Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И.Ульянова (Ленина)
(СПбГЭТУ «ЛЭТИ»)

Направление	01.04.02 – Прикладная математика и информатика
Программа	Математическое и программное обеспечение вычислительных машин
Факультет	КТИ
Кафедра	МО ЭВМ

К защите допустить

И.о. зав. кафедрой А.А. Лисс

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРА

ТЕМА: РАЗРАБОТКА МОДЕЛИ ОБНАРУЖЕНИЯ НЕДОСТОВЕРНЫХ НОВОСТЕЙ В СОЦИАЛЬНОЙ СЕТИ В ФОРМАТЕ ГРАФА ЗНАНИЙ

Студент		_____ *подпись*	А.А. Головин
Руководитель	д.т.н., профессор	_____ *подпись*	Н.А. Жукова
Консультанты		_____ *подпись*	И.А. Куликов
	к.с.н., доцент	_____ *подпись*	А.С. Курапова
	к.т.н.	_____ *подпись*	М.М. Заславский

Санкт-Петербург

2023

ЗАДАНИЕ

НА ВЫПУСКНУЮ КВАЛИФИКАЦИОННУЮ РАБОТУ

Утверждаю

И.о. зав. кафедрой МО ЭВМ

_____ А.А. Лисс

«___»_____2023 г.

Студент Головин А.А. Группа 7381

Тема работы: Разработка модели обнаружения недостоверных новостей в социальной сети в формате графа знаний

Место выполнения ВКР: СПбГЭТУ «ЛЭТИ», кафедра МО ЭВМ

Исходные данные (технические требования):

Необходимо разработать и реализовать на языке Python модель бинарной классификации новостных статей в социальной сети на достоверные и недостоверные с использованием технологии графов знаний

Содержание ВКР:

Введение, Обзор предметной области, Обзор набора данных, Постановка задачи, Описание метода решения, Исследование свойств решения, Заключение

Перечень отчетных материалов: пояснительная записка, иллюстративный материал.

Дополнительные разделы: анализ социального содержания заказа и социально-политических условий реализации работы.

Дата выдачи задания	Дата представления ВКР к защите
«06» февраля 2023 г.	«30» мая 2023 г.

Студент _____ А.А. Головин

Руководитель д.т.н., профессор _____ Н.А. Жукова

Консультант _____ И.А. Куликов

КАЛЕНДАРНЫЙ ПЛАН ВЫПОЛНЕНИЯ
ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

Утверждаю
И.о. зав. кафедрой МО ЭВМ
_____ А.А. Лисс
«___»_____2023 г.

Студент Головин А.А. Группа 7381
Тема работы: Разработка модели обнаружения недостоверных новостей в социальной сети в формате графа знаний

№ п/п	Наименование работ	Срок выполнения
1	Обзор литературы по теме работы	06.02 – 28.02
2	Обзор предметной области	01.03 – 05.03
3	Обзор набора данных	06.03 – 16.03
4	Постановка задачи	17.03 – 19.03
5	Описание метода решения	20.03 – 24.04
6	Исследование свойств решения	25.04 – 04.05
7	Анализ социального содержания заказа и социально-политических условий реализации работы	05.05 – 08.05
8	Оформление пояснительной записки	09.05 – 13.05
9	Оформление иллюстративного материала	14.05 – 16.05
10	Предзащита	20.05.2023

Студент _____ А.А. Головин

Руководитель д.т.н., профессор _____ Н.А. Жукова

Консультант _____ И.А. Куликов

РЕФЕРАТ

Пояснительная записка 83 стр., 25 рис., 1 табл., 71 ист.

НОВОСТИ, СОЦИАЛЬНЫЕ СЕТИ, ГРАФЫ ЗНАНИЙ, МАШИННОЕ ОБУЧЕНИЕ, BERT, GNN, GAT.

Объектом исследования являются недостоверные новости в социальных сетях.

Предметом исследования являются методы обнаружения недостоверных новостей в формате графа знаний.

Цель работы: разработка и реализация модели бинарной классификации новостных статей в социальной сети на достоверные и недостоверные с использованием технологии графов знаний.

В ходе выполнения данной работы были изучены методы обнаружения недостоверных новостей в социальной сети, а также проведен анализ существующих решений. Описан выбранный набор данных и его представление в виде графа знаний. Разработана модель обнаружения недостоверных новостей в социальной сети в формате графа знаний. Реализован программный комплекс данной модели на языке Python с использованием библиотек PyTorch и PyG. Было выполнено исследование свойств разработанного решения и его сравнение с другими методами. Исследована социальная значимость разработанной модели.

ABSTRACT

In this paper, methods for detecting fake news in a social network were studied, as well as an analysis of existing solutions. The dataset and its representation in the form of a knowledge graph are described. The model for detecting fake news in a social network in the knowledge graph format has been developed. The software of the algorithm was implemented in Python using the PyTorch and PyG libraries. A study of the properties of the developed solution and its comparison with other algorithms was carried out. The social significance of the developed model was investigated.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Развитие интернета, а также социальных сетей позволило быстро и эффективно обмениваться информацией. В настоящее время люди более склонны узнавать новости из социальных сетей, а не из традиционных СМИ [1]. Это привело к широкому распространению недостоверных новостей или, как их принято сейчас называть, «fake news» [2]. «fake news» могут представлять собой как просто низкокачественную или неполною информацию, так и заведомо ложные политически замотивированные новости.

Распространение недостоверных новостей крайне негативно влияет на людей и общество. Проблема особенно актуальна, так как легко ввести людей в заблуждение и заставить поверить в ложную информацию, что способно привести их к фатальному исходу [3]. Поэтому поиск и прогнозирование распространения недостоверных новостей являются актуальными и важными областями исследований [4]. Вовремя обнаруживая и предсказывая распространение «fake news» можно своевременно применять меры по предотвращению их дальнейшего продвижения [5].

Целью данной работы является разработка и реализация модели бинарной классификации новостных статей в социальной сети на достоверные и недостоверные с использованием технологии графов знаний.

Задачи данной работы:

1. Анализ существующих методов решения проблемы обнаружения недостоверных новостей в социальной сети. Выбор подходящей модели.
2. Выбор и анализ подходящего набора данных.
3. Разработка модели выявления недостоверных новостей в социальной сети в формате графа знаний.
4. Реализация программного комплекса для разработанной модели.
5. Исследование полученных результатов.

Объектом исследования являются недостоверные новости в социальных сетях.

Предметом исследования являются методы обнаружения недостоверных новостей в формате графа знаний.

Практическая ценность работы: внедрение разработанной модели позволит своевременно обнаруживать недостоверные новости в социальной сети.

Новизна работы и значимость результатов заключаются в следующем:

- Разработанная модель обнаружения недостоверных новостей в социальной сети построена на основе графовой модели с вниманием GAT, что отличает ее от существующих, в которых важность узлов в графе не учитывается. В отличие от традиционных методов обработки естественного языка, использование данной модели позволяет программному комплексу сохранять значение точности и F1-меры с течением времени и повышает качество обнаружения недостоверных новостей.
- Методика оценки эффективности графовых моделей, обеспечивающая корректное сравнение с другими методами. Исследование проводилось с использованием одного и того же набора данных в двух видах: табличном и предварительно преобразованном к формату графа знаний с помощью технологии RML. Данный подход позволил корректно проанализировать различные методы решения проблемы и оценить эффективность и перспективность графового подхода по сравнению с аналогами.
- Разработанный программный комплекс на основе предложенной модели обеспечил значительное улучшение практического результата по сравнению с аналогами. Полученное значение F1-меры, равное 96%, свидетельствует о высоком качестве и эффективности разработанной модели.

Результаты работы были представлены в виде доклада на конференции НТС-2023 и опубликованы в сборнике [6].

1 Обзор предметной области

1.1 Графы знаний

Граф знаний представляет собой структуру данных, которая используется для моделирования отношений между понятиями, объектами или событиями в области знаний. Это мощный инструмент, который позволяет организовывать и систематизировать информацию, а также улучшать процесс ее поиска и анализа.

Граф знаний – это современная форма представления и хранения знаний. Он состоит из трёх основных компонент: узлов, ребер и меток (рис. 1) [7].

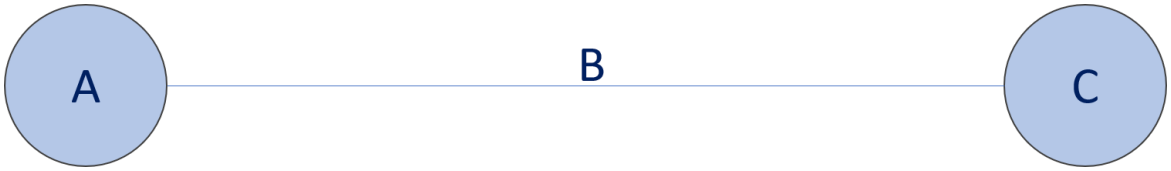

Рисунок 1 – Пример связи предикатом B субъекта A с объектом C

Графы знаний позволяют не только сохранять данные о конкретных сущностях, но и учитывать связи между ними. Благодаря такому подходу графы знаний могут быть использованы как базы знаний, которые содержат огромное количество информации. Одной из ключевых особенностей графов знаний является возможность объединения данных из разных источников в один граф. Таким образом, можно создавать графы знаний на основе «классических» датасетов, которые содержат информацию о различных объектах.

Для успешной интеграции данных в графы знаний существует множество технологий и подходов. Одним из наиболее эффективных и широко используемых является технология RML [8], которая позволяет автоматически генерировать RDF [9] представление графа знаний на основе различных исходных данных.

Исходные данные для интеграции могут быть представлены в различных форматах, таких как реляционные базы данных, CSV, JSON и

XML файлы. С помощью RML можно легко и быстро преобразовать эти данные в RDF формат, который позволяет представить информацию в виде графа знаний и обеспечивает ее более эффективное использование и анализ.

Одним из главных преимуществ технологии RML является ее гибкость и универсальность. Она может быть использована для интеграции данных из различных источников и форматов, что делает ее идеальным инструментом для работы с большими объемами информации.

Для обработки графов знаний можно пользоваться специфическими методами этой структуры. Графы знаний могут иметь векторные представления, существует область машинного обучения на графах (Graph ML), графовые нейронные сети (GNNs) [10].

Технология графов знаний является перспективной и быстро растущей. Её используют Google, Wikipedia и многие другие [11, 12]. Графы знаний могут быть использованы в различных областях, таких как наука, бизнес, медицина и т.д. Они позволяют выявлять скрытые связи и зависимости между объектами и событиями, что помогает принимать более эффективные решения. Кроме того, граф знаний является важным инструментом для разработки искусственного интеллекта и его приложений.

Пример визуализации графа знаний представлен на рис. 2.

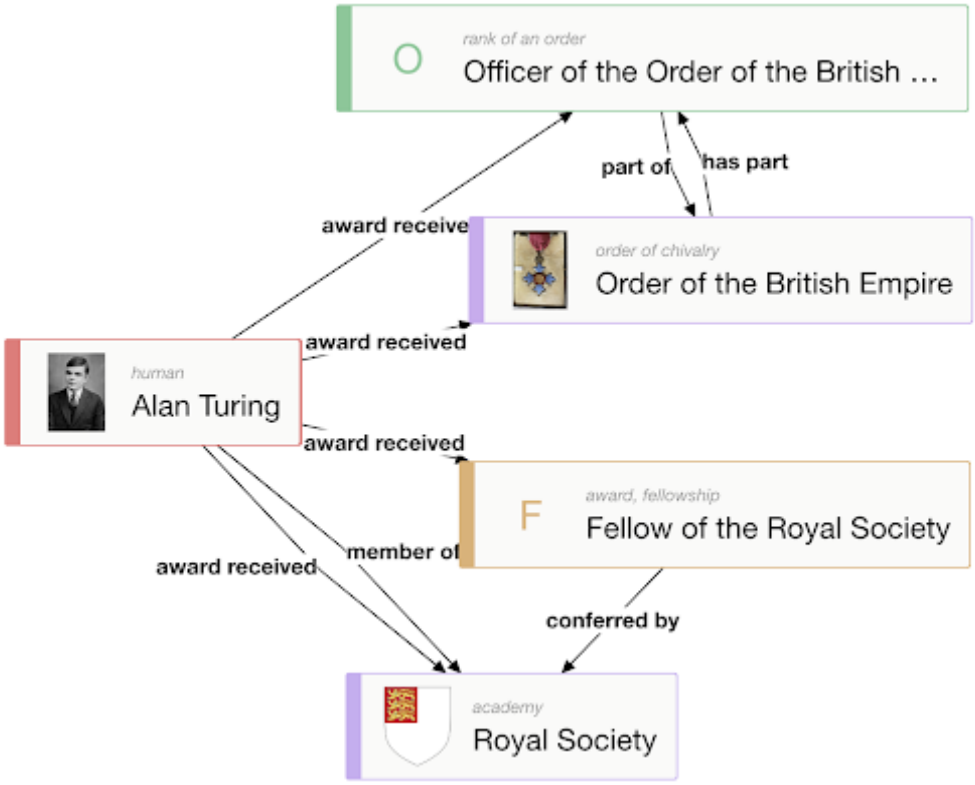

Рисунок 2 – Пример визуализации графа знаний

Среди преимуществ технологии графов знаний можно выделить следующие [13, 14, 15]:

- Возможность интеграции данных из разных источников.
- Удобство поиска информации, хранящейся в графе знаний.
- Возможность анализа связей между сущностями и структуры графа.
- Может использоваться в различных прикладных областях.

В контексте данной работы также перспективно использовать граф знаний в качестве базовой структуры данных для последующей классификации новостных статей на достоверность.

1.2 Бинарная классификация

Бинарная классификация – это задача разделения элементов заданного множества на две группы по какому-то правилу или критерию [16]. Бинарная классификация часто используется в машинном обучении, когда нужно научить модель распознавать объекты или ситуации по набору признаков. Для этого обычно требуется обучающая выборка, то есть набор данных с известными метками классов. Затем модель машинного обучения строит

модель, которая пытается предсказать метку класса для новых примеров. Такая модель называется бинарным классификатором.

В качестве бинарных классификаторов используются различные алгоритмы машинного обучения, такие как логистическая регрессия, деревья решений, методы опорных векторов и нейронные сети.

Бинарная классификация широко используется в различных областях, таких как медицина, финансы, маркетинг и др. Например, в медицине бинарная классификация может использоваться для диагностики заболеваний, а в финансах для кредитного скоринга и принятия инвестиционных решений.

Решаемую проблему обнаружения недостоверных новостей также можно представить как проблему бинарной классификации новостных статей на два класса: достоверные и недостоверные.

1.3 Методы решения

Для достижения цели по решению поставленной прикладной задачи целесообразно начать работу с тщательного изучения и анализа существующих методов, чтобы определить их потенциальную применимость в данном контексте. Оценка различных подходов позволит выбрать наиболее оптимальный способ решения задачи.

1.3.1 Наивный байесовский классификатор

Наивный байесовский классификатор – это модель машинного обучения, основанный на применении теоремы Байеса с допущением о независимости признаков [17]. Теорема Байеса формулируется как (1):

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)}, \quad (1)$$

где:

- $P(h|d)$ – вероятность гипотезы h при данных d, также называемая апостериорной вероятностью.
- $P(d|h)$ – вероятность данных d при условии, что гипотеза h верна.
- $P(h)$ – вероятность того, что гипотеза h верна (независимо от данных), также называемая априорной вероятностью h.

- $P(d)$ – вероятность данных d (независимо от гипотезы).

Для бинарной классификации, когда нужно определить, к какому из двух классов относится объект, наивный байесовский классификатор работает следующим образом:

- Для каждого класса вычисляется априорная вероятность $P(h)$, то есть частота встречаемости этого класса в обучающей выборке.
- Для каждого признака вычисляется условная вероятность $P(d|h)$, то есть частота встречаемости данного признака среди объектов данного класса.
- Для нового объекта с заданным набором признаков d вычисляется апостериорная вероятность каждого класса по формуле Байеса (1), где $P(d)$ – общая вероятность признаков, которая может быть опущена при сравнении классов.
- Выбирается класс с наибольшей апостериорной вероятностью.

Наивный байесовский классификатор имеет ряд преимуществ и недостатков. Среди преимуществ можно выделить его простоту, скорость обучения и удобство применения. Для обучения наивного байесовского классификатора не требуется большой объем обучающих данных, что позволяет использовать его в условиях ограниченного доступа к данным.

Среди главных недостатков можно выделить нереалистичность допущения о независимости признаков, которое может приводить к плохой обобщающей способности на сложных данных. Алгоритм также чувствителен к нулевым вероятностям, когда некоторые признаки не встречаются в обучающей выборке для определенного класса. Это может быть исправлено с помощью сглаживания Лапласа.

Несмотря на недостатки метод широко используется в современных научных исследованиях и разработках в качестве «baseline» модели, с которой в дальнейшем сравниваются предлагаемые авторами методы [18, 19].

1.3.2 Логистическая регрессия

Логистическая регрессия является статистической моделью, которая используется для прогнозирования вероятности возникновения некоторого события на основе одной или нескольких переменных [20]. Алгоритм широко применяется в машинном обучении и статистике для классификации и прогнозирования.

Модель логистической регрессии относится к классу обобщенных линейных моделей, в которых зависимая переменная имеет дискретное распределение. В отличие от линейной регрессии, которая используется для предсказания непрерывных значений, логистическая регрессия используется для предсказания принадлежности определенному классу. Например, она может быть использована для определения вероятности того, что клиент купит продукт или не купит.

Основная идея логистической регрессии заключается в том, что она моделирует логарифм отношения вероятностей принадлежности к определенному классу [21]. Однако, в отличие от линейной регрессии, она использует логистическую функцию (сигмоид), чтобы ограничить значения от 0 до 1 (рис. 3) и преобразовать вероятности в логарифмические шансы (лог-шансы). Эта логистическая функция представлена формулой (2):

$$P = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}}{1+e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}}. \quad (2)$$

После преобразований получим (3):

$$\frac{P}{1-P} = e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}. \quad (3)$$

Прологарифмировав обе части, можно увидеть, что (4):

$$ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k, \quad (4)$$

где:

- P – вероятность принадлежности к классу.
- X_i – входные переменные (предикторы).

- β_i – неизвестные коэффициенты, подлежащие оцениванию по доступным обучающим данным.

Левая часть уравнения (4) называется логитом, или логарифмом риска.

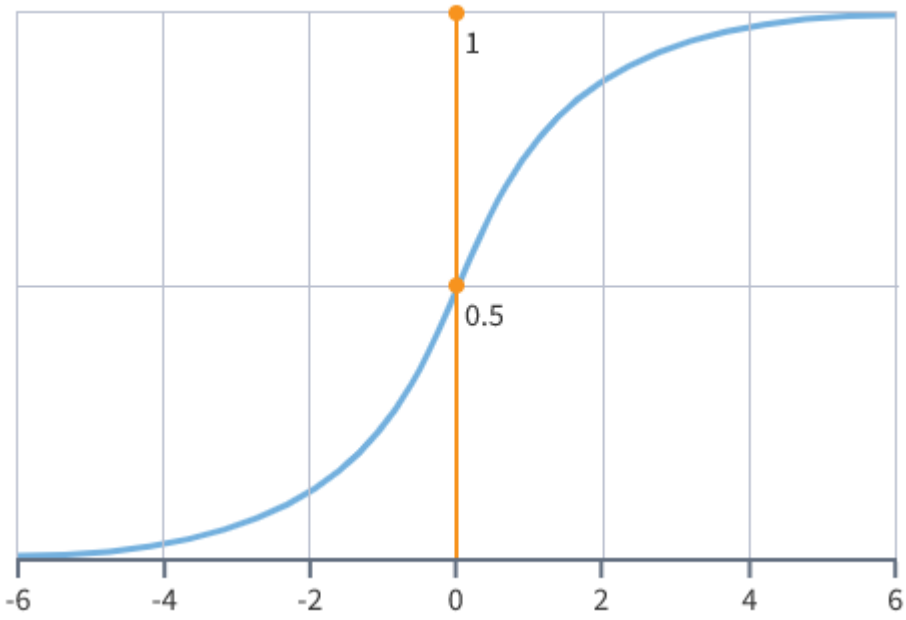

Рисунок 3 – График сигмоидной функции

Логистическая регрессия строит линейную комбинацию независимых переменных (предикторов) и подбирает коэффициенты таким образом, чтобы минимизировать ошибку между наблюдаемыми и предсказанными значениями лог-шансов. Для этого используются различные методы оптимизации, такие как метод максимального правдоподобия и градиентный спуск. Модель может быть применена к различным типам данных, включая числовые, бинарные и категориальные переменные. Коэффициенты регрессии показывают, насколько сильно каждый предиктор влияет на вероятность события. Чем больше коэффициент по модулю, тем больше вклад переменной в модель. Знак коэффициента указывает на направление влияния: положительный знак означает, что увеличение предиктора повышает вероятность события, а отрицательный – что уменьшает.

Логистическая регрессия имеет множество применений в различных областях, таких как маркетинг, медицина, финансы и др. Метод позволяет не

только оценивать вероятности событий, но и проводить классификацию объектов по заданным категориям. Например, логистическая регрессия может помочь определить, является ли электронное письмо спамом или нет, или принадлежит ли изображение к определенному классу (например, кошка или собака).

Таким образом, логистическая регрессия является мощным инструментом для классификации и прогнозирования вероятностей. Модель может быть применена к различным типам данных и имеет множество применений в различных областях [22, 23].

1.3.3 Метод опорных векторов

Метод опорных векторов (SVM) – это модель машинного обучения с учителем, которая используется для решений задач классификации и регрессии. Основная идея метода состоит в том, что для разделения объектов разных классов можно построить гиперплоскость (линию, плоскость или многомерную поверхность), которая максимально отстоит от ближайших к ней объектов (рис. 4). Эти объекты называются опорными векторами, так как они определяют положение и ориентацию гиперплоскости [24].

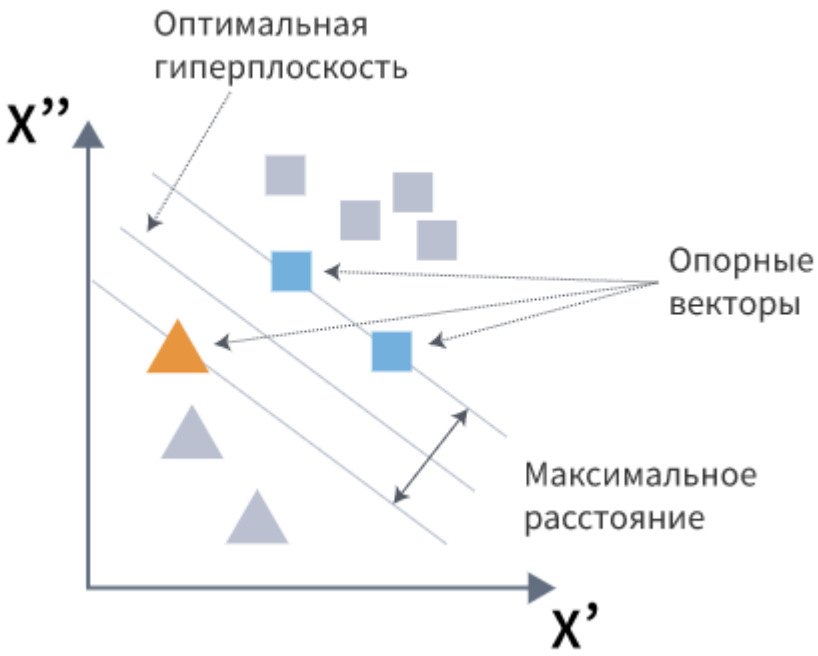

Рисунок 4 – Оптимальная разделяющая гиперплоскость в $\mathbb{R}^2$

Метод опорных векторов имеет ряд преимуществ, таких как:

- Высокая обобщающая способность, то есть способность хорошо работать на новых данных, которые не были использованы при обучении.
- Возможность работать с данными высокой размерности, то есть с большим количеством признаков или переменных.
- Возможность использовать различные ядра (kernel), которые позволяют преобразовывать данные в другое пространство, где они лучше разделимы гиперплоскостью.

Метод опорных векторов также обладает некоторыми недостатками, такими как высокая вычислительная сложность и затраты памяти при работе с большими объемами данных, а также неустойчивость к шуму.

1.3.4 BERT

BERT – это модель глубокого обучения, которая была разработана и выпущена командой исследователей из Google [25, 26]. Она представляет собой одну из самых мощных и эффективных моделей для обработки естественного языка.

BERT использует технологию трансформеров для построения своей архитектуры. Эта технология позволяет модели обрабатывать последовательности данных в более эффективной и точной форме. Трансформеры используют механизм внимания, который позволяет модели фокусироваться на наиболее важных частях входных данных. BERT отличается от других моделей тем, что он двунаправленный, то есть он учитывает контекст с обеих сторон слова при предсказании его значения или роли. Это позволяет BERT лучше понимать смысл и структуру текста.

Одной из ключевых особенностей BERT является то, что он обучается на больших объемах неструктурированных данных. Это позволяет модели получить более широкое понимание естественного языка и улучшить ее способность к обработке сложных задач, таких как ответы на вопросы, классификация текстов, анализ тональности и многое другое.

Перед тем как подать текст на вход нейронной сети, необходимо выполнить его токенизацию [27]. Токены представляют собой слова из словаря или их компоненты. Если слово отсутствует в словаре, то оно разбивается на части, которые уже присутствуют в словаре.

В своей изначальной форме трансформеры включают в себя два отдельных механизма – кодировщик, который считывает введенный текст, и декодировщик, который выдает прогноз для задачи. Поскольку целью BERT является создание языковой модели, необходим только механизм кодировщика.

При обучении языковых моделей возникает сложность определения цели прогнозирования, так как многие модели ограничивают контекстное обучение, предсказывая следующее слово в последовательности. Для решения этой проблемы BERT использует две стратегии обучения: генерацию пропущенного токена и предсказание следующего предложения.

Перед подачей последовательностей слов в BERT, 15% слов в каждой последовательности заменяются на токен «MASK». Затем модель пытается предсказать исходное значение замаскированных слов на основе контекста, создаваемого другими словами в последовательности, которые не были замаскированы. Чтобы предсказать пропущенные слова, необходимо выполнить следующие действия: добавить слой классификации поверх вывода кодировщика, умножить выходные векторы на матрицу вложения и преобразовать их в размерность словаря, а также рассчитать вероятность каждого слова в словаре с помощью softmax.

В отличие от однонаправленных моделей, функция потерь BERT учитывает только предсказание замаскированных значений и игнорирует предсказание не замаскированных слов, что может замедлить сходимость модели, но компенсируется её более высокой осведомленностью о контексте.

Во время обучения, модель BERT использует пары предложений в качестве входных данных и обучается определять, является ли второе предложение в паре следующим предложением в оригинальном документе.

Половину входных данных составляют пары, где второе предложение является последующим в оригинальном документе, а в другой половине в качестве второго предложения выбирается случайное предложение из корпуса. Предполагается, что случайное предложение будет отсоединено от первого.

Чтобы помочь модели различать два предложения при обучении, входные данные обрабатываются следующим образом (рис. 5):

- В начало первого предложения вставляется маркер «CLS», а в конец каждого предложения маркер «SEP».
- Каждому токену добавляется вложение предложения, указывающее на то, относится ли токен к предложению A или предложению B.
- Для каждого токена добавляется позиционное вложение, которое указывает его положение в последовательности.

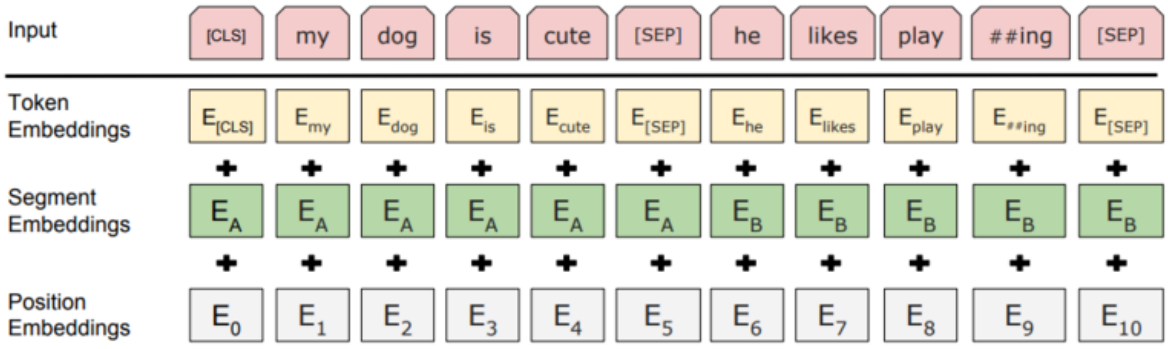

Рисунок 5 – Представление входных данных модели BERT

Для проверки связи второго предложения с первым выполняются следующие действия: сначала входная последовательность проходит через модель трансформера, затем выход токена «CLS» преобразуется в вектор формы 2×1 с помощью простого слоя классификации (используя выученные матрицы весов и смещений), и, наконец, рассчитывается вероятность того, является ли следующим предложением, с помощью функции softmax.

При обучении модели BERT, генерация пропущенного токена и предсказание следующего предложения обучаются вместе с целью минимизирования комбинированной функции потерь двух стратегий.

BERT также имеет возможность предварительного обучения, что позволяет модели получить более высокую точность при работе с новыми задачами. В процессе предварительного обучения модель обрабатывает большие объемы текстовых данных, что позволяет ей научиться распознавать общие паттерны и закономерности в языке.

Можно сделать вывод, что BERT является одной из самых мощных и эффективных моделей для обработки естественного языка. Она имеет широкий спектр применения и может использоваться для решения различных задач, связанных с обработкой текста. Эффективность и точность делают её одной из наиболее востребованных моделей в области машинного обучения и искусственного интеллекта.

На сегодняшний день стало уже классическим использование нейросетевой языковой модели BERT для решения широкого спектра задач обработки естественного языка. BERT нашел применение и для решения проблемы определения достоверности новости [28].

При этом существует целое семейство моделей, являющихся модифицированными версиями оригинального трансформера BERT. Эти алгоритмы настолько популярны, что существует область исследований BERTology, связанная с изучением их внутренней работы и сравнением производительности для разных задач [29].

1.3.5 DistilBERT

DistilBERT представляет собой один из алгоритмов ранее упомянутого семейства нейросетевых архитектур BERT. DistilBERT основан на идее дистилляции [30]. Её суть в имитации легковесной моделью поведения более сложной модели-учителя, в данном случае BERT.

Основная идея модели заключается в том, чтобы уменьшить количество параметров в модели, не ухудшая ее качество и производительность. DistilBERT сохраняет около 95% производительности оригинального BERT на задачах понимания языка, но имеет в 2 раза меньше параметров и обучается на 60% быстрее [31]. Модель также потребляет

меньше памяти. Это делает её более доступной для использования на устройствах с ограниченными ресурсами, таких как мобильные устройства или IoT-устройства.

DistilBERT также имеет свои ограничения. Она не может обрабатывать более сложные задачи, такие как генерация текста или машинный перевод, которые требуют более глубокого понимания языка. Кроме того, DistilBERT не может обучаться на таких больших наборах данных, как BERT, что может привести к снижению ее точности на некоторых задачах.

Таким образом, DistilBERT – это эффективная модель языкового моделирования, которая показывает хорошие результаты на многих задачах и может быть использована в различных приложениях, где требуется анализ текстовых данных.

1.3.6 KG-BERT

Технология графов знаний, в свою очередь, также применяется для улучшения результатов моделей, использующих BERT [32, 33] и позволяет улучшить результаты классификации [34]. Сочетание языковой модели BERT и технологии графов знаний нашло свое применение в алгоритме KG-BERT [33].

Главная причина применения данного подхода заключается в том, что несмотря на то, что предварительно обученные языковые модели, такие как BERT обучаются на огромных текстовых наборах данных, им не хватает предметно-ориентированных знаний. При чтении текста эксперты в предметной области делают выводы, исходя из соответствующих знаний, которыми они обладают. KG-BERT реализовывает эту возможность, добавляя языковой модели (BERT) интеграцию с графом знаний (KG).

В модели KG-BERT триплеты из графа знаний вводятся в предложения как знания из предметной области.

Данная модель значительно превосходит BERT, особенно в задачах, которые тесно связаны с предметной областью (включая медицину, финансы и юриспруденцию), свидетельствуя о том, что KG-BERT является

подходящим выбором для решения проблем, которые требуют привлечения экспертов.

Однако технология KG-BERT также имеет существенную проблему, называемую шумом знаний (KN), заключающейся в слишком большом включении знаний, которое приводит к отклонению предложения от его правильного значения [34].

1.3.7 GNN

Анализ графов является трудоемким процессом, так как граф представлен в неевклидовом пространстве, что затрудняет интерпретацию данных графов в сравнении с другими типами данных, такими как изображения или сигналы временных рядов, которые могут быть преобразованы в двумерное или трехмерное евклидово пространство [35]. Текст также может рассматриваться как временной ряд.

Обычные инструменты машинного и глубокого обучения не подходят для обработки графовых данных, так как они специализируются на простых типах данных с одинаковой структурой и размером. Тем временем, существуют достаточно сложные графы без фиксированной формы и с переменным размером неупорядоченных узлов, имеющих разное количество соседей (рис. 6). Также не помогает тот факт, что существующие модели машинного обучения имеют предположение о том, что экземпляры независимы друг от друга [36]. Это неверно для данных графа, потому что каждый узел связан с другими узлами связями различных типов.

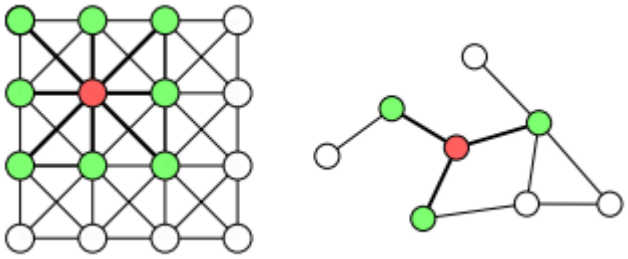

Рисунок 6 – Структуры данных в евклидовом (слева) и в неевклидовом (справа) пространствах

Так как в данной работе рассматриваются данные, представленные в графовом формате, целесообразно рассмотреть модели, разработанные специально для этой задачи.

Таким подходом является использование графовых нейронных сетей. Графовые нейронные сети (GNN) – это тип нейронных сетей, которые способны работать с данными, представленными в виде графов [37].

GNN основаны на идее распространения информации по графу с помощью локальных сообщений между соседними узлами. Каждый узел имеет свой вектор признаков, который обновляется на каждом шаге алгоритма в зависимости от признаков своих соседей. Таким образом, GNN способны агрегировать информацию из разных частей графа и учитывать его структуру. Это позволяет им учитывать контекст и зависимости между объектами в графе.

GNN могут применяться для различных задач на графах, таких как:

- Классификация графов: определение категории или свойства всего графа. Например, определение того, является ли химическое соединение токсичным или нет.
- Классификация узлов: определение категории или свойства отдельных узлов графа. Например, определение того, к какой группе принадлежит пользователь социальной сети или какая роль у слова в предложении.
- Классификация ребер: определение категории или свойства отдельных ребер графа. Например, определение того, какая связь существует между двумя пользователями социальной сети или какое отношение имеют два слова в предложении.
- Генерация графов: создание новых графов на основе заданных критериев или образцов. Например, создание новых химических соединений или текстов.

Существует большое количество различных подвидов графовых нейронных сетей, например:

- Рекуррентные графовые нейронные сети (RecGNN).
- Свёрточные графовые нейронные сети (GCN).
- Графовые нейронные сети с механизмом внимания (GAT).

В целом, GNN являются мощным инструментом для работы с графами и могут использоваться в различных областях науки и технологий. Они позволяют учитывать контекст и зависимости между объектами в графе, что может привести к более точным результатам и более эффективной обработке данных.

1.3.8 RecGNN

Рекуррентные графовые нейронные сети (RecGNN) относятся к ранним работам в области графовых нейронных сетей. Эти методы итеративно применяют одну и ту же параметризированную функцию к значениям узла для извлечения паттернов информации высокого уровня. Информация распространяется по ребрам графа до тех пор, пока не будет достигнуто равновесие [38].

RecGNN строится с учетом теоремы Банаха о неподвижной точке [41]. Если применить отображение T к x k раз, x^k должно быть почти равно x^{k-1}, то есть (5):

$$x^k = T(x^{k-1}), k \in (1, n). \quad (5)$$

Модель RecGNN определяет параметризированную функцию f_w (6):

$$x_n = f_w\big(l_n, l_{co[n]}, x_{ne[n]}, l_{ne[n]}\big), \quad (6)$$

где $l_n, l_{co}, x_{ne}, l_{ne}$ представляют характеристики текущего узла $[n]$, ребра узла $[n]$, состояние соседних узлов и свойства соседних узлов (рис. 7).

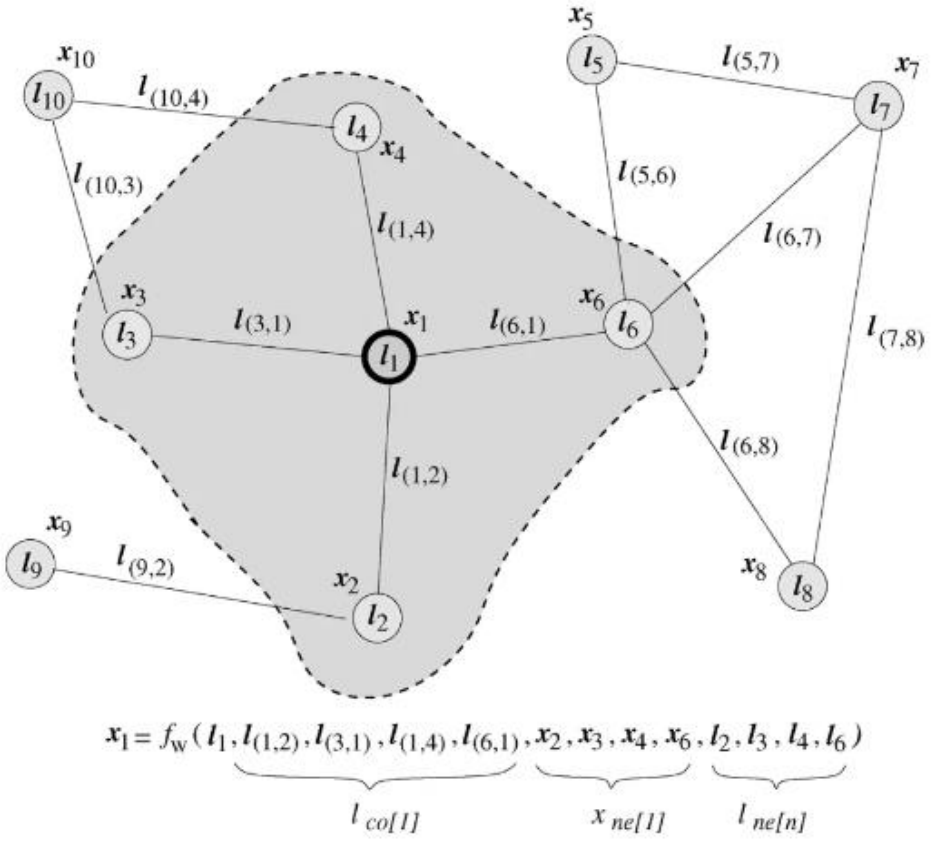

Рисунок 7 – Пример обновления состояния узла на основе информации о его соседях

В заключении, после k итераций конечное состояние узла используется для получения выходных данных для принятия решения о каждом узле. Выходная функция определяется как (7):

$$o_n = g_w(x_n, l_n). \quad (7)$$

Таким образом, архитектура RecGNN – это мощный и гибкий метод глубокого обучения на графах, стоящий у истоков графовых нейронных сетей, который может быть полезен в различных областях. Данный метод активно используется в современных научных работах наряду с более современными архитектурами [40].

1.3.9 GCN

Свёрточные нейронные сети (CNN) часто используются для решения таких задач как классификация изображений, обнаружение объектов. CNN используют операцию свёртки для обработки данных [41]. Свертка – это математическая операция, которая позволяет извлекать признаки из данных,

например, изображений, звуков или текстов. Свёрточные нейронные сети состоят из нескольких слоев, каждый из которых применяет свертку к входным данным и передает результат на следующий слой. Свёрточные нейронные сети имеют различную архитектуру в зависимости от задачи и данных. Однако обычно они состоят из трех основных типов слоев: свёрточных слоев, пулинговых слоев и полносвязных слоев (рис. 8). Свёрточный слой применяет набор фильтров к входным данным и получает карты признаков. Пулинговый слой уменьшает размерность карт признаков путем выбора максимального или среднего значения в небольшом окне. Полносвязный слой соединяет все узлы предыдущего слоя с узлами следующего слоя и выполняет линейную комбинацию входов с весами и смещениями. Последний полносвязный слой обычно имеет столько узлов, сколько классов в задаче классификации, и использует функцию активации softmax для получения вероятностей принадлежности к каждому классу.

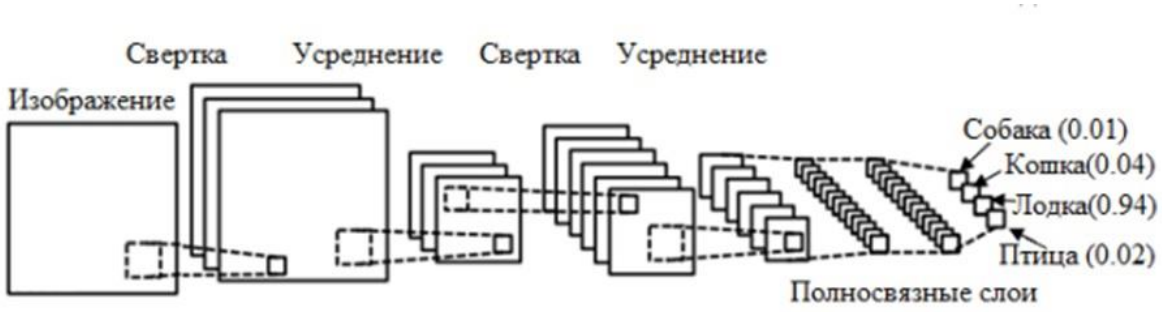

Рисунок 8 – Архитектура свёрточной нейронной сети

Свёрточные нейронные сети могут автоматически находить оптимальные признаки для решения различных задач, таких как распознавание объектов, классификация изображений, сегментация, генерация текста и другие.

Использование CNN для графов очень затруднительно из-за произвольного размера графа и сложной топологии, что означает отсутствие пространственной локальности. Также есть проблема нефиксированного порядка узлов. Например, если сначала узлы были обозначены как «A, B, C, D, E», а во второй раз как «B, D, E, A, C», то входы матрицы в нейросеть

изменятся. Графы инвариантны к порядку узлов, поэтому необходимо получать один и тот же результат независимо от того, как упорядочены узлы.

GCN (Graph Convolutional Networks) – это класс моделей глубокого обучения, которые применяются для анализа графовых структур [42]. Они представляют собой расширение свёрточных нейронных сетей на графы, позволяющее использовать информацию о структуре графа в процессе обучения (рис. 9).

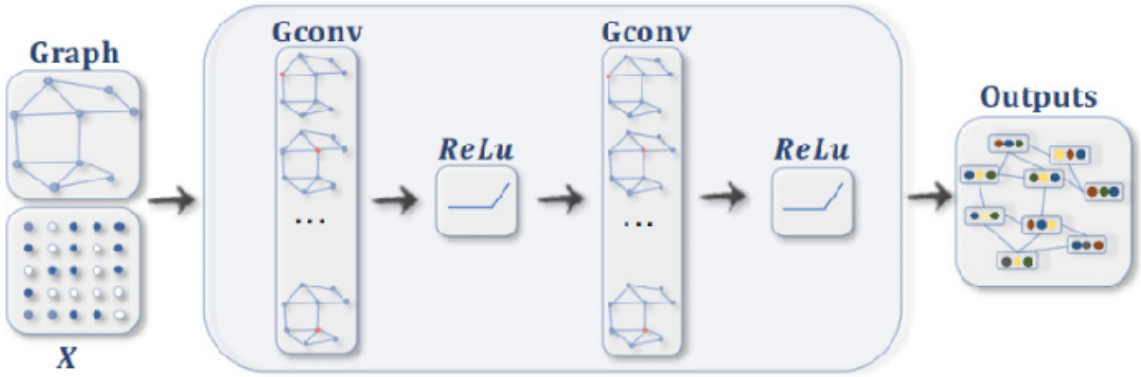

Рисунок 9 – Архитектура графовой свёрточной нейронной сети

Основная идея GCN заключается в том, что каждый узел графа представляется в виде вектора признаков, а связи между узлами – в виде матрицы смежности. Затем производится свертка по матрице смежности, которая позволяет учитывать информацию о соседних узлах при обработке каждого узла. Таким образом, GCN позволяет извлекать информацию о структуре графа и использовать ее для более точного предсказания.

Графовые свёрточные нейронные сети пытаются обобщить свёртки из евклидового пространства для графовых структур. Аналогично RecGNN, GCN вычисляют скрытые состояния путем агрегирования соседних скрытых состояний. Главным отличием от RecGNN является то, что GCN объединяет несколько слоёв свёрток графа для извлечения информации об узлах высокого уровня. Таким образом, они заменяют итерацию фиксированным числом слоёв, несущих разные наборы параметров. GCN различают двух основных видов:

- Спектральные свёрточные графовые нейронные сети (Spectral ConvGNN). В значительной степени полагаются на обработку сигнала графа. Фильтры спектральных GCN можно рассматривать как глобальные фильтры.
- Пространственные свёрточные графовые нейронные сети (Spatial ConvGNN). Фильтры работают непосредственно в локальной окрестности графов. Поэтому они обычно более масштабируемы для больших графов, а также подходят для задач с различными топологиями в ситуациях индуктивного обучения.

Спектральные GCN полагаются на преобразование Фурье графа для выполнения обработки сигналов, путем упрощения и аппроксимации свёртки графа.

Пространственные GCN определяют фильтры непосредственно в окрестности графа, складывая нелинейные функции агрегации, определенные в локальной окрестности узлов. Поскольку топология графа обычно не определяет явный порядок соседей, данные функции агрегации должны быть инвариантными к перестановкам. Обучаемые фильтры соответствуют симметричным вариантам евклидовых ядер. По сравнению со спектральными подходами пространственные GCN избегают глобальной обработки, что делает этот метод масштабируемым даже для больших графов. Кроме того, пространственные подходы подходят для изучения индуктивных моделей, которые можно применять даже к изменяющимся топологиям графа. RecGNN и пространственные GCN разделяют схожие идеи, но их отличает тот факт, что пространственные методы исключают разделение весов и заменяют итеративную обработку фиксированным числом слоев [43].

Все доступные в настоящее время нейронные сети свёрточных графов используют один и тот же формат. Они пытаются обучить функцию для передачи информации об узле и обновления состояния узла с помощью процесса передачи сообщений. Любая графовая нейронная сеть может быть

рассмотрена как нейронная сеть передачи сообщений с функциями передачи сообщений, обновления узла и считывания.

1.3.10 GAT

Graph Attention Networks (GAT) – это графовая нейронная сеть, которая использует механизм внимания для обработки графовых данных. GAT является одной из наиболее эффективных моделей для задач, связанных с графами [44].

Одной из ключевых особенностей GAT является использование механизма внимания, который помогает учитывать важность каждого соседнего узла, присваивая ребрам весовые коэффициенты (рис. 10). Это позволяет более точно моделировать связи между узлами и учитывать их влияние на результаты модели. Принципиально новым для данной архитектуры является использование технологии «самовнимания», для которой можно провести аналогию с важностью узла.

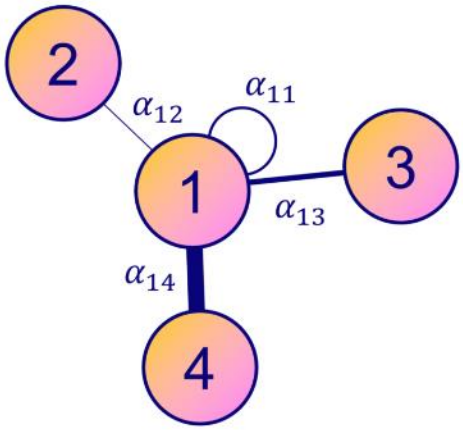

Рисунок 10 – Иллюстрация механизма внимания

Несмотря на то, что графовые сети внимания обучаются дольше, точность у них существенно выше, чем у обычных графовых свёрточных нейросетей.

1.4 Метрики

Метрики бинарной классификации – это инструменты, которые используются для оценки качества моделей машинного обучения, которые

работают с двумя классами. Они позволяют оценить, насколько точно модель может отнести объекты к одному из двух классов [45].

Одной из наиболее распространенных метрик бинарной классификации является точность (accuracy). Она определяется как отношение числа правильно классифицированных объектов к общему числу объектов. Хотя точность может быть полезной метрикой, она может ввести в заблуждение, если классы несбалансированы. Формула для точности (8):

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (8)$$

где:

- TP – истинно положительные прогнозы.
- TN – истинно отрицательные прогнозы.
- FP – ложно положительные прогнозы.
- FN – ложно отрицательные прогнозы.

Для учета несбалансированных классов используются другие метрики, такие как precision и recall.

Precision определяется как отношение числа правильно классифицированных положительных объектов к общему числу положительных объектов. Precision показывает, насколько модель избегает ложных срабатываний, но не учитывает ложные пропуски. Формула для precision (9):

$$precision = \frac{TP}{TP + FP}. \quad (9)$$

Recall определяется как отношение числа правильно классифицированных положительных объектов к общему числу положительных объектов в выборке. Recall показывает, насколько модель избегает ложных пропусков, но не учитывает ложные срабатывания. Формула для recall (10):

$$recall = \frac{TP}{TP + FN}. \quad (10)$$

F1-мера (F1-score) – это среднее гармоническое precision и recall. F1-мера показывает компромисс между $precision$ и $recall$ и является более сбалансированной метрикой, чем каждая из них по отдельности. Формулы для F1-меры (11-12):

$$F1 = 2 \times \frac{precision \times recall}{precision + recall}. \quad (11)$$

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN}. \quad (12)$$

Матрица ошибок (confusion matrix) – это таблица, которая показывает количество правильных и неправильных прогнозов для каждого класса (рис. 11). Матрица ошибок позволяет легко увидеть, какие типы ошибок делает модель и как они влияют на общую производительность.

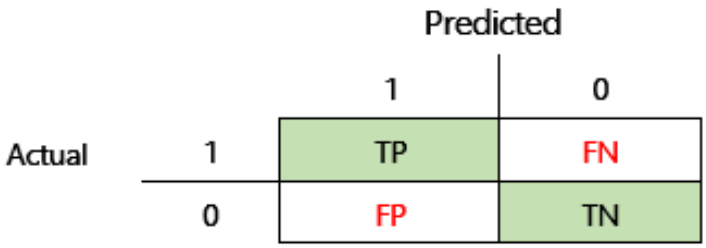

Рисунок 11 – Иллюстрация матрицы ошибок

ROC-кривая – это график, который показывает зависимость между долей истинно положительных прогнозов (TPR) и долей ложно положительных прогнозов (FPR) при различных порогах бинарной классификации. ROC-кривая показывает, насколько хорошо модель разделяет два класса при различных уровнях чувствительности и специфичности. Площадь под кривой ROC (AUC) является метрикой, которая показывает общую производительность модели бинарной классификации независимо от порога. AUC ROC равна вероятности того, что модель правильно отранжирует случайно выбранный положительный объект выше случайно выбранного отрицательного объекта. AUC ROC принимает значения от 0 до 1, где 1 означает идеальную модель, а 0,5 соответствует случайному угадыванию.

Таким образом, можно сделать вывод, что метрики бинарной классификации играют важную роль в оценке качества моделей машинного обучения. Выбор наиболее подходящих метрик зависит от конкретной задачи и требований к модели.

2 Обзор набора данных

В качестве набора данных для решения поставленной задачи выбран датасет «FakeNewsNet» [46]. «FakeNewsNet» создан для исследования распространения фейковых новостей в социальных сетях. Датасет содержит данные о реальных и недостоверных новостях: источники, заголовки и упоминания в социальной сети Twitter. «FakeNewsNet» сформирован на основе информации о новостных публикациях, размеченной журналистами из организаций GossipCop и PolitiFact, являющихся авторитетными в области разоблачения недостоверных новостей [47, 48].

Датасет состоит из нескольких частей. В первой части содержится информация о новостных статьях, такие как заголовки, тексты и ссылки на источники. Вторая часть содержит информацию о пользователях, которые опубликовали статьи, а также о тех, кто сделал репосты или комментарии.

Для каждой новостной статьи в датасете также указывается ее статус – настоящая новость или фейковая. Это позволяет использовать набор данных для обучения моделей машинного обучения для определения недостоверных новостей.

Датасет «FakeNewsNet» представляет собой ценный инструмент для исследования феномена фейковых новостей. Этот набор данных содержит множество примеров статей, которые были признаны недостоверными или ложными, а также соответствующую информацию о том, как они распространялись в социальных сетях. «FakeNewsNet» также содержит различные метаданные, такие как дата публикации статьи и количество поделившихся пользователей в социальных сетях. Эти данные могут быть использованы для анализа тенденций в распространении недостоверных новостей и определения наиболее эффективных методов борьбы с ними.

Данные были предварительно обработаны, очищены и опубликованы на платформе Kaggle [49]. Полученный набор данных представляет собой CSV файл, в котором каждый элемент является публикацией новостной статьи (рис. 11).

	title	news_url	source_domain	tweet_num	real
0	Kandi Burruss Explodes Over Rape Accusation on...	http://toofab.com/2017/05/08/real-housewives-a...	toofab.com	42	1
1	People's Choice Awards 2018: The best red carp...	https://www.today.com/style/see-people-s-choic...	www.today.com	0	1
2	Sophia Bush Sends Sweet Birthday Message to 'O...	https://www.etonline.com/news/220806_sophia_bu...	www.etonline.com	63	1
3	Colombian singer Maluma sparks rumours of inap...	https://www.dailymail.co.uk/news/article-33655...	www.dailymail.co.uk	20	1
4	Gossip Girl 10 Years Later: How Upper East Sid...	https://www.zerchoo.com/entertainment/gossip-g...	www.zerchoo.com	38	1

Рисунок 11 – Фрагмент используемого датасета

В датасете содержится более двадцати тысяч записей. Атрибутами являются: заголовок, ссылка на статью, источник новости, количество пользователей, поделившихся новостью, а также метка достоверности статьи. Эти данные используются для классификации новостных публикаций с целью выявления недостоверных новостей.

Используемый набор данных также преобразован к графовому виду [50], что полезно для дальнейшего использования графовых подходов. В данной версии каждая новость представляется отдельным графом, у которого корневой узел – это сама новость, а последующие узлы – это пользователи Twitter, которые делились друг с другом этой новостью (рис. 12). Кроме того, узел пользователя является родительским для узлов, содержащих текстовую информацию о его профиле в социальной сети (предварительно обработанную с помощью BERT).

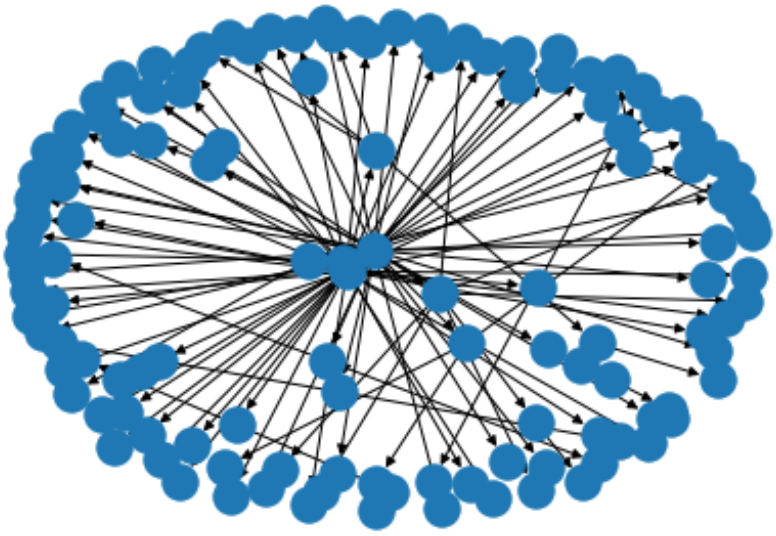

Рисунок 12 – Граф распространения новостной статьи

Таким образом, на данный момент данные полностью готовы для их дальнейшего использования. Применение одного датасета в двух видах позволит провести корректную оценку графовой модели по сравнению с традиционными подходами обработки текста.

3 Постановка задачи

Необходимо разработать модель обнаружения недостоверных новостей в социальной сети, выполняющую бинарную классификацию новостных статей на достоверные и недостоверные с использованием технологии графов знаний.

Для обоснованного выбора подходящей модели классификации требуется реализовать и сравнить результаты рассмотренных ранее методов на выбранном наборе данных «FakeNewsNet».

Для этого важно выполнить следующие шаги:

- Разработка «baseline» модели бинарной классификации, то есть используемой как ориентир для оценки качества работы основной модели. Предполагалось использование привычного для данного типа задач метода обработки естественного языка, например, BERT.
- Разработка основной модели, использующей графовые нейросетевые подходы.
- Оценка и сравнение полученных результатов.

Для достижения указанных целей необходимо провести тщательный анализ выбранного набора данных, а также учитывать возможные ограничения и особенности моделей. Успешное выполнение поставленных задач позволит получить значимые результаты и улучшить качество анализа и классификации фейковых новостей.

Должна быть создана программная реализация разработанного модели с использованием языка Python 3.10 и фреймворка PyTorch. Для работы с графом знаний, который представляет собой сеть связанных узлов и отношений между ними, будут использоваться специальные расширения библиотеки PyTorch, такие как PyG. Данная библиотека позволяет удобно работать с графами, выполнять операции над узлами и ребрами, а также применять различные алгоритмы и модели машинного обучения к графам.

Итоговая модель с демонстрацией работы и сравнением с другими моделями будет представлена в формате Jupyter Notebook. Это позволит

наглядно продемонстрировать все этапы работы модели, а также провести сравнительный анализ с другими подходами.

Результаты данной работы могут быть использованы для более эффективного борьбы с фейковыми новостями и повышения достоверности информации в социальных сетях, что является актуальной задачей в современном информационном обществе.

4 Описание метода решения

4.1 Graph Attention Networks

Графовые сети внимания являются одним из самых популярных типов графовых нейросетей [51]. Их широкая применимость обусловлена рядом причин. В графовых свёрточных сетях у всех соседних узлов важность одинаковая, что не всегда является правильным подходом. Так как каждый узел может иметь свою уникальную значимость и влиять на результат работы сети. Для решения этой проблемы были разработаны графовые сети с механизмом внимания. Они позволяют учитывать важность каждого соседнего узла, присваивая каждому соединению весовой коэффициент. Благодаря этому, сеть может более точно определять, какие узлы следует учитывать при выполнении задачи.

Механизм внимания является важным инструментом в различных областях, где требуется работа с графами данных. Он позволяет повысить эффективность и точность работы нейросетей, что делает его незаменимым инструментом для многих исследователей и разработчиков [52].

В архитектуре GAT (Graph Attention Network) был введен принципиально новый подход, основанный на использовании технологии «самовнимания». Эта технология позволяет моделировать важность узлов в графах, что является ключевым аспектом для решения многих задач в области машинного обучения. Аналогией к этому подходу можно привести важность узла в графе. Таким образом, использование технологии «самовнимания» в архитектуре GAT открывает новые возможности для эффективной работы с графами и повышения точности моделей машинного обучения [53].

В первую очередь необходимо рассмотреть строительные блоки, используемые для построения нейросетей GAT – слои внимания графа (Graph Attentional Layer) [44, 54].

Входными данными для слоя внимания является набор признаков узла, $h = \{\vec{h}_1, \vec{h}_2, \cdots, \vec{h}_N\}, \vec{h}_i \in \mathbb{R}^F$, где N – количество узлов, а F – количество

признаков в каждом узле. Слой на выходе создает новый набор признаков узлов, $h' = \{\vec{h}'_1, \vec{h}'_2, \cdots, \vec{h}'_N\}, \vec{h}'_i \in \mathbb{R}^{F'}$.

Чтобы получить достаточную выразительную силу для преобразования входных признаков в признаки более высокого уровня, требуется по крайней мере одно обучаемое линейное преобразование. С этой целью в качестве начального шага к каждому узлу применяется общее линейное преобразование, параметризированное весовой матрицей, $W \in \mathbb{R}^{F' \times F}$. Общий механизм внимания $a : \mathbb{R}^{F'} \times \mathbb{R}^{F} \to \mathbb{R}$, вычисляющий коэффициенты внимания (13):

$$e_{ij} = a\left(W\,\vec{h}_i, W\,\vec{h}_j\right). \tag{13}$$

Для определения значимости каждого соединения необходимы пары скрытых векторов, которые можно получить путем конкатенации векторов из обоих узлов. Только после этого возможно применение нового линейного преобразования с использованием весовой матрицы W_{att}^{T} (14):

$$e_{ij} = W_{att}^{T}\left[W\vec{h}_i || W\vec{h}_j\right], \tag{14}$$

где T – операция транспонирования, а $||$ – операция конкатенации.

Коэффициент внимания e_{ij} указывает на важность признаков узла j для узла i. В самой общей формулировке модель позволяет каждому узлу уделять внимание каждому другому узлу, отбрасывая всю структурную информацию. Структура графа вводится в механизм с помощью выполнения маскированного внимания. То есть e_{ij} вычисляются только для узлов $j \in \mathcal{N}_i$, где $\mathcal{N}_i$ – некоторая окрестность узла i в графе. В данной работе под этим будут иметься ввиду соседи первого порядка узла i (включая i). Для того чтобы коэффициенты были легко сопоставимы для разных узлов, они нормализуются для всех вариантов j с помощью функции $softmax$ (15):

$$\alpha_{ij} = softmax\left(e_{ij}\right) = \frac{exp\left(e_{ij}\right)}{\sum_{k \in \mathcal{N}_i} exp(e_{ik})}. \tag{15}$$

В качестве функции активации используется нелинейность $LeakyReLU$ [55]. Полностью развернутые коэффициенты, вычисленные механизмом внимания (рис. 13) могут быть выражены как (16):

$$\alpha_{ij} = \frac{exp\left(LeakyReLU\left(W_{att}^{T}\left[W\vec{h}_i || W\vec{h}_j\right]\right)\right)}{\sum_{k\in\mathcal{N}_i} exp\left(LeakyReLU\left(W_{att}^{T}\left[W\vec{h}_i || W\vec{h}_k\right]\right)\right)}. \quad (16)$$

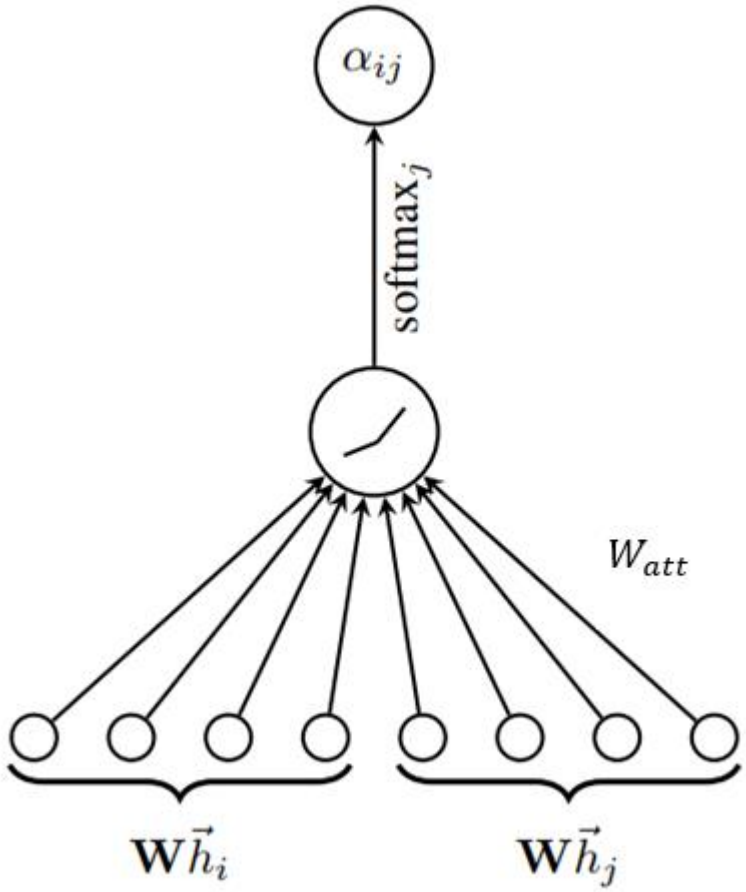

Рисунок 13 – Иллюстрация механизма внимания

После получения нормализованные коэффициенты внимания используются для вычисления линейной комбинации соответствующих им признаков, которые служат окончательными выходными признаками для каждого узла (после возможного применения нелинейности σ) (17):

$$\vec{h}_i' = \sigma\left(\sum_{j\in\mathcal{N}_i} \alpha_{ij}\, W\vec{h}_j\right). \quad (17)$$

Для того чтобы стабилизировать процесс обучения «самовниманию» используется технология многоголового внимания [56]. В частности, K независимых механизмов внимания выполняют преобразование (17), а затем их признаки объединяются, в результате чего получается следующее представление выходных признаков (18):

$$\vec{h}_i' = ||_{k=1}^{K} \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^k W^k \vec{h}_j \right), \tag{18}$$

где α_{ij}^k – нормализованные коэффициенты внимания, вычисленные k-м механизмом внимания (a_k), а W^k – соответствующая входная матрица весов линейного преобразования.

При применении многоголового внимания к последнему (прогнозирующему) слою сети, конкатенация больше не имеет смысла. Вместо этого используется усреднение (19), и только после этого применяется окончательная нелинейность:

$$\vec{h}_i' = \sigma\left(\frac{1}{K} \sum_{k=1}^{K} \sum_{j \in \mathcal{N}_i} \alpha_{ij}^k W^k \vec{h}_j \right). \tag{19}$$

Процесс агрегирования многоголового графового слоя внимания проиллюстрирован на рис. 14.

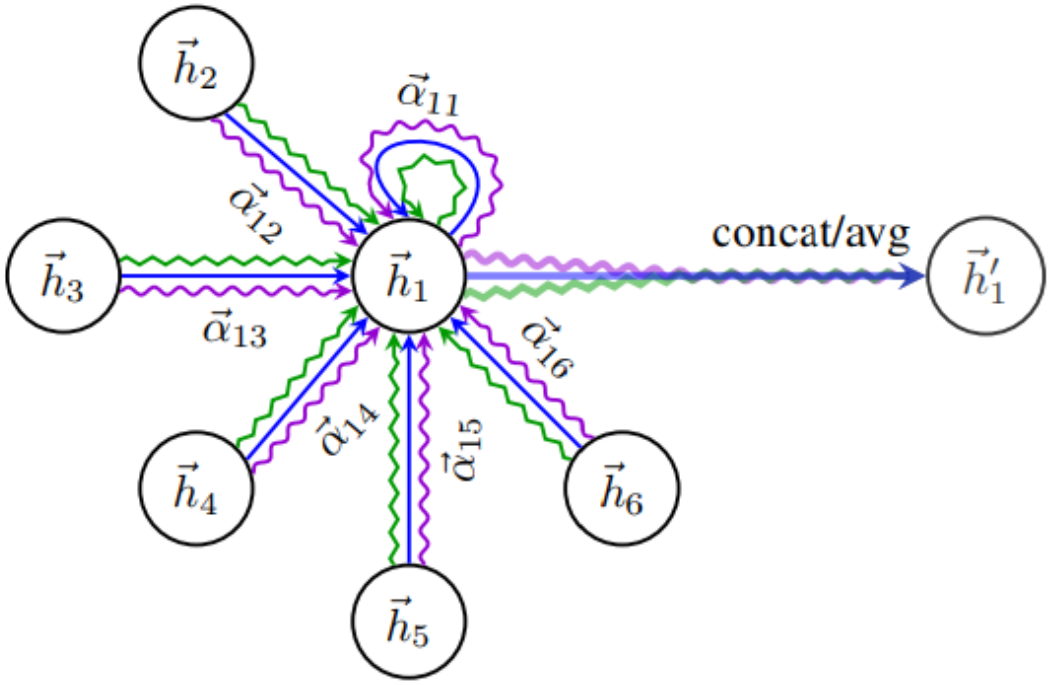

Рисунок 14 – Иллюстрация многоголового внимания ($K = 3$) узла 1 к своей окрестности

4.2 Описание модели

В качестве архитектуры графовой нейронной сети использовалась ранее упомянутая модель с механизмом внимания GAT с тремя свёрточными слоями GATv2Conv.

В классическом варианте свёрточного слоя GATConv коэффициенты внимания вычислялись следующим образом (20):

$$e_{ij} = LeakyReLU\left(W_{att}^{T}\left[W\vec{h}_i || W\vec{h}_j\right]\right). \quad (20)$$

В новой модифицированной версии GATv2Conv весовая матрица W применяется после конкатенации, а весовая матрица внимания W_{att} – после функции $LeakyReLU$. Формула коэффициентов внимания для GATv2Conv (21):

$$e_{ij} = W_{att}^{T} LeakyReLU\left(W\left[\vec{h}_i || \vec{h}_j\right]\right). \quad (21)$$

GATv2Conv считается более эффективным чем классическая версия и позволяет обеспечить лучшие результаты на бенчмарках [57].

В процессе выполнения работы слои графового внимания использованы в двух вариациях. В первом и втором слое применяется

многоголовое внимание, а в третьем голова только одна, в ней и вычисляется окончательный результат перед нелинейностью.

Для ускорения и эффективной векторизации обучение происходило на GPU. В качестве оптимизатора использовался алгоритм Adam (Adaptive Moment Estimation), основанный на градиентном спуске. Он имеет адаптивный шаг обучения, что означает, что шаг обучения изменяется в зависимости от градиента. Это позволяет быстрее сходиться к оптимальному решению и избежать затухания градиента. В качестве функции ошибок была выбрана BCELoss (Binary Cross Entropy Loss), которая является одной из наиболее распространенных функций потерь в задачах бинарной классификации. Она используется для оценки качества работы модели, которая должна определить, принадлежит ли входной объект к одному из двух классов. BCELoss вычисляет ошибку между предсказанными моделью значениями и фактическими метками классов. Она применяется к бинарным данным, где каждый объект может быть отнесен только к одному из двух классов. Функция BCELoss является дифференцируемой, что позволяет использовать оптимизаторы, основанные на градиентном спуске, для обучения модели.

4.3 Программная реализация

Для программной реализации разрабатываемой модели обнаружения недостоверных новостей в социальной сети в формате графа знаний был использован высокоуровневый интерпретируемый язык программирования Python 3.10 [58]. Данный выбор объясняется большой популярностью данного языка, в том числе и в современных научных исследованиях. Язык Python обладает огромным количеством готовых и хорошо реализованных сторонних библиотек. Основными используемыми в данном программном комплексе библиотеками являются NumPy [59], Pandas [60], Matplotlib [61], PyTorch [62] и PyG [63].

NumPy – это библиотека для языка программирования Python, которая предоставляет мощные инструменты для работы с массивами и матрицами.

Она является одной из самых популярных библиотек для научных вычислений и обработки данных. Библиотека NumPy широко используется в научных и инженерных приложениях, таких как обработка изображений, машинное обучение, обработка сигналов, статистика и т.д. Она также является необходимым инструментом для работы с другими библиотеками Python, такими как Pandas, SciPy и Matplotlib. Одним из ключевых преимуществ NumPy является то, что она обеспечивает высокую производительность при работе с большими объемами данных. Это достигается благодаря использованию оптимизированных алгоритмов и структур данных. Необходимость использования данной библиотеки заключается в том, что математические алгоритмы, реализованные с помощью компилируемых языков, работают гораздо быстрее чем в Python. NumPy помогает решить данную проблему и работает также быстро, как аналогичный код, реализованный в MATLAB [64].

Библиотека Pandas – это одна из наиболее популярных библиотек для работы с данными в языке программирования Python. Она предоставляет удобный и мощный инструментарий для обработки, анализа и визуализации различных типов данных. Основными структурами данных в Pandas являются серии (Series) и таблицы (DataFrame). Серия – это одномерный массив данных, который может содержать любой тип данных. Таблица – это двумерный массив данных, который представляет собой набор серий, объединенных в одну структуру. Pandas предоставляет множество функций для работы с данными, таких как сортировка, фильтрация, группировка, агрегирование и многое другое. Она также позволяет работать с данными в различных форматах, включая CSV, Excel, SQL и другие. Одной из основных возможностей Pandas является работа с пропущенными данными. Библиотека предоставляет функции для обнаружения и заполнения пропущенных значений, а также для удаления строк или столбцов, содержащих пропущенные данные. Библиотека Pandas является неотъемлемой частью многих проектов, связанных с анализом данных и машинным обучением.

Она позволяет быстро и эффективно обрабатывать и анализировать большие объемы данных, что делает ее незаменимой для многих задач.

Для визуализации полученных результатов использовалась библиотека Matplotlib. Она позволяет создавать графики, диаграммы, гистограммы, круговые диаграммы и многое другое. Библиотека Matplotlib является одной из наиболее популярных библиотек для визуализации данных в Python. Matplotlib использует объектно-ориентированный подход к созданию графиков. Это означает, что каждый элемент графика (оси, линии, текст и т.д.) представлен в виде объекта Python, который можно настроить и изменить по своему усмотрению. Также Matplotlib позволяет расширить возможности библиотеки NumPy.

PyTorch – это библиотека машинного обучения с открытым исходным кодом, разработанная компанией Facebook. Она предоставляет простой и гибкий интерфейс для создания глубоких нейронных сетей и других моделей машинного обучения. PyTorch также предоставляет множество инструментов для работы с данными, включая загрузчики данных и предобработку. Библиотека также поддерживает работу с различными форматами данных, включая изображения, звук и текст. Таким образом, PyTorch – это мощная и гибкая библиотека машинного обучения, которая позволяет создавать сложные модели.

PyG (PyTorch Geometric) – это библиотека для работы с графовыми нейронными сетями в PyTorch. Она предоставляет широкий спектр инструментов для работы с графами, включая загрузку и преобразование данных, создание моделей и обучение. Одним из ключевых преимуществ PyG является то, что данная библиотека создана как расширение библиотеки PyTorch. Это позволяет использовать все возможности PyTorch для работы с графовыми нейронными сетями, а также хорошо интегрируется с другими инструментами PyTorch.

Для реализации данного программного комплекса была выбрана интерактивная среда разработки Jupyter Notebook (рис. 15). Исходный код опубликован в свободный доступ на платформе GitHub [65].

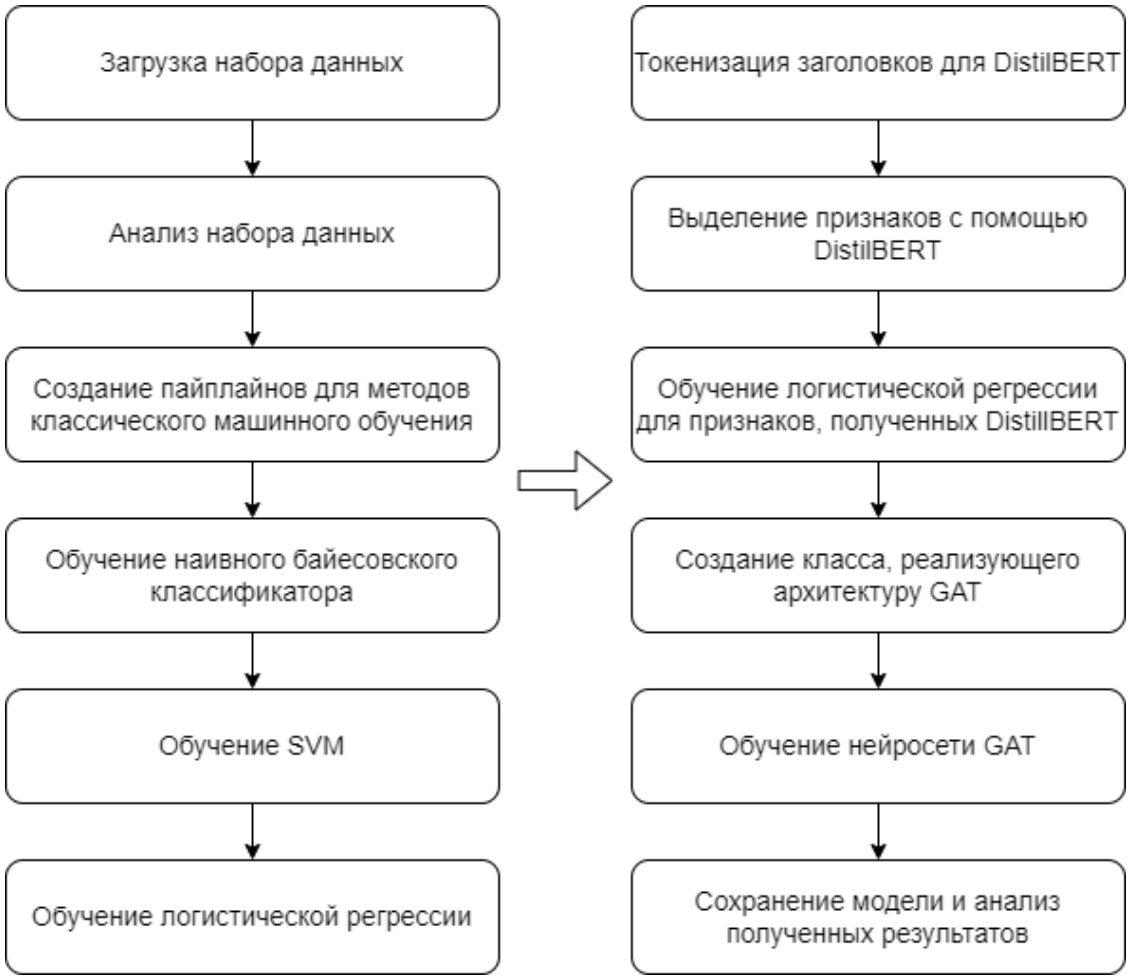

Рисунок 15 – Иллюстрация структуры программного комплекса, реализованного в среде Jupyter Notebook

Рекомендуемыми системными требованиями разработанного программного обеспечения являются:

- 64-битная версия Microsoft Windows 8, 10, 11;
- RAM: 4 ГБ;
- Python 3.10 или более поздняя версия.

5 Исследование свойств решения

5.1 Методика оценки эффективности разработанной модели

Для выполнения эксперимента по исследованию свойств полученного решения проблемы обнаружения недостоверных новостей в социальной сети в формате графа знаний использовались входные данные в двух форматах. Классический табличный формат датасета «FakeNewsNet» использовался для реализации «baseline» моделей и изучения эффективности классических методов машинного обучения и модели обработки естественного языка. Для обучения и исследования эффективности классификации разработанной модели использовалась версия этого датасета предварительно преобразованная в формат графа знаний с помощью технологии RML.

В рамках данного исследования был использован один и тот же набор данных, представленный в двух форматах: табличном и графовом. Такая методика позволила корректно провести анализ различных методов решения задачи и оценить эффективность и перспективность графового подхода по сравнению с аналогами.

Основной метрикой оценки качества и сравнения эффективности моделей выбрана F1-мера. Выбор объясняется тем, что точность не подойдет для оценки результатов решаемой задачи, так как может привести к заблуждению из-за несбалансированности классов. Тем временем, F1-мера является сбалансированной метрикой и представляет собой компромисс между $precision$ и $recall$.

В качестве основной модели использовалась графовая сеть с вниманием GAT (с тремя свёрточными слоями). Для обучения использовались ранее упомянутые оптимизатор Adam и функция ошибок BCELoss. Данные были предварительно разделены на обучающую, валидационную и тестовую выборки. Для предотвращения переобучения реализована ранняя остановка, прерывающая цикл обучения, если значение функции ошибки на валидационной выборке не уменьшается в течении 10 эпох. Наилучшая модель сохраняется с помощью сериализации. Обучение

производилось с помощью платформы Google Colab на GPU. Графики зависимостей значения функции ошибки и F1-меры от номера эпохи представлены на рис. 16 и рис. 17 соответственно.

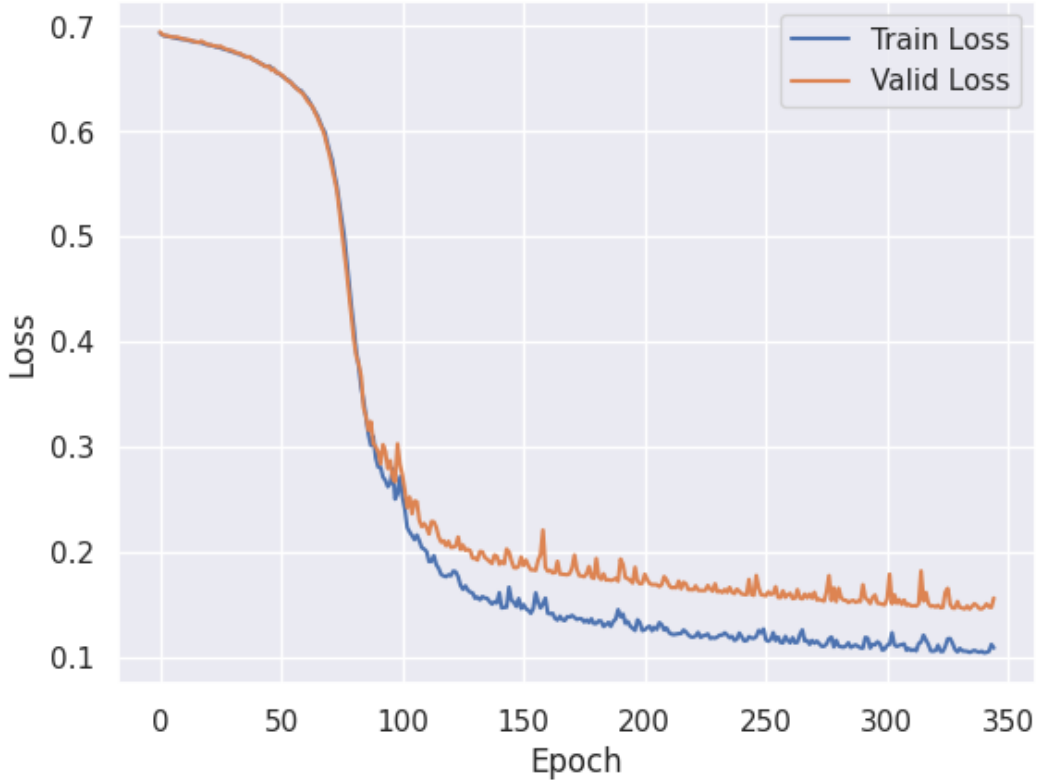

Рисунок 16 – График зависимости значения функции ошибки от номера эпохи для обучающей и валидационной выборок

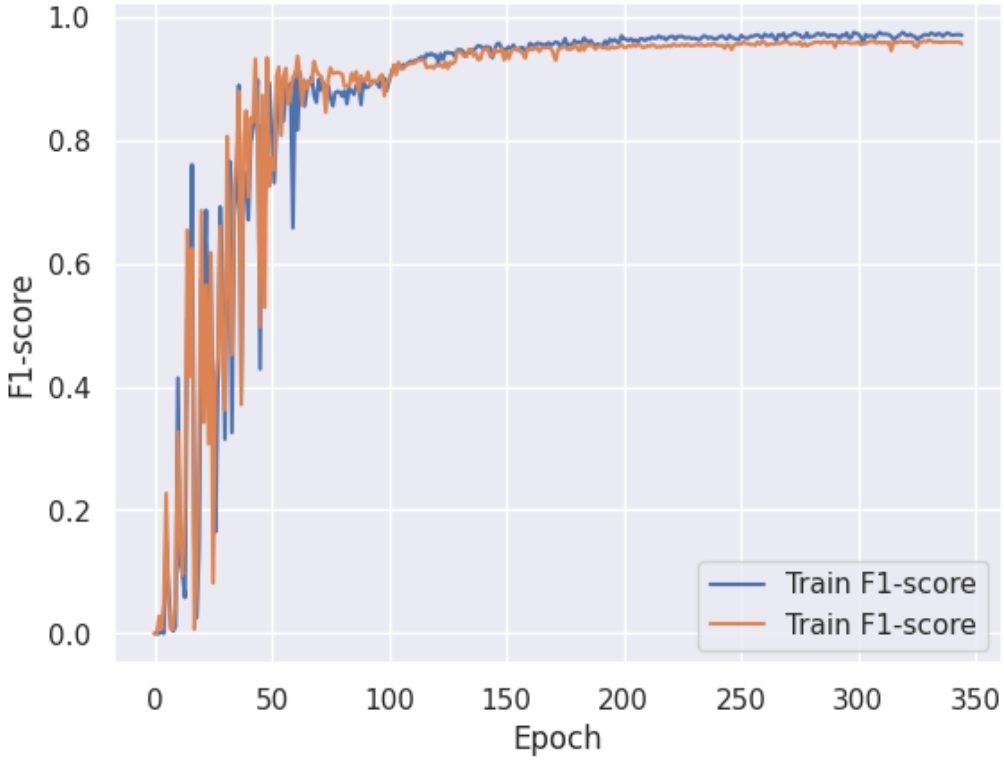

Рисунок 17 – График зависимости значения F1-меры от номера эпохи для обучающей и валидационной выборок

Изучив графики, можно сделать вывод, что модель не переобучилась и достигла локального оптимума.

На тестовых данных модель достигла точности 96,8% и F1-меры 96,7%, что является крайне высокими результатами.

5.2 Сравнение с результатами аналогов

Для оценки качества результатов, полученных с помощью разработанной модели, были реализованы модели классического машинного обучения, такие как наивный байесовский классификатор, логистическая регрессия и метод опорных векторов. Также была реализована языковая модель DistilBERT.

В процессе разработки моделей, основанных на методах классического машинного обучения, были созданы пайплайны, реализующие последовательные стадии преобразования данных. Это необходимо из-за того, что для работы с текстом на естественным языке его нужно предварительно представить в понятном для моделей численном виде.

Первым шагом пайплайнов является модуль CountVectorizer, который преобразовывает поданный на вход текст в матрицу количеств вхождений слов в текст. Далее использовался метод TfidfTransformer, который применяется для преобразования числовых векторов, полученных с помощью CountVectorizer, в векторы, учитывающие важность слов в тексте. Tfidf расшифровывается как Term Frequency-Inverse Document Frequency и является одним из наиболее популярных методов векторизации текстовых данных. TfidfTransformer использует формулу, которая учитывает как частоту встречаемости слова в документе (term frequency), так и обратную частоту встречаемости слова во всех документах (inverse document frequency). Это позволяет выделить ключевые слова, которые наиболее характерны для данного документа и могут помочь в его классификации.

Результаты метрики F1 для классических моделей машинного обучения, обученных на классической версии «FakeNewsNet», составили примерно 60%.

Далее была разработана языковая модель DistilBERT, являющаяся одним из алгоритмов ранее упомянутого семейства нейросетевых архитектур BERT.

DistilBERT основан на идее дистилляции. Её суть в имитации легковесной моделью поведения более сложной модели-учителя, в данном случае BERT. Языковая модель DistilBERT оптимизирует процесс обучения за счет уменьшения размера и увеличения скорости BERT, при этом сохраняя до 97% производительности.

Модель была использована для получения признаков из новостных заголовков с целью дальнейшей их обработки с помощью алгоритмов логистической регрессии для бинарной классификации новостей на достоверные и не достоверные.

DistilBERT получает на вход заголовки новостных статей, токенизирует их и, далее, на GPU выделяет признаки и подает их в следующую модель для обработки. Преимущество использования моделей

семейства BERT в двустороннем анализе контекста, что дает возможность легко получить достаточно хорошие результаты.

Для задачи определения достоверности новости модель, использующая связку DistilBERT и логистическую регрессию, показала хорошие результаты и достигла F1-меры 82% на тестовом наборе данных.

Значения метрик точности и F1-меры для реализованных моделей проиллюстрированы в таблице 1.

Таблица 1 – Сравнения методов по метрикам

Название метода	Точность	F1-мера
Наивный Байес	85%	64%
SVM	85%	65%
Логистическая регрессия	83%	59%
DistilBERT	84%	82%
GAT	96%	96%

Таким образом, можно сделать вывод, что использование графовых подходов, а в частности архитектуры GAT способно значительно улучшить качество бинарной классификации новостных статей на достоверные и не достоверные. При этом графовая нейронная сеть с вниманием GAT обладает дополнительными преимуществами из-за использования графового подхода и структуры данных [66, 67]. Модель сохраняет свою эффективность с течением времени по сравнению с моделями, основанными только на обработке текста.

6 Анализ социального содержания заказа и социально-политических условий реализации работы

6.1 Социальная значимость проблемы обнаружения недостоверных новостей

Социальные сети стали неотъемлемой частью повседневной жизни, и многие люди получают новости и информацию преимущественно из социальных сетей. Однако, с ростом количества пользователей, возникает проблема распространения недостоверных новостей. Это может привести к серьезным последствиям, таким как паника, неправильное принятие решений и даже насилие [1, 2, 3].

Использование искусственного интеллекта для обнаружения недостоверных новостей в социальных сетях является перспективным способом борьбы с этой проблемой [68]. ИИ может быстро анализировать большие объемы данных и выявлять недостоверные новости, основываясь на различных критериях, таких как источник, контекст и достоверность информации. Эта технология может быть особенно полезна в периоды кризисов, когда люди ищут информацию о происходящих событиях. Например, в случае пандемии Covid19, недостоверные новости могут привести к распространению ложной информации о лечении и профилактике болезни, что может привести к серьезным последствиям для здоровья людей.

Кроме того, использование ИИ для обнаружения недостоверных новостей может помочь в борьбе с дезинформацией и фейковыми новостями, которые используются в политических целях. Это может помочь сохранить демократические ценности, такие как свобода слова и право на информацию.

Таким образом, можно сделать вывод, что использование искусственного интеллекта для обнаружения недостоверных новостей в социальных сетях имеет большую социальную актуальность и является важным шагом в направлении борьбы с дезинформацией и сохранением достоверности информации. Это может помочь сохранить здоровье и безопасность людей.

6.2 Проблема этики искусственного интеллекта

Проблема этики искусственного интеллекта является одной из самых актуальных в настоящее время. Развитие технологий и науки приводит к появлению новых возможностей в области искусственного интеллекта, что может привести к серьезным последствиям для общества.

При использовании искусственного интеллекта для обнаружения недостоверных новостей может возникнуть проблема дискриминации и стереотипов [69]. Если модели обучаются на данных, которые содержат предвзятость, то их результаты могут быть неправильными и несправедливыми. Например, если система обучается на данных, где определенные группы людей часто связываются с негативными новостями, то модель может начать ассоциировать новости про них с недостоверностью в случае их упоминания в положительном контексте.

Таким образом, проблема этики искусственного интеллекта в целом и в задаче обнаружения недостоверных новостей в социальной сети, в частности, является очень актуальной. Необходимо разрабатывать модели, которые не будут подвержены проблемам дискриминации и стереотипам.

6.3 Социологические методики

Для оценки актуальности решаемой задачи и общественного мнения о проблеме обнаружения недостоверных новостей в социальных сетях необходимо получить социально значимую информацию с помощью специальных социологических методик. Важно также проанализировать мнение экспертов и пользователей социальных сетей о свойствах, которыми должно обладать решение поставленной задачи, и сопоставить их ожидания с разработанной моделью.

Интервью с экспертами и социологические опросы являются важными социологическими методиками, которые позволяют получить ценную информацию о различных социальных явлениях и процессах.

Интервью с экспертами – это методика, при которой исследователь общается с экспертом в определенной области знаний, чтобы получить

информацию о теме своего исследования [70]. Эксперты могут быть учеными, практикующими специалистами, представителями общественных организаций и т.д. Их мнение и опыт могут помочь исследователю понять сложные социальные явления и процессы, а также дать рекомендации по дальнейшему исследованию.

Социологический опрос – это методика, при которой исследователь задает определенный набор вопросов респондентам, чтобы получить информацию о социальных явлениях и процессах [71]. Опросы могут проводиться как в форме личного интервью, так и через телефон, интернет или почту. Результаты опросов могут быть представлены в виде статистических данных, которые позволяют исследователю делать обобщения о социальных явлениях и процессах.

Обе методики имеют свои преимущества и недостатки. Интервью с экспертами позволяет получить глубокое понимание темы исследования, однако может быть ограничено мнением нескольких человек. Социологический опрос, с другой стороны, позволяет получить данные от большого количества людей, что увеличивает репрезентативность результатов, но может быть ограничен формой и темой вопросов.

Тем не менее, обе методики являются важными инструментами для социологического исследования и позволяют получить ценную информацию о социальных явлениях и процессах. Кроме того, сочетание различных методик может улучшить качество и точность исследования.

В процессе выполнения данной работы было проведено социологическое исследование с использованием двух упомянутых ранее методик: интервью с экспертами и социологического опроса. В качестве экспертов были выбраны:

- Журналист регионального СМИ с опытом работы более 6 лет.
- Программист в области ИИ с опытом работы более 5 лет.

Данный набор экспертов поможет оценить с двух различных точек зрения социальную значимость обнаружения недостоверных новостей в

социальной сети, а также понять, насколько актуальна проблема этики ИИ для решаемой задачи.

Также был проведен социологический опрос среди пользователей социальной сети ВКонтакте. В качестве выборки использовались участники различных сообществ с разной тематикой для обеспечения наибольшей репрезентативности. В ходе опроса своим мнением поделились 167 человек.

Перед интервью и опросом были получены согласия респондентов на публикацию исследования, а также сохранена их анонимность.

6.4 Анализ интервью с экспертами

В процессе социологического исследования было взято интервью у эксперта в области журналистики. Его мнение крайне актуально в вопросах социальной значимости обнаружения недостоверных новостей и этики использования ИИ для решения этой задачи, так как ему часто приходится сталкиваться с недостоверными новостями в рабочей практике.

Анализ интервью с журналистом показывает, что недостоверные новости являются серьезной проблемой для общества, и их проверка является важной задачей. Журналист подчеркивает необходимость использования разных критериев для проверки новостей, включая проверку источников, кросс-проверку информации и анализ контекста. Эксперт положительно относится к использованию искусственного интеллекта для обнаружения недостоверных новостей, так как это позволяет сократить время на проверку новостей и улучшить качество информирования аудитории, но при этом отмечает, что есть риск, что искусственный интеллект может быть неправильно настроен и будет обнаруживать недостоверные новости, которые на самом деле являются правдивыми, а также может возникнуть риск дискриминации и манипуляции общественным мнением.

Журналист выделяет преимущества и недостатки использования искусственного интеллекта в задаче обнаружения недостоверных новостей по сравнению с работой журналистов. Искусственный интеллект может обрабатывать большие объемы информации и быстро определять

недостоверные новости, что экономит время журналистов. Однако, он не всегда может учесть контекст и эмоциональный подтекст новостей, что может привести к ошибкам.

Эксперт в сфере журналистики также отмечает, что для регулирования использования искусственного интеллекта в борьбе с фейками может потребоваться изменение законодательства в области защиты персональных данных, ответственности за распространение ложной информации, а также прозрачности алгоритмов работы систем искусственного интеллекта. Кроме того, возможно потребуется разработка новых правил и стандартов для оценки эффективности и безопасности таких систем.

Также было проведено интервью с экспертом в области разработки ИИ. Его мнение является актуальным, так как он понимает алгоритмы работы и обучения моделей и знаком с проблемами этики в искусственном интеллекте.

Анализ интервью с программистом показывает, что использование искусственного интеллекта в борьбе с недостоверными новостями является актуальной и серьезной проблемой. Программист положительно относится к использованию ИИ для обнаружения недостоверных новостей, так как это позволяет повысить качество информационного потока и защитить людей от ложной информации. Однако, он также указывает на возможность ложных срабатываний и ошибок в классификации.

Программист указывает на возможные последствия неправильной классификации новостей как недостоверных, такие как распространение ложной информации, что может негативно повлиять на общественное мнение и принятие решений, а также ухудшить доверие к системам искусственного интеллекта в целом. Для предотвращения таких последствий необходимо разработать более точные модели классификации и проводить регулярную проверку систем.

Разработчик выделяет проблему этики в области искусственного интеллекта в задаче обнаружения недостоверных новостей. Он подчеркивает необходимость разработки этических стандартов для использования ИИ в

таких задачах, чтобы избежать возможных негативных последствий для общества. Кроме того, важно учитывать возможность искажения результата работы ИИ из-за предвзятости данных, на которых он обучается.

Эксперт в области искусственного интеллекта обращает внимание на возможные этические проблемы при использовании ИИ в задаче обнаружения недостоверных новостей, такие как дискриминация на основе расовой, половой или иной принадлежности. Он подчеркивает необходимость убедиться, что данные, используемые для обучения, не содержат скрытых предубеждений и соответствуют разнообразию населения.

В ходе интервью эксперт предложил меры для уменьшения рисков этических нарушений при использовании искусственного интеллекта в задаче обнаружения недостоверных новостей, такие как обеспечение прозрачности работы алгоритмов ИИ и организация мониторинга работы моделей для быстрого выявления и исправления ошибок.

Таким образом, оба эксперта признают преимущества использования искусственного интеллекта для решения проблемы обнаружения недостоверных новостей в социальной сети. Однако подчеркивают этические проблемы, которые могут возникать при использовании ИИ. Разработанный программный комплекс полностью соответствует требованиям экспертов, так как модель, которая лежит в его основе, производит классификацию на основе структуры графа распространения новости и не учитывает новостной заголовок, что помогает избежать дискриминации и стереотипов при работе модели.

6.5 Анализ социологического опроса

Для определения социальной значимости разработанной модели для целевой аудитории был проведен анонимный социологический опрос пользователей социальной сети ВКонтакте, в ходе которого своим мнением поделилось 167 человек.

Анализ результатов социологического опроса показывает, что большинство респондентов сталкиваются с недостоверными новостями в

социальных сетях несколько раз в месяц, а 21% прошедших опрос наблюдают фейковые новости почти каждый день (рис. 18).

Рисунок 18 – Диаграмма ответов на первый вопрос социологического опроса

Наиболее важными мерами для решения проблем распространения и обнаружения недостоверных новостей в социальных сетях большинство респондентов считает: применение ИИ-технологий и усиление модерации контента (рис. 19).

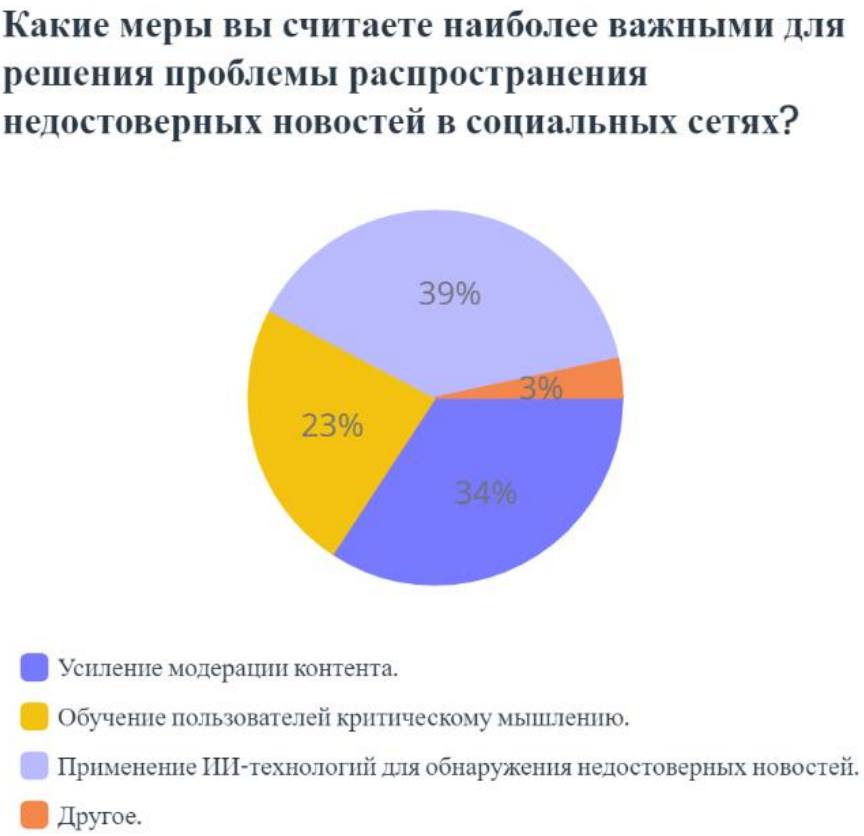

Рисунок 19 – Диаграмма ответов на второй вопрос социологического опроса

Большая часть пользователей, прошедших опрос, относится к использованию искусственного интеллекта для обнаружения недостоверных новостей положительно, что говорит о одобрении целевой аудиторией использования таких моделей (рис. 20).

Рисунок 20 – Диаграмма ответов на третий вопрос социологического опроса

Почти все респонденты считают, что искусственный интеллект способен сыграть важную роль в борьбе с распространением недостоверных новостей в социальной сети (рис. 21).

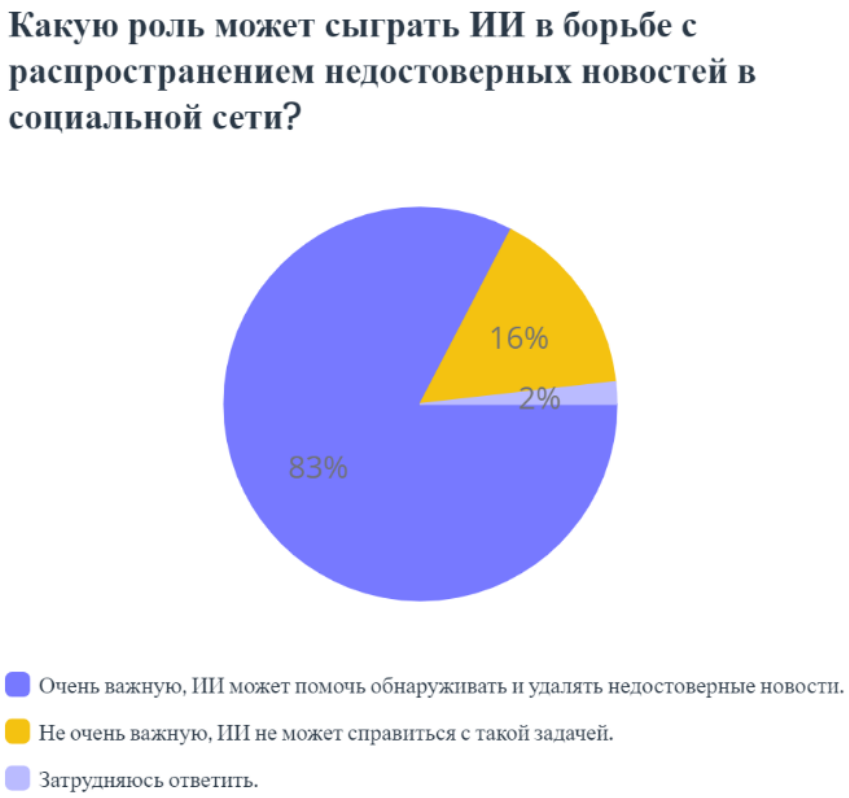

Рисунок 21 – Диаграмма ответов на четвертый вопрос социологического опроса

Большинство прошедших опрос считают, что обладают базовыми представлениями о вопросах этики в сфере искусственного интеллекта (рис. 22).

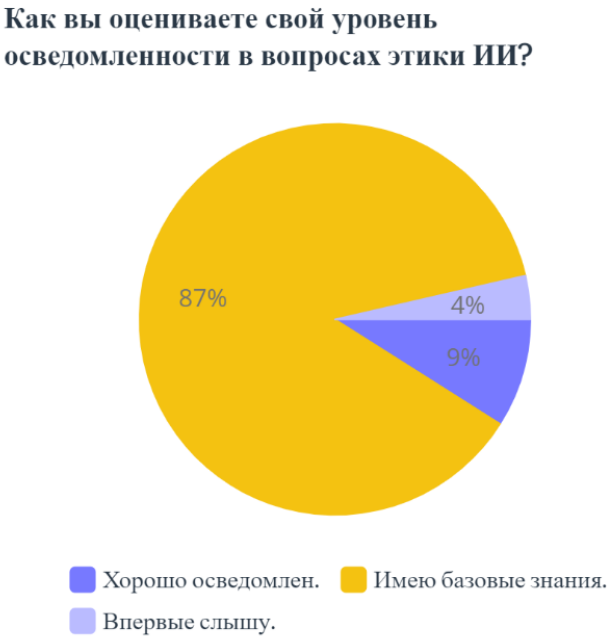

Рисунок 22 – Диаграмма ответов на пятый вопрос социологического опроса

Около половины прошедших опрос пользователей считают проблему этики ИИ в задаче обнаружения недостоверных новостей важной (рис. 23)

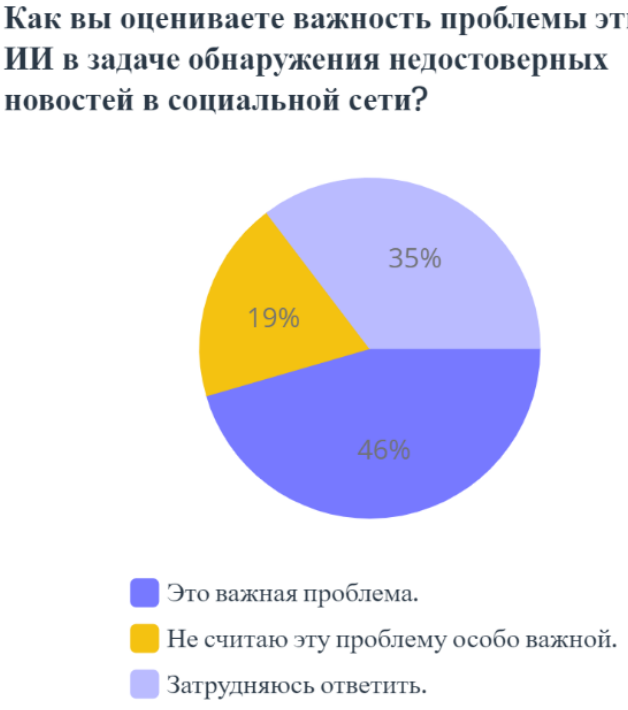

Рисунок 23 – Диаграмма ответов на шестой вопрос социологического опроса

57% респондентов считают, что алгоритмы искусственного интеллекта способны дискриминировать отдельные категории людей (рис. 24).

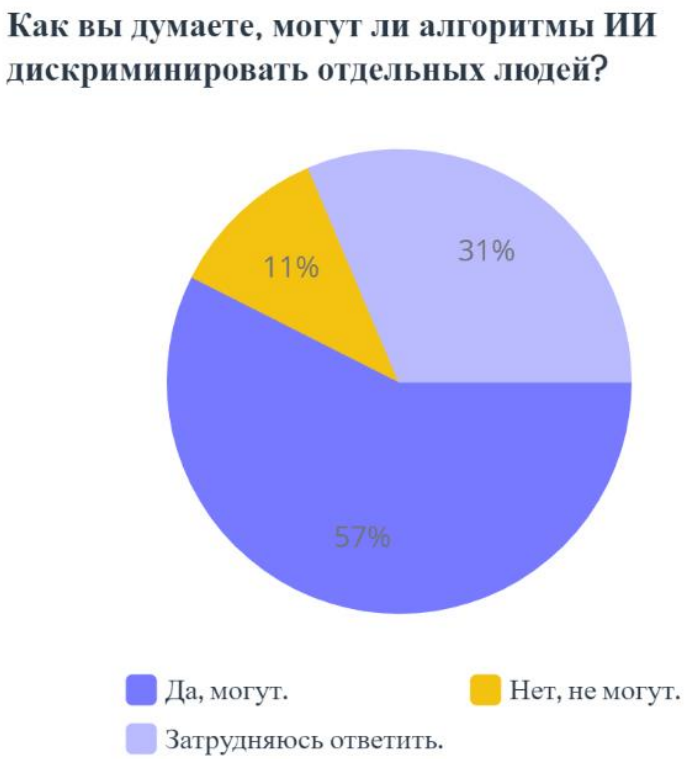

Рисунок 24 – Диаграмма ответов на седьмой вопрос социологического опроса

Почти все респонденты, поучаствовавшие в социологическом опросе согласны с утверждением, что искусственный интеллект, использующийся для обнаружения недостоверных новостей, должен основываться только на моделях и данных, которые исключают возможность дискриминации отдельных категорий лиц и другие этические проблемы (рис. 25).

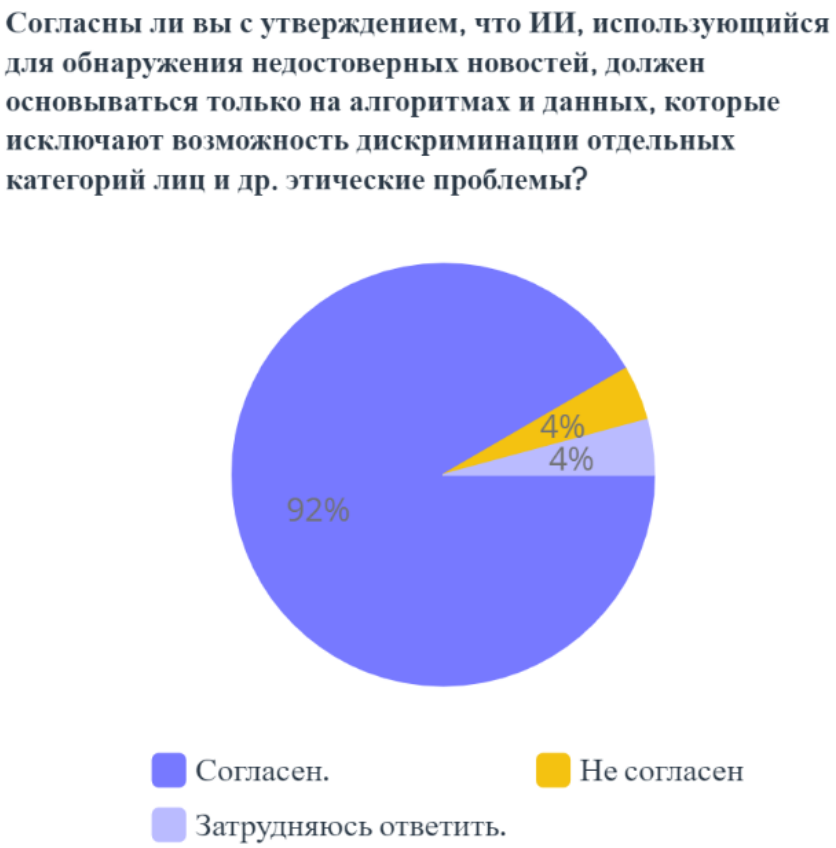

Рисунок 25 – Диаграмма ответов на восьмой вопрос социологического опроса

Таким образом, большинство респондентов сталкивается с недостоверными новостями на регулярной основе и относится положительно к использованию искусственного интеллекта для обнаружения недостоверных новостей в социальной сети и считают, что ИИ может сыграть очень важную роль в борьбе с их распространением. Большая часть респондентов имеет базовые знания в вопросах этики искусственного интеллекта, но только небольшое количество хорошо осведомлено. Тем не менее, большинство респондентов считают, что алгоритмы ИИ могут дискриминировать отдельных людей, но большинство также согласны с тем, что искусственный интеллект, использующийся для обнаружения недостоверных новостей, должен основываться только на алгоритмах и данных, которые исключают возможность дискриминации отдельных категорий лиц и другие этические проблемы.

Разработанный программный комплекс полностью соответствует требованиям респондентов, так как решает актуальную проблему обнаружения недостоверных новостей в социальной сети, применяя модель, избегающую использование информации, которая может потенциально привести к этическим проблемам.

6.6 Вывод

Таким образом, можно сделать вывод, что разработанная модель является актуальной для целевой аудитории и решает социально значимую проблему обнаружения недостоверных новостей, а также, что немаловажно, использует алгоритмы, основанные на структуре графа распространения новости в социальной сети, что помогает избежать этических проблем. Эксперты в двух прикладных областях, журналистике и разработке искусственного интеллекта, также показали свою заинтересованность и положительное отношение к таким моделям. Разработанный программный комплекс полностью соответствует требованиям как респондентов, представляющих целевую аудиторию, так и экспертов.

ЗАКЛЮЧЕНИЕ

В результате выполнения выпускной квалификационной работы достигнута поставленная цель – разработана модель обнаружения недостоверных новостей в социальной сети в формате графа знаний.

В процессе выполнения работы были изучены различные методы обнаружения недостоверных новостей в социальной сети.

Была разработана и реализована на языке Python с использованием библиотек PyTorch и PyG модель бинарной классификации недостоверных новостей в социальной сети в формате графа знаний.

Основным показателем сравнения эффективности модели выбрана метрика F1-мера. Графовая нейронная сеть с механизмом внимания GAT обеспечила значение F1-меры 96% на датасете «FakeNewsNet». При этом результаты классических моделей машинного обучения на этом наборе данных составляют около 60%, а использование DistilBERT позволяет достичь F1-меры 82%. Решения, использующие графовые подходы без механизма внимания, на используемом наборе данных обеспечивают результат 80-85%.

Дополнительным преимуществом графового подхода является способность модели сохранять свою эффективность с течением времени по сравнению с методами, основанными только на обработке естественного языка. Текстовые признаки в заголовках изменяются при появлении новых тенденций в мире. Графовая модель основана на структуре графа распространения новости, а также на информации о пользователях, что позволяет ей эффективно работать в новых условиях.

Новизна работы заключается в использовании графовой модели с вниманием, учитывающей важность узлов, а также позволяющей сохранять эффективность обнаружения недостоверных новостей с течением времени. Результаты исследования являются достоверными так как применяется методика оценки эффективности графовых моделей, предусматривающая корректное сравнение с другими методами решения проблемы за счёт

применения одного и того же набора данных в двух видах: табличном и предварительно преобразованном к формату графа знаний с помощью технологии RML.

Решение данной задачи имеет большое прикладное значение в сфере журналистики и является социально значимым. В ходе данной работы удалось существенно улучшить результаты, полученные с помощью других моделей.

Результаты работы были представлены в виде доклада на конференции НТС-2023 и опубликованы в сборнике [6].

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Shu K., Wang S., Liu H. Exploiting Tri-Relationship for Fake News Detection // Computer Science and Engineering. – 2017. – P. 1-10.
2. Fake News Detection Using Machine Learning Ensemble Methods / I. Ahmad, M. Yousaf, S. Yousaf, et al. // Hindawi Complexity. – 2020. – P. 1-11.
3. Evaluating Deep Learning Approaches for Covid19 Fake News Detection / A. Wani, I. Joshi, S. Khandve, et al. // Communications in Computer and Information Science. – 2021, Vol. 1402. – P. 153-163.
4. Fake News Detection on Social Media: A Data Mining Perspective / K. Shu, A. Sliva, S. Wang, et al. // ACM SIGKDD Exploration Newsletter. – 2017, Vol. 19, № 1. – P. 22-36.
5. Tian L., Zhang X., Peng M. FakeFinder: Twitter Fake News Detection on Mobile // Companion Proceedings of the Web Conference. – 2020. – P. 79-80.
6. Головин А.А., Куликов И.А., Жукова Н.А. Разработка модели обнаружения недостоверных новостей в социальной сети в формате графа знаний // Научно-технический семинар кафедры МО ЭВМ. – 2023. – С. 17-19.
7. What is Knowledge Graph? | IBM [Электронный ресурс]. URL: https://www.ibm.com/cloud/learn/knowledge-graph (дата обращения: 07.02.2023).
8. RDF Mapping Language (RML) [Электронный ресурс]. URL: https://rml.io/specs/rml/ (дата обращения: 08.02.2023).
9. Resource Description Framework (RDF) Model and Syntax Specification [Электронный ресурс]. URL: https://www.w3.org/TR/PR-rdf-syntax/Overview.html (дата обращения: 08.02.2023)
10. KG Course 2021 [Электронный ресурс]. URL: https://migalkin.github.io/kgcourse2021/ (дата обращения: 08.02.2023).

11. Google Knowledge Graph Search API [Электронный ресурс]. URL: https://developers.google.com/knowledge-graph (дата обращения: 15.02.2023).
12. Wikidata: Introduction [Электронный ресурс]. URL: https://www.wikidata.org/wiki/Wikidata:Introduction (дата обращения: 15.02.2023).
13. Баланова Л.А., Ющенко Е.В. Модели представления знаний: виды, применение, достоинства и недостатки // Материалы XII Международной студенческой научной конференции «Студенческий научный форум». – 2020.
14. Тельнов В.П., Коровин Ю.А. Семантический веб и графы знаний как образовательная технология подготовки кадров для ядерной энергетики // Известия вузов – Ядерная энергетика. – 2019. – С. 219-229.
15. Как графы знаний увеличивают продажи в реальном времени [Электронный ресурс]. URL: https://blog.salesai.ru/how-knowledge-graphs-increase-sales-in-real-time (дата обращения: 12.03.2023).
16. What is Binary Classification [Электронный ресурс]. URL: https://deepchecks.com/glossary/binary-classification/ (дата обращения: 15.03.2023).
17. Naive Bayes Classifier From Scratch in Python [Электронный ресурс]. URL: https://machinelearningmastery.com/naive-bayes-classifier-scratch-python/ (дата обращения: 15.03.2023).
18. Ala’raj M., Majdalawieh M., Abbod M.F. Improving binary classification using filtering based on k-NN proximity graphs // J Big Data. – 2020, Vol. 7. – P. 1-18.
19. Alkhateem Y.N.S, Mejri M. Auto Encoder Fixed-Target Training Features Extraction Approach for Binary Classification Problems // Asian Journal of Research in Computer Science. – 2023, Vol. 15. – P. 32-43.
20. Tolles J., Meurer W.J. Logistic Regression: Relating Patient Characteristics to Outcomes // JAMA. – 2016.

21. What is Logistic regression? [Электронный ресурс]. URL: https://www.ibm.com/topics/logistic-regression (дата обращения: 16.03.2023).
22. Rymarczyk T., Kozlowski E., Klosowski G. Logistic Regression for Machine Learning in Process Tomography // Sensors. – 2019, Vol. 19.
23. Shah K., Patel H., Sanghvi D. A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification // Augment Hum Res. – 2020, Vol. 5.
24. Support Vector Machine — Introduction to Machine Learning Algorithms [Электронный ресурс]. URL: https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47 (дата обращения: 20.03.2023).
25. Devlin J., Chang M.W., Lee K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Computation and Language. – 2019.
26. Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing [Электронный ресурс]. URL: https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html (дата обращения: 21.03.2023).
27. BERT (языковая модель) [Электронный ресурс]. URL: https://neerc.ifmo.ru/wiki/index.php?title=BERT_(языковая_модель) (дата обращения: 22.03.2023).
28. Kaliyar R.K., Goswami A., Narang P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach // Multimed Tools Appl. – 2021, Vol. 80. – P. 11765-11788.
29. BERTology [Электронный ресурс]. URL: https://huggingface.co/docs/transformers/bertology (дата обращения: 23.03.2023).
30. Sanh V., Debut L., Chaumond J. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter // Computation and Language. – 2020.

31. DistilBERT [Электронный ресурс]. URL: https://huggingface.co/docs/transformers/model_doc/distilbert (дата обращения: 24.03.2023).
32. Enriching BERT with Knowledge Graph Embeddings for Document Classification / M. Ostendorff, P. Bourgonje, M. Berger, et al. // Proceedings of the 15th Conference on Natural Language Processing. – 2019. – P. 1-8.
33. Yao L., Mao C., Luo Y. KG-BERT: BERT for knowledge graph completion // arXiv preprint arXiv:1909.03193. – 2019.
34. K-BERT: Enabling Language Representation with Knowledge Graph / W. Liu, P. Zhou, Z. Zhao, et al. // The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20). – 2020. – P. 2901-2908.
35. An Introduction to Graph Neural Network (GNN) For Analysing Structured Data [Электронный ресурс]. URL: https://towardsdatascience.com/an-introduction-to-graph-neural-network-gnn-for-analysing-structured-data-afce79f4cfdc (дата обращения: 25.03.2023).
36. Graph Neural Network and Some of GNN Applications: Everything You Need to Know [Электронный ресурс]. URL: https://neptune.ai/blog/graph-neural-network-and-some-of-gnn-applications (дата обращения: 26.03.2023).
37. The Graph Neural Network Model / F. Scarselli, M. Gori, A. C. Tsoi, et al. // IEEE Transactions on Neural Networks. – 2009, Vol. 20. – P. 61-80.
38. Heindl C. Graph Neural Networks for Node-Level Predictions // eprint arXiv. – 2020. – P. 1-9.
39. Rec-GNN: Research on Social Recommendation based on Graph Neural Networks / G. Si, S. Xu, Z. Li, et al. // Proceedings of the 2022 International Conference on Computer Science, Information Engineering and Digital Economy (CSIEDE 2022). – 2022. – P. 478-485.
40. A compact review of molecular property prediction with graph neural networks / O. Wieder, S. Kohlbacher, M. Kuenemann, et al. // Drug Discovery Today: Technologies. – 2020, Vol. 37. – P. 1-12.

41. Improved Handwritten Digit Recognition Using Convolutional Neural Networks (CNN) / S. Ahlawat, A. Chouldhary, A. Nayyar // Sensors. – 2020, Vol. 20. – P. 1-18.
42. Kipf T.N., Welling M. Semi-Supervised Classification with Graph Convolutional Networks // ICLR 2017. – 2017. – P. 1-14.
43. Bridging the Gap Between Spectral and Spatial Domains in Graph Neural Networks / M. Balciar, G. Rentom, P. Heroux, et al. // eprint arXiv. – 2020. – P. 1-24.
44. Graph Attention Networks / P. Velickovic, G. Cucurull, A. Casanova, et al. // ICLR 2018. – 2018. – P. 1-12.
45. 24 Evaluation Metrics for Binary Classification (And When to Use Them) [Электронный ресурс]. URL: https://neptune.ai/blog/evaluation-metrics-binary-classification (дата обращения: 27.03.2023).
46. Shu K. FakeNewsNet: A Data Repository with News Content, Social Context and Spatiotemporal Information for Studying Fake News on Social Media // arXiv preprint arXiv:1809.01286. – 2018.
47. Gossip Cop [Электронный ресурс]. URL: http://gossipcop.com/ (дата обращения 15.04.2023).
48. PolitiFact [Электронный ресурс]. URL: https://www.politifact.com/ (дата обращения 15.04.2023).
49. Fake News [Электронный ресурс]. URL: https://www.kaggle.com/datasets/algord/fake-news (дата обращения: 20.04.2023).
50. Dou Y. User Preference-aware Fake News Detection // Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. – 2021.
51. Graph Ordering Attention Networks / M. Chatzianastasis, J.F. Lutzeyer, G. Dasoulas, et al. // arXiv preprint arXiv:2204.05351. – 2022.

52. Graph Neural Networks: Taxonomy, Advances, and Trends / Y. Zhou, H. Zheng, X. Huang, et al. // ACM Trans. Intell. Syst. Technol. – 2022, Vol. 13. – P. 1-54.
53. Improving Graph Attention Networks with Large Margin-based Constraints / G. Wang, R. Ying, J. Huang, et al. // arXiv preprint arXiv:1910.11945v1. – 2019.
54. Knyazev B., Tailor G.W., Amer M.R. Understanding Attention and Generalization in Graph Neural Networks // arXiv preprint arXiv:1905.02850. – 2019.
55. Reluplex made more practical: Leaky ReLU / J. Xu, Z. Li, B. Du, et al. // 2020 IEEE Symposium on Computers and Communications (ISCC). – 2020. – P. 1-7.
56. Attention Is All You Need / A. Vaswani, N. Shazeer, N. Parmar, et al. // arXiv preprint arXiv:1706.03762v5. – 2017.
57. Brody S., Alon U., Yahav E. How Attentive are Graph Attention Networks? // arXiv preprint arXiv:2105.14491v3. – 2022.
58. Lubanovic, B. Introducing Python, 2nd Edition. – S.: O'Reilly Media, 2019. – 630 p.
59. NumPy v1.24 Manual [Электронный ресурс]. URL: https://numpy.org/doc/stable/ (дата обращения: 25.04.2023).
60. pandas - Python Data Analysis Library [Электронный ресурс]. URL: https://pandas.pydata.org/ (дата обращения: 25.04.2023).
61. Matplotlib 3.7.1 documentation [Электронный ресурс]. URL: https://matplotlib.org/stable/contents.html (дата обращения: 25.04.2023).
62. PyTorch Documentation [Электронный ресурс]. URL: https://pytorch.org/docs/stable/index.html (дата обращения: 25.04.2023).
63. PyG Documentation [Электронный ресурс]. URL: https://pytorch-geometric.readthedocs.io/en/latest/ (дата обращения: 25.04.2023).
64. Ranjani J., Sheela A., Meena K.P. Combination of NumPy, SciPy and Matplotlib/Pylab - a good alternative methodology to MATLAB - A

Comparative analysis // 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT) (Singapore). – 2019. – P. 1-5.

65. A1gord/GraduationProject [Электронный ресурс]. URL: https://github.com/A1gord/GraduationProject (дата обращения: 27.04.2023).
66. Song C., Shu K., Wu B. Temporally evolving graph neural network for fake news detection // Information Processing & Management. – 2021, Vol.58. – P. 1-18.
67. Matsumoto H., Yoshida S., Muneyasu M. Propagation-Based Fake News Detection Using Graph Neural Networks with Transformer // 2021 IEEE 10th Global Conference on Consumer Electronics. – 2021. – P. 19-20.
68. Роскомнадзор оценил перспективы использования ИИ для борьбы с фейками [Электронный ресурс]. URL: https://www.sostav.ru/publication/gosudarstvo-i-ii-60234.html (дата обращения: 01.05.2023).
69. Allein L., Moens M.F., Perrotta D. Preventing profiling for ethical fake news detection // Information Processing & Management. – 2023, Vol.60. – P. 1-14.
70. Интервьюирование как метод социологического опроса [Электронный ресурс]. URL: https://spravochnick.ru/sociologiya/sociologicheskie_issledovaniya/intervyuirovanie_kak_metod_sociologicheskogo_oprosa/ (дата обращения 02.05.2023).
71. Зерчанинова Т.Е. Опрос социологический. – Е.: УрАГС, 2006. – 64 с.

ПРИЛОЖЕНИЕ А
РАСШИФРОВКА ИНТЕРВЬЮ С ЖУРНАЛИСТОМ

Здравствуйте! Расскажите, пожалуйста, о своем месте работы и опыте.

Журналист: «Здравствуйте! Я работаю в региональном СМИ уже более 6 лет».

1. Как часто вы сталкиваетесь с недостоверными новостями в своей работе и какие последствия они имеют для аудитории?

Журналист: «К сожалению, такие случаи происходят достаточно часто. Недостоверные новости могут иметь серьезные последствия для аудитории, включая недоверие к СМИ и ухудшение качества информирования общества».

2. Как вы определяете недостоверные новости и какие критерии вы используете для их проверки?

Журналист: «Для проверки новостей мы используем разные критерии, включая проверку источников, кросс-проверку информации и анализ контекста. Если есть сомнения в достоверности новости, то мы не публикуем ее до тех пор, пока не получим подтверждение ее достоверности».

3. Как вы считаете, какую роль играет искусственный интеллект в борьбе с фейками и недостоверными новостями?

Журналист: «Искусственный интеллект может играть важную роль в борьбе с фейками и недостоверными новостями. Он может помочь автоматизировать процесс проверки новостей и быстро обнаруживать недостоверную информацию».

4. Как вы относитесь к использованию искусственного интеллекта для обнаружения недостоверных новостей и какие преимущества и недостатки вы видите в этом подходе?

Журналист: «Я отношусь к использованию искусственного интеллекта для обнаружения недостоверных новостей положительно. Это позволяет сократить время на проверку новостей и улучшить качество информирования аудитории. Однако, есть риск, что искусственный интеллект может быть неправильно настроен и обнаруживать недостоверные новости, которые на самом деле являются правдивыми».

5. Какие этические проблемы могут возникнуть при использовании искусственного интеллекта в задаче обнаружения недостоверных новостей?

Журналист: «Использование искусственного интеллекта может привести к дискриминации, так как алгоритмы могут быть обучены на небольшом количестве данных, которые не учитывают разнообразие культурных и социальных контекстов. Это может привести к искажению результатов и дискриминации определенных групп людей. Кроме того, возможна этическая проблема в том, что искусственный интеллект может использоваться для манипуляции общественным мнением и распространения фейковых новостей».

6. Какие меры могут быть приняты для того, чтобы избежать негативных последствий использования искусственного интеллекта в борьбе с фейками?

Журналист: «Для того, чтобы избежать негативных последствий использования искусственного интеллекта в борьбе с фейками, необходимо установить строгие правила использования технологии и контролировать его работу».

7. Какие преимущества и недостатки имеет использование искусственного интеллекта в задаче обнаружения недостоверных новостей по сравнению с работой журналистов?

Журналист: «Искусственный интеллект может обрабатывать большие объемы информации и быстро определять недостоверные новости, что экономит время журналистов. Однако, ИИ не всегда может учесть контекст и эмоциональный подтекст новостей, что может привести к ошибкам. Кроме того, ИИ не может заменить журналистскую этику и моральные принципы, которые являются важными аспектами работы журналистов».

8. Какие изменения в законодательстве могут быть необходимы для регулирования использования искусственного интеллекта в борьбе с фейками?

Журналист: «Для регулирования использования искусственного интеллекта в борьбе с фейками может потребоваться изменение законодательства в области защиты персональных данных, ответственности за распространение ложной информации, а также прозрачности алгоритмов работы систем искусственного интеллекта. Кроме того, возможно потребуется разработка новых правил и стандартов для оценки эффективности и безопасности таких систем».

Спасибо за подробные ответы, было интересно узнать ваше мнение!

ПРИЛОЖЕНИЕ Б
РАСШИФРОВКА ИНТЕРВЬЮ С ПРОГРАММИСТОМ

Здравствуйте! Расскажите, пожалуйста, о своем месте работы и опыте.

Программист: «Здравствуйте! Я работаю программистом в области искусственного интеллекта примерно 5 лет».

1. Как вы считаете, какую роль играет искусственный интеллект в борьбе с недостоверными новостями?

Программист: «Искусственный интеллект играет очень важную роль в борьбе с недостоверными новостями. Он позволяет автоматически анализировать большие объемы информации, выделять ключевые факты и определять степень достоверности новостей».

2. Как вы относитесь к использованию искусственного интеллекта для обнаружения недостоверных новостей и какие преимущества и недостатки вы видите в этом подходе?

Программист: «Я отношусь положительно к использованию искусственного интеллекта для обнаружения недостоверных новостей, так как это позволяет повысить качество информационного потока и защитить людей от ложной информации. Однако, есть и некоторые недостатки, например, возможность ложных срабатываний и ошибок в классификации».

3. Как вы оцениваете проблему этики искусственного интеллекта в задаче обнаружения недостоверных новостей?

Программист: «Проблема этики в области искусственного интеллекта в задаче обнаружения недостоверных новостей является крайне актуальной и серьезной. Необходимо разработать этические стандарты для использования ИИ в таких задачах, чтобы избежать возможных негативных последствий для

общества. Кроме того, важно учитывать возможность искажения результата работы ИИ из-за предвзятости данных, на которых он обучается».

4. Какие последствия могут возникнуть, если искусственный интеллект неправильно классифицирует новости как недостоверные?

Программист: «Неправильная классификация новостей может привести к распространению ложной информации, что может негативно повлиять на общественное мнение и принятие решений. Также это может ухудшить доверие к системам искусственного интеллекта в целом. Для предотвращения таких последствий необходимо разработать более точные алгоритмы классификации и проводить регулярную проверку и обучение системы».

5. Какие этические проблемы могут возникнуть при использовании искусственного интеллекта в задаче обнаружения недостоверных новостей?

Программист: «При использовании ИИ в задаче обнаружения недостоверных новостей возможны проблемы с дискриминацией на основе расовой, половой или иной принадлежности. Алгоритмы могут быть обучены на данных, которые не являются репрезентативными для всего населения, что может привести к неправильному определению новостей как недостоверных. Необходимо убедиться, что данные, используемые для обучения, не содержат скрытых предубеждений и соответствуют разнообразию населения».

6. Какие меры можно предпринять для уменьшения рисков этических нарушений при использовании искусственного интеллекта в задаче обнаружения недостоверных новостей?

Программист: «Для уменьшения рисков этических нарушений при использовании искусственного интеллекта в задаче обнаружения

недостоверных новостей можно предпринять следующие меры: обеспечить прозрачность работы алгоритмов искусственного интеллекта, а также организовать мониторинг работы алгоритмов искусственного интеллекта, чтобы быстро выявлять и исправлять ошибки, и предотвращать возможные этические нарушения».

7. Какие требования должны быть установлены для разработчиков искусственного интеллекта, чтобы минимизировать возможность этических нарушений в задаче обнаружения недостоверных новостей?

Программист: «Разработчики искусственного интеллекта должны быть обучены этическим принципам и осведомлены о возможных негативных последствиях своих разработок. Также необходимо установить четкие правила использования ИИ в задаче обнаружения недостоверных новостей, а также проводить регулярные проверки на соответствие этим правилам».

8. Какие механизмы контроля за использованием искусственного интеллекта в задаче обнаружения недостоверных новостей могут быть введены для предотвращения этических нарушений?

Программист: «Можно ввести систему ответственности за публикацию недостоверной информации, которая будет включать штрафы и другие санкции. Необходимо также учитывать мнение общественности и экспертов при разработке и внедрении систем обнаружения недостоверных новостей на базе ИИ».

Спасибо за подробные ответы, было интересно узнать ваше мнение!

ПРИЛОЖЕНИЕ В
АНКЕТА ОПРОСА

1. Как часто вы сталкиваетесь с недостоверными новостями в социальных сетях?
   a) Почти каждый день.
   b) Несколько раз в месяц.
   c) Редко, но иногда.
   d) Не сталкиваюсь.

2. Какие меры вы считаете наиболее важными для решения проблемы распространения недостоверных новостей в социальных сетях?
   a) Усиление модерации контента.
   b) Обучение пользователей критическому мышлению.
   c) Применение ИИ-технологий для обнаружения недостоверных новостей.
   d) Другое.

3. Как вы относитесь к использованию искусственного интеллекта для обнаружения недостоверных новостей?
   a) Положительно.
   b) Нейтрально.
   c) Отрицательно.

4. Какую роль может сыграть ИИ в борьбе с распространением недостоверных новостей в социальной сети?
   a) Очень важную, ИИ может помочь обнаруживать и удалять недостоверные новости.
   b) Не очень важную, ИИ не может справиться с такой задачей.
   c) Затрудняюсь ответить.

5. Как вы оцениваете свой уровень осведомленности в вопросах этики ИИ?
   a) Хорошо осведомлен.
   b) Имею базовые знания.
   c) Впервые слышу.

6. Как вы оцениваете важность проблемы этики ИИ в задаче обнаружения недостоверных новостей в социальной сети?
   a) Это важная проблема.
   b) Не считаю эту проблему особо важной.
   c) Затрудняюсь ответить.

7. Как вы думаете, могут ли алгоритмы ИИ дискриминировать отдельных людей?
   a) Да, могут.
   b) Нет, не могут.
   c) Затрудняюсь ответить.

8. Согласны ли вы с утверждением, что ИИ, использующийся для обнаружения недостоверных новостей, должен основываться только на алгоритмах и данных, которые исключают возможность дискриминации отдельных категорий лиц и др. этические проблемы?
   a) Согласен.
   b) Не согласен.
   c) Затрудняюсь ответить.