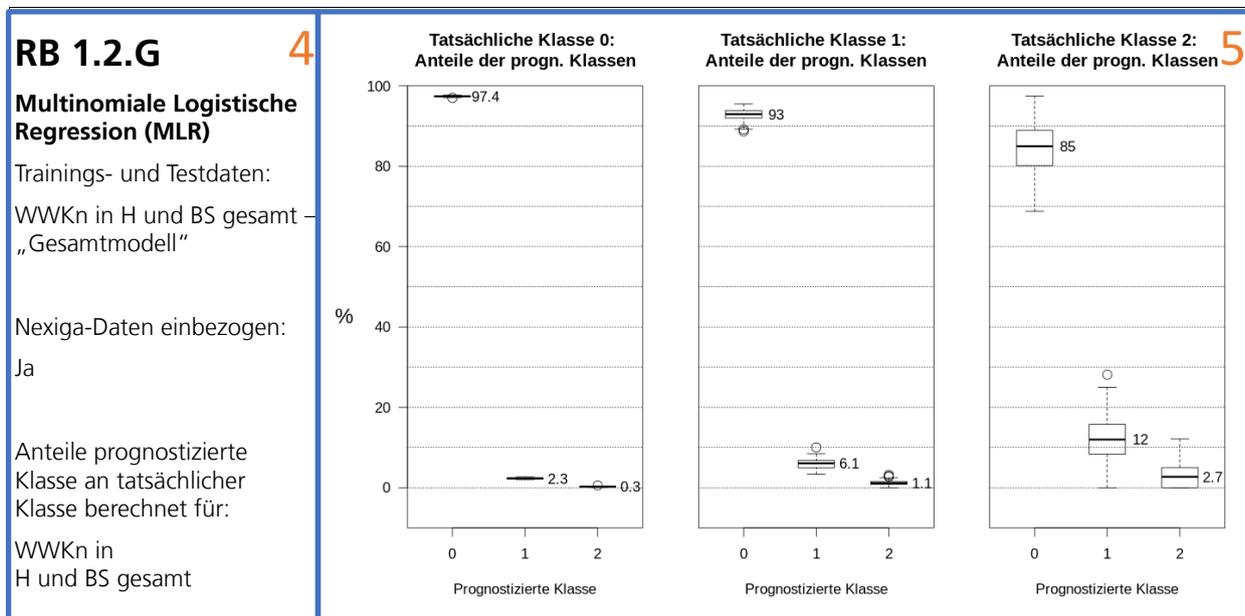
Tabelle RB.1: Kennzahlen zum Delikttyp 1 (Datenlage und Prognosequalität)¹

Gebiet	Anzahl Wohnquartiere	Wohnquartier-Woche-Kombinationen (WWK _n)	darunter		
			Klasse 0 (0 Delikte)	Klasse 1 (1 Delikt)	Klasse 2 (>1 Delikte)
Braunschweig, Stadt	289	45.084	44.290	728	66
Hannover, Landeshauptstadt	671	104.676	101.365	2.992	319
Gesamt	960	149.760	145.655	3.720	385
	PPV (Anteil der zutreffenden an allen Nicht-0-Prognosen, berechnet über alle drei tatsächlichen Klassen zusammen) bei Prognoseansatz ... (in %, Medianbetrachtung)				
	RB 1.0. ...	RB 1.1. ...	RB 1.2. ...	RB 1.3. ...	RB 1.4. ...
... .BS	5,2		3,8		3,6
... .H	8,5		8,5	7,9	
... .G	7,9	8,1	7,9		

- Die raumbezogenen Prognoseansätze werden durch das Kürzel „RB“ gekennzeichnet. Sie werden nach Delikttypen gegliedert (z. B. Delikttyp 1: „WED / TWE“). Jeder Abschnitt zu den einzelnen Delikttypen ist mit einer Übersicht zu relevanten Kennzahlen und Größen versehen.
- In der zweiteiligen Tabelle wird im oberen Teil für die Städte Braunschweig und Hannover sowie als Summe der beiden dargestellt, wie viele Wohnquartiere vorhanden sind und wie viele insgesamt zu prognostizierende Einheiten (Wohnquartier-Woche-Kombinationen, WWK_n) sich daraus ergeben. Da für die Berechnungen Daten eines Dreijahreszeitraums von 2017 bis 2019 genutzt wurden, ergibt sich die Zahl der WWK_n aus folgender Multiplikation: 3 Jahre × 52 Wochen × Anzahl der Wohngebiete). Ferner ist angegeben, wie sich die WWK_n in den drei Jahren auf die drei Delikthäufigkeitsklassen verteilt haben. Im vorliegenden Fall war es so, dass insgesamt 145.655mal in einem bestimmten Wohnquartier in einer bestimmten Woche gar keine Tat passierte (Klasse 0). 3.720mal wurde eine Tat verzeichnet (Klasse 1) und 385mal mehr als eine (Klasse 2). Die WWK-Anzahlen in jeder Klasse sind jeweils für die beiden Städte Hannover und Braunschweig sowie für beide insgesamt angegeben.
- Der untere Block der Tabelle gibt für verschiedene Prognoseansätze als Qualitätskennzahl den Anteil der zutreffenden an allen Alarmen wieder (PPV in einem weiteren Sinn als oben bei den tatbezogenen Ansätzen, vgl. Abschnitt 3.3.2 und die Erläuterungen unten zu Bildabschnitt 5). Beispielsweise findet sich der PPV des Ansatzes RB 1.2, der sich in der Einzelstadt-Auswertung für Braunschweig ergibt, in der Spalte „RB 1.2. ...“ und der Zeile „... .BS“ und beträgt 3,8 %.²² Die genauere Beschreibung der Ansätze ist den zugehörigen Abbildungen

²² Die zweite Ziffer beim Ansatz RB.1.2.BS bezeichnet den jeweiligen Prognoseansatz, und die Buchstabenkombination gibt an, ob die Kennzahlen für Hannover, Braunschweig oder beide zusammen berechnet wurden.

zu entnehmen. Aus folgenden Gründen ist der untere Tabellenteil ggf. nicht vollständig gefüllt: Für einige Ansätze wurde auf eine Einzelstadt-Auswertung verzichtet, und bei Prognoseansätzen, die nur die Daten einer der Städte einbeziehen (Hannover- bzw. Braunschweig-Modelle), gibt es nur PPV-Angaben für diese Stadt.



4) Zu jedem PPV-Eintrag im unter Punkt 3 erläuterten Tabellenteil gibt es einen Abbildungsblock, der den jeweiligen Ansatz genau charakterisiert (4) und die Ergebnisse des Ansatzes graphisch darstellt (5). Im linken Teil des Blocks (4) findet sich eine Charakterisierung des Ansatzes und die Angabe, ob sich die in den Abbildungen dargestellten Anteile auf Hannover, auf Braunschweig oder auf beide zusammen beziehen:

- a. Trainings- und Testdaten: Hier ist angegeben, ob der Prognoseansatz mit Daten für Hannover und Braunschweig gemeinsam trainiert wurde („Gesamtmodell“, auch im Beispiel) oder nur mit solchen aus einer der beiden Städte („Hannover-Modell“ bzw. „Braunschweig-Modell“).
- b. Nexiga-Daten einbezogen: Hier ist angegeben, ob zusätzlich zu den NIVADIS-Daten die zu erprobenden sozioökonomischen Daten in den Ansatz eingeflossen sind. Bei „Nein“ wurden ausschließlich Daten aus NIVADIS verwendet. Welche Daten genau in welchem Ansatz verwendet wurden, geht aus Anhang A.2 hervor. Die Kombination der Angaben unter a und b spiegelt sich in der zweiten Ziffer der Ansatzbezeichnung wieder, hier der „2“ in „RB 1.2.G“.
- c. Anteil prognostizierte Klasse an tatsächlicher Klasse berechnet für: Hier ist angegeben, ob die im Abbildungsblock dargestellten Anteile in den Testmengen für Hannover, für Braunschweig oder für beide zusammen berechnet wurden. Diese Information spiegelt sich im Buchstabencode am Ende der Ansatzbezeichnung wieder, hier dem „G“ für „Gesamt“ in „RB 1.2.G“.²³

²³ Bei Einzelstadt-Auswertungen zu Gesamtmodellen, also beispielsweise bei RB 1.2.H, sind die betrachteten Prognoseergebnisse eine Teilmenge der gesamten Prognoseergebnisse (im Beispiel aus RB 1.2.G). Die

- 5) Im rechten Teil des Abbildungsblocks (5) wird in einem Kasten für jede der drei tatsächlichen Klassen (von links nach rechts: Klasse 0 = keine Tat, Klasse 1 = eine Tat, Klasse 2 = mehr als eine Tat) durch drei Boxplots dargestellt, mit welchem Anteil die WWK_n aus dieser Klasse durch den Prognoseansatz den drei Klassen zugeordnet wurden.²⁴ Zum grundsätzlichen Informationsgehalt der Boxplots (Median, Boxober- und -unterkante, Whiskers und Ausreißer) sei auf die Erläuterungen oben in Unterabschnitt 4.1.1 verwiesen.

Im Beispiel gab es gemäß der oben erläuterten Tabelle insgesamt 149.760 Wohnquartier-Woche-Kombinationen (WWK_n) für Hannover und Braunschweig gemeinsam, und bei 145.655 davon war kein WED / TWE zu verzeichnen (d. h., sie gehörten zur tatsächlichen Klasse 0, linker Kasten). Der Prognoseansatz hätte nun im Mittel 97,4 % dieser WWK_n korrekt der Klasse 0 zugeordnet (Medianbeschriftung beim linken Boxplot im linken Kasten), für 2,4 % dieser WWK_n hätte er fälschlich die Klasse 1 und für 0,2 % fälschlich die Klasse 2 prognostiziert (mittlerer und rechter Boxplot im linken Kasten). In 3.720 WWK_n hatte es in der Vergangenheit exakt einen Fall gegeben (tatsächliche Klasse 1, mittlerer Kasten); hier lag der Ansatz im Mittel bei 5,9 % der WWK_n richtig und sagte die Klasse 1 voraus (mittlerer Boxplot). Für 93,0 % dieser WWK_n prognostizierte der Ansatz fälschlich die Klasse 0 (linker Boxplot) und für 1,2 % fälschlich die Klasse 2 (rechter Boxplot). Die Werte des rechten Kastens (für die 382 WWK_n mit mehr als einem WED / TWE – tatsächliche Klasse 2) sind analog zu lesen.

Eine Interpretation der Werte ist auch in dem weiteren Sinne möglich, dass die prognostizierten Klassen 1 und 2 als „Nicht-0-Prognosen“ bzw. als „Alarime im weiteren Sinne“ zusammengefasst werden: In dieser Betrachtung würde der mittlere Kasten im Beispiel zeigen, dass für 7,2 % (= 6,1 % + 1,1 %) der WWK_n, bei denen genau eine Tat vorkam (tatsächliche Klasse 1), auch eine oder mehrere Taten prognostiziert wurden (prognostizierte Klasse 1 oder 2). Der entsprechende Wert im rechten Boxplot beträgt 14,7 % und gibt an, für welchen Anteil der WWK_n mit mindestens zwei Taten (tatsächliche Klasse 2) mindestens eine Tat prognostiziert wurde. In beiden Fällen kann man die Zusammenfassung der prognostizierten Klassen 1 und 2 als „zutreffende Alarime im weiteren Sinne“ ansehen.

Umgekehrt treten bei 2,6 % der WWK_n ohne aufgetretenes Delikt (tatsächliche Klasse 0, linker Kasten) fälschlich die prognostizierten Klassen 1 oder 2 auf, hier handelt es sich um „falsche Alarime im weiteren Sinne“. Auf diese zutreffenden und falschen Alarime im weiteren Sinne beziehen sich die PPV-Angaben in der oben erläuterten Tabelle.

Anteilberechnung wird lediglich auf die jeweilige Einzelstadt eingeschränkt. Beispielsweise könnte in einer der zufällig gezogenen 10-%-Testmengen in RB 1.2.G von 360 WWK_n (davon 300 in Hannover), die tatsächlich zur Klasse 1 gehören, für 22 auch die Klasse 1 prognostiziert werden, also ein Anteil von 6,1 %. Wenn von diesen 22 WWK_n 15 in Hannover liegen, ergibt sich in der Hannover-Betrachtung für dieselbe Testmenge ein Anteil von 10 prognostizierten an 300 tatsächlichen, also von 5,0 %, in RB 1.2.H.

²⁴ Aufgrund von Ungenauigkeiten bei der Medianbildung und bei der Rundung auf eine Nachkommastelle addieren sich die angegebenen Anteils-Mediane nicht immer zu 100 %.

4.2.2 Delikttyp 1 „WED / TWE“ (Deliktschlüssel 435 und 436)

In diesem und den folgenden Unterabschnitten zu den weiteren Delikttypen werden die wesentlichen Beobachtungen aus den entsprechenden Tabellen und Abbildungen beschrieben.

Tabelle RB.1: Kennzahlen zum Delikttyp 1 (Datenlage und Prognosequalität)

Gebiet	Anzahl Wohnquartiere	Wohnquartier- Woche-Kombi- nationen (WWK _n)	darunter		
			Klasse 0 (0 Delikte)	Klasse 1 (1 Delikt)	Klasse 2 (>1 Delikte)
Braunschweig, Stadt	289	45.084	44.290	728	66
Hannover, Landeshauptstadt	671	104.676	101.365	2.992	319
Gesamt	960	149.760	145.655	3.720	385
	PPV (Anteil der zutreffenden an allen Nicht-0-Prognosen, berechnet über alle drei tatsächlichen Klassen zusammen) bei Prognoseansatz ... (in %, Medianbetrachtung)				
	RB 1.0. ...	RB 1.1. ...	RB 1.2. ...	RB 1.3. ...	RB 1.4. ...
... .BS	5,2		3,8		3,6
... .H	8,5		8,5	7,9	
... .G	7,9	8,1	7,9		

Abbildung RB.1 zeigt die Prognosequalität von zwei Varianten des MLR-Ansatzes (RB 1.1 ohne, RB 1.2 mit Einbeziehung der Nexiga-Daten) für den Delikttyp 1 „WED / TWE“ im Vergleich zu der des Benchmark-Ansatzes RB 1.0, und zwar ausgewertet für Hannover und Braunschweig zusammen. Dazu werden je Ansatz in drei Kästen für jede tatsächliche WWK-Klasse (0, 1, >1 Delikte) drei Boxplots zu den Anteilen dargestellt, mit denen der Prognoseansatz diese WWK_n einer der drei Klassen zugeordnet hat.²⁵ Beispielsweise hat der Ansatz RB 1.2.G von denjenigen WWK_n, die tatsächlich zu Klasse 1 gehörten, 93,0 % der Klasse 0, außerdem 6,1 % der Klasse 1, also der korrekten Klasse, und 1,1 % der Klasse 2 zugeordnet (Mediane der drei Boxplots im mittleren Kasten in der unteren Zeile).

Es zeigt sich, dass der „daten-ignorierende“ Benchmark-Ansatz in den tatsächlichen Klassen 0 und 1 sehr ähnlich abschneidet wie beide Varianten des methodisch anspruchsvolleren MLR-Ansatzes. Dass der Ansatz RB 1.1 in der tatsächlichen Klasse 0 (linker Kasten in jeder Zeile) ein Quäntchen weniger (-0,1 Prozentpunkte) falsche Klasse-1-Prognosen macht, darf nicht überbewertet werden, zumal der Anteils-Median für die zutreffenden Klasse-0-Prognosen dann doch für alle drei Ansätze 97,4 % beträgt. Zwar würden 0,1 Prozentpunkte der tatsächlichen Klasse 0 bei diesem Delikttyp etwa 146 WWK_n in beiden Städten zusammen entsprechen, aber gleichzeitig liegen 0,1 Prozentpunkte auch immer im Bereich der Rundungsungenauigkeit.

In der tatsächlichen Klasse 1 (mittlerer Kasten in jeder Zeile) liefern immerhin beide MLR-Ansätze gegenüber dem Benchmark-Ansatz einen Anstieg beim Anteil der korrekten Klasse-1-Prognosen um 0,2 Prozentpunkte bei gleichzeitiger Anteils-Verringerung um 0,1 Prozentpunkte bei den

²⁵ Beim Benchmark-Ansatz „degeneriert“ der Boxplot zu einer waagerechten Linie, weil hier nicht 100 verschiedene Modelle mit 100 zufällig gezogenen Trainingsmengen zugrunde liegen, die dann auch 100 verschiedene Werte für jeden Prognoseanteil in jeder Klasse aufweisen. Stattdessen wurde der Benchmark-Ansatz einmal mit allen vorhandenen WWK_n ausgeführt und liefert damit genau einen Wert für jeden Prognoseanteil in jeder Klasse. Auf eine auch hier denkbare Randomisierung durch Einschränkung des Benchmark-Ansatzes auf 100 zufällig gezogene WWK_n-Teilmengen des Datenbestandes wurde in unserer Untersuchung verzichtet.

Klasse-2-Prognosen. Auch hier ist wieder Vorsicht bei der Interpretation geboten, zumal gleichzeitig in beiden Ansätzen der Anteils-Median der falschen Klasse-0-Prognosen ebenfalls um 0,1 Prozentpunkte höher liegt. Zudem entsprechen 0,2 Prozentpunkte der tatsächlichen Klasse 1 bei diesem Delikttyp nur etwa 7 WWK_n in beiden Städten zusammen.

In der tatsächlichen Klasse 2 (rechter Kasten in jeder Zeile) schneidet der Benchmark-Ansatz sogar sichtbar besser ab als die beiden MLR-Ansätze, und zwar sowohl bei den korrekten Klasse-2-Prognosen (3,4 % gegenüber 2,4 % und 2,7 %) als auch bei den weiteren Nicht-0-Prognosen als „zutreffenden Alarmen im weiteren Sinn“, also den Klasse-1-Prognosen (12,8 % gegenüber 11,3 % und 12,0 %). Hier zeigt sich zum ersten Mal ein positiver Effekt der Einbeziehung der Nexiga-Daten in den MLR-Ansatz, denn bei RB 1.2.G liegt der Anteil der korrekten Klasse-2-Prognosen um 0,3 Prozentpunkte (\cong 1 WWK) und der der Klasse-1-Prognosen um 0,7 Prozentpunkte (\cong 3 WWK_n) höher²⁶ als bei RB 1.1.G. Beide Werte bleiben aber noch sichtbar unter denen des Benchmark-Ansatzes.

Zusammen bewirken beim Ansatz RB 1.1.G der um -0,1 Prozentpunkte bessere Anteil der falschen Nicht-0-Prognosen (definitionsgemäß in Klasse 0) und der um +0,1 Prozentpunkte bessere Anteil der zutreffenden Nicht-0-Prognosen in Klasse 1 (trotz der anteilmäßig, aber nicht anzahlmäßig sichtbaren Verschlechterung in Klasse 2) einen um 0,2 Prozentpunkte höheren PPV als beim Benchmark-Ansatz RB 1.0.G (8,1 % im Vergleich zu 7,9 %, vgl. Tabelle RB.1). In Verbindung mit den obigen Überlegungen, die für eine höhere Prognosequalität des Benchmark-Ansatzes sprechen, zeigt sich also hier, dass ein höherer PPV wie dieser nicht zwingend auf eine eindeutig bessere Prognosequalität schließen lässt.

Es kann festgehalten werden, dass die Prognosequalität des „daten-ignorierenden“ Benchmark-Ansatzes in der Gesamtbetrachtung für Hannover und Braunschweig zusammen durch einen MLR-Ansatz nicht übertroffen werden kann, auch nicht durch die Einbeziehung der sozioökonomischen Nexiga-Daten. Überraschenderweise führt bei den MLR-Ansätzen die Berücksichtigung von zusätzlichen Informationen von erwartbarer Relevanz, nämlich der Nexiga-Daten, nicht durchgängig zu besseren Ergebnissen. Immerhin schneidet der MLR-Ansatz mit den zusätzlichen sozioökonomischen Daten bei den zutreffenden Klasse-2-Prognosen besser ab als der ohne diese Daten.

²⁶ Auf den ersten Blick ist ein höherer Anteil von Klasse-1-Prognosen, also Fehlprognosen, für Wohnquartiere, die in der betreffenden Woche zu Klasse 2 gehören, zwar keine gute Nachricht. Unter dem Blickwinkel der polizeilichen Maßnahmen, die aus einer solchen Prognose resultieren, ist es aber günstiger, wenn ein tatsächliches Klasse-2-Wohnquartier (mit mehr als einem Delikt) zumindest wie ein prognostiziertes Klasse-1-Wohnquartier (mit einem erwarteten Delikt) behandelt wird und nicht wie ein Klasse-0-Wohnquartier (ganz ohne Delikt). Diese Überlegung korrespondiert wieder mit dem Konzept der Nicht-0-Prognosen als „zutreffende Alarme im weiteren Sinn“. Diese ambivalente Bewertung der falschen Klasse-1-Prognosen in der tatsächlichen Klasse 2 spielt auch weiter unten wieder eine Rolle. In gewisser Hinsicht sind die falschen Klasse-2-Prognosen in der tatsächlichen Klasse 1 schlimmer, weil hier eine Klasse-1-WWK die – plausiblerweise aufwendigere – Klasse-2-Behandlung erhalten könnte. Bei der PPV-Berechnung wird dieser Unterschied ausgeblendet.

RB 1.0.G

Benchmark-Ansatz (Klasse wie in Vorwoche)

Trainings- und Testdaten:

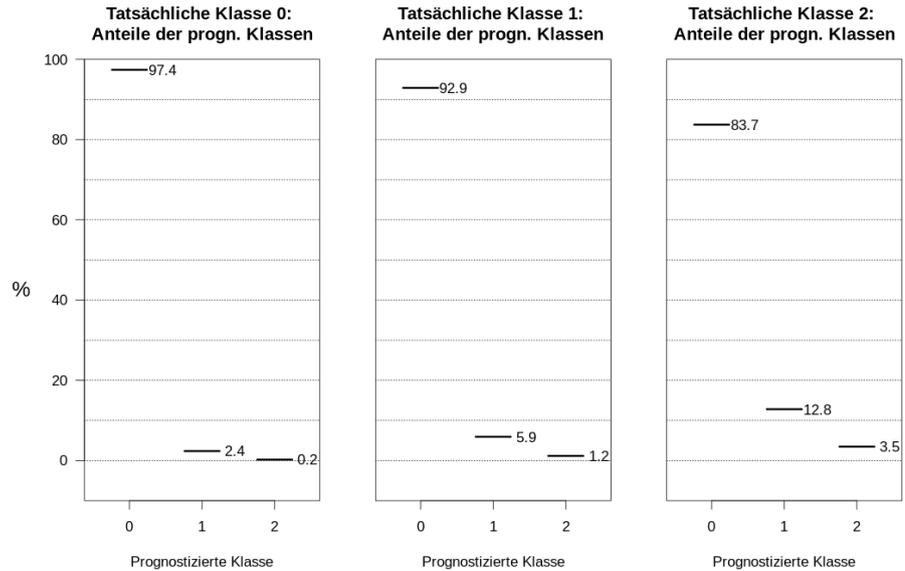
Nicht erforderlich

Nexiga-Daten einbezogen:

Nicht erforderlich

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in
H und BS gesamt



RB 1.1.G

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

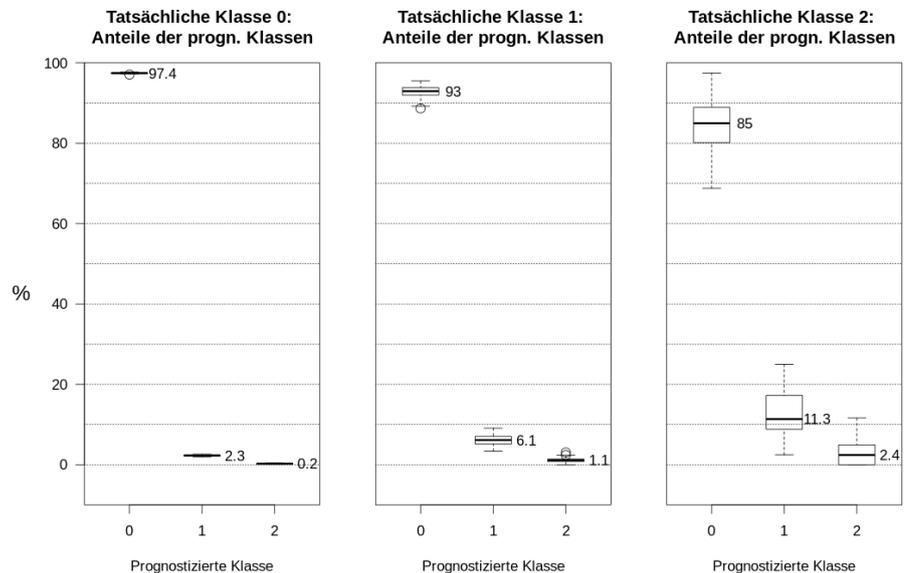
WWKn in H und BS gesamt –
„Gesamtmodell“

Nexiga-Daten einbezogen:

Nein

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in
H und BS gesamt



RB 1.2.G

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

WWKn in H und BS gesamt –
„Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in
H und BS gesamt

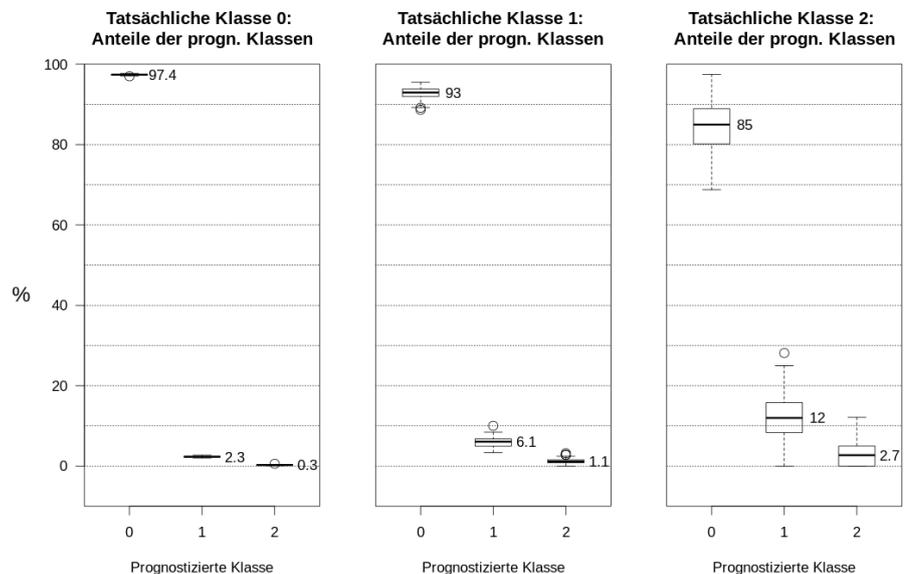


Abbildung RB.1: Prognosequalität verschiedener raumbezogener Ansätze für Delikttyp 1 für H und BS zusammen

Abbildung RB.2 zeigt wieder die Prognosequalität von zwei Varianten des MLR-Ansatzes für den Delikttyp 1 im Vergleich zum Benchmark-Ansatz. Diesmal beziehen sie beide die Nexiga-Daten ein und werden beide für Hannover alleine ausgewertet, weshalb hier auch der Benchmark-Ansatz für Hannover allein ausgewertet wurde (RB 1.0.H). Der eine MLR-Ansatz wird mit Daten für beide Städte trainiert (RB 1.2.H), der andere nur mit Daten für Hannover („Hannover-Modell“, RB 1.3.H). Das bedeutet, die mittlere Zeile hier korrespondiert mit der unteren in Abbildung RB.1 und die untere ist neu dazugekommen.

Es zeigt sich ein ähnliches Bild wie beim obigen Vergleich der Ergebnisse für beide Städte zusammen: Bei den Anteilen der prognostizierten Klassen treten in der tatsächlichen Klasse 0 Unterschiede zwischen den drei Ansätzen von maximal $\pm 0,1$ Prozentpunkten auf (also innerhalb der Rundungsungenauigkeit und daher nicht aussagekräftig), in der tatsächlichen Klasse 1 Unterschiede von maximal $\pm 0,3$ Prozentpunkten. Dabei sind beim dritten Ansatz die zutreffenden Klasse-1-Prognosen um 0,3 Prozentpunkte schlechter und beim zweiten Ansatz die Klasse-2-Prognosen um (nicht unbedingt aussagekräftige, s. o.) 0,1 Prozentpunkte höher als bei den anderen Ansätzen. In der tatsächlichen Klasse 2 liefert der Benchmark-Ansatz mit sichtbarem Abstand die besten Klasse-2-Prognosen und nur einen minimal niedrigeren (also im Rahmen der oben beschriebenen Ambivalenz etwas schlechteren) Anteil an Klasse-1-Prognosen als der bessere der beiden MLR-Ansätze.

Der MLR-Ansatz mit reinen Hannover-Trainingsdaten (RB 1.3.H) schneidet bei den Prognosen in der tatsächlichen Klasse 2 besonders schlecht ab (niedrigster Anteil von Klasse-1-Prognosen, korrekte Klasse-2-Prognosen genau so niedrig wie bei RB 1.2.H), und auch in den beiden anderen tatsächlichen Klassen liefert er immer minimal schlechtere Prognosen als die beiden anderen dargestellten Ansätze. Die Prognosequalität des MLR-Ansatzes mit Nexiga-Daten verschlechtert sich also durch die Einschränkung auf Trainingsdaten aus Hannover, d. h., dass für diesen Delikttyp die Daten aus Braunschweig einen hilfreichen Beitrag für MLR-Prognosen in Hannover leisten. Das schlägt sich auch in einem sichtbar niedrigeren PPV des Ansatzes mit reinem Hannover-Modell nieder (7,9 % gegenüber je 8,5 % bei den beiden anderen Ansätzen, vgl. Tabelle RB.1).

Alle drei Einzelstadtauswertungen für Hannover zeigen in allen drei tatsächlichen Klassen etwas günstigere Nicht-Klasse-0-Prognoseanteile (in der Medianbetrachtung) als die drei oben betrachteten Auswertungen für beide Städte gemeinsam, die beiden betrachteten MLR-Ansätze außerdem auch einen höheren Anteil von korrekten Klasse-2-Prognosen (definitionsgemäß an der tatsächlichen Klasse 2). Das PPV-Niveau liegt etwas höher als bei den Auswertungen für beide Städte gemeinsam.

Abbildung RB.3 fasst die entsprechenden Auswertungen für Braunschweig zusammen: Benchmark-Ansatz sowie zwei MLR-Ansätze mit einbezogenen Nexiga-Daten, alle ausgewertet für Braunschweig alleine, wobei einer der MLR-Ansätze mit Daten aus beiden Städten (RB 1.2.BS) und der andere mit reinen Braunschweig-Daten trainiert wurde (RB 1.4.BS).

Die Ergebnisdarstellung ähnelt stark der obigen Einzelstadt-Darstellung für Hannover: Bei den Anteilen der prognostizierten Klassen treten in der tatsächlichen Klasse 0 Unterschiede zwischen den drei Ansätzen von maximal $\pm 0,1$ Prozentpunkten, also innerhalb der Rundungsungenauigkeit auf, in der tatsächlichen Klasse 1 Unterschiede von immerhin schon maximal $\pm 0,8$ Prozentpunkten, und zwar aufgrund des Ergebnisses, dass bei beiden MLR-Ansätzen (in der Medianbetrachtung) alle Nicht-0-Prognosen korrekte Klasse-1-Prognosen sind, dabei aber auch geringere Anteile an der tatsächlichen Klasse 1 haben als die Nicht-0-Prognosen des Benchmark-Ansatzes. In der tatsächlichen Klasse 2 liegen bei den MLR-Ansätzen beide Boxen für die Anteile der Nicht-0-Prognosen komplett auf der Nulllinie, also treten für weniger als ein Viertel der Trainingsmen-

RB 1.0.H

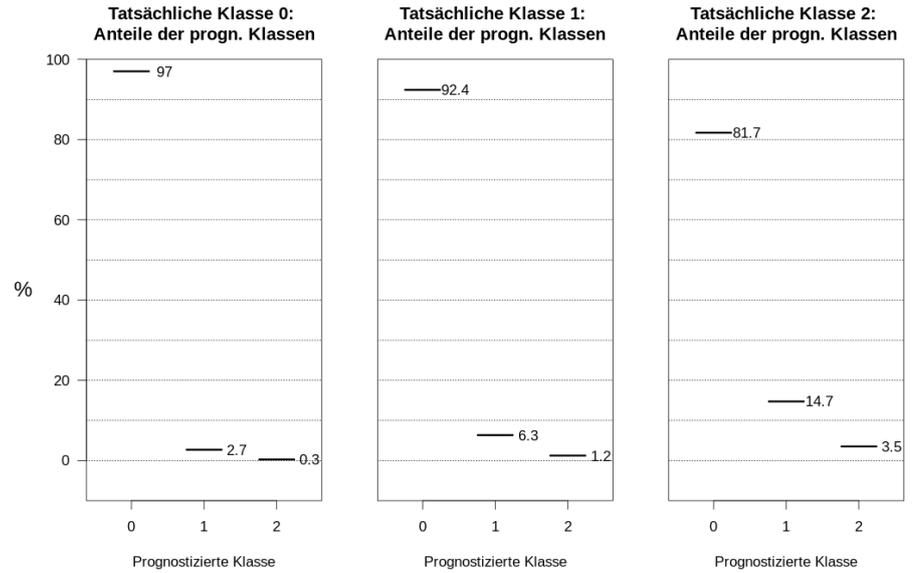
Benchmark-Ansatz (Klasse wie in Vorwoche)

Trainings- und Testdaten:
Nicht erforderlich

Nexiga-Daten einbezogen:
Nicht erforderlich

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in H



RB 1.2.H

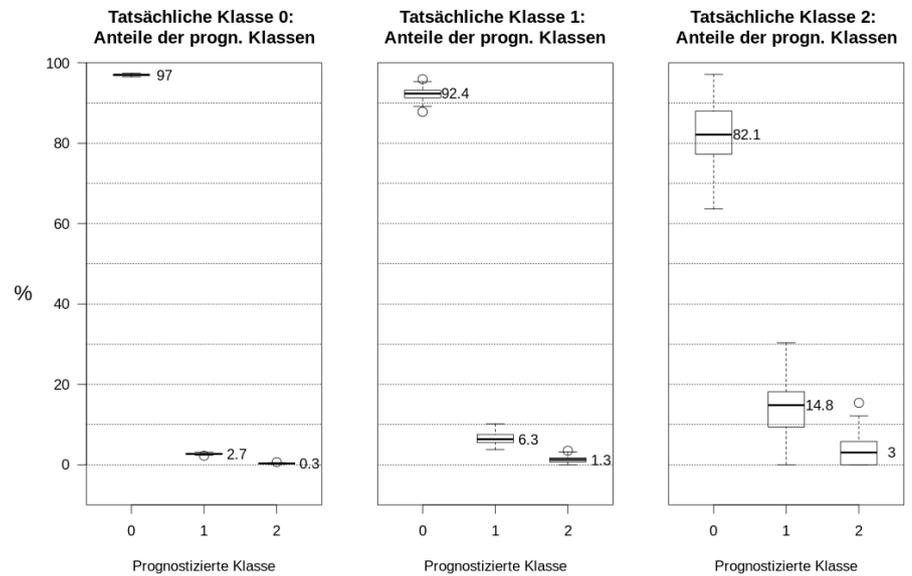
Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:
WWKn in H und BS gesamt –
„Gesamtmodell“

Nexiga-Daten einbezogen:
Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in H



RB 1.3.H

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:
WWKn in H –
„Hannover-Modell“

Nexiga-Daten einbezogen:
Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in H

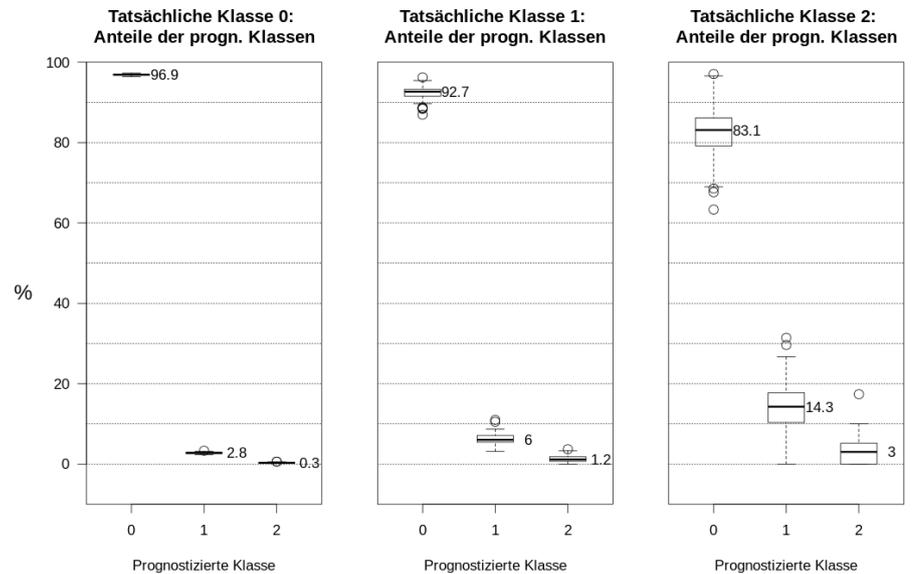


Abbildung RB.2: Prognosequalität verschiedener raumbezogener Ansätze für Delikttyp 1 für H alleine

gen überhaupt Klasse-1-Prognosen bzw. Klasse-2-Prognosen auf.^{27 28} Das ist nur zum Teil ein Beleg für schlechte Prognosequalität der Ansätze. Zumindest bei den korrekten Klasse-2-Prognosen dürfte es auch eine Rolle spielen, dass bei nur 66 WWKn in der tatsächlichen Klasse 2 in mehreren der 100 Testmengen keine oder nur sehr wenige Klasse-2-WWKn enthalten waren und damit auch keine oder kaum korrekte Klasse-2-Prognosen möglich waren.

In der tatsächlichen Klasse 2 fällt außerdem auf, dass beim Benchmark-Ansatz erstmals die Anteile der Klasse-1- und der Klasse-2-Prognosen gleich hoch (aber eigentlich eher gleich niedrig) sind. In Braunschweig sind also immerhin die Hälfte der Nicht-0-Prognosen dieses Ansatzes in der tatsächlichen Klasse 2 korrekte Klasse-2-Prognosen.

Auch in der Einzelstadtbetrachtung der drei Prognoseansätze für Braunschweig liefert der Benchmark-Ansatz die insgesamt beste Prognosequalität, was auch sein PPV von 5,2 % anzeigt, der hier sichtbar höher ist als die PPV-Ergebnisse für die beiden MLR-Ansätze (3,8 % beim Gesamtmodell, 3,6 % beim Braunschweig-Modell, vgl. Tabelle RB.1). Wie bei den Einzelstadtauswertungen für Hannover schneidet auch hier der MLR-Ansatz mit dem Einzelstadt-Modell etwas schlechter ab als der mit dem Gesamtmodell, also leisten die Daten aus der anderen Stadt wieder einen nützlichen Beitrag zu den Prognosen in dieser Stadt.

Im Vergleich zu den Auswertungen für beide Städte zusammen (Abbildung RB.1) zeigt sich für alle Einzelstadtauswertungen für Braunschweig folgendes Bild: in der Betrachtung der Anteils-Mediane etwas weniger falsche Nicht-0-Prognosen (in der tatsächlichen Klasse 0), sichtbar weniger zutreffende Nicht-0-Prognosen in der tatsächlichen Klasse 1 und deutlich weniger zutreffende Nicht-0-Prognosen in der tatsächlichen Klasse 2. Das PPV-Niveau liegt sichtbar niedriger.

Zusammenfassend kann für die raumbezogenen Prognosen zum Delikttyp 1 „WED / TWE“ festgehalten werden, dass hier die Prognosequalität des einfachen, „daten-ignorierenden“ Benchmark-Ansatzes – mit der direkten Übernahme der Delikthäufigkeits-Klasse jedes Wohnquartiers in einer Woche als Prognose für die Folgewoche – durch die untersuchten MLR-Ansätze *nicht* übertroffen wird, also weder mit noch ohne Einbeziehung der sozioökonomischen Daten zusätzlich zu den polizeilichen Daten in ein Gesamtmodell und auch nicht durch Trainieren von reinen Hannover- bzw. Braunschweig-Modellen mit Einbeziehung der sozioökonomischen Daten.

²⁷ Das müssen nicht dieselben Trainingsmengen sein. Beispielsweise kann ein MLR-Ansatz für eine Trainingsmenge nur Klasse-0 und korrekte Klasse-2-Prognosen liefern und für eine andere Trainingsmenge nur Klasse-0- und Klasse-1-Prognosen.

²⁸ Bei einer Boxhöhe von 0 wie hier werden aufgrund der Whisker-Definition zwingend alle Werte oberhalb der Box (und unterhalb, hier irrelevant) als „Ausreißer“ dargestellt, d. h., es gibt keine Whiskers.

RB 1.0.BS

Benchmark-Ansatz (Klasse wie in Vorwoche)

Trainings- und Testdaten:

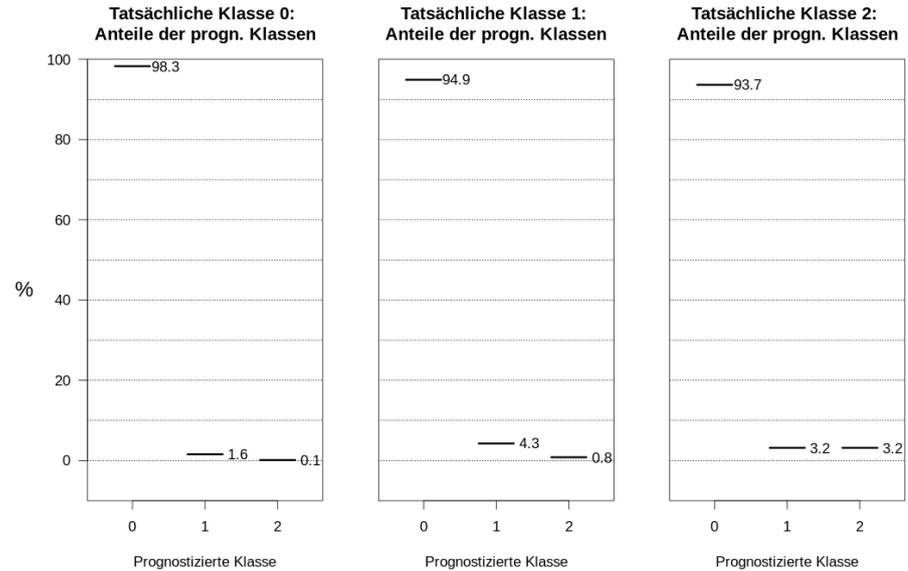
Nicht erforderlich

Nexiga-Daten einbezogen:

Nicht erforderlich

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in BS



RB 1.2.BS

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

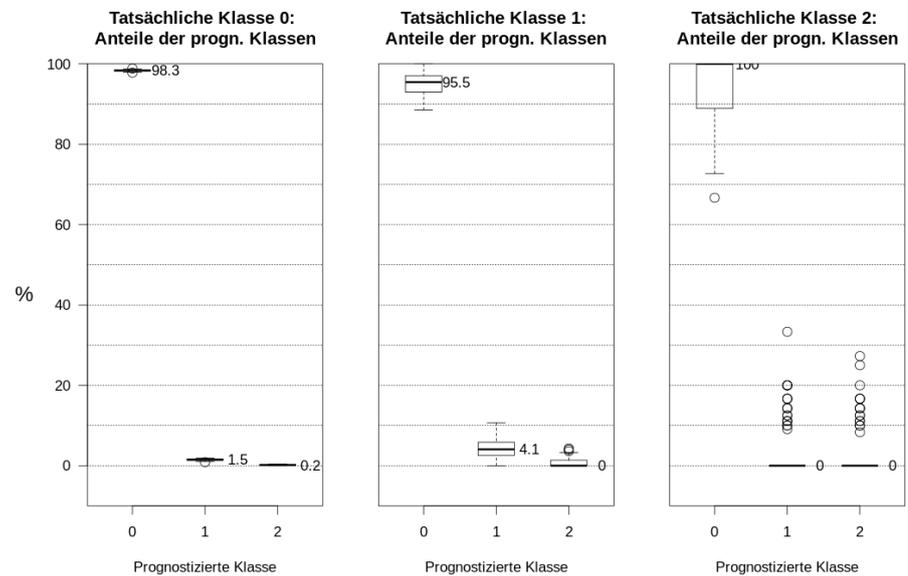
WWKn in H und BS gesamt –
„Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in BS



RB 1.4.BS

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

WWKn in BS – „Braun-
schweig-Modell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in BS

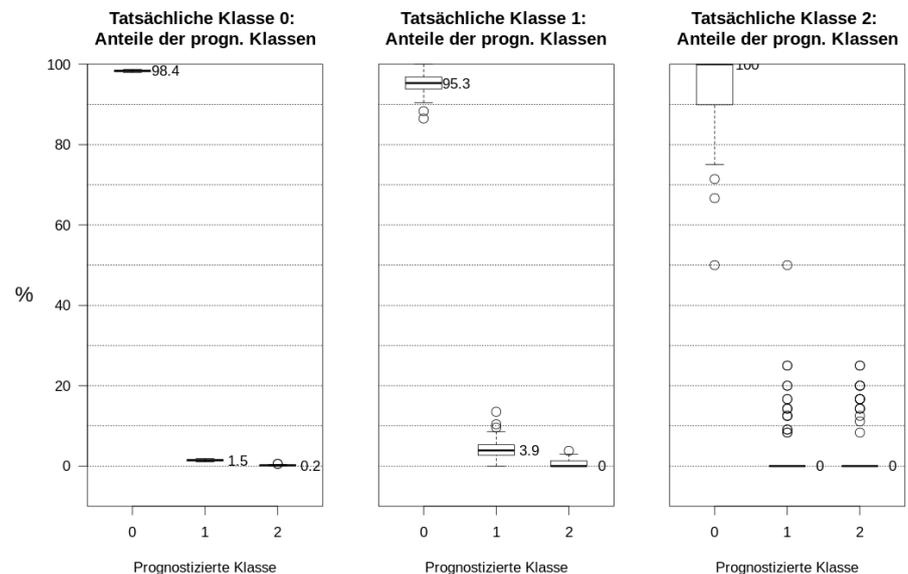


Abbildung RB.3: Prognosequalität verschiedener raumbezogener Ansätze für Delikttyp 1 für BS alleine

4.2.3 Delikttyp 2 „Schwerer Diebstahl aus Nichtwohnungen (ohne Ladendiebstahl)“ (Deliktschlüssel 410, 415, 425)

Tabelle RB.2 zeigt analog zu Tabelle RB.1 die Gegebenheiten in den Daten (Wohnquartiere und WWKns wie dort, tatsächliche Klassenzugehörigkeiten anders) und die PPV-Ergebnisse für verschiedene untersuchte Prognoseansätze, diesmal für den Delikttyp 2 „Schwerer Diebstahl aus Nichtwohnungen (ohne Ladendiebstahl)“. Bei den PPV-Angaben fällt sofort ins Auge, dass sie durchweg sichtbar höher liegen als beim Delikttyp 1 (vgl. Tabelle RB.1). Dabei zeigen sich in den Hannover-Auswertungen (Zeile „... .H“) wieder etwas höhere PPV-Ergebnisse als in den Auswertungen für beide Städte zusammen (Zeile „... .G“). In den Braunschweig-Auswertungen für diesen Delikttyp (Zeile „... .BS“) sind die PPV-Ergebnisse wieder deutlich niedriger als für beide Städte zusammen.

Tabelle RB.2: Kennzahlen zum Delikttyp 2 (Datenlage und Prognosequalität)

Gebiet	Anzahl Wohnquartiere	Wohnquartier- Woche-Kombi- nationen (WWKns)	darunter		
			Klasse 0 (0 Delikte)	Klasse 1 (1 Delikt)	Klasse 2 (>1 Delikte)
Braunschweig, Stadt	289	45.084	44.311	703	70
Hannover, Landeshauptstadt	671	104.676	101.391	2.763	522
Gesamt	960	149.760	145.702	3.466	592
	PPV (Anteil der zutreffenden an allen Nicht-0-Prognosen, berechnet über alle drei tatsächlichen Klassen zusammen) bei Prognoseansatz ... (in %, Medianbetrachtung))				
	RB 2.0. ...	RB 2.1. ...	RB 2.2. ...	RB 2.3. ...	RB 2.4. ...
... .BS	8,2		6,2		7,3
... .H	13,6		13,1	12,0	
... .G	12,8	13,0	11,7		

Beim Blick auf die Anteile der prognostizierten Klassen in beiden Städten zusammen für den Benchmark-Ansatz sowie den MLR-Ansatz ohne die und mit den sozioökonomischen Daten (Abbildung RB.4) fällt Folgendes auf: Die Prognosequalität der drei untersuchten Ansätze in den tatsächlichen Klassen 0 und 1 ist sehr ähnlich, die meisten Unterschiede betragen nur $\pm 0,1$ Prozentpunkte und bewegen sich damit im Rahmen der Rundungsungenauigkeit. Bei optimistischer Interpretation kann man bei den Prognoseanteilen des Ansatzes RB 2.1.G (MLR ohne Nexiga-Daten) gegenüber dem Benchmark-Ansatz in der tatsächlichen Klasse 1 eine aussagekräftige Verschiebung von den Klasse-2-Prognosen (-0,2 Prozentpunkte) zu den korrekten Klasse-1-Prognosen (+0,1 Prozentpunkte) feststellen. Die falschen Nicht-0-Prognosen in Klasse 0 (2,6 %, Summe des Klasse-1- und des Klasse-2-Medians) und die zutreffenden Nicht-0-Prognosen in Klasse 1 (10,1 %) fallen bei RB 2.2.G jeweils schlechter aus als bei den beiden anderen Ansätzen, das heißt, dass in diesen tatsächlichen Klassen auch bei diesem Delikttyp die Einbeziehung der Nexiga-Daten die Prognosequalität des MLR-Ansatzes nicht verbessert.

RB 2.0.G

Benchmark-Ansatz (Klasse wie in Vorwoche)

Trainings- und Testdaten:

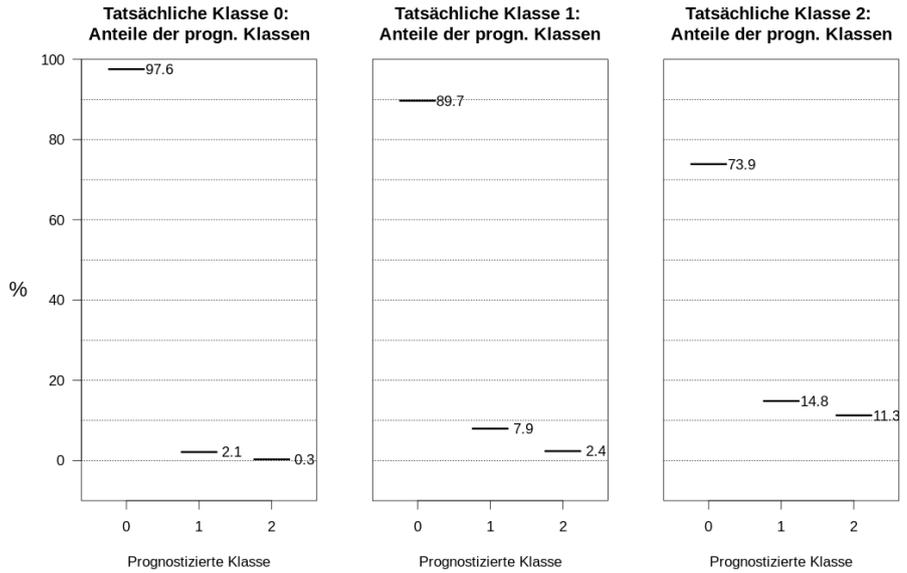
Nicht erforderlich

Nexiga-Daten einbezogen:

Nicht erforderlich

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in
H und BS gesamt



RB 2.1.G

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

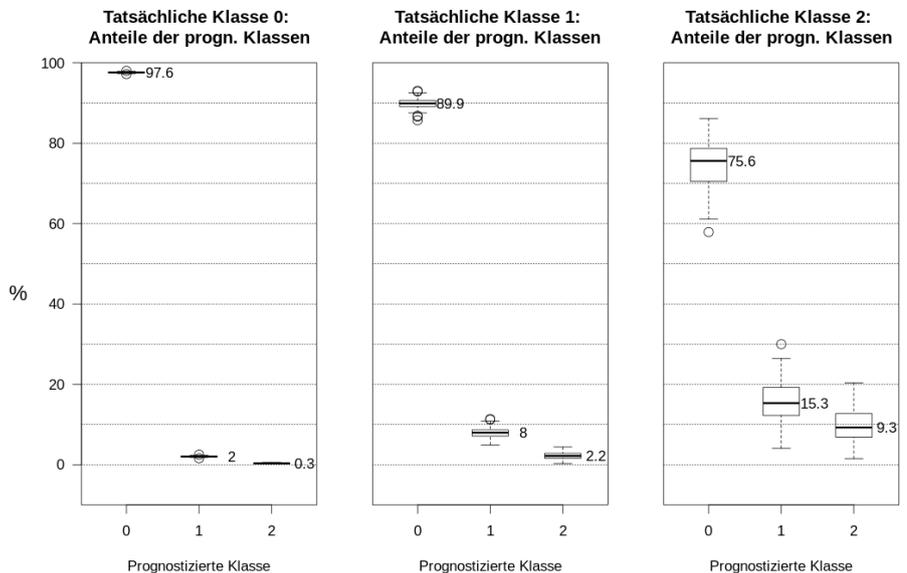
WWKn in H und BS gesamt –
„Gesamtmodell“

Nexiga-Daten einbezogen:

Nein

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in
H und BS gesamt



RB 2.2.G

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

WWKn in H und BS gesamt –
„Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in
H und BS gesamt

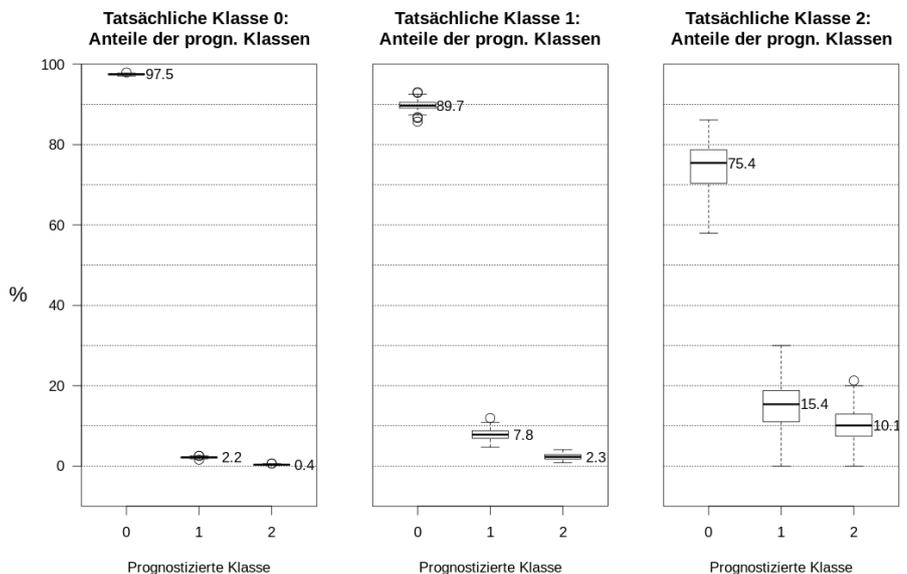


Abbildung RB.5: Prognosequalität verschiedener raumbezogener Ansätze für Delikttyp 2 für H und BS zusammen

In der tatsächlichen Klasse 2 tut sie dies – wie schon beim Delikttyp 1 – aber sehr wohl, denn hier liegt der Anteil der korrekten Klasse-2-Prognosen um ganze 0,8 Prozentpunkte höher als beim MLR-Ansatz ohne Nexiga-Daten (und der Anteil der Klasse-1-Prognosen bleibt im Rahmen der Rundungsungenauigkeit gleich groß). Die Verbesserung wird hier – anders als bei Delikttyp 1 – auch durch den niedrigeren Anteils-Median der Klasse-0-Prognosen bestätigt.²⁹ Der Anteil der zutreffenden Nicht-0-Prognosen in der tatsächlichen Klasse 2 bleibt allerdings auch beim MLR-Ansatz mit den sozioökonomischen Daten hinter dem Benchmark-Ansatz zurück, der insbesondere für die zutreffenden Klasse-2-Prognosen – wie schon beim Delikttyp 1 – noch besser abschneidet.

Zusammenfassend ist also der Benchmark-Ansatz in der Zwei-Städte-Betrachtung den beiden MLR-Ansätzen vorzuziehen.³⁰

Die Abbildung RB.5 stellt analog zu Abbildung RB.2 die Anteile der prognostizierten Klassen in der Einzelstadtauswertung für Hannover für den Benchmark-Ansatz sowie für zwei MLR-Ansätze mit den sozioökonomischen Daten dar, auch hier einmal mit einem Gesamtmodell (RB 1.2.H) und einmal mit einem reinen Hannover-Modell (RB 1.3.H). Zwischen dem MLR-Gesamtmodell und dem Benchmark-Ansatz zeigen sich in den tatsächlichen Klassen 0 und 1 wieder keine aussagekräftigen Unterschiede. Das MLR-Hannover-Modell verlässt diesen Rahmen und weist einen deutlich schlechteren Anteil der falschen Nicht-0-Prognosen (in Klasse 0) auf, der fast vollständig auf die Klasse-2-Prognosen entfällt (+0,4 Prozentpunkte, Anteil bei den Klasse-1-Prognosen nur noch 2,3 %). Das schlechtere Abschneiden wird auch durch einen schlechteren Anteil zutreffender Klasse-0-Prognosen (96,9 %) gestützt. Auch in der tatsächlichen Klasse 1 zeigt sich hier eine Verschlechterung in ähnlicher Größenordnung gegenüber den beiden anderen Absätzen (Anteils-Median der Klasse-1-Prognosen sichtbar geringer, der der Klasse-2-Prognosen minimal höher³¹).

In der tatsächlichen Klasse 2 liefern die beiden MLR-Ansätze sichtbar bzw. minimal mehr Klasse-1-Prognosen als Beitrag zu den zutreffenden Nicht-0-Prognosen, aber bei den korrekten Klasse-2-Prognosen liegt auch hier der Benchmark-Ansatz vorne. Im Hinblick auf dies beiden MLR-Ansätze zeigt sich hier – anders als bei den Einzelstadtauswertungen zum Delikttyp 1 – in Klasse 2 zum ersten Mal eine sichtbar bessere Prognosequalität des Einzelstadt-Modells in einer Klasse, denn der Anteils-Median der korrekten Klasse-2-Prognosen ist sichtbar höher (+0,5 Prozentpunkte) als der des Gesamtmodells, ausgewertet für Hannover (RB 2.2.H). Ein Blick auf die Klasse-2-Prognosen in den tatsächlichen Klassen 0 und 1 legt aber nahe, dass das Hannover-Modell möglicherweise einfach eine generelle Tendenz zu mehr Klasse-2-Prognosen hat, eventuell sogar aus dem Grund, dass in Hannover der Anteil der Klasse-2-WWKn (0,5 %) höher ist als in beiden Städten zusammen (0,4 %).

²⁹ Der Unterschied zwischen den -0,2 Prozentpunkten hier und den insgesamt +0,9 Prozentpunkten bei den zutreffenden Nicht-0-Prognosen ist wieder auf Effekte der Medianbildung und der Rundung zurückzuführen.

³⁰ Dass der MLR-Ansatz ohne Nexiga Daten dennoch den höchsten PPV der drei Ansätze aufweist, beruht nur auf dem minimal kleineren Anteil von falschen Nicht-0-Prognosen (also denen in der tatsächlichen Klasse 0), der aber im Rahmen der Rundungsungenauigkeit als (potentiell) fast gleich hoch anzusehen ist.

³¹ Wie oben bereits angemerkt, sind Klasse-2-Prognosen als Nicht-0-Prognosen in Klasse 1 und damit als zutreffende Alarme nicht ganz so begrüßenswert wie Klasse-1-Prognosen in Klasse 2, weil Klasse-2-Prognosen die Art von Alarm darstellen, der die aufwendigeren polizeilichen Maßnahmen verursachen kann und sollte.

RB 2.0.H

Benchmark-Ansatz (Klasse wie in Vorwoche)

Trainings- und Testdaten:

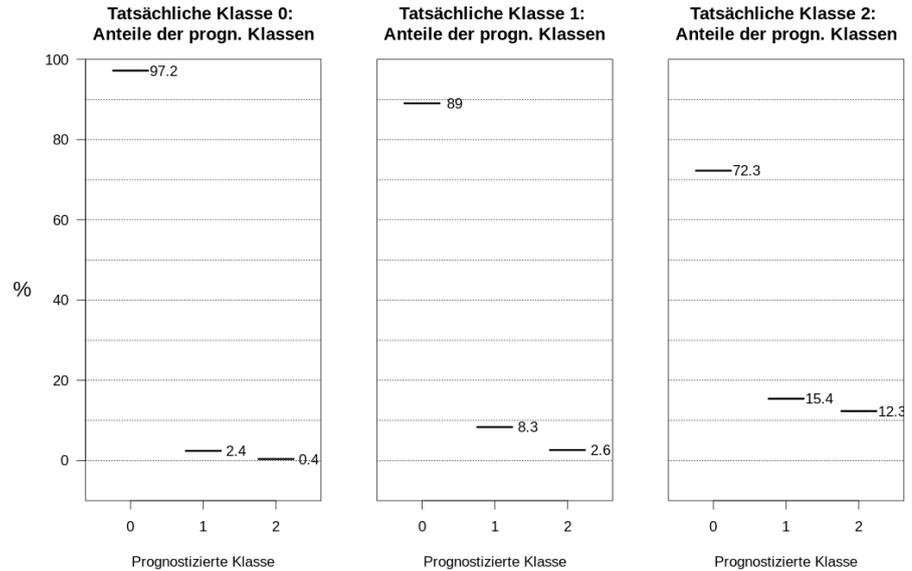
Nicht erforderlich

Nexiga-Daten einbezogen:

Nicht erforderlich

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWK_n in H



RB 2.2.H

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

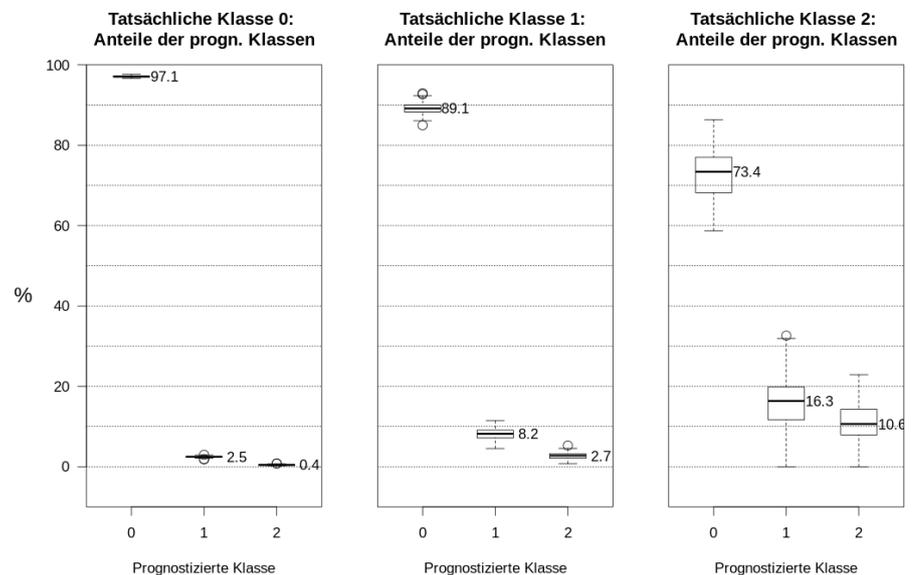
WWK_n in H und BS gesamt –
„Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWK_n in H



RB 2.3.H

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

WWK_n in H – „Hannover-
Modell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWK_n in H

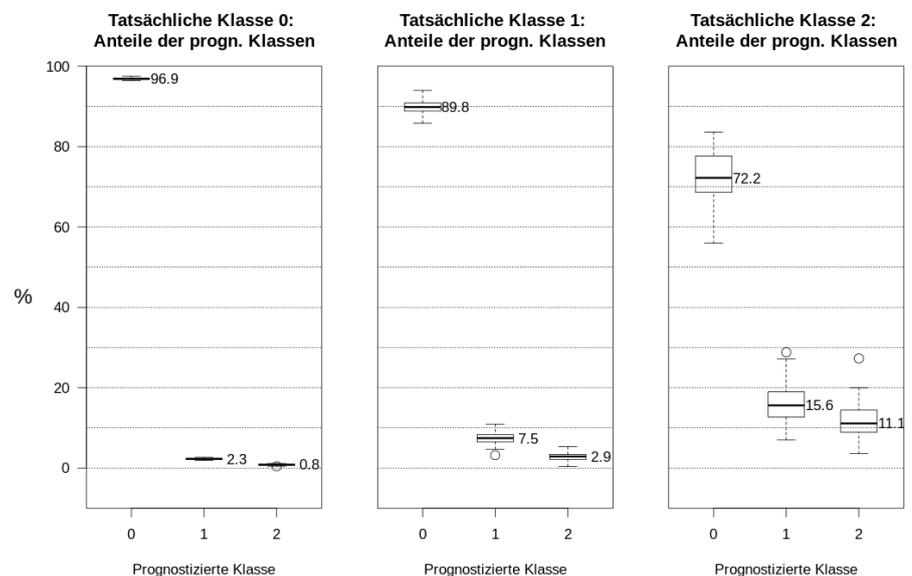


Abbildung RB.5: Prognosequalität verschiedener raumbezogener Ansätze für Delikttyp 2 für H alleine

Insgesamt ist festzuhalten, dass auch beim Delikttyp 2 „Schwerer Diebstahl aus Nichtwohnungen (ohne Ladendiebstahl)“ in der Einzelstadtauswertung für Hannover der Benchmark-Ansatz eine höhere Prognosequalität – vor allem in der tatsächlichen Klasse 2 – aufweist als die beiden untersuchten MLR-Ansätze, von denen der mit dem Gesamtmodell insgesamt besser abschneidet als der mit dem Hannover-Modell. Dies spiegelt sich diesmal vollständig in der Rangfolge der PPV-Ergebnisse wider (vgl. Tabelle RB.2).

Die Abbildung RB 6 stellt entsprechend der Abbildung RB 5 die Anteile der prognostizierten Klassen in der Einzelstadtauswertung für Braunschweig dar, wieder für den Benchmark-Ansatz sowie für zwei MLR-Ansätze mit den sozioökonomischen Daten, einmal mit einem Gesamtmodell (RB 1.2.BS) und einmal mit einem reinen Braunschweig-Modell (RB 1.4.BS).

In der tatsächlichen Klasse 0 zeigen sich zwischen allen drei Ansätzen wieder einmal fast keine Abweichungen außerhalb der Rundungsungenauigkeit von $\pm 0,1$ Prozentpunkten. In der tatsächlichen Klasse 1 zeichnet sich der MLR-Ansatz mit Gesamtmodell durch sichtbar bessere korrekte Klasse-1-Prognosen aus. Dass diese mit einem Anteils-Median von 0 bei den Klasse-2-Prognosen einhergehen, ist erfreulich für die exakte Klasse-1-Erkennung, aber ungünstig für die Nicht-0-Prognosen insgesamt. Der MLR-Ansatz mit Braunschweig-Modell schneidet in der tatsächlichen Klasse 1 sehr ähnlich ab wie der Benchmark-Ansatz.

In der tatsächlichen Klasse 2 zeigt sich bei beiden MLR-Ansätzen ein Anteils-Median von 0 für die prognostizierten Klassen 1 und 2. Anders als bei den Einzelstadtauswertungen für Braunschweig zu Delikttyp 1 liegen aber hier zumindest Boxoberkanten oberhalb von 0 vor, und zwar bei beiden Ansätzen bei den Klasse-1-Prognosen und beim MLR-Braunschweig-Modell sogar auch bei den Klasse-2-Prognosen. Das heißt, bei einem Viertel der Trainingsmengen tritt ein Anteil von Klasse-1-Prognosen auf, der bei ca. 18 % oder höher (RB 2.2.BS) bzw. bei ca. 14 % oder höher liegt (RB 2.4.BS), und eine unbestimmte Anzahl³² aus dem „zweithöchsten Viertel“ der Werte liegt immerhin oberhalb der 0 und unterhalb der Boxoberkante. Außerdem liegen beim MLR-Braunschweig-Modell auch bei den Klasse-2-Prognosen ein Viertel der bei 100 Trainingsmengen aufgetretenen Anteilswerte bei ca. 11 % oder höher. Damit ist die Prognosequalität der beiden MLR-Ansätze im Hinblick auf die Nicht-0-Prognosen in Klasse 2 nicht ganz so schlecht, wie die reine Medianbetrachtung und der (mit den Medianen berechnete) PPV befürchten lassen, und das MLR-Braunschweig-Modell schneidet sogar bei der korrekten Klasse-2-Erkennung besser (bzw. weniger schlecht) ab als das MLR-Gesamtmodell.

Zusammenfassend kann man feststellen, dass beim Delikttyp 2 auch in der Einzelstadtbetrachtung für Braunschweig die Prognosequalität des Benchmark-Ansatzes durch die beiden untersuchten MLR-Ansätze nicht übertroffen wird. Wie bei der Einzelstadtbetrachtung für Hannover übertrifft bei den MLR-Ansätzen das Braunschweig-Modell das Gesamtmodell bei den korrekten Klasse-2-Prognosen (in Klasse 2), und wie dort ist es bei den korrekten Klasse-1-Prognosen (in Klasse 1) umgekehrt.

Damit liefert auch beim Delikttyp 2 „Schwerer Diebstahl aus Nichtwohnungen (ohne Ladendiebstahl)“ der „daten-ignorierende“ Benchmark-Ansatz eine bessere Prognosequalität als alle untersuchten MLR-Ansätze.

³² Aus dem Boxplot geht nur hervor, dass ein Viertel der Werte zwischen 0 einschließlich (Median) und der Boxoberkante liegen. Da der Wert 0 potenziell mit zu diesem Viertel gehört, kann die genaue Anzahl der Werte, die tatsächlich über der 0 liegen, aus dem Boxplot nicht abgelesen werden.

RB 2.0.BS

Benchmark-Ansatz (Klasse wie in Vorwoche)

Trainings- und Testdaten:

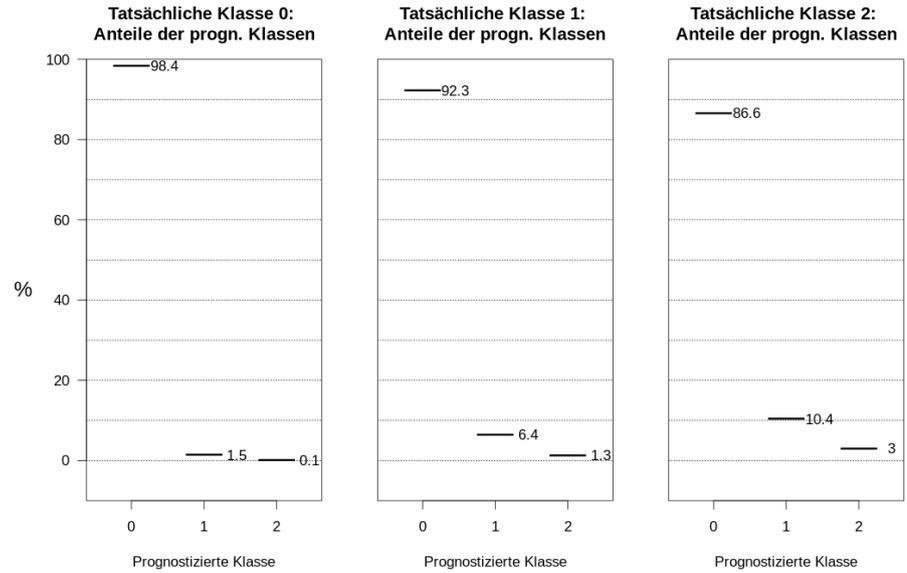
Nicht erforderlich

Nexiga-Daten einbezogen:

Nicht erforderlich

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in BS



RB 2.2.BS

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

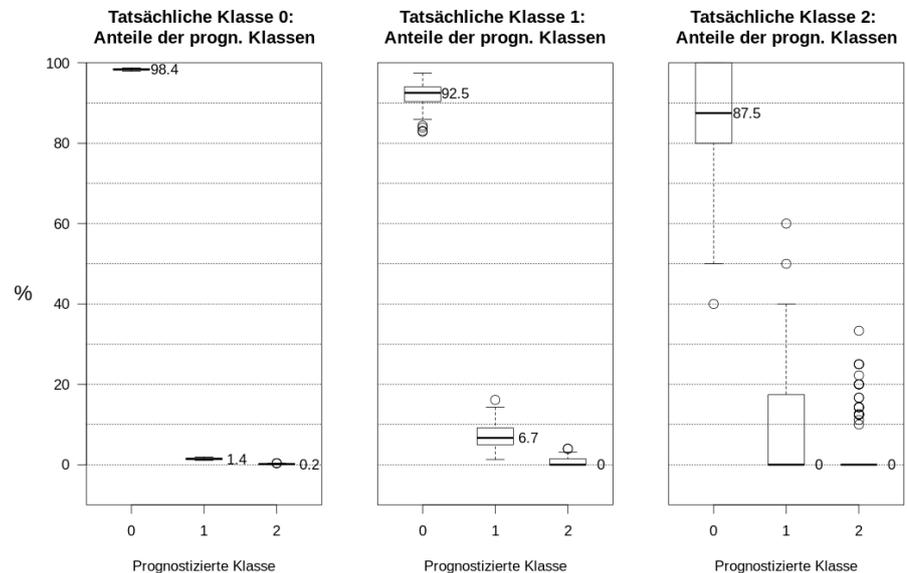
WWKn in H und BS gesamt –
„Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in BS



RB 2.4.BS

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

WWKn in BS – „Braun-
schweig-Modell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in BS

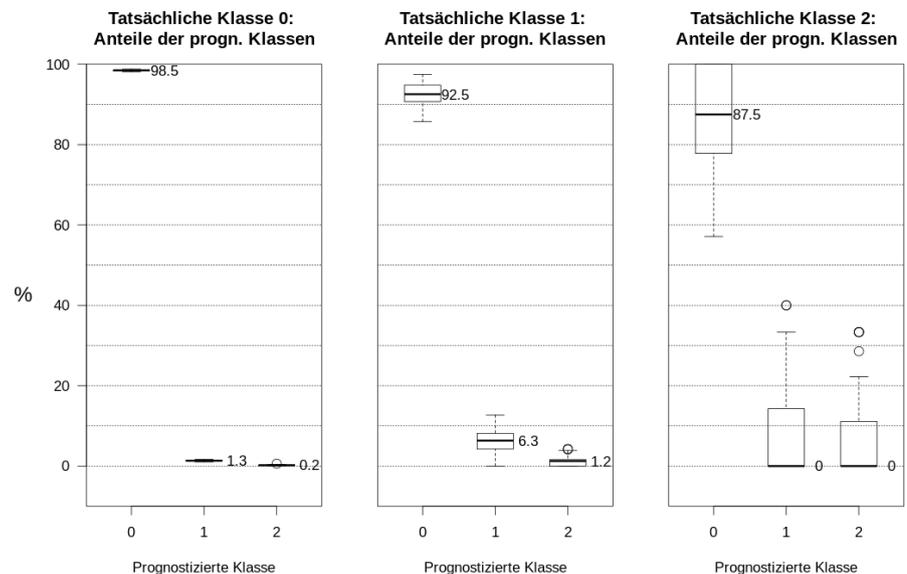


Abbildung RB.6: Prognosequalität verschiedener raumbezogener Ansätze für Delikttyp 2 für BS alleine

4.2.4 Delikttyp 3 „Diebstahl in / aus KFZ“ (Deliktschlüssel 350 und 450)

Die Tabelle RB.3 stellt analog zu Tabelle RB.2 die Gegebenheiten in den Daten (Wohnquartiere und WWKn wie dort, tatsächliche Klassenzugehörigkeiten anders) und die PPV-Ergebnisse für verschiedene untersuchte Prognoseansätze dar, diesmal für den Delikttyp 3 „Diebstahl in / aus KFZ“. Die PPV-Ergebnisse liegen hier für alle Ansätze und Auswertungsregionen etwas höher als beim Delikttyp 2 und damit deutlich höher als beim Delikttyp 1. Wieder sind sie bei den Einzelstadtbetrachtungen für Hannover noch etwas höher als bei den Gesamtbetrachtungen, und wieder sind sie bei den Einzelstadtbetrachtungen für Braunschweig deutlich niedriger.

Dennoch bedeuten diese Werte, dass auch hier nur um 15,0 % (bzw. in Braunschweig um 9,5 %) der Nicht-0-Prognosen, die die betrachteten Ansätze für bestimmte Wohnquartiere in bestimmten Wochen liefern, zutreffend sind in dem „weiteren Sinn“, dass in diesem Wohnquartier in dieser Woche auch wirklich ein Delikt (Klasse 1) bzw. mehrere Delikte (Klasse 2) zu verzeichnen sind.

Tabelle RB.3: Kennzahlen zum Delikttyp 3 (Datenlage und Prognosequalität)

Gebiet	Anzahl Wohnquartiere	Wohnquartier-Woche-Kombinationen (WWKn)	darunter				
			Klasse 0 (0 Delikte)	Klasse 1 (1 Delikt)	Klasse 2 (>1 Delikte)		
Braunschweig, Stadt	289	45.084	43.470	1.450	164		
Hannover, Landeshauptstadt	671	104.676	98.219	5.463	994		
Gesamt	960	149.760	141.689	6.913	1.158		
	PPV (Anteil der zutreffenden an allen Nicht-0-Prognosen, berechnet über alle drei tatsächlichen Klassen zusammen) bei Prognoseansatz ... (in %, Medianbetrachtung))						
	RB 3.0. ...	RB 3.1. ...	RB 3.2. ...	RB 3.3. ...	RB 3.4. ...	RB 3.5. ...	RB 3.6. ...
... .BS	9,9		9,7		9,1	9,6	9,6
... .H	15,7		15,3	15,3		15,4	15,3
... .G	14,3	14,7	14,1			14,4	14,4

Die Abbildungen RB.7a und RB.7b stellen analog zu Abbildung RB.4 die Anteile der prognostizierten Klassen in der Gesamtauswertung dar, und zwar diesmal zusätzlich zu den drei bisher betrachteten Ansätzen noch für die Multinomiale Logistische Ridge-Regression (MLRR) als Variante des MLR-Ansatzes (RB 3.5.G) sowie für den Multinomial-Functional-Tree-(MFT-)Ansatz (RB 3.6.G), und zwar beide mit Einbeziehung der Nexiga-Daten.

RB 3.0.G

Benchmark-Ansatz (Klasse wie in Vorwoche)

Trainings- und Testdaten:

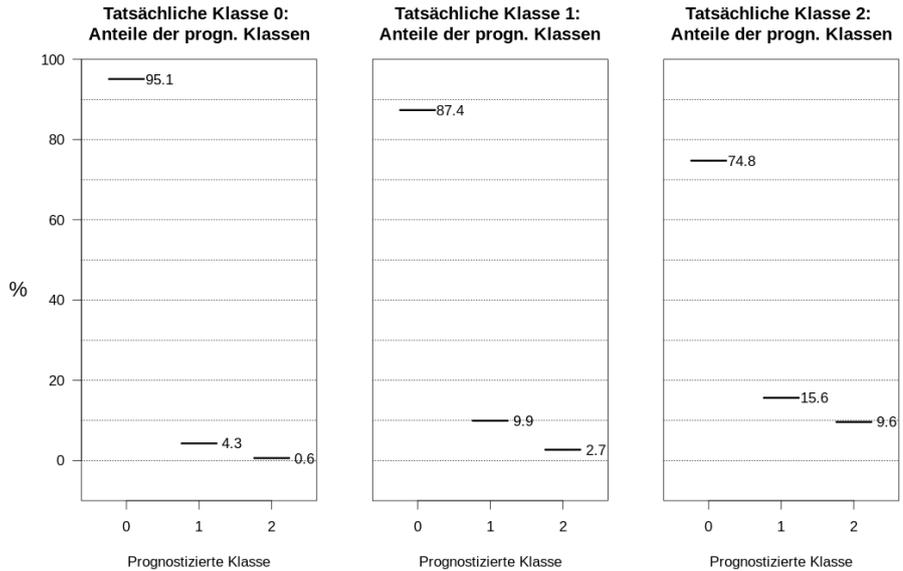
Nicht erforderlich

Nexiga-Daten einbezogen:

Nicht erforderlich

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in
H und BS gesamt



RB 3.1.G

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

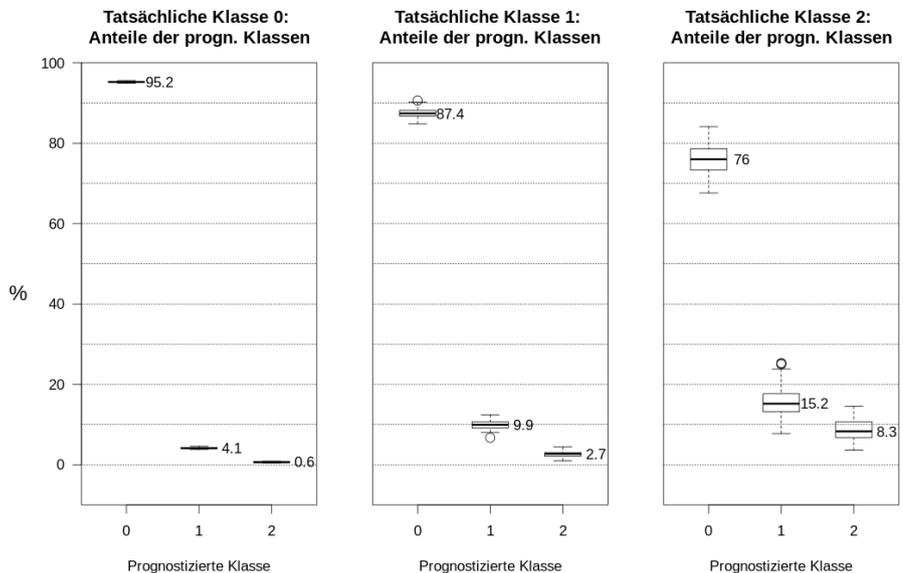
WWKn in H und BS gesamt –
„Gesamtmodell“

Nexiga-Daten einbezogen:

Nein

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in
H und BS gesamt



RB 3.2.G

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

WWKn in H und BS gesamt –
„Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in
H und BS gesamt

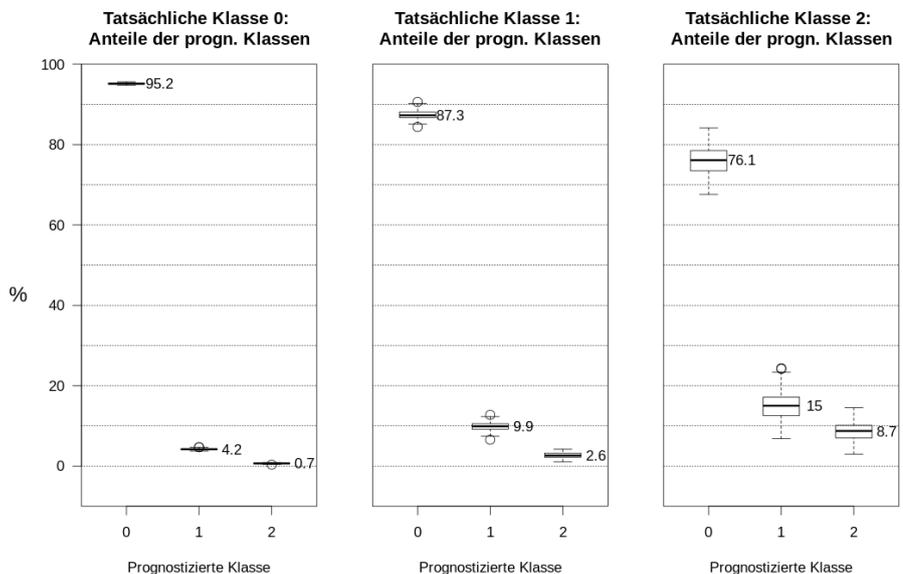


Abbildung RB.7a: Prognosequalität verschiedener raumbezogener Ansätze für Delikttyp 3 für H und BS zusammen

RB 3.5.G

Multinomiale Logistische Ridge-Regression (MLRR)

Trainings- und Testdaten:

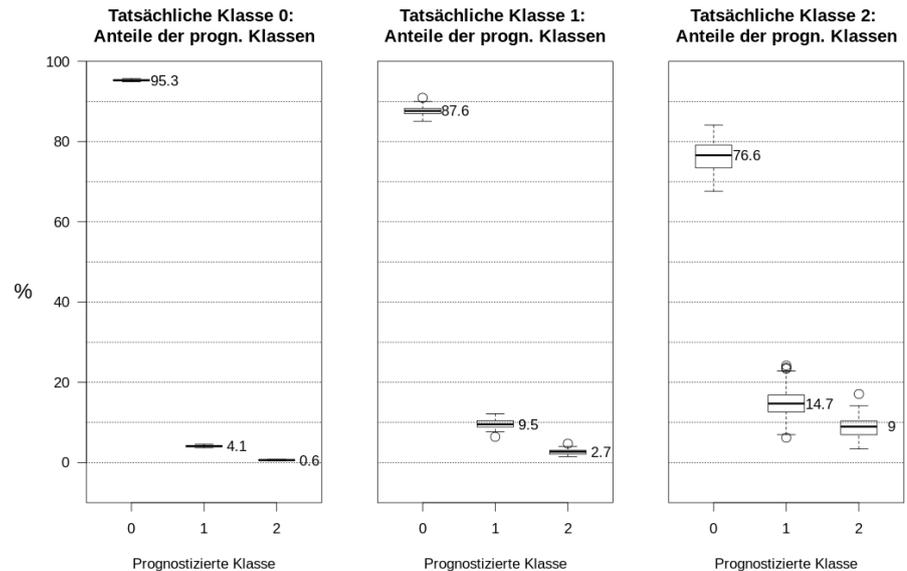
WWK_n in H und BS gesamt – „Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte Klasse an tatsächlicher Klasse berechnet für:

WWK_n in H und BS gesamt



RB 3.6.G

Multinomial Functional Tree (MFT)

Trainings- und Testdaten:

WWK_n in H und BS gesamt – „Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte Klasse an tatsächlicher Klasse berechnet für:

WWK_n in H und BS gesamt

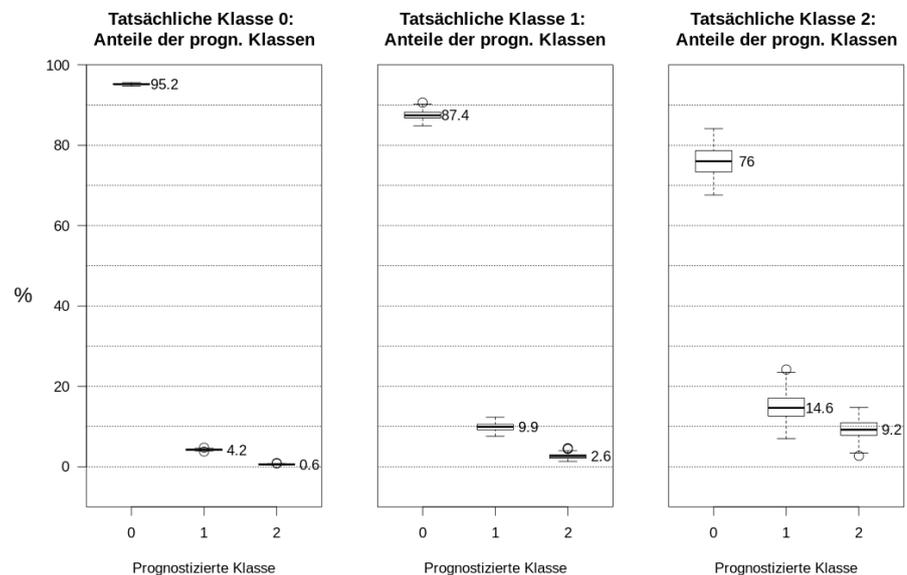


Abbildung RB.7b: Prognosequalität verschiedener raumbezogener Ansätze für Delikttyp 3 für H und BS zusammen

Es zeigt sich folgendes Gesamtbild: In der tatsächlichen Klasse 0 finden sich nur minimale Abweichungen zwischen den fünf Ansätzen, die sich fast alle im Rahmen der Rundungsungenauigkeit bewegen. In der tatsächlichen Klasse 1 gilt dies ebenfalls für alle Ansätze mit Ausnahme des MLRR-Ansatzes, der bei den zutreffenden Klasse-1-Prognosen sichtbar schlechter abschneidet. In der tatsächlichen Klasse 2 liefern die beiden neu betrachteten Ansätze sichtbar bessere korrekte Klasse-2-Prognosen als die beiden MLR-Ansätze, in beiden Fällen verbunden mit einer Verminderung der falschen Klasse-1-Prognosen, wobei sich als Saldo immer noch ein höherer oder wenigstens gleich hoher Anteil von zutreffenden Nicht-0-Prognosen ergibt. Der Benchmark-Ansatz liefert allerdings noch mehr korrekte Klasse-2-Prognosen und in Verbindung mit dem höchsten Anteil von Klasse-1-Prognosen auch die meisten Nicht-0-Prognosen aller fünf Ansätze und hat damit über alle drei Klassen die höchste Prognosequalität.

Dennoch erreicht der Benchmark-Ansatz bei diesem Delikttyp in der Gesamtbetrachtung nicht den höchsten PPV, sondern wird gerade eben von den beiden neuen Ansätzen und noch deutlicher vom MLR-Ansatz ohne Nexiga-Daten übertroffen.

RB 3.0.H

Benchmark-Ansatz (Klasse wie in Vorwoche)

Trainings- und Testdaten:

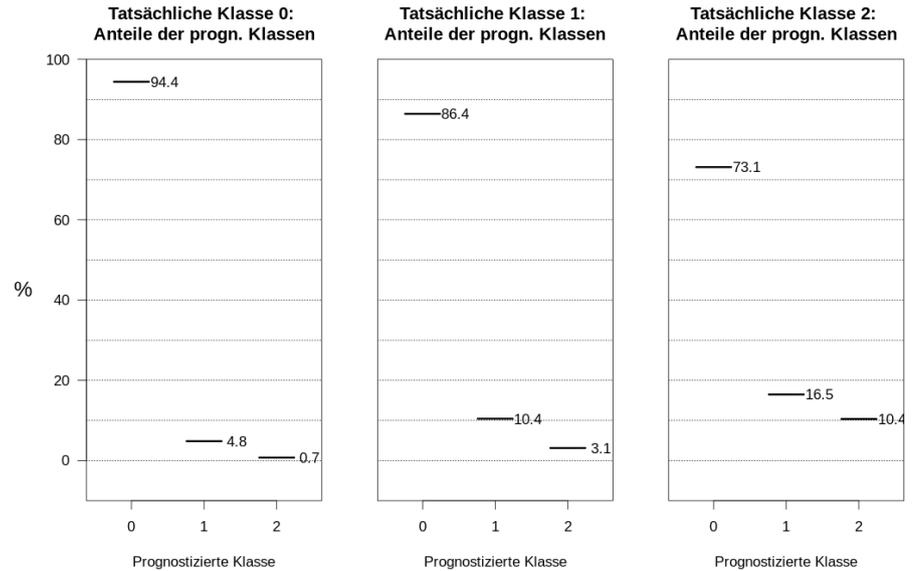
Nicht erforderlich

Nexiga-Daten einbezogen:

Nicht erforderlich

Anteile prognostizierte Klasse an tatsächlicher Klasse berechnet für:

WWKn in H



RB 3.2.H

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

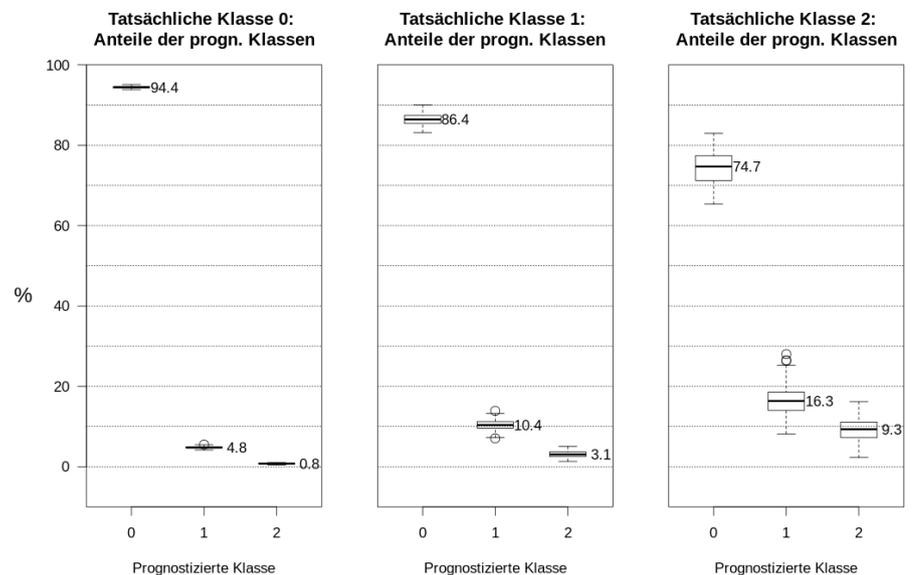
WWKn in H und BS gesamt – „Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte Klasse an tatsächlicher Klasse berechnet für:

WWKn in H



RB 3.3.H

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

WWKn in H – „Hannover-Modell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte Klasse an tatsächlicher Klasse berechnet für:

WWKn in H

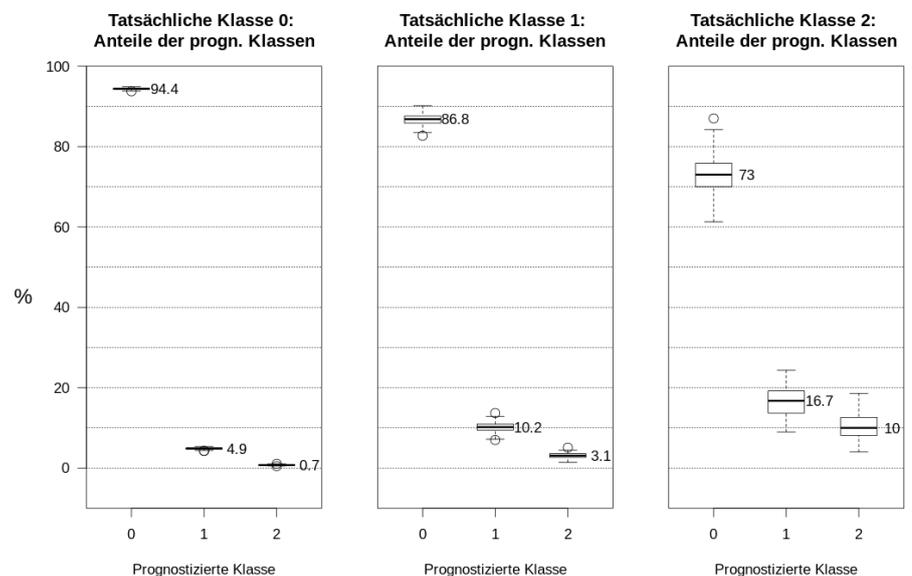


Abbildung RB.8a: Prognosequalität verschiedener raumbezogener Ansätze für Delikttyp 3 Hannover alleine

RB 3.5.H

Multinomiale Logistische Ridge-Regression (MLRR)

Trainings- und Testdaten:

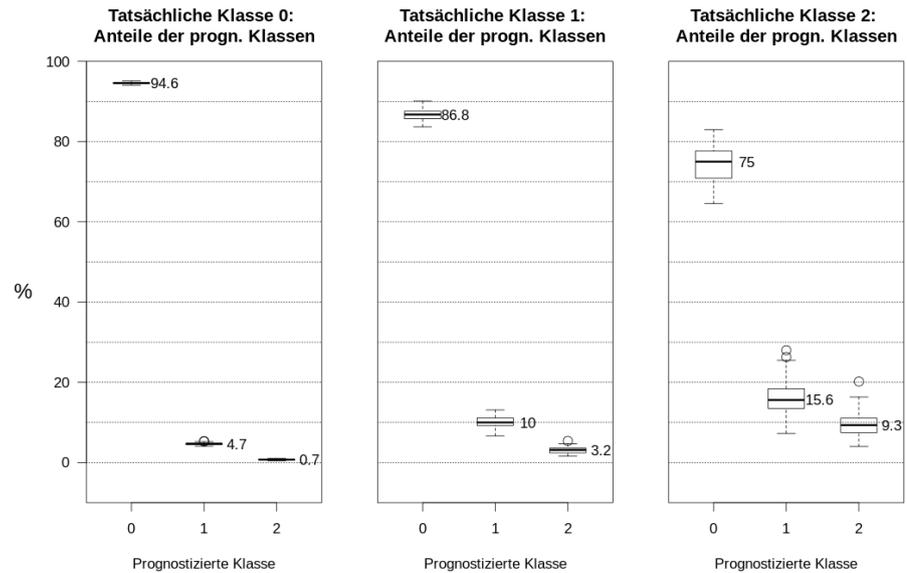
WWK_n in H und BS gesamt – „Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte Klasse an tatsächlicher Klasse berechnet für:

WWK_n in H



RB 3.6.H

Multinomial Functional Tree (MFT)

Trainings- und Testdaten:

WWK_n in H und BS gesamt – „Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte Klasse an tatsächlicher Klasse berechnet für:

WWK_n in H

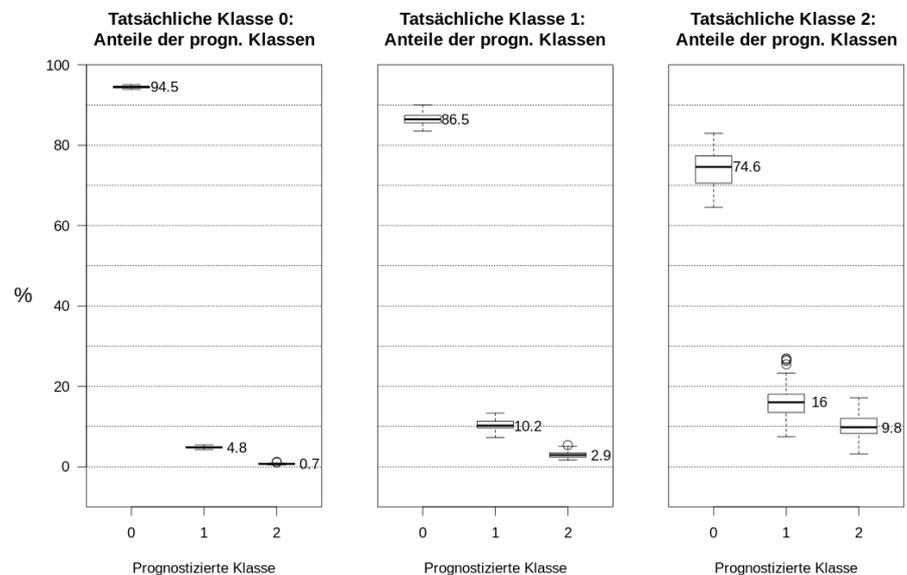


Abbildung RB.8b: Prognosequalität verschiedener raumbezogener Ansätze für Delikttyp 3 für H alleine

Die Abbildungen RB.8a und RB.8b stellen analog zu Abbildung RB.5 die Anteile der prognostizierten Klassen in der Hannover-Auswertung dar. Der MLR-Ansatz ohne Nexiga-Daten fehlt wieder, dafür kommen die prognostizierten Anteile eines MLR-Ansatzes mit den Nexiga-Daten als reines Hannover-Modell hinzu (RB 3.3.H). Auch für die beiden neu betrachteten Ansätze gibt es eine Hannover-Auswertung.

Auch hier zeigt sich eine große Ähnlichkeit der prognostizierten Anteile in der tatsächlichen Klasse 0 über alle fünf Ansätze. Das gilt überwiegend auch für die Prognoseanteile in der tatsächlichen Klasse 1. Hier liegen die korrekten Klasse-1-Prognosen aber beim Benchmark-Ansatz und beim MLR-Gesamtmodell-Ansatz mit Nexiga-Daten sichtbar oben, und die falschen Klasse-2-Prognosen beim MFT-Ansatz sichtbar unten. Bei den korrekten Klasse-2-Prognosen schneidet hier unter den datengestützten Ansätzen der MLR-Hannover-Ansatz mit Nexiga-Daten am besten ab, dicht gefolgt vom MFT-Ansatz. Der Benchmark-Ansatz weist einen noch besseren Anteil von korrekten Klasse-2-Prognosen auf, wird aber im Hinblick auf die Nicht-0-Prognosen insgesamt vom MLR-Hannover-Ansatz nahezu eingeholt. Insgesamt ergibt sich auch in der Hannover-

RB 3.0.BS

Benchmark-Ansatz (Klasse wie in Vorwoche)

Trainings- und Testdaten:

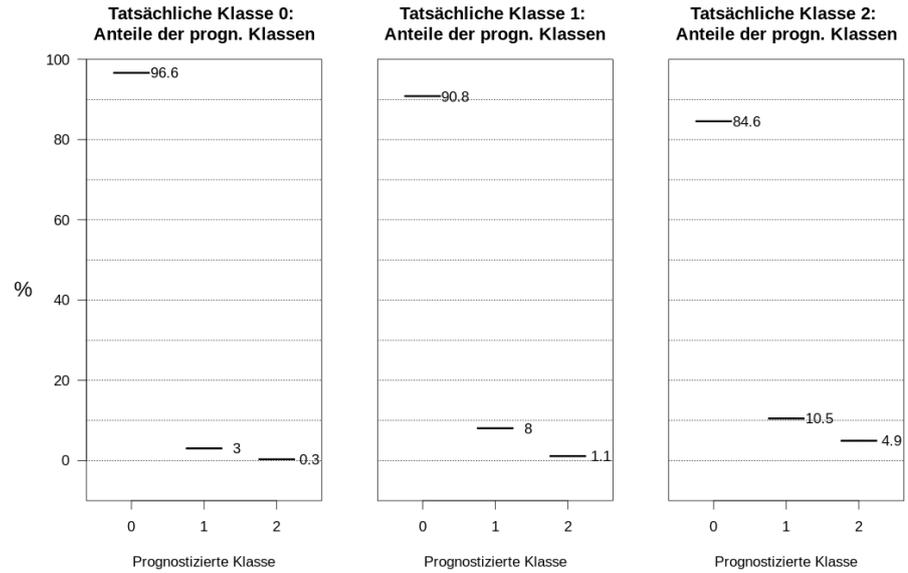
Nicht erforderlich

Nexiga-Daten einbezogen:

Nicht erforderlich

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWK_n in BS



RB 3.2.BS

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

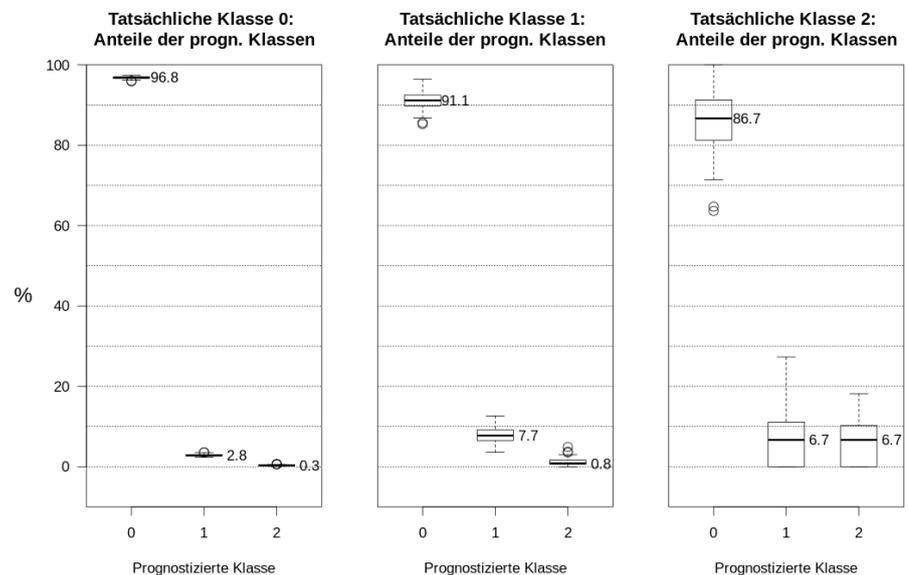
WWK_n in H und BS gesamt –
„Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWK_n in BS



RB 3.4.BS

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

WWK_n in BS – „Braun-
schweig-Modell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWK_n in BS

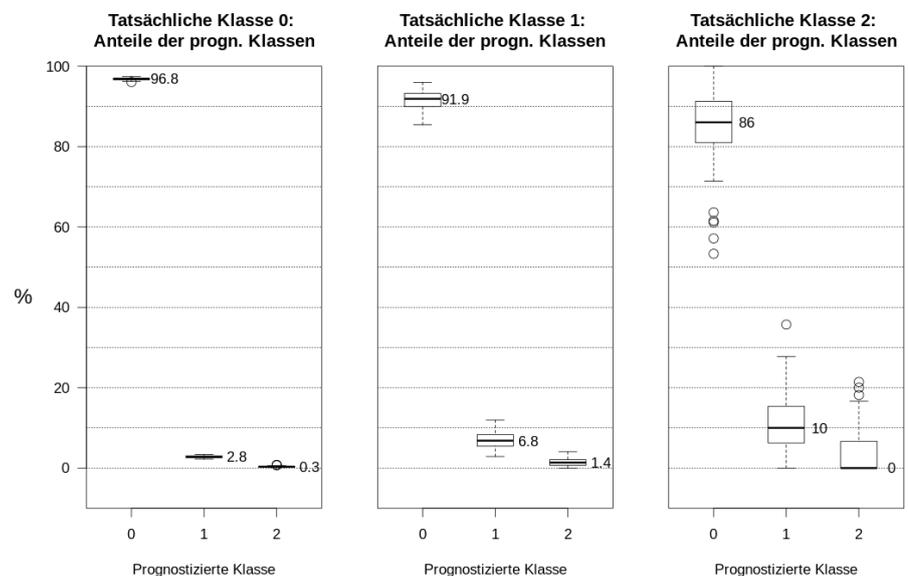


Abbildung RB.9a: Prognosequalität verschiedener raumbezogener Ansätze für Delikttyp 3 für BS alleine

RB 3.5.BS

Multinomiale Logistische Ridge-Regression (MLRR)

Trainings- und Testdaten:

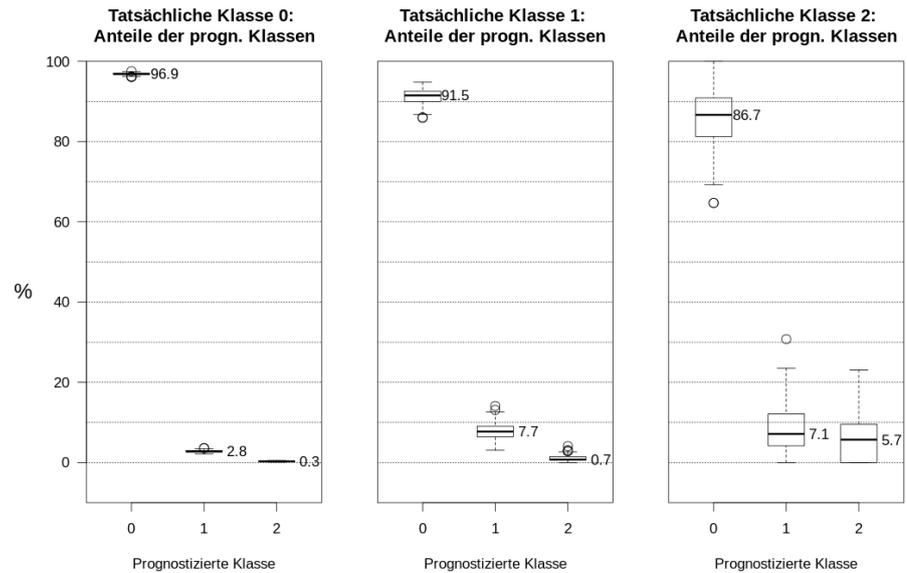
WWK_n in H und BS gesamt – „Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte Klasse an tatsächlicher Klasse berechnet für:

WWK_n in BS



RB 3.6.BS

Multinomial Functional Tree (MFT)

Trainings- und Testdaten:

WWK_n in H und BS gesamt – „Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte Klasse an tatsächlicher Klasse berechnet für:

WWK_n in BS

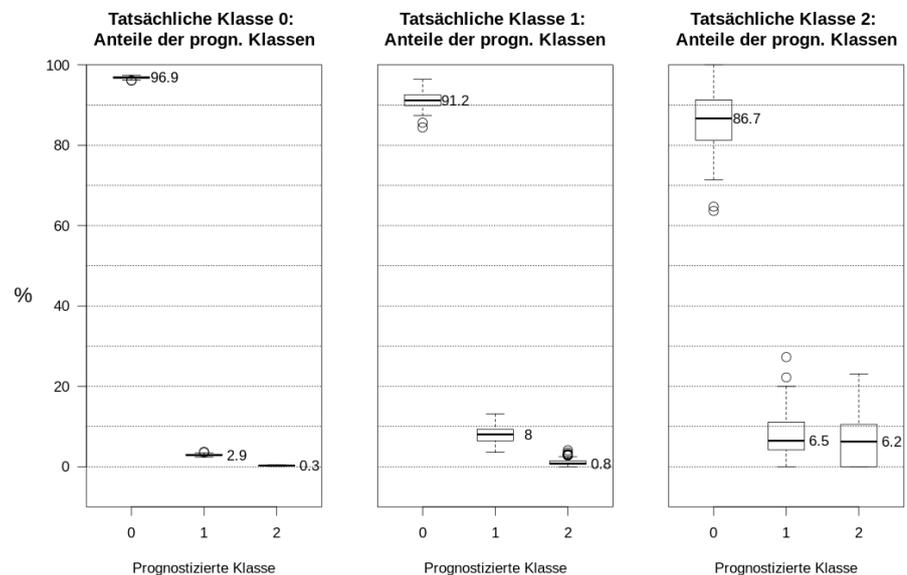


Abbildung RB.9b: Prognosequalität verschiedener raumbezogener Ansätze für Delikttyp 3 für BS alleine

Betrachtung eine klare Überlegenheit des Benchmark-Ansatzes gegenüber allen anderen Ansätzen, die hier auch durch den PPV bekräftigt wird.

Die Abbildungen RB.9a und RB.9b stellen entsprechend den Abbildungen RB.8a und RB.8b die Anteile der prognostizierten Klassen in der Braunschweig-Auswertung dar. Hier sind in der tatsächlichen Klasse 0 die vier datengestützten Ansätze etwas besser als der Benchmark-Ansatz (weniger falsche Nicht-0-Prognosen, bestätigt durch höhere Mediane bei den zutreffenden Klasse-0-Prognosen). In der tatsächlichen Klasse 1 zeigt sich ein differenziertes Bild: Der Benchmark-Ansatz liefert von allen Ansätzen die meisten zutreffenden Nicht-0-Prognosen, und mit einem Anteils-Median von 8,0 % auch die meisten korrekten Klasse-1-Prognosen. Diese Platzierung teilt er sich aber mit dem MFT-Ansatz, der hier außerdem sichtbar weniger falsche Klasse-2-Prognosen liefert – also weniger von den etwas unerwünschteren zutreffenden Alarmen. Der MLR- und der MLRR-Gesamtansatz schneiden sichtbar schlechter ab, der MLR-Ansatz mit Braunschweig-Modell noch schlechter.

In der tatsächlichen Klasse 2 fallen die Ergebnisse der fünf Ansätze ebenfalls relativ uneinheitlich aus: Der Benchmark-Ansatz liefert einen Anteil der zutreffenden Nicht-0-Prognosen von 15,4 %,

davon ein knappes Drittel korrekte Klasse-2-Prognosen. Die drei datengestützten Ansätze mit Gesamtmodellen schneiden bei den korrekten Klasse-2-Prognosen im Median alle besser ab – beim MLR-Gesamtansatz umfassen die Klasse-2-Prognosen sogar die Hälfte aller Nicht-0-Prognosen, und auch bei den beiden anderen Ansätzen ist der Abstand der beiden Anteils-Mediane für Klasse-1- und Klasse-2-Prognosen auffallend gering. Diese gute Klasse-2-Erkennung in Relation zur gesamten Nicht-Klasse-0-Erkennung, die hier zum ersten Mal bei den datengestützten Ansätzen auftritt, geht in allen drei Fällen einher mit einem Gesamtanteil der Nicht-0-Prognosen, der schlechter ausfällt als beim Benchmark-Ansatz. Eine weitere ungünstige Beobachtung sind die bei allen drei Ansätzen auftretenden Boxunterkanten von 0 für die Klasse-2-Prognosen, denn sie bedeuten, dass für jeweils mindestens ein Viertel der geprüften Trainingsmengen keine einzige Klasse-2-WWK in der zugehörigen Testmenge zutreffend prognostiziert wurde.

Der MLR-Ansatz mit Braunschweig-Modell liefert im Median 0,0 % korrekte Klasse-2-Prognosen – wieder einmal mit der „tröstenden“ Boxoberkante über 0 – und auch nur 10,0 % Klasse-1-Prognosen und schneidet damit deutlich am schlechtesten ab.

In der Einzelstadtbetrachtung für Braunschweig liegt der Benchmark-Ansatz also in den tatsächlichen Klassen 1 und 2 vor den vier untersuchten datengestützten Ansätzen, aber in der tatsächlichen Klasse 0 liefern diese alle etwas weniger falsche Klasse-0-Prognosen – hier entsprechen 0,2 Prozentpunkte weniger immerhin 283 WWKn. Daraus resultieren – mit Ausnahme des MLR-Braunschweig-Ansatzes, der in den tatsächlichen Klassen 1 und 2 sehr schlecht abschneidet – PPV-Ergebnisse, die nur minimal unter dem PPV des Benchmark-Ansatzes liegen.

Als Fazit aus der Gesamt- und den beiden Einzelstadtbetrachtungen ist festzuhalten, dass auch beim Delikttyp 3 „Diebstahl in / aus KFZ“ der Benchmark-Ansatz die höchste Prognosequalität aller untersuchten Ansätze aufweist.

4.2.5 Delikttyp 4 „Diebstahl von KFZ“ (Deliktschlüssel 300, 345, 400, 410 und 445)

Die Tabelle RB.4 enthält die entsprechenden Informationen wie die Tabellen RB.1 bis RB.3, diesmal für den Delikttyp 4 „Diebstahl von KFZ“. Es fällt auf, dass in den tatsächlichen Klassen 1 und 2 deutlich weniger WWKn vorliegen als bei den anderen Delikttypen. Mit anderen Worten: KFZ-Vollentwendungen sind in beiden Städten deutlich seltener als die drei anderen Delikttypen, und Häufungen in einem Wohnquartier innerhalb einer Woche treten – nicht zuletzt auch wegen dieser vergleichsweise Seltenheit dieser Delikte – nur in Einzelfällen auf (nämlich nur 52mal bei 149.760 WWKn³³). Damit liegen so wenige Klasse-2-WWKn in den Daten vor, dass die Datenbasis schon für das Trainieren eines Gesamtmodells für die drei Klassen sehr dünn ist – und zudem dürften in den Testmengen oft kaum Klasse-2-WWKn enthalten sein.

Daher wurden bei diesem Delikttyp keine MLR-Einzelstadt-Modelle untersucht, dafür aber zusätzlich zu den MLR-Gesamtmodellen ein MFT-Gesamtmodell mit Einbeziehung der Nexiga-Daten (RB.4.3), für das die Gesamtauswertung und die beiden Einzelstadtauswertungen vorliegen. Die PPV-Ergebnisse sind für alle Ansätze und Auswertungsregionen deutlich niedriger als bei allen anderen Delikttypen, mit etwas höheren Werten in der Hannover- als in der Gesamtbetrachtung, und in der Braunschweig-Betrachtung noch einmal sichtbar niedriger oder sogar 0,0 %.

Die Abbildungen RB.10a und RB.10b zeigen wieder die Boxplots zu den Anteils-Verteilungen der prognostizierten Klassen für jede der tatsächlichen Klassen, hier ausgewertet für Braunschweig und Hannover zusammen. In der tatsächlichen Klasse 0 zeigt sich wie bei den anderen Delikttypen ein weitgehend einheitliches Bild für alle untersuchten Ansätze. Auffällig ist, dass bei diesem Delikttyp anstelle der auch sonst schon sehr kleinen Anteils-Mediane für die Klasse-2-Prognosen

³³ Aus diesem Grund sind auch die Angaben in der Klasse-2-Spalte im unteren Tabellenteil ausnahmsweise nicht auf ganze WWKn gerundet, sondern auf eine Nachkommastelle.

im Rahmen der Rundungsungenauigkeit tatsächlich fast immer 0 ist. Die einzige aussagekräftige Abweichung nach oben zeigt sich beim MLR-Ansatz mit den Nexiga-Daten (RB 4.2.G) mit einem Anteils-Median von 0,4 %, so dass dieser im Hinblick auf die falschen Nicht-0-Prognosen schlechter abschneidet als die drei anderen (bekräftigt auch durch den Anteils-Median der korrekten Klasse-0-Prognosen).

Tabelle RB.4: Kennzahlen zum Delikttyp 4 (Datenlage und Prognosequalität)

Gebiet	Anzahl Wohnquartiere	Wohnquartier-Woche-Kombinationen (WWKn)	darunter		
			Klasse 0 (0 Delikte)	Klasse 1 (1 Delikt)	Klasse 2 (>1 Delikte)
Braunschweig, Stadt	289	45.084	44.677	395	12
Hannover, Landeshauptstadt	671	104.676	103.770	866	40
Gesamt	960	149.760	148.447	1.261	52
	PPV (Anteil der zutreffenden an allen Nicht-0-Prognosen, berechnet über alle drei tatsächlichen Klassen zusammen) bei Prognoseansatz ... (in %, Medianbetrachtung))				
	RB 4.0. ...	RB 4.1. ...	RB 4.2. ...	RB 4.3. ...	
... .BS	1,0		0,0	0,0	
... .H	3,1		1,4	2,1	
... .G	2,5	1,6	1,1	1,7	

In der tatsächlichen Klasse 1 schneidet der Benchmark-Ansatz bei den zutreffenden Klasse-1-Prognosen mit einem Anteil von 1,9 % sichtbar am besten ab, dabei tritt aber auch ein Anteil der Klasse-2-Prognosen von 0,2 % auf (bei den drei datengestützten Ansätzen: im Median 0,0 %). Dies ist hier besonders ungünstig, weil der Benchmark-Ansatz bei den korrekten Klasse-2-Prognosen (in Klasse 2) einen Anteil von 0 aufweist. Das heißt, über die drei tatsächlichen Klassen betrachtet sind alle Klasse-2-Prognosen des Benchmark-Ansatzes in Wirklichkeit unzutreffend (aber immerhin zutreffende Nicht-0-Prognosen in Klasse 1). Noch schlimmer ist es beim MLR-Ansatz mit Nexiga-Daten, denn hier treten die einzigen Klasse-2-Prognosen wie oben angemerkt in der tatsächlichen Klasse 0 auf, sind also auch noch falsche Nicht-0-Prognosen.

In der tatsächlichen Klasse 2 tritt bei allen drei datengestützten Ansätzen hier zum ersten Mal ein Problemfeld in der Gesamtbetrachtung auf, dass bisher nur vereinzelt bei den Braunschweig-Betrachtungen vorkam: Anteils-Mediane oder – schlimmer – ganze Boxen auf der Nulllinie, und zwar bei den Klasse-1- und den Klasse-2-Prognosen. Wie oben gilt, dass eine Boxoberkante oberhalb der 0 eine gewisse Hoffnung auf Anteile (hier nur von Klasse-1-Prognosen) oberhalb

RB 4.0.G

Benchmark-Ansatz (Klasse wie in Vorwoche)

Trainings- und Testdaten:

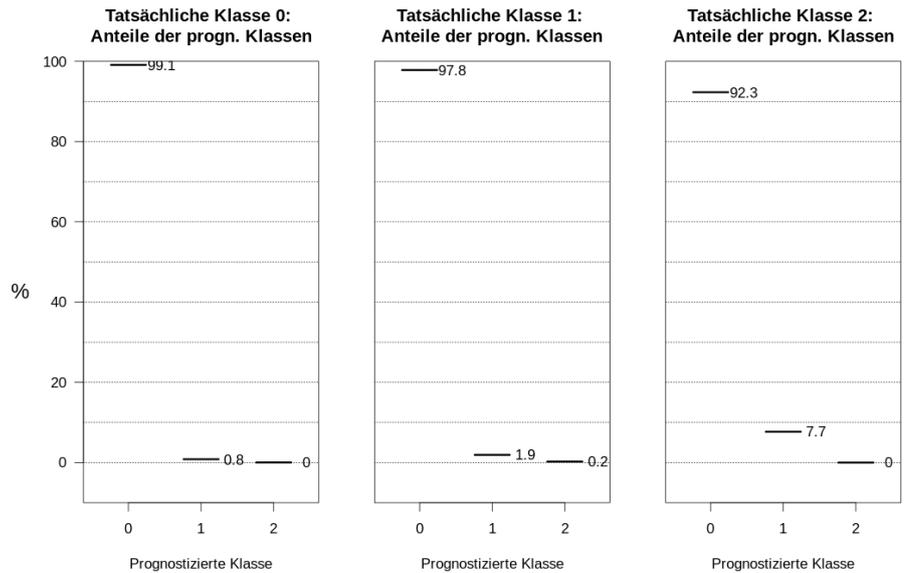
Nicht erforderlich

Nexiga-Daten einbezogen:

Nicht erforderlich

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWK_n in
H und BS gesamt



RB 4.1.G

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

WWK_n in H und BS gesamt –
„Gesamtmodell“

Nexiga-Daten einbezogen:

Nein

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWK_n in
H und BS gesamt

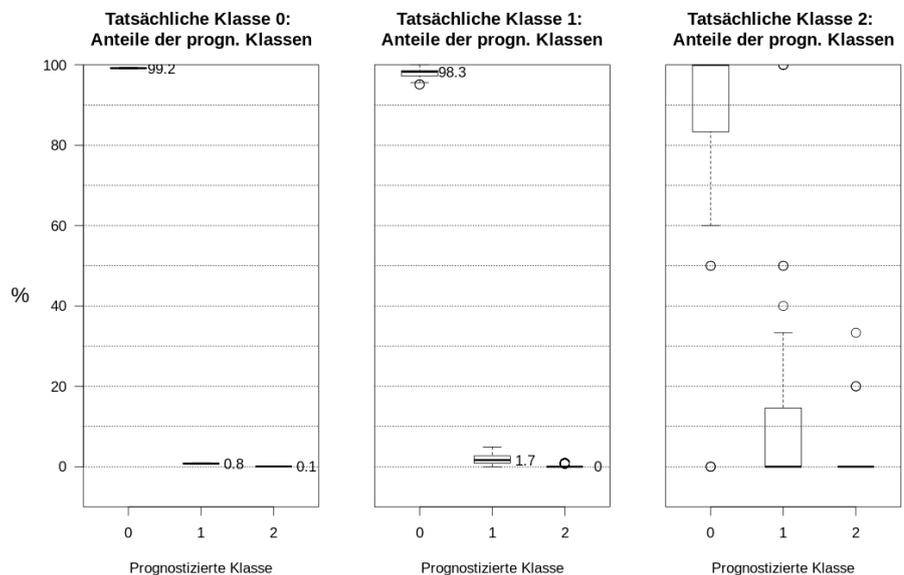


Abbildung RB.10a: Prognosequalität verschiedener raumbezogener Ansätze für Delikttyp 4 für H und BS zusammen

der 0 rechtfertigt, aber „im Mittel“ (also in der Medianbetrachtung) liefern die betreffenden Ansätze (hier die beiden MLR-Varianten) eben keine einzige Klasse-1-Prognose. Bei den korrekten Klasse-2-Prognosen sieht es noch schlechter aus, weil hier die genannte Hoffnung nur noch durch vereinzelte positive Anteilswerte genährt wird – und beim MFT-Ansatz gibt es nicht einmal die. Auch im Benchmark-Ansatz liegt der Anteil der korrekten Klasse-2-Prognosen bei 0, aber hier gibt es mit einem Anteil von 7,7 % wenigstens Klasse-1-Prognosen als zutreffende Nicht-0-Prognosen.

Die festgestellte überlegene Prognosequalität des Benchmark-Ansatzes für diesen Delikttyp in der Gesamtbetrachtung spiegelt sich auch im mit Abstand höchsten PPV wider, die angemerkten Schwächen des MLR-Ansatzes mit Nexiga-Daten auch in dessen PPV, der noch niedriger liegt als der der beiden übrigen Ansätze (vgl. Tabelle RB.4). Es zeigt sich aber wie bei den übrigen Delikttypen, dass der MLR-Ansatz mit Nexiga-Daten bei den korrekten Klasse-2-Prognosen etwas besser abschneidet als ohne Nexiga-Daten (hier sind es allerdings nur noch einige Ausreißer mehr oberhalb der 0), allerdings wieder um den Preis von mehr Klasse-2-Prognosen in der tatsächlichen Klasse 0 mit den entsprechenden Folgen für den PPV. Daher wird im Folgenden wie

RB 4.2.G

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

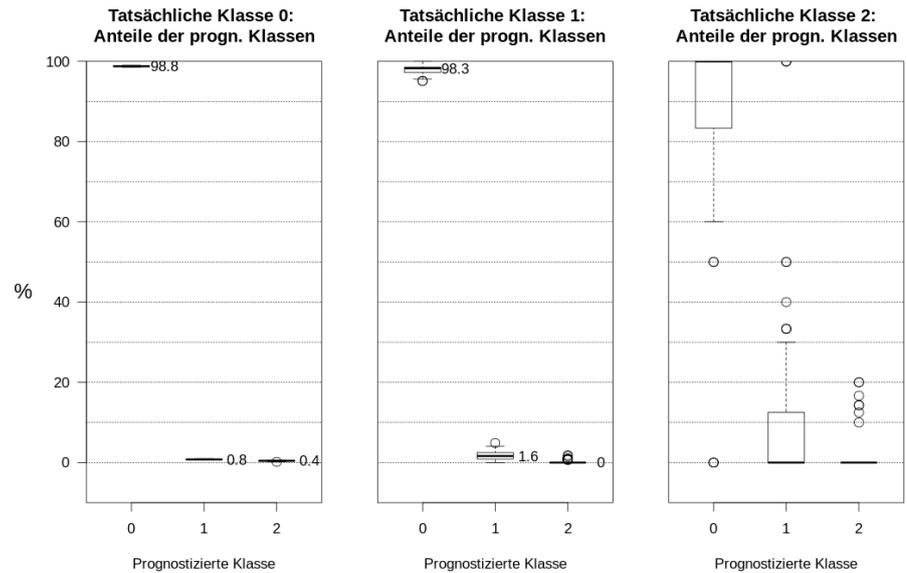
WWK_n in H und BS gesamt – „Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte Klasse an tatsächlicher Klasse berechnet für:

WWK_n in H und BS gesamt



RB 4.3.G

Multinomial Functional Tree (MFT)

Trainings- und Testdaten:

WWK_n in H und BS gesamt – „Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte Klasse an tatsächlicher Klasse berechnet für:

WWK_n in H und BS gesamt

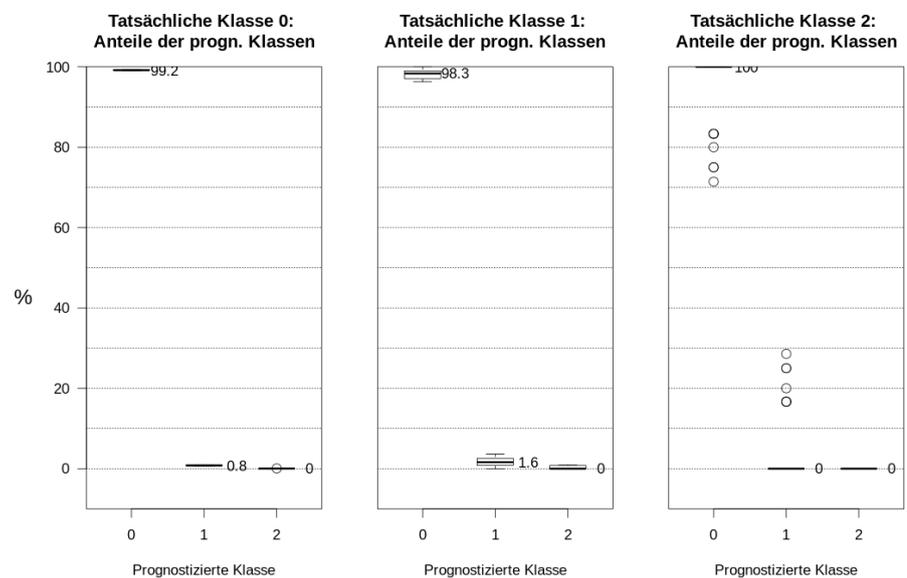


Abbildung RB.10b: Prognosequalität verschiedener raumbezogener Ansätze für Delikttyp 4 für H und BS zusammen

bei den übrigen Delikttypen wieder der MLR-Ansatz mit den sozioökonomischen Daten in die Einzelstadtauswertungen einbezogen.

Die Abbildung RB.11 zeigt die entsprechenden Boxplots für die Einzelstadtauswertung für Hannover, und zwar für den Benchmark-Ansatz, den MLR-Ansatz und den MFT-Ansatz, die letzteren beiden jeweils mit den Nexiga-Daten. Die Befunde ähneln denen in der Gesamtauswertung: In der tatsächlichen Klasse 0 hebt sich der MLR-Ansatz nur durch sichtbar höhere falsche Klasse-2-Prognosen von den beiden anderen Ansätzen ab, in der tatsächlichen Klasse 1 liefern jetzt der Benchmark-Ansatz und der MLR-Ansatz gleich viele korrekte Klasse-1-Prognosen und der MFT-Ansatz minimal weniger. Die beiden datengestützten Ansätze liefern in dieser Klasse als Anteils-Mediane wieder günstige 0,0 % der ambivalenten Klasse-2-Prognosen. In der tatsächlichen Klasse 2 schließlich liefern sie Anteils-Mediane von 0,0 % bei allen Nicht-0-Prognosen und bis auf die Klasse-1-Prognosen beim MLR-Ansatz auch eine Boxoberkante von 0,0 %. Auch hier treten beim MLR-Ansatz noch vereinzelt Anteile über 0 bei den Klasse-2-Prognosen auf, beim MFT-Ansatz nicht einmal die. Der Benchmark-Ansatz liefert ebenfalls keine einzige korrekte Klasse-2-Prognose, aber mit 10 % Nicht-0-Prognosen in Klasse 2 noch sichtbar

RB 4.0.H

Benchmark-Ansatz (Klasse wie in Vorwoche)

Trainings- und Testdaten:

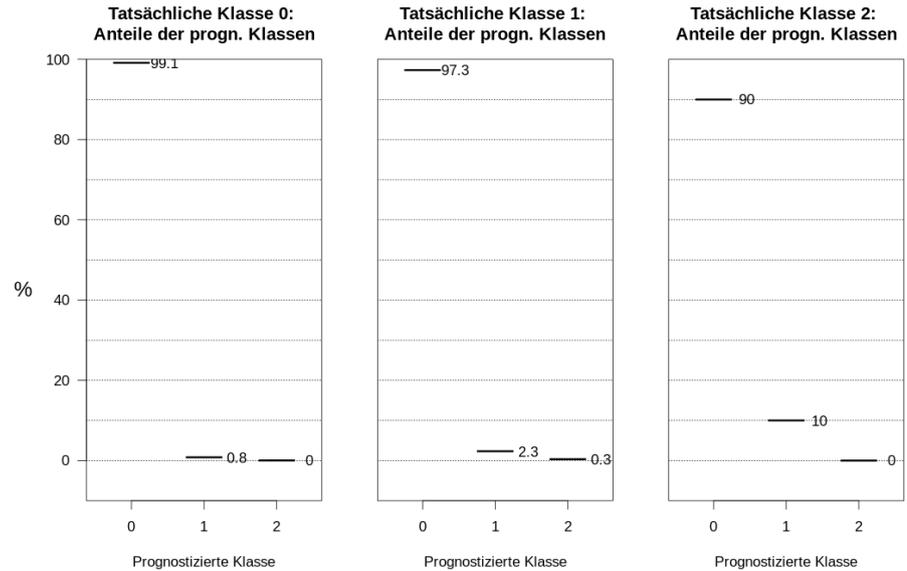
Nicht erforderlich

Nexiga-Daten einbezogen:

Nicht erforderlich

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in H



RB 4.2.H

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

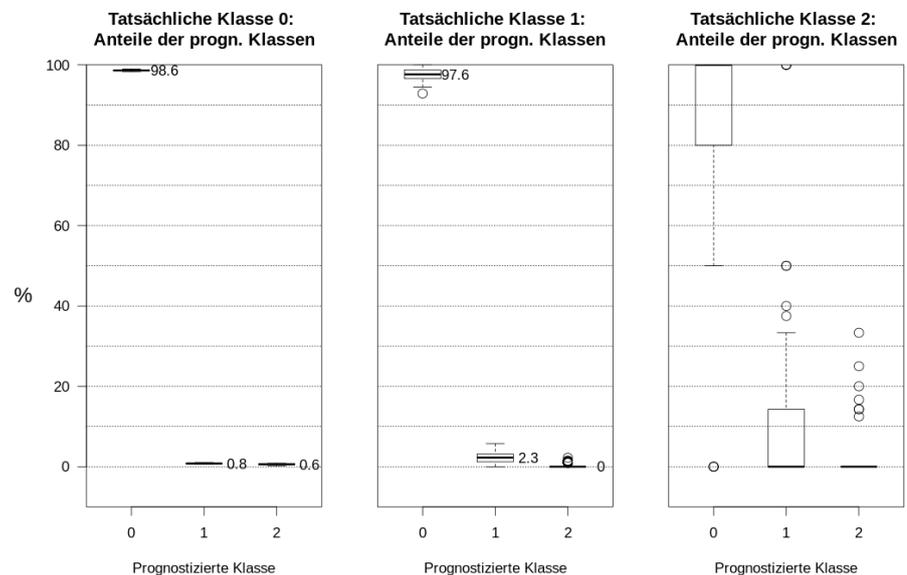
WWKn in H und BS gesamt –
„Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in H



RB 4.3.H

Multinomial Functional Tree (MFT)

Trainings- und Testdaten:

WWKn in H und BS gesamt –
„Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWKn in H

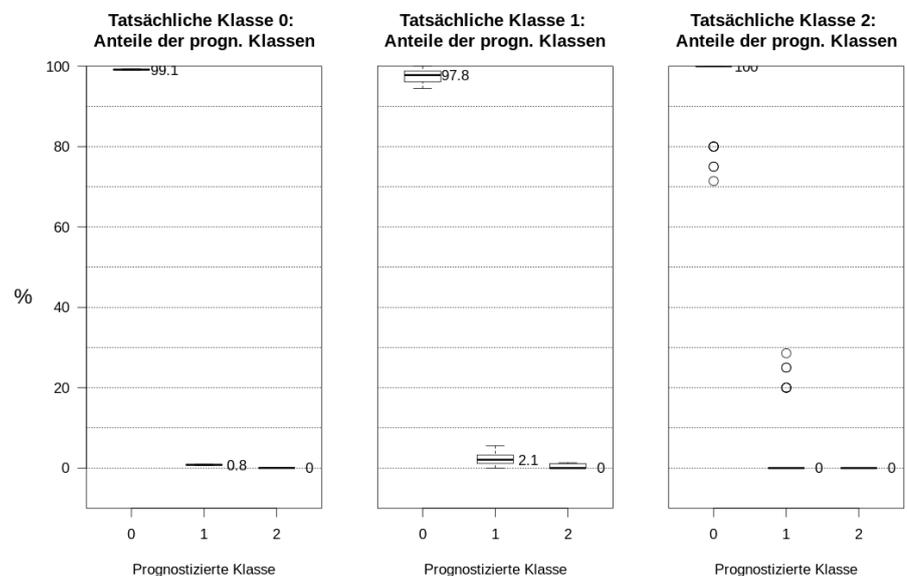


Abbildung RB.11: Prognosequalität verschiedener raumbezogener Ansätze für Delikttyp 4 für H alleine

RB 4.0.BS

Benchmark-Ansatz (Klasse wie in Vorwoche)

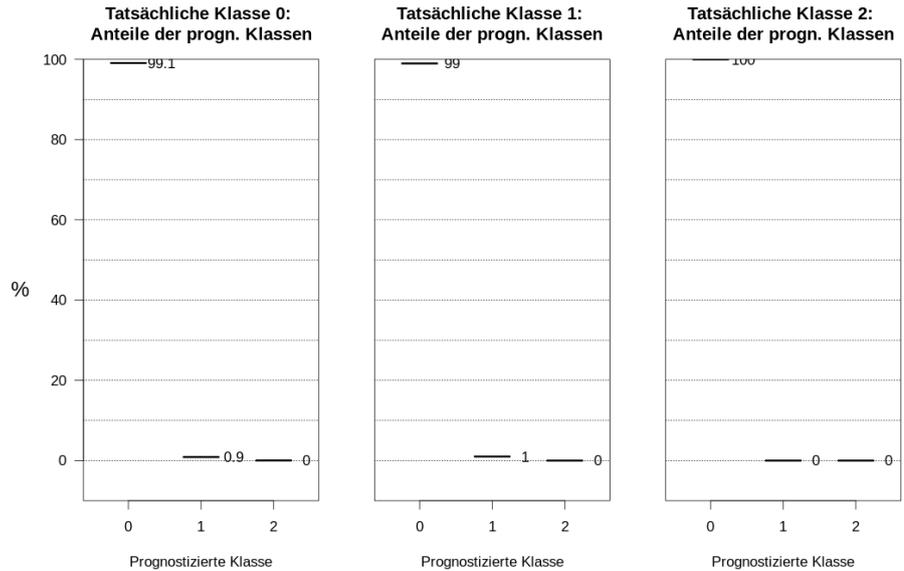
Trainings- und Testdaten:

Nicht erforderlich

Nexiga-Daten einbezogen:

Nicht erforderlich

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:



RB 4.2.BS

Multinomiale Logistische Regression (MLR)

Trainings- und Testdaten:

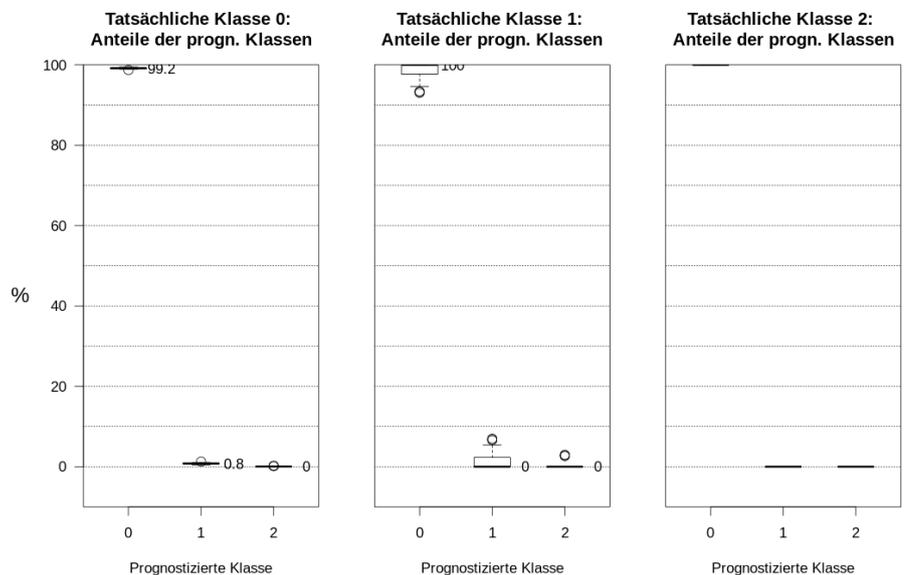
WWK_n in H und BS gesamt –
„Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWK_n in BS



RB 4.3.BS

Multinomial Functional Tree (MFT)

Trainings- und Testdaten:

WWK_n in H und BS gesamt –
„Gesamtmodell“

Nexiga-Daten einbezogen:

Ja

Anteile prognostizierte
Klasse an tatsächlicher
Klasse berechnet für:

WWK_n in BS

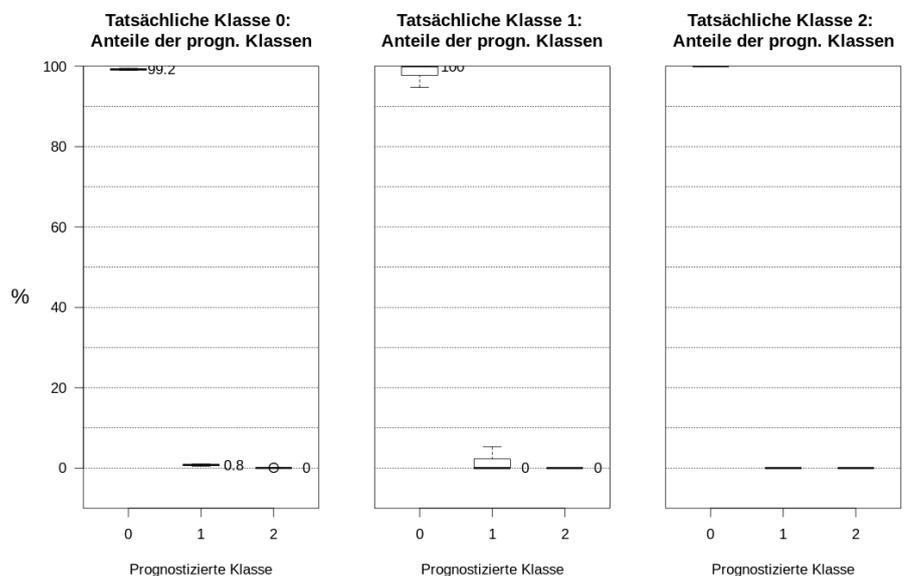


Abbildung RB.12: Prognosequalität verschiedener raumbezogener Ansätze für Delikttyp 4 für BS alleine

mehr zutreffenden Nicht-0-Prognosen als in der Gesamtbetrachtung. Aufgrund des hohen Anteils falscher Nicht-0-Prognosen schneidet auch hier der MLR-Ansatz beim PPV am schlechtesten ab, und der Benchmark-Ansatz liegt deutlich vorne.

Die Abbildung RB.12 zeigt die Einzelstadtbetrachtung für Braunschweig für die gleichen drei Ansätze. Hier kommen WWK_n, die tatsächlich zur Klasse 2 gehören, in den verwendeten Deliktdateien nur noch zwölfmal vor (also bei 0,03 % der 45.084 WWK_n in Braunschweig), dementsprechend dürften sie in vielen Testmengen gar nicht vorkommen. Tatsächlich liefert hier keiner der drei Ansätze einen Anteils-Median über 0 bei den Klasse-2-Prognosen, lediglich in der tatsächlichen Klasse 1 sieht man mehrere Ausreißer-Anteile (fetter Kringel, ein einzelner Ausreißer hätte einen dünneren Kringel) oberhalb der 0. Die falschen Nicht-0-Prognosen (also die in Klasse 0) sehen für alle drei Ansätze sehr ähnlich aus, in der tatsächlichen Klasse 2 sind hier erstmals alle drei Ansätze gleich schlecht, und in der tatsächlichen Klasse 1 bringt es selbst der Benchmark-Ansatz nur noch auf einen Anteil von 1,0 % korrekter Klasse-1-Prognosen. Bei den beiden datengestützten Ansätzen treten demzufolge hier erstmals PPV-Ergebnisse von 0,0 % auf, und der PPV des Benchmark-Ansatz beträgt hier auch nur noch 1,0 %.

Beim Delikttyp 4 „Diebstahl von KFZ“ erweist sich die Datenlage als sehr ungünstig für das Trainieren von datengestützten Modellen für 3-Klassen-Prognosen. Daher schneiden die entsprechenden Ansätze hier schlecht ab. Der Benchmark-Ansatz kommt in der Gesamtbetrachtung immerhin noch auf einen PPV von 2,5 % und in der Hannover-Betrachtung sogar auf einen PPV von 3,1 %, aber auch diese Werte sind sehr dürftig im Vergleich mit den Ergebnissen für die übrigen Delikttypen.

4.2.6 Fazit zu den raumbezogenen Prognosen

Die geprüften Prognoseansätze weisen für jeden Delikttyp drei relativ einheitliche Niveaus bei den PPV-Ergebnissen auf, also bei den Anteilen der zutreffenden an allen „Alarmen“ für ein oder mehrere zu erwartende Delikte (im Sinne von Nicht-Klasse-0-Prognosen für ein Wohnquartier in einer Woche). Das PPV-Niveau für Hannover liegt stets etwas über dem PPV-Niveau für beide Städte zusammen, das für Braunschweig stets deutlich niedriger. Nach aufsteigender so gemessener Qualität der Prognostizierbarkeit ergibt sich die Reihenfolge „Diebstahl von KFZ“, „WED / TWE“, „Schwerer Diebstahl aus Nichtwohnungen (ohne Ladendiebstahl)“ und „Diebstahl in / aus KFZ“. Es hat sich aber auch gezeigt, dass eine über die PPV-Betrachtung hinausgehende differenzierte Inaugenscheinnahme der Prognosequalität in den einzelnen tatsächlichen Klassen eine genauere Bewertung der einzelnen Ansätze ermöglicht.

Insgesamt schneidet immer ein „daten-ignorierender“ Prognoseansatz am besten ab, der lediglich die tatsächliche Häufigkeitsklasse jedes Wohnquartiers in einer Woche in die Folgeweche fortschreibt. Lediglich für einen Teil der für jeden datengestützten Alternativansatz untersuchten 100 Trainingsmengen (Boxoberkanten, obere Whiskers oder Ausreißer) wird dieser Benchmark-Ansatz im Hinblick auf korrekte Klasse-1- oder Klasse-2-Prognosen vereinzelt übertroffen, ebenso auch im Hinblick auf die „gröberen“ korrekten Nicht-0-Prognosen. Dabei zeigt sich auch gelegentlich, dass die Einbeziehung von sozioökonomischen Daten oder das Trainieren eines Einzelstadt-Modells die Prognosequalität – zumindest in Teilbereichen – erhöht, aber im Allgemeinen nie bis auf das Niveau des Benchmark-Ansatzes.

5. Zusammenfassung und Handlungsempfehlungen

Bislang lag der primäre Fokus im Projekt PreMAP darauf, einen ersten Ansatz inhaltlich wie technisch zu realisieren und eine entsprechende Aufbau- und Ablauforganisation zu beschreiben. Im Rahmen der erweiterten Pilotierung bot sich dem Teilprojekt Wissenschaft nunmehr die Gelegenheit, weitere Prognoseansätze für verschiedene Delikte unter Verwendung verschiedenster Einflussvariablen zu testen.

Maßgeblich bei der Bewertung von Prognoseansätzen ist vor allem das Zusammenspiel zwischen der Prognose und dem tatsächlich empirisch eintretenden Ereignis. Ideal sind jene Konstellationen, bei denen ein Ansatz ein Ereignis vorhersagt – eine Tat tritt ein oder nicht – und dann auch genau dieses Ereignis eintritt. Problematisch aus Ressourcensicht sind dagegen die sogenannten False Positives, also „falsche Alarmer“, bei denen der Ansatz eine Tat voraussagt, diese aber nicht geschieht. In dieser Konstellation würden Ressourcen zum Durchführen von Maßnahmen in das erkannte Risikogebiet entsandt werden, dort könnten sie aber die prognostizierte Tat nicht verhindern, da diese ja gar nicht eintritt. Sehr wohl würde in dem entsprechenden Gebiet aber immerhin Präsenz gezeigt. Aus Sicht der Bevölkerung vor allem problematisch sind schließlich die sogenannten False Negatives, also jene Fälle, bei denen der Ansatz kein Risiko prognostiziert, dann aber Straftaten zu verzeichnen sind. Das Gebiet wäre aufgrund der Vorhersage polizeilich nicht auf dem Radar und dadurch quasi unvorbereitet. Nachfolgende Tabelle zeigt die verschiedenen Konstellationen noch einmal:

TABELLE 2: KONSTELLATIONEN VON VORHERSAGEN UND TATSÄCHLICH EINTRETENDEN EREIGNISSEN

	Tat passiert	Tat passiert nicht
Tat vorhergesagt (Positives)	<i>True Positives (TP), „zutreffende Alarmer“</i>	<i>False Positives (FP) „falsche Alarmer“</i>
Tat nicht vorhergesagt (Negatives)	<i>False Negatives „falsche Entwarnungen“</i>	<i>True Negatives „zutreffende Entwarnungen“</i>

Empirisch sind bestimmte seltene Delikte, wie sie Gegenstand der Vorhersagen bei PreMAP waren, selbst in hochbelasteten Gebieten eher die Ausnahme als die Regel. Insofern weist die in der Tabelle rechte Spalte („Tat passiert nicht“) immer die meisten Fälle auf. Das hat regelmäßig zur Folge, dass selbst bei Prognoseansätzen mit einer günstigen False Positive Rate, also einem niedrigen Anteil der falschen Alarmer an der gesamten rechten Spalte als Qualitätskennzahl, immer noch eine hohe absolute Zahl von falschen Alarmen verbunden ist.

Doch selbst, wenn die Problematik der False Positives und False Negatives außer Acht gelassen wird, stellt sich die Frage, wie viele der durch den Prognoseansatz statistisch korrekt vorhergesagten Fälle denn tatsächlich verhindert werden können. Für Maßnahmen im Zuge von PreMAP stehen, wie der Evaluationsbericht der ersten Pilotierung zeigen konnte, nur sehr begrenzte Ressourcen im Einsatzgeschehen zur Verfügung, die nicht zuletzt ob der Größe der Risikogebiete kaum in der Lage sind, eine abschreckende Wirkung zu entfalten oder gar Täter / -innen auf frischer Tat zu stellen.

Grundsätzlich leiden alle hier untersuchten Prognoseansätze – auch der aktuelle PreMAP-Ansatz – darunter, dass für jede korrekte Vorhersage ein durchaus erkleckliches Maß an falschen Vorhersagen, in deren Rahmen dennoch polizeiliche Maßnahmen getroffen werden müssen, in Kauf genommen werden muss. Überdies kann nicht quantifiziert werden, wie viele der richtig

vorhergesagten Taten tatsächlich verhindert werden können. Eine Implementierung grundsätzlich neuer Prognoseansätze wird daher insbesondere vor dem Hintergrund zu erwartender Kosten (für jährlich zu erwerbende polizeiexterne Daten, für Programmierarbeit in der ZPD) nicht vorgeschlagen. Es sollten weiterhin tatbezogene Prognosen unter Nutzung eines Scores genutzt werden.

Entsprechende Near-Repeat-Prognosen kommen aber auch für die Delikttypen „Schwerer Diebstahl aus Nichtwohnungen (ohne Ladendiebstahl)“ und „Diebstahl in / aus KFZ“ analog zum „WED / TWE“ in Frage. Ferner könnte die Verwendung standardisierter NIVADIS-Daten für die Score-Berechnung für den Bereich „WED / TWE“ in Betracht gezogen werden. Die hierfür auf Seiten der ZPD entstehenden Aufwände für eine entsprechende Umsetzung dürften geringer sein als bei der Implementierung gänzlich neuer Prognoseansätze, wären aber freilich dennoch vorhanden und müssten dem nicht konkret in beispielsweise verhinderten Taten messbaren Nutzen gegenübergestellt werden.

Sollte eine Weiterentwicklung von PreMAP erfolgen, sollte zukünftig weniger auf das Prognoseelement gesetzt werden. Diesem sind, das haben die hier präsentierten Ergebnisse sowie die Ergebnisse der Evaluation gezeigt, Grenzen gesetzt. Vielmehr sollte die Komponente einer datengestützten Lagedarstellung gestärkt werden. Diese hatte von Beginn an den größten Zuspruch in der Praxis erfahren und sollte durch Einbeziehungen weiterer Daten (Spuren, Hinweise, Kontrollmeldungen, pp.) und Kartendarstellungen über die Landesgrenzen hinaus weiterentwickelt werden.

Variable	Datengestützter Prognoseansatz																											
	Ausw. TB 1.1	Ausw. TB 1.2	Ausw. TB 1.3	Ausw. TB 1.4	Ausw. TB 1.5	Ausw. TB 1.6	Ausw. TB 1.7	Ausw. TB 2.1	Ausw. TB 2.2	Ausw. TB 2.3	Ausw. TB 2.4	Ausw. TB 2.5	Ausw. TB 2.6	Ausw. TB 2.7	Ausw. TB 3.1	Ausw. TB 3.2	Ausw. TB 3.3	Ausw. TB 3.4	Ausw. TB 3.5	Ausw. TB 3.6	Ausw. TB 3.7	Ausw. TB 4.1	Ausw. TB 4.2	Ausw. TB 4.3	Ausw. TB 4.4	Ausw. TB 4.5	Ausw. TB 4.6	Ausw. TB 4.7
SG_ANDERER_GEGENSTAND	•																											
SG_ARBEITS_SONDERFAHRZEUG																												
SG_AUTOMAT_AUTOMATENTEIL																												
SG_BEHAELTER_PFLANZE_TIER_SONSTIGES																												
SG_BEHAELTNIS																												
SG_BEKLEIDUNG_STOFF_FELL_LEDER_PELZ																												
SG_BTM																												
SG_CHEM_ERZEUGNIS_KOSMETIK_PHYSIKAL_MITTEL																												
SG_DOKUMENT_URKUNDE																												
SG_DRUCKERZEUGNIS																												
SG_EINRICHTUNGSGEGENSTAND																												
SG_FAHRRAD																												
SG_FAHRZEUGZUBEHOER_ALLGEMEIN																												
SG_FAUSTFEUERWAFFE																												
SG_GERAET_MASCHINE_INSTRUMENTE																												
SG_HAUSHALTSARTIKEL																												
SG_HAUSHALTSGERAET																												
SG_HOHEITSZEICHEN																												
SG_KFZ_ANHAENGER																												
SG_KRAD																												
SG_KRAFTFAHRZEUG_FAHRZEUG																												
SG_KUNST_ANTIQUAET																												
SG_LKW																												
SG_LUFTVERKEHR																												
SG_MASKIERUNGSMITTEL																												
SG_METALL																												
SG_MUSIKINSTRUMENT																												
SG_NAHRUNGS_GENUSSMITTEL																												
SG_OEFF_FERNSPRECHER																												
SG_OPTISCHES_GERAET																												
SG_PFLANZE																												
SG_PKW																												
SG_PROPAGANDAMITTEL																												
SG_SAMMLERGEGENSTAND																												
SG_SCHIENENVERKEHR																												
SG_SCHMUCK_UHR_EDELSTEIN																												
SG_SICHERHEITSBELAETER																												
SG_SPIEL_SPORT_CAMPINGARTIKEL																												
SG_SPRENGMITTEL_BRENNBARE_FLUESSIGKEIT																												
SG_STEHLGUT_UNBEKANNT																												
SG_STEMPEL_SIEGEL_PLAKETTE																												
SG_TEPPICH_BRUECKE																												
SG_TIER																												
SG_UNTERHALTUNGSELEKTRONIK																												
SG_VERSANDGUT																												
SG_WAFFEN_ZUBEHOER_MUNITION																												
SG_WAFFE_DOKUMENT																												
SG_WASSERVERKEHR																												
SG_WERKZEUG_BAU_MATERIALIEN																												
SG_WERTDOKUMENT																												
SG_WOHNUNTERKUNFT																												
SG_ZAHLUNGSMITTEL																												
TM_AKKUSCHRAUBER																												
TM_AXT_BEIL																												
TM_BANKVERBINDUNGEN_MEDIEN_IT_GERAET_NACHRICHTENMITTEL																												
TM_BEHAELTER																												
TM_BEKLEIDUNG_STOFF_FELL_LEDER_PELZ																												
TM_BOHRMASCHINE_HAMMER																												
TM_BOHRWERKZEUG																												
TM_BOLZENSCHNEIDER																												
TM_BRECHSTANGE																												
TM_DIETRICH																												
TM_FAHRZEUGTEILE_ZUBEHOER																												
TM_GEISSFUSS																												
TM_GELD_WERTPAPIER_DOKUMENT																												
TM_GERAETE_KUNST_SCHMUCK_TEPPICH_METALL																												
TM_GERAET_MASCHINE																												
TM_GLASSCHNEIDER																												
TM_HACKE																												
TM_HAMMER																												
TM_HANDBOHRGERAET																												
TM_HARKE																												
TM_HEBELWERKZEUG																												
TM_MAGNET																												
TM_MASKIERUNGSMITTEL																												
TM_MEISSEL																												
TM_METALL																												

SG_...: Stehlgut, TM_...: Tatmittel

Variable	Datengestützter Prognoseansatz																											
	Ausw. TB 1.1	Ausw. TB 1.2	Ausw. TB 1.3	Ausw. TB 1.4	Ausw. TB 1.5	Ausw. TB 1.6	Ausw. TB 1.7	Ausw. TB 2.1	Ausw. TB 2.2	Ausw. TB 2.3	Ausw. TB 2.4	Ausw. TB 2.5	Ausw. TB 2.6	Ausw. TB 2.7	Ausw. TB 3.1	Ausw. TB 3.2	Ausw. TB 3.3	Ausw. TB 3.4	Ausw. TB 3.5	Ausw. TB 3.6	Ausw. TB 3.7	Ausw. TB 4.1	Ausw. TB 4.2	Ausw. TB 4.3	Ausw. TB 4.4	Ausw. TB 4.5	Ausw. TB 4.6	Ausw. TB 4.7
TM_MOEBEL	•																											
TM_PFLANZE																												
TM_SAEGE																												
TM_SAUERSTOFFLANZE																												
TM_SCHAUFEL																												
TM_SCHLUESSEL_NACHSCHLUESSEL																												
TM_SCHRAUBENDREHER	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
TM_SCHWEISSGERAET																												
TM_SONSTIGES_BOHRWERKZEUG																												
TM_SONSTIGES_BRECHWERKZEUG																												
TM_SONSTIGES_SCHLAGWERKZEUG																												
TM_SONSTIGES_SCHNEIDWERKZEUG																												
TM_SONSTIGES_TATMITTEL	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
TM_SONSTIGES_TRANSPORTWERKZEUG																												
TM_SONSTIGES_WERKZEUG	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
TM_SPATEN																												
TM_SPERRWERKZEUG																												
TM_SPORTGERAET_SPIELAUTOMAT																												
TM_SPRENG_BRANDMITTEL																												
TM_STECHWERKZEUG																												
TM_STEMMEISEN																												
TM_STRASSENVERKEHRSRAUM_FAHRZEUGE																												
TM_SUBSTANZEN																												
TM_TECHNISCHES_HILFSMITTEL																												
TM_TRENNSCHEIBE																												
TM_WAFFE	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
TM_WAFFE_DOKUMENT																												
TM_WERKZEUG_ALLGEMEIN	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
TM_WERKZEUG_ALLGEMEIN_A_G																												
TM_WERKZEUG_ALLGEMEIN_H_R																												
TM_WERKZEUG_ALLGEMEIN_S_Z																												
TM_WINKELSCHLEIFER																												
TM_WOHNUNTERKUNFT																												
TM_ZANGE																												
TM_ZIEH_FIX	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Nexiga-Daten																												
HHp1 (Anteil Haushalte mit 1 Person)				•	•	•																						
HHp2 (Anteil Haushalte mit 2 Personen)																												
HHpEK900 (Anteil Haushalte mit monatlichen Nettoeinkommen von 0 bis 900 EUR)																												
HHpEK1500 (Anteil Haushalte mit monatlichen Nettoeinkommen von 900 bis 1.500 EUR)																												
HHpEK2600 (Anteil Haushalte mit monatlichen Nettoeinkommen von 1.500 bis 2.600 EUR)																												
HHpEK3600 (Anteil Haushalte mit monatlichen Nettoeinkommen von 2.600 bis 3.600 EUR)																												
HHpEK5000 (Anteil Haushalte mit monatlichen Nettoeinkommen von 3.600 bis 5.000 EUR)																												
KK_EW (Kaufkraft je Einwohner in EUR)																												
PpEIN (Anteil Einpendler)																												
PpAUS (Anteil Auspendler)																												
MIETE (Durchschnittliche Angebotskaltmiete je qm)																												
PKW_GES (PKW, insgesamt)																												
PKW_GEW (PKW, gewerblich)																												
Baudichte (Anteil der bebauten Fläche in %)																												
Garagen (Anzahl)																												
EW_je_QKM (Einwohner je qkm)																												
HH_E_je_QKM (In [selbstgenutztem] Eigentum wohnende Haushalte je qkm)																												
ZZG_je_EW (Zuzüge [an Einwohnern] je Einwohner)																												
FZG_je_EW (Fortzüge [an Einwohnern] je Einwohner)																												
ALOp (Arbeitslose je Einwohner)																												
Grgn_je_HH (Anzahl Garagen je Haushalt)																												
Grgn_je_QKM (Anzahl Garagen je qkm)																												
Grgn_je_PKW (Anzahl Garagen je PKW)																												
PKW_je_HH (Anzahl PKW je Haushalt)																												
PKWg_je_HH (Anzahl gewerbliche PKW je Haushalt)																												
PKWg_je_QKM (Anzahl gewerbliche PKW je qkm)																												
ENTF_BAHN36: Routingdistanz Straßenabschnitt zum nächstgelegenen Bahnhof in Metern (ggf. Median der Gesamtstraße)																												
ENTF_ICE36: Routingdistanz Straßenabschnitt zum nächstgelegenen ICE-Bahnhof in Metern (ggf. Median der Gesamtstraße)																												
ENTF_FLUG3: Routingdistanz Straßenabschnitt zum nächstgelegenen Flughafen in Metern (ggf. Median der Gesamtstraße)																												
ENTF_BUSTR: Routingdistanz Straßenabschnitt zur nächstgelegenen Bundesstraße in Metern (ggf. Median der Gesamtstraße)																												
ENTF_BAB36: Routingdistanz Straßenabschnitt zum nächstgelegenen Autobahnanschluss in Metern (ggf. Median der Gesamtstraße)																												
ENTF_OEPNV: Routingdistanz Straßenabschnitt zur nächstgelegenen ÖPNV-Haltestelle in Metern (ggf. Median der Gesamtstraße)																												

A.2 Raumbezogene Prognoseansätze

Variable	Datengestützter Prognoseansatz																
	Ausw. RB 1.1	Ausw. RB 1.2	Ausw. RB 1.3	Ausw. RB 1.4	Ausw. RB 2.1	Ausw. RB 2.2	Ausw. RB 2.3	Ausw. RB 2.4	Ausw. RB 3.1	Ausw. RB 3.2	Ausw. RB 3.3	Ausw. RB 3.4	Ausw. RB 3.5	Ausw. RB 3.6	Ausw. RB 4.1	Ausw. RB 4.2	Ausw. RB 4.3
NIVADIS-Daten																	
Haeufigk_KI
NR_Anteil
Sum(NEAR_REPEAT)
Sum(BS_ALARMANLAGE)
Sum(BS_BALKON_TUER)
Sum(BS_EINFRIEDUNG)
Sum(BS_FEMSTERDICHTUNG)
Sum(BS_FENSTER)
Sum(BS_FENSTERBAND_DICHTUNG_GITTER)
Sum(BS_FENSTERRAHMEN)
Sum(BS_FENSTERSCHEIBE)
Sum(BS_FENSTER_BAND_DICHTUNG_RAHMEN)
Sum(BS_FENSTER_TUER_TERRASSENTUER_BALKONTUER)
Sum(BS_KEINE)
Sum(BS_KELLERROST)
Sum(BS_KFZ_AUSSENBEREICH)
Sum(BS_KFZ_GESAMT)
Sum(BS_KFZ_SCHEIBE)
Sum(BS_KFZ_SCHIEBEDACH_SONNENDACH_PLANE_VERDECK)
Sum(BS_KFZ_SCHLOSS)
Sum(BS_KFZ_SCHLUESSEL)
Sum(BS_KFZ_TUER_INCL_HECKKLAPPE)
Sum(BS_KLINGEL)
Sum(BS_LEERSTEHEND_UNBEWOHNT)
Sum(BS_ROLLADEN)
Sum(BS_SCHACHT)
Sum(BS_SCHLOSS)
Sum(BS_SCHLUESSEL)
Sum(BS_TERRASSENTUER)
Sum(BS_TUERBAND_TUERDICHTUNG_TUERGITTER)
Sum(BS_TUERRAHMEN)
Sum(BS_TUERSCHEIBE)
Sum(BS_TUER_TOR)
Sum(BS_TUER_TUERBAND_TUERDICHTUNG_TUERRAHMEN)
Sum(BS_WAND_DECKE_DACH_DACHABDECKUNG)
Sum(BW_ABREISSEN)
Sum(BW_ABSCHRAUBEN)
Sum(BW_ABWESENHEIT_AUSNUTZEN)
Sum(BW_AUFBOHREN)
Sum(BW_AUFSCHLITZEN)
Sum(BW_AUFSCHNEIDEN_ZERSCHNEIDEN)
Sum(BW_AUSBAUEN_ABBAUEN)
Sum(BW_AUSHAENGEN)
Sum(BW_BESONDERE_DELIKTSBEGEHUNG)
Sum(BW_BOHREN)
Sum(BW_DROHUNG)
Sum(BW_DURCHBRECHEN)
Sum(BW_DURCHKNEIFEN)
Sum(BW_DURCHSTECHEN)
Sum(BW_EINBRUCH_ZUR_TAGESZEIT)
Sum(BW_EINSCHLAGEN)
Sum(BW_EINSTEIGEN)
Sum(BW_EINWIRKEN)
Sum(BW_ENRFFERNEN)
Sum(BW_ERSTEIGEN)
Sum(BW_GELEGENHEIT_NUTZEN)
Sum(BW_GEWALT_GEGEN_PERSONEN_I_V_M_FAHRZEUG)
Sum(BW_GEWALT_GEGEN_PERSONEN_I_V_M_SEXUALMISSBRAUCH)
Sum(BW_GEWALT_GEGEN_PERSONEN_I_V_M_WAFFE)
Sum(BW_GEWALT_GEGEN_SACHEN)
Sum(BW_HEBELN_BRECHEN)
Sum(BW_HERAUSNEHMEN_BRECHEN_ZIEHEN)
Sum(BW_HOCHBIEGEN_SCHIEBEN_ZIEHEN_HEBEN)
Sum(BW_KEINE)
Sum(BW_OHNE_ERKENNBARE_AUFBRUCHSPUR_HAUS)
Sum(BW_OHNE_ERKENNBARE_AUFBRUCHSPUR_KFZ)
Sum(BW_POLENSCHLUESSEL_RUMAENENKNICK)
Sum(BW_SONSTIGE_GEWALT_GEGEN_PERSONEN)
Sum(BW_TATSICHERUNG)
Sum(BW_TATSICHERUNG_I_V_M_KFZ_KENNZEICHEN)
Sum(BW_TATSICHERUNG_I_V_M_SACHEN)
Sum(BW_TATSICHERUNG_I_V_M_TAETER)
Sum(BW_UEBERDREHEN)
Sum(BW_UEBERKLETTERN_UEBERWINDEN)
Sum(BW_UNPROFESSIONELLES_VERHALTEN)
Sum(BW_VERBEULEN)
Sum(BW_VERSTOPFEN)
Sum(BW_VERWENDEN)

Sum (...): Bei den 0-1-Variablen aus NIVADIS liefert die Summenbildung die Anzahl der „Ja“-Ausprägungen für alle Delikte zusammen für das jeweilige Wohnquartier.

BS_...: Begehungsstelle, BW_...: Begehungsweise

Variable	Datengestützter Prognoseansatz																
	Ausw. RB 1.1	Ausw. RB 1.2	Ausw. RB 1.3	Ausw. RB 1.4	Ausw. RB 2.1	Ausw. RB 2.2	Ausw. RB 2.3	Ausw. RB 2.4	Ausw. RB 3.1	Ausw. RB 3.2	Ausw. RB 3.3	Ausw. RB 3.4	Ausw. RB 3.5	Ausw. RB 3.6	Ausw. RB 4.1	Ausw. RB 4.2	Ausw. RB 4.3
Sum(SG_ANDERER_GEGENSTAND)
Sum(SG_ARBEITS_SONDERFAHRZEUG)
Sum(SG_AUTOMAT_AUTOMATENTEIL)
Sum(SG_BEHALTER_PFLANZE_TIER_SONSTIGES)
Sum(SG_BEHALTNIS)
Sum(SG_BEKLEIDUNG_STOFF_FELL_LEDER_PELZ)
Sum(SG_BTM)
Sum(SG_CHEM_ERZEUGNIS_KOSMETIK_PHYSIKAL_MITTEL)
Sum(SG_DOKUMENT_URKUNDE)
Sum(SG_DRUCKERZEUGNIS)
Sum(SG_EINRICHTUNGSGEGENSTAND)
Sum(SG_FAHRRAD)
Sum(SG_FAHRZEUGZUBEHOER_ALLGEMEIN)
Sum(SG_FAUSTFEUERWAFFE)
Sum(SG_GERAET_MASCHINE_INSTRUMENTE)
Sum(SG_HAUSHALTSARTIKEL)
Sum(SG_HAUSHALTSGERAET)
Sum(SG_HOHEITSZEICHEN)
Sum(SG_KFZ_ANHAENGER)
Sum(SG_KRAD)
Sum(SG_KRAFTFAHRZEUG_FAHRZEUG)
Sum(SG_KUNST_ANTIQUITAET)
Sum(SG_LKW)
Sum(SG_LUFTVERKEHR)
Sum(SG_MASKIERUNGSMITTEL)
Sum(SG_METALL)
Sum(SG_MUSIKINSTRUMENT)
Sum(SG_NAHRUNGS_GENUSSMITTEL)
Sum(SG_OFF_FERNSPRECHER)
Sum(SG_OPTISCHES_GERAET)
Sum(SG_PFLANZE)
Sum(SG_PKW)
Sum(SG_PROPAGANDAMITTEL)
Sum(SG_SAMMLERGEGENSTAND)
Sum(SG_SCHIENENVERKEHR)
Sum(SG_SCHMUCK_UHR_EDELSTEIN)
Sum(SG_SICHERHEITSBEHALTER)
Sum(SG_SPIEL_SPORT_CAMPINGARTIKEL)
Sum(SG_SPRENGMITTEL_BRENNBARE_FLUESSIGKEIT)
Sum(SG_STEHLGUT_UNBEKANNT)
Sum(SG_STEMPEL_SIEGEL_PLAKETTE)
Sum(SG_TEPPICH_BRUECKE)
Sum(SG_TIER)
Sum(SG_UNTERHALTUNGSELEKTRONIK)
Sum(SG_VERSANDGUT)
Sum(SG_WAFFEN_ZUBEHOER_MUNITION)
Sum(SG_WAFFE_DOKUMENT)
Sum(SG_WASSERVERKEHR)
Sum(SG_WERKZEUG_BAU_MATERIALIEN)
Sum(SG_WERTDOKUMENT)
Sum(SG_WOHNUNTERKUNFT)
Sum(SG_ZAHLUNGSMITTEL)
Sum(TM_AKKUSCHRAUBER)
Sum(TM_AXT_BEIL)
Sum(TM_BANKVERBINDUNGEN_MEDIEN_IT_GERAET_NACHRICHTENMITTEL)
Sum(TM_BEHALTER)
Sum(TM_BEKLEIDUNG_STOFF_FELL_LEDER_PELZ)
Sum(TM_BOHRMASCHINE_HAMMER)
Sum(TM_BOHRWERKZEUG)
Sum(TM_BOLZENSCHNEIDER)
Sum(TM_BRECHSTANGE)
Sum(TM_DIETRICH)
Sum(TM_FAHRZEUGTEILE_ZUBEHOER)
Sum(TM_GEISSFUSS)
Sum(TM_GELD_WERTPAPIER_DOKUMENT)
Sum(TM_GERAETE_KUNST_SCHMUCK_TEPPICH_METALL)
Sum(TM_GERAET_MASCHINE)
Sum(TM_GLASSCHNEIDER)
Sum(TM_HACKE)
Sum(TM_HAMMER)
Sum(TM_HANDBOHRGERAET)
Sum(TM_HARKE)
Sum(TM_HEBELWERKZEUG)
Sum(TM_MAGNET)
Sum(TM_MASKIERUNGSMITTEL)
Sum(TM_MEISSEL)
Sum(TM_METALL)

SG_...: Stehlgut, TM_...: Tatmittel

Variable	Datengestützter Prognoseansatz															
	Ausw. RB 1.1	Ausw. RB 1.2	Ausw. RB 1.3	Ausw. RB 1.4	Ausw. RB 2.1	Ausw. RB 2.2	Ausw. RB 2.3	Ausw. RB 2.4	Ausw. RB 3.1	Ausw. RB 3.2	Ausw. RB 3.3	Ausw. RB 3.4	Ausw. RB 3.5	Ausw. RB 4.1	Ausw. RB 4.2	Ausw. RB 4.3
Sum(TM_MOEBEL)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_PFLANZE)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SAEGE)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SCHAUFEL)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SCHLUESSEL_NACHSCHLUESSEL)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SCHRAUBENDREHER)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SCHWEISSGERAET)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SONSTIGES_BOHRWERKZEUG)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SONSTIGES_BRECHWERKZEUG)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SONSTIGES_SCHLAGWERKZEUG)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SONSTIGES_SCHNEIDWERKZEUG)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SONSTIGES_TATMITTEL)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SONSTIGES_TRANSPORTWERKZEUG)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SONSTIGES_WERKZEUG)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SPATEN)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SPERRWERKZEUG)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SPORTGERAET_SPIELAUTOMAT)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SPRENG_BRANDMITTEL)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_STECHWERKZEUG)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_STEMMEISEN)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_STRASSENVERKEHRSRAUM_FAHRZEUGE)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_SUBSTANZEN)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_TECHNISCHES_HILFSMITTEL)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_TRENNSCHEIBE)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_WAFFE)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_WAFFE_DOKUMENT)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_WERKZEUG_ALLGEMEIN)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_WERKZEUG_ALLGEMEIN_A_G)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_WERKZEUG_ALLGEMEIN_H_R)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_WERKZEUG_ALLGEMEIN_S_Z)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_WINKELSCHLEIFER)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_WOHNUNTERKUNFT)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_ZANGE)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Sum(TM_ZIEH_FIX)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Nexiga-Daten																
HHp1 (Anteil Haushalte mit 1 Person)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
HHp2 (Anteil Haushalte mit 2 Personen)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
HHpEK900 (Anteil Haushalte mit monatlichen Nettoeinkommen von 0 bis 900 EUR)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
HHpEK1500 (Anteil Haushalte mit monatlichen Nettoeinkommen von 900 bis 1.500 EUR)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
HHpEK3600 (Anteil Haushalte mit monatlichen Nettoeinkommen von 2.600 bis 3.600 EUR)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
HHpEK5000 (Anteil Haushalte mit monatlichen Nettoeinkommen von 3.600 bis 5.000 EUR)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
HHpEK5000M (Anteil Haushalte mit monatlichen Nettoeinkommen von 5.000 EUR u.m.)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
KK_EW (Kaufkraft je Einwohner in EUR)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
PpEIN (Anteil Einpendler)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
PpAUS (Anteil Auspendler)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
MIETE (Durchschnittliche Angebotskaltmiete je qm)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
PKW_GES (PKW, insgesamt)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
PKW_GEW (PKW, gewerblich)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Baudichte (Anteil der bebauten Fläche in %)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Garagen (Anzahl)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
EW_je_QKM (Einwohner je qkm)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
HH_E_je_QKM (In [selbstgenutztem] Eigentum wohnende Haushalte je qkm)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
ZZG_je_EW (Zuzüge [an Einwohnern] je Einwohner)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
FZG_je_EW (Fortzüge [an Einwohnern] je Einwohner)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
ALOp (Arbeitslose je Einwohner)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Grgn_je_HH (Anzahl Garagen je Haushalt)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Grgn_je_QKM (Anzahl Garagen je qkm)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
PKW_je_HH (Anzahl PKW je Haushalt)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
PKWg_je_HH (Anzahl gewerbliche PKW je Haushalt)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
PKWg_je_QKM (Anzahl gewerbliche PKW je qkm)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Median(ENTF_BAHN36): Routingdistanz Straßenabschnitt zum nächstgelegenen Bahnhof in Metern	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Median(ENTF_ICE36): Routingdistanz Straßenabschnitt zum nächstgelegenen ICE-Bahnhof in Metern	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Median(ENTF_FLUG3): Routingdistanz Straßenabschnitt zum nächstgelegenen Flughafen in Metern	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Median(ENTF_BUSTR): Routingdistanz Straßenabschnitt zur nächstgelegenen Bundesstraße in Metern	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Median(ENTF_BAB36): Routingdistanz Straßenabschnitt zum nächstgelegenen Autobahnanschluss in Metern	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Median(ENTF_OEPNV): Routingdistanz Straßenabschnitt zur nächstgelegenen ÖPNV-Haltestelle in Metern	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Median(LCHH): Anzahl der Haushalte je Haus	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Median(LCKREI): Affinität für Urlaubsreisen (Skala von 1 bis 9)	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

Median (...): Die Variable ... liegt unterhalb der Wohnquartierebene vor. Um sie für die raumbezogene Prognose nutzen zu können, wird sie durch Medianbildung für das jeweilige Wohnquartier aggregiert.

Anhang B: Evaluationsblätter zu den tatbezogenen Prognoseansätzen

TB 1.1.G Delikttyp 1: WED / TWE (Deliktschlüssel 435 und 436)

Prognoseansatz: Merkmalspezifische Score-Beiträge (NRE, aktueller PreMAP-Ansatz)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	831	60	891
Hannover, Landeshauptstadt	3.359	419	3.778
Gesamt	4.190	479	4.669

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	H u. BS gesamt ³⁴

Gesamtmodell, Gesamtauswertung

Nexiga-Daten einbezogen:	Nein (konzeptionell nur eingeschränkt möglich)
Korrelationsschwelle:	0,8
Score-Schwelle (keine Wahrscheinlichkeit):	0,155

True-Positive-Rate (Testdaten)

False-Positive-Rate (Testdaten)

Charakteristische Beobachtungen:

- ➔ Mediane: 13,8 % (TPR) etwas höher als 11,4 % (FPR), das entspricht aber nur 66 True Positives (TP, „Zutreffende Alarme“) versus 477 False Positives (FP, „Falsche Alarme“), resultierender PPV siehe unten
- ➔ TPR-Boxunterkante: Für drei Viertel der 100 Trainingsmengen liegt die TPR in der Testmenge bei 9,5 % oder höher
- ➔ FPR-Boxoberkante: Für drei Viertel der 100 Trainingsmengen liegt die FPR in der Testmenge bei 15,7 % oder niedriger
- ➔ Untere „Whiskers“ (gestrichelte Linien mit Querstrich am Ende): Sowohl bei TPR als auch bei FPR kommt auch der Wert 0 vor (ungünstig bei TPR, unproblematisch bei FPR)
- ➔ Obere Whiskers bzw. obere Ausreißer: Im höchsten Viertel der TPR-Werte kommt ein Maximalwert von ca. 31 % vor („schön aber selten“), im höchsten Viertel der FPR-Werte ein Maximalwert von ca. 29 % („unschön aber selten“)
- ➔ Prozentanteil der zutreffenden an allen Alarmen³⁵ (PPV, Medianbetrachtung) = $TP / (TP + FP) * 100 = 66 / (66 + 477) * 100 = 12,2 \%$

Fazit: TPR-Verteilung liegt nur etwas höher als die FPR-Verteilung, das bedeutet nur mäßig gute Prognosequalität, was sich auch beim PPV niederschlägt.

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 1.2.G (gleicher Ansatz mit standardisierten Input-Variablen, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Charakteristische Beobachtungen“.
- TB 1.n.G für n = 3, 4, 7 (alternative Ansätze, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Charakteristische Beobachtungen“ bzw. unter „Bes. rel. Ansatzvarianten zum Vergleich“.

³⁴ Die TPR-Werte werden als Prozentanteil der TP in der jeweiligen Testmenge an den tatsächlichen Delikten mit NR in der Testmenge berechnet und die FPR-Werte entsprechend. Die TP- und FP-Anzahl, die in Klammern unter den Medianen angegeben sind, werden rein rechnerisch durch Multiplikation des jeweiligen Medians (Dezimalbruch statt Prozentwert, z. B. 0,99 statt 99 %) mit der Mit-NR-Anzahl bzw. mit der Ohne-NR-Anzahl im gesamten Datenbestand (nicht in der Testmenge!) ermittelt. Sie geben Näherungswerte für die TP- und die FP-Anzahl an, die bei einer NR-Ja/Nein-Prognose für die Gesamtheit von Trainings- und Testmenge durch das (fiktiv mit anderen Daten) fertig trainierte „Median-Modell“ zu erwarten wären.

³⁵ Andere übliche, aber nicht selbsterklärende Bezeichnungen für diesen Anteil: Genauigkeit, positiver Vorhersagewert, positiver prädiktiver Wert, englisch: precision, positive predictive value (PPV, Abkürzung hier verwendet)

TB 1.2.G Delikttyp 1: WED / TWE (Deliktschlüssel 435 und 436)

Prognoseansatz: Merkmalspezifische Score-Beiträge mit standardisierten Input-Variablen

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	831	60	891
Hannover, Landeshauptstadt	3.359	419	3.778
Gesamt	4.190	479	4.669

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	H u. BS gesamt

Gesamtmodell, Gesamtauswertung

Nexiga-Daten einbezogen:	Nein (konzeptionell nur eingeschränkt möglich)
Korrelationsschwelle:	0,8
Score-Schwelle (keine Wahrscheinlichkeit):	0,65

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- ➔ TPR-Median höher, FPR-Median niedriger als bei TB 1.1.G, d. h. insgesamt besser im Hinblick auf die Mediane
- ➔ ebenfalls bessere TPR-Boxunterkante (höher als bei TB 1.1.G) und FPR-Boxoberkante (niedriger als dort)
- ➔ FPR: Box und Whiskers wenig ausgedehnt, d. h. nur geringe FPR-Schwankungen, also ebenfalls besser als bei TB 1.1.G
- ➔ untere Whiskers nicht bis zur 0, d. h. TPR und FPR für alle Trainingsmenge größer als 0
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 13,6 \%$
(etwas höher, also besser als bei TB 1.1.G)

Fazit: Bei diesem Delikttyp kann ggü. dem aktuellen PreMAP-Ansatz eine Verbesserung der Prognosequalität (TPR-Verteilung sichtbar oberhalb der FPR-Verteilung, etwas höherer PPV) durch Verwendung der standardisierten Input-Variablen erreicht werden.

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 1.2.H und TB 1.2.BS (gleicher Ansatz, TPR/FPR-Auswertung für regionale Teil-Testmengen – nur H bzw. nur BS – der jeweiligen Testmenge): Siehe dort unter „Charakteristische Beobachtungen“.
- TB 1.1.G für n = 1, 3, 4, 7 (alternative Ansätze, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- TB 1.n.G für n = 3, 4, 7 (alternative Ansätze, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Charakteristische Beobachtungen“.

TB 1.2.H Delikttyp 1: WED / TWE (Deliktschlüssel 435 und 436)

Prognoseansatz: Merkmalspezifische Score-Beiträge mit standardisierten Input-Variablen

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	831	60	891
Hannover, Landeshauptstadt	3.359	419	3.778
Gesamt	4.190	479	4.669

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	H ³⁶

Gesamtmodell, Hannover-Auswertung

Nexiga-Daten einbezogen:	Nein (konzeptionell nur eingeschränkt möglich)
Korrelationsschwelle:	0,8
Score-Schwelle (keine Wahrscheinlichkeit):	0,65

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- ➔ insgesamt etwas ungünstiger als bei TB 1.2.G (beide Mediane, TPR-Boxunterkante, unterer TPR-Whisker): d. h., der Prognoseansatz TB 1.2 funktioniert in H etwas schlechter als in H und BS zusammen
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 13,1 \%$
(minimal schlechter als bei TB 1.2.G)

Fazit: Für Hannover alleine funktioniert der Ansatz TB 1.2 schlechter als für H und BS zusammen.

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 1.2.G (Gleicher Ansatz, TPR/FPR-Auswertung für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- TB 1.n.H für n = 4, 5 (alternativer Ansatz mit Einbeziehung der Nexiga-Daten, Training des Modells mit Gesamtdaten bzw. nur mit Hannover-Daten in TB 1.5.H, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für H): Siehe dort unter „Charakteristische Beobachtungen“ bzw. unter „Bes. rel. Ansatzvarianten zum Vergleich“.

³⁶ Diesem Auswertungsblatt liegen dieselben Prognoseergebnisse für dieselben 100 Testmengen wie bei TB 1.2.G zugrunde, aber nur die Teil-Testmenge für Hannover wird ausgewertet. D. h., die TPR-Werte werden hier als Prozentanteil der TP-Anzahl für Hannover an den tatsächlichen Delikten mit NR für Hannover (in der Testmenge!) berechnet und die FPR-Werte entsprechend. Die TP- und FP-Anzahl, die in Klammern unter den Medianen angegeben sind, werden rein rechnerisch aus dem jeweiligen Median und der Mit-NR-Anzahl bzw. Ohne-NR-Anzahl für Hannover im gesamten Datenbestand (nicht in der Testmenge!) ermittelt.

TB 1.2.BS Delikttyp 1: WED / TWE (Deliktschlüssel 435 und 436)

Prognoseansatz: Merkmalspezifische Score-Beiträge mit standardisierten Input-Variablen

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	831	60	891
Hannover, Landeshauptstadt	3.359	419	3.778
Gesamt	4.190	479	4.669

Trainings- und Testdaten: H u. BS gesamt
 TPR, FPR berechnet für: BS³⁷

Gesamtmodell, Braunschweig-Auswertung

Nexiga-Daten einbezogen: Nein (konzeptionell nur eingeschränkt möglich)
 Korrelationsschwelle: 0,8
 Score-Schwelle (keine Wahrscheinlichkeit): 0,65

True-Positive-Rate (Testdaten)

False-Positive-Rate (Testdaten)

Charakteristische Beobachtungen:

- ➔ Beide Mediane günstiger als bei TB 1.2.G (d. h., „im Mittel“, also im Hinblick auf die Mediane funktioniert der Prognoseansatz TB 1.2 in BS etwas besser als in H und BS zusammen)
- ➔ Ausdehnung von Boxen und Whiskers: größere Streuung bei TPR (auffallend größer, nach oben unproblematisch) und FPR (etwas größer, nach unten unproblematisch) als bei TB 1.2.G und TB 1.2.H
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 12,5 \%$ (etwas schlechter als bei TB 1.2.G)

Fazit: Für Braunschweig alleine funktioniert der Ansatz TB 1.2 im Hinblick auf die Verteilungen von TPR und FPR auffallend besser als für H und BS zusammen (wenn auch mit einer größeren Streuung bei den TPR-Werten), aber weil es so wenige tatsächliche NR-Delikte in BS gibt, resultiert daraus kein besserer PPV.

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 1.2.G (Gleicher Ansatz, TPR/FPR-Auswertung für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- TB 1.n.BS für n = 4, 6 (alternativer Ansatz mit Einbeziehung der Nexiga-Daten, Training des Modells mit Gesamtdaten bzw. nur mit Braunschweig-Daten in TB 1.6.BS, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS): Siehe dort unter „Charakteristische Beobachtungen“ bzw. unter „Bes. rel. Ansatzvarianten zum Vergleich“.

³⁷ Diesem Auswertungsblatt liegen dieselben Prognoseergebnisse für dieselben 100 Testmengen wie bei TB 1.2.G zugrunde, aber nur die Teil-Testmenge für Braunschweig wird ausgewertet. D. h., die TPR-Werte werden hier als Prozentanteil der TP-Anzahl für Braunschweig an den tatsächlichen Delikten mit NR für Braunschweig (in der Testmenge!) berechnet und die FPR-Werte entsprechend. Die TP- und FP-Anzahl, die in Klammern unter den Medianen angegeben sind, werden rein rechnerisch aus dem jeweiligen Median und der Mit-NR-Anzahl bzw. Ohne-NR-Anzahl für Braunschweig im gesamten Datenbestand (nicht in der Testmenge!) ermittelt.

TB 1.3.G Delikttyp 1: WED / TWE (Deliktschlüssel 435 und 436)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	831	60	891
Hannover, Landeshauptstadt	3.359	419	3.778
Gesamt	4.190	479	4.669

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	H u. BS gesamt

Gesamtmodell, Gesamtauswertung

Nexiga-Daten einbezogen:	Nein
Korrelationsschwelle:	0,8
Wahrscheinlichkeitsschwelle:	0,95

**True-Positive-Rate
(Testdaten)**

10.4 (62 TP)

**False-Positive-Rate
(Testdaten)**

11.5 (437 FP)

Charakteristische Beobachtungen:

- ➔ Mediane: eindeutig schlechter als bei TB 1.2.G (TPR-Median niedriger, FPR-Median höher); Vergleich mit TB 1.1.G uneindeutig (beide niedriger als dort)
- ➔ Aber: TPR-Boxunterkante sowie unteres TPR-Whisker-Ende höher und FPR-Boxoberkante niedriger als bei TB 1.2.G, FPR-Streuung insgesamt ähnlich wie dort
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 12,4 \%$
(sichtbar schlechter als bei TB 1.2.G, minimal besser als bei 1.3.G)

Fazit: Durch den BLR-Ansatz ohne Einbeziehung der Nexiga-Daten kann der Ansatz TB 1.1 hinsichtlich des PPV minimal übertroffen werden, aber nicht in gleichem Maße wie durch die Einbeziehung standardisierter Inputvariablen (Ansatz TB 1.2).

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 1.n.G für n = 1, 2 (alternative Ansätze, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- TB 1.4.G (gleicher Ansatz mit Einbeziehung der Nexiga-Daten, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Bes. rel. Ansatzvarianten zum Vergleich“.
- TB 1.7.G (alternativer Ansatz mit Einbeziehung der Nexiga-Daten, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Charakteristische Beobachtungen“.

TB 1.4.G Delikttyp 1: WED / TWE (Deliktschlüssel 435 und 436)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	831	60	891
Hannover, Landeshauptstadt	3.359	419	3.778
Gesamt	4.190	479	4.669

Trainings- und Testdaten: H u. BS gesamt
 TPR, FPR berechnet für: H u. BS gesamt

Gesamtmodell, Gesamtauswertung

Nexiga-Daten einbezogen: Ja
 Korrelationsschwelle: 0,8
 Wahrscheinlichkeitsschwelle: 0,975

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- ➔ Mediane: besser als bei TB 1.2.G (also beim besten Ansatz ohne Nexiga-Daten), Abstand größer als bei allen anderen „TB 1.G“-Auswertungen
- ➔ TPR-Boxunterkante, unterer TPR-Whisker, FPR-Boxoberkante: schlechter als bei TB 1.2.G
- ➔ Boxhöhe bei beiden, Whiskers bei TPR: Streuung größer als bei TB 1.2.G
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 14,4 \%$
(etwas besser als bei TB 1.2.G)

Fazit: Durch den BLR-Ansatz mit Einbeziehung der Nexiga-Daten kann für beide Städte zusammen der Ansatz TB 2.1 (und damit auch der Ansatz TB 1.1) hinsichtlich der Mediane für TPR und FPR und hinsichtlich des PPV übertroffen werden. Allerdings sind die TPR-Boxunterkante und die FPR-Boxoberkante ungünstiger als bei TB 2.1.G

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 1.4.H und TB 1.4.BS (gleicher Ansatz, TPR/FPR-Auswertung für regionale Teil-Testmengen – nur H bzw. nur BS – der jeweiligen Testmenge): Siehe dort unter „Charakteristische Beobachtungen“.
- TB 1.n.G für n = 1, 2 (alternative Ansätze, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- TB 1.3.G (gleicher Ansatz ohne Einbeziehung der Nexiga-Daten, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): TPR- und FPR-Median besser als dort (also auch Abstand der Mediane größer), aber TPR- Boxunterkante und -Minimum sowie FPR- Boxoberkante und -Maximum schlechter als dort
- TB 1.7.G (alternativer Ansatz mit Einbeziehung der Nexiga-Daten, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Charakteristische Beobachtungen“.

TB 1.4.H Delikttyp 1: WED / TWE (Deliktschlüssel 435 und 436)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	831	60	891
Hannover, Landeshauptstadt	3.359	419	3.778
Gesamt	4.190	479	4.669

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	H

Gesamtmodell, Hannover-Auswertung

Nexiga-Daten einbezogen:	Ja
Korrelationsschwelle:	0,8
Wahrscheinlichkeitsschwelle:	0,975

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- ➔ Mediane: TPR sichtbar höher als bei TB 1.2.H, aber FPR auch höher als dort (damit besteht Spielraum, durch Anheben der Wahrscheinlichkeitsschwelle auch im Hinblick auf die FPR eine Verbesserung zu erzielen; wegen zeitlich begrenzter Datenverfügbarkeit nicht geprüft)
- ➔ TPR-Boxunterkante: schlechter als bei TB 1.2.H
- ➔ FPR-Boxhöhe, FPR-Whiskers: Streuung größer als bei TB 1.2.H (ungünstig, insb. nach oben)
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 13,0 \%$ (minimal schlechter als bei TB 1.2.H)

Fazit: Für Hannover alleine funktioniert der Ansatz TB 1.4 in der hier untersuchten Form im Hinblick auf den PPV minimal schlechter als der Ansatz TB 1.2 (bei Uneindeutigkeit hinsichtlich der Verteilungen von TPR und FPR), aber durch eine Anpassung der Wahrscheinlichkeitsschwelle wäre eventuell eine Verbesserung bei den Verteilungen von TPR und FPR und damit auch beim PPV ggü. TB 1.2 möglich.

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 1.4.G (Gleicher Ansatz, TPR/FPR-Auswertung für die Gesamt-Testmengen – H und BS zusammen): beide Mediane schlechter als dort, TPR-Box liegt etwas niedriger, FPR-Box etwas höher als dort (beides schlechter), PPV etwas schlechter als dort
- TB 1.2.H (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für H): Siehe oben.
- TB 1.5.H (gleicher Ansatz ausschließlich mit Hannover-Daten, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für H – dort identisch mit den Gesamt-Testmengen): Siehe dort unter „Charakteristische Beobachtungen“.

TB 1.4.BS Delikttyp 1: WED / TWE (Deliktschlüssel 435 und 436)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	831	60	891
Hannover, Landeshauptstadt	3.359	419	3.778
Gesamt	4.190	479	4.669

Trainings- und Testdaten: H u. BS gesamt
 TPR, FPR berechnet für: BS

Gesamtmodell, Braunschweig-Auswertung

Nexiga-Daten einbezogen: Ja
 Korrelationsschwelle: 0,8
 Wahrscheinlichkeitsschwelle: 0,975

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- ➔ Mediane: TPR schlechter als bei TB 1.2.BS, aber FPR auffallend besser als dort (damit besteht Spielraum, durch Absenken der Wahrscheinlichkeitsschwelle auch im Hinblick auf die TPR eine Verbesserung zu erzielen; wegen zeitlich begrenzter Datenverfügbarkeit nicht geprüft)
- ➔ TPR-Boxunterkante ist 0: für ein Viertel der Testmengen oder mehr (!) ist die TPR 0 (da 0 der kleinste mögliche Wert ist, sind Werte „unter der Box“ und Werte „auf dem unteren Rand der Box“ nicht zu unterscheiden) – denkbare Erklärung: da die 100 Testmengen als Zufallsstichprobe aus den Gesamtdaten (H und BS) gezogen werden, sind möglicherweise in einigen davon gar keine bzw. sehr wenige tatsächlichen Near-Repeat-Delikte für Braunschweig enthalten (nur 60 von 4.669, also 1,3 % im Gesamtdatensatz) und damit gar keine bzw. sehr wenige True Positives möglich (aber: gleiches Argument gilt auch für TB 1.2.BS, wo trotz identischer Testmengen die TPR-Boxunterkante dennoch über 0 liegt)
- ➔ FPR-Boxhöhe: geringere Streuung, also etwas besser als bei TB 1.2.BS
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 20,0 \%$ (auffallend höher als bei TB 1.2.BS, aber die zahlreichen TPR-Werte von 0 konterkarieren die Medianbetrachtung)

Fazit: Für Braunschweig alleine funktioniert der Ansatz TB 1.4 besser als der Ansatz TB 1.2 (für die TRP-Verteilung nur potentiell, aber in der hier untersuchten Form auffallend für den PPV aufgrund der deutlich günstigeren FPR-Verteilung).

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 1.4.G (Gleicher Ansatz, TPR/FPR-Auswertung für die Gesamt-Testmengen – H und BS zusammen): ähnlich wie Vergleich mit 1.2.BS, Siehe oben. (Mediane, FPR-Streuung, höherer PPV konterkariert durch TPR-0-Problematik)
- TB 1.2.BS (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS): Siehe oben.
- TB 1.6.BS (gleicher Ansatz ausschließlich mit Braunschweig-Daten, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS – dort identisch mit den Gesamt-Testmengen): Siehe dort unter „Charakteristische Beobachtungen“.

TB 1.5.H Delikttyp 1: WED / TWE (Deliktschlüssel 435 und 436)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Hannover, Landeshauptstadt	3.359	419	3.778
Trainings- und Testdaten:	Nur H		
TPR, FPR berechnet für:	H		
Hannover-Modell			
Nexiga-Daten einbezogen:	Ja		
Korrelationsschwelle:	0,8		
Wahrscheinlichkeitsschwelle:	0,94		

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- ➔ Mediane: eindeutig schlechter als bei TB 1.4.H (TPR-Median etwas kleiner, FPR-Median größer)
- ➔ aber: TPR-Boxunterkante ist auffallend besser, FPR-Boxoberkante ist minimal schlechter als bei TB 1.4.H
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 11,8 \%$
(etwas schlechter als bei TB 1.4.H)

Fazit: Trainieren des Ansatzes TB 1.4 nur mit Hannover-Daten bringt keine Verbesserung, sondern eine sichtbare Verschlechterung. Das heißt, die Braunschweig-Daten liefern dem Ansatz nützliche Informationen für die Hannover-Prognosen.

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 1.2.H (alternativer Ansatz, Training des Modells mit Gesamtdaten, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für H): TPR: Median besser als dort, Boxunterkante sehr ähnlich, daher etwas größere Boxhöhe unproblematisch; FPR: Median und Boxoberkante auffallend schlechter, Streuung schlechter (Boxhöhe etwas, Maximum deutlicher); PPV etwas schlechter als dort
- TB 1.4.H (gleicher Ansatz, Training des Modells mit Gesamtdaten, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für H): Siehe oben.

TB 1.6.BS Delikttyp 1: WED / TWE (Deliktschlüssel 435 und 436)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	831	60	891
Trainings- und Testdaten:	Nur BS		
TPR, FPR berechnet für:	BS		
Braunschweig-Modell			
Nexiga-Daten einbezogen:	Ja		
Korrelationsschwelle:	0,8		
Wahrscheinlichkeitsschwelle:	0,999		

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- ➔ TPR-Boxplot sehr ähnlich wie bei TB 1.4.BS (ein zusätzlicher Ausreißer nach oben von ca. 75 % und die höhere Boxoberkante sind Verbesserungen)
- ➔ FPR-Boxplot etwas schlechter als bei TB 1.4.BS: Median und Streuung nach oben größer als dort
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 18,4 \%$
(etwas schlechter als bei 1.4.BS, aber die zahlreichen TPR-Werte von 0 konterkarieren auch hier die Medianbetrachtung)

Fazit: Trainieren des Ansatzes TB 1.4 nur mit Braunschweig-Daten bringt keine Verbesserung, sondern eine sichtbare Verschlechterung (mit Ausnahme der TPR-Boxoberkante). Das heißt, die Hannover-Daten liefern dem Ansatz nützliche Informationen für die Braunschweig-Prognosen.

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 1.2.BS (alternativer Ansatz, Training des Modells mit Gesamtdaten, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS): TPR: Median und Boxunterkante (=0!) schlechter als dort, Streuung ähnlich; FPR: Median und Boxoberkante besser, Streuung ähnlich; PPV auffallend besser als dort (Aber: viele TPR-Werte sind 0!)
- TB 1.4.BS (gleicher Ansatz, Training des Modells mit Gesamtdaten, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS): Siehe oben.

TB 1.7.G Delikttyp 1: WED / TWE (Deliktschlüssel 435 und 436)

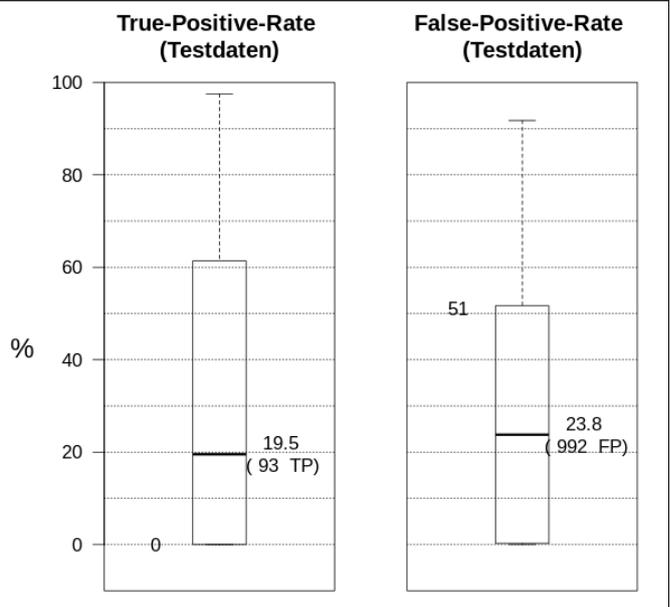
Prognoseansatz: Tree Ensemble

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	831	60	891
Hannover, Landeshauptstadt	3.340	415	3.755
Gesamt	4.171	475	4.646

Trainings- und Testdaten: H u. BS gesamt
 TPR, FPR berechnet für: H u. BS gesamt

Gesamtmodell, Gesamtauswertung

Nexiga-Daten einbezogen: Nein
 Korrelationsschwelle: 0,8
 Wahrscheinlichkeitsschwelle: 0,85



Bei diesem Ansatz wurde die Ortszuordnung anhand der NIVADIS-Daten vorgenommen, daher können die Anzahlen in der Tabelle von denen bei den übrigen Ansätzen TB 1. ... abweichen.

Charakteristische Beobachtungen:

- Mediane: für TPR sichtbar niedriger als für FPR
- ungünstig: TPR-Boxunterkante ist 0
- Streuung: für FPR und TPR relativ hoch (Boxhöhe bei beiden, zusätzlich FPR-Maximalwert – TPR-Maximalwert irrelevant)
- PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 8,6 \%$
(sichtbar niedriger, also schlechter als die schlechtesten Werte bei den übrigen Modellen)

Fazit: Der TPR-Median liegt sichtbar unterhalb des FPR-Medians, woraus ein sehr ungünstiger PPV resultiert. Damit schneidet dieser Ansatz bei diesem Delikttyp deutlich schlechter ab als die anderen Ansätze.

(In einem früheren Prüfungslauf des Tree-Ensemble-Ansatzes für diesen Delikttyp mit Einbeziehung der Nexiga-Daten, allerdings ohne die neu entwickelte Kategorisierung der NIVADIS-Daten, konnte ein TPR-Median oberhalb des FPR-Medians erreicht werden, allerdings in Verbindung mit einer FPR-Box, die insgesamt oberhalb der TPR-Box lag. Der dort resultierende PPV von 10,7 % war immer noch niedrig im Vergleich mit den hier dargestellten anderen Ansätzen.)

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 1.n.G für n = 1, 2, 3, 4 (alternative Ansätze, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Vergleich nur begrenzt möglich, siehe oben.

TB 2.1.G Delikttyp 2: Schwere Diebstahl aus Nichtwohnungen (ohne Landdiebstahl – Deliktschlüssel 410, 415, 425)

Prognoseansatz: Merkmalspezifische Score-Beiträge (NRE, aktueller PreMAP-Ansatz)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	818	76	894
Hannover, Landeshauptstadt	3.476	896	4.372
Gesamt	4.294	972	5.266
Trainings- und Testdaten:	H u. BS gesamt		
TPR, FPR berechnet für:	H u. BS gesamt		
Gesamtmodell, Gesamtauswertung			
Nexiga-Daten einbezogen:	Nein (konzeptionell nur eingeschränkt möglich)		
Korrelationsschwelle:	0,8		
Score-Schwelle (keine Wahrscheinlichkeit):	0,0575		

True-Positive-Rate (Testdaten)

False-Positive-Rate (Testdaten)

Charakteristische Beobachtungen:

- ➔ Mediane: 23,3 % (TPR) auffallend höher als 11,2 % (FPR), das entspricht 226 True Positives (TP, „Zutreffende Alarme“) und damit fast der Hälfte der 480 False Positives (FP, „Falsche Alarme“), resultierender PPV siehe unten
- ➔ TPR-Boxunterkante: Für drei Viertel der 100 Trainingsmengen liegt die TPR in der Testmenge bei 20,3 % oder höher
- ➔ FPR-Boxoberkante: Für drei Viertel der 100 Trainingsmengen liegt die FPR in der Testmenge bei 13,0 % oder niedriger
- ➔ FPR-Maximum unterhalb der TPR-Box (!)
- ➔ TPR-Minimum in der Größenordnung des FPR-Medians
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 32,0 \%$ (auffallend besser als bei allen Modellen TB 1. ... zum Delikttyp 1)

Fazit: Verteilung der TPR-Werte auffallend oberhalb der Verteilung der FPR-Werte, also relativ gute Prognosequalität, die sich auch im PPV von über 30 % niederschlägt.

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 2.n.G für n = 2, 3, 4, 7 (alternative Ansätze, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Charakteristische Beobachtungen“.
- ➔ TB 2.1.H und TB 2.1.BS (gleicher Ansatz, TPR/FPR-Auswertung für regionale Teil-Testmengen – nur H bzw. nur BS – der jeweiligen Testmenge): Siehe dort unter „Charakteristische Beobachtungen“.

TB 2.1.H Delikttyp 2: Schwere Diebstahl aus Nichtwohnungen (ohne Landdiebstahl – Deliktschlüssel 410, 415, 425)

Prognoseansatz: Merkmalspezifische Score-Beiträge (NRE, aktueller PreMAP-Ansatz)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	818	76	894
Hannover, Landeshauptstadt	3.476	896	4.372
Gesamt	4.294	972	5.266
Trainings- und Testdaten:	H u. BS gesamt		
TPR, FPR berechnet für:	H		
Gesamtmodell, Hannover-Auswertung			
Nexiga-Daten einbezogen:	Nein (konzeptionell nur eingeschränkt möglich)		
Korrelationsschwelle:	0,8		
Score-Schwelle (keine Wahrscheinlichkeit):	0,0575		

True-Positive-Rate (Testdaten)

22 24.4
(219 TP)

False-Positive-Rate (Testdaten)

13.5 11.7
(406 FP)

Charakteristische Beobachtungen:

- ➔ FPR-Median und -Boxunterkante etwas besser als bei TB 2.1.G
- ➔ TPR-Median und -Boxoberkante minimal schlechter als bei TB 2.1.G (d. h., im Hinblick auf die falschen Alarme funktioniert der Prognoseansatz TB 2.1 in H minimal schlechter als in H und BS zusammen)
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 35,0 \%$
(sichtbar besser als bei TB 2.1.G)

Fazit: Im Hinblick auf die zutreffenden Alarme funktioniert der Prognoseansatz TB 2.1 in H alleine etwas besser, im Hinblick auf die falschen Alarme minimal schlechter als in H und BS zusammen. Insgesamt ergibt sich eine sichtbare Verbesserung beim PPV ggü. TB 2.1.G.

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 2.1.G (Gleicher Ansatz, TPR/FPR-Auswertung für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- TB 2.n.H für n = 4, 5 (alternativer Ansatz mit Einbeziehung der Nexiga-Daten, Training des Modells mit Gesamtdaten bzw. nur mit Hannover-Daten in TB 2.5.H, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für H): Siehe dort unter „Charakteristische Beobachtungen“ bzw. unter „Bes. rel. Ansatzvarianten zum Vergleich“.

TB 2.1.BS Delikttyp 2: Schwere Diebstahl aus Nichtwohnungen (ohne Landdiebstahl – Deliktschlüssel 410, 415, 425)

Prognoseansatz: Merkmalspezifische Score-Beiträge (NRE, aktueller PreMAP-Ansatz)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	818	76	894
Hannover, Landeshauptstadt	3.476	896	4.372
Gesamt	4.294	972	5.266

Trainings- und Testdaten: H u. BS gesamt
 TPR, FPR berechnet für: BS
Gesamtmodell, Braunschweig-Auswertung

Nexiga-Daten einbezogen: Nein (konzeptionell nur eingeschränkt möglich)
 Korrelationsschwelle: 0,8
 Score-Schwelle (keine Wahrscheinlichkeit): 0,0575

True-Positive-Rate (Testdaten)

False-Positive-Rate (Testdaten)

Charakteristische Beobachtungen:

- ➔ TPR-Median ist 0: sehr ungünstig! – denkbare Erklärung: da die 100 Testmengen als Zufallsstichprobe aus den Gesamtdaten (H und BS) gezogen werden, sind möglicherweise in einigen davon gar keine bzw. sehr wenige tatsächlichen Near-Repeat-Delikte für Braunschweig (nur 76 von 5.266, also 1,4 % im Gesamtdatensatz) enthalten und damit gar keine bzw. sehr wenige True Positives möglich
- ➔ FPR-Median etwas besser, FPR-Boxoberkante minimal besser als bei TB 2.1.G (also wenigstens im Hinblick auf die falschen Alarme funktioniert der Prognoseansatz TB 2.1 in BS besser als in H und BS zusammen)
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 0,0 \%$

Fazit In der Medianbetrachtung liefert dieser Ansatz für BS mit einem TPR-Median von 0 und einem daraus resultierenden PPV von 0 die schlechtestmögliche Prognosequalität. Immerhin liegt die TPR-Boxoberkante bei ca. 12 %, so dass auch TPR-Werte über 0 eine nennenswerte Rolle spielen.

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 2.1.G (Gleicher Ansatz, TPR/FPR-Auswertung für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- TB 2.n.BS für n = 2, 4, 6 (alternative Ansätze, Training des Modells mit Gesamtdaten bzw. nur mit Hannover-Daten in TB 2.6.BS, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS): Siehe dort unter „Charakteristische Beobachtungen“ bzw. unter „Bes. rel. Ansatzvarianten zum Vergleich“.

TB 2.2.G Delikttyp 2: Schwere Diebstahl aus Nichtwohnungen (ohne Landdiebstahl – Deliktschlüssel 410, 415, 425)

Prognoseansatz: Merkmalspezifische Score-Beiträge mit standardisierten Input-Variablen

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	818	76	894
Hannover, Landeshauptstadt	3.476	896	4.372
Gesamt	4.294	972	5.266

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	H u. BS gesamt
Gesamtmodell, Gesamtauswertung	
Nexiga-Daten einbezogen:	Nein (konzeptionell nur eingeschränkt möglich)
Korrelationsschwelle:	0,8
Score-Schwelle (keine Wahrscheinlichkeit)	0,837

True-Positive-Rate (Testdaten)

19.8 22.2
(216 TP)

False-Positive-Rate (Testdaten)

13.4 11.1
(475 FP)

Charakteristische Beobachtungen:

- ➔ TPR-Median etwas schlechter, TPR-Boxunterkante minimal schlechter als bei TB 2.1.G
- ➔ FPR-Verteilung minimal schlechter als bei TB 2.1.G: Median zwar minimal besser, aber Boxoberkante minimal und Maximum etwas schlechter
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 31,3 \%$
(etwas schlechter als bei TB 2.1.G)

Fazit: Bei diesem Delikttyp keine Verbesserung, sondern leichte Verschlechterung der Prognose durch Verwendung der standardisierten Input-Variablen (dennoch im Folgenden TPR-FPR-Auswertung für Braunschweig)

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 2.1.G (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe oben
- TB 2.2.BS (gleicher Ansatz, TPR/FPR-Auswertung für regionale Teilmenge – nur BS – der jeweiligen Testmenge): Siehe dort unter „Bes. rel. Ansatzvarianten zum Vergleich“.

TB 2.2.BS Delikttyp 2: Schwere Diebstahl aus Nichtwohnungen (ohne Landdiebstahl – Deliktschlüssel 410, 415, 425)

Prognoseansatz: Merkmalspezifische Score-Beiträge mit standardisierten Input-Variablen

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	818	76	894
Hannover, Landeshauptstadt	3.476	896	4.372
Gesamt	4.294	972	5.266

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	BS

Gesamtmodell, Braunschweig-Auswertung

Nexiga-Daten einbezogen:	Nein (konzeptionell nur eingeschränkt möglich)
Korrelationsschwelle:	0,8
Score-Schwelle (keine Wahrscheinlichkeit)	0,837

True-Positive-Rate (Testdaten)

False-Positive-Rate (FPR) (Testdaten)

Obwohl TB 2.2.G schlechter als TB 2.1.G ist, wurde diese regionale Auswertung durchgeführt, da die regionale Auswertung TB 2.1.BS eine so ungünstige TPR-Verteilung aufweist. (Denkbar wäre es, Prognosen für Hannover mit dem Ansatz TB 2.1 und für Braunschweig mit dem Ansatz TB 2.2 zu machen.)

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung: Median jetzt 10,0 % (ggü. 0,0 % bei TB 2.1.BS), damit nur knapp halb so hoch wie bei TB 2.2.G, außerdem Boxunterkante immer noch 0,0 %.
- ➔ FPR-Verteilung: Median etwas höher als TPR-Median, Verteilung insgesamt etwas schlechter (Maximum minimal schlechter) als bei TB 2.1.BS und minimal schlechter (Maximum auffallend schlechter) als bei TB 2.2.G
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 8,0 \%$
(naturgemäß auffallend besser als bei TB 2.1.BS mit 0,0 %, aber auffallend schlechter als bei TB 2.2.G, zudem konterkarieren die zahlreichen TPR-Werte von 0 die Medianbetrachtung)

Fazit: Ansatz TB 2.2 für Braunschweig etwas besser als TB 2.1 (bessere TPR-Verteilung bei minimal schlechterer FPR-Verteilung), aber TPR-Verteilung bleibt problematisch

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 2.2.G (Gleicher Ansatz, TPR/FPR-Auswertung für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- TB 2.1.BS (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS): Siehe oben.
- TB 2.n.BS für n = 4, 6 (alternativer Ansatz mit Einbeziehung der Nexiga-Daten, Training des Modells mit Gesamtdaten bzw. nur mit Braunschweig-Daten in TB 2.6.BS, TPR/FPR-Auswertung dementsprechend nur für BS): Siehe dort unter „Charakteristische Beobachtungen“.

TB 2.3.G Delikttyp 2: Schwerer Diebstahl aus Nichtwohnungen (ohne Landendiebstahl – Deliktschlüssel 410, 415, 425)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	818	76	894
Hannover, Landeshauptstadt	3.476	896	4.372
Gesamt	4.294	972	5.266

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	H u. BS gesamt

Gesamtmodell, Gesamtauswertung

Nexiga-Daten einbezogen:	Nein
Korrelationsschwelle:	0,8
Wahrscheinlichkeitsschwelle:	0,9945

True-Positive-Rate (Testdaten)

False-Positive-Rate (Testdaten)

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung minimal schlechter als bei TB 2.1.G (Median, Boxunterkante), Minimum sogar besser
- ➔ FPR-Verteilung: Median ganz minimal schlechter (1 FP!), Boxoberkante und Maximum etwas besser als bei TB 2.1.G
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 31,5 \%$
(minimal schlechter als bei TB 2.1.G)

Fazit: sehr ähnlich wie bei TB 2.1.G, etwas besser als dort im Hinblick auf die FPR-Verteilung

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 2.1.G (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe oben
- TB 2.4.G (gleicher Ansatz mit Einbeziehung der Nexiga-Daten, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Charakteristische Beobachtungen“.
- TB 2.7.G für n = (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Charakteristische Beobachtungen“.

TB 2.4.G Delikttyp 2: Schwerer Diebstahl aus Nichtwohnungen (ohne Landdiebstahl – Deliktschlüssel 410, 415, 425)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	818	76	894
Hannover, Landeshauptstadt	3.476	896	4.372
Gesamt	4.294	972	5.266

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	H u. BS gesamt

Gesamtmodell, Gesamtauswertung

Nexiga-Daten einbezogen:	Ja
Korrelationsschwelle:	0,8
Wahrscheinlichkeitsschwelle:	0,9996

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung: Median und Boxunterkante mindestens 11 Prozentpunkte besser als bei TB 2.3.G, Minimum ca. 8 Prozentpunkte
- ➔ FPR-Verteilung sehr ähnlich wie bei TB 2.3.G (Median minimal besser, Boxoberkante minimal schlechter, Maximum ähnlich)
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 41,2 \%$
(auffallend höher als bei TB 2.3.G)

Fazit: hinsichtlich der TPR-Verteilung auffallende Verbesserung ggü. TB 2.3.G (und entsprechend ggü. TB 2.1.G), FPR-Verteilung sehr ähnlich wie bei diesen beiden Vergleichsmodellen, daraus resultierend sehr viel besserer PPV als dort

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 2.3.G (gleicher Ansatz ohne Einbeziehung der Nexiga-Daten, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- TB 2.1.G (alternative Ansätze, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Sehr ähnlich zu TB 2.3.G, Vergleich also analog zu TB 2.3.G
- TB 2.4.H und TB 2.4.BS (gleicher Ansatz, TPR/FPR-Auswertung für regionale Teil-Testmengen – nur H bzw. nur BS – der jeweiligen Testmenge): Siehe dort unter „Charakteristische Beobachtungen“.

TB 2.4.H Delikttyp 2: Schwerer Diebstahl aus Nichtwohnungen (ohne Landdiebstahl – Deliktschlüssel 410, 415, 425)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	818	76	894
Hannover, Landeshauptstadt	3.476	896	4.372
Gesamt	4.294	972	5.266

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	H

Gesamtmodell, Hannover-Auswertung

Nexiga-Daten einbezogen:	Ja
Korrelationsschwelle:	0,8
Wahrscheinlichkeitsschwelle:	0,9996

True-Positive-Rate (Testdaten)

False-Positive-Rate (Testdaten)

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung: Median und Boxunterkante mindestens 9,9 Prozentpunkte besser als bei TB 2.1.H, Minimum ca. 10 Prozentpunkte
- ➔ FPR-Verteilung sehr ähnlich wie bei TB 2.1.H (Median minimal schlechter, Boxoberkante minimal besser, Maximum ähnlich)
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 42,7 \%$ (auffallend besser als bei TB 2.1.H)

Fazit: hinsichtlich der TPR-Verteilung auffallende Verbesserung ggü. TB 2.1.H, FPR-Verteilung sehr ähnlich wie bei diesem Vergleichsmodell, daraus resultierend sehr viel besserer PPV als dort

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 2.4.G (Gleicher Ansatz, TPR/FPR-Auswertung für die Gesamt-Testmengen – H und BS zusammen): TPR-Verteilung etwas besser, FPR-Verteilung etwas schlechter, resultierender PPV sichtbar besser als dort
- TB 2.1.H (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für H): Siehe oben.
- TB 2.5.H (gleicher Ansatz mit Einbeziehung der Nexiga-Daten, Training des Modells nur mit Hannover-Daten, TPR/FPR-Auswertung dementsprechend nur für H): Siehe dort unter „Charakteristische Beobachtungen“.

TB 2.4.BS Delikttyp 2: Schwere Diebstahl aus Nichtwohnungen (ohne Landdiebstahl – Deliktschlüssel 410, 415, 425)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	818	76	894
Hannover, Landeshauptstadt	3.476	896	4.372
Gesamt	4.294	972	5.266

Trainings- und Testdaten: H u. BS gesamt
 TPR, FPR berechnet für: BS

Gesamtmodell, Braunschweig-Auswertung

Nexiga-Daten einbezogen: Ja
 Korrelationsschwelle: 0,8
 Wahrscheinlichkeitsschwelle: 0,9996

True-Positive-Rate (Testdaten)

False-Positive-Rate (Testdaten)

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung: Median auffallend besser als bei TB 2.2.BS, ungünstige Boxunterkante 0 wie dort
- ➔ FPR-Verteilung: Median, Boxoberkante und Maximum sichtbar besser als bei TB 2.2.BS
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 19,4 \%$
(auffallend besser als bei TB 2.2.BS)

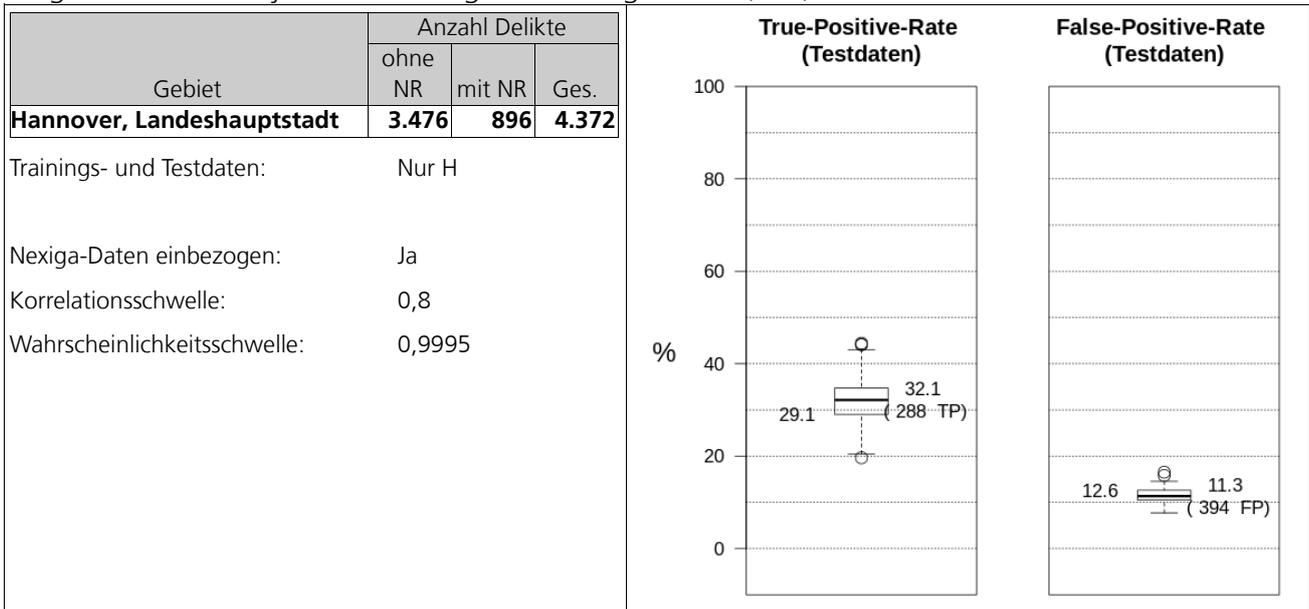
Fazit: BLR-Ansatz mit Nexiga-Daten liefert für diesen Delikttyp die besten Ergebnisse für die Teil-Testmengen für BS, aber Problem der häufigen TPR-0-Werte besteht auch hier.

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 2.4.G (Gleicher Ansatz, TPR/FPR-Auswertung für die Gesamt-Testmengen – H und BS zusammen): TPR-Verteilung auffallend schlechter (Median nur knapp halb so hoch wie dort, Boxunterkante 0), FPR-Verteilung sichtbar besser als dort, resultierender PPV nur knapp halb so hoch wie dort (Vorsicht mit FPR-0-Werten)
- TB 2.2.BS (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS): Siehe oben.
- TB 2.6.BS für $n =$ (gleicher Ansatz mit Einbeziehung der Nexiga-Daten, Training des Modells nur mit Braunschweig-Daten, TPR/FPR-Auswertung dementsprechend nur für BS): Siehe dort unter „Charakteristische Beobachtungen“.

TB 2.5.H Delikttyp 2: Schwere Diebstahl aus Nichtwohnungen (ohne Landendiebstahl – Deliktschlüssel 410, 415, 425)

Prognoseansatz: Bayesianische Logistische Regression (BLR)



Charakteristische Beobachtungen:

- ➔ TPR-Verteilung: Median sichtbar, Boxunterkante und Minimum etwas schlechter als bei TB 2.4.H
- ➔ FPR-Verteilung: Median und Boxoberkante etwas besser als bei TB 2.4.H
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 42,2 \%$
(minimal schlechter als bei TB 2.4.H)

Fazit: Trainieren des Modells zum Ansatz TB 2.4 ausschließlich mit H-Daten („Hannover-Modell“ mit BLR einschl. Nexiga-Daten) bringt für die Prognose in H keine Verbesserung gegenüber dem Trainieren mit BS- und H-Daten, sondern eine Verschlechterung

(Das heißt, die Delikte in Braunschweig liefern dem Prognoseansatz nützliche Informationen für die Prognosen in Hannover. Umgekehrt muss das nicht zwingend der Fall sein, ist es aber, vgl. TB 2.6.BS.)

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 2.1.H (alternativer Ansatz, Training des Modells mit Gesamtdaten, TPR/FPR-Auswertung ebenfalls für die Teilmengen für H)
- TB 2.2.H (alternativer Ansatz, Training des Modells mit Gesamtdaten, TPR/FPR-Auswertung ebenfalls für die Teilmengen für H)
- TB 2.4.H (gleicher Ansatz, Training des Modells mit Gesamtdaten, TPR/FPR-Auswertung ebenfalls für die Teilmengen für H): Siehe oben.

TB 2.6.BS Delikttyp 2: Schwere Diebstahl aus Nichtwohnungen (ohne Landdiebstahl – Deliktschlüssel 410, 415, 425)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	818	76	894
Trainings- und Testdaten:	Nur BS		
TPR, FPR berechnet für:	BS		
Braunschweig-Modell			
Nexiga-Daten einbezogen:	Ja		
Korrelationsschwelle:	0,8		
Wahrscheinlichkeitsschwelle:	0,9997		

True-Positive-Rate (Testdaten)

False-Positive-Rate (Testdaten)

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung: Median etwas schlechter als bei TB 2.4.BS, ungünstige Boxunterkante 0 wie dort
- ➔ FPR-Verteilung: Median, Boxoberkante und Maximum auffallend schlechter als bei TB 2.4.BS
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 9,1 \%$
(auffallend schlechter als bei TB 2.4.BS)

Fazit: Trainieren des Modells zum Ansatz TB 2.4 ausschließlich mit BS-Daten („Braunschweig-Modell“ mit BLR einschl. Nexiga-Daten) bringt für die Prognose in BS keine Verbesserung gegenüber dem Trainieren mit BS- und H-Daten, sondern eine Verschlechterung

(Das heißt, die Delikte in Hannover liefern dem Prognoseansatz nützliche Informationen für die Prognosen in Braunschweig. Umgekehrt ist das auch der Fall, vgl. TB 2.5.H)

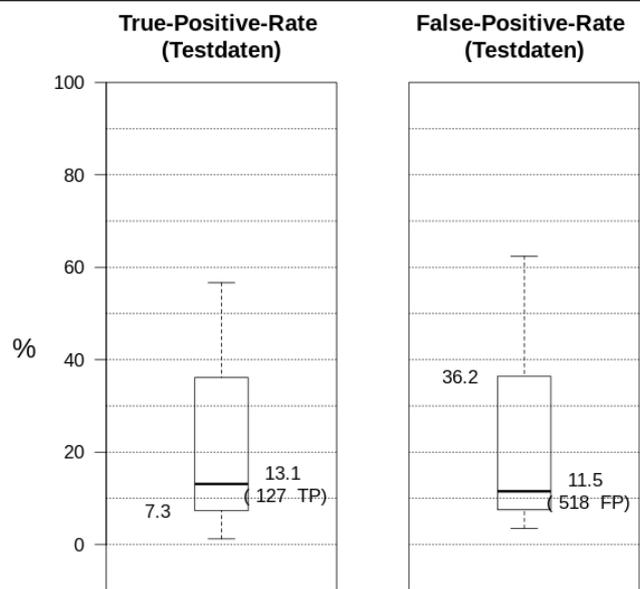
Besonders relevante Ansatzvarianten zum Vergleich:

- TB 2.4.BS (gleicher Ansatz, Training des Modells mit Gesamtdaten, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS), bisher bester Ansatz für diesen Delikttyp und diese Teil-Testmengen: Siehe oben.

TB 2.7.G Delikttyp 2: Schwere Diebstahl aus Nichtwohnungen (ohne Landdiebstahl – Deliktschlüssel 410, 415, 425)

Prognoseansatz: Tree Ensemble

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	818	75	893
Hannover, Landeshauptstadt	3.678	894	4.572
Gesamt	4.496	969	5.465
Trainings- und Testdaten:	H u. BS gesamt		
TPR, FPR berechnet für:	H u. BS gesamt		
Gesamtmodell, Gesamtauswertung			
Nexiga-Daten einbezogen:	Nein		
Korrelationsschwelle:	0,8		
Wahrscheinlichkeitsschwelle:	0,8		



Bei diesem Ansatz wurde die Ortszuordnung anhand der NIVADIS-Daten vorgenommen, daher können die Anzahlen in der Tabelle von denen bei den übrigen Ansätzen TB 2. ... abweichen.

Charakteristische Beobachtungen:

- FPR-Verteilung: Median in ähnlicher Größenordnung wie bei den anderen betrachteten Ansätzen für diesen Delikttyp, aber Boxoberkante deutlich höher als dort
- TPR-Verteilung: Median nur etwas höher als der FPR-Median, hohe Boxoberkante hier unproblematisch
- PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 19,7 \%$
(auffallend schlechter als bei allen anderen Ansätzen TB 2.G)

Fazit: Der Tree-Ensemble-Ansatz ohne Nexiga-Daten liefert für diesen Delikttyp eine auffallend schlechtere Prognosequalität als die übrigen Ansätze.

(In einem früheren Prüfungslauf des Tree-Ensemble-Ansatzes für diesen Delikttyp ohne die neu entwickelte Kategorisierung der NIVADIS-Daten und ohne die Nexiga-Daten lag der TPR-Median etwas unterhalb des FPR-Medians – allerdings bei sichtbar flacheren Boxen – und der PPV bei 17,5 %)

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 2.n.G für n = 1, 2, 3, 4 (alternative Ansätze, für TB 2.4.G mit Einbeziehung der Nexiga-Daten, Training des Modells mit Gesamtdaten, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.

TB 3.1.G Delikttyp 3: Diebstahl in / aus KFZ (Deliktschlüssel 350 und 450)

Prognoseansatz: Merkmalspezifische Score-Beiträge (NRE, aktueller PreMAP-Ansatz)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	1.553	312	1.865
Hannover, Landeshauptstadt	5.290	2.743	8.033
Gesamt	6.843	3.055	9.898
Trainings- und Testdaten:	H u. BS gesamt		
TPR, FPR berechnet für:	H u. BS gesamt		
Gesamtmodell, Gesamtauswertung			
Nexiga-Daten einbezogen:	Nein (konzeptionell nur eingeschränkt möglich)		
Korrelationsschwelle:	0,8		
Score-Schwelle (keine Wahrscheinlichkeit):	0,2		

True-Positive-Rate (Testdaten)

False-Positive-Rate (Testdaten)

Charakteristische Beobachtungen:

- Mediane: 16,2 % (TPR) auffallend höher als 9,4 % (FPR), das entspricht 494 True Positives (TP, „Zutreffende Alarme“) und damit mehr als drei Vierteln der 646 False Positives (FP, „Falsche Alarme“), resultierender PPV siehe unten
- TPR-Boxunterkante: Für drei Viertel der 100 Trainingsmengen liegt die TPR in der Testmenge bei 14,9 % oder höher
- FPR-Boxoberkante: Für drei Viertel der 100 Trainingsmengen liegt die FPR in der Testmenge bei 10,2 % oder niedriger
- FPR-Maximum unterhalb der TPR-Box (!)
- TPR-Minimum oberhalb des FPR-Medians
- PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 43,3 \%$

Fazit: Verteilung der TPR-Werte sichtbar oberhalb der Verteilung der FPR-Werte, also relativ gute Prognosequalität

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 3.n.G für n = 2, 3, 4, 7 (alternative Ansätze, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Charakteristische Beobachtungen“.
- ➔ TB 3.1.H und TB 3.1.BS (gleicher Ansatz, TPR/FPR-Auswertung für regionale Teil-Testmengen – nur H bzw. nur BS – der jeweiligen Testmenge): Siehe dort unter „Charakteristische Beobachtungen“.

TB 3.1.H Delikttyp 3: Diebstahl in / aus KFZ (Deliktschlüssel 350 und 450)

Prognoseansatz: Merkmalspezifische Score-Beiträge (NRE, aktueller PreMAP-Ansatz)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	1.553	312	1.865
Hannover, Landeshauptstadt	5.290	2.743	8.033
Gesamt	6.843	3.055	9.898

Trainings- und Testdaten: H u. BS gesamt
 TPR, FPR berechnet für: H
Gesamtmodell, Hannover-Auswertung

Nexiga-Daten einbezogen: Nein (konzeptionell nur eingeschränkt möglich)
 Korrelationsschwelle: 0,8
 Score-Schwelle (keine Wahrscheinlichkeit): 0,2

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- TPR-Verteilung: Median und Boxunterkante etwas besser als bei TB 3.1.G (d. h., im Hinblick auf die zutreffenden Alarme funktioniert der Prognoseansatz TB 3.1 in H etwas besser als in H und BS zusammen)
- FPR-Verteilung: Median und Boxoberkante etwas schlechter als bei TB 3.1.G (d. h., im Hinblick auf die falschen Alarme funktioniert der Prognoseansatz TB 3.1 in H etwas schlechter als in H und BS zusammen)
- PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 45,0 \%$
(etwas besser als bei TB 3.1.G)

Fazit: Im Hinblick auf die zutreffenden Alarme funktioniert der Prognoseansatz TB 3.1 in H alleine etwas besser, im Hinblick auf die falschen Alarme etwas schlechter als in H und BS zusammen. Insgesamt ergibt sich eine sichtbare Verbesserung beim PPV ggü. TB 3.1.G.

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 3.1.G (Gleicher Ansatz, TPR/FPR-Auswertung für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- ➔ TB 3.n.H für n = 4, 5 (alternativer Ansatz mit Einbeziehung der Nexiga-Daten, Training des Modells mit Gesamtdaten bzw. nur mit Hannover-Daten in TB 3.5.H, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für H): Siehe dort unter „Charakteristische Beobachtungen“ bzw. unter „Bes. rel. Ansatzvarianten zum Vergleich“.

TB 3.1.BS Delikttyp 3: Diebstahl in / aus KFZ (Deliktschlüssel 350 und 450)

Prognoseansatz: Merkmalspezifische Score-Beiträge (NRE, aktueller PreMAP-Ansatz)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	1.553	312	1.865
Hannover, Landeshauptstadt	5.290	2.743	8.033
Gesamt	6.843	3.055	9.898

Trainings- und Testdaten: H u. BS gesamt
 TPR, FPR berechnet für: BS

Gesamtmodell, Braunschweig-Auswertung

Nexiga-Daten einbezogen: Nein
 Korrelationsschwelle: 0,8
 Score-Schwelle (keine Wahrscheinlichkeit): 0,2

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- TPR-Verteilung: Median und Boxunterkante auffallend schlechter als bei TB 3.1.G (d. h., im Hinblick auf die zutreffenden Alarme funktioniert der Prognoseansatz TB 3.1 in BS auffallend schlechter als in H und BS zusammen)
- FPR-Verteilung: Median und Boxoberkante sichtbar besser als bei TB 3.1.G (d. h., im Hinblick auf die falschen Alarme funktioniert der Prognoseansatz TB 3.1 in BS sichtbar besser als in H und BS zusammen)
- PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 23,7 \%$
(nur gut die Hälfte des PPV bei TB 3.1.G, also auffallend schlechter als dort)

Fazit: Ansatz TB 3.1 liefert generell für Delikte in BS weniger Alarme (sowohl weniger zutreffende als auch weniger falsche) als in H – und damit weniger als in H und BS zusammen. Eine Erhöhung der TPR (ungünstigerweise einhergehend mit einer Erhöhung der FPR) durch eine niedrigere Score-Schwelle speziell für BS wäre zu erwägen (Vgl. unten TB 3.4a.BS als Variante von TB 3.4.BS mit einer niedrigeren Wahrscheinlichkeitsschwelle.)

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 3.1.G (Gleicher Ansatz, TPR/FPR-Auswertung für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- ➔ TB 3.n.BS für n = 4a, 6 (alternativer Ansatz mit Einbeziehung der Nexiga-Daten, Training des Modells mit Gesamtdaten bzw. nur mit Braunschweig-Daten in TB 3.6.BS, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS): Siehe dort unter „Charakteristische Beobachtungen“ bzw. unter „Bes. rel. Ansatzvarianten zum Vergleich“.

TB 3.2.G Delikttyp 3: Diebstahl in / aus KFZ (Deliktschlüssel 350 und 450)

Prognoseansatz: Merkmalspezifische Score-Beiträge mit standardisierten Input-Variablen

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	1.553	312	1.865
Hannover, Landeshauptstadt	5.290	2.743	8.033
Gesamt	6.843	3.055	9.898

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	H u. BS gesamt

Gesamtmodell, Gesamtauswertung

Nexiga-Daten einbezogen:	Nein (konzeptionell nur eingeschränkt möglich)
Korrelationsschwelle:	0,8
Score-Schwelle (keine Wahrscheinlichkeit):	0,915

True-Positive-Rate (Testdaten)

13.2 15.6
(476 TP)

False-Positive-Rate (Testdaten)

12.2 10.1
(693 FP)

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung: Median minimal schlechter, Boxunterkante etwas schlechter als bei TB 3.1.G
- ➔ FPR-Verteilung etwas schlechter als bei TB 3.1.G: Median, Boxoberkante und Maximum etwas schlechter
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 40,7 \%$
(sichtbar schlechter als bei TB 3.1.G)

Fazit: Bei diesem Delikttyp keine Verbesserung, sondern leichte Verschlechterung der Prognose durch Verwendung der standardisierten Input-Variablen

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 3.1.G (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen):
Siehe oben

TB 3.3.G Delikttyp 3: Diebstahl in / aus KFZ (Deliktschlüssel 350 und 450)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	1.553	312	1.865
Hannover, Landeshauptstadt	5.290	2.743	8.033
Gesamt	6.843	3.055	9.898

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	H u. BS gesamt

Gesamtmodell, Gesamtauswertung

Nexiga-Daten einbezogen:	Nein
Korrelationsschwelle:	0,8
Wahrscheinlichkeitsschwelle:	0,805

**True-Positive-Rate
(Testdaten)**

12.6 14.9
(456 TP)

**False-Positive-Rate
(Testdaten)**

12.1 9.6
(656 FP)

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung weitgehend schlechter als bei TB 3.1.G: Median, Boxunterkante und Minimum etwas schlechter als dort.
- ➔ FPR-Verteilung ebenfalls weitgehend schlechter als bei TB 3.1.G: Median minimal, Boxoberkante etwas und Maximum sichtbar schlechter als dort.
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 41,0 \%$
(etwas schlechter als bei TB 3.1.G)

Fazit: Ansatz TB 2.1 liefert für H und BS zusammen etwas schlechtere Prognosen als Ansatz TB 3.1

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 3.1.G (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe oben
- TB 3.4.G (gleicher Ansatz mit Einbeziehung der Nexiga-Daten, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Charakteristische Beobachtungen“.
- TB 3.7.G für n = (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Charakteristische Beobachtungen“.

TB 3.4.G Delikttyp 3: Diebstahl in / aus KFZ (Deliktschlüssel 350 und 450)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	1.553	312	1.865
Hannover, Landeshauptstadt	5.290	2.743	8.033
Gesamt	6.843	3.055	9.898

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	H u. BS gesamt

Gesamtmodell, Gesamtauswertung

Nexiga-Daten einbezogen:	Ja
Korrelationsschwelle:	0,8
Wahrscheinlichkeitsschwelle:	0,9925

**True-Positive-Rate
(Testdaten)**

25.6 27.1
(827 TP)

**False-Positive-Rate
(Testdaten)**

9.8 9
(615 FP)

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung auffallend besser als bei TB 3.1.G: Median, Boxunterkante und Minimum mindestens 10,5 Prozentpunkte höher als dort.
- ➔ FPR-Verteilung ebenfalls minimal besser als bei TB 3.1.G: Median und Boxoberkante minimal höher, Maximum sehr ähnlich
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 57,4 \%$
(14,1 Prozentpunkte besser als der schon sehr hohe PPV bei TB 3.1.G)

Fazit: hinsichtlich der TPR-Verteilung auffallende Verbesserung ggü. TB 3.1.G, FPR-Verteilung sehr ähnlich wie dort, daraus resultierend sehr viel besserer PPV als dort

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 3.1.G (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- TB 3.3.G (gleicher Ansatz ohne Einbeziehung der Nexiga-Daten, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen)
- TB 3.4.H und TB 3.4.BS (gleicher Ansatz, TPR/FPR-Auswertung für regionale Teil-Testmengen – nur H bzw. nur BS – der jeweiligen Testmenge): Siehe dort unter „Charakteristische Beobachtungen“.

TB 3.4.H Delikttyp 3: Diebstahl in / aus KFZ (Deliktschlüssel 350 und 450)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	1.553	312	1.865
Hannover, Landeshauptstadt	5.290	2.743	8.033
Gesamt	6.843	3.055	9.898

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	H

Gesamtmodell, Hannover-Auswertung

Nexiga-Daten einbezogen:	Ja
Korrelationsschwelle:	0,8
Wahrscheinlichkeitsschwelle:	0,9925

**True-Positive-Rate
(Testdaten)**

28.3 29.7
(816 TP)

**False-Positive-Rate
(Testdaten)**

12.1 11
(583 FP)

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung auffallend besser als bei TB 3.1.H: Median und Boxunterkante mindestens 12,4 Prozentpunkte höher, Minimum in der Größenordnung des Maximums bei TB 3.1.H
- ➔ FPR-Verteilung minimal schlechter als bei TB 3.1.H: Median 0,2 Prozentpunkte (\pm 10 FP) und Boxoberkante 0,4 Prozentpunkte (\pm 20 FP) höher, Maximum ähnlich
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 58,3 \%$
(Mehr als die Hälfte der Alarme sind zutreffende; 13,3 Prozentpunkte höher, also auffallend besser als bei TB 3.1.H)

Fazit: hinsichtlich der TPR-Verteilung auffallende Verbesserung ggü. TB 3.1.H, FPR-Verteilung sehr ähnlich wie dort, daraus resultierend sehr viel besserer PPV als dort

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 3.4.G (Gleicher Ansatz, TPR/FPR-Auswertung für die Gesamt-Testmengen – H und BS zusammen): TPR-Verteilung etwas besser, FPR-Verteilung etwas schlechter als dort, resultierender PPV etwas besser
- TB 3.1.H (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für H): Siehe oben.
- TB 3.5.H (alternativer Ansatz, Training des Modells nur mit Hannover-Daten, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für H): Siehe dort unter „Charakteristische Beobachtungen“.

TB 3.4.BS Delikttyp 3: Diebstahl in / aus KFZ (Deliktschlüssel 350 und 450)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	1.553	312	1.865
Hannover, Landeshauptstadt	5.290	2.743	8.033
Gesamt	6.843	3.055	9.898

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	BS

Gesamtmodell, Braunschweig-Auswertung

Nexiga-Daten einbezogen:	Ja
Korrelationsschwelle:	0,8
Wahrscheinlichkeitsschwelle:	0,9925

**True-Positive-Rate
(Testdaten)**

0 3
(9 TP)

**False-Positive-Rate
(Testdaten)**

2.6 2
(31 FP)

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung: Werte sehr nahe an 0, mindestens ein Viertel der Werte exakt 0 (Boxunterkante), Boxoberkante nur ca. 5 %
- ➔ FPR-Verteilung: ebenfalls sehr niedrige Werte, das ist aber irrelevant angesichts der sehr ungünstigen TPR-Verteilung
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 22,5 \%$
(irrelevant angesichts der sehr ungünstigen TPR-Verteilung)

Fazit: Ansatz TB 3.4 (mit dieser Wahrscheinlichkeitsschwelle) liefert für die Delikte in Braunschweig sehr unbefriedigende Ergebnisse, das bedeutet eine Verschlechterung ggü. allen relevanten Vergleichsmodellen

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 3.4.G (Gleicher Ansatz, TPR/FPR-Auswertung für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- TB 3.1.BS (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS): Siehe oben.
- TB 3.6.BS (alternativer Ansatz, Training des Modells nur mit Hannover-Daten, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS – dort identisch mit den Gesamt-Testmengen): Siehe oben.

TB 3.4a.BS Delikttyp 3: Diebstahl in / aus KFZ (Deliktschlüssel 350 und 450)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	1.553	312	1.865
Hannover, Landeshauptstadt	5.290	2.743	8.033
Gesamt	6.843	3.055	9.898

Trainings- und Testdaten: H u. BS gesamt
 TPR, FPR berechnet für: BS

**Gesamtmodell, Braunschweig-Auswertung
(Variante von TB 3.4.BS mit anderer Wahrscheinlichkeitsschwelle)**

Nexiga-Daten einbezogen: Ja
 Korrelationsschwelle: 0,8
 Wahrscheinlichkeitsschwelle: 0,6

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Aufgrund der ungünstigen TPR-Verteilung bei 3.4.BS (fast alle Werte nahe 0, mindestens ein Viertel exakt 0) wurde das sonst identische Modell hier für BS mit einer niedrigeren Wahrscheinlichkeitsschwelle ausgewertet, so dass sich für jede Testmenge mehr Alarme, also erwartbar³⁸ sowohl mehr zutreffende (TP) als auch mehr falsche Alarme (FP) und damit höhere TPR- und FPR-Werte ergeben.

Charakteristische Beobachtungen:

- ➔ TPR- und FPR-Verteilung: deutlich besser als bei TB 3.4.BS (Median liegt deutlich über dem FPR-Median, TPR-Boxoberkante deutlich über 0, TPR-Boxunterkante oberhalb der FPR-Boxunterkante)
- ➔ Vergleich mit TB 3.1.BS nicht aussagekräftig (beide Verteilungen deutlich weiter oben als dort)
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 24,5 \%$
(Etwas besser als bei TB 3.4.BS, auch besser als bei TB 3.1.BS)

Fazit: Durch eine Anpassung der Wahrscheinlichkeitsschwelle können die Einzelstadt-Ergebnisse des Ansatzes TB 3.4 für Braunschweig verbessert werden, so dass keine TPR-Boxunterkante von 0 mehr auftritt und der PPV jetzt über dem Wert bei TB 3.1.BS liegt.

Besonders relevante Ansatzvarianten zum Vergleich:

- TB 3.4.BS (Gleicher Ansatz, TPR/FPR-Auswertung für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- TB 3.1.BS (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS): Siehe oben.
- TB 3.6.BS (alternativer Ansatz, Training des Modells nur mit Hannover-Daten, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS – dort identisch mit den Gesamt-Testmengen): Siehe dort unter „Charakteristische Beobachtungen“ .

³⁸ In sehr unwahrscheinlichen Fällen könnten theoretisch nur zutreffende oder nur falsche Alarme dazukommen.

TB 3.5.H Delikttyp 3: Diebstahl in / aus KFZ (Deliktschlüssel 350 und 450)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Hannover, Landeshauptstadt	5.290	2.743	8.033
Trainings- und Testdaten:	Nur H		
TPR, FPR berechnet für:	H		
Hannover-Modell			
Nexiga-Daten einbezogen:	Ja		
Korrelationsschwelle:	0,8		
Wahrscheinlichkeitsschwelle:	0,975		

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung etwas schlechter als bei TB 3.4.H: Median minimal kleiner, Boxunterkante etwas kleiner als dort (Minimum ähnlich wie dort)
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 58,1 \%$
(minimal schlechter als bei TB 3.4.H)

Fazit: Kaum Veränderung der Prognosegüte für H ggü dem Ansatz TB 3.4 durch Einschränkung der Trainingsmengen auf Daten für Hannover, dementsprechend besteht auch hier die Verbesserung ggü. TB 3.1.H.

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 3.4.H (gleicher Ansatz, Training des Modells mit Gesamtdaten, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für H): Siehe oben.
- ➔ TB 3.1.H (alternativer Ansatz, Training des Modells mit Gesamtdaten, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für H): Siehe oben.

TB 3.6.BS Delikttyp 3: Diebstahl in / aus KFZ (Deliktschlüssel 350 und 450)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	1.553	312	1.865
Trainings- und Testdaten:	Nur BS		
TPR, FPR berechnet für:	BS		
Braunschweig-Modell			
Nexiga-Daten einbezogen:	Ja		
Korrelationsschwelle:	0,8		
Wahrscheinlichkeitsschwelle:	0,99		

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung etwas besser als bei TB 3.4a.BS: Median etwas größer, aber: Boxunterkante gleich (aber Minimum = 0)
- ➔ FPR-Verteilung etwas besser als bei TB 3.4a.BS: Median etwas kleiner, Boxoberkante minimal kleiner (aber: Maximum etwas größer)
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 28,3 \%$
(3,8 Prozentpunkte besser als bei TB 3.4a.BS)

Fazit: Für die Delikte in Braunschweig ist es sichtbar besser, beim BLR-Ansatz mit Einbeziehung der Nexiga-Daten (TB 3.4) das Modell allein mit BS-Daten zu trainieren, als das mit den Gesamtdaten trainierte Modell durch eine BS-spezifische Wahrscheinlichkeitsschwelle (TB 3.4a.BS) zu modifizieren (d. h., bei diesem Delikttyp liefern die H-Daten diesem Ansatz *keine* nützlichen Informationen über die BS-Delikte).

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 3.1.BS (alternativer Ansatz, Training des Modells mit Gesamtdaten, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS)
- ➔ TB 3.4a.BS (gleicher Ansatz, Training des Modells mit Gesamtdaten, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS): Siehe oben.

TB 3.7.G Delikttyp 3: Diebstahl in / aus KFZ (Deliktschlüssel 350 und 450)

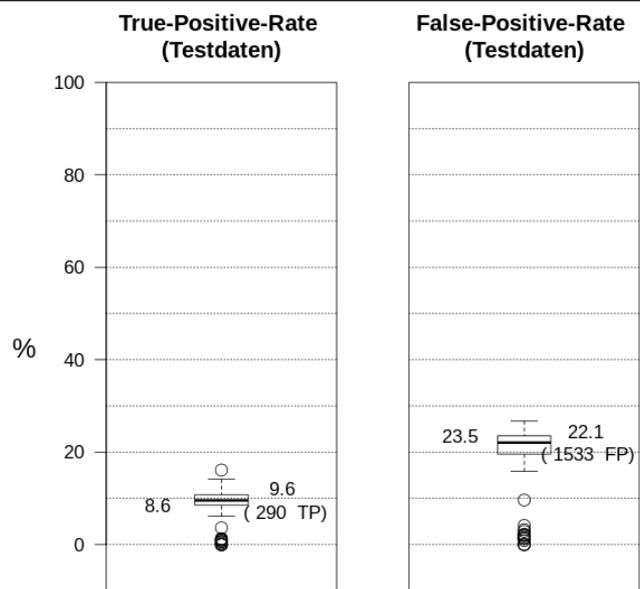
Prognoseansatz: Tree Ensemble

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	1.587	312	1899
Hannover, Landeshauptstadt	5.365	2.722	8.087
Gesamt	6.952	3.034	9.986

Trainings- und Testdaten: H u. BS gesamt
 TPR, FPR berechnet für: H u. BS gesamt

Gesamtmodell, Gesamtauswertung

Nexiga-Daten einbezogen: Nein
 Korrelationsschwelle: 0,8
 Wahrscheinlichkeitsschwelle: 0,8



Bei diesem Ansatz wurde die Ortszuordnung anhand der NIVADIS-Daten vorgenommen, daher können die Anzahlen in der Tabelle von denen bei den übrigen Ansätzen TB 3. ... abweichen.

Charakteristische Beobachtungen:

- Sehr ungünstig: TPR-Verteilung liegt auffallend unterhalb der FPR-Verteilung
- PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 15,9 \%$
(auffallend schlechter als bei allen hier betrachteten Ansätzen TB 3.n.G)

Fazit: Der Tree-Ensemble-Ansatz ohne Nexiga-Daten liefert eine auffallend schlechtere Prognosequalität als die übrigen Ansätze.

(In einem früheren Prüfungslauf des Tree-Ensemble-Ansatzes für diesen Delikttyp mit Einbeziehung der Nexiga-Daten, allerdings ohne die neu entwickelte Kategorisierung der NIVADIS-Daten, ergab sich ein geringerer Abstand zwischen dem TPR- und dem FPR-Median, allerdings lag die ganze TPR-Verteilung immer noch sichtbar niedriger als die FPR-Verteilung. Der dort resultierende PPV von 25,5 % war immer noch niedrig im Vergleich mit den hier dargestellten anderen Ansätzen.)

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 3.n.G für n = 1, 2, 3, 4 (alternative Ansätze, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.

TB 4.1.G Delikttyp 4: Diebstahl von KFZ (Deliktschlüssel 300, 345, 400, 410 und 445)

Prognoseansatz: Merkmalspezifische Score-Beiträge (NRE, aktueller PreMAP-Ansatz)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	400	24	424
Hannover, Landeshauptstadt	862	105	967
Gesamt	1.262	129	1.391

Trainings- und Testdaten: H u. BS gesamt
 TPR, FPR berechnet für: H u. BS gesamt

Gesamtmodell, Gesamtauswertung

Nexiga-Daten einbezogen: Nein (konzeptionell nur eingeschränkt möglich)
 Korrelationsschwelle: 0,8
 Score-Schwelle (keine Wahrscheinlichkeit): 0,09

True-Positive-Rate (Testdaten)

False-Positive-Rate (Testdaten)

Charakteristische Beobachtungen:

- ➔ Insgesamt geringe Deliktanzahl (BS hier auffallend näher an H als bei den übrigen untersuchten Delikttypen), also schwache Datenbasis für alle Ansätze aufgrund relativ kleiner Trainingsmengen, außerdem sehr kleine Testmengen (je 139 Delikte) mit potentiell stark schwankenden H- und BS-Anteilen
- ➔ Mediane: 10,0 % (TPR) minimal kleiner als 10,4 % (FPR), das entspricht 13 True Positives (TP, „Zutreffende Alarme“) und damit einem knappen Zehntel der 131 False Positives (FP, „Falsche Alarme“), resultierender PPV siehe unten
- ➔ TPR-Boxunterkante sehr ungünstig: Für mindestens ein Viertel der 100 Trainingsmengen liegt die TPR in der Testmenge bei 0,0 % – eventuell enthalten einige der sehr kleinen Testmengen nur vereinzelte bzw. gar keine Delikte mit NR, was eine TPR von 0 begünstigt bzw. erzwingt.
- ➔ FPR-Boxoberkante: Für drei Viertel der 100 Trainingsmengen liegt die FPR in der Testmenge bei 25,4 % oder niedriger
- ➔ Ungünstig: TPR-Verteilung liegt tendenziell (bis auf oberen Whisker und Maximum) unterhalb der FPR-Verteilung
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 9,0 \%$
(Vorsicht: begrenzt aussagekräftig wegen der hohen Anzahl von TPR-0-Werten)

Fazit: Keine gute Prognosequalität, schlägt sich auch nieder im geringen PPV.

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 4.n.G für n = 2, 3, 4, 7 (alternative Ansätze, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Charakteristische Beobachtungen“.

TB 4.2.G Delikttyp 4: Diebstahl von KFZ (Deliktschlüssel 300, 345, 400, 410 und 445)

Prognoseansatz: Merkmalspezifische Score-Beiträge mit standardisierten Input-Variablen

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	400	24	424
Hannover, Landeshauptstadt	862	105	967
Gesamt	1.262	129	1.391
Trainings- und Testdaten:	H u. BS gesamt		
TPR, FPR berechnet für:	H u. BS gesamt		
Gesamtmodell, Gesamtauswertung			
Nexiga-Daten einbezogen:	Nein (konzeptionell nur eingeschränkt möglich)		
Korrelationsschwelle:	0,8		
Score-Schwelle (keine Wahrscheinlichkeit):	0,09		

True-Positive-Rate (Testdaten)

False-Positive-Rate (Testdaten)

Charakteristische Beobachtungen:

- ➔ Mediane: 13,3 % (TPR) etwas größer als 10,9 % (FPR), das entspricht 17 True Positives (TP, „Zutreffende Alarme“) und damit mehr als einem Zehntel der 131 False Positives (FP, „Falsche Alarme“), resultierender PPV siehe unten
- ➔ TPR-Boxunterkante: Für drei Viertel der 100 Trainingsmengen liegt die TPR in der Testmenge bei 8,3 % oder höher (Boxunterkante nicht wie bei TB 4.1.G beim Wert 0 – damit ist empirisch erwiesen, dass die Boxunterkante bei TB 4.1.G auf die schlechte Prognosequalität dort und nicht auf keine oder sehr wenige Delikte mit NR in den Testmengen zurückzuführen ist.)
- ➔ ungünstig: unterer TPR-Whisker reicht bis zur 0
- ➔ FPR-Boxoberkante: Für drei Viertel der 100 Trainingsmengen liegt die FPR in der Testmenge bei 12,7 % oder niedriger
- ➔ Günstig: TPR-Verteilung liegt tendenziell (bis auf Boxunterkante und Minimum) unterhalb der FPR-Verteilung
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 11,0 \%$
(etwas besser als bei TB 4.1.G, dortige begrenzte Aussagekraft wegen der Boxunterkante von 0 entfällt hier, Wert liegt dennoch nur in der Größenordnung der schlechtesten Werte bei den bisher betrachteten Delikttypen)

Fazit: Bei diesem Delikttyp Verbesserung der Prognosequalität durch Verwendung der standardisierten Input-Variablen.

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 4.1.G (gleicher Ansatz ohne Standardisierung der Input-Variablen, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- ➔ TB 4.n.G für n = 3, 4, 7 (alternative Ansätze ohne und mit Einbeziehung der Nexiga-Daten, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Charakteristische Beobachtungen“.

TB 4.3.G Delikttyp 4: Diebstahl von KFZ (Deliktschlüssel 300, 345, 400, 410 und 445)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	400	24	424
Hannover, Landeshauptstadt	862	105	967
Gesamt	1.262	129	1.391

Trainings- und Testdaten: H u. BS gesamt
 TPR, FPR berechnet für: H u. BS gesamt

Gesamtmodell, Gesamtauswertung

Nexiga-Daten einbezogen: Nein
 Korrelationsschwelle: 0,8
 Wahrscheinlichkeitsschwelle: 0,9

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung: Median und Boxunterkante sichtbar besser als bei TB 4.2.G (Minimum ungünstigerweise immer noch 0,0 %)
- ➔ FPR-Verteilung: Median etwas, Boxoberkante sichtbar schlechter als bei TB 4.2.G, außerdem viele Werte oberhalb des oberen Whiskers (damit eigentlich keine „Ausreißer“ mehr – hier nicht ablesbar: es handelt sich um 10 FPR-Werte, davon gehören 5 zu einer Testmenge mit einem noch höheren TPR-Wert – davon 4 im Bereich des oberen TPR-Whiskers –, aber auch 4 zu einer Testmenge mit einem auffallend niedrigeren TPR-Wert)
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 13,5 \%$
(2,5 Prozentpunkte besser als bei TB 4.2.G)

Fazit: Dieser Ansatz liefert keine eindeutige Verbesserung ggü. TB 4.2.G

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 4.2.G (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- ➔ TB 4.4.G (gleicher Ansatz mit Einbeziehung der Nexiga-Daten, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe dort unter „Charakteristische Beobachtungen“.
- ➔ TB 4.7.G (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen).

TB 4.4.G Delikttyp 4: Diebstahl von KFZ (Deliktschlüssel 300, 345, 400, 410 und 445)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	400	24	424
Hannover, Landeshauptstadt	862	105	967
Gesamt	1.262	129	1.391

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	H u. BS gesamt

Gesamtmodell, Gesamtauswertung

Nexiga-Daten einbezogen:	Ja
Korrelationsschwelle:	0,8
Wahrscheinlichkeitsschwelle:	0,995

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung: Median auffallend, Boxunterkante sichtbar besser als bei TB 4.2.G (Minimum ungünstigerweise immer noch 0,0 %)
- ➔ FPR-Verteilung: Median minimal besser, Boxoberkante etwas schlechter als bei TB 4.2.G (außerdem Maximum sichtbar kleiner)
- ➔ TPR- liegt auffallend über der FPR-Verteilung (TPR-Boxunterkante sogar oberhalb des FPR-Medians)
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 18,0 \%$
(auffallend höher als bei TB 4.2.G)

Fazit: Durch die Einbeziehung der Nexiga-Daten lässt sich der BLR-Ansatz sichtbar verbessern, so dass er nun eine auffallend höhere Prognosequalität als der Ansatz TB 4.2 aufweist.

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 4.2.G (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- ➔ TB 4.3.G (gleicher Ansatz ohne Einbeziehung der Nexiga-Daten, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): etwas besser als dort.
- ➔ TB 4.4.H und TB 4.4.BS (gleicher Ansatz, TPR/FPR-Auswertung für regionale Teil-Testmengen – nur H bzw. nur BS – der jeweiligen Testmenge): Siehe dort unter „Charakteristische Beobachtungen“.
- ➔ TB 4.7.G (alternativer Ansatz, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen).

TB 4.4.H Delikttyp 4: Diebstahl von KFZ (Deliktschlüssel 300, 345, 400, 410 und 445)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	400	24	424
Hannover, Landeshauptstadt	862	105	967
Gesamt	1.262	129	1.391

Trainings- und Testdaten:	H u. BS gesamt
TPR, FPR berechnet für:	H

Gesamtmodell, Hannover-Auswertung

Nexiga-Daten einbezogen:	Ja
Korrelationsschwelle:	0,8
Wahrscheinlichkeitsschwelle:	0,995

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung: Median und Boxunterkante sichtbar höher als bei TB 4.4.G (Maximum auch, Minimum leider 0 wie dort)
- ➔ FPR-Verteilung schlechter als bei TB 4.4.G: Median und Boxoberkante mindestens 4,1 Prozentpunkte höher als dort
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 17,1 \%$
(etwas schlechter als bei TB 4.4.G)

Fazit: Für Hannover allein liefert der Ansatz TB 4.4 eine bessere TPR-Verteilung als für H und BS zusammen, aber in Verbindung mit einer schlechteren FPR-Verteilung resultiert ein ungünstigerer PPV.

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 4.4.G (gleicher Ansatz, TPR/FPR-Auswertung für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.
- ➔ TB 4.5.H (gleicher Ansatz, Training des Modells nur mit Hannover-Daten, TPR/FPR-Auswertung ebenfalls): Siehe dort unter „Charakteristische Beobachtungen“ .

TB 4.4.BS Delikttyp 4: Diebstahl von KFZ (Deliktschlüssel 300, 345, 400, 410 und 445)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	400	24	424
Hannover, Landeshauptstadt	862	105	967
Gesamt	1.262	129	1.391

Trainings- und Testdaten: H u. BS gesamt
 TPR, FPR berechnet für: BS

Gesamtmodell, Braunschweig-Auswertung

Nexiga-Daten einbezogen: Ja
 Korrelationsschwelle: 0,8
 Wahrscheinlichkeitsschwelle: 0,995

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung: die ganze Box liegt auf dem Wert 0, d. h., bei weniger als einem Viertel der 100 Testmengen ist überhaupt mindestens 1 zutreffender Alarm für BS aufgetreten (hier nicht ablesbar: tatsächlich sind es nur 12 Testmengen, in denen sich eine TPR über 0 ergibt).
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 0,0 \%$ (sehr ungünstig)

Fazit: Ansatz TB 4.4 weist für Braunschweig allein eine katastrophal schlechte Prognosequalität auf, die Verwendung des Ansatzes TB 4.4 für Braunschweig kann eigentlich nur durch die „Hoffnung auf Glückstreffer“ gerechtfertigt werden.

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 4.4.G (gleicher Ansatz, TPR/FPR-Auswertung für die Gesamt-Testmengen – H und BS zusammen): unnötig, da TB 4.4.BS eine sehr schlechte Prognosequalität aufweist.
- ➔ TB 4.6.BS (gleicher Ansatz, Training des Modells nur mit Braunschweig-Daten, TPR/FPR-Auswertung ebenfalls): Siehe dort unter „Charakteristische Beobachtungen“ .

TB 4.5.H Delikttyp 4: Diebstahl von KFZ (Deliktschlüssel 300, 345, 400, 410 und 445)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Hannover, Landeshauptstadt	862	105	967
Trainings- und Testdaten:	Nur H		
TPR, FPR berechnet für:	H		
Hannover-Modell			
Nexiga-Daten einbezogen:	Ja		
Korrelationsschwelle:	0,8		
Wahrscheinlichkeitsschwelle:	0,995		

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- ➔ TRP-Verteilung etwas schlechter als bei TB 4.4.H: Median wie dort, aber Boxunterkante und Boxoberkante (und außerdem das Maximum) etwas schlechter als dort.
- ➔ FPR-Verteilung etwas besser als bei TB 4.4.H: Median etwas, Boxoberkante (und außerdem das Maximum) sichtbar niedriger als dort.
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 19,0 \%$
(etwas besser als bei TB 4.4.H)

Fazit: Trainieren des Ansatzes aus TB 4.4 allein mit den Hannover-Daten bewirkt insgesamt eine leichte Verbesserung der Prognosequalität durch eine nach unten verschobene FPR-Verteilung, allerdings bei einer leichten Verschlechterung der TPR-Verteilung. Durch eine kleine Absenkung der Wahrscheinlichkeitsschwelle könnte hier aber sicherlich eine Verbesserung hinsichtlich der FPR- und der TPR-Verteilung erreicht werden. D. h., die Braunschweig-Daten liefern dem BLR-Ansatz mit Einbeziehung der Nexiga-Daten keine nützlichen Informationen über die Delikte in Hannover.

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 4.4.H (gleicher Ansatz, Training des Modells mit Gesamtdaten, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für H): Siehe oben.

TB 4.6.BS Delikttyp 4: Diebstahl von KFZ (Deliktschlüssel 300, 345, 400, 410 und 445)

Prognoseansatz: Bayesianische Logistische Regression (BLR)

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	400	24	424
Trainings- und Testdaten:	Nur BS		
TPR, FPR berechnet für:	BS		
Braunschweig-Modell			
Nexiga-Daten einbezogen:	Ja		
Korrelationsschwelle:	0,8		
Wahrscheinlichkeitsschwelle:	0,995		

**True-Positive-Rate
(Testdaten)**

**False-Positive-Rate
(Testdaten)**

Charakteristische Beobachtungen:

- ➔ TPR-Verteilung etwas besser als bei TB 4.4.BS: Median immer noch bei 0, Boxoberkante jetzt auffallend höher als 0 (ca. 25 %), d. h., bei einem Viertel der 100 Testmengen liegt der TPR-Wert bei 25 % oder höher (hier nicht ablesbar: es sind 24 Testmengen mit $TPR \geq 25\%$, bei weiteren 6 Testmengen liegt der TPR-Wert über 0 und unter 25 %, dabei ergibt sich rechnerisch eine Boxoberkante knapp unter 25 %)
- ➔ FPR-Verteilung auffallend schlechter als bei TB 4.4.BS, aber durchaus im Rahmen der hier untersuchten (und letztlich durch Wahl von Score- und Wahrscheinlichkeitsschwellen angestrebten) FPR-Verteilungen
- ➔ PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 0,0\%$
(Wie bei TB 4.4.BS nicht wirklich aussagekräftig, da TPR-Median wie dort bei 0,0 % liegt.)

Fazit: Trainieren des Ansatzes TB 4.4 allein mit den Braunschweig-Daten behebt nur teilweise die katastrophal schlechte Prognosequalität bei TB 4.4.BS. Immerhin treten bei 30 der 100 Testmengen TPR-Werte über 0 auf. Die unter TB 4.4.BS genannte „Hoffnung auf Glückstreffer“ bekommt also durch die Einschränkung der Trainingsmenge auf die BS-Daten etwas mehr Nahrung. Auch hier kann man festhalten, dass die Prognosequalität ohne Einbeziehung der Hannover-Daten besser ist, dass diese also dem Ansatz keine nützlichen Informationen über die Braunschweig-Delikte liefern.

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 4.4.BS (gleicher Ansatz, Training des Modells mit Gesamtdaten, TPR/FPR-Auswertung ebenfalls für die Teil-Testmengen für BS): Siehe oben.

TB 4.7.G Delikttyp 4: Diebstahl von KFZ (Deliktschlüssel 300, 345, 400, 410 und 445)

Prognoseansatz: Tree Ensemble

Gebiet	Anzahl Delikte		
	ohne NR	mit NR	Ges.
Braunschweig, Stadt	406	24	430
Hannover, Landeshauptstadt	863	104	967
Gesamt	1.269	128	1.397

Trainings- und Testdaten: H u. BS gesamt

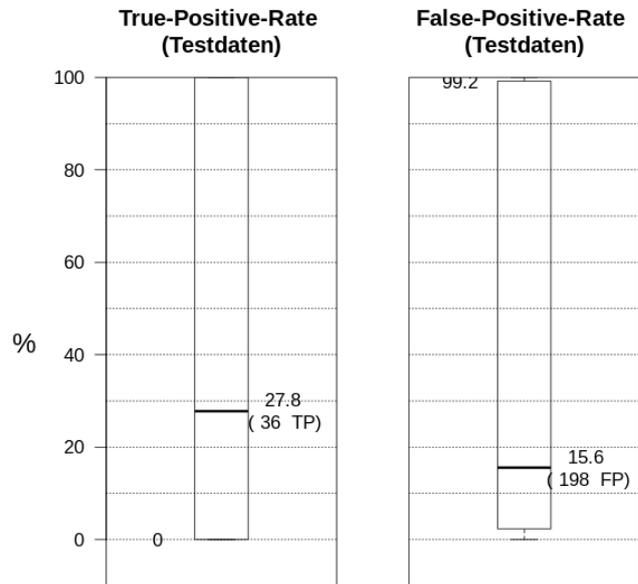
TPR, FPR berechnet für: H u. BS gesamt

Gesamtmodell, Gesamtauswertung

Nexiga-Daten einbezogen: Nein

Korrelationsschwelle: 0,8

Wahrscheinlichkeitsschwelle: 0,7



Bei diesem Ansatz wurde die Ortszuordnung anhand der NIVADIS-Daten vorgenommen, daher können die Anzahlen in der Tabelle von denen bei den übrigen Ansätzen TB 4. ... abweichen.

Charakteristische Beobachtungen:

- sehr ungünstig: TPR-Box reicht exakt, FPR-Box annähernd von 0 % bis 100 % (d. h., mindestens ein Viertel aller aufgetretenen TPR-Werte sind 0 % und mindestens ein Viertel der FPR-Werte betragen annähernd 100 %)
- PPV (Medianbetrachtung) = $TP / (TP + FP) * 100 = 15,4 \%$
(zweitbesten Wert nach TB 4.4.G, aber begrenzt aussagekräftig wegen der großen TPR- und FPR-Streuung)

Fazit: Der TPR-Median liegt in wünschenswerter Weise oberhalb des FPR-Medians, aber dieses Ergebnis und der daraus resultierende günstige PPV werden konterkariert durch die großen Streuungen der TPR- und FPR-Werte.

(Durch eine Erhöhung der Wahrscheinlichkeitsschwelle auf 0,75 reduzieren sich beide Mediane auf Werte unter 10 %, aber beide Boxoberkanten liegen dann immer noch bei 90 %. Der resultierende PPV beträgt dann nur 8,9 %)

Besonders relevante Ansatzvarianten zum Vergleich:

- ➔ TB 4.n.G für n = 1, 2, 3, 4 (alternative Ansätze, TPR/FPR-Auswertung ebenfalls für die Gesamt-Testmengen – H und BS zusammen): Siehe oben.