Международные процессы, Том 17, № 4 (59), сс. 153–177
DOI 10.17994/IT.2019.17.4.59.9

ПИСЬМО В РЕДАКЦИЮ

# МЕЖДУНАРОДНЫЙ ОПЫТ ПРИМЕНЕНИЯ МАТЕМАТИКО-СТАТИСТИЧЕСКИХ АЛГОРИТМОВ ПРОГНОЗИРОВАНИЯ ПРЕСТУПНОСТИ

АЛЕКСЕЙ ТУРОБОВ
МАРИЯ ЧУМАКОВА
АЛЕКСАНДР ВЕЧЕРИН
Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

**Резюме**

Сфера обеспечения безопасности наполняется новыми элементами (например, кибербезопасность, информационная безопасность, безопасность компьютерных сетей и т.д.); расширяется арсенал средств обеспечения безопасности (технологии, а также технические и организационные средства, включая телекоммуникационные каналы для сбора, формирования, обработки, передачи или приёма информации об угрозах безопасности и мерах по её укреплению), которые значительно укрепляются за счёт использования цифровых технологий. В данной работе проводится анализ современных методов и технологий прогнозирования преступности, применяемых в области национальной безопасности. Достижения в сфере науки о данных (Data Science) и работы с большими данными (Big Data) заложили научную основу для развития интеллектуального анализа данных (Intellectual Analysis, Predictive Analysis), на основании которого сформировалось математико-статистическое прогнозирование общественно опасных преступных деяний (антитеррористические алгоритмы, алгоритмы прогнозирования деятельности организованной преступности/банд). Цель статьи заключается в выявлении основных тенденций и потенциальных выгод применения цифровых технологий, а также определение вызовов, стоящих перед государствами при использовании математико-статистических методов прогнозирования преступности. Посредством мета-анализа научных разработок и практического применения алгоритмов прогнозирования преступности в разных странах (США, Китай, Япония, Сингапур, Индия) демонстрируется разнообразие подходов в применении прогностических систем. В первой части статьи представлены методологические и технические аспекты применения алгоритмов. Вторая часть содержит обзор национальных практик использования алгоритмов прогнозирования преступности в Индии, Японии и Сингапуре. Третья и четвёртая части посвящены более детальному рассмотрению стратегий и практик применения алгоритмов в США и Китае соответственно. Выбор стран-кейсов Индии, Японии и Сингапура определяется высокими показателями в различных инновационных и технологических рейтингах стран мира. Китай и

Дата поступления рукописи в редакцию: 13.08.2019
Дата принятия к публикации: 30.11.2019

*Для связи с авторами / Corresponding author:*
Email: alturobov@yahoo.com

США имеют большие технологические экономики, располагающие наиболее развитыми цифровыми технологиями. В результате метаанализа выявлены риски и выгоды применения математико-статистических алгоритмов прогнозирования преступности, в числе которых: «милитаризация» гражданской сферы; игнорирование социальных, культурных и политических аспектов жизни обществ, из-за чего утрачивается точность статистического прогноза; использование исторических данных (зарегистрированные преступления) содержат изначально заложенные расовые, половые, конъектурные предрассудки; существующие подходы не учитывают личностные особенности субъекта, также процессы принятия решения о совершении противоправных действий; отсутствие государственного контроля за соблюдением баланса между использованием алгоритмов и соблюдением прав граждан.

**Ключевые слова:**

безопасность; прогнозирование преступности; данные; алгоритмы; цифровизация.

В современной международной повестке дня всё, что связано с цифровизацией (или «цифровой трансформацией»), занимает особое место. Несмотря на продолжающиеся дискуссии о содержании и последствиях цифровизации на различных академических, экспертных и профессиональных площадках[1], сложился консенсус относительно того, что цифровизация создаёт большие возможности[2] для позитивных изменений в разных сферах жизни общества[3]. Принятие и реализация решений политического характера не исключение. Правительства становятся всё более осведомлёнными о технологических возможностях и стремятся адаптироваться к ним, и наука о данных (являющаяся одной из основ стремительной цифровизации) начинает играть важную роль в государственной политике и управлении [Cornish 2010]. Взаимодействия достижений научно-технического прогресса и сферы политики и государственного управления являются самостоятельным объектом исследовательского интереса политической науки [Tufekci 2014; Wanna 2018; Hecker, Haklay и др. 2018], причём этот интерес несоразмерен (существенно меньше) уже оказываемому влиянию цифровых технологий на государства и общества.

Мир стремительно меняется, и далеко не все (или *редкие*) изменения ожидаемы (как показала, например, пандемия 2020 года). При этом одной из основных функций современного государства, причём вне зависимости от типа политического режима, остаётся обеспечение безопасности граждан и самого себя в целом, что предусматривает обеспечение безопасных для жизни и развития условий, в том числе предупреждение и недопущение смертельных рисков и опасностей. Демократические государства стремятся обеспечивать безопасность всех граждан, декларируя приоритетность этой задачи и расширяя спектр угроз для противодействия; практики недемократических государств более сложные и разнообразные. Эффективность реализации функции обеспечения безопасности влияет на восприятие гражданами легитимности правительства, политических институтов и игроков. Более того, успехи в обеспечении национальной безопасности транслируются государствами в международное пространство для демонстрации своих политических, управленческих, технологических возможностей.

В последние годы всё больше внимания при формировании, регламентации и непосредственном обеспечении безопасно-

[1] Например, Culture Digitally project (National Science Foundation) — дискуссия о термине и его содержании для будущей Международной энциклопедии теории и философии коммуникации. URL: http://culturedigitally.org/2014/09/digitalization-and-digitization/

[2] World Economic Forum. Digital Transformation Initiative. URL: http://reports.weforum.org/digital-transformation

[3] Обзоры ключевых отраслей и рынков, Институт «Центр развития» НИУ ВШЭ. URL: https://dcenter.hse.ru/otrasli

сти уделяется информационной и цифровой безопасности. Результат – сфера обеспечения безопасности наполняется новыми элементами, такими как информационная безопасность, кибербезопасность, безопасность компьютерных сетей. Соответственно, расширяется арсенал средств обеспечения безопасности (технологии, а также технические, программные, лингвистические, правовые, организационные средства, включая телекоммуникационные каналы, используемые в системе обеспечения национальной безопасности для сбора, формирования, обработки, передачи или приёма информации о состоянии национальной безопасности и мерах по её укреплению), которые значительно укрепляются за счёт использования цифровых технологий. Способность государства самостоятельно (что не исключает разумное и взаимовыгодное сотрудничество с другими государствами и частным бизнесом) обеспечивать безопасность – это элемент его суверенности и независимости [Vohra 2012], в связи с чем цифровые технологии и способности осуществлять цифровую трансформацию могут рассматриваться в качестве факторов государственной состоятельности [Besley, Persson 2009; Roy, Shah, Srikisha 2018; Peeters, Widlak 2018].

Данная статья фокусируется исключительно на одном элементе цифровых технологий – алгоритмах, и на одном элементе системы обеспечения национальной безопасности – предупреждении преступности. Рассматривая «вовлечённость» цифровых технологий в сферу обеспечения безопасности, следует помнить, что обеспечение безопасности, которая волнует обычного гражданина в первую очередь, напрямую связано с предупреждением и недопущением совершения преступлений. Предупреждение преступности преследует две базовые цели: обеспечение безопасного общества и недопущение наступления негативных последствий от преступной деятельности.

В целом литература, освещающая проблематику применения математико-статистических алгоритмов в борьбе с преступностью, не содержит системного обзора разнообразных подходов и их результатов. Главная цель настоящей статьи – определить основные тенденции и выгоды применения цифровых технологий, а также вызовы, с которыми сталкиваются государства при использовании математико-статистических методов прогнозирования преступности. Для достижения указанной цели будут проанализированы существующие научно-исследовательские разработки и свидетельства практического применения алгоритмов в разных странах мира. Первая часть статьи характеризует методологические и технические аспекты применения алгоритмов. Вторая часть содержит обзор национальных практик прогнозирования преступности в Индии, Японии и Сингапуре. Третья и четвёртая части посвящены более детальному рассмотрению стратегий и практик применения алгоритмов в США и Китае соответственно. Выбор Индии, Японии и Сингапура в качестве стран-кейсов определяется тем, что они имеют высокие показатели в различных рейтингах[4] инновационных и технологических стран мира, а Китай и США – это большие технологические экономики с наибольшим опытом применения цифровых технологий в разных сферах жизни. Отметим, что данное исследование не ставит своей целью реализацию сравнительного исследования, для этого потребовался бы иной дизайн и крайне сложный сбор эмпирических данных (на сегодняшний день отсутствует возможность сбора полноценных и релевантных данных по странам по данной тематике).

[4] Например: Рейтинг Всемирной организации интеллектуальной собственности (WIPO) – Global Innovation Index 2019. URL: https://www.wipo.int/global_innovation_index/en/2019/ (Индия, Япония и Сингапур признаются лидерами по инновациям в регионе); Рейтинг Всемирного экономического форума (WEF) – The Most Innovative Economies in the World 2019 и Global Competitiveness Index 2019. URL: https://www.weforum.org/agenda/2020/02/most-innovative-economies-global/ (Сингапур и Япония входят в десятку стран-лидеров, а Индия определяется как страна с высоким потенциалом).

Безусловно, анализ одних только алгоритмов не позволяет предложить универсальные выводы о социально-политических эффектах применения технологий. В этой связи обзор конкретных алгоритмов подкрепляется анализом национальных стратегий (как проявления актов политической воли).

Таким образом, статья представляет собой метаанализ научных разработок и практического применения алгоритмов прогнозирования преступности в разных странах с определением рисков и выгод, присущих большинству рассматриваемых технологий. Тем самым она поможет структурировать имеющиеся подходы и обратить внимание на важные дискуссии относительно практического отправления политики в сфере обеспечения безопасности.

## 1

Интеллектуальный анализ данных преступности (Intelligent Crime Data Analysis[5]) призван оценить динамику противоправных действий, выявить закономерности преступного поведения, которые помогают ответить на три вопроса – где, когда и почему могут возникать конкретные преступления. В силу своей постулируемой прикладной полезности интеллектуальный анализ данных преступности и привлекает внимание учёных и практиков[6] [Mande et al 2012]. Один из вызовов работы с большими объёмами информации состоит в определении наиболее эффективных способов предоставления данных сотрудникам правоохранительных органов и специальных служб. Несвоевременная передача или недостаточная систематизация, неполнота сведений создают трудности для их работы. Напротив, сбор и обработка данных современными техническими способами на основе корректной методологии повышают раскрываемость преступлений, способствуя выполнению задач государства по поддержанию порядка и обеспечению безопасности граждан.

Непосредственно методологические, технические и технологические новеллы работы с данными всё активнее проявляются в предиктивной полицейской деятельности. Предиктивная полицейская деятельность определяется как «применение аналитических методов для определения вероятных целей вмешательства полиции, предотвращения преступности либо для раскрытия преступлений прошлых лет путём составления статистических прогнозов» [Lum, Isaac 2016]. Хотя обсуждаемый термин имеет различные концептуальные определения, предиктивная полицейская деятельность обычно состоит из двух элементов: модели прогнозирования, которая использует алгоритм для выявления случаев повышенного риска преступности, и связанной с моделью стратегией профилактики для смягчения и/или снижения этих рисков. Используя расширенную аналитику, полицейские подразделения и специальные службы могут более эффективно определять будущие цели преступников для превентивного вмешательства.

Более конкретно в данной статье под прогностической (предиктивной) полицейской деятельностью будет пониматься использование данных для прогнозирования и предотвращения преступлений и снижения общего уровня преступности посредством проведения профилактических мероприятий. Прогнозы часто касаются времени и места, в которых может произойти преступление, но также могут заключаться в установлении того, кто может быть преступником или жертвой (демографи-

[5] Interpol. Crime Intelligence analysis. URL: https://www.interpol.int/INTERPOL-expertise/Criminal-Intelligence-analysis

[6] United Nations Office on Drugs and Crime. Criminal Intelligence Manual for Analysts. URL: https://www.unodc.org/documents/organized-crime/Law-Enforcement/Criminal_Intelligence_for_Analysts.pdf

International Journal of Engineering Research and Applications (IJERA). AN INTELLIGENT ANALYSIS OF CRIME DATA USING DATA MINING & AUTO CORRELATION MODELS. URL: https://www.ijera.com/papers/Vol2_issue4/U24149153.pdf

ческое и девиантное профилирование). С помощью методик обработки, анализа и прогнозирования данных правоохранительные органы и специальные службы имеют возможность отслеживать собственные действия и действия преступников с высоким разрешением и с высокой точностью, в результате получая в своё распоряжение современные аналитические методы для разработки оптимальных стратегий по обеспечению безопасности.

Интеллектуальный анализ данных как таковой представляет собой совокупность математико-статистических методов поиска закономерностей в данных для выявления значимых переменных (предикторов), связанных с целевой переменной. На основании выявленных взаимосвязей делается прогноз целевой переменной. Важным результатом такого анализа также является оценка значимости каждого предиктора в итоговой модели. Предсказательный интеллектуальный анализ данных обычно представляет собой итерационный трёхэтапный процесс, включающий в себя построение, оценку и использование модели:

– этап построения модели: тренировочный набор данных, состоящий из элементов с известной классовой принадлежностью (например, преступник/непреступник), подаётся на алгоритм прогнозирования для подбора таких параметров модели («обучение модели»), чтобы результаты её предсказания классовой принадлежности максимально соответствовали реальности (первоначальному классу элемента). Примеры подобных алгоритмов – различные искусственные нейронные сети (ANN), деревья решений (DT)[7] и классификаторы (SVM), используемые для прогнозирования или классификации данных;

– этап оценки модели: результаты предсказания моделью классовой принадлежности элемента сравниваются с его изначально известной классовой принадлежностью для определения точности и, следовательно, полезности оцениваемой модели для дальнейшего практического применения. Следует заметить, что для такой проверки точности модели существуют ограничения: 1) модель может очень точно описать существующие данные, но будет показывать низкую точность на новых данных, так как учитывает редкие взаимосвязи («переобученная модель»); 2) модель не учитывает социальные и политические изменения, так как опирается на предшествующие данные, в которых изменение не отражено должным образом (проблема архивных/исторических данных); 3) модель учитывает социально-демографические данные (например, этническую преступность), что потенциально может привести к устойчивой дискриминации отдельных категорий населения;

– этап непосредственного использования модели на практике, например в работе полицейских департаментов. Для данного этапа крайне важным представляется качественный анализ выявленных взаимосвязей, который должен осуществляться при тесном взаимодействии разработчиков прогнозной модели, сотрудников правоохранительных органов и экспертов в области анализа социальных и политических процессов.

Сегодня в открытых источниках можно найти информацию о более чем двух тысячах математико-статистических алгоритмов прогнозирования и предупреждения преступности[8]. Многие из них были апробированы в реальной работе полицейских управлений в разных странах мира. Порог точности алгоритмов составляет 80–93%[9].

[7] Алгоритм дерева решений является распространённым средством поддержки принятия решений в статистике, анализе данных и машинном обучении. Выделяют множество видов алгоритмов, но все они относятся к деревьям классификации и/или регрессии (Classification and Regression Tree).

[8] Ресурс GitHub. https://github.com/search?q=predictive+crimes; ресурс Kaggle. https://www.kaggle.com/c/predict-10-key-crime-incidences-for-10-districts-in-karnataka
https://www.kaggle.com/c/sf-crime

[9] Статистическая точность – степень соответствия величины какого-либо показателя, определённого по материалам статистического наблюдения, действительной его величине.

При анализе применения математических и статистических моделей для прогнозирования преступлений выявляется множественность подходов. Существуют четыре типа алгоритмов, которые используются для проведения анализа в области построения прогностических моделей:

1) алгоритмы классификации используют атрибуты (предикторы) в наборе данных для прогнозирования значений одной или нескольких дискретных (качественных) переменных;

2) алгоритмы регрессии используют атрибуты в наборе данных для прогнозирования значений одной или нескольких переменных, принимающих непрерывные значения (например, прибыль/убыток). Данный статистический инструмент используется для исследования отношений между переменными;

3) алгоритмы сегментации (кластеризации) разделяют данные на группы или кластеры элементов, имеющие похожие свойства;

4) алгоритмы ассоциации осуществляют нахождение корреляций между различными атрибутами в наборе данных. Типичное применение таких алгоритмов предполагает создание правил ассоциативных связей, которые могут быть использованы при анализе общей совокупности данных.

Каждый из предложенных алгоритмов имеет как достоинства, так и недостатки, которые проявляются при решении каждой конкретной задачи. Некоторые типы методов интеллектуального анализа данных могут быть более подходящими для определённых типов проблем, явлений или наборов данных. Например, при использовании технологии распознавания лиц используются алгоритмы классификации (соответствие меток-структуры лица с фотовидеозаписи с метками-структурой лица из базы розыска). Оценки точности прогнозирования в различных сравнительных исследованиях [Ang, Goh 2013] варьируются от сравнительно низких 80% до экстремально высоких 98%, и эти методы интеллектуального анализа данных могут быть полезными инструментами для изучения проблем преступности.

Ниже представлена более подробная последовательность действий при апробации моделей прогнозирования преступности:

– извлечение данных из различных источников статистической информации правоохранительных органов, государственных статистических бюро, социальных сетей и др.;

– предварительная обработка данных (preprocessing – систематизация и структурирование) с целью построения качественной и точной матрицы данных (очистка, объединение и масштабирование извлечённых данных о преступлениях в структурированные данные с выделением определённых атрибутов-предикторов преступности);

– непосредственная обработка данных алгоритмическими методами (одними из самых распространённых в данной отрасли является кластеризация – разделение данных на группы и элементы, имеющие схожие свойства, и классификация – разделение данных на классы и элементы по чётко определённым (заданным алгоритмом) свойствам);

– графическая визуализация данных;

– проверка точности результатов при помощи отдельной (тестовой, валидационной) выборки данных.

Такое описание процесса создания и работы со статистическими моделями нужно для того, чтобы зафиксировать важные аспекты цифровизации вообще и в сфере безопасности в частности: (1) каждая задача/проблема требует своей специфичной технологии; (2) в большинстве современных цифровых технологий речь идёт о продвинутых алгоритмах, но не о системах искусственного интеллекта (алгоритмы не мыслят, не осознают, а действуют сугубо в рамках заданных математических правил и на основании конкретных данных). Все последующие примеры алгоритмов в конкретных странах соответствуют указанной «типологии» и процессу апробации алгоритмов – это унифицированный процесс для всех.

## 2

Обзор национальных практик использования алгоритмов прогнозирования преступности мы начнём с практик Сингапура, Японии и Индии. Для каждой страны сначала представлено краткое описание основных политических стратегий и некоторые организационные аспекты их реализации с последующими примерами исследований и/или практического применения алгоритмов прогнозирования преступности.

В основе официально декларируемого государственного курса Сингапура – концепция «умной нации» (Smart Nation[10]), предполагающая, что всем гражданам доступны технологии, предоставляющие значительные возможности. С этой целью создано Правительственное технологическое агентство (GovTech[11]), целью которого является обслуживание граждан и бизнеса путём создания цифровой инфраструктуры, платформ и приложений «Smart Nation», развёртывания ИКТ и развития интернета вещей (Internet of Things – IoT). GovTech разрабатывает общенациональную сенсорную сеть «Smart Nation Sensor Platform»[12] (SNSP), ранее известную как «Smart Nation Platform», с общей инфраструктурой и услугами. По сути целью данной программы является развёртывание сети датчиков (видеокамеры, датчики движения/сейсмической активности, наружного наблюдения) с последующим централизованным анализом для оперативного планирования и обслуживания городской инфраструктуры. Кроме этого, в Сингапуре создан Правительственный центр оперативной безопасности (новый государственный орган – Government Security Operation Centre (SOC), который заменит «Cyber-Watch Centre» с 2019 года[13]), призванный разрабатывать системы искусственного интеллекта для обнаружения киберугроз[14]. При этом новые разработки искусственного интеллекта, кроме улучшения городской инфраструктуры, также должны заниматься прогнозированием и предотвращением преступности в Сингапуре[15]. Следует отметить, что национальная стратегия по искусственному интеллекту Сингапура закладывает теоретико-методологические основания для усиления прогностической аналитики[16].

Сингапур является одной из пока немногих стран мира, которая, кроме разработок и изучения возможностей математико-статистического прогнозирования преступности, реформирует законодательство с целью введения достижений современных технологий в правовое русло. В частности, законы о защите персональных данных (Personal Data Protection Act)[17] и злоупотреблении компьютерными технологиями и кибербезопасности (Computer Misuse and Cybersecurity Act)[18] в основном направлены на правовое регулирование вопросов кибербезопасности, однако содержат и нормы в сфере больших данных и анализа данных. Правительственная инициатива увеличения объёмов данных из

[10] Официальная презентация Правительства Сингапура. URL: https://www.tech.gov.sg/files/media/speeches/2017/05/Factsheet%20Smart%20Nation%20Sensor%20Platform.pdf

[11] Официальный сайт Singapore Government Agency https://www.tech.gov.sg

[12] Официальный сайт Правительства Сингапура. URL: https://www.tech.gov.sg/products-and-services/smart-nation-sensor-platform/

Официальный сайт платформы. URL: https://www.tech.gov.sg/products-and-services/smart-nation-sensor-platform/

[13] Официальный сайт. URL: https://www.csa.gov.sg

[14] Tham I. New govt centre to detect cyber threats. The Strait Times. 25.05.2017. https://www.straitstimes.com/tech/new-govt-centre-to-detect-cyber-threats

[15] https://www.straitstimes.com/singapore/artificial-intelligence-plays-key-role-in-securing-smart-cities-but-people-are-sensors-too

[16] Официальный сайт стратегии. URL: https://www.aisingapore.org

[17] Официальный сайт Государственной комиссии Сингапура по персональным данным: https://www.pdpc.gov.sg

[18] Актуальная официальная версия документа на официальном сайте Правительства Сингапура: https://sso.agc.gov.sg/Act/CMA1993

различных источников, централизованный анализ получаемых данных, постепенное нормативное регулирование этой области, а также пропаганда посредством СМИ превентивной профилактики преступности при помощи алгоритмов прогнозирования – всё это элементы планомерной подготовки граждан Сингапура к IT-реформированию сферы обеспечения безопасности (и, вероятно, государственного управления вообще).

Сингапурское исследование оценки риска подросткового рецидива на примере насильственных действий сексуального характера [Chu et al. 2012] установило, что разработанные методики «ERASOR», «J-SOAP-II», «YLS / CMI» демонстрируют статистически значимые показатели прогнозирования вероятности совершения рецидивного преступления. На протяжении 4,5 лет исследования сравнение прогностической достоверности указанных методик продемонстрировало высокий уровень точности. Более того, в результате исследователями делается вывод, что оценочные меры (аналогичные исследуемым) могут использоваться при прогнозировании иных составов преступлений.

***Япония*** также имеет национальную стратегию цифровой трансформации[19] как экономики, так и сферы политики и государственного управления. В сфере безопасности в стране функционирует Центр безопасности системы управления (Control System Security Center[20] – CSSC), сформированный по указанию министра экономики, торговли и промышленности в соответствии с «Актом о партнёрстве в области исследований и разработок». В сферу компетенций центра входит проверка безопасности цифровых и инфраструктурных систем, создание цифровых систем и технологий для повышения безопасности (в том числе повышение безопасности систем управления), сертификация информационных и цифровых систем, а также международная стандартизация технологий в сфере безопасности. Отдельно внимание уделено кибербезопасности: с 2015 г. функционирует Национальный центр готовности к инцидентам и Стратегия кибербезопасности (National Center of Incident Readiness and Strategy for Cybersecurity[21] – NISC) для создания «свободного, справедливого и безопасного киберпространства». Указанный центр примечателен тем, что он предоставляет правоохранительным органам аналитическую информацию для предупреждения преступлений и вне интернет-пространства.

Японские исследователи изучили возможность трёхмерного отображения событий-преступлений в пространстве-кубе с помощью пространственно-временных вариантов оценки плотности [Nakaya, Yano 2010]. Уникальность подхода заключается в следующем: создаётся пространственно-временной куб, в котором пространственно-временное распределение плотности преступности можно визуализировать в объёме (3-D модель), в отличие от классического двухмерного представления преступности при помощи зон – «горячих точек»[22]. Предлагаемое трёхмерное отобра-

[19] Japan Digital Transformation Strategies. URL: https://www.idc.com/getdoc.jsp?containerId=IDC_P38604; Report "Digital Transformation & Innovation in Japan" from Deloitte. URL: https://www2.deloitte.com/jp/en/pages/financial-services/solutions/lc/en-dti.html; Report «Digital Economy in Japan and the Eu» from EU-Japan Centre for Industrial Cooperation. URL: https://www.eu-japan.eu/sites/default/files/publications/docs/digitaleconomy_final.pdf

[20] Control System Security Center. URL: http://www.css-center.or.jp/en/

[21] National Center of Incident Readiness and Strategy for Cybersecurity. URL: https://www.nisc.go.jp/eng/

[22] Карты «горячих точек» (англ. hot spot) – традиционный метод анализа и визуализации распределения преступлений в пространстве и времени. Соответствующие методы включают оценку плотности ядра (KDE), которая соответствует двухмерной пространственной функции плотности вероятности для архивной записи о преступлениях. Этот подход позволяет визуализировать районы с высокой концентрацией преступности в прошлом. Будущие преступления часто происходят в непосредственной близости от предыдущих, что делает карты «горячих точек» ценным инструментом

жение преступлений позволяет эффективно визуализировать геопространственную протяжённость и временную продолжительность для локализации преступлений.

По сравнению с традиционным временным отображением преступности, выполненным с использованием карт поперечного сечения с произвольными временными интервалами, 3D-метод с использованием пространственно-временного куба особенно полезен для отслеживания динамики перемещения преступности. Иными словами, каждое преступление рассматривается уже не как статичная точка на карте, а учитывается возможность территориальной протяжённости преступного деяния.

В результате применения алгоритма были выявлены последовательные стабильные кластеры (зоны) в центральной части города и вокруг вокзала Киото, а также «переходные» (случайные/временные) кластеры вокруг нескольких пригородных железнодорожных станций. Временные различия в «переходных» кластерах указывают на тот факт, что преступность в масштабе мигрирует, а не жёстко закреплена в единой зоне, определяемой правоохранительными органами как «горячая точка».

Разработка и внедрение математико-статистических моделей прогнозирования преступности доступны не только развитым странам с высоким уровнем жизни и качеством государственного управления. В частности, в ***Индии*** запущен масштабный национальный план «Цифровая Индия» (Digital India[23]), разработанный в 2014 г. и ставящий целью цифровую трансформацию страны и расширение возможностей граждан в этом процессе. В рамках данного плана разработано и внедрено множество технологических решений, начиная от развития доступа к государственным услугам при помощи Интернета – *Aadhaar*[24] и заканчивая цифровой идентификацией всего населения в рамках проекта *India Stack*[25] (представляет собой набор функционально совместимых программных уровней, поддерживающих цифровые платежи, проверенные безбумажные записи, деловые и сервисные транзакции, а также всю пользовательскую информацию из *Aadhaar*). Несмотря на невысокие практические результаты реализации плана «Цифровая Индия» большинство отчётов и исследователей указывают на стремительное развитие технологий и их внедрение в социальную, экономическую и политическую сферы.

В контексте применения цифровых технологий в сфере обеспечения безопасности в Индии учреждены различные центры и ведомства по работе с данными, технологиями и Интернетом (Data Security Council of India[26], Indian Computer Emergency Response Team[27], Cyber Coordination Centre[28], Cyber and Information Security (C&Is) Division[29]), а также многочисленные

прогнозирования преступности. Более продвинутые методы, такие как самовозбуждающиеся модели точечных процессов, также фиксируют пространственно-временную кластеризацию преступных событий. Эти технологии полезны, но имеют ограничения. *Во-первых*, они локальны, следовательно, модель «горячей точки» для одной географической области нельзя автоматически перенести для характеристики другой географической области. *Во-вторых*, им требуются данные о предыдущих преступлениях. То есть они не могут быть построены для областей, которые не имеют таких данных. *В-третьих*, они не рассматривают комплексный социальный ландшафт области при анализе криминального поведения.

[23] Digital India Plan. URL: https://www.digitalindia.gov.in

[24] Официальный сайт: https://uidai.gov.in/my-aadhaar/avail-aadhaar-services.html

Обзор о возможностях и целях Aadhaar: https://economictimes.indiatimes.com/wealth/personal-finance-news/aadhaar-everything-you-need-to-know-about-it/articleshow/60173210.cms

[25] https://www.indiastack.org/about/

[26] Официальный сайт: https://www.dsci.in

[27] Официальный сайт: https://www.cert-in.org.in/s2cMainServlet?pageid=PRESSLIST

[28] Официальный пресс-релиз Правительства Индии: https://pib.gov.in/Pressreleaseshare.aspx?PRID=1575751

[29] Официальный сайт: https://www.mha.gov.in/division_of_mha/cyber-and-information-security-cis-division

организации для обеспечения безопасности отдельных объектов. Например, Национальный центр защиты критической информационной инфраструктуры (National Critical Information Infrastructure Protection Centre[30] – NCIIPC) активно внедряет цифровые технологии в прогнозирование и превентивное недопущение нарушений безопасности на основании Закона об информационных технологиях. Совет национальной безопасности Индии (National Security Council[31] – NSC) с 2014 года обязан (во исполнение плана «Цифровая Индия») активно применять весь возможный потенциал цифровых технологий для реализации стратегии национальной безопасности.

Исследователи из Индийского университета науки и технологии разработали алгоритм прогнозирования преступности *Apriori* [Bansal, Bhambhu 2013] с последующей локальной апробацией. На наборе данных, содержащих информацию о преступлениях исключительно против женщин (дополнительно использовался инструмент *WEKA*[32] для извлечения результатов статистических анализов из государственных репозиториев), алгоритмом были установлены характеристики потенциальных субъектов преступных деяний: возрастная группа, фактор знакомства преступника и жертвы (незнаком или известен жертве), возрастная группа женщин. В результате были выявлены ассоциации между возрастом преступника, возрастом жертв и фактором их знакомства друг с другом с порогом точности более 80%: преступления совершаются мужчинами 20–24 лет в отношении женщин в возрасте 16–22 лет, и в основном обвиняемые хорошо известны жертве. Последующие тесты алгоритма по составлению прогноза в конкретных регионах продемонстрировали более высокую статистическую точность, что позволило предупредить и не допустить ряд насильственных преступлений против женщин.

Алгоритм *CDCI* [Tayal, Jain, Arora, Agarwal, Gupta, Tyagi 2014] построен по модульной схеме и включает в себя шесть элементов: извлечение данных из статистических репозиториев; предварительная обработка данных (за период 2000–2012 годов все данные представлены в 5038 блоках преступлений); кластеризация обработанных данных с выявлением 35 атрибутов (предикторов) преступности; отображение полученных данных на картах городов и регионов (выделение «горячих точек»); проверка достоверности прогнозов и тестирование результатов. В результате для алгоритма *CDMI* заявляется (его создателями) высочайшая прогностическая мощность – 93,62–93,99% по прогнозированию всех видов насильственных преступлений.

## 3

В странах, являющихся технологическими лидерами, можно найти самые разные примеры использования технических решений науки о данных для обеспечения безопасности[33]. Прежде чем перейти к непосредственному анализу общенациональной стратегии применения алгоритмов в ***США***, кратко рассмотрим отдельные (особые) случаи использования конкретных систем в отдельных городах и штатах. Например, с целью снижения количества ДТП в Сан-Франциско (США)[34] были внедрены непрерывные картографирование и визуализация инцидентов, связанных с дорожным движением в городе, через онлайн-платформу *TransBase*[35] и разрабо-

[30] Официальный сайт: https://nciipc.gov.in

[31] http://www.allgov.com/india/departments/ministry-of-youth-affairs-and-sports/national-security-council?agencyid=7599

[32] Waikato Environment for Knowledge Analysis (Weka) – https://www.cs.waikato.ac.nz/~ml/weka/

[33] Local government. How cities score (2016). The Economist. URL: https://www.economist.com/special-report/2016/03/23/how-cities-score

[34] "Data-Driven Policy": San Francisco just showed us how it should work. URL: https://medium.com/@abhinemani/data-driven-policy-san-francisco-just-showed-us-how-it-should-work-c7725e0e2b40

[35] URL: http://transbasesf.org/transbase/

тана сеть *Vision Zero High Injury Network*[36] на основе полученных данных для локализации основных проблем и определения возможных действий городских властей. Полученные знания были преобразованы в решения о создании «защищённых перекрёстков» (с выделением специальных зон для пешеходов) и защищённых велосипедных дорожек, разграничений «пересечений» опасных участков.

Указанный пример снижения количества ДТП в Сан-Франциско напрямую связан с технологиями анализа геоданных. Подавляющее большинство данных (80%) имеет пространственную природу [Гохман, Третьяченко 2019], в связи с чем использование этих данных алгоритмами приводит к созданию продвинутых неконвенциональных решений, в том числе в области прогнозирования преступности. Современные тенденции (распространённость мобильных устройств и возможность быстрого обмена гиперлокальной информацией) влияют на все направления развития алгоритмов. Мобильные приложения и социальные сети превращаются в платформы, предлагающие большие массивы данных пользователей в сочетании с геолокацией, открывают новые возможности для обнаружения противоправных ситуаций. Таким образом, объединение алгоритмических техник и методов и геопространственной аналитики позволяет расширять возможности принятия социально значимых решений и использования предсказательной аналитики в полиции [Гохман, Третьяченко 2019].

Одним из направлений новой стратегии правоохранительной деятельности США является интеллектуальная полицейская деятельность, предусматривающая широкое внедрение современных цифровых технологий. Принимая во внимание, что отделы полиции, особенно крупные, ежедневно генерируют значительные объёмы данных, используется технология *Auto Vehicle Locator* (AVL) [Wang, Zhao 2016], которая кодирует всю входящую информацию с учётом геопространственного положения в реальном времени. Другим примером служит программное обеспечение для патрульных автомобилей от исследователей из Калифорнийского университета в Лос-Анджелесе[37], осуществляющее предсказания возникновения противоправных ситуаций в пределах 46,5 квадратного метра в течение 24 часов (в основе закладывается прогностическая модель землетрясений).

При содействии Национального института правосудия исследователи из Университета Ратгерса разработали программное обеспечение «Модель рельефа местности»[38] для прогнозирования концентрации преступных событий в ближайшие несколько месяцев. Такая работа с полицейской информацией позволяет формировать временные кластеры преступлений в географической области. Отметим, что данные программные продукты являются усовершенствованными системами геоинформационной аналитики с фокусировкой на обеспечении безопасности. Несмотря на то что сами по себе модели рельефов местности алгоритмически не столь интересны, источники данных, а также «чувствительность» обрабатываемой информации (кластеры преступлений с информацией о потенциальных правонарушителях) вызывает множество правовых и этических дискуссий [Lum, Isaac 2016].

Калифорнийский университет в Лос-Анджелесе при финансовой поддержке министерства обороны в начале 2000-х годов адаптировал исследования о прогнозировании потерь на поле боя в Ираке для прогнозирования преступности с последующим созданием компании под названием

[36] URL: http://sfgov.maps.arcgis.com/apps/webappviewer/index.html?

[37] Исследование UCLA разработало метод для снижения преступности в Лос-Анджелесе. URL: http://newsroom.ucla.edu/releases/predictive-policing-substantially-reduces-crime-in-los-angeles-during-months-long-test

[38] Risk Terrain Modelling for Public Security. URL: http://techfinder.rutgers.edu/tech/Risk_Terrain_Modeling_Diagnostics_Software_(RTMDx)

*PredPol llc*[39]. Она[40] быстро стала одним из лидеров в зарождающейся области прогнозирования преступности в 2012 году, но также подверглась сильной критике активистов, которые утверждали, что фирма предоставила своего рода «техническую легитимацию» для расово предвзятых и неэффективных методов полицейской деятельности[41].

В настоящее время *PredPol* разрабатывает методики и программное обеспечение с использованием машинного обучения для анализа криминальных данных полицейских департаментов с целью автоматизации классификации преступлений, связанных с бандами[42]. Прогностическое программное обеспечение этой компании предполагает, что преступления следуют модели «землетрясения после землетрясения» – районы, в которых ранее совершались преступления, скорее всего будут криминализированы, то есть на данной территории с незначительными изменениями будут совершаться схожие преступления. В основе разработанного алгоритма используется процедура максимизации ожиданий для определения параметров модели.

Такая модель использует только данные об инцидентах (включая как обнаруженные, так и сообщенные инциденты) для каждого региона с целью определения истинного уровня преступности и не использует никакого контекста в виде демографических данных, арестов и иных статистических данных. Более конкретно *PredPol* предсказывает, где будут обнаруживаться преступления и в какой локации о преступлениях будут сообщать в правоохранительные органы, а не там, где будет происходить преступление [Ensign, Friedler, Neville, Scheidegger, Venkatasubramanian 2017]. Каждый день сотрудники полиции отправляются в районы с наивысшей прогнозируемой интенсивностью, и полученные данные об обнаруженных инцидентах возвращаются в систему.

Другой проект по прогнозированию преступности – *CrimeScan Program*[43] – реализован в Питтсбурге (США). На ноутбуках в полицейских автомобилях отображаются карты, показывающие места, где может произойти преступление, на основе алгоритмов хэширования данных, разработанных учёными из Университета Карнеги-Меллона. *CrimeScan Program* имеет геопространственный фокус для определения зон-«горячих точек», но опирается на более широкий спектр показателей. В основе работы алгоритма лежит теория, согласно которой преступники, как правило, подчиняются тенденции к «переходу» от незначительных к более тяжким преступлениям. В результате сообщения о мелких преступлениях могут помочь предсказать потенциальные вспышки преступлений большей тяжести. Программа обрабатывает большой объём информации о мелких правонарушениях и звонков в экстренные службы о противоправном поведении, наркотрафике и беспорядках, и такие данные помогают прогнозировать динамику насилия в течение следующих нескольких дней или недель в определённой локации.

Примечателен алгоритм прогнозирования преступности, основанный на принципе частично генерирующей нейронной сети *PGNN* [Seo et al 2018], при помощи которой изучается вопрос о классификации того,

[39] *Winston A., Burrington I.* A pioneer in predictive policing is starting a troubling new project. The Verge. 26.042018. URL: https://www.theverge.com/2018/4/26/17285058/predictive-policing-predpol-pentagon-ai-racial-bias

[40] Official site PredPol. URL: https://www.predpol.com

[41] *Reynolds M.* Biased policing is made worse by errors in pre-crime algorithms. NewScientist. 04.10.2017. URL: https://www.newscientist.com/article/mg23631464-300-biased-policing-is-made-worse-by-errors-in-pre-crime-algorithms/

[42] *Hutson M.* Artificial intelligence could identify gang crimes—and ignite an ethical firestorm. Science. 28.02.2018. URL: http://www.sciencemag.org/news/2018/02/artificial-intelligence-could-identify-gang-crimes-and-ignite-ethical-firestorm

[43] *Hvistendahl M.* Can 'predictive policing' prevent crime before it happens? Science. 28.09.2016. URL: http://www.sciencemag.org/news/2016/09/can-predictive-policing-prevent-crime-it-happens

связано ли конкретное насильственное преступление с деятельностью банд. Такая нейронная сеть способна точно классифицировать преступления, связанные с бандами, как при наличии полной информации, так и при наличии частичной информации. Используя набор данных о противоправных действиях из Лос-Анджелеса, охватывающий 2014–2016 года, экспериментально показано, что *PGNN* превосходит все другие типично используемые классификаторы в решении проблем классификации преступлений, связанных с бандами. Разделив набор сведений на обучающие и тестовые выборки, для решения проблемы несбалансированных данных произвольно выбирается 10% преступлений, связанных с бандами, и такое же количество преступлений, не связанных с бандами (для тестового набора), а последующие массивы данных используются для обучения. Оценка повторяется 100 раз, чтобы получить надёжные результаты. В результате *PGNN* выдаёт прогностическую точность и надёжность на 5–7% выше существующих алгоритмов (по состоянию на 2018–2019), уменьшает временные затраты полицейских управлений для определения, совершаются ли преступления с участием банд, или преступления совершены в отношении членов банд по личным мотивам.

Немало американских городов уже применяют аналогичные системы, которые включают в себя различные данные – от небольших отчётов о преступлениях до профилей преступников в социальных сетях. Опираясь на подходы, происходящие из таких разнообразных областей, как сейсмология и эпидемиология, алгоритмы могут внести лепту в снижение уровня преступности, а также в уменьшение предвзятости в полицейской деятельности.

Важным направлением повышения качества прогностической полицейской деятельности в США стала интеграция данных социальных сетей в проводимый анализ. Временные и пространственные координаты, доступные из пользовательских данных в социальных сетях, отражают деятельность, которая имеет отношение ко многим общественным событиям. Большинство социальных сетей имеют индикаторы геолокации (конкретное место, где было сделано фото/видео или опубликован текст), времени, а также персональные индикаторы пользователя. Такие данные расширяют возможности идентификации места, времени и лиц, присутствующих или находящихся поблизости с «горячей точкой» и/или потенциальным местом совершения преступления. Иными словами, онлайн-социальная активность может значительно повысить эффективность прогностической полицейской деятельности, особенно в местах большого скопления людей в связи с массовыми празднествами, митингами и тому подобными общественными мероприятиями. Использование смартфонов с возможностью загрузки фотографий или видео с точным местоположением наиболее полезно. Примерами платформ для наблюдений за криминогенным состоянием являются *Postacrime*[44], *Spotcrime*[45] или *CrimeReports*[46].

Полезным источником данных для поддержки принятия решений становится *Twitter*. Безусловно, применяется аналитика и из других популярных социальных сетей, однако *Twitter* (1) имеет специфику контента (публикуются короткие и содержательно ёмкие впечатления, мысли и описание конкретных действий/событий); (2) по умолчанию учитывает данные геолокации; (3) пользуется огромной популярностью в США для максимально быстрого распространения информации (яркий пример политической деятельности – аккаунт Президента США Дональда Трампа). Кроме того, благодаря доступности непосредственно сбора данных *Twitter* – один из самых популярных источников для создания прогностических моделей (с опорой на

[44] https://mashable.com/2008/04/08/postacrime/
[45] https://spotcrime.com
[46] https://www.crimereports.com

доработанный принцип распределения Пуассона[47]) и представляет доказательства взаимосвязи между данными твитов и криминальной деятельностью в тесной временной и пространственной смежности. Имея дополнительные данные *Twitter* (геолокация, время и объём твита), происходит углубление прогностического анализа и детально рассматриваются дополнительные функции: геопространственные точки интереса, поведенческие характеристики людей в городах или темы сообщений. Данные социальных сетей, представленные сообщениями *Twitter*, для объяснения и прогнозирования преступной деятельности в «мелкозернистых» (сильно сегментированных) временных и пространственных отношениях могут выявить ценную информацию, превосходящую информационный прирост общего анализа, основанного только на геолокации.

Исследование совокупного трафика твитов по сопоставлению с типологией преступлений [Bendler et al 2014] указывает, что, например, убийства демонстрируют специфические шаблоны твитов[48] до того, как преступление было совершено. Анализ таких специфических шаблонов может повысить точность объяснения преступной деятельности в городских районах. В работе с данными *Twitter* предложен эмпирический подход для анализа взаимосвязи между интерактивным социальным взаимодействием пользователей и преступными инцидентами. Результаты исследования установили корреляцию данных, собранных и проанализированных из *Twitter*, с зафиксированными преступлениями, по факторам времени и местоположения в районах Сан-Франциско. Первичная модель предсказания преступлений, построенная на данных о часовом поле и местоположении, при добавлении данных *Twitter* в алгоритм повысила точность прогнозирования. Таким образом, *Twitter* как дополнительный инструмент анализа общественной деятельности в городской местности оказался ценным дополнением для объяснения и прогнозирования преступных инцидентов.

Похожие результаты повышения точности прогнозирования при помощи анализа данных из *Twitter* продемонстрированы в исследовании, посвящённом использованию твитов с пространственно-временными метками для прогнозирования преступности [Gerber 2014]. Используя лингвистический анализ и статистическое моделирование для автоматического определения темы обсуждения в одном из крупных городов США, удалось повысить эффективность прогноза в среднем на 10%. Полученные массивы данных твитов добавили в модель прогнозирования преступности (стандартная модель, рассмотренная выше, основанная на оценке плотности ядра – геотипология преступности), и в результате для 19 из 25 типов преступлений продемонстрировано повышение пика качества прогноза.

Потенциально развитие и реализация качественных предсказательных моделей преступности могут улучшить распределение всегда ограниченных правоохранительных ресурсов (полицейское патрулирование, временные затраты офицеров), что приведёт к сокращению затраченных усилий и уменьшению времени реагирования на преступления. Однако, несмотря на большое количество реализуемых алгоритмов на территории США, некоторые эксперты утверждают, что алгоритм не должен полагаться только на статистические методы, а комбинировать их с другими подходами. В пример приводится «Стратегия сокращения насилия в Чикаго»[49] – програм-

[47] Распределение Пуассона – вероятностное распределение дискретного типа, моделирует случайную величину, представляя собой число событий, произошедших за фиксированное время, при условии, что данные события происходят с помощью фиксированной средней интенсивности и независимо друга от друга.

[48] Шаблон – определённая структура твит-сообщения.

[49] Официальный сайт программы, Национальная программа безопасного общества. URL: https://nnscommunities.org/impact/city/chicago

ма, идентифицирующая людей, которые рискуют стать либо преступниками, либо жертвами преступлений, а затем получает доступ к оказанным социальным услугам. Другим примером может служить инициатива полицейского департамента Пенсильвании: по мере запуска *CrimeScan* полицейский департамент стремится наладить отношения с сообществами с высоким уровнем преступности и обеспечить, чтобы большие данные использовались для решения проблем, а не просто для того, чтобы сменить фокус деятельности полиции.

## 4

Стратегия борьбы с преступностью в ***КНР*** состоит из четырёх элементов, которые, в свою очередь, основаны на анализе временных факторов событий совершения преступлений [Wang, Zhao 2016]. *Первый элемент* – работа в местных сообществах для мобилизации ресурсов в области предупреждения противоправной деятельности и формирования партнёрства между полицией и обществом. *Вторым элементом* выступает сбор информации от граждан и об их непосредственном окружении с последующим построением прогнозов о возможных противоправных действиях в отдельных микрорайонах (использование методов превентивной (упреждающей) полицейской деятельности). Такие сведения фиксируются в электронной системе, которая выявляет вероятностные характеристики совершения преступления и сигнализирует (устанавливаются «красные флажки») о потенциальной возможности возникновения в определённом районе того или иного преступного деяния. Правоохранительные органы КНР исходят из того, что некоторые виды преступлений (например, кража со взломом) могут быть предсказаны на основе информации о составе населения и прошлых событиях.

*Третья составляющая* – вмешательство полиции, обеспечивающее предотвращение наступления последствий путём своевременного реагирования с использованием всех ресурсов (законодательных, технических, физического присутствия) непосредственно во время преступления. Таким образом, правоохранительные органы осуществляют контроль за текущими событиями и активными действиями стремятся нарушить цикл криминальной активности (в отношении длящихся или структурно сложных преступлений полиция реагирует на этапе подготовки преступления). *Четвёртый элемент* – максимально активное, оперативное и своевременное полицейское реагирование после факта совершения преступления с целью скорейшего разрешения ситуации. Основная цель данного элемента стратегии состоит в повышении «цены» совершения преступления, направленного на недопущение совершения противоправных деяний в будущем. Иными словами, преступному сообществу транслируется сигнал, что полицейские департаменты не допустят не только факты совершения противоправных деяний, но и минимизируют их последствия (а также максимизируют цену совершения преступления для нарушителя).

Реализация такой стратегии требует формирования и поддержания единой базы данных – Интегрированной платформы совместных операций (Integrated Joint Operations Platform[50]). Платформа объединяет разнообразную информацию и данные о гражданах, включая финансово-банковские записи лиц, юридическое прошлое, полицейские отчёты, геолокацию, факты перемещения транспортных средств, данные контрольно-пропускных пунктов, а также данные с различных технических устройств, в том числе с камер видеонаблюдения с функцией распознавания лиц.

Технологии распознавания лиц используются в Китае в сочетании с алгоритмами искусственного интеллекта с целью уведомления полиции о потенциальных пре-

[50] *Ma A.* China uses an intrusive surveillance app to track its Muslim minority, with technology that could be exported to the rest of the world. Here's how it works. Business Insider. 11.05.2019. URL: https://www.businessinsider.com/how-ijop-works-china-surveillance-app-for-muslim-uighurs-2019-5

ступниках на основе различных моделей поведения. Например, исходя из анализа перемещений конкретного лица, штаб-квартира *Cloud Walk*[51] в Гуанчжоу формирует отчёты о потенциальном риске совершения лицом противоправного деяния. Данная технология устанавливается на ключевых объектах критической инфраструктуры и отслеживает людей «с высоким уровнем риска» для оперативного предоставления информации полиции[52].

Китайская компания *UniView*[53] отслеживает людей, которые часто путешествуют в «чувствительные» страны (страны, с которыми у Китая напряжённые отношения), такие как Мьянма и Вьетнам, и автоматически маркируют их. *UniView* утверждает, что может помечать подозрительных субъектов с помощью поведенческого анализа и предупреждать полицию, если на камере наблюдения появился человек, находящийся в «чёрном списке».

Китайские власти стремятся использовать прогностические и оценочные алгоритмы по отслеживанию поведения граждан во многих регионах страны. В столице Чжэнчжоу провинции Хэнань активно применяются технологии распознавания лиц, в Циндао благодаря алгоритмам, анализирующим всё видео с камер наблюдения в городе в реальном времени, обеспечивается безопасность массовых мероприятий (отчёты свидетельствуют как о пресечении массовых беспорядков, так и об успешном задержании лиц, находящихся в розыске)[54]. Более того, результативность мониторинга населения в районах с особым режимом легли в основу разработки и применения масштабного социального рейтинга на территории всей страны [Hoffman 2018]. Система социального рейтинга (Social Credit System) является масштабной инфраструктурой, включающей в себя множество доступных цифровых технологий, начиная от сбора и анализа данных городской инфраструктуры и заканчивая «чувствительными» персональными данными каждого лица[55]. В данной статье мы не будем останавливаться на данной системе (и по причине недостатка достоверных данных, и по причине объёма, сложности самой системы, а также масштаба политических и правовых проблем её применения), лишь обратим внимание на развитие концепции отслеживания и мониторинга населения при помощи цифровых технологий для обеспечения безопасности государства и граждан.

Учитывая специфику политического устройства Китая, представляется сложным систематически проанализировать деятельность правоохранительных органов в сфере профилактики преступности в связи с недостатком открытых и достоверных сведений. На этом фоне отчёт международной неправительственной организации *Human Rights Watch*[56] выглядит самым масштабным исследованием «прогностической политики» правительства Китая на примере округа Кашгар в Синьцзян-Уйгурском автономном районе КНР. Безусловно, опираться только на данные организации, которая систематически критикует Пекин, недостаточно, поэтому в рамках нашего исследования демонстрируется и общий стратегический подход китайских властей

[51] *Burt C.* Universal Beijing to use facial recognition throughout theme park as CloudWalk VP urges privacy balance. 20.11.2019. URL: https://www.biometricupdate.com/201911/universal-beijing-to-use-facial-recognition-throughout-theme-park-as-cloudwalk-vp-urges-privacy-balance

[52] Официальный сайт: http://www.cloudwalk.cn

[53] *Huang E.* What do China's police collect on citizens in order to predict crime? Everything. Quartz. 20.11.2017. URL: https://qz.com/1133504/to-predict-crime-chinas-tracking-medical-histories-cafe-visits-supermarket-membership-human-rights-watch-warns/

[54] *Mozur P.* Inside China's Dystopian Dreams: A.I., Shame and Lots of Cameras. The New York Times. 08.07.2018. URL: https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html

[55] *Botsman R.* Big data meets Big Brother as China moves to rate its citizens. The Wired. 21.10.2017. URL: https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion

[56] China: Big Data Fuels Crackdown in Minority Region. Human Rights Watch. 26.02.2018. URL: https://www.hrw.org/news/2018/02/26/china-big-data-fuels-crackdown-minority-region

(как указано выше на примере национальной стратегии), так и конкретная информация по отдельным нюансам реализации прогностических алгоритмов (данные о компаниях-разработчиках алгоритмов; национальная платформа *Cloud Police*; изменения в ряде нормативно-правовых актов по вопросам оперирования данными). Таким образом, анализ примера Китая подкреплён не только информацией в СМИ, но и официальными источниками и данными исследований.

Вся информация в Синьцзян-Уйгурском автономном районе анализируется Интегрированной платформой совместных операций. Объём обрабатываемой информации платформой необычайно велик: (1) данные с камер видеонаблюдения с распознаванием лиц и инфракрасными датчиками; (2) данные уникальных идентификационных адресов компьютеров, смартфонов и других сетевых устройств; (3) информация об идентификационных номерах граждан и о данных транспортных средств; (4) данные из систем «управления посетителями» в различных социальных группах-сообществах; (5) информация о зарегистрированных правах собственности; (6) данные организаций здравоохранения и планирования семьи; (7) финансово-банковские данные и юридические документы; (8) информация от должностных лиц о любой необычной деятельности; (9) массовый сбор ДНК и биометрии у лиц в возрасте от 12 до 65 лет в определённых регионах (в отчёте указывается Синьцзянь).

Уведомления о закупках для Интегрированной платформы совместных операций показывают, что она тесно связана (в рамках соглашений о поставках и закупках) с компанией «Синьцзян-Лянхай Цанчжи»[57]. Эта фирма является дочерней компанией *China Electronics Technology Group Corporation* (CETC[58]), крупного государственного подрядчика в Китае. На пресс-конференции в марте 2016 г. представители этой компании объявили о получении правительственного контракта на создание единой базы данных по всей стране, которая будет сопоставлять повседневное поведение граждан и фиксировать «необычные действия» для прогнозирования терроризма. Учитывая, что в Китае реализуется национальная программа цифровой системы идентификационных карточек гражданина (ключ доступа ко многим государственным и частным услугам, а также является идентификатором для обширных баз данных личной информации), формируется единая по стране правительственная система цифрового профиля гражданина с возможностями предиктивного анализа криминальной ситуации.

Для работы с большими объёмами данных в сфере общественной безопасности создаётся система «Полицейское облако» (Police Cloud[59]). Эта система опирается на технологии больших данных (Big Data) для функционирования умного города (Smart City) [Yang 2019]. Она основана на алгоритме выявления отношений между событиями и людьми (их реакциями и поведением), «скрытых» от полиции, анализирует и информирует полицию о деятельности, которая может показаться необычной (потенциально противоправной). Огромные массивы информации «Полицейского облака» (согласно отчёту о городе Вэйхай в провинции Шаньдун, интегрируются 63 типа полицейских данных, 115 видов социальных данных, в том числе IP-адреса от телекоммуникационных компаний и учётные данные социальных сетей, а также 43 типа данных государственных ведомств и отраслей) в конечном итоге поступают в качестве кратких сведений на устройства сотрудников полиции с информацией о местоположении подозрительно-

[57] Информация о регистрационных данных компании URL: https://www.emis.com/php/company-profile/CN/Xinjiang_Lianhai_Chuangzhi_Information_Technology_Co_Ltd__新疆联海创智信息科技有限公司__en_5358422.html

[58] Официальный сайт компании URL: http://en.cetc.com.cn

[59] Human Rights Watch, ноябрь 2017 г.: https://www.hrw.org/news/2017/11/19/china-police-big-data-systems-violate-privacy-target-dissent

го субъекта и возможном правонарушении с целью превентивного предотвращения преступления.

Подобные технические и цифровые проекты реализуются параллельно с реформированием законодательства Китая. В последние годы власти издали ряд директив и нормативно-правовых актов в сферах сбора, интеграции, хранения и обмена данными для «улучшения социальной стабильности», включая: (1) создание единой системы интеллектуальных персональных данных с 2016 по 2020; (2) принятие принципов «Укрепления сбора базовой информации о полицейской деятельности» (2015)[60]; Уведомление об укреплении строительства систем социального обеспечения и контроля (2015)[61]. При этом в Китае отсутствует нормативно-правовая база о защите персональных данных, конфиденциальности и частной информации. Мало сведений о том, как и насколько безопасно хранятся данные, собранные для указанных систем, кто может получать или передавать их, при каких обстоятельствах они удаляются. Более того, отсутствует правовой способ узнать, какая информация содержится о гражданах в указанных системах, и, соответственно, нет способа защищать права и свободы, связанные с частными/персональными сведениями. Впрочем, нет и достоверных свидетельств об утечках данных такого рода.

Власти Китая целенаправленно поддерживают научно-исследовательские разработки в области цифровых технологий. Например, в июле 2017 г. был создан Национальный исследовательский институт в Урумчи с целью улучшения технического и аналитического оснащения региональных властей. Основной научной деятельностью называется изучение и обнаружение «скрытых инцидентов в области социального обеспечения». Осуществляемые в Китае исследования преступности посредством математико-статистического моделирования (создания алгоритмов прогнозирования) находят свое отражение в англоязычных публикациях в высокорейтинговых журналах.

В частности, исследование влияния временных и погодных факторов на имущественные преступления [Peng et al. 2011] выявило системную взаимосвязь влияния сезонных факторов на изменение динамики криминальной активности; было предложено создавать индексы сезонности преступности. Полученные результаты указывают, что значительное влияние на уровень грабежей оказывают временные переменные, но не погодные, в то время как на возникновение краж со взломом влияют не только временные переменные, но также продолжительность светового дня. Такие выводы используются как дополнительные переменные (информация о погодных условиях, учёт временных данных) для увеличения качества алгоритмического прогнозирования преступности.

Другое исследование содержит сравнение факторов имущественных преступлений (преступлений против собственности) и насильственных преступлений [Liu 2006]. Указывается, что имущественные преступления в течение последних двух с половиной десятилетий (во время интенсивной экономической модернизации Китая) увеличивались быстрее, чем насильственные преступления. Выявленные изменения динамики совершения преступлений стали основой для выделения статистически значимых предикторов для алгоритмов прогнозирования преступности.

Исследование специалистов Китайского университета в Гонконге установило, что теории когнитивного развития, социального контроля и социального обучения имеют статистическую значимость для

[60] Уведомление о расширении использования больших данных и облачных вычислений в работе полиции. URL: http://www.xinhuanet.com//politics/2015-06/02/c_1115490346.htm

[61] Использование технологических средств, облачных вычислений и больших данных для обеспечения социальной стабильности. URL: https://baike.baidu.com/item/关于加强社会治安防控体系建设的意见

прогнозирования вероятности преступного поведения среди маргинальной молодёжи [Ngai et al, 2005]. Значительное влияние на формирование преступного поведения оказывают следующие факторы: социальные проблемы, привязанность к труду, моральные убеждения и одобрение проступка индивида его друзьями. Менее значимы чувство разочарования и осознанное социальное неравенство. Данное исследование послужило основой для формирования «карты факторов» – данных из социальных сетей, которые используются для прогностического анализа противоправных деяний.

Представленные примеры свидетельствуют, что правительство КНР не только на декларативном уровне заявляет о внедрении различных цифровых технологий, в том числе продвинутых алгоритмических систем, для обеспечения безопасности, но и активно тестирует их в разных регионах страны. Более того, стремление государства обеспечить максимальный контроль и безопасность проявляется и на политическом, и на юридическом, и на техническом уровнях с масштабной модернизацией инфраструктуры для сбора данных и научно-исследовательскими разработками.

* * *

Стремительное развитие цифровых технологий, достижения науки о данных, увеличение вычислительных мощностей, равно как и увеличение количества информации и данных, модернизация правоохранительных органов, качественные изменения их кадрового состава и наличие запроса на трансформацию сферы обеспечения безопасности создали область математико-статистических моделей алгоритмического прогнозирования преступности и обеспечили их внедрение на практике.

Геоинформационные системы (ГИС) получили активное применение и развитие не только на уровне «картографии», хотя цифровые карты используются повсеместно, но и в области прогнозирования преступности. Экспоненциальный рост количества методов сбора данных, объёма генерируемых и собираемых данных привёл к большему количеству усовершенствований как самих ГИС, так и производных продуктов в сфере определения местоположения. Пространственный анализ используется в отслеживании и прогнозировании за счёт широкого спектра геолокационных возможностей. Социальные сети и повседневные мобильные приложения превращаются в платформы, предлагая ещё больше данных, что создало условия для улучшения алгоритмов прогнозирования преступности.

Представленный обзор позволяет сформулировать ряд выгод для общества и государства от реализации алгоритмического подхода. *Во-первых*, совершенствование алгоритмических моделей позволяет установить каузальную связь между явлениями, выступающими в качестве факторов возникновения криминогенных ситуаций, и непосредственно криминогенными ситуациями. *Во-вторых*, преступность и определяющие её факторы рассматриваются как взаимодействующие системы, в связи с чем возможен системный подход к борьбе с преступностью в больших и сложных сообществах. *В-третьих*, выявление и эмпирическая валидизация факторов позволяет вводить новые характеристики при создании алгоритмов (например, коэффициент сезонности, коэффициент преступной активности/интенсивности преступности). Указанное, в свою очередь, позволяет проверять практическую результативность прогностических алгоритмов при их реальной апробации, а также своевременно корректировать данные алгоритмы. *В-четвёртых*, геопространственные маркеры («горячие точки») позволяют решать проблемы пространственно-временного распределения преступности (девиантности и делинквентности) и улучшать на этой основе региональное и городское планирование, а также корректировать стратегии полицейской деятельности (разумно распределять ресурсы полиции для предотвращения и своевременного пресечения преступности). Выделение демографических характеристик позволяет получать более качественные прогностические данные по

возможным преступлениям с учётом конкретизации жертв и/или субъектов противоправного деяния.

Математико-статистические модели (алгоритмы) могут оказывать позитивное влияние на деятельность правоохранительных органов, сокращая ресурсные затраты, улучшая логистику, тактику и стратегию полицейской деятельности. Применение алгоритмов усиливает её доказательную базу (Evidence Based Approach), снижает риски (как физические, так и ресурсные) и может быть использовано для расширения профилактики правонарушений.

Наряду с выгодами возникают и риски, в том числе (1) риски злоупотреблений, систематических ошибок в работе с данными (включая угрозу персональным данным), а также предвзятость алгоритмов; (2) алгоритмы могут утрачивать точность статистического расчета, так как основываются исключительно на геопространственном анализе и демографическом составе населения в отрыве от социальных, политических и культурных компонентов; (3) инструменты, построенные исключительно на данных полицейских департаментов за предшествующие периоды, прогнозируют не вероятность совершения преступлений, а результативность работы полиции. Иными словами, если массивы данных, на которых строится анализ, учитывают исключительно зарегистрированные полицией преступления, то результаты такого анализа охватывают лишь те преступления, которые потенциально может зарегистрировать полиция; (4) существующие техники и методики прогнозирования преступности обходят стороной личностные особенности субъекта и индивидуальные факторы преступной модели личности, а также анализ преступного поведения субъекта через призму принятия решения о совершении противоправного деяния.

Риски имеют два базовых источника. Первый связан с технологиями в основе применяемых алгоритмов. В частности, методологической ошибкой выступает применение специфических данных, то есть набора сведений, хранящихся в полиции, для которых характерны систематические искажения. Программное обеспечение по прогнозированию преступности предназначено для изучения и воспроизведения закономерностей, найденных в данных. Если в модели загружаются предвзятые данные, они воспроизводят и усиливают те же самые предубеждения. С целью повышения точности результатов целесообразно расширять источники получения данных. Полиция традиционно использует интеллектуальные и субъективные оценки для выявления и мониторинга лиц с высоким риском преступного поведения, тогда как актуальные оценки риска являются относительно стандартной практикой в других частях правоохранительной и судебной системы, например в исправительных учреждениях, судебных инстанциях и службах исполнения наказания. Кроме того, необходимо эмпирически проверить возможности выделения факторов (предикторов) субъективных предпосылок преступного поведения, учитывая достижения психологии преступного поведения. Современные научные исследования указывают, что для точности и соизмеримости с реальными показателями преступности алгоритмы должны быть гибкими в настройках. Соответственно, расширение факторов (предикторов) посредством учета субъективных (индивидуализированных) показателей субъекта-преступника, а также показателей компонентов личностного принятия решения о совершении преступления, *во-первых*, персонифицирует субъектный состав прогнозов, *во-вторых*, повысит их точность. Показательно, что последние исследования в значительной мере рассматривают возможности включения маркеров преступного поведения в математико-статистические модели. При должном уровне экспериментального подтверждения включение индексов субъективных характеристик преступников может значительно улучшить прогностические модели.

Второй источник рисков имеет отнюдь не технический характер. Речь идет о разработке этико-ценностных рамок регулирования алгоритмического прогнозирова-

ния. Алгоритмы внедряются в различных политических системах для предупреждения и профилактики преступности. С одной стороны, статистический и криминологический анализ, а также профилирование субъекта преступления являются инструментами, помогающими правоохранительным органам, а не заменяющими конкретных сотрудников. С другой стороны, присутствуют риски предвзятости, ангажированности как в работе с данными и над алгоритмами, так и непосредственно в рамках функционирования самой прогностической модели. Более того, отсутствие контроля (как политического, правового, так и социального) порождает нарушение баланса между прогнозированием преступлений и соблюдением прав граждан: вопросы информационной свободы, расово (или национально) ангажированного профилирования и сопутствующих практик дискриминации замалчиваются, остаются «в тени».

Отсутствие систематических нетехнических исследований алгоритмизации, *во-первых*, вызывает скепсис в отношении применения алгоритмов, а когда они применяются и об этом становится известно, возникает общественное недовольство. *Во-вторых*, без эмпирически обоснованных стратегий реализации политики внедрения прогностических алгоритмов прогнозирования преступности невозможна практическая апробация и верификация результативности такого элемента предупреждения преступности. *В-третьих*, без изучения социально-политических аспектов применения алгоритмов прогнозирования преступности потенциальный риск от реализации указанных стратегий может принести больше вреда, чем потенциальной выгоды.

Приведённые примеры не свидетельствуют о том, что «бездушные машины» (алгоритмы, дроны-убийцы, «искусственный интеллект») в обозримом будущем поработят (или освободят) человечество. Нет оснований опасаться или надеяться на реализацию антиутопического в духе «Особого мнения» Ф. Дика или, напротив, утопического сценария создания идеальных алгоритмов прогнозирования преступности в версии сериала «В поле зрения».

Важно другое. Алгоритмическое прогнозирование преступности – это различные технологии обработки данных, внедрение которых (при наличии воли, эффективного государственного управления, кадров) уже сегодня даёт вполне ощутимые результаты и позволяет решать конкретные задачи обеспечения безопасности граждан. Успехи в этой области демонстрируют не только отдельные демократии, но и некоторые государства с режимными характеристиками, не позволяющими классифицировать их в качестве демократий. По мере совершенствования алгоритмов государства, не являющиеся технологическими лидерами, могут стать их покупателями, поскольку успешное обеспечение безопасности – это фактор обеспечения легитимности в глазах избирателей (или селектората) и конкурентное преимущество на выборах. Вполне вероятна ситуация, когда алгоритмы и технические решения «под ключ» станут объектом экспортно-импортных операций в такой же мере, в какой сейчас ими являются любые программные продукты, компьютерная техника или «специальные средства», включая оружие. Иными словами, программные и технические средства, поддерживающие алгоритмы прогнозирования преступности, станут ещё одной составляющей международной конкуренции государств–поставщиков технологий.

С учётом масштабов и востребованности прогностической полицейской деятельности всё более актуальным становится политико-правовое регулирование применения алгоритмов прогнозирования. Его отсутствие или слабость порождают теневой характер апробации и доработки алгоритмов, их потенциальное использование для иных целей, нежели прогнозирование и предотвращение преступности. Вместо инструмента, который уже сейчас может работать на общественное благо, алгоритмы вызывают скорее обеспокоенность и недовольство отдельных общественных активистов и продолжают оставаться «чёрными ящиками».

**Список литературы**

*Гохман В.В., Третьяченко Д.А.* Восемь прорывных технологий и их связь с геопространством // ArcReview 2019. №2 (89). URL: https://www.dataplus.ru/news/arcreview/detail.php?ID=27212&SECTION_ID=1117

*Ang R.P., Goh D.H.* Predicting juvenile offending: A comparison of data mining methods // International Journal of Offender Therapy and Comparative Criminology. 2013. Vol. 57. No. 2. P. 191–207. https://doi.org/10.1177/0306624X11431132

*Bansal D., Bhambhu L.* Execution of APRIORI Algorithm of Data Mining Directed Towards Tumultuous Crimes Concerning Women // International Journal of Advanced Research in Computer Science and Software Engineering. 2013. No. 3(9). 54 p.

*Bendler J., Brandt T., Wagner S., Neumann D.* Investigating Crime-To-Twitter Relationships in Urban Environments – Facilitating a Virtual Neighborhood Watch // Ecis. 2014. P. 1–16.

*Besley T., Persson. T.* The origins of state capacity: Property rights, taxation, and politics // American Economic Review. 2009. Vol. 99. No. 4. P. 1218–1244 https://doi.org/10.1257/aer.99.4.1218

*Chu C.M., Ng K., Fong J., Teoh J.* Assessing Youth Who Sexually Offended: The Predictive Validity Of The ERASOR, J-SOAP-II, and YLS/CMI // Non-Western Context. Sexual Abuse: Journal of Research and Treatment. 2012. Vol. 24. No. 2. P. 153–174. https://doi.org/10.1177/1079063211404250

*Cornish P.* Technology, strategy and counterterrorism // International Affairs. 2010. Vol. 86. No. 4. P. 875–888 https://doi.org/10.1111/j.1468-2346.2010.00917.x

*Ensign D., Friedler S.A., Neville S., Scheidegger C., Venkatasubramanian S.* Runaway Feedback Loops in Predictive Policing // eprint arXiv. 2017. P. 1–12. Retrieved from http://arxiv.org/abs/1706.09847

*Gerber M.S.* Predicting crime using Twitter and kernel density estimation // Decision Support Systems. 2014. Vol. 61. No. 1. P. 115–125. https://doi.org/10.1016/j.dss.2014.02.003

*Hecker S., Haklay M., Bowser A., Makuch Z., Vogel J., Bonn A.* Citzen Science: Innovation in Open Science, Society and Policy // Citzen Science. London: UCL Press 2018. https://doi.org/10.14324/111.9781787352339

*Hoffman S.* Managing the State: Social Credit, Surveillance and the CCP's Plan for China // AI, China, Russia, and the Global Order: Technological, Political, Global, and Creative. 2018. P. 42–47.

*Liu J.* Modernization and crime patterns in China // Journal of Criminal Justice. 2006. Vol. 34. No. 2. P. 119–130. https://doi.org/10.1016/j.jcrimjus.2006.01.009

*Lum K., Isaac W.* To predict and serve // Significance. 2016. Vol. 13. P. 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x

*Mande U., Srinivas Y., Murthy J.V.R.* An Intelligent Analysis of Crime Data Using Data Mining & Auto Correlation Models // International Journal of Engineering Research and Applications (IJERA). 2012. Vol. 2. No 4. P. 149–153. URL: https://www.ijera.com/papers/Vol2_issue4/U24149153.pdf

*Nakaya T., Yano K.* Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics // Transactions in GIS. 2010. Vol. 14. No. 3. P. 223–239. https://doi.org/10.1111/j.1467-9671.2010.01194.x

*Ngai N.P., Cheung C.K.* Predictors of the likelihood of delinquency: A study of marginal youth in Hong Kong, China // Youth and Society. 2005. Vol. 36. No. 4. P. 445–470. https://doi.org/10.1177/0044118X04265090

*Peeters R., Widlak A.* The digital cage: Administrative exclusion through information architecture – The case of the Dutch civil registry's master data management system // Government Information Quarterly. 2018. Vol. 35. No. 2. P. 175–183 https://doi.org/10.1016/j.giq.2018.02.003

*Peng C., Xueming S., Hongyong Y., Dengsheng L.* Assessing temporal and weather influences on property crime in Beijing, China // Crime, Law and Social Change. 2011. Vol. 55. No. 1. P. 1–13. https://doi.org/10.1007/s10611-010-9264-3

*Roy S., Shah A., Srikrishna B.N., Sundaresan S.* Building State capacity for regulation in India // Working paper No. 237. National Institute of Public Finance and Policy New Delhi. In: Regulation in India: Design, Capacity, Performance / ed. by D. Kapur, M. Khosla. Oxford: Hart Publishing, forthcoming 2019.

*Seo S., Chan H., Brantingham P.J., Leap J., Vayanos P., Tambe M., & Liu Y.* Partially Generative Neural Networks for Gang Crime Classification with Partial Information // ACM Conference on Artificial Intelligence, Ethics, and Society. 2018. https://doi.org/10.1145/3278721.3278758

*Tayal D.K., Jain A., Arora S., Agarwal S., Gupta T., Tyagi N.* Crime detection and criminal identification in India using data mining techniques // AI and Society. 2014. Vol. 30. No. 1. P. 117–127. https://doi.org/10.1007/s00146-014-0539-6

*Tufekci Z.* Social Movements and Governments in the Digital Age: Evaluating a Complex Landscape // Journal of International Affairs. 2014. Vol. 68. No. 1. P. 1–18. https://www.jstor.org/stable/24461703

*Vohra N.N.* National Security Concerns. India International Centre Quarterly // India International Centre Quarterly 2012. Vol. 38. No. 3/4. The Golden Thread: Essays in Honour of C.D. Deshmukh. P. 370–385. https://www.jstor.org/stable/41803992

*Wang L., Zhao J.S.* Contemporary police strategies of crime control in U.S. and China: a comparative study // Crime, Law and Social Change. 2016. Vol. 66. No. 5. P. 525–537. https://doi.org/10.1007/s10611-016-9641-7

*Wanna J.* Opening government: Transparency and engagement in the information age // Opening Government: Transparency and Engagement in the Information Age. 2018. Canberra: ANU Press. https://doi.org/10.22459/og.04.2018.01

*Yang Qiaomei.* The smart city of Changsha, China // Smart City Emergence. 2019. P. 219–241. ISBN 9780128161692. https://doi.org/10.1016/B978-0-12-816169-2.00010-9.

# WORLD BEST PRACTICES IN APPLYING MATHEMATICAL AND STATISTICAL CRIME PREDICTION ALGORITHMS

ALEKSEY TUROBOV
MARIA CHUMAKOVA
ALEKSANDR VECHERIN
National Research University "Higher School of Economics", Moscow 101000, Russia

**Abstract**

The sphere of security provision is expanding and constantly bringing in new elements, including cybersecurity, information security, computer network security, etc.). The arsenal of security tools is also growing due to the ongoing proliferation of digital technologies (e.g. different technologies and telecommunication channels for collecting, forming, processing, transmitting or receiving information related to security of the state). The article provides an analysis of current methods and technologies for crime forecasting in the national security domain. Achievements in the Data Science and Big Data generated the scientific basis for the development of Intellectual Data Analysis (Intellectual Analysis, Predictive Analysis), based on which mathematical and statistical forecasting of socially dangerous, criminal acts was designed (e.g. anti-terrorism algorithms, algorithms for predicting the activities of organized crime/gangs). The article aims to identify major trends and potential benefits of digital technologies proliferation as well as the challenges that states face while using mathematical and statistical methods for predicting crime. The meta-analysis of scientific researches and implementation of crime forecasting algorithms in different countries (such as USA, China, Japan, Singapore, India) helps to demonstrate a pluralism of approaches in the application of forecasting systems. The first part of the article presents the methodological and technical aspects of criminal data mining. The second part provides an overview of national practices in using crime prediction algorithms by the examples of Singapore, Japan, and India. The third and fourth parts are devoted to a more detailed analysis of the strategies and tactics of using algorithms in the USA and China, respectively. The analysis reveals the risks and benefits inherent in the most frequently applied mathematical and statistical crime forecasting algorithms. First, it is the "militarization" of the civilian sphere. Second, the algorithms, which do not take into account the social, cultural and political features of a given society, lead to the loss of statistical significance of forecasting. Third, historical data (recorded crimes) often contain racial, sexual, and contextual biases. Fourth, existing

approaches do not pay heed to personal characteristics of a subject, as well as decision-making processes not infrequently resulting in wrongful conduct. Finally, there is no state control over the balance between the use of algorithms and respect for human rights.

**Keywords:**
security; crime forecasting; data; algorithms; digitalization.

**References**

Ang R.P., Goh D.H. (2013). Predicting juvenile offending: A comparison of data mining methods. *International Journal of Offender Therapy and Comparative Criminology*. Vol. 57. No. 2. P. 191–207. https://doi.org/10.1177/0306624X11431132

Bansal D., Bhambhu L. (2013). Execution of APRIORI Algorithm of Data Mining Directed Towards Tumultuous Crimes Concerning Women. *International Journal of Advanced Research in Computer Science and Software Engineering*. No. 3(9). 54 p.

Bendler J., Brandt T., Wagner S., Neumann D. (2014). Investigating Crime-To-Twitter Relationships in Urban Environments – Facilitating a Virtual Neighborhood Watch. *Ecis*. P. 1–16.

Besley T., Persson T. (2009). The origins of state capacity: Property rights, taxation, and politics. *American Economic Review*. Vol. 99. No. 4. P. 1218–1244. https://doi.org/10.1257/aer.99.4.1218

Chu C.M., Ng K., Fong J., Teoh J. (2012). Assessing Youth Who Sexually Offended: The Predictive Validity Of The ERASOR, J-SOAP-II, and YLS/CMI in a Non-Western Context. *Sexual Abuse: Journal of Research and Treatment*. Vol. 24. No. 2. P. 153–174. https://doi.org/10.1177/1079063211404250

Cornish P. (2010). Technology, strategy and counterterrorism. *International Affairs*. Vol. 86. No. 4. P. 875–888. https://doi.org/10.1111/j.1468-2346.2010.00917.x

Ensign D., Friedler S.A., Neville S., Scheidegger C., Venkatasubramanian S. (2017). *Runaway Feedback Loops in Predictive Policing*. URL: http://arxiv.org/abs/1706.09847

Gerber M.S. (2014). Predicting crime using Twitter and kernel density estimation. *Decision Support Systems*. Vol. 61. No. 1. P. 115–125. https://doi.org/10.1016/j.dss.2014.02.003

Gokhman V.V., Tret'yachenko D.A. (2019) Vosem' proryvnykh tekhnologiy i ikh svyaz' s geoprostranstvom [Eight Disrupting Technologies and Their Connection to Geospace]. *ArcReview*. No. 2. URL: https://www.dataplus.ru/news/arcreview/detail.php?ID=27212&SECTION_ID=1117

Hecker S., Haklay M., Bowser A., Makuch Z., Vogel J., Bonn A. (2018). Innovation in Open Science, Society and Policy. In: Citizen Science / ed. by S. Hecker, M. Haklay, A. Bowser, Z. Makuch, J. Vogel, A. Bonn. London: UCL Press. P. 1–26. https://doi.org/10.14324/111.9781787352339

Hoffman S. (2018) Managing the State: Social Credit, Surveillance and the CCP's Plan for China. *AI, China, Russia, and the Global Order: Technological, Political, Global, and Creative* / ed. by S. Ahmed. NSI Boston United States. P. 42–47.

Liu J. (2006). Modernization and crime patterns in China. *Journal of Criminal Justice*. Vol. 34. No. 2. P. 119–130. https://doi.org/10.1016/j.jcrimjus.2006.01.009

Lum K., Isaac W. (2016). To predict and serve? *Significance*. Vol. 13. No. 5. P. 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x

Mande U., Srinivas Y., Murthy J.V.R. (2012). An Intelligent Analysis of Crime Data Using Data Mining & Auto Correlation Models. *International Journal of Engineering Research and Applications (IJERA)*. Vol. 2. No. 4. P. 149–153. URL: https://www.ijera.com/papers/Vol2_issue4/U24149153.pdf

Nakaya T., Yano K. (2010). Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics. *Transactions in GIS*. Vol. 14. No. 3. P. 223–239. https://doi.org/10.1111/j.1467-9671.2010.01194.x

Ngai N.P., Cheung C.K. (2005). Predictors of the likelihood of delinquency: A study of marginal youth in Hong Kong, China. *Youth and Society*. Vol. 36. No. 4. P. 445–470. https://doi.org/10.1177/0044118X04265090

Peeters R., Widlak A. (2018). The digital cage: Administrative exclusion through information architecture – The case of the Dutch civil registry's master data management system. *Government Information Quarterly*. Vol. 35. No. 2. P. 175–183. https://doi.org/10.1016/j.giq.2018.02.003

Peng C., Xueming S., Hongyong Y., Dengsheng L. (2011). Assessing temporal and weather influences on property crime in Beijing, China. *Crime, Law and Social Change*. Vol. 55. No. 1. P. 1–13. https://doi.org/10.1007/s10611-010-9264-3

Roy S., Shah A., Srikrishna B.N., Sundaresan S. (2018). Building State capacity for regulation in India. Working paper No. 237. National Institute of Public Finance and Policy New Delhi. In: *Regulation in India: Design, Capacity, Performance* / ed. by D. Kapur, M. Khosla. Oxford: Hart Publishing, 2019. Forthcoming.

Seo S., Chan H., Brantingham P.J., Leap J., Vayanos P., Tambe M., Liu Y. (2018). Partially Generative Neural Networks for Gang Crime Classification with Partial Information. *ACM Conference on Artificial Intelligence, Ethics, and Society*. https://doi.org/10.1145/3278721.3278758

Tayal D. K., Jain A., Arora S., Agarwal S., Gupta T., Tyagi N. (2014). Crime detection and criminal identification in India using data mining techniques. *AI and Society*. Vol. 30. No. 1. P. 117–127. https://doi.org/10.1007/s00146-014-0539-6

Tufekci Z. (2014). Social Movements and Governments in the Digital Age: Evaluating a Complex Landscape. *Journal of International Affairs*. Vol. 68. No. 1. P. 1–18. https://doi.org/10.1017/CBO9781107415324.004

Vohra N.N. (2012). National Security Concerns. *India International Centre Quarterly*. Vol. 38. No. 3/4, The Golden Thread: Essays in Honour of C.D. Deshmukh. P. 370–385. https://www.jstor.org/stable/41803992

Wang L., Zhao J.S. (2016). Contemporary police strategies of crime control in U.S. and China: a comparative study. *Crime, Law and Social Change*. Vol. 66. No. 5. P. 525–537. https://doi.org/10.1007/s10611-016-9641-7

Wanna J. (2018). Opening government: Transparency and engagement in the information age. In: *Opening Government: Transparency and Engagement in the Information Age*. Canberra: ANU Press. P. 3–26. https://doi.org/10.22459/og.04.2018.01

Yang Qiaomei. (2019). The smart city of Changsha, China. In: Anthopoulos L. (ed.) *Smart City Emergence*. Elsevier. 2019. P. 219–241. https://doi.org/10.1016/B978-0-12-816169-2.00010-9.