

ДОСЛІДЖЕННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ВИЗНАЧЕННЯ ТОНАЛЬНОСТІ ТЕКСТОВИХ ДАНИХ

Гученко В.В.

e-mail: vsevolod.huchenko@nure.ua

Науковий керівник – д.т.н, проф. Хаханова Г.В.

Харківський національний університет радіоелектроніки, каф. АПОТ
м. Харків, Україна

Machine learning methods for text sentiment detection are studied in this work. Four classification algorithms were implemented and compared: Naive Bayes, Support Vector Machine (SVM), Logistic Regression, and a neural network. The models were trained and tested on a labeled dataset of 1000 text samples. The neural network achieved the highest accuracy of 89.2% and F1-score of 88.9%. The results of this study can be used for automated analysis of customer feedback, social media monitoring, and other applications requiring sentiment detection.

Вступ. Стрімке зростання обсягів текстової інформації в цифровому середовищі зумовлює необхідність створення ефективних інструментів для її автоматизованої обробки. Одним із актуальних завдань у галузі обробки природної мови є визначення тональності текстів – класифікація повідомлень за емоційним забарвленням (позитивне, негативне або нейтральне) [1].

Такі системи широко застосовуються для аналізу відгуків клієнтів, моніторингу соціальних мереж, оцінки репутації брендів та автоматизації зворотного зв'язку.

Метою даного дослідження є порівняльний аналіз ефективності різних методів машинного навчання для визначення тональності текстових даних та вибір оптимального підходу для практичного застосування.

Зміст дослідження. Для досягнення поставленої мети було виконано такі завдання: проаналізовано сучасні методи аналізу тональності текстів; створено розмічений набір даних із 1000 текстових прикладів українською мовою; реалізовано чотири класифікатори: наївний байєсів класифікатор (Naive Bayes), метод опорних векторів (SVM), логістичну регресію (Logistic Regression) та нейронну мережу (MLPClassifier); проведено тестування та порівняльний аналіз точності роботи алгоритмів [2].

Програмну реалізацію виконано мовою Python з використанням бібліотек scikit-learn, NLTK та pandas. Для оцінки якості класифікації застосовувались метрики точності (accuracy), повноти (recall), F1-міри та матриця невідповідностей (confusion matrix). Навчання моделей проводилось на 80% даних, тестування – на 20%.

Результати експериментального дослідження наведено в таблиці 1.

Таблиця 1 – Порівняльний аналіз ефективності методів класифікації

Метод	Точність (Accuracy)	Точність (Precision)	Повнота (Recall)	F1-міра
Naive Bayes	82.1%	82.3%	81.5%	81.9%
SVM	87.4%	87.6%	86.9%	87.2%
Логістична регресія	85.3%	85.4%	84.8%	85.1%
Нейронна мережа	89.0%	89.2%	88.7%	88.9%

Як видно з таблиці, найкращі результати показала нейронна мережа (F1-міра 88.9%), дещо нижчі – SVM (87.2%). Найпростіший класифікатор Naive Bayes очікувано продемонстрував найнижчу точність (81.9%).

Висновки. У результаті проведеного дослідження розроблено та протестовано чотири методи машинного навчання для визначення тональності текстових даних, встановлено, що найкращу ефективність забезпечує використання нейронної мережі з F1-мірою 88.9%, визначено, що метод опорних векторів (SVM) також показує високі результати (F1-міра 87.2%) і може бути рекомендований для задач, де важлива швидкість навчання, підтверджено практичну придатність розроблених методів для автоматизованого аналізу відгуків клієнтів, моніторингу соціальних мереж та обробки звернень до служб підтримки, а отримані результати можуть бути використані при створенні систем аналізу тональності текстів українською мовою.

Список використаних джерел:

1. Bird S., Klein E., Loper E. Natural Language Processing with Python. Walnut Creek : O'Reilly Media, 2021. 512 p.
2. Pedregosa F. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011. Vol. 12. P. 2825–2830.
3. Hahanov V. I., Chumachenko S. V., Litvinova E. I., Hahanova H. V., Obrizan V. I., Maksymova N. H. Vector logic structures and functions of computing. *Applied Aspects of Information Technology*. 2025. Vol. 8, No. 4. P. 442–452. DOI:<https://doi.org/10.15276/aait.08.2025.28>.