

## АВТОМАТИЧЕСКОЕ ОБНАРУЖЕНИЕ ФЕЙКОВЫХ НОВОСТЕЙ КАК ЗАДАЧА КЛАССИФИКАЦИИ ТЕКСТОВ

*Авагян Давид Арменович*

*Студент*

*Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия*

*E-mail: david\_avagyan@list.ru*

*Научный руководитель — Волкова Ирина Анатольевна*

Главными свойствами информации являются актуальность и достоверность. Примером популярного источника информации может служить новостной ресурс. Объёмы новостных потоков уже не позволяют производить модерацию новостей вручную; также не формализовано понятие фейка.

Цель исследования — описание системы автоматического распознавания и классификации фейковых новостей на русском языке. В ходе исследования подтвердилась гипотеза о том, что зачастую для определения достоверности новости достаточно её лингвистического анализа. Существующие определения и классификации фейков [1, 2] не подходят для автоматической обработки текстов. В статье [3] предложены определение и классификация фейковых новостей, послужившие основой данного исследования.

В настоящей работе *новость* — текст публицистического стиля, созданный с целью освещения событий и явлений действительности. Важные свойства достоверных новостей — наличие официального проверенного источника и объективность изложения. *Фейковая новость (фейк)* — веб-контент, распространяемый под видом новостей, но содержащий ложную информацию или преследующий иные цели, в отличие от достоверных новостей. В частности, фейки могут создаваться для дезинформации, выражения субъективной позиции автора, в рекламных или юмористических целях.

Выделяются следующие разновидности фейковых новостей: *ложные новости, поляризованный контент, сатирические новости, непреднамеренно искажённые факты, комментарии и мнения, нативная реклама и гражданская журналистика* [3]. В рамках исследования автоматически собраны и размечены фейки четырёх типов: сатира, комментарии, нативная реклама и гражданская журналистика — для них не требуется решение сложной задачи проверки фактов. Более 57 000 новостей на русском языке, собранных из

десяти источников<sup>1</sup>, объединены в датасет **TopFakeSet**, доступный в репозитории на **GitLab**<sup>2</sup>. **TopFakeSet** стал одной из немногих открытых коллекций новостей на русском языке и превосходит по объёму многие датасеты фейковых новостей.

**TopFakeSet** послужил выборкой для обучения и тестирования классификаторов новостей. Основной метрикой качества выбрана (англ. *soft accuracy*) — модификация метрики точности, частично поощряющая модель в случае отнесения фейков одного типа к другому. Лучшим классификатором<sup>3</sup> стал ансамбль из градиентного бустинга, логистической регрессии и перцептрона — качество на тестовой выборке превысило 96.5%.

Возможны несколько направлений дальнейшего исследования: построение классификатора новостей, выделяющего в тексте  $n$ -граммы и конструкции, указывающие на его фейковость, или уточнение классификации фейков, в которой большую роль сыграют лингвистические признаки и база знаний, позволяющая судить о справедливости фактов и логичности их изложения.

### Литература

1. Zellers R., Holtzman A., Rashkin H., Bisk Y., Farhadi A., Roesner F., Choi Y. Defending against neural fake news // *Advances in Neural Information Processing Systems* 32, 2019. P. 9054–9065.
2. Shu K., Silva A., Wang S., Tang J., Liu H. Fake news detection on social media: a data mining perspective // *ACM SIGKDD Explorations Newsletter*, 2017. Vol. 19, № 1, P. 22–36.
3. Molina M. D., Sundar S. S., Le T., Lee D. is not simply false information: a concept explication and taxonomy of online content // *American Behavioral Scientist*, 2019. Vol. 65, № 2, P. 180–212.

---

<sup>1</sup>[panorama.pub](#) (сатира), [wikinews.org](#) (гражданская журналистика), [vz.ru](#), [newsru.com](#), [ria.ru](#), [iz.ru](#) (комментарии), [pikabu.ru](#), [meduza.io](#) (нативная реклама), [interfax.ru](#), [lenta.ru](#) (достоверные новости).

<sup>2</sup>[gitlab.com/Mathematician2000/topfakeset](#)

<sup>3</sup>[top-fake-set.herokuapp.com](#)