УДК 004.4:004.8

# СПЕЦИФІКАЦІЯ ВИМОГ ТА ПРОЄКТУВАННЯ ПРОГРАМНОЇ СИСТЕМИ ДЛЯ АНАЛІЗУ РЕАКЦІЙ КОРИСТУВАЧІВ НА ПУБЛІКАЦІЇ В СОЦІАЛЬНИХ МЕРЕЖАХ

Халюза Я.Д.[1]; Матвіюк Д.Д.[2]; Винярський В.А.[3]
[1,2,3]Національний університет «Одеська політехніка»
[1,2,3]Україна, Одеса.
[1]6193586@stud.op.edu.ua; [2]7453967@stud.op.edu.ua; [3]7433222@stud.op.edu.ua

***Анотація.*** *В даній роботі була створена специфікація вимог до програмної системи для аналізу реакцій користувача на публікації в соціальних мережах, а також проведене проєктування даної системи. Основна мета системи полягає у підвищенні охоплення публікацій в соціальних мережах та збільшенні частки позитивних відгуків до них за допомогою аналізу реакцій користувачів.*
***Ключові слова:*** *програмна система; компонент штучного інтелекту; аналіз тональності тексту; соціальна мережа; реакції користувачів.*

***Вступна частина.*** В сучасному світі стрімкий розвиток соціальних мереж призвів до того, що вони стали основним інструментом комунікації для різних ЗМІ, компаній, підприємств, магазинів, незалежних авторів тощо. Завдяки соціальним мережам перераховані вище суб'єкти ведуть свою діяльність та взаємодіють зі своєю аудиторією. Важливим аспектом такої взаємодії через соціальні мережі є публікації та реакції на них. Від того, наскільки добре буде сприйнята публікація, залежить чи будуть користувачі читати саме цей новинний ресурс, купляти саме цей товар, стежити саме за цим автором тощо. Тому вміння правильно оцінити реакції користувачів на публікації стає дуже важливим завданням.

Для вирішення цієї задачі існує декілька шляхів. Першим варіантом є використання послуг спеціальних аналітиків, які надають готовий аналіз реакцій на публікації та загальних настроїв аудиторії. Але такий варіант є досить дорогим та повільним, адже обробка публікацій з тисячами коментарів може займати дуже багато часу людської роботи. Іншим варіантом є використання відповідних програмних систем, але традиційні методи можуть працювати з поганою точністю через особливості людської мови.

Рішенням такої проблеми є створення програмної системи для аналізу реакцій користувачів на публікації в соціальних мережах, що буде використовувати компонент штучного інтелекту. Це дозволить підвищити швидкість проведення аналізу подібних реакцій, а компонент штучного інтелекту має підвищити точність результатів завдяки врахування особливостей людської мови та спеціальним методам її обробки.

***Метою роботи*** є створення специфікації вимог та проєктування програмної системи для аналізу реакцій користувачів на публікації в соціальних мережах.

***Основна частина.*** Першим етапом роботи над будь-якою програмною системою є створення специфікації вимог до неї. Основними елементами структури специфікації вимог є функціональні та нефункціональні вимоги до програмної системи [1]. Функціональні вимоги дозволяють побачити, що буде робити система, а нефункціональні – як вона буде працювати.

Визначення функціональних вимог до програмної системи складається з визначення основних функцій, які має виконувати система, а також опису цілей цих функцій.

Майбутня програмна система матиме наступні функції: збір даних з соціальних мереж; попередня очистка та підготовка даних; аналіз публікації; створення звітів щодо результатів

аналізу; порівняння результатів; експорт результатів; авторизація та аутентифікація. Усі ці функції необхідні для досягнення цілей програмної системи, а отже їх наявність у системі є обов'язковою вимогою. Самі ж цілі поділяються на цілі організації, провідні індикатори, результати для користувачів та властивості моделі штучного інтелекту.

Основною ціллю організації є формування широкої клієнтської бази, що дозволить ефективно монетизувати продукт. Провідними індикаторами є кількість зареєстрованих користувачів системи, кількість активних користувачів, а також їх відгуки.

Основним результатом для користувачів буде розв'язання задачі аналізу успішності публікацій на основі тональності коментарів. Саме з розв'язанням цієї задачі пов'язані усі функції системи, що визначені у функціональних вимогах.

Властивості моделі штучного інтелекту відображають вимоги за якістю, що висуваються до цієї моделі. Перш за все, модель має бути точною та робити правильні висновки щодо успішності публікації у 90% випадках або більше. Швидкість роботи є також важливою для користувачів, тому модель має проводити аналіз публікації не більше ніж за 1 хвилину. Модель має бути надійною, точність її роботи на нових даних має складати не менше ніж 90% від результатів на тренувальних даних.

Також важливим є визначення вимог за якістю вже до системи загалом. Система має працювати безвідмовно не менш ніж 98% часу. Система повинна відбивати не менш ніж 99% сторонніх атак, а дані користувачів мають бути захищені. Система має бути сумісною з соціальними мережами, кількість помилок при обміні через API не повинна перевищувати 1% від всіх запитів.

За основний патерн проєктування системи було обрано патерн MVC. Через представлення користувач буде взаємодіяти з системою. Модель буде містити усі необхідні компоненти, що будуть опрацьовувати отримані дані та виконувати запити. Контролер буде зв'язком між представленням та моделлю.

Функції представлення виконує компонент веб-інтерфейсу користувача. Модель складається з компоненту одержання даних, компоненту очистки та підготовки даних, компоненту аналізу та бази даних.

Компонент одержання даних здійснює процес отримання даних про публікацію, а саме коментарів та інших можливих реакцій. Через те, що компонент буде працювати з різними соціальними мережами, кожна з яких вимагає особливого підходу, внутрішня реалізація компоненту базується на основі патерну проєктування стратегії [2]. В результаті роботи модулю, всі уніфіковані дані передаються далі на обробку.

Після отримання даних про публікацію, компонент очистки та підготовки даних отримує їх і здійснює попередню перевірку на правильність даних, застосовує граматичні виправлення та очищає дані він знаків та послідовностей, які можуть заважати аналізу. Результат роботи даного компоненту зберігається за допомогою контролера та передається компоненту аналізу даних.

Компонент аналізу даних, отримавши очищені та підготовлені дані, здійснює обробку застосовуючи засоби штучного інтелекту. Основною задачею компоненту є проведення сентимент-аналізу всіх коментарів та здійснення класифікації кожного коментаря. На основі аналізу формується звіт в спеціальному форматі, який передається далі контролеру.

Отриманий результат контролер зберігає у базі даних з ціллю подальшого відображення результатів аналізу користувачу. Далі контролер здійснює формування звіту для користувача та передає даний звіт веб-інтерфейсу. Через веб-інтерфейс користувач переглядає отриманий аналіз публікації.

Реалізація компоненту штучного інтелекту є найбільш об'ємною задачею серед з усіх інших. Даний компонент для системи представляється як чорний ящик, на вхід якого подається текст, а на виході надається числова оцінка в заданих межах. Реалізація компоненту є комплексною і складається з наступних підзадач: формування навчальної вибірки; попередня обробка тексту; представлення тексту; проєктування та навчання нейронної мережі; оцінювання моделі [3].

Формування навчальної вибірки є найважливішим етапом при побудові моделі. Основною вимогою до даних є збалансована кількість позитивних та негативних коментарів. Інакше, результати класифікації моделі будуть мати зміщення моделі в бік домінуючого класу.

Після формування навчальної вибірки здійснюється попередня обробка тексту [4], під час якої видаляються шумові дані – теги розмітки, URL-адреси, електронні скриньки, зайві символи, числа. Далі проводиться приведення до єдиного регістру, видалення стоп-слів і рідкісних слів. Наприкінці здійснюватися лематизація – приведення, де можливо, кожного слова до початкової форми.

Попередньо оброблений текст проходить процес токенізації – поділу тексту на фрагменти з унікальними індексами. У сентимент-аналізі коментарів доцільно використовувати токенізацію на рівні частин слів, що забезпечує стійкість до орфографічних помилок та невідомих слів. Далі токени будуть передаватися шару Embedding, який перетворює токени у векторні представлення методом fasttext, який розбиває токени на частини, що мають свій вектор, та подає наступним шарам середній вектор або вектор суму.

Для обробки послідовності векторів було обрано рекурентні нейронні мережі, оскільки вони ефективно працюють із послідовними даними, враховуючи контекст попередніх елементів. У реалізації моделі було застосовано архітектуру GRU (Gated Recurrent Unit), яка є компромісним варіантом між LSTM (Long Short-Term Memory) та класичною рекурентною нейронною мережею із зворотним зв'язком, забезпечуючи спрощену структуру при збереженні здатності моделювати довгострокові залежності.

Вихід рекурентної нейронної мережі обробляється багатошаровою повнозв'язною мережею. Кінцевий шар моделі містить один нейрон із сигмоїдною функцією активації, що забезпечує формування вихідного значення в діапазоні від 0 до 1, яке інтерпретується як оцінка для відповідного коментарю.

Для оцінювання якості роботи компоненту як класифікатора буде використовуватися метрика f1-score. Дана метрика використовується як поєднання метрик precision та recall. Чим більше її значення, тим краща виходить модель.

Остаточна архітектура нейронної мережі, зокрема оптимальна кількість шарів та доцільність використання додаткових типів шарів, буде визначена на етапі програмної реалізації та експериментального навчання моделі. Ці параметри розглядатимуться як гіперпараметри, які підлягають емпіричному підбору та валідації в процесі розробки.

***Висновки.*** В результаті проведеного дослідження була визначена специфікація вимог до програмної системи, а саме визначені функціональні та нефункціональні вимоги до неї. Була визначена архітектура системи та її компоненти, а також описана модель компоненту штучного інтелекту, що буде виконувати завдання аналізу публікацій. На основі отриманих результатів проєктування буде виконуватись програмна реалізація системи. Система збиратиме дані про публікацію та проводитиме аналіз тональності тексту коментарів. На основі отриманих результатів система робитиме висновок щодо успішності публікації.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Stellman, A., & Greene, J. (2005). *Applied Software Project Management*. Sebastopol, CA: O'Reilly Media.
[2] Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994*). Design Patterns: Elements of Reusable Object-Oriented Software*. Boston, MA: Addison-Wesley Professional.
[3] Chollet, F. (2021). *Deep learning with Python* (2nd ed.). Shelter Island, NY: Manning Publication.
[4] Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed.). London: Pearson.

Khaliuza Y.D., Matviiuk D.D., Vyniarskyi V.A.

**Requirements specification and design of a software system for analyzing user reactions to social media posts**

***Abstract.*** *In this work, a specification of requirements for a software system for analyzing user reactions to publications on social networks was created, and the design of this system was completed. The main goal of the system is to increase the reach of publications on social networks and increase the share of positive responses to them by analyzing user reactions.*

***Keywords:*** *software system; artificial intelligence component; text tone analysis; social network; user reactions.*