

ЗАЛУЦЬКА ОЛЬГА

Хмельницький національний університет

<https://orcid.org/0000-0003-1242-3548>e-mail: [zalutska.olha@gmail.com](mailto:zalutska.olha@gmail.com)

МОЛЧАНОВА МАРІНА

Хмельницький національний університет

<https://orcid.org/0000-0001-9810-936X>e-mail: [m.o.molchanova@gmail.com](mailto:m.o.molchanova@gmail.com)

МАЗУРЕЦЬ ОЛЕКСАНДР

Хмельницький національний університет

<https://orcid.org/0000-0002-8900-0650>e-mail: [exe.chong@gmail.com](mailto:exe.chong@gmail.com)

МЕЛЬНИК ОЛЕГ

Хмельницький національний університет

<https://orcid.org/0009-0001-6367-8922>e-mail: [melnik77oleg@gmail.com](mailto:melnik77oleg@gmail.com)

СКРИПНИК ТЕТЯНА

Хмельницький національний університет

<https://orcid.org/0000-0002-8531-5348>e-mail: [tskripnik1970@gmail.com](mailto:tskripnik1970@gmail.com)

## МЕТОД ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ЕМОЦІЙНОЇ ТОНАЛЬНОСТІ ТЕКСТОВОЇ ІНФОРМАЦІЇ ДЛЯ ВИЗНАЧЕННЯ ПОВЕДІНКОВИХ НАМІРІВ НЕЙРОМЕРЕЖЕВИМИ ЗАСОБАМИ

У роботі за результатом аналізу сучасного стану проблеми інтелектуального аналізу емоційної тональності текстової інформації визначено, що є актуальним застосування нейронних мереж для аналізу емоційної тональності текстів, оскільки це забезпечує вищу точність класифікації, ніж альтернативні підходи. Було запропоновано для інтелектуального аналізу емоційної тональності текстової інформації використати нейромережу архітектури BERT як одну із найбільш точних, в той час як для аналізу коротких документів запропоновано використовувати її модифікацію, RoBERTa.

ZALUTSKA OLHA, MOLCHANOVA MARYNA,  
MAZURETS OLEKSANDR, MELNYK OLEH, SKRYPNYK TETIANA  
Khmelnytskyi National University

### METHOD FOR INTELLECTUAL ANALYSIS OF TEXTUAL INFORMATION EMOTIONAL TONALITY FOR DETERMINE THE BEHAVIORAL INTENTIONS BY NEURAL NETWORKS MEANS

In this paper, based on analysis results of current state of problem of intellectual analysis of the information texts emotional tonality, was determined that the use of neural networks for the analysis of texts emotional tonality is relevant, as it provides higher classification accuracy than alternative approaches. It was proposed to use the neural network of BERT architecture as one of the most accurate for the intellectual analysis of emotional tonality of texts, while it was proposed to use its modification RoBERTa for analysis of short documents.

The paper proposes the method for intellectual analysis of textual information emotional tonality for determine the behavioral intentions of users of socially oriented services and electronic commerce tools by neural networks means. The method uses neural network of the RoBERTa architecture. Conducted studies of the effectiveness of the method established that for combined multilingual texts it was possible to obtain an accuracy of 0.92, while the loss function had value of 0.29. This method should be used to determine the emotional tonality of short texts up to 500 words long, presented in Ukrainian language. At the same time, texts may contain surzhik and foreign words. The obtained results confirm the possibility and effectiveness of using method for intellectual analysis of textual information emotional tonality by neural networks means for determine the behavioral intentions of users, in particular, in socially oriented services and electronic commerce tools.

Keywords: BERT, RoBERTa, emotional tonality analysis, behavioral intentions, sentiment classification, sentiment analysis, emotion detection, neural networks.

#### Аналіз предметної області

Емоційна тональність тексту вказує на емоційний характер або емоційний забарвлення текстового висловлення. Ця характеристика визначає, які емоції чи почуття виражені в тексті, і чи є вони позитивними, негативними або нейтральними. Емоційна тональність тексту важлива для розуміння того, як текст сприймається читачами або як він може впливати на їхні емоції та настрої.

На сучасному етапі аналіз емоційної тональності текстових повідомлень, що входить до задач обробки природної мови привертає значну увагу науковців. Це пов'язано із зростанням сфер можливого застосування, до яких зокрема належать:

– Аналіз соціальних медіа. У соціальних медіа та онлайн-форумах публікується величезна кількість текстової інформації, яка може містити вказівки на наміри та емоційний стан користувачів. Інтелектуальний аналіз тональності дозволяє розуміти, які теми актуальні та які дії користувачів можна очікувати.

– Бізнес та маркетинг. Визначення тональності текстової інформації, пов'язаної з продуктами,

послугами та брендами, допомагає компаніям розуміти відгуки та думки клієнтів. Це може впливати на прийняття бізнес-рішень, покращення продуктів та створення маркетингових стратегій.

– Фінанси та інвестиції. Аналіз тональності новин та повідомлень у фінансових та інвестиційних спільнотах може допомогти інвесторам та трейдерам передбачити зміни на ринку й приймати обґрунтовані рішення.

– Безпека та контроль. Інтелектуальний аналіз тональності може використовуватись у сфері безпеки та правопорядку для моніторингу соціальних мереж та пошуку індикаторів загроз чи намірів.

– Політика та громадська думка. Оцінка тональності політичних дебатів та громадської думки може допомогти передбачити вибори, визначити настрої у суспільстві та розробити політичні стратегії.

– Охорона здоров'я. Аналіз тональності текстових відгуків про медичні послуги та ліки може допомогти покращити якість медичного обслуговування та управління охороною здоров'я.

– Освіта та психологія. В освітніх та психологічних дослідженнях аналіз тональності текстів може використовуватися для вивчення поведінки та психологічних характеристик людей.

– Клієнтський сервіс та підтримка. Аналіз тональності текстових звернень клієнтів може допомогти організаціям надавати більш ефективне обслуговування та реагувати на проблеми та запити клієнтів.

– Правоохоронні органи. Будь-які натяки на злочинні наміри чи погрози, виражені у текстовій формі, потребують уваги правоохоронних органів, та інтелектуальний аналіз тональності може допомогти виявити такі натяки.

– Інтернет-безпека. Моніторинг та аналіз тональності текстової інформації також є важливими інструментами для боротьби з фейками, кібербулінгом та негативним контентом в інтернеті.

Отже, у сучасному інформаційному суспільстві вміння розуміти емоційний стан та наміри людей на основі тексту має стратегічне значення для багатьох сфер діяльності. Емоційна тональність тексту може бути важливою для багатьох застосувань, таких як аналіз відгуків користувачів, визначення настроїв ринку, виявлення відгуків у соціальних мережах, фільтрація інформації тощо. Також вона може бути корисною для покращення якості комунікації та взаємодії з користувачами в різних додатках і системах.

#### Останні публікації

Оцінка емоційної тональності тексту зазвичай виконується з використанням засобів обробки природної мови (Natural Language Processing, NLP). Деякі методи і техніки NLP дозволяють автоматично визначати емоційну оцінку тексту на основі слів, фраз, контексту та інших ознак [1, 2].

Так, у статті [3] запропоновано фреймворк «двонаправлений емоційний рекурентний блок» що використовується для аналізу розмовних настроїв. У запропонованій системі узагальнений нейронний тензорний блок, за яким слідує двоканальний класифікатор, призначений для виконання контекстної композиції та класифікації настроїв відповідно.

Дослідниками [4] запропоновано метод аналізу настроїв у Twitter, заснований на словнику, який дав відповідні результати щодо настроїв щодо вакцин проти COVID-19 AstraZeneca/Oxford, Moderna та Pfizer/BioNTech за 4 місяці. Натомість, у [5] запропоновано використовувати для оцінки настрою TextBlob із векторизацією TF-IDF і моделлю класифікації LinearSVC, що дало змогу отримати точність 0.96752 для англомовних твітів.

У роботі [6] показано, що сучасні маркетингові дослідження переважно поклалися на словникові інструменти для вилучення настроїв із текстових даних, які мають явну перевагу з точки зору інтерпретації, проте явно втрачають в точності. Також авторами надано досить всебічну оцінку доступних методів аналізу настроїв, та показано, що методи на основі машинного навчання мають вищу точність класифікації, проте мають нижчий рівень інтерпретації.

У роботі [2] досліджується використання розширених моделей BERT для розпізнавання настроїв твітів. Для успішного оцінювання за допомогою Enhanced BERT розглядається набір даних Kaggle SMILE, який перевіряється на такі емоції, як «щастя», «смуток» тощо, і класифікується відповідно до таких категорій. Експерименти показують, що ця версія моделі досягає точності 0,96.

Таким чином, напрямок автоматичного інтелектуального аналізу емоційної тональності текстової інформації для визначення поведінкових намірів є актуальним напрямом, проте для української мови таких досліджень значно менше, ніж для легко формалізованих мов типу англійської. Це пов'язано з недостатньою кількістю датасетів та з досить важкою формалізацією мови, адже розмовна українська мова характеризується значною кількістю запозичень, та окрім них ще й містить фрагменти, які запозичені з інших мов [1, 7].

**Метою роботи** є розробка методу інтелектуального аналізу емоційної тональності текстової інформації для визначення поведінкових намірів користувачів соціально-орієнтованих сервісів та засобів електронної комерції.

#### Основна частина

Задача автоматичного інтелектуального аналізу емоційної тональності текстів для визначення поведінкових намірів їх авторів зводиться до задачі класифікації:

Позитивна тональність – вказує на наявність позитивних емоцій у тексті. Це можуть бути радість, задоволення, захоплення тощо. Наприклад: «Я дуже задоволений цим результатом».

Негативна тональність – вказує на наявність негативних емоцій у тексті. Це можуть бути обурення, розчарування, гнів тощо. Наприклад: «Я розчарований такою поведінкою».

В межах даного дослідження, оцінка емоційної тональності текстів виконувалась відносно відгуків у засобах електронної комерції [8]. У свою чергу, відгуки електронної комерції мають наступні особливості:

- обмежений обсяг контенту (до 500 слів);
- малий обсяг контенту (1-3 слова);
- використання суржику, слів-покручів, професіоналізмів, жаргонів та інтегрованого мультимовного контенту.

Щодо обмеженого обсягу контенту, переважна більшість відгуків не перевищує 100 слів, а більш довгими, як правило, є негативні відгуки.

Тому і якості набору експериментальних даних було використано набір даних відгуків з платформи «Hotline». Такий вибір експериментальних даних обумовлено тим, що цікавить саме розмовний україномовний контент, який до того ж повинен бути розміченим. Оцінками слугуватимуть оцінки клієнтів, які залишають відгуки, де оцінка «Не рекомендую» – негативні відгуки, а «Рекомендую» – позитивні. Для видобутку відгуків було створене відповідне програмне забезпечення на базі бібліотеки Crawllee [9], та оброблені в подальшому засобами мови C#, розподілені на 2 каталоги – «позитив» та «негатив». Загалом датасет складається із 7656 документів, де в навчальній вибірці знаходиться 6655 документів, і з них 1331 документ використано для валідації (що складає 20% навчальної вибірки). Розподіл відгуків в датасеті проілюстровано на рис. 1, 2.



Рис. 1. Кількісний розподіл позитивних відгуків



Рис. 2. Кількісний розподіл негативних відгуків

### Вибір нейронної мережі

Для бінарної класифікації настроїв україномовних відгуків електронної комерції розглядалися як нейромережеві варіанти, так і інші варіанти розв'язку поставленої задачі. Однак, виходячи з проведеного аналізу публікацій, в якому показано, що дослідження які переважно поклалися на словникові інструменти для вилучення настроїв із текстових даних мають явну перевагу з точки зору інтерпретації, але явно втрачають в точності. Серед розглянутих вище нейромережевих засобів на сьогоднішній день BERT-подібні мережі вважаються найкращими.

BERT було розроблено, щоб допомогти комп'ютерам зрозуміти значення неоднозначної мови в тексті, використовуючи навколишній текст, щоб зрозуміти контекст, у якому цей текст міг бути написаний [10]. Проте, як вже було досліджено авторами [11], ukr-RoBERTa, ukr-ELECTRA та XLM-R large мають тенденцію демонструвати найвищу продуктивність, хоча XLM-R large та ukr-ELECTRA мають тенденцію працювати краще на довших текстах, тоді як ukr-RoBERTa значно перевершує інші моделі на коротших послідовностях. Оскільки дослідження проводиться на текстах відгуків інтернет-платформи «Hotline» [12], які, як правило, є короткими текстовими повідомленнями та опираючись на проведені дослідження, було прийнято рішення використовувати нейромережу RoBERTa.

### Підбір семантичної моделі мови

Варіація нейронної мережі RoBERTa (скорочення від «Надійно оптимізований підхід BERT») є варіантом моделі BERT (Bidirectional Encoder Representations from Transformers), яку розробили дослідники Facebook AI [13]. Як і BERT, RoBERTa є мовною моделлю на основі трансформера, яка використовує самоувагу для обробки вхідних послідовностей і створення контекстуалізованих представлень слів у реченні.

Однією з ключових відмінностей між RoBERTa та BERT є те, що RoBERTa навчався на значно більшому наборі даних і з використанням ефективнішої процедури навчання. Під час навчання RoBERTa використовує техніку динамічного маскуванню, що допомагає моделі вивчати більш надійні та узагальнені представлення слів.

Так як семантичний аналіз на основі нейромережевого підходу є сьогодні актуальним напрямом наукових досліджень, для української мови на сьогодні також є деякі напрацювання. Одними з яких є попередньо навчена мультимовна модель препроцесингу, що працює також і з українською мовою та ще з

понад 50 іншими мовами [14] та ембендингу [15] автора Ukjae Jeong, та входить до складу моделей бібліотеки Tensorflow\_hub мови Python. На базі цих моделей пропонується створити модель, що буде донаведено на вищеописаній вибірці експериментальних даних. Вибір мультимовних моделей обумовлено тим, що як вже було вище наведено, відгуки можуть містити текст не тільки літературною українською мовою.

Конфігурація нейронної мережі для інтелектуального аналізу емоційної тональності текстової інформації на базі обраного датасету та типу нейромережі має структуру, показану на рис. 3.

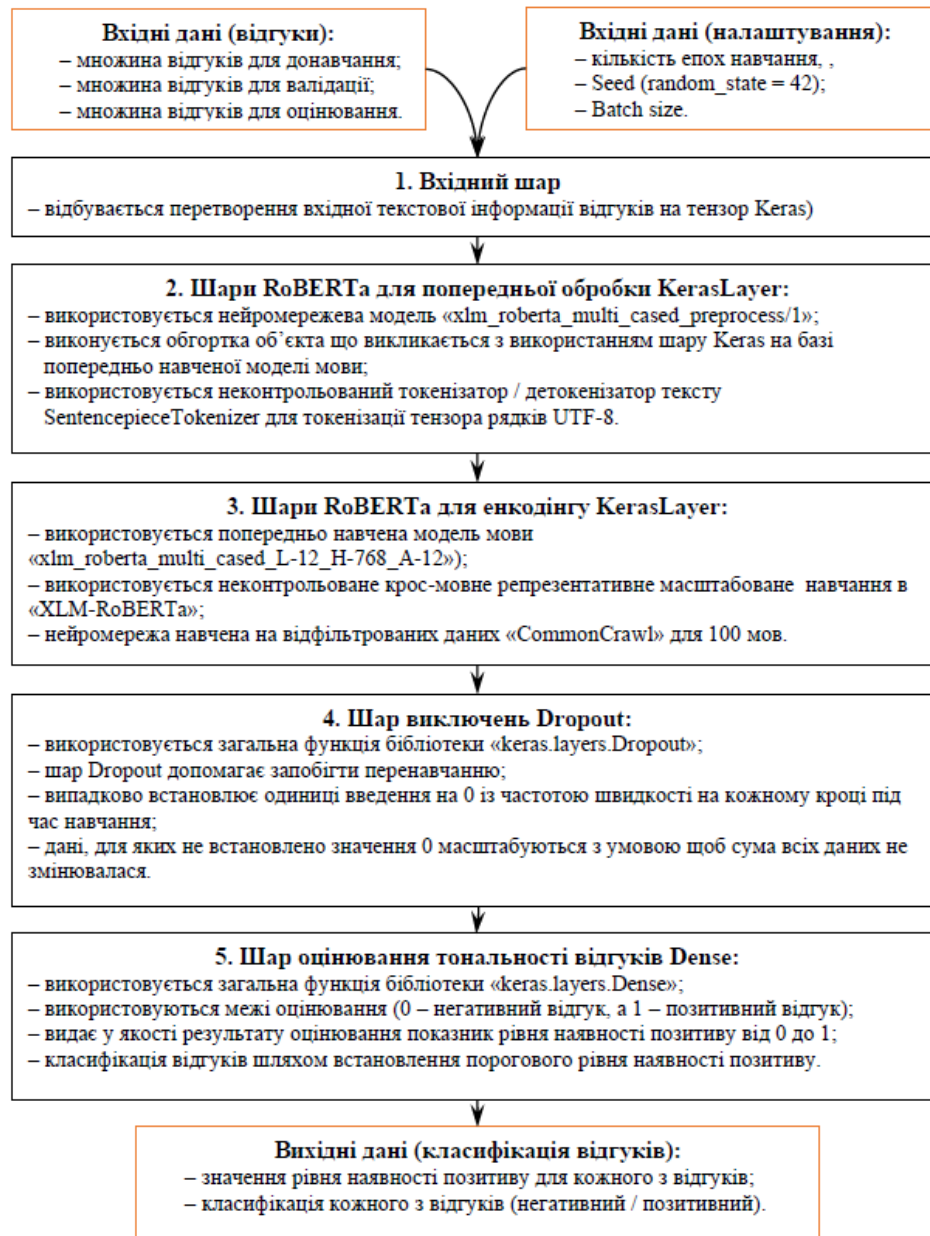


Рис. 3. Схема роботи класифікатора на основі RoBERTa для інтелектуального аналізу емоційної тональності текстової інформації

На вхідному шарі відбувається перетворення вхідної текстової інформації на тензор Keras, тобто символічний тензороподібний об'єкт, який доповнюється атрибутами, які дозволяють побудувати модель Keras за вхідним та вихідними даними моделі. Надалі тензор подається на вхід шару попередньої обробки, яка включає в себе обгортку об'єкта, що викликається, для використання як шару Keras на базі попередньо навченої моделі попередньої обробки тексту [14]. Дана модель використовує SentencepieceTokenizer [15], що токенизує тензор рядків UTF-8 та є неконтрольованим токенизатором і детокенизатором тексту.

Наступним шаром є RoBERTa енкодер. Цей шар працює на основі попередньо навченої моделі «xlm\_roberta\_multi\_cased\_L-12\_H-768\_A-12» [16], що є результатом неконтрольованого крос-мовного репрезентативного навчання в масштабі (XLM-RoBERTa) [16], та попередньо навчена на 2,5 ТБ відфільтрованих даних CommonCrawl, що містять 100 мов [17].

Наступним шаром є шар dropout, що випадково встановлює одиниці введення на 0 із частотою швидкості на кожному кроці під час навчання, що допомагає запобігти перенавчанню. Вхідні дані, для яких

не встановлено значення 0 масштабуються таким чином, щоб сума всіх вхідних даних не змінювалася.

Останнім кроком в моделі є безпосередньо класифікація, що здійснюється з використанням функції Dense та видає результат від 0 до 1, що є мірою позитиву в україномовних відгуках електронної комерції. Де 0 – негативний відгук, а 1 – позитивний відгук.

Далі запропонована модель проходить донавчання під вищеписану вибірку. Донавчання проводилось із різною комбінацією кількісних показників параметрів, таких як: кількість епох навчання, Seed, Batch size [18].

Кількість епох навчання показує, скільки разів модель підлягає навчанню. Параметр Seed буде взято 42, з огляду на те, що якщо не встановити для random\_state значення 42, щоразу, коли знову буде запускатись програмний код, він створюватиме інший тестовий набір. Batch size – кількість навчальних прикладів, що використовуються в межах однієї ітерації. Дуже важко відразу визначити, який ідеальний розмір партії для потреб конкретної задачі, тому даний параметр буде підібрано експериментальним шляхом.

Відповідно до обраних параметрів визначались показники оцінки функціональності моделі, такі як: час навчання в секундах, точність та втрати. У якості функції втрат використовувалась бінарна крос-ентропічна функція, що виражається формулою [19]:

$$Loss = -\frac{1}{N} \left[ \sum_{j=1}^N [t_j \log(p_j + (1 - t_j) \log(1 - p_j))] \right],$$

де  $N$  – кількість зразків даних,  $t_j$  є істинним значенням, яке приймає значення 0 або 1,  $p_j$  імовірність Softmax для  $i$ -ї точки даних.

Точність для проведеного дослідження визначається як ділення кількості правильних відповідей на загальну кількість відповідей.

Одержані показники оцінки функціональності (час навчання, точність та втрати) за обраних параметрів моделей налаштування (кількість епох навчання, seed, batch size) нейромережевого класифікатора наведено у табл. 1. Дослідження проводилось на базі процесора Intel Core I7 8th gen, ОЗУ 16 ГБ, NVIDIA GeForce MX150.

Таблиця 1

Параметри донавчання класифікатора

Параметри	Значення
Кількість епох навчання	5
Seed	42
Batch size	32
Час навчання (сек.)	15894
Точність	0.91
Loss	0.32

Графік ілюстрації проходження процесу донавчання по епохам показано на рис. 4, 5. Оскільки досліджувана версія RoBERTa є мультимовним трансформером, донавченим на білінгвістичних даних, в цілому нейромережа не має проблем з аналізом емоційної тональності текстової інформації.

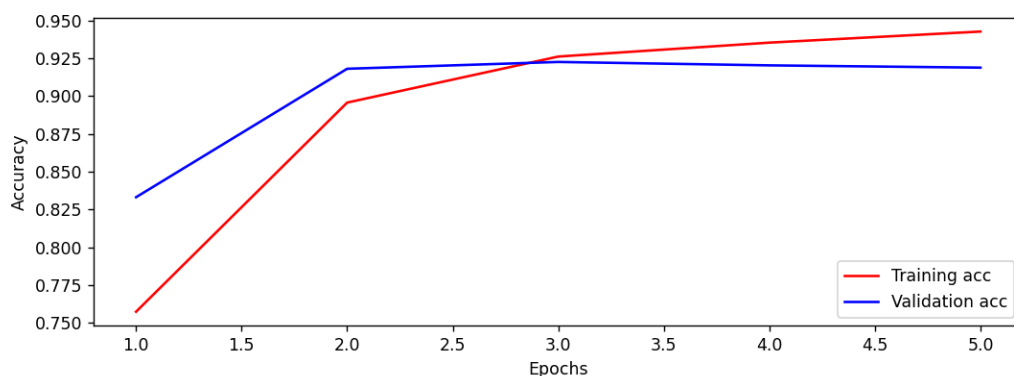


Рис. 4. Ілюстрація процесу навчання за епохами за показником точності



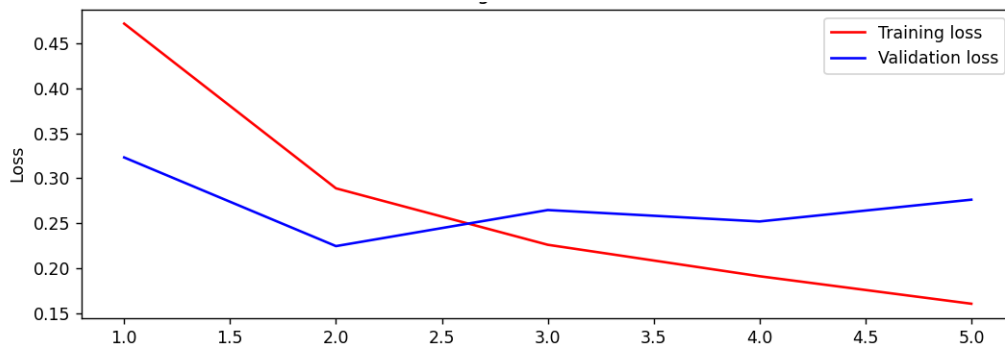


Рис. 5. Ілюстрація процесу навчання за епохами за показником функції втрат

При дослідженні відгуків, яких немає в навчальній та тестовій вибірках показано високу ефективність запропонованої архітектури. Отримані результати свідчать, що при використанні вибірки для валідації точність класифікації не росте. А функція втрат після 3ї ітерації для вибірки для валідації мала тенденцію до незначного зростання.

### Висновки

У роботі було розглянуто сучасний стан напряму семантичної обробки тексту, а саме інтелектуального аналізу емоційної тональності текстової інформації. Проведений аналіз показав, що даний напрямок є актуальним, зокрема, застосування нейронних мереж для аналізу емоційної тональності текстової інформації, що дає вищу точність класифікації, ніж альтернативні підходи. Однією із найбільш точних нейромереж визначили архітектуру BERT, в той час як для аналізу коротких документів краще себе показала її модифікація – RoBERTa.

При розробці методу досліджувались: формування розміченого датасету для навчання нейромережі, підбір та налаштування нейромережевого класифікатора, побудову семантичної моделі мови. Оскільки метою дослідження було саме аналіз емоційної тональності текстової інформації на прикладі україномовних відгуків електронної комерції, а такі відгуки мають певні характеристики, було створено власний датасет, що налічує 7656 відгуків для донавчання обраної нейронної мережі RoBERTa. Зібрані відгуки були розподілені на 2 вибірки – навчальну та тестову, кожна з яких мала негативні коментарі та позитивні коментарі. Для оцінки роботи запропонованої архітектури було використано точність та функцію втрат. Для комбінованих мультимовних відгуків вдалося отримати точність 0.92, в той час як функція втрат мала значення 0.29.

Запропонований підхід має певні обмеження. Доцільно його застосовувати до визначення емоційної тональності коротких текстових відгуків (довжиною до 500 слів), представлених на українській мові та можуть містити суржик та іншомовні вкладення слів. Зміна вмісту навчальної вибірки впливає на результат навчання нейронної мережі, і відповідно впливає на ефективність аналізу емоційної тональності текстів. З часом в побутовій мові можуть відбуватися зміни, які також впливають на хід та результати аналізу емоційної тональності текстової інформації.

### Література

1. Slobodzian V., Kovalchuk O., Molchanova M., Sobko O., Mazurets O., Barmak O., Krak I. Text Data Vectorization Model of Ukrainian-Language Internet Communication Content. CEUR Workshop Proceedings, 2022, vol. 3171, pp. 561–571. <https://ceur-ws.org/Vol-3171/paper45.pdf>
2. Mann S., Arora J., Bhatia M., Sharma R., Taragi R. Twitter Sentiment Analysis Using Enhanced BERT, in: Kulkarni, A.J., Mirjalili, S., Udgata, S.K. Intelligent Systems and Applications. Lecture Notes in Electrical Engineering, vol. 959. Springer, Singapore, 2023, pp. 263-271. DOI:10.1007/978-981-19-6581-4\_21.
3. Wei Li, Wei Shao, Shaoxiong Ji, Erik Cambria. BiERU: Bidirectional emotional recurrent unit for conversational sentiment analysis. Neurocomputing (2022): 73-82. DOI: 10.1016/j.neucom.2021.09.057.
4. Robert Marcec, Robert Likic. Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines. Postgraduate Medical Journal, Volume 98, Issue 1161, (2022): 544–550. DOI: 10.1136/postgradmedj-2021-140685
5. Miftahul Qorib, Timothy Oladunni, Max Denis, Esther Ososanya, Paul Cotaе. Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset. Expert Systems with Applications (2023). DOI: 10.1016/j.eswa.2022.118715.
6. Jochen Hartmann, Mark Heitmann, Christian Siebert, Christina Schamp. More than a Feeling: Accuracy and Application of Sentiment Analysis. International Journal of Research in Marketing (2022). DOI: 10.1016/j.ijresmar.2022.05.005.
7. Лазоренко Я., Сініцин І., Шевченко В. Ідентифікація переважної мови спілкування людини. Проблеми програмування. 2022. № 3-4. Спеціальний випуск, с. 271-280. DOI: 10.15407/pp2022.03-04.271
8. Ковальчук О.В., Слободзян В.О., Мазурець О.В., Бармак О.В. Метод формування бінарного класифікатора україномовного інтернет-контенту. Збірник наукових праць за матеріалами XIV

Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2022». Хмельницький, 2022.

9. Crawlee.Dev. A web scraping and browser automation library. 2023. <https://crawlee.dev>.
10. Sentiment Analysis of Movie Reviews with Google's BERT. 2023. <https://medium.com/mlearning-ai/sentiment-analysis-of-movie-reviews-with-googles-bert-c2b97f4217f>
11. Panchenko D., Maksymenko D., Turuta O., Yerokhin A., Daniil Y., Turuta O. Evaluation and Analysis of the NLP Model Zoo for Ukrainian Text Classification, in: Information and Communication Technologies in Education, Research, and Industrial Applications. ICTERI 2021, Communications in Computer and Information Science, vol. 1698. Springer, Cham. DOI: 10.1007/978-3-031-20834-8\_6
12. Hotline.Ua. Reviews of the store Rozetka. 2023. <https://hotline.ua/ua/yp/2476/reviews>
13. Ai.Facebook.Com., RoBERTa: An optimized method for pretraining self-supervised NLP systems. 2022. <https://ai.facebook.com/blog/roberta-an-optimized-method-for-pretraining-self-supervised-nlp-systems>.
14. Tfhub.Dev., Text preprocessing model xlm\_roberta\_multi\_cased\_preprocess. 2022. [https://tfhub.dev/jeongukjae/xlm\\_roberta\\_multi\\_cased\\_preprocess/1](https://tfhub.dev/jeongukjae/xlm_roberta_multi_cased_preprocess/1).
15. Tensorflow.Org., Text.SentencepieceTokenizer. 2023. [https://www.tensorflow.org/text/api\\_docs/python/text/SentencepieceTokenizer](https://www.tensorflow.org/text/api_docs/python/text/SentencepieceTokenizer).
16. Tfhub.Dev., Unsupervised Cross-lingual Representation Learning at Scale. xlm\_roberta\_multi\_cased\_L-12\_H-768\_A-12. 2023. [https://tfhub.dev/jeongukjae/xlm\\_roberta\\_multi\\_cased\\_L-12\\_H-768\\_A-12/1](https://tfhub.dev/jeongukjae/xlm_roberta_multi_cased_L-12_H-768_A-12/1).
17. Huggingface.Co. XLM-RoBERTa (base-sized model). URL: <https://huggingface.co/xlm-roberta-base>.
18. Zalutska O., Molchanova M., Sobko O., Mazurets O., Pasichnyk O., Barmak O., Krak I. Method for Sentiment Analysis of Ukrainian-Language Reviews in E-Commerce Using RoBERTa Neural Network. CEUR Workshop Proceedings, 2023, vol. 3387, pp. 561–571. 2023. <https://ceur-ws.org/Vol-3387/paper26.pdf>
19. Kiprono Elijah Koech, Cross-Entropy Loss Function. 2023. <https://towardsdatascience.com/cross-entropy-loss-function-f38c4ec8643e>.

#### References

1. Slobodzian V., Kovalchuk O., Molchanova M., Sobko O., Mazurets O., Barmak O., Krak I. Text Data Vectorization Model of Ukrainian-Language Internet Communication Content. CEUR Workshop Proceedings, 2022, vol. 3171, pp. 561–571. <https://ceur-ws.org/Vol-3171/paper45.pdf>
2. Mann S., Arora J., Bhatia M., Sharma R., Taragi R. Twitter Sentiment Analysis Using Enhanced BERT, in: Kulkarni, A.J., Mirjalili, S., Udgata, S.K. Intelligent Systems and Applications. Lecture Notes in Electrical Engineering, vol. 959. Springer, Singapore, 2023, pp. 263-271. DOI:10.1007/978-981-19-6581-4\_21.
3. Wei Li, Wei Shao, Shaoxiong Ji, Erik Cambria. BiERU: Bidirectional emotional recurrent unit for conversational sentiment analysis. *Neurocomputing* (2022): 73-82. DOI: 10.1016/j.neucom.2021.09.057.
4. Robert Marcec, Robert Likic. Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines. *Postgraduate Medical Journal*, Volume 98, Issue 1161, (2022): 544–550. DOI: 10.1136/postgradmedj-2021-140685
5. Miiftahul Qorib, Timothy Oladunni, Max Denis, Esther Ososanya, Paul Cota. Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset. *Expert Systems with Applications* (2023). DOI: 10.1016/j.eswa.2022.118715.
6. Jochen Hartmann, Mark Heitmann, Christian Siebert, Christina Schamp. More than a Feeling: Accuracy and Application of Sentiment Analysis. *International Journal of Research in Marketing* (2022). DOI: 10.1016/j.ijresmar.2022.05.005.
7. Lazorenko Ya., Sinityn I., Shevchenko V. Identyfikatsiia perevazhnoi movy cpilkuvannia liudyny. *Problemy prohramuvannia*. 2022. № 3-4. Spetsialnyi vypusk, c. 271-280. DOI: 10.15407/pp2022.03-04.271
8. Kovalchuk O.V., Slobodzian V.O., Mazurets O.V., Barmak O.V. Metod formuvannia binarnoho klasyfikatoru ukrainomovnoho internet-kontentu. *Zbirnyk naukovykh prats za materialamy XIV Vseukrainskoi nauково-praktychnoi konferentsii «Aktualni problemy kompiuternykh nauk APKN-2022»*. Khmelnytskyi, 2022.
9. Crawlee.Dev. A web scraping and browser automation library. 2023. <https://crawlee.dev>.
10. Sentiment Analysis of Movie Reviews with Googles BERT. 2023. <https://medium.com/mlearning-ai/sentiment-analysis-of-movie-reviews-with-googles-bert-c2b97f4217f>
11. Panchenko D., Maksymenko D., Turuta O., Yerokhin A., Daniil Y., Turuta O. Evaluation and Analysis of the NLP Model Zoo for Ukrainian Text Classification, in: Information and Communication Technologies in Education, Research, and Industrial Applications. ICTERI 2021, Communications in Computer and Information Science, vol. 1698. Springer, Cham. DOI: 10.1007/978-3-031-20834-8\_6
12. Hotline.Ua. Reviews of the store Rozetka. 2023. <https://hotline.ua/ua/yp/2476/reviews>
13. Ai.Facebook.Com., RoBERTa: An optimized method for pretraining self-supervised NLP systems. 2022. <https://ai.facebook.com/blog/roberta-an-optimized-method-for-pretraining-self-supervised-nlp-systems>.
14. Tfhub.Dev., Text preprocessing model xlm\_roberta\_multi\_cased\_preprocess. 2022. [https://tfhub.dev/jeongukjae/xlm\\_roberta\\_multi\\_cased\\_preprocess/1](https://tfhub.dev/jeongukjae/xlm_roberta_multi_cased_preprocess/1).
15. Tensorflow.Org., Text.SentencepieceTokenizer. 2023. [https://www.tensorflow.org/text/api\\_docs/python/text/SentencepieceTokenizer](https://www.tensorflow.org/text/api_docs/python/text/SentencepieceTokenizer).
16. Tfhub.Dev., Unsupervised Cross-lingual Representation Learning at Scale. xlm\_roberta\_multi\_cased\_L-12\_H-768\_A-12. 2023. [https://tfhub.dev/jeongukjae/xlm\\_roberta\\_multi\\_cased\\_L-12\\_H-768\\_A-12/1](https://tfhub.dev/jeongukjae/xlm_roberta_multi_cased_L-12_H-768_A-12/1).
17. Huggingface.Co. XLM-RoBERTa (base-sized model). URL: <https://huggingface.co/xlm-roberta-base>.
18. Zalutska O., Molchanova M., Sobko O., Mazurets O., Pasichnyk O., Barmak O., Krak I. Method for Sentiment Analysis of Ukrainian-Language Reviews in E-Commerce Using RoBERTa Neural Network. CEUR Workshop Proceedings, 2023, vol. 3387, pp. 561–571. 2023. <https://ceur-ws.org/Vol-3387/paper26.pdf>
19. Kiprono Elijah Koech, Cross-Entropy Loss Function. 2023. <https://towardsdatascience.com/cross-entropy-loss-function-f38c4ec8643e>.