

**ПРИМЕНЕНИЕ МОДЕЛЕЙ АРХИТЕКТУРЫ ТРАНСФОРМЕР
ДЛЯ РАСПОЗНАВАНИЯ ФИКТИВНЫХ ОТЗЫВОВ
ПРИ ОГРАНИЧЕННОМ ОБЪЕМЕ ДАННЫХ**

Фрид З.А. (ИТМО)

**Научный руководитель – кандидат технических наук Русак А.В.
(ИТМО)**

Введение. В современном информационном обществе огромное количество потребителей полагается на интернет-рейтинги и отзывы при принятии решений о покупках или выборе услуг. В связи с этим, фиктивные отзывы представляют из себя не только морально-этическую проблему, но и имеют серьезное влияние на экономику. Дезинформация, вызванная фиктивными отзывами, может серьезно исказить рыночную конкуренцию и доверие потребителей [1]. Одним из перспективных направлений борьбы с фиктивными отзывами является применение машинного обучения. Современные трансформерные модели [2] семейства BERT демонстрируют высокую эффективность в задачах обработки текста, однако их потенциал в детекции фиктивных отзывов остается недостаточно изученным. Особый интерес вызывают подходы, позволяющие эффективно применять предобученные модели для решения этой задачи в условиях дефицита размеченных данных и вычислительных ресурсов.

Основная часть. В рамках данного исследования были рассмотрены и протестированы четыре модели на основе трансформеров для классификации фиктивных отзывов: BERT [3], RoBERTa [4], DistilBERT [5] и ALBERT [6]. В качестве набора данных использовались отзывы на товары с платформы Amazon [7], содержащие около 490 тысяч записей с настоящими и поддельными отзывами.

Для оценки точности моделей проводились эксперименты с разными объемами данных — 5% и 20% от исходного набора (около 24 и 98 тысяч отзывов соответственно). Данные были предварительно обработаны, включая конвертацию формата, очистку текста от лишних символов и токенизацию, встроенную в соответствующие модели. Наборы данных были разделены на обучающую, валидационную и тестовую выборки в пропорции 80:10:10. Оценка качества классификации проводилась с использованием взвешенного F1-score. Кроме того, фиксировалось время обучения моделей.

По результатам экспериментов на 5% выборке наилучшую точность показали BERT с 0.917 и RoBERTa с 0.932, на 20% выборке лучшими также оказались BERT и RoBERTa с точностью 0.922 и 0.935 соответственно. Модели ALBERT и DistilBERT показали наименьшее время обучения на 5% выборке, а на 20% быстрее всего обучились RoBERTa и DistilBERT. Модель RoBERTa демонстрирует наиболее оптимальные результаты по балансу точности и времени обучения.

Выводы. В данном исследовании был проведен анализ и оценка эффективности предобученных моделей трансформеров в контексте решения задачи автоматического распознавания фиктивных отзывов в условиях ограниченного объема размеченных данных и вычислительных ресурсов.

Список использованных источников:

1. Mohawesh R., Xu S., Springer M., Jararweh Y., Al-Hawawreh M., Maqsood S. An explainable ensemble of multi-view deep learning model for fake review detection //Journal of King Saud University-Computer and Information Sciences. – 2023. – Т. 35. – №. 8. – С. 101644.
2. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser L., Polosukhin I. Attention is all you need //Advances in neural information processing systems. – 2017. – Т. 30.
3. Devlin J., Chang M. W., Lee K., Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding //arXiv:1810.04805. – 2018.

4. Liu Y., Ott M., Goyal N., Du J., Joshi M., Chen D., Stoyanov V. Roberta: A robustly optimized bert pretraining approach //arXiv:1907.11692. – 2019.
5. Sanh V., Debut L., Chaumond J., Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter //arXiv:1910.01108. – 2019.
6. Lan Z., Chen M., Goodman S., Gimpel K., Sharma P., Soricut R. Albert: A lite bert for self-supervised learning of language representations //International Conference on Learning Representations (ICLR). – 2020.
7. Hussain N., Mirza H. T., Hussain I., Iqbal F., Memon I. Spam review detection using the linguistic and spammer behavioral methods //IEEE Access. – 2020. – T. 8. – C. 53801-53816.