

УДК 621.317

ВПРОВАДЖЕННЯ АВТОМАТИЗАЦІЇ АНАЛІЗУ ВІДГУКІВ КОРИСТУВАЧІВ ШЛЯХОМ ЗАСТОСУВАННЯ ОБРОБКИ ПРИРОДНОЇ МОВИ

Скорін Юрій, Петренко Богдан

Харківський національний економічний університет ім. Семена Кузнеця

У цифрову епоху взаємодія користувачів з продуктами та послугами призвела до величезного числа відгуків та коментарів. Компанії, як малі стартапи, так і великі корпорації, щодня отримують сотні, навіть тисячі текстових відгуків через різні канали — електронну пошту, соціальні мережі, спеціалізовані платформи та внутрішні системи зворотного зв'язку. У цих неструктурованих даних приховано найважливішу інформацію про задоволеність клієнтів, їхні потреби та очікування. За даними досліджень, більше 90% керівників компаній вважають аналіз відгуків користувачів важливим для прийняття стратегічних рішень, але лише приблизно 30% організацій мають системи, які дозволяють ефективно обробляти та аналізувати ці дані. Традиційний ручний аналіз відгуків давно втратив актуальність для сучасного бізнесу. Перегляд тисяч коментарів потребує багато часу та людини, тому цей процес відкрити дуже дорогим. Крім цього, особистість людини часто призводить до того, що різні особи можуть розуміти однакові думки по-різному. Даже найкращі аналітики не можуть швидко обробляти такі великі обсяги даних, щоб компанії могли швидко відповісти на зміни в настроях покупців. Обробка природної мови (NLP) допомагає у цьому, використовуючи алгоритми, які автоматично аналізують зміст тексту, визначають його ставлення та вирізують основні теми з великих наборів даних. З розвитком технологій машинного навчання та глибоких нейронних мереж сучасні системи NLP стають все точнішими в розумінні контексту та тонкощів мовлення. Нещодавні досягнення в NLP, включаючи розробку моделей-трансформерів, таких як BERT, GPT та RoBERTa, значно

розширили можливості обробки природної мови. Ці моделі здатні розуміти контекст слів у реченні, враховувати полісемію та ідіоматичні вирази, а також успішно обробляти тексти різними мовами. Їх застосування до аналізу відгуків користувачів відкриває нові горизонти для розуміння потреб клієнтів та покращення продуктів і послуг [1, 4]. Впровадження автоматизованого аналізу відгуків користувачів за допомогою NLP не тільки значно скорочує час та ресурси, але й забезпечує глибші аналітичні результати. NLP-системи здатні виявляти приховані тенденції та закономірності, недоступні під час ручного аналізу, а також забезпечують масштабованість рішення відповідно до зростання обсягів даних.

Актуальність теми дослідження полягає в тому, що бізнес постійно потребує ефективних інструментів для аналізу великих обсягів текстової інформації. Такі інструменти мають допомогти витягувати корисні висновки із відгуків клієнтів та перетворювати їх на реальні поради для покращення товарів та послуг. Особливу важливість ця тема має для електронної комерції, сфери послуг та розробки програмного забезпечення, оскільки якість та швидкість реакції на відгуки впливають на здатність компаній змагатися на ринку.

Основні завдання включають вивчення сучасних підходів до обробки природної мови, аналіз алгоритмів класифікації тексту та аналізу настроїв, а також розробку та оцінку ефективності системи автоматизованого аналізу відгуків користувачів. У статті розглядаються теоретичні аспекти обробки природної мови, досліджуються різні моделі та алгоритми аналізу відповідей, експериментально оцінюється їх ефективність на реальних даних.

Методологічною основою дослідження є методи машинного навчання, глибокі нейронні мережі, статистичні методи аналізу тексту та методи оцінки якості моделей машинного навчання. В експериментальній частині роботи використовується мова програмування Python та бібліотеки обробки природної мови, такі як NLTK, spaCy, Transformers та TensorFlow. Результати дослідження можуть бути використані для розробки програмних рішень, які

дозволять компаніям краще розуміти потреби своїх клієнтів, швидко реагувати на проблеми та покращувати якість своїх продуктів і послуг.

Практичне значення роботи полягає в можливості впровадження розроблених методів і моделей у бізнес-процеси організацій різного масштабу для покращення взаємодії з клієнтами та підвищення їхньої лояльності. У сучасному цифровому світі аналіз відгуків користувачів став критичним елементом для будь-якого бізнесу, який прагне довести якість своїх продуктів і послуг. Однак масштабність і неструктурований характер цих даних створюють значні труднощі для традиційного аналізу. Основною проблемою, що розглядається в цьому дослідженні, є неефективність та обмеження ручного аналізу великих обсягів текстових відгуків користувачів. Ця проблема має кілька важливих аспектів: експоненціальне зростання обсягів даних; суб'єктивність та непослідовний аналіз; витрати критичного часу; обмежена здатність виявляти приховані закономірності; зростання складності багатомовного аналізу; відсутність стандартизації та суб'єктивність оцінок; зростання вартості аналітики відгуків.

Метою дослідження є розроблення та оцінювання ефективності комплексної системи автоматизованого аналізу відгуків користувачів на основі сучасних методів обробки природної мови, яка забезпечить високу точність визначення тону, ключових аспектів та тем, можливість роботи зі відгуками різними мовами та ефективну інтеграцію в бізнес-процеси організацій різного масштабу.

Для досягнення мети необхідно вирішити такі конкретні *завдання*: розроблення ефективних методів попередньої обробки відгуків користувачів; дослідження та порівняльний аналіз сучасних моделей класифікації тону відповідей; розроблення методу аспектно-орієнтованого аналізу тональності для різних областей; створення ефективних методів визначення тем та ключових аспектів у корпусах відповідей; розробка підходу до багатомовного аналізу відгуків; створення методів інтерпретації та візуалізації результатів аналізу зворотного зв'язку; розроблення та

оцінювання комплексної автоматизованої системи аналізу зворотного зв'язку; експериментальна оцінка ефективності розроблених методів на реальних даних. Останні два пункти саме й визначають *новизну* проведених досліджень.

Для вирішення поставлених завдань буде використано комплексний методологічний підхід, що поєднує машинне навчання, глибоке навчання, обробку природної мови та аналіз даних. Дослідження буде проведено в кілька етапів: аналітичний етап; етап проектування та розробки; експериментальний етап; аналітично-узагальнювальний етап.

У ході дослідження здійснено повний аналіз проблем та сучасних методів автоматизації аналізу відгуків користувачів. Підготовлено теоретичні та методичні засади використання NLP для аналізу текстів, зокрема комплексний підхід до попередньої обробки даних з урахуванням особливостей користувацьких текстів — емотиконів, сленгу та помилок. Вірна попередня обробка тексту дозволяє збільшити якість аналізу на 8–15%. Створено набір моделей різного рівня складності для різних випадків використання. Надано повну методику оцінки ефективності NLP-моделей, яка враховує не лише показники точності (F1-score, precision, recall), але й обчислювальну ефективність (час навчання, швидкість аналізу, використання пам'яті), статистичну достовірність результатів та стійкість до різних типів даних. Методичне забезпечення дослідження включає сучасний технологічний стек на базі Python 3.10+ з використанням спеціалізованих бібліотек NLTK, spaCy, Hugging Face Transformers, PyTorch, TensorFlow. Систематичний підхід до організації експериментів через Weights & Biases, Jupyter Notebooks та Docker забезпечує відтворюваність результатів та ефективну ітерацію при розробці моделей. Практична цінність результатів дослідження полягає у можливості їх безпосереднього застосування для створення систем автоматизованого аналізу відгуків у різних галузях: електронній комерції, сервісних компаніях, розробці програмного забезпечення, маркетингу та аналітиці. Впровадження розроблених рішень

дозволяє скоротити час аналізу в 120-150 разів порівняно з ручною обробкою, забезпечити об'єктивність результатів та виявляти на 45% більше потенційно критичних проблем. Рекомендації, які було розроблено, допомагають компаніям вибирати найкращі моделі під конкретні ситуації бізнесу. Це дозволяє підприємствам різних розмірів створювати ефективні системи аналізу відгуків, враховуючи доступні ресурси та вимоги до точності. На наступному етапі планують перевірити розроблені методи на реальних даних та створити приклади, як застосовувати отримані результати у роботі організацій.

Література:

1. Bird S., Klein E., Loper E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, 2019. 504 p.
2. Goldberg Y. Neural Network Methods for Natural Language Processing. Synthesis Lectures on Human Language Technologies. 2017. Vol. 10, №. 1. P. 1–309.
3. Jurafsky D., Martin J. H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 3rd ed. draft. Pearson, 2023. URL: <https://web.stanford.edu/~jurafsky/slp3/>
4. Acheampong F. A., Wenyu C., Nunoo-Mensah H. Text-based emotion detection: Advances, challenges, and opportunities. Engineering Reports. 2020. Vol. 2, No. 7. e12189.
5. Chen Z., Qian T. Transfer Capsule Network for Aspect Level Sentiment Classification. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019. P. 547–556.