

Kooperationsprojekt "Algorithmen fürs Gemeinwohl"

Intersektoraler Workshop zu Predictive Policing

20.4.2018, 12 bis 16:30 Uhr, Stiftung Neue Verantwortung (Berlin)

Ergebnisse

Dr. Tobias Knobloch, Mai 2018

Zusammenfassung & Evaluation

In Deutschland werden derzeit ausschließlich ortsbezogene Predictive-Policing-Systeme zur Bekämpfung von Wohnungseinbrüchen im urbanen Raum eingesetzt. Die Gründe hierfür sind zum einen, dass für diese Form von Delikten wiederholt statistisch signifikante Near-Repeat-Muster nachgewiesen werden konnten, die im ländlichen Raum wesentlich schwächer sind, und des Weiteren, dass die Dunkelziffer im Bereich der Wohnungseinbruchskriminalität grundsätzlich niedrig ist. Dadurch wird das Risiko einer systematischen Verzerrung verringert. Das Problem der Isolierung von Einflussfaktoren und des Aufdeckens von Wirkzusammenhängen besteht jedoch auch hier und stellt eine strukturelle Herausforderung für die Evaluierung dar. Es wären dazu langfristig-vergleichende Studiendesigns erforderlich.

Der intersektorale Austausch hat weitere gesellschaftspolitische, kriminologische, psychologische und methodologische Fragen aufgeworfen. Offenkundige Mängel bestehen derzeit bei der sozialen Einbettung von Maßnahmen der vorausschauenden Polizeiarbeit; diese sind im Wesentlichen auf **mangelnde Kommunikation auf mehreren Ebenen** zurückzuführen. Während Systemtransparenz (zumindest für Experten) in den betrachteten Anwendungsfällen (Precobs in Bayern und Baden-Württemberg, SKALA in Nordrhein-Westfalen) gegeben ist, mangelt es an einer öffentlich nachvollziehbaren **Systemnutzungstransparenz**. Man stehe hier, so die teilnehmenden Kriminalbeamten, zwischen den beiden Polen Abschottung und totaler Offenheit – letzteres mit der möglichen Folge, den Erfolg der eigenen Arbeit zu gefährden.

Die von den Teilnehmenden bemängelte politische und mediale Über-Aufmerksamkeit, die dem Thema Predictive Policing gemessen an seiner tatsächlichen Bedeutung in der Tat geschenkt wird, geht auch auf die zu passive Informationspolitik der Polizeibehörden zurück – mit der Folge, dass der Kommunikationsaufwand in der Regel stark unterschätzt wird. Zwar ist der folgenden Aussage eines Teilnehmers zuzustimmen: „Vertrauensaufbau in Technologie ist eine gesamtgesellschaftliche Aufgabe, die nicht allein bei der Polizei abgeladen werden kann.“ Allerdings verkennt diese Haltung, dass die Pflicht zur Information Betroffener zunächst bei demjenigen liegt, der algorithmische Systeme einsetzt. Vertrauen in Technologie beginnt mit dem Wissen darüber. Die größten Ängste schüren gewöhnlich Dinge, die unbekannt sind und daher potentiell bedrohlich wirken. Predictive Policing ist

jedoch kein Hexenwerk und in der Form, in der es derzeit in Deutschland angewendet wird, (noch) mit geringen Risiken verbunden. Es ist allenfalls mehr oder weniger wirksam, weshalb es mehr oder weniger sinnvoll ist, dafür Ressourcen einzusetzen. Darüber kann und sollte man einen **offenen gesellschaftlichen Diskurs unter Einbezug aller Sektoren** führen. Resultate wären vermutlich eine erhöhte Akzeptanz innerhalb der Bevölkerung, weniger parlamentarische Anfragen und eine geringere mediale Aufmerksamkeit. Die am Workshop teilnehmenden Behörden haben aufgrund ihrer sehr reflektierten Vorgehensweise eigentlich keinen Grund, einen proaktiven gesellschaftlichen Diskurs zu scheuen.

Alle Teilnehmenden waren sich darin einig, dass der intersektorale Austausch über die Vor- und Nachteile sowie die Chancen und Gefahren des Einsatzes prädiktiver Algorithmen zur Verbrechensbekämpfung und -vorbeugung gerade zum jetzigen, frühen Zeitpunkt sehr wichtig ist. Denn es sei davon auszugehen, dass die Methode auf andere Deliktsfelder ausgeweitet und dabei eine größere Datenbasis als heute herangezogen werden wird. **Unbedingt geboten sei daher, sich mit der Frage zu befassen, was sich durch den Einsatz von Algorithmen zukünftig an der Polizeiarbeit strukturell verändere.** Auch müsse vor jeder bewussten Gestaltung eines solchen Veränderungsprozesses jetzt geklärt werden, wo man hinwolle und was man auf keinen Fall wolle (die berüchtigte rote Linie). Dies ist auch in Anbetracht eines derzeit insgesamt eher unproblematischen Einsatzes prädiktiver Algorithmen durch die Landespolizeibehörden angebracht, denn das sogenannte Shifting-Baselines-Phänomen kann durchaus für eine unmerkliche Unterwanderung demokratischer Prinzipien durch Technikeinsatz im Polizeialltag sorgen.

Nachfolgend werden nun zunächst die Thesen reflektiert, die dem Workshop zur Überprüfung vorangestellt waren¹, und dabei weitere Detailergebnisse aufgezeichnet. Am Ende stehen einige allgemeine Schlussfolgerungen für das Design und die Nutzung von teilhaberelevanten algorithmischen Analyse- und Entscheidungssystemen.

Block A: Die Funktionsweise deutscher Predictive-Policing-Systeme

Zu prüfende Thesen waren:

1. Ortsbezogenes Predictive Policing eignet sich für manche Straftaten (z.B. Wohnungseinbrüche) besser als für andere (z.B. Raubüberfälle); dass eine geringe Dunkelziffer von Vorteil und die Anwendbarkeit der Near-Repeats-Theorie Voraussetzung ist, sind die Gründe dafür.

Diese These konnte **bestätigt** werden, wobei insbesondere NRW mit seiner Eigenentwicklung SKALA, die neben den Polizeidaten weitere Datenquellen einbezieht, auf eine deliktische Weiterentwicklung sowie auf Anwendbarkeit auch im ländlichen Raum abzielt (letzteres gilt auch für die Länder, die Precobs einsetzen).

2. Die Datenbasis ist für die mit Predictive-Policing-Systemen erzielbaren Effekte zentral. Deshalb braucht es Data Literacy in den Behörden und einen sehr bewussten, sorgsamem Umgang mit Datenquellen und datenverarbeitenden Systemen; hier besteht vermutlich Nachholbedarf.

¹ Vgl. das Input-Papier, das der Autor am 18.4. an die Teilnehmenden geschickt hat.

These 2 konnte ebenfalls **bestätigt** werden, wurde in ihrem zweiten Teil des Workshops (Nachholbedarf bzgl. Data Literacy) jedoch **relativiert**, weil das Kompetenzniveau bei Implementierenden und Anwendern (Lagesachbearbeiter:innen) anscheinend bereits recht hoch ist. In diesem Punkt ergab die Diskussion jedoch einen anderen sehr wichtigen Aspekt: Die Datenerfassung selbst wird in öffentlichen Debatten kaum thematisiert. Die Vorgangsbearbeitungssysteme, mit denen sämtliche Polizeidaten erfasst werden, sind komplexe Anwendungen, deren Voreinstellungen mit darüber entscheiden, welche Daten zu einem Vorgang überhaupt angelegt und gespeichert und später weiterverarbeitet werden können (bezogen auf das hier behandelte Thema z.B. Einbrucharten in einem Drop-down-Menü). Zudem kosten diese Systeme in der Anschaffung und Pflege weitaus mehr als jede auf dem Markt verfügbare Predictive-Policing-Software, so dass hier sehr viel mehr Aufmerksamkeit durchaus angebracht erscheint. Dass die Datenbasis allgemein ein für den Erfolg oder Misserfolg von Predictive Policing ausschlaggebender Faktor ist, ist bei den im Workshop vertretenen Anwendern hinreichend bekannt und anerkannt. Nicht zuletzt die Vorstellung des SKALA Systems machte deutlich, dass die mangelnde Zugänglichkeit von öffentlichen Daten, vor allem aufgrund unterschiedlicher Daten-Formate, eine Hürde für die Eigenentwicklung algorithmischer Systeme darstellt. Dies gilt insbesondere für Software, die neben den Polizeidaten weitere Datenquellen einbezieht und somit auf die Kooperation anderer Behörden angewiesen ist. Zur Entwicklung von SKALA hat das Ministerium des Inneren NRW beispielsweise Daten des auf Geomarketing spezialisierten Unternehmens nexiga GmbH eingekauft.

Block B: Institutionelle Verankerung, soziale Einbettung und Evaluierung

Zu prüfende Thesen waren:

1. Institutionelle Verankerung und soziale Einbettung von Predictive-Policing-Systemen müssen von Anfang bis Ende systematisch erfolgen; Anzeichen deuten darauf hin, dass das derzeit nicht immer der Fall ist.

Diese These wurde in ihrem ersten Teil **bestätigt** und in ihrem zweiten Teil **teilweise bestätigt**. Die Lagesachbearbeiter:innen entscheiden nach wie vor autonom und betrachten Predictive-Policing-Systeme als ein weiteres Tool in ihrem Werkzeugkoffer; allerdings fiel auch die Aussage, dass derlei Systeme durchaus an der Ehre der Polizei kratzen. Zwischen den drei Ebenen Entscheider/Beschaffer, Anwender/Lagesachbearbeiter und den mit den Ergebnissen vor Ort arbeitenden Polizeibeamten gibt es anscheinend fest etablierte Feedback-Wege, was sehr zu begrüßen ist. Allerdings unterscheidet sich die Akzeptanz von und der Umgang mit Predictive-Policing-Systemen nach Alter, Ausbildungsgrad und Dienstrang z.T. erheblich. Notwendige Schulungen im Umgang mit den Systemen brauchen Zeit und erreichen nicht alle betroffenen Beamten zugleich. Hier wäre ein systematischeres Vorgehen mit größeren Anstrengungen wünschenswert, um für ein vergleichbares Kompetenzniveau innerhalb der Beamtenschaft zu sorgen. Drei große Herausforderungen stehen einer guten sozialen Einbettung derzeit im Weg: erstens die fehlende Information der Bevölkerungsteile, für die sich die Polizeipräsenz aufgrund von Predictive Policing verändert; zweitens die politischen und medialen Übertreibungen, die im Umgang mit dem Thema Predictive Policing zu verzeichnen sind; drittens das Fehlen einer proaktiven polizeilichen Kommunikationsstrategie rund um das Thema.

2. Die Wirkungen derzeitig eingesetzter Predictive-Policing-Systeme sind schwierig und nur mit erheblichem Aufwand messbar. Evaluierung ist anspruchsvoll, aber absolut notwendig; bislang wird sie insgesamt vernachlässigt.

These 2 konnte **voll bestätigt** werden (vgl. dazu auch die Zusammenfassung oben). Die Vernachlässigung der Evaluierung resultiert vor allem aus der Schwierigkeit und Aufwändigkeit des Designs: Für eine State-of-the-art Evaluierung müssten die Genauigkeit der algorithmischen Prognosen und die Effektivität der polizeilichen Maßnahmen voneinander unabhängig und in ihrem Zusammenwirken, unter Kontrolle anderer Einflussfaktoren und über eine Langzeitlaufzeit untersucht werden. Nur so kann beispielsweise festgestellt werden, ob der Einsatz von Predictive Policing Software tatsächlich Kriminalität senkt, oder ob Kriminalität lediglich an andere Orte verdrängt wird. Durch mehrere Teilnehmer wurde angemerkt, dass die Forderungen nach solch aufwändigen Evaluierungen von Predictive-Policing-Systemen überzogen und durch das politisch-mediale Interesse induziert sein könnten. Derlei Forderungen werden für andere Datenverarbeitungssysteme, die viel kostspieliger und im Grunde für die gesamte Polizeiarbeit erfolgskritisch sind, üblicherweise nicht erhoben. Dennoch ist festzuhalten, dass eine exemplarische State-of-the-art-Evaluierung eines Predictive-Policing-Systems für alle Polizeien in Deutschland von großem Interesse zu sein scheint. Von daher stellt sich die Frage: Wer übernimmt diese Kärnerarbeit?² Insgesamt war sich die Gruppe einig, dass es im Grunde unmöglich ist, Wirkungen von Predictive Policing im streng wissenschaftlichen Sinne nachzuweisen.

3. Auch wenn ortsbezogene Modelle keine Individual-Daten nutzen, besteht aufgrund der relativen Homogenität von Wohnquartieren ein Risiko im Hinblick auf gruppenbezogene Diskriminierung; historische Kriminalitätsdaten können entsprechende Verzerrungen aufweisen.

Diese These konnte **teilweise bestätigt** werden. Es wurde darauf hingewiesen, dass solche Effekte auch schon bei der eher intuitiven Polizeiarbeit vorgekommen sein dürften. Ferner wurde darauf verwiesen, dass Einbrüche und Einbruchsserien nicht ausschließlich in bestimmten sozialen Milieus und Wohnquartieren vorkommen. Das Risiko für Diskriminierung besteht vor allem, wenn systematische Verzerrungen in der Datenbasis aufgrund hoher Dunkelziffern besteht. Da dies bei Wohnungseinbrüchen nicht der Fall ist, ist das Risiko der gruppenbezogenen Diskriminierung eher niedrig.

4. Transparenz und Nachvollziehbarkeit von Systemarchitektur und Systemfunktionen sind von Software, die in einem derart sensiblen Bereich von der öffentlichen Hand eingesetzt wird, grundsätzlich zu verlangen; dagegen sprechen könnten allein:
 - a. die Wahrung von Geschäftsgeheimnissen (auch andere Software-Anbieter müssen aus diesem Grund die Funktionsweise ihrer Programme nicht offenlegen);
 - b. die Gefahr, dass die Systeme umgangen werden bzw. dass Täter ihr Verhalten entsprechend anpassen („gaming the system“);

² Einen vielversprechenden Anfang stellt die Evaluierung des Precobs-Einsatzes in Baden-Württemberg durch das Max-Planck-Institut für internationales Strafrecht in Freiburg dar, wobei hier eben beispielsweise keine Kontrollgruppenexperimente durchgeführt werden konnten. Vgl. Gerstner, >>>

- c. allgemeine staatliche Geheimhaltungsinteressen wie Geheimschutz oder die Vermeidung des Vorgehens behördlicher Entscheidungen.

These 4 wurde mit dem Hinweis **zugestimmt**, dass die im Einsatz befindlichen Systeme mindestens hinsichtlich ihrer Funktionalität und ihrer Datenquellen transparent und nachvollziehbar seien. Das in NRW durch das Ministerium des Inneren entwickelt SKALA ist für die Behörden natürlich vollständig transparent. Dieser privilegierte Zugang wurde u.a. dafür genutzt, den ursprünglichen Entscheidungsbaum-Algorithmus gegen ein neuronales Netz auf Performanz zu testen; da das neue, auf neuronalen Netzen basierende System nicht signifikant besser war, hat man sich für die Beibehaltung des Entscheidungsbaum-Algorithmus entschieden, weil dessen Funktion besser nachvollziehbar ist. Dieses Vorgehen, das den Behörden in NRW nur deshalb möglich ist, weil es sich erstens um eine Eigenentwicklung handelt und weil man zweitens Mitarbeiter mit profundem technischen Sachverstand (Programmierer und Data Scientists) eingestellt hat, kann als Best Practice gelten.

Block C: Erweiterungs- und Übertragungsmöglichkeiten

Zu prüfende These war:

1. In seiner aktuell praktizierten Form ist Predictive Policing nicht viel mehr als gut aufbereitete Einbruchstatistik. Dennoch ist es sinnvoll, sich damit jetzt intensiv kritisch zu befassen, weil davon auszugehen ist, dass derlei Techniken verbreitet zum Einsatz kommen, erweitert und auf noch sensiblere Bereiche ausgeweitet werden.

Der erste Teil der These trifft nur auf einen Teil der im Einsatz befindlichen Systeme zu, nicht auf die komplexeren Anwendungen wie z.B. SKALA, das bereits recht weitgehende prognostische Ergebnisse liefert. Und es ist sehr wahrscheinlich, dass die Entwicklung flächendeckend in Richtung solcher Systeme gehen wird. In der Schweiz sind insgesamt 20 Systeme im Einsatz, mit denen die Gefahr körperlicher Gewaltanwendung, die von Individuen ausgeht, vorausgesagt wird.³ In den USA hat es eine intensive Debatte um eine Prognosesoftware zur Einstufung des häuslichen Gefahrenpotentials für Kinder gegeben.⁴ Diese Beispiele deuten an, wohin die Reise geht und wie wichtig es ist, dass wir uns jetzt kritisch Aufschluss darüber geben, was solche Systeme leisten können und was nicht und welchen Design- und Implementierungsprinzipien sie folgen sollten – womit wir beim diesbezüglichen Schlussabschnitt sind.

Verallgemeinerbare Befunde

Ziel des Projekts „Algorithmen fürs Gemeinwohl“ ist es, anhand der näheren Betrachtung dreier Anwendungsbereiche von prädiktiven, teilhaberelevanten Algorithmen gemeinwohlorientierte Designprinzipien für solche Systeme zu formulieren. Predictive Policing war das erste dieser Vertiefungsthemen, aus dem wir folgende, unserer Einschätzung nach möglicherweise verallgemeinerbare Befunde⁵ mitnehmen:

³ Vgl. <https://www.srf.ch/news/schweiz/predictive-policing-polizei-software-verdaechtigt-zwei-von-drei-personen-falsch>

⁴ Vgl. <https://algorithmenethik.de/2018/01/22/die-algorithmische-kristallkugel-der-us-kinderschutzbehoerden/>

⁵ Dies ist natürlich im Zuge der Bearbeitung der beiden weiteren Themen im Sinne des SNV-Ansatzes der „Radikalen Prüfung“ auf die Probe zu stellen.

- Die mediale und/oder politische Mode eines Themas kann der Entwicklung und Anwendung guter prädiktiver Software im Weg stehen.
- Der systematische Einbezug von Ethiker:innen, Menschenrechtler:innen und Sozialwissenschaftler:innen erfolgt oft in einer zu späten Projektphase bzw. erst ex post im Zuge der Evaluierung; die Demystifizierung von Algorithmen im intersektoralen Austausch ex ante ist jedoch von großer Bedeutung.
- Der Aufwand für die institutionelle Verankerung und soziale Einbettung wird häufig unterschätzt, was sowohl der internen als auch der gesellschaftlichen Akzeptanz und letztlich auch dem Erfolg prädiktiver algorithmischer Systeme abträglich ist.
- Algorithmische Analyse- und Entscheidungssysteme sind immer nur so gut wie ihre Datenbasis; daher lohnt es sich, diese stets kritisch auf Verbesserungspotenziale zu prüfen (Datenmenge und -qualität, möglichst leichter permanenter Zugang, Aktualität, Erkennen und Eliminieren von verstecktem Bias etc.). Auch die Art und Weise der Datenerhebung ist relevant – besonders wenn die entsprechenden Systeme eine häufige Aktualisierung der Daten erfordern.
- Der Aufwand, der für die Bereinigung und Aufbereitung von Daten betrieben werden muss, wird leicht unterschätzt; es sollten dafür aber ausreichend Ressourcen und Expertise einplant werden.
- Es zahlt sich sehr aus, bei jedem der Software-Entwicklungsprozessschritte – Modellierung und Kodierung (Algorithmenentwicklung im engeren Sinne), System-Kalibrierung und System-Implementierung – jeweils große Sorgfalt walten zu lassen und gegebenenfalls auch einen hohen Aufwand nicht zu scheuen.
- Der Aufbau von technischer In-house-Expertise – durch angestellte Data Scientists und Programmierer – ist in diesem Zusammenhang wünschenswert (und durchaus möglich, wie das Beispiel NRW zeigt).
- Eine Vielfalt von unterschiedlichen im Einsatz befindlichen bzw. getesteten Systemen ist dann von großem gesellschaftlichen Nutzen, wenn ein wechselseitiger Erfahrungsaustausch seitens der Systemhersteller, Anwender und Betroffenen stattfindet.
- Eine umfassende Evaluation muss sowohl die Prognose-Genauigkeit des Algorithmus, als auch die Art und Weise, wie das System genutzt wird, untersuchen; dabei müssen soziale Wechselwirkungen betrachtet werden, weshalb eine Evaluation mit Langzeitlaufzeit und Kontrollgruppen das Ideal darstellt.
- Eine sinnvolle Wirkungsmessung kann nur dann vorgenommen werden, wenn zuvor konkrete Ziele definiert wurden; das ist zwar trivial, wird aber oft vernachlässigt.