

УДК 004.8

Кашперук Т.Р., Тищенко О.О., Мазурець О.В., Петровський С.С.

*Хмельницький національний університет*

## **ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ МЕТОДУ ВИЗНАЧЕННЯ РІВНЯ ЗАДОВОЛЕНОСТІ ЖИТТЯМ ЛЮДИНИ ЗА ТЕКСТОВИМ ОПИСОМ ЗАСОБАМИ NLP**

*Досліджений метод визначення рівня задоволеності життям людини за текстовим описом вхідними даними має текстовий опис користувача та набори навчених класифікаторів, а вихідними даними має оцінку рівня задоволеності життям в залежності від емоційної забарвленості опису. Наведено результати дослідження ефективності методу визначеності рівня задоволеності життям людини за текстовим описом засобами NLP на основі використаних датасетів, яке встановило, що підхід є перспективним, оскільки задоволеність життям та щастя людини прямо впливає на психологічний добробут і загальне здоров'я.*

*Investigated method of determining the level of satisfaction with a person's life by text description has as input data a text description of the user and sets of trained classifiers, and as output data an assessment of the level of satisfaction with life depending on the emotional coloring of the description. The results of study of the effectiveness of the method of determining the level of satisfaction with a person's life by text description using NLP tools based on the datasets used are presented, which established that the approach is promising, since life satisfaction and human happiness directly affect psychological well-being and general health.*

Вже протягом тривалого часу люди усвідомлюють важливість розуміння свого емоційного стану та стеження за ним, оскільки емоційний стан безпосередньо впливає на психологічне благополуччя особи [1]. Це усвідомлення призвело до створення окремої науки, яка вивчає та аналізує емоційний стан людини – психології. Наукові дослідження різних аспектів людської психіки та їх впливу на відчуття комфорту і задоволення є важливими елементами визначення психологічного комфорту [2]. Центральними аспектами для визначення рівня задоволеності життям є стан і психологічний комфорт людини, а фахівці, які допомагають досягти цього комфорту, є психологами [3].

Однією з важливих галузей досліджень в сфері ШІ є NLP, яка спеціалізується на аналізі та генерації цифрових текстів [4-9]. Метою цього напрямку є класифікація та генерація текстів, аналіз тональності текстів, розпізнавання мови та інші завдання, що пов'язані з обробкою мови [10]. У сфері NLP існує різноманіття методів розв'язку завдань, серед яких можна виділити використання нейронних мереж, пошукових методів та моделей машинного навчання [11, 12].

Нейронні мережі є моделями комп'ютерних систем, які моделюють структуру та функціонування нейронних мереж у людському мозку [13]. Вони відіграють ключову роль у машинному та глибокому навчанні, застосовуючись для розв'язання різноманітних завдань, таких як класифікація, регресія, кластеризація, розпізнавання образів, машинний переклад, генерація тексту та інші [14]. Нейронні мережі також відзначаються високою адаптивністю до різноманітних типів даних та завдань, що робить їх універсальним інструментом у багатьох галузях [15]. Крім того, завдяки своїй здатності до автоматичного визначення корисних ознак у вхідних даних, нейронні мережі можуть ефективно вирішувати завдання, для яких складно вручну проєктувати ознаки [16]. Однак вони мають певні обмеження, зокрема складність у розумінні та інтерпретації отриманих результатів.

Машинне навчання дозволяє розглядати нові можливості для автоматизації цього процесу. Зокрема, шляхом аналізу описів життя людини можна визначити її рівень задоволеності, що спростить процес його оцінки [17, 18]. Проаналізувавши різні методи у сфері NLP та пошукових алгоритмів, виявлено, що для реалізації програми визначеності рівня задоволеності життям доцільно застосувати метод SVM. Цей вибір обумовлений його здатністю ефективно працювати з великим обсягом даних, забезпечуючи надійні результати класифікації. Крім того, враховуючи високу адаптивність методу до різноманітних типів даних та завдань, а також його здатність гнучкого керування параметрами, можна зробити висновок про вибір SVM як оптимального методу для даної задачі в рамках сфери NLP.

Метою роботи є дослідження ефективності методу визначення рівня задоволеності життям людини за текстовим описом засобами NLP, де вхідними даними є текстовий опис користувача та набори навчених класифікаторів, з яким у наступних етапах будуть здійснюватися операції. На першому етапі відображено підготовку даних, які включає в себе видалення стоп-символів та слів. Цей етап є обов'язковим перед використанням алгоритму SVM для аналізу текстових даних. Попередня обробка тексту дозволяє забезпечити зручні умови для подальшого використання алгоритму та оптимальну роботу класифікатора. Далі відбувається перетворення текстових даних у числовий векторний формат – векторизація даних, а саме: побудова векторного простору, де кожне слово представлене унікальним індексом. Третій крок – оцінка текстового опису класифікаторами. Програма видає оцінку згідно навчених моделей та проаналізованого тексту. Останнім етапом є формування оцінки рівня задоволеності життям людини за наступною формулою:

$$\text{Оцінка} = SVM1 + SVM2 + (1-SVM3) + (1-SVM4)/4,$$

де  $SVM1$  та  $SVM2$  – це моделі, які навчені на емоційно позитивних датасетах, а  $SVM3$  і  $SVM4$  – навпаки, на емоційно негативних даних.

Вихідними даними є оцінка рівня задоволеності життям за шкалою від нуля до одиниці, в залежності від емоційної забарвленості опису. Якщо текст більше негативний, ніж позитивний – бал буде наближений до нуля, а якщо більш позитивний – до одиниці.

За результатами досліджень ефективності методу визначеності рівня задоволеності життям людини за текстовим описом засобами NLP було створено графіки та таблиці, що відображають результат функціонування програми. Для написання методу визначення рівня задоволеності життям людини було використано метод SVM, що має для навчання декілька типів ядра, і було протестовано різні варіанти ядра та їх вплив на кінцевий результат.

З рисунка 1 видно, що характеристики у різних ядрах однакові, що свідчить що дані, що аналізуються, добре розділяються у просторі ознак за допомогою кожного з використовуваних ядер.

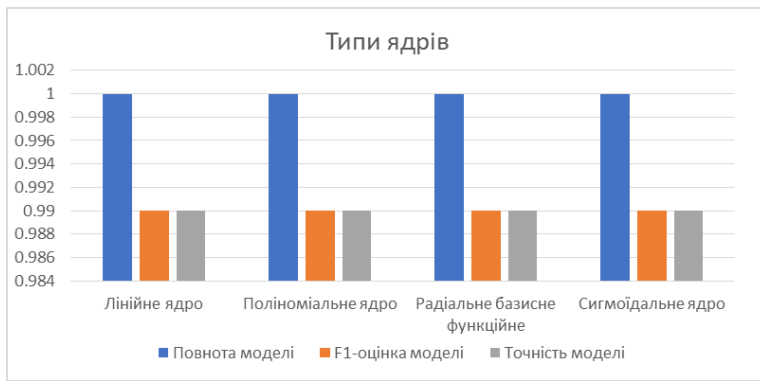


Рисунок 1 – Діаграма характеристики ядерів

Також було проаналізовано довжину чотирьох датасетів, що зображено на рисунку 2.

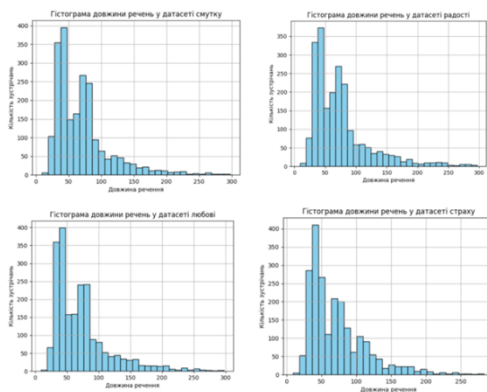


Рисунок 2 – Гістограми довжини речень датасетів

Усі чотири гістограми подібні одна до одного, що говорить про добре розподілення речень у датасетах та їх якісне навчання. На рисунку 3 зображено діаграми найпопулярніших слів у датасетах згідно проведеного аналізу.

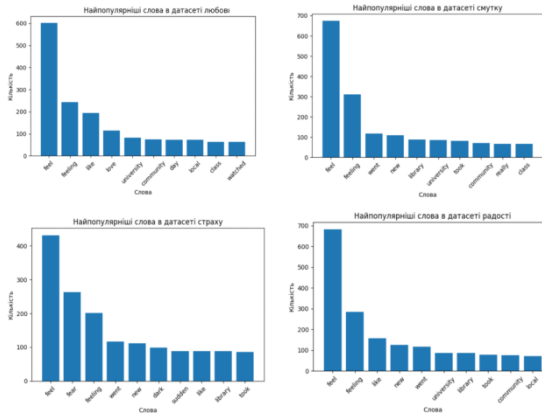


Рисунок 3 – Діаграми найпопулярніших слів у датасетах

Як видно з рисунка 3, кожна діаграма має свій набір слів. Далі було створено хмари слів на основі датасетів, які відображені на рисунку 4.



Рисунок 4 – Хмари слів

У кожній хмарі слів на рисунку 4 відображені від найбільш повторювані слова у датасетах до менш повторюваних, що видно з розмірів слів.

Таким чином, було проведено дослідження ефективності методу визначення рівня задоволеності життям людини за текстовим описом засобами NLP на основі використаних датасетів, яке встановило, що підхід є перспективним, оскільки задоволеність життям та щастя людини прямо впливає на психологічний добробут і загальне здоров'я. Один із напрямків застосування може бути психологія. У майбутньому є шляхи удосконалення підходу, такі як збільшення обсягу даних для більш точної оцінки, збільшення кількості емоцій та інше.

### Перелік посилань

1. Liga.net. Психологія майбутнього. Як штучний інтелект допоможе у боротьби зі стресом та депресією. URL: <https://life.liga.net/poyasnennya/news/psihologiya-buduschego-kak-iskusstvennyu-intellekt-pomojet-v-borbe-so-stressom-i-depressiey>
2. Solso R. L. Cognitive Psychology: Pearson New International. United Kingdom: Pearson Education, 2013. 532 p.
3. Eztests. Шкала психологічного благополуччя. URL: [https://www.eztests.xyz/tests/personality\\_ryff/](https://www.eztests.xyz/tests/personality_ryff/)
4. 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. 344–356.
5. Krak I., Didur V., Molchanova M., Mazurets O., Zalutska O., Manziuk E., Barmak O. Method for Political Propaganda Detection in Internet Content Using Recurrent Neural Network Models Ensemble. CEUR Workshop Proceedings, 2024, vol. 3806, pp. 312-324.
6. Залуцька О.О., Молчанова М.О., Віт Р.В., Мазурець О.В. Конфігурування нейронної мережі для класифікації емоційної тональності текстової інформації за показниками семантичної зв'язності. Збірник наукових праць за матеріалами XV Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2023». Хмельницький, 2023. с. 102-107.
7. Мазурець О.В., Віт Р.В. Дослідження ефективності методу виявлення цільових об'єктів предметної області. Інформаційні технології і автоматизація. Матеріали XVII міжнародної науково-практичної конференції. Одеса, ОНТУ. 2024. С.650-653.
8. Mazurets O., Tymofiev I., Dydo R. Approach for Using Neural Network BERT-GPT2 Dual Transformer Architecture for Detecting Persons Depressive State. Ricerche scientifiche e metodi della loro realizzazione: esperienza mondiale e realtà domestiche. Raccolta di articoli scientifici con gli atti della VI Conferenza scientifica e pratica internazionale. 15 novembre, 2024. Bologna, Repubblica Italiana. 2024. Pp. 147-151.
9. Мазурець О.В., Молчанова М.О., Клименко В.І., Собко О.В., Супрун П.К. Даталогічна модель бази даних для виявлення гендерної приналежності за SVM-аналізом дописів інтернет-мереж з використанням об'єктно-орієнтованого проектування. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2024. №3, Т.2 (337). С. 197-204.
10. Mazurets O., Molchanova M., Klimentko V., Prosvitliuk M Practice Implementation of Neural Network Model BART-Large-CNN for Text Annotation. Prospects of Scientific Research in the Conditions of the Modern World. Proceedings of XXVII International scientific and practical conference. Rotterdam, Netherlands. 2024. Pp. 97-102.

11. Krak I., Zalutska O., Molchanova M., Mazurets O., Bahrii R., Sobko O., Barmak O. Abusive Speech Detection Method for Ukrainian Language Used Recurrent Neural Network. CEUR Workshop Proceedings, 2024, vol. 3688, pp. 16-28.
12. Sobko O., Mazurets O., Didur V., Chervonchuk I. Recurrent Neural Network Model Architecture for Detecting a Tendency to Atypical Behavior Of Individuals by Text Posts. Theoretical and Practical Aspects of Modern Research. Proceedings of XXVI International scientific and practical conference. International Scientific Unity. Ottawa, Canada. 2024. Pp. 113-117.
13. Mazurets O.V., Sobko O.V., Molchanova M.O., Zalutska O.O., Yurchak A.V. Practical Implementation of Neural Network Method for Stress Features Detection by Social Internet Networks Posts. Global Science: Prospects and Innovations. Proceedings of the II International Scientific and Theoretical Conference «Scientific Review of the Actual Events, Achievements and Problems». Berlin, Federal Republic of Germany: International Center of Scientific Research. 2024. Pp. 160-167.
14. Molchanova M., Mazurets O., Sobko O., Boiarchuk I. Object-Oriented Approach for Ethnic Enmity Detection in Text Messages by NLP. Proceedings of XXI International Scientific and Practical Conference «Scientific Achievements and Innovations as a Way to Success». Vilnius, Lithuania. 2024. Pp. 73-77.
15. Blazhuk V., Mazurets O., Zalutska O. An Approach to Using the mBERT Deep Learning Neural Network Model for Identifying Emotional Components and Communication Intentions. The Impact of Scientific Research on the Development of the Modern World. Proceedings of the XLIV International scientific and practical conference. Dubrovnik, Croatia. 2024. Pp. 79-84.
16. Молчанова М.О., Мазурець О.В., Собко О.В., Кліменко В.І., Андрощук В.І. Метод нейромережевого виявлення кібербулінгу з використанням хмарних сервісів та об'єктно-орієнтованої моделі. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2024. №2 (333). С. 200-206.
17. Бармак О.В., Молчанова М.О., Денисенко Б.О. Нейромережева модель для виявлення дезінформації в текстовому контенті. Інформаційні технології і автоматизація. Матеріали XVII міжнародної науково-практичної конференції. 31 жовтня – 1 листопада 2024 р. Одеса, ОНТУ. 2024. С.583-585.
18. Залуцька О.О., Кліменко В.І., Гладун О.В. Нейромережева модель для визначення емоційного стану людини у режимі реального часу. Інформаційні технології і автоматизація. Матеріали XVII міжнародної науково-практичної конференції. Одеса, ОНТУ. 2024. С.614-617.