

АНАЛИЗ СЕНТИМЕНТА ФИНАНСОВЫХ НОВОСТНЫХ ПУБЛИКАЦИЙ С ИСПОЛЬЗОВАНИЕМ FinBERT

САЯПИН Артём Вячеславович
аспирант

ЯМАШКИН Станислав Анатольевич
кандидат технических наук, доцент

Мордовский государственный университет им. Н.П. Огарева
г. Саранск, Россия

В данной статье анализируется эффективность применения предобученных моделей анализа сентимента финансовых новостей. Произведен обзор основных исследований и моделей в данной области, собран набор данных, произведена автоматическая разметка новостей и оценены результаты качества моделей.

Ключевые слова: анализ сентиментов, обработка естественного языка, большие языковые модели, классификация текстов, финансовые новости, машинное обучение.

С развитием Интернета и всеобщего доступа к информационным ресурсам у инвесторов появился мгновенный доступ к новой информации о компаниях, что позволяет им почти моментально принимать решения. Информация о дивидендах [5], новостные публикации [3] оказывают существенное влияние на стоимость ценных бумаг компании.

В данном исследовании рассматривается использование предобученных больших языковых моделей (LLM, Large Language Models) для классификации сентимента новостей на русском языке. Целью исследования является оценка применимости больших языковых моделей для русскоязычных текстов и анализ эффективности моделей в предсказании динамики цены как прокси сентимента.

Исследования, связывающие текстовую новостную информацию и динамику финансовых рынков, получили значительное развитие в XXI в. В статье [6] было показано, что негативный тон в колонках The Wall Street Journal статистически значимо предшествует понижению рыночных индексов.

Развитие моделей машинного обучения архитектуры Transformer [8] привело к созданию доменно-специфичных решений для анализа новостной информации. Одной из ключевых и наиболее цитируемых работ в области анализа сентимента финансовых новостей является FinBERT [1] – версия BERT, обученная на наборе данных FinancialPhraseBank [4].

В 2023 г. был представлен BloombergGPT [9] – большая языковая модель с 50 миллиардами параметров, обученная на внутренних данных Bloomberg, включая новости, пресс-релизы и регуляторные документы. В отличие от FinBERT данная модель не находится в открытом доступе, поэтому оценить качество её ответов не представляется возможным.

Основной проблемой для анализа новостных текстов экономической направленности является отсутствие больших размеченных наборов данных для обучения моделей машинного обучения и тематических словарей тональности на русском языке. В данной работе новостные данные были размечены на основе динамики цен акций после публикации в дневной временной интервал.

Исторические данные по ценам акций были получены с помощью API Московской Биржи за период с 2011 по 2025 гг.

Для анализа эффективности модели FinBERT и её дообученной версии FinBERT-tone [2] был собран набор новостных данных путем парсинга новостей с сайта новостного интернет-издания Лента.ру из разделов Экономика, Финансы и Бизнес. Модель FinBERT обучалась на англоязычных данных, поэтому текст новостей был автоматически переведен на английский с помощью модели OPUS-MT [7].

Итоговый набор данных включает новости о компаниях Газпром, Роснефть, Сбер, ВТБ, Аэрофлот, МТС, Яндекс, Северсталь, Совкомфлот, Интер-РАО, Лукойл, Московская биржа,

НЛМК, X5, Вуш, Т-банк, Татнефть, Сургут-нефтегаз, Русал, Ростелеком, Ренессанс страхование, Полюс-золото, Пик, Новатэк.

Для сравнения качества оценки тональности новости была использована выборка новостей за период с 1 января 2011 г. до 31 декабря 2024 г. В

выборке 1600 наблюдений негативного класса и 1847 позитивного класса.

В качестве метрик качества использовались accuracy, ROC-AUC, precision, recall, PR-AUC и F1-мера. Сравнение метрик бинарной классификации для моделей приведено в таблице 1.

Таблица 1

СРАВНЕНИЕ МОДЕЛЕЙ АНАЛИЗА СЕНТИМЕНТА НОВОСТЕЙ

| Метрики\Модели | FinBERT | FinBERT-tone |
|----------------|---------|--------------|
| Accuracy | 0.54 | 0.54 |
| Precision | 0.57 | 0.58 |
| Recall | 0.59 | 0.44 |
| F1 | 0.58 | 0.51 |
| ROC-AUC | 0.55 | 0.55 |

Оригинальный FinBERT показывает лучшие значения метрик по сравнению с его дообученной версией. Результаты на наборе данных с автоматической разметкой заметно ниже значений из тестового набора на основе ручной разметки из оригинальной статьи. Это

может быть объяснено тем, что для оценки изменений цены акции на основе новостной информации по компании дополнительно необходимо учитывать контекст, финансовые показатели, макроэкономическую ситуацию.

СПИСОК ЛИТЕРАТУРЫ

1. *Araci D.* Finbert: Financial sentiment analysis with pre-trained language models //arXiv preprint arXiv:1908.10063. 2019.
2. *Huang A.H., Wang H., Yang Y.* FinBERT: A large language model for extracting information from financial text //Contemporary Accounting Research. 2023. Т. 40. № 2. – P. 806-841.
3. *Li X. et al.* News impact on stock price return via sentiment analysis //Knowledge-Based Systems. 2014. Т. 69. P. 14-23.
4. *Malo P. et al.* Good debt or bad debt: Detecting semantic orientations in economic texts //Journal of the Association for Information Science and Technology. 2014. Т. 65. № 4. P. 782-796.
5. *Suwanna T.* Impacts of dividend announcement on stock return //Procedia-Social and Behavioral Sciences. 2012. Т. 40. P. 721-725.
6. *Tetlock P.C.* Giving content to investor sentiment: The role of media in the stock market //The Journal of finance. 2007. Т. 62. № 3. P. 1139-1168.
7. *Tiedemann J., Thottingal S.* OPUS-MT--Building open translation services for the World //Annual Conference of the European Association for Machine Translation. European Association for Machine Translation, 2020. P. 479-480.
8. *Vaswani A. et al.* Attention is all you need //Advances in neural information processing systems. 2017. Т. 30.
9. *Wu S. et al.* Bloomberggpt: A large language model for finance // arXiv preprint arXiv: 2303.17564. 2023.

SENTIMENT ANALYSIS OF FINANCIAL NEWS PUBLICATIONS WITH FinBERT

SAYAPIN Artem Vyacheslavovich

Postgraduate

YAMASHKIN Stanislav Anatolyevich

Candidate of Sciences in Technology, Associate Professor

N.P. Ogarev Mordovia State University

Saransk, Russia

This article analyzes the performance of using natural language processing methods to analyze sentiment contained in financial news. A review of the literature of research in this area was carried out, a dataset was collected, data was manually labeled, and a comparison of models was made.

Keywords: sentiment analysis, natural language processing, large language models, text classification, financial news, machine learning.
