DOI 10.37882/2223-2966.2024.7-2.11

РАЗРАБОТКА АНАЛИТИКО-ПРОГНОСТИЧЕСКОЙ МОДЕЛИ ДЛЯ СОЗДАНИЯ СИСТЕМЫ ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ НА ОСНОВЕ АВТОМАТИЧЕСКОГО ВЫЯВЛЕНИЯ АНОМАЛИЙ И ФЕЙКОВЫХ ТЕКСТОВ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

DEVELOPMENT OF SOFTWARE SYSTEM FOR SECURITY ASSURANCE BASED ON AUTOMATIC DETECTION OF ANOMALIES AND FAKE TEXTS USING MACHINE LEARNING ALGORITHMS

M. Zolotukhina

Summary. The anomaly detection system in text data, developed using modern technologies and algorithms, is an innovative approach to identifying hidden threats, including fake text data in various information sources. The development plays a key role in ensuring the national security of the state, offering innovative methods for detecting and preventing fake data in the digital sphere. The project covers a wide range of threats, including phishing attacks, disinformation, and data leaks. The unique method of this system includes an in-depth analysis of textual features to identify the characteristic features and patterns inherent in fake or manipulated texts. The effectiveness of the system is ensured using an ML model learning algorithm, which ensures the accurate identification of entities indicating the fake nature of the text. The integration of the system with corporate information systems allows the analysis of data from various sources, including websites, social media, and many others. This approach is an integral component for ensuring a high level of protection of companies (organizations) from fraud and maintaining trust, which is of critical importance in the modern information environment. The development provides high adaptability to new threats and changes in text strategies, which makes it a powerful tool in an ever-changing environment to combat anomalies.

Keywords: ML algorithms, nlp tasks, anomalies, software, security, fake, autocorrelation, static analysis, EIS.

Золотухина Мария Александровна
Аспирант,
МИРЭА Российский Технологический Университет
rtu_mary@mail.ru

Аннотация. Система обнаружения аномалий в текстовых данных, разработанная с использованием современных технологий и алгоритмов, представляет собой инновационный подход к выявлению скрытых угроз, включая фейковые текстовые данные в различных источниках информации. Разработка играет ключевую роль в обеспечении национальной безопасности государства, предлагая инновационные методы выявления и предотвращения фейковых данных в цифровой сфере. Проект охватывает широкий спектр угроз, включая фишинговые атаки, дезинформацию, утечку данных. Уникальный метод этой системы включает глубокий анализ текстовых признаков с целью выявления характерных особенностей и паттернов, свойственных фейковым или манипулированным текстам. Эффективность системы обеспечивается применением алгоритма обучения ML моделей, что обеспечивает точное выявление сущностей, указывающих на фейковый характер текста. Интеграция системы с корпоративными информационными системами позволяет проводить анализ данных из различных источников, включая веб-сайты, социальные медиа и мн. др. Данный подход является неотъемлемым компонентом для обеспечения высокого уровня защиты компаний(организаций) от мошенничества и поддержания доверия, что приобретает критическое значение в условиях современной информационной среды. Разработка предоставляет высокую адаптивность к новым угрозам и изменениям в текстовых стратегиях, что делает его мощным инструментом в постоянно меняющейся среде борьбы с аномалиями.

Ключевые слова: ML алгоритмы, nlp-задачи, аномалии, ПО, обеспечение безопасности, фейк, автокорреляция, статический анализ, КИС.

Введение

Текстовые признаки идентифицирующие, себя как аномалии позволяют создать характеристику сообщения, главы книги или статьи, тем самым помогая обнаружить скрытые носимые угрозы. Под скрытыми угрозами подразумевается нанесение ущерба компании в контексте фишинговых писем, фейковых текстов и раскрытия конфиденциальной информации, т. е. потенциальные риски, которые могут оставаться незамеченными или недооцененными, но при этом могут нанести серьезный ущерб репутации, финансам или бизнес-процессам компании. Скрытые угрозы могут возникать из-за различных факторов, и их выявление и управление являются важными задачами для компаний и предприятий [1].

Сама методика по обнаружению скрытых угроз в виде разработки системы для автоматического выявления аномалий и фейковых текстов дает возможность

управления скрытыми угрозами в организации. Фейковые тексты, созданные с целью навредить, могут вызвать различные проблемы и негативные последствия, а именно:

- Фейковые тексты искажают аналитические данные, которые компания использует для принятия стратегических решений. Следовательно, это влияет на качество анализа и, как следствие, на качество решений.
- Повышенная нагрузка на службу поддержки при отправке дополнительных запросов клиентов.
- Для борьбы с фейковыми рассылками компании могут потребоваться дополнительные затраты, расследования и противодействие мошенничеству.
- Мошеннические тексты оказывают воздействие на решения клиентов о сотрудничестве с организацией. Клиенты могут отказаться от услуг или продуктов компании, опираясь на дезинформацию.
- Несанкционированное раскрытие конфиденциальных данных организаций и частных лиц. Последствия — угроза личной конфиденциальности, потери репутации компаний, возможные правовые последствия.
- Фишинговые письма содержащие текстовые аномалии скрывают истинное предназначение и водят в заблуждение, например пользователей почтовых аккаунтов.

Создать фейковый текст достаточно легко, но тяжелее его обнаружить и поэтому требуются сложные многофункциональные средства идентификации. Эффективный метод идентификации фейка включает в себя комбинацию нескольких методов и средств для максимальной точности и надежности. Данный способ является необходимым для анализа, который использует алгоритмы машинного обучения с разработанным модулем выявления нетипичных, необычных или аномальных текстовых данных среди большого объема нормальных данных. Это важный процесс в различных областях, включая безопасность информации, медицинскую диагностику, финансы и многие другие [2].

КИС и система для выявления фейковых текстов взаимодействуют таким образом, что позволяют компании обнаруживать и управлять фейковыми текстами в данных и на веб-сайтах. КИС интегрирует данные из различных источников, включая веб-сайты компании, онлайн-платформы, социальные медиа [3]. Эти данные затем передаются в ПО для анализа на признаки и составляющие аномалий.

Материалы и методы

Для организации процесса предварительной работы по созданию данных и программы для анализа текстов с целью обеспечения безопасности на основе автоматического выявления аномалий и фейковых текстов был написан алгоритм выполняемых действий, показан в таблице 1. Это помогает структурировать этапы работы, определить необходимые параметры и задачи, а также обеспечивает лучшее понимание промежуточный работы, процессов разработки системы и анализа данных.

Таблица 1.

Алгоритм предварительной работы создания данных и программы

Шаг	Действие	Параметры	Результаты
1	Загрузка данных из различных источников.	сайты, веб-приложения, социальные сети и т. д.	Набор данных для анализа
2	Предварительная обработка данных	Удаление ненужных символов и форматирование, реорганизация текста, уникальные значения	Обработанные данные для анализа
3	Создание датасета	Тексты с аномалиями, описания, сценарии, логические схемы, благоприятные и неблагоприятные события,	Датасет с пометками рейтинговой системой, анализ
4	Поиск уникальных характеристик и кластеризация	Применение алгоритма кластеризации к найденным параметрам	Кластеры общих характеристик текстов
5	Определение цели программы и выбор технологий	описание цели программы, выбор языка программирования и алгоритма машинного обучения. Применение алгоритма обучения.	Цель программы: обнаружение аномалий и фейков в текстовых данных. Язык программирования: Python
6	Категории для обнаружения аномалий	Реорганизация текста по категориям:	Категории: тональность, эмоции, лексика, детализация
7	Анализ взаимосвязей и зависимостей	Применение методов машинного обучения и статистических методов	Аномалии выявлены
8	Исследование данных и дальнейший анализ	Анализ данных на предмет структуры, корреляций, зависимостей	Корреляции и зависимости выявлены, структура и характеристики текстов изучены
9	Результаты	Оценка результатов алгоритма.	Результаты оценены, намечена траектория для дальнейшего исследования.

Для корректного анализа текста из различных областей сайтов, веб-приложений и социальных сетей следует создать датасет состоящий из благоприятных и неблагоприятных описаний естественным языком и [4]. Датасет состоит из текстов с аномалиями, описаний, логических схем, написанных, интеллектуальным чатом, благоприятных и не благоприятных отзывов, написанных человеком, причем все данные помечены рейтинговой системой анализа. Проводится поиск уникальных характеристик для данного стиля создания текста. Он заключается в идентификации ключевых слов и терминов, которые часто встречаются, применение алгоритмов кластеризации по тематическим группам для выделения общих характеристик. Необходимо определить кластеры, фокусирующиеся на качестве описанного объекта естественным языком так, как они могут содержать определенные ключевые слова и фразы. Еще один датасет собирается для определения частей речи, который также оформляется в главную программу для распознавания аномалий в отзывах. Теперь соответственно требуется определить цель программы, на каком языке ее целесообразно писать и какие алгоритмы применять чтобы результат имел высокую точность идентификации.

Цель программы и выбор технологий

Цель программы обнаружить аномалии или несоответствия в тексте по найденным признакам, используемый язык Python имеет большое количество библиотек и фреймворков для алгоритмов машинного обучения и соответственно требуется найти алгоритм, который предоставит функции для сравнения, поиска и анализа составляющих текст.

Для обнаружения аномалий и признаков фейковых действий нужно реорганизовать текст по категориям [5]. А именно:

- Тональность текста — Позитивные аномалии, которые содержат чрезмерное количество положительных эпитетов и мало конкретных деталей.
- Использование эмоций — Текст, содержащий чрезмерное количество эмоций или неадекватно сильные выражения.
- Лексика и стиль — Нестандартное использование лексики, чрезмерная формальность или неестественность стиля.
- Детализация текста — Слишком детализированные или, наоборот, чрезмерно общие, могут быть подвергнуты сомнению.
- Синтаксис и грамматика — Тексты с явными нарушениями синтаксиса или грамматики могут быть аномальными.
- Временные аспекты — Сравнение временных аспектов в текстах, таких как слишком быстрые изменения тональности и др.
- Сравнение или приведение примера с другими текстами — Сравнение структуры и содержания текста с другими о том же виде описываемом объекте поможет выявить аномалии.
- Длина — Крайне короткие или, наоборот, чрезмерно длинные являются подозрительными.
- Оценка качества, рейтинговая система — Если рейтинг на текст является несоразмерно большим, вероятнее всего здесь присутствует признак мошенничества или раскрытия данных, т. к. важно чтобы скрытые аномалии не привлекали внимание со стороны программ для обнаружения.

Отсутствие конструктивной критики означает, что в объекте исследования отсутствует содержательная, обоснованная критика или замечания, направленные на улучшение ситуации или продукта. Обычно конструктивная критика предполагает не просто указание на недостатки, но и предложение вариантов улучшения или рекомендации [6,7,8].

"Дорогой коллега,

Надеюсь, вы в порядке. Просто хотел поделиться новостью о нашем проекте. Вот ссылка на наш обновленный бизнес-план: [ссылка]

Буду рад услышать ваши мысли по этому поводу. С нетерпением жду обсуждения на следующей неделе.

С наилучшими пожеланиями

Рис. 1. Пример исследуемого текста

На рис. 1 показан пример сообщения фишинговой рассылки, которая подвергается распределению по частям речи, частоте, количеству и кластерам для каждого анализируемого модуля используемые в качестве объекта исследования на корреляцию и автокорреляцию. Распределение данных осуществляется с помощью специально написанной программы на Python и конструктивной особенности Natural Language Toolkit(nltk), stopwords, pymorphy2 и т. д.

Метод анализа текстовых данных

Рассмотрим зачем применять анализ частей речи для обнаружения фейковых текстов. Данный анализ влияет на выявление аномалий среди писем в социальных сетях или рассылках, а именно: первая причина — это частота использования негативных слов и конструкций: NOUN (существительные), ADJF (прилагательные): если в тексте существенно увеличивается частота использования негативных существительных и прилагательных, следовательно, это указывает на попытку ухудшить восприятие. Далее синтаксическая сложность и структура предложений: CONJ (союзы), PRCL (частицы), ADVB (наречия): Аномалии в структуре предложений или частое использование союзов и частиц свидетельствуют о попытке искусственного улучшения впечатления или, наоборот, создания излишне сложных и запутанных текстов.

Частота использования глаголов и наречий: VERB (глаголы), ADVB (наречия): чрезмерное использование действенных глаголов и наречий служит признаком попытки создать в тексте ощущение активности или динамичности. Оценка синтаксической сложности: NOUN (существительные), VERB (глаголы), ADJF (прилагательные): фейковые тексты либо излишне сложные и длинные, чтобы внушить авторитетность, либо чрезмерно простые и короткие, чтобы не вызывать подозрений.

Частота и контекст использования этих частей речи варьируется в зависимости от специфики фейковой активности, и их анализ требует определенного контекста и подхода. Комбинация методов, таких как анализ сентимента, машинное обучение и статистические методы достаточно эффективны при выявлении аномалий в текстовых данных. Чтобы расширить границы анализа текста были добавлены в датасет данные, созданные интеллектуальным помощником, причем есть обязательная пометка кем написаны. Также для данного исследования включен модуль анализа текста по корреляционным и автокорреляционным результатам значений [9].

Подробное описание и разбор фейкового создания рассылки или текста используется для максимального использования конкретных деталей и более подробного анализа текстовой структуры. Данный разбор показывает необходимость умело оперировать мелкими и неточными фактами, чтобы сформировать правдоподобный текст. Пример фейковой рассылки показан на рисунке 1. Текст является фейковым и представляют собой удачную подделку, включающую в себя конкретные детали, тем самым создавая впечатление реальности. Примеры фейковых текстов для построения датасета были использованы из источника [10].

Данный анализ помог найти признаки определения аномалий, здесь представлены некоторые признаки из общего числа найденных, которые могут указывать на фейковую аномалию, такие:

- Присутствие критики: Утверждение о том, что наличие критики делает текст более приближенным к настоящему, является признаком, того, что автор старается создать впечатление реальности, внедряя элементы негативной оценки.
- Решение проблемы: Указание на обязательность решения проблемы, даже маленькой, как признак поддельности, подчеркивает, что автор стремится создать сюжет с развитием, что не всегда характерно для настоящих отзывов.
- Разделение на части речи: Упоминание о разделении датасета на части речи может быть признаком структурированности текста, что не всегда характерно для естественной речи.
- Эмоциональная окраска: Использование слова «емко» в описании датасета подчеркивает стремление автора придать своим данным эмоциональную окраску, что характерно для выдуманных отзывов.

Такие данные позволяют провести качественный семантический анализ составляющих текстовую структуру информации. Информативный текст используются по структуре эмоционального состава, как положительные, отрицательные так и нейтральные, но все они могут быть неправдоподобные. Существует необходимость при анализе частей текста, при распределении текстовых данных и выявлении нетипичных или подозрительных паттернов использовать методы, основанные на частоте, распределении и статистики [11, 12]. Это позволит подготовить количественные и описательные характеристики для алгоритма машинного обучения.

A	B	C	D	E
zak_1	col_1	chastota_1	chasti_rechi_1	claster_1
Компани	1	2,777778	NOUN	0
отлично	1	7,407407	ADVB	0
чуткое	1	3,703704	VERB	0
высоте	1	2,777778	NOUN	0

Рис. 2. Пример обрабатываемых данных для статистического анализа

На рис. 2 показан пример характеристик для анализа текстовых данных. Каждое слово в строке помечено кластером, единый кластер относится к одному смысловому тексту. Колонка «col» показывает сколько раз использовалась часть речи в конкретном тексте. Количество таких данных только по модулю «Настоящие тексты» представляет 48127 значений, для остальных модулей 16679 и 17593 соответственно. Чтобы увеличить время обработки и представления данных в итоговом виде как на рис. 2 была реализована программа для расчета основных показателей и анализа ядра. Программа подготовлена на языке Python и имеет конструктив в виде nltk, pandas, stopwords, word_tokenize, MorphAnalyzer. После привидения текстовых данных к табличному виду как показано на рисунке 2, далее необходимо подготовить еще одну программу для расчета корреляции между входными данными. На рис. 3 продемонстрированы слова, которые больше всего коррелируют, также показана частота появления слова в отзыве и связь с количеством слов.

Поиск уникальных характеристик и кластеризация

Благодаря определению коэффициентов Спирмена и Пирсона были сформированы отдельные текстовые выдержки, где собраны количественные данные [13]. Здесь они не показаны, т. к. имеют еще более сложную структуру. Так колонки частота и кластеры обозначенных текстов, используемых в отдельном для этого тексте, показали, как монотонную зависимость, так и линейную корреляцию в одинаков исследуемых данных. Что по-

зволяет собрать достаточно мощную первоначальную картину объяснения признаков и аномалий для определения фейковых рассылок, текстов и информации, публикуемой злоумышленниками.

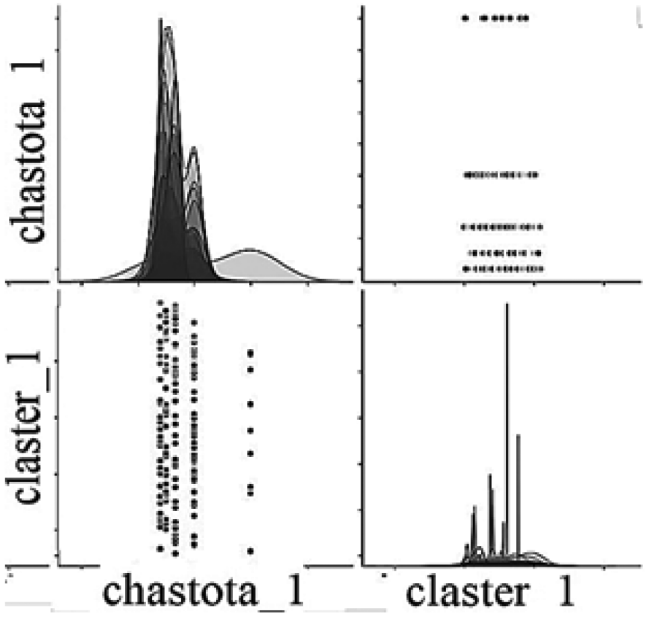

Рис. 3. Графики корреляции частоты и идентификатора кластера

Были выделены структуры, которые имеют больший вес для последующего исследования. После проведенного поиска зависимостей и получения новых данных для трех модулей проведен анализ в направлении применения автокорреляции этих модулей. Результаты, описание, а также категории показаны на рис. 4 и в таблице 2.

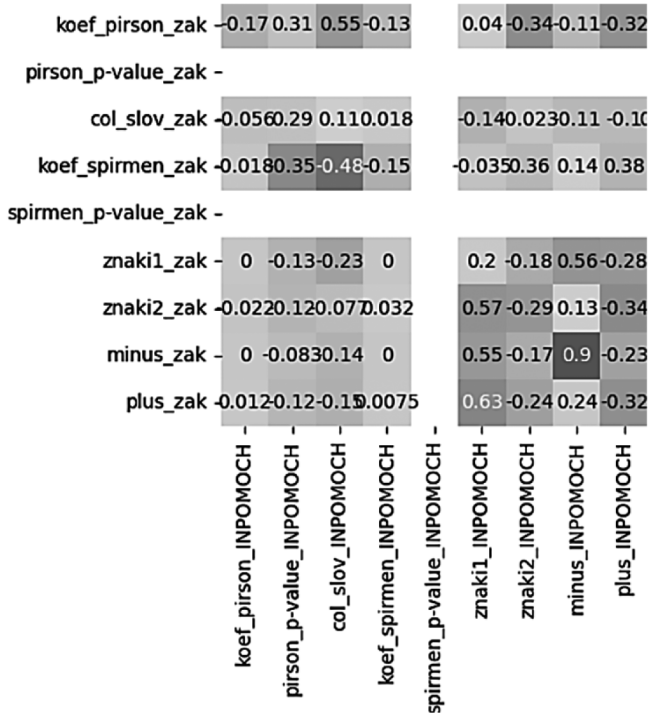

Рис. 4. График автокорреляции модулей

Результаты указывают на наличие внутренних зависимостей по всем определенным признакам. Автокорреляция, проведенная над числовыми значениями для модулей, показала признаки с корреляцией и позволила определить составляющие данных для качественного обучения модели, рис. 4. Выбраны характеристики, которые при анализе автокорреляции имели высокое значение и соответственно большую связь между ними, следовательно такие данные позволяют определить сценарий фейкового отзыва или рассылки и т.д. с точностью в 99 %. Данная схема в таблице 2 строиться для каждого сценария угрозы подаваемого на вход текста.

Таблица 2.

Исследование сценариев и паттернов

1	Корреляционный анализ	
1.1	Модель данных «Настоящие тексты»	
	Отрицательная корреляция	Выявлены отрицательные корреляции между коэффициентами Пирсона и Спирмена и некоторыми признаками.
	Влияние количества слов	Наблюдается влияние количества слов и некоторых эмоциональных показателей на коэффициенты корреляции.
	Отсутствие корреляции	–
1.2	Модель данных «Фейковые тексты»	
	Отрицательная корреляция	Наличие сильной отрицательной корреляции между коэффициентом Пирсона и количеством слов, указывающее на влияние длины текста на его характеристики.
	Влияние количества слов	–
	Отсутствие корреляции	–
1.3	Модель данных «генеративные тексты»	
	Отрицательная корреляция	Наблюдается сильная отрицательная корреляция между коэффициентами Пирсона и Спирмена с некоторыми признаками, что указывает на взаимосвязь между ними.
	Отсутствие корреляции	Отсутствие явных корреляций с показателями, такими как количество слов и знаки препинания.
	Влияние количества слов	–
2	Автокорреляция	
2.1	Модель данных «Настоящие тексты»	
	Внутренние зависимости	Внутренние зависимости, выявленные с использованием автокорреляции, могут предоставить дополнительные сведения о структуре отзывов.
2.2	Модель данных «Фейковые тексты»	

	Внутренние зависимости	Возможные внутренние закономерности, выявленные методом автокорреляции, могут служить основой для дальнейшего анализа и идентификации особенностей данного модуля.
2.3	Модель данных «генеративные тексты»	
	Внутренние зависимости	Наблюдаются внутренние зависимости между некоторыми признаками, указывает на наличие паттернов или структур внутри модуля.
3	Обнаружение аномалий и фейковых текстов	
3.1	Аномалии	Высокие значения коэффициентов корреляции.
		Необычное количество слов.
3.2	Фейковые тексты	Модели обнаружения аномалий выделяют информацию с необычными характеристиками.
		Сценарии фейковых текстов включают в себя попытки манипуляции структурой данных, изменение количества слов и корреляций между признаками.

Для выявления аномалий и фейковых текстов необходимо учитывать не только схему построения, но и внутренние взаимосвязи между характеристиками и их зависимостями. Дальнейший анализ и определение сценариев фейковых текстов требуют дополнительных исследований и экспертного анализа, для этого необходимо задействовать алгоритмы машинного обучения и структурное программирование [14].

Результаты

Применение алгоритмов машинного обучения в новой интерпретации, созданной на основе соединения нескольких алгоритмов, позволяет определить инновационную стратегию поиска аномалий и скрытых признаков фейковых текстов. Для такого поиска потребуются исследовательские данные из п. Материалы и методы, алгоритмы, которые предполагают анализ признаков, имеющие сильные отличия в своих корреляционных коэффициентах или количестве слов, также изолированные составляющие аномалии в тексте. Чтобы такие аномалии определить с высокой точностью необходимо рассмотреть аномальные тексты, которые имеют высокие значения коэффициентов корреляции.

Недостаточно просто идентифицировать признак, следует прописать в алгоритме обнаружения сценарий, позволяющий описать характеристики таких текстов. Используя отклонения, которые были исследованы в прошлом пункте и показаны на рис. 3 и 4 можно найти определенные значения на основе аномалий для таких сценариев.

Например, количество восклицательных знаков — сценарий вида:

- Провокационных отзывов, которые содержат восклицательные знаки для придания эмоциональной насыщенности.
- Фейковых отзывов для управления репутацией, использующие много восклицательных знаков для подчеркивания положительных моментов или отрицательных если присутствуют знаки отрицания.

Данный пример показывает, как один из признаков, как количество знаков препинания дает возможность приблизиться к реальному сценарию текста, при котором он создавался. В программе для автоматического выявления аномальных или фейковых текстов используется статическая модель машинного обучения, которая использовала более 70000 признаков и зависимостей.

Недостоверность метрик успеха приводит к искажению результатов метрик, таких как уровень удовлетворенности клиентов или конверсия. Организация недооценивает проблемы, с которыми сталкиваются реальные клиенты, и упускать возможности улучшения. Также потеря рыночной доли на прямую зависит от качества результатов, выполненных алгоритмами машинного обучения на основе таких недостоверных данных, а именно упущенные возможности улучшения. То же самое касается рекомендательных систем, которые обучаются на предпочтениях пользователей, следовательно, не адекватное построение данных для матрицы алгоритмов приведет к неверным результатам [15].

Для управления безопасностью и выявления аномальных или фейковых текстов и данных аналитики, организации могут использовать программное обеспечение, которое реализует защиту, мониторинг, поиск изменений в составе данных на основе методов обнаружения с помощью ансамблевых nlp-алгоритмов [16]. Ниже показан фрагмент программы, который входит в комплекс ПО для автоматического определения аномалий в тексте.

Из данного, рис. 5, показан фрагмент ПО проекта можно увидеть модели заключенные в ансамблевый вид, также оценку по трем метрикам Precision, Recall и т. д.

Все три модели (обнаружение аномалий, обнаружение фейковых текстов и их ансамбль результатов) проявили выдающуюся производительность с точностью и полнотой до 99 %, показано на рис. 6. Это свидетельствует об их способности эффективно выявлять аномалии и фейковые информационные тексты в представленных данных. Такие высокие показатели точности делают эти модели надежными инструментами для защиты.

Модели анализирует, используя методы NLP для выделения характеристик текста и алгоритмы машинного

```
# Оценка модели обнаружения аномалий (Random Forest для аномалий)
anomaly_predictions = best_anomaly_model.predict(anomaly_data)
print("Anomaly Detection Model Results:")
print(classification_report(df['IsFake'], anomaly_predictions))
anomaly_accuracy = accuracy_score(df['IsFake'], anomaly_predictions)
print("\nAnomaly Detection Model Accuracy:", anomaly_accuracy)
```

Рис. 5. Программный код на Python, обученные ансамблевые модели, построенные на матрице признаков, оценка точности

обучения для матрицы признаков аномалий. Если модели обнаруживают текст, который сильно выбивается из статистики, содержит аномальные языковые конструкции или имеет экстремальное количество значений, то они помечают его как потенциально фейковый. Команда безопасности может затем проверить эти помеченные тексты для принятия дополнительных мер.

```
26              znaki1_zak    0.025687
27              znaki2_zak    0.039675
28               minus_zak    0.052211
29                plus_zak    0.013519

Anomaly Prediction: [ True]
Fake Prediction: [ True]
Ensemble Prediction: [ True]
```

Рис. 6. Результаты работы программы автоматического определения аномалий в тексте

Реальные тексты обычно более сбалансированы и содержат детали, но также исследование проведенное в п. Материалы и методы на рисунке 1 показало, что признаки, определяющие зависимости между фейковыми текстами (почтовыми сообщениями и т. д.) и стадией определения такого отзыва сильно размыты. Поэтому существует необходимость внести корректировки в период тестирования программы, а именно реализовать поиск сценария создания текстовых данных. Для этого были введены признаки, которые характеризуют каждый сценарий. Для проверки обученных моделей был подан на вход текст с наибольшими признаками классифицирующиеся, как аномальные. Результаты важности признаков для моделей обнаружения аномалий и фейковых текстов представлены на рис. 6. В «Anomaly detection model feature importance», «Fake review detection model feature importance» и «Ensemble model accuracy» наибольший вклад вносят признаки, связанные с эмоциональной окраской и структурой текста, такие как vopros_znaki, voskl_znaki, emostia_minus, emostia_plus и др. Обе модели и ансамбль предсказывают, что предоставленные данные соответствуют как сценарию аномалий, так и фейку что и требовалось доказать.

Общий результат проверок моделей указывает на высокую точность классификации, показано на рис. 7. Метрики precision, recall и f1-score для каждого класса демонстрируют высокие значения, что свидетельствует о хорошем балансе между точностью и полнотой классификации. Значение accuracy также довольно высоко, составляя 0.976657329598506, это указывает на общую эффективность модели.

Макроусредненные и взвешенные средние метрик подтверждают высокую производительность модели в целом. Взвешенное среднее учитывает различное количество экземпляров в каждом классе, а макроусредненные значения дают общую картину производительности модели, усредняя результаты для каждого класса без учета их размера.

	precision	recall	f1-score
0	0.96	1.00	0.98
1	1.00	0.94	0.97
accuracy			0.98
macro avg	0.98	0.97	0.98
weighted avg	0.98	0.98	0.98

```
Accuracy: 0.976657329598506
```

Рис. 7. Проверка модели по стандартным метрикам

Интеграция системы автоматического обнаружения в КИС для последующей работы в организации может быть реализована с использованием библиотек pandas и SQLAlchemy для эффективной обработки и хранения данных. Автоматизация обновления данных должна быть организована с использованием Celery, а мониторинг и уведомления реализованы с использованием Grafana или ELK Stack. Также, интеграция с существующими приложениями рассматривается через REST API, используя Flask, Django и др. Аспекты безопасности обеспечиваются с помощью SSL и библиотеками для аутентификации и авторизации, такими как Flask-Security. Тестирование системы включает как модульные, так и интеграционные тесты, реализуемые с использованием pytest. Для поддержки предусмотрены каналы обратной связи, создание подробной документации для пользователя и администратора системы.

Исследование аналогов

Одним из рассмотренных аналогов проекта «Разработка системы ПО для обеспечения безопасности на основе автоматического выявления аномалий и фейковых текстов с использованием алгоритмов машинного обучения», является система анализа аномального поведения пользователей на основе данных о перемещениях и текстовых данных [17]. Данное исследование также ис-

пользует методы машинного обучения и алгоритмы анализа данных для выявления отклонений от нормального поведения пользователей. В отличие от системы анализа пользователей, проект поиска аномалий в текстовой информации фокусируется на анализе текстовых данных с целью выявления фейковых или манипулированных текстов. Проект предлагает инновационный подход, который не только обнаруживает аномалии в текстах, но и способен определять скрытые коррупционные схемы, предотвращать распространение дезинформации и предоставлять эффективные меры по поиску утечек конфиденциальной информации. Кроме того, разработанный проект способен анализировать разнообразные источники текстовой информации, включая веб-сайты, социальные медиа и онлайн-платформы. Это позволяет более широко охватывать угрозы в цифровой среде, такие как фишинговые атаки, дезинформация и утечка данных, что делает проект более универсальным и эффективным инструментом для обеспечения информационной безопасности.

Дополнительным преимуществом разработанного проекта является его способность к более глубокому анализу текстовых данных и выявлению характерных особенностей, свойственных аномальным или манипулированным текстам. Данная система использует современные технологии и алгоритмы машинного обучения для обнаружения структурных аномалий, специфических лингвистических приемов и других атипичных черт, которые могут быть связаны с намеренным искажением информации. Это позволяет обученным моделям выявлять угрозы с более высокой точностью и надежностью, чем аналогичные методики, описанные здесь [18]. Также разработка предоставляет комплексный механизм для управления обнаруженными угрозами, что делает его не только инструментом для выявления аномалий, но и эффективным средством для управления их последствиями. Это важно для компаний и организаций, которые сталкиваются с различными угрозами в текстовых данных и нуждаются в инструменте, способном не только идентифицировать эти угрозы, но и принимать меры по их управлению.

Заключение

Разработанная методика, основанная на объединении алгоритмов машинного обучения для определения сценариев информационной текстовой составляющей, демонстрирует высокую эффективность в повышении вероятности точного обнаружения фейка. Использование данной разработки позволяет системе более точно определить особенности, характерные для неправдоподобных сообщений, и выявить их сценарии с высокой степенью достоверности. Анализ сценариев по предшествующей имитационной модели, является ключевым элементом в применении этой методики. Путем внедрения дополнительных сценариев, которые описывают характеристики фейковых отзывов, алгоритм становится более чувствительным и адаптированным к разнообразным техникам создания фейка. Разработка предоставляет комплексный механизм для управления обнаруженными угрозами, что делает его не только инструментом для выявления аномалий, но и эффективным средством для управления их последствиями. Это важное решение для компаний и организаций, которые сталкиваются с различными угрозами в текстовых данных и нуждаются в инструменте, способном не только идентифицировать эти угрозы, но и принимать меры по их управлению. Применение всех этих аспектов позволяет более эффективно выявлять и анализировать сущности, указывающие на фейковый характер текста, такие как структурные аномалии, использование определенных лингвистических приемов, иных атипичных черт, которые могут быть связаны с намеренным искажением информации. Программа способна анализировать разнообразные источники текстовой информации, включая веб-сайты, социальные медиа и онлайн-платформы. Следовательно, позволяет более широко охватывать угрозы в цифровой среде, такие как фишинговые атаки, дезинформация и утечка данных, что делает проект более универсальным и эффективным инструментом для обеспечения информационной безопасности.

ЛИТЕРАТУРА

1. Зыков С.В. Семантическая интеграция данных для безопасности и целостности корпоративных систем // Безопасность информационных технологий. — 2009. — №3. — с.16–19.
2. Бутакова М.А., Чернов А.В., Говда А.Н., Верескун В.Д., Карташов О.О. Метод представления знаний для проектирования интеллектуальной системы ситуационного информирования. В: Абрахам А., Ковалев С., Тарасов В., Снасель В., Суханов А. (ред.) Материалы Третьей Международной научной конференции «Интеллектуальные информационные технологии для промышленности» (IITI'18). 2018. Достижения в области интеллектуальных систем и вычислений, том 875. — Springer, Cham. стр. 225–235. doi: 10.1007/978-3-030-01821-4_24.
3. Луизи Дж.В. «Прагматичная архитектура предприятия: стратегии преобразования информационных систем в эпоху больших данных», Morgan Kaufmann, 2014. 372 с. ISBN: 9780128005026
4. Исобоев Ш.И., Везарко Д.А., Чечельницкий А.С. Интеллектуальная система мониторинга безопасности сети беспроводной связи на основе машинного обучения. Экономика и качество систем связи. 2022:1;44–48.
5. Бачотти А. Стабильность и управление линейными системами. Cham: Springer, 2019. — 200p. ISBN 978-3-030-02405-5

6. Бурнашев Р.А. и др. Исследования по разработке экспертных систем с использованием искусственного интеллекта //Международная конференция по архитектуре и технологиям информационных систем. — Springer, Cham, 2019. — С. 233–242.
7. Дей Р., Рэй Г., Балас В.Е. Устойчивость и стабилизация линейных и нечетких систем с временной задержкой. Подход с Линейным Матричным Неравенством. Нью-Йорк: Спрингер, 2018. — 274 с. — DOI:10.1007/978-3-319-70149-3 ISBN: 978-3-319-70147-
8. Лин Чжан, Бернард П. Зиглер, Юаньцзюнь Лайли. Разработка моделей для моделирования / Elsevier; 1-е издание — 2019 г. — 453 с.
9. Хинкель Г. NMF: мультиплатформенный фреймворк моделирования //Международная конференция по теории и практике преобразований моделей. — Springer, Cham, 2018. — С. 184–194. — DOI:10.1007/978-3-319-93317-7_10
10. Skillbox Media. Как отличить заказной фейковый отзыв от настоящего: 10 признаков. 10.07.2020. Режим доступа: https://skillbox.ru/media/marketing/kak_otlichit_zakaznoy_feykovyy_otzyv/ [дата обращения: 27.05.2024].
11. Хасти Тревор и Тибширани Роберт. Основы статистического обучения: интеллектуальный анализ данных, логический вывод и прогнозирование. [2-е изд.] — Springer. 2020. — 770 с.
12. Хасти Т., Тибширани Р., Фридман Дж. Элементы статистического обучения. Интеллектуальный анализ данных, логический вывод и прогнозирование. 2-е изд. Springer, 2009. — 745 с.
13. Виттен И.Х., Фрэнк Э., Холл М.А., Пэл К.Дж. Интеллектуальный анализ данных. Практические инструменты и методы машинного обучения. 4-е изд. Elsevier, 2017. 621 с. ISBN: 0120884070
14. Бринк Х. Ричардс Дж. Феверолф М. Машинное обучение в реальном мире. — Санкт-Петербург: Питер, 2017. — 336 с. — ISBN: 978-5-496-02989-6
15. Шолле Ф. Глубокое обучение на Python / СПб.: Питер, 2018. — 400 с.
16. Шелухин О.И. Сетевые аномалии. Обнаружение, локализация, прогнозирование. — М.: Горячая линия — Телеком, 2019. 448 с. ISBN 978-5-9912-0756-0
17. Савенков П.А., Трегубов П.С. Сохранение целостности данных при помощи анализа аномалий в поведенческой деятельности пользователей // Известия ТулГУ. Технические науки. 2021. № 2 с. 45–49
18. Ениколопов С.Н., Ковалёв А.К., Кузнецова Ю.М., Чудова Н.В., Старостина Е.В. Признаки, характерные для текстов, написанных в состоянии фрустрации // Вестник Московского университета. Серия 14. Психология. 2019. № 3. С. 66–85. doi: 10.11621/vsp.2019.03.66

Журнал «Современная наука: актуальные проблемы теории и практики»