

УДК 004.056

РАЗРАБОТКА АЛГОРИТМА ОБНАРУЖЕНИЯ ЛОЖНЫХ НОВОСТЕЙ НА ОСНОВЕ АНАЛИЗА ТЕКСТОВ И МУЛЬТИМЕДИА С ИСПОЛЬЗОВАНИЕМ КОЛЛАБОРАТИВНОЙ ФИЛЬТРАЦИИ НА БАЗЕ БЛОКЧЕЙНА

Гаврилова В. В. (Университет ИТМО)

Научный руководитель – Менщиков А. А.
(Университет ИТМО)

Введение. Развитие технологий, доступность Интернета и растущая популярность социальных сетей стали непреднамеренными путями распространения ложных новостей. Во всем мире люди используют ложные новости в своих корыстных целях, либо представляя информацию в неверном виде для привлечения большего числа зрителей, либо манипулируя мнением общественности, чтобы повлиять на ее поведение во время важных событий в обществе, например, выборов. Для решения этой проблемы были предложены различные алгоритмы с использованием искусственного интеллекта и методов машинного обучения, включающих обработку естественного языка, методы ансамбля, классификацию и т. д. Несмотря на все усилия, большинство систем могут анализировать только текст на английском языке и непригодны для анализа новостей на русском языке. Так же важно не только обучить и протестировать модель, но и найти ей практическое применение для улучшения медиа пространства каждого пользователя.

Основная часть. В исследовании решаются следующие проблемы:

1. Произведены анализ существующих методов определения ложных новостей и оценка их точности.
2. Были составлены датасеты на русском языке, а также предобработаны все текстовые данные.
3. Разработан алгоритм обнаружения ложных новостей на основе анализа текстов и мультимедиа с использованием коллаборативной фильтрации на базе блокчейна.
4. Проведен эксперимент и анализ результатов, а также создан макет платформы оценки новостей на базе блокчейна.

В результате проведенного исследования была разработана модель обнаружения ложных новостей на основе анализа текстов и мультимедиа с использованием коллаборативной фильтрации, содержащая многослойный перцептрон для бинарной классификации текстов (ложные/правдивые), многослойный перцептрон для мультиклассовой классификации типа ложных новостей, рекомендательная система на основе сети ResNet50.

Тренировка и оценка моделей была проведена на собранных русскоязычных текстовых данных и изображениях из русскоязычных новостей. Все собранные данные выложены на открытый ресурс Kaggle.

Предложенная модель показывает достойный уровень точности, по сравнению с аналогичными англоязычными системами [1],[2],[3],[4],[5]. Рассчитанная точность составляет 95,73%.

Разработанный метод не только определяет правдивость новости с высокой точностью, но и может быть использован на практике для улучшения общего медийного пространства. Так как точность глубокого обучения не может достигать 100%, разработанная система предоставляет возможность независимым и анонимным модераторам определить статус новости на основе предсказания алгоритма глубокого обучения. Опираясь на результаты анализа текста и рекомендательной системы анализа изображения модератор может подтвердить или опровергнуть оценку новости.

Система может использоваться для безопасного и анонимного распространения правдивого новостного контента, гарантируя отсутствие цензуры и предвзятости.

Выводы. По результатам исследований был произведен анализ полученных результатов и их сравнение с работами других авторов. Точность предложенного алгоритма уступает на 0,2% самому точному алгоритму, однако предложенная система обучена и протестирована на русскоязычном датасете, что является преимуществом. Так же был разработан макет пользовательской стороны платформы на базе блокчейн и описан весь функционал предлагаемой платформы [6].

Список использованных источников:

1. V. M. Krešňáková, M. Sarnovský and P. Butka, "Deep learning methods for Fake News detection," 2019 IEEE 19th International Symposium on Computational Intelligence and Informatics and 7th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics (CINTI-MACRo), Szeged, Hungary, 2019, pp. 000143-000148, doi: 10.1109/CINTI-MACRo49179.2019.9105317.
2. R. K. Kaliyar, P. Kumar, M. Kumar, M. Narkhede, S. Namboodiri and S. Mishra, "DeepNet: An Efficient Neural Network for Fake News Detection using News-User Engagements," 2020 5th International Conference on Computing, Communication and Security (ICCCS), Patna, India, 2020, pp. 1-6, doi: 10.1109/ICCCS49678.2020.9277353.
3. H. Saleh, A. Alharbi and S. H. Alsamhi, "OPCNN-FAKE: Optimized Convolutional Neural Network for Fake News Detection," in IEEE Access, vol. 9, pp. 129471-129489, 2021, doi: 10.1109/ACCESS.2021.3112806.
4. F. Najar, N. Zamzami and N. Bouguila, "Fake News Detection Using Bayesian Inference," 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI), Los Angeles, CA, USA, 2019, pp. 389-394, doi: 10.1109/IRI.2019.00066.
5. S. Ni, J. Li and H. -Y. Kao, "MVAN: Multi-View Attention Networks for Fake News Detection on Social Media," in IEEE Access, vol. 9, pp. 106907-106917, 2021, doi: 10.1109/ACCESS.2021.3100245.
6. <https://pravdaru.my.canva.site>

Гаврилова В. В. (автор)

Подпись

Менщиков А. А. (научный руководитель)

Подпись