

УДК 004.4

Трохимчук В.Є.¹, Тіменко А.В.², Куликовська Н.А.²

¹ студ. гр. КНТ-522 НУ «Запорізька Політехніка»

² старш. викл. НУ «Запорізька Політехніка»

АНАЛІЗ ТОНАЛЬНОСТІ, КЛЮЧОВИХ СЛІВ І АГРЕСИВНИХ ПОВІДОМЛЕНЬ У ТЕКСТОВИХ ДАНИХ ЗА ДОПОМОГОЮ PYTHON ТА NLP

Дослідження спрямоване на вивчення сучасних методів аналізу тональності текстових даних з використанням бібліотек Python для обробки природної мови. Зростаючі обсяги неструктурованих текстових даних в Інтернеті та потреба в ефективних методах аналізу тональності для розуміння громадської думки, реакцій споживачів, політичних настроїв тощо визначають актуальність теми.

Аналіз останніх досліджень показує активний розвиток методів машинного навчання та обробки природної мови для аналізу тональності текстів, зокрема в соціальних мережах і відгуках споживачів. Широко вивчається ефективність різних підходів, включаючи NLTK, TextBlob, spaCy та ін. Увага приділяється також розробці гібридних методів.

Для аналізу перелічених бібліотек були взяті кількісні показники їх використання на GitHub (рис. 1):

- а) відсоток використання бібліотек для обробки тексту та виконання основних NLP завдань;
- б) відсоток використання бібліотек для розробки пошукової системи;
- в) відсоток використання бібліотек для аналізу сентименту;
- г) відсоток використання бібліотек для векторного представлення слів та тематичне моделювання;
- д) відсоток використання бібліотек для створення продуктів, які вимагають обробки тексту в реальному часі;
- е) відсоток використання для векторного представлення слів та тематичного моделювання у текстових даних.

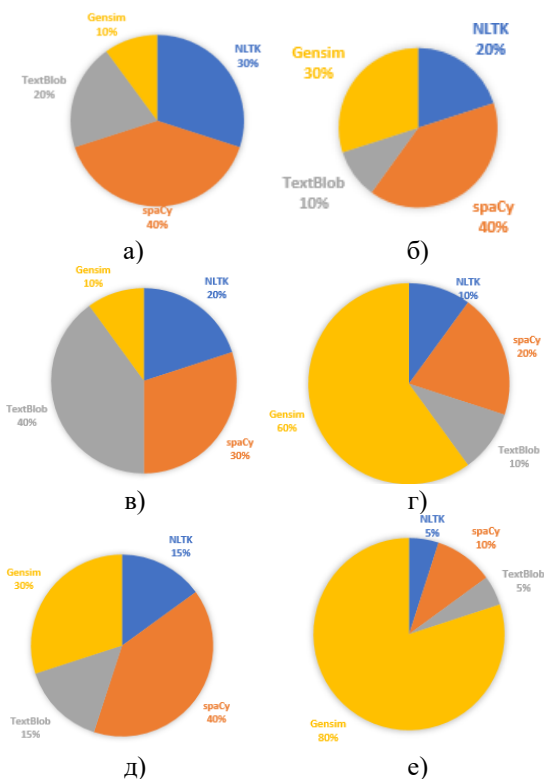


Рисунок 1 – Порівняння методів обробки природної мови як NLTK, spaCy, TextBlob і Gensim

В роботі проведено порівняльний аналіз можливостей бібліотек NLTK і TextBlob для класифікації тональності україномовних текстів. Експериментально оцінено їх точність, швидкодію, зручність та гнучкість..

Таблиця 1 – Порівняння методів за точністю, зручністю, гнучкістю та налаштованістю

Метод	Точність виявлення	Зручність у використанні	Гнучкість та налаштованість
NLTK	80%	70%	90%
TextBlob	70%	90%	60%

Вибір оптимального методу аналізу тональності залежить від конкретних вимог. Для складних досліджень рекомендується NLTK, для прикладних задач з високою продуктивністю - spaCy, для простих завдань - TextBlob. Перспективним є розробка спеціалізованих алгоритмів для української мови та інтеграція аналізу тональності в системи підтримки прийняття рішень.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Ivokhin E., Про один спосіб аналізу тональності текстів за допомогою штучних нейронних мереж / E.Ivokhin, M.Makhno, V.Rets // Системи управління, навігації та зв'язку. Збірник наукових праць. 2022. – Т. 3, № 69. – С. 71–74. URL: <https://doi.org/10.26906/sunz.2022.3.071>
2. Orel A. Social media analyzing for evaluation opinions determination based on sentiment analysis. / A. Orel // International scientific journal "Internauka". 2018. – No. 10. URL: <https://doi.org/10.25313/2520-2057-2018-10-3858>.
3. Deng Y. Research on sentiment analysis methods for text-oriented data. / Y. Deng // Frontiers in computing and intelligent systems. 2023. – Vol. 3, no. 1. – P. 42–47. URL: <https://doi.org/10.54097/fcis.v3i1.6022>.
4. Mukasheva A. Tasks and methods of text sentiment analysis. / A.Mukasheva // Scientific journal of astana IT university. 2021. – No. 7. – P. 55–62. URL: <https://doi.org/10.37943/aitu.2021.57.68.005>.
5. Abonizio H. Q., Toward text data augmentation for sentiment analysis. / H.Q.Abonizio, P.E.Caraiso, S.Barbon Junior // IEEE transactions on artificial intelligence. 2021. – P. 1. URL: <https://doi.org/10.1109/tai.2021.3114390>.