УДК 002.304

Зиязетдинов Руслан Наилевич.

студент

4 курс, Национальный исследовательский университет

«Высшая школа экономики»,

департамент компьютерной инженерии

МИЭМ НИУ ВШЭ

АНАЛИЗ ЭФФЕКТИВНОСТИ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ КЛАССИФИКАЦИИ ТЕКСТОВЫХ КОММЕНТАРИЕВ О CHATGPT

Аннотация. В данной статье представлены результаты сравнительного исследования различных моделей машинного обучения для анализа тональности текстовых комментариев пользователей о ChatGPT. В работе оцениваются классические алгоритмы (Naive Bayes, Logistic Regression, SVM, Random Forest, CatBoost), нейросетевые модели (MLP, LSTM) и трансформеры (DistilBERT) на размеченном наборе данных, содержащем 219286 твитов. Исследовано влияние различных подходов к предобработке текста и количества классов на качество классификации. Модель DistilBERT показала наилучшие результаты с точностью 94,79% для трехклассовой задачи и 97% для бинарной классификации. Выявлено, что для нейросетевых моделей минимальная очистка текста дает лучшие результаты, в то время как для классических моделей более эффективна стандартная очистка. Представлены рекомендации по выбору оптимальной модели в зависимости от требований к качеству классификации и доступных вычислительных ресурсов.

Ключевые слова: анализ тональности текста, машинное обучение, глубокое обучение, нейронные сети, трансформеры, обработка естественного языка, классификация текста.

Abstract. This paper presents the results of a comparative study of various machine learning models for sentiment analysis of user text comments about ChatGPT. The research evaluates classical algorithms (Naive Bayes, Logistic Regression, SVM, Random Forest, CatBoost), neural network models (MLP, LSTM), and transformers (DistilBERT) on a labeled dataset containing 219,286 tweets. The study investigates the impact of different text preprocessing approaches and the number of classes on classification quality. The DistilBERT model demonstrated the best results with 94.79% accuracy for the three-class task and 97% for the binary classification task. The research revealed that minimal text cleaning provides better results for neural network models, while standard cleaning is more effective for classical models. Recommendations for selecting the optimal model based on classification quality requirements and available computational resources are provided.

Keywords: sentiment analysis, machine learning, deep learning, neural networks, transformers, natural language processing, text classification.

Введение

В современном мире диалоговые агенты становятся неотъемлемой частью цифровой экосистемы, революционизируя способы взаимодействия между людьми и компьютерами. Особую актуальность приобретает анализ пользовательских комментариев, который позволяет оценить качество взаимодействия и выявить проблемные аспекты диалоговых систем. Данная статья представляет результаты исследования эффективности различных моделей машинного обучения для анализа тональности текстовых комментариев пользователей о ChatGPT.

Актуальность исследования

Стремительный рост использования диалоговых агентов наблюдается в различных сферах человеческой деятельности, включая обслуживание клиентов, маркетинг и образование. По данным исследований, интеграция

ИИ-чатботов в маркетинговые платформы малых и средних предприятий привела к революционным изменениям в обслуживании клиентов за счет автоматизации рутинных задач, обеспечения круглосуточной доступности и персонализации взаимодействий.

Однако современные диалоговые системы часто страдают от недостатка правдивости, надежности и способности анализировать ход диалога, что подчеркивает критическую важность анализа пользовательских комментариев для их совершенствования. Анализ тональности в диалогах играет ключевую роль в определении эмоционального тона на протяжении всего разговора, что может значительно улучшить взаимодействие человека с компьютером.

С ростом использования диалоговых агентов увеличивается и объем генерируемых пользовательских комментариев, что создает потребность в автоматизации их анализа. Традиционные методы ручного анализа становятся неэффективными в условиях масштабных данных, что обуславливает необходимость разработки эффективных моделей машинного обучения для этой задачи.

Методология исследования

В исследовании использовался набор данных, содержащий 219286 твитов о ChatGPT с метками тональности, полученный с платформы Kaggle[1]. Выбор данного датасета обусловлен популярностью Twitter как платформы для анализа тональности и актуальностью изучения общественного мнения о ChatGPT [2].

Анализ распределения классов выявил следующую картину:

- "bad" (негативные) - 107796 записей (49%)

- "good" (позитивные) - 56011 записей (26%)

- "neutral" (нейтральные) - 55487 записей (25%)

Для повышения качества работы моделей были реализованы два подхода к предобработке текста:

1. *Стандартная очистка*, включающая:

- Приведение текста к нижнему регистру
- Удаление URL-адресов, упоминаний пользователей, хэштегов
- Удаление специальных символов и пунктуации
- Удаление лишних пробелов

2. *Минимальная очистка*, сохраняющая больше исходной информации:

- Приведение текста к нижнему регистру
- Удаление только URL-адресов
- Удаление лишних пробелов

Для векторизации текстовых данных использовались методы Bag of Words (BoW) и TF-IDF. Для нейросетевых моделей применялась токенизация с ограничением словаря до 10000 наиболее частых слов.

Исследуемые модели

В рамках исследования были разработаны и проанализированы следующие модели:

1. Классические модели машинного обучения:

- Наивный байесовский классификатор (Naive Bayes)
- Логистическая регрессия (Logistic Regression)
- Метод опорных векторов (SVM)
- Случайный лес (Random Forest)

- CatBoost

2. Нейросетевые модели:

- Многослойный персептрон (MLP)

- Рекуррентные нейронные сети с долгой краткосрочной памятью (LSTM)

3. Трансформеры:

- DistilBERT

Для каждой модели были проведены эксперименты с различными подходами к предобработке текста и различным количеством классов (трехклассовая и двухклассовая задачи).

Результаты исследования

Результаты исследования приведены на *Рисунок 1*.

```
--- Сводная таблица результатов (Accuracy и F1 Weighted) ---
```

Эксперимент	Модель	Accuracy	F1 Weighted
2class_minimal_clean	CatBoost (TF-IDF Small)	0.8759	0.8758
	DistilBERT	0.9689	0.9689
	LSTM	0.9537	0.9536
	Logistic Regression (TF-IDF)	0.9271	0.9271
	MLP (TF-IDF Small)	0.9086	0.9086
	Naive Bayes (BoW)	0.8434	0.8431
	Random Forest (TF-IDF Small)	0.8020	0.7995
	SVM (TF-IDF)	0.9398	0.9398
2class_standard_clean	CatBoost (TF-IDF Small)	0.8666	0.8663
	DistilBERT	0.9626	0.9626
	LSTM	0.9416	0.9416
	Logistic Regression (TF-IDF)	0.9179	0.9179
	MLP (TF-IDF Small)	0.8965	0.8965
	Naive Bayes (BoW)	0.8366	0.8363
	Random Forest (TF-IDF Small)	0.7962	0.7933
	SVM (TF-IDF)	0.9285	0.9285
3class_minimal_clean	CatBoost (TF-IDF Small)	0.7422	0.7254
	DistilBERT	0.9479	0.9479
	LSTM	0.9303	0.9306
	Logistic Regression (TF-IDF)	0.8143	0.8042
	MLP (TF-IDF Small)	0.8444	0.8435
	Naive Bayes (BoW)	0.7366	0.7280
	Random Forest (TF-IDF Small)	0.7176	0.6955
	SVM (TF-IDF)	0.8312	0.8230
3class_standard_clean	CatBoost (TF-IDF Small)	0.7355	0.7171
	DistilBERT	0.9392	0.9390
	LSTM	0.9107	0.9109
	Logistic Regression (TF-IDF)	0.8032	0.7919
	MLP (TF-IDF Small)	0.8270	0.8253
	Naive Bayes (BoW)	0.7309	0.7208
	Random Forest (TF-IDF Small)	0.7074	0.6839
	SVM (TF-IDF)	0.8177	0.8081

Рисунок 1 - Результаты экспериментов

Наилучшие результаты среди всех моделей показала модель DistilBERT с точностью (Accuracy) 94,79% и F1-score 94,79% для трехклассовой задачи с минимальной предобработкой текста. Для двухклассовой задачи DistilBERT достигает точности 97% и F1-score 97%.

Модель LSTM также продемонстрировала высокую эффективность с точностью 93,03% и F1-score 93,03% для трехклассовой задачи, что значительно превосходит результаты классических моделей.

Среди классических моделей наилучшие результаты показал SVM с точностью 83,10% и F1-score 83,10% для трехклассовой задачи со стандартной предобработкой текста.

Влияние предобработки текста

Исследование показало, что для нейросетевых моделей (DistilBERT, LSTM) минимальная очистка текста, сохраняющая пунктуацию, эмодзи и другие специальные символы, дает лучшие результаты. Это подтверждает гипотезу о том, что эти элементы несут эмоциональную окраску, полезную для определения тональности комментария.

Для классических моделей, напротив, более эффективной оказалась стандартная очистка текста. Это объясняется тем, что классические алгоритмы лучше работают с более структурированными и очищенными данными, в то время как нейросетевые модели способны извлекать полезную информацию из "шумных" данных.

Влияние количества классов

Переход от трехклассовой к двухклассовой задаче улучшает результаты всех моделей, но это улучшение менее значительно для нейросетевых моделей (2-3% для DistilBERT и LSTM) по сравнению с классическими моделями (10-14% для SVM, Logistic Regression и CatBoost).

Это свидетельствует о лучшей способности нейросетевых моделей различать нейтральные комментарии, которые представляют наибольшую сложность для классификации.

Анализ ошибок классификации

Все модели показали наилучшие результаты для класса "bad" (негативные комментарии) и наихудшие для класса "neutral" (нейтральные комментарии). Это может быть связано как с большим количеством примеров негативного класса в обучающей выборке, так и с размытостью границ между нейтральными и другими комментариями.

Типичные ошибки классификации включают:

- Комментарии со смешанной тональностью

- Сарказм и ирония

- Технические термины

- Контекстуально-зависимые выражения

Рекомендации по выбору оптимальной модели

На основе проведенного исследования можно сформулировать следующие рекомендации по выбору оптимальной модели для анализа текстовых комментариев о ChatGPT:

1. При высоких требованиях к качеству классификации рекомендуется использовать модель DistilBERT с минимальной предобработкой текста. Эта модель обеспечивает наилучшие результаты для всех классов, особенно для сложного класса "neutral".

2. При ограниченных вычислительных ресурсах оптимальным выбором будет SVM или логистическая регрессия со стандартной предобработкой

текста. Эти модели обеспечивают хороший баланс между качеством и вычислительной эффективностью.

3. Для задач, требующих быстрой обработки большого количества комментариев в реальном времени, рекомендуется использовать логистическую регрессию, которая обеспечивает высокую скорость предсказания при достаточно высоком качестве классификации.

4. Если важна максимальная точность классификации, рекомендуется использовать двухклассовую постановку задачи (negative, non-negative), которая обеспечивает более высокие значения метрик качества для всех моделей.

5. Если требуется более детальный анализ тональности комментариев, рекомендуется использовать трехклассовую постановку задачи (bad, good, neutral) с применением моделей DistilBERT или LSTM, которые показывают высокие результаты даже для сложного класса "neutral".

Заключение

Проведенное исследование демонстрирует высокую эффективность современных моделей машинного обучения, особенно трансформеров, для задачи анализа тональности текстовых комментариев пользователей о ChatGPT. Модель DistilBERT показала наилучшие результаты с точностью 94,79% для трехклассовой задачи и 97% для двухклассовой задачи.

Важным результатом исследования является выявление влияния предобработки текста на качество моделей: для нейросетевых моделей минимальная очистка текста дает лучшие результаты, в то время как для классических моделей более эффективна стандартная очистка.

Практическая значимость работы заключается в разработке и сравнительном анализе различных моделей машинного обучения для

анализа тональности текстовых комментариев, что может быть использовано для мониторинга общественного мнения о ChatGPT и других технологиях искусственного интеллекта, улучшения пользовательского опыта, выявления проблемных аспектов и оценки эффективности внесенных изменений.

Направления дальнейших исследований могут включать применение более сложных архитектур трансформеров, исследование методов обработки несбалансированных данных, разработку многоязычных моделей и применение методов объяснимого искусственного интеллекта для интерпретации решений моделей.

Использованные источники

1. ChatGPT sentiment analysis dataset [Электронный ресурс]. URL: https://www.kaggle.com/datasets/charunisa/chatgpt-sentiment-analysis (дата обращения: 11.05.2025).
2. Social Media Sentiment Analysis Using Twitter Datasets [Электронный ресурс]. URL: https://www.datasciencecentral.com/social-media-sentiment-analysis-using-twitter-datasets/ (дата обращения: 11.05.2025).
3. Зиязетдинов Р.Н. Разработка моделей машинного обучения для анализа текстовых комментариев, часть 2: Python-ноутбук. Google Colab. [Электронный ресурс] URL: https://colab.research.google.com/drive/1BbkVN5pT-H5q2KffE3BZot4mFmedg30-?usp=sharing (дата обращения: 11.05.2025).

Информация о себе: senior.ziyazetdinov@gmail.com