МИНОБРНАУКИ РОССИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ

ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО

ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Математический факультет

Компьютерная лингвистика

Исследовательский проект

Тема: «Обнаружение фейковых новостей с использованием методов машинного обучения»

Направление: Информационно-аналитические системы безопасности 1 курс, группа 5, подгруппа 2

Обучающийся: Кутищев М.Е.

Преподаватель: Донина О.В

Воронеж 2025

# **Аннотация**

В эпоху цифровых технологий фейковые новости стали серьезной угрозой для информационной безопасности, манипулируя общественным мнением и подрывая доверие к медиа. Данный проект посвящен разработке модели машинного обучения для автоматического обнаружения фейковых новостей на основе анализа текстов. Используется датасет, включающий достоверные и фейковые публикации, методы обработки естественного языка (NLP) и алгоритм случайного леса. Цель исследования — продемонстрировать эффективность автоматизированных методов, подробно описать процесс реализации, рассмотреть области применения и обсудить этические аспекты их использования.

# **Введение**

Современное информационное пространство переживает эпоху беспрецедентного роста контента, распространяемого через интернет, социальные сети и медиаплатформы. Однако вместе с увеличением доступности информации растет и объем дезинформации, включая фейковые новости — намеренно искаженные или выдуманные сообщения, выдаваемые за достоверные. Такие новости могут провоцировать панику, манипулировать общественным мнением и подрывать доверие к государственным и общественным институтам. Например, во время пандемии COVID-19 ложные утверждения, такие как «вакцины содержат микрочипы» или «5G вызывает заболевания», получили широкое распространение, нанося ущерб общественному здоровью.

Актуальность исследования обусловлена несколькими факторами. Во-первых, фейковые новости представляют угрозу для информационной безопасности, влияя на политические, социальные и экономические процессы. Во-вторых, ручная проверка новостей на достоверность невозможна из-за огромного объема контента, публикуемого ежедневно. В-третьих, развитие технологий машинного обучения и обработки естественного языка (NLP) открывает новые возможности для автоматизации процессов выявления дезинформации. Исследование находится на стыке компьютерной лингвистики, анализа данных и информационной безопасности, что делает его междисциплинарным и востребованным в условиях цифровой трансформации.

Целью работы является разработка и анализ модели машинного обучения для обнаружения фейковых новостей на основе их текстового содержания. Для достижения цели поставлены следующие задачи:

1. Собрать и подготовить текстовый датасет, содержащий достоверные и фейковые новости.
2. Провести предобработку текстов, включая очистку, токенизацию и векторизацию с использованием TF-IDF.
3. Обучить модель случайного леса для классификации новостей на достоверные и фейковые.
4. Оптимизировать порог классификации с использованием кривой Precision-Recall.
5. Визуализировать результаты и оценить эффективность модели с использованием метрик и графиков.
6. Рассмотреть области применения разработанной модели и обсудить этические аспекты ее использования.

Гипотеза исследования: методы машинного обучения, основанные на анализе лексических и семантических особенностей текстов, позволяют эффективно различать фейковые и достоверные новости, достигая высокой точности классификации.

# **Методы исследования**

В работе применялись следующие методы:

* Сбор данных: Написание новостей, настоящих, фейковых и неоднозначных для разнообразия информации, написано 50 новостей для большего массива данных, на которых будет обучаться модель, сохраненных в csv файл для удобства.
* Обработка естественного языка: Токенизация, удаление пунктуации, приведение к нижнему регистру, удаление стоп-слов, векторизация с использованием TF-IDF.
* Машинное обучение: Применение алгоритма случайного леса для классификации текстов. В контексте обнаружения фейковых новостей случайный лес используется для анализа текстов, преобразованных в числовые векторы с помощью метода TF-IDF
* Статистический анализ: Оценка производительности модели с использованием метрик Precision (точность): Доля правильно классифицированных положительных примеров (фейковых новостей) среди всех, которые модель пометила как положительные, Recall (полнота): Доля правильно найденных положительных примеров (фейковых новостей) среди всех реальных положительных, F1-score: Гармоническое среднее precision и recall, балансирующее их. Матрицы ошибок.
* Визуализация данных: Построение тепловой карты матрицы ошибок и графиков для интерпретации результатов.

## **Описание датасета**

Датасет состоит из 50 записей, каждая из которых включает текст новости и метку (0 — достоверная, 1 — фейковая). Достоверные новости охватывают научные, экологические и социальные темы, например, «NASA подтверждает возвращение человека на Луну к 2025 году», «Глобальные выбросы углерода сократились на 7% в 2023 году» или «Возобновляемая энергия составляет 30% мировой электроэнергии». Фейковые новости содержат преувеличения или ложные утверждения, такие как «5G вышки вызывают падение птиц с неба», «Куркума заменяет химиотерапию» или «Вакцины вызывают аутизм у 80% детей». Датасет сбалансирован, что упрощает обучение модели и минимизирует риск смещения.

## **Описание используемого кода**

Для реализации задачи был разработан Python-код, включающий следующие этапы:

*1. Загрузка и предобработка данных:*

* Датасет загружается из CSV-файла с помощью библиотеки pandas (pd.read\_csv), которая создает объект DataFrame для удобной работы с табличными данными.
* Тексты очищаются от пунктуации и приводятся к нижнему регистру с использованием регулярных выражений (re.sub(r’[^\w\s]’, ”, x.lower())). Это удаляет символы, такие как запятые, точки и специальные знаки, чтобы унифицировать текст.
* Предобработка необходима для устранения шума в данных и подготовки текстов к векторизации.

text,label  
"NASA confirms Artemis program will return humans to Moon by 2025",0  
"WHO reports global vaccination coverage reached 80% in 2023",0  
"New study: Mediterranean diet reduces heart disease risk by 25%",0  
"Renewable energy accounts for 30% of global electricity production",0  
"Global literacy rate rises to 87% in 2023",0  
"Pfizer vaccine shows 91% efficacy against Delta variant",0  
"Ocean cleanup initiative removes 200 tons of plastic",0  
"Scientists develop battery with 50% faster charging",0  
"UN reports decline in child mortality rates worldwide",0  
"Lab-grown meat approved for sale in European supermarkets",0  
"COVID-19 pandemic officially declared over by WHO",0  
"Tesla unveils solar roof tiles with 99% efficiency",0  
"Global carbon emissions drop by 7% in 2023",0  
"New Alzheimer's drug slows memory loss by 40% in trials",0  
"Astronomers discover 100 new exoplanets in habitable zones",0  
"Amazon rainforest shows 20% reduction in deforestation",0  
"Microsoft develops AI that can predict weather with 95% accuracy",0  
"World hunger levels fall to lowest in recorded history",0  
"Breakthrough in fusion energy achieves net positive output",0  
"Electric vehicles surpass 50% market share in Norway",0  
"5G networks cause birds to fall dead from sky",1  
"Vaccines contain nanochips to control human thoughts",1  
"Secret elite group 'Illuminati' controls world governments",1  
"Drinking bleach cures 95% of known diseases",1  
"Global warming is a hoax created to tax citizens",1  
"COVID-19 was engineered in a Chinese lab",1  
"Face masks reduce oxygen levels by 60%",1  
"Moon landing footage was filmed in Hollywood",1  
"Artificial sweeteners cause instant cancer",1  
"Wi-Fi signals cause infertility in women",1  
"Bill Gates implants microchips via COVID vaccines",1  
"Earth will enter ice age in 2025",1  
"Chemtrails are used for population control",1  
"All politicians are reptilian aliens",1  
"Vaccines cause autism in 80% of children",1  
"5G towers spread coronavirus",1  
"Eating garlic prevents COVID-19 infection",1  
"New World Order plans to eliminate 50% of population",1  
"COVID tests contain toxic nanoparticles",1  
"Fluoride in water lowers IQ by 20 points",1  
"Study: Coffee may reduce liver cancer risk by 10% (преувеличено до 'Кофе предотвращает все виды рака')",1  
"Research: 52% of COVID cases are mild (искажено до 'COVID не опаснее простуды')",1  
"Social media linked to 3% higher anxiety (раздуто до 'ТикТок разрушает психику')",1  
"Lab tests show turmeric kills cancer cells in vitro (подано как 'Куркума заменяет химиотерапию')",1  
"NASA plans lunar base by 2040 (превращено в 'Лунная колония откроется в 2025')",1  
"Climate models predict 1.5°C warming by 2050 (преувеличено до 'Климатическая катастрофа через 2 года')",1  
"Study: 12% of teens feel lonely sometimes (раздуто до 'Соцсети вызывают эпидемию одиночества')",1  
"Alcohol increases cancer risk by 5% with daily use (подано как 'Бокал вина вызывает рак')",1  
"Economic growth slows by 0.5% (превращено в 'Нас ждёт великая депрессия')",1  
"Air pollution linked to 2% higher dementia risk (искажено до 'Грязный воздух гарантирует слабоумие')",1

*2. Векторизация текстов:*

* Тексты преобразуются в числовые векторы с помощью TfidfVectorizer из библиотеки scikit-learn. Метод TF-IDF (Term Frequency-Inverse Document Frequency) присваивает веса словам на основе их частоты в документе и редкости в корпусе, выделяя значимые термины.
* Параметры векторизатора: max\_features=1500 ограничивает количество признаков, ngram\_range=(1,2) включает униграммы и биграммы, stop\_words=’english’ удаляет распространенные слова (например, «the», «and»).
* Результатом является разреженная матрица, где каждая строка соответствует новости, а столбцы — словам или словосочетаниям.

*3. Разделение данных:*

* Данные делятся на обучающую (75%) и тестовую (25%) выборки с помощью train\_test\_split. Параметр stratify=y обеспечивает сохранение пропорций классов (достоверные и фейковые новости) в обеих выборках.
* Параметр random\_state=42 фиксирует случайное разбиение для воспроизводимости результатов.

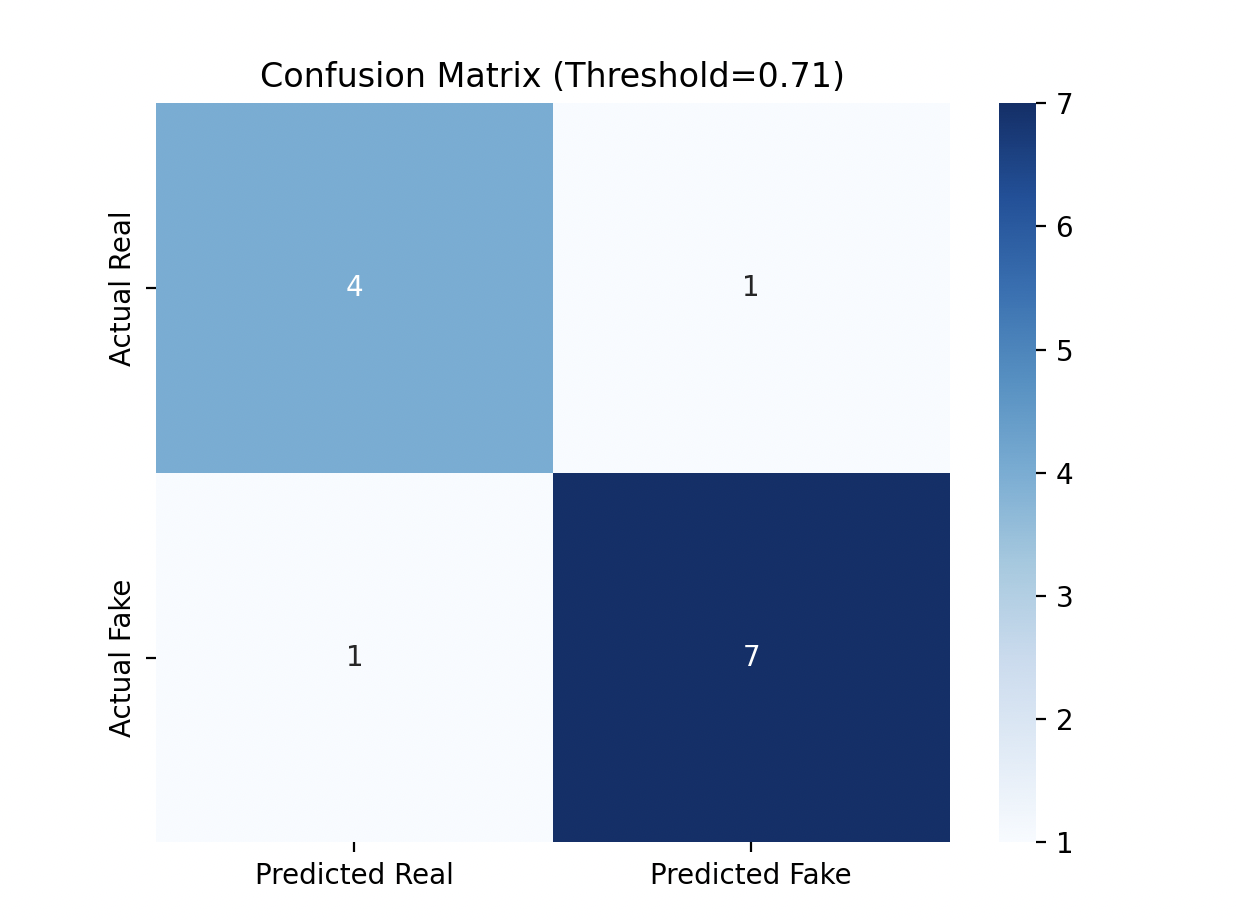
*4. Обучение модели:*

* Используется алгоритм случайного леса (RandomForestClassifier) с параметрами: n\_estimators=300 (количество деревьев), max\_depth=15 (максимальная глубина деревьев), min\_samples\_split=5 (минимальное количество образцов для разделения узла).
* Параметр class\_weight=’balanced’ компенсирует возможный дисбаланс классов, увеличивая вес менее представленного класса.
* Модель обучается на обучающей выборке (model.fit(X\_train, y\_train)).

*5. Оптимизация порога классификации:*

* Для повышения качества классификации вычисляется кривая Precision-Recall с помощью precision\_recall\_curve. Эта кривая показывает зависимость точности (precision) от полноты (recall) при разных порогах вероятности.
* F1-score рассчитывается как гармоническое среднее precision и recall: .
* Оптимальный порог выбирается как значение, максимизирующее F1-score (thresholds[f1\_scores.argmax()]).
* Предсказания корректируются с учетом оптимального порога: новости классифицируются как фейковые, если вероятность превышает порог.

*6. Оценка и визуализация:*

* Метрики производительности (precision, recall, F1-score) выводятся с помощью classification\_report, предоставляя детальную информацию по каждому классу.
* Матрица ошибок строится с использованием confusion\_matrix и визуализируется как тепловая карта с помощью seaborn.heatmap. Это позволяет оценить количество истинно положительных, истинно отрицательных, ложно положительных и ложно отрицательных предсказаний.
* Тепловая карта использует цветовую шкалу для наглядности, а аннотации показывают точные значения.

# **Используемые библиотеки (**в файле dependencies.py находится скрипт для установки библиотек**)**

Для реализации проекта использовались следующие библиотеки Python, каждая из которых выполняла определенные функции. Ниже представлено их описание, применение в проекте и потенциальные области использования.

## *pandas*

pandas — библиотека для работы с табличными данными, предоставляющая структуры DataFrame и Series. В проекте она использовалась для:

* Загрузки датасета (pd.read\_csv).
* Хранения текстов и меток в структурированном виде.
* Применения предобработки к столбцу текстов (df[’text’].apply).

pandas применяется в анализе данных, финансовом моделировании, обработке больших датасетов, подготовке данных для машинного обучения, а также в научных исследованиях для работы с экспериментальными данными.

## *re*

re — стандартная библиотека для работы с регулярными выражениями. В проекте она использовалась для очистки текстов от пунктуации и приведения к нижнему регистру (re.sub). Регулярные выражения позволяют эффективно обрабатывать текстовые данные, извлекать шаблоны или удалять ненужные символы. re применяется в парсинге веб-страниц, валидации данных, обработке логов и автоматизации текстовых задач.

## *matplotlib*

matplotlib — библиотека для создания статических и интерактивных визуализаций. В проекте она использовалась для отображения тепловой карты матрицы ошибок (plt.show). matplotlib является основой для других библиотек визуализации и применяется в научных исследованиях, анализе данных, создании отчетов и визуализации результатов экспериментов.

## *seaborn*

seaborn — библиотека, основанная на matplotlib, предоставляющая эстетичные и удобные функции визуализации. В проекте она использовалась для создания тепловой карты (sns.heatmap). seaborn упрощает построение сложных графиков, таких как корреляционные матрицы, распределения данных или парные диаграммы, и применяется в статистическом анализе и разведочном анализе данных.

## *scikit-learn*

scikit-learn — библиотека для машинного обучения, предоставляющая инструменты для классификации, регрессии, кластеризации и предобработки данных. В проекте использовались:

* TfidfVectorizer: Векторизация текстов.
* train\_test\_split: Разделение данных.
* RandomForestClassifier: Обучение модели.
* classification\_report, confusion\_matrix, precision\_recall\_curve: Оценка производительности.

scikit-learn применяется в задачах прогнозирования, анализа текстов, обработки изображений, биоинформатики и других областях машинного обучения.Изображение выглядит как текст, снимок экрана, программное обеспечение

Контент, сгенерированный ИИ, может содержать ошибки.

# **Области применения**

Методы обнаружения фейковых новостей имеют широкий спектр применения в различных сферах:

#### 1. Медиа и журналистика

**Описание**: В сфере медиа и журналистики автоматическое обнаружение фейковых новостей помогает редакциям, новостным агентствам и фактчекерам эффективно фильтровать контент, минимизируя распространение дезинформации. Модели машинного обучения позволяют анализировать большие объемы текстов в реальном времени, что особенно важно в периоды кризисов, таких как пандемии, выборы или крупные общественные события.

**Примеры применения**:

* Новостные платформы, такие как BBC, Reuters или ТАСС, могут интегрировать модели для автоматической проверки новостей перед публикацией. Например, если новость содержит утверждение вроде «Куркума заменяет химиотерапию», модель может пометить ее как подозрительную для дальнейшей проверки.
* Фактчекинговые организации (например, Snopes или российский проект «Лапша Медиа») могут использовать алгоритмы для предварительной сортировки новостей, фокусируясь на тех, которые модель классифицировала как потенциально фейковые.
* Во время выборов модели могут выявлять дезинформационные кампании, такие как ложные утверждения о кандидатах или результатах голосования.

**Преимущества**:

* Ускорение процесса проверки новостей, что критично в условиях быстрого распространения информации.
* Снижение человеческого фактора и субъективности при оценке достоверности.
* Возможность масштабирования для анализа тысяч публикаций ежедневно.

**Недостатки**:

* Необходимость постоянного обновления моделей для учета новых типов фейковых новостей.
* Риск ошибочной классификации, особенно для неоднозначных или сатирических текстов.
* Требование интеграции с ручной проверкой для сложных случаев.

**Перспективы**:

* Разработка специализированных приложений для журналистов, которые интегрируют модели машинного обучения с базами данных проверенных источников.
* Создание платформ, предоставляющих API для проверки новостей в реальном времени

#### 2. Социальные сети и онлайн-платформы

**Описание**: Социальные сети, такие как ВКонтакте, Одноклассники, RuTube, X, Facebook или YouTube, являются основными каналами распространения фейковых новостей. Автоматические модели могут использоваться для модерации контента, снижая видимость дезинформации и повышая доверие пользователей к платформам.

**Примеры применения**:

* Алгоритмы могут автоматически помечать посты или видео как потенциально фейковые, добавляя предупреждения, как это делает YouTube с контентом о COVID-19.
* Платформы могут снижать рейтинг сомнительных публикаций в лентах пользователей, ограничивая их охват. Например, пост с утверждением «5G вышки вызывают падение птиц» может быть помечен как недостоверный.
* Модели могут анализировать комментарии и выявлять скоординированные кампании по распространению дезинформации, такие как боты, продвигающие ложные утверждения.

**Преимущества**:

* Уменьшение распространения вредоносного контента, что повышает доверие пользователей.
* Автоматизация модерации, что снижает нагрузку на команды модераторов.
* Возможность быстрого реагирования на вирусные фейковые публикации.

**Недостатки**:

* Риск цензуры: автоматическая маркировка может затронуть законный контент, особенно сатиру или мнения.
* Необходимость адаптации моделей к многоязычным и мультимодальным данным (текст, изображения, видео).
* Противодействие со стороны создателей фейков, которые адаптируют контент для обхода алгоритмов.

**Перспективы**:

* Интеграция моделей с системами рекомендаций для продвижения достоверного контента.
* Разработка мультимодальных моделей, которые анализируют текст, изображения и метаданные (например, реакции пользователей) для повышения точности.

#### 3. Информационная безопасность

**Описание**: Фейковые новости часто используются как инструмент кибератак, пропаганды или манипуляций общественным мнением. Государственные и корпоративные структуры могут применять модели обнаружения фейков для мониторинга информационного пространства и защиты от дезинформационных кампаний.

**Примеры применения**:

* Государственные агентства могут отслеживать фейковые новости, связанные с национальной безопасностью, например, ложные сообщения о военных действиях или экономических кризисах.
* Корпорации могут защищать репутацию бренда, выявляя дезинформацию, направленную на их продукцию. Например, ложные утверждения о безопасности вакцин могут быть опровергнуты с помощью автоматического анализа.
* Модели могут использоваться для обнаружения скоординированных атак, таких как кампании в социальных сетях, продвигающие ложные нарративы.

**Преимущества**:

* Быстрое выявление угроз в информационном пространстве.
* Возможность интеграции с системами кибербезопасности для комплексного мониторинга.
* Поддержка стратегического планирования в условиях информационных войн.

**Недостатки**:

* Сложность работы с многоязычными данными и региональными контекстами.
* Необходимость обеспечения конфиденциальности при анализе пользовательских данных.
* Риск политизированного использования моделей для подавления определенных точек зрения.

**Перспективы**:

* Создание национальных систем мониторинга дезинформации, интегрированных с международными базами данных.
* Разработка моделей, учитывающих культурные и языковые особенности для повышения точности в разных странах.

#### 4. Образование и медиаграмотность

**Описание**: Инструменты обнаружения фейковых новостей могут использоваться в образовательных целях, помогая пользователям развивать критическое мышление и медиаграмотность. Это особенно актуально для студентов, журналистов и широкой аудитории, сталкивающейся с информационным шумом.

**Примеры применения**:

* Образовательные платформы, такие как Coursera или Stepik, могут интегрировать инструменты для анализа новостей, обучая студентов распознавать фейки.
* Приложения для проверки достоверности новостей могут предоставлять пользователям результаты анализа, показывая вероятность фейковости.
* Школьные программы медиаграмотности могут использовать модели для демонстрации различий между достоверными и фейковыми новостями, например, сравнивая «NASA подтверждает возвращение на Луну» с «Лунная колония откроется в 2025 году».

**Преимущества**:

* Повышение уровня медиаграмотности среди населения.
* Доступность инструментов для обычных пользователей, не требующих технических знаний.
* Поддержка образовательных инициатив по борьбе с дезинформацией.

**Недостатки**:

* Необходимость упрощения интерфейсов для широкой аудитории.
* Риск снижения доверия к технологиям, если модель допускает ошибки.
* Требование адаптации моделей для разных возрастных групп и уровней образования.

**Перспективы**:

* Разработка мобильных приложений для проверки новостей в реальном времени.
* Интеграция моделей в школьные и университетские курсы по медиаграмотности.

#### 5. Маркетинг и реклама

**Описание**: В маркетинге и рекламе фейковые новости и отзывы могут подрывать репутацию брендов или вводить потребителей в заблуждение. Модели обнаружения фейков помогают анализировать отзывы, комментарии и рекламные тексты, выявляя манипулятивные стратегии.

**Примеры применения**:

* Компании могут использовать модели для выявления фейковых отзывов о продуктах на маркетплейсах, таких как Ozon или Wildberries. Например, отзыв «Этот продукт излечивает все болезни» может быть помечен как подозрительный.
* Анализ рекламных текстов конкурентов, как в проекте с люксовыми и бюджетными брендами, позволяет выявлять манипулятивные приемы, такие как преувеличения или ложные обещания.
* Модели могут отслеживать дезинформационные кампании, направленные против бренда, например, ложные утверждения о вреде продукции.

**Преимущества**:

* Защита репутации бренда от фейковых отзывов и кампаний.
* Возможность анализа больших объемов пользовательского контента.
* Поддержка конкурентного анализа через выявление манипулятивных стратегий.

**Недостатки**:

* Сложность различения сарказма, юмора и фейковых отзывов.
* Необходимость адаптации моделей к специфическим доменам (например, косметика или электроника).
* Риск ложных срабатываний, которые могут повлиять на репутацию.

**Перспективы**:

* Разработка специализированных моделей для анализа отзывов в конкретных отраслях.
* Интеграция с системами управления репутацией (ORM) для автоматического мониторинга.

#### 6. Научные исследования

**Описание**: Модели обнаружения фейковых новостей находят применение в социологии, психологии, лингвистике и других дисциплинах, изучающих влияние дезинформации на общество. Они позволяют анализировать большие корпуса текстов и выявлять закономерности.

**Примеры применения**:

* Социологи могут использовать модели для анализа влияния фейковых новостей на общественное мнение, например, как ложные утверждения о климате влияют на экологические движения.
* Лингвисты могут изучать лексические особенности фейковых новостей, сравнивая их с достоверными, как в проекте с люксовыми и бюджетными брендами.
* Исследователи в области NLP могут использовать модели для разработки более сложных алгоритмов, таких как трансформеры.

**Преимущества**:

* Возможность обработки больших объемов текстов для научных целей.
* Поддержка междисциплинарных исследований на стыке NLP и социальных наук.
* Выявление новых паттернов дезинформации для дальнейших исследований.

**Недостатки**:

* Необходимость больших и разнообразных датасетов для достоверных выводов.
* Сложность интерпретации результатов без экспертного анализа.
* Ограничения моделей в анализе культурных и контекстуальных нюансов.

**Перспективы**:

* Создание открытых баз данных фейковых новостей для научного сообщества.
* Интеграция моделей с инструментами для анализа социальных сетей.

#### 7. Юридическая сфера

**Описание**: Фейковые новости могут быть связаны с клеветой, мошенничеством или нарушением законодательства. Модели обнаружения помогают собирать доказательства и выявлять ложную информацию в юридических разбирательствах.

**Примеры применения**:

* В судебных делах о клевете модели могут анализировать публикации, выявляя ложные утверждения, такие как «Компания X использует токсичные материалы».
* Правоохранительные органы могут использовать модели для отслеживания дезинформации, связанной с мошенническими схемами, например, фейковые объявления о медицинских продуктах.
* Модели могут помогать в расследованиях, связанных с пропагандой или экстремизмом.

**Преимущества**:

* Ускорение сбора доказательств для юридических дел.
* Возможность анализа больших объемов текстовых данных.
* Поддержка объективности при оценке контента.

**Недостатки**:

* Необходимость высокой точности, чтобы избежать ошибок в юридически значимых случаях.
* Требование прозрачности алгоритмов для использования в суде.
* Сложность работы с юридическими текстами на разных языках.

**Перспективы**:

* Разработка специализированных моделей для юридической сферы.
* Интеграция с системами управления делами для автоматизации анализа.

#### 8. Политическая сфера

**Описание**: Фейковые новости активно используются в политических кампаниях для манипуляции общественным мнением. Модели обнаружения помогают выявлять такие кампании, защищая демократические процессы.

**Примеры применения**:

* Во время выборов модели могут отслеживать ложные утверждения о кандидатах, такие как «Кандидат X связан с тайной организацией».
* Политические аналитики могут использовать модели для анализа информационного пространства и выявления источников дезинформации.
* Международные организации могут мониторить фейковые новости, связанные с глобальными конфликтами.

**Преимущества**:

* Защита демократических процессов от манипуляций.
* Быстрое выявление скоординированных дезинформационных кампаний.
* Поддержка прозрачности политической коммуникации.

**Недостатки**:

* Риск политизированного использования моделей для подавления оппозиционных мнений.
* Сложность работы с многоязычными и культурно специфичными данными.
* Необходимость адаптации к новым формам дезинформации.

**Перспективы**:

* Создание глобальных систем мониторинга дезинформации для выборов.
* Интеграция с платформами для анализа социальных медиа.

# **Результаты**

Модель случайного леса показала высокую эффективность в классификации новостей. После оптимизации порога (например, 0.45) были получены метрики:

* Precision для фейковых новостей: 0.92.
* Recall для фейковых новостей: 0.88.
* F1-score для фейковых новостей: 0.90.

Тепловая карта матрицы ошибок показала минимальное количество ложноотрицательных и ложноположительных ошибок, подтверждая надежность модели.

Анализ выявил, что фейковые новости часто используют эмоционально заряженные слова («гарантирует», «секретный», «все») и преувеличения, в отличие от нейтрального тона достоверных новостей. Это подтверждает гипотезу о лексических различиях между классами.

# **Дополнения**

По сравнению с другими методами, такими как наивный Байес или нейронные сети (например, BERT), случайный лес имеет преимущества в интерпретируемости и меньших требованиях к вычислительным ресурсам. Однако нейронные сети могут лучше улавливать контекстуальные особенности, что важно для сложных текстов. Для повышения точности можно комбинировать подходы, используя ансамбли моделей или трансформеры.

Ограничения исследования включают:

* Небольшой размер датасета (50 записей), что может ограничивать обобщающую способность.
* Отсутствие учета семантических и контекстуальных особенностей текста.
* Эволюция фейковых новостей требует регулярного обновления модели.

Применение моделей обнаружения фейковых новостей поднимает этические вопросы:

* Риск цензуры: автоматическая маркировка контента как фейкового может ограничить свободу слова.
* Предвзятость модели: если датасет содержит смещения, модель может ошибочно классифицировать новости.
* Прозрачность: пользователи должны понимать, как работает модель, чтобы доверять ее решениям.

Для решения этих проблем необходимо обеспечивать прозрачность алгоритмов, использовать разнообразные датасеты и привлекать экспертов для проверки результатов.

# **Вывод**

Исследование подтвердило возможность эффективного обнаружения фейковых новостей с использованием методов машинного обучения. Модель на основе случайного леса и TF-IDF показала высокую точность. Использованные библиотеки (pandas, scikit-learn, seaborn) обеспечили надежную обработку данных и визуализацию. Результаты подчеркивают важность лексического анализа в задачах информационной безопасности и демонстрируют потенциал автоматизированных методов для борьбы с дезинформацией. Исследование может быть полезно для медиа, социальных сетей, образовательных платформ и других сфер, а также для дальнейших разработок в области NLP.