

Юрій Мочернюк, Ольга Мокрицька*

¹ НЛТУ України, Львів, Україна. ORCID: <https://orcid.org/0009-0002-1370-363X>, mocherniuk.iurii@nltu.lviv.ua

² НУЛП, Львів, Україна. ORCID: <https://orcid.org/0000-0002-2887-9585>, olha.v.mokrytska@lpnu.ua

Використання інтелектуального аналізу даних для оптимізації комунікаційних процесів у месенджерах

Анотація. Роль месенджерів у сучасному цифровому середовищі є важливим інструментом для особистої, бізнес-комунікації та навчання. Зокрема, увага акцентується на впливі користувацького досвіду (UX) на ефективність взаємодії з платформами, а також на необхідності дослідження та оптимізації UX за допомогою інтелектуального аналізу даних. Методи інтелектуального аналізу даних, такі як аналіз настроїв та кластерний аналіз в контексті дослідження та оптимізації користувацьких взаємодій в комунікаційних додатках дають змогу оптимізувати ці аспекти та підвищити продуктивність користувачів.

Ключові слова – аналіз, машинне навчання, месенджер, комунікація, аналіз настроїв, логістична регресія, кластерний аналіз, TF-IDF, UX.

У сучасному світі, де цифрові технології стали невід'ємною частиною повсякденного життя, месенджери відіграють важливу роль у спілкуванні між людьми. Вони є не лише засобом для особистих розмов, але й важливим інструментом для бізнес-комунікації, навчальних процесів та соціальної взаємодії. Відтак, якість користувацького досвіду (UX) у месенджерах безпосередньо впливає на ефективність користувачів та їхню взаємодію з платформою.

Саме тому є простір для досліджень зосереджених на розробці та застосуванні методів інтелектуального аналізу даних для виявлення та оптимізації аспектів UX у месенджерах. Покращення користувацького досвіду може значно підвищити ефективність комунікацій, продуктивність користувачів, а також сприяти зростанню користувацької бази та загальному успіху платформи.

Ключовими методами в арсеналі інтелектуального аналізу даних, які розглядаються: аналіз настроїв [3], машинне навчання [4], кластерний аналіз [7] та аналіз послідовностей [6]. Існує і ряд інших на зразок аналізу A/B тестування,

аналізу потоку кліків, регресійного аналізу, когортного аналізу, аналізу часових рядів, семантичного аналізу та інших [1, 2].

Аналіз настроїв. Аналіз настроїв (Sentiment Analysis) є потужним методом для визначення емоційної забарвленості тексту, що може бути особливо корисним при вивченні відгуків, повідомлень та коментарів користувачів месенджерів [3]. В цьому методі спершу проводиться попередня обробка тексту, що містить токенизацію (розбиття тексту на слова або фрази), лемматизацію або стемінг (зведення слів до базової форми) та видалення стоп-слів (загальних слів, які не несуть значного сенсу для аналізу) [5]. Текст тут представляється у вигляді векторів за допомогою методів Bag-of-Words (BoW), TF-IDF (Term Frequency-Inverse Document Frequency) або Word Embeddings. Формула TF-IDF для слова t в документі d виглядає так:

$$TF-IDF(t, d) = TF(t, d) \times \log\left(\frac{N}{|\{d \in D : t \in d\}|}\right), \quad (1)$$

де $TF(t, d)$ – частота слова t в документі d , N – загальна кількість документів, $|\{d \in D : t \in d\}|$ – кількість документів, що містять слово t .

Для класифікації настроїв використовуються моделі машинного навчання, такі як логістична регресія, SVM або нейронні мережі, зокрема LSTM, GRU, BERT [6]. Наприклад, для логістичної регресії можна використати функцію:

$$P(y = 1 \vee x) = \sigma(w \cdot x + b), \quad (2)$$

де σ – сигмоїдна функція, w – ваговий вектор, x – вектор ознак, b – зміщення. Навчання моделі далі здійснюється на навчальній вибірці з подальшою оцінкою точності за допомогою метрик, таких як Precision, Recall та F1-score.

Процес дослідження містить збір текстових даних з повідомлень та відгуків користувачів месенджера. Потім проводиться попередня обробка текстів та створення векторних представлень. Навчання обраної моделі виконується на навчальній вибірці, з оптимізацією гіперпараметрів для досягнення найкращих результатів. Продуктивність моделі оцінюється на тестовій вибірці за допомогою метрик точності [4].

Отримані результати аналізу дають змогу визначити загальний настрій користувачів, виявити основні проблемні точки та позитивні аспекти. Це забез-

печує об'єктивний спосіб покращення UX месенджера, сприяючи підвищенню ефективності комунікацій, продуктивності користувачів та зростанню користувацької бази.

Кластерний аналіз. Аналіз настроїв можна ефективно доповнити кластерним аналізом (Cluster Analysis), який дає змогу групувати користувачів за подібними поведінковими ознаками та емоційними реакціями. Це допомагає краще зрозуміти різні сегменти користувачів і адаптувати UX для кожної групи [7].

Кластерний аналіз містить кілька етапів. Спочатку проводиться підготовка даних, яка містить попередню обробку тексту (токенізація, лемматизація, видалення стоп-слів) і представлення тексту у вигляді векторів за допомогою методів BoW, TF-IDF або Word Embeddings, аналогічно до аналізу настроїв [5].

Основний математичний інструмент для кластерного аналізу – метод k -середніх (k -means). Він полягає у розділенні n об'єктів на k кластерів, де кожен об'єкт належить кластеру з найближчим середнім значенням. Формально, цільова функція мінімізується так:

$$\min \sum_{i=1}^k \sum_{x \in C_i} x - \mu_i^2 \quad (3)$$

де C_i – кластер, μ_i – центроїд кластера C_i , а x – об'єкт.

Процес містить наступні етапи. Спочатку задається кількість кластерів k та ініціалізуються центроїди. Потім для кожного об'єкта визначається найближчий центроїд і об'єкт призначається відповідному кластеру. Після цього центроїди перераховуються як середнє значення всіх об'єктів у кластері. Процес повторюється до досягнення конвергенції [7].

Висновки. Застосування кластерного аналізу для оптимізації UX месенджерів дає змогу виявити групи користувачів з подібними патернами поведінки та настроями. Наприклад, одні групи можуть частіше висловлювати задоволення певними функціями, тоді як інші можуть скаржитися на конкретні аспекти інтерфейсу. Це дає змогу проводити цілеспрямовані інтервенції для кожного сегмента, оптимізуючи UX відповідно до потреб кожної групи користувачів.

В поєднанні з аналізом настроїв, кластерний аналіз забезпечує багатовимірний підхід до розуміння користувацького досвіду, допомагаючи виявляти не лише емоційні реакції, але й поведінкові патерни, що дає змогу приймати більш обґрунтовані рішення щодо покращення UX месенджера.

У підсумку ці методи як в ізоляції, так і, для підвищення загальної точності та ефективності, у поєднанні з іншими можуть використовуватись для реалізації досліджень різного рівня масштабування та деталізації для виведення цифрових комунікаційних процесів на новий рівень.

Список використаних літературних джерел

- [1] Jurafsky, D., & Martin, J. H. (2020). "Speech and Language Processing" (3rd Edition). Pearson.
- [2] Manning, C. D., Raghavan, P., & Schütze, H. (2008). "Introduction to Information Retrieval." Cambridge University Press.
- [3] Liu, B. (2012). "Sentiment Analysis and Opinion Mining." Morgan & Claypool Publishers.
- [4] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,... & Duchesnay, E. (2011). "Scikit-learn: Machine Learning in Python." Journal of Machine Learning Research, 12, 2825-2830.
- [5] Bird, S., Klein, E., & Loper, E. (2009). "Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit." O'Reilly Media.
- [6] Goldberg, Y. (2016). "A Primer on Neural Network Models for Natural Language Processing." Journal of Artificial Intelligence Research, 57, 345-420.
- [7] Aggarwal, C. C., & Zhai, C. (2012). "*Mining text data*." Springer Science & Business Media.

Using data mining to optimize communication processes in messengers

Yurii Mocherniuk, Olha Mokrytska

The role of messengers in the modern digital environment is an important tool for personal, business communication and learning. In particular, attention is focused on the impact of user experience (UX) on the effectiveness of interaction with plat-

forms, as well as the need to study and optimize UX using data mining. Data mining techniques such as sentiment analysis and cluster analysis in the context of researching and optimizing user interactions in communication applications will allow optimizing these aspects and increasing user productivity.