

ЯРОШЕНКО ОЛЕКСАНДР

Національний технічний університет України  
"Київський політехнічний інститут імені Ігоря Сікорського"  
ORCID ID: [0000-0003-1871-3810](https://orcid.org/0000-0003-1871-3810)  
e-mail: [o.yaroshenko.ip12f@kpi.ua](mailto:o.yaroshenko.ip12f@kpi.ua)

## СЕНТИМЕНТ АНАЛІЗ ВІДГУКІВ ПРО ПРОДУКТИ ЗА ДОПОМОГОЮ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ

*У роботі наведено результати досліджень щодо використання підходів глибокого навчання для задачі сентимент аналізу відгуків про продукти з комерційних платформ. Було проведено аналіз існуючих підходів для цього завдання. Побудовано моделі на основі BERT та GPT-3, проаналізовано їх доцільність для практичного використання, виділено переваги та недоліки різних типів моделей.*

*Ключові слова: сентимент аналіз, відгуки про продукти, NLP, BERT, GPT-3*

YAROSHENKO OLEKSANDR

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

### SENTIMENT ANALYSIS OF PRODUCT REVIEWS WITH DEEP LEARNING MODELS

*This article researches the usage of modern deep learning models for sentiment analysis of product reviews gathered from online commercial platforms. For companies that launch and sell products, feedback analysis is valuable for gathering intel on their customers' opinions. Reviews allow companies to understand exactly what customers like or dislike about their products. This information helps businesses identify their products' strengths and weaknesses and use this knowledge to improve product quality, introduce new features, or solve problems that may arise. With the rapid growth of consumer review data available on the internet, opinion mining has become a complex task that can benefit from automation based on machine learning techniques.*

*This article describes the hands-on approach of review analysis using modern NLP models based on transformers, such as BERT and GPT-3. During previous work analysis, it was concluded that those models are trending in NLP nowadays.*

*Models were trained using pre-processed Amazon Product Reviews dataset. During the research, a few minor tasks were accomplished, such as GPT-3 prompt generation and selection of main metrics for accuracy evaluation. Essential steps of development were documented and described. Research results show that both models can accurately analyze text sentiment and are eligible for practical use, but GPT-3 based model shows higher balanced accuracy on provided dataset. The caveats of each approach were described and concluded.*

*The research results will help obtain more accurate and comprehensive answers about customers' attitudes to products and contribute to improving interaction in the aspect of company-client in this matter. The developed models can be applied to automate the analysis of feedback from those data sources where there is no scoring functionality, for example, social networks.*

*Keywords: sentiment analysis, product reviews, NLP, BERT, GPT-3.*

### Постановка проблеми

За останні роки інтернет став не тільки незамінним джерелом інформації, але й платформою для вираження думок і вражень користувачів щодо різноманітних продуктів та послуг. За допомогою онлайн-ресурсів, таких як соціальні мережі, веб-сайти, форуми і електронні комерційні платформи, люди активно діляться своїми досвідами та відгуками про продукти, надаючи важливу інформацію для користувачів і компаній, які ці продукти створюють.

Для компаній, які запускають та продають продукти, аналіз відгуків є цінним інструментом для збору зворотного зв'язку від своїх клієнтів. Відгуки надають компаніям можливість зрозуміти, що саме сподобалося або не сподобалося клієнтам у їхніх продуктах. Ця інформація дозволяє компаніям виявити сильні та слабкі сторони своїх продуктів і використовувати ці знання для поліпшення якості товарів, впровадження нових функцій або вирішення проблем, які можуть виникнути. Наприклад, якщо багато відгуків показують наявність певних проблем або недоліків у продукті, компанія може приступити до розробки плану вдосконалення і вирішення цих проблем. Відгуки також надають компаніям можливість впроваджувати маркетингові стратегії, спрямовані на залучення нових клієнтів та підтримку існуючих.

Для користувачів, що купують продукти, аналіз відгуків є цінним джерелом інформації перед покупкою. Користувачі шукають об'єктивну оцінку продукту від реальних людей, які вже його випробували. Відгуки допомагають користувачам отримати додаткову інформацію про продукт, включаючи його переваги, недоліки, якість, функціонал та загальне задоволення попередніх покупців. Ця інформація дозволяє користувачам зробити більш обґрунтований вибір під час покупки і уникнути ризику придбання низькоякісного або невлаштовуючого продукту. Крім того, відгуки можуть допомогти користувачам знайти продукти, що відповідають їхнім потребам та вимогам, і отримати підтримку від інших споживачів у виборі найкращого варіанту.

Зі зростанням обсягів цифрових даних, що їх потрібно аналізувати, стає складною задачею визначити загальний настрій відгуків. Тому, для компаній, які запускають та продають продукти, важливо мати ефективні методи аналізу відгуків, які допоможуть їм отримати цінну інформацію про враження та задоволення клієнтів, а також про можливі проблеми або недоліки продукту [1]. Автоматизація цього процесу може значно підвищити ефективність бізнес-процесів, пов'язаних із розробкою ідей, запуском та підтримкою наявних на ринку продуктів.

Останнім часом виникли нові підходи до аналізу відгуків, зокрема використання моделей

машинного навчання, таких як BERT (Bidirectional Encoder Representations from Transformers) та GPT-3 (Generative Pre-trained Transformer 3). Ці моделі, які були запущені в 2018 та 2020 роках відповідно, використовують глибокі нейронні мережі та трансформерну архітектуру для досягнення високої точності в обробці природної мови [2].

BERT та GPT-3 відкривають нові можливості в аналізі відгуків, оскільки вони можуть ефективно враховувати контекст та семантику тексту, розуміти відтінки сенсу та допомогти розрізнити позитивні, негативні та нейтральні відгуки. Ці моделі здатні працювати з великими обсягами даних і швидко аналізувати текстову інформацію, що робить їх ідеальними кандидатами для задачі сентимент аналізу у великому масштабі.

Метою цієї статті є розгляд задачі аналізу настрою відгуків про продукти з комерційних платформ і визначення потенціалу моделей машинного навчання, зокрема BERT та GPT-3, у вирішенні цієї задачі. В рамках цієї статті буде проведено попередню обробку та векторизацію тексту на основі готового розміченого датасету, а також розробку рішення з використанням моделей ШІ для класифікації настрою відгуків. Також будуть розглянуті існуючі рішення в галузі сентимент аналізу та вказані на їхні можливі обмеження та потенційні шляхи покращення. Результати дослідження допоможуть отримати більш точні та комплексні відповіді про ставлення клієнтів до продуктів та сприяти покращенню взаємодії в аспекті компанія-клієнт в цьому питанні. Розроблені моделі можуть бути застосовані для автоматизації аналізу відгуків з тих джерел даних, де немає функціоналу скорингу, наприклад, соціальних мереж.

### Аналіз останніх джерел

За останні 5 років спостерігається кілька важливих трендів у вирішенні задачі аналізу сентименту. Один з них - це розвиток архітектур моделей глибокого навчання для досягнення кращої точності та універсальності у сентимент аналізі. Transformers моделі, такі як BERT, RoBERTa, або GPT-3, користуються з успіхом у багатьох дослідженнях, оскільки вони можуть автоматично вивчати складні залежності між словами та контекстуалізувати інформацію з великими обсягами тексту [2].

Ще одним трендом є використання методів активного навчання та підходів з учителем для поліпшення якості сентимент аналізу. Це означає, що модель може взаємодіяти з експертами або користувачами, щоб отримувати повторну зворотну інформацію про класифікацію сентименту і вдосконалювати свою роботу в режимі реального часу [3]. Цей підхід виявляється особливо корисним в ситуаціях, коли змінюються стилі виразу сентименту або коли виникають нові види відгуків, які не були враховані в початковому тренувальному наборі даних.

Інший тренд, який можна помітити, - це розширення області застосування сентимент аналізу. Окрім класичних випадків, таких як аналіз відгуків про фільми або ресторани, сентимент аналіз стає все більш популярним у нових контекстах, таких як аналіз соціальних медіа, політичних дебатів, бізнес-відгуків та інших сфер життя. Наприклад, у статті "Sentiment analysis meets social media – Challenges and solutions of the field in view of the current information sharing context " [4] розглянуто застосування сентимент аналізу в соціальних мережах та його вплив на традиційні дослідження.

Ще однією значною тенденцією є розробка систем сентимент аналізу з урахуванням культурних та мовних особливостей. Оскільки сентимент може бути виражений різними способами в залежності від культурного тла та мови, дослідники працюють над створенням моделей, які можуть враховувати ці особливості. Наприклад, у статті [5] автори аналізують методи та виклики, пов'язані з врахуванням перекладу, мовних відмінностей та культурних нюансів у сентимент аналізі. Впровадження сучасних методів аналізу тексту для інших мовних груп має значну практичну цінність, адже більшість нових досліджень беруть за основу найпоширеніші мови та латинський алфавіт, що призводить до ускладнення впровадження результатів у певних регіонах.

Загалом, тренди в області сентимент аналізу орієнтовані на досягнення більшої точності, універсальності та контекстуалізації в аналізі сентименту. Враховуючи розширення джерел даних, використання новітніх моделей машинного навчання та розвиток нових підходів до аналізу мультимодальних та культурно зумовлених даних, можна очікувати подальше зростання в якості та поширення застосування сентимент аналізу в різних галузях, включаючи маркетинг, медицину, політичний аналіз та багато інших.

**Метою роботи є:** налаштування моделей Bert та GPT-3 для задачі оцінки сентименту та проведення порівняльного аналізу цих моделей.

### Виклад основного матеріалу

У цьому розділі описано методи та матеріали, що були використані для виконання завдання даного дослідження. Першим етапом у процесі визначення тону тексту за допомогою моделей машинного навчання є збір даних. За основу для виконання завдання в представленому дослідженні був взятий набір даних [6], що містить відгуки про продукти з онлайн-платформи Amazon, з оцінкою відгуку за 5-ти бальною шкалою.

Після отримання даних, була проведена їх обробка, а саме – значення сентименту для кожного відгуку було перетворено у 2 класи – "positive" (значення Score  $\in [4;5]$ ) та "negative" (Score  $\in [1;2]$ ). Нейтральні відгуки (зі значенням Score=3) було відкинуто. Було встановлено, що датасет, який використовується для даної моделі - незбалансований, тобто кількість записів зі значенням позитивного сентименту значно перевищує кількість записів з негативним сентиментом. Для того щоб збалансувати

датасет, було обрано однакову кількість відгуків з різним значенням сентименту у фінальному представленні даних для обробки.

Наступним етапом є очистка та трансформація тексту відгуку – відкидання спецсимволів, знаків пунктуації, видалення стоп-слів, токенизація та лемматизація. Для цих задач були використані stopwords, word\_tokenize та WordNetLemmatizer з модулю nltk. Наступними кроками є тренування моделі та аналіз точності прогнозів.

За одну з основних метрик для аналізу точності моделі було взято F1-score, який поєднує в собі метрики Precision (чутливість), та Recall (повнота) та відповідає за малу кількість хибних позитивних та негативних класифікацій [8]. Але оскільки обраний датасет був збалансованим, є доцільним аналізувати також метрику balanced\_accuracy\_score, що включає в себе показники Sensitivity та Specificity. Sensitivity вимірює частку реальних позитивних результатів, які правильно передбачені серед усіх позитивних прогнозів, які може зробити модель і розраховується за такою формулою:  $Sensitivity = TP / (TP + FN)$ , де TP – правильно визначений позитивний клас, FN – хибно визначений негативний клас. Specificity вимірює частку правильно визначених негативів у загальній кількості негативних прогнозів, які може зробити модель і розраховується за такою формулою:  $Specificity = TN / (TN + FP)$ , де TN – правильно визначений негативний клас, FP – хибно визначений позитивний клас. Формула для обчислювання збалансованої точності має вигляд:

**Ошибка! Объект не может быть создан из кодов полей редактирования.,**

Для аналізу відгуків за допомогою моделі BERT був використаний модуль transformers, у якості оптимізатора використовувався AdamW - це варіант оптимізатора Adam, спрямований на вирішення проблеми розпаду ваги. Розпад ваги — це техніка регуляризації, яка штрафує великі ваги в нейронній мережі, щоб запобігти перетренуванню. Навчання проводилось на 10 епохах з використанням підмножини датасету з 2000 відгуків. Під час запуску даної моделі було досягнуто наступних результатів:

Таблиця 1

Результати оцінки сентименту з використанням моделі BERT

Class	Precision	Recall	F1-Score
Negative	0.94	0.85	0.89
Positive	0.83	0.93	0.88
Accuracy			0.89
Macro Avg	0.89	0.89	0.88
Weighted Avg	0.89	0.89	0.89

Для дослідження вище було отримано наступне значення Balanced accuracy: 0.8846. Таке значення можна вважати прийнятним для аналізу реальних відгуків та використовувати розроблену модель в практичних цілях для даних зі структурою, аналогічною навчальному набору даних.

Для аналізу GPT-3 архітектури було використано модуль openai та модель Curie, адже вона є оптимальним вибором з точки зору потужності та витрат. Варіація Curie була основою аналізу даних з огляду GPT-3 API [7]. Хоча ця модель і показує демонструє відставання від варіації Davinci в загальному, при цьому вона є в рази вигідніше з економічної точки зору при використанні у практичних цілях. Також вказано, що Curie є моделлю загального призначення, але виконує задачі класифікації та сентимент аналізу помітно краще, ніж Ada та Babbage.

Таблиця 2

Результати оцінки сентименту з використанням моделі GPT-3 з використанням необробленого тесту відгуків

Class	Precision	Recall	F1-Score
Negative	0.83	0.77	0.80
Positive	0.78	0.84	0.81
Accuracy			0.80
Macro Avg	0.81	0.80	0.80
Weighted Avg	0.81	0.80	0.80

Для побудови prompt-запиту було протестовано декілька варіантів і виявлено, що найбільш оптимальним варіантом є структура запиту з двох частин, що задають контекст роботи моделі та приклади

для доповнення. “This is a Sentence sentiment classifier with 2 classes: 'Positive' and 'Negative.’” - ця частина задає контекст. Наступна частина, що має вигляд повторюваних блоків “\nSentence: {review}\nSentiment: {sentiment}”, де review – приклад відгуку, а {sentiment} – значення “Positive” та “Negative” – оголошує моделі формат передачі даних, які потрібно аналізувати. В цій частині перші блоки мають бути зафіксованими, з однаковим контентом. Останнім блоком йде відгук, що має бути проаналізованим, для нього блок {sentiment} залишається пустим – саме його поверне модель у результаті запиту.

В результаті запуску моделі оцінки настрою для підмасиву відгуків розміром в 2000 елементів було отримано наступні результати:

Для дослідження вище було отримано наступне значення Balanced accuracy: 0.8035

Таблиця 3

**Результати оцінки настрою з використанням моделі GPT-3 з використанням обробленого тесту відгуків**

Class	Precision	Recall	F1-Score
Negative	0.96	0.93	0.94
Positive	0.93	0.96	0.94
Accuracy			0.94
Macro Avg	0.94	0.94	0.94
Weighted Avg	0.94	0.94	0.94

Для запуску вище було отримано наступне значення Balanced accuracy: 0.942

Як виявилось, для моделі GPT-3 очистка тексту від спецсимволів, пунктуації і іншого не є необхідним і модель GPT-3 показує кращі результати при використанні вхідних текстів без обробки. Це можна пояснити тим, що ця модель була натренована на величезних об'ємах даних з текстом у вихідному вигляді і може оброблювати дані з шумним або неструктурованим текстом. GPT-3 може ефективно працювати з текстом, що містить різні знаки пунктуації, аббревіатури та помилки, які зазвичай зустрічаються в природній мові.

Отже, з наведених вище результатів можна побачити, що модель GPT-3 краще підходить для задачі визначення настрою тексту. Проте, недоліком цієї моделі є те, що вона доступна через платне API і використання її на практиці для аналізу великих масивів даних є менш економічно доцільним ніж модель BERT, яка має меншу, але відносно гарну точність у виконанні задачі визначення настрою.

Загалом, розроблені моделі можна вважати прийнятними для використання у практичних цілях для аналізу відгуків на комерційних платформах. Буде доречним також використання результатів дослідження для обробки текстів з соціальних мереж, хоча лексика і контекст повідомлень з джерел цього типу є відмінним від даних в досліджуваному датасеті.

### Висновки

В роботі було продемонстровано підхід до вирішення задачі аналізу настрою для продуктових відгуків за допомогою сучасних NLP моделей.

Для виконання завдання було використано відкритий датасет з даними з онлайн-платформи Amazon, що має систему для оцінки Product Score, що був використаний при оцінці настрою.

У ході дослідження було представлено комплексний підхід до вирішення задачі настрою-аналізу з описом етапів попередньої обробки тексту та векторизації, навчання та аналізу результатів моделей. Для архітектури GPT-3 було описано логіку вибору моделі та промпту для генерації потрібної відповіді. Ця модель і показала найкращий результат зі значенням метрики balanced accuracy 0.942.

Як можна побачити з результатів оцінки точності прогнозування, моделі на основі BERT та GPT-3 архітектури показали результати в межах допустимої похибки точності; а отже підходять для завдання оцінки настрою відгуків та можуть використовуватись у комерційних цілях.

Подальшими кроками можуть бути: аналіз відгуків з використанням інших мов та джерел даних або заглиблення у aspect-based sentiment analysis для більш точного розподілу настрою відповідно до окремих якостей товарів, що надасть змогу більш точно аналізувати ставлення клієнтів до них.

### Література

1. Dang N. C. Sentiment Analysis Based on Deep Learning: A Comparative Study [Електронний ресурс] / Nhan Cach Dang, Maria N. Moreno-Garcia, Fernando De la Prieta // Electronics. – 2020. – Т. 9, № 3. – С. 483. – Режим доступу: <https://doi.org/10.3390/electronics9030483> (дата звернення: 30.07.2023). – Назва з екрана.
2. Transformers: State-of-the-Art Natural Language Processing / Thomas Wolf [та ін.] // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. – 2020. – С. 38–45.

3. Survey on reinforcement learning for language processing [Електронний ресурс] / Victor Uc-Cetina [та ін.] // Artificial Intelligence Review. – 2022. – Режим доступу: <https://doi.org/10.1007/s10462-022-10205-5> (дата звернення: 30.07.2023). – Назва з екрана.
4. Balahur A. Sentiment analysis meets social media – Challenges and solutions of the field in view of the current information sharing context [Електронний ресурс] / Alexandra Balahur, Guillaume Jacquet // Information Processing & Management. – 2015. – Т. 51, № 4. – С. 428–432. – Режим доступу: <https://doi.org/10.1016/j.ipm.2015.05.005> (дата звернення: 30.07.2023). – Назва з екрана.
5. Wan Y. A Sentiment Analysis of Star-rating: a Cross-Cultural Perspective / Yun Wan, Makoto Nakayama // Hawaii International Conference on System Sciences. – 2022.
6. Amazon Product Reviews [Електронний ресурс] // Kaggle: Your Machine Learning and Data Science Community. – Режим доступу: <https://www.kaggle.com/datasets/saurav9786/amazon-product-reviews> (дата звернення: 30.07.2023). – Назва з екрана.
7. Tingiris S. Exploring GPT-3: An Unofficial First Look at the General-Purpose Language Processing API from OpenAI / Steve Tingiris, Bret Kinsella. – [Б. м.] : Packt Publishing, Limited, 2021. – 296 с.
8. Rudzevych A.-M. P. Machine learning in sentiment-analysis of text information on the example of user attitudes regarding candidates for Ukrainian presidential elections 2019 [Електронний ресурс] / Anna-Mariia P. Rudzevych // System research and information technologies. – 2020. – № 3. – С. 78–88. – Режим доступу: <https://doi.org/10.20535/srit.2308-8893.2020.3.06> (дата звернення: 30.07.2023). – Назва з екрана.

#### References

1. Dang N. C. Sentiment Analysis Based on Deep Learning: A Comparative Study [Electronic resource] / Nhan Cach Dang, María N. Moreno-García, Fernando De la Prieta // Electronics. – 2020. – Vol. 9, no. 3. – P. 483. – Mode of access: <https://doi.org/10.3390/electronics9030483> (date of access: 30.07.2023). – Title from screen.
2. Transformers: State-of-the-Art Natural Language Processing / Thomas Wolf [et al.] // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. – 2020. – P. 38–45.
3. Survey on reinforcement learning for language processing [Electronic resource] / Victor Uc-Cetina [et al.] // Artificial Intelligence Review. – 2022. – Mode of access: <https://doi.org/10.1007/s10462-022-10205-5> (date of access: 30.07.2023). – Title from screen.
4. Balahur A. Sentiment analysis meets social media – Challenges and solutions of the field in view of the current information sharing context [Electronic resource] / Alexandra Balahur, Guillaume Jacquet // Information Processing & Management. – 2015. – Vol. 51, no. 4. – P. 428–432. – Mode of access: <https://doi.org/10.1016/j.ipm.2015.05.005> (date of access: 30.07.2023). – Title from screen.
5. Wan Y. A Sentiment Analysis of Star-rating: a Cross-Cultural Perspective / Yun Wan, Makoto Nakayama // Hawaii International Conference on System Sciences. – 2022.
6. Amazon Product Reviews [Electronic resource] // Kaggle: Your Machine Learning and Data Science Community. – Mode of access: <https://www.kaggle.com/datasets/saurav9786/amazon-product-reviews> (date of access: 30.07.2023). – Title from screen.
7. Tingiris S. Exploring GPT-3: An Unofficial First Look at the General-Purpose Language Processing API from OpenAI / Steve Tingiris, Bret Kinsella. – [S. l.] : Packt Publishing, Limited, 2021. – 296 p.
8. Rudzevych A.-M. P. Machine learning in sentiment-analysis of text information on the example of user attitudes regarding candidates for Ukrainian presidential elections 2019 [Electronic resource] / Anna-Mariia P. Rudzevych // System research and information technologies. – 2020. – No. 3. – P. 78–88. – Mode of access: <https://doi.org/10.20535/srit.2308-8893.2020.3.06> (date of access: 30.07.2023). – Title from screen.