



Отчет о проверке

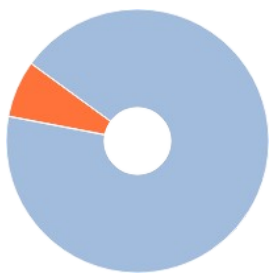
Автор: Соломон Фейт Осэби

Название документа: Chapter1-Rus

Проверяющий: Ушканова Елена Васильевна

Организация: Южный федеральный университет

РЕЗУЛЬТАТЫ ПРОВЕРКИ



Совпадения:
7,07%



Оригинальность:
92,93%



Цитирования:
0%



Самоцитирования:
0%



«Совпадения», «Цитирования», «Самоцитирования», «Оригинальность» являются отдельными показателями, отображаются в процентах и в сумме дают 100%, что соответствует проверенному тексту документа.



Есть подозрения на следующие группы маскировки заимствований: Сгенерированный текст на страницах: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10... еще на 6 стр.



Проверено: 98,83% текста документа, исключено из проверки: 1,17% текста документа. Разделы, отключенные пользователем: Титульный лист, Содержание, Приложение

- **Совпадения** — фрагменты проверяемого текста, полностью или частично сходные с найденными источниками, за исключением фрагментов, которые система отнесла к цитированию или самоцитированию. Показатель «Совпадения» — это доля фрагментов проверяемого текста, отнесенных к совпадениям, в общем объеме текста.
- **Самоцитирования** — фрагменты проверяемого текста, совпадающие или почти совпадающие с фрагментом текста источника, автором или соавтором которого является автор проверяемого документа. Показатель «Самоцитирования» — это доля фрагментов текста, отнесенных к самоцитированию, в общем объеме текста.
- **Цитирования** — фрагменты проверяемого текста, которые не являются авторскими, но которые система отнесла к корректно оформленным. К цитированиям относятся также шаблонные фразы; библиография; фрагменты текста, найденные модулем поиска «СПС Гарант: нормативно-правовая документация». Показатель «Цитирования» — это доля фрагментов проверяемого текста, отнесенных к цитированию, в общем объеме текста.
- **Текстовое пересечение** — фрагмент текста проверяемого документа, совпадающий или почти совпадающий с фрагментом текста источника.
- **Источник** — документ, проиндексированный в системе и содержащийся в модуле поиска, по которому проводится проверка.
- **Оригинальный текст** — фрагменты проверяемого текста, не обнаруженные ни в одном источнике и не отмеченные ни одним из модулей поиска. Показатель «Оригинальность» — это доля фрагментов проверяемого текста, отнесенных к оригинальному тексту, в общем объеме текста.

Обращаем Ваше внимание, что система находит текстовые совпадения проверяемого документа с проиндексированными в системе источниками. При этом система является вспомогательным инструментом, определение корректности и правомерности совпадений или цитирований, а также авторства текстовых фрагментов проверяемого документа остается в компетенции проверяющего.

ИНФОРМАЦИЯ О ДОКУМЕНТЕ

Номер документа: 662

Тип документа: Выпускная квалификационная работа

Дата проверки: 04.06.2024 19:33:34

Дата корректировки: 04.06.2024 20:34:17

Количество страниц: 17

Символов в тексте: 49141

Слов в тексте: 5801

Число предложений: 324

Комментарий: не указано

ПАРАМЕТРЫ ПРОВЕРКИ

Выполнена проверка с учетом редактирования: Нет

Выполнено распознавание текста (OCR): Нет

Выполнена проверка с учетом структуры: Да

Модули поиска: Перефразирования по СПС ГАРАНТ: аналитика, Библиография, Издательство Wiley, Публикации РГБ, Диссертации НББ, Патенты СССР, РФ, СНГ, Цитирование, Публикации eLIBRARY, Перефразирования по Интернету (EN), ИПС Адилет, СМИ России и СНГ, IEEE, Коллекция НБУ, Перефразирования по коллекции издательства Wiley, Интернет Плюс*, Перефразирования по коллекции IEEE, Медицина, Переводные заимствования по Интернету (EnRu), Переводные заимствования издательства Wiley, Сводная коллекция ЭБС, Переводные заимствования IEEE, СПС ГАРАНТ: аналитика, СПС ГАРАНТ: нормативно-правовая документация, Переводные заимствования (RuEn), Перефразированные заимствования по коллекции Интернет в английском сегменте, Публикации eLIBRARY (переводы и перефразирования), Переводные заимствования по коллекции Интернет в русском сегменте, Переводные заимствования по коллекции Интернет в английском сегменте, Перефразирования по Интернету, Перефразированные заимствования по коллекции Интернет в русском сегменте, Переводные заимствования по коллекции Гарант: аналитика, Собственная коллекция компании

ИСТОЧНИКИ

№	Доля в тексте	Доля в отчете	Источник	Актуален на	Модуль поиска	Комментарий
[01]	2,28%	2,28%	Применение инновационных те...	17 Янв 2024	Публикации eLIBRARY (переводы и перефразирования)	
[02]	2,19%	2,19%	Цифровые технологии и сферы ... https://elibrary.ru	31 Дек 2021	Публикации eLIBRARY (переводы и перефразирования)	
[03]	1,89%	1,72%	СТАТИЧЕСКИЕ СТРУКТУРЫ ДАНН... http://elibrary.ru	26 Янв 2019	Публикации eLIBRARY (переводы и перефразирования)	
[04]	0,68%	0%	Статистические методы оценки ... https://gomelagro.ru	04 Июн 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[05]	0,61%	0%	5 основных навыков, необходим... https://evogeeek.ru	04 Июн 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[06]	0,49%	0,49%	Перевернет ли ChatGPT нашу ИТ... http://elibrary.ru	01 Янв 2023	Публикации eLIBRARY (переводы и перефразирования)	
[07]	0,49%	0,39%	УНИФИКАЦИЯ ДАННЫХ БУХГАЛТ...	17 Янв 2024	Публикации eLIBRARY (переводы и перефразирования)	
[08]	0,46%	0%	Право цифровой среды (моногра... http://ivo.garant.ru	20 Авг 2022	Перефразирования по СПС ГАРАНТ: аналитика	Источник исключен. Причина: Маленький процент пересечения.
[09]	0,46%	0%	%D0%92%D1%8B%D0%BF%D1%8... https://gpa.cfuv.ru	30 Апр 2024	Перефразированные заимствования по коллекции Интернет в русском сегменте	Источник исключен. Причина: Маленький процент пересечения.
[10]	0,41%	0%	Искусственный интеллект и ана... https://nauchniestati.ru	21 Фев 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[11]	0,33%	0%	X5_Annual_Report_2019_RUS_com... https://rspp.ru	18 Фев 2024	Перефразированные заимствования по коллекции Интернет в русском сегменте	Источник исключен. Причина: Маленький процент пересечения.
[12]	0,32%	0%	https://accounting.fa.ru/jour/articl... https://accounting.fa.ru	07 Апр 2022	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[13]	0,31%	0%	Анализ данных Excel - Краткое р... https://coderlessons.com	17 Ноя 2020	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[14]	0,31%	0%	Монография Боровская МА, Шев...	25 Дек 2020	Собственная коллекция компании	Источник исключен. Причина: Маленький процент пересечения.
[15]	0,29%	0%	Машинное обучение: как приме... https://fastercapital.com	26 Мая 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[16]	0,26%	0%	Полнота данных: все, что вам ну... https://astera.com	19 Мая 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[17]	0,25%	0%	Достижения в полууправляемом ... https://machinelearningmastery.ru	02 Янв 2021	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[18]	0,24%	0%	Услуги по обучению цифровой т... https://fastercapital.com	22 Мая 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[19]	0,24%	0%	Административно-правовое рег... http://elibrary.ru	05 Авг 2016	Публикации eLIBRARY	Источник исключен. Причина: Маленький процент пересечения.
[20]	0,24%	0%	МЕЖДУНАРОДНОЕ СОТРУДНИЧЕ...	22 Фев 2017	Сводная коллекция ЭБС	Источник исключен. Причина: Маленький процент пересечения.

[21]	0,24%	0%	Метелев А.В., Образцов В.А., Поз... http://ivo.garant.ru	13 Окт 2018	СПС ГАРАНТ: аналитика	Источник исключен. Причина: Маленький процент пересечения.
[22]	0,24%	0%	Чапчиков С.Ю. Комментарий к Ф... http://ivo.garant.ru	22 Окт 2011	СПС ГАРАНТ: аналитика	Источник исключен. Причина: Маленький процент пересечения.
[23]	0,22%	0%	Преимущества и недостатки раб... https://bim-portal.ru	26 Авг 2023	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[24]	0,19%	0%	Topic - Publications - National Cri... https://ncjrs.gov	23 Мар 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[25]	0,19%	0%	Интерпретация данных: как сде... https://fastercapital.com	27 Мая 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[26]	0,18%	0%	Введение в аналитику данных, В... https://bi-data.ru	30 Апр 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[27]	0,18%	0%	Введение в аналитику данных, В... https://bi-data.ru	29 Апр 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[28]	0,18%	0%	Национальный стандарт РФ ГОС... http://ivo.garant.ru	22 Мар 2022	СПС ГАРАНТ: нормативно-правовая документация	Источник исключен. Причина: Маленький процент пересечения.
[29]	0,17%	0%	КРИМИНАЛИСТИКА. Учебник дл...	06 Мар 2017	Сводная коллекция ЭБС	Источник исключен. Причина: Маленький процент пересечения.
[30]	0,15%	0%	Статистический анализ данных: ... https://apni.ru	16 Апр 2023	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[31]	0,15%	0%	Постановление Администрации ... http://ivo.garant.ru	27 Июн 2013	СПС ГАРАНТ: нормативно-правовая документация	Источник исключен. Причина: Маленький процент пересечения.
[32]	0,15%	0%	https://bigdataminsk.bsuir.by/file... https://bigdataminsk.bsuir.by	20 Апр 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[33]	0,15%	0%	ИНФОРМАЦИОННЫЕ ТЕХНОЛОГ... http://elibrary.ru	01 Янв 2016	Публикации eLIBRARY	Источник исключен. Причина: Маленький процент пересечения.
[34]	0,14%	0%	ИСПОЛЬЗОВАНИЕ ДАШБОРДОВ ... http://elibrary.ru	01 Янв 2023	Публикации eLIBRARY	Источник исключен. Причина: Маленький процент пересечения.
[35]	0,14%	0%	СОЦИАЛЬНЫЕ СЕТИ: ВЛИЯНИЕ Н... http://elibrary.ru	01 Янв 2023	Публикации eLIBRARY	Источник исключен. Причина: Маленький процент пересечения.
[36]	0,14%	0%	Информационные технологии в ... https://cchgeu.ru	26 Ноя 2022	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[37]	0,14%	0%	https://cchgeu.ru/upload/science/... https://cchgeu.ru	06 Мар 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[38]	0,14%	0%	Принципы технологии больших ... https://b2b2c.market	08 Ноя 2023	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[39]	0,14%	0%	10 рекомендаций по обеспечени... https://appmaster.io	26 Апр 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[40]	0,13%	0%	Перевернет ли ChatGPT нашу ИТ... https://bit.samag.ru	12 Дек 2023	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[41]	0,13%	0%	Эффективность управленческой ... http://elibrary.ru	раньше 2011	Публикации eLIBRARY	Источник исключен. Причина: Маленький процент пересечения.
[42]	0,13%	0%	Вестник Санкт-Петербургского у... http://bibliorossica.com	26 Мая 2016	Сводная коллекция ЭБС	Источник исключен. Причина: Маленький процент пересечения.
[43]	0,13%	0%	Приказ МВД РФ от 20 марта 1996... http://ivo.garant.ru	раньше 2011	СПС ГАРАНТ: нормативно-правовая документация	Источник исключен. Причина: Маленький процент пересечения.
[44]	0,13%	0%	Агамагомедова С.А., Гречкина О.... http://ivo.garant.ru	21 Дек 2019	СПС ГАРАНТ: аналитика	Источник исключен. Причина: Маленький процент пересечения.
[45]	0,12%	0%	Большие данные: трансформаци... https://fastercapital.com	01 Мая 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[46]	0,12%	0%	Exploratory Analysis of Energy Us... https://frontiersin.org	12 Ноя 2020	СМИ России и СНГ	Источник исключен. Причина: Маленький процент пересечения.
[47]	0,12%	0%	Topic - Publications - National Cri... https://ncjrs.gov	23 Мар 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[48]	0,12%	0%	Topic - Publications - National Cri... https://ncjrs.gov	24 Мар 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[49]	0,12%	0%	OJP Publications - National Crimin... https://ncjrs.gov	28 Дек 2023	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[50]	0,12%	0%	http://www.iadlest.org/Portals/0/Fi... http://iadlest.org	28 Дек 2023	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[51]	0,12%	0%	https://safety.fhwa.dot.gov/ped_bi... https://safety.fhwa.dot.gov	09 Янв 2021	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[52]	0,12%	0%	http://www.iadlest.org/Portals/0/Fi... http://iadlest.org	28 Дек 2023	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[53]	0,12%	0%	ИСПОЛЬЗОВАНИЕ ИСКУССТВЕН... https://cyberleninka.ru	05 Мая 2024	Интернет Плюс*	Источник исключен. Причина: Маленький процент пересечения.
[54]	0,12%	0%	Постановление Правительства П... http://ivo.garant.ru	31 Янв 2011	СПС ГАРАНТ: нормативно-правовая документация	Источник исключен. Причина: Маленький процент пересечения.
[55]	0,11%	0%	Современные образовательные ... http://elibrary.ru	26 Дек 2015	Публикации eLIBRARY	Источник исключен. Причина: Маленький процент пересечения.

[56]	0,11%	0%	Проблемы борьбы с организа... http://ivo.garant.ru	17 Сен 2022	СПС ГАРАНТ: аналитика	Источник исключен. Причина: Маленький процент пересечения.
[57]	0,11%	0%	Киселев А.П., Васильев О.А., Бел... http://ivo.garant.ru	13 Фев 2016	СПС ГАРАНТ: аналитика	Источник исключен. Причина: Маленький процент пересечения.
[58]	0,11%	0%	Распоряжение Правительства Р... http://ivo.garant.ru	16 Мар 2013	СПС ГАРАНТ: нормативно-правовая документация	Источник исключен. Причина: Маленький процент пересечения.

Содержание

1. Аннотация
2. Введение
 - 2.1. Постановка проблемы и цели исследования.
 - 2.2. Обзор литературы
 - 2.3. Методология
 - 2.3.1. Дизайн исследования
 - 2.3.2. Методы сбора данных
3. Глава первая:
 - 3.1. Концепция и принципы анализа больших данных.
 - 3.2. Обзор ситуации в правоохранительной сфере
 - 3.3. Предыстория и значение анализа больших данных в правоохранительной деятельности
 - 3.3.1. Применение анализа больших данных в правоохранительной деятельности во всем мире
 - 3.3.2. Существующие исследования или инициативы, связанные с анализом больших данных в правоохранительных органах

Аннотация

В этом исследовании изучается использование аналитики больших данных в стратегиях правоохранительных органов, уделяя особое внимание ее потенциалу для улучшения усилий по предотвращению и расследованию преступлений. Используя как качественные, так и количественные методы, исследование направлено на изучение применения методов анализа данных для решения современных проблем, с которыми сталкиваются правоохранительные органы. Используя отзывы сотрудников правоохранительных органов, политиков и экспертов по технологиям, а также анализ наборов данных, полученных с таких платформ, как Kaggle, исследование направлено на выявление возможностей и препятствий в использовании анализа больших данных для правоохранительных органов. Изучая этические соображения, социальные и практические последствия интеграции анализа данных в правоохранительную практику, исследование призвано внести вклад в принятие обоснованных решений и разработку политики в этой важной области.

Введение

В последние годы распространение цифровых технологий и экспоненциальный рост данных произвели революцию в различных аспектах жизни общества, включая правоохранительную деятельность. Правоохранительные органы все чаще обращаются к аналитике больших данных, чтобы улучшить свои стратегии по предотвращению и расследованию преступлений, используя огромные объемы данных, полученных из различных источников, таких как отчеты о преступлениях, записи камер наблюдения, социальные сети и датчики. Этот переход к полицейской деятельности, основанной на данных, обещает повысить эффективность, результативность и оперативность усилий правоохранительных органов.

Аналитика больших данных действительно представляет собой сложную концепцию, которая предполагает применение передовых аналитических методов к огромным объемам данных. По своей сути аналитика больших данных направлена на извлечение ценной информации из больших и разнообразных наборов данных. Используя методы анализа высокого уровня, организации могут выявить закономерности, тенденции и корреляции, которые могут способствовать принятию обоснованных решений.

2

В сфере больших данных данные поступают из разных источников, в разных размерах и форматах. Управление и обработка таких данных может оказаться сложной задачей из-за их огромного объема и неоднородности. Задача заключается в том, чтобы эффективно справиться с притоком данных и извлечь из них значимую информацию. Аналитика больших данных решает эту проблему, предоставляя средства для навигации по огромному объему данных и поиска актуальной информации, соответствующей конкретным целям.

В отличие от традиционных подходов к обработке данных, где запросы формулируются на основе имеющихся данных, анализ больших данных использует другой подход. Вместо создания запросов на основе существующих данных аналитика больших данных фокусируется на понимании требований организации, а затем пропускает большую часть данных через эти требования для получения значимых результатов. Этот сдвиг в перспективе позволяет организациям получать ценную информацию и принимать обоснованные решения на основе всего набора данных, а не полагаться на подмножество данных.

Процесс анализа больших данных включает в себя несколько ключевых этапов. Во-первых, данные собираются из различных источников, которые могут включать структурированные базы данных,

платформы социальных сетей, датчики и многое другое. После сбора данные подвергаются процессу очистки для удаления дубликатов, ошибок и ненужной информации. Далее очищенные данные обрабатываются и анализируются с использованием специальных алгоритмов или аналитических инструментов. Эти инструменты могут включать в себя анализ текста, контент-анализ, статистический анализ и прогнозный анализ. Применяя эти методы, из данных можно получить ценную информацию. Дальнейшее объяснение аналитики больших данных будет представлено в последующих разделах этого исследования.

Машинное обучение играет решающую роль в анализе больших данных. Алгоритмы машинного обучения могут учиться на данных и обобщать их, используя статистические подходы. Эта возможность делает машинное обучение эффективным средством анализа больших данных. Автоматизируя процесс анализа больших и сложных наборов данных, машинное обучение значительно снижает проблемы и трудности, связанные с традиционными методами.

Однако традиционные методы машинного обучения, основанные на определенных входных и выходных данных, могут оказаться неэффективными в контексте анализа больших данных, где потоки данных представлены в различных формах и типах. Разнообразие, характерное для больших данных, требует использования алгоритмов глубокого машинного обучения, которые могут обрабатывать сложности и нюансы данных. Модели глубокого обучения с их способностью изучать иерархические представления подходят для решения задач, связанных с анализом больших данных.

Еще одним важным соображением в аналитике больших данных является необходимость непрерывного обучения моделей машинного обучения. По мере появления новых параметров, типов и областей данных модели необходимо постоянно обновлять и обучать, чтобы обеспечить точную и актуальную информацию. Динамичный характер больших данных требует от организаций постоянной адаптации своих моделей к меняющемуся ландшафту данных.

Интеграция анализа больших данных в правоохранительную практику поднимает сложные этические, социальные и практические вопросы. Проблемы, связанные с конфиденциальностью данных, гражданскими свободами, предвзятостью, подотчетностью и прозрачностью, должны быть тщательно решены, чтобы обеспечить ответственное и этическое использование анализа данных в правоохранительных органах. Кроме того, технические проблемы сбора, хранения, анализа и интерпретации данных создают серьезные препятствия для эффективной реализации стратегий анализа данных в правоохранительных органах.

На этом фоне данное исследование направлено на критический анализ роли анализа больших данных в стратегиях правоохранительных органов с акцентом на его потенциальные преимущества, проблемы и последствия. Используя смешанный подход, который сочетает в себе качественные интервью с количественным анализом наборов данных, полученных с таких платформ, как Kaggle, исследование направлено на предоставление информации о возможностях и препятствиях в использовании анализа данных для предотвращения и расследования преступлений. Посредством углубленного изучения этических соображений, социальных последствий и практических последствий полицейской деятельности, основанной на данных, исследование стремится предоставить информацию для принятия решений и разработки политики в этой быстро развивающейся области.

Постановка проблемы и цели исследования.

В последние годы правоохранительные органы (LEA) были завалены огромными объемами данных, полученных из разных источников, включая отчеты о преступлениях, записи камер наблюдения, исторические данные, социальные сети и так далее. Проблема заключается в необходимости эффективно использовать потенциал анализа больших данных и повышать операционную эффективность, оптимизировать распределение ресурсов и улучшать результаты общественной безопасности. Несмотря на достижения в области технологий и инструментов анализа данных, многие правоохранительные органы с трудом могут эффективно использовать большие данные из-за таких ограничений, как неадекватная инфраструктура, ограниченные аналитические возможности, а также юридические и этические проблемы, связанные с конфиденциальностью данных и гражданскими свободами.

Поэтому существует острая необходимость в комплексных исследованиях для решения этих проблем и разработки действенных стратегий и решений для максимизации преимуществ анализа больших данных в правоохранительных органах, одновременно снижая потенциальные риски и обеспечивая соблюдение правовых и этических стандартов.

Научно-исследовательские цели

1. Изучить текущее состояние внедрения и использования аналитики больших данных в правоохранительных органах во всем мире.
2. Определить ключевые проблемы и барьеры, препятствующие эффективному внедрению аналитики больших данных в правоохранительных органах.
3. Оценить потенциальные выгоды и возможности, связанные с использованием анализа больших данных для предотвращения, расследования и реагирования на преступления.
4. Изучить передовой опыт и успешные тематические исследования инициатив по анализу больших данных в правоохранительных органах.
5. Анализировать юридические и этические соображения, связанные с использованием анализа больших данных в правоохранительных органах, включая вопросы, связанные с конфиденциальностью, прозрачностью и подотчетностью данных.
6. Разработать практические рекомендации и рекомендации для правоохранительных органов по расширению их возможностей в эффективном и ответственном использовании анализа больших данных.
7. Вносить вклад в научную литературу и политический дискурс по анализу больших данных в правоохранительных органах, информируя будущие исследования и принятие решений в этой области.

Решая эти исследовательские задачи, данное исследование призвано предоставить ценную информацию и рекомендации для правоохранительных органов, политиков, исследователей и других заинтересованных сторон, стремящихся использовать потенциал анализа больших данных для улучшения общественной безопасности и повышения эффективности усилий правоохранительных органов.

Обзор литературный

Распространение цифровых технологий и экспоненциальный рост данных существенно повлияли на различные сектора, включая правоохранительную деятельность. Правоохранительные органы (LEA) все чаще обращаются к анализу больших данных, чтобы улучшить свои стратегии по предотвращению преступлений, расследованию и общим полицейским стратегиям. В этом обзоре рассматриваются потенциальные преимущества, проблемы и этические соображения, связанные с интеграцией анализа больших данных в правоохранительную практику.

Анализ больших данных позволяет правоохранительным органам выявлять модели преступности, прогнозировать зоны повышенного риска и более эффективно распределять ресурсы. Такие исследования, как Zheng и др., 2019: демонстрируют, как анализ данных о преступности наряду с тенденциями в социальных сетях может помочь в прогнозировании потенциальных вспышек насилия.

Анализируя обширные наборы данных из различных источников (например, отчетов о преступлениях, записей с камер видеонаблюдения, социальных сетей, сенсорных сетей), правоохранительные органы могут генерировать следственные версии, выявлять подозреваемых и более эффективно реконструировать места преступлений. Исследование Chen и др., 2014 г.: подчеркивает потенциал анализа больших данных для выявления скрытых связей в данных о преступности, которые могут быть упущены традиционными методами расследования.

Аналитика больших данных позволяет правоохранительным органам принимать решения, основанные на фактах. Анализ данных может помочь в распределении ресурсов, стратегиях патрулирования и тактике развертывания, что потенциально приведет к более эффективным и целенаправленным усилиям полиции.

Сбор, хранение и анализ огромных объемов персональных данных вызывают серьезные проблемы с конфиденциальностью. Чтобы сбалансировать потребности в безопасности с индивидуальными правами на конфиденциальность, необходимо тщательно учитывать правила защиты данных, такие как GDPR (Общие правила защиты данных) в ЕС. Исследования Ohm, 2010 г.: изучить юридические и этические сложности, связанные с использованием больших данных в правоохранительных органах.

Алгоритмы анализа данных могут увековечить существующие предвзятости, если они обучены на предвзятых наборах данных. Это может привести к дискриминационным последствиям, таким как несправедливое профилирование определенных демографических групп. Исследование Selbst и др., 2016: подчеркивает опасность алгоритмической предвзятости в инструментах оценки рисков, используемых правоохранительными органами.

Большие наборы данных уязвимы для кибератак, которые могут поставить под угрозу конфиденциальную информацию и нарушить работу правоохранительных органов. Надежные меры безопасности данных имеют решающее значение для снижения этих рисков.

Методология

Дизайн исследования

В этом исследовании «Использование анализа больших данных в правоохранительной деятельности» я использовал количественные методы, которые играют решающую роль в

предоставлении эмпирических данных и статистическом анализе, дополняющих наши качественные результаты.

Один из аспектов нашей количественной методологии включает в себя проведение всестороннего обзора литературы существующих исследований и эмпирических данных, связанных с анализом больших данных и его использованием в правоохранительной деятельности. Этот обзор литературы позволяет нам синтезировать количественные данные, такие как статистические данные, тенденции данных и эмпирический анализ, проведенный исследователями в этой области. Систематически анализируя и обобщая количественные исследования, мы можем выявить закономерности, тенденции и эмпирические выводы, имеющие отношение к нашему исследованию.

Кроме того, мы используем сравнительный анализ как метод количественного исследования для сравнения нормативной базы, правовых стандартов и политических подходов, регулирующих использование больших данных в разных юрисдикциях. Этот сравнительный анализ включает количественную оценку различий, сходств и различий в том, как большие данные регулируются и реализуются в различных контекстах. Анализируя количественные данные, связанные с нормативно-правовой базой и политическими мерами, мы можем выявить глобальные тенденции, передовой опыт и области для улучшения управления анализом больших данных в правоохранительных органах.

В целом, интеграция количественных методов в дизайн нашего исследования позволяет нам дополнить наши качественные результаты эмпирическими данными, статистическим анализом и сравнительными оценками. Такой смешанный подход позволяет нам предложить всестороннее и детальное исследование этических и социальных последствий технологии распознавания лиц в правоохранительных органах, способствуя принятию обоснованных решений и разработке политики в этой быстро развивающейся области.

Методы сбора данных

Для этого исследования мы используем многогранный метод сбора данных для сбора разнообразных идей и точек зрения. Был проведен комплексный обзор литературы для обобщения существующих исследований, исследований и эмпирических данных, связанных с анализом больших данных и его использованием в правоохранительной деятельности. Этот обзор литературы позволяет нам выявить соответствующие количественные данные, статистические данные и эмпирический анализ, проведенный исследователями в этой области, предоставляя дополнительный контекст и информацию для информирования нашего исследования. Использование сравнительного анализа для изучения нормативно-правовой базы, правовых стандартов и политических подходов, регулирующих использование больших данных в разных юрисдикциях. Этот сравнительный анализ включает в себя рассмотрение и анализ количественных данных, связанных с мерами регулирования, политическими мерами и стратегиями реализации, что позволяет выявить глобальные тенденции, различия и лучшие практики в управлении анализом больших данных в правоохранительных органах. Поиск онлайн-данных через такие платформы, как Kaggle, служит жизненно важным методом сбора данных. Kaggle хранит разнообразные наборы данных, имеющие отношение к правоохранительным органам: от отчетов о преступлениях и данных о происшествиях до демографической информации и активности в социальных сетях. Подключаясь к хранилищу наборов данных Kaggle, мы получаем доступ к огромному количеству количественных данных, которые могут помочь в наших исследованиях в области анализа больших данных в правоохранительных органах. Эти наборы данных могут включать исторические данные

о преступности, географическую информацию, демографические профили сообществ и другие соответствующие переменные.

Объединив эти методы сбора данных, данное исследование направлено на формирование всестороннего понимания этических соображений и социальных последствий анализа больших данных в правоохранительной деятельности. Такой многогранный подход позволяет нам собирать различные точки зрения, эмпирические данные и количественные данные для обоснования нашего анализа и внесения вклада в принятие обоснованных решений и разработку политики в этой сложной и быстро развивающейся области.

Концепция и принципы анализа больших данных

Прежде всего нам нужно знать, что такое большие данные. Что такое большие данные?

Как правило, большие данные — это огромные объемы структурированных, полуструктурированных и неструктурированных данных, которые генерируются с огромной скоростью. Эти данные могут иметь разную форму: от текста до аудио, видео, данных, полученных от датчиков, торговых графиков Форекс и многого другого.

Большинство людей говорят, что любые данные, которые имеют три V: объем, скорость и разнообразие, называются большими данными.[7] Эти наборы данных быстро поступают в больших объемах в различных формах и продолжают расти в геометрической прогрессии с течением времени.[6]

Согласно Википедии, большие данные в первую очередь относятся к наборам данных, которые слишком велики или сложны, чтобы с ними можно было справиться с помощью традиционного прикладного программного обеспечения для обработки данных.[8]

Помимо трех характеристик больших данных, в последние годы появились еще две V, они называются Value, Veracity.[9]

Характеристики больших данных

1. Объем. Характеристика объема относится к размеру или количеству данных или набора данных. Любой набор данных, называемый большими данными, должен быть большим по размеру, как следует из названия. Данные количественно выражаются или описываются единицей, называемой «байт», наборы данных в больших данных могут занимать терабайты, петабайты, эксабайты, зеттабайты или йоттабайты. Например: Facebook может генерировать около миллиарда сообщений, в 4,5 миллиарда раз регистрируется кнопка «Мне нравится», и каждый день загружается более 350 миллионов новых сообщений.[10]

2. Скорость: скорость — это скорость создания данных. Это описывает скорость, с которой данные передаются из источника.[10] Это то, насколько быстро генерируются, собираются и анализируются данные.[11] В прошлые времена создание данных было ограничено, данные можно было получать только пакетами, затем они переходили в близкое к реальному времени, затем в реальном времени, а теперь — в потоковом режиме.[1]

3. Разнообразие: это относится к форме **3** ли формату больших данных. Данные могут храниться в разных форматах **3** или формах: структурированных, полуструктурированных и неструктурированных.

- Структурированные данные: это данные, которые организованы, их легче читать и искать, поскольку они обычно содержатся в строках и столбцах, составляющих таблицу.[12] Данные, хранящиеся в электронной таблице Excel или в базе данных, представляют собой структурированные данные.

- Полу-структурированные данные: это информация, которая не хранится в реляционной базе данных, но имеет некоторые организационные свойства, которые облегчают анализ.[13] Пример: XML, CSV, JSON. Такие данные содержат теги или другие маркеры для разделения семантических элементов и обеспечения иерархии записей и полей в данных.[14]

3

- Неструктурированные данные: это данные, которые не организованы, они не упорядочены в соответствии с заданной или предопределенной моделью или схемой данных. Такие данные нельзя использовать в СУБД. Пример: текстовые документы, файлы журналов, резюме.

Большие данные — это смесь всех этих форм данных, а набор данных, содержащий все эти формы данных, — это большие данные. Это усложняет анализ и обработку больших данных.

4. Достоверность: эта характеристика определяет качество, точность, целостность, достоверность и надежность больших данных. Это относится к уровню доверия к набору данных. Насколько точны данные? Полны ли данные?[15]

5. Ценность: это относится к преимуществам, которые предлагают большие данные.[15] Ценностная характеристика больших данных отвечает на следующие вопросы: вписываются ли эти данные в рамки нашей проблемы? Подходят ли эти данные для нашего бизнеса? Ценность больших данных возрастает благодаря той информации, которую можно из них извлечь.

Аналитика больших данных

Большие данные сами по себе — это всего лишь форма материи, имеющая вес и занимающая пространство. Чем хороши большие данные? Это тот факт, что из этого можно извлечь информацию, которая поможет будущим потребностям пользователей. Вот тут-то и приходит аналитика, аналитика больших данных.

Аналитика больших данных в основном применяет высокие аналитические методы к большим данным.[1] Аналитика больших данных описывает процесс 2 выявления тенденций, закономерностей и корреляций в больших объемах необработанных данных, чтобы помочь принимать решения на основе данных. В этих процессах используются знакомые методы статистического анализа, такие как кластеризация и регрессия, и применяются они к более обширным наборам данных.[16]

Учитывая огромный объем данных, генерируемых сегодня, обычное программное обеспечение для обработки больше не может их обрабатывать, равно как и человек не может просмотреть их и заметить детали, чтобы получить представление. Современные аналитические инструменты могут лучше обрабатывать большие данные.

Как работает аналитика больших данных?

Чтобы проанализировать большие данные, они должны пройти через некоторые процессы, мы называем этот процесс методами больших данных. Теперь эти процессы можно сгруппировать в два подпроцесса: управление данными и аналитика.[17]

1. Управление данными

Управление большими данными — это организация, администрирование и управление большими объемами как структурированных, так и неструктурированных данных.[18] Эти большие неорганизованные наборы данных должны быть доступны и иметь необходимое качество для использования, поэтому ими необходимо управлять. Когда управление большими данными эффективно, оно позволяет организациям находить необходимую информацию из большого пула данных и делает информацию легко идентифицируемой. При управлении большими данными данные необходимо собирать, хранить, извлекать, очищать, аннотировать, интегрировать, агрегировать, где это необходимо, и представлять в доступной форме.[19]

- Сбор больших данных

Сбор больших данных включает в себя сбор или получение дополнительных объемов данных из разных источников, которые содержат необходимую информацию для использования организацией. Эти данные бывают разных видов: их можно купить у компаний, предоставляющих данные как услугу, например, Infor Bist, Oracle DaaS, или получить с веб-сайтов с открытым исходным кодом, таких как Kaggle. Данные также можно удалить с веб-сайтов с помощью инструментов очистки, таких как Octoparse, веб-краулер, 80legs, ParseHub и т. д.

- Хранение больших данных

Хранение больших данных — это процесс сохранения или хранения больших данных, созданных различными источниками, такими как социальные сети, устройства IoT и другие платформы. Хранение больших данных может оказаться сложной задачей из-за огромного объема, разнообразия и скорости обработки данных.[20] Наборы данных больших данных неструктурированы, хранилище для них обычно создается с использованием объектного и файлового хранилища.[21] При сборе больших данных их необходимо хранить, и по своей природе их нельзя хранить в обычных системах хранения или базах данных. Существуют архитектуры, разработанные для размещения таких объемов и разнообразия данных, мы назвали их озерами данных и хранилищами данных. Основная разница между этими двумя хранилищами заключается в том, что в озерах данных хранятся необработанные, необработанные данные, а в хранилищах данных хранятся уточненные, обработанные, отфильтрованные или очищенные данные, то есть готовые к анализу. Примерами программного обеспечения для озер данных являются Amazon Web Services (AWS), Cloudera, Snowflake, Microsoft Azure, Google Cloud и т. д. Примерами программного обеспечения для хранилищ данных являются Apache Hive, Teradata Vantage, Amazon Redshift и т. д.

- Извлечение, очистка и аннотирование.

Сохраненные данные извлекаются, мы используем pandas в Python для чтения данных файлов из локальных файлов и используем библиотеку запросов для извлечения данных из Интернета. Затем данные очищаются, обрабатываются шумы, избыточность, дубликаты, шумы и нулевые значения. Существуют разные методы очистки данных: мы можем использовать метод Cleanix, предложенный Hongzhi Wang и др. [30], SCARE (Масштабируемое автоматическое восстановление) Mohamed Yakout [31] или KATARA Zhu Chu и др. [32].

Управление данными можно свести к методу ETL, который представляет собой извлечение, загрузку и преобразование [18].

2. Аналитика

Аналитика — это процесс обнаружения существенных тенденций или закономерностей в данных. Необработанные данные обрабатываются и исследуются, чтобы предоставить ценную информацию и сделать выводы. Аналитика происходит после этапа управления данными: данные должны быть собраны, очищены, преобразованы, а затем на основе таких данных можно провести анализ. Аналитика включает моделирование и анализ данных: это взаимосвязанные концепции, которые тесно взаимодействуют друг с другом, но отличаются друг от друга.

- Моделирование данных. Моделирование данных — это процесс создания концептуального представления данных и их взаимосвязей. Это включает в себя разработку структуры или схемы, которая определяет, как данные будут организованы, сохранены и доступны в базе данных или хранилище данных. Модель данных помогает обеспечить точность, согласованность и целостность данных, а также обеспечивает основу для эффективного управления и анализа данных [33] [34].

- Анализ данных. С другой стороны, анализ данных — это процесс проверки, очистки, преобразования и моделирования данных для обнаружения полезной информации, получения выводов и поддержки принятия решений. Он включает в себя применение статистических и аналитических методов для выявления закономерностей, тенденций и понимания данных. Анализ данных играет решающую роль в различных областях, включая бизнес-аналитику, исследования и принятие решений на основе данных [35].

Моделирование данных создает структуру данных, а анализ данных использует эту структуру для извлечения знаний и ценности из данных. У нас есть разные типы анализа; прогнозный анализ, описательный анализ, диагностический анализ, предписывающий анализ.

- Предиктивная аналитика использует инструменты машинного обучения и исторические данные для прогнозирования и выявления будущих тенденций и событий.

— Описательная аналитика — это самая базовая форма анализа данных. Основное внимание в нем уделяется использованию исторических данных для выявления тенденций и взаимосвязей, обобщению данных и описанию их ключевых характеристик [39].

- Диагностический анализ позволяет глубже изучить данные, чтобы выявить коренные причины проблем или тенденций.

- Предписывающая аналитика использует передовые процессы для оптимизации принятия решений путем рассмотрения потенциальных последствий различных вариантов.

Методы анализа данных

Методы анализа данных — это широкий спектр подходов или методов, которые используются для извлечения информации и знаний из данных. Эти многочисленные методы можно разделить на два основных типа: качественный и количественный анализ.

1. Качественный анализ данных. Качественный анализ данных включает в себя изучение нечисловых данных, чтобы получить представление и понять основной смысл и контекст. Он фокусируется на словах, описаниях, концепциях и идеях, а не на числовых значениях. Методы качественного анализа данных используются для изучения и описания явлений, понимания перспектив и выявления социальных структур или процессов. Метод качественного анализа данных извлекает данные из символов, текстов или слов, фотографий и наблюдений [40].

- Контент-анализ: этот метод включает в себя группировку больших объемов текста в коды, их обобщение по категориям и, возможно, табулирование данных для расчета частоты определенных понятий или переменных. Контент-анализ сочетает в себе элементы качественного и количественного мышления в рамках качественного метода [41].

- Анализ атрибутов: атрибуты — это характеристики или переменные, связанные с качественными данными. Сходства внутри групп или различия между ними можно определить с помощью атрибутов. Анализ атрибутов помогает понять детали выборочной группы и не ставит своей целью сделать предположения или прогнозы в отношении всей совокупности [42].

- Обоснованная теория: Обоснованная теория – это индуктивный подход к качественному анализу данных, который включает разработку теорий или объяснений, основанных на самих данных. Его цель — генерировать новые идеи и теории на основе данных, а не проверять уже существующие гипотезы [43].

- Анализ текста. Анализ текста включает в себя организацию и анализ неупорядоченных текстовых документов для определения их базовой структуры [44]. Он состоит из таких методов, как анализ настроений, тематическое моделирование, обобщение текста, классификация текста, обработка естественного языка, языковой перевод.

2. Количественный анализ данных. Количественный анализ данных касается числовых данных и использует статистические методы для измерения различий, выявления закономерностей и выявления коренных причин проблем. Основное внимание в нем уделяется ответам на такие вопросы, как «сколько» или «какой процент», чтобы предложить широкую информацию, которую можно применить к более крупным группам.

- Регрессионный анализ: изучение взаимосвязи между зависимой переменной и одной или несколькими независимыми переменными для выявления закономерностей, прогнозирования результатов и вывода причинно-следственной связи. Регрессионный анализ включает в себя такие методы, как линейная регрессия, логистическая регрессия и множественная регрессия.

- Описательная статистика: суммирование и описание характеристик набора данных с использованием таких показателей, как среднее значение, медиана, мода, стандартное отклонение и частотное распределение. Описательная статистика дает представление о центральной тенденции, изменчивости и распределении данных.

- Инференциальная статистика: Делание выводов и обобщений о населении на основе выборки данных. Инференциальная статистика включает в себя такие методы, как проверка гипотез, доверительные интервалы и дисперсионный анализ (ANOVA), позволяющие делать выводы и делать прогнозы для более крупных групп населения.

- Факторный анализ: изучение базовой структуры или размеров набора данных путем выявления скрытых факторов, которые объясняют корреляции между наблюдаемыми переменными. Факторный анализ помогает уменьшить сложность данных за счет выявления основных закономерностей и взаимосвязей.

- Кластерный анализ: группировка точек данных в кластеры или сегменты на основе сходства или близости. Кластерный анализ включает в себя такие методы, как кластеризация k-средних и иерархическая кластеризация, для выявления закономерностей и структур в наборах данных.

Анализ временных рядов: анализ данных, собранных с течением времени, для выявления тенденций, сезонности и закономерностей. Анализ временных рядов включает такие методы, как скользящие средние, экспоненциальное сглаживание и модели авторегрессионного интегрированного скользящего среднего (ARIMA) для прогнозирования будущих значений и обнаружения аномалий.

Большие данные: приложения

Большие данные имеют множество практических применений в различных отраслях и секторах. Вот некоторые примеры:

1. Бизнес и маркетинг. Большие данные используются для анализа поведения, предпочтений и тенденций клиентов, что позволяет компаниям принимать решения на основе данных, персонализировать маркетинговые кампании и повышать удовлетворенность клиентов.
2. Финансы и банковское дело. Большие данные помогают обнаруживать мошенничество, оценивать риски, кредитный рейтинг и улучшать качество обслуживания клиентов. Это позволяет финансовым учреждениям анализировать большие объемы транзакционных данных и делать точные прогнозы.
3. Здравоохранение. Большие данные используются для медицинских исследований, наблюдения за заболеваниями, мониторинга пациентов и персонализированной медицины. Это помогает анализировать крупномасштабные данные о состоянии здоровья для выявления закономерностей, улучшения диагностики и разработки эффективных методов лечения.
4. Цепочка поставок и логистика. Анализ больших данных оптимизирует операции цепочки поставок, управление запасами и транспортную логистику. Это позволяет компаниям отслеживать поставки, прогнозировать спрос и повышать общую эффективность.
5. Умные города. Большие данные используются для управления городской инфраструктурой, транспортными потоками, потреблением энергии и общественной безопасностью. Это помогает сделать города более устойчивыми, эффективными и пригодными для жизни.
6. Социальные сети и онлайн-сервисы. Большие данные используются для анализа поведения пользователей, анализа настроений и систем рекомендаций. Это позволяет компаниям персонализировать пользовательский опыт, таргетировать рекламу и повышать вовлеченность пользователей.
7. Исследования и разработки. Большие данные играют решающую роль в научных исследованиях, геномике, открытии лекарств и разработке новых технологий. Это помогает анализировать большие наборы данных, выявлять закономерности и ускорять инновации.
8. Преступность и правоохранительная деятельность. Правоохранительные органы все чаще используют большие данные в своих повседневных операциях и наблюдениях, включая патрулирование, расследование и анализ преступлений. Аналитика больших данных может помочь правоохранительным органам более эффективно использовать ресурсы, прогнозировать и предупреждать поведение и, в итоге, предотвращать и перехватывать преступления, что приведет к снижению уровня преступности [26].

В нашем обществе существует множество других применений анализа больших данных. В этой исследовательской работе мы сосредоточимся на его применении в сфере преступности и правоохранительной деятельности.

Большие данные: проблемы

Большие данные создают ряд проблем, которые организациям необходимо решить, чтобы эффективно использовать свой потенциал.

1. Объем. Огромный объем данных, полученных из различных источников, может быть огромным. Хранение, управление и обработка больших объемов данных требуют масштабируемой инфраструктуры и технологий, способных эффективно обрабатывать такие объемы.

2. Скорость. Данные генерируются с беспрецедентной скоростью, и организациям необходимо обрабатывать и анализировать их в режиме реального времени или почти в реальном времени, чтобы получать своевременную информацию. Для обработки данных с высокой скоростью необходимы надежные технологии обработки и потоковой передачи данных.

3. Разнообразие. Большие данные существуют в различных форматах, включая структурированные, неструктурированные и полуструктурированные данные. Работа с различными типами данных и их интеграция для анализа может оказаться сложной задачей. Организациям необходимо использовать методы и инструменты, которые могут эффективно обрабатывать различные форматы данных.

4. Достоверность. Достоверность означает качество и надежность данных. Большие данные часто содержат шум, противоречия и неточности, которые могут повлиять на точность и достоверность результатов анализа. Обеспечение качества данных и решение проблем их достоверности имеют решающее значение для достоверной информации.

5. Безопасность и конфиденциальность. Большие данные подразумевают обработку чувствительной и конфиденциальной информации. Обеспечение безопасности данных, защита конфиденциальности и соблюдение нормативных требований являются серьезными проблемами. Организациям необходимы надежные меры безопасности и системы конфиденциальности для защиты данных.

6. Навыки и знания. Извлечение информации из больших данных требует специальных навыков и опыта в области анализа данных, науки о данных и инженерии данных. Нехватка квалифицированных специалистов в этих областях представляет собой проблему для организаций.

7. Инфраструктура и стоимость. Хранение, обработка и анализ больших данных требуют значительных вычислительных ресурсов и инфраструктуры. Масштабирование инфраструктуры для обработки растущих объемов данных может оказаться дорогостоящим. Организациям необходимо тщательно планировать и инвестировать в правильную инфраструктуру для эффективного управления большими данными.

Интеграция и управление данными. Интеграция данных из различных источников и обеспечение согласованности и качества данных может быть сложной задачей. Организациям необходимы надежные стратегии интеграции и управления данными, чтобы справиться с проблемами интеграции данных.

Обзор ситуации в правоохранительной сфере

Сегодня в нашем обществе правоохранительная деятельность быстро меняется, она становится все более разнообразной и сложной для разных стран и регионов, отражая их различные политические, культурные и социальные контексты. Несмотря на разнообразие их структур и операционных подходов, между ними существуют некоторые общие тенденции. Это:

1. Агентства: в большинстве, если не во всех странах, есть одно или несколько правительственных агентств, ответственных за соблюдение закона [22]. Примеры агентств включают в себя:

- Полицейские силы: например, Полицейские силы Нигерии (НПФ) в Нигерии и Полиция в России.
- Специализированные подразделения: борьба с наркотиками, киберпреступность, борьба с терроризмом и т. д., например Национальное агентство по борьбе с наркотиками (NDLEA) в Нигерии и Федеральная служба безопасности (ФСБ) в России.

- Пограничные и таможенные органы: предотвращение незаконного оборота наркотиков, людей и оружия.

- Федеральная и местная юрисдикции: их структура варьируется в зависимости от централизованной и децентрализованной системы.

2. Полицейская деятельность: это деятельность по поддержанию общественного порядка и безопасности, обеспечению соблюдения закона, регулированию или контролю со стороны сотрудников полиции [23].

3. Проблемы преступности и безопасности: глобализация, распространение цифровых технологий и другие изменения в обществе способствовали распространению и росту преступности и терроризма.

4. Правовые основы: страны имеют разные правовые системы, включая общее право, гражданское право, шариат и обычное право, которые влияют на структуру и функции правоохранительных органов и отправление правосудия. Кроме того, права человека и нормы права являются правовыми рамками и требуют соблюдения международных правовых стандартов и механизмов надзора.

5. Международное сотрудничество: такая организация, как Международная организация уголовной полиции (Интерпол), помогает сотрудничеству между правоохранительными органами по всему миру. Аналогичным образом, двусторонние и многосторонние соглашения помогают решать проблемы взаимной безопасности.

6. Проблемы и возможности: сегодня проблемы правоохранительных органов включают:

- Недостаточные ресурсы, обучение и оборудование.

- Коррупция и подотчетность

Возникающие угрозы: развивающиеся тенденции преступности, быстрый технологический прогресс и геополитическая динамика создают новые проблемы для правоохранительных органов, что требует постоянной адаптации и инноваций для их эффективного решения.

Предыстория и значение анализа больших данных в правоохранительной деятельности

Большие данные используются в разных областях и секторах общества, в которых не исключена правоохранительная деятельность. Правоохранительная деятельность – это сектор, требующий сбора существенных и разрозненных объемов информации [24]. Таким образом, интеграция аналитики больших данных делает процесс правоохранительной деятельности эффективным и действенным.

Значение:

1. Предотвращение и обнаружение преступности. Аналитика больших данных позволяет правоохранительным органам выявлять закономерности, тенденции и аномалии в больших наборах данных, что позволяет им более эффективно предвидеть и предотвращать преступную деятельность. Анализируя исторические данные о преступности и используя информацию в режиме реального времени, правоохранительные органы могут активно распределять ресурсы и выявлять горячие точки преступности.

2. Возможности расследования. Аналитика больших данных расширяет возможности расследования, позволяя правоохранительным органам анализировать и сопоставлять данные из нескольких источников для выявления подозреваемых, раскрытия преступных сетей и более эффективного сбора доказательств. Передовые аналитические методы, такие как анализ ссылок и анализ социальных сетей, помогают следователям соединить точки и раскрыть сложные дела.

3. Оптимизация ресурсов. Аналитика больших данных помогает правоохранительным органам оптимизировать распределение ресурсов, предоставляя информацию о наиболее эффективном использовании персонала, оборудования и финансовых ресурсов. Расставляя приоритеты ресурсов на основе анализа данных, агентства могут максимизировать свое воздействие и повысить операционную эффективность.

4. Предиктивная полицейская деятельность. Аналитика больших данных облегчает стратегии прогнозирующей полицейской деятельности, которые используют модели, основанные на данных, для прогнозирования того, где и когда могут произойти преступления. Выявляя прогностические факторы и индикаторы риска, правоохранительные органы могут активно вмешиваться в предотвращение преступности и стратегически распределять ресурсы.

Повышение общественной безопасности. В итоге использование анализа больших данных в правоохранительных органах способствует улучшению результатов общественной безопасности, позволяя ведомствам более эффективно реагировать на возникающие угрозы, распределять ресурсы на основе анализа данных и строить более прочные партнерские отношения с сообществами.

Применение анализа больших данных в правоохранительных органах по всему миру

Большие данные все чаще используются в правоохранительных органах во всем мире для улучшения различных аспектов предотвращения преступности, расследования и общественной безопасности. Вот несколько приложений:

- Предиктивная полицейская деятельность ¹ анализ больших данных может помочь правоохранительным органам прогнозировать, когда и где могут произойти преступления, на основе исторических данных о преступности, социально-экономических факторов, погодных условий и других соответствующих переменных. Это позволяет департаментам полиции более эффективно распределять ресурсы и активно предотвращать преступления [36].

- Анализ моделей преступности: анализируя большие наборы данных, правоохранительные органы могут выявить закономерности и тенденции в преступной деятельности, такие как образ действий преступников или появление новых горячих точек преступности. Эта информация помогает в разработке целевых стратегий борьбы с преступностью.

- Поддержка расследований: инструменты анализа больших данных позволяют следователям анализировать огромные объемы структурированных и неструктурированных данных, включая записи камер наблюдения, сообщения в социальных сетях, записи телефонных разговоров и финансовые операции, чтобы собирать доказательства и выстраивать дела против подозреваемых.

- Анализ разведывательной информации: правоохранительные органы используют большие данные для анализа огромных объемов разведывательных данных из различных источников, включая правительственные базы данных, социальные сети и перехваты сообщений, для выявления потенциальных угроз, преступных сетей и террористической деятельности.

- Распознавание лиц и биометрия. Технологии больших данных в сочетании с достижениями в области распознавания лиц и биометрической идентификации используются для идентификации

подозреваемых по записям наблюдения, сравнения фотографий с криминальными базами данных и повышения безопасности границ.

- Обнаружение киберпреступлений: анализ больших данных помогает выявлять закономерности подозрительного поведения, аномалии и потенциальные киберугрозы в огромных объемах сетевого трафика и данных журналов. Это имеет решающее значение для обнаружения и предотвращения киберпреступлений, таких как взлом, утечка данных и онлайн-мошенничество.

- Распределение и размещение ресурсов. Департаменты полиции используют большие данные для оптимизации распределения ресурсов путем анализа тенденций преступности, времени реагирования и потребностей в услугах. Это гарантирует эффективное использование ресурсов правоохранительных органов для решения проблем наиболее нуждающихся областей.

- Общественная безопасность и реагирование на чрезвычайные ситуации. Аналитика больших данных может улучшить реагирование на чрезвычайные ситуации за счет анализа данных в реальном времени из различных источников, таких как социальные сети, датчики и вызовы служб экстренной помощи, для выявления инцидентов, оценки их серьезности и координации усилий по реагированию.

- Картирование и визуализация преступности: географические информационные системы (ГИС) и технологии больших данных используются для создания карт преступности и визуализаций, которые дают представление о моделях преступности, пространственных отношениях и социально-экономических факторах, влияющих на преступную деятельность. Это помогает понять и устранить основные причины преступности.

- Программы оценки и предотвращения рисков: анализ больших данных используется для оценки риска повторных правонарушений среди лиц в системе уголовного правосудия, помогая адаптировать меры вмешательства, программы реабилитации и стратегии надзора для снижения уровня рецидивизма.

Эти приложения демонстрируют, как большие данные меняют практику правоохранительных органов во всем мире, позволяя агентствам стать более активными, эффективными и действенными в борьбе с преступностью и обеспечении общественной безопасности. Однако важно отметить, что при использовании больших данных в правоохранительной деятельности также возникают важные вопросы, касающиеся конфиденциальности, этики и потенциальных предубеждений, которые необходимо тщательно учитывать.

Существующие исследования или инициативы, связанные с анализом больших данных в правоохранительных органах.

Существующие исследования и инициативы, связанные с анализом больших данных в правоохранительных органах, включают:

- Sintelix: Sintelix — это решение для работы с большими данными и аналитикой для правоохранительных органов, которое позволяет осуществлять поиск и анализ баз данных, объединенных с данными социальных сетей и Интернета. Это помогает выявить местных подозреваемых и сузить возможности рассмотрения уголовных дел [25].

- Palantir: Palantir — компания-разработчик программного обеспечения, которая предоставляет аналитическое программное обеспечение, используемое правоохранительными органами, в том числе федеральными агентствами, такими как ЦРУ, и местными правоохранительными органами, такими как Департамент полиции Лос-Анджелеса. Их программное обеспечение помогает анализировать большие данные для расследований и сбора разведывательной информации [26].

- Инициатива «Наука данных для социального блага»: Департамент полиции Атланты сотрудничал с Oracle и Технологическим институтом Джорджии в рамках инициативы «Наука данных для социального блага». Эта инициатива поручает стажерам применять науку о данных к социальным проблемам, включая изучение скорости ответа на звонки 911 с использованием собранных данных [27].

- IBM: IBM предлагает решения для правоохранительных органов, такие как Хранилище криминальной информации (CIW) и Центр интеллектуальных операций для умных городов. Эти решения используют анализ больших данных и прогнозную полицейскую деятельность, чтобы обеспечить целостное представление информации и помочь предотвратить, и раскрыть преступления [28].

- PredPol (США): это коммерчески доступная программа для прогнозирования полицейской деятельности, используемая некоторыми полицейскими управлениями в Соединенных Штатах. PredPol анализирует исторические данные о преступности, чтобы выявить районы с более высокой вероятностью преступности [29].

- Национальный институт юстиции (National Institute of Justice, NIJ) - Подходы к преступности и безопасности дорожного движения, основанные на данных (Data-Driven Approaches to Crime and Traffic Safety, DDACTS): NIJ поддержал исследования и внедрение подходов к полицейской деятельности, основанных на данных, посредством таких инициатив, как DDACTS. Эта программа поощряет правоохранительные органы использовать анализ данных для целенаправленных усилий по обеспечению правопорядка в районах с высоким уровнем преступности и дорожно-транспортных происшествий с целью снижения преступности и повышения общественной безопасности [45].