

# МАШИННОЕ ОБУЧЕНИЕ

## ПРИ АНАЛИЗЕ ГОСУДАРСТВЕННОГО РЕГУЛИРОВАНИЯ И УПРАВЛЕНИЯ СУБЪЕКТАМИ НЕПРОИЗВОДСТВЕННОЙ СФЕРЫ

Монография



Санкт-Петербург,  
Наукоемкие технологии, 2026

Научно-издательский центр  
ООО «Актуальность РФ»

# **МАШИННОЕ ОБУЧЕНИЕ ПРИ АНАЛИЗЕ ГОСУДАРСТВЕННОГО РЕГУЛИРОВАНИЯ И УПРАВЛЕНИЯ СУБЪЕКТАМИ НЕПРОИЗВОДСТВЕННОЙ СФЕРЫ**

***Монография***

*Электронное текстовое издание*

Санкт-Петербург  
Научно-технологические  
2026

© Бондарь В. В., Босюк В. Н., Горобиевская С. В.,  
Гощина Н. Н., Надькин Л. Ю., Салко Д. А., Скитская Л. В., 2026  
ISBN 978-5-00271-111-6

УДК 343.9 : 004.8  
ББК Х511+А663.1  
М38

*Авторы:*

**В. В. Бондарь, В. Н. Босюк, С. В. Горобиевская, Н. Н. Гощина, Л. Ю. Надькин, Д. А. Салко, Л. В. Скитская**

*Рецензенты:*

**В. А. Щербаков**, доктор физико-математических наук, профессор специальность ВАК – 01.01.06 «Математическая логика, алгебра и теория чисел», главный научный сотрудник Института математики и информатики АНМ (академии наук Молдавии), г. Кишинев

**А. П. Погребняк**, доктор сельскохозяйственных наук, профессор, НИЛ «Экономические исследования», Приднестровский государственный университет им. Т. Г. Шевченко, г. Тирасполь

**Машинное обучение при анализе государственного регулирования и управления субъектами непроеизводственной сферы** [Электронный ресурс]: монография / В. В. Бондарь, В. Н. Босюк, С. В. Горобиевская, Н. Н. Гощина, Л. Ю. Надькин, Д. А. Салко, Л. В. Скитская. – Санкт-Петербург : Научные технологии, 2026. – 48 с. – URL: <https://publishing.intelgr.com/archive/Mashinnoe-obuchenie-pri-analize-gosudarstvennogo-regulirovaniya-i-upravleniya-subektami-neproizvodstvennoi-sferi.pdf>.

ISBN 978-5-00271-111-6

*Монография посвящена исследованию возможностей применения методов машинного обучения в системе мониторинга и анализа уровня преступности. В работе рассматриваются теоретические основы криминологического мониторинга, анализируются традиционные статистические подходы и их ограничения при работе с большими и разнородными массивами данных. Особое внимание уделено интерпретируемым моделям машинного обучения – деревьям решений, ансамблевым методам и алгоритму случайного леса, а также инструментам объяснимого искусственного интеллекта (SHAP), позволяющим количественно оценивать вклад факторов в формирование прогнозов. Представлена программная реализация аналитической системы на базе современных цифровых инструментов, включающая обработку, моделирование и визуализацию данных. Результаты исследования демонстрируют практическую применимость интеллектуальных методов при принятии управленческих решений в сфере обеспечения общественной безопасности.*

*Монография адресована специалистам в области криминологии, государственного управления и аналитики данных.*

**УДК 343.9 : 004.8  
ББК Х511+А663.1**

ISBN 978-5-00271-111-6

© Бондарь В. В., Босюк В. Н., Горобиевская С. В.,  
Гощина Н. Н., Надькин Л. Ю., Салко Д. А., Скитская Л. В., 2026

*Научное издание*

**Бондарь** Виталий Вячеславович  
**Босюк** Виктор Николаевич  
**Горобиевская** Светлана Васильевна  
**Гощина** Наталия Николаевна  
**Надькин** Леонид Юрьевич  
**Салко** Дмитрий Анатольевич  
**Скитская** Лариса Валерьевна

МАШИННОЕ ОБУЧЕНИЕ ПРИ АНАЛИЗЕ ГОСУДАРСТВЕННОГО РЕГУЛИРОВАНИЯ  
И УПРАВЛЕНИЯ СУБЪЕКТАМИ НЕПРОИЗВОДСТВЕННОЙ СФЕРЫ

*Монография*

Электронное текстовое издание

Публикуется из оригинал-макета, подготовленного авторами

Подписано к использованию 12.02.2026.

Объем издания – 2,0 Мб.

*Издательство «Наукоемкие технологии»*

*ООО «Корпорация «Интел Групп»*

*<https://publishing.intelgr.com>*

*E-mail: [publishing@intelgr.com](mailto:publishing@intelgr.com)*

*Тел.: +7 (812) 945-50-63*

*Интернет-магазин издательства*

*<https://shop.intelgr.com/>*

ISBN 978-5-00271-111-6



9 785002 711116 >

## ВВЕДЕНИЕ

В условиях цифровизации современного общества проблема обеспечения общественной безопасности приобретает особую значимость. Рост объёмов информации требует пересмотра традиционных подходов к её контролю и анализу. Мониторинг уровня преступности перестаёт быть исключительно статистической процедурой и рассматривается как аналитический процесс, который направлен на выявление закономерностей, прогнозирование и влияет на принятие управленческих решений в правоохранительной сфере деятельности.

Вопросы применения количественных и статистических методов к анализу социально-экономических и криминологических процессов достаточно подробно рассматривались в трудах отечественных исследователей. Так, в работах С. А. Айвазяна представлены методологические основы анализа временных рядов и многомерных статистических данных, которые могут быть использованы при исследовании динамики преступности и факторов, влияющих на её уровень [1; 3].

Проблемам статистического анализа преступности и интерпретации криминологических показателей посвящены исследования Ю. Д. Блувштейна, в которых показаны возможности и ограничения традиционных методов описательной статистики при анализе криминальных данных [6].

Современные подходы к оценке криминогенной ситуации с использованием математического моделирования получили развитие в работах В. Н. Босюка и соавторов, где преступность рассматривается как многофакторный объект анализа с выраженными нелинейными и пространственно-временными зависимостями [8; 11; 12].

На протяжении многих лет анализ уровня преступности основывался на использовании методов описательной статистики, корреляционно-регрессионного анализа. Несмотря на их научную обоснованность и широкое распространение, эти подходы обладают рядом существенных ограничений. Также они неэффективны при работе с большими массивами разнородных данных. Эти подходы не всегда позволяют учитывать нелинейные зависимости и часто оказываются недостаточно гибкими в условиях быстро меняющейся криминогенной обстановки. Классические методы ориентированы на ретроспективный анализ и в меньшей степени ориентированы на прогнозирование.

Развитие цифровых технологий и накопление значительных объёмов криминологических данных создают предпосылки для внедрения более удобных аналитических инструментов. В предлагаемом исследовании мы используем методы машинного обучения, которые позволяют автоматически выявлять скрытые закономерности в данных, работать с большим числом признаков. Использование таких методов открывает новые возможности для мониторинга уровня преступности: например, это помогло выявить территориальные «горячие точки», влияющие криминальную ситуацию.

В последние годы методы машинного обучения всё активнее применяются в самых разных областях – от экономики и медицины до маркетинга и информационной безопасности. При этом в криминологическом анализе и задачах мониторинга преступности подобные методы используются заметно реже. Во многом это объясняется спецификой исходных данных. Информация о преступлениях часто оказывается неполной, содержит ошибки и отражает лишь зарегистрированную часть правонарушений, тогда как латентная преступность остаётся за рамками статистики.

При решении задач анализа преступности особенно важно использовать такие модели, которые позволяют не только получать количественные оценки, но и понимать причины полученных результатов. Для практического применения принципиальное значение имеет возможность объяснить, какие факторы оказывают наибольшее влияние на уровень преступности и каким образом они взаимодействуют между собой. В этом отношении интерпретируемые модели машинного обучения обладают существенными преимуществами по сравнению со сложными алгоритмами, работа которых остаётся непрозрачной для пользователя. Деревья решений и ансамблевые методы, в том числе алгоритм случайного леса, представляют собой удачное сочетание приемлемой точности и наглядности, что делает их особенно востребованными при анализе криминологических данных.

Актуальность настоящего исследования определяется необходимостью поиска и обоснования эффективных подходов к применению методов машинного обучения в системе мониторинга преступности. Использование современных аналитических инструментов позволяет более глубоко анализировать криминогенную обстановку, выявлять скрытые закономерности. В практическом плане это создаёт условия для принятия более взвешенных и обоснованных управленческих решений в деятельности правоохранительных органов.

В рамках данной работы основное внимание уделяется анализу возможностей и ограничений применения методов машинного обучения

при мониторинге преступности, а также разработке практических подходов к их использованию в составе информационно-аналитической системы. В ходе исследования рассматриваются теоретические основы мониторинга преступности и существующие подходы к анализу данных, изучается структура и специфика информации о преступности, а также проблемы, возникающие при её обработке. Особое внимание уделяется оценке возможностей применения методов машинного обучения для анализа уровня преступности, а также принципам работы деревьев решений и ансамблевых моделей, включая алгоритм случайного леса.

Практическая часть работы связана с реализацией выбранных моделей и оценкой их эффективности на примере реальных данных криминологической ситуации в республике. Полученные результаты анализируются с точки зрения их применимости в задачах мониторинга преступности.

Практическая значимость исследования определяется возможностью использования его результатов при разработке и внедрении информационно-аналитической системы мониторинга преступности. Материалы работы могут быть использованы аналитическими подразделениями правоохранительных органов.

В современных условиях обеспечение общественной безопасности и противодействие преступности остаются важнейшими направлениями деятельности государства. Уровень преступности формируется под воздействием множества социально-экономических, территориальных и организационно-управленческих факторов, влияние которых носит сложный и нелинейный характер. В связи с этим возрастает значение количественных методов анализа, позволяющих не только описывать текущую ситуацию, но и оценивать вклад отдельных факторов в формирование криминогенной обстановки.

Традиционные подходы к анализу преступности, основанные преимущественно на описательной статистике, имеют определённые ограничения и не всегда позволяют учитывать сложные взаимосвязи между показателями. Развитие методов математической статистики и машинного обучения в сочетании с цифровыми технологиями открывает новые возможности для более глубокого анализа криминогенной ситуации. Использование корреляционного и факторного анализа совместно с методами машинного обучения, такими как деревья решений и случайный лес, позволяет количественно оценивать влияние различных факторов на уровень преступности и использовать полученные результаты в практической деятельности.

Важную роль в реализации аналитических исследований играет применение современных программных средств. Язык программирования

Python, благодаря развитой экосистеме библиотек, широко используется для обработки данных и построения аналитических моделей. Применение web-фреймворка Flask позволяет создавать интерактивные информационно-аналитические системы, а библиотека Leaflet – территориально представлять криминогенную обстановку с использованием интерактивных карт. Использование системы управления базами данных PostgreSQL обеспечивает надёжное хранение и обработку аналитической информации.

# ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ПОДХОДЫ К МОНИТОРИНГУ УРОВНЯ ПРЕСТУПНОСТИ

Теоретические основы мониторинга преступности достаточно широко представлены в трудах отечественных и зарубежных криминологов. В работах В. В. Лунеева преступность рассматривается как социальное явление, формирующееся под воздействием комплекса социально-экономических, демографических факторов, при этом подчёркивается необходимость систематического анализа её динамики и структуры [27].

Существенный вклад в развитие методологии криминологического анализа внесли исследования Ю. М. Антоняна и Д. А. Шестакова, в которых раскрываются социальные детерминанты преступности и подчёркивается значение комплексного подхода к оценке криминогенной ситуации [4; 33].

В работах Ю. Д. Блувштейна и Н. М. Добрынина подробно рассматриваются вопросы статистического учёта преступности, анализируются возможности применения традиционных количественных методов и указываются их ограничения при изучении сложных криминологических процессов [6; 19].

В то же время ряд авторов отмечает, что в условиях усложнения социально-экономических процессов и роста объёма информации традиционные методы анализа преступности требуют методологического расширения и дополнения более гибкими аналитическими инструментами [26].

## 1.1. Мониторинг преступности в системе криминологических исследований

Мониторинг преступности является одним из ключевых инструментов криминологического анализа и играет важную роль в системе обеспечения общественной безопасности. В широком смысле под мониторингом преступности понимается систематический процесс сбора, обработки, анализа и интерпретации информации о состоянии, структуре

и динамике преступности с целью выявления устойчивых тенденций и закономерностей её развития.

В отличие от разовых статистических наблюдений, мониторинг носит непрерывный и комплексный характер. Он ориентирован не только на фиксацию уже совершённых преступлений, но и на анализ факторов, влияющих на криминогенную обстановку, а также на прогнозирование возможных изменений в будущем. Именно эта особенность делает мониторинг преступности важным элементом научного обеспечения деятельности правоохранительных органов и органов государственной власти.

В криминологической науке мониторинг преступности традиционно рассматривается как часть более широкой системы криминологического анализа. Он опирается на данные официальной статистики, результаты социологических исследований, сведения оперативно-служебного характера и иные источники информации. При этом основной задачей мониторинга является не простое накопление данных, а их аналитическая обработка с целью формирования объективного представления о реальном состоянии преступности.

Сегодня социально-экономические процессы заметно усложняют анализ преступности, поскольку она становится более разнородной: меняется по структуре, отличается от территории к территории и всё сильнее зависит от множества факторов – уровня урбанизации, миграции, экономической нестабильности и развития цифровых технологий. В такой ситуации привычные методы изучения, основанные только на усреднённых показателях, уже не дают полного понимания происходящего.

В криминологических исследованиях мониторинг преступности выполняет сразу несколько важных задач. Прежде всего, он формирует информационную базу для науки: позволяет видеть долгосрочные тенденции, отслеживать структурные сдвиги и понимать, как меняется картина с течением времени. Кроме того, мониторинг имеет практическое значение – он даёт органам управления актуальную аналитику, необходимую для принятия решений в сфере профилактики и противодействия преступности. Наконец, у него есть прогностическая функция: на основе накопленных данных можно оценивать криминальные риски и моделировать вероятные сценарии развития ситуации.

В случаях, когда влияние локальных социально-экономических и демографических условий проявляется наиболее отчетливо, мониторинг особенно важен. Пространственный анализ преступности помогает выявлять территории с повышенной криминальной активностью и оценивать, насколько эффективны уже принятые профилактические меры. В этом смысле мониторинг становится не просто инструментом учёта, а полноценным механизмом территориального анализа и планирования.

При этом результативность мониторинга напрямую зависит от методов анализа. Традиционно в криминологии используются описательная статистика, сравнительный анализ и корреляционные методы. Они полезны для получения общей картины и выявления базовых зависимостей, но их возможностей часто недостаточно, когда нужно исследовать сложные нелинейные связи и большие массивы данных.

В условиях цифровизации объём и разнообразие информации стремительно растут. К официальной статистике всё чаще добавляются данные из смежных сфер: социально-экономические индикаторы, геоинформационные данные, материалы из открытых источников. Работа с такими разнородными наборами требует более гибких и мощных аналитических подходов.

Поэтому сегодня мониторинг преступности становится одним из наиболее динамично развивающихся направлений криминологических исследований. Он требует постоянного обновления методологии и инструментов анализа. Современные интеллектуальные методы обработки данных позволяют глубже понять природу криминогенных процессов и повысить эффективность профилактики.

В связи с этим всё большую значимость приобретает внедрение методов машинного обучения в систему мониторинга преступности. Эти методы расширяют аналитические возможности, делают выводы более обоснованными и позволяют быстрее реагировать на современные вызовы и угрозы.

## **1.2. Характеристика и источники данных о преступности**

Мониторинг преступности занимает важное место в системе криминологического анализа и фактически является одним из основных инструментов обеспечения общественной безопасности. В самом общем виде под мониторингом преступности понимается регулярное и целенаправленное наблюдение за состоянием, структурой и динамикой преступности, основанное на сборе, обработке и анализе различной информации. Его основная задача заключается не просто в фиксации статистических показателей, а в выявлении устойчивых тенденций и закономерностей развития криминогенной ситуации.

Мониторинг преступности в системе криминологического анализа включает в себя следующие процессы: сбор данных, анализ связей, построение моделей, обработку информации и принятие управленческих решений (рис. 1).



Рис. 1. Этапы криминологического анализа

В отличие от разовых статистических наблюдений, мониторинг преступности носит непрерывный характер и предполагает комплексный подход к анализу данных. Он ориентирован не только на учёт уже совершённых преступлений, но и на изучение факторов, оказывающих влияние на криминогенную обстановку, а также на оценку возможных изменений в будущем. Последовательность основных этапов данного процесса наглядно представлена на рис. 2.

В рамках криминологической науки мониторинг преступности рассматривается как составная часть более широкой системы криминологического анализа. Он опирается на данные официальной статистики,



Рис. 2. Основные этапы процесса мониторинга преступности

результаты социологических исследований, материалы оперативно-служебной деятельности и другие источники информации. При этом ключевая цель мониторинга заключается не в накоплении массивов данных, а в их аналитической интерпретации, позволяющей сформировать объективное представление о реальном состоянии преступности и факторах, определяющих её уровень.

Современные социально-экономические процессы существенно усложняют характер преступности. Она становится более неоднородной по структуре, различается по территории и всё сильнее зависит от совокупности внешних факторов. Существенное влияние оказывают процессы миграции населения, экономической нестабильности, а также развитие цифровых технологий. В таких условиях возрастает значение пространственного анализа, позволяющего учитывать территориальные особенности преступности.

В пространственном анализе применяется интерактивная карта, показывающая уровень преступности по районам за 2015–2025 годы.

Цветовая шкала карты, отражает интенсивность уровня преступности в каждой районе, что позволяет наглядно сравнивать районы с низким, средним и высоким уровнем преступности. Согласно цветовой градации красный цвет имеет значение от 4 до 5, оранжевый – от 3 до 4, желтый – от 2,5 до 3, Зеленый – от 2 до 2,5.

Карта построена на основе нормированных данных, что обеспечивает корректное сопоставление районов различной численности населения.

Анализ пространственного распределения преступлений позволяет выявлять территории с повышенной криминальной активностью, оценивать эффективность принимаемых мер и корректировать направления профилактической работы. На основе это анализа мониторинг становится инструментом территориального анализа и планирования.

Эффективность мониторинга преступности во многом определяется методами, используемыми для анализа данных. Традиционно в криминологии применяются методы описательной статистики, сравнительного анализа и корреляционных исследований. Эти методы позволяют получать обобщённые оценки и выявлять базовые зависимости, однако они мало эффективны при анализе сложных нелинейных взаимосвязей и работе с большими объёмами разнородных данных.

В условиях цифровизации существенно расширяется объём информации, используемой при мониторинге преступности. Помимо официальной статистики всё чаще привлекаются социально-экономические показатели. Это требует применения более гибких аналитических инструментов и приводит к постепенному переходу от исключительно опи-

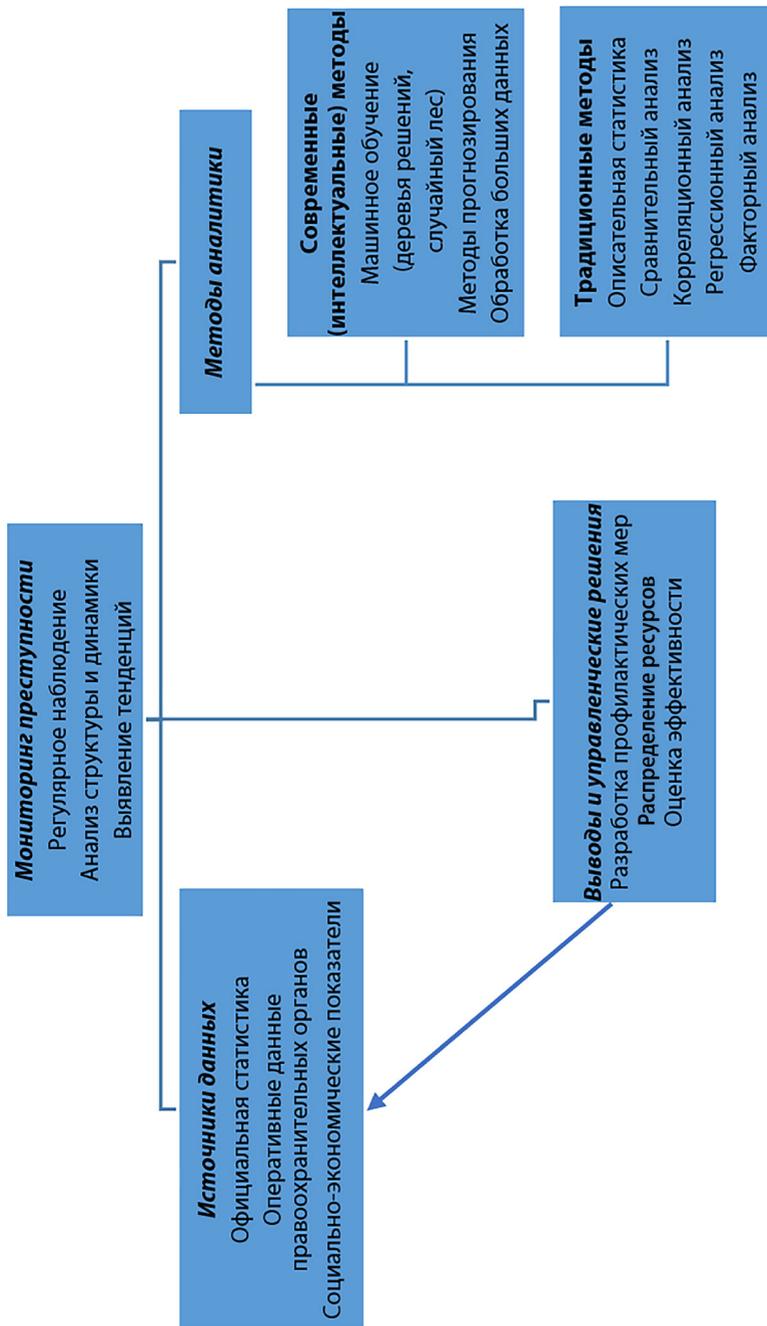


Рис. 3. Интеграция аналитических методов в систему мониторинга преступности

сательного анализа к использованию более сложных методов обработки данных, включая методы искусственного интеллекта. Общая схема интеграции аналитических методов в систему мониторинга преступности представлена на рис. 3.

В системе криминологических исследований мониторинг преступности выполняет как научные, так и прикладные функции. Он позволяет отслеживать долгосрочные тенденции и структурные изменения преступности, а также обеспечивает органы управления аналитической информацией, необходимой для разработки и корректировки мер профилактики.

Таким образом, мониторинг преступности представляет собой многоуровневый и динамично развивающийся процесс, эффективность которого напрямую зависит от выбранных методов анализа. На первом этапе развития криминологических исследований ключевую роль играли традиционные статистические подходы, позволявшие формировать общее представление о состоянии преступности и её динамике. Несмотря на определённые ограничения, именно они заложили основу для дальнейшего развития аналитических методов в криминологии.

### **1.3. Традиционные методы анализа криминальной статистики**

Традиционные методы анализа криминальной статистики на протяжении длительного времени составляли основу эмпирических криминологических исследований. В трудах С. А. Айвазяна и В. Е. Гмурмана подробно изложены методы описательной статистики, корреляционного и регрессионного анализа, широко применяемые для изучения социально-экономических процессов, включая анализ преступности [1; 2; 15].

В работах М. Кендалла и А. Стюарта, а также Б. Дюрана и П. Оделла показаны возможности многомерного и кластерного анализа при исследовании сложных социальных явлений, однако подчёркивается необходимость осторожной интерпретации результатов при наличии нелинейных зависимостей и мультиколлинеарности признаков [20; 22].

Ряд исследователей указывает, что классические статистические модели, ориентированные преимущественно на линейные связи, обладают ограниченной прогностической способностью и слабо приспособлены для анализа больших и разнородных массивов данных, характерных для современной криминологической статистики [19; 32].

Анализ криминальной статистики на протяжении длительного времени основывался на использовании традиционных методов математической и статистической обработки данных. Эти методы сформировали методологическую основу криминологических исследований и до настоящего времени широко применяются в практике анализа преступности. Их основная задача заключается в обобщении статистической информации, выявлении тенденций и сравнительном анализе показателей за различные периоды времени или по различным территориям.

К числу наиболее распространённых относятся методы описательной статистики, включающие расчёт абсолютных и относительных показателей, средних величин, индексов и коэффициентов.

С их помощью оцениваются уровень преступности, её структура и динамика, а также изменения отдельных категорий преступлений. Данные методы отличаются простотой применения и наглядностью результатов, что делает их удобными для подготовки аналитических отчётов и статистических обзоров.

На рис. 4. представлен уровень преступности объекта № 3 с 2013 по 2022 г.: с 2013 по 2016 г. – достаточно высокий, в 2017 г. произошел достаточно резкий спад, а далее наблюдается тенденция роста, с 2019 по 2021 г. – небольшой спад, однако в 2022 г. уровень преступности значительно увеличился и стал выше, чем в 2013 г.

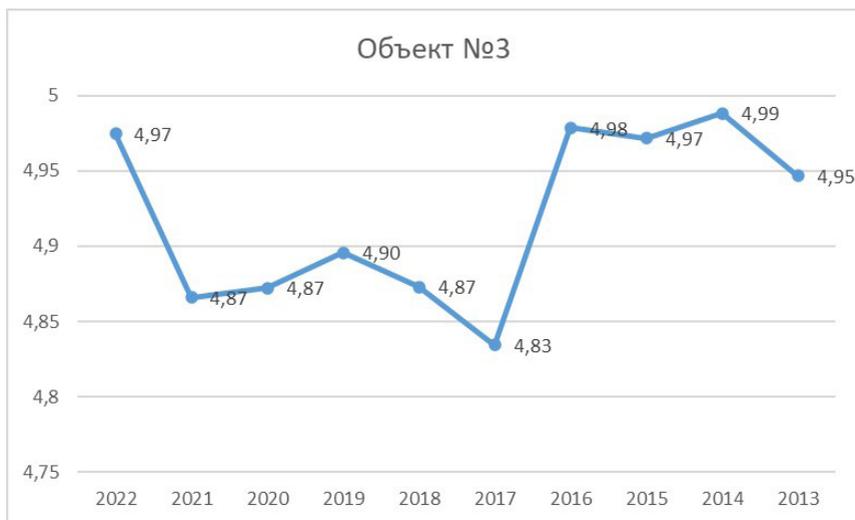


Рис. 4. Уровень преступности объекта № 3 в период с 2013 по 2022 г.

Широкое распространение в криминологических исследованиях получили методы сравнительного анализа. Они позволяют сопоставлять показатели преступности между районами, временными периодами. Такой подход используется для выявления территориальных различий, оценки эффективности профилактических мер и анализа влияния отдельных факторов на уровень преступности.

Примером может послужить сравнение коэффициентов, влияющих на уровень преступности. В объекте № 2 по сравнению с объектом № 1 коэффициент, отражающий количество штатных единиц, отрицательный, что свидетельствует о более рациональном планировании штатных единиц (рис. 5, 6).

Сравнение коэффициентов объектов № 3 и № 4 представлено на рис. 7. В объекте № 3 количество сотрудников, их средний возраст оказы-

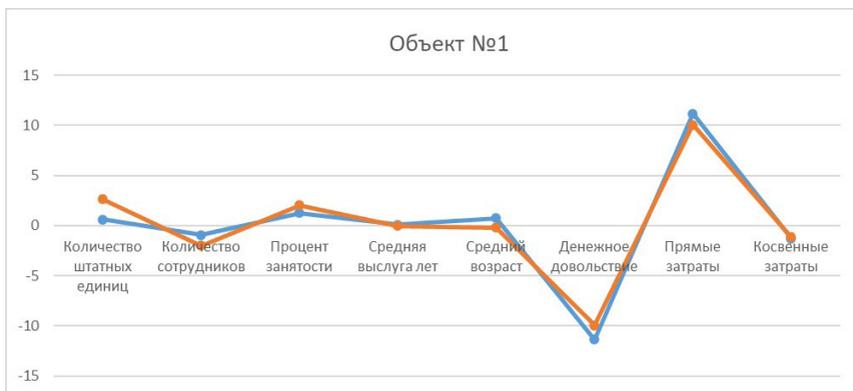


Рис. 5. Коэффициенты, влияющие на уровень преступности в объекте № 1



Рис. 6. Коэффициенты, влияющие на уровень преступности в объекте № 2

вают отрицательное влияние, количество штатных единиц, процент занятости, средняя выслуга лет – положительное влияние.

В объекте № 4 наблюдается слабое влияние большинства коэффициентов. Количество штатных единиц, процент занятости – слабо выраженное положительное влияние, количество сотрудников – слабо выраженное отрицательное влияние. Средний возраст, средняя выслуга лет – почти нейтральное влияние (рис. 7).

Однако сравнительный анализ, как правило, ограничивается поверхностными выводами и не позволяет выявлять сложные причинно-следственные связи.

Важное место в нашем исследовании занимают методы корреляционного и регрессионного анализа. Корреляционный анализ применяется для оценки степени взаимосвязи между уровнем преступности и социально-экономическими показателями. Регрессионные модели, в свою очередь, используются для количественной оценки влияния отдельных факторов на показатели преступности.

В работе особое внимание уделено было корреляционному анализу, который позволил определить, какие показатели связаны между собой и какие факторы могут влиять на уровень преступности. На рис. 8. изображена тепловая карта корреляций за 2015 год. Тепловая карта наглядно показывает, между какими показателями существует сильная положи-

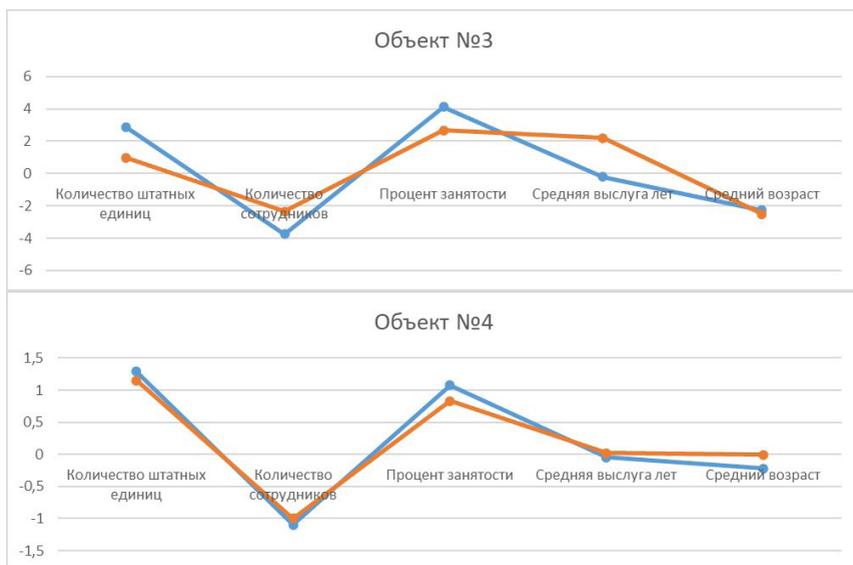


Рис. 7. Коэффициенты, влияющие на уровень преступности в объекте № 3 и объекте № 4 (синяя линия с 2013 по 2017 г., оранжевая с 2018 по 2022 г.)

тельная или отрицательная связь. Этот метод помогает выделить факторы, на которые стоит обратить внимание при построении прогнозов. Визуализация тепловой карты стала необходимой в исследовании для быстрого выявления зависимостей в больших массивах данных.

В большинстве случаев регрессионные модели предполагают линейный характер зависимостей между переменными, что не всегда соответствует реальным криминологическим процессам. Кроме того, такие модели чувствительны к выбросам и мультиколлинеарности признаков, а также требуют предварительного отбора факторов, что может приводить к потере значимой информации.

Отдельной проблемой традиционных методов является их ограниченная применимость к большим и разнородным массивам данных. По мере увеличения объёма информации и числа учитываемых признаков эффективность классических статистических подходов снижается. Анализ становится трудоёмким, а интерпретация результатов – менее одно-

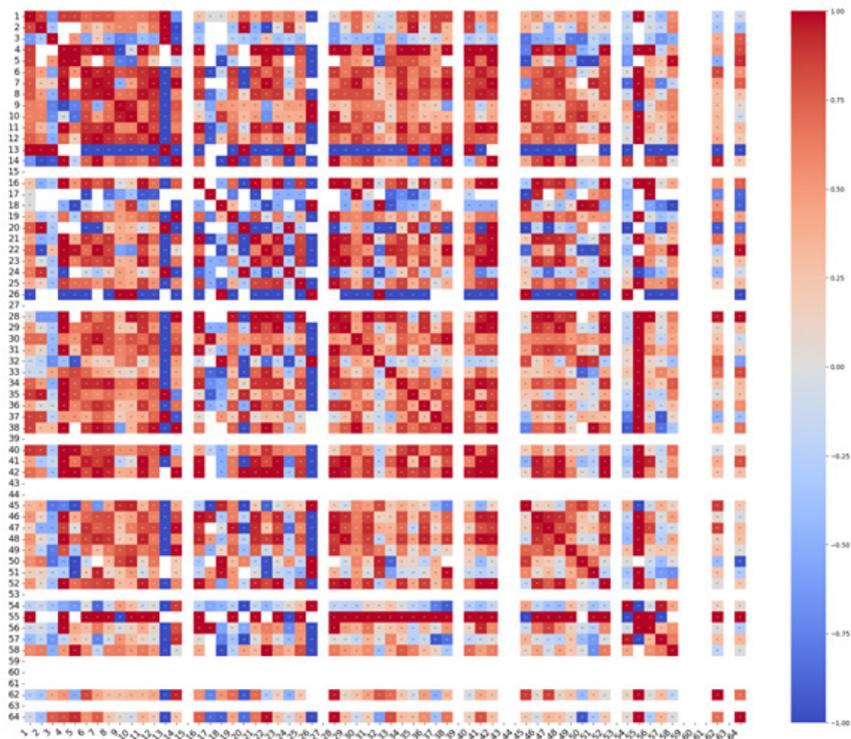


Рис. 8. Тепловая карта корреляций за 2015 г.

значной. Это особенно было заметно при работе с пространственно-временными данными и сложными многомерными структурами.

Следует также отметить, что традиционные методы анализа криминальной статистики в основном ориентированы на ретроспективное описание уже произошедших событий. Возможности прогнозирования в рамках таких подходов ограничены и, как правило, сводятся к экстраполяции выявленных тенденций. В условиях нестабильной социально-экономической обстановки и быстрого изменения характера преступности подобные прогнозы часто оказываются недостаточно точными.

Таким образом, традиционные методы анализа криминальной статистики сохраняют своё значение как базовый инструмент криминологических исследований, обеспечивающий общее представление об уровне и динамике преступности.

В связи с этим возникает необходимость перехода к более гибким и адаптивным методам анализа, способным эффективно работать с большими массивами разнородных данных и выявлять сложные закономерности. Именно такими возможностями обладают методы машинного обучения.

## **ГЛАВА 2. ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В АНАЛИЗЕ УРОВНЯ ПРЕСТУПНОСТИ**

В современных научных исследованиях всё более широкое распространение получают методы машинного обучения, ориентированные на анализ сложных и многомерных данных. В работах Ф. Босче, У. Маккинни и Дж. Вандер Пласа подробно рассматриваются алгоритмы классификации и регрессии, а также их применение для анализа социально-экономических процессов [7; 28; 31].

Возможности использования методов искусственного интеллекта и машинного обучения в непромышленной сфере, включая анализ социально значимых процессов, освещены в исследованиях В. Н. Босюка и соавторов. Авторы подчёркивают потенциал интеллектуальных моделей при работе с разнородными данными и их преимущество по сравнению с традиционными статистическими подходами [8; 29].

При этом в ряде работ отмечается, что применение методов машинного обучения в криминологических исследованиях требует особого внимания к интерпретируемости моделей и корректности выводов, поскольку результаты анализа используются при принятии управленческих решений в сфере общественной безопасности [11; 12].

### **2.1. Машинное обучение в анализе преступности**

Развитие машинного обучения радикально изменило подход к анализу данных. В отличие от традиционных статистических методов, здесь не нужно заранее задавать форму зависимостей между переменными – алгоритмы сами находят закономерности в данных. Это особенно полезно в сферах, где происходят сложные и многомерные процессы, например, при изучении преступности.

Машинное обучение можно рассматривать как набор практических методов, с помощью которых на основе реальных данных строятся модели для анализа, классификации и прогнозирования. В криминологических исследованиях эти методы не заменяют классические подходы, а дополняют их, позволяя по-новому взглянуть на процессы, связанные с преступностью.

Сложность анализа преступности связана с тем, что на её уровень и структуру влияет сразу несколько групп факторов. К ним относятся социально-экономические условия, демографические характеристики населения, территориальные особенности и временные изменения. Использование традиционных моделей часто оказывается недостаточным, поскольку они плохо работают при большом количестве взаимосвязанных показателей. Методы машинного обучения, напротив, позволяют учитывать одновременно множество признаков и их сочетаний.

В рамках нашего исследования были проанализированы данные по районам за 10-летний период. Для решения поставленных задач применялись методы классификации, что позволило распределить преступления по основным типам и оценить уровень преступности в каждом районе. Применение машинного обучения дало возможность учесть нелинейный характер влияния факторов и их сложные взаимосвязи, которые не выявляются при использовании простых методов экстраполяции.

Существенным преимуществом таких моделей является возможность их регулярного обновления. По мере накопления новых данных модели могут переобучаться, что позволяет своевременно отражать изменения криминогенной ситуации. Это особенно актуально в условиях нестабильной социальной и экономической обстановки, когда ранее выявленные зависимости быстро теряют свою актуальность.

Вместе с тем применение машинного обучения требует учета специфики криминологических данных. На практике такие данные часто оказываются несбалансированными, содержат пропуски и ошибки, а также большое количество категориальных показателей. В связи с этим особое значение приобретает этап предварительной обработки данных и обоснованный выбор используемых алгоритмов.

Отдельного внимания заслуживает интерпретация полученных результатов. Для криминологических исследований важно не только получить формально точный результат, но и понимать причины, лежащие в основе выводов модели. Поэтому в работе предпочтение отдавалось интерпретируемым методам, в частности деревьям решений и ансамблевым моделям.

В целом использование методов машинного обучения расширяет возможности анализа и прогнозирования преступности, повышает точность мониторинга и позволяет быстрее реагировать на изменения криминогенной обстановки. При этом их применение требует аккуратной настройки моделей и внимательного анализа результатов с учетом особенностей используемых данных.

## 2.2. Деревья решений как интерпретируемые модели анализа преступности

Бинарные деревья решений реализуют алгоритмический процесс последовательного разбиения пространства признаков на области с однородными значениями целевой переменной. Каждое разбиение основывается на выборе такого показателя и порогового значения, которые обеспечивают максимальное различие между группами наблюдений по уровню преступности.

Этот процесс формирует древовидную модель, которая показывает иерархию факторов и их сочетаний.

Одно из ключевых методологических преимуществ бинарных деревьев решений – возможность находить пороговые значения факторов. Для анализа преступности это особенно важно: можно определить критические уровни социально-экономических показателей, после которых криминогенная обстановка меняется резко и заметно.

Работа дерева регрессии основана на последовательном разбиении пространства признаков на более однородные области. На каждом шаге алгоритм выбирает такой признак и пороговое значение, которые минимизируют ошибку прогнозирования внутри полученных подмножеств.

В качестве критерия разбиения чаще всего используется:

- минимизация дисперсии целевой переменной;
- минимизация среднеквадратичной ошибки (MSE).

Таким образом, каждое последующее разбиение повышает однородность наблюдений внутри узлов, что позволяет получать более точные прогнозы уровня преступности.

При исследовании уровня преступности дерево регрессии позволяет учитывать одновременно множество факторов различной природы.

Дерево регрессии автоматически определяет наиболее значимые факторы и их пороговые значения, влияющие на уровень преступности.

Выявление таких порогов имеет практическую ценность, поскольку помогает разрабатывать более точные превентивные меры и обоснованные управленческие решения.

Структура дерева также позволяет оценить относительную значимость факторов по их месту в иерархии разветвлений. Факторы верхних уровней, как правило, отражают базовые детерминанты преступности и задают общее направление криминогенных процессов. Факторы нижних уровней уточняют, как именно эти базовые причины проявляются в конкретных социально-экономических условиях.

Такой подход даёт наглядную картину сложных причинно-следственных связей и делает результаты анализа более понятными для практического применения.

Дерево регрессии (рис. 9) – это метод машинного обучения из класса деревьев решений, который используется для прогнозирования непрерывной числовой величины. В исследованиях преступности такой величиной может быть, например, уровень преступности.

Модель дерева регрессии имеет иерархическую структуру из узлов и ветвей: каждый внутренний узел задаёт проверку условия по одной из объясняющих переменных, а конечные узлы (листья) содержат прогнозное числовое значение.

Применение дерева регрессии в анализе преступности обладает рядом преимуществ:

- 1) интерпретируемость – модель легко визуализируется и позволяет выявить логические зависимости между факторами;
- 2) отсутствие строгих предположений о распределении данных;
- 3) возможность работы с нелинейными зависимостями;
- 4) устойчивость к взаимодействию факторов без явного задания формы модели.

При этом у метода есть и ограничения, которые важно учитывать на практике.

Во-первых, дерево может переобучаться, особенно если его структура становится слишком глубокой.

Во-вторых, модель чувствительна к выбросам, из-за чего отдельные аномальные значения могут заметно влиять на результат.

В-третьих, структура дерева может меняться даже при небольших изменениях исходных данных, что снижает его устойчивость.

Особенно хорошо деревья решений работают в сочетании с другими методами машинного обучения. Например, ансамблевые подходы, такие как случайный лес, объединяют сразу много деревьев. За счёт этого обычно удаётся повысить точность прогноза и сделать модель более стабильной. При этом сохраняется интерпретируемость: исследователь по-прежнему может понять логику отдельных деревьев, но получает более надёжный итоговый результат.

Таким образом, дерево регрессии – это наглядный и достаточно эффективный инструмент анализа уровня преступности. Он позволяет учитывать сложные связи между социальными, экономическими и демографическими факторами и строить более обоснованные прогнозы развития криминогенной ситуации.

В целом деревья решений остаются удобным инструментом криминологического анализа. Они помогают не только классифицировать объекты и прогнозировать события, но и объяснять, почему модель пришла именно к таким выводам. Для сферы профилактики преступности это особенно важно, поскольку управленческие решения должны быть не

только точными, но и понятными. Простота, наглядность и способность работать с разными типами данных делают деревья решений логичным переходом от классической статистики к более гибким и адаптивным методам машинного обучения.

Использование деревьев в работе позволило выявить наиболее значимые статьи расходов, влияющие на уровень преступности по разным подразделениям.

#### Подразделение № 1

Статьи, оказывающие наибольшее влияние: «Оплата работ и услуг, переданных на аутсорсинг», «Оплата текущего ремонта оборудования и инвентаря». Также отмечается влияние статей «Медикаменты» и «Стипендии».

#### Подразделение № 2

Статьи, оказывающие наибольшее влияние: «Оплата текущего ремонта оборудования и инвентаря», «Прочие специальные расходы». Также неким образом влияют «Медикаменты», «Оплата услуг связи», «Оплата топлива», «Стипендии».

#### Подразделение № 3

Статьи, оказывающие наибольшее влияние: «Оплата текущего ремонта оборудования и инвентаря», «Прочие трансферты на продукцию и услуги», «Переподготовка кадров», «Информационно-вычислительные работы». Также неким образом влияют «Медикаменты», «Книги и периодические издания», «Стипендии», «Транспортные услуги».

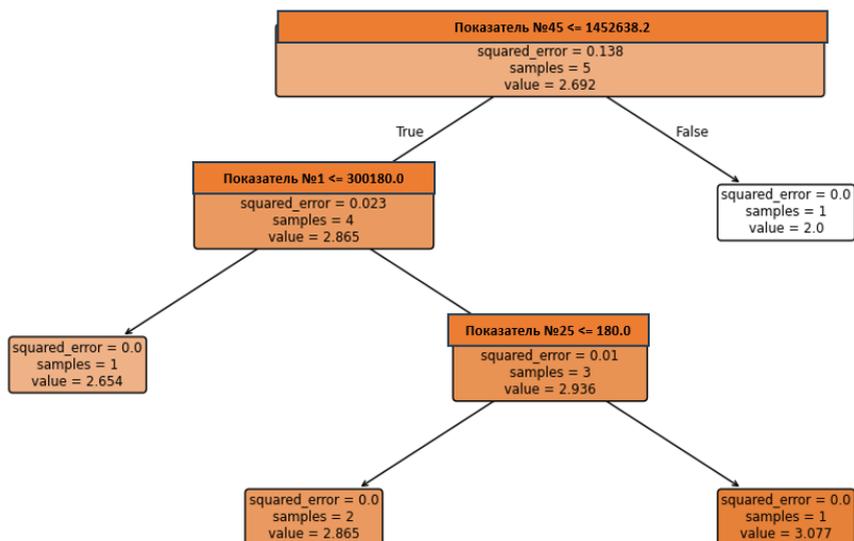


Рис. 9. Дерево регрессии

Анализируя влияние корреляционного анализа, можно прийти к следующим выводам:

1. После корреляционного анализа сдвиг по финансированию позволяет уточнить приоритеты расходов: статьи, которые считались незначительными до анализа, стали ключевыми после коррекции. Например, «Оплата текущего ремонта оборудования и инвентаря» повторяется как важный фактор, что говорит о его устойчивом влиянии.

2. Корреляционный анализ позволяет исключить «шум» и выявить реальные причинно-следственные связи.

Таким образом метод деревьев позволяет определить приоритет статей расходов, на которые следует обратить внимание при планировании бюджета и мер по снижению уровня преступности. Это помогает ведомствам распределять средства более эффективно и целенаправленно. Результаты показывают стабильность выявленных факторов: одни и те же статьи расходов повторяются в нескольких методах анализа, что подтверждает их значимость и устойчивость влияния на уровень преступности.

### **2.3. Ансамблевые методы и случайный лес в криминологических исследованиях**

Среди методов машинного обучения особое место занимают деревья решений и ансамблевые алгоритмы. Их методологические основы и практические возможности подробно рассматриваются в работах, посвящённых анализу данных и построению интерпретируемых моделей [7; 21].

В исследованиях, ориентированных на социально-экономические и криминологические данные, подчёркивается, что деревья решений позволяют выявлять пороговые значения факторов и наглядно представлять сложные причинно-следственные связи [11].

Ансамблевые методы, в частности алгоритм случайного леса, рассматриваются как эффективный инструмент повышения устойчивости и точности моделей за счёт объединения множества деревьев решений. Отмечается, что данные методы хорошо адаптированы к анализу зашумлённых и неполных данных, характерных для криминологической статистики [12].

В последние годы методы ансамблевого машинного обучения приобрели особую значимость в аналитике данных, включая криминологию. Основная идея ансамблевых методов заключается в объединении нескольких моделей с целью повышения точности и устойчивости прогнозных оценок. Совместное использование ряда относительно простых мо-

делей позволяет компенсировать их индивидуальные ограничения и получить более надёжный результат по сравнению с применением каждой модели по отдельности. Данный подход особенно актуален при анализе преступности, поскольку криминологические данные характеризуются высокой размерностью, сложной структурой и наличием пропусков.

Одним из наиболее распространённых ансамблевых алгоритмов является метод случайного леса (Random Forest). Он основан на построении множества деревьев решений, каждое из которых обучается на случайной подвыборке исходных данных с использованием случайного подмножества признаков. Итоговое решение формируется путём агрегирования результатов отдельных деревьев на основе принципа большинства голосов. Такая схема позволяет снизить эффект переобучения и повысить устойчивость модели к шуму и искажениям в данных, что имеет принципиальное значение для анализа криминологической статистики.

Случайный лес обладает рядом свойств, определяющих его практическую применимость в криминологических исследованиях.

Во-первых, этот алгоритм позволяет одновременно учитывать большое количество разнородных факторов: социально-экономические показатели, демографические характеристики, территориальные особенности и временные параметры.

Во-вторых, модель оценивает относительную важность каждого признака. Благодаря этому можно определить, какие факторы сильнее всего влияют на формирование криминогенной обстановки.

Кроме того, случайный лес хорошо выявляет сложные нелинейные связи и взаимодействия между признаками. В отличие от классических статистических моделей, где обычно предполагаются линейные зависимости и независимость факторов, этот подход учитывает их совместное действие. В результате динамика преступности описывается точнее и ближе к реальным условиям изменяющейся социальной и экономической среды.

Дополнительным преимуществом ансамблевых методов является их устойчивость к зашумлённым данным. Для криминологических исследований это особенно важно, поскольку статистическая информация о преступности может содержать ошибки учёта, пропущенные наблюдения и неоднородность классификаций. Благодаря своей структуре случайный лес снижает влияние подобных факторов и обеспечивает более стабильные и воспроизводимые результаты анализа.

Тем не менее применение ансамблевых моделей связано с определёнными ограничениями. Главная проблема – уменьшение интерпретируемости по сравнению с отдельными деревьями решений. Множество деревьев создают сложную структуру, которая трудно воспринимается визуально и аналитически. Для решения этой проблемы используются методы оценки

важности признаков, визуализация отдельных деревьев и разбор наиболее значимых решений. Такие подходы позволяют понять, какие факторы действительно влияют на прогнозируемый результат, что критически важно при принятии управленческих решений в правоохранительных органах.

Случайный лес также удобен для интеграции с другими аналитическими инструментами. Результаты модели могут быть визуализированы на карте для выявления «горячих точек» преступности.

В конечном итоге ансамблевые методы и случайный лес являются логическим развитием идей, заложенных в деревьях решений. Они сохраняют структуру и интуитивную понятность отдельных деревьев, но одновременно дают более точные и устойчивые результаты, позволяют работать с большими и разнородными массивами данных, выявлять скрытые закономерности и учитывать сложные взаимосвязи факторов.

Метод случайного леса позволяет получить количественные показатели значимости факторов, отражающие их вклад в объяснение вариации уровня преступности. Такие оценки используются для ранжирования факторов и выявления устойчивых детерминант криминогенной ситуации.

В отличие от одиночных моделей, данные оценки обладают высокой степенью надёжности и воспроизводимости.

Применение этих методов в криминологических исследованиях открывает новые возможности для построения интеллектуальных систем мониторинга преступности. Они позволяют анализировать влияние различных факторов на уровень преступности, что создаёт основу для более обоснованных и эффективных управленческих решений.

Random Forest – это не одно большое дерево, а множество маленьких случайных деревьев, которые вместе дают более стабильный и точный прогноз.

#### *Подразделение № 1*

Статьи, оказывающие наибольшее влияние: «Оплата текущего ремонта оборудования и инвентаря», «Командировки», «Оплата работ и услуг, переданных на аутсорсинг», «Оплата услуг связи».

#### *Подразделение № 2*

Статьи, оказывающие наибольшее влияние: «Приобретение производственного оборудования», «Командировки внутри республики», «Командировки за пределы республики», «Прочие специальные расходы», Оплата работ и услуг, переданных на аутсорсинг, Оплата услуг связи.

#### *Подразделение № 3*

Статьи, оказывающие наибольшее влияние: «Оплата текущего ремонта оборудования и инвентаря», «Переподготовка кадров», «Расходы на содержание автотранспорта», «Информационно-вычислительные работы».

Рассмотрим общие тенденции (рис. 10).

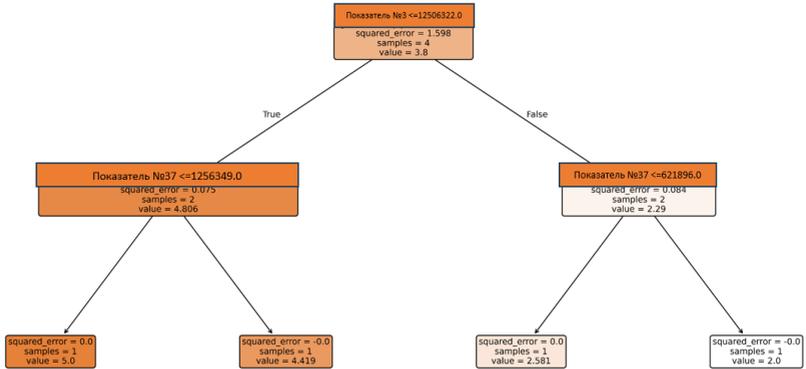


Рис. 10. Случайное дерево

Случайный лес позволяет выявить более широкий спектр значимых факторов по сравнению с простым деревом решений.

Статьи расходов, связанные с ремонтом оборудования, командировками, стипендиями и оплатой услуг, стабильно появляются как важные факторы в нескольких методах анализа.

Корреляционный анализ со сдвигом по финансированию выявляет дополнительные значимые статьи (например, приобретение производственного оборудования), которые не были очевидны без проведения корреляционного анализа. Это дает основания для целевого перераспределения бюджета и оптимизации расходов.

Методы случайного леса в совокупности с корреляционным анализом позволяют определить статьи расходов, оказывающие наибольшее влияние на эффективность работы подразделений и уровень преступности.

В этом и заключается практическая значимость метода в данном исследовании.

Использование случайного леса показало устойчивость выявленных факторов и уменьшило вероятность пропуска важных влияющих статей.

## 2.4. Интерпретация моделей с использованием SHAP Values в анализе уровня преступности

Вопросы интерпретации моделей машинного обучения в последние годы приобрели особую актуальность. В ряде исследований подчёркивается, что высокая точность прогнозов не является достаточным условием практического применения моделей без понимания механизма формирования результатов [21].

Метод SHAP, основанный на теории кооперативных игр, получил широкое распространение как универсальный инструмент объяснимого искусственного интеллекта. Он позволяет количественно оценивать вклад отдельных факторов в прогноз модели и использовать полученные результаты для анализа сложных социально-экономических процессов, включая уровень преступности [31].

Современные модели машинного обучения, такие как случайный лес и градиентный бустинг, демонстрируют высокую точность прогнозов уровня преступности. Однако для практического применения в правоохранительной деятельности крайне важно понимать вклад каждого фактора в прогнозируемый результат. Для этого исследователи используют метод SHAP (SHapley Additive exPlanations), который позволяет количественно оценить влияние каждого признака на предсказание модели, обеспечивая прозрачность и интерпретируемость.

Метод SHAP основан на концепции ценности Шепли из теории кооперативных игр. Прогноз модели рассматривается как сумма вкладов отдельных признаков, где каждый признак получает своё значение SHAP, показывающее, насколько он увеличивает или уменьшает прогнозируемый уровень преступности относительно среднего прогноза.

Применение SHAP в криминологических исследованиях включает несколько этапов:

1. Подготовка данных и обучение модели.

Данные о преступности объединяются с социально-экономическими, демографическими и территориальными признаками. Обучается машинная модель (случайный лес, градиентный бустинг), которая способна прогнозировать уровень преступности в каждом районе.

2. Расчёт SHAP Values.

Для каждого района рассчитываются значения SHAP для всех признаков. Положительные значения показывают факторы, увеличивающие вероятность преступлений, отрицательные – факторы, снижающие риск.

3. Визуализация влияния признаков.

Summary plot: отображает важность всех признаков и распределение их влияния на прогноз по всем районам. Например, видно, что высокий уровень безработицы оказывает наибольшее положительное влияние на прогнозируемый уровень преступности.

Dependence plot: показывает зависимость значения SHAP от признака и взаимодействие с другими признаками. Например, рост плотности населения усиливает влияние безработицы на прогнозируемый уровень краж.

Force plot: визуализирует вклад каждого признака для конкретного района. Например, для района с высокой преступностью force plot показывает, что причинами являются высокие безработица и плотность населения.

Карта SHAP: пространственная визуализация вкладов признаков по районам города, где цветовые шкалы показывают зоны повышенного риска, что позволяет выявлять «горячие точки» преступности.

В рамках работы была рассмотрена SHAP-диаграмма важности признаков (рис.11.), которая демонстрирует, какой вклад каждый показатель вносит в предсказание модели.

Ось X показывает среднее значение  $|\text{SHAP value}|$  для каждого признака, т. е. средний вклад данного признака в предсказание модели. Чем больше значение, тем сильнее влияние признака на итоговый прогноз.

Ось Y показывает названия признаков.

Признаки расположены сверху вниз по убыванию влияния: сверху – самые значимые, снизу – менее важные.

При построении диаграммы также использовались lag-показатели.

Lag означает «запаздывание» – это значение признака из прошлого периода времени.

Например:

- показатель\_2\_lag1 – значение показателя 2 за предыдущий год ( $t-1$ ).
- показатель\_12\_lag2 – значение показателя 12 два года назад ( $t-2$ ).

Использование лагов позволяет модели учитывать временные зависимости: прошлые показатели могут влиять на текущий уровень преступности.

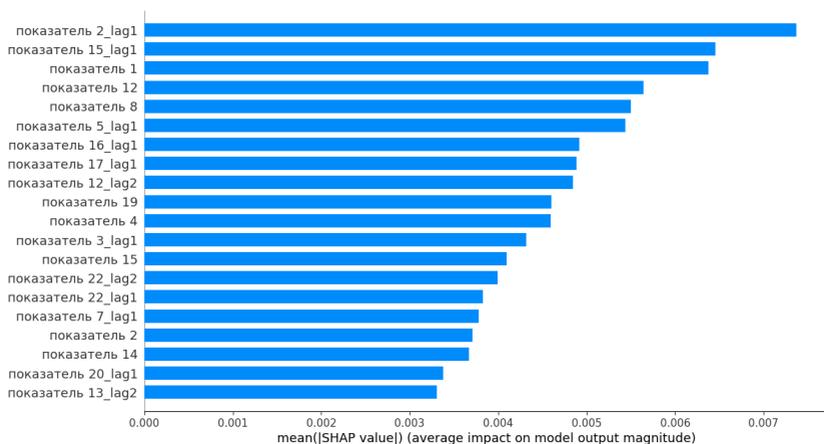


Рис. 11. SHAP-диаграмма важности признаков

SHAP summary plot (рис.12.) – один из самых информативных графиков для интерпретации модели. Признаки отсортированы сверху вниз по убыванию важности. Самый верхний (показатель 2\_lag1) – наиболее влияющий, самый нижний – наименее влияющий.

Для большинства признаков красные точки находятся справа, синие – слева.

Это значит:

чем больше значение показателя – тем выше прогноз.

Но есть и признаки, где красные точки встречаются и слева, и справа, что означает немонотонное влияние (возможны пороги, нелинейности).

Диапазон SHAP: примерно от  $-0.011$  до  $+0.008$ .

Это говорит нам о том, что модель достаточно стабильная, нет признаков, которые «ломают» прогноз.

Проведённый SHAP-анализ (рис. 13, 14) позволил оценить относительный вклад факторов в формирование прогнозируемого уровня преступности и выявить структуру их влияния. Полученные результаты свидетельствуют об отсутствии факторов с выражено доминирующим вкладом: ни один из показателей не продемонстрировал критически высокого значения относительной значимости. Это указывает на многофакторный характер исследуемого процесса и отсутствие единственного определяющего элемента уровня преступности.

Наибольшее относительное влияние среди всех рассмотренных переменных оказывают показатели со средним уровнем значимости, в част-

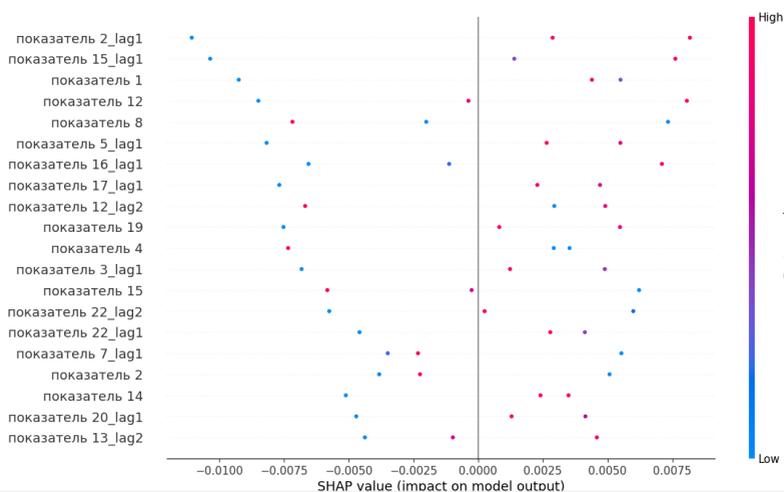


Рис. 12. SHAP summary plot

### Направление влияния факторов на уровень преступности

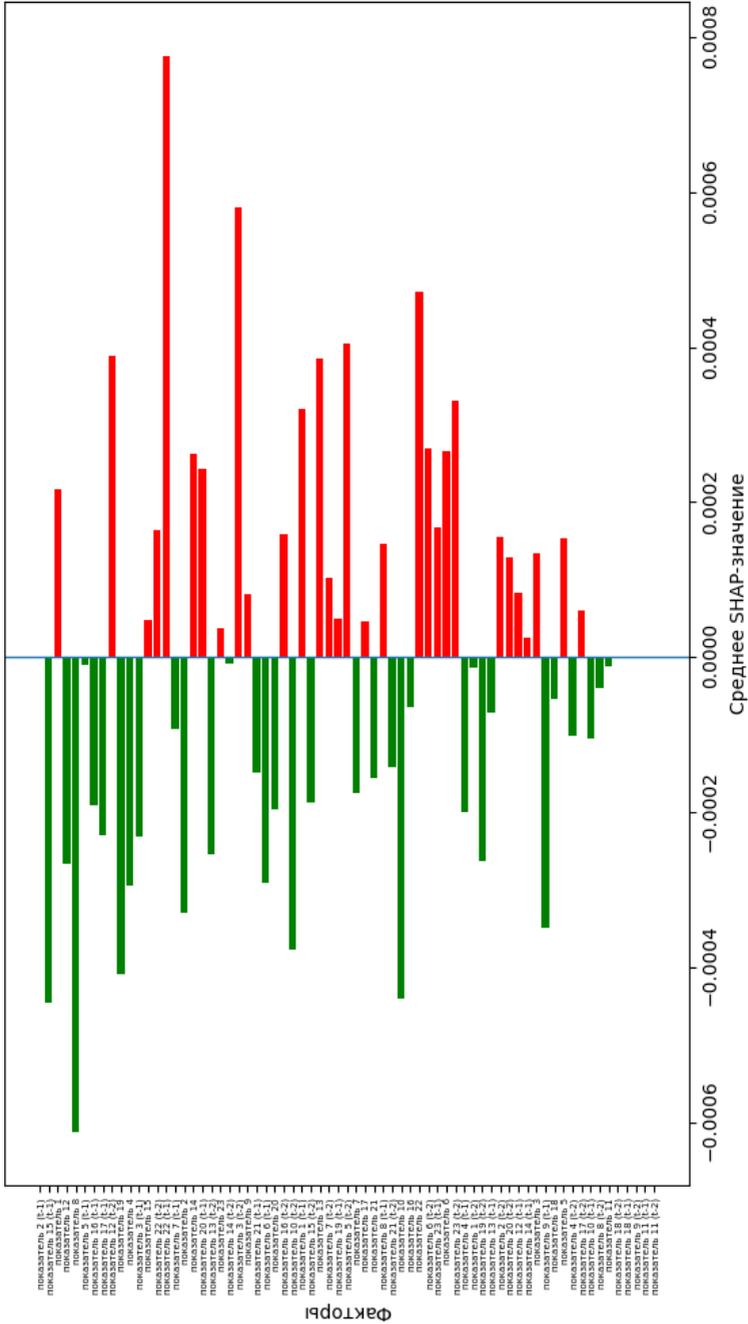


Рис. 13. Направление влияния факторов на уровень преступности



ности показатель 2 с лагом первого порядка ( $t-1$ ), показатель 15 с лагом первого порядка ( $t-1$ ), а также показатель 1 без лага. Их вклад в совокупное формирование прогноза превышает вклад остальных факторов, что позволяет рассматривать данные показатели как ключевые в рамках используемой модели, несмотря на умеренный характер их индивидуального воздействия.

Значительная часть факторов относится к группе со слабым влиянием. Данные показатели вносят ограниченный вклад в изменение прогнозируемого уровня преступности. Вместе с тем их суммарное влияние не является пренебрежимо малым, что подтверждает кумулятивный характер воздействия второстепенных факторов и необходимость их учёта при построении прогностических моделей.

Отдельного внимания заслуживает роль лагированных переменных. Существенная представленность факторов с лагами первого и второго порядка указывает на выраженный отложенный характер влияния социально-экономических показателей на уровень преступности. Это свидетельствует о наличии временной инерции криминогенных процессов, при которой изменения условий проявляются в динамике преступности не мгновенно, а с определённым временным запаздыванием.

В целом результаты SHAP-анализа подтверждают, что уровень преступности формируется под воздействием совокупности факторов с преимущественно умеренным и слабым индивидуальным вкладом, а ключевую роль в модели играют показатели с лагированной структурой.

*Преимущества применения SHAP:*

- прозрачность модели. SHAP Values показывают влияние каждого признака, повышая доверие к прогнозам;
- идентификация ключевых факторов. Позволяет выделять социальные, экономические и территориальные характеристики, которые существенно влияют на уровень преступности;
- анализ локальных закономерностей. Помогает исследовать влияние признаков на конкретные районы или группы населения.
- Поддержка принятия решений. Данные SHAP могут быть использованы для оптимального распределения ресурсов, планирования патрулирования и разработки социальных программ профилактики.

*Ограничения:*

- высокие вычислительные ресурсы для больших ансамблей;
- SHAP показывает влияние признаков в рамках модели, а не причинно-следственные связи;

- требуется экспертная интерпретация для корректного применения в практических целях.

Метод SHAP обеспечивает глубокую интерпретацию моделей машинного обучения и позволяет анализировать уровень преступности с точки зрения влияния отдельных факторов. Он сочетает точность прогнозов и прозрачность моделей, что особенно важно для органов правопорядка. Использование SHAP в криминологических исследованиях позволяет выявлять критические зоны, планировать профилактические меры и обоснованно распределять ресурсы, создавая основу для интеллектуальных систем.

## ГЛАВА 3. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ

### 3.1. Программные средства и архитектура аналитической системы

Создание современной аналитической системы для мониторинга преступности невозможно без тщательно продуманной архитектуры и подбора соответствующих программных средств. Архитектура такой системы должна обеспечивать надёжный сбор, хранение, обработку и анализ данных, а также визуализацию и интерпретацию результатов. При этом основными требованиями к системе являются масштабируемость, гибкость и возможность интеграции с различными источниками информации.

Программные средства для построения аналитической платформы играют ключевую роль. В качестве основного языка программирования часто выбирается Python, благодаря его универсальности, богатой экосистеме библиотек и возможности работы с различными типами данных. Python поддерживает работу с таблицами, базами данных, геоинформационными системами, а также предоставляет мощные инструменты машинного обучения и визуализации.

В рамках программной реализации аналитической системы использовались современные библиотеки языка Python, обеспечивающие полный цикл обработки и анализа данных. Для работы с табличными данными и выполнения численных операций применялись библиотеки Pandas и NumPy. Обучение машинных моделей осуществлялось с использованием библиотеки scikit-learn. Визуализация результатов анализа и представление статистических зависимостей выполнялись с помощью библиотек Matplotlib и Seaborn.

Для обработки и анализа пространственно привязанных данных, а также их картографической визуализации, использовались библиотеки GeoPandas и Folium.

Архитектура аналитической системы разработана на основе модульного принципа, что обеспечивает логическое разделение функциональных компонентов и повышает надёжность программного решения. Модульная структура упрощает сопровождение системы, позволяет локализовать изменения и расширять функциональность без существенной переработки остальных компонентов.

В состав системы входят основные функциональные модули, каждый из которых ориентирован на решение определённого класса задач и взаимодействует с другими компонентами через формализованные интерфейсы.

*Модуль сбора данных* отвечает за интеграцию информации из различных источников. К ним относятся официальная статистика правоохранительных органов, результаты социологических исследований, данные открытых источников и геоинформационных систем. Важной задачей модуля является обеспечение целостности и актуальности информации, а также предварительная фильтрация и очистка данных от явных ошибок и пропусков.

*Модуль хранения данных* представляет собой базу данных или систему хранилища, способную обрабатывать большие объёмы информации. В современных аналитических платформах широко используются реляционные базы данных (PostgreSQL, MySQL) для структурированных данных.

*Модуль аналитики* основной блок, в котором реализуются алгоритмы машинного обучения, статистического анализа. Он включает разработку и обучение моделей, выявление значимых факторов, влияющих на уровень преступности. Особое внимание уделяется использованию интерпретируемых моделей, таких как деревья решений и ансамблевые методы, включая случайный лес.

*Модуль визуализации и отчётности* обеспечивает наглядное представление результатов анализа и построение карт, графиков, диаграмм и отчётных документов. В криминологических исследованиях визуализация пространственных данных позволяет выявлять «горячие точки» преступности.

Архитектура системы строится с учётом возможности масштабирования и интеграции с внешними источниками. Это позволяет оперативно добавлять новые виды данных, расширять функционал аналитических модулей и внедрять современные алгоритмы машинного обучения по мере их появления. Кроме того, модульная структура облегчает сопровождение и модернизацию системы без необходимости полной переработки архитектуры.

Важной составляющей архитектуры является взаимодействие между модулями. Модуль сбора данных обеспечивает непрерывное поступление информации, которая затем поступает в модуль хранения для структурирования и оптимизации доступа. Далее аналитический модуль обрабатывает данные, обучает модели, передаёт результаты в модуль визуализации. Такой поток данных обеспечивает непрерывность анализа, минимизирует задержки и повышает оперативность принимаемых решений.

В целом правильно спроектированная архитектура аналитической системы обеспечивает:

- надёжность и целостность данных;
- гибкость в расширении функционала;
- возможность интеграции с разнородными источниками информации;
- поддержку современных методов машинного обучения;
- наглядную визуализацию результатов для пользователей различного уровня – от аналитиков до управленцев.

Создание такой платформы позволяет перейти от традиционного анализа преступности к интеллектуальной системе мониторинга, способной выявлять закономерности, прогнозировать изменения криминогенной обстановки и поддерживать принятие обоснованных решений. Таким образом, архитектура и программные средства становятся фундаментом для интеграции теории, данных и методов машинного обучения в единое аналитическое решение.

### **3.2. Реализация и обучение моделей машинного обучения**

После проектирования архитектуры аналитической платформы следующим ключевым этапом является реализация и обучение моделей машинного обучения, которые составляют аналитическое ядро системы мониторинга преступности. Этот процесс включает подготовку данных, выбор и настройку алгоритмов, обучение моделей, а также оценку их качества и интерпретацию результатов.

Эффективность работы моделей напрямую зависит от качества входных данных. В криминологических исследованиях данные могут быть разнородными, неполными и зашумлёнными. Поэтому первоочередной задачей является их очистка и стандартизация. Процедуры подготовки данных включают:

1. Удаление или корректировку пропусков и ошибок. Например, отсутствующие сведения о месте совершения преступления могут быть дополнены через сопоставление с другими источниками или анкетными данными.

2. Кодирование категориальных признаков, таких как типы преступлений, категории нарушителей, территориальная принадлежность.

3. Нормализация и масштабирование числовых данных, чтобы различия в диапазоне значений признаков не оказывали чрезмерного влияния на обучение моделей.

4. Формирование обучающей и тестовой выборок, что позволяет оценивать качество модели на независимых данных и предотвращает переобучение.

#### *Выбор моделей и алгоритмов*

В рамках исследования были использованы различные методы машинного обучения, каждый из которых обладает специфическими преимуществами с точки зрения точности прогнозирования и интерпретируемости результатов:

1) деревья решений, обеспечивающие высокую интерпретируемость и наглядность модели за счёт иерархической структуры правил принятия решений, что позволяет проследивать влияние отдельных факторов на формирование прогноза;

2) случайный лес, повышающий точность и устойчивость прогнозов за счёт ансамблирования большого числа деревьев решений и снижения риска переобучения. Данный метод позволяет учитывать нелинейные зависимости и сложные взаимодействия между факторами, сохраняя при этом приемлемый уровень интерпретируемости;

3) метод SHAP (SHapley Additive exPlanations), применяемый в качестве инструмента интерпретации моделей машинного обучения. SHAP-анализ позволяет количественно оценить вклад каждого фактора в формирование прогнозов как на уровне отдельных наблюдений, так и в среднем по всей выборке. Использование данного подхода обеспечивает сопоставимость интерпретаций для различных моделей, включая ансамблевые методы, и позволяет выявить ключевые факторы и структуру их отложенного влияния во времени. Выбор алгоритма определяется как характеристиками данных, так и практическими требованиями: например, необходимость интерпретируемости модели для управленческих решений или способность работать с большим числом признаков.

#### *Обучение моделей*

Обучение моделей машинного обучения включает этап настройки параметров, задаваемых до начала обучения и оказывающих существенное влияние на качество, устойчивость и обобщающую способность модели. Данные параметры определяют структуру алгоритма и характер его работы с данными, в связи с чем их корректный выбор является необходимым условием получения надёжных прогнозов. В частности, для деревьев решений к таким параметрам относятся глубина дерева, минимальное количество наблюдений в листе и критерий разбиения, тогда как в случайном лесу существенную роль играют количество деревьев и число признаков, используемых при построении каждого дерева.

Для подбора оптимальных значений настроечных параметров используется процедура кросс-валидации, основанная на разбиении обу-

чающей выборки на несколько подвыборок. В рамках данной процедуры модель поочередно обучается на части данных и проверяется на отложенной выборке, что позволяет получить более устойчивую и объективную оценку её качества. Такой подход снижает риск переобучения, связанный с чрезмерной подгонкой модели под обучающие данные.

В процессе обучения особое внимание уделяется оценке качества построенных моделей. Выбор показателей качества осуществляется с учётом специфики решаемой задачи. При решении задач классификации используются метрики точности, полноты, precision и F1-мера, которые характеризуют способность модели корректно различать классы, в том числе выявлять редкие и слабо представленные события. Для задач регрессии применяются среднеквадратичная ошибка и коэффициент детерминации, позволяющие оценить точность прогнозирования и степень соответствия результатов моделирования фактическим наблюдениям.

Применение адекватных метрик обеспечивает комплексную оценку эффективности моделей и позволяет выявлять их ограничения, включая склонность к переобучению и недостаточную чувствительность к отдельным группам наблюдений.

#### *Интерпретация результатов*

После обучения моделей важным этапом является анализ полученных результатов и интерпретация. Для интерпретируемых моделей, таких как деревья решений, можно визуально отследить последовательность условий, приводящих к конкретному прогнозу. Для ансамблевых методов применяются методы оценки важности признаков: они позволяют определить, какие социально-экономические, демографические или пространственные характеристики оказывают наибольшее влияние на прогнозируемые события.

Интерпретация результатов играет ключевую роль при принятии управленческих решений.

Для практической реализации модели прогнозирования уровня преступности и последующей интерпретации факторов использовался язык программирования Python и специализированные библиотеки анализа данных и машинного обучения. Основной акцент был сделан на построении интерпретируемой регрессионной модели и выявлении как силы, так и направления влияния социально-экономических факторов во времени.

На первом этапе осуществлялась загрузка исходных данных из файла формата Excel с использованием библиотеки Pandas. Структура исходного файла предполагала, что в первом столбце содержатся наименования показателей, а в последующих – значения по годам. Такой формат является стандартным для задач временного анализа и обучения моделей машинного обучения.

В качестве модели машинного обучения использовался алгоритм CatBoostRegressor, который относится к классу градиентных ансамблевых методов. Выбор данного алгоритма обусловлен его высокой устойчивостью к переобучению, способностью эффективно работать с небольшими и средними выборками, а также корректно моделировать нелинейные зависимости между признаками. Модель обучалась с фиксированными параметрами глубины деревьев, скорости обучения и количества итераций, обеспечивающими баланс между точностью прогнозирования и обобщающей способностью.

Для повышения интерпретируемости результатов после обучения модели была проведена процедура SHAP-анализа с использованием библиотеки SHAP. Данный метод основан на теории кооперативных игр и позволяет количественно оценить вклад каждого признака в формирование прогнозируемого значения. SHAP-значения рассчитывались для всех наблюдений обучающей выборки, что обеспечивало возможность анализа как индивидуальных прогнозов, так и глобальной структуры влияния факторов.

Для удобства визуального и аналитического восприятия наименования факторов были приведены к сокращённому виду, при этом лаговые переменные явно маркировались как значения предыдущих периодов ( $t-1$ ,  $t-2$ ). Это позволило напрямую интерпретировать временную структуру влияния факторов.

Глобальное влияние факторов оценивалось с использованием агрегированных SHAP-значений, представленных в виде столбчатых диаграмм. Данный анализ позволил ранжировать признаки по степени их значимости и выделить ключевые факторы, оказывающие наибольшее влияние на уровень преступности. Дополнительно строились детализированные SHAP-графики рассеяния, отражающие распределение влияния каждого признака и характер его воздействия в зависимости от величины значения.

Отдельное внимание уделялось анализу индивидуальных наблюдений. Для выбранного года применялся waterfall-график, наглядно демонстрирующий, какие именно факторы и в какой степени повышали или снижали прогнозируемый уровень преступности относительно базового значения модели.

На основе рассчитанных SHAP-значений формировалась аналитическая таблица важности факторов, включающая среднее абсолютное значение SHAP, относительный вклад каждого признака в процентах и качественную классификацию силы влияния (сильное, среднее, слабое). Такая группировка позволила не только количественно, но и содержательно интерпретировать полученные результаты.

Дополнительно был проведён анализ направления влияния факторов. Для этого использовались средние SHAP-значения со знаком, что позволило разделить признаки на факторы, увеличение которых связано с ростом уровня преступности, и факторы, способствующие её снижению. Результаты данного анализа были визуализированы с помощью горизонтальных диаграмм, где цветовая кодировка отражала направление воздействия. Для повышения наглядности подписи факторов размещались непосредственно на графических элементах.

В целом представленный программный модуль реализует полный цикл анализа: от загрузки и трансформации данных до построения интерпретируемой модели и формирования содержательных выводов. Использование CatBoost в сочетании с SHAP-анализом позволяет не только достигать высокой точности прогнозирования, но и обеспечивать обоснованность и объяснимость результатов, что является важным при исследовании уровня преступности.

## ЗАКЛЮЧЕНИЕ

В ходе исследования была проведена комплексная оценка уровня преступности с использованием современных методов анализа данных. Основной целью работы являлась количественная и качественная оценка влияния факторов на уровень преступности.

Использование бинарных деревьев решений и деревьев регрессии позволило выявить нелинейные зависимости между факторами и уровнем преступности, определить значимость отдельных признаков и выявить критические пороговые значения, при которых наблюдаются существенные изменения криминогенной обстановки. Применение случайного леса обеспечило повышение точности прогнозирования и устойчивость модели к переобучению, что позволило получать более надёжные результаты при анализе больших и разнородных данных.

На этапе моделирования были рассмотрены и апробированы различные алгоритмы машинного обучения. В частности, использование деревьев решений позволило получить наглядную и логически интерпретируемую модель, отражающую иерархию факторов и последовательность условий, при которых формируется тот или иной уровень преступности. Данный подход оказался полезным для первичного анализа данных и выявления ключевых зависимостей, однако его применение в чистом виде ограничено высокой чувствительностью к структуре выборки и риском переобучения.

Для повышения устойчивости и точности прогнозов в рамках исследования был применён метод случайного леса, основанный на ансамблировании большого числа деревьев решений. Использование случайного леса позволило учесть нелинейные зависимости и сложные взаимодействия между факторами, а также снизить влияние случайных колебаний данных. Результаты моделирования показали, что данный алгоритм обеспечивает более стабильные оценки по сравнению с одиночными деревьями, сохраняя при этом приемлемый уровень интерпретируемости.

Дополнительно в работе была реализована модель на основе алгоритма CatBoost, продемонстрировавшая высокую эффективность при анализе временных рядов с ограниченным количеством наблюдений и наличием лаговых признаков. Включение временных лагов позволило выявить отложенное влияние финансирования и иных социально-экономических показателей на уровень преступности, что подтверждает инер-

ционный характер процессов формирования криминогенной обстановки.

Применение SHAP-анализа позволило количественно оценить вклад каждого фактора в формирование прогнозируемого уровня преступности как в среднем по выборке, так и для отдельных периодов времени. Кроме того, данный подход обеспечил возможность определения направления влияния факторов – повышающего или снижающего уровень преступности, что существенно расширяет аналитические и практические возможности модели.

Применение методов сравнительного и интегративного анализа позволило объединить разнородные показатели в единую аналитическую систему, провести оценку территориальных различий между районами и выявить общие закономерности пространственного распределения преступности. Использованный подход обеспечил не только количественную оценку уровня преступности, но и более глубокое понимание факторов, определяющих его изменение во времени и пространстве.

В ходе исследования были выделены основные статьи затрат, оказывающие наибольшее влияние на формирование криминогенной ситуации, что позволило уточнить роль финансовых и социально-экономических факторов в динамике преступности.

Проведённый анализ показывает, что сочетание количественных методов обработки данных с их содержательной интерпретацией позволяет получать обоснованные и практико-ориентированные выводы. Разработанная аналитическая система, в свою очередь, представляет собой удобный и наглядный инструмент для принятия решений в сфере мониторинга и анализа преступности.

В целом проведённое исследование подтвердило целесообразность использования ансамблевых методов машинного обучения в сочетании с инструментами объяснимого искусственного интеллекта для анализа и мониторинга преступности. Разработанная аналитическая модель может быть использована в практической деятельности аналитических и правоохранительных подразделений, а также служить основой для дальнейших исследований, направленных на расширение набора факторов и пространственный анализ.

## ЛИТЕРАТУРА

1. Айвазян С. А. Анализ временных рядов. – М.: Финансы и статистика, 2013. – 432 с.
2. Айвазян С. А. Методы эконометрики. – М.: Магистр; Инфра-М, 2016. – 512 с.
3. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности данных. – М.: Финансы и статистика, 2010. – 607 с.
4. Антонян Ю. М. Криминология. – М.: Юрайт, 2020. – 448 с.
5. Бейдер А. SQL и PostgreSQL для анализа данных. – СПб.: Питер, 2019. – 352 с.
6. Блувштейн Ю. Д. Статистика преступности. – М.: ЮНИТИ-ДАНА, 2012. – 231 с.
7. Босче Ф. Машинное обучение и анализ данных на Python. – М.: Диалектика, 2020. – 416 с.
8. Босюк В. Н., Бондарь В. В., Гощина Н. Н., Надькин Л. Ю., Салко Д. А. Криминогенность как объект математического анализа: модельно-прогностический подход. – СПб., 2025.
9. Босюк В. Н., Гощина Н. Н. Пакеты прикладных программ и их использование в экономических исследованиях // Цифровизация как фактор экономического развития: материалы междунар. науч.-практ. конф. – СПб., 2023. – С. 7–18.
10. Босюк В. Н., Надькин Л. Ю., Гощина Н. Н. Методы и примеры использования парной корреляции в эконометрическом моделировании // Теоретико-прикладные аспекты экономики и управления: материалы республ. науч.-практ. конф. обучающихся магистерских программ. – СПб., 2024. – С. 7–14.
11. Босюк В. Н., Саломатина Е. В., Нягу В. Н., Салко Д. А., Кирман С. Н., Гощина Н. Н. Мультифакторный корреляционный анализ оценки криминогенной ситуации в регионе // Экономика и предпринимательство. – 2023. – № 11 (160). – С. 591–595.
12. Босюк В. Н., Узун И. Н., Гощина Н. Н. Оценка факторов, влияющих на криминогенную ситуацию в регионе: методологический аспект // Экономика и предпринимательство. – 2023. – № 12 (161). – С. 358–363.

13. Босюк В. Н., Узун И. Н., Гощина Н. Н. Сравнительный анализ использования типовых программных продуктов при компьютерном моделировании хозяйственных ситуаций // Финансовый бизнес. – 2022. – № 12 (234). – С. 47–51.
14. Босюк В. Н., Узун И. Н., Гощина Н. Н. Язык программирования Python в прикладных исследованиях // Экономика и управление: теория, методология, практика: материалы XVIII Междунар. науч.-практ. конф. / отв. ред. Л. С. Валинурова, Н. А. Кузьминых. – Уфа, 2023. – С. 43–48.
15. Гмурман В. Е. Теория вероятностей и математическая статистика. – М.: Высшая школа, 2014. – 479 с.
16. Гощина Н. Н., Босюк В. Н. Статистические методы оценки влияния различных факторов на результаты работы субъектов непроеизводственной сферы // Экономика и предпринимательство. – 2022. – № 12 (149). – С. 686–689.
17. Гринберг М. Flask. Разработка веб-приложений на Python. – М.: ДМК Пресс, 2019. – 258 с.
18. Грубер М. Python. Создание веб-приложений с Flask. – СПб.: Питер, 2021. – 304 с.
19. Добрынин Н. М. Математические методы анализа преступности. – М.: Юридический центр Пресс, 2014. – 289 с.
20. Дюран Б., Оделл П. Кластерный анализ. – М.: Статистика, 1977. – 128 с.
21. Зандерс Э. Анализ данных и визуализация на Python. – М.: ДМК Пресс, 2021. – 368 с.
22. Кендалл М., Стюарт А. Многомерный статистический анализ. – М.: Наука, 1976. – 736 с.
23. Клэр С., Кунц М. PostgreSQL. Основы и практика. – СПб.: Питер, 2018. – 416 с.
24. Кочетов А. В. Веб-картография и геоинформационные системы. – М.: Академия, 2015. – 256 с.
25. Кочетов В. В. Геоинформационные системы в криминологии. – М.: Академия, 2016. – 272 с.
26. Кудрявцев В. Н. Причинность в криминологии. – М.: Юридическая литература, 1988. – 176 с.
27. Лунеев В. В. Преступность XX века: мировые, региональные и российские тенденции. – М.: Норма, 2005. – 912 с.
28. Маккинни У. Python и анализ данных. – М.: ДМК Пресс, 2019. – 544 с.

29. Нягу В. Н., Салко Д. А., Кирман С. Н., Босюк В. Н., Гощина Н. Н. Факторный анализ и искусственный интеллект в непроеизводственной сфере // Актуальные вопросы и проблемы экономики и управления: материалы Республ. науч.-практ. конф. обучающихся магистерских программ. – СПб., 2023. – С. 7–15.

30. Тейлор М. Прикладной анализ данных на Python. – СПб.: Питер, 2018. – 320 с.

31. Вандер Плас Дж. Python для анализа данных. – М.: ДМК Пресс, 2017. – 480 с.

32. Харрисон М., Хоуэлл Д. Статистические методы в социальных науках. – М.: ЮНИТИ-ДАНА, 2008. – 512 с.

33. Шестаков Д. А. Социальные факторы преступности. – М.: Норма, 2011. – 304 с.

# ОГЛАВЛЕНИЕ

<b>ВВЕДЕНИЕ.....</b>	<b>4</b>
<b>ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ПОДХОДЫ К МОНИТОРИНГУ УРОВНЯ ПРЕСТУПНОСТИ.....</b>	<b>8</b>
1.1. Мониторинг преступности в системе криминологических исследований.....	8
1.2. Характеристика и источники данных о преступности.....	10
1.3. Традиционные методы анализа криминальной статистики .....	14
<b>ГЛАВА 2. ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В АНАЛИЗЕ УРОВНЯ ПРЕСТУПНОСТИ.....</b>	<b>20</b>
2.1. Машинное обучение в анализе преступности .....	20
2.2. Деревья решений как интерпретируемые модели анализа преступности .....	22
2.3. Ансамблевые методы и случайный лес в криминологических исследованиях .....	25
2.4. Интерпретация моделей с использованием SHAP Values в анализе уровня преступности.....	28
<b>ГЛАВА 3. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ .....</b>	<b>36</b>
3.1. Программные средства и архитектура аналитической системы.....	36
3.2. Реализация и обучение моделей машинного обучения .....	38
<b>ЗАКЛЮЧЕНИЕ .....</b>	<b>43</b>
<b>ЛИТЕРАТУРА .....</b>	<b>45</b>