




Хмельницький національний університет
Факультет інформаційних технологій
Кафедра комп'ютерних наук

КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

на тему Метод визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів


Галузь знань 12 – Інформаційні технології
Шифр і назва галузі знань
Спеціальність 122 – Комп'ютерні науки
Шифр і назва спеціальності
Освітня програма Комп'ютерні науки
Назва освітньої програми

Виконав: студент 2 курсу, група КНм-22-1  О.І. Мельник
Курс, група виконавця Підпис Ініціали, прізвище
Керівник: ст.викладач кафедри КН  Т.К. Скрипник
Науковий ступінь, посада Підпис Ініціали, прізвище
Нормоконтроль: к.т.н., доцент кафедри КН  Р.О. Багрії
Науковий ступінь, посада Підпис Ініціали, прізвище

До захисту допускаю:

Зав. кафедри КН, д.т.н., професор

12 12 2023 р.

 О.В. Бармак
Підпис Ініціали, прізвище

Хмельницький 2023

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет інформаційних технологій

Кафедра комп'ютерних наук

Освітній ступінь магістр

Галузь знань 12 – Інформаційні технології

Спеціальність 122 – Комп'ютерні науки

ЗАТВЕРДЖУЮ

Завідувач кафедри комп'ютерних наук



(підпис)

д.т.н., професор О.В. Бармак

« 01 » вересня 2023 року

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ МАГІСТРА**

1. Тема кваліфікаційної роботи магістра: «Метод визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів»

2. Завдання видано студенту Мельнику Олегу Івановичу

(прізвище, ім'я, по батькові)

3. Керівник роботи ст. викладач кафедри КН Скрипник Тетяна Казимирівна

(прізвище, ім'я, по батькові)

4. Затверджені наказом університету від « 15 » серпня 2023 р. № 30

5. Зміст пояснювальної записки (перелік задач) та вихідні дані:

Метою кваліфікаційної роботи магістра є вирішення задачі автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на прикладі споживчих поведінкових намірів клієнтів у межах предметної області відгуків щодо товарів щоденного вжитку в засобах електронної комерції, для чого слід розробити відповідний метод визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів, виконати проектування інформаційної системи, що буде використовувати розроблений метод, після чого слід виконати програмну реалізацію та дослідити ефективність застосування розробленого методу.

Реферат

Кваліфікаційна робота магістра присвячена розв'язку науково-технічної задачі автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на прикладі споживчих поведінкових намірів клієнтів у межах предметної області відгуків щодо товарів щоденного вжитку в засобах електронної комерції. Результатом роботи є метод, що призначений для визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів, а саме коментарів інтернет-магазину, та перетворює вхідні дані у вигляді вибірки коментарів, навченої нейромережевої моделі визначення тональності допису та вхідного допису для аналізу в вихідну інформацію у вигляді оцінки поведінкового наміру схильності до повторної купівлі обраного товару та числових оцінок компонентів споживчого поведінкового наміру.

Актуальність теми. На сьогоднішній день спостерігається зростання кількості взаємодій між споживачами та інтернет-магазинами. З поглибленням цифрової трансформації в сфері роздрібної торгівлі зростає важливість розуміння поведінкових намірів клієнтів для покращення якості обслуговування, підвищення ефективності маркетингових стратегій та підтримки прийняття рішень. Автоматизоване визначення поведінкових намірів через семантичний аналіз дописів клієнтів важливе для вирішення завдання персоналізації обслуговування, виявлення та аналізу тенденцій споживацького попиту та розробки інтелектуальних систем рекомендацій. Застосування такого методу у сфері відгуків щодо товарів щоденного вжитку дозволяє покращити якість обробки великого обсягу неструктурованих даних та забезпечити підтримку прийняття рішень відповідно до потреб споживачів.

Мета і задачі роботи. Метою кваліфікаційної роботи магістра є вирішення задачі автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на прикладі споживчих поведінкових намірів клієнтів у межах предметної області відгуків щодо товарів щоденного вжитку в

засобах електронної комерції, а також створення відповідної програмної реалізації. Для досягнення мети необхідно виконати такі завдання дослідження:

- Провести аналіз предметної області щодо визначення поведінкових намірів клієнтів електронної комерції, виділити актуальні напрями для автоматизації визначення поведінкових намірів.

- Дослідити теоретичні підходи щодо автоматизації визначення поведінкових намірів за семантичним аналізом дописів та обрати комбінацію для подальшої розробки інформаційної системи.

- Створити метод автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів щодо товарів електронної комерції категорії повсякденного вжитку.

- Спроекувати інформаційну систему на основі методу автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів щодо товарів електронної комерції.

- Обрати засоби розробки для спроектованої архітектури інформаційної системи.

- Розробити відповідну програмну реалізацію методу автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів щодо товарів у системах електронної комерції.

- Дослідити ефективність програмно реалізованого методу автоматизованого визначення поведінкових намірів клієнтів.

Об’єкт дослідження – процес визначення поведінкових намірів клієнтів щодо готовності повторного придбання товару за дописом до товару.

Предмет дослідження – моделі, методи, алгоритми визначення поведінкових намірів клієнтів за семантичним аналізом.

Методи дослідження, що застосовані для вирішення поставлених завдань, наступні: положення методів аналізу даних та теорії множин, методології проектування інформаційних систем, а також об’єктно-орієнтований підхід для розробки інформаційної системи.

Наукова новизна одержаних результатів. Результати виконання кваліфікаційної роботи магістра містять інновації та наукову новизну, зокрема було створено метод автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів щодо товарів у інтернет-магазині на основі перетворення вхідних даних у вигляді навчальної вибірки коментарів, рекурентної нейромережевої моделі та вхідного допису для аналізу у вихідні дані у вигляді числової та описової оцінки споживчого поведінкового наміру щодо купівлі обраного товару. Створений метод, будучи застосований для визначення споживчих поведінкових намірів клієнтів щодо купівлі товарів категорій повсякденного вжитку, має інноваційну властивість враховувати когнітивну (компетентність і розуміння функціональності об'єкту), афективну (емоційне схвалення об'єкту) та конативну (поведінкова дієва рішучість) компоненти, визначені за семантичним аналізом відповідних коментарів товарів системи електронної комерції. Результати дослідження сприяють підвищенню ефективності процесу купівлі-продажу товарів і послуг внаслідок покращення інформованості категорій користувачів засобів електронної комерції щодо маркетингової ситуації, визначаючи для клієнтів та власників узагальнені та локальні споживчі поведінкові наміри щодо придбання окремих товарних позицій. Розроблений метод опосередковано забезпечує для клієнтів засобів електронної комерції підвищення довіри до відгуків, що дозволяє орієнтуватись на відгуки при здійсненні покупки, а для власників покращення зворотного зв'язку для підвищення ефективності процесу купівлі-продажу товарів і послуг.

Практичне значення одержаних результатів. Було створено інформаційну систему автоматизованого визначення поведінкових намірів клієнтів, яка є прикладною програмною реалізацією метода визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів і використовувалась для проведення експериментальних досліджень з метою апробації запропонованого метода.

Проведені дослідження ефективності розробленого в роботі методу визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів свідчать про те, що даний підхід можна використовувати для аналізу маркетингової ситуації шляхом визначення для клієнтів та власників узагальнених та локальних споживчих поведінкових намірів щодо придбання окремих товарних позицій за клієнтськими дописами. Це в свою чергу сприятиме підвищенню ефективності процесу купівлі-продажу товарів і послуг, так як дозволяє для клієнтів засобів електронної комерції підвищити рівень довіри до відгуків. Що сприятиме орієнтуванню на відгуки при здійсненні покупки, а для власників сприятиме покращення зворотного зв'язку для підвищення ефективності процесу купівлі-продажу товарів і послуг.

Апробація результатів кваліфікаційної роботи магістра та публікації.

За темою роботи автором виконано наукову публікацію у фаховому виданні: Залуцька О.О., Молчанова М.О., Мазурець О.В., Мельник О.І., Скрипник Т.К. Метод інтелектуального аналізу емоційної тональності текстової інформації для визначення поведінкових намірів нейромережевими засобами. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2023. №5 (325). Т.1. С. 67-73. URL: <http://journals.khnu.km.ua/vestnik/wp-content/uploads/2023/11/325-ts-2023-5-1-3-8.pdf>

Структура і обсяг роботи. Кваліфікаційна робота магістра складається з наступного: реферату, завдання, змісту, переліку скорочень, вступу, 4 розділів, висновків, переліку посилань з 41-го найменування й 9-х додатків. Обсяг основного тексту кваліфікаційної роботи магістра становить 88 сторінок. В роботі наведено 48 світлин і 5 таблиць.

Ключові слова: поведінковий намір клієнта, SVC, LSTM, показник когнітивності, показник афектності, показник конативності, інформаційна система, візуальна аналітика.

Зміст

Перелік скорочень	4
Вступ.....	5
Розділ 1 Дослідження предметної області автоматизації визначення поведінкових намірів	9
1.1 Електронна комерція та властивості дописів щодо побутових товарів щоденного вжитку	9
1.2 Аналіз існуючих підходів до визначення поведінкових намірів	12
1.3 Вибір нейромережі для визначення афективного прояву емоційного схвалення у дописах користувача	14
1.4 Вибір підходу до визначення конативного компоненту впевненості для визначення поведінкових намірів клієнта	20
1.5 Аналіз існуючих програмних рішень та наукових підходів.....	21
1.6 Постановка задачі.....	24
Висновки до розділу 1	25
Розділ 2 Компоненти методу визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів	27
2.1 Складові показника споживчого поведінкового наміру клієнта.....	27
2.2 Узагальнена схема методу визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів.....	29
2.3 Формування моделі конативності методом SVC та визначення показника конативності	31
2.4 Навчання нейронної мережі та формування моделі афектності, визначення показника афектності нейромережею LSTM	34
2.5 Підхід до визначення показника компетентності.....	38
2.6 Визначення показника споживчого поведінкового наміру клієнта.....	40
2.7 Формування множини навчальних даних	42
2.7.1 Набір даних для визначення прояву афектності.....	42
2.7.2 Набір даних для визначення прояву конативності.....	44
Висновки до розділу 2	48

Розділ 3 Інформаційна система автоматизованого визначення поведінкових намірів клієнтів.....	50
3.1 Схема інформаційної системи	50
3.2 Формування рекомендованої комбінації засобів розробки інформаційної системи.....	52
3.3 Спеціалізовані програмні розширення для розробки інформаційної системи.....	55
3.4 Метрики для оцінки роботи класифікаторів	58
Висновки до розділу 3	59
Розділ 4 Дослідження ефективності методу автоматизованого визначення поведінкових намірів клієнтів.....	61
4.1 Програмна архітектура прикладної інформаційної системи.....	61
4.2 Особливості розробки прикладних компонентів інформаційної системи... ..	63
4.3 Прикладне тестування інформаційної системи	69
4.4 Дослідження ефективності та інтерпретація отриманих результатів засобами візуальної аналітики.....	75
Висновки до розділу 4	81
Загальні висновки.....	83
Перелік посилань.....	85
Додатки	

Перелік скорочень

Скорочення, термін, позначення	Пояснення
E-Commerce	Електронна комерція
IT	Інформаційні технології
KPM	Кваліфікаційна робота магістра
KH	Комп'ютерні науки
B2B	Електронна комерція для бізнесу
B2C	Електронна комерція для споживача
C2C	Електронна комерція від споживача до споживача
C2B	Електронна комерція від споживача до бізнесу
B2A	Бізнес-адміністрування
C2A	Електронна комерція між споживачами та адміністрацією
SVC	Support vector classifier
LSTM	Long short term memory
TF	Term frequency
TF-IDF	Term frequency - Inverse document frequency
ШІ	Штучний інтелект
ХНУ	Хмельницький національний університет
IDE	Інтегроване середовище розробки

Вступ

Кваліфікаційна робота магістра присвячена розв'язку науково-технічної задачі автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на прикладі споживчих поведінкових намірів клієнтів у межах предметної області відгуків щодо товарів щоденного вжитку в засобах електронної комерції. Результатом роботи є метод, що призначений для визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів, а саме коментарів інтернет-магазину, та перетворює вхідні дані у вигляді вибірки коментарів, навченої нейромережевої моделі визначення тональності допису та вхідного допису для аналізу в вихідну інформацію у вигляді оцінки поведінкового наміру схильності до повторної купівлі обраного товару та числових оцінок компонентів споживчого поведінкового наміру.

Актуальність теми. На сьогоднішній день спостерігається зростання кількості взаємодій між споживачами та інтернет-магазинами. З поглибленням цифрової трансформації в сфері роздрібної торгівлі зростає важливість розуміння поведінкових намірів клієнтів для покращення якості обслуговування, підвищення ефективності маркетингових стратегій та підтримки прийняття рішень. Автоматизоване визначення поведінкових намірів через семантичний аналіз дописів клієнтів важливе для вирішення завдання персоналізації обслуговування, виявлення та аналізу тенденцій споживацького попиту та розробки інтелектуальних систем рекомендацій. Застосування такого методу у сфері відгуків щодо товарів щоденного вжитку дозволяє покращити якість обробки великого обсягу неструктурованих даних та забезпечити підтримку прийняття рішень відповідно до потреб споживачів.

Мета і задачі роботи. Метою кваліфікаційної роботи магістра є вирішення задачі автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на прикладі споживчих поведінкових намірів клієнтів у межах предметної області відгуків щодо товарів щоденного вжитку в

засобах електронної комерції, а також створення відповідної програмної реалізації. Для досягнення мети необхідно виконати такі завдання дослідження:

- Провести аналіз предметної області щодо визначення поведінкових намірів клієнтів електронної комерції, виділити актуальні напрями для автоматизації визначення поведінкових намірів.

- Дослідити теоретичні підходи щодо автоматизації визначення поведінкових намірів за семантичним аналізом дописів та обрати комбінацію для подальшої розробки інформаційної системи.

- Створити метод автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів щодо товарів електронної комерції категорії повсякденного вжитку.

- Спроекувати інформаційну систему на основі методу автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів щодо товарів електронної комерції.

- Обрати засоби розробки для спроектованої архітектури інформаційної системи.

- Розробити відповідну програмну реалізацію методу автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів щодо товарів у системах електронної комерції.

- Дослідити ефективність програмно реалізованого методу автоматизованого визначення поведінкових намірів клієнтів.

Об’єкт дослідження – процес визначення поведінкових намірів клієнтів щодо готовності повторного придбання товару за дописом до товару.

Предмет дослідження – моделі, методи, алгоритми визначення поведінкових намірів клієнтів за семантичним аналізом.

Методи дослідження, що застосовані для вирішення поставлених завдань, наступні: положення методів аналізу даних та теорії множин, методології проектування інформаційних систем, а також об’єктно-орієнтований підхід для розробки інформаційної системи.

Наукова новизна одержаних результатів. Результати виконання кваліфікаційної роботи магістра містять інновації та наукову новизну, зокрема було створено метод автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів щодо товарів у інтернет-магазині на основі перетворення вхідних даних у вигляді навчальної вибірки коментарів, рекурентної нейромережевої моделі та вхідного допису для аналізу у вихідні дані у вигляді числової та описової оцінки споживчого поведінкового наміру щодо купівлі обраного товару. Створений метод, будучи застосований для визначення споживчих поведінкових намірів клієнтів щодо купівлі товарів категорій повсякденного вжитку, має інноваційну властивість враховувати когнітивну (компетентність і розуміння функціональності об'єкту), афективну (емоційне схвалення об'єкту) та конативну (поведінкова дієва рішучість) компоненти, визначені за семантичним аналізом відповідних коментарів товарів системи електронної комерції. Результати дослідження сприяють підвищенню ефективності процесу купівлі-продажу товарів і послуг внаслідок покращення інформованості категорій користувачів засобів електронної комерції щодо маркетингової ситуації, визначаючи для клієнтів та власників узагальнені та локальні споживчі поведінкові наміри щодо придбання окремих товарних позицій. Розроблений метод опосередковано забезпечує для клієнтів засобів електронної комерції підвищення довіри до відгуків, що дозволяє орієнтуватись на відгуки при здійсненні покупки, а для власників покращення зворотного зв'язку для підвищення ефективності процесу купівлі-продажу товарів і послуг.

Практичне значення одержаних результатів. Було створено інформаційну систему автоматизованого визначення поведінкових намірів клієнтів, яка є прикладною програмною реалізацією метода визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів і використовувалась для проведення експериментальних досліджень з метою апробації запропонованого метода.

Проведені дослідження ефективності розробленого в роботі методу визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів свідчать про те, що даний підхід можна використовувати для аналізу маркетингової ситуації шляхом визначення для клієнтів та власників узагальнених та локальних споживчих поведінкових намірів щодо придбання окремих товарних позицій за клієнтськими дописами. Це в свою чергу сприятиме підвищенню ефективності процесу купівлі-продажу товарів і послуг, так як дозволяє для клієнтів засобів електронної комерції підвищити рівень довіри до відгуків. Що сприятиме орієнтуванню на відгуки при здійсненні покупки, а для власників сприятиме покращення зворотного зв'язку для підвищення ефективності процесу купівлі-продажу товарів і послуг.

Апробація результатів кваліфікаційної роботи магістра та публікації.

За темою роботи автором виконано наукову публікацію у фаховому виданні: Залуцька О.О., Молчанова М.О., Мазурець О.В., Мельник О.І., Скрипник Т.К. Метод інтелектуального аналізу емоційної тональності текстової інформації для визначення поведінкових намірів нейромережевими засобами. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2023. №5 (325). Т.1. С. 67-73. URL: <http://journals.khnu.km.ua/vestnik/wp-content/uploads/2023/11/325-ts-2023-5-1-3-8.pdf>

Структура і обсяг роботи. Кваліфікаційна робота магістра складається з наступного: реферату, завдання, змісту, переліку скорочень, вступу, 4 розділів, висновків, переліку посилань з 41-го найменування й 9-х додатків. Обсяг основного тексту кваліфікаційної роботи магістра становить 88 сторінок. В роботі наведено 48 світлин і 5 таблиць.

Розділ 1 Дослідження предметної області автоматизації визначення поведінкових намірів

1.1 Електронна комерція та властивості дописів щодо побутових товарів щоденного вжитку

Автоматизація процесів продажу побутових товарів щоденного вжитку є надзвичайно корисною з позицій підвищення ефективності, зручності та конкурентоспроможності підприємства. Дописи щодо товарів є основним засобом зворотного зв'язку клієнтів з засобами електронної комерції.

Електронна комерція (E-Commerce) визначається як галузь економіки, в якій реклама, просування продуктів, торговельні угоди та фінансові транзакції здійснюються безпосередньо через Інтернет. Тобто здійснення покупок чи продажів у мережі – діяльність, що репрезентує собою E-Commerce [1].

З точки зору виробників та постачальників, електронна комерція є стратегією просування та реалізації своїх товарів чи послуг через Інтернет. З позиції покупців, або клієнтів, це означає перегляд торгових пропозицій, вибір, замовлення та проведення оплати безпосередньо в електронному середовищі.

Існує шість основних типів електронної комерції, які відрізняються основною стороною взаємодії: постачальником, споживачем та адміністрацією [2]:

– B2B є формою електронної комерції, що функціонує за принципом "від бізнесу до бізнесу". У цьому випадку одна компанія надає свої товари або послуги іншій компанії, виключаючи роздрібного споживача. Наприклад, співпраця між компаніями, такими як Sony та Apple, є ілюстративним прикладом цього типу електронної комерції.

– B2C є однією з найпоширеніших форм електронної комерції, де компанія пропонує свої товари, послуги та сервіси безпосередньо роздрібному споживачеві. Наприклад, продаж кросівок Nike на їхньому офіційному веб-сайті визначається як типовий приклад B2C електронної комерції.

– C2C передбачає електронну комерцію в мережі, де один споживач продає товари безпосередньо іншому споживачеві, незалежно від їхнього виробництва. Наприклад, платформа OLX в Україні виступає яскравим прикладом C2C електронної комерції, де користувачі можуть обмінюватися товарами та укладати угоди без посередництва виробників або постачальників.

– C2B є протилежністю B2C, де споживач надає щось бізнесу. Наприклад, індивідуальний майстер може створювати вироби, такі як гаманці чи портмоне, і продавати їх не безпосередньо споживачеві, але через інтернет-магазин, що є прикладом C2B електронної комерції.

– B2A включає електронну комерцію, що охоплює онлайн-транзакції між компаніями та державними структурами. Це застосовується у фіскальному та соціальному забезпеченні, а також при обробці різноманітних юридичних документів.

– C2A є не таким поширеним видом електронної комерції і включає транзакції між споживачами та органами державного управління. Прикладами можуть служити запис на прийом до лікаря, оплата медичних послуг, формування податкових декларацій та інші взаємодії з адміністрацією.

Моделі реалізації комерційних відносин у сфері цифрової комерції.

Інтернет-магазини. Цей вид електронної комерції є одним із найсуттєвіших у сучасному контексті. Останнім часом інтернет став платформою для купівлі та продажу практично будь-яких товарів. Інтернет-магазин представляє собою засіб для реалізації товарів, роботи або послуг шляхом укладання електронного правочину з покупцями, замовниками [3].

Одним із відомих інтернет-магазинів є Розетка (стилізовано англ. ROZETKA), що є українським інтернет-магазином, що започатковувався з 2004 року. Раніше спеціалізувався як інтернет-магазин з продажу парфумерії та косметики, та вийшов в онлайн з серпня 2004 року. Вже з червня 2005 почав працювати як інтернет-магазин із продажу техніки, що можна вмикати в розетку. З 2019 року повноцінно почав роботу як маркетплейс, а в 2020 запусився як франшиза. Наразі має понад 400 відділень у 122 містах і селищах України.

Магазин «Розетка» у 2020 році був сьомим найвідвідуванішим сайтом в Україні [4].

Сервіси та майданчики оголошень. Включаючи різноманітні ресурси, такі як OLX, ці платформи дозволяють користувачам розміщувати та переглядати оголошення про купівлю та продаж різноманітних товарів та послуг.

Сайти зі знижками та промокодами. Цей популярний напрямок дозволяє створювати агрегатори різноманітних бонусів від різних компаній, що має популярність серед споживачів.

Інтернет-аукціони. Платформи, як «eBay» та інші аналогічні сервіси, забезпечують унікальну форму купівлі та продажу, привертаючи увагу як покупців, так і продавців.

Системи онлайн-платежів. Ці сервіси надають можливість легко та зручно здійснювати різноманітні платежі через Інтернет, такі як оплата комунальних послуг чи сплата штрафів. Це розширює сферу фінансових транзакцій та підвищує рівень зручності для користувачів.

Споживачем товарів електронної комерції є особа або організація, яка використовує Інтернет для здійснення покупок товарів та послуг через електронні платформи, такі як вебсайти, мобільні додатки, або інші онлайн-канали. Такий споживач може мати різний профіль, і його покупки можуть варіюватися від товарів щоденного вжитку до великих покупок, залежно від його потреб, інтересів та можливостей.

Своє ставлення до товарів споживачі залишають у формі відгуків. Відгуки на товари характеризуються деякими властивостями [5]:

- Короткі тексти (текст відгуків на більшості платформ обмежений).
- Побутова мова (відгуки зазвичай пишуться побутовою мовою, що далека від ідеалу)
- Сленг (можуть містити сленг)
- Професіоналізми (можуть містити професіоналізми відносно товару та його характеристик).

Приклад відгуку на антисептик наведено на рисунку 1.1.

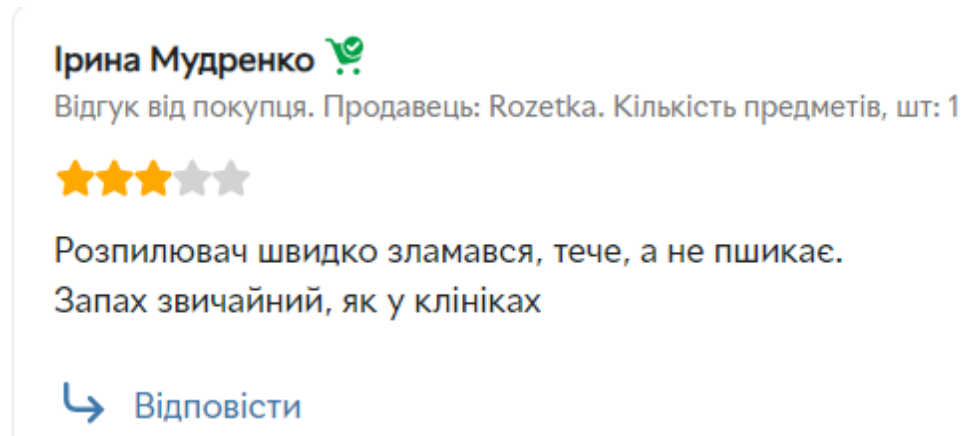


Рисунок 1.1 – Приклад відгуку з інтернет-магазину «Розетка» [6]

Отже, автоматизація процесів торгівлі сприяє підвищенню продуктивності, зменшенню помилок та витрат, поліпшенню обслуговування клієнтів та зміцненню позицій бізнесу в конкурентному середовищі. Для проведення дослідження надалі буде розглядатись електронна комерція у форматі інтернет-магазину, а саме на прикладі продажів товарів «Розетка». Також було визначено властивості дописів клієнтів щодо товарів, які формуються клієнтом залежно від його потреб, інтересів та можливостей.

1.2 Аналіз існуючих підходів до визначення поведінкових намірів

Показано, що відгуки дійсно допомагають продавати товари, викликаючи високий рівень довіри користувачів [7].

Для визначення консьюмерного наміру автора допису потрібно визначити його поведінкову установку комерційної схильності.

Під установкою комерційної схильності розуміється цілісний стан особистості, сформований на основі досвіду, і виражає готовність стійко реагувати на очікувані об'єкти або ситуації. Це вибіркова активність, спрямована на задоволення потреби [8].

Важливою формою установки визначення поведінкових намірів є соціальна установка, що також відома як атитюд. Атитюд представляє собою внутрішній стан готовності людини до дії, що передує її поведінці.

Атитюд формується на основі попереднього досвіду та проявляється на двох рівнях – усвідомленому і неусвідомленому. Він впливає на регулювання поведінки індивіда щодо визначення поведінкових намірів, надаючи йому стійкість, послідовність та цілеспрямованість в змінюваних ситуаціях.

Атитюд спрямовує і контролює дії особи, звільняючи її від постійного прийняття рішень і довільного контролю над поведінкою в стандартних ситуаціях. Однак це може також призводити до інертності у діях та уповільнення адаптації до нових ситуацій, які вимагають зміни стандартної програми поведінки.

Г. Олпорт визначає поведінкову установку як психонервову готовність особистості реагувати на всі об'єкти і ситуації, пов'язані з її життям. Ця готовність впливає на спрямованість та динаміку її поведінки і завжди корелює з минулим досвідом [9].

Шкали установок, розроблені американським психологом Луїсом Терстоном (1887 – 1955), ґрунтуються на принципі біполярного континууму з полюсами "дуже добре" – "дуже погано", "цілком згоден" – "цілком не згоден" і так далі.

Атитюд включає рівень розуміння, рівень позитивності відношення автора (емоційний) та рівень впевненості (поведінковий) компоненти за принципом біполярного континууму. Це дозволяє поведінкову установку як знання суб'єкта про об'єкт, а також як емоційну оцінку та програму дій, пов'язану з конкретним об'єктом.

Отже, визначення поведінкових намірів клієнтів за дописами щодо товарів повсякденного вжитку що продаються засобами електронної комерції необхідно визначати як складову з компонентів когнітивності, афектності та конативності.

1.3 Вибір нейромережі для визначення афективного прояву емоційного схвалення у дописах користувача

Наявність схвального і відсутність відторгаючого емоційного контексту є основною ознакою афективного прояву у дописах клієнта в системах електронної комерції. Оскільки вище було визначено, що у роботі доцільно використати підхід на основі семантичного аналізу для визначення поведінкових намірів клієнтів за їх дописами, є також доцільним окреслення нейромережевих методів ШІ для визначення рівню позитиву.

Аналіз прояву емоційного схвалення можна розглядати як більш просту форму аналізу емоцій. Завдяки цьому з тексту можна отримати суб'єктивну інформацію, яку зазвичай можна класифікувати на позитивну, нейтральну та негативну полярності [10].

Класифікація настрою може здійснюватися на трьох рівнях: на рівні речення, на рівні ознак і на рівні документа. Ці функції можна отримати за допомогою лексиконних, машинного навчання та гібридних підходів.

1. На основі лексики: лексичні методи є найпростішими підходами до визначення настрою, які поділяються на дві категорії: на основі словника та на основі корпусу. У підході, заснованому на словнику, аналіз настрою виконується на основі попередньо визначеного словника термінів, яким присвоєно значення настрою. Це використовується, наприклад, у SentiWordNet і WordNet. Корпусні методи базуються не на словниках, а на статистичному аналізі документів. Використовуються такі методи, як K-NN або приховані марковські моделі.

2. Засновані на машинному навчанні: методи, засновані на машинному навчанні, зазвичай поділяються на традиційні моделі та моделі глибокого навчання. Традиційні моделі використовують байєсівські класифікатори, класифікатори максимальної ентропії, такі як дерева рішень або опорний вектор машини. Текстові дані мають бути векторизовані, щоб уможливити використання цих класифікаторів. Точність сильно залежить від попередньої обробки. Як правило, моделі глибокого навчання дають кращі результати, ніж

традиційні моделі. Для цього використовуються CNN, DNN, RNN і мережі архітектури трансформер.

3. Гібридні підходи: гібридні підходи поєднують підходи лексикону та машинного навчання в спробі досягти кращого результату.

Оскільки на сучасному етапі розвитку галузі обробки природної мови найкращі результати показують нейромережеві методи, у роботі буде використано саме цей підхід.

Основні види нейромереж для роботи з текстом описано нижче.

RNN. Порівняно зі звичайними мережами з перенаправленням каналів, рекурентні нейронні мережі або RNN розроблені для отримання серії залежних вхідних даних. Більш конкретно, ці вхідні дані складаються з послідовностей, які взаємопов'язані, але мають різну довжину. Завдяки базовій структурі цього типу мережі він особливо підходить для обробки речень, оскільки мережа також обробляє базову структуру. З цієї причини цей тип мережі також добре підходить для розпізнавання емоцій з текстів.

На рисунку 1.2 зображена структура RNN [10]. Це представлення окремих кроків однієї комірки RNN у спрощеному вигляді. V позначає повторювані з'єднання.

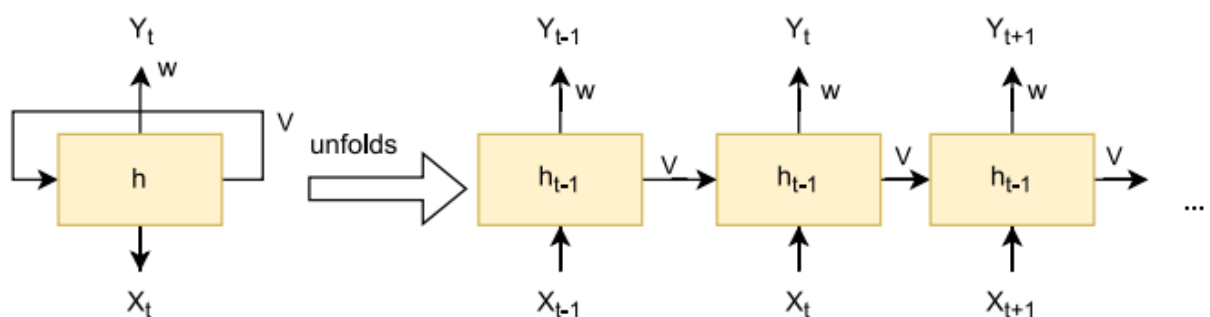


Рисунок 1.2 – Архітектура RNN [10]

Кожен шар у цьому випадку відображає конкретне слово x_i , яке представлено вкладенням. У цьому представленні легко побачити, що попередні слова впливають на наступні. Однак це працює лише в одному напрямку.

Слід також зазначити, що, на відміну від звичайних мереж, ваговий параметр w є спільним. Як і в мережах прямого зв'язку, вагові параметри регулюються за допомогою зворотного поширення.

LSTM. Мережа довгострокової короткочасної пам'яті або LSTM є більш досконалим варіантом RNN, який вирішив низку проблем [11]. Цей тип мережі дуже добре працює з великою кількістю проблем. Вони призначені для запобігання довготривалим проблемам, що дозволяє зберігати інформацію протягом більш тривалого періоду часу. LSTM також були створені для вирішення проблеми зникнення градієнта, яка може виникнути під час навчання традиційних RNN. Однак ці зміни роблять кожен нейрон складнішим. Архітектура нейромережі зображена на рисунку 1.3.

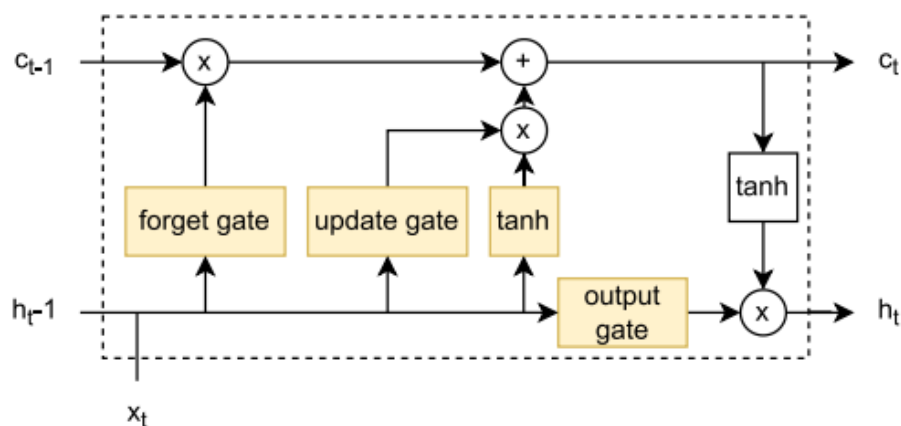


Рисунок 1.3 – Архітектура LSTM [10]

Щоб подолати проблеми стандартної комірки RNN, комірка зазвичай стає складнішою. Проте повторювана ланцюжкова структура залишається. Було введено ряд додаткових вентилів порівняно зі стандартною коміркою RNN.

Стан комірки представлено лінією, яка починається від $ct-1$ і тягнеться до ct . Тут передається інформація та відбувається легка взаємодія. За допомогою серії вентилів, що складається з рівня сигмоїдної нейронної мережі та оператора поточкового множення, можна контролювати цей інформаційний потік. За допомогою оператора *sigmoid* можна керувати потоком інформації, оскільки ця функція виводить значення від 0 до 1.

Неймережі архітектури трансформера. Архітектура була запропонована як рішення проблем, з якими стикаються традиційні мережі LSTM, наприклад, повільне навчання та складність обробки довгострокових залежностей. Загальна архітектура неймережевої архітектури трансформера зображено на рисунку 1.4. У той час як мережі LSTM представили відділення для передачі старішої інформації безпосередньо в наступній клітинці для вирішення цієї проблеми, збереження інформації в дуже довгих послідовностях залишалося складним. Ці проблеми також намагалися вирішити за допомогою варіанту LSTM. З LSTM вводиться гілка, яка пересилає старішу інформацію безпосередньо до наступної комірці. Це вирішує проблему та дозволяє зберігати інформацію довше, ніж у стандартних мережах RNN. Однак інформація не зберігається протягом дуже довгих послідовностей.

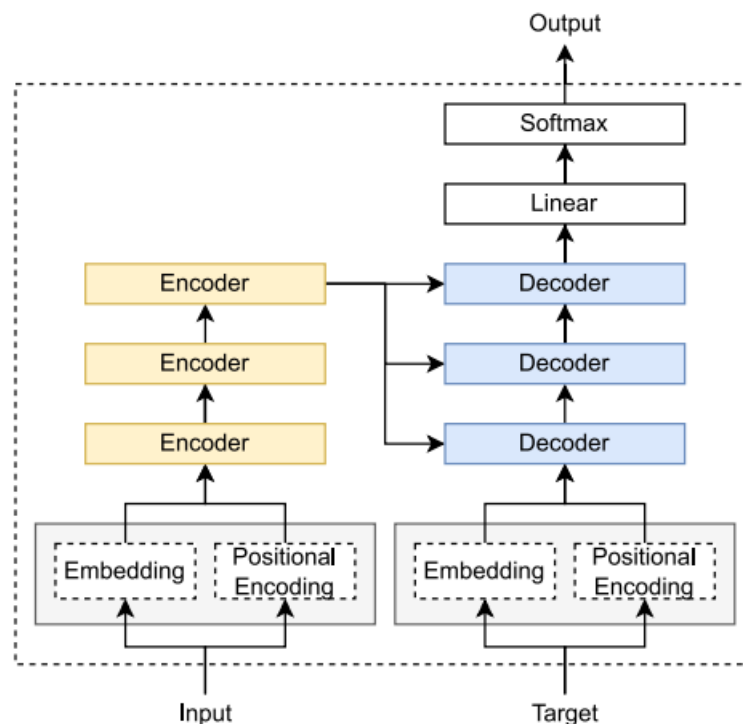


Рисунок 1.4 – Архітектура трансформера [10]

Крім того, мережі на основі RNN не можна навчати паралельно, що перешкоджає використанню хмарних обчислень і кількох графічних процесорів під час навчання. Мережа архітектури трансформер усуває ці обмеження, використовуючи механізм привертання уваги та усуваючи рекурентну

структуру, дозволяючи проводити паралельне навчання та досягати результатів, порівнянних із рекурентними моделями [12].

Як вказує назва Bidirectional Encoder Representations from Transformers, BERT є моделлю представлення мови, заснована на трансформерах. Завдяки двонаправленій структурі мережа була навчена краще розуміти контекст речень природної мови.

Основна структура дає змогу точно налаштувати попередньо підготовлені моделі BERT для конкретних завдань NLP, просто додавши вихідний рівень. Вони включають передбачення наступного речення, класифікацію тексту та класифікацію за кількома мітками. Оскільки для класифікації емоцій потрібна в основному класифікація тексту, для цієї мети можна використовувати BERT [10].

Загалом BERT ділиться на дві фази. Фаза попереднього навчання та тонкої настройки [10]. На етапі попереднього навчання BERT спочатку навчається на немаркованих даних. На етапі finetuning спочатку попередньо навчена модель додатково налаштовується за допомогою даних, що стосуються предметної області. Попереднє навчання зазвичай вимагає багато ресурсів і даних. З цієї причини в основному використовується BERT попередньо навчений і лише налаштований.

BERT використовує багатопарові двонаправлені трансформаторні мережі у фоновому режимі, які мають механізми уваги для встановлення контекстуальних зв'язків між словами у всьому тексті. У порівнянні зі звичайними трансформаторними мережами BERT не використовує механізм декодування його остаточної форми, необхідний лише механізм кодера. Двонаправлена структура дозволяє моделі вивчати контекст слів на основі всього оточення.

Для того, щоб мати можливість обробляти різні речення та слова, їх потрібно спочатку перетворити на правильне представлення. У випадку BERT використовуються так звані вбудовування WordPiece. Вони містять словниковий запас із 30 000 лексем. Також вводяться спеціальні жетони.

Перший маркер послідовності завжди є маркером класифікації [CLS]. Цей стан можна використовувати для класифікаційних завдань. Така послідовність може складатися з одного або двох речень. Пари речень з'єднуються спеціальною розділовою лексемою [SEP]. Крім того, вбудовування додається до кожного токена, який показує, чи належать відповідні токени до першого чи другого набору. Архітектура BERT наведена на рисунку 1.5.

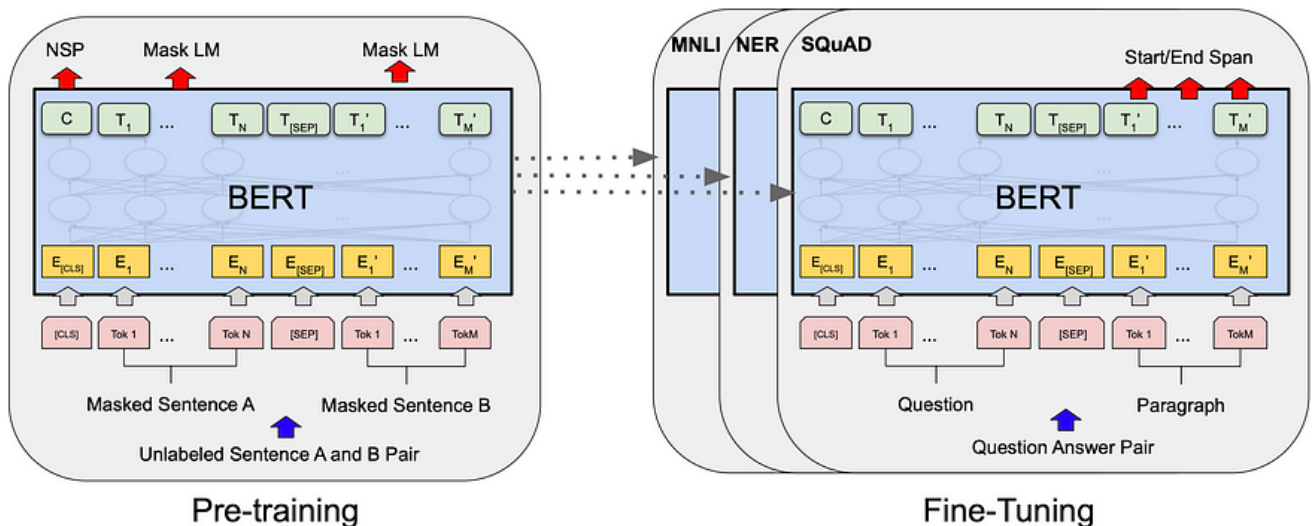


Рисунок 1.5 – Архітектура нейромережі BERT [13]

BERT пройшов попередню підготовку з двох корпусів: BooksCorpus (800 млн слів) та англійської Вікіпедії (2500 млн слів) [14]. У попередній підготовці BERT в основному використовуються два неконтрольованих завдання. Замаскований LM і передбачення наступного речення. Проте, BERT має складності в навчанні, і сам процес навчання є досить довготривалим.

Отже, зважаючи на специфіку відгуків, які зазвичай є короткими послідовностями, з проведених досліджень, у якості нейромережі для визначення прояву емоційного схвалення було обрано LSTM, оскільки його архітектура дозволяє розуміти контекст, а також показала високі результати для задачі визначення позитивності.

1.4 Вибір підходу до визначення конативного компонента впевненості для визначення поведінкових намірів клієнта

Наявність специфічних конативних слів і виразів, без врахування емоційної компоненти, є основною ознакою конативного прояву в дописах клієнтів у системах електронної комерції. Враховуючи розглянуту вище специфічність відгуків, для автоматизованого визначення рівня компонента впевненості для визначення поведінкових намірів клієнта доцільно застосовувати засоби інтелектуального аналізу даних.

Одними з найпоширеніших алгоритмів класифікації є класифікатор наївного Байєса та SVM. Класифікатор наївного Байєса – це ефективний алгоритм машинного навчання для завдань класифікації. Він ґрунтується на теоремі Байєса, яка визначає ймовірність події через урахування попередніх знань, помножену на ймовірність цих знань. Ця теорема застосовується для розрахунку ймовірності певного класу мітки для заданого набору ознак [15].

Наївний класифікатор Байєса функціонує, обчислюючи ймовірність для кожної мітки класу, враховуючи ознаки. Далі він використовує максимальну апостеріорну (MAP) оцінку для призначення мітки класу з найвищою ймовірністю. Оцінка MAP ґрунтується на припущенні, що всі ознаки незалежні, тому класифікатор отримав назву "наївний".

Support Vector Machine (SVM) – це алгоритм машинного навчання, який використовується для завдань як бінарної, так і багатокласової класифікації, а також для регресії. Основною ідеєю SVM є знаходження оптимальної гіперплощини, яка розділяє дані в просторі ознак [16].

SVM використовує концепцію ядра для відображення даних в більш високорозмірний простір ознак, де розділ між класами може бути легше здійснений. Популярні ядра включають лінійне, поліноміальне та радіально-базисне функціональне ядро.

Згідно з дослідженням [17], найкращі результати в класифікації текстів показав метод SVM, так як може працювати ефективно в просторах високої

розмірності, і не вимагає сильної незалежності між ознаками. А також має меншу чутливість до шуму, і може працювати добре навіть у випадках, коли дані мають аномалії. Тому для визначення рівня компоненти впевненості з метою визначення подальших поведінкових намірів клієнта буде використано SVM.

1.5 Аналіз існуючих програмних рішень та наукових підходів

У роботі [18] запропоновано різні методи оцінки емоційно забарвлених слів, включаючи їх ранжування, визначення коефіцієнта важливості за допомогою різних методик, таких як методика Фішберна, парне порівняння, гіпотеза Пурто і т.д. Ці методи різняться у використанні коефіцієнтів, норм та використанні логарифмічних шкал, але їхньою основною метою є встановлення порядку слів і фраз без глибокого аналізу їхньої емоційної природи та взаємозв'язків між ними. Цей підхід призначений для створення платформи, яка здатна здійснювати інтегровану оцінку, що допоможе визначити думку користувача.

У роботі [19] було запропоновано метод із застосуванням моделей глибокого навчання до задачі аналізу тональності текстових даних, що відрізняється від існуючих своєю структурою, та дозволяє підвищити точність виявлення інформаційно-психологічних впливів у контенті соціальних мереж.

Також окрім уваги науковців, дана тема також знайшла своє відображення у розроблених програмних рішеннях. Одним із таких популярних рішень є ChatGPT [20].

Чат-бот, створений на основі штучного інтелекту, розроблений компанією OpenAI, обладнаний рядом функцій. Він вміє взаємодіяти з користувачем, відповідаючи на запитання, виявляти та виправляти помилки, відмовляти від недоречних пропозицій, генерувати текстовий контент за необхідними вимогами, здійснювати пошук ключових слів за заданою тематикою, виявляти помилки в коді та виконувати інші завдання. Має веб-версію (рисунок 1.6).

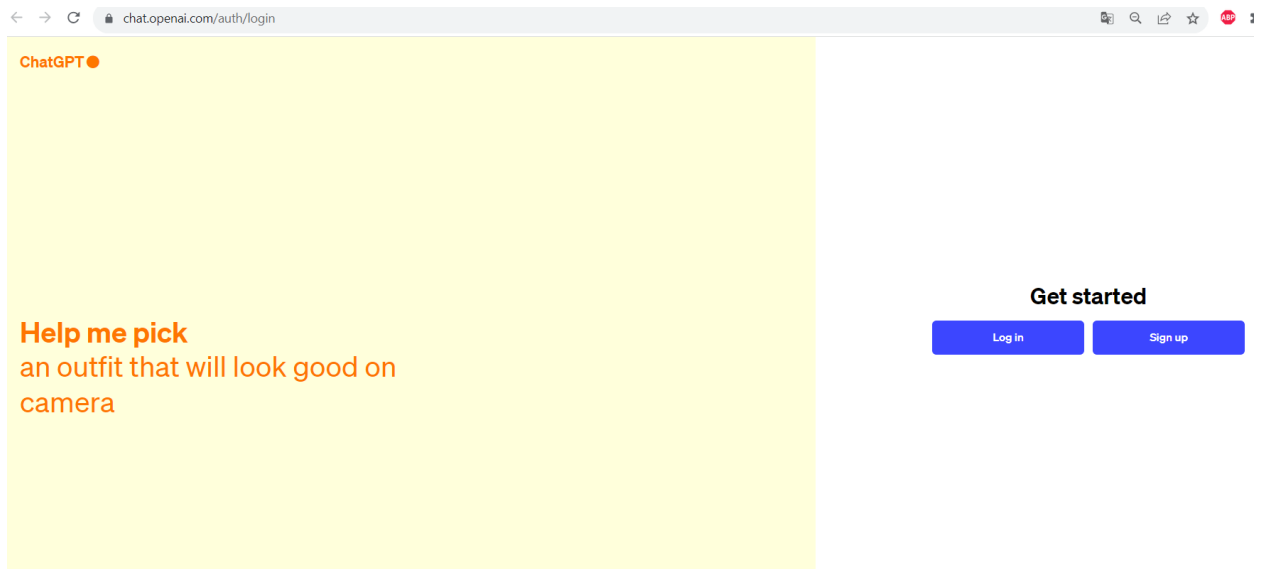


Рисунок 1.6 – Вебверсія ChatGPT

Dynamics 365 Customer Service – це рішення для управління взаємодією з клієнтами (CRM), яке надає компаніям інструменти для автоматизації, персоналізації та оптимізації обслуговування клієнтів [21].

Dynamics 365 Customer Service включає в себе функцію визначення тональності повідомлень, яка використовує машинне навчання для визначення емоційного відтінку тексту. Ця функція може бути використана для виявлення позитивних, негативних і нейтральних тональностей в чатах, електронних листах, дзвінках і інших формах спілкування з клієнтами.

Функція визначення тональності повідомлень в Dynamics 365 Customer Service працює за допомогою наступного алгоритму:

- Текст повідомлення розбивається на слова і фрази.
- Кожен слово або фраза аналізується на наявність слів і фраз, які зазвичай пов'язані з позитивними, негативними або нейтральними емоціями.
- На основі результатів аналізу кожному слову або фразі присвоюється певна оцінка тональності.
- Оцінки тональності для всіх слів і фраз в повідомленні об'єднуються для отримання загальної оцінки тональності повідомлення.

Оцінка тональності повідомлення може бути використана для прийняття рішень про те, як реагувати на повідомлення. Наприклад, якщо повідомлення має негативну тональність, компанія може надати клієнту додаткову допомогу або підтримку. Інтерфейс наведено на рисунку 1.7.

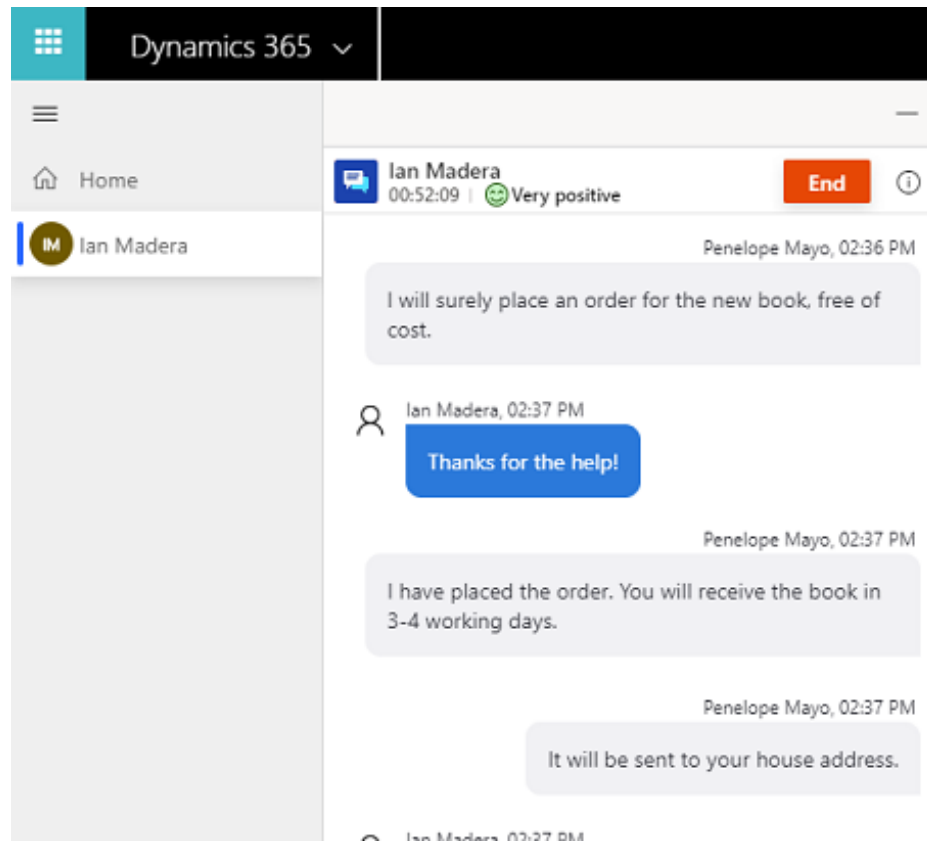


Рисунок 1.7 – Dynamics 365 [21]

Функція визначення тональності повідомлень в Dynamics 365 Customer Service може бути корисна для компаній, які хочуть [21]:

- Краще розуміти емоції своїх клієнтів.
- Швидко реагувати на негативні відгуки клієнтів.
- Покращити якість обслуговування клієнтів.

Отже, із виконаного аналізу програмного забезпечення, можна зробити висновок, що є програмні засоби для виявлення компоненти емоційного схвалення, проте відсутні програмні засоби, які б враховували додатково компоненти когнітивності та конативності. Тому подальша робота у даному напрямку є актуальною.

1.6 Постановка задачі

Метою кваліфікаційної роботи магістра є вирішення задачі автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на прикладі споживчих поведінкових намірів клієнтів у межах предметної області відгуків щодо товарів щоденного вжитку в засобах електронної комерції, а також створення відповідної програмної реалізації. Для досягнення мети необхідно виконати такі завдання дослідження:

1. Провести аналіз предметної області щодо визначення поведінкових намірів клієнтів електронної комерції, виділити практичні підходи для автоматизації визначення поведінкових намірів.

2. Дослідити теоретичні підходи щодо автоматизації визначення поведінкових намірів за семантичним аналізом дописів та обрати комбінацію теоретичних підходів для використання в методі.

3. Створити метод автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів щодо товарів електронної комерції категорії повсякденного вжитку.

4. Спроекувати інформаційну систему на основі методу автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів щодо товарів електронної комерції.

5. Обрати засоби розробки для спроектованої архітектури інформаційної системи.

6. Розробити відповідну програмну реалізацію методу автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів щодо товарів електронної комерції.

7. Дослідити ефективність програмно реалізованого методу автоматизованого визначення поведінкових намірів клієнтів.

Висновки до розділу 1

У першому розділі кваліфікаційної роботи магістра було виконано аналіз предметної області автоматизованого визначення поведінкових намірів, у рамках якого було здійснено аналіз засобів електронної комерції, зокрема в сегменті засобів продажу побутових товарів щоденного вжитку. У рамках проведеного аналізу було зазначено, що автоматизація процесів торгівлі сприяє підвищенню продуктивності, зменшенню помилок та витрат, поліпшенню обслуговування клієнтів та зміцненню позицій бізнесу в конкурентному середовищі.

Також було проведено аналіз предметної області визначення поведінкових намірів клієнтів електронної комерції за їх дописами, де було зроблено висновок, що показник визначення поведінкових намірів клієнтів за дописами щодо товарів повсякденного вжитку, які продаються засобами електронної комерції, необхідно визначати як трикомпонентну складову з компонентів когнітивності, афектності та конативності.

Було здійснено огляд існуючих варіантів нейромереж для визначення прояву емоційного схвалення у дописах користувача, де звертаючи увагу на специфіку відгуків, які зазвичай є короткими послідовностями, з проведених досліджень, у якості нейромережі для визначення прояву емоційного схвалення було обрано LSTM, так як його архітектура дозволяє розуміти контекст, а також показала високі результати для задачі виявлення прояву емоційного схвалення.

Здійснено вибір підходу визначення компоненту впевненості для визначення поведінкових намірів клієнта, та показано з досліджень, що найкращі результати в класифікації текстів показав метод SVM, так як він може працювати ефективно в просторах високої розмірності, і не вимагає сильної незалежності між ознаками. А також має меншу чутливість до шуму, і може працювати добре навіть у випадках, коли дані мають аномалії. Тому SVM було

обрано для визначення рівня компоненту впевненості з метою визначення подальших поведінкових намірів.

Також було проведено аналіз існуючих програмних рішень та наукових підходів, де було показано, що наразі існують програмні засоби для виявлення компоненти емоційного схвалення, проте відсутні програмні засоби, які б враховували додатково компоненти когнітивності та конативності. Тому було зроблено висновок, що робота у даному напрямку є актуальною.

Також визначено мету кваліфікаційної роботи магістра, якою є вирішення задачі автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на прикладі предметної області відгуків щодо товарів щоденного вжитку, а також створення відповідної програмної реалізації. А також визначено перелік задач, які потрібно виконати для досягнення поставленої мети, які містять, зокрема, необхідність розробити відповідну програмну реалізацію методу автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів щодо товарів електронної комерції для дослідження ефективності розробленого методу автоматизованого визначення поведінкових намірів клієнтів.

Розділ 2 Компоненти методу визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів

2.1 Складові показника споживчого поведінкового наміру клієнта

Як вже було вище зазначено, споживча поведінкова установка є психонервовою готовністю індивіда до привласненні та придбання всіх об'єктів в поточній ситуації, з якими поведінкова установка пов'язана. При цьому, високий рівень споживчої поведінкової установки не гарантує її здійснення в вигляді купівлі товару, оскільки можуть бути нездоланні контрфактори.

Основною психологічною ознакою споживчої поведінкової установки є інтенсивність індивідуального консьюмерного афекту – ставлення до ідеї привласнення товарного об'єкта. В психологічних дослідженнях споживчу поведінкову установку вимірюють на підставі вербальних самозвітів опитуваних, які є узагальненою оцінкою особистості власного відчуття схильності або несхильності до придбання конкретного об'єкта. Отже, споживча поведінкова установка – це міра відчуття бажання до придбання товару, викликаного конкретним товарним об'єктом («за» чи «проти»). Споживча поведінкова установка справляє спрямовуючий і динамічний вплив на споживчу поведінкові наміри, вона завжди залежна від минулого досвіду.

Структуру споживчого поведінкового наміру (за М. Смітом) [8] утворюють когнітивний (компетентність і розуміння функціональності об'єкту), афективний (емоційне схвалення об'єкту) та конативний (поведінкова дієва рішучість) компоненти (рисунок 2.1). Це дає підстави розглядати споживчий поведінковий намір одночасно як знання суб'єкта про товар і як емоційну оцінку та програму дій щодо конкретного товару. Багато вчених відзначає, що в реальному житті дуже важко відокремити когнітивний та конативний компоненти споживчого поведінкового наміру від афективного.

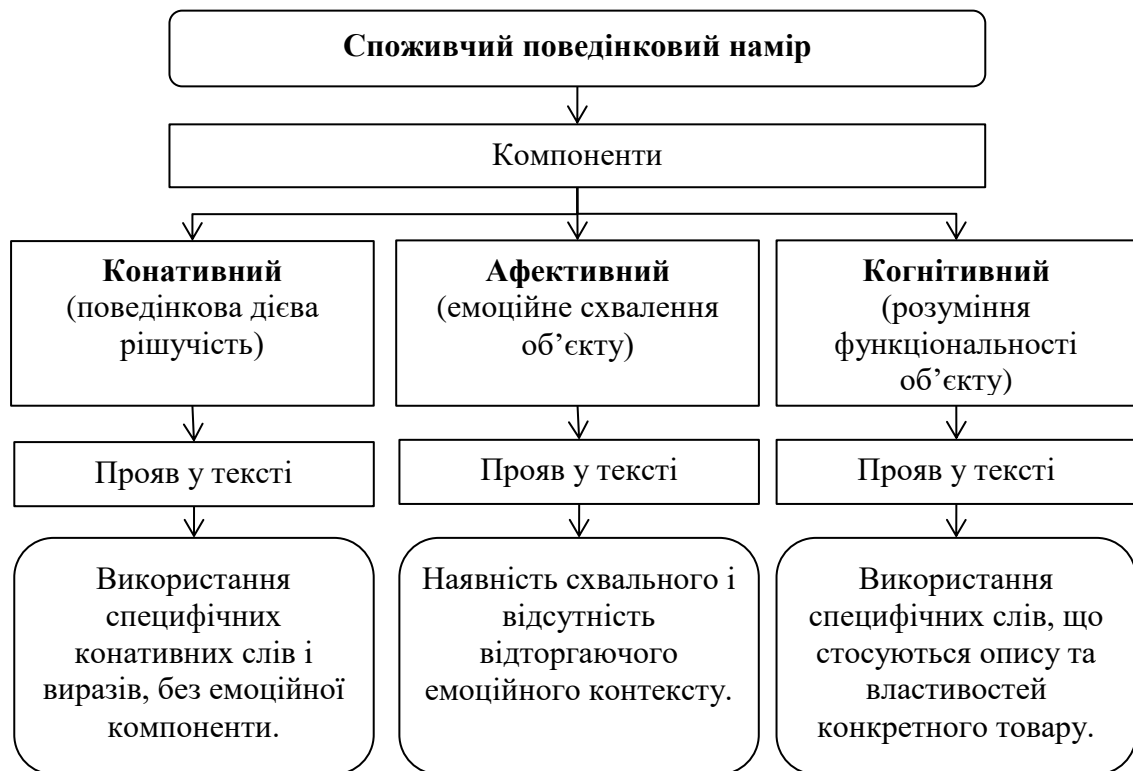


Рисунок 2.1 – Компоненти показника споживчого поведінкового наміру клієнта

Відповідно до рисунку 2.1, інформаційну модель визначення споживчого поведінкового наміру можна представити таким чином:

$$SPN = k_1 \cdot PDR + k_2 \cdot EM + k_3 \cdot FO,$$

де PDR – рівень поведінкової дієвої рішучості, EM – рівень емоційного схвалення об'єкту, FO – рівень розуміння функціональності об'єкту. k_1 , k_2 , k_3 - коефіцієнти впливу кожної з компонент. Коефіцієнти підбираються емпіричним шляхом, за замовчуванням рівні 1.

Отже, була сформована зважена трикомпонентна інформаційна модель визначення споживчого поведінкового наміру за компонентами прояву емоційного схвалення об'єкту, прояву розуміння функціональності об'єкту та прояву поведінкової дієвої рішучості.

2.2 Узагальнена схема методу визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів

Метод автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів дає можливість визначити поведінковий намір клієнта на основі перетворення вхідних даних у вигляді навченої нейромережевої моделі для визначення прояву афектності за дописом, навченої моделі класифікатора для визначення прояву конативності, вхідного клієнтського допису для аналізу у вихідну інформацію у вигляді показника споживчого поведінкового наміру клієнта та показників його складників: конативності, афектності та компетентності. Також необхідною є візуалізація отриманих результатів. Узагальнена схема методу наведена на рисунку 2.2

Вхідними даними методу визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів є навчальні вибірки для моделі SVC та нейромережевої моделі LSTM, а також користувацький допис для аналізу. Навчальні вибірки формуються із загальних підготовлених наборів даних у розмірі 80 % від загальної кількості зразків у датасетах.

Першим етапом є підготовка навчальних вибірок для SVC та LSTM. На цьому етапі здійснюється препроцесинг та токенізація текстових даних. Для SVC моделі дані проходять етап перетворення у векторне представлення з використанням підходу TF-IDF.

Другим етапом є етап формування моделі конативності методом SVC. Класифікатор навчається на навчальному наборі даних, що були отримані на першому етапі.

На третьому етапі здійснюється навчання нейронної мережі LSTM та формування відповідної моделі афектності для визначення рівня емоційного схвалення об'єкту за його текстовим дописом, що здійснюється на четвертому етапі. Після навчання модель зберігається для подальшого використання. Окрім моделі також зберігається токенізатор.

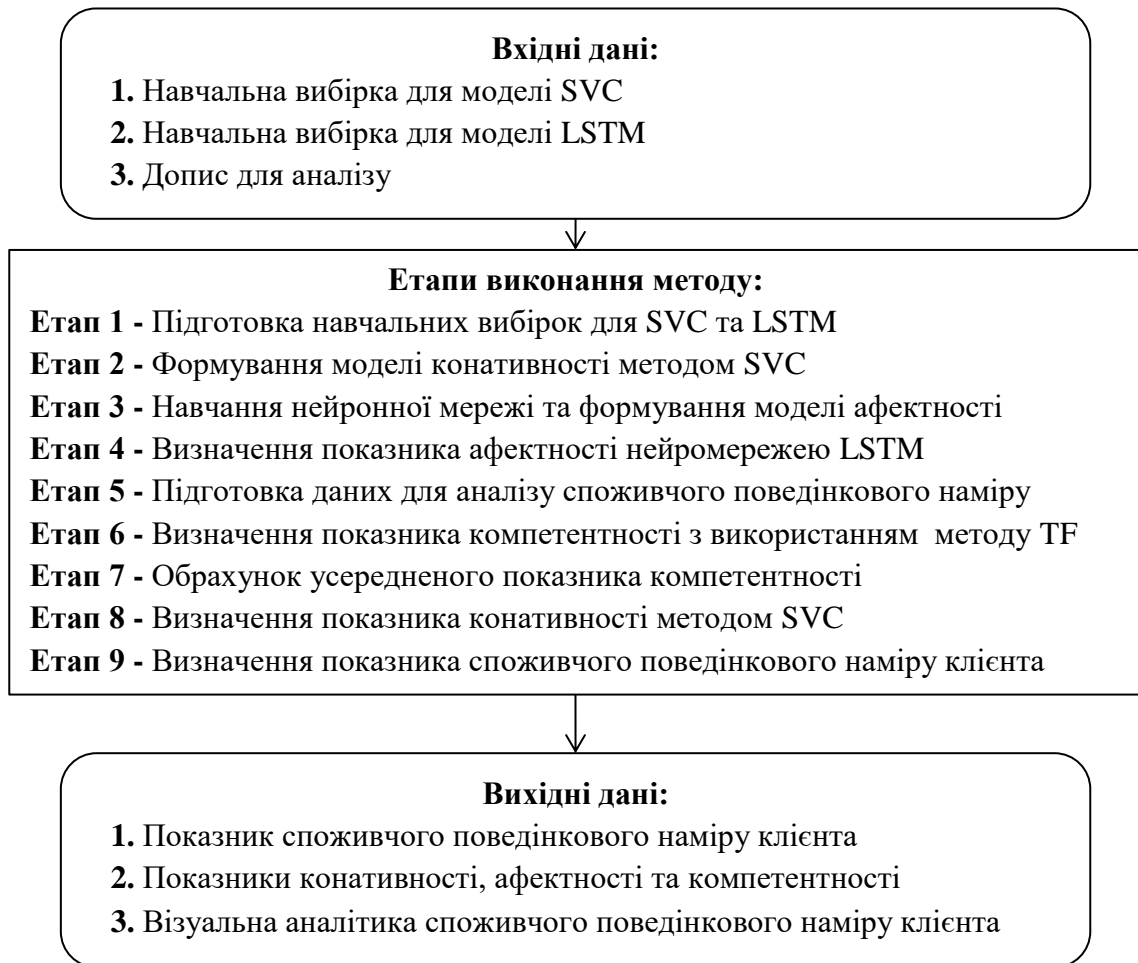


Рисунок 2.2 – Узагальнена схема методу визначення поведінкових намірів клієнтів

Наступним, п'ятим етапом, є підготовка даних для аналізу споживчого поведінкового наміру. Отримані текстові дані проходять препроцесинг, лематизацію, токенізацію та векторизацію.

Шостим етапом є визначення показника компетентності з використанням методу TF, для чого лематизований текст користувацького відгуку порівнюється із лематизованим описом товару.

Наступним етапом є обрахунок усередненого показника компетентності. Він обраховується як число співпадінь лем з відгука та лем з опису товару, без урахувань стоп-слів.

Восьмим етапом є визначення показника конативності методом SVC. Оброблені дані з п'ятого етапу подаються як вхідні дані навченому

класифікатору SVC, та отримується відповідь у вигляді числової оцінки рівня поведінкової дієвої рішучості.

Останнім етапом є визначення показника споживчого поведінкового наміру клієнта, що є взваженим складеним результатом трикомпонентної інформаційної моделі з показників конативності, афектності та компетентності.

Вихідними даними вищеописаного методу є показник споживчого поведінкового наміру клієнта, показники конативності, афектності та компетентності та візуальна аналітика споживчого поведінкового наміру клієнта.

Отже, було наведено узагальнену схему методу визначення поведінкових намірів клієнтів, який дає можливість визначити поведінковий намірів клієнтів на основі перетворення вхідних даних у вигляді навченої нейромережевої моделі для визначення прояву афектності за дописом, навченої моделі класифікатора для визначення прояву конативності, вхідного клієнтського допису для аналізу у вихідну інформацію у вигляді показника споживчого поведінкового наміру клієнта.

2.3 Формування моделі конативності методом SVC та визначення показника конативності

Схема формування моделі конативності методом SVC та визначення показника конативності на базі сформованої навченої моделі наведена на рисунку 2.3.

Вхідними даними підходу до формування моделі конативності методом SVC є навчальна вибірка текстів для тренування класифікатора та відгук для тестового визначення показника конативності.

Першим етапом у процесі формування моделі конативності здійснюється підбір параметрів класифікатора для подальшого навчання. Ядро SVC буде використано лінійне, оскільки класифікатор буде працювати з текстами, а згідно до проведених вище досліджень, автором [22] було показано, що найкращу

точність у класифікації текстових даних було отримано саме SVC з лінійним ядром. Результати з якості класифікації наведено на рисунку 2.4.

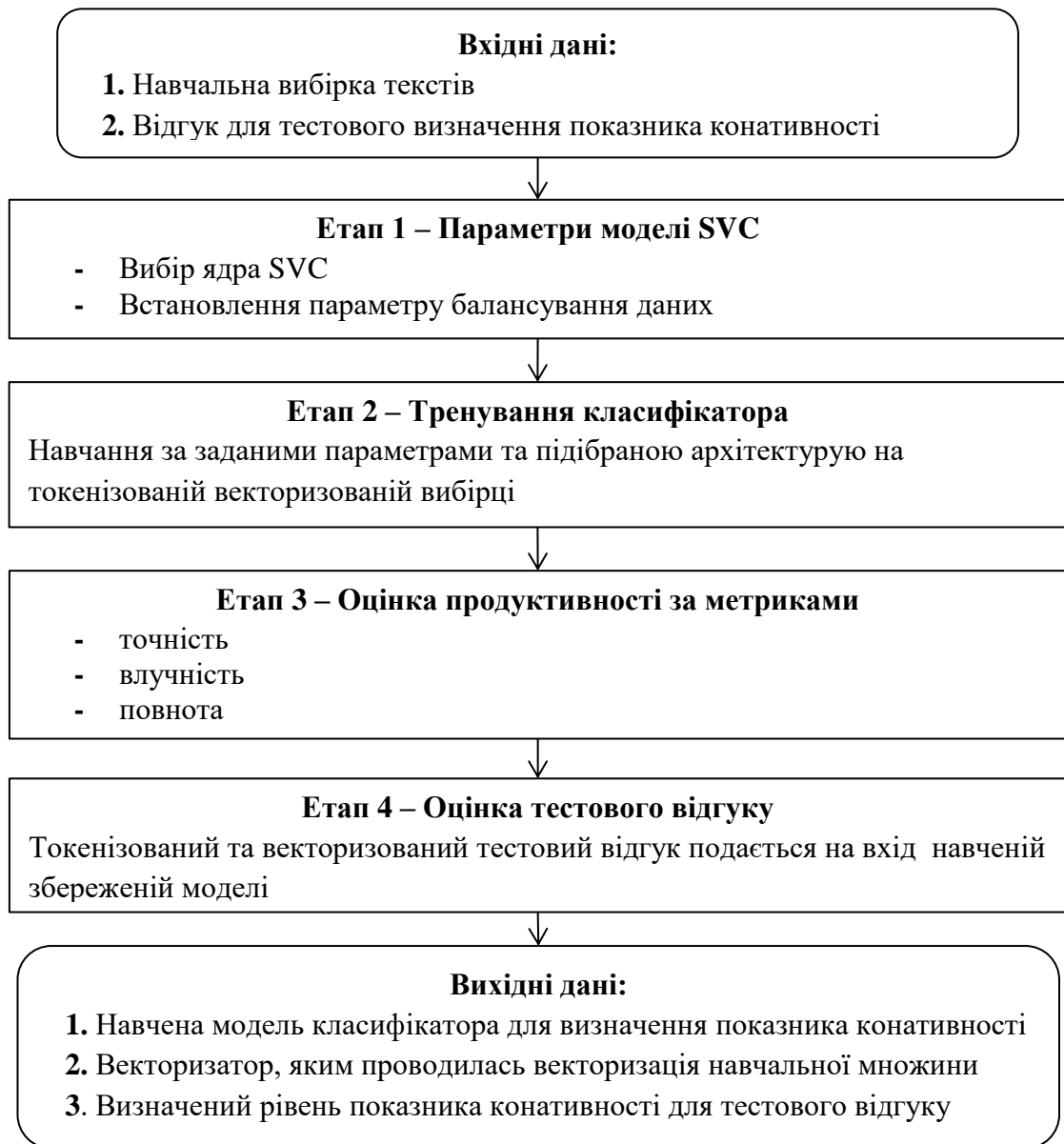


Рисунок 2.3 – Формування моделі конативності методом SVC

Оскільки розміри вибірки з відгуками можуть бути нерівномірні за класами, є потреба в установці параметру балансування даних.

Другим етапом є тренування класифікатора. Згідно до сформованої архітектуру на токенізованій векторизованій вибірці навчальних даних здійснюється процес навчання.

Після процесу навчання наступним етапом проходить оцінка продуктивності навченого класифікатора за метриками точності, влучності та повноти. Після чого навчена модель зберігається, а також зберігається векторизатор TF-IDF, який використовувався для попередньої обробки навчальної множини.

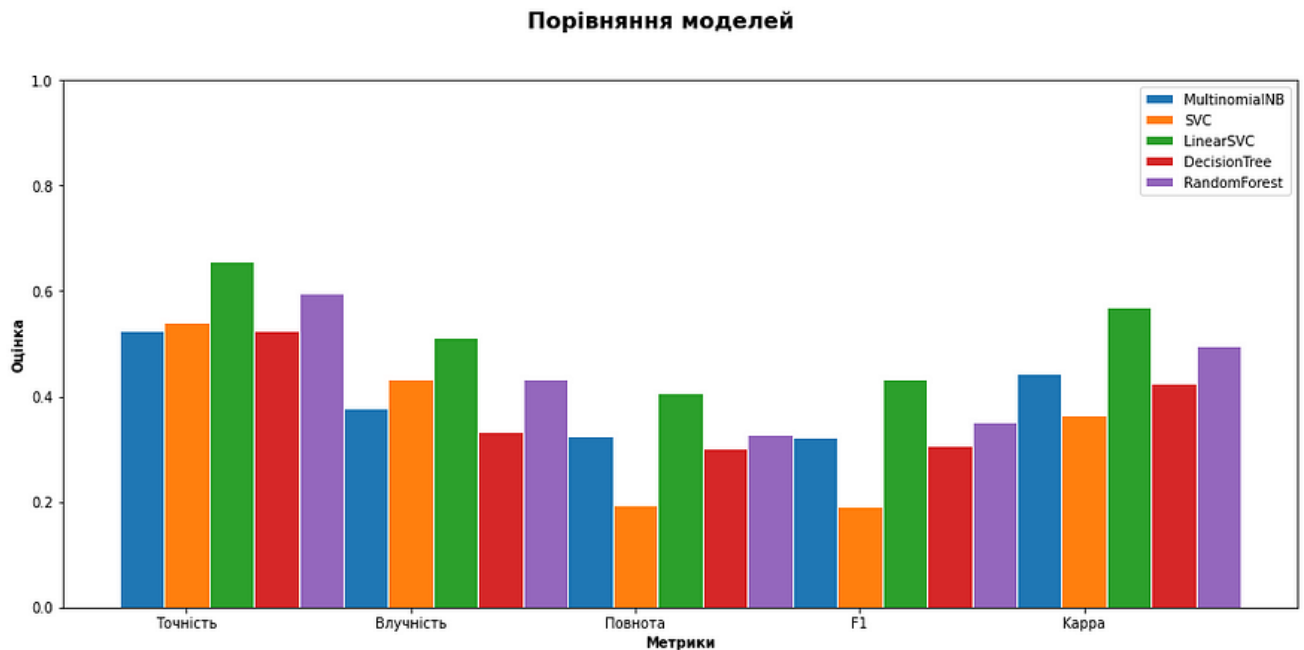


Рисунок 2.4 – Використання різних видів класифікаторів для задачі класифікації текстів [22]

Четвертим етапом проводиться оцінка тестового відгуку. Для цього токенований та векторизований тестовий відгук подається на вхід навченій збереженій моделі, яка в свою чергу дасть відповідь у вигляді числової оцінки рівня поведінкової дієвої рішучості користувача.

Вихідними даними є навчена модель класифікатора для визначення показника конативності, векторизатор, яким проводилась векторизація навчальної множини TF-IDF, а також визначений рівень показника конативності для тестового відгуку.

Таким чином було описано етап формування моделі конативності методом SVC та визначення показника конативності на основі навченої моделі.

2.4 Навчання нейронної мережі та формування моделі афектності, визначення показника афектності нейромережею LSTM

Схему формування моделі афектності на основі нейронної мережі LSTM наведено на рисунку 2.5. Вхідними даними у процесі формування моделі ефектності є навчальна вибірка текстів.

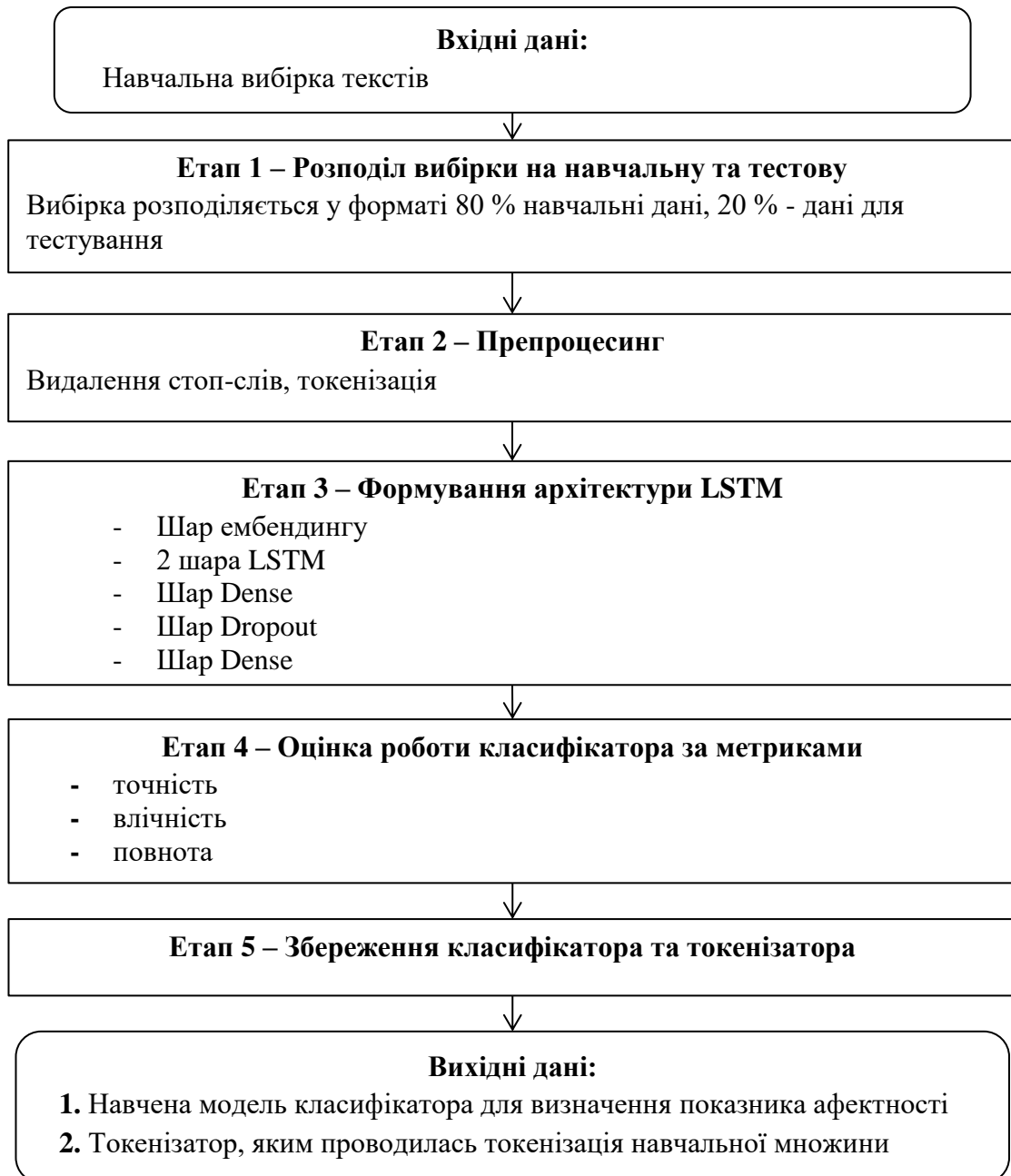


Рисунок 2.5 – Схема формування моделі афектності на основі нейронної мережі LSTM

Першим етапом здійснюється аналіз отриманих навчальних даних та розподіл їх у відношенні 80 % навчальні дані, 20 % - дані для тестування. Виконується балансування класів, у випадку, якщо навчальна вибірка має більше даних у рамках одного з класів. Другим етапом є виконання репроцесингу. Видалення стоп-слів та стоп-символів, а також токенізація.

Наступним етапом є формування архітектури нейромережі. Згідно до запропонованого підходу, першим шаром буде шар ембендингу, з кількістю унікальних слів у корпусі тексту 1500. При цьому довжина вхідної послідовності буде взято у розмірі 200 слів, так близько 80% навчальних текстів мають довжину до 150 слів, а понад 200 слів є лише 1% навчальної вибірки.

Наступним шаром моделі є шар LSTM, який має 64 нейрона з імовірністю випадкового відключення нейрона під час тренування для уникнення перенавчання 0,2. Виходом шару буде повний рядок вихідних послідовностей, який стане вхідним для наступного дублюючого LSTM шару.

Наступним шаром є Dense з кількістю нейронів 64, та функцією активації ReLU, графік якої наведено на рисунку 2.6. Вибір даної функції активації обумовлено тим, що її похідна дорівнює або одиниці, або нулю, отже не може відбутись розростання або загасання градієнтів. Також використання даної функції приводить до проріджування ваг.

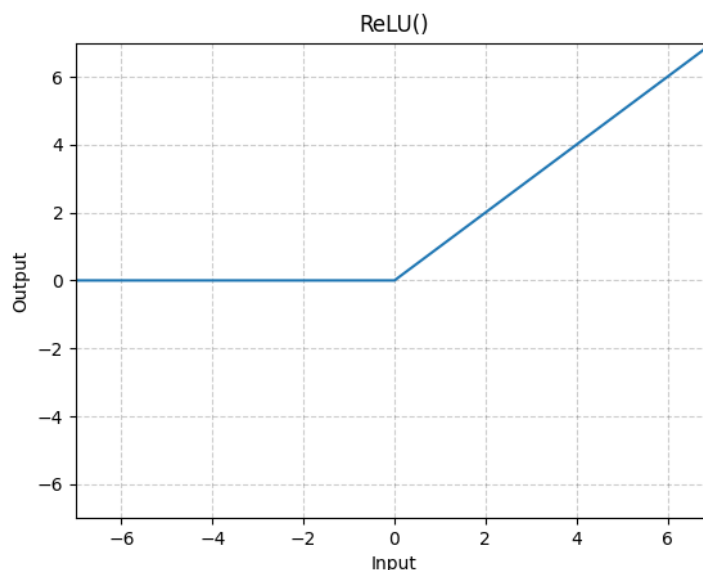
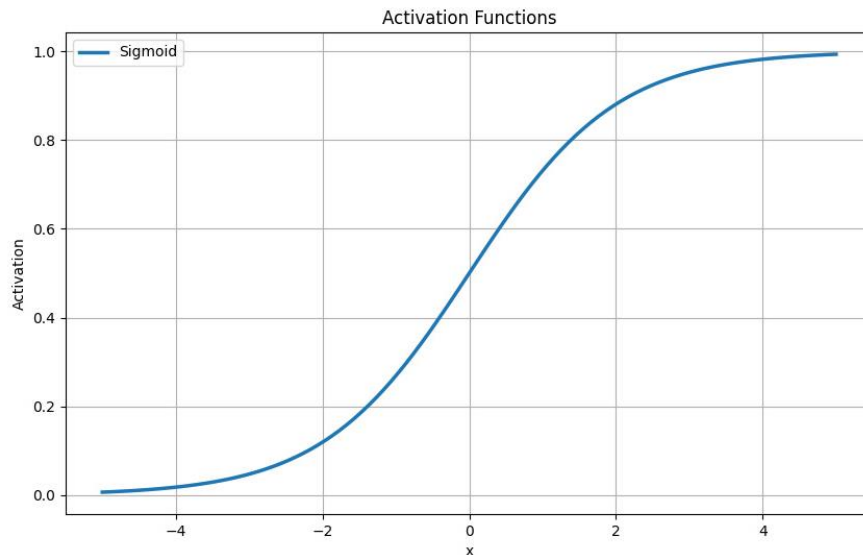


Рисунок 2.6 – Графік функції активації Dense шару [23]

Наступним шаром є шар Dropout. Для даного шару буде взято параметр імовірності випадкового відключення нейрона у розмірі 50%.

Останній шар для запропонованої архітектури є шар Dense, він буде з одним нейроном та сигмоїдною функцією активації, графік якої наведено на рисунку 2.7.



Рисунк 2.7 – Графік функції активації вихідного Dense шару [24]

Вибір сигмоїдної функції активації обумовлено тим, що рівень афектності є видом бінарної класифікації, що вказує на рівень схвального і відсутність відторгаючого емоційного контексту (або навпаки).

Після процесу навчання наступним етапом є етап оцінки продуктивності навченої нейронної мережі LSTM за метриками точності, влучності та повноти.

На п'ятому етапі після процесу оцінки за метриками навчена модель зберігається, а також зберігається токенизатор, який використовувався для попередньої обробки навчальної множини.

Відповідно, вихідними даними є навчена модель нейронної мережі архітектури LSTM для визначення показника ефектності, та токенизатор, яким проводилась токенизація навчальної множини.

Схема визначення показника афектності нейромережею LSTM для користувацького допису наведена на рисунку 2.8.

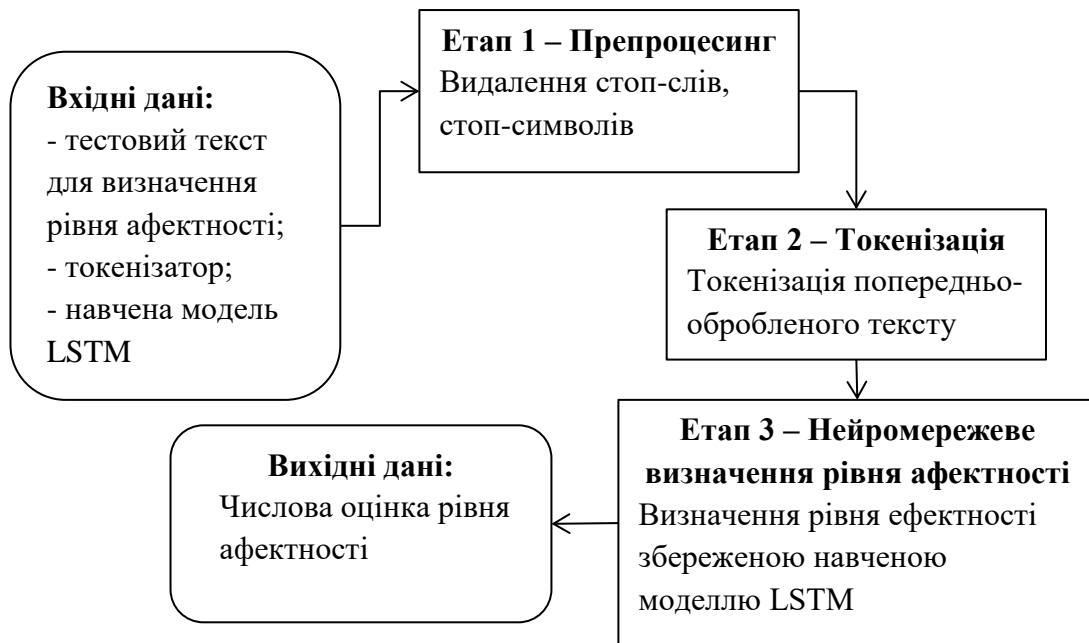


Рисунок 2.8 – Схема визначення показника афектності нейромережею LSTM

Вхідними даними є тестовий текст для визначення рівня афектності, токенизатор, а також навчена модель LSTM. Важливим аспектом є те, що токенизатор повинен бути там самим, на якому відбувалось навчання моделі нейромережі.

Першим етапом є попередня обробка тестового тексту. До неї належить видалення стоп-слів та стоп-символів. Наступним етапом є токенизація попередньо-обробленого тексту токенизатором з вхідних даних.

Третім етапом є нейромережеве визначення показника афектності нейромережею LSTM, що буде наведено у числовому вигляді оцінки від 0 до 1, де 1 – наявність схвального і відсутність відторгаючого емоційного контексту, а 0 – навпаки, практично повна відсутність схвального і наявність відторгаючого емоційного контексту, яке у свою чергу буде вихідними даними.

Отже, таким чином наведено процес навчання нейронної мережі та формування моделі ефектності, а також проілюстровано процес визначення показника афектності нейромережею LSTM, що свідчить про рівень схвального або відторгаючого емоційного контексту.

2.5 Підхід до визначення показника компетентності

Схему визначення показника компетентності з використанням методу TF наведено на рисунку 2.9. Вхідними даними у процесі визначення показника компетентності є текст з описом товару та клієнтський відгук на товар.

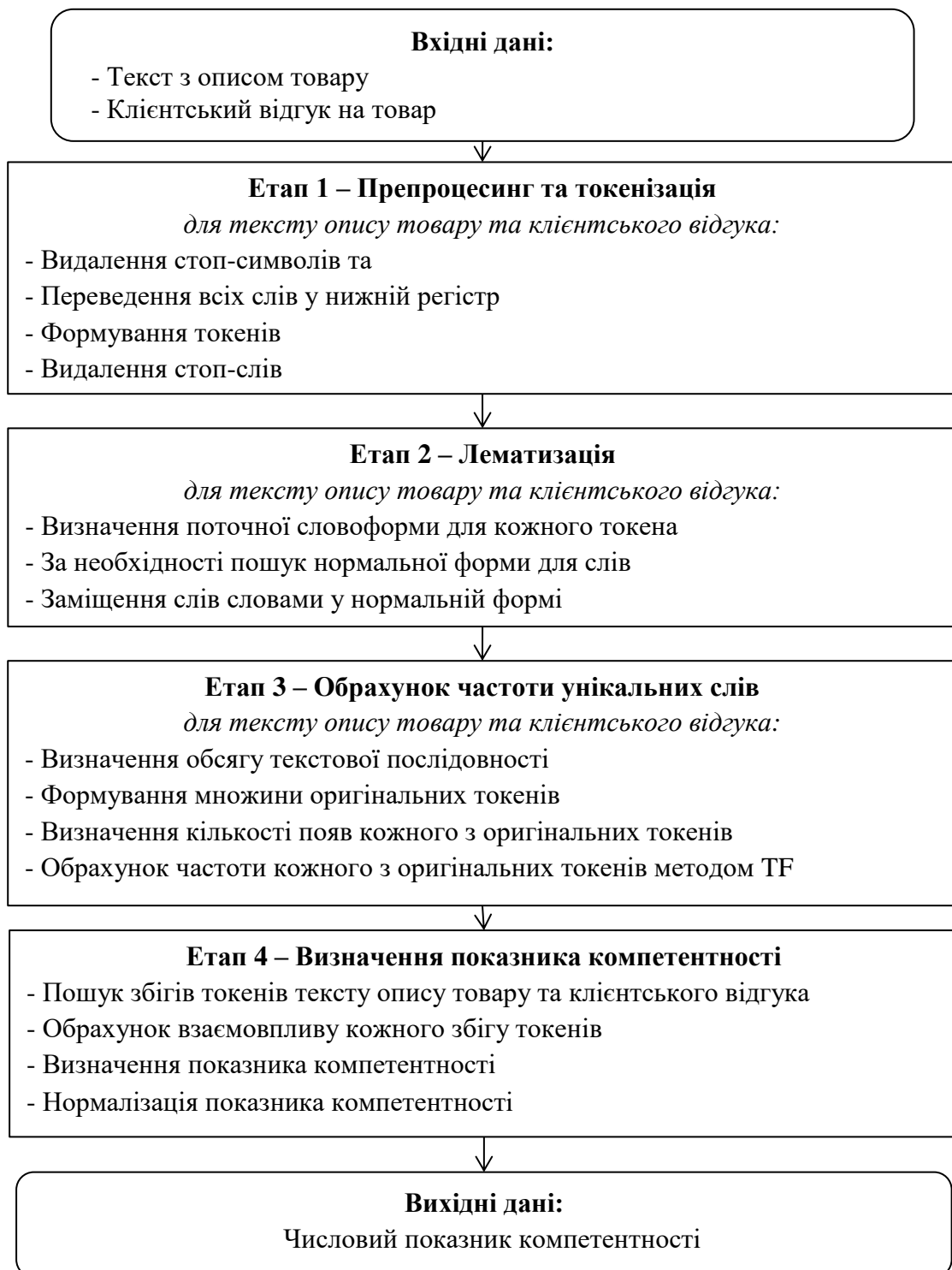


Рисунок 2.9 – Схема визначення показника компетентності

Перший етап призначений для виконання препроцесингу та токенизації як для текстового представлення опису товару, так і для текстового представлення клієнтського відгука. Включає в себе такі складові, як видалення стоп-символів, переведення всіх слів у нижній регістр, формування множини токенів та видалення стоп-слів. Результатом виконання першого етапу є токенизовані очищені дані.

Наступним етапом є етап лематизації. Лематизація – процес групування відмінюваних форм слова таким чином, щоб надалі їх можна було аналізувати як єдиний елемент, що ідентифікується за лемою слова або словниковою формою [25]. Приклад лематизації користувачького відгуку «Найкращі серветки на українському ринку» – «найкращий серветка український ринок». Лематизація застосовується аналогічно попередньому етапу, як до опису товару так і до клієнтського відгука. Ключовими діями цього етапу є визначення поточної словоформи для кожного точена, згідно до якої здійснюється за потреби пошук нормальної форми слова. Якщо слово перебувало не в нормальній формі, здійснюється заміщення слів словами у нормальній формі.

На етапі 3 здійснюється обрахунок частоти унікальних слів для тексту опису товару та клієнтського відгука. Для цього відбувається визначення обсягу текстової послідовності для тексту опису товару та клієнтського відгука, і з визначених текстових послідовностей формується множина оригінальних токенів. Обраховується кількість появ кожного з оригінальних токенів у множинах опису товару та клієнтського відгука та здійснюється обрахунок частоти кожного з оригінальних токенів методом TF.

Наступним етапом є етап визначення показника компетентності, що включає в себе пошук збігів токенів тексту опису товару та клієнтського відгука, згідно до якого обраховується взаємовплив кожного збігу токенів, який визначається як частота появи токена у описі товару помножена на кількість зустрічань у множині токенів тексту відгуку.

Визначений таким чином показник компетентності нормалізується з використанням обмеження. Якщо отримане значення перевищує 1, то показник вважається рівним 1.

Вихідними даними буде відповідно обрахований числовий показник компетентності клієнта.

Отже, таким чином відбувається визначення показника компетентності за відгуком клієнта, що є складовою частиною методу визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів.

2.6 Визначення показника споживчого поведінкового наміру клієнта

Показник Визначення показника споживчого поведінкового наміру клієнта. Схема логічного зв'язку етапів визначення споживчого поведінкового наміру клієнта наведена на рисунку 2.10.

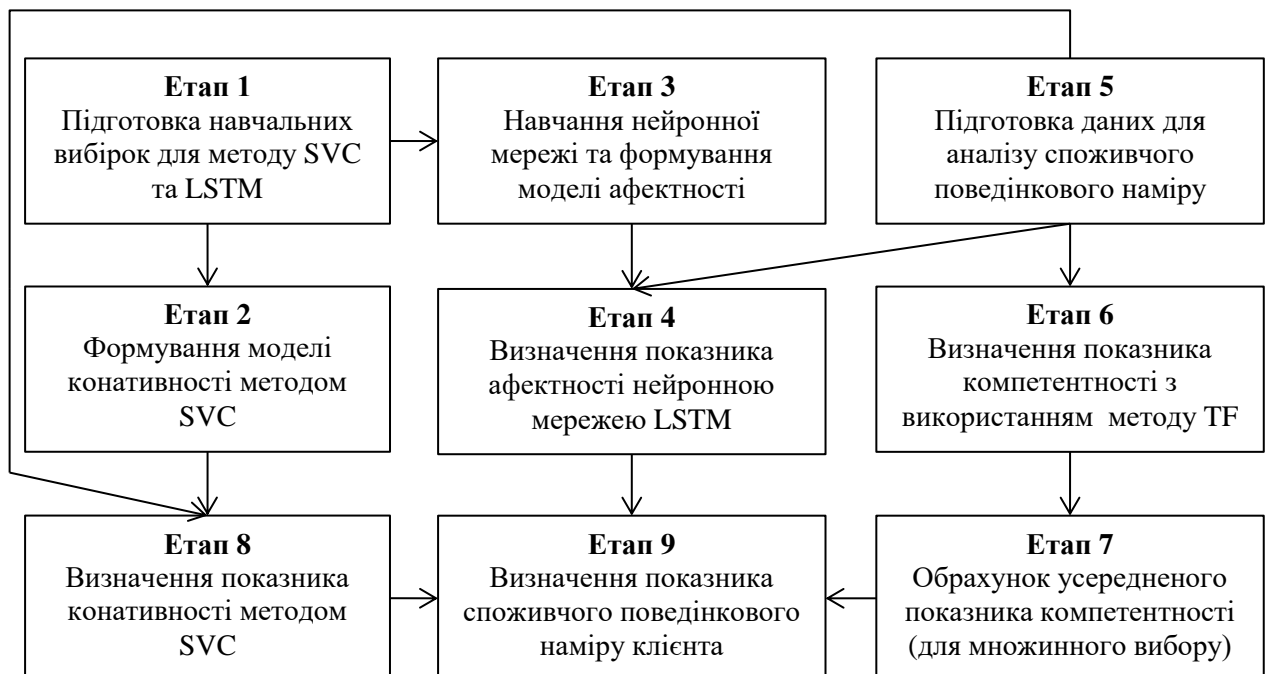


Рисунок 2.10 – Логічний зв'язок етапів визначення споживчого поведінкового наміру клієнта

Етап 1 слугує для підготовки навчальних вибірок для методу SVC та LSTM. На даному етапі відбувається препроцесинг (попередня обробка даних) та токенизація. Для SVC моделі дані проходять етап перетворення у векторне представлення з використанням підходу TF-IDF. Важливою деталлю даного етапу є те, що потрібно зберегти вектори затор та токенизатор, оскільки вони в подальшому будуть потрібні для обробки тестових даних. Перший етап тісно взаємопов'язаний з другим та третім етапами, оскільки готує для них вхідні дані.

Етап формування моделі конативності методом SVC в свою чергу є другим етапом, що тісно пов'язаний з етапом 8, на якому здійснюється визначення показника конативності методом SVC на базі сформованої на даному етапі моделі.

Етап 3, що відповідає за навчання нейронної мережі та формування моделі афектності, тісно пов'язаний з етапом 4, що відповідає за визначення показника афектності нейронною мережею LSTM, тим самим слугує для нього вхідними даними. Також вхідними даними для 4-го етапу є етап 5, що відповідає за підготовку даних для аналізу споживчого поведінкового наміру.

В свою чергу, етап 5 взаємопов'язаний з етапом 6, де відбувається визначення показника компетентності з використанням методу TF, та етапом 8, який відповідає за визначення показника конативності методом SVC, а саме є для цих етапів вхідними даними.

Етап 6 пов'язаний з етапом 7, де відбувається обрахунок усередненого показника компетентності. Відповідно 7 етап на ряду з 4-м та 8-м формують 9 етап, де відбувається на базі отриманих на попередніх етапах даних визначення показника споживчого поведінкового наміру клієнта.

Отриманий на етапі 9 показник споживчого поведінкового наміру клієнта можна інтерпретувати за таблицею 2.1.

Таблиця 2.1 – Шкала інтерпретації прояву споживчого поведінкового наміру клієнта

Числова оцінка прояву наміру	Інтерпретація
Від 0 до 0.2	Низький прояв купівельного поведінкового наміру клієнта
Від 0.21 до 0.4	Прояв купівельного поведінкового наміру клієнта нижче середнього
Від 0.41 до 0.6	Середній прояв купівельного поведінкового наміру
Від 0.6 до 0.8	Прояв купівельного поведінкового наміру клієнта вище середнього
Від 0.81 до 1	Високий прояв купівельного поведінкового наміру клієнта

Отже, наведеним чином відбувається логічний зв'язок етапів визначення споживчого поведінкового наміру клієнта та визначається інтерпретація отриманої оцінки отриманого прояву споживчого поведінкового наміру клієнта.

2.7 Формування множини навчальних даних

Для роботи методу визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів необхідно мати навчальні набори даних для нейромережевого визначення прояву афектності користувача та для автоматизованого визначення прояву конативності, що є складовими частинами визначення поведінкових намірів клієнта.

2.7.1 Набір даних для визначення прояву афектності

Навчальний датасет для навчання нейронної мережі LSTM використано з ресурсу IMDb з перекладом на українську.

Дані містять 40 000 відгуків про фільми, зібраних з сайту IMDb. Кожен відгук містить наступну інформацію [26]:

- ID відгуку: Унікальний ідентифікатор відгуку.
- Тональність відгуку: Має 2 категорії, схвалення та відторгнення.
- Текст відгуку: Текст відгуку.

Це набір даних для двійкової класифікації схвалення, який містить понад 250 000 відгуків, що розбиті у відношенні 53% схвалюючі та 47 % відторгаючі. Вигляд датасету наведено на рисунку 2.11.

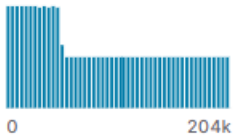
movie_reviews.csv (316.31 MB)				
Detail Compact Column				
#		review	sentiment	
		253039 unique values	positive 53%	negative 47%
0		Hakuna Matata. What a wonderful phrase. Hakuna Matata. Ain't no passing phrase. The Lion King (2019)...	positive	
1		Obviously a second attempt at the Hugh Glass story as portrayed by Richard Harris in Man in the Wild...	positive	
2		If some studio head, or person who was involved with this film ever reads this, I must apologize for ...	negative	
3		Rocknrolla rocked me big time. I fall for everything that has to do with mafias and organized crime....	positive	
4		Finally, harry potter has a better movie than the book	positive	

Рисунок 2.11 – Вигляд датасету

Отже, використання даного датасету дозволить виконати навчання нейромережі LSTM для подальшого виявлення прояву афектності.

2.7.2 Набір даних для визначення прояву конативності

Для визначення прояву компоненту конативності було власноруч створено набір даних, який має наступну структуру:

```
{
  "categories": [
    {
      "category_name"
      "products": [
        {
          "name"
          "description"
          "reviews": [
            {
              "author"
              "comment"
              "mark"
            },
            {
              "author"
              "comment"
              "mark "
            }
          ]
        }
      ]
    }
  ]
  ...
}
```

Дані для датасету були взяті із товарів повсякденного вжитку інтернет-магазину «Розетка» [4] засобом вебскрапінгу з використанням бібліотеки «crawlee» [29].

Набір даних налічує 5 категорій, та загальну кількість товарів 483. Кількість відгуків рівна 4947. Кількість відгуків, що говорять про високий показник конативності клієнта та свідчить його про дієву рішучість щодо наміру наступної повторної покупки складає 3958. Кількість відгуків, що свідчить про відсутність дієвої рішучості автора щодо наміру здійснити повторну покупку складає 742 і говорить про низький показник конативності. Кількість нейтральних відгуків, що не можуть свідчити однозначно про дієву рішучість автора щодо наміру наступної повторної покупки – 247.

Статистичний розподіл маркованих клієнтських відгуків для категорії «Засоби для миття посуду» наведено на рисунку 2.12

Статистика розподілу користувацьких оцінок у категорії "Засоби для миття посуду"

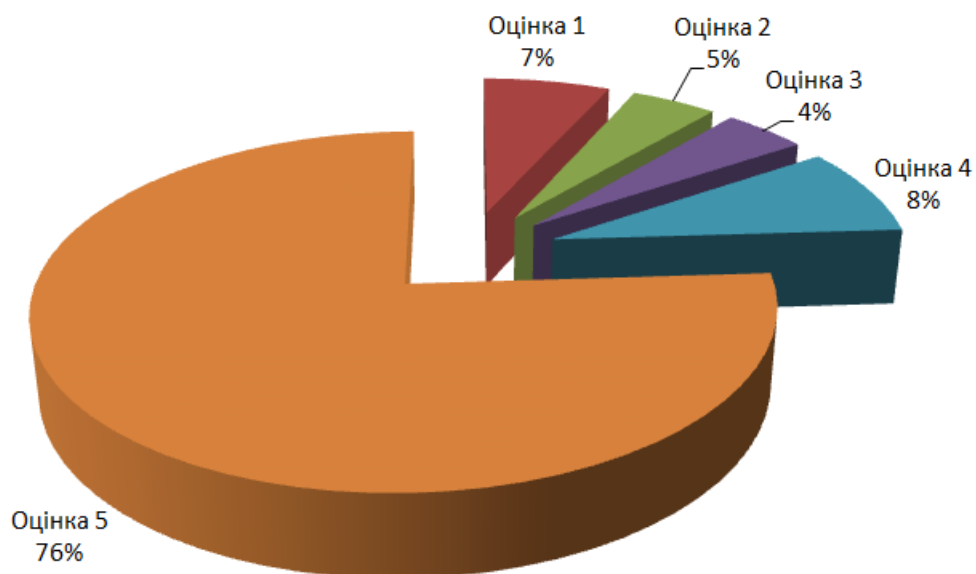


Рисунок 2.12 – Статистика розподілу оцінок для категорії «Засоби для миття посуду»

Статистика розподілу кількості товарів за категоріями наведена на рисунку 2.13



Рисунок 2.13 – Розподіл товарів за кількістю по категоріям

Статистика розподілу кількості відгуків за категоріями наведена на рисунку 2.4



Рисунок 2.14 – Розподіл відгуків за кількістю по категоріям



Рисунок 2.15 – Розподіл відгуків за довжиною з високим показником конативності

Розподіл відгуків, що говорять про високий показник конативності клієнта та свідчить його про дієву рішучість щодо наміру наступної повторної покупки наведено на рисунку 2.15. Розподіл відгуків, що свідчить про відсутність дієвої рішучості автора щодо наміру здійснити повторну покупку за довжиною текстового представлення наведено на рисунку 2.16.



Рисунок 2.16 – Розподіл відгуків за довжиною з низьким рівнем конативності

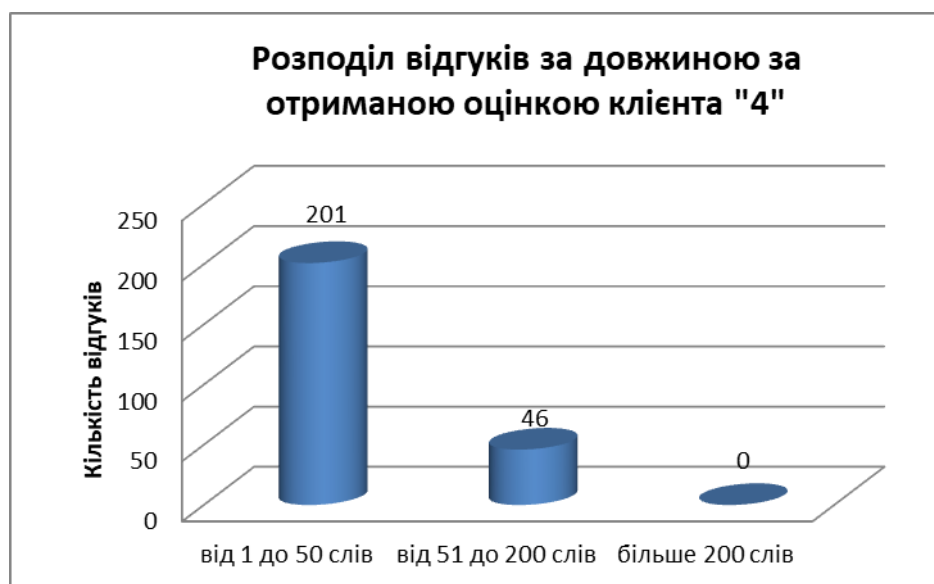


Рисунок 2.17 – Розподіл відгуків за довжиною з неоднозначним рівнем конативності

Розподіл неоднозначних відгуків, що не можуть свідчити однозначно про дієву рішучість автора щодо наміру наступної повторної покупки наведено на рисунку 2.17.

Отже, таким чином було виконано формування навчальних вибірок для визначення показників конативності та афектності, що будуть застосовані для навчання класифікаторів LSTM та SVC.

Висновки до розділу 2

У розділі розглянуто показники споживчого поведінкового наміру клієнта, якими є прояви конативності, ефективності та когнітивності, що в сукупності визначають міру бажання клієнта щодо придбання товару. За вищенаведеними показниками була сформована зважена трикомпонентна інформаційна модель визначення споживчого поведінкового наміру.

На основі запропонованої трикомпонентної інформаційної моделі було сформовано узагальнену схему методу визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів та детальний розгляд складових методу. Метод автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів дає можливість визначити поведінковий намірів клієнта на основі перетворення вхідних даних у вигляді навченої нейромережевої моделі для визначення прояву афектності за дописом, навченої моделі класифікатора для визначення прояву конативності, вхідного клієнтського допису для аналізу у вихідну інформацію у вигляді показника споживчого поведінкового наміру клієнта та показників його складових: конативності, афектності та компетентності.

Наведено детальний опис основних етапів роботи методу, таких як етап формування моделі конативності методом SVC та визначення показника конативності на основі навченої моделі; наведено процес навчання нейронної мережі та формування моделі ефектності, а також проілюстровано процес визначення показника афектності нейромережею LSTM, що свідчить про рівень

схвального або відторгаючого емоційного контексту; описано процес визначення показника компетентності за відгуком клієнта; описано логічний зв'язок етапів визначення споживчого поведінкового наміру клієнта та визначається інтерпретація отриманої оцінки отриманого прояву.

Обрано набір даних, що налічують 40 000 відгуків про фільми, зібраних з сайту IMDb, що дозволить виконати навчання нейромережі LSTM для подальшого виявлення прояву афектності з автоматизованим перекладом на українську мову.

Сформовано набір даних для визначення прояву конативності, що охоплює 5 категорій та загальну кількість товарів 483. Відповідно, кількість відгуків рівна 4947. Кількість відгуків, що говорять про високий показник конативності клієнта та свідчить його про дієву рішучість щодо наміру наступної повторної покупки складає 3958. Кількість відгуків, що свідчить про відсутність дієвої рішучості автора щодо наміру здійснити повторну покупку складає 742 і говорить про низький показник конативності. Кількість нейтральних відгуків, що не можуть свідчити однозначно про дієву рішучість автора щодо наміру наступної повторної покупки рівна 247.

Розділ 3 Інформаційна система автоматизованого визначення поведінкових намірів клієнтів

3.1 Схеми інформаційної системи

Інформаційна система автоматизованого визначення поведінкових намірів клієнтів є прикладною програмною реалізацією метода визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на основі перетворення вхідних даних у формі навченої нейромережевої моделі LSTM для визначення прояву афектності за дописом, навченої моделі класифікатора SVC для визначення прояву конативності, вхідного клієнтського допису для аналізу у вихідні дані у вигляді показника споживчого поведінкового наміру клієнта та показників його складників: конативності, афектності та компетентності.

Пропонована інформаційна система автоматизованого визначення поведінкових намірів клієнтів складається із 5 підсистем: «Підсистема визначення показника конативності», «Підсистема визначення показника афектності», «Підсистема визначення показника компетентності», «Підсистема визначення показника споживчого поведінкового наміру», «Підсистема візуальної аналітики» і набору даних. Схему інформаційна система автоматизованого визначення поведінкових намірів клієнтів наведено на рисунку 3.1.

Підсистема визначення показника конативності призначена для визначення поведінкової дієвої рішучості клієнта за дописом. Показник конативності визначається за допомогою використання навченого класифікатора SVC.

Підсистема визначення показника афектності призначена для оцінки емоційного схвалення клієнтом об'єкту за дописом. Показник афектності визначається за допомогою нейронної мережі LSTM.

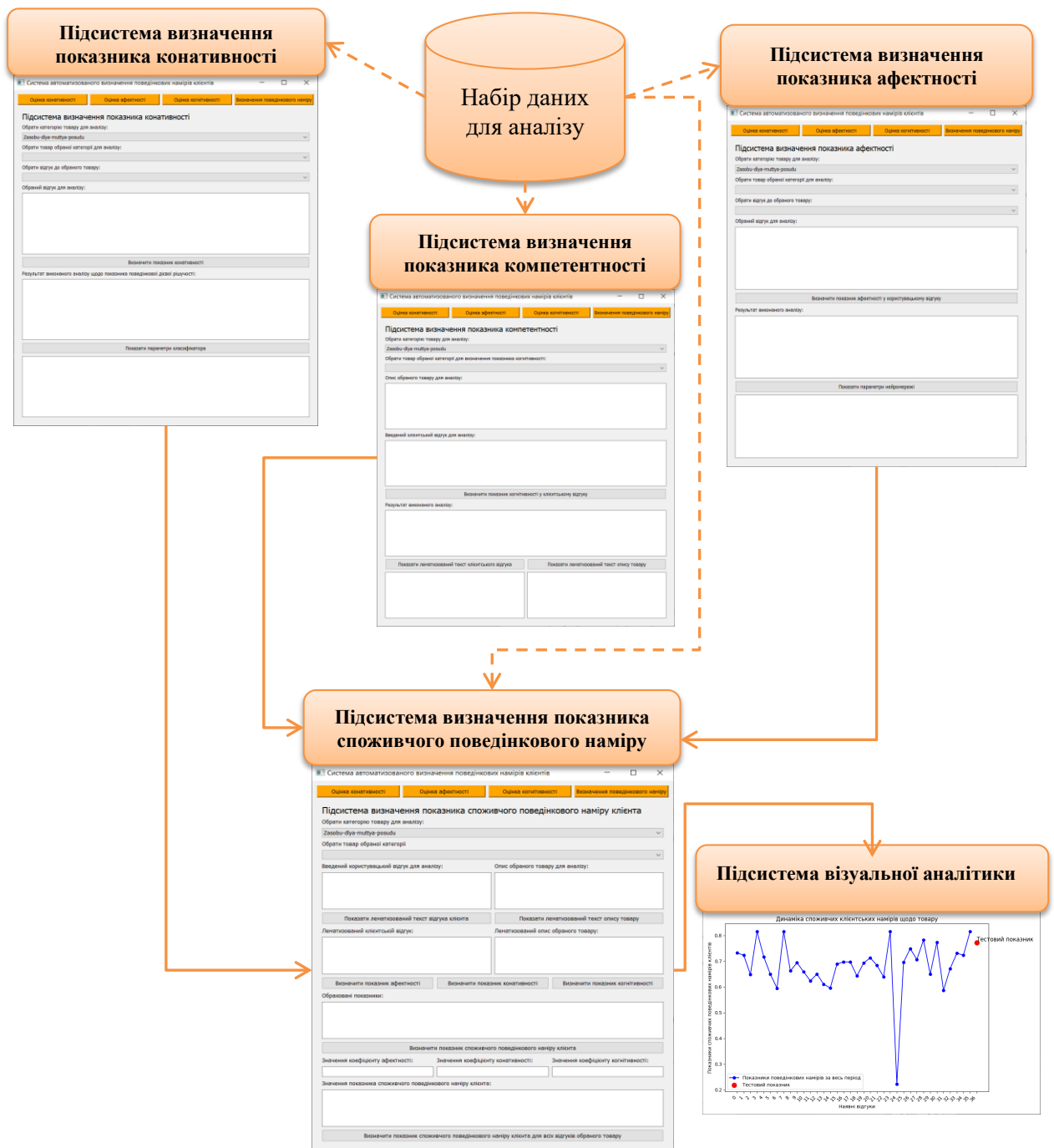


Рисунок 3.1 – Схема інформаційна система автоматизованого визначення поведінкових намірів клієнтів

Підсистема визначення показника компетентності дозволяє визначити прояв розуміння функціональності об'єкту за дописом клієнта та описом об'єкту. Показник компетентності визначається з використанням методу TF.

Підсистема визначення показника споживчого поведінкового наміру є головною підсистемою інформаційної системи, та дозволяє оцінити, а також інтерпретувати обчислений зважений показник споживчого поведінкового

наміру на основі показників конативності, афектності та компетентності. Також дана підсистема взаємодіє із підсистемою візуальної аналітики, та дає можливість візуалізувати вже наявні показники споживчих намірів клієнтів за дописами та показати візуально рівень споживчого наміру тестового допису щодо обраного об'єкту.

Таким чином, було створено архітектуру інформаційної системи автоматизованого визначення поведінкових намірів клієнтів, що складається із головної підсистеми визначення показника споживчого поведінкового наміру, та підсистем визначення показника афектності, визначення показника конативності та визначення показника компетентності, а також множини впорядкованих даних для аналізу. Розроблена інформаційна система є прикладною програмною реалізацією відповідного методу визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів.

3.2 Формування рекомендованої комбінації засобів розробки інформаційної системи

Засоби розробки інформаційної системи визначення поведінкових купівельних намірів щодо товарів повсякденного вжитку поділені за призначенням на такі категорії:

1. Програмні засоби для формування датасету щодо визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів.
2. Програмні засоби для навчання класифікаторів та побудови інтерфейсу користувача.

Для формування датасету щодо визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів у якості редактору програмного коду буде використано Visual Studio Code, так як він легкий, потужний та безкоштовний редактор коду, розроблений компанією Microsoft. Він є однією з найпопулярніших інтегрованих середовищ розробки (IDE) серед програмістів та

розробників [27]. Зовнішній вигляд редактора програмного коду наведено на рисунку 3.2.

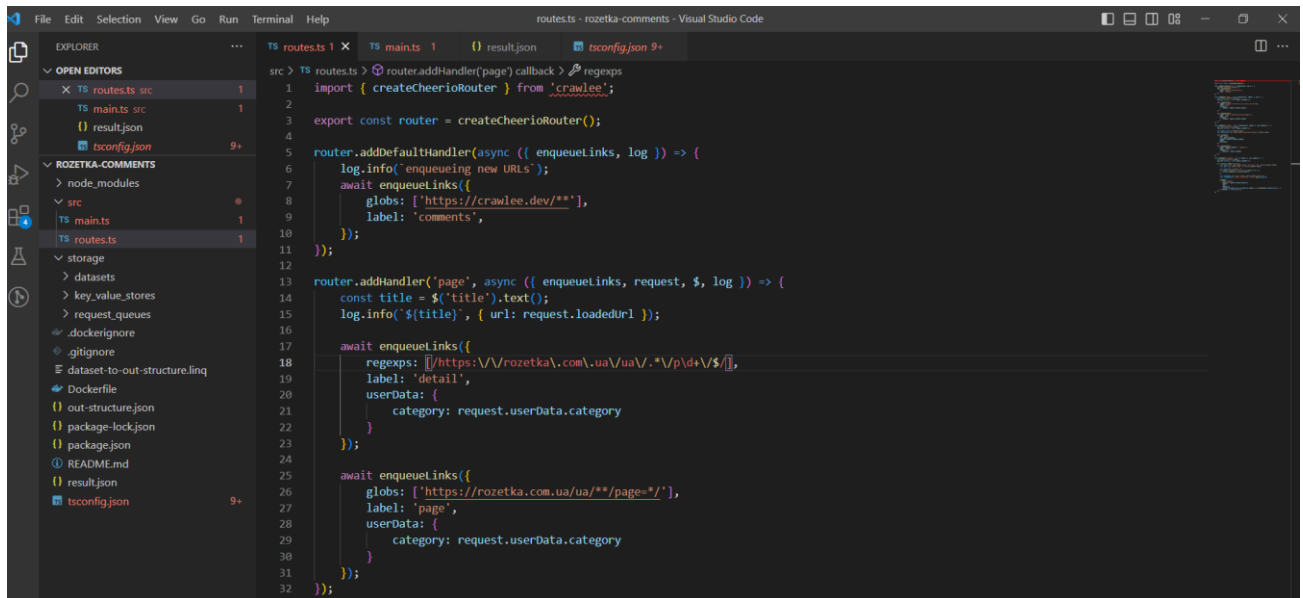


Рисунок 3.2 – Редактор коду Visual Studio Code

У якості мови програмування для вебскрапінгу буде використано JavaScript, оскільки вона є потужним інструментом для парсингу вебдокументів. Вона має ряд функцій, які роблять її придатною для цієї задачі, зокрема [28] вбудований доступ до DOM, що дозволяє отримувати доступ до елементів вебсторінки та їхніх властивостей. Ще однією перевагою є можливість використовувати регулярні вирази.

Також існує ряд бібліотек JavaScript, які були спеціально розроблені для парсингу вебдокументів, однією з яких є бібліотека Crawler, що є потужним інструментом для вебскрейпінгу, який використовує JavaScript для парсингу вебдокументів [29].

У якості програмних засобів для навчання класифікаторів (для визначення прояву конативності та для визначення прояву афектності за дописом клієнта) буде використана мова програмування Python та редактор програмного коду PyCharm.

Мова програмування Python показала високі результати для виконання таких завдань, так як Python досить простий у порівнянні з іншими мовами.

Більше того, Python має високі показники обробки даних. Ця мова програмування була створена ще в 1991 році Гвідо ван Россумом. Python має ефективні структури даних високого рівня, але поряд з цим простий, але ефективний підхід до ООП. Простий синтаксис, динамічне введення тексту, поряд з інтерпретованою природою, ставлять її на високу позицію з точки зору роботи зі сценаріями та швидкої розробки застосувань у багатьох областях на більшості платформах. Python працює на багатьох основних платформах (Windows, Linux, Mac, Raspberry Pi тощо) та має простий зрозумілий англійський синтаксис, що дозволяє девелоперам писати програми з меншою кількістю рядків коду, ніж у порівнянні це можна зробити іншими мовами програмування. Python можна вважати як процедурною, об'єктно-орієнтованою, так і функціональною мовою програмування.

Основною причиною, використання Python для машинного навчання, є те, що ця мова має багато фреймворків, що значно спрощують процес кодування та скорочують відповідно час розробки [30].

Працюючи з мовою Python, програмісту не потрібно приділяти велику увагу безпосередньому написанню коду: програміст може зосередити всю свою увагу на вирішенні більш складних проблем, пов'язаних з машинним навчанням. Водночас синтаксис Python завдяки своїй простоті допомагає програмістам перевірити складні алгоритми з мінімальним часом реалізації. Ще однією вагомою перевагою Python у машинному навчанні є його гнучкість: наприклад, розробник має вибір між ООП підходом та сценаріями. Python дозволяє поєднувати різні типи даних. Для того, щоб стати фахівцем з Machine learning, необхідно знати не тільки як систематизувати та аналізувати дані, розуміти як працюють базові алгоритми, освоїти реальні кейси, а й знати мову програмування Python.

Отже, таким чином для формування датасету щодо визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів у якості редактору програмного коду буде використано Visual Studio Code, у якості мови програмування JavaScript. У якості програмних засобів для навчання

класифікаторів SVC та LSTM, а також для побудови інтерфейсу користувача буде використана мова програмування Python та редактор програмного коду PyCharm.

3.3 Спеціалізовані програмні розширення для розробки інформаційної системи

Зважаючи на вибір комбінації засобів розробки, а саме на мову програмування Python, доцільно обрати спеціалізовані програмні розширення, що допоможуть спростити процес створення інформаційної системи.

Для навчання класифікаторів SVC та LSTM буде використана бібліотека Scikit-learn, яка є бібліотекою Python для машинного навчання, та містить інструменти для вирішення різноманітних завдань, таких як класифікація, регресія, кластеризація та зменшення розмірності [31]. Основні можливості по використанню бібліотеки наведені на офіційному сайті (рисунок 3.3)

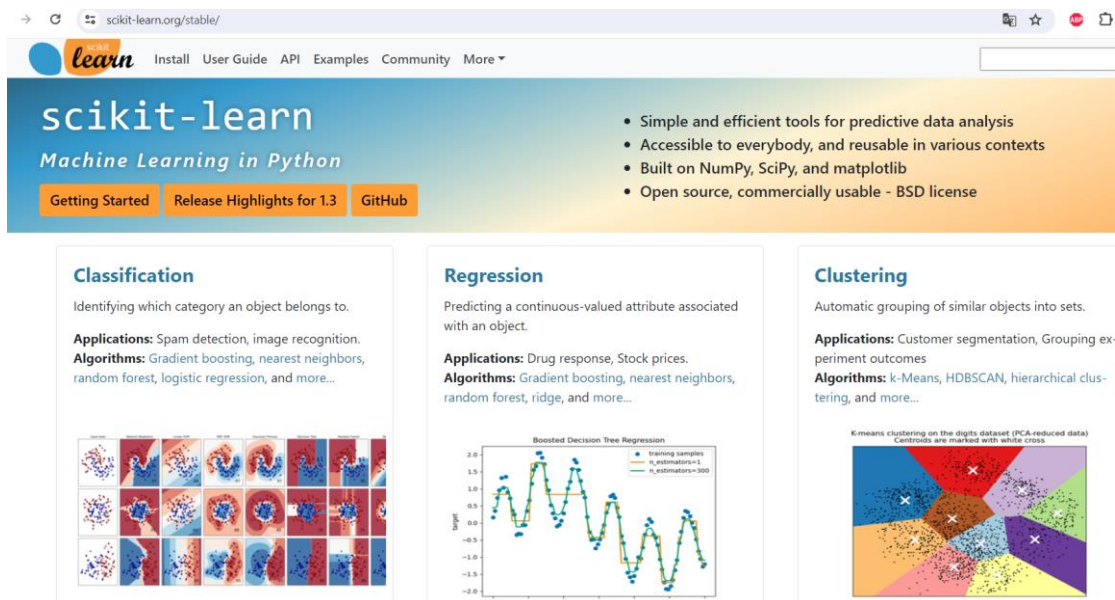


Рисунок 3.3 – Офіційний сайт Scikit-learn [32]

Поєднання Scikit-learn і нейронних мереж має ряд переваг. По-перше, Scikit-learn надає широкий вибір попередньо створених моделей і алгоритмів, які можна використовувати для швидкого створення прогнозних моделей. По-друге,

нейронні мережі здатні вивчати складні зв'язки між входами та виходами. По-третє, поєднання Scikit-learn і нейронних мереж дозволяє створювати більш потужні моделі.

Для попередньої обробки тексту та роботи з текстовими послідовностями буде використано Keras, що є API високого рівня для платформи TensorFlow. Загалом Keras надає доступний, високопродуктивний інтерфейс для вирішення проблем машинного навчання з акцентом на сучасне глибинне навчання. Keras охоплює кожен етап робочого процесу машинного навчання, від обробки даних до налаштування гіперпараметрів і розгортання. Він був розроблений з упором на можливість швидкого експериментування [33].

Для лематизації допису та опису об'єкту, що відноситься до допису для визначення показника компетентності буде використано бібліотеку Stanza, яка має мовні моделі для різних мов світу (понад 60 мов), зокрема серед них є підтримка української мови. Бібліотека Stanza заснована на PyTorch, що дозволяє їй використовувати потужні можливості машинного навчання для обробки природної мови.

Вона надає набори даних, інструменти та API для виконання різних завдань обробки природної мови, таких як токенізація, лемматизація, визначення частин мови, аналіз залежностей, вилучення іменованих сутностей та переклад. Основні елементи пайплайну Stanza наведені на рисунку 3.4.

Отже, за допомогою бібліотеки Stanza та наданого нею алгоритму лематизації, слова будуть лематизовані до кореневих форм слів.

Графічна візуалізація результатів буде виконана засобами matplotlib.pyplot що є набором функцій, завдяки яким matplotlib працює як MATLAB. Кожна функція pyplot вносить певні зміни у фігуру: наприклад, створює фігуру, створює область побудови на фігурі, малює деякі лінії в області побудови, прикрашає графік мітками тощо [35].

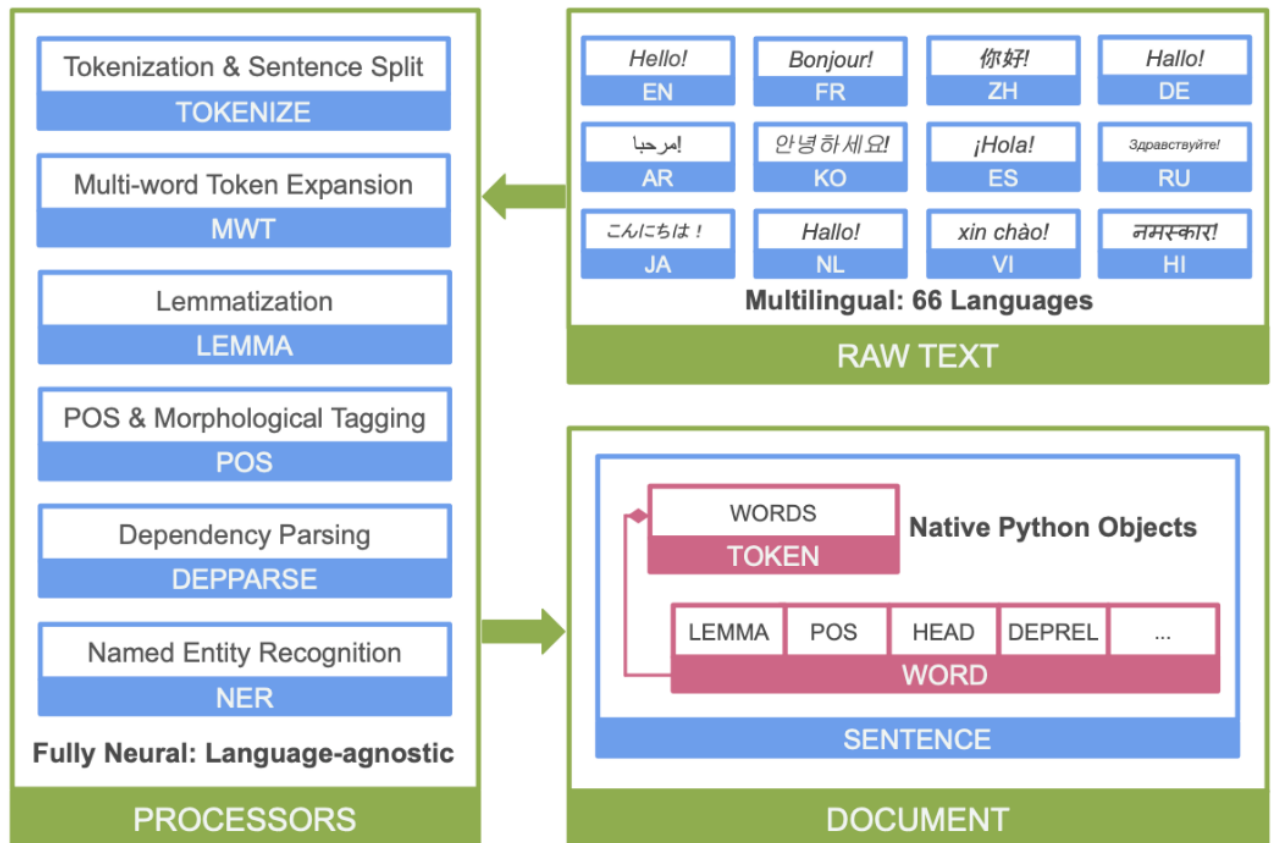


Рисунок 3.4 – Основні елементи пайплайну Stanza [34]

У `matplotlib.pyplot` зберігаються різні стани під час викликів функцій, тому він відстежує такі речі, як поточна фігура та область побудови, а функції побудови спрямовані на поточні осі (зверніть увагу, що «осі» тут і в більшості місць у документація відноситься до осей частини фігури, а не до суворого математичних термінів для кількох осей).

Отже, у якості спеціалізованих програмних розширень будуть використані: бібліотека `Scikit-learn` для навчання класифікаторів `SVC` та `LSTM`, `Keras` для попередньої обробки тексту та роботи з текстовими послідовностями, бібліотека `Stanza` для лематизації допису та опису об'єкту, що відноситься до допису для визначення показника компетентності, `matplotlib.pyplot` для графічної візуалізації результатів.

3.4 Метрики для оцінки роботи класифікаторів

Робота класифікаторів оцінюється спеціальними методами, які називаються метриками. Найпоширенішими метриками є: матриця сплутувань, точність, влучність, повнота [36].

Матриця сплутувань – це таблиця, яка відображає результати класифікації. Матриця сплутаності має чотири основні поля: істинно позитивний, хибно позитивний, істинно негативний, хибно негативний. Ілюстрація матриці сплутувань наведено на рисунку 3.5

		Прогнозований клас	
		Predicted Positive (PP)	Predicted Negative (PN)
Загальна кількість = P + N			
Справжній клас	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

Рисунок 3.5 – Матриця сплутувань [37]

Матриця сплутувань дозволяє визначити деталі класифікаційної продуктивності, зокрема помилки, що виникають при визначенні кожного класу.

Точність, влучність і повнота – три метрики, які можна використовувати для оцінки продуктивності моделі класифікації.

Точність (Accuracy) – це загальна кількість зразків, які були правильно класифіковані, поділена на загальну кількість зразків [38].

$$Accuracy = (TP + TN) / (TP + FP + FN + TN),$$

TP – істинно позитивні, TN – істинно негативні, FP – хибно позитивні, FN – хибно негативні ідентифіковані зразки.

Влучність (Precision) – це кількість зразків, які були правильно класифіковані як позитивні, поділена на загальну кількість зразків, які були класифіковані як позитивні [39].

$$Precision = TP / (TP + FP),$$

TP – істино позитивні, FP – хибно позитивні ідентифіковані зразки.

Повнота (Recall) – це кількість зразків, які були правильно класифіковані як позитивні, поділена на загальну кількість зразків, які є позитивними [40].

$$Recall = TP / (TP + FN),$$

TP – істино позитивні, FN – хибно негативні ідентифіковані зразки.

Точність є загальною мірою продуктивності, але вона може бути заплутаною, якщо в наборі даних є нерівномірний розподіл класів. Наприклад, якщо в наборі даних переважають позитивні зразки, то модель, яка просто класифікує всі зразки як позитивні, матиме високу точність, навіть якщо вона не дуже добре класифікує негативні зразки.

Влучність і повнота є більш специфічними мірами продуктивності, які враховують розподіл класів. Влучність є мірою того, наскільки добре модель класифікує позитивні зразки. Повнота є мірою того, наскільки добре модель виявляє позитивні зразки.

Отже, для оцінки роботи класифікатора буде використано вищеописаний набір метрик, а саме: матрицю сплутувань, точність, влучність і повноту.

Висновки до розділу 3

У третьому розділі сформовано схему інформаційної системи автоматизованого визначення поведінкових намірів клієнтів, що налічує 5 підсистем та відповідний набір даних. А саме: : «Підсистема визначення показника конативності», «Підсистема визначення показника афектності», «Підсистема визначення показника компетентності», «Підсистема визначення показника споживчого поведінкового наміру», «Підсистема візуальної аналітики». Головною підсистемою є підсистема визначення показника

споживчого поведінкового наміру, яка дозволяє оцінити, а також інтерпретувати обчислений зважений показник споживчого поведінкового наміру на основі показників конативності, афектності та компетентності. Головна підсистема взаємодіє із підсистемою візуальної аналітики, та дає можливість візуалізувати вже наявні показники споживчих намірів клієнтів за дописами та показати візуально рівень споживчого наміру тестового допису відносно обраного об'єкту.

Для подальшої розробки інформаційної системи було сформовано комбінацію засобів розробки, а саме: Visual Studio Code у якості редактору програмного коду, для формування датасету щодо визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів, мову JavaScript у якості мови програмування. У якості програмних засобів для навчання класифікаторів SVC та LSTM, а також для побудови інтерфейсу користувача було обрано для використання мову програмування Python та редактор програмного коду PyCharm.

На базі сформованої комбінації засобів розробки було обрано набір спеціалізованих програмних розширень, а саме: бібліотека Scikit-learn для навчання класифікаторів SVC та LSTM, Keras для попередньої обробки тексту та роботи з текстовими послідовностями, бібліотека Stanza для лематизації допису та опису об'єкту, що відноситься до допису для визначення показника компетентності, matplotlib.pyplot для графічної візуалізації результатів. Для оцінки роботи класифікаторів SVC та LSTM буде використано набір метрик: матрицю сплутувань, точність, влучність і повноту.

Розділ 4 Дослідження ефективності методу автоматизованого визначення поведінкових намірів клієнтів

4.1 Програмна архітектура прикладної інформаційної системи

Етап створення складових частин прикладного програмного забезпечення починається з проектування програмної архітектури. Діаграма класів, що реалізує метод визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів наведено на рисунку 4.1.

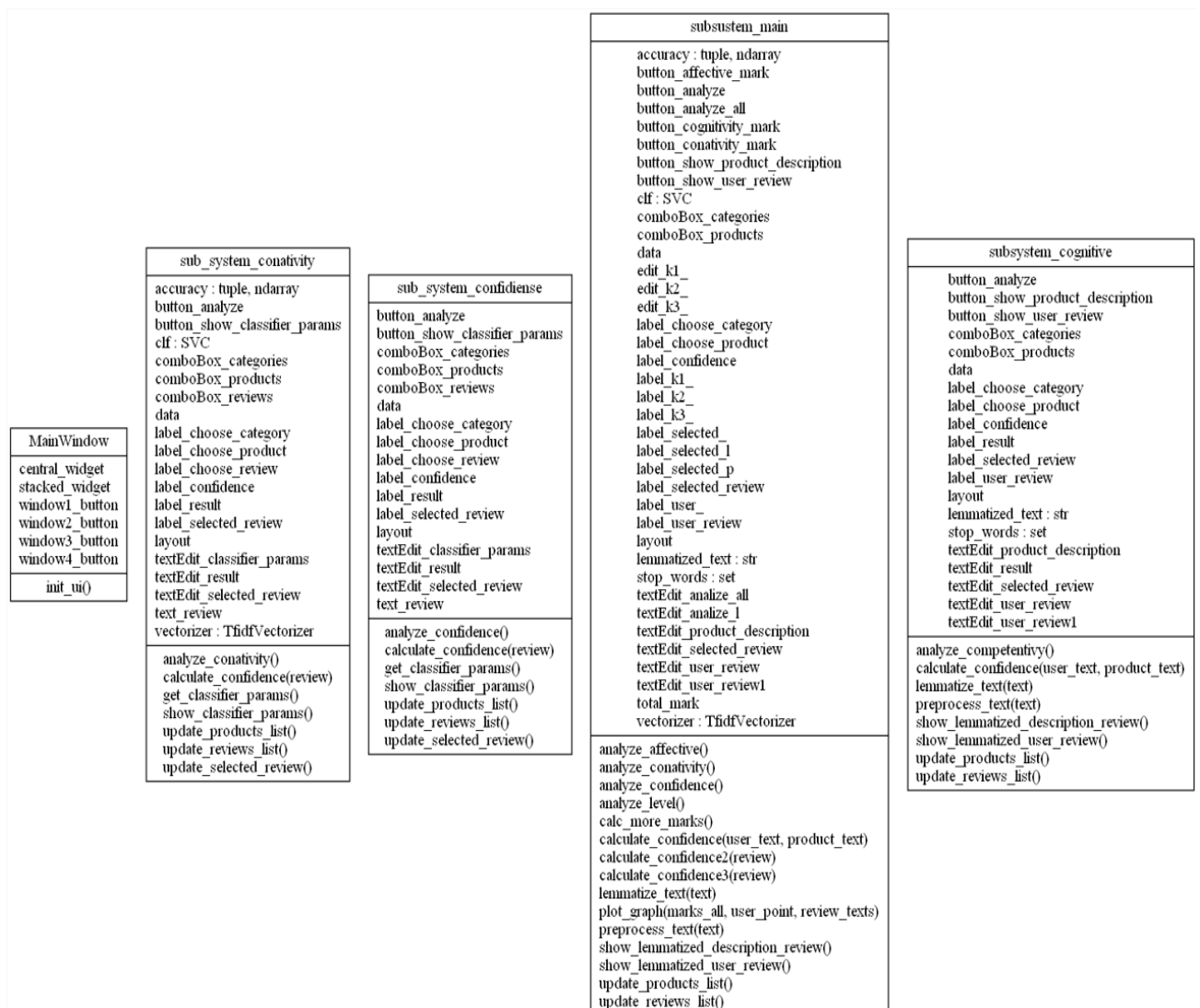


Рисунок 4.1 – Діаграма класів інформаційної системи автоматизованого визначення поведінкових намірів клієнтів

Головним класом є «`subsystem_main`», який використовує основні можливості трьох допоміжних підсистем, що включають аналіз показників конативності, когнітивності та афектності. За обрахунок показника афектності відповідає метод `analyze_conativity()`. За обрахунок показника афектності відповідає метод `analyze_affective()`, а за обрахунок показника когнітивності – метод `analyze_confidence()`. Текстові поля `edit_k1_`, `edit_k2_`, `edit_k3_` відповідають за можливість уведення коефіцієнтів впливу кожного показника на обрахунок загального показника споживчого поведінкового наміру клієнта. Метод `calc_more_marks()` відповідає за обрахунок оцінок інших клієнтів за їх відгуками на обраний товар, та дає вихідні дані для побудови графіку, що буде використовуватись для візуальної аналітики стану показника споживчого поведінкового наміру клієнта.

Підсистема `sub_system_conativity` призначена для визначення показника конативності за обраним товаром для обраного клієнтського відгуку. Дозволяє оцінити прояв поведінкової дієвої рішучості за текстовим представленням клієнтського відгуку. Метод для визначення показника конативності `analyze_conativity()`, що повертає числове значення показника. Метод `show_classifier_params()` дозволяє переглянути параметри класифікатора SVC, що використовується для визначення показника конативності.

Підсистема `sub_system_confidence` призначена для визначення показника афектності за обраним товаром для обраного клієнтського відгуку. Дозволяє оцінити прояв наявного схвального емоційного контексту за текстовим представленням клієнтського відгуку. Метод для визначення показника афектності `analyze_confidence()`, що повертає числове дробове значення показника. Метод `show_classifier_params()` дозволяє переглянути параметри нейромережевого класифікатора LSTM, що використовується для визначення показника конативності.

Підсистема `subsystem_cognitive` призначена для визначення показника когнітивності за обраним товаром та уведеним клієнтом дописом. Дозволяє оцінити прояв розуміння функціональності об'єкту за лематизованим текстовим

представленням клієнтського відгуку та лематизованим текстовим представленням опису до товару. Метод для визначення показника когнітивності `analyze_competentivny()`, що повертає числове дробове значення показника когнітивності. Метод `lemmatize_text()` виконує лематизацію вхідного тексту, метод `preprocess_text()` виконує попередню обробку тексту, прибирає стоп-символи, стоп-слова тощо.

Отже, було створено програмну архітектуру інформаційної системи автоматизованого визначення поведінкових намірів клієнтів, що дозволяє в подальшому використати створену архітектуру для програмної розробки описаний складових.

4.2 Особливості розробки прикладних компонентів інформаційної системи

За створеною програмною архітектурою інформаційної системи надалі створюються окремі її складові. Для компоненту інформаційної системи, що відповідає за показник афектності необхідно мати переднавчену нейромережу LSTM. Для цього був створений окремий консольний застосунок, який приймав на вхід датасет з даними, що містять 40 000 відгуків про фільми, зібраних з сайту IMDb, попередньо перекладених засобами автоматизованого перекладу з використанням бібліотеки «Googletrans», що є безкоштовною бібліотекою Python, у якій реалізовано API Google Translate, що використовує API Google Translate Ajax для викликів метод перекладу [41]. Для навчання кількість відгуків було обмежено та використано випадкові 5000 представників одного класу, що відповідає за емоційне схвалення, та 5000 зразків протилежного йому класу. Навчалась нейромережа протягом 10 етапів, ілюстрація процесу з показниками точності та функції втрат на рисунку 4.2.

Після процесу навчання було збережено токенизатор та навчену модель нейромережі, які стануть вхідними даними віконного застосування.

```

Epoch 1/10
527/527 [=====] - 100s 185ms/step - loss: 0.4386 - accuracy: 0.8190 - val_loss: 0.3728 - val_accuracy: 0.8512
Epoch 2/10
527/527 [=====] - 91s 172ms/step - loss: 0.3417 - accuracy: 0.8660 - val_loss: 0.3628 - val_accuracy: 0.8616
Epoch 3/10
527/527 [=====] - 91s 172ms/step - loss: 0.3183 - accuracy: 0.8750 - val_loss: 0.3654 - val_accuracy: 0.8483
Epoch 4/10
527/527 [=====] - 91s 173ms/step - loss: 0.2874 - accuracy: 0.8873 - val_loss: 0.3669 - val_accuracy: 0.8455

```

Рисунок 4.2 – Процес навчання нейромережі LSTM

Приклад роботи підсистеми, що відповідає за показник афектності наведено на рисунку 4.3.

Система автоматизованого визначення поведінкових намірів клієнтів
— □ ×

Оцінка конативності
Оцінка афектності
Оцінка когнітивності
Визначення поведінкового наміру

Підсистема визначення показника афектності

Обрати категорію товару для аналізу:

Вологі серветки

Обрати товар обраної категорії для аналізу:

Упаковка вологих серветок Naturelle Antibacterial із D-пантенолом, іонами срібла та вітаміном Е 3 пачки по 72 шт.

Обрати відгук до обраного товару:

О. Ю. Н. - Не сподобалось. Маленькі і слабка пропітка

Обраний відгук для аналізу:

Автор: О. Ю. Н.
Текст відгуку: Не сподобалось. Маленькі і слабка пропітка
Користувачька оцінка автора: 2

Визначити показник афектності у користувачькому відгуку

Результат виконаного аналізу:

Показник афектності допису: 0.14504112303256989

Показати параметри нейромережі

name: LSTM
Total params: 1313089(5.01 MB)
Trainable params: 1313089(5.01 MB)
Non-trainable params: 0.00 Byte
Acuracy: 87.5

Рисунок 4.3 – Приклад роботи підсистеми з визначення показника афектності

Для складової інформаційної системи, що відповідає за показник конативності також необхідно мати вхідними даними навчений класифікатор SVC, який проходить навчання на власноруч створеному датасеті, описаному у другому розділі. Процес добування даних за допомогою створеного парсеру наведено на рисунку 4.4.

```

C:\WINDOWS\system32\cmd.exe
rwill-9000101577952/p366914688"}
INFO CheerioCrawler: Пральний порошок для дитячої білизни Frosch 1215 г (4009175958181) - відгуки покупців | ROZETKA {"url": "https://rozetka.com.ua/ua/frosch_4009175958181/p301608753/comments/" }
INFO CheerioCrawler: Капсули для прання Persil Color Power Caps 46 шт (9000101537529) - відгуки покупців | ROZETKA {"url": "https://rozetka.com.ua/ua/persil-9000101537529/p366914592/comments/" }
INFO CheerioCrawler: Пральний порошок Savex Premium Color 8 кг (3800024047961) - фото, відгуки, характеристики в інтернет-магазині ROZETKA | Купити в Україні: Києві, Харкові, Дніпрі, Одесі, Запоріжжі, Львові {"url": "https://rozetka.com.ua/ua/savex_3800024047961/p373259469/" }
INFO CheerioCrawler: Засіб для делікатного прання Perwoll Renew для білих речей 2970 мл 54 цикли прання (9000101578171) - фото, відгуки, характеристики в інтернет-магазині ROZETKA | Купити в Україні: Києві, Харкові, Дніпрі, Одесі, Запоріжжі, Львові {"url": "https://rozetka.com.ua/ua/perwoll_9000101578171/p366023550/" }
INFO Statistics: CheerioCrawler request statistics: {"requestAvgFailedDurationMillis":113, "requestAvgFinishedDurationMi..."}
70, "crawlerRuntimeMillis":780577, "retryHistogram": [40, 20, 16, 94]}
  
```

Рисунок 4.4 – Процес вебскрапінгу відгуків

Через те, що фізично відгуки та опис товару знаходяться на різних сторінках, необхідно потім по ідентифікатору товару приєднати відгуку до відгуки до опису. Всі дані будуть зібрані в файл json, з яким надалі буде взаємодіяти класифікатор для навчання моделі SVC. Вигляд процесу автоматизованого об'єднання даних наведено на рисунку 4.5

category	name	description	reviews																								
Paperovirushnuku	Паперові рушники Smart Classic 2 шари 365 відривів (594582100169)	Первинна високоякісна целюлоза.Сертифікат FSC (деревина з поновлюваних джерел).365 дуже широких відривів = 545 звичайних відривів шириною 15.7 см.Дуже довгий рулон - 84 м.Дуже широкий відрив - 230 мм.Високий рулон - 220 мм.Розмір відриву 23 см x 22 см.Висока щільність паперу - 36 г/м2.Паперові рушники Smart Classic з двома шарами та 365 дуже широкими відривами - це практичний і зручний варіант для різноманітних господарських потреб. Вони виготовлені з високоякісного паперу та мають два шари, що робить їх міцними та здатними впоратися з великою кількістю завдань. Це дуже ефективний паперовий рушник для вологого прибирання, приготування їжі або інших ситуацій, де потрібно швидко вібрати рідину.Паперові рушники Smart Classic - це надійний та економічний вибір для повсякденних господарських завдань.	(0 items)																								
Paperovirushnuku	Рушники паперові Диво Premio Grande XL 2-шарові 1 рулон 500 відривів Білі (4820003837603)	Паперові рушники Диво Premio "Гранд XL" - універсальні целюлозні рушники на 500 аркушів, яких вистачить надовго. Виготовлені зі 100% первинної целюлози і тому безпечні для контакту з харчовими продуктами.	▲ List<Review> (36 items) *** <table border="1"> <thead> <tr> <th>author</th> <th>comment</th> <th>mark</th> </tr> </thead> <tbody> <tr> <td>Тарас Ключко</td> <td>Великий рулон якісних рушників ще й за таку вартість. Брав по акційній ціні, тому взагалі ідеально</td> <td>5</td> </tr> <tr> <td>Інна Рябоконе</td> <td></td> <td>5</td> </tr> <tr> <td>Евгенія Жижина</td> <td>Рушник найкращий у співвідношенні ціна якість. Рекомендую. Дякую</td> <td>5</td> </tr> <tr> <td>Евгенія Жижина</td> <td>Якісний товар. Рушник найкращий.</td> <td>5</td> </tr> <tr> <td>Веронічка</td> <td>Бумага чудова, дівна, плотна</td> <td>5</td> </tr> <tr> <td>Тетяна Шипелик</td> <td>Постійно використовую ці рушники і поки ,на мою думку, вони найкращі на ринку</td> <td>5</td> </tr> <tr> <td>Тетяна Кравченко</td> <td>Чудові рушники.</td> <td>5</td> </tr> </tbody> </table>	author	comment	mark	Тарас Ключко	Великий рулон якісних рушників ще й за таку вартість. Брав по акційній ціні, тому взагалі ідеально	5	Інна Рябоконе		5	Евгенія Жижина	Рушник найкращий у співвідношенні ціна якість. Рекомендую. Дякую	5	Евгенія Жижина	Якісний товар. Рушник найкращий.	5	Веронічка	Бумага чудова, дівна, плотна	5	Тетяна Шипелик	Постійно використовую ці рушники і поки ,на мою думку, вони найкращі на ринку	5	Тетяна Кравченко	Чудові рушники.	5
author	comment	mark																									
Тарас Ключко	Великий рулон якісних рушників ще й за таку вартість. Брав по акційній ціні, тому взагалі ідеально	5																									
Інна Рябоконе		5																									
Евгенія Жижина	Рушник найкращий у співвідношенні ціна якість. Рекомендую. Дякую	5																									
Евгенія Жижина	Якісний товар. Рушник найкращий.	5																									
Веронічка	Бумага чудова, дівна, плотна	5																									
Тетяна Шипелик	Постійно використовую ці рушники і поки ,на мою думку, вони найкращі на ринку	5																									
Тетяна Кравченко	Чудові рушники.	5																									

Рисунок 4.5 – Об'єднання опису товару з відгуками

Для самого процесу навчання класифікатора необхідно здійснити програмне балансування класів, оскільки вибірка не є рівномірно збалансованою. Після навчання зберігається сам навчений класифікатор SVC, а також векторизатор TF-IDF, навчений на вказаному наборі даних.

Приклад роботи підсистеми, що відповідає за показник конативності наведено на рисунку 4.6.

Система автоматизованого визначення поведінкових намірів клієнтів

Оцінка конативності Оцінка афектності Оцінка когнітивності **Визначення поведінкового наміру**

Підсистема визначення показника конативності

Обрати категорію товару для аналізу:
Паперові-рушнику

Обрати товар обраної категорії для аналізу:
Паперові рушники Ruta Jumbo Roll 2 шари 500 відривів (4820202895503)

Обрати відгук до обраного товару:
Яна Семакина - Все супер. Рушник чудовий. Завжди купую саме такий. Вистачає дуже на довго.

Обраний відгук для аналізу:
Автор: Яна Семакина
Текст відгуку: Все супер. Рушник чудовий. Завжди купую саме такий. Вистачає дуже на довго.
Клієнтська оцінка автора: 5

Визначити показник конативності

Результат виконаного аналізу щодо показника поведінкової дієвої рішучості:
Показник поведінкової дієвої рішучості: Клієнт впевнений, що при нагоді придбає вказаний товар з імовірністю: 100.00%

Показати параметри класифікатора

```
name: SVC
kernel: linear
C: 1.0
Accuracy: 0.9134078212290503
n_support_: [ 423 1090]
class_weight: balanced
```

Рисунок 4.6 – Приклад роботи підсистеми з визначення показника конативності

Для реалізації підсистеми «subsystem_cognitive», що відповідає за силу прояву показника когнітивності було виконано препроцесинг текстів відгуку та опису, які були лематизовані для подальшого аналізу. Приклад виконання лематизації наведено на рисунку 4.7.

Опис обраного товару для аналізу:

Зубна паста Colgate Total 12 Professional Clean Mint містить активний інгредієнт, який надає 12-годинний захист від бактерій та забруднень. Забезпечує відчуття свіжості та чистоти протягом довгого часу.

Введений клієнтський відгук для аналізу:

Люблю цю пасту, вона має приємну свіжість.

Визначити показник когнітивності у клієнтському відгуку

Результат виконаного аналізу:

Показник когнітивності у користувацькому відгуку:
 Кількість співпадінь за лемами: 2
 Спільні лема: {'свіжість', 'паста'}
 Усереднений показник компетентності: 0.2857142857142857

Показати лематизований текст клієнтського відгуку

Показати лематизований текст опису товару

любити цей паста вона мати приємний свіжість

зубний паста Colgate Total 12 Professional Clean Mint містити активний інгредієнт який надавати 12годинний захист від бактерія забруднення забезпечувати відчуття свіжість чистота протягом довгий час

Рисунок 4.7 – Лематизація очищених текстів

Головна підсистема включає в себе проміжні етапи роботи кожної з підсистем, та дозволяє оцінити, а також інтерпретувати обчислений зважений показник споживчого поведінкового наміру на основі показників конативності, афектності та компетентності. Також дана підсистема взаємодіє із підсистемою візуальної аналітики, та дає можливість візуалізувати вже наявні показники споживчих намірів клієнтів за дописами та показати візуально рівень

споживчого наміру тестового допису щодо обраного об'єкту. Приклад роботи головного модуля наведено на рисунку 4.8.

Система автоматизованого визначення поведінкових намірів клієнтів

Оцінка конативності Оцінка афектності Оцінка когнітивності Визначення поведінкового наміру

Підсистема визначення показника споживчого поведінкового наміру клієнта

Обрати категорію товару для аналізу:
Вологі серветки

Обрати товар обраної категорії
Упаковка вологих серветок Superfresh Antibacterial з клапаном 2 пачки по 72 шт.

Введений користувачський відгук для аналізу:
Користуюсь цими серветками 3й місяць. Висока якість, хороша ціна.

Опис обраного товару для аналізу:
Вологі серветки Superfresh Antibacterial м'яко та дбайливо очищують ніжну шкіру мам та малюка. Завдяки особливому поєднанню очищувальних і антибактеріальних компонентів, серветки знімають зі шкіри різні забруднення, усувають бактерії, а

Показати лематизований текст відгука клієнта Показати лематизований текст опису товару

Лематизований клієнтський відгук:
користуватися цей серветка 3й місяць високий якість хороший ціна

Лематизований опис обраного товару:
вологий серветка Superfresh Antibacterial який дбайливо очищати ніжний шкіра мама малюка завдяки особливий поєднання очищувальний антибактеріальний компонент серветка знімати зі шкіра різний забруднення усувати бактерія збалансований формула забезпечувати

Визначити показник афектності Визначити показник конативності Визначити показник когнітивності

Обраховані показники:
Показник афектності допису: 0.6551871299743652
Показник поведінкової дієвої рішучості: Клієнт впевнений, що при нагоді придбає вказаний товар з імовірністю: 99.73%
Показник когнітивності у відгуку клієнта: 0.1111111111111111

Визначити показник споживчого поведінкового наміру клієнта

Значення коефіцієнту афектності: 0.3 Значення коефіцієнту конативності: 0.6 Значення коефіцієнту когнітивності: 0.1

Значення показника споживчого поведінкового наміру клієнта:
Значення показника споживчого поведінкового наміру клієнта: 0.6971666212509157

Визначити показник споживчого поведінкового наміру клієнта для всіх відгуків обраного товару

Рисунок 4.8 – Підсистема визначення показника поведінкового наміру клієнта

Отже, таким чином було розроблено компоненти інформаційної системи автоматизованого визначення поведінкових намірів клієнтів та описано особливості їх створення.

4.3 Прикладне тестування інформаційної системи

Для перевірки коректності інформаційної системи було проведено ряд тест-кейсів. Оскільки програмний продукт містить тісну взаємодію з відгуками клієнтів, потрібно перевірити коректність завантаження обраного відгуку за обраною категорією та обраним товаром для підсистеми визначення показника конативності. Кроки тест-кейса наведено у таблиці 4.1.

Таблиця 4.1 – Тест-кейс M001

Тест-кейс ID: M001	Пріоритет: 1	Створено: 1.12.2023, Олег МЕЛЬНИК
Назва: Перевірка коректності відображення обраного відгуку клієнта за обраною категорією та обраним товаром для підсистеми визначення показника конативності.		
Кроки		Очікуваний результат
<ol style="list-style-type: none"> 1. Запустити програму; 2. Перейти на підсистему «Підсистема визначення показника конативності» 3. Обрати категорію «Паперові рушники» 4. Обрати товар «Паперові рушники Ruta Soft & Strong 3 шари 8 рулонів (4820202891079_1)» 5. Обрати відгук «Хороші рушники, якщо по акції, взагалі супер» 		<p>Відкрився головний екран застосунку</p> <p>Відкрилась підсистема визначення показника конативності</p> <p>У випадаючому списку для вибору товару обраної категорії відображаються товари з категорії «Паперові рушники»</p> <p>У випадаючому списку відгуків на обраний товар відображаються відгуки, що відносяться до товару «Паперові рушники Ruta Soft & Strong 3 шари 8 рулонів (4820202891079_1)»</p> <p>У текстове поле «Обраний відгук для аналізу» Виводиться текст відгука , автор та клієнтська оцінка.</p>
Результат виконання тест-кейсу: пройдено успішно		

Результат успішного виконання тест-кейсу наведено на рисунку 4.9.

Система автоматизованого визначення поведінкових намірів клієнтів

Оцінка конативності Оцінка афектності Оцінка когнітивності Визначення поведінкового наміру

Підсистема визначення показника конативності

Обрати категорію товару для аналізу:
Паперові-rushnuku

Обрати товар обраної категорії для аналізу:
Паперові рушники Ruta Soft & Strong 3 шари 8 рулонів (4820202891079_1)

Обрати відгук до обраного товару:
Гончар Ірина - Хороші рушники, якщо по акції, взагалі супер

Обраний відгук для аналізу:
Автор: Гончар Ірина
Текст відгуку: Хороші рушники, якщо по акції, взагалі супер
Клієнтська оцінка автора: 5

Рисунок 4.9 – Результат проходження тест-кейсу M001

Наступною буде перевірка працездатності визначення показника конативності за обраним відгуком клієнта до товару для підсистеми визначення показника конативності. Кроки тест-кейса наведено у таблиці 4.2.

Таблиця 4.2 – Тест-кейс M002

Тест-кейс ID: M002	Пріоритет: 1	Створено: 2.12.2023, Олег МЕЛЬНИК
Назва: Перевірка коректності визначення показника конативності за обраним відгуком клієнта.		
Кроки	Очікуваний результат	
<ol style="list-style-type: none"> Запустити програму; Перейти на підсистему «Підсистема визначення показника конативності» Обрати категорію «Паперові рушники» Обрати товар «Паперові рушники Ruta Soft & Strong 3 шари 8 рулонів (4820202891079_1)» Обрати відгук «Хороші рушники, якщо по акції, взагалі супер» Натиснути кнопку «Визначити показник конативності» 	<p>Відкрився головний екран застосунку</p> <p>Відкрилась підсистема визначення показника конативності</p> <p>У випадаючому списку для вибору товару обраної категорії відображаються товари з категорії «Паперові рушники»</p> <p>У випадаючому списку відгуків на обраний товар відображаються відгуки, що відносяться до товару «Паперові рушники Ruta Soft & Strong 3 шари 8 рулонів (4820202891079_1)»</p> <p>У текстове поле «Обраний відгук для аналізу» Виводиться текст відгука , автор та клієнтська оцінка.</p> <p>За даним текстом відгука показник конативності має становити понад 90%.</p>	
Результат виконання тест-кейсу: пройдено успішно		

Результат успішного виконання тест-кейсу наведено на рисунку 4.10.

Система автоматизованого визначення поведінкових намірів клієнтів

Оцінка конативності Оцінка афектності Оцінка когнітивності **Визначення поведінкового наміру**

Підсистема визначення показника конативності

Обрати категорію товару для аналізу:
Паперові-rushnuku

Обрати товар обраної категорії для аналізу:
Паперові рушники Ruta Soft & Strong 3 шари 8 рулонів (4820202891079_1)

Обрати відгук до обраного товару:
Гончар Ірина - Хороші рушники, якщо по акції, взагалі супер

Обраний відгук для аналізу:
Автор: Гончар Ірина
Текст відгуку: Хороші рушники, якщо по акції, взагалі супер
Клієнтська оцінка автора: 5

Визначити показник конативності

Результат виконаного аналізу щодо показника поведінкової дієвої рішучості:

Показник поведінкової дієвої рішучості: Клієнт впевнений, що при нагоді придбає вказаний товар з імовірністю 99.43%

Рисунок 4.10 – Результат проходження тест-кейсу M001

Наступним тестовим випадком буде перевірено працездатності лематизації відгука клієнта для підсистеми визначення показника когнітивності. Кроки тест-кейса наведено у таблиці 4.3.

Таблиця 4.3 – Тест-кейс M003

Тест-кейс ID: M003	Пріоритет: 1	Створено: 2.12.2023, Олег МЕЛЬНИК
Назва: Перевірка коректності лематизації відгука клієнта для підсистеми визначення показника когнітивності.		
Кроки	Очікуваний результат	
<ol style="list-style-type: none"> Запустити програму; Перейти на підсистему «Підсистема визначення показника компетентності» Увести у поле «Введений клієнтський відгук для аналізу» текст: <i>«Чудові рушники постійно ними користуємось для ванної кімнати.»</i> Натиснути кнопку «Показати лематизований текст клієнтського відгука» 	<p>Відкрився головний екран застосунку</p> <p>Відкрилась підсистема визначення показника компетентності</p> <p>У текстовому полі відображається уведений текст</p> <p>Вивід: <i>«чудовий рушник постійно вони користуватися для ванний кімната»</i></p>	
Результат виконання тест-кейсу: пройдено успішно		

Результат успішного виконання тест-кейсу наведено на рисунку 4.11.

Введений клієнтський відгук для аналізу:

Чудові рушники постійно ними користуємось для ванної кімнати.

Визначити показник когнітивності у клієнтському відгуку

Результат виконаного аналізу:

Показати лематизований текст клієнтського відгука

чудовий рушник постійно вони користуватися для ванний кімната

Показати лематизований текст опису товару

Рисунок 4.11 – Лематизація клієнтського відгука

Наступним тестовим випадком буде перевірено функцію визначення показника когнітивності для підсистеми визначення показника споживчого поведінкового наміру клієнта. Кроки тест-кейса наведено у таблиці 4.4.

Таблиця 4.4 – Тест-кейс M004

Тест-кейс ID: M004	Пріоритет: 1	Створено: 3.12.2023, Олег МЕЛЬНИК
Назва: Перевірка коректності функції визначення показника когнітивності для підсистеми визначення показника споживчого поведінкового наміру клієнта		
Кроки		Очікуваний результат
<ol style="list-style-type: none"> 1. Запустити програму; 2. Перейти на підсистему «Підсистема визначення показника споживчого поведінкового наміру клієнта» 3. Обрати категорію товару «Засоби для миття посуду» 4. Обрати з випадаючого списку товарів «Таблетки для миття посуду у посудомийній машині Somat All in one Все в одному 110 шт (9000101577044)» 5. Увести у поле «Введений користувачький відгук для аналізу» текст: <i>«Найкращі таблетки що пробували. Посуд блищить, розводів немає. Ціна приємна, будемо купувати ще».</i> 6. Натиснути кнопку «Лематизувати клієнтський відгук» 7. Натиснути на кнопку «Лематизувати опис обраного товару» 8. Натиснути кнопку «Визначити показник когнітивності» 		<p>Відкрився головний екран застосунку</p> <p>Відкрилась підсистема визначення показника споживчого поведінкового наміру клієнта</p> <p>У випадаючому списку з'явилися товари, що належать даній категорії</p> <p>У текстовому полі «Опис обраного товару для аналізу» з'явився текст опису обраного товару .</p> <p>У текстовому полі відображається уведений текст</p> <p>Виводиться лематизований текст відгуку</p> <p>Виводиться лематизований текст опису</p> <p>Виводиться обрахований показник когнітивності</p>
Результат виконання тест-кейсу: пройдено успішно		

Результат успішного виконання тест-кейсу, що присвячений перевірці коректності визначення показника когнітивності наведено на рисунку 4.12.

Система автоматизованого визначення поведінкових намірів клієнтів

Оцінка конативності Оцінка афектності Оцінка когнітивності Визначення поведінкового наміру

Підсистема визначення показника споживчого поведінкового наміру клієнта

Обрати категорію товару для аналізу:
Zasobu-dlya-muttya-posudu

Обрати товар обраної категорії
Таблетки для миття посуду у посудомийній машині Somat All in one Все в одному 110 шт (9000101577044)

Введений користувачький відгук для аналізу:
Найкращі таблетки що пробували. Посуд блистить, розвідів немає. Ціна приємна, будемо купувати ще

Опис обраного товару для аналізу:
Таблетки для посудомийної машини Somat Все в одному для чистого і блискучого посуду, що надають потужне очищення від засохлих плям навіть при низькій температурі на режимі Eco та на коротких циклах миття посуду. Таблетки Somat All in 1 виконують 6 функцій:

Показати лематизований текст відгука клієнта Показати лематизований текст опису товару

Лематизований клієнтський відгук:
найкращий таблетка пробувати посуд блистити розвід немати ціна приємний бути купувати

Лематизований опис обраного товару:
таблетка для посудомийний машина Somat все один для чистий блискучий посуд надавати потужний очищення від засохлий пляма навіть при низький температура режим Eco короткий цикл миття посуд таблетка Somat All in виконувати функція функція очищення догляд за

Визначити показник афектності Визначити показник конативності **Визначити показник когнітивності**

Обраховані показники:
Показник когнітивності у відгуку клієнта: 0.18

Рисунок 4.12 – Перевірка коректності визначення показника когнітивності

Згідно до проведено тестування програмного забезпечення, можна зробити висновок, що розроблена інформаційна система автоматизованого визначення поведінкових намірів клієнтів працює коректно, реалізовані методи інформаційної системи повертають очікувані прогнозовані значення.

4.4 Дослідження ефективності та інтерпретація отриманих результатів засобами візуальної аналітики

Для дослідження ефективності розробленого в роботі методу визначення поведінкових намірів клієнтів за їх дописами було використано розроблену інформаційну систему автоматизованого визначення поведінкових намірів клієнтів, що є прикладною програмною реалізацією розробленого методу.

Система автоматизованого визначення поведінкових намірів клієнтів

Оцінка конативності Оцінка афектності Оцінка когнітивності **Визначення поведінкового наміру**

Підсистема визначення показника споживчого поведінкового наміру клієнта

Обрати категорію товару для аналізу:
Паперови-gushnuku

Обрати товар обраної категорії
Рушники паперові Диво Premio Grande XL 2-шарові 1 рулон 500 відривів Білі (4820003837603)

Введений користувачський відгук для аналізу:
Паперові рушники Диво Premio це кращий мій вибір. Добре впітують вологу, міцні. Рекомендую.

Опис обраного товару для аналізу:
Паперові рушники Диво Premio "Гранд XL" - універсальні целюлозні рушники на 500 аркушів, яких вистачить надовго. Виготовлені зі 100% первинної целюлози і тому безпечні для контакту з харчовими продуктами.

Показати лематизований текст відгука клієнта Показати лематизований текст опису товару

Лематизований клієнтський відгук:
паперовий рушник диво Premio це кращий мій вибір добре впітувати волога міцні рекомендувати

Лематизований опис обраного товару:
паперовий рушник диво Premio Гранд XL універсальний целюлозний рушник 500 аркуш який вистачити надовго виготовлений зі 100 первинний целюлоз тому безпечний для контакт харчовий продукт

Визначити показник афектності **Визначити показник конативності** **Визначити показник когнітивності**

Обраховані показники:

Показник афектності допису: 0.9405445456504822
Показник поведінкової дієвої рішучості: Клієнт впевнений, що при нагоді придбає вказаний товар з імовірністю: 95.66%
Показник когнітивності у відгуку клієнта: 0.3076923076923077

Визначити показник споживчого поведінкового наміру клієнта

Значення коефіцієнту афектності: 0.3 Значення коефіцієнту конативності: 0.6 Значення коефіцієнту когнітивності: 0.1

Значення показника споживчого поведінкового наміру клієнта:
Значення показника споживчого поведінкового наміру клієнта: 0.7597847280746151

Рисунок 4.13 – Обрахунок показників

Для товару «Рушники паперові Диво Premio Grande XL» було написано тестовий відгук, що містить високі показники когнітивності, афектності та

компетентності. Текст відгука був наступним: «Паперові рушники Диво Premіо це кращий мій вибір. Добре впитують вологу, міцні. Рекомендуюю». Отримані програмою показники наведені на рисунку 4.13.

Далі було сформовано графік, що відображає динаміку споживчих намірів відносно товару «Рушники паперові Диво Premіо Grande XL», та нанесено позиціонування червоним маркером тестового клієнтського відгуку (рисунок 4.14).

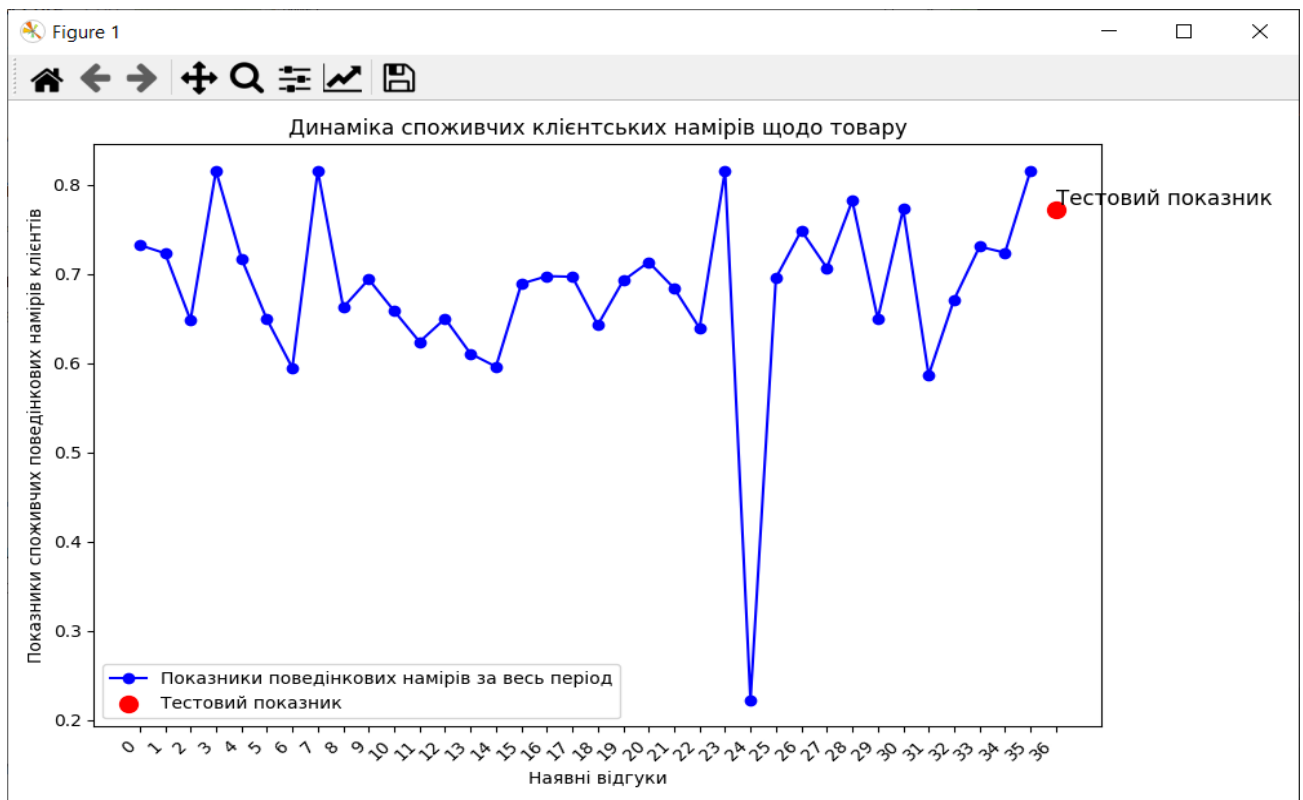


Рисунок 4.14 – Динаміка споживчих купівельних намірів щодо товару «Рушники паперові Диво Premіо Grande XL»

Згідно графіку на рисунку 4.10, що відображає динаміку споживчих клієнтських намірів щодо товару «Рушники паперові Диво Premіо Grande XL», можна зробити висновок, що відносно даного товару клієнти мають прояв купівельного поведінкового наміру вище середнього та високий, проте відгук під №24 має прояв купівельного поведінкового наміру нижче середнього.

Однак, при виконаному аналізі, на даний товар відсутні погані оцінки (рисунок 4.15), однак, відгук, що показав найнижчу оцінку має такий текст: «Великий рулон полотенець, не особо м'які, але не критично».

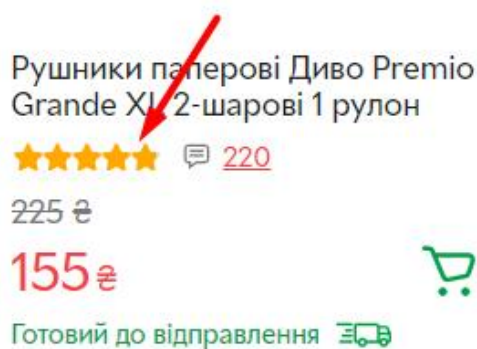


Рисунок 4.15 – Рейтинг товару «Рушники паперові Диво Premio Grande XL» [4]

Відповідно до тексту, користувач має сумніви щодо повторної купівлі, також оцінка прояву афектності також буде доволі низькою, нейромережа визначила її як 0,102. На рисунку 4.16 наведено детальний розбір даного відгуку за параметрами з використанням розробленого програмного забезпечення.

Система автоматизованого визначення поведінкових намірів клієнтів

Оцінка конативності Оцінка афектності Оцінка когнітивності Визначення поведінкового наміру

Підсистема визначення показника споживчого поведінкового наміру клієнта

Обрати категорію товару для аналізу:
Паперови-rushnuku

Обрати товар обраної категорії
Рушники паперові Диво Premio Grande XL 2-шарові 1 рулон 500 відривів Білі (4820003837603)

Введений користувачський відгук для аналізу:
Великий рулон полотенець, не особо м'які, але не критично.

Опис обраного товару для аналізу:
Паперові рушники Диво Premio "Гранд XL" - універсальні целюлозні рушники на 500 аркушів, яких вистачить надовго. Виготовлені зі 100% первинної целюлози і тому безпечні для контакту з харчовими продуктами.

Показати лематизований текст відгука клієнта Показати лематизований текст опису товару

Лематизований клієнтський відгук:
великий рулон полотенець особо який критично

Лематизований опис обраного товару:
паперовий рушник диво Premio Гранд XL універсальний целюлозний рушник 500 аркуш який вистачити надовго виготовлений зі 100 первинний целюлоз тому безпечний для контакт харчовий продукт

Визначити показник афектності Визначити показник конативності Визначити показник когнітивності

Обраховані показники:

Показник поведінкової дієвої рішучості: Клієнт має сумніви щодо повторного придбання товару. Імовірність: 93.24%
Показник когнітивності у відгуку клієнта: 0.16666666666666666
Показник афектності допису: 0.10211749374866486

Визначити показник споживчого поведінкового наміру клієнта

Значення коефіцієнту афектності: 0.3 Значення коефіцієнту конативності: 0.6 Значення коефіцієнту когнітивності: 0.1

Значення показника споживчого поведінкового наміру клієнта:
Значення показника споживчого поведінкового наміру клієнта: 0.10076383265906283

Рисунок 4.16 – Аналіз відгуку з показником споживчого поведінкового наміру нижче середнього

Наступним було досліджено споживчі поведінкові наміри відносно товару «Таблетки для посудомийних машин Fairy Platinum Plus Все-в-одному 84 шт». Для даного товару був написаний позитивний відгук, який був проаналізований програмою, результат аналізу наведено на рисунку 4.17.

Система автоматизованого визначення поведінкових намірів клієнтів

Оцінка конативності Оцінка афектності Оцінка когнітивності **Визначення поведінкового наміру**

Підсистема визначення показника споживчого поведінкового наміру клієнта

Обрати категорію товару для аналізу:
 Zasobu-dlya-muttya-posudu

Обрати товар обраної категорії:
 Таблетки для посудомийних машин Fairy Platinum Plus Все-в-одному 84 шт (8001841893693)

Введений користувачський відгук для аналізу:
 Капсули відповідають опису на всі 100! Відмінний результат

Опис обраного товару для аналізу:
 Капсули для посудомийної машини Fairy Platinum Plus Все-в-одному забезпечують найкраще очищення від Fairy і допомагають видалити навіть утворений з часом наліт, повертаючи посуду його первинний блиск. Fairy Platinum Plus обладнаний вбудованою системою

Показати лематизований текст відгука клієнта Показати лематизований текст опису товару

Лематизований клієнтський відгук:
 капсула відповідати опис весь 100 відмінний результат

Лематизований опис обраного товару:
 капсула для посудомийний машина Fairy Platinum Plus всеодний забезпечувати найкраще очищення від Fairy допомагати видалити навіть утворений час наліт повертати посуд його первинний блиск Fairy Platinum Plus обладнаний вбудований система попередній

Визначити показник афектності **Визначити показник конативності** **Визначити показник когнітивності**

Обраховані показники:
 Показник афектності допису: 0.6455559730529785
 Показник поведінкової дієвої рішучості: Клієнт впевнений, що при нагоді придбає вказаний товар з імовірністю: 95.02%
 Показник когнітивності у відгуку клієнта: 0.14285714285714285

Визначити показник споживчого поведінкового наміру клієнта

Значення коефіцієнту афектності: 0.3 **Значення коефіцієнту конативності: 0.6** Значення коефіцієнту когнітивності: 0.1

Значення показника споживчого поведінкового наміру клієнта:
 Значення показника споживчого поведінкового наміру клієнта: 0.6743556222631304

Визначити показник споживчого поведінкового наміру клієнта для всіх відгуків обраного товару

Рисунок 4.17 – Аналіз відгуку з показником споживчого поведінкового наміру вище середнього

Також за допомогою створеної підсистеми візуальної аналітики було сформовано графік що відображає динаміку споживчих намірів відносно товару «Таблетки для посудомийних машин Fairy Platinum Plus Все-в-одному 84 шт», та

нанесено позиціонування червоним маркером тестового клієнтського відгуку (рисунок 4.18).

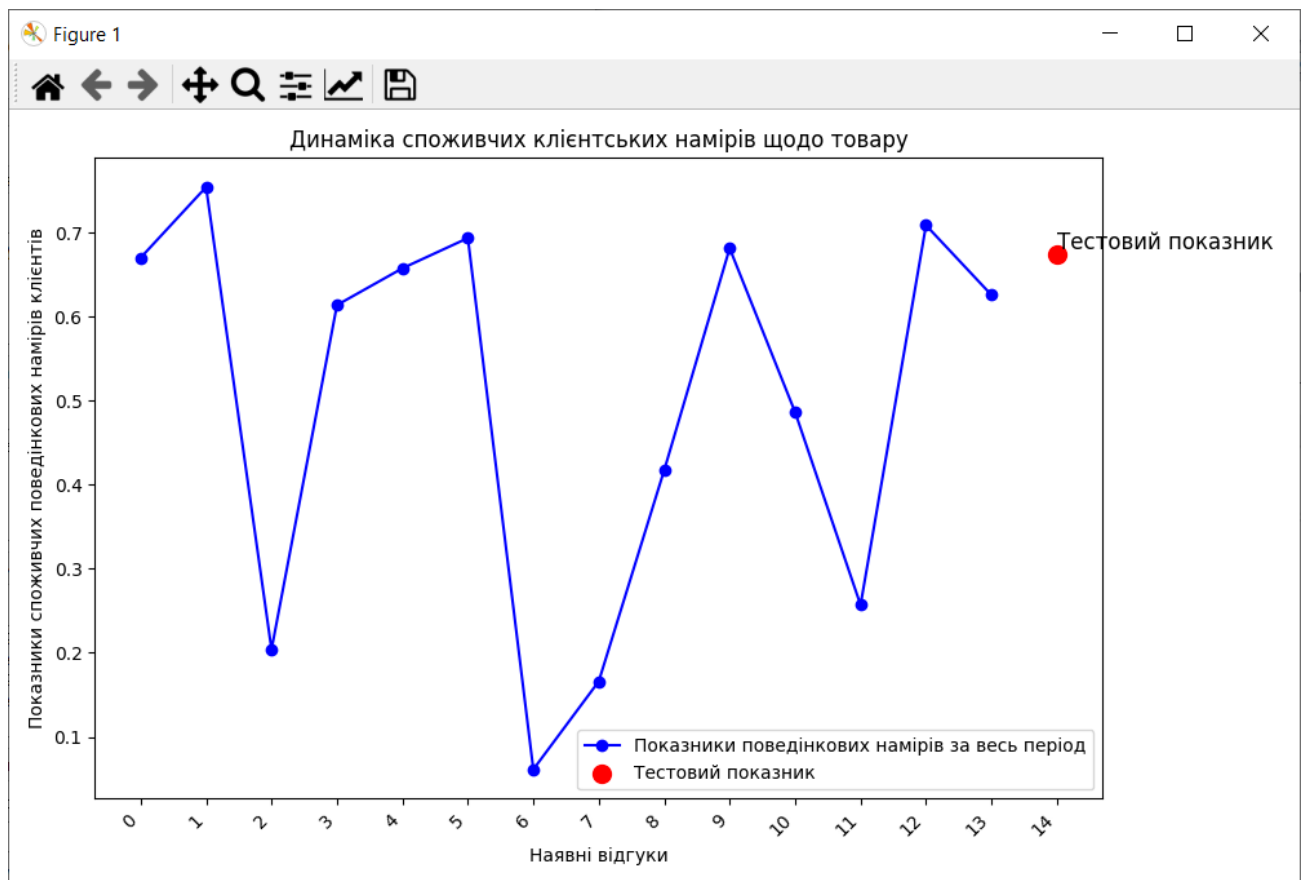


Рисунок 4.18 – Динаміка споживчих купівельних намірів щодо товару «Таблетки для посудомийних машин Fairy Platinum Plus Все-в-одному 84 шт»

Графік наведений на рисунку 4.18 в цілому показує, що споживчі наміри щодо купівлі даного товару мають здебільшого рівень вище середнього, хоча і є декілька відгуків з низьким та середнім показниками споживчого купівельного наміру. Також наведений на рисунку 4.14 графік корелює з відгуками на платформі «Розетка» (рисунок 4.19)



Рисунок 4.19 – Рейтинг товару «Таблетки для посудомийних машин Fairy Platinum Plus Все-в-одному 84 шт» [4]

Отже, розроблена інформаційна система автоматизованого визначення поведінкових намірів клієнтів є ефективною та допомагає забезпечити для клієнтів засобів електронної комерції підвищення довіри до відгуків, що дозволяє орієнтуватись на відгуки при здійсненні покупки, а для власників покращення зворотного зв'язку для підвищення ефективності процесу купівлі-продажу товарів і послуг.

Висновки до розділу 4

У рамках виконання четвертого розділу було виконано дослідження ефективності інформаційної системи автоматизованого визначення поведінкових намірів клієнтів, що є прикладною програмною реалізацією метода визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на основі

перетворення вхідних даних у формі навченої нейромережевої моделі LSTM для визначення прояву афектності за дописом, навченої моделі класифікатора SVC для визначення прояву конативності, вхідного клієнтського допису для аналізу у вихідні дані у вигляді показника споживчого поведінкового наміру клієнта та показників його складників: конативності, афектності та компетентності.

Були описані основні моменти з розробки прикладних компонентів, процес навчання класифікаторів та процес створення всіх підсистем із архітектури програмного комплексу. Після створення прикладних компонентів, було проведено тестування програмного забезпечення, з якого можна зробити висновок, що розроблена інформаційна система автоматизованого визначення поведінкових намірів клієнтів працює коректно, реалізовані методи інформаційної системи повертають очікувані прогнозовані значення.

Безпосередньо для дослідження ефективності розробленого в роботі методу визначення поведінкових намірів клієнтів за їх дописами було використано розроблену інформаційну систему автоматизованого визначення поведінкових намірів клієнтів, що є прикладною програмною реалізацією розробленого методу. Інформаційна система показала, що визначення поведінкових намірів клієнтів є ефективною та допомагає забезпечити для клієнтів засобів електронної комерції підвищення довіри до відгуків, що дозволяє орієнтуватись на відгуки при здійсненні покупки, а для власників покращення зворотного зв'язку для підвищення ефективності процесу купівлі-продажу товарів і послуг.

Загальні висновки

Кваліфікаційна робота магістра розв'язує науково-технічну задачу автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на прикладі предметної області відгуків щодо товарів щоденного вжитку. Розроблений метод призначений для визначення споживчих поведінкових намірів клієнтів за семантичним аналізом коментарів товарів інтернет-магазину.

Вхідними даними методу визначення споживчих поведінкових намірів клієнтів за семантичним аналізом їх дописів є навчальна вибірка для моделі класифікатора SVC для визначення прояву афектності за дописом та нейромережевої моделі LSTM для визначення прояву афектності, а також користувацький допис для аналізу. Одержана вихідна інформація є оцінкою поведінкового наміру схильності клієнта до купівлі обраного товару й додатково містить показники конативності, афектності та компетентності споживчого поведінкового наміру клієнта.

На шляху до вирішення задачі автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на прикладі споживчих поведінкових намірів клієнтів у межах предметної області відгуків щодо товарів щоденного вжитку в засобах електронної комерції було виконано такі завдання дослідження:

1. Проведено аналіз предметної області щодо визначення поведінкових намірів клієнтів електронної комерції, виділено актуальні практичні підходи для автоматизації визначення поведінкових намірів.

2. Досліджено теоретичні підходи щодо автоматизації визначення поведінкових намірів за семантичним аналізом дописів та обрано комбінацію теоретичних підходів для використання в методі.

3. Розроблено метод автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на прикладі предметної області відгуків щодо товарів повсякденного вжитку.

4. Спроектовано інформаційну систему на основі методу автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів.

5. Обрано засоби розробки для спроектованої архітектури інформаційної системи та розроблено відповідну програмну реалізацію методу автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів на прикладі визначення споживчих поведінкових намірів клієнтів у межах предметної області відгуків щодо товарів щоденного вжитку в засобах електронної комерції.

6. Досліджено ефективність програмно реалізованого методу автоматизованого визначення поведінкових намірів клієнтів.

Дослідження ефективності методу автоматизованого визначення поведінкових намірів клієнтів за використання його програмної реалізації виявило, що визначення поведінкових намірів клієнтів є ефективною практикою та допомагає забезпечити для клієнтів засобів електронної комерції підвищення довіри до відгуків, що дозволяє орієнтуватись на відгуки при здійсненні покупки, а для власників покращення зворотного зв'язку для підвищення ефективності процесу купівлі-продажу товарів і послуг. Автоматизоване визначення поведінкових намірів через семантичний аналіз дописів клієнтів важливе для вирішення завдання персоналізації обслуговування, виявлення та аналізу тенденцій споживацького попиту та розробки інтелектуальних систем рекомендацій.

За темою роботи автором у фаховому виданні виконано наукову публікацію [5], яка присвячена визначенню поведінкових намірів клієнтів нейромережевими засобами.

Перелік посилань

1. Що таке електронна комерція URL:
<https://wezom.com.ua/ua/blog/elektronna-komertsiya>
2. Основні типи електронної комерції. URL:
<https://interkassa.com/blog/shho-take-elektronna-komerciya-e-commerce-dlya-rochatkivciv>
3. Право регулювання організації діяльності Інтернет-магазину. URL:
https://wiki.legalaid.gov.ua/index.php/Право_регулювання_організації_діяльності_Інтернет-магазину
4. Розетка (інтернет-магазин). URL: [https://uk.wikipedia.org/wiki/Розетка_\(інтернет-магазин\)](https://uk.wikipedia.org/wiki/Розетка_(інтернет-магазин))
5. Залуцька О.О., Молчанова М.О., Мазурець О.В., Мельник О.І., Скрипник Т.К. Метод інтелектуального аналізу емоційної тональності текстової інформації для визначення поведінкових намірів нейромережевими засобами. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2023. №5 (325). Т.1. С. 67-73. URL:
<http://journals.khnu.km.ua/vestnik/wp-content/uploads/2023/11/325-ts-2023-5-1-3-8.pdf>
6. Відгуки покупців про Антисептик спрей для дезінфекції рук, тіла та поверхонь Touch Protect 250 мл (4823109400894). URL:
https://rozetka.com.ua/ua/touch_protect_4823109400894/p197774893/comments/
7. 10 прикладів відгуків в інтернет-магазині. URL:
<https://lemarbet.com/ua/seo-instrumenty/10-primerov-otzyvov-v-im/>
8. Соціальна установка. URL:
https://pidru4niki.com/1292052239107/psihologiya/sotsialna_ustanovka
9. Екзистенція страху життя і смерті в сучасних реаліях. URL:
https://dspace.lvduvs.edu.ua/bitstream/1234567890/5060/3/Монографія_Католик%20%284%29.pdf

10. Emotion classification of text. URL: <https://epub.jku.at/obvulihs/download/pdf/8620963?originalFilename=true>
11. Sepp Hochreiter and Jürgen Schmidhuber. “Long short-term memory”. In: *Neural computation* 9.8 (1997), pp. 1735–1780 (cit. on pp. 3, 22)
12. Ashish Vaswani et al. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017) (cit. on pp. 3, 25, 30, 36)
13. BERT: Why it’s been revolutionizing NLP. URL: <https://towardsdatascience.com/bert-why-its-been-revolutionizing-nlp-5d1bcae76a13>
14. Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. “BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding”. In: (2019), pp. 4171–4186 (cit. on pp. 3, 32, 33, 34, 36, 60, 67)
15. Наївний байєсівський класифікатор Apache Spark MLlib: глибоке занурення. URL: <https://ts2.space/uk/наивний-байєсівський-класифікатор-apache-spar/#gsc.tab=0>
16. Використання методу опорних векторів для розпізнавання літер друкованого тексту. URL: <https://naukam.triada.in.ua/index.php/konferentsiji/47-simnadtsyata-vseukrajinska-praktichno-piznavalna-internet-konferentsiya/371-vikoristannya-metodu-opornikh-vektoriv-dlya-rozpiznavannya-liter-drukovanogo-tekstu>
17. Обробка природної мови (NLP) у Python з кодом (Частина 2. Класифікація текстів). URL: <https://oleg-dubetcky.medium.com/обробка-природної-мови-nlp-у-python-з-кодом-частина-2-класифікація-текстів-b168878ba32d>
18. Шкалювання емоційно забарвлених слів для використання у методах класифікації тональності. URL: <https://science.lpnu.ua/sites/default/files/journal-paper/2018/jun/13016/ilovepdfcom-195-203.pdf>
19. Виявлення негативних впливів у соціальних інтернет-сервісах. URL: <http://infotech-soccult.knukim.edu.ua/article/view/155667/157353>
20. ChatGPT. URL: <https://sendpulse.ua/support/glossary/chatgpt>

21. Відстеження тону. URL: <https://learn.microsoft.com/uk-ua/dynamics365/customer-service/supervisor-sentiment-monitoring>
22. Обробка природної мови (NLP) у Python з кодом (Частина 2. Класифікація текстів). URL: <https://oleg-dubetcky.medium.com/обробка-природної-мови-nlp-у-python-з-кодом-частина-2-класифікація-текстів-b168878ba32d>
23. RELU. URL: <https://pytorch.org/docs/stable/generated/torch.nn.ReLU.html>
24. Функції активації. URL: <https://robotdreams.cc/uk/blog/327-funkciji-aktivaciji-stupinchasta-liniyna-sigmojida-relu-ta-tanh#item4>
25. Лематизація. URL: <https://uk.wikipedia.org/wiki/Лематизація>
26. IMDB Dataset of 50K Movie Reviews. URL: <https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>
27. Visual Studio Code. URL: <https://visualstudio.microsoft.com/#vscode-section>
28. Початок роботи з веб-скрапінгом у JavaScript. URL: <https://techukraine.net/початок-роботи-з-веб-скрапінгом-у-javascript/>
29. Crawlee is a web scraping and browser automation library. URL: <https://crawlee.dev/>
30. Переваги використання мови python для глибокого машинного навчання. URL: <https://openarchive.nure.ua/server/api/core/bitstreams/889aeb40-b609-47b7-a5a3-80a543b9e90a/content>
31. Дослідження переваг поєднання Scikit-learn і нейронних мереж. URL: <https://ts2.space/uk/scikit-learn-i-нейронні-мережі-інтеграція-scikit-learn-із-б/#gsc.tab=0>.
32. Scikit-learn. URL: <https://scikit-learn.org/stable/>
33. Keras: The high-level API for TensorFlow. URL: <https://www.tensorflow.org/guide/keras>

34. Інструменти для Natural Language Understanding: поради, особливості роботи, та українська мова в NLP. URL: <https://dou.ua/forums/topic/38785/>
35. Pyplot tutorial. URL: <https://matplotlib.org/stable/tutorials/pyplot.html>
36. Матриця невідповідностей і метрики. URL: <https://dumnyj.eu/blog/matrytsia-nevidpovidnostej-i-metryky/>
37. Матриця невідповідностей. URL: https://uk.wikipedia.org/wiki/Матриця_невідповідностей
38. Точність. URL: <https://uk.wikipedia.org/wiki/Точність>
39. Влучність та повнота. URL: https://uk.wikipedia.org/wiki/Влучність_та_повнота
40. Класифікаційні метрики. URL: http://www.andriystav.cc.ua/Downloads/MITER/Lecture_04.pdf
41. Googletrans 3.0.0 . URL: <https://pypi.org/project/googletrans/>

ДОДАТКИ

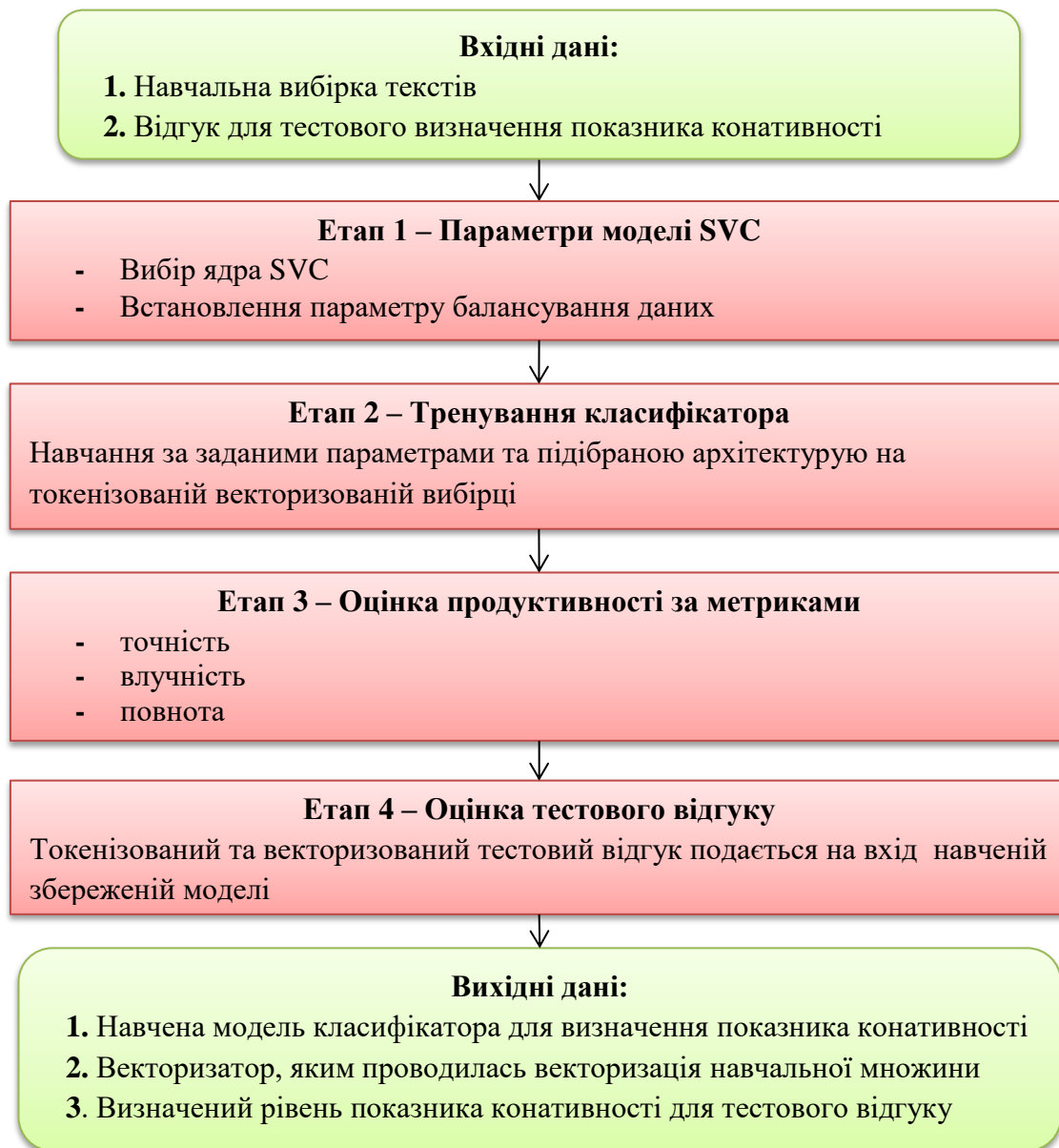
Додаток А

Узагальнена схема методу визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів



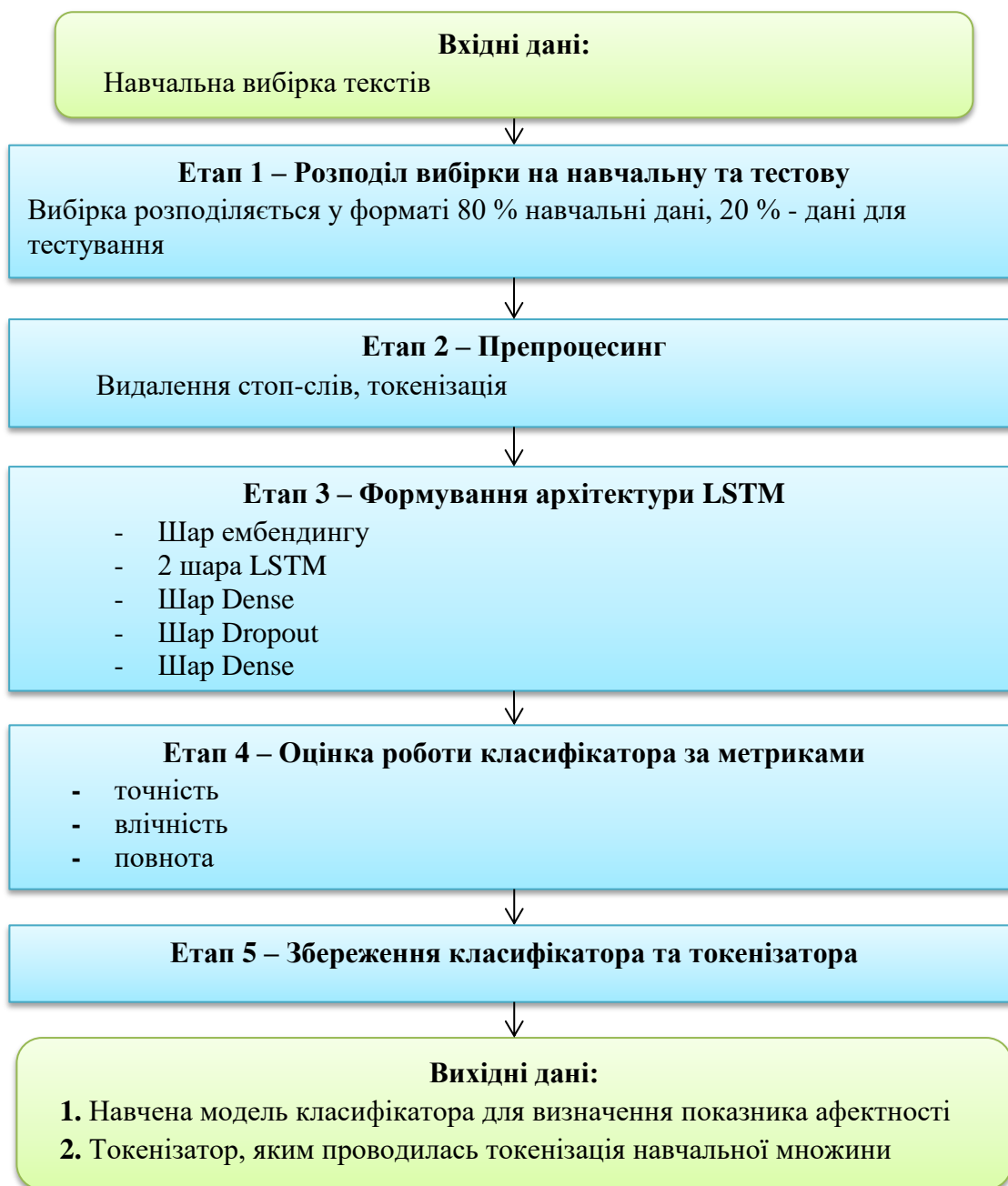
Додаток Б

Формування моделі конативності методом SVC у методі визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів



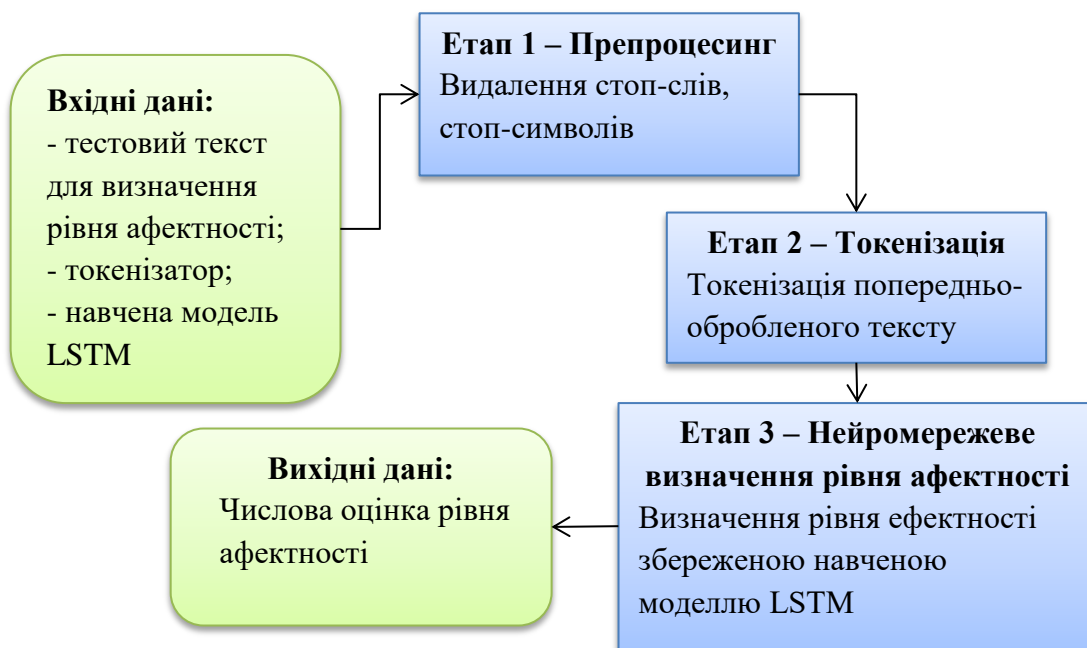
Додаток В

Формування моделі афектності на основі нейронної мережі LSTM у методі визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів



Додаток Г

**Визначення показника афектності неймережею LSTM у методі
визначення поведінкових намірів клієнтів за семантичним аналізом їх
дописів**



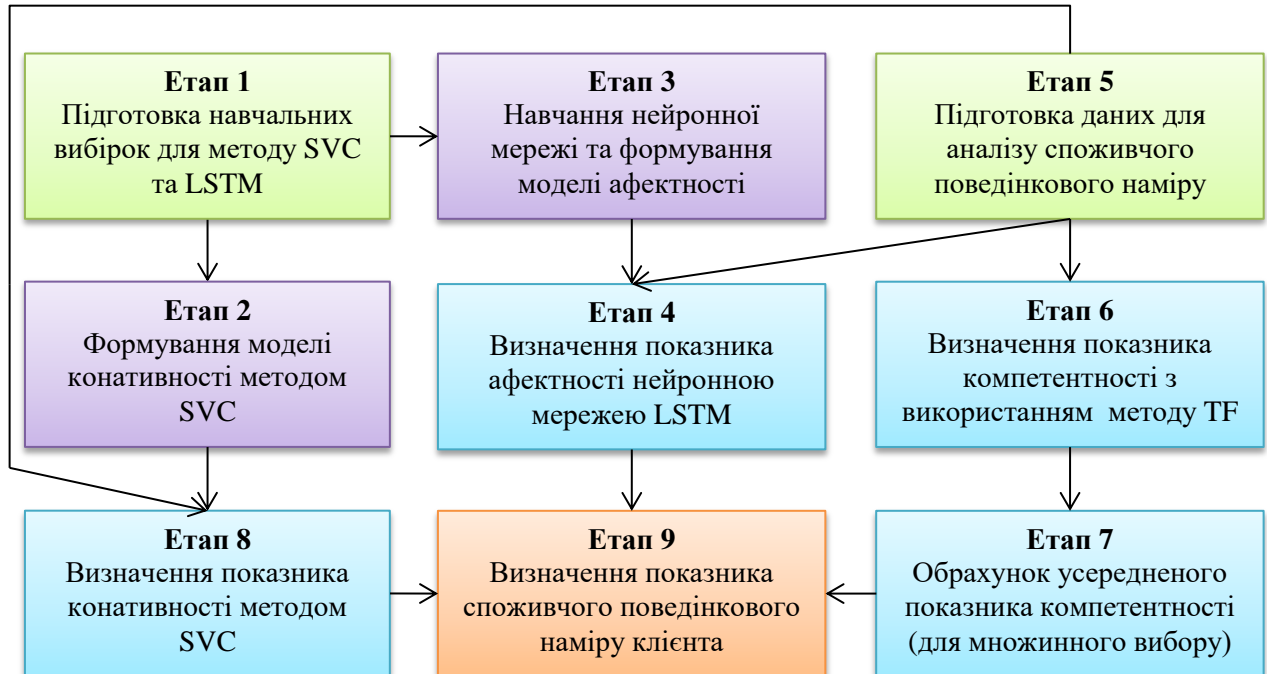
Додаток Д

Схема визначення показника компетентності у методі визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів



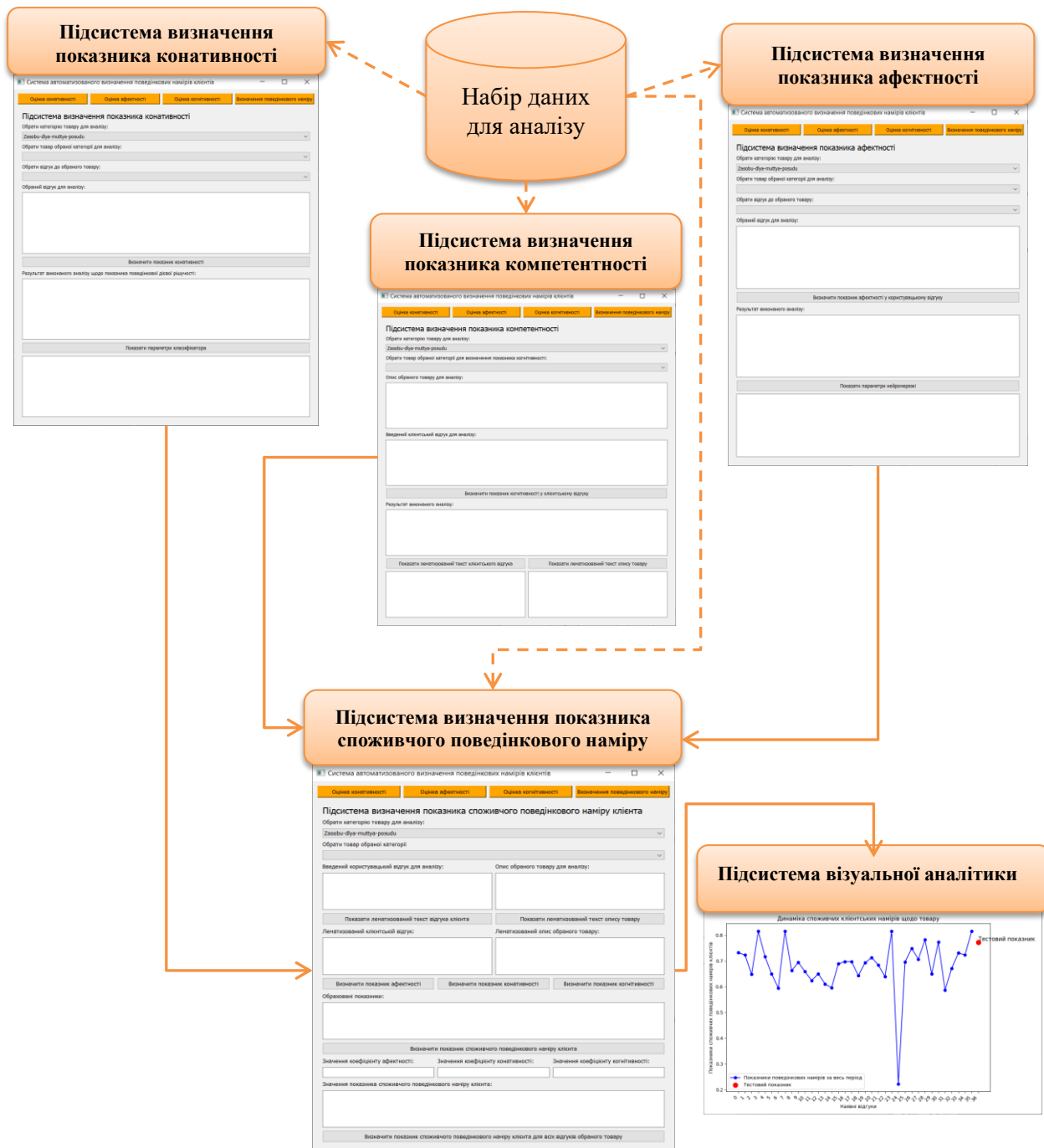
Додаток Е

Логічний зв'язок етапів визначення споживчого поведінкового наміру клієнта



Додаток Ж

Схема інформаційної системи автоматизованого визначення поведінкових намірів клієнтів



Додаток З

Світлини наукових публікацій, виконаних при роботі над кваліфікаційною роботою магістра

(ксерокопії титульної сторінки, сторінки змісту та всіх сторінок із публікацією)

Наукова публікація:

Залуцька О.О., Молчанова М.О., Мазурець О.В., Мельник О.І., Скрипник Т.К. Метод інтелектуального аналізу емоційної тональності текстової інформації для визначення поведінкових намірів нейромережевими засобами. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2023. №5 (325). Т.1. С. 67-73.

ISSN 2307-5732

DOI 10.31891/2307-5732

НАУКОВИЙ ЖУРНАЛ

5.2023

ВІСНИК

**Хмельницького
національного
університету**

Том 1

Технічні науки

Technical sciences

SCIENTIFIC JOURNAL

HERALD OF KHMELNYTSKYI NATIONAL UNIVERSITY

2023, Issue 5, Volume 325, Part 1

Хмельницький

ЗМІСТ

АНДРІЙЧУК ВОЛОДИМИР, НАКОНЕЧНИЙ МИРОСЛАВ, ФІЛЮК ЯРОСЛАВ, КОСТИК ЛЮБОВ, ОСАДЦА ЯРОСЛАВ ДОСЛІДЖЕННЯ КІНЕТИКИ СВІТЧЕННЯ СВІТЛОДІОДНИХ ДЖЕРЕЛ СВІТЛА	9
АФТАНАЗІВ І.С., СТРОГАН О.І., ШЕВЧУК Л.І., СТРУТИНСЬКА Л.Р. КІНЕМАТИЧНЕ ПРОЄКТУВАННЯ ЯК ЗАСІБ ВДОСКОНАЛЕННЯ ПОШУКУ МОРСЬКИХ МІН	16
БАБИН І.А., БУРЛАКА С.А., ХОЛОДЮК О.В. ЕФЕКТИВНІСТЬ ВИКОРИСТАННЯ СТРІЧКОВОЇ СУШАРКИ	26
БЕРДНІК Д., ПЕЛІШКО Д. ВІДНОВЛЕННЯ ЗОБРАЖЕНЬ ЗА ДОПОМОГОЮ ГЕНЕРАТИВНИХ НЕЙРОННИХ МЕРЕЖ	30
БЛАЖЕНКО МАРІЯ, ФАЛЕНДИШ НАТАЛІЯ ВИКОРИСТАННЯ ПРОДУКТІВ ПЕРЕРОБКИ НАСІННЯ КОНОПЕЛЬ У ВИРОБНИЦТВІ ХЛІБА	35
БОЙКО ЮЛІЙ, КАРПОВА ЛЕСЯ, СЕМЕНЮК ВІТАЛІЙ ВИСОКОРІВНЕВА ОДНОПЛОЩИННА АНТЕНА МІМО ДЛЯ ДЕВАЙСІВ 5G	40
БУРЕНКО В. О. АНАЛІЗ НАПОВНЕНОСТІ ЗУПИНОК ПАСАЖИРСЬКОГО ТРАНСПОРТУ ЗА ДОПОМОГОЮ АЛГОРИТМІВ ОБРОБКИ ЗОБРАЖЕНЬ З ІР-КАМЕР «РОЗУМНОГО МІСТА»	47
ГУРКОВСЬКА ОЛЕНА, АНДРЕЄВА ОЛЬГА ПОРІВНЯЛЬНЕ ОЦІНЮВАННЯ СИСТЕМ ЕКСПРЕСІ У ВИРОБНИЦТВІ РЕКОМБІНАНТНОГО ІНСУЛІНУ	53
ДОРОГІЙ Я. Ю., КОЛІСНІЧЕНКО В. Ю. ЗАСТОСУВАННЯ ЛОГУВАННЯ РІЗНИМИ УЧАСНИКАМИ БЛОКЧЕЙН-МЕРЕЖ ДЛЯ ДЕАНОНІМІЗАЦІЇ КІНЦЕВОГО КОРИСТУВАЧА	60
ЗАЛУЦЬКА ОЛЬГА, МОЛЧАНОВА МАРІНА, МАЗУРЕЦЬ ОЛЕКСАНДР, МЕЛЬНИК ОЛЕГ, СКРИПНИК ТЕТЯНА МЕТОД ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ЕМОЦІЙНОЇ ТОНАЛЬНОСТІ ТЕКСТОВОЇ ІНФОРМАЦІЇ ДЛЯ ВИЗНАЧЕННЯ ПОВЕДІНКОВИХ НАМІРІВ НЕЙРОМЕРЕЖЕВИМИ ЗАСОБАМИ	67
ЗАЛЮБОВСЬКИЙ МАРК, ПАНАСЮК ІГОР, КОШЕЛЬ ОЛЕКСАНДР ВИЗНАЧЕННЯ ЕКСТРЕМАЛЬНИХ ЗНАЧЕНЬ РЕАКЦІЙ У КІНЕМАТИЧНИХ ПАРАХ ГАЛТУВАЛЬНОЇ МАШИНИ, У ЯКІЙ ЄМНІСТЬ ЗДІЙСНЮЄ СКЛАДНИЙ ПРОСТОРОВИЙ РУХ	74
ЗОЛОТУХА Р.А., ГЛАЗУНОВА О.Г. РОЗРОБКА МАТЕМАТИЧНОЇ АЛГОРИТМУ ДЛЯ ПІДБОРУ КОМАНДИ В ІТ ПРОЄКТАХ	81
ІВАНШЕНА ТЕТЯНА, МАНДЗЮК ІГОР, ТРУХІНА ОКСАНА, ПЕКАРСЬКА ВАЛЕРІЯ ОЦІНКА ЖИТТЄВОГО ЦИКЛУ МАТЕРІАЛІВ ЯК ІНСТРУМЕНТ УДОСКОНАЛЕННЯ ПРОЦЕСІВ ЛЕГКОЇ ПРОМИСЛОВОСТІ ТА ВПРОВАДЖЕННЯ ПРИНЦИПІВ КРУГОВОЇ ЕКОНОМІКИ ВИРОБНИЦТВ	89
ІЛЛЯШ О.Е., БРЕДУН В.І. ОБҐРУНТУВАННЯ ВИБОРУ МІСЦЯ ДОСЛІДЖЕННЯ МОРФОЛОГІЧНОГО СКЛАДУ ПОБУТОВИХ ВІДХОДІВ	98
КАМІНСЬКИЙ РОМАН, ПШЕНИЧНИЙ ОЛЕКСАНДР, ХУДНІЙ АНДРІЙ РОЛЬ ТА ВИЗНАЧЕННЯ ФРАКТАЛЬНОЇ КОНСТАНТИ ЦИКЛІЧНОГО ЧАСОВОГО РЯДУ	103
КАМІНСЬКИЙ РОМАН, ПШЕНИЧНИЙ ОЛЕКСАНДР, ХУДНІЙ АНДРІЙ ФРАКТАЛЬНИЙ АНАЛІЗ ЦИКЛІЧНОГО ЧАСОВОГО РЯДУ ЧИСЕЛ ВОЛЬФА (ПОКАЗНИКА СОНЯЧНОЇ АКТИВНОСТІ)	108

ЗАЛУЦЬКА ОЛЬГАХмельницький національний університет
<https://orcid.org/0000-0003-1242-3548>
e-mail: zalutcka.olha@gmail.com**МОЛЧАНОВА МАРИНА**Хмельницький національний університет
<https://orcid.org/0000-0001-9810-936X>
e-mail: m.o.molchanova@gmail.com**МАЗУРЕЦЬ ОЛЕКСАНДР**Хмельницький національний університет
<https://orcid.org/0000-0002-8900-0650>
e-mail: olex.chomaz@gmail.com**МЕЛЬНИК ОЛЕГ**Хмельницький національний університет
<https://orcid.org/0009-0001-6367-8922>
e-mail: melnik.77oleg@gmail.com**СКРИПНИК ТЕТЯНА**Хмельницький національний університет
<https://orcid.org/0000-0002-8531-5348>
e-mail: tkskripnik1970@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, а той час як для аналізу коротких документів запропоновано використовувати її модифікацію, RoBERTa.

ZALUTSKA OLHA, MOLCHANOVA MARYNA,
MAZURETS OLEKSANDR, MELNYK OLEG, SKRYPNYK TETIANA
Kmelnytskyi 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 вводить 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 використовуватися 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 англійських твітів.

У роботі [6] показано, що сучасні маркетингові дослідження переважно покладалися 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, напрямком 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 суржиків, 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 Grawlee [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 покладалися 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

Варіанти 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 відних послідовностей і створення контекстуалізованих представлень слів у реченні.

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

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

понад 50 іншими мовами [14] та ембедінгу [15] автора Ukjae Jeong, та входить до складу моделей бібліотеки TensorFlow_hub мови Python. На базі цих моделей пропонується створити модель, що буде донавлено на вищеприказаній 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.

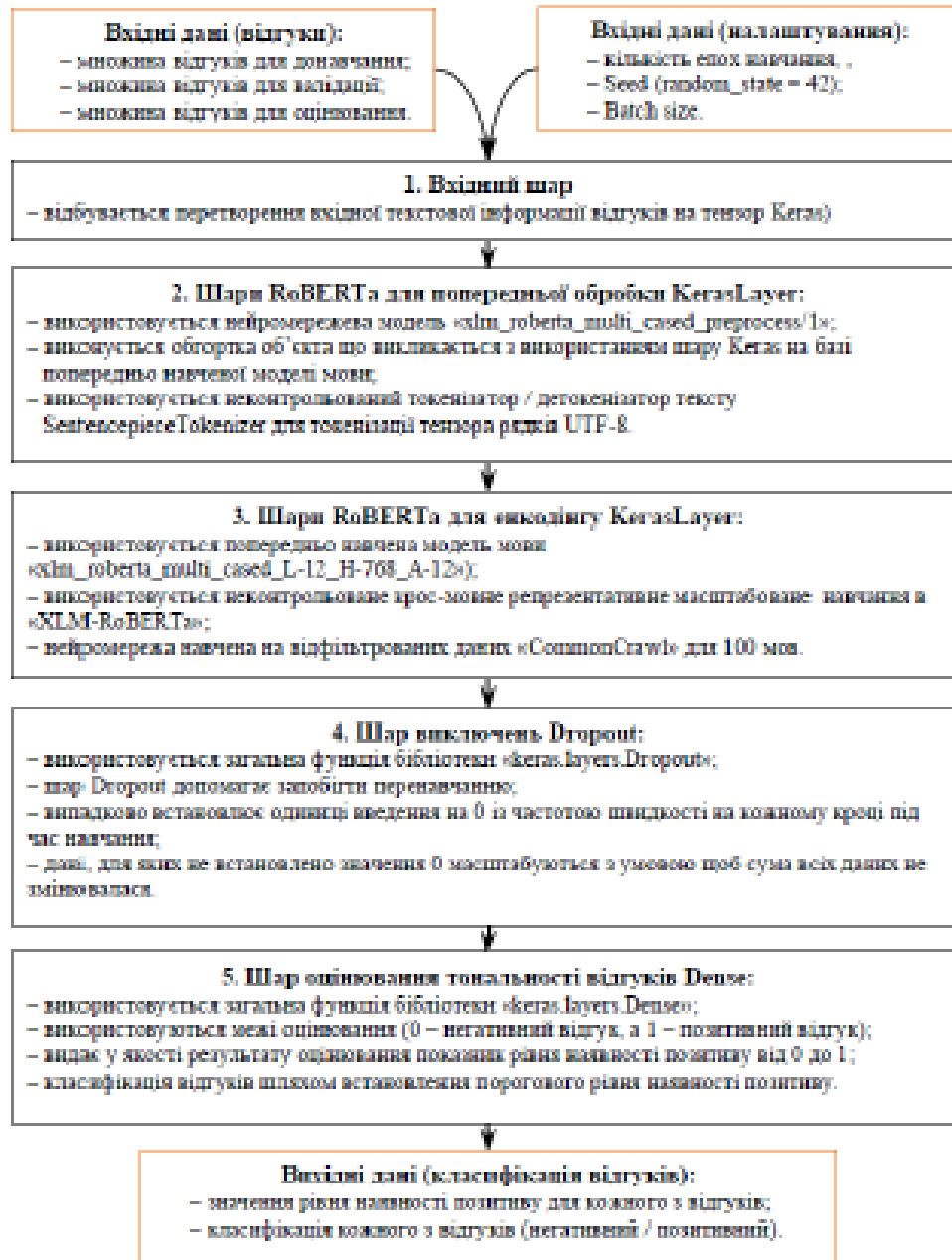


Рис. 3. Схема роботи класифікатора на основі RoBERTa для інтелектуального аналізу емоційної тональності текстової інформації

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 енкадер. 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-ентропійна 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, донавченням на білінгвістичних даних, в цілому нейромережа не має проблем з аналізом емоційної тональності текстової інформації.

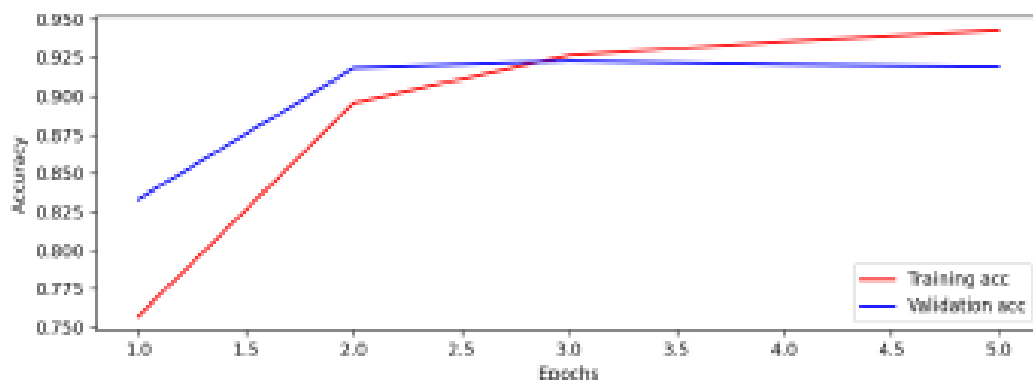


Рис. 4. Ілюстрація процесу навчання за епохами за показником точності

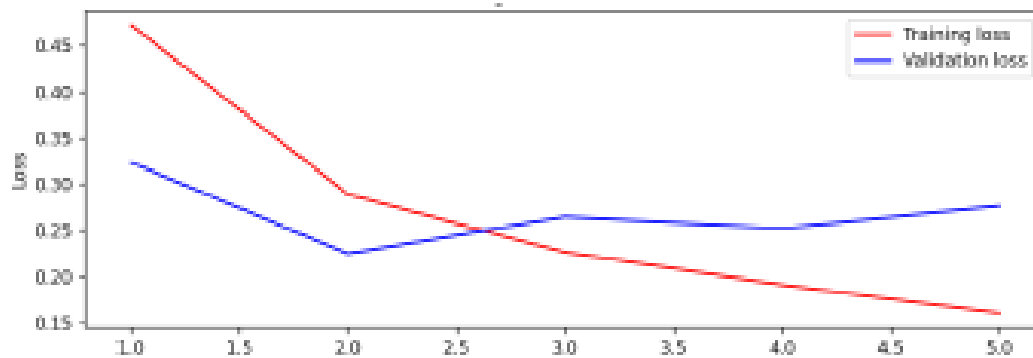


Рис. 5. Ілюстрація процесу навчання за словами за показником функції втрат

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 напрямку 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 вкладки 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 Cotae. 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., Сінцін 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., Слободян 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, 2023.

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., Daniel 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.
15. Tensorflow.Org., Text.SentencepieceTokenizer. 2023. 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.
17. Huggingface.Co. XLM-RoBERTa (base-sized model). URL: <https://huggingface.co/xlm-roberta-base>.
18. Zalutsk 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. Slobodczian 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, Shaosong 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 Marce, Robert Lkic. 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, Timofiy Oladumi, Max Denis, Esther Oosanya, Paul Cotae. 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. Identifikatsiia perevazhnoi moyi epilkuvannia liudyny. *Problemy programuvannia*. 2022. № 3–4. Spetsialnyi vypusk, s. 271–280. DOI: 10.15407/pp2022.03-04.271
8. Kovalchuk O.V., Slobodczian V.O., Mazurets O.V., Barmak O.V. Metod formuvannia binarnoho klasyfikatora ukrainomovnoho internet-kontentu. *Zbirnyk naukovykh prats za materialamy XIV Vsenkrainskoi naukovo-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., Daniel 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.
15. Tensorflow.Org., Text.SentencepieceTokenizer. 2023. 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.
17. Huggingface.Co. XLM-RoBERTa (base-sized model). URL: <https://huggingface.co/xlm-roberta-base>.
18. Zalutsk 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>.

Додаток И

Презентаційний матеріал

КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

Метод визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів

Виконав студент 2 курсу, гр. КНм-22-1
Мельник Олег Іванович

Керівник
ст.викладач каф.КН Скрипник Тетяна Казимирівна

Мета роботи

На сьогоднішній день спостерігається зростання кількості взаємодій між споживачами та інтернет-магазинами. З поглибленням цифрової трансформації в сфері роздрібної торгівлі зростає важливість розуміння поведінкових намірів клієнтів для покращення якості обслуговування, підвищення ефективності маркетингових стратегій та підтримки прийняття рішень.

Автоматизоване визначення поведінкових намірів через семантичний аналіз дописів клієнтів важливе для вирішення завдання персоналізації обслуговування, виявлення та аналізу тенденцій споживацького попиту та розробки інтелектуальних систем рекомендацій. Застосування такого методу у сфері відгуків щодо товарів щоденного вжитку дозволяє покращити якість обробки великого обсягу неструктурованих даних та забезпечити підтримку прийняття рішень відповідно до потреб споживачів.

Об'єкт дослідження - процес визначення поведінкових намірів клієнтів щодо готовності повторного придбання товару за дописом до товару.

Предмет дослідження - моделі, методи, алгоритми визначення поведінкових намірів клієнтів за семантичним аналізом.

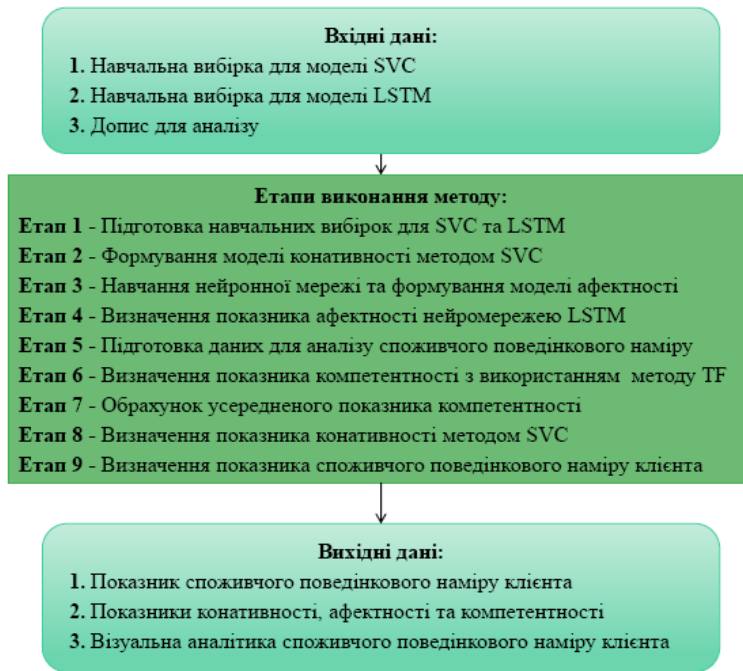
Мета роботи

Метою кваліфікаційної роботи магістра є вирішення задачі автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на прикладі споживчих поведінкових намірів клієнтів у межах предметної області відгуків щодо товарів щоденного вжитку в засобах електронної комерції, а також створення відповідної програмної реалізації. Для досягнення мети необхідно виконати такі завдання дослідження:

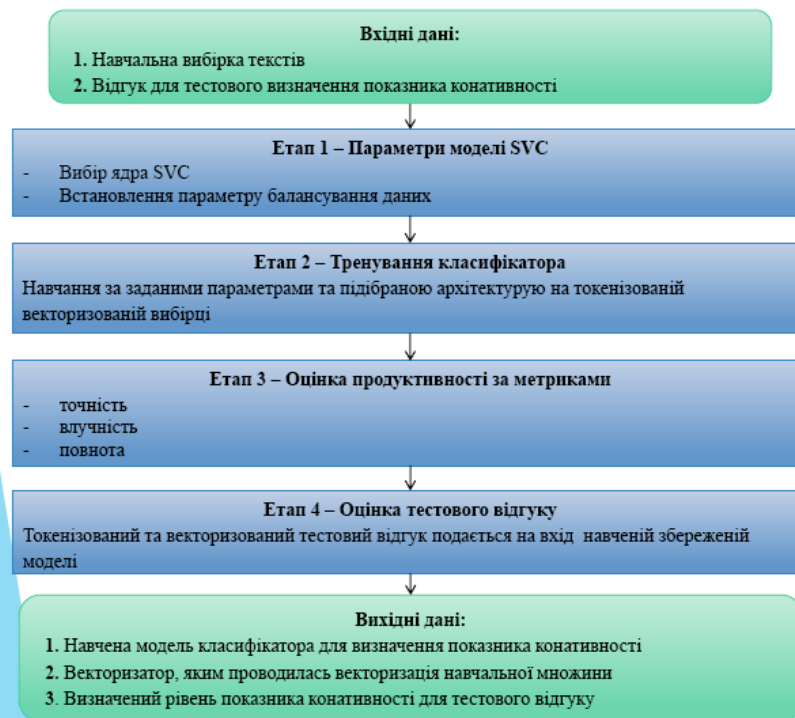
1. Провести аналіз предметної області щодо визначення поведінкових намірів клієнтів електронної комерції, виділити практичні підходи для автоматизації визначення поведінкових намірів.
2. Дослідити теоретичні підходи щодо автоматизації визначення поведінкових намірів за семантичним аналізом дописів та обрати комбінацію теоретичних підходів для використання в методи.
3. Створити метод автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів щодо товарів електронної комерції категорії повсякденного вжитку.
4. Спроекувати інформаційну систему на основі методу автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів щодо товарів електронної комерції.
5. Обрати засоби розробки для спроектованої архітектури інформаційної системи.
6. Розробити відповідну програмну реалізацію методу автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів щодо товарів електронної комерції.
7. Дослідити ефективність програмно реалізованого методу автоматизованого визначення поведінкових намірів клієнтів.

Компоненти показника споживчого поведінкового наміру клієнта





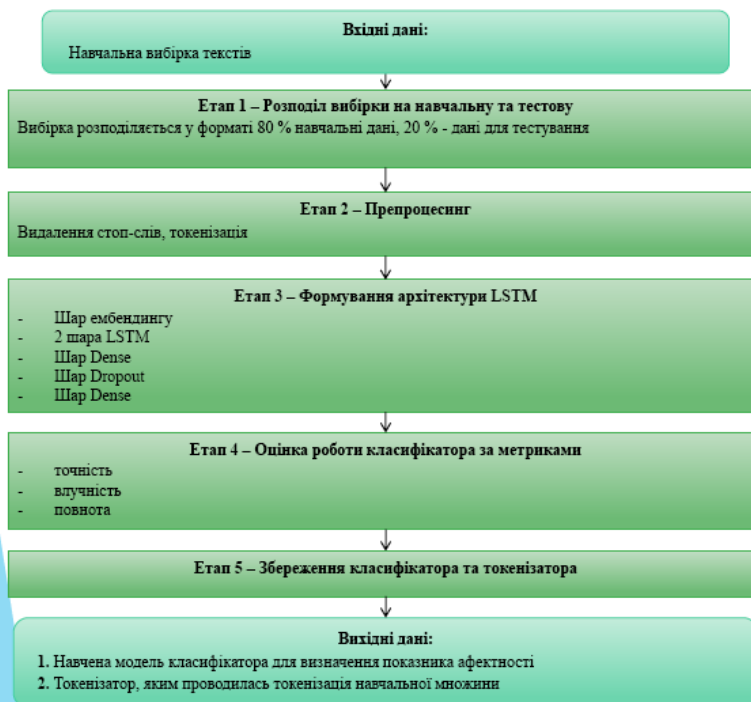
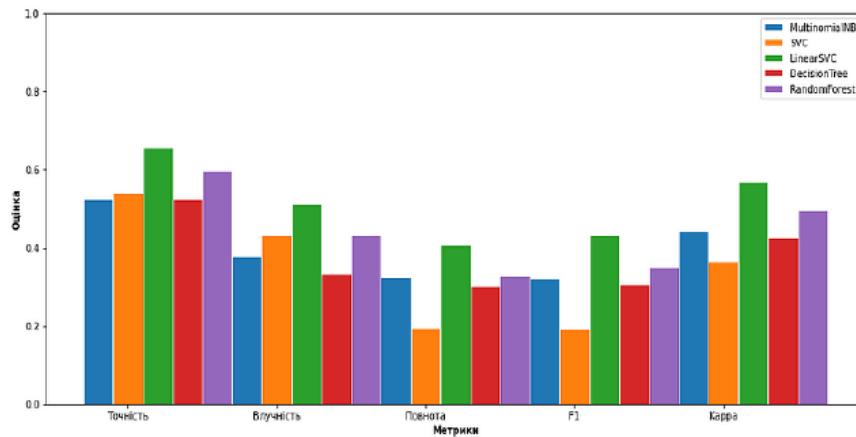
Узагальнена схема методу визначення поведінкових намірів клієнтів



Формування моделі конативності методом SVC у методі визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів

Використання різних видів класифікаторів для задачі класифікації текстів

Порівняння моделей



Формування моделі афектності на основі нейронної мережі LSTM у методі визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів

Визначення показника афектності нейромережею LSTM у методі визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів

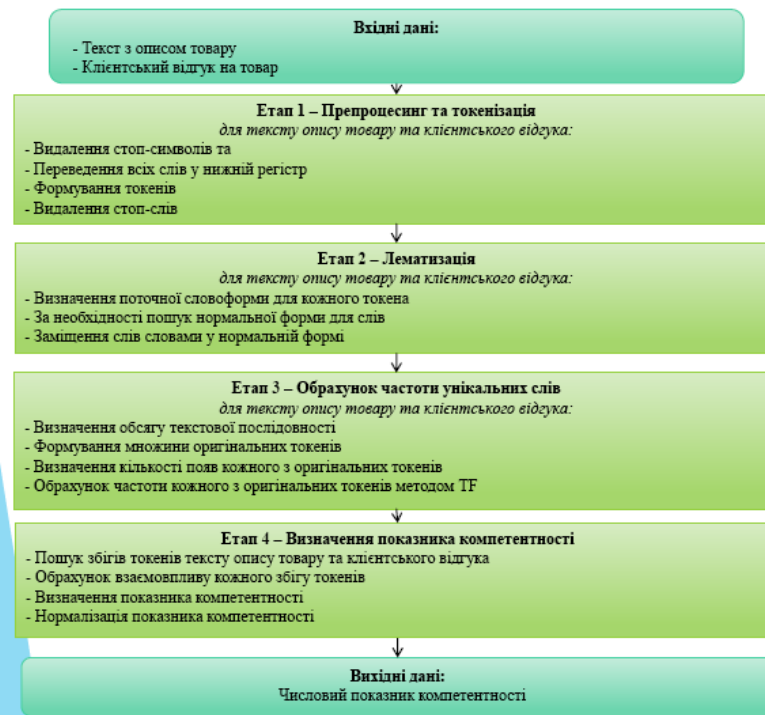
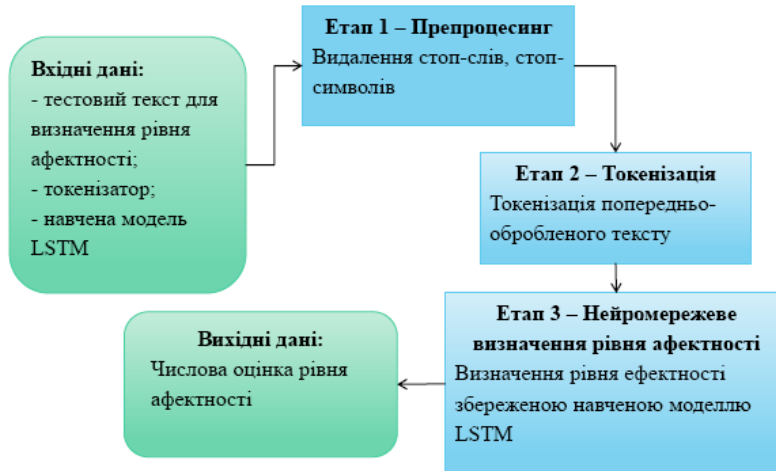


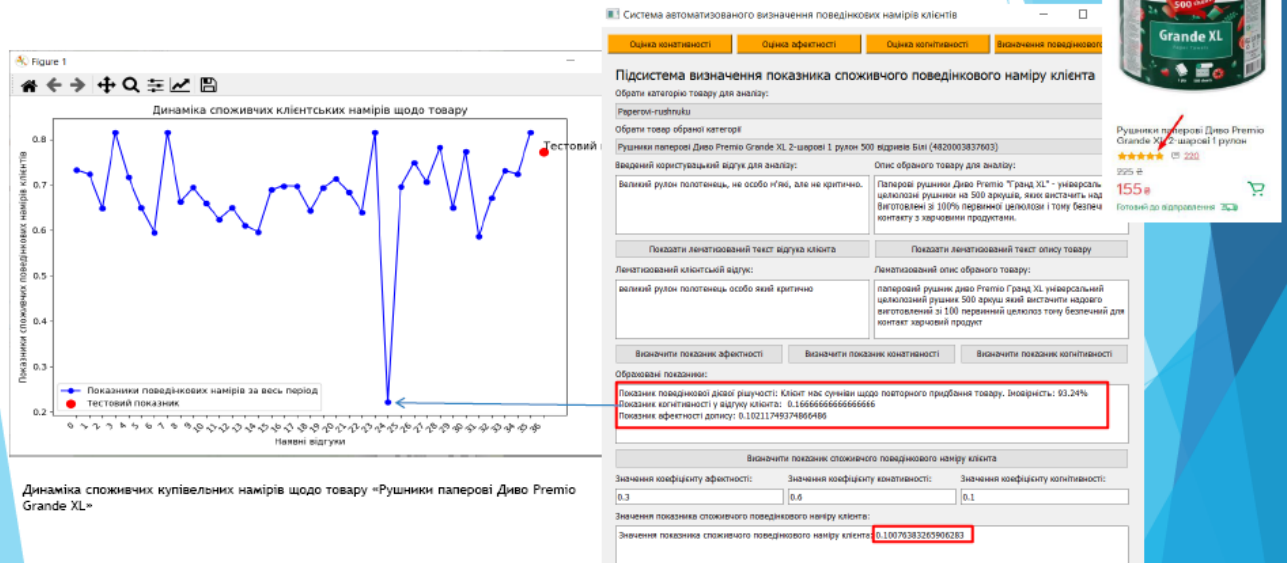
Схема визначення показника компетентності у методі визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів

Логічний зв'язок етапів визначення споживчого поведінкового наміру клієнта



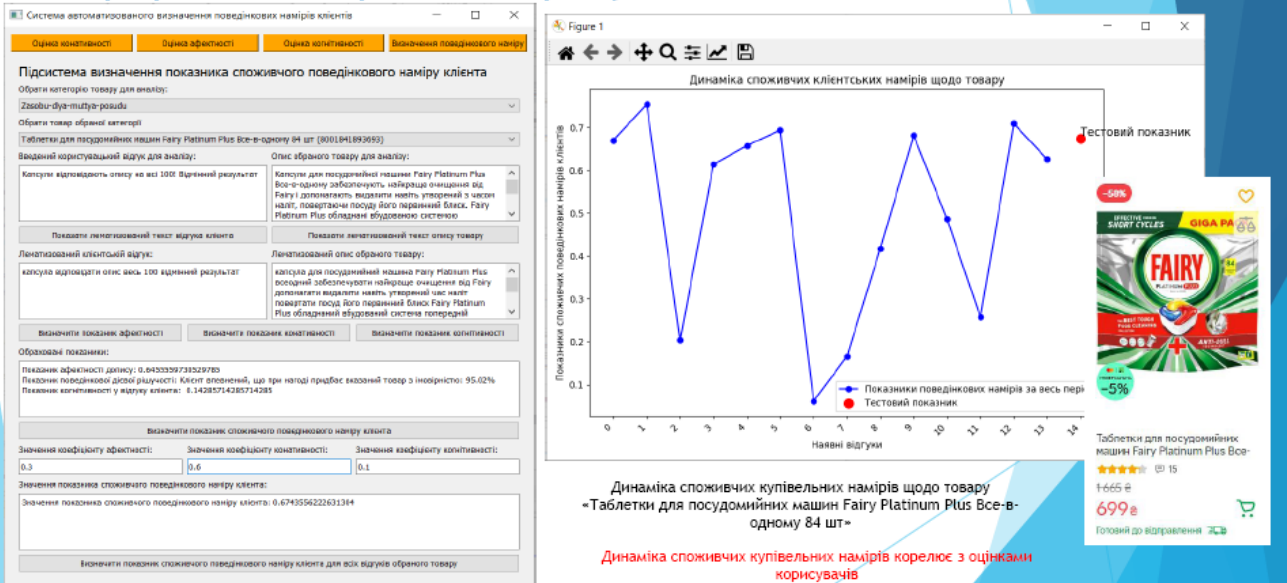
Схема інформаційної системи автоматизованого визначення поведінкових намірів клієнтів

Дослідження ефективності та інтерпретація отриманих результатів



Аналіз відгуку з невідповідністю до користувацьких оцінок

Дослідження ефективності та інтерпретація отриманих результатів



Динаміка споживчих купівельних намірів корелює з оцінками користувачів

Аналіз відгуку з показником споживчого поведінкового наміру вище середнього

Загальні висновки

На шляху до вирішення задачі автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на прикладі споживчих поведінкових намірів клієнтів у межах предметної області відгуків щодо товарів щоденного вжитку в засобах електронної комерції було виконано такі завдання дослідження:

1. Проведено аналіз предметної області щодо визначення поведінкових намірів клієнтів електронної комерції, виділено актуальні практичні підходи для автоматизації визначення поведінкових намірів.
2. Досліджено теоретичні підходи щодо автоматизації визначення поведінкових намірів за семантичним аналізом дописів та обрано комбінацію теоретичних підходів для використання в методі.
3. Розроблено метод автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на прикладі предметної області відгуків щодо товарів повсякденного вжитку.
4. Спроектовано інформаційну систему на основі методу автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом дописів.
5. Обрано засоби розробки для спроектованої архітектури інформаційної системи та розроблено відповідну програмну реалізацію методу автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на прикладі визначення споживчих поведінкових намірів клієнтів у межах предметної області відгуків щодо товарів щоденного вжитку в засобах електронної комерції.
6. Досліджено ефективність програмно реалізованого методу автоматизованого визначення поведінкових намірів клієнтів.

Загальні висновки

- ❖ Кваліфікаційна робота магістра розв'язує науково-технічну задачу автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів на прикладі предметної області відгуків щодо товарів щоденного вжитку.
- ❖ За виконання роботи були одержано результати, що містять інновації й наукову новизну, зокрема був розроблений метод автоматизованого визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів, що призначений для визначення споживчих поведінкових намірів клієнтів за семантичним аналізом коментарів товарів інтернет-магазину.
- ❖ Вхідними даними методу визначення споживчих поведінкових намірів клієнтів за семантичним аналізом їх дописів є навчальна вибірка для моделі класифікатора SVC для визначення прояву афектності за дописом та нейромережевої моделі LSTM для визначення прояву афектності, а також користувацький допис для аналізу. Одержана вихідна інформація є оцінкою поведінкового наміру схильності клієнта до купівлі обраного товару й додатково містить показники конативності, афектності та компетентності споживчого поведінкового наміру клієнта.
- ❖ Дослідження ефективності методу автоматизованого визначення поведінкових намірів клієнтів за використання його програмної реалізації виявило, що визначення поведінкових намірів клієнтів є ефективною практикою та допомагає забезпечити для клієнтів засобів електронної комерції підвищення довіри до відгуків, що дозволяє орієнтуватись на відгуки при здійсненні покупки, а для власників покращення зворотного зв'язку для підвищення ефективності процесу купівлі-продажу товарів і послуг. Автоматизоване визначення поведінкових намірів через семантичний аналіз дописів клієнтів важливе для вирішення завдання персоналізації обслуговування, виявлення та аналізу тенденцій споживацького попиту та розробки інтелектуальних систем рекомендації.
- ❖ За темою роботи автором у фаховому виданні виконано наукову публікацію, яка присвячена визначенню поведінкових намірів клієнтів нейромережевими засобами.



Ім'я користувача:
Кафедра КН

Дата перевірки:
12.12.2023 13:13:59 EET

Дата звіту:
12.12.2023 13:16:40 EET

ID перевірки:
1015997328

Тип перевірки:
Doc vs Internet + Library

ID користувача:
100005671

Назва документа: КНм-22-1 Мельник

Кількість сторінок: 86 Кількість слів: 13753 Кількість символів: 108266 Розмір файлу: 2.75 MB ID файлу: 101568029

Виявлено модифікації тексту (можуть впливати на відсоток схожості)

8.41% Схожість

Найбільша схожість: 2.2% з джерелом з Бібліотеки (ID файлу: 1015678871)

6.84% Джерела з Інтернету 687

Сторінка 88

4.74% Джерела з Бібліотеки 126

Сторінка 92

0% Цитат

Вилучення цитат вимкнене

Вилучення списку бібліографічних посилань вимкнене

0% Вилучень

Немає вилучених джерел

Модифікації

Виявлено модифікації тексту. Детальна інформація доступна в онлайн-звіті.

Підозріле форматування 27 сторінок

Anti-Plagiarism v-15.257

Максимальне співпадіння з одним документом 2.0%

Словники перевірки: en_US, ru_RU, ua_UA. Помилки в документах: 12%

ID: 122718 Назва: КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА на тему Метод визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів Додано в БД: 2023-12-12 Автора: О.І. Мельник Керівники: Т.К. Скрипник Консультанти: Опоненти:	Документ		Сумарний збіг по Базі Даних	
	Символи	Лексеми	Символи	Лексеми
	91927	1332	4411 (5%)	72 (5%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

РІШЕННЯ ЕКСПЕРНОЇ КОМІСІЇ
КАФЕДРИ КОМП'ЮТЕРНИХ НАУК
ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ МАГІСТРА ДО ЗАХИСТУ
ЗА РЕЗУЛЬТАТАМИ АНАЛІЗУ ЗВІТУ ПОДІБНОСТІ

Підтверджуємо ознайомлення з результатом звіту подібності щодо роботи, генерованого системою виявлення текстових збігів/ідентичності/схожості:

Назва: Метод визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів

Автор: Мельник Олег Іванович

Спеціальність: 122 – Комп'ютерні науки

Освітня програма: освітньо-професійна

Науковий керівник: старший викладач кафедри КН Скрипник Тетяна Казимирівна

Після аналізу звіту подібності зроблено такий висновок:

№	Висновок	Позначка про відповідність
1	Запозичення, виявлені в роботі, є законними і не є плагіатом. Робота приймається до захисту.	відповідає
2	Виявлені запозичення не є плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи. Робота приймається до захисту, але має бути відкоригована. Відкоригований варіант має бути поданий на кафедру за 2 дні до захисту, разом із заявою щодо самостійності виконання письмової роботи та ідентичності друкованої та електронної версії роботи	—
3	Виявлені запозичення не є плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. В зв'язку з цим мета роботи та поставлені завдання не були досягнені. Робота може бути допущена до захисту (наступного року) після того як буде відкоригована та допрацьована і успішно пройде повторну перевірку на академічний плагіат.	—
4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	—


Підтвердження:

Запозичення, виявлені в роботі, є законними і не є плагіатом, оскільки:

- 1) За програмою Anti-Plagiarism виявлені 2%, які є фрагментарними – містять поширені конструкції, загальновідомі терміни та визначення.
- 2) За програмою UNICHECK виявлені 8,41%, які є фрагментарними на джерело – містять поширені конструкції, загальновідомі терміни та визначення.

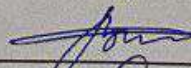
Сумарний обсяг всіх запозичень, визначений системою виявлення збігів/ідентичності/схожості, складає 2% і 8,41% відповідно, що, з урахуванням наведених обґрунтувань, відповідає характеру наукового дослідження і свідчить на користь кваліфікаційної роботи.

Керівник роботи




Тетяна СКРИПНИК

Гарант ОП



Руслан БАГРІЙ

Завідувач кафедри КН



Олександр БАРМАК



ВІДГУК НАУКОВОГО КЕРІВНИКА

на кваліфікаційну роботу магістра

гр. *КНм-22-1 Мельника Олега Івановича* за темою: *Метод визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів*

1. Актуальність обраної теми

З поглибленням цифрової трансформації в сфері роздрібної торгівлі зростає важливість розуміння поведінкових намірів клієнтів для покращення якості обслуговування, підвищення ефективності маркетингових стратегій та підтримки прийняття рішень. Автоматизоване визначення поведінкових намірів через семантичний аналіз дописів клієнтів важливе для вирішення завдання персоналізації обслуговування, виявлення та аналізу тенденцій споживацького попиту та розробки інтелектуальних систем рекомендацій. Застосування такого методу у сфері відгуків щодо товарів щоденного вжитку дозволяє покращити якість обробки великого обсягу неструктурованих даних та забезпечити підтримку прийняття рішень відповідно до потреб споживачів, тому тема кваліфікаційної роботи є актуальною.

2. Відповідність роботи предметній області спеціальності 122 Комп'ютерні науки та загальним вимогам до наукових робіт

Поставлена у кваліфікаційній роботі магістра мета, пов'язана з створенням методу визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів, повною мірою відповідає предметній області спеціальності 122 «Комп'ютерні науки» та вимогам до кваліфікаційної роботи.

3. Професійні та особистісні якості магістранта

Під час виконання своєї кваліфікаційної роботи на ступінь магістра, Мельник Олег Іванович виявив високі професійні якості, проявивши себе як дисциплінований та висококваліфікований фахівець. Він з великою увагою підійшов до кожного завдання, забезпечуючи їх виконання на високому рівні якості, з дотриманням встановлених термінів.

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Результати, отримані за виконання кваліфікаційної роботи бакалавра, є результатом самостійної діяльності студента. Отримані положення наукової новизни та інновації, описані в роботі, дозволили покращити існуючі методи в галузі визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів

5. Наукова новизна та оригінальність запропонованих підходів

Результати кваліфікаційної роботи магістра включають інноваційний метод для автоматизованого визначення поведінкових намірів клієнтів через семантичний аналіз їх дописів на прикладі коментарів у інтернет-магазинах. Цей метод, використовуючи рекурентну нейромережу та семантичний аналіз коментарів, дозволяє оцінити когнітивні, афективні та конативні аспекти споживчих намірів.

6. Ступінь оволодіння методами дослідження

Магістрант виявив високий ступінь оволодіння необхідними методами дослідження.

7. Повнота та якість розкриття теми роботи

Тема роботи в повній мірі обґрунтована й розкрита, проведено аналіз актуальності та відомих досліджень в межах обраної теми, поставлені завдання у роботі виконані, а також проведено аналіз результатів прикладного застосування розробленого методу визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів

8. Логічність, послідовність, аргументованість, літературна грамотність викладу матеріалу


Структура роботи й послідовність викладення логічні та відповідні поставленій меті. Викладення матеріалу грамотне та виявляє високий ступінь відповідності стилю.

9. Можливість практичного застосування кваліфікаційної роботи, окремих її частин

Було створено інформаційну систему автоматизованого визначення поведінкових намірів клієнтів, яка є прикладною програмною реалізацією метода визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів і використовувалась для проведення експериментальних досліджень з метою апробації запропонованого метода. Проведені дослідження ефективності свідчать про те, що даний підхід можна використовувати для аналізу маркетингової ситуації шляхом визначення для клієнтів та власників узагальнених та локальних споживчих поведінкових намірів щодо придбання окремих товарних позицій за клієнтськими дописами.

10. Висновок про можливість допуску кваліфікаційної роботи до захисту, на яку оцінку заслуговує робота

Враховуючи високий рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка «відмінно».

Науковий керівник _____  ст. викладач каф. КН Скрипник Тетяна Казимирівна



ВІДГУК ОПОНЕНТА

на кваліфікаційну роботу магістра

гр. *КНм-22-1 Мельника Олега Івановича* за темою: Метод визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів.

1. Актуальність обраної теми

Автоматизований семантичний аналіз дописів клієнтів стає ключовим для персоналізації сервісу, аналізу трендів споживчого попиту та розробки систем рекомендацій. Адже на сучасному етапі спостерігається збільшення взаємодій між споживачами та онлайн-магазинами, особливо у контексті цифрової трансформації у роздрібній торгівлі. Це підкреслює необхідність розуміння клієнтських намірів для оптимізації обслуговування та маркетингу. Такий підхід у сфері відгуків на товари дозволяє ефективніше обробляти великі обсяги неструктурованих даних і відповідати на потреби споживачів. Відповідно, тема «Метод визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів» є досить актуальною, тому кваліфікаційна робота має значну наукову та практичну цінність.

2. Відповідність роботи предметній області спеціальності 122 Комп'ютерні науки та загальним вимогам до наукових робіт

Обрана тема «Метод визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів», в межах якої виконані поставлені задачі, повною мірою відповідає предметній області спеціальності 122 «Комп'ютерні науки» та вимогам до кваліфікаційної роботи магістра.

3. Повнота розкриття мети та завдань дослідження

В роботі автор повністю розкриває мету дослідження та поставленні в межах теми завдання.

4. Наявність наукової новизни

У кваліфікаційній роботі магістра розроблено інноваційний метод автоматизованого аналізу поведінкових намірів клієнтів через семантичний аналіз коментарів у інтернет-магазинах. Цей метод використовує рекурентну нейромережу для перетворення текстових даних у оцінку споживчих намірів, враховуючи когнітивні, афективні та конативні компоненти. Як наслідок, дослідження сприяє підвищенню ефективності електронної

комерції, забезпечуючи клієнтам більшу довіру до відгуків та власникам – кращий зворотний зв'язок для оптимізації процесу купівлі-продажу.

5. Зміст кожного розділу роботи

Робота містить чотири розділи. У першому розділі виконано аналіз сучасного стану області визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів. Другий розділ присвячено розробці методу визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів та компонентів методу. У третьому розділі виконано розробку прикладного програмного застосунку на базі розробленого методу. У четвертому розділі виконано дослідження ефективності методу визначення поведінкових намірів клієнтів за семантичним аналізом їх дописів

6. Ступінь розкриття теми роботи

Тема кваліфікаційної роботи повною мірою розкрита та обґрунтована, проведено аналіз актуальності та відомих досліджень в межах обраної теми, поставлені завдання, які у роботі виконані, та проведено аналіз результатів прикладного застосування розроблених методу і засобів.

7. Якість оформлення кваліфікаційної роботи

Оформлення роботи цілком відповідає необхідним нормам та вимогам, які ставляться до оформлення кваліфікаційних робіт.

8. Недоліки кваліфікаційної роботи

Перелік скорочень, поданий в кваліфікаційній роботі наведено не повною мірою, рекомендовано доповнити його іншими скороченнями, що зустрічається в роботі. Не розглянуто перспективи прикладного використання компонентів розробленого методу, які поокремо теж мають наукову та практичну цінність.

9. Загальний висновок (допускається чи не допускається до захисту), якої оцінки заслуговує кваліфікаційна робота

Враховуючи високий рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка відмінно.

Опонент

Гимельс О.В., доц. кафедри ІТІЗ