

Научная статья

УДК 343.01:343.9

EDN ZQKNOC

DOI 10.17150/2500-4255.2025.19(2).126-136



ТЕОРИЯ И МЕТОДОЛОГИЯ БОЛЬШИХ ДАННЫХ: КРИМИНОЛОГО-ЭПИСТЕМОЛОГИЧЕСКИЙ РАКУРС

Т.М. Судакова*Байкальский государственный университет, г. Иркутск, Российская Федерация*

Информация о статье

Дата поступления

2 апреля 2025 г.

Дата принятия в печать

9 июня 2025 г.

Дата онлайн-размещения

17 июня 2025 г.

Ключевые слова

Большие данные; эпистемология БД;
интеллектуальный анализ данных;
предиктивная аналитика; прогнозная
деятельность; противодействие
преступности

Аннотация. Цифровые технологии изменили способы поиска и производства информации. Сегодня остро стоит вопрос о возможностях социальных, социально-правовых наук адекватно реагировать на цифровизацию социума, на технологические изменения и распространение информационно-телекоммуникационных технологий, виртуального пространства, больших данных. Это и вопрос практической целесообразности и возможностей разработки инструментов использования больших данных при проведении социогуманитарных исследований. Проблема эпистемологии больших данных сопутствует появлению огромного потока структурированных и неструктурированных данных, активно используемых в научных исследованиях, в предиктивной аналитике. Это проблема понимания природы современного знания. Для криминологического мышления и его прикладной значимости большие данные не являются совершенно новым параметром исследований. Существуют применяемые в работе правоохранительной сферы, но не используемые повсеместно механизмы использования больших данных для измерения и установления корреляции между различными факторами, значимыми в сфере противодействия преступному поведению и выработки прогнозных решений. В рамках настоящего исследования рассматривается методологический ракурс проблемы использования больших данных с позиции такой возможности, целесообразности и надежности. Это проблема «методологической чистоты» и научной корректности в их интерпретации для выработки научно обоснованных решений. Машинное обучение как разновидность интеллектуального анализа больших данных является мощным и эффективным инструментом научного анализа, требуемого при этом постоянного совершенствования используемых алгоритмов. Философское осмысление проблемы использования больших данных заключается в формировании постклассической эпистемологической парадигмы с основной идеей о причинно-следственных связях как необходимом элементе любого исследовательского проекта. Информационные издержки, технические сложности структурирования и интеграции данных, перспективы и направления монетизации больших данных, сложность создания хранилищ данных и выбора способа их хранения неотъемлемо сопровождают технологии использования инструментов интеллектуального анализа в социальных и криминологических исследованиях. Перестройка всей системы аналитической и прогностической работы в сфере противодействия преступности и перестройка психологии самих исследователей и правоохранителей — стратегически важная задача в развитии криминологии.

Original article

THE THEORY AND METHODOLOGY OF BIG DATA: A CRIMINOLOGICAL-EPISTEMOLOGICAL PERSPECTIVE

Tatiana M. Sudakova*Baikal State University, Irkutsk, the Russian Federation*

Article info

Received

2025 April 2

Accepted

2025 June 9

Available online

2025 June 17

Abstract. Digital technologies have changed the ways of searching and producing information. An urgent question today is the ability of social, socio-legal sciences to provide an adequate reaction to the digitization of society, technological changes and the spread of information-telecommunication technologies, virtual environment, big data. It also concerns the practical value and the possibilities of developing instruments for using big data in socio-humanitarian research. The problem of big data epistemology contributes to the emergence of an enormous stream of structured and unstructured data, actively used in scientific research, in predictive analytics. It is the problem of understanding

Keywords

Big data; database epistemology; data mining; predictive analytics; predictive activity; crime counteraction

the nature of modern knowledge. Big data is not a completely new research parameter for criminological cognition and its applications. There are mechanisms, used in law enforcement but not universally widespread, of applying big data to measuring and establishing correlations between various factors relevant for counteracting unlawful behavior and developing predictive solutions. Within the framework of the current research, the author examines the methodological aspect of the problem of using big data from the standpoint of possibility, expediency and reliability. It is a problem of “methodological purity” and scientific integrity of their interpretation for the purpose of developing scientifically grounded solutions. Machine learning as a variety of intellectual analysis of big data is a powerful and effective instrument of research analysis requiring a constant improvement of the used algorithms. The philosophical interpretation of the problems of using big data consists in forming a post-classical epistemological paradigm with the central idea of causal relationships as a necessary element of any research project. Information costs, technical difficulties of data structuring and integration, the prospects and trends of monetizing big data, the difficulties of creating data storages and selecting the storage format are inalienable companions of the technology of using data mining in social and criminological studies. The transformation of the whole system of analytical and predictive work in the sphere of crime counteraction, as well as the psychological adjustment of researchers and law enforcement employees themselves is a strategically paramount task in the development of criminology.

Принципы формализованного описания тенденций развития общества и создания прогнозных моделей рассматриваются учеными в качестве нового фактора его существования и развития науки [1]. Применительно к социальным и социально-правовым явлениям и современной специфике их изучения, как и к разработке прогнозов, этот факт озвучен специалистами через методологию вычислительных социальных теорий [2–5]. Сегодня получила развитие теория вычислительной социологии как перспективная и развивающаяся отрасль социальной науки, разрабатываемая для изучения сложных социальных явлений и синтезирующая такие предметные отрасли, как информатика, наука о данных, социальная теория, статистика.

Речь идет о междисциплинарном подходе как принципе синергетической методологии, который позволяет извлечь информацию о сложных процессах социальной жизни, их динамике и перспективах на основе интеллектуального анализа больших массивов информации — больших данных (БД) и моделировании этих социальных явлений. Специалисты расценивают эти вычислительные процессы как мощный исследовательский инструмент цифровой эпохи. Однако, как и у любой новой исследовательской парадигмы, потенциал ее оценивается неоднозначно, и для этого существует ряд причин.

Обращение к технологическим инновационным процессам закономерно подводит ученого к теории и практике применения БД, перестраивающих в определенной мере традиционные методы научного исследования.

Использование БД при изучении и оценке социальных проблем обозначило появление новой междисциплинарной области исследований, изменившей способы поиска и производства информации. Большие данные — это огромный массив данных, для работы с которым требуются технологии, информационная архитектура и проектирование систем, а также аналитические методы. Постоянный рост массивов разнообразной разнородной неструктурированной информации и существующие сегодня возможности компьютерных наук выступили основой развития вычислительной или цифровой социологии.

Влияние цифровизации на социум, на саму социологию поставило перед специалистами вопросы о роли технологических изменений, способствующих оперативному сбору, анализу информации и ее структурированию. Как и вопросы о направлениях участия социологов в практике разработки требуемых ресурсов-технологий, инструментов для исследования и интерпретации больших данных, выступающих сегодня важной основой восприятия социальных явлений [2].

Научная корректность в условиях информационного изобилия — актуальнейший вопрос, получивший освещение и критическое осмысление в последнее десятилетие. Почему этот вопрос вообще возникает?

Искусственный интеллект, машинное обучение (МО) как инструменты познания социальной реальности стали основой возникновения теории осмысления иной научной парадигмы.

Огромные объемы информации, доступные для обработки и анализа, способствовали их использованию для формализованного описания различных моделируемых объектов. Эпоха больших данных (BIG DATA) — терминология, не только распространенная в науке, но и активно разрабатываемая в аспекте понимания и принципов использования огромного массива информации. Процесс анализа больших объемов данных не только активно разрабатывается во многих сферах научного анализа, но и имеет широкую прикладную значимость. Вместе с тем однозначности в вопросах о целесообразности повсеместного применения анализа больших данных в разных социальных сферах, как и в вопросах методологического обеспечения этого процесса и последующего его осмысления в социальных, социально-правовых науках, пока нет. Более того, эта категория представляет непосредственный интерес как источник новой эпистемологии [6; 7].

Интернет в условиях интенсивного проникновения технологической революции трансформируется из хранилища данных в пространство знаний, где искусственный интеллект и его системы выполняют функции преобразования данных в информацию, и в последующем — в знания. А знания выступают источником решения разнообразных задач [8, с. 267]. Искусственный интеллект (автоматизированный системно-когнитивный анализ) позволяет без предварительной подготовки осмысливать эти данные, преобразовывая их в информацию, а затем в применимые на практике знания [8].

Методы обработки таких данных и более эффективные алгоритмы, разработанные с помощью искусственного интеллекта и обеспечивающие более продвинутые инструменты для анализа и визуализации, включают в себя машинное обучение, нейронные сети, глубокое обучение, которые могут автоматически извлекать информацию из больших объемов данных и прогнозировать результаты.

Сегодня заявлено об активном развитии нового междисциплинарного направления, цифровые гуманитарные науки — Digital Humanities, работающие с большими данными. Это послужило, в свою очередь, триггером для формирования иной, четвертой парадигмы научного знания, базирующейся на интенсивном использовании больших данных и пришедшей на смену третьей парадигме, основанной на крупномасштабном компьютерном моделиро-

вании. И этот процесс рассматривается сегодня, при всей неоднозначности присутствующих оценок, как эпистемологическая революция в познании. Ее преломление в изучении характера и специфики социальных явлений и процессов заключается в том, что цифровые данные позволяют заменить поиск причинно-следственных связей на установление корреляции между ними [9].

Эпоха больших данных, таким образом, серьезным образом повлияла на процесс получения и интерпретации научного знания, поставив вопрос об эпистемологии больших данных, об изменении методологии научных исследований. При этом дискуссии об эпистемологии БД в основном посвящены проблемам целесообразности новой структуры знания, обусловленной особенностями процесса его получения. Это и вопросы о том, что считать знанием в контексте их использования, какова роль теории в таком случае, поскольку гипотеза заменяется анализом.

И вполне закономерно возникает еще ряд сопутствующих вопросов, требующих разрешения: о чистоте самих данных, их объективности, о надежности и эффективности стандартов сбора и обработки, о соотношении корреляции и причинно-следственных зависимостей, о роли гипотезы и методах ее проверки в данном анализе.

Если традиционный научный метод предполагал построение гипотезы, ее проверку путем наблюдений и экспериментов, то БД и их использование представили альтернативный метод, базирующийся на анализе огромных массивов информации и последующей постановке вопросов для выявления закономерностей в исследовании. Вместо теоретических предпосылок, исследование начинается с изучения данных. И именно такой подход и поставил вопрос о традиционной логике формирования научного знания.

Философский анализ происходящих перемен помогает осмыслить изменения методологии получения и интерпретации научных данных и сформировать принципы, обеспечивающие научную корректность в условиях автоматизации процессов получения знания и информационного изобилия. Что считать знанием в контексте БД, если теория заменяется анализом и могут ли корреляция и причинно-следственная связь не подменять друг друга. Проблема достоверности данных, этические основы и последствия их применения обладают чрезвычай-

ной значимостью в контексте объективности и ценности получаемых выводов [1; 9].

Философское осмысление, таким образом, помогает адаптироваться науке к новым условиям ее существования при сохранении фундаментальных принципов построения науки. Новая философская сфера — постгуманистическая эпистемология — развивается как учение о фундаментальных принципах пересмотра природы знания. В отличие от классической эпистемологии, где человек выступал главным источником получения знания, постгуманистическая эпистемология источником знания видит сеть взаимодействий, где технологии, ИИ, биология и материальные объекты выступают равноправными источниками знания. Машинное обучение и его алгоритмы, к примеру, обрабатывая огромные массивы данных, способны определить закономерности, недоступные человеку при статистическом анализе. Их решения не подчинены контролю исследователя и зачастую вызывают необходимость пересматривать классические подходы к истине научного знания. Постгуманистическая философия предлагает также рассматривать знание не как фиксированный результат, а как процесс, постоянно меняющийся и адаптирующийся под влиянием новых факторов, которых может быть множество, имеющих самую различную природу и исследуемых на уровне междисциплинарного взаимодействия [9].

Искусственный интеллект и машинное обучение выступают инструментами, позволяющими находить связи, не обнаруживаемые человеком. Они значительно продвигают обработку данных и развиваются, улучшая свой потенциал, и способны создавать мощные алгоритмы. Эти технологические инструменты меняют, таким образом, подход к исследованию и интерпретации получаемых результатов и становятся неотъемлемой частью научного процесса, связанного с их использованием и влекущего новые вызовы и возможности в науке.

Однако стремительно растущие объемы информации постоянно требуют новых стандартов ее сбора и обработки для обеспечения надежности получаемых результатов. Эти технологии не защищены от ошибки при наличии неоднозначных исходных данных [10]. Кроме того, процессы автоматизации способны усиливать предвзятость или исказить данные, поскольку алгоритмы могут давать разные результаты в зависимости от используемых моделей [11]. Таким образом, важность человеческого участия в

анализе и наличие особого контроля лишь повышается вследствие описанных обстоятельств. Учет контекста исследования и способность увидеть более сложные взаимосвязи делает обязательным субъективный фактор. Кроме того, БД часто используют информацию частного характера, и этот факт ставит вполне закономерно вопрос об этике такого рода исследований.

Сегодня, несмотря на всю неоднозначность оценок и получаемых результатов, в науке заявлено о высоком потенциале исследований, выполненных с использованием вычислительных методов анализа БД, как и о необходимости фундаментальных наработок для создания теоретико-методологической основы вычислительной или цифровой социологии.

О вычислительной криминологии, изучающей социально-правовую реальность, сегодня заявлено как о сфере криминологической теории, требующей стремительного развития, соответствующего тенденциям цифровизации общества и цифровизации преступности [12–15].

Теория вычислительной криминологии или кримиометрика не является чем-то новым для криминологии. О целесообразности использования языка математического моделирования, об алгоритмической криминологии ученые разных специальностей заявляли неоднократно, подтверждая свои выводы конкретными эмпирическими исследованиями с применением алгоритмов, в том числе в целях выработки прогнозных тенденций преступности [16–18]. Не является примечательным и тот факт, что алгоритмические инструменты применялись профильными специалистами — математиками, информатиками, эконометриками, владеющими данным «искусством». И на основе таких расчетов формулировались выводы о тесноте корреляций между различными факторами, имеющими определенную взаимосвязь с преступностью. Профильные знания в вопросах постановки гипотезы, сбора необходимых источников для построения модели и последующий анализ полученных выводов использовались как закономерный процесс и итог всего исследования. Гипотеза получала свое разрешение.

Развитие технологий и появление ИИ обозначили проблему использования массивов информации в проведении исследований, в изучении преступного поведения, его причинно-следственных связей и корреляций на новом уровне. Следует отметить, что зарубежная криминология использует новейшие техноло-

гии в исследовании преступности, преступного поведения весьма активно [16–18], однако этот факт сегодня нельзя признать состоявшимся в российской криминологии. Есть ли шанс и методологическая возможность у криминологии использовать теорию и принципы БД для своих исследований и какова в целом целесообразность таких инноваций?

Относительно недавно известный криминолог заявил о необходимости оперативного реагирования отечественной криминологии на изменяющуюся социальную реальность путем внедрения механизмов вычислительной криминологии и нейрокогнитивистики, использования формальных методов определения количественных показателей преступности как противовеса процессам стагнации науки о преступности. Обоснование целесообразности такого направления развития ученых связывает с теорией и методологией цифровых технологий и теми постулатами, на которых сегодня развивается зарубежная криминология [19].

Вопросы, которые следует обозначить как составляющие суть исследуемой проблематики, сводятся к трем основным группам. Это вопрос о целесообразности и возможности перехода к новой парадигме социального и социально-правового знания в связи с развитием методологии анализа больших данных и новой эпистемологической ситуации. Это вопрос о целесообразности сочетанного подхода к использованию больших массивов информации в контексте изучаемого явления. И это вопрос о техническом компоненте проблемы — а именно о возникающих технических сложностях структурирования и интеграции данных, создания хранилищ данных и выбора способа их хранения.

Вычислительная социология стремительно развивается как совокупность имплементированных в ее принципы исследований методов математической логики при изучении социальных негативных процессов. Нужно ли ассимилироваться российской криминологии с этими процессами, и какова ситуация сегодня в данном контексте в криминологии?

Рассуждения о вычислительной социологии не характеризуются однозначностью подходов к оценке потенциала такого междисциплинарного взаимодействия. Сегодня заявлено о необходимости приоритетного и параллельно протекающего процесса создания теоретико-методологической базы для вычислительной социологии через множество имеющихся фун-

даментальных наработок. Данный факт признается социологами через целесообразность разработки принципов междисциплинарного анализа с помощью вычислительных и математических методов, позволяющего изучать сложную динамику социальной жизни через исследование социальных сетей и применение методов моделирования и машинного обучения. Эти возможности значительно расширяют исследовательское поле социологии [4].

Зарубежная социальная и криминологическая теория о принципах использования больших данных «высококочувствительна» к данной проблеме, и социогуманитарные исследовательские проекты, социологические, криминологические работы насыщены множеством суждений как о методологии БД и новой парадигме научных исследований, так и о применении алгоритмов машинного обучения и интеллектуального анализа данных в различных сферах социальной действительности.

В качестве предварительных суждений следует сделать акцент на самом механизме использования БД в науке. Большие данные рассматриваются как обеспечивающие большую точность и предсказательную силу для повышения эффективности, безопасности и создания ресурсов или управления ими. Способность цифровых технологий собирать, добывать, хранить и анализировать огромные массивы данных представляется превосходящей другие формы знаний, поскольку предлагается больше возможностей, чем когда-либо, для более глубокого изучения человеческого поведения, в том числе преступного.

С момента популяризации теории и практики применения БД они были применены в самых разных областях социологии. Исследователи проблем применимости технологий БД в политических науках подчеркивают, что до сих пор не сложилось единодушия в понимании того, как исследования, использующие БД, соотносятся с существующей эпистемологией знания. Дискуссии, касающиеся эпистемологии больших данных, в основном затрагивают вопросы о необходимости новой эпистемологической структуры для понимания механизма использования больших данных в исследованиях. Так, исследователи особенностей влияния внешних факторов на организм человека (экспозомики) утверждают о важности причинно-следственных связей в этой сфере, необходимых как обязательный этап применения анализа БД.

Доказывается, что причинно-следственные связи являются необходимым элементом любого исследовательского проекта, в том числе базирующегося на использовании БД в качестве источника формирования основы исследования и в качестве его результата. Таким образом утверждается, что причинно-следственные связи должны рассматриваться как необходимый элемент науки о больших данных [20].

Углубление в суть этой проблемы позволяет говорить о двух полярных позициях авторов, одна из которых сводима к установке о переориентации науки на новую форму, бросающую вызов традиционным представлениям о связи между теоретическими и эмпирическими исследованиями. И вторая, акцентирующая внимание на неплотности такой переориентации и недооценке роли теории в данном ракурсе. Размежевание противоположных суждений исследователи видят в признании эпистемологической концепции, учитывающей фактическое использование технологий использования БД применительно к различным научным контекстам. Эпистемологические дискуссии ориентированы на конкретный научный контекст и различны в разных научных сферах¹.

Потенциал в изучении преступности и противодействии преступному поведению основывается в науке преимущественно на оценках эффективности таких практик, которые, в свою очередь, имеют предметом исследования уже используемые в полицейской деятельности технологии определения криминальной активности с помощью интеллектуального анализа. Кроме того, предмет расширяется и за счет так называемой предиктивной аналитики [21–23]. Зарубежные аналитики важным направлением обеспечения безопасности и поддержания общественного порядка, в частности в работе полицейских структур, активно обсуждают проблемы использования больших данных, оказывающих значительное влияние на развитие стратегий в данной сфере.

Значимость БД и связанных с ними технологий познания криминологической реальности, подчеркивают специалисты, рассматривается в двух основных аспектах: использование БД в криминологических исследованиях в реальном или практически реальном временном режиме,

что соответствует направлению развития социальных наук в целом, и использование моделирования и алгоритмов в качестве одного из прогнозных инструментов для их последующей оценки и учета в работе правоохранительной системы [24].

Исследовательские зарубежные научные проекты и предиктивная аналитика развиваются уже не первое десятилетие, и, более того, ее механизмы достаточно активно внедрены в деятельность правоохранительных органов, не имея при этом, однако, однозначных оценок в контексте эффективности и необходимости этой сферы научной и прикладной деятельности. С одной стороны, границы прогнозирования в аспекте выявления неизвестных закономерностей в данных, без учета ценности их причинно-следственных объяснений и в условиях недостаточной интерпретируемости моделей, расширяются [24]. С другой стороны, именно различные способы получения информации о будущих тенденциях преступности, их сочетанный характер позволят эффективно развиваться криминологии как научной обоснованной теории о разработке мер противодействия преступности, полагают ученые.

Обсуждения машинных прогнозов, имеющиеся в зарубежной криминологии, затрагивают вопросы, к примеру, анализа процессов, лежащих в основе судебных решений. Частым акцентом является оценка ошибок прогнозирования, дискуссии о том, как методы машинного обучения должны «взвешивать» такие ошибки [25–28].

Исследование социальных сетей путем применения машинного обучения и выработка прогнозных тенденций определенных видов преступлений, как и анализ проблем, сопровождающих применение таких методик, остается объектом внимания [29; 30]. Попытки прогноза преступного поведения на основе анализа ежедневных сообщений в Твиттере, на основе ежедневной полярности настроений и особенностей демографических факторов показали, по мнению авторов, возможность оценивать криминальные риски в режиме реального времени, оценивая быстро меняющуюся тактику преступников. При этом исследователи рассматривают в качестве удобного для интерпретации инструмента МО алгоритм XGboost с настройкой [31].

Исследователи предлагают и нейронно-нечеткие модели для прогнозирования преступлений, используя время совершения преступления, а также его долготу и широту, основываясь

¹ Юхо Пяякконен. Обзор использования больших данных в политических и социальных науках. Политическая методология и интернет-методы. Университет Хельсинки.

на алгоритмах k-средних для кластеризации записей о преступлениях по их местоположению. Закономерность преступлений в каждом кластере изучается с помощью собственных нейронно-нечетких сетей, а также с помощью алгоритма «прыгающая лягушка» для определения оптимального радиуса кластера. Результаты такого эксперимента показали, что предложенная модель является наиболее точной и имеет наименьшую погрешность при прогнозировании совершения преступлений в сравнении с результатами регрессии по методу Гаусса, дерева классификации. Поэтому ее можно использовать в качестве эффективного инструмента в реальных условиях [32].

Предиктивная аналитика рассматривается как система определенных методик и технологических инструментов, позволяющих на основе заданных параметров прогнозировать будущие тенденции преступного поведения и иных, непосредственно связанных с ним, факторов и последствий. Ее эффективность зависит в значительной мере от качества самих данных, от специфики используемых алгоритмов и других параметров исследования.

Зарубежная парадигма, таким образом, связана с критическим осмыслением языка алгоритмического прогнозирования преступности как более эффективного и привлекательного с точки зрения и затрат, и объективности базирующейся на нем предупредительной деятельности, однако знания о вычислительных методах, методах машинного обучения, позволяющих делать прогнозы, остаются неопределенными. Исследователи эпистемологии, прогнозирующей деятельности правоохранительных органов, также неоднозначны в своих оценках критериев достоверности прогнозов о преступлениях и условий подотчетности при принятии решений правоохранительными органами. Отношение к социально-техническим практикам алгоритмического прогнозирования преступности, таким образом, остается неоднозначным [13].

О ценности *машинного обучения* в исследованиях заявлено более десятилетия назад и сегодня оно активно применяется. Машинное обучение позволяет автоматизировать исследование данных, что является прорывом в области компьютерных наук. Данная интеллектуальная система «учится» на основе данных и оценивает сложные функции для последующих прогнозов по новым данным. Инструменты МО весьма разнообразны и существуют их классификации

по механизму «обучения» или по типу обретаемого им опыта в процессе обучения. В управляемом машинном обучении (УМО), по свидетельству специалистов, алгоритм наблюдает выводы (Y) для всех входящих данных (X). Этот вывод дает алгоритму цель для прогнозирования. В неконтролируемом же машинном обучении (НМО) алгоритм наблюдает только входные данные (X). Он должен расшифровать данные без информации о правильных выводах [31; 32]. Различные технологические подходы к исследованию БД, различные виды алгоритмов, таким образом, используются и совершенствуются.

Существует множество различных алгоритмов, которые анализируются и сравниваются на основе их эффективности с точки зрения точности, разборчивости, запоминания и оценки. Некоторые из этих алгоритмов включают логистическую регрессию, метод K-ближайших соседей (KNN), метод опорных векторов (SVM), Случайный лес, Наивный Байесовский алгоритм, Дерево решений, многослойный персептрон (MLP) и XG Boost [33].

В настоящее время благодаря внедрению новых алгоритмов обнаружения аномалий обеспечивается снижение финансовых потерь посредством выявления данных о мошеннических транзакциях в режиме реального времени. Есть много других типов мошенничества, в частности, страховые компании используют алгоритмы кластеризации для разделения данных и определения свойств объектов внутри кластеров. Поэтому обнаружение вторжений в телекоммуникационной отрасли стало важной практикой интеллектуального анализа данных, включающей в себя анализ ассоциаций и корреляции, методы агрегирования, визуализации и инструменты запросов, которые могут эффективно обнаруживать любые аномалии или отклонения [34].

В социологических и криминологических современных исследованиях используются эвристические инструменты для оценки полученных с помощью методов машинного обучения результатов. При этом специалисты в сфере их применения заявляют о том, что выходные данные из МО обычно становятся входными данными, позволяющими *последующий анализ или теоретизирование*. Цель их использования в социологии — обработка и группировка данных социальных сетей, других новых видов данных (изображений или текста). Методы МО предоставляют широкий спектр и набор инструментов

для изучения данных, которые могут информировать о разных проблемах.

Вместе с тем, опираясь на традиционный классический подход, базирующийся на проверке гипотез, большинство ученых используют МО для оценки влияния некоторых входных параметров на выводы. МО улучшает стратегический потенциал исследования, оценивает наборы данных и получаемые прогнозы, их отклонение от наблюдаемых результатов, оценивая базовую структуру некоторых входных данных и их связь с внешними факторами. Ответы на такие вопросы выступают основой совершенствования теории и генерирования новых гипотез. Получаемые результаты выступают отправной точкой для дальнейшего анализа и *дополняют, а не заменяют существующие методы в социологии* [35].

Сегодня активно исследуются вопросы и об информационных издержках, сопряженных с созданием и использованием информационной технологии больших данных. Специалисты рассуждают о технических сложностях структурирования и интеграции данных. Концепция больших данных предполагает совместный анализ структурированных и неструктурированных данных, однако этот механизм осложнен необходимостью согласования схем описания и правил отнесения данных к определенной предметной области, т.е. необходимостью приведения данных к единому формату [36]. Консолидация, понимаемая через централизацию их хранения, федерализация или унификация доступа и распространение данных составляют методы интеграции данных для обеспечения возможности использования технологии БД. Предлагаемая учеными концептуальная функциональная модель технологии больших данных исследует в том числе структуру и характер инвестиционных издержек и затрат при создании и внедрении в предметную область технологии БД, а также издержек на стадии предподготовки данных. Структура таких затрат традиционна для концептуального проектирования информационных систем. И, оценивая перспективы и направления монетизации БД, т.е. превращения их в источник дохода, авторы связывают это с возможностью снижения характерных для технологии информационных издержек, через специальные научные исследования в технологической и экономической областях. Такие разработки приводят к оценке перспективы широкого применения технологий БД как неопределенной, зависящей от успеш-

ности ее монетизации, обусловленной разработкой методологической и методической базы технико-экономического проектирования соответствующих систем.

Среди проблем технологического сопровождения механизма использования БД узкие специалисты, рассматривая современные подходы к разработке хранилищ данных и технологий работы с БД, отмечают, что задача выбора способов хранения является одной из основных и трудозатратных, но имеющей важное практическое значение. Математическое обеспечение выбора эффективного формата файлов для хранения витрин в озерах данных, разработка методик анализа влияния на параметры обработки больших объемов данных в программной инфраструктуре, выбор формата хранения данных в условиях, определяемых функциями и компонентами информационно-аналитической среды остаются актуальными задачами в данном аспекте. Создание информационных структур — методов и алгоритмов, и использование их для обработки данных являются целями научных исследований больших объемов данных. Использование БД — это процесс управления, обработки запросов и анализа данных, содержащихся в информационных структурах для научных исследований.

Таким образом, перестройка всей системы аналитической и прогностической работы в сфере анализа преступности и перестройка психологии самих исследователей и правоохранителей — стратегически важнейшая задача в развитии криминологии.

Значимость БД и связанных с ними технологий познания криминологической реальности проявляется в двух основных аспектах: использование БД в криминологических исследованиях в реальном или практически реальном временном режиме, что соответствует направлению развития социальных наук в целом, и использование моделирования и алгоритмов в качестве одного из прогностических инструментов для их последующей оценки и учета в работе правоохранительной системы.

Границы прогнозирования в аспекте выявления неизвестных закономерностей в данных, без учета ценности их причинно-следственных объяснений и в условиях недостаточной интерпретируемости моделей, расширяются. С другой стороны, именно различные способы получения информации о будущих тенденциях преступности, их сочетанный характер позволяют эффективно развиваться криминологии как на-

учной обоснованной теории о разработке мер противодействия преступности.

Методы машинного обучения предоставляют широкий спектр и набор инструментов для изучения данных, которые могут информировать о разных проблемах. Машинное обучение улучшает стратегический потенциал исследования, оценивает наборы данных и получаемые прогнозы, их отклонение от наблюдаемых результатов, оценивая базовую структуру некоторых входных данных и их связь с внешними

факторами. Ответы на такие вопросы выступают основой совершенствования теории и генерирования новых гипотез. Получаемые результаты выступают отправной точкой для дальнейшего анализа и *дополняют, а не заменяют существующие методы в социолого-криминологических исследованиях.*

Эпистемологическая концепция учитывает и ориентирует использование технологий анализа больших данных применительно к различным научным контекстам.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Малинецкий Г.Г. Синергетика, междисциплинарность и постнеклассическая наука XXI века / Г.Г. Малинецкий. — Москва : ИПМ им. М.В. Келдыша РАН, 2013. — 36 с.
2. Сорочан В.В. Цифровая социология: от цифровизации общества к цифровизации науки / В.В. Сорочан, НП Гаврилюк. — EDN IINTPP // Этносоциум и межнациональная культура. — 2023. — № 1 (175). — С. 37–46.
3. Горошникова Т.А. Вычислительная социология: сдвиг парадигмы или "эконометрика" социологии / Т.А. Горошникова. — DOI 10.26794/2226-7867-2021-11-1-37-42. — EDN IZOPMZ // Гуманитарные науки. Вестник Финансового университета. — 2021. — Т. 11, № 1. — С. 37–42.
4. Рафикова К.Ф. Методология вычислительной социологии / К.Ф. Рафикова. — EDN OLKPFC // Социологический нарратив 2023: новая социальная реальность: жизнь на пороге киберпанка : материалы XXII Всерос. науч. конф., Москва, 14 апр. 2023 г. — Москва, 2023. — С. 7–10.
5. Платонова С.И. Большие данные: создание вызовов и возможностей в социальных науках / С.И. Платонова. — DOI 10.30853/manuscript.2020.4.24. — EDN RSHXAS // Манускрипт. — 2020. — Т. 13, № 4. — С. 119–123.
6. Карчагин Е.В. Эпистемология и эвристические возможности Big Data / Е.В. Карчагин. — EDN VCEAXJ // Концепт. — 2015. — № 12. — С. 151–155.
7. Горошко И.В. Авторская методика анализа данных. (на примере анализа данных в органах внутренних дел) / И.В. Горошко. — EDN YDGANR // Диссертационное исследование: технологии подготовки : в 2 ч. — Москва, 2023. — Ч. 2. — С. 201–226.
8. Луценко Е.В. Революция начала XXI века в искусственном интеллекте: глубинные механизмы и перспективы / Е.В. Луценко, Н.С. Головин. — Краснодар : Изд-во КубГАУ, 2024. — 497 с. — EDN OMIPIL.
9. Постклассические исследования права: перспективы научно-практической программы / ред. Е.И. Тонков, И.Л. Честнов. — Санкт-Петербург : Алетей, 2023. — 498 с. — EDN KLUMOD.
10. Туробов А.В. Международный опыт применения математико-статистических алгоритмов прогнозирования преступности / А.В. Туробов, М.А. Чумакова, А.В. Вечерин. — DOI 10.17994/IT.2019.17.4.59.9. — EDN XMF0EI // Международные процессы. — 2019. — Т. 17, № 4 (59). — С. 153–177.
11. Ильницкий А.С. Технологии искусственного интеллекта в криминологической науке: компьютерное зрение / А.С. Ильницкий. — EDN DXUSVZ // Алтайский юридический вестник. — 2024. — № 4 (48). — С. 145–151.
12. Грибанов Е.В. Перспективные направления развития кибертехнологий предупреждения преступлений / Е.В. Грибанов. — EDN UNVRKW // Общество и право. — 2021. — № 4. — С. 21–27.
13. Соловьев В.С. Криминологические исследования преступности несовершеннолетних в условиях цифровизации общества: ожидания и реальность / В.С. Соловьев. — EDN QAFBFP // Алтайский юридический вестник. — 2024. — № 1 (45). — С. 92–99.
14. Ищук Я.Г. Цифровая криминология / Я.Г. Ищук, Т.В. Пинкевич, Е.С. Смольянинов. — Москва : Изд-во Акад. упр. МВД РФ, 2021. — 244 с. — EDN VCKOEF.
15. Big Data in Investigating and Preventing Crimes / E. Bulgakova, V. Bulgakov, I. Trushchenkov [et al] // Big Data-driven World: Legislation Issues and Control Technologies / ed. A.G. Kravets. — Springer, 2019. — P. 61–69.
16. Манифест новой количественной криминологии «Уголовная политика с опорой на данные» / К.Д. Титаев, Д.А. Скугаревский, А.В. Кнорре [и др.]. — DOI 10.21638/11701/spbu14.2018.107. — EDN YTYBJA // Вестник Санкт-Петербургского университета. Право. — 2018. — Т. 9, № 1. — С. 91–101.
17. Молотникова А.А. Использование панельных данных для анализа криминогенной обстановки в федеральных округах России / А.А. Молотникова, Д.В. Звонкова. — EDN UEATNP // Наука и образование: хозяйство и экономика; предпринимательство; право и управление. — 2019. — № 1 (104). — С. 92–103.
18. Багманова Л.Р. Моделирование социальных факторов, влияющих на уровень преступности в г. Казани / Л.Р. Багманова. — EDN XTGKUP // Вестник современных исследований. — 2018. — № 5.1 (20). — С. 324–332.
19. Овчинский В. Как изучать организованную преступность в XXI веке / В. Овчинский // Изборский клуб. — 2018. — 14 мая. — URL: <https://izborsk-club.ru/15222>.
20. Canali S. Big Data, epistemology and causality: Knowledge in and knowledge out in Exposomics / S. Canali // Big Data & Society. — 2016. — URL: <https://philpapers.org/archive/CANBDE.pdf>.
21. Bonikowski B. Varieties of American Popular Nationalism / B. Bonikowski, P. DiMaggio // American Sociological Review. — 2016. — Vol. 81, no. 5. — P. 949–980.

22. Cederman, L.E. Predicting armed conflict : Time to adjust our expectations? / L.E. Cederman, N.B. Weidmann // Science. — 2017. — Vol. 355. — P. 474–476.
23. DiMaggio P. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding / P. DiMaggio, M. Nag, D. Blei // Poetics. — 2013. — Vol. 41, no. 6. — P. 570–606.
24. Chan J. Is Big Data Challenging Criminology? / J. Chan, L.B. Moses // University of New South Wales Law Research Series. — 2020. — No. 20-81. — URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3742536#.
25. Halterlein J. Epistemologies of predictive policing: Mathematical social science, social physics and machine learning / J. Halterlein. — DOI 10.1177/20539517211003118 // Big Data & Society. — 2021. — Vol. 8, no. 1. — P. 1–13.
26. Domingos P. The master algorithm : How the quest for the ultimate learning machine will remake our world / P. Domingos. — New York : Basic Books, 2015. — 360 p.
27. Donoho D. 50 Years of Data Science / D. Donoho // Journal of Computational and Graphical Statistics. — 2017. — Vol. 26, no. 4. — P. 745–766.
28. Crowdsourcing City Government : Using Tournaments to Improve Inspection Accuracy / E.L. Glaeser, A. Hillis, S.D. Kominers, M. Luca // American Economic Review. — 2016. — Vol. 106, no. 5. — P. 114–118.
29. Monika. Predictive Analytics of Crime Data in Social Media: A Systematic Review, Incorporating Framework, and Future Investigation Schedule / Monika, A. Bhat // SN Computer Science. — 2025. — Vol. 6, 11 March. — P. 270.
30. Taiwo G.A. Crime Prediction Using Twitter Sentiments and Crime Data / G.A. Taiwo, M. Saraee, J. Fatai. — DOI 10.31449/inf.v48i6.4749 // Informatica. — 2024. — Vol. 48, no. 6. — P. 35–42.
31. Jaber M. A model for predicting crimes using big data and neural-fuzzy networks / M. Jaber, R. Sheibani, H. Shakeri. — DOI 10.1002/cpe.6985 // Concurrency and Computation: Practice and Experience. — 2022. — Vol. 34, no. 17.
32. Ключева И.А. Области применения интеллектуального анализа данных / И.А. Ключева. — DOI 10.5281/zenodo.5529200. — EDN IXAIWI // Инновационные научные исследования. — 2021. — № 7 (9). — С. 173–194.
33. Магомедтагиров М.М. Применение машинного обучения в социологии / М.М. Магомедтагиров, С.З. Айгубов, С.Р. Омарова. — DOI 10.26726/1812-7096-2019-12-233-239. — EDN ORMTTA // Региональные проблемы преобразования экономики. — 2019. — № 12 (110). — С. 233–239.
34. Шальнова Ю.П. Монетизация больших данных: технико-экономический анализ драйверов роста и издержек / Ю.П. Шальнова. — DOI 10.18413/2687-0932-2020-43-3-491-500. — EDN HMKZPG // Экономика. Информатика. — 2020. — 47 (3). С. 491–500.
35. Нгуен Нгок Зьонг. Методики и алгоритмы извлечения знаний из реляционных баз данных на основе семантики предметной области : автореф. дис. ... канд. техн. наук : 05.13.17 / Нгуен Нгок Зьонг. — Пенза, 2020. — 17 с.
36. Белов В.А. Методики анализа форматов хранения и глобально распределенной обработки больших объемов данных : автореф. дис. ... канд. техн. наук : 2.3.5 / В.А. Белов. — Москва, 2023. — 17 с.

REFERENCES

1. Malinetskii G.G. *Synergetics, Multidisciplinarity and Post-neoclassical Science of the 21st Century*. Moscow, Keldysh Institute of Applied Mathematics of the Russian Academy of Sciences Publ., 2013. 36 p.
2. Sorochan V.V., Gavriluk N.P. Digital Sociology: from the Digitalization of Society to the Digitalization of Science. *Etnosotium i mezhnatsional'naya kul'tura = Etnosocium (Multinational Society)*, 2023, no. 1, pp. 37–46. (In Russian). EDN: IINTPP.
3. Goroshnikova T.A. Computational Sociology: a Paradigm Shift or "Econometrics" of Sociology. *Gumanitarnye nauki. Vestnik Finansovogo universiteta = Humanities and Social Sciences. Bulletin of the Financial University*, 2021, vol. 11, no. 1, pp. 37–42. (In Russian). EDN: IZOPMZ. DOI: 10.26794/2226-7867-2021-11-1-37-42.
4. Rafikova K.F. Methodology of Computational Sociology. In *Sociological Narrative 2023: New Social Reality: Life at the Verge of Cyberpunk*. Proceedings of the XXII All-Russian Scientific Conference, Moscow, April 14, 2023. Moscow, 2023, pp. 7–10. (In Russian). EDN: OLKPFC.
5. Platonova S.I. Big Data: Challenges and Opportunities in Social Sciences. *Manuskript = Manuscript*, 2020, vol. 13, no. 4, pp. 119–123. (In Russian). EDN: RSHXAS. DOI: 10.30853/manuskript.2020.4.24.
6. Karchagin E.In. Epistemology and Heuristic Big Data. *Kontsept = Concept*, 2015, no. 12, pp. 151–155. (In Russian). EDN: VCEAXJ.
7. Goroshko I.V. The Author's Method of Data Analysis (Using the Example of Data Analysis in Internal Affairs' Bodies). In *Dissertation Research: Technologies of Preparation*. Moscow, 2023. Pt. 2, pp. 201–226. (In Russian). EDN: YDGANR.
8. Lutsenko E.V., Golovin N.S. *The AI Revolution of the Early 21st Century: Deep Mechanisms and Prospects*. Krasnodar, Kuban State Agrarian University Publ., 2024. 497 p. EDN: OMIPIL.
9. Tonkov E.I., Chestnov I.L. *Postclassical Studies of Law*. Saint Petersburg, Aleteiya Publ., 2023. 498 p. EDN: KLUMOD.
10. Turobov A.V., Chumakova M.A., Vecherin A.V. World Best Practices in Applying Mathematical and Statistical Crime Prediction Algorithms. *Mezhdunarodnye protsessy = International Trends*, 2019, vol. 17, no. 4, pp. 153–177. (In Russian). EDN: XM-FOEI. DOI: 10.17994/IT.2019.17.4.59.9.
11. Ilnitsky A.S. Artificial Intelligence Technologies in Criminological Science: Computer Vision. *Altaiiskii yuridicheskii vestnik = Altai Law Journal*, 2024, no. 4, pp. 145–151. (In Russian). EDN: DXUSVZ.
12. Gribanov E.V. Promising Areas of Development of Cyber Technologies for Crime Prevention. *Obshchestvo i pravo = Society and Law*, 2021, no. 4, pp. 21–27. (In Russian). EDN: UNVRKW.
13. Solovov V.S. Criminological Research of Juvenile Crime under Digitalization of Society: Expectations and Reality. *Altaiiskii yuridicheskii vestnik = Altai Law Journal*, 2024, no. 1, pp. 92–99. (In Russian). EDN: QAFBFP.
14. Ishchuk Ya.G., Pinkevich T.V., Smol'yaninov E.S. *Digital Criminology*. Moscow, Management Academy of the Ministry of the Interior of Russian Federation Publ., 2021. 244 p. EDN: VCKOEF.

15. Bulgakova E., Bulgakov V., Trushchenkov I., Vasilev D., Kravets E. Big Data in Investigating and Preventing Crimes. In Kravets A.G. (ed.). *Big Data-Driven World: Legislation Issues and Control Technologies*. Springer, 2019, pp. 61–69.
16. Titaev K.D., Skougarevskiy D.A., Knorre A.V., Kudryavtsev V.E., Shklyaruk M.S. A New Quantitative Criminology Manifesto: Evidence Based Criminal Justice Policymaking. *Vestnik Sankt-Petersburgskogo universiteta. Pravo = Vestnik of Saint-Petersburg University. Law*, 2018, vol. 9, no. 1, pp. 91–101. (In Russian). EDN: YTYBJA. DOI: 10.21638/11701/spbu14.2018.107.
17. Molotnikova A.A., Zvonkova D.V. The Panel Data Use for the Analysis of the Crime Situation in the Federal Districts of Russia. *Nauka i obrazovanie: khozyaistvo i ekonomika, predprinimatel'stvo; pravo i upravlenie = Science and Education: Economy and Economics, Entrepreneurship; Law and Management*, 2019, no. 1, pp. 92–103. (In Russian). EDN: UEATNP.
18. Bagmanova L.R. Modeling of Social Factors Influencing the Crime Rate in Kazan. *Vestnik sovremennykh issledovaniy = Bulletin of Modern Studies*, 2018, no. 5.1, pp. 324–332. (In Russian). EDN: XTGKUP.
19. Ovchinskii V. How to Study Organized Crime in the 21st Century. *Izbornskii Club*, 2018, May 14. URL: <https://izborsk-club.ru/15222>. (In Russian).
20. Canali S. Big Data, Epistemology and Causality: Knowledge in and Knowledge out in Exposomics. *Big Data & Society*, 2016. URL: <https://philpapers.org/archive/CANBDE.pdf>.
21. Bonikowski B., DiMaggio P. Varieties of American Popular Nationalism. *American Sociological Review*, 2016, vol. 81, no. 5, pp. 949–980.
22. Cederman, L.E., Weidmann, N.B. Predicting Armed Conflict: Time to Adjust Our Expectations? *Science*, 2017, vol. 355, pp. 474–476.
23. DiMaggio P., Nag M., Blei D. Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government arts Funding. *Poetics*, 2013, vol. 41, no. 6, pp. 570–606.
24. Chan J., Moses L.B. Is Big Data Challenging Criminology? *University of New South Wales Law Research Series*, 2020, no. 20-81. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3742536#.
25. Halterlein J. Epistemologies of Predictive Policing: Mathematical Social Science, Social Physics and Machine Learning. *Big Data & Society*, 2021, vol. 8, no. 1, pp. 1–13. DOI: 10.1177/20539517211003118.
26. Domingos P. *The Master Algorithm: How the Quest for the Ultimate Learning Machine will Remake our World*. New York, Basic Books, 2015. 360 p.
27. Donoho D. 50 Years of Data Science. *Journal of Computational and Graphical Statistics*, 2017, vol. 26, no. 4, pp. 745–766.
28. Glaeser E.L., Hillis A., Kominers S.D., Luca M. Crowdsourcing City Government: Using Tournaments to Improve Inspection Accuracy. *American Economic Review*, 2016, vol. 106, no. 5, pp. 114–118.
29. Monika, Bhat A. Predictive Analytics of Crime Data in Social Media: A Systematic Review, Incorporating Framework, and Future Investigation Schedule. *SN Computer Science*, 2025, vol. 6, 11 March, pp. 270.
30. Taiwo G.A., Saraee M., Fatai J. Crime Prediction Using Twitter Sentiments and Crime Data. *Informatica*, 2024, vol. 48, no. 6, pp. 35–42. DOI 10.31449/inf.v48i6.4749.
31. Jaber M., Sheibani R., Shakeri H. A Model for Predicting Crimes Using Big Data and Neural-Fuzzy Networks. *Concurrency and Computation: Practice and Experience*, 2022, vol. 34, no. 17. DOI: 10.1002/cpe.6985.
32. Klyueva I.A. The Areas of Application of Data Mining. *Innovatsionnye nauchnye issledovaniya = Innovative Scientific Research*, 2021, no. 7, pp. 173–194. (In Russian). EDN: IXAIWI. DOI: 10.5281/zenodo.5529200.
33. Magomedtagirov M.M., Aigubov S.Z., Omarova S.R. Application of Machine Learning in Sociology. *Regional'nye problemy preobrazovaniya ekonomiki = Regional Problems of Economic Transformation*, 2019, no. 12, pp. 233–239. (In Russian). EDN: ORM-TTA. DOI: 10.26726/1812-7096-2019-12-233-239.
34. Shalnova Yu.P. Big Data Monetization: Qualitative Technical and Economic Analysis of Drivers of Growth and Costs. *Ekonomika. Informatika = Economics. Information Technologies*, 2020, vol. 47, no. 3, pp. 491–500. (In Russian). EDN: HMKZPG. DOI: 10.18413/2687-0932-2020-43-3-491-500.
35. Nguen Ngok Zyong. *The Methods and Algorithms of Extracting Knowledge from Relative Databases on the Basis of Data Domain's Semantics. Cand. Diss. Thesis*. Penza, 2020. 17 p.
36. Belov V.A. *Methods of Analyzing the Formats of Storing and Distributed Global Processing of Big Data. Cand. Diss. Thesis*. Moscow, 2023. 17 p.

ИНФОРМАЦИЯ ОБ АВТОРЕ

Судакова Татьяна Михайловна — заведующий кафедрой уголовного права и криминологии Института юстиции Байкальского государственного университета, кандидат юридических наук, доцент, г. Иркутск, Российская Федерация; e-mail: krime@mail.ru.

ДЛЯ ЦИТИРОВАНИЯ

Судакова Т.М. Теория и методология больших данных: криминологическо-эпистемологический ракурс / Т.М. Судакова. — DOI 10.17150/2500-4255.2025.19(2).126-136. — EDN ZQKNOC // Всероссийский криминологический журнал. — 2025. — Т. 19, № 2. — С. 126–136.

INFORMATION ABOUT THE AUTHOR

Sudakova, Tatiana M. — Head, Department of Criminal Law and Criminology, Institute of Justice, Baikal State University, Ph.D. in Law, Ass. Professor, Irkutsk, the Russian Federation; e-mail: krime@mail.ru.

FOR CITATION

Sudakova T.M. The Theory and Methodology of Big Data: A Criminological-Epistemological Perspective. *Vserossiiskii kriminologicheskii zhurnal = Russian Journal of Criminology*, 2025, vol. 19, no. 2, pp. 126–136. (In Russian). EDN: ZQKNOC. DOI: 10.17150/2500-4255.2025.19(2).126-136.