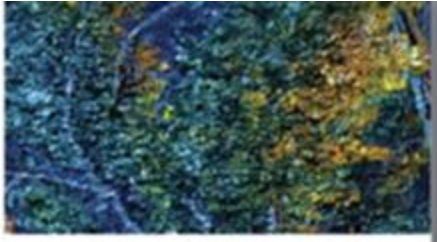

№1 (6) 2025

INTERNATIONAL SCIENCE REVIEWS

AIU

Natural Sciences and Technologies series

www.aiu.kz

ISSN: 2707-4862

ASTANA INTERNATIONAL UNIVERSITY

INTERNATIONAL SCIENCE REVIEWS
Natural Sciences and Technologies series

Has been published since 2020

№1 (6) 2025

Astana

INTERNATIONAL SCIENCE REVIEWS. NATURAL SCIENCES AND TECHNOLOGIES SERIES ЖУРНАЛЫНЫҢ РЕДАКЦИЯСЫ

БАС РЕДАКТОР

Қалимолдаев Мақсат Нұрадилович, техникалық ғылымдар докторы, ҚР ҰҒА академигі, профессор, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты бас директорының кеңесшісі, бас ғылыми қызметкері (*Қазақстан*)

БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ

Мырзағалиева Анар Базаровна, биология ғылымдарының докторы, профессор, бірінші вице-президент, Астана халықаралық университеті (*Қазақстан*);

РЕДАКТОРЛАР:

- **Сейткан Айнур Сейтканқызы**, техника ғылымдарының кандидаты, PhD, жаратылыстану ғылымдары жоғары мектебінің деканы, Астана халықаралық университеті (*Қазақстан*);
- **Муканова Асель Сериковна,** PhD, Ақпараттық технологиялар және инжерения жоғары мектебінің деканы, Астана халықаралық университеті (*Қазақстан*);
- **Абдилдаева Асель Асылбековна,** PhD, қауымдастырылған профессор, Әл-Фараби атындағы ҚазҰУ (Қазақстан);
- **Хлахула Иржи** PhD, профессор, Познаньдағы Адам Мицкевич атындағы университет (Польша);
- **Редферн Саймон А.Т.,** PhD, профессор, Наньян технологиялық университеті (Сингапур);
- **Сяолей Фенг,** PhD, Наньян технологиялық университеті (Сингапур);
- **Шуджаул Мулк Хан,** PhD, профессор, Каид-және-Азам университеті (Пакистан);
- **Базарнова Наталья Григорьевна,** химия ғылымдарының докторы, профессор, Химия және химиялық-фармацевтикалық технологиялар институты (Ресей);
- **Черёмушкина Вера Алексеевна,** биология ғылымдарының докторы, профессор, РҒА СБ Орталық Сібір ботаникалық бағы (Ресей);
- **Тасболатұлы Нұрболат,** PhD, Ақпараттық технологиялар және инжерения жоғары мектебі деканының орынбасары, Астана халықаралық университеті (Қазақстан);
- **Байшоланов Сакен Советович,** география ғылымдарының кандидаты, доцент, Астана халықаралық университеті (*Қазақстан*);
- **Нуркенов Серик Амангельдинович,** PhD, қауымдастырылған профессор, Астана халықаралық университеті (*Қазақстан*).

РЕДАКЦИЯ ЖУРНАЛА INTERNATIONAL SCIENCE REVIEWS. NATURAL SCIENCES AND TECHNOLOGIES SERIES

ГЛАВНЫЙ РЕДАКТОР

Калимолдаев Максат Нурадилович, доктор технических наук, академик НАН РК, профессор, ГНС, советник генерального директора Института информационных и вычислительных технологии КН МНВО РК (*Казахстан*)

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА

Мырзагалиева Анар Базаровна, доктор биологических наук, профессор, первый вице-президент, Международный университет Астана (*Казахстан*)

РЕДАКТОРЫ:

- **Сейткан Айнур Сейтканкызы**, кандидат технических наук, PhD, декан высшей школы естественных наук, Международный университет Астана (Казахстан);
- **Муканова Асель Сериковна,** PhD, декан Высшей школы информационных технологии и инженерии, Международный университет Астана (Казахстан);
- **Абдилдаева Асель Асылбековна,** PhD, ассоциированный профессор, Казахский национальный университет имени Аль-Фараби (Казахстан);
- **Хлахула Иржи** PhD, профессор, Университет имени Адама Мицкевича в Познани (Польша);
- **Редферн Саймон А.Т.,** PhD, профессор, Наньянский технологический университет (Сингапур);
- **Фенг Сяолей,** PhD, Наньянский технологический университет (Сингапур);
- **Шуджаул Мулк Хан,** PhD, профессор, Университет Каид-и Азама (Пакистан);
- **Базарнова Наталья Григорьевна,** доктор химических наук, профессор, Институт химии и химико-фармацевтических технологий (Россия);
- **Черёмушкина Вера Алексеевна,** доктор биологических наук, профессор, Центральный Сибирский Ботанический сад СО РАН (Россия);
- **Тасболатұлы Нұрболат,** PhD, заместитель декана Высшей школы информационных технологии и инженерии, Международный университет Астана (Казахстан);
- **Байшоланов Сакен Советович,** кандидат географических наук, доцент, Международный университет Астана (Казахстан);
- **Нуркенов Серик Амангельдинович,** PhD, ассоциированный профессор, Международный университет Астана (Казахстан);

EDITORIAL TEAM OF THE JOURNAL INTERNATIONAL SCIENCE REVIEWS. NATURAL SCIENCES AND TECHNOLOGIES SERIES

CHIEF EDITOR

Maksat Kalimoldayev, Doctor of Technical Sciences, Academician of NAS RK, Professor, SRF, CEO's councelor «The Institute of Information and Computational Technologies» CS MSHE RK (Kazakhstan)

DEPUTY CHIEF EDITOR

Anar Myrzagaliyeva, Doctor of Biological Sciences, Professor, First Vice-President, Astana International University (Kazakhstan)

EDITORS:

- **Ainur Seitkan**, Candidate of Technical Sciences, PhD, Dean of the Higher School of Natural Sciences, Astana International University (Kazakhstan);
- **Assel Mukanova,** PhD, Dean of the Higher School of Information Technology and Engineering, Astana International University (Kazakhstan);
- **Assel Abdildayeva,** PhD, Associate Professor, of the Department of Artificial Intelligence and Big Data, Al-Farabi Kazakh National University (Kazakhstan);
- **Jiri Chlachula,** PhD, Dr.Hab., Full Professor, Adam Mickiewicz University, Poznań (Poland);
- **Simon A.T. Redfern,** PhD, Professor, Nanyang Technological University (Singapore);
- **Xiaolei Feng,** PhD, Nanyang Technological University (Singapore);
- **Khan Shujaul Mulk,** PhD, Professor, Quaid-i-Azam University (Pakistan);
- **Natal'ya Bazarnova,** Doctor of Chemical Sciences, Professor, Institute of Chemistry and Chemical-Pharmaceutical Technologies (Russia);
- **Vera Cheryomushkina,** Doctor of Biological Sciences, Professor, Central Siberian Botanical Garden SB RAS (Russia);
- **Nurbolat Tasbolatuly,** PhD, Deputy Dean of the Higher School of Information Technology and Engineering, Astana International University (Kazakhstan);
- **Saken Baisholanov,** Candidate of Geographical Sciences, Associate Professor, Astana International University (Kazakhstan);
- **Serik Nurkenov,** PhD, Associate Professor, Astana International University (Kazakhstan).

Editorial address: 8, Kabanbay Batyr avenue, of.316, Nur-Sultan,
Kazakhstan, 010000
Tel.: (7172) 24-18-52 (ext. 316)
E-mail: natural-sciences@aiu.kz

International Sciense Reviews NST - 76153
International Science Reviews
Natural Sciences and Technologies series
Owner: Astana International University
Periodicity: quarterly
Circulation: 500 copies

CONTENT

1. A.A. Kussainova, O.Bulgakova Blood mtDNA Copy Number as a Potential Indicator of X-ray Radiation Exposure in Animals.....7

2. А.Маллер, Ж.Адамжанова Идентификации почвенных бактерий города Астана с последующей оценкой их антибиотикорезистентности.....14

3. А. М. Султанкулов К вопросу о формировании устойчивых университетов ……...32

4. А. М. Султанкулов, А. С. Сейткан, М.К. Карибаева Анализ экологической осведомленности студентов МУА …….....…..... 38

5. Ж.Сугурбаев Обоснование использования метода проб Питерсена в образовательной практике ……..….....……47

6. А.Тыныкулова, Б.Таңырыс Проектирование платформы для автоматизированной генерации и оценки тестовых заданий ……...…....53

7. Д.Байғожанова, Н.Тасболатұлы, Қ.Нартай Архитектура интеллектуального управления в управленческих веб-приложениях с использованием MLOps и машинного обучения …….…….……62

8. М.И.Есбота, Г.Ж.Таганова Выбрать правильный стек технологий для разработки приложений искусственного интеллекта …….....…71

9. Н. Ж.Жарасхан, С.А.Наурызбаева Постквантты криптографиялық алгоритмдерді қауіпсіздік пен стандарттауға қолдану және бағалау критерийлері.....81

10. А.Н.Сұлтанғазиева, Ж.Р.Абдуханова , Д.С.Молдаш Обзор методов и технологий для выявления ложных новостей на основе анализа текста и машинного обучения...90

11. Т.Ә Талдықбаева, Ж.Т.Абдуллаева Определение фейковых отзывов с помощью машинного обучения.....98

12. Т.Б.Бекбосынова Применение методов машинного обучения для автоматического распознавания опухолей головного мозга на мрт-снимках.....104

13. А.Нарынбай Разработка и исследование алгоритмов сегментации и распознавания объектов на медицинских изображениях на основе нейронных сетей.....111

14. Н.Г.Турсынбек Методы стратегического прогнозирования в финансовом менеджменте.....117

МРНТИ 28.23.15

Определение фейковых отзывов с помощью машинного обучения

Т.Ә. Талдықбаева, Ж.Т.Абдуллаева*

Высшая школа информационных технологий и инженерии, Международный университет Астана, Астана, Казахстан
*Автор-корреспондент

Аннотация. В данной работе исследуется задача автоматического выявления фейковых отзывов о товарах и услугах с использованием базовых методов машинного обучения. Развитие онлайн-торговли поставило задачу разработки эффективных инструментов для выявления ложных отзывов, которые искажают восприятие потребителей и наносят ущерб репутации бизнеса. В исследовании рассматриваются методы текстового анализа и классификации, такие как Naive Bayes и Random Forest, с использованием представления текста в формате TF-IDF. Приводится сравнительный анализ этих методов, а также обсуждается их эффективность с точки зрения точности, точности предсказания, полноты и F1-меры.
Ключевые слова: фейковые отзывы, машинное обучение, Naive Bayes, Random Forest, TF-IDF, обработка текста, анализ тональности, классификация текстов, искусственный интеллект.

Введение

Онлайн-отзывы играют важную роль при выборе товаров и услуг, поэтому наличие фейковых отзывов (намеренно ложных обзоров, вводящих потребителей в заблуждение) становится серьёзной проблемой [1]. Недобросовестные продавцы могут публиковать хвалебные отзывы о своих продуктах или негативные - о товарах конкурентов, чтобы искусственно улучшить репутацию или нанести ущерб бизнесу оппонента [1]. По оценкам отраслевых исследований, доля поддельных отзывов может достигать 10% на платформах вроде Google и ~5-7% на сайтах типа Yelp или TripAdvisor [6]. Это подрывает доверие пользователей и может привести к ошибочным решениям о покупке. В связи с масштабом проблемы необходимы эффективные методы автоматического обнаружения фейкового контента [4].

Однако выявление фейковых отзывов является нетривиальной задачей. Авторы обманных текстов стараются сделать их максимально «реалистичными», из-за чего такие отзывы зачастую неотличимы от настоящих при ручной модерации. Более того, возможности человека в распознавании обмана ограничены: как показывают исследования, даже эксперты выявляют фейковые отзывы лишь на уровне случайного угадывания (~50% точности) [4]. С другой стороны, алгоритмы машинного обучения способны обнаруживать скрытые закономерности и лингвистические маркеры обмана, достигая высокой точности - вплоть до ~90% на тестовых наборах данных [4].

Ранние работы в этой области рассматривали «opinion spam» - злоупотребления пользовательскими отзывами - и показали, что фейковые отзывы заметно отличаются от спама в email или веб-комментариях [1]. Например, в работе Джиндал и Лю было выявлено широкое распространение поддельных отзывов на товарных платформах и предложены первые подходы к их обнаружению с использованием текстовых и поведенческих признаков авторов. В дальнейшем исследователи сосредоточились на алгоритмах классификации текстов: так, в работе Ott и соавт. [2] был создан первый золотой стандарт корпуса обманных отзывов (800 отзывов о отелях, сбалансированных по классам), на котором методы машинного обучения (n-граммные модели + классификаторы) достигли ~89% точности, значительно превосходя

человека [4]. В последние годы для задачи выявления фейковых отзывов применяются как классические алгоритмы (Naive Bayes, SVM, деревья решений и др.), так и более сложные методы глубокого обучения, включая нейросети [4]. Тем не менее, несмотря на активное развитие глубоких нейросетевых подходов, более простые модели остаются ценными благодаря своей интерпретируемости и невысоким требованиям к вычислительным ресурсам.

Настоящая работа посвящена построению базовой модели классификации отзывов на честные и фейковые с использованием классических методов машинного обучения - Naive Bayes и Random Forest - на признаках TF-IDF. В статье представлены описания использованных данных, методы обработки текста и обучения моделей, а также экспериментальные результаты и обсуждение, на основе которых сделаны выводы о перспективах развития системы.

Материалы и методы

Для обучения и тестирования моделей в данной работе был использован открытый датасет Fake Reviews Dataset, доступный на платформе Kaggle [7]. Этот набор данных содержит 20 000 фейковых и 20 000 реальных отзывов о товарах, размеченных как фейковые (CG) и настоящие (OR). Отзывы представлены на английском языке и сбалансированы по классам, что упрощает обучение моделей, исключая смещение в сторону большинства класса. Важные столбцы в наборе данных включают: category (категория товара), rating (рейтинг отзыва), label (метка отзыва - CG (фейковый) или OR (реальный)), и text_ (текст отзыва).

Для предобработки текстовых данных была выполнена стандартная очистка и нормализация. На этом этапе из текстов удалялись неинформативные символы, такие как HTML-разметка, знаки пунктуации и цифры. Текст приводился к нижнему регистру, устранялись лишние пробелы и спецсимволы. Для упрощения модели не рассматривались эмодзи и редкие Unicode-символы. Важным шагом было оставление стоп-слов (например, предлогов и союзов), поскольку некоторые из них могут нести значимую информацию (например, отрицание "не" или местоимения "я", "мы"), что полезно для анализа стиля текста. В отличие от некоторых подходов, стемминг и лемматизация не использовались, так как в проведенных экспериментах это не дало значительного прироста качества модели. Вместо этого использовались необработанные формы слов для улучшения точности классификации.

Очищенные тексты были преобразованы в численные признаки с использованием метода TF-IDF (Term Frequency-Inverse Document Frequency). Этот метод помогает определить важность каждого слова в тексте относительно всего корпуса текстов. Метод оценивает важность слова, учитывая его частоту в документе и редкость в корпусе. Сначала строится словарь всех слов, встречающихся в отзывах, после очистки. Затем каждый отзыв представляется как вектор, где каждый компонент пропорционален частоте слова в тексте и обратно пропорционален количеству отзывов, в которых это слово встречается. В работе использовались уникраммы (unigrams), то есть отдельные слова, которые считаются признаками. Общий размер словаря для данного датасета составляет порядка 5-6 тысяч слов после очистки. Для формирования признаков TF-IDF и обучения моделей использовалась библиотека scikit-learn на языке Python.

Для разделения данных на обучающую и тестовую выборки использовалось соотношение 80/20. Обучение проводилось на 80% случайно выбранных отзывов, а оставшиеся 20% использовались для оценки качества моделей (имитируя ситуацию новых, ранее не виденных моделью отзывов). Для повышения надежности оценки применялась k-блочная перекрёстная проверка (k-fold cross-validation) с параметром k=5. Данные случайным образом разбивались на 5 равных частей, каждая из которых по очереди использовалась как тестовая, а остальные - как обучающие. Результаты усреднялись по всем складам.

Naive Bayes (Наивный Байес) - вероятностный классификатор, основанный на теореме Байеса и предположении независимости признаков. Этот алгоритм часто демонстрирует высокую эффективность в задачах классификации текста благодаря способности учитывать частотность слов. В данной работе использовалась модель Multinomial Naive Bayes из библиотеки scikit-learn. Особых гиперпараметров этот алгоритм не требует, кроме сглаживания Лапласа с параметром α=1 (по умолчанию), что предотвращает нулевые вероятности для слов, отсутствующих в обучении.

Random Forest (Случайный лес) - ансамблевый метод классификации, использующий несколько решающих деревьев для улучшения точности и уменьшения переобучения. Каждый решающий дерево строится на случайной подвыборке обучающих данных и случайном подмножестве признаков. В результате деревья голосуют за класс большинства. В данном исследовании использовался классификатор Random Forest из библиотеки scikit-learn с количеством деревьев n_estimators=100. Глубина деревьев не ограничивалась явно, модель строила деревья до чистого разбиения листьев или до минимального количества объектов в листе (1). Остальные параметры принимались по умолчанию.

Для количественной оценки качества классификации использовались следующие метрики: accuracy - доля правильно классифицированных отзывов, precision - точность, доля реально фейковых среди всех, предсказанных моделью как фейковые, recall - полнота, доля выявленных моделью фейковых отзывов среди всех реально фейковых, и F1-score - гармоническое среднее precision и recall. Метрики были вычислены для класса фейковых отзывов и усреднены (macro-avg) для получения интегральной оценки моделей.

Результаты

Модели были обучены и протестированы с использованием метрик точности (accuracy), точности предсказания (precision), полноты (recall) и F1-меры.

Таблица 1 – Средние значения метрик качества моделей Naive Bayes и Random Forest

Модель	Accuracy	Precision (Класс 0)	Precision (Класс 1)	Recall (Класс 0)	Recall (Класс 1)	F1-Score (Класс 0)	F1-Score (Класс 1)
Naive Bayes	0.8432	0.86	0.82	0.82	0.87	0.84	0.85
Random Forest	0.8395	0.86	0.82	0.82	0.86	0.84	0.84

Обе модели показали схожие результаты по всем метрикам. Naive Bayes немного опередил Random Forest по полноте (Recall = 0.87 против 0.86) и F1-score (0.85 против 0.84), в то время как точность и precision были одинаковыми. Это иллюстрируется на (рис 1.)

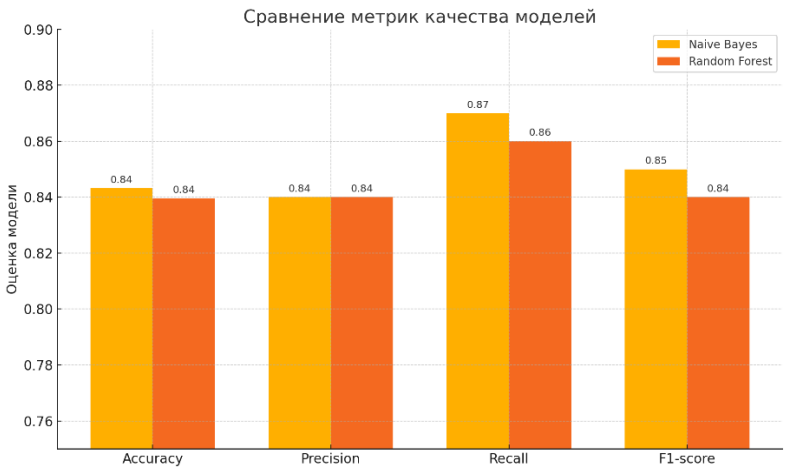

Рисунок 1 – Сравнение метрик качества моделей Naive Bayes и Random Forest

Обсуждение

В ходе эксперимента обе модели, Naive Bayes и Random Forest, продемонстрировали схожие результаты с точностью классификации около 84%. Это подтверждает, что оба алгоритма являются эффективными для задачи классификации фейковых отзывов, несмотря на различие в принципах работы.

Naive Bayes, как вероятностный классификатор, оказался успешным в выявлении фейковых отзывов за счет учета частотности слов, что позволило ему показать высокие значения для precision и recall, особенно для класса "фейк". Этот алгоритм также продемонстрировал высокую скорость обучения и предсказания, что делает его полезным в задачах, где важна быстрота обработки данных. Однако его зависимость от предположения о независимости признаков может ограничивать его эффективность в более сложных задачах с взаимозависимыми признаками.

Random Forest, использующий ансамбль решающих деревьев, показал схожие результаты по точности, но отличается тем, что может учитывать более сложные взаимозависимости между признаками. Такой подход позволил модели немного лучше выявлять фейковые отзывы, что проявилось в слегка улучшенных показателях recall для класса "фейк". Однако Random Forest требует больше вычислительных ресурсов для обучения, что может быть важным фактором при работе с большими объемами данных.

Обе модели показали одинаковые значения для accuracy, macro avg и weighted avg, что указывает на их сходную способность к обобщению. Это подтверждает, что оба алгоритма обеспечивают хороший баланс между точностью и полнотой классификации, что является ключевым для решения задачи обнаружения фейковых отзывов.

Заключение

Результаты эксперимента подтверждают, что классические методы машинного обучения, такие как Naive Bayes и Random Forest, могут эффективно решать задачу классификации фейковых отзывов, достигая точности порядка 84%. Оба алгоритма продемонстрировали хорошие показатели по меткам precision, recall и F1-score, что свидетельствует о стабильности их работы и способности правильно классифицировать как фейковые, так и реальные отзывы.

Литература
[1] Jindal N., Liu B. Opinion spam and analysis // *Proc. of Intl. Conf. on Web Search and Data Mining (WSDM)*. – 2008. – P. 219–230.
[2] Ott M., Choi Y., Cardie C., Hancock J. T. Finding deceptive opinion spam by any stretch of the imagination // *Proc. of ACL: Human Language Tech.* – 2011. – P. 309–319.
[3] Гаряева Э.О. Обнаружение заказных отзывов с использованием методов машинного обучения: ВКР. – НИУ ВШЭ, 2023. – 64 с.
[4] Логвиненко Ю.А., Вовк Л.П. Модели и методы машинного обучения для распознавания фейкового контента // *Вестник науки*. – 2024. – URL: https://cyberleninka.ru/article/n/modeli-i-metody-mashinnogo-obucheniya-dlya-raspoznavaniya-feykovogo-kontenta
[5] Abd M. J., Hussein M. H. Fake reviews detection in e-commerce using machine learning techniques: a comparative survey // *BIO Web of Conferences*. – 2024. – Vol. 97. – Art. 00099.
[6] Invesp. The State of Fake Online Reviews – Statistics and Trends [Электронный ресурс]. – 2025. – Режим доступа: *Invesp Blog*, URL: https://www.invespcro.com/blog/fake-reviews-statistics/
[7] Fake Reviews Dataset, URL: https://www.kaggle.com/datasets/mexwell/fake-reviews-dataset/data

References
[1] Jindal N., Liu B. Opinion spam and analysis // Proc. of Intl. Conf. on Web Search and Data Mining (WSDM). – 2008. – P. 219–230.
[2] Ott M., Choi Y., Cardie C., Hancock J. T. Finding deceptive opinion spam by any stretch of the imagination // Proc. of ACL: Human Language Tech. – 2011. – P. 309–319.
[3] Garyaeva E.O. Obnaruzhenie zakaznyh otzyvov s ispol'zovaniem metodov mashinnogo obucheniya: VKR. – NIU VSHE, 2023. – 64 s.
[4] Logvinenko YU.A., Vovk L.P. Modeli i metody mashinnogo obucheniya dlya raspoznavaniya fejkovogo kontenta // Vestnik nauki. – 2024. – URL: https://cyberleninka.ru/article/n/modeli-i-metody-mashinnogo-obucheniya-dlya-raspoznavaniya-feykovogo-kontenta
[5] Abd M. J., Hussein M. H. Fake reviews detection in e-commerce using machine learning techniques: a comparative survey // BIO Web of Conferences. – 2024. – Vol. 97. – Art. 00099.
[6] Invesp. The State of Fake Online Reviews – Statistics and Trends [Elektronnyj resurs]. – 2025. – Rezhim dostupa: Invesp Blog, URL: https://www.invespcro.com/blog/fake-reviews-statistics/
[7] Fake Reviews Dataset, URL: https://www.kaggle.com/datasets/mexwell/fake-reviews-dataset/data

Жалған пікірлерді анықтау үшін машиналық оқыту әдістері

Т.Ә. Талдықбаева, Ж.Т.Абдуллаева *

Ақпараттық технологиялар және инженерия жоғары мектебі, Астана халықаралық университеті, Астана, Қазақстан

Аңдатпа. Бұл жұмыс машиналық оқыту әдістерін пайдалана отырып, тауарлар мен қызметтер туралы жалған пікірлерді автоматты түрде анықтау мәселесі қарастырылады. Онлайн сауданың дамуы, тұтынушылардың пікірлерін бұрмалаған және компаниялардың беделіне нұқсан келтіретін жалған пікірлерді анықтаудың сенімді құралдарын дамыту қажеттілігін туғызды. Зерттеу мәтіндерді талдау және классификациялау әдістерін қарастырады, соның ішінде Naive Bayes және

Random Forest алгоритмдері пайдаланылады, және TF-IDF мәтіндерін көрсету әдісі қолданылады. Мақалада осы әдістердің тиімділігін салыстыру, оларды дәлдік, дәлдік көрсеткіші, қайталану және F1-мера сияқты метрикалар бойынша бағалау ұсынылады.
Түйін сөздер: жалған пікірлер, машиналық оқыту, Naive Bayes, Random Forest, TF-IDF, мәтінді өңдеу, сезімталдықты талдау, мәтіндерді классификациялау, жасанды интеллект.

Machine Learning Methods for Fake Review Detection

T.A Taldykbayeva , Zh.T. Abdullayeva*

Higher School of Information Technology and Engineering, Astana International University, 010000, Astana, Kazakhstan

Abstract. This paper investigates the problem of automatically detecting fake reviews of products and services using machine learning methods. The rise of online commerce has made it necessary to develop reliable tools for identifying deceptive reviews that distort consumer perceptions and damage businesses' reputations. The study explores text analysis and basic classification methods using algorithms such as Naive Bayes and Random Forest, employing TF-IDF text representation. The paper presents a comparative analysis of these methods and discusses their effectiveness in terms of accuracy, precision, recall, and F1-score.
Keywords: fake reviews, machine learning, Naive Bayes, Random Forest, TF-IDF, text processing, sentiment analysis, text classification, artificial intelligence.