

**БАКАЛАВРСЬКА РОБОТА**

**БР. ІІ - 40.00.00.000 ІІЗ**

**Група ІІ-23-1К**

**Дмитрук Ярема**

**2025**

**Івано-Франківський національний технічний університет нафти і газу**

Факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

**Дмитрук Ярема Володимирович**

(прізвище, ім'я, по батькові)

УДК 004  
(індекс)

## **БАКАЛАВРСЬКА РОБОТА**

**Застосування методів машинного навчання для обробки даних з user**

**feedback**

(назва роботи)

**Інженерія програмного забезпечення**

(назва освітньої програми)

**121 - Інженерія програмного забезпечення**

(шифр і назва спеціальності)

**Робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело**

Здобувач освітнього рівня Дмитрук Я.В.  
(підпис, ініціали та прізвище здобувача)

Науковий керівник Піх Володимир Ярославович, к.т.н., доцент  
(підпис, прізвище, ім'я, по батькові, науковий ступінь, вчене звання керівника)

Допущено до захисту  
Завідувач кафедри

доц. Бандура В.В.  
(посада) (підпис) (дата) (ініціали та прізвище)

**Івано-Франківськ – 2025**

**Івано-Франківський національний технічний університет нафти і газу**

Інститут, факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Ступінь вищої освіти бакалавр

Спеціальність 121 – Інженерія програмного забезпечення

**ЗАТВЕРДЖУЮ:**

Зав. кафедрою ІІЗ

доц.

В.В. Бандура

“     ”     2025 р.

## **ЗАВДАННЯ**

### **НА БАКАЛАВРСЬКУ РОБОТУ СТУДЕНТОВІ**

**Дмитруку Яремі Володимировичу**

(прізвище, ім'я, по-батькові)

**1. Тема проекту (роботи) “Застосування методів машинного навчання для обробки даних з user feedback”**

керівник проекту (роботи) Піх В.Я., доцент

затвержені наказом закладу вищої освіти від “ 28 ” квітня 2025 р. № 264/7

**2. Строк подання студентом проекту (роботи) 10 червня 2025 р.**

**3. Вихідні дані до проекту (роботи) Результати і матеріали отримані під час проходження переддипломної практики**

**4. Зміст розрахунково - пояснювальної записки (перелік питань, які потрібно розробити)**

1. Аналіз предметної області використання машинного навчання класифікації відгуків

2. Проектування архітектури системи обробки даних з user feedback

3. Моделювання системи з використанням діаграми варіантів використання

4. Програмна реалізація системи обробки даних з user feedback на основі машинного навчання

5. Виконання процесу тестування

**5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)**

1. Основні способи збору user feedback (рис. 1.1)

2. Алгоритм онтологічного методу (рис. 1.2)

3. Переклад термінів однієї предметної області термінами іншої предметної області (рис. 1.3)

4. Сервісно-орієнтована архітектура аналізу тональності (рис. 1.4)

5. Дашборд платформи Brandwatch (рис. 1.5)

## 6. Консультанти розділів проекту (роботи)

Розділ	Консультант	Підпис, дата	
		Завдання видав	Завдання прийняв

7. Дата видачі завдання 28 квітня 2025 р.

Керівник \_\_\_\_\_

(підпис)

Завдання прийняв до виконання \_\_\_\_\_

(підпис)

## КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту (роботи)	Строк виконання етапів проекту	Примітка
1	Аналіз предметної області використання машинного навчання класифікації відгуків	03.05.2025	виконано
2	Проектування архітектури системи обробки даних з user feedback	10.05.2025	виконано
3	Моделювання системи з використанням діаграми варіантів використання	22.05.2025	виконано
4	Програмна реалізація системи обробки даних з user feedback на основі машинного навчання	02.06.2025	виконано
5	Виконання процесу тестування	07.06.2025	виконано
6	Оформлення пояснювальної записки дипломної роботи завідувачем кафедри	10.06.2025	виконано

Студент – дипломник \_\_\_\_\_

(підпис)

Керівник роботи \_\_\_\_\_

(підпис)

## АНОТАЦІЯ

Бакалаврська робота містить 77 сторінок, 19 рисунків, список використаних джерел із 39 найменуваннями, 1 додаток.

**Мета роботи:** розробка програмної системи для автоматизованої обробки та класифікації користувацьких відгуків із використанням методів машинного навчання

**Об'єкт дослідження:** процес автоматизованої обробки текстових відгуків користувачів у цифрових системах

**Предмет дослідження:** методи машинного навчання, які застосовуються для аналізу та класифікації тональності користувацьких текстових даних

**Результати дослідження:** розроблено програмну архітектуру системи, реалізовано її функціональні модулі.

**В першому розділі** проведено системний аналіз предметної області, вивчено сучасні підходи до обробки відгуків та визначено обмеження існуючих рішень.

**В другому розділі** розроблено архітектуру програмної системи, побудовано діаграми UML та визначено функціональні компоненти з урахуванням вимог до апаратного і програмного забезпечення..

**В третьому розділі** реалізовано програмну систему з модулями аналізу тональності, проведено експериментальне тестування, яке підтвердило ефективність обраних алгоритмів машинного навчання.

**Висновок:** Реалізовано інтеграцію класичних методів машинного навчання у розроблену систему з аналізу тональності user feedback.

**КЛЮЧОВІ СЛОВА:** АНАЛІЗ ТОНАЛЬНОСТІ, ВІДГУКИ КОРИСТУВАЧІВ, МАШИННЕ НАВЧАННЯ, МЕТОД ОПОРНИХ ВЕКТОРІВ, НАЇВНИЙ БАЙЄС, SENTIWORDNET, ОБРОБКА ПРИРОДНОЇ МОВИ

## ANNOTATION

The Bachelor's thesis comprises 77 pages, 19 figures, a list of 39 references, and 1 appendix.

**Purpose of the work:** Development of a software system for automated processing and classification of user reviews using machine learning methods.

**Object of research:** The process of automated processing of user text reviews in digital systems.

**Subject of research:** Machine learning methods applied for analyzing and classifying the sentiment of user textual data.

**Research results:** A software system architecture was developed, and its functional modules were implemented.

**The first chapter** conducted a systemic analysis of the subject domain, examined modern approaches to review processing, and identified limitations of existing solutions.

**The second chapter** developed the software system architecture, constructed UML diagrams, and defined functional components considering hardware and software requirements.

**The third chapter** implemented the software system with sentiment analysis modules and conducted experimental testing, which confirmed the effectiveness of the chosen machine learning algorithms.

**Conclusion:** The integration of classical machine learning methods into the developed system for user feedback sentiment analysis has been successfully implemented.

**KEYWORDS:** SENTIMENT ANALYSIS, USER REVIEWS, MACHINE LEARNING, SUPPORT VECTOR MACHINE, NAIVE BAYES, SENTIWORDNET, NATURAL LANGUAGE PROCESSING

## ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ .....	9
ВСТУП .....	10
РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ВИКОРИСТАННЯ МАШИННОГО НАВЧАННЯ КЛАСИФІКАЦІЇ ВІДГУКІВ КОРИСТУВАЧІВ .....	13
1.1. Основні акценти розробки системи класифікації user feedback із застосування машинного навчання.....	13
1.1.1. Типи та методи збору user feedback .....	14
1.1.2. Методологія аналізу та оцінки user feedback .....	15
1.1.3. Застосування алгоритмів машинного навчання .....	15
1.2. Проблематика та обмеження існуючих підходів до аналізу онлайн- відгуків.....	16
1.2.1. Аналіз існуючих підходів та обмежень SentiWordNet .....	17
1.3. Концепція аналізу тональності user feedback соціальних мереж з використанням онтології SentiWordNet .....	19
1.3.1. Групи методів для аналізу тональності текстових даних .....	19
1.3.2. Архітектура програмного забезпечення для аналізу тональності в соціальних мережах .....	23
1.4. Аналіз спеціалізованих SaaS-платформ для аналізу відгуків/соціальних медіа .....	25
1.4.1. Платформа Brandwatch .....	25
1.4.2. Talkwalker .....	27
1.4.3. Qualtrics (Customer Experience Management) .....	29

					<b>БР.ІІ – 40.00.00.000 ПЗ</b>			
Змн.	Арк.	№ докум.	Підпис	Дата				
Розроб.		Дмитрук Я.В.			Застосування методів машинного навчання для обробки даних з user feedback  <b>Пояснювальна записка</b>	Літ.	Арк.	Акрушіє
Перевір.		Піх В.Я.					6	
Реценз.						<b>ІФНТУНГ ІІІ-23-1К</b>		
Н. Контр.		Піх М.М.						
Затверд.		Бандура В.В.						

## РОЗДІЛ 2. ПРОЕКТУВАННЯ АРХІТЕКТУРИ СИСТЕМИ ОБРОБКИ

ДАНИХ З USER FEEDBACK .....	32
2.1. Опис функціональності запропонованої системи для аналізу тональності user feedback .....	32
2.2. Специфікації вимог до системи .....	34
2.2.1. Апаратні вимоги .....	34
2.2.2. Програмні вимоги.....	34
2.3. Опис використаних програмних засобів та алгоритмів.....	35
2.3.1. Графічний інтерфейс користувача (GUI).....	35
2.3.2. Метод опорних векторів (SVM).....	36
2.4. Моделювання системи з використанням діаграми варіантів використання .....	40
2.5. Розробка діаграми послідовності .....	42
2.6. Проектування діаграми потоків даних (DFD) .....	45
2.7. Розробка діаграми активності .....	48

## РОЗДІЛ 3. ПРОГРАМНА РЕАЛІЗАЦІЯ СИСТЕМИ ОБРОБКИ ДАНИХ З USER FEEDBACK З ВИКОРИСТАННЯМ МЕТОДІВ МАШИННОГО

НАВЧАННЯ.....	52
3.1. Розробка діаграми компонентів системи .....	52
3.2. Розробка інтерфейсу входу та головної сторінки системи .....	54
3.2.1. Екран входу (Login Screen).....	54
3.2.2. Головна сторінка (Main Page) .....	55
3.3. Опис функціональних модулів та експериментальних результатів .....	57
3.3.1. Модуль завантаження словника .....	57
3.3.2. Модуль аналізу тональності за допомогою методу опорних векторів (SVM).....	58
3.3.3. Модуль аналізу тональності за допомогою Наївного Байєса .....	60
3.4. Результати експериментів та аналіз ефективності алгоритмів .....	62

3.4.1. Матриця плутанини (Confusion Matrix) .....	62
3.4.2. Точність SVM (SVM Accuracy) .....	63
3.4.3. Точність Наївного Байєса (Naive Bayes Accuracy) .....	65
3.4.4. Діаграма правильно класифікованих (Correctly Classified Chart) ...	67
3.5. Тестування системи .....	68
3.5.1. Тестування прийняття користувачем (UAT) .....	68

ВИСНОВКИ.....	72
---------------	----

ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ .....	74
---------------------------------------	----

ДОДАТКИ

БІБЛІОГРАФІЧНА ДОВІДКА

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						8
Змн.	Арк.	№ докум.	Підпис	Дата		

## ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

AWT – Abstract Window Toolkit (Абстрактний інструментарій вікон)

CX – Customer Experience (Клієнтський досвід)

CXM – Customer Experience Management (Управління клієнтським досвідом)

DFD – Data Flow Diagram (Діаграма потоків даних)

JFC – Java Foundation Classes (Базові класи Java Foundation)

ML – Machine Learning (Машинне навчання)

NPS – Net Promoter Score (Індекс чистої прихильності)

NLP – Natural Language Processing (Обробка природної мови)

SaaS – Software as a Service (Програмне забезпечення як послуга)

sp – Strongly Positive (Сильно-позитивний)

p – Positive (Позитивний)

wp – Weakly Positive (Слабо-позитивний)

sn – Strongly Negative (Сильно-негативний)

n – Negative (Негативний)

wn – Weakly Negative (Слабо-негативний)

nt – Neutral (Нейтральний) - виведений з контексту семи категорій тональності

SVM – Support Vector Machine (Метод опорних векторів)

t.n – Total Negative (Загальна негативна оцінка)

t.p – Total Positive (Загальна позитивна оцінка)

UAT – User Acceptance Testing (Тестування прийняття користувачем)

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						9
Змн.	Арк.	№ докум.	Підпис	Дата		

## ВСТУП

У сучасних умовах цифровізації та зростання ролі користувацького досвіду в розвитку бізнесу, соціальних платформ та сервісних систем, аналіз відгуків користувачів (user feedback) набуває вирішального значення. Користувацькі відгуки містять важливу інформацію про якість продуктів, послуг, взаємодію з інтерфейсами та очікування споживачів, тому їх своєчасна та коректна інтерпретація є стратегічно важливою для прийняття обґрунтованих рішень.

Однак, з огляду на великий обсяг та різноманітність даних, традиційні методи аналізу не завжди здатні забезпечити ефективну обробку текстової інформації. У зв'язку з цим, дедалі ширше застосовуються методи машинного навчання, які дозволяють автоматизувати процеси класифікації, виявлення тональності, тематичного групування та фільтрації інформації. Застосування таких алгоритмів як SVM, Наївний Байєс, дерева рішень тощо, відкриває нові можливості для побудови адаптивних систем, здатних до самонавчання та гнучкої обробки текстових даних.

Дана дипломна робота присвячена розробці та реалізації програмного рішення для автоматизованого аналізу тональності користувацьких відгуків з використанням алгоритмів машинного навчання. У роботі досліджено існуючі інструменти аналізу відгуків, здійснено критичний огляд методів обробки текстів, розроблено програмну архітектуру системи, реалізовано її функціональні модулі та проведено експериментальне тестування із застосуванням реальних даних.

### **Актуальність роботи**

Зростаючі обсяги цифрових даних, що генеруються користувачами у вигляді текстових відгуків, потребують ефективних інструментів для їх інтерпретації. Методи машинного навчання забезпечують точну й автоматизовану обробку текстових даних, що робить цю тему надзвичайно

					БР.ІП – 40.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		10

актуальною для бізнесу, маркетингу, соціології, UX-дизайну та розробки програмного забезпечення.

**Метою дипломної роботи** є розробка програмної системи для автоматизованої обробки та класифікації користувацьких відгуків із використанням методів машинного навчання для аналізу тональності текстів.

### **Завдання дослідження**

1. Провести огляд існуючих підходів до збору та обробки user feedback.
2. Дослідити алгоритми машинного навчання, що використовуються для класифікації текстових даних.
3. Розробити архітектуру програмного забезпечення для аналізу тональності.
4. Реалізувати основні модулі системи на основі методів SVM та Наївного Байєса.
5. Провести експериментальну оцінку ефективності реалізованих алгоритмів.
6. Здійснити тестування програмної системи та оцінити її функціональність.

**Об'єкт дослідження** - процес автоматизованої обробки текстових відгуків користувачів у цифрових системах.

**Предмет дослідження** - методи машинного навчання, які застосовуються для аналізу та класифікації тональності користувацьких текстових даних.

### **Методи дослідження**

У роботі застосовувалися методи:

- машинного навчання (SVM, Наївний Байєс);
- обробки природної мови (NLP);
- моделювання програмного забезпечення (UML);
- статистичного аналізу результатів експериментів;
- структурного та порівняльного аналізу існуючих систем.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		11

### **Наукова новизна**

Запропоновано та реалізовано інтеграцію класичних методів машинного навчання (SVM, Наївний Байєс) у розроблену систему з аналізу тональності user feedback з можливістю візуалізації результатів класифікації та адаптації до різних джерел даних, що підвищує гнучкість і точність обробки текстової інформації.

### **Практичне застосування**

Розроблена система може бути використана в бізнес-аналітиці, управлінні клієнтським досвідом (СЕМ), маркетингових дослідженнях, соціальних платформах та у сфері підтримки користувачів для автоматизованого виявлення проблем та аналізу реакцій споживачів у реальному часі.

Бакалаврська робота містить 77 сторінок, 19 рисунків, 3 розділи список використаних джерел із 39 найменуванням, 1 додаток.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						12
Змн.	Арк.	№ докум.	Підпис	Дата		

# РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ВИКОРИСТАННЯ МАШИННОГО НАВЧАННЯ КЛАСИФІКАЦІЇ ВІДГУКІВ КОРИСТУВАЧІВ

## 1.1. Основні акценти розробки системи класифікації user feedback із застосування машинного навчання

У сучасних умовах стрімкого технологічного розвитку та збільшення обсягів електронної комерції, онлайн-відгуки користувачів (user feedback) відіграють ключову роль у процесі прийняття рішень споживачами щодо придбання товарів та послуг. User feedback (зворотний зв'язок від користувачів / відгук користувача) — це будь-яка інформація, дані, думки, враження або коментарі, які користувачі надають щодо свого досвіду використання продукту, послуги, веб-сайту, програми тощо. Ця інформація може бути як позитивною, так і негативною, і надаватися різними способами. Головна мета збору user feedback полягає в отриманні інсайтів (цінних даних), які допомагають розробникам, дизайнерам та власникам продуктів краще зрозуміти потреби, вподобання та проблеми своєї цільової аудиторії. Це дозволяє приймати обґрунтовані рішення щодо покращення продукту, виправлення помилок, впровадження нових функцій та оптимізації загального користувацького досвіду (UX).

Проте, значна кількість і часто суперечливий характер цих відгуків створюють серйозні виклики для ефективної оцінки якості продукту. Споживачі стикаються з труднощами у формуванні об'єктивного уявлення про продукт, ґрунтуючись на сотнях різноспрямованих думок, що вимагає значних часових витрат на їх опрацювання. Це підкреслює нагальну потребу в автоматизованих методах аналізу тональності (sentiment analysis) для ефективної обробки та інтерпретації відгуків.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		13

### 1.1.1. Типи та методи збору user feedback

User feedback може бути:

- Якісним (Qualitative) - описує думки, емоції, враження. Це, наприклад, текстові відгуки, коментарі, відповіді на відкриті питання, інтерв'ю.

- Кількісним (Quantitative) - виражається числовими показниками, статистикою. Сюди входять рейтинги (наприклад, 1-5 зірок), оцінки задоволеності (CSAT), індекс чистої лояльності (NPS), кількість використань функції тощо.



Рисунок 1.1 – Основні способи збору user feedback

Існує безліч способів збору зворотного зв'язку (рис. 1.1), серед яких:

- Опитування та анкети. Зазвичай проводяться онлайн і містять структуровані питання.

- Відгуки та коментарі. На платформах електронної комерції, у соціальних мережах, на спеціалізованих сайтах відгуків.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		14

- Інтерв'ю з користувачами. Глибинні бесіди для виявлення неочевидних потреб та проблем.
- Тестування юзабіліті. Спостереження за тим, як користувачі взаємодіють із продуктом, для виявлення проблем у використанні.
- Аналіз поведінки користувачів. Використання аналітичних інструментів для відстеження дій користувачів (наприклад, куди вони клікають, скільки часу проводять на сторінці).
- Системи підтримки клієнтів. Запити, скарги та пропозиції, що надходять через службу підтримки.

### *1.1.2. Методологія аналізу та оцінки user feedback*

Даний проєкт спрямований на розробку та імплементацію системи, що забезпечує автоматичне перетворення текстових відгуків про продукт у кількісний рейтинг. Основною метою є розробка надійного та ефективного методу оцінки цих рейтингів за допомогою алгоритмів машинного навчання.

Процес перетворення user feedback у рейтинги передбачає використання лексичних ресурсів, таких як SentiWordNet, для створення оцінних слів (sentiment lexicon). SentiWordNet дозволяє визначити рівень позитивності, негативності та об'єктивності для кожного слова, що є критично важливим для визначення загальної тональності відгуку. Після цього, відгуки класифікуються у сім дискретних категорій, що охоплюють спектр від "дуже позитивних" до "дуже негативних". Ця багатокатегорійна система забезпечує більш гранульовану та точну оцінку порівняно з бінарною класифікацією.

### *1.1.3. Застосування алгоритмів машинного навчання*

Для оцінки та валідації отриманих рейтингів будуть застосовані дві ключові моделі машинного навчання:

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						15
Змн.	Арк.	№ докум.	Підпис	Дата		

- Наївний Байєс (Naive Bayes) - цей імовірнісний класифікатор, заснований на теоремі Байєса з припущенням про умовну незалежність ознак, ефективно використовується для завдань класифікації тексту завдяки своїй простоті та швидкості обчислень.

- Метод опорних векторів (Support Vector Machines - SVM) - є потужним алгоритмом для класифікації та регресії, що будує гіперплощину або набір гіперплощин у багатовимірному просторі, які розділяють об'єкти різних класів з максимальною межею. Його висока ефективність у роботі з високорозмірними даними робить його ідеальним для аналізу текстових відгуків.

Порівняльний аналіз продуктивності цих алгоритмів дозволить визначити найбільш оптимальний підхід для точного та надійного прогнозування рейтингів на основі текстових відгуків. Результати дослідження матимуть значну практичну цінність для розробки систем рекомендацій, покращення клієнтського сервісу та оптимізації продуктів на основі комплексного аналізу зворотного зв'язку від користувачів.

У контексті даного проекту, "онлайн-відгуки (user feedback)" є текстовими даними, які потрібно аналізувати за допомогою машинного навчання для перетворення їх у кількісні рейтинги.

## **1.2. Проблематика та обмеження існуючих підходів до аналізу онлайн-відгуків**

У сучасному інформаційному просторі, де споживачі стикаються з величезною кількістю онлайн-відгуків про продукти, що походять з різноманітних веб-ресурсів, виникає значна складність у формуванні однозначного висновку щодо якості товару або послуги. Істотна розбіжність у думках користувачів, коли один і той же продукт сприймається частиною аудиторії як високоякісний, а іншою – як незадовільний, ускладнює

					БР.ІП – 40.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		16

об'єктивну оцінку. Це підкреслює необхідність розробки ефективних методів для індивідуальної оцінки та агрегації відгуків з метою прийняття обґрунтованого кінцевого рішення.

### *1.2.1. Аналіз існуючих підходів та обмежень SentiWordNet*

Існуючі системи аналізу тональності, такі як SentiWordNet, позиціонуються як лексичні ресурси для визначення настроїв у тексті. SentiWordNet присвоює чисельні оцінки частинам мови, зокрема прислівникам та прикметникам, відображаючи їхню позитивну, негативну чи об'єктивну спрямованість. Проте, емпіричні дослідження показують, що ефективність SentiWordNet у наданні точних та інтерпретованих результатів аналізу тональності є обмеженою.

SentiWordNet — це лексичний ресурс (або лексикон), спеціально розроблений для завдань аналізу тональності (sentiment analysis) та видобування думок (opinion mining). Він є розширенням загальновідомої лексичної бази даних WordNet. WordNet — це велика лексична база даних для англійської мови, яка групує слова в набори синонімів, так звані синсети (synsets), кожен з яких представляє одне лексичне поняття. Різні синсети пов'язані між собою смисловими та лексичними зв'язками (наприклад, "є видом", "є частиною", "протилежний").

SentiWordNet бере кожен синсет з WordNet і присвоює йому три числові оцінки (бали), що відображають його тональність:

- Позитивна оцінка (Positivity Score) - відображає, наскільки позитивним є значення синсета.
- Негативна оцінка (Negativity Score) - відображає, наскільки негативним є значення синсета.
- Об'єктивна оцінка (Objectivity Score) - відображає, наскільки об'єктивним (нейтральним) є значення синсета.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		17

Кожна з цих оцінок знаходиться в діапазоні від 0 до 1, і сума всіх трьох оцінок для одного синсета завжди дорівнює 1.

SentiWordNet був створений за допомогою напівавтоматичного процесу анотування. Це означає, що початкові оцінки для синсетів були отримані алгоритмічно (наприклад, за допомогою машинного навчання або евристик), а потім могли бути допрацьовані вручну для підвищення точності.

Виявлені недоліки існуючої системи SentiWordNet включають:

1. Проблеми з інтерпретацією тональності.

Чисельні оцінки, надані SentiWordNet, не завжди дозволяють однозначно визначити загальну тональність відгуку як "позитивну" чи "негативну". Це створює труднощі в автоматизованій класифікації та подальшому використанні цих даних.

2. Складність агрегації.

Перетворення сукупності числових оцінок окремих слів на загальну тональність речення або документа часто є нетривіальним завданням і може призвести до неточних результатів.

3. Відсутність метрик продуктивності.

Існуюча система не передбачає механізмів для оцінки власної продуктивності або точності своїх висновків. Це унеможливорює об'єктивну валідацію результатів та їх порівняння з іншими методами аналізу тональності.

Ці обмеження вказують на необхідність розробки більш досконалих та надійних методологій аналізу відгуків, які б забезпечували як точну класифікацію тональності, так і можливість кількісної оцінки ефективності використовуваних алгоритмів.

Саме через ці обмеження у даному проєкті пропонується використовувати SentiWordNet як один з компонентів для створення "оцінних слів", але основну класифікацію та оцінку виконувати за

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						18
Змн.	Арк.	№ докум.	Підпис	Дата		

допомогою алгоритмів машинного навчання, таких як Наївний Байєс та SVM, які можуть краще враховувати контекст та навчатися на даних.

### **1.3. Концепція аналізу тональності user feedback соціальних мереж з використанням онтології SentiWordNet**

За допомогою аналізу тональності текстових повідомлень користувачів дослідник може зробити висновки про:

- емоційну оцінку користувачами різних подій та об'єктів;
- індивідуальні переваги користувачів;
- деякі особливості характеру користувачів [2].

Тональність — це емоційне ставлення автора до певного об'єкта у формі тексту. Наразі існує набір методів аналізу тональності текстової інформації. Перш за все, це методи, засновані на використанні словників та машинного навчання з учителем.

Основним компонентом концепції Web 2.0 є розвиток соціальних мереж. Web 2.0 передбачає формування вмісту електронних ресурсів (включаючи текст) користувачами через їхні профілі (пости, коментарі, назви файлів, підписи файлів різних форматів тощо). Тому текстові дані в соціальних мережах мають такі особливості:

- Використання сленгових виразів, неологізмів, діалектних форм.
- Використання цілих та неповних речень.
- Наявність мовленнєвих та орфографічних помилок.
- Використання смайлів, емодзі для надання повідомленню певного емоційного забарвлення.

#### *1.3.1. Групи методів для аналізу тональності текстових даних*

Існують дві основні групи методів для автоматичного аналізу тональності текстових даних:

					БР.ІП – 40.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		19

## А. Статистичні методи

Основою цих методів є використання машинного класифікатора. Цей класифікатор навчається на попередньо розмічених текстах на початкових етапах. Потім класифікатор будує модель для аналізу нових документів, використовуючи отримані знання. Алгоритм складається з:

- Збір документів для навчання машинного класифікатора;
- Кожен документ розкладається на вектор ознак;
- Вказується правильний тип тональності для кожного документа;
- Вибір алгоритму класифікації та методу навчання класифікатора;
- Отримана модель використовується для визначення тональності документів нової колекції.

Недоліком таких методів є необхідність великої кількості даних для навчання.

Статистичний підхід широко використовує метод опорних векторів (SVM), байєсові моделі, різні типи регресії, методи Word2Vec, Doc2Vec, CRF, згорткові та рекурентні нейронні мережі.

## Б. Методи на основі словників

Тональні словники та правила складаються за допомогою лінгвістичного аналізу. Ці словники шукають емоційно забарвлені слова та вирази. Далі, набір емоційно забарвлених слів оцінюється за шкалою, що містить кількість негативної та позитивної лексики.

Методи на основі словників передбачають наявність лінгвіста-експерта, який складає точний довідник емоційно забарвлених слів та виразів. Кожен вираз (слово, смайл або стиль) називається маркером. При знаходженні маркера емоція враховується в алгоритмі. Результатом алгоритму є середнє емоційне забарвлення тексту. Зазвичай використовується наступний алгоритм:

- Призначити оцінку тональності з словника кожному слову в тексті;

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						20
Змн.	Арк.	№ докум.	Підпис	Дата		

- Розрахувати загальну оцінку тональності всього тексту, додавши оцінки тональності окремих слів.

Недоліком цього методу є значна кількість праці, оскільки метод вимагає створення великої кількості правил.

Крім того, іноді використовується змішаний метод (комбінація першого та другого підходів).

#### В. Онтологічний метод

Онтологічний метод аналізу тональності текстових даних з соціальних мереж є модифікацією методів, заснованих на лексичних словниках.

Загальна схема онтологічного алгоритму представлена на рисунку 1.2.

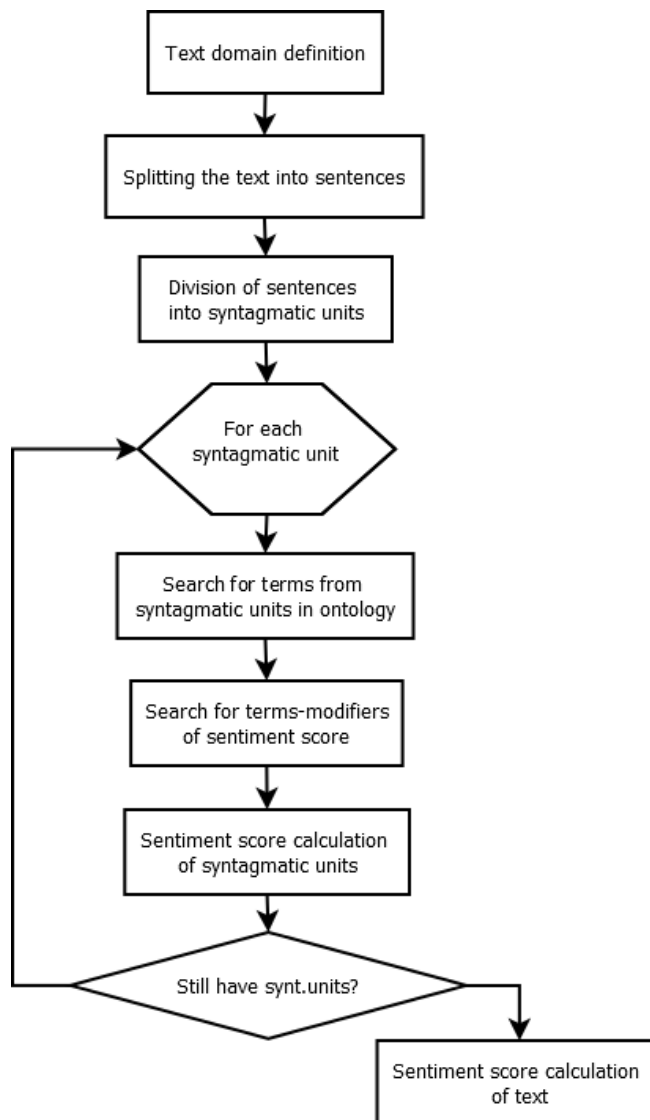


Рисунок 1.2 – Алгоритм онтологічного методу

Основні особливості онтологічного методу:

- Врахування особливостей предметної області. Оцінка тональності кожної синтагматичної одиниці завжди залежить від характеру предметної області. На рисунку 1.3 Pr<sub>1</sub>, Pr<sub>2</sub> — предметні області, S<sub>1</sub>-S<sub>6</sub> — оцінки тональності термінів.

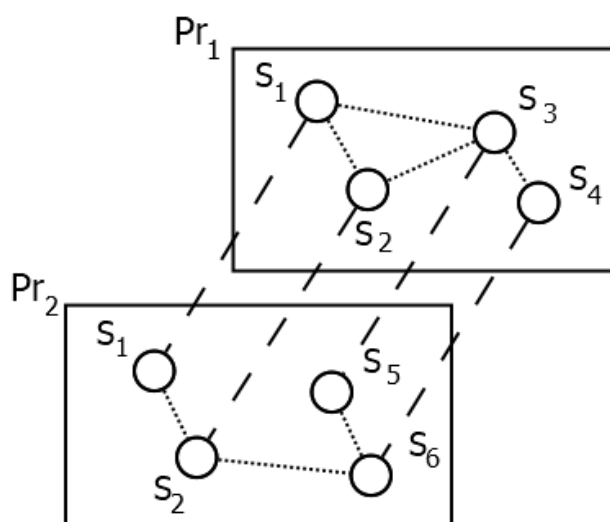


Рисунок 1.3 - Переклад термінів однієї предметної області термінами іншої предметної області

- Врахування тональності синтагматичних одиниць, а не окремих слів. Синтагматична одиниця — це сукупність кількох слів, об'єднаних за принципом семантичної, граматичної та фонетичної сумісності.

- Класифікація предметної області за набором використаних термінів.

- Врахування семантичних відношень між об'єктами (термінами). Перш за все враховуються відношення синонімії, гіпонімії тощо. Ця функціональність допомагає оцінити тональність сленгових виразів, неологізмів, що використовуються в текстах соціальних мереж, та врахувати орфографічні та мовленнєві помилки.

- Врахування впливу термінів, які змінюють рівень емоційного забарвлення синтагматичної одиниці. Деякі терміни мають а пріорі коефіцієнти, які змінюють загальну оцінку тональності синтагматичної

одиниці в поєднанні з іншими термінами. Наприклад, слова з коефіцієнтами посилення: "дуже", "над", "сильніше" тощо, зменшення — "менше", "нижче", заперечення — "не" тощо.

$$O = \{ \text{Synt}^{PrA}, PrA, \rho_V^{\text{Synt}}, V^{\text{Synt}} \}$$

де Synt — набір синтагматичних одиниць аналізованого текстового фрагмента; PrA — предметна область,  $\rho_V$  — набір оцінок тональності синтагматичних текстових одиниць, де

$$\rho_V^{\text{Synt}} = k \times \rho_V$$

де  $k$  — апріорі коефіцієнт модифікації оцінки тональності синтагматичної одиниці,  $\rho_V$  — оцінка тональності терміна предметної області.

### *1.3.2. Архітектура програмного забезпечення для аналізу тональності в соціальних мережах*

Інтелектуальні інструменти для аналізу думок у соціальних мережах базуються на гібридному підході, що поєднує онтологічний аналіз з методами обробки природної мови (НЛП). Ця синергія дозволяє ефективно виокремлювати семантичні та емоційні компоненти з напівструктурованих та неструктурованих ресурсів. Такий підхід значно підвищує ефективність аналізу контенту соціальних медіа, враховуючи специфіку представлення даних та властиву природним мовам нечіткість.

На рисунку 1.4 представлена сервісно-орієнтована архітектура програмного забезпечення. Вона передбачає обмін даними та виклик сервісів за допомогою HTTP-запитів.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						23
Змн.	Арк.	№ докум.	Підпис	Дата		

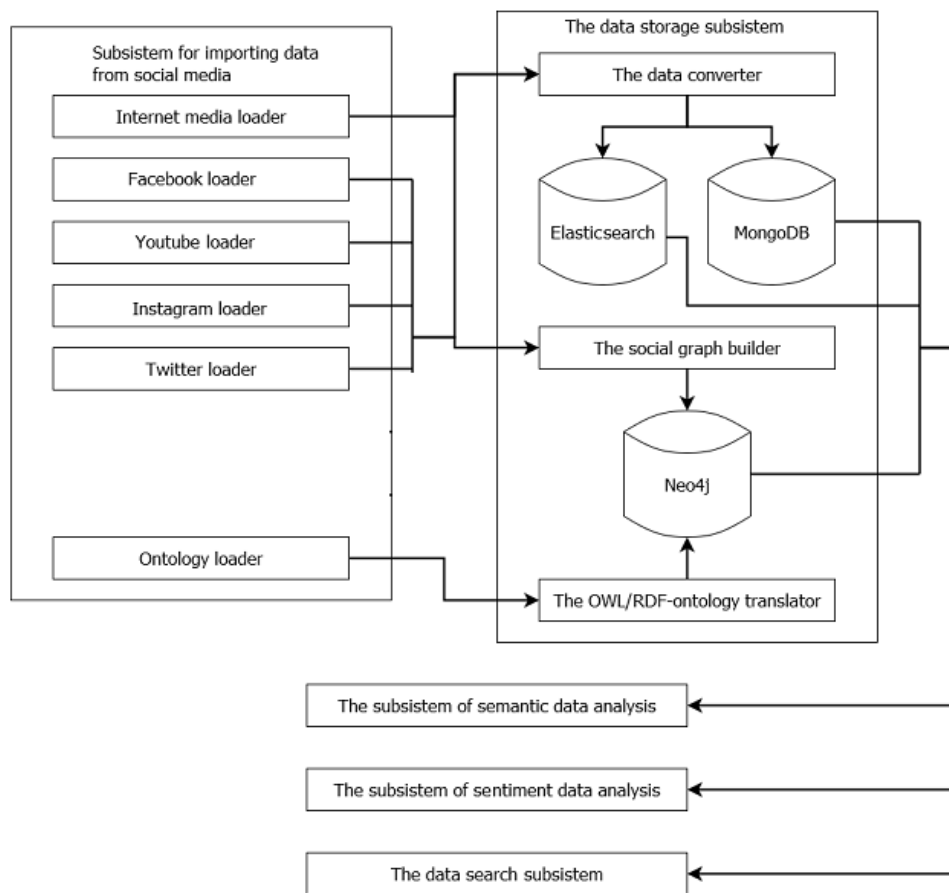


Рисунок 1.4 - Сервісно-орієнтована архітектура аналізу тональності

Як видно з рисунку, програмна система комплекс складається з наступних ключових підсистем:

- Підсистема імпорту даних: Забезпечує збір та надходження даних із соціальних мереж.
- Підсистема зберігання: Відповідає за ефективне та структуроване зберігання отриманих даних.
- Конвертер даних: Здійснює перетворення даних до уніфікованого формату для подальшої обробки.
- Підсистема семантичного аналізу даних: Виконує глибокий аналіз змісту для виявлення значущих сутностей та зв'язків.
- Підсистема аналізу тональності даних: Оцінює емоційне забарвлення аналізованого тексту. Ця підсистема є однією з ключових послуг системи.

- Підсистема пошуку даних: Забезпечує ефективний доступ та пошук необхідної інформації.

У рамках реалізації одного з алгоритмів аналізу думок всередині підсистеми аналізу тональності даних було проведено серію експериментальних досліджень. Метою цих експериментів була оцінка ефективності кожного з найпопулярніших підходів до аналізу тональності, що дозволило визначити оптимальні рішення для інтеграції в програмний комплекс. Для оцінки ефективності онтологічного алгоритму була використана лексична онтологія. Основою цієї онтології є словник оціночних слів SentiWordNet. SentiWordNet — це лексичний ресурс для аналізу думок. SentiWordNet присвоює три значення кожній лексичній одиниці WordNet: позитивність, негативність, об'єктивність. Найкращий результат показав метод машинного навчання з використанням наївного байєсівського класифікатора. Байєсівський класифікатор навчався на уніграмах та біграмах: в обох випадках результати класифікації були кращими за всі інші алгоритми. Даний метод на основі словника, отриманого з SentiWordNet, показав ефективність 77%. Рекурентні нейронні мережі показали ефективність 76%.

#### **1.4. Аналіз спеціалізованих SaaS-платформ для аналізу відгуків/соціальних медіа**

Ці платформи зазвичай пропонують не тільки API для аналізу тональності, а й повний цикл інструментів для збору, агрегації, візуалізації та звітності по відгуках.

##### *1.4.1. Платформа Brandwatch*

Комплексна платформа для моніторингу соціальних медіа та аналізу споживчих настроїв. Brandwatch — це провідна глобальна платформа для

					БР.ІП – 40.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		25

моніторингу та аналізу цифрових даних, яка спеціалізується на аналізі соціальних медіа та споживчих настроїв (sentiment analysis). Вона надає компаніям інструменти для збору, обробки, аналізу та візуалізації величезних обсягів публічних онлайн-даних, що дозволяє отримувати глибокі інсайти щодо репутації бренду, сприйняття продукту, трендів ринку та поведінки споживачів.

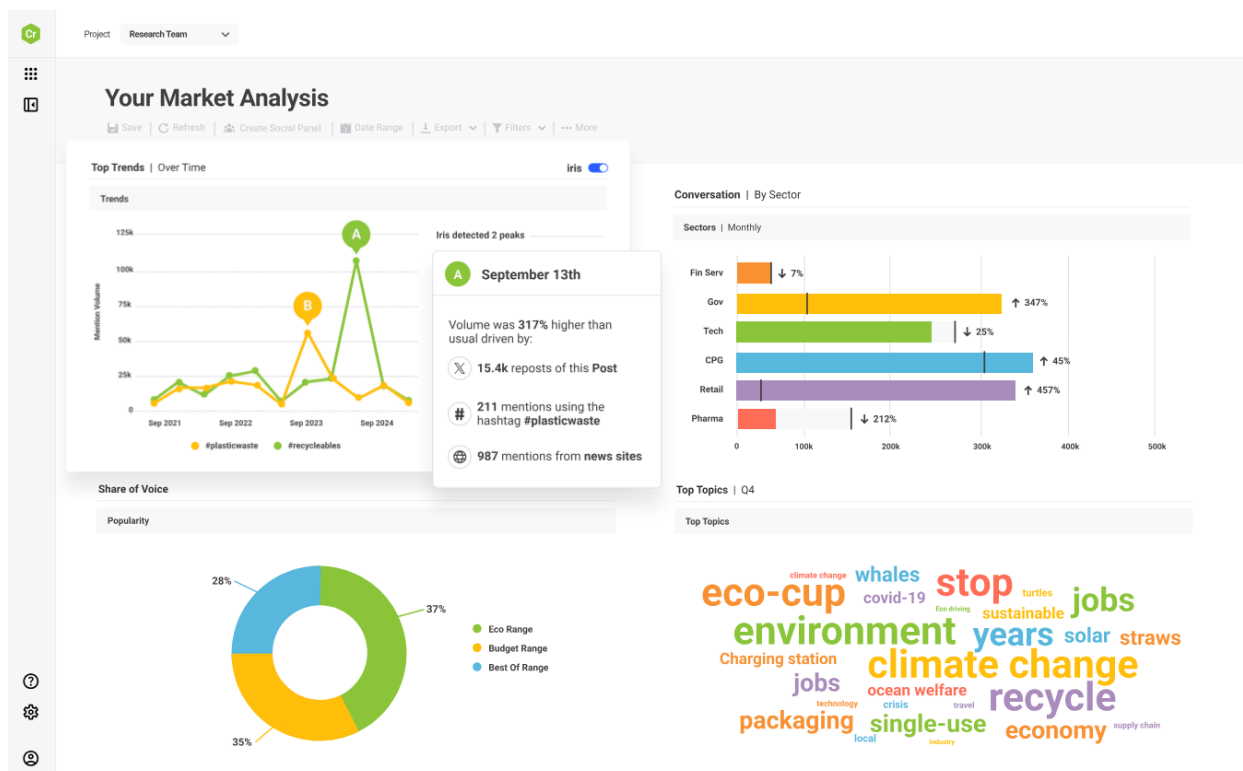


Рисунок 1.5 – Дашборд платформи Brandwatch

Головна мета Brandwatch – допомогти компаніям зрозуміти, що говорять люди про їхній бренд, продукти чи послуги в Інтернеті, і як ці розмови впливають на їхній бізнес. Це досягається шляхом агрегації даних з різних джерел, таких як соціальні мережі (Twitter, Facebook, Instagram, YouTube), блоги, форуми, новинні сайти, сайти відгуків тощо.

Ключові можливості та функції:

1. Збирає дані з мільйонів онлайн-джерел, включаючи соціальні мережі, новинні портали, форуми, блоги, сайти відгуків.

									Арк.
									26
Змн.	Арк.	№ докум.	Підпис	Дата	БР.ІП – 40.00.00.000 ПЗ				

2. Дозволяє створювати складні пошукові запити для моніторингу конкретних ключових слів, фраз, хештегів або згадок брендів.

3. Автоматично визначає емоційне забарвлення згадок (позитивне, негативне, нейтральне).

4. Забезпечує можливість більш детального аналізу тональності, іноді з розпізнаванням емоцій (гнів, радість, сум тощо).

5. Дозволяє користувачам навчати систему розпізнавати специфічну для галузі або бренду тональність, підвищуючи точність.

6. Автоматично визначає ключові теми та суб'єкти розмов, що дозволяє зрозуміти, що найбільше хвилює аудиторію.

7. Виявляє ключових лідерів думок та впливових користувачів, які говорять про бренд або галузь.

8. Надає інформацію про аудиторію, яка генерує згадки (вік, стать, місцезнаходження тощо).

Пропонує настроювані дашборди з графіками, діаграмами та іншими візуалізаціями для швидкого перегляду даних. Є можливість генерувати регулярні або спеціальні звіти для внутрішнього використання або для клієнтів.

Платформа надає сповіщення в реальному часі, налаштування автоматичних сповіщень про раптові зміни в тональності або зростання негативних згадок, що дозволяє швидко реагувати на потенційні кризи.

#### *1.4.2. Talkwalker*

Talkwalker — це провідна платформа для моніторингу соціальних медіа та аналітики, яка надає компаніям всебічні інструменти для збору, обробки, аналізу та візуалізації величезних обсягів даних з онлайн-джерел. Її основна функція полягає у забезпеченні глибокого розуміння споживчих настроїв, репутації бренду, ринкових тенденцій та ефективності маркетингових кампаній.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		27



Рисунок 1.6 – Інтерфейс Talkwalker

Talkwalker допомагає організаціям слухати та розуміти онлайн-розмови, що стосуються їхнього бренду, конкурентів, галузі чи будь-яких інших тем, які є важливими для їхнього бізнесу. Це дозволяє приймати більш інформовані рішення у сферах маркетингу, PR, розвитку продукту та управління клієнтським досвідом.

Збирає дані з понад 150 мільйонів джерел, включаючи соціальні мережі (Facebook, X (раніше Twitter), Instagram, YouTube, LinkedIn, TikTok), блоги, форуми, новинні сайти, сайти відгуків, телевізійні та радіопередачі, подкасти, а також власні джерела даних компаній.

Підтримує моніторинг та аналіз контенту більш ніж 187 мовами. Надає доступ до великих архівів історичних даних, що дозволяє відстежувати тенденції та аналізувати минулі події.

Класифікує згадки не лише як позитивні, негативні чи нейтральні, а й може розпізнавати специфічні емоції, такі як радість, злість, подив тощо.

Можливість налаштування та навчання моделей тональності під специфіку конкретної галузі або бренду для підвищення точності.

Автоматично ідентифікує основні теми, що обговорюються, та висхідні тренди в режимі реального часу. Виявляє ключових інфлюенсерів, лідерів думок та їхній вплив на дискусії. Надає деталі про демографію та психографіку аудиторії, яка генерує згадки.

Надає настроювані дашборди, інтуїтивно зрозумілі та настроювані панелі, які дозволяють користувачам візуалізувати ключові метрики та інсайти. Можливість планування та генерації звітів у різних форматах (PDF, Excel, PowerPoint) для зручного обміну інформацією.

### 1.4.3. Qualtrics (Customer Experience Management)

Qualtrics CXM (Customer Experience Management) — це комплексна програмна платформа, призначена для управління клієнтським досвідом. Вона інтегрує збір, аналіз та активацію даних зворотного зв'язку від клієнтів з різних джерел.

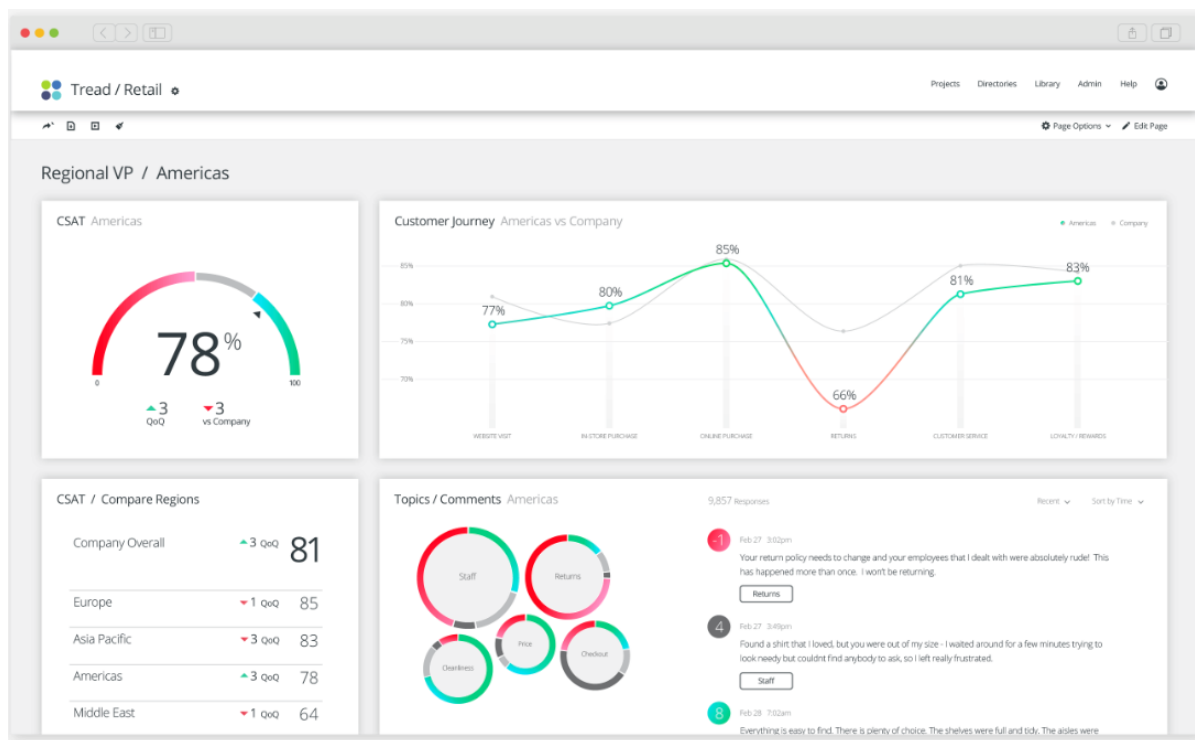


Рисунок 1.7 – Платформа Qualtrics CXM

Метою Qualtrics CXM є надання організаціям інструментів для:

- Збору інформації про клієнтський досвід.
- Аналізу цієї інформації для виявлення інсайтів.
- Прийняття обґрунтованих рішень для покращення клієнтського досвіду.
- Автоматизації дій на основі отриманих даних.

Ключові функціональні можливості:

- Опитування: Розробка та розповсюдження різноманітних опитувань (традиційні, Net Promoter Score (NPS), Customer Satisfaction (CSAT), Customer Effort Score (CES)).

- Багатоканальність: Збір даних через веб-сайти, мобільні додатки, електронну пошту, SMS, чат-боти, соціальні мережі, контакт-центри та фізичні точки взаємодії.

- Unsolicited Feedback: Моніторинг та аналіз неструктурованих відгуків з відкритих джерел (соціальні медіа, відгуки на сайтах).

- Аналіз тональності (Sentiment Analysis): Автоматичне визначення емоційного забарвлення текстових відгуків (позитивне, негативне, нейтральне), а також ідентифікація конкретних емоцій.

- Аналіз тексту (Text Analysis): Виявлення ключових тем, фраз та сутностей у неструктурованих текстових даних.

- Автоматизація робочих процесів на основі зворотного зв'язку (наприклад, створення тикетів підтримки).

- Персоналізовані взаємодії і використання інсайтів для адаптації комунікацій та пропозицій для кожного клієнта.

- Статистичний аналіз: Використання статистичних методів для виявлення кореляцій та причинно-наслідкових зв'язків.

- Предиктивна аналітика: Прогнозування поведінки клієнтів (наприклад, відтоку) на основі історичних даних.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						30
Змн.	Арк.	№ докум.	Підпис	Дата		

- Дашборди та візуалізація: Створення інтерактивних дашбордів та звітів для візуалізації ключових метрик та тенденцій.

Є функції налаштування автоматичних сповіщень для відповідних команд або відділів у випадку виявлення негативних відгуків або проблем.

Платформа Qualtrics CXM є комплексним рішенням для CX, інтеграція з опитуваннями. До недоліків можна віднести, те що це є не суто платформою для аналізу тональності, і вимагає інвестицій у всю екосистему CX.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						31
Змн.	Арк.	№ докум.	Підпис	Дата		

## РОЗДІЛ 2. ПРОЕКТУВАННЯ АРХІТЕКТУРИ СИСТЕМИ ОБРОБКИ ДАНИХ З USER FEEDBACK

### 2.1. Опис функціональності пропонованої системи для аналізу тональності user feedback

Даний проєкт передбачає розробку графічного інтерфейсу користувача (GUI), призначеного для автоматизованої оцінки продукту (фільму, серіалу) на основі аналізу user feedback (користувацьких відгуків). Користувач взаємодіятиме з системою шляхом вибору відповідного набору даних відгуків та ініціації процесу аналізу, що зрештою призведе до відображення агрегованого рейтингу продукту.

Опишемо структура та функціонал користувацького інтерфейсу.

Додаток починає свою роботу з екрану автентифікації, де користувачеві пропонується ввести ім'я користувача та пароль. Система передбачає валідацію введених даних; у випадку їх відсутності або некоректності буде відображено відповідне повідомлення.

Після успішної автентифікації користувач отримає доступ до основного інтерфейсу, що містить назву проєкту у верхній частині та набір з семи функціональних кнопок у нижній панелі:

1. Завантажити словник (Load Dictionary) - ця функція є основоположною. Вона ініціює попередню обробку текстових даних, що включає видалення стоп-слів, застосування алгоритму стеммінгу Портера та інтеграцію SentiWordNet. Завантаження словника є обов'язковою процедурою при кожному вході користувача до системи.

2. Аналіз настроїв SVM (SVM Sentiment Analysis) - ця кнопка активує процес трансформації тональних оцінок, отриманих з SentiWordNet, у категоріальний рейтинг за допомогою методу опорних векторів (SVM). Аналіз здійснюється на вибраному користувачем наборі даних. Результати

									Арк.
									32
Змн.	Арк.	№ докум.	Підпис	Дата	БР.ІП – 40.00.00.000 ПЗ				

відображаються у лівій панелі інтерфейсу у вигляді таблиці, що містить назву набору даних, текст відгуку, класифікований SVM рейтинг (позитивний, негативний або нейтральний) та початкову оцінку SentiWordNet.

3. Аналіз настроїв Наївного Байєса (Naive Bayes Sentiment Analysis) - подібно до попередньої функції, ця кнопка ініціює перетворення оцінок SentiWordNet у рейтинг, але вже із застосуванням алгоритму Наївного Байєса. Результати аналізу відображаються у правій панелі інтерфейсу у форматі таблиці, яка включає назву набору даних, текст відгуку, рейтинг, визначений Наївним Байєсом (позитивний, негативний або нейтральний), та відповідну оцінку SentiWordNet.

4. Матриця плутанини (Confusion Matrix) - ця функція генерує таблицю, що забезпечує порівняльний аналіз класифікаційних результатів, отриманих за допомогою SVM та Наївного Байєса. Обидва алгоритми категоризують оцінки SentiWordNet за сім'ю тональними категоріями, що варіюються від "дуже позитивних" до "дуже негативних".

5. Точність SVM (SVM Accuracy) - розраховує метрику точності для моделі SVM, визначаючи відношення кількості правильно класифікованих позитивних рейтингів до загальної кількості оцінок у вибраному наборі даних.

6. Точність Наївного Байєса (Naive Bayes Accuracy) - обчислює метрику точності для моделі Наївного Байєса, аналогічно до SVM, визначаючи відношення кількості правильно класифікованих позитивних рейтингів до загальної кількості оцінок у наборі даних.

7. Діаграма правильно класифікованих (Correctly Classified Chart) - Надає графічне представлення порівняльної точності алгоритмів SVM та Наївного Байєса, візуалізуючи їхню ефективність.

Вихід (Exit): Ця кнопка завершує поточну сесію користувача.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		33

Ця архітектура забезпечує комплексну платформу для інтерактивного аналізу тональності відгуків, дозволяючи користувачеві не тільки отримувати кінцеві рейтинги, але й глибоко аналізувати ефективність різних алгоритмів машинного навчання.

## 2.2. Специфікації вимог до системи

Для забезпечення коректного функціонування та ефективного розгортання запропонованого програмного рішення оцінки тональності user feedback користувачі повинні дотримуватися наведених нижче апаратних та програмних вимог.

### 2.2.1. Апаратні вимоги

Апаратні вимоги (Hardware Requirements) визначають мінімальну або рекомендовану конфігурацію комп'ютерного обладнання, необхідного для запуску програмного забезпечення.

Для успішного запуску та експлуатації додатка необхідні наступні мінімальні апаратні конфігурації:

Тип пристрою: Ноутбук

Операційна система: Windows

Інтегроване середовище розробки (IDE): Eclipse IDE

Оперативна пам'ять (RAM): Мінімум 8 ГБ

### 2.2.2. Програмні вимоги

Програмне забезпечення, необхідне для функціонування системи, включає:

Операційна система: Windows 10

Мова програмування: Java

Графічна бібліотека (Toolkit): Java Swing

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						34
Змн.	Арк.	№ докум.	Підпис	Дата		

## 2.3. Опис використаних програмних засобів та алгоритмів

Реалізація програмного забезпечення передбачає використання специфічних технологій для розробки графічного інтерфейсу користувача та імплементації алгоритмів аналізу тональності.

### 2.3.1. Графічний інтерфейс користувача (GUI)

Для забезпечення графічної взаємодії користувача із системою, на противагу текстовим командним інтерфейсам, застосовується інструментарій Java Swing. Swing є частиною Java Foundation Classes (JFC), що надає набір легковагових GUI-компонентів. Ключовою перевагою Swing є його платформна незалежність, що забезпечує уніфікований зовнішній вигляд та поведінку компонентів на різних операційних системах. До типових елементів керування Swing, використаних у системі, належать кнопки, мітки, а також компоненти для роботи зі списками, деревами та таблицями. Ієрархія API Java Swing представлена на рисунку 2.1.

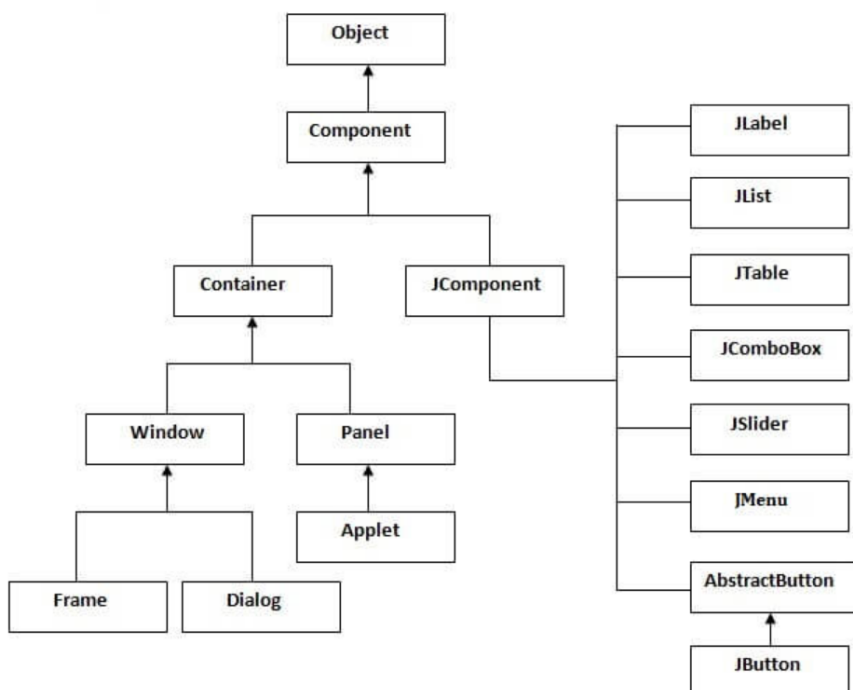


Рисунок 2.1 - Ієрархія API Java Swing

Змн.	Арк.	№ докум.	Підпис	Дата

У системі інтегровано два ключові алгоритми машинного навчання для класифікації текстових відгуків за тональністю: метод опорних векторів (SVM) та наївний байєс.

### 2.3.2. Метод опорних векторів (SVM)

Метод опорних векторів (SVM, Support Vector Machine) — це потужний алгоритм машинного навчання, який широко використовується для задач класифікації, зокрема для класифікації тональності текстів (sentiment analysis). SVM може бути використаний як для класифікаційних задач (наприклад, визначення, до якого класу належить зображення), так і для регресійних задач (наприклад, прогнозування цін на житло). Його основна мета — знайти оптимальну гіперплощину, яка найкраще розділяє дані різних класів.

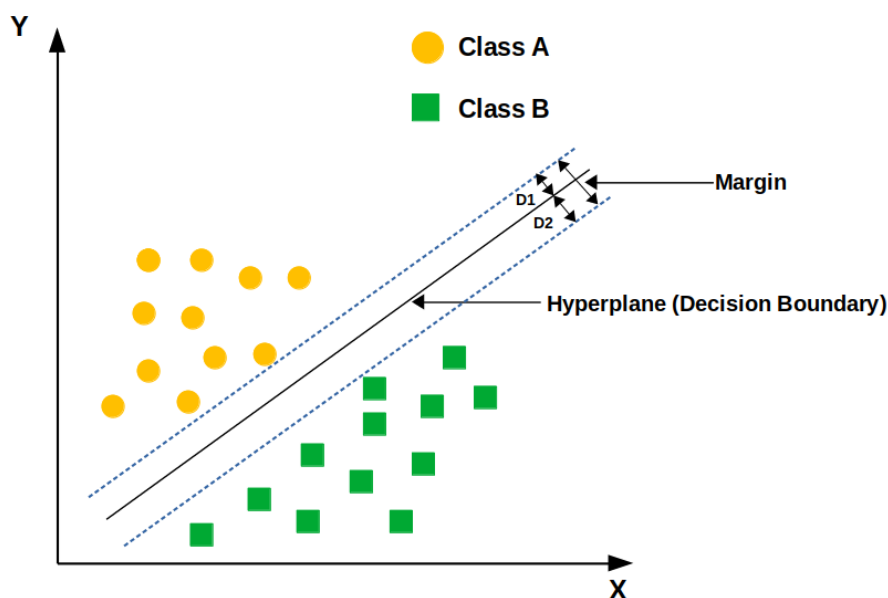


Рисунок 2.2 – Ілюстрація роботи методу SVM

Алгоритм роботи SVM для класифікації тональності текстів:

1. Представлення тексту у вигляді числових ознак:

- Перед застосуванням SVM текстові дані потрібно перетворити у векторну форму за допомогою методів Bag of Words (BoW) — фіксує частоту

слів, TF-IDF (Term Frequency-Inverse Document Frequency) — враховує важливість слів у текстах або Word Embeddings (наприклад, Word2Vec, GloVe) — семантичне представлення слів у вигляді векторів.

## 2. Навчання SVM:

- Після векторизації текстів алгоритм SVM навчається розділяти їх за тональністю: позитивна, негативна, нейтральна (залежно від задачі — бінарна або багатокласова класифікація).

- Він шукає гіперплощину у багатовимірному просторі, яка максимально розділяє тексти різної тональності.

## 3. Класифікація нових текстів:

- Нові тексти також перетворюються на вектори.

- SVM визначає, по який бік гіперплощини знаходиться текст, і таким чином класифікує його тональність.

Метод опорних векторів (SVM) застосовується для класифікації текстів на позитивні, негативні або нейтральні категорії. Цей алгоритм демонструє високу ефективність при обробці великих обсягів текстових даних. Індивідуальні тональні оцінки, отримані для кожного оцінного слова за допомогою SentiWordNet, класифікуються SVM у сім дискретних категорій:

- Сильно-позитивний (sp): оцінка  $>0.4$
- Позитивний (p):  $0.3 < \text{оцінка} < 0.4$
- Слабо-позитивний (wp):  $0.2 < \text{оцінка} < 0.3$
- Сильно-негативний (sn):  $0.1 < \text{оцінка} < 0.2$
- Негативний (n):  $0 < \text{оцінка} < 0.1$
- Слабо-негативний (wn): оцінка  $< 0$
- Нейтральний (nt): оцінка  $= 0$

Загальна позитивна оцінка для відгуку (t.p) розраховується як сума часткових оцінок:

$$t.p = svm.sp + svm.p + svm.wp$$

Загальна негативна оцінка для відгуку (t.n) розраховується аналогічно:

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						37
Змн.	Арк.	№ докум.	Підпис	Дата		

$$t.n = svm.sn + svm.n + svm.wn$$

Точність SVM визначається як відношення кількості позитивно класифікованих відгуків до загальної кількості відгуків у наборі даних.

$$\text{Точність SVM} = \frac{\text{Кількість позитивних відгуків}}{\text{Загальна кількість відгуків}}$$

SVM є надійним і точним методом для класифікації тональності текстів, особливо якщо дані добре підготовлені й векторизовані.

### 2.3.3. Наївний Байєс

Наївний Байєсівський класифікатор (Naive Bayes) — це простий, але ефективний алгоритм машинного навчання, який часто використовується для класифікації текстів за тональністю (sentiment analysis). Його основа — теорема Байєса, яка дозволяє обчислювати ймовірність того, що текст належить до певного класу (наприклад, позитивного чи негативного), з урахуванням слів, які в ньому містяться.

Алгоритм роботи Наївного Байєса у класифікації тональності текстів:

#### 1. Попередня обробка текстів:

- Текст очищається, видаляються стоп-слова, пунктуація, застосовується лематизація або стемінг.

- Потім текст представляється у вигляді числових ознак — зазвичай це з використанням Bag of Words (BoW) або TF-IDF.

#### 2. Модель і припущення:

- Алгоритм використовує теорему Байєса для обчислення ймовірності приналежності тексту до певного класу:

$$P(C | X) = \frac{P(X | C) \cdot P(C)}{P(X)}$$

де:

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						38
Змн.	Арк.	№ докум.	Підпис	Дата		

C — клас (наприклад, позитивна або негативна тональність),

X — набір слів у тексті.

- "Наївне" припущення полягає в тому, що всі ознаки (тобто слова) незалежні одне від одного, що значно спрощує обчислення.

### 3. Класифікація:

Для кожного класу обчислюється ймовірність  $P(C | X)$ , і текст відносять до того класу, де ця ймовірність максимальна.

Наївний Байєс також використовується для класифікації текстів за тональністю (позитивні, негативні або нейтральні). Цей алгоритм особливо ефективний при роботі з відносно невеликими текстовими фрагментами. Подібно до SVM, індивідуальні оцінки SentiWordNet класифікуються Наївним Байєсом у ті самі сім тональних категорій, але з дещо іншими пороговими значеннями:

1. Сильно-позитивний (sp): оцінка  $>0.25$
2. Позитивний (p):  $0.2 < \text{оцінка} < 0.25$
3. Слабо-позитивний (wp):  $0.15 < \text{оцінка} < 0.2$
4. Сильно-негативний (sn):  $0.1 < \text{оцінка} < 0.15$
5. Негативний (n):  $0 < \text{оцінка} < 0.1$
6. Слабо-негативний (wn): оцінка  $< 0$
7. Нейтральний (nt): оцінка  $= 0$

Загальна позитивна оцінка для відгуку (t.p) розраховується аналогічно до SVM:

$$t.p = nb.sp + nb.p + nb.wp$$

Загальна негативна оцінка для відгуку (t.n) також розраховується аналогічно:

$$t.n = nb.sn + nb.n + nb.wn$$

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						39
Змн.	Арк.	№ докум.	Підпис	Дата		

Точність Наївного Байєса обчислюється як відношення кількості позитивно класифікованих відгуків до загальної кількості відгуків у наборі даних

$$\text{Точність Наївного Байєса} = \frac{\text{Кількість позитивних відгуків}}{\text{Загальна кількість відгуків}}$$

Наївний Байєс — це базова, але дуже ефективна модель для аналізу тональності текстів. Його часто використовують як перший підхід у задачах NLP через його швидкість, простоту і достатню точність.

#### **2.4. Моделювання системи з використанням діаграми варіантів використання**

На рисунку 2.3 подана діаграма варіантів використання, що ілюструє функціональні можливості системи визначення тональності користувацьких відгуків з точки зору актора "User".

На діаграмі представлено одного актора — "User" (Користувач) — і низку варіантів використання, що відображають основний функціонал системи.

1. Login (Вхід до системи) - цей варіант використання є початковою точкою взаємодії користувача із системою. Він передбачає процес автентифікації користувача (введення імені користувача та пароля) для отримання доступу до функціоналу.

2. Load Dictionary (Завантажити словник). Після успішного входу користувач може ініціювати завантаження та попередню обробку лексичного словника. Цей процес, як правило, включає видалення стоп-слів, застосування алгоритму стеммінгу та ініціалізацію SentiWordNet, що є критично важливим для подальшого аналізу тональності.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						40
Змн.	Арк.	№ докум.	Підпис	Дата		

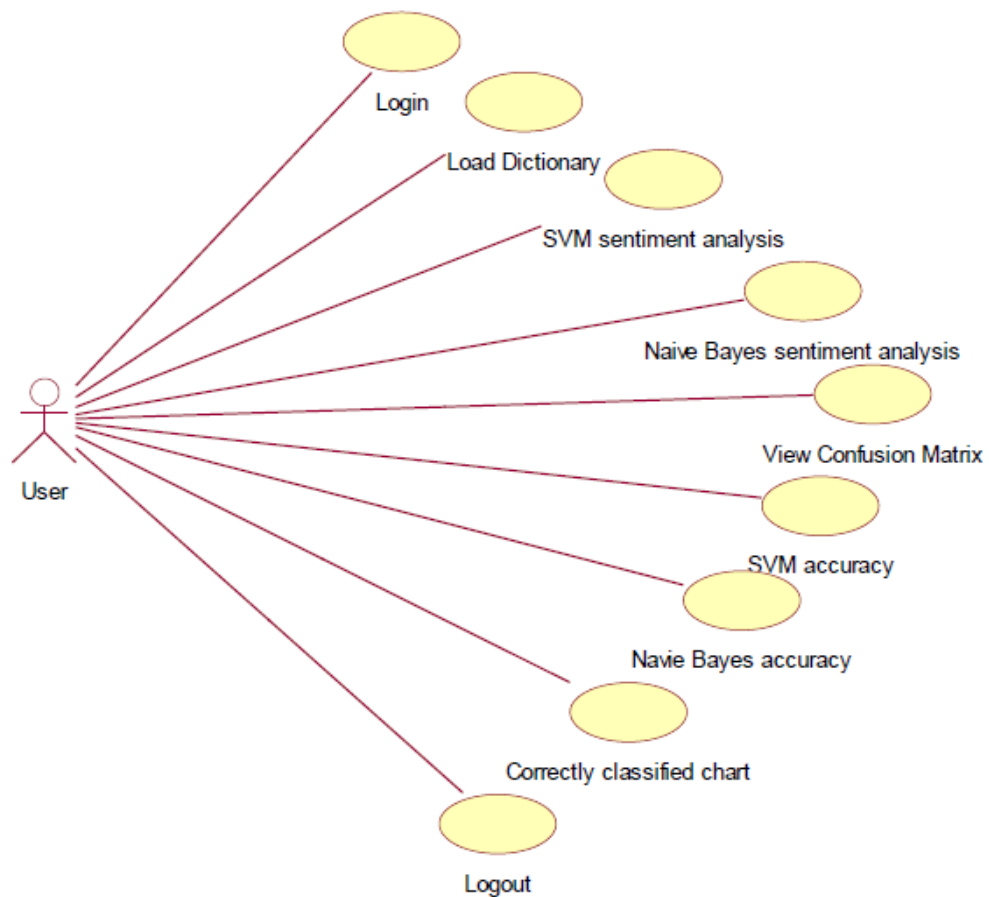


Рисунок 2.3 - Діаграма варіантів використання для системи визначення тональності user feedback

3. SVM sentiment analysis (Аналіз тональності за допомогою SVM). Користувач може виконати аналіз тональності відгуків, використовуючи алгоритм Метод опорних векторів (SVM). Цей варіант використання передбачає вибір набору даних для обробки та відображення результатів класифікації тональності (позитивна, негативна, нейтральна).

4. Naive Bayes sentiment analysis (Аналіз тональності за допомогою Наївного Байєса). Аналогічно до попереднього, цей варіант використання дозволяє користувачеві провести аналіз тональності, але вже з використанням алгоритму Наївного Байєса. Результати класифікації також відображаються для обраного набору даних.

5. View Confusion Matrix (Переглянути матрицю плутанини) - цей варіант використання дозволяє користувачеві отримати доступ до матриці

плутанини, яка надає порівняльний аналіз результатів класифікації, отриманих за допомогою SVM та Наївного Байєса, демонструючи їхню взаємну відповідність або розбіжності.

6. SVM accuracy (Точність SVM): Користувач може переглянути показник точності роботи алгоритму SVM. Цей варіант використання відображає ефективність моделі SVM у правильній класифікації тональності.

7. Naive Bayes accuracy (Точність Наївного Байєса). Аналогічно до SVM, цей варіант використання дозволяє переглянути показник точності алгоритму Наївного Байєса, оцінюючи його ефективність у класифікації тональності.

8. Correctly classified chart (Діаграма правильно класифікованих) - цей варіант використання надає графічне представлення порівняльної точності обох алгоритмів (SVM та Наївного Байєса), візуалізуючи їхню продуктивність.

9. Logout (Вихід з системи) - завершальний варіант використання, що дозволяє користувачеві безпечно вийти з системи та завершити свою сесію.

Хоча діаграма варіантів використання не показує пряму послідовність виконання, логічно, що "Login" є першим кроком, а "Logout" — останнім. "Load Dictionary" є важливим попереднім кроком для виконання будь-яких операцій аналізу тональності ("SVM sentiment analysis", "Naive Bayes sentiment analysis"). Інші варіанти використання ("View Confusion Matrix", "SVM accuracy", "Naive Bayes accuracy", "Correctly classified chart") можуть бути викликані після проведення відповідних аналізів.

## 2.5. Розробка діаграми послідовності

На рисунку 2.4 представлена діаграма послідовності, яка ілюструє покрокову взаємодію між актором "User" (Користувач) та "System" (Система) для процесу аналізу тональності користувацьких відгуків. Ця діаграма

					БР.ІП – 40.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		42

показує порядок виклику операцій та обміну повідомленнями між об'єктами протягом певного сценарію.

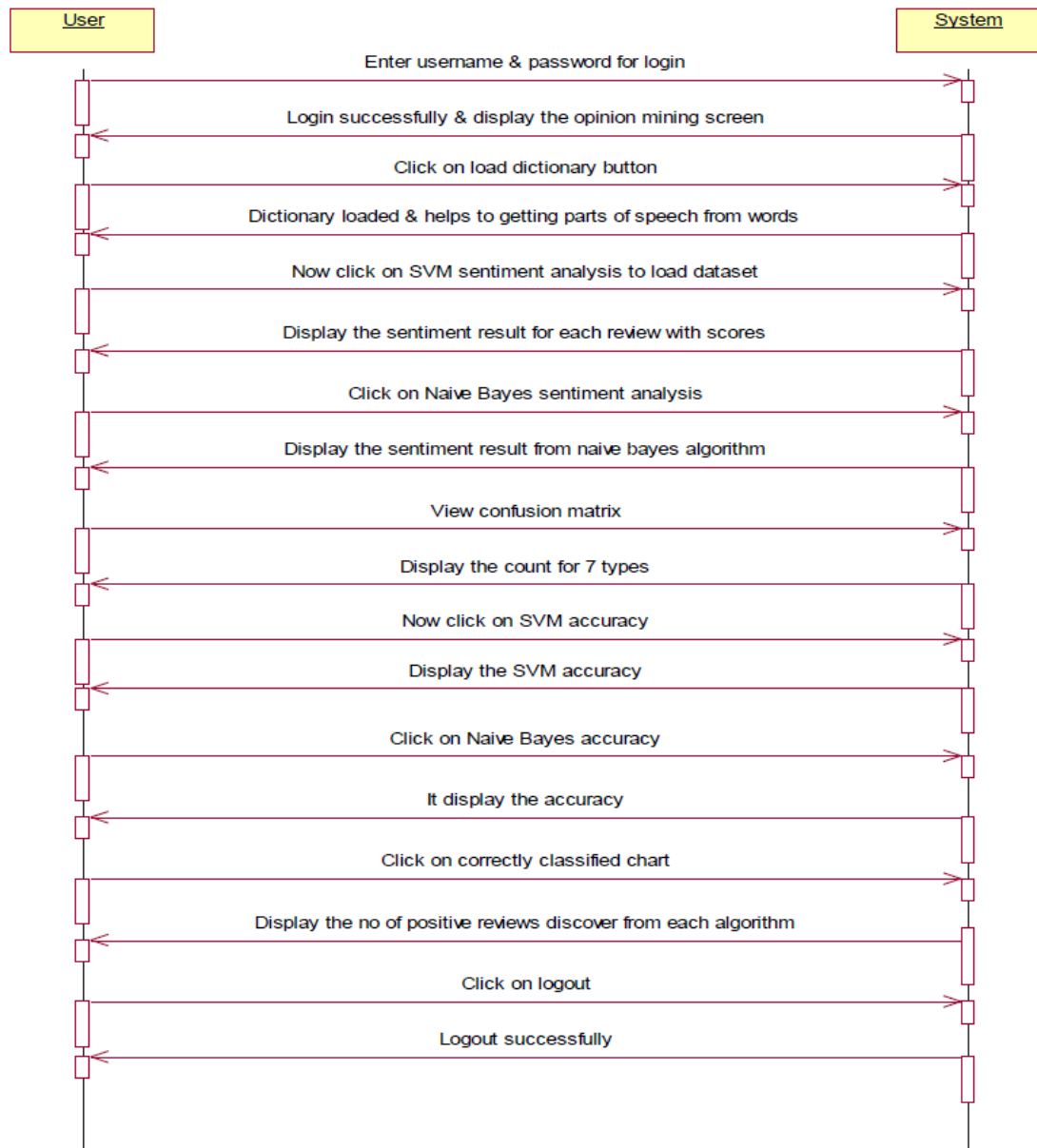


Рисунок 2.4 – Розроблена діаграма послідовності

Сценарій взаємодії розгортається наступним чином:

1. Введення облікових даних:

- User -->System: Користувач вводить ім'я користувача та пароль для входу.

## 2. Вхід до системи:

- System -->User: Система успішно виконує вхід і відображає екран аналізу думок (opinion mining screen).

## 3. Завантаження словника:

- User -->System: Користувач натискає кнопку "Load Dictionary".

- System -->User: Система завантажує словник, що допомагає отримати частини мови зі слів (попередня обробка тексту та ініціалізація лексичних ресурсів, таких як SentiWordNet).

## 4. Аналіз тональності за допомогою SVM:

- User -->System: Користувач натискає кнопку "SVM sentiment analysis" для завантаження набору даних (dataset).

- System -->User: Система відображає результат аналізу тональності для кожного відгуку разом з оцінками (scores).

## 5. Аналіз тональності за допомогою Наївного Байєса:

- User -->System: Користувач натискає кнопку "Naive Bayes sentiment analysis".

- System -->User: Система відображає результат аналізу тональності, отриманий за допомогою алгоритму Наївного Байєса.

## 6. Перегляд матриці плутанини:

- User -->System: Користувач натискає "View Confusion Matrix".

- System -->User: Система відображає підрахунок (count) для 7 типів (категорій тональності, таких як "сильно-позитивний", "сильно-негативний" тощо) у матриці плутанини, порівнюючи результати обох алгоритмів.

## 7. Перегляд точності SVM:

- User -->System: Користувач натискає кнопку "SVM accuracy".

- System -->User: Система відображає показник точності SVM.

## 8. Перегляд точності Наївного Байєса:

- User -->System: Користувач натискає кнопку "Naive Bayes accuracy".

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						44
Змн.	Арк.	№ докум.	Підпис	Дата		

- System -->User: Система відображає показник точності Наївного Байєса.

9. Перегляд діаграми правильно класифікованих:

- User -->System: Користувач натискає кнопку "Correctly classified chart".

- System -->User: Система відображає кількість позитивних відгуків, виявлених кожним алгоритмом (графічне представлення точності).

10. Вихід з системи:

- User -->System: Користувач натискає кнопку "Logout".

- System -->User: Система успішно виконує вихід.

## 2.6. Проектування діаграми потоків даних (DFD)

Рисунок 2.4 представляє діаграму потоків даних (Data Flow Diagram), що показує, як дані рухаються через систему аналізу тональності користувацьких відгуків.

Основні компоненти DFD:

- Сутності (External Entities) - це User (Користувач) та System (Система).

- Процеси (Processes) - Login, Load Dictionary, SVM Analysis.

- Сховища даних (Data Stores) - Dataset (Набір даних).

- Потіки даних (Data Flows) - зображуються стрілками.

Наведемо детальний опис потоків даних та процесів:

1. Login (Вхід):

- User --> Login: Користувач надає облікові дані для входу.

- Login --> System: Система отримує дані для аутентифікації та обробляє їх.

- System --> User: Потік даних, що вказує на успішний/невдалий вхід або повідомлення.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						45
Змн.	Арк.	№ докум.	Підпис	Дата		

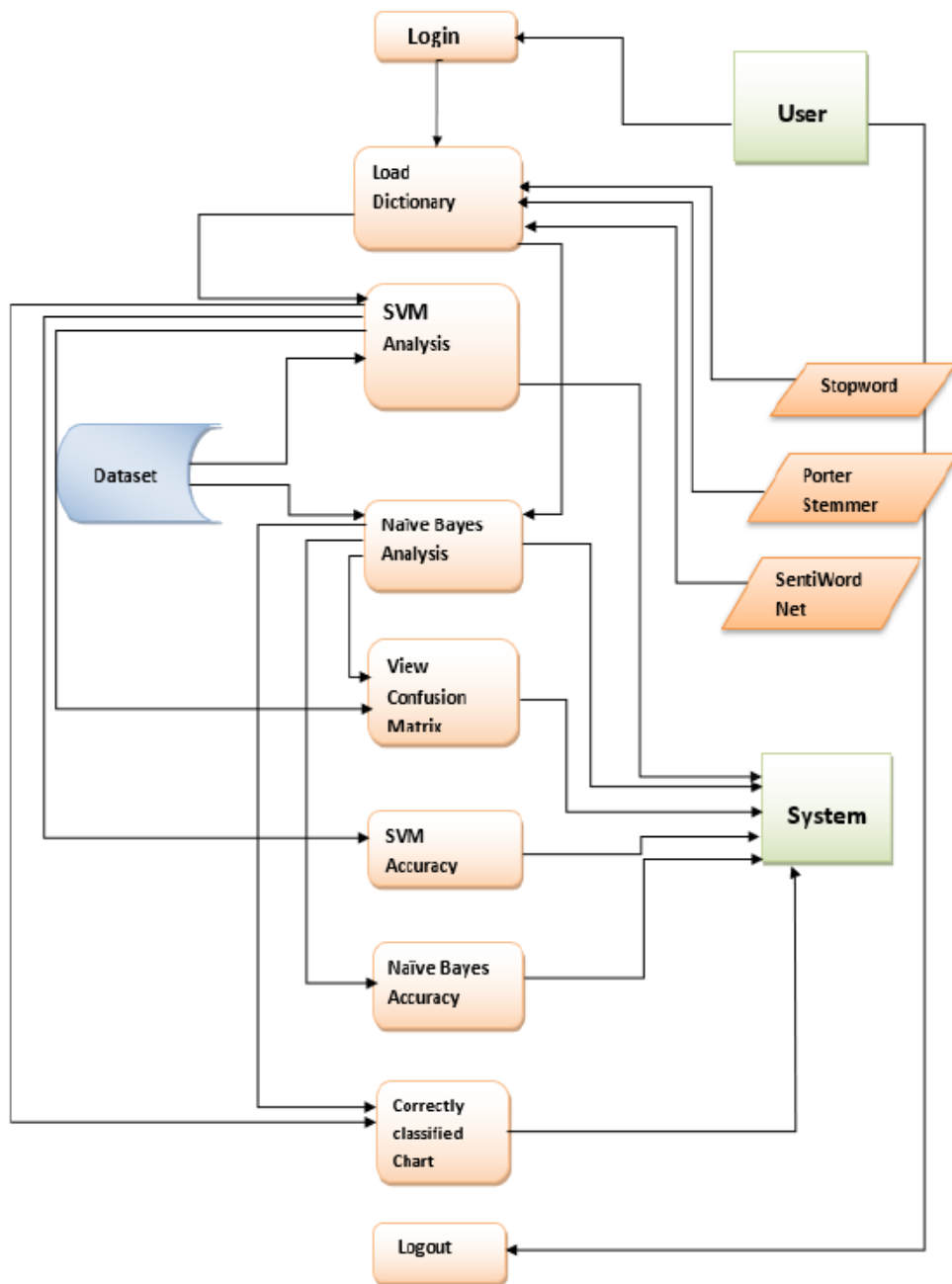


Рисунок 2.5 – Діаграма потоків даних

## 2. Load Dictionary (Завантаження словника):

- User --> Load Dictionary: Користувач ініціює процес завантаження словника.
- Load Dictionary --> Stopword, Porter Stemmer, SentiWordNet: Процес "Load Dictionary" взаємодіє з окремими джерелами/компонентами даних:
- Stopword: Дані про стоп-слова використовуються для їх видалення.

Змн.	Арк.	№ докум.	Підпис	Дата

- Porter Stemmer: Алгоритм стеммінгу Портера використовується для нормалізації слів.

- SentiWordNet: Лексичний ресурс SentiWordNet завантажується та використовується для оцінки тональності.

- Load Dictionary --> SVM Analysis, Naive Bayes Analysis: Оброблені дані (словник, знання про SentiWordNet, стеммінг) передаються до процесів аналізу для подальшого використання.

### 3. SVM Analysis (Аналіз SVM):

- User --> SVM Analysis: Користувач ініціює аналіз за допомогою SVM.

- Dataset <--> SVM Analysis: Процес "SVM Analysis" отримує дані для обробки з "Dataset" і може записувати або оновлювати щось у сховищі.

- SVM Analysis --> System: Результати аналізу SVM передаються системі для подальшого відображення або обробки.

- SVM Analysis --> View Confusion Matrix, SVM Accuracy, Correctly classified Chart: Результати аналізу SVM є вхідними даними для процесів відображення матриці плутанини, розрахунку точності SVM та побудови діаграми.

### 4. Naive Bayes Analysis (Аналіз Наївного Байєса):

- User --> Naive Bayes Analysis: Користувач ініціює аналіз за допомогою Наївного Байєса.

- Dataset <--> Naive Bayes Analysis: Процес "Naive Bayes Analysis" отримує дані з "Dataset".

- Naive Bayes Analysis --> System: Результати аналізу Наївного Байєса передаються системі.

- Naive Bayes Analysis --> View Confusion Matrix, Naive Bayes Accuracy, Correctly classified Chart: Результати аналізу Наївного Байєса також є вхідними даними для відповідних процесів відображення та розрахунків.

### 5. View Confusion Matrix (Перегляд матриці плутанини):

- User --> View Confusion Matrix: Користувач ініціює перегляд матриці.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						47
Змн.	Арк.	№ докум.	Підпис	Дата		

- View Confusion Matrix --> System: Результати (матриця) передаються системі для відображення користувачу.

#### 6. SVM Accuracy (Точність SVM):

- User --> SVM Accuracy: Користувач ініціює розрахунок точності SVM.

- SVM Accuracy --> System: Розрахована точність передається системі.

#### 7. Naive Bayes Accuracy (Точність Наївного Байєса):

- User --> Naive Bayes Accuracy: Користувач ініціює розрахунок точності Наївного Байєса.

- Naive Bayes Accuracy --> System: Розрахована точність передається системі.

#### 8. Correctly classified Chart (Діаграма правильно класифікованих):

- User --> Correctly classified Chart: Користувач ініціює побудову діаграми.

- Correctly classified Chart --> System: Дані для діаграми передаються системі для візуалізації.

#### 9. Logout (Вихід):

- User --> Logout: Користувач ініціює вихід із системи.

- Logout --> System: Система обробляє запит на вихід.

Дана DFD демонструє логічний потік даних у системі визначення тональності. Вона показує, як вхідні дані від користувача та сховищ обробляються різними функціональними модулями (процесами), такими як аналіз SVM або Наївного Байєса, і як кінцеві результати передаються назад до користувача через "System".

## 2.7. Розробка діаграми активності

На рисунку 2.6 показано діаграму активності (Activity Diagram), яка відображає потік управління або послідовність дій у системі визначення

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						48
Змн.	Арк.	№ докум.	Підпис	Дата		

тональності користувацьких відгуків. Вона схожа на блок-схему, але орієнтована на дії та рішення, які приймаються в процесі.

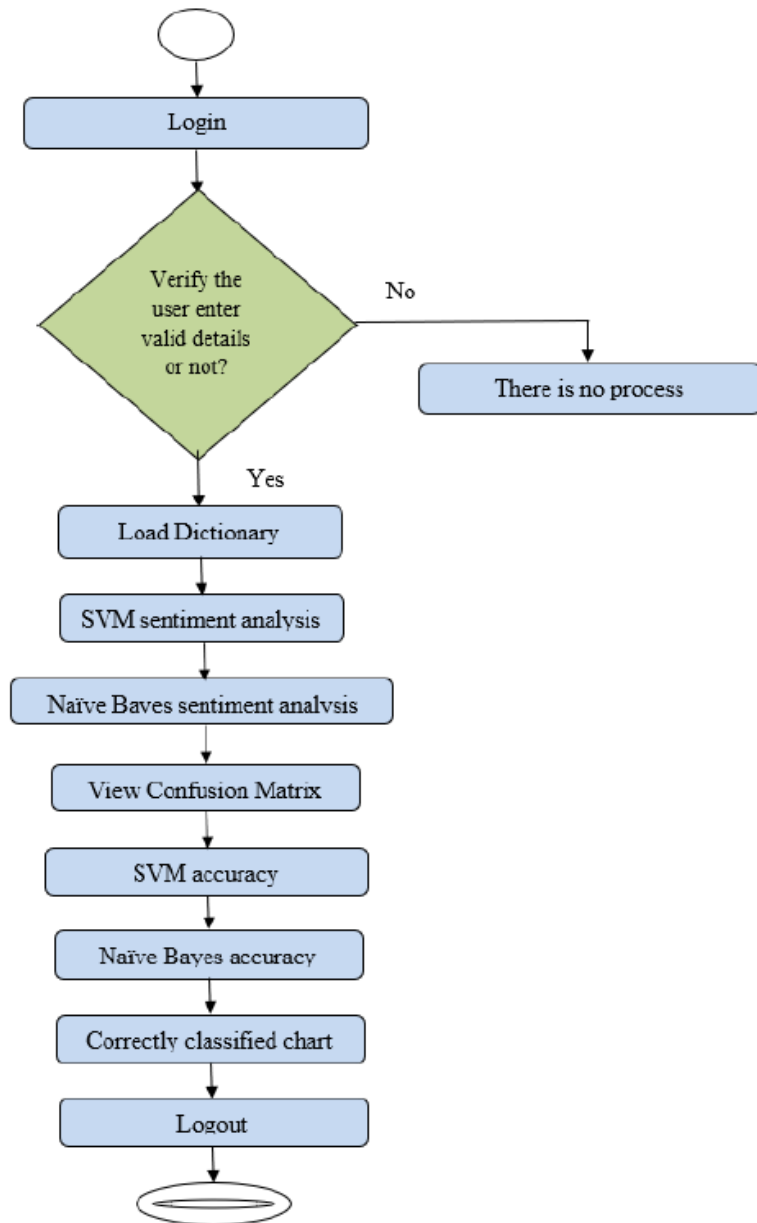


Рисунок 2.6 – Розроблена діаграма активності

Виконаємо детальний опис потоку активності:

1. Початок: Потік активності починається з початкового вузла.
2. Login (Вхід до системи): Перша дія — це вхід до системи.
3. Verify the user enter valid details or not? (Перевірка введених даних користувача): Після спроби входу система переходить до вузла рішення.

3.1. "No" (Ні): Якщо введені дані недійсні, потік переходить до стану "There is no process" (Немає процесу), що вказує на припинення подальших дій до коректного входу.

3.2. "Yes" (Так): Якщо дані дійсні, потік продовжується до наступної дії.

4. Load Dictionary (Завантажити словник): Після успішної аутентифікації виконується дія завантаження словника, що включає попередню обробку даних.

5. SVM sentiment analysis (Аналіз тональності за допомогою SVM): Наступна дія — виконання аналізу тональності із застосуванням алгоритму Метод опорних векторів (SVM).

6. Naive Bayes sentiment analysis (Аналіз тональності за допомогою Наївного Байєса): Після аналізу SVM виконується аналіз тональності за допомогою алгоритму Наївного Байєса.

7. View Confusion Matrix (Переглянути матрицю плутанини): Далі відбувається дія перегляду матриці плутанини, яка порівнює результати обох алгоритмів.

8. SVM accuracy (Точність SVM): Виконується дія розрахунку та відображення точності моделі SVM.

9. Naive Bayes accuracy (Точність Наївного Байєса): Після цього розраховується та відображається точність моделі Наївного Байєса.

10. Correctly classified chart (Діаграма правильно класифікованих): Наступна дія — побудова та відображення діаграми, що візуалізує порівняльну точність класифікації.

11. Logout (Вихід з системи): Завершальна дія, що дозволяє користувачеві вийти з системи.

12. Кінець: Потік активності завершується кінцевим вузлом.

Ця діаграма активності ілюструє послідовність виконання основних операцій у системі. Вона показує, які дії повинні бути виконані і в якому

					БР.ІП – 40.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		50

порядку, а також відображає точку прийняття рішення (перевірка входу), яка може змінити потік виконання. Це допомагає в розумінні бізнес-логіки та робочого процесу системи обробки user feedback.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		51

# РОЗДІЛ 3. ПРОГРАМНА РЕАЛІЗАЦІЯ СИСТЕМИ ОБРОБКИ ДАНИХ З USER FEEDBACK З ВИКОРИСТАННЯМ МЕТОДІВ МАШИННОГО НАВЧАННЯ

## 3.1. Розробка діаграми компонентів системи

Діаграма компонентів (Component Diagram), що подана на рисунку 3.1 відображає структурні відносини між фізичними компонентами пропонованої системи обробки даних user feedback.

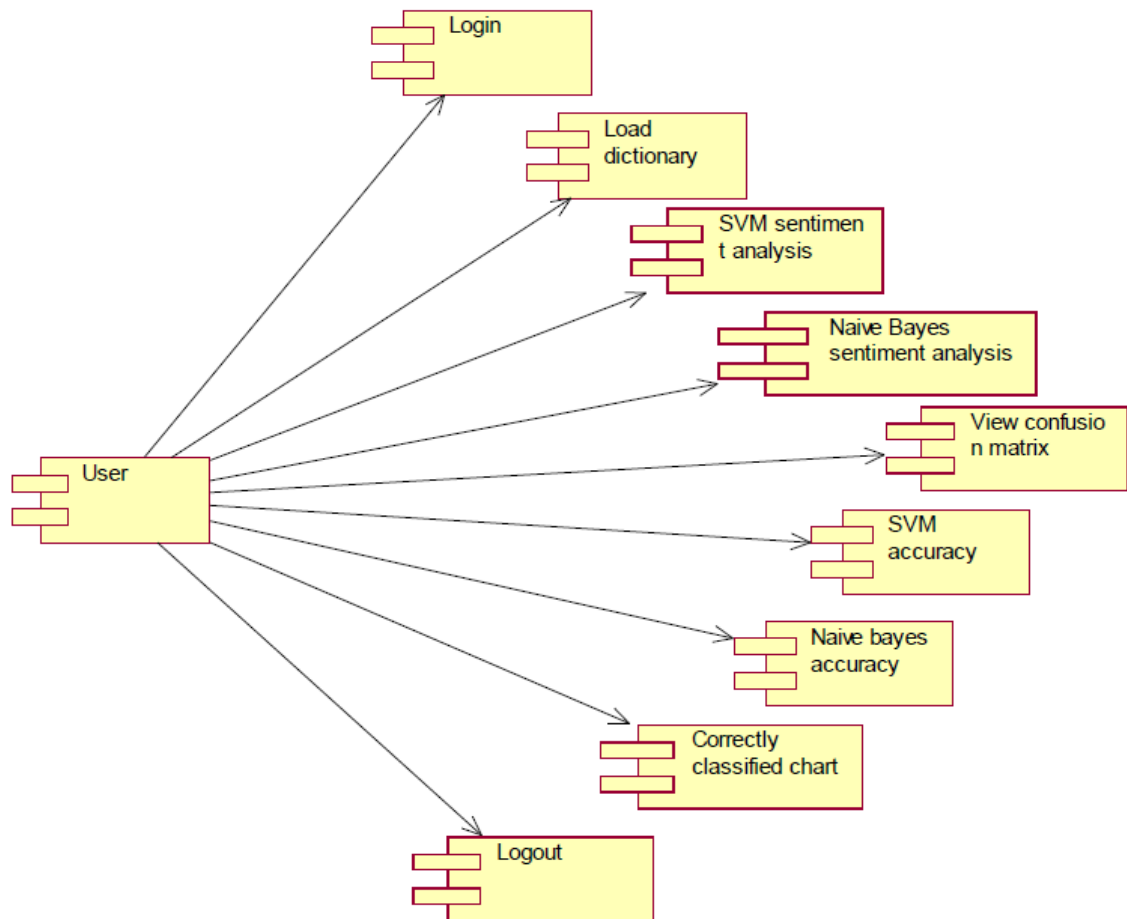


Рисунок 3.1 – Діаграма компонентів

На цій діаграмі представлено один основний компонент — "User", який, по суті, виступає як адаптер або інтерфейс користувача, що взаємодіє з

іншими функціональними компонентами системи. Інші компоненти представляють основні модулі або сервіси системи аналізу тональності.

1. User (Компонент інтерфейсу користувача):

- Цей компонент є центральним і представляє рівень Presentation Layer або GUI-компонент, через який користувач взаємодіє з системою.

- Він має вихідні зв'язки (стрілки) до всіх інших функціональних компонентів, що вказує на те, що він ініціює виклики або використовує сервіси, які надаються цими компонентами.

2. Login (Компонент автентифікації):

- Відповідає за обробку входу користувача. "User" ініціює цей компонент.

3. Load Dictionary (Компонент завантаження словника):

- Відповідає за завантаження та попередню обробку необхідних лінгвістичних ресурсів (стоп-слів, стеммера Портера, SentiWordNet). "User" ініціює цей компонент.

4. SVM Sentiment Analysis (Компонент аналізу тональності SVM):

- Реалізує логіку аналізу тональності із застосуванням методу опорних векторів. "User" ініціює цей аналіз.

5. Naive Bayes Sentiment Analysis (Компонент аналізу тональності Наївного Байєса):

- Реалізує логіку аналізу тональності із застосуванням алгоритму Наївного Байєса. "User" ініціює цей аналіз.

6. View Confusion Matrix (Компонент відображення матриці плутанини):

- Відповідає за генерацію та відображення матриці плутанини, що порівнює результати SVM та Наївного Байєса. "User" ініціює цей компонент.

7. SVM Accuracy (Компонент розрахунку точності SVM):

- Розраховує та надає показники точності для моделі SVM. "User" ініціює цей розрахунок.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						53
Змн.	Арк.	№ докум.	Підпис	Дата		

8. Naive Bayes Accuracy (Компонент розрахунку точності Наївного Байєса):

- Розраховує та надає показники точності для моделі Наївного Байєса. "User" ініціює цей розрахунок.

9. Correctly Classified Chart (Компонент побудови діаграми):

- Відповідає за візуалізацію (графічне представлення) результатів точності. "User" ініціює цей компонент.

10. Logout (Компонент виходу):

- Обробляє процес виходу користувача із системи. "User" ініціює цей компонент.

Представлення "User" як компонента, що безпосередньо взаємодіє з іншими модулями, вказує на те, що він є основним інтерфейсним компонентом системи, який агрегує та надає доступ до всього функціоналу. Зв'язки переважно односторонні, від "User" до функціональних компонентів, що означає про ініціювання дій з боку користувацького інтерфейсу.

## 3.2. Розробка інтерфейсу входу та головної сторінки системи

### 3.2.1. Екран входу (Login Screen)

Екран входу є початковою точкою взаємодії користувача із програмною системою.

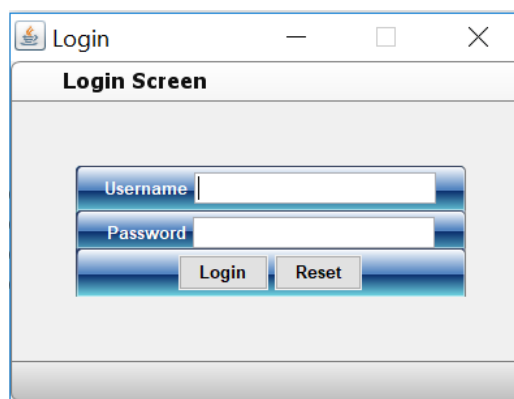


Рисунок 3.2 – Екран входу

Він надає функціонал для автентифікації користувача, дозволяючи йому або увійти до програми, або ініціювати процедуру скидання імені користувача та/або пароля. У випадку некоректного введення облікових даних (неправильне ім'я користувача або пароль, або їх відсутність), система генерує та відображає відповідне повідомлення про помилку. Візуальне представлення інтерфейсу екрану входу зображено на рисунку 3.2.

### 3.2.2. Головна сторінка (Main Page)

Після успішної автентифікації користувача система перенаправляє його на головну сторінку. Верхня панель цього інтерфейсу містить назву проекту, що забезпечує чітку ідентифікацію програми. Нижня панель головної сторінки містить вісім функціональних кнопок, кожна з яких відповідає за активацію окремого етапу або алгоритму, що використовуються для перетворення користувацьких відгуків у кількісний рейтинг. Детальний вигляд головної сторінки представлений на рисунку 3.3.

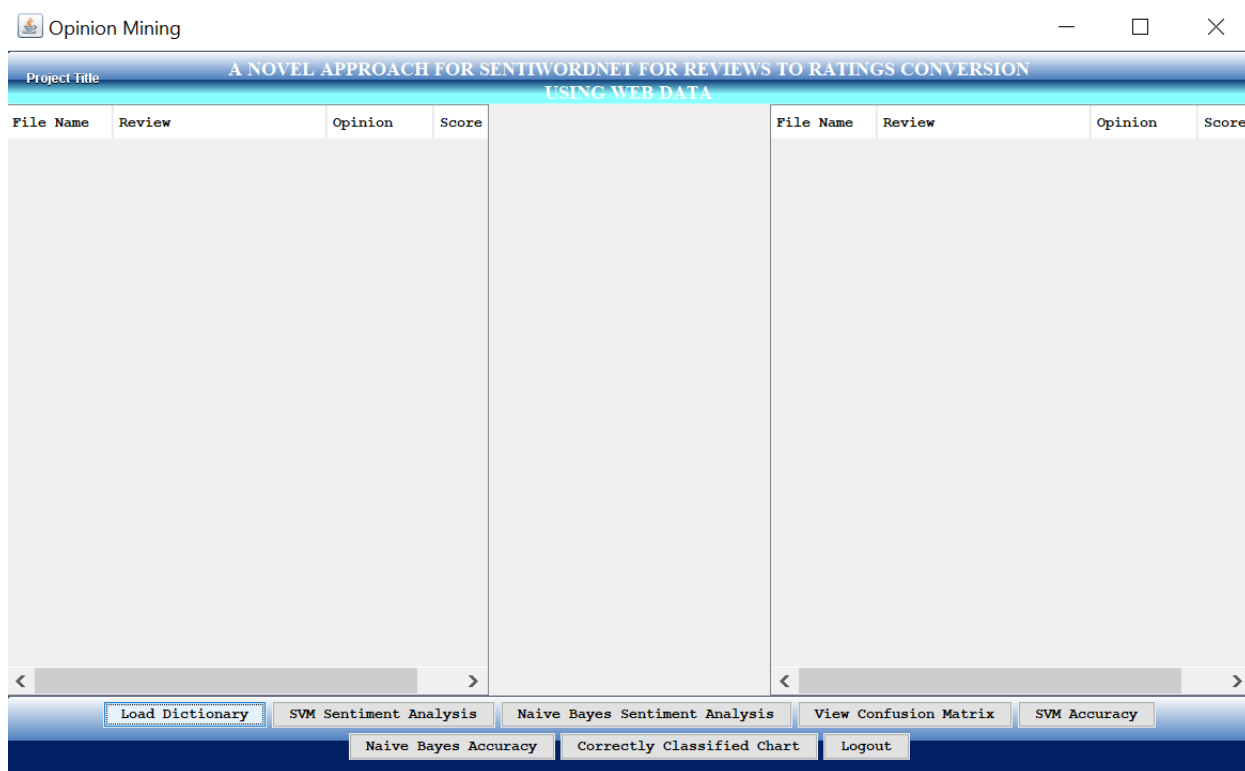


Рисунок 3.3 – Вигляд головної сторінки

Центральна частина екрану містить дві симетричні панелі відображення даних, призначені для представлення результатів аналізу. Кожна з цих панелей має ідентичні заголовки колонок:

- File Name (Ім'я файлу) - для відображення назви файлу або ідентифікатора набору даних, з якого взято відгук.

- Review (Відгук) - призначено для відображення текстового змісту користувацького відгуку.

- Opinion (Думка) - тут буде відображатися класифікована тональність відгуку (наприклад, "позитивний", "негативний", "нейтральний"), визначена відповідним алгоритмом.

- Score (Оцінка) - призначено для відображення числової оцінки тональності, отриманої від SentiWordNet або внутрішнього механізму оцінювання.

У нижній частині інтерфейсу розташована панель з вісьмома кнопками, які відповідають за основний функціонал системи, як це було описано в діаграмі варіантів використання та специфікації вимог:

- Load Dictionary (Завантажити словник) - для ініціалізації лінгвістичних ресурсів.

- SVM Sentiment Analysis (Аналіз тональності SVM) - для запуску аналізу тональності за допомогою алгоритму SVM.

- Naive Bayes Sentiment Analysis (Аналіз тональності Наївного Байєса) - для запуску аналізу тональності за допомогою алгоритму Наївного Байєса.

- View Confusion Matrix (Переглянути матрицю плутанини) - для відображення порівняльної ефективності класифікації.

- SVM Accuracy (Точність SVM) - для відображення показника точності алгоритму SVM.

- Naive Bayes Accuracy (Точність Наївного Байєса) - для відображення показника точності алгоритму Наївного Байєса.

- Correctly Classified Chart (Діаграма правильно класифікованих) - для візуалізації порівняльної точності.

Дві робочі панелі дозволяють зручно порівнювати результати, отримані різними методами аналізу.

### 3.3. Опис функціональних модулів та експериментальних результатів

#### 3.3.1. Модуль завантаження словника

Функція "Завантаження словника" ініціює критично важливий етап попередньої обробки даних, що включає три ключові операції: видалення стоп-слів (виключення з тексту частотних, але малоінформативних слів), застосування алгоритму стеммінгу Портера (редукція слів до їхньої кореневої форми) та завантаження лексичного ресурсу SentiWordNet. Після успішного завершення цих операцій користувачеві відображається системне повідомлення, що підтверджує готовність словника до використання. Інтерфейс модуля завантаження словника представлений на рисунку 3.4.

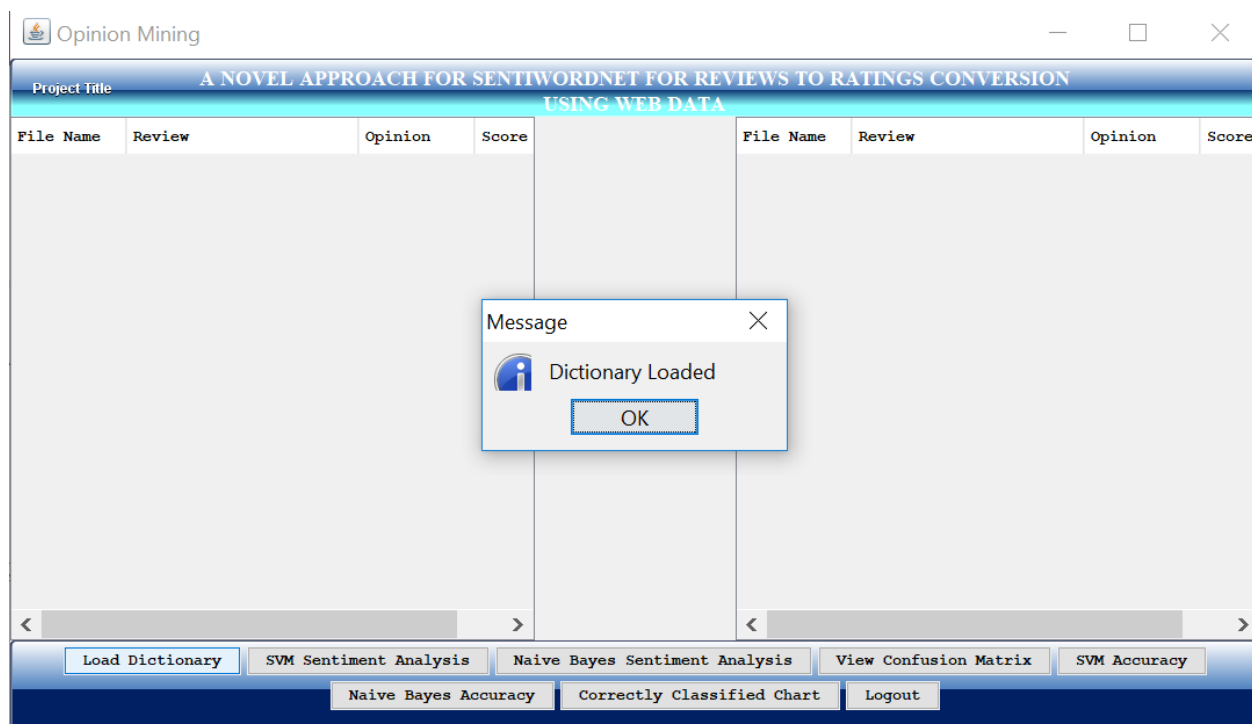


Рисунок 3.4 - Інтерфейс модуля завантаження словника

Змн.	Арк.	№ докум.	Підпис	Дата

### 3.3.2. Модуль аналізу тональності за допомогою методу опорних векторів (SVM)

Модуль "Аналіз тональності SVM" дозволяє користувачеві вибрати конкретний набір даних із доступного списку для розрахунку тональних рейтингів. Нижче представлено результати експериментів, проведених для різних наборів даних:

File Name	Review	Opinion	Score
0_9.txt	bromwell high cartoon...	Negative	0.2
10_9.txt	male women movies spe...	Negative	1.4
11_9.txt	stanley iris acting g...	Positive	1.4
12_9.txt	stanley iris acting g...	Positive	1.4
13_7.txt	production quality ca...	Negative	0.9
14_10.txt	film special place he...	Positive	2.4
15_7.txt	guess film magic don ...	Negative	0.4
16_7.txt	found romance drama n...	Negative	3.2
17_9.txt	complex film explores...	Negative	-2.
18_7.txt	stanley iris heart wa...	Positive	1.3
19_10.txt	read comments tomreyn...	Negative	2.1
1_7.txt	adult comedy cartoons...	Negative	0.5
20_9.txt	stanley iris show tri...	Positive	3.1

Рисунок 3.5 - Аналіз тональності SVM для серіалу "Bromwell High"

На цьому етапі було проаналізовано відгуки, що стосуються анімаційного розважального серіалу "Bromwell High", демонструючи здатність системи обробляти специфічний контент.

На рисунку 3.6 показано аналіз тональності SVM для пристрою iPhone 8+. Результати цього аналізу ілюструють застосування системи до відгуків про технічний пристрій, підтверджуючи її універсальність.

На рисунку 3.7 подано аналіз тональності SVM для фільму "Темний лицар". Цей приклад показує ефективність алгоритму SVM у визначенні тональності відгуків, що стосуються художніх фільмів.

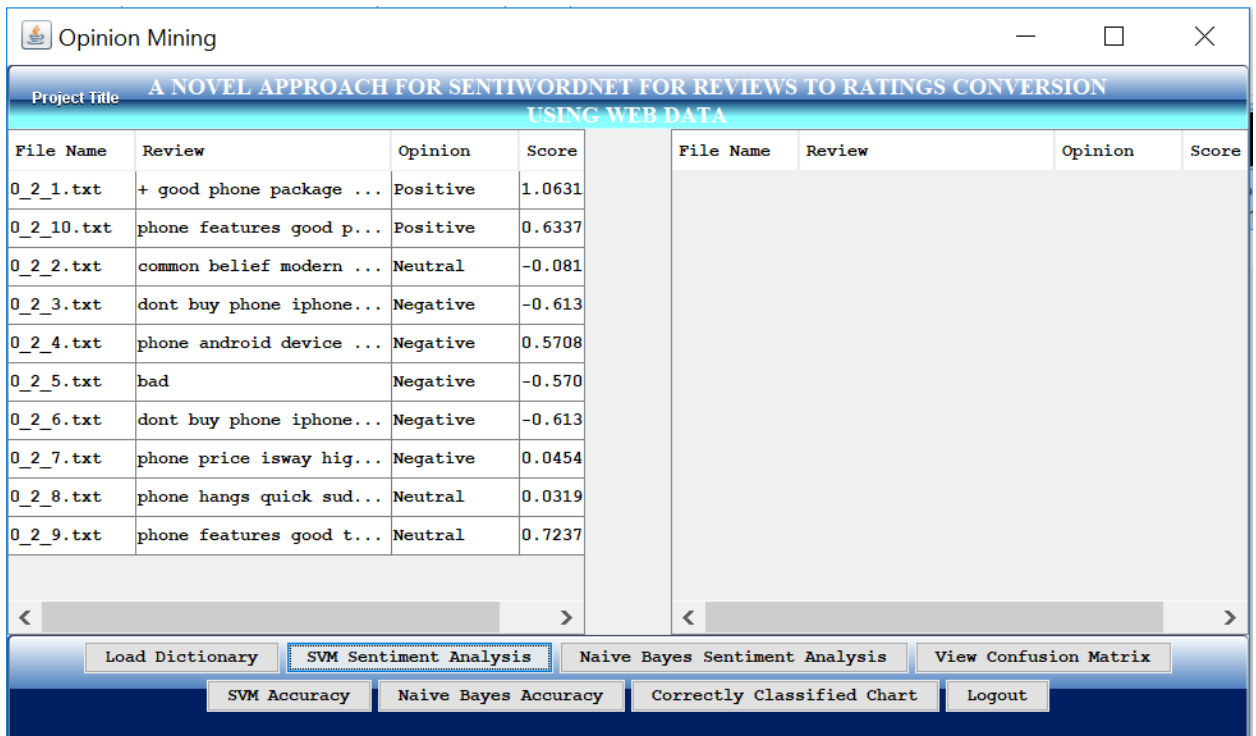


Рисунок 3.6 - Аналіз тональності SVM для пристрою iPhone

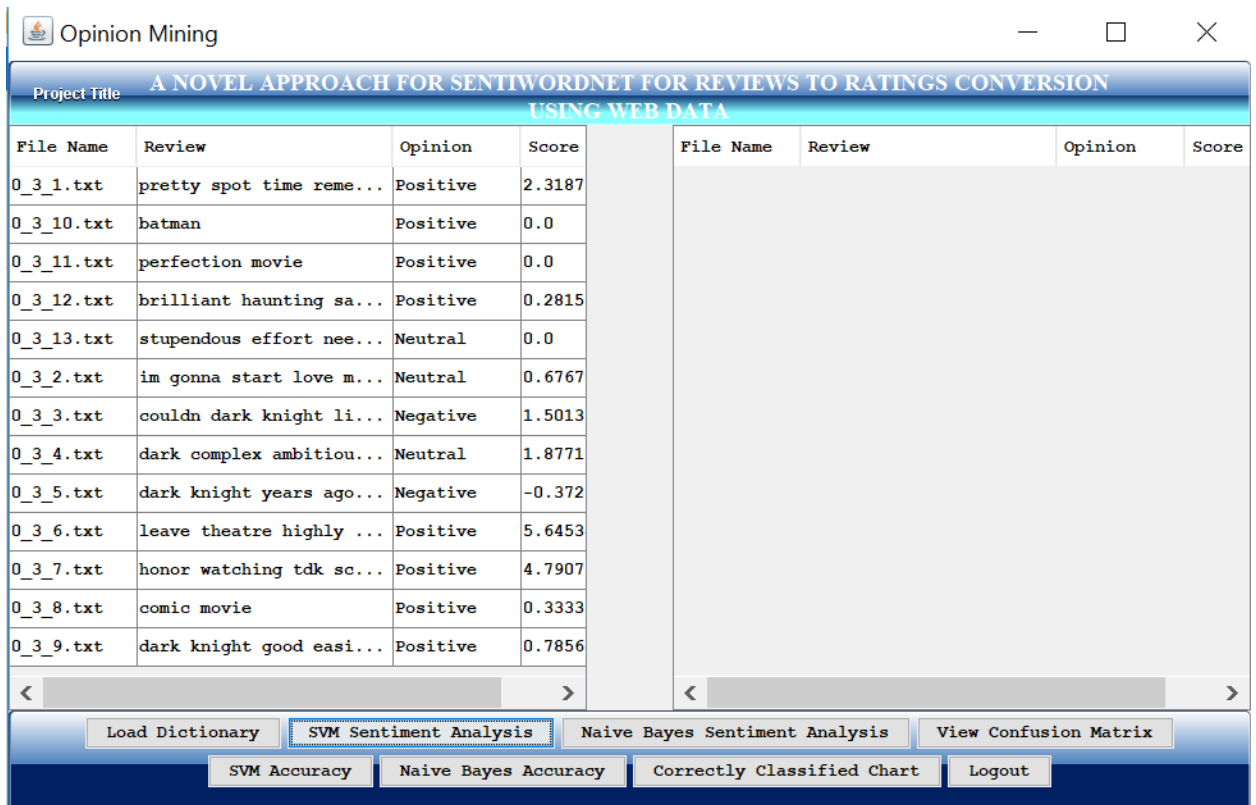


Рисунок 3.7 - Аналіз тональності SVM для фільму "Темний лицар"

### 3.3.3. Модуль аналізу тональності за допомогою Наївного Байєса

Модуль "Аналіз тональності Наївного Байєса" також надає можливість вибору набору даних для розрахунку тональних рейтингів. Нижче наведені приклади застосування цього алгоритму.

На рисунку 3.8 подано аналіз тональності Наївного Байєса для пристрою iPhone 8+. На цьому етапі було проведено аналіз відгуків про iPhone 8+, що дозволяє порівняти результати з тими, що були отримані за допомогою SVM для того ж набору даних.

File Name	Review	Opinion	Score	File Name	Review	Opinion	Score
0_2_1.txt	+ good phone package ...	Positive	1.0631	0_2_1.txt	+ good phone package ...	Positive	1.0631
0_2_10.txt	phone features good p...	Positive	0.6337	0_2_10.txt	phone features good p...	Positive	0.6337
0_2_2.txt	common belief modern ...	Neutral	-0.081	0_2_2.txt	common belief modern ...	Neutral	-0.081
0_2_3.txt	dont buy phone iphone...	Negative	-0.613	0_2_3.txt	dont buy phone iphone...	Negative	-0.613
0_2_4.txt	phone android device ...	Negative	0.5708	0_2_4.txt	phone android device ...	Positive	0.5708
0_2_5.txt	bad	Negative	-0.570	0_2_5.txt	bad	Negative	-0.570
0_2_6.txt	dont buy phone iphone...	Negative	-0.613	0_2_6.txt	dont buy phone iphone...	Negative	-0.613
0_2_7.txt	phone price isway hig...	Negative	0.0454	0_2_7.txt	phone price isway hig...	Negative	0.0454
0_2_8.txt	phone hangs quick sud...	Neutral	0.0319	0_2_8.txt	phone hangs quick sud...	Neutral	0.0319
0_2_9.txt	phone features good t...	Neutral	0.7237	0_2_9.txt	phone features good t...	Neutral	0.7237

Рисунок 3.8 - Аналіз тональності Наївного Байєса для пристрою iPhone

На рисунку 3.9 подано аналіз тональності Наївного Байєса для серіалу "Bromwell High". Результати аналізу відгуків про "Bromwell High" за допомогою Наївного Байєса.

Opinion Mining

Project Title: A NOVEL APPROACH FOR SENTIWORDNET FOR REVIEWS TO RATINGS CONVERSION USING WEB DATA

File Name	Review	Opinion	Score
0_9.txt	bromwell high cartoon...	Negative	0.2
10_9.txt	male women movies spe...	Negative	1.4
11_9.txt	stanley iris acting g...	Positive	1.4
12_9.txt	stanley iris acting g...	Positive	1.4
13_7.txt	production quality ca...	Negative	0.9
14_10.txt	film special place he...	Positive	2.4
15_7.txt	guess film magic don ...	Negative	0.4
16_7.txt	found romance drama n...	Negative	3.2
17_9.txt	complex film explores...	Negative	-2.
18_7.txt	stanley iris heart wa...	Positive	1.3
19_10.txt	read comments tomreyn...	Negative	2.1
1_7.txt	adult comedy cartoons...	Negative	0.5
20_9.txt	stanley iris show tri...	Positive	3.1

Buttons: Load Dictionary, SVM Sentiment Analysis, Naive Bayes Sentiment Analysis, View Confusion Matrix, SVM Accuracy, Naive Bayes Accuracy, Correctly Classified Chart, Logout

Рисунок 3.9 - Аналіз тональності Наївного Байєса для серіалу "Bromwell High"

Opinion Mining

Project Title: A NOVEL APPROACH FOR SENTIWORDNET FOR REVIEWS TO RATINGS CONVERSION USING WEB DATA

File Name	Review	Opinion	Score
0_3_1.txt	pretty spot time reme...	Positive	2.3187
0_3_10.txt	batman	Positive	0.0
0_3_11.txt	perfection movie	Positive	0.0
0_3_12.txt	brilliant haunting sa...	Positive	0.2815
0_3_13.txt	stupendous effort nee...	Neutral	0.0
0_3_2.txt	im gonna start love m...	Neutral	0.6767
0_3_3.txt	couldn dark knight li...	Negative	1.5013
0_3_4.txt	dark complex ambitiou...	Neutral	1.8771
0_3_5.txt	dark knight years ago...	Negative	-0.372
0_3_6.txt	leave theatre highly ...	Positive	5.6453
0_3_7.txt	honor watching tdk sc...	Positive	4.7907
0_3_8.txt	comic movie	Positive	0.3333
0_3_9.txt	dark knight good easi...	Positive	0.7856

Buttons: Load Dictionary, SVM Sentiment Analysis, Naive Bayes Sentiment Analysis, View Confusion Matrix, SVM Accuracy, Naive Bayes Accuracy, Correctly Classified Chart, Logout

Рисунок 3.10 - Аналіз тональності Наївного Байєса для фільму "Темний лицар"

На рисунку 3.10 представлено аналіз тональності Наївного Байєса для фільму "Темний лицар". Цей приклад демонструє застосовність алгоритму Наївного Байєса для аналізу відгуків про кінематографічні твори.

Представлені приклади свідчать про гнучкість системи у застосуванні різних алгоритмів машинного навчання для аналізу тональності широкого спектра користувацьких відгуків про різні типи продуктів та контенту.

### 3.4. Результати експериментів та аналіз ефективності алгоритмів

#### 3.4.1. Матриця плутанини (Confusion Matrix)

Обидва алгоритми аналізу тональності, Метод опорних векторів (SVM) та Наївний Байєс, виконують класифікацію користувацьких відгуків за семи категоріями тональності, що варіюються від "сильно позитивних" до "сильно негативних". Матриця плутанини в даному контексті є інструментом для порівняльного аналізу рейтингів, присвоєних кожному відгуку алгоритмами SVM та Наївного Байєса для заданого набору даних. Це дозволяє візуалізувати відповідність або розбіжності у класифікації між двома моделями. Приклад матриці плутанини для відгуків про серіал "Bromwell High" представлений на Рисунку 3.11.

Algorithm Type	Neutral	Positive	Weak Positive	Strong Positive	Negative	Weak Negative	Strong Negative
SVM	4	0	2	7	0	0	1
Naive Bayes	4	2	0	7	0	0	1

Рисунок 3.11 - Приклад матриці плутанини для відгуків про серіал "Bromwell High"

### 3.4.2. Точність SVM (SVM Accuracy)

Точність SVM визначається як відношення кількості позитивно класифікованих відгуків до загальної кількості відгуків у заданому наборі даних. Результати розрахунку точності SVM для різних наборів даних демонструють наступні показники на рисунку.

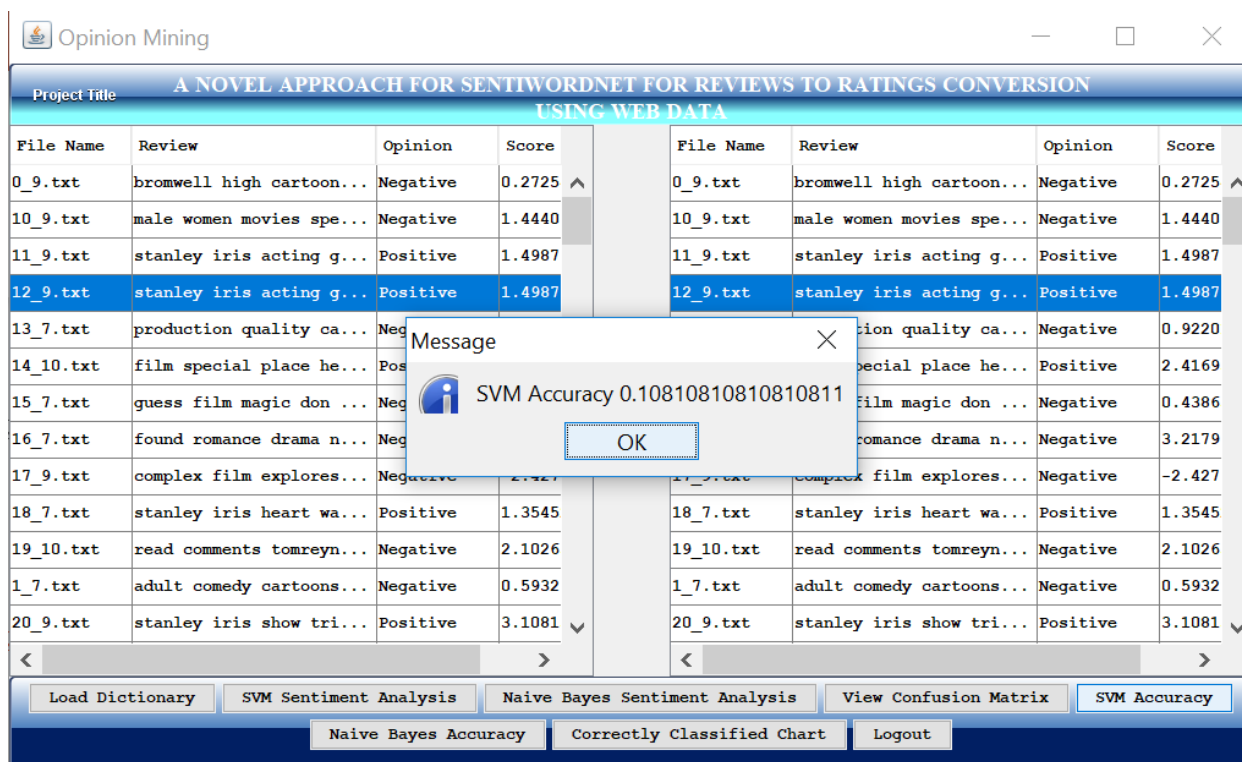


Рисунок 3.12 - Точність SVM для анімаційного серіалу "Bromwell High"

Рисунок 3.12 показує для анімаційного серіалу "Bromwell High" точність SVM, що склала 10.8% позитивних відгуків. Цей низький показник може свідчити про переважання негативних/нейтральних відгуків або про обмеження моделі SVM у класифікації саме цього типу контенту.

Рисунок 3.13 представляє точність SVM для відгуків про фільм "Темний лицар".

Рисунок 3.14 демонструє точність SVM для відгуків про пристрій "iPhone 8+".

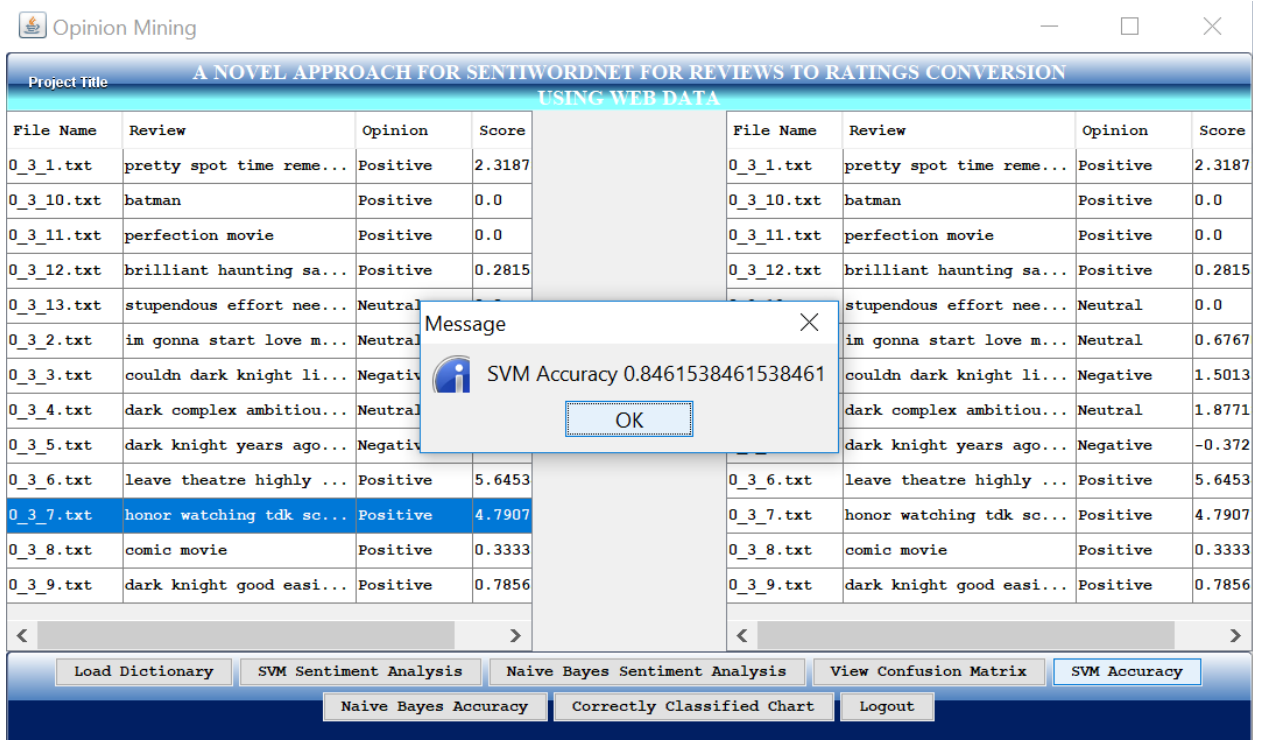


Рисунок 3.13 – Представлення точності SVM для відгуків про фільм "Темний лицар"

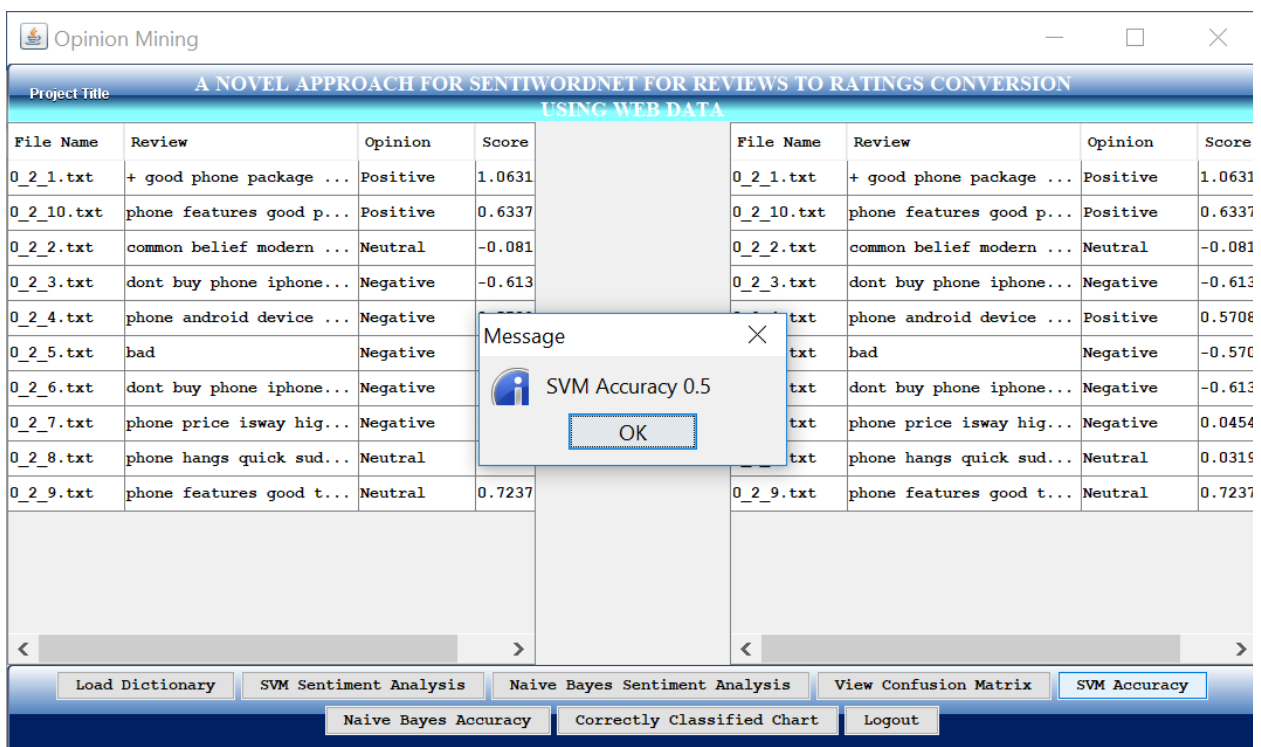


Рисунок 3.15 – Представлення точності SVM для відгуків про пристрій "iPhone 8+"

### 3.4.3. Точність Наївного Байєса (Naive Bayes Accuracy)

Точність Наївного Байєса аналогічно визначається як відношення кількості позитивно класифікованих відгуків до загальної кількості відгуків у наборі даних. Результати для Наївного Байєса є наступними

На рисунку 3.16 для фільму "Темний лицар" показана точність Наївного Байєса яка показала, що 84.61% відгуків є позитивними. Цей значно вищий показник порівняно з "Bromwell High" підкреслює варіативність ефективності алгоритмів залежно від специфіки даних.

The screenshot shows the 'Opinion Mining' application window. The title bar reads 'Opinion Mining'. The main window title is 'A NOVEL APPROACH FOR SENTIWORDNET FOR REVIEWS TO RATINGS CONVERSION USING WEB DATA'. A message box is overlaid on the interface, displaying 'Naive Bayes Accuracy 0.8461538461538461' with an 'OK' button. The background table contains the following data:

File Name	Review	Opinion	Score
0_3_1.txt	pretty spot time reme...	Positive	2.3187
0_3_10.txt	batman	Positive	0.0
0_3_11.txt	perfection movie	Positive	0.0
0_3_12.txt	brilliant haunting sa...	Positive	0.2815
0_3_13.txt	stupendous effort nee...	Neutral	0.0
0_3_2.txt	im gonna start love m...	Neutral	0.6767
0_3_3.txt	couldn dark knight li...	Negative	1.5013
0_3_4.txt	dark complex ambitiou...	Neutral	1.8771
0_3_5.txt	dark knight years ago...	Negative	-0.372
0_3_6.txt	leave theatre highly ...	Positive	5.6453
0_3_7.txt	honor watching tdk sc...	Positive	4.7907
0_3_8.txt	comic movie	Positive	0.3333
0_3_9.txt	dark knight good easi...	Positive	0.7856

Рисунок 3.16 - Точність Наївного Байєса для фільму "Темний лицар"

Відповідно рисунку 3.17 представляє точність Наївного Байєса для відгуків про серіал "Bromwell High".

Рисунок 3.18 демонструє точність Наївного Байєса для відгуків про пристрій "iPhone 8+".

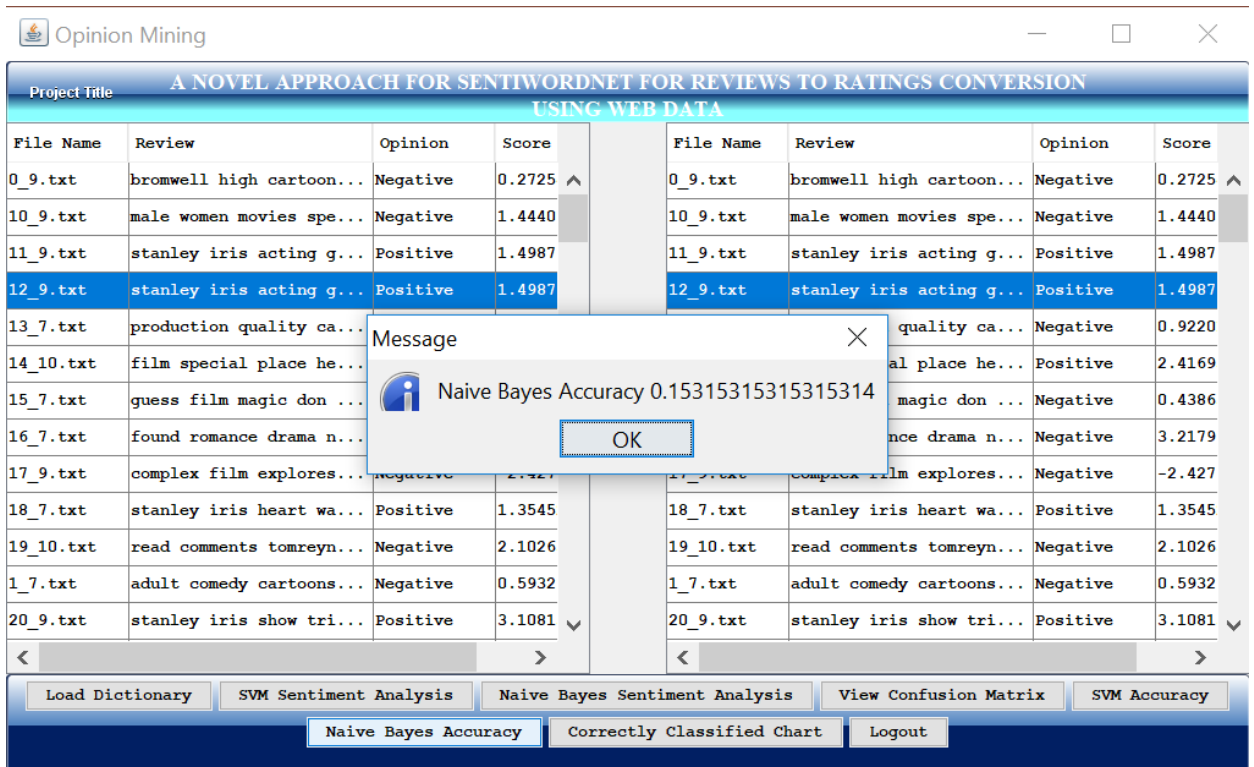


Рисунок 3.17 - Точність Наївного Байєса для відгуків про серіал "Bromwell High"

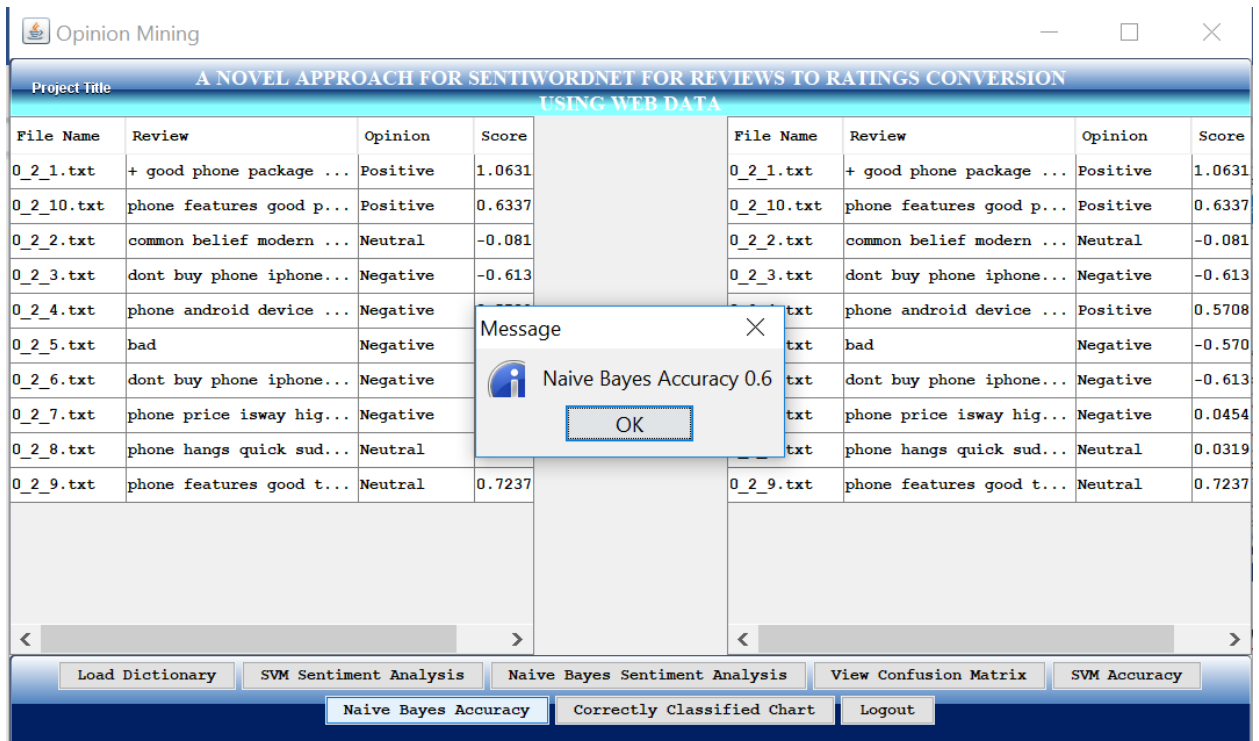


Рисунок 3.18 - Точність Наївного Байєса для відгуків про пристрій "iPhone 8+"

### 3.4.4. Діаграма правильно класифікованих (Correctly Classified Chart)

Діаграма правильно класифікованих надає графічне представлення порівняльної точності алгоритмів SVM та Наївного Байеса. Ця візуалізація дозволяє користувачеві швидко оцінити відносну ефективність обох моделей у класифікації тональності. Приклад такої діаграми для анімаційного серіалу "Bromwell High" наведений на рисунку 3.19. Це графічне представлення є ключовим для інтегрованої оцінки продуктивності системи.

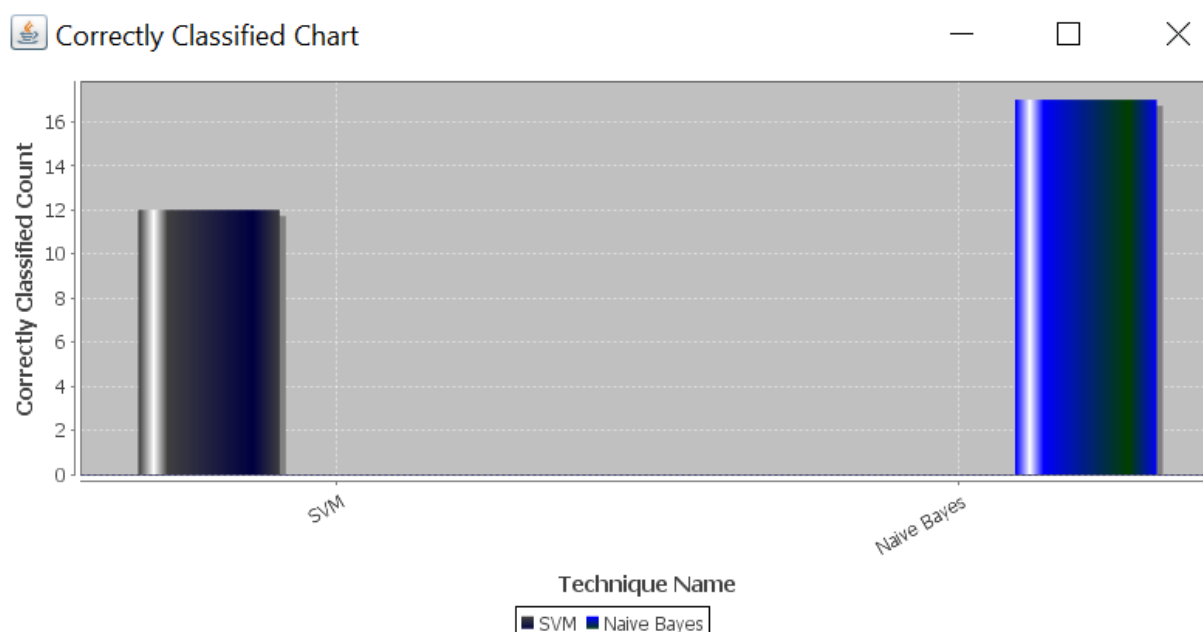


Рисунок 3.19 - Діаграма правильно класифікованих

Рисунок 3.19 представляє "Correctly Classified Chart", що є стовпчастою діаграмою, яка візуально порівнює кількість коректно класифікованих елементів (відгуків) двома різними алгоритмами: SVM (Метод опорних векторів) та Naive Bayes (Наївний Байес).

Стовпець для SVM досягає значення близько 12 на осі Y ("Correctly Classified Count"), стовпець для Naive Bayes досягає значення близько 16.5-17 на осі Y, що помітно вище, ніж у SVM.

Діаграма чітко демонструє, що алгоритм Наївного Байєса значно перевершує SVM за кількістю правильно класифікованих відгуків для цього конкретного набору даних.

Для даного сценарію та набору даних, Наївний Байєс виявився більш ефективним у правильній класифікації відгуків. Це узгоджується з інформацією, яку показано вище, де для "Темного лицаря" точність Наївного Байєса була 84.61%, а для "Bromwell High" точність SVM лише 10.8%.

### 3.5. Тестування системи

Тестування є фундаментальним та інтегральним етапом процесу розробки програмного забезпечення, що має вирішальне значення для забезпечення якості продукту та оптимізації витрат на його подальше обслуговування. В контексті даного проєкту, для валідації коректності функціонування розробленого додатку було проведено тестування прийняття користувачем (User Acceptance Testing - UAT).

#### 3.5.1. Тестування прийняття користувачем (UAT)

Тестування прийняття користувачем є критичним етапом тестування, який проводиться для перевірки відповідності системи вимогам кінцевого користувача в реальних умовах експлуатації. У рамках даного тестування було всебічно перевірено всі компоненти програмного забезпечення. Загалом, було розроблено та виконано дев'ять тестових випадків, які охоплювали функціонал додатку на кожному рівні, починаючи від початкової сторінки входу і до фінального процесу виходу з системи.

Представлені тестові випадки та їх результати підтверджують, що всі основні функціональні компоненти системи коректно виконують свої передбачені операції, що є важливим показником готовності програмного продукту до розгортання та експлуатації.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		68

Детальний перелік тестових випадків, їх опис та очікувані результати представлені в таблиці 3.1.

Таблиця 3.1 - Тестування прийняття користувачем

Ідентифікатор тестового випадку	Назва тестового випадку	Опис тестового випадку	Кроки тестування (Очікуваний результат)	Фактичний результат	Статус тестового випадку	Пріоритет тесту
01	Вхід	Перевірка завантаження даних користувача	Якщо дані не завантажені, подальші операції недоступні; Вхід має бути успішним.	Вхід успішний	Високий	Високий
02	Завантаження словника	Перевірка успішності завантаження словника	Якщо словник не завантажений, неможливо отримати частини мови зі слів; Словник має бути завантажений.	Словник завантажено	Високий	Високий
03	Аналіз тональності SVM	Перевірка завантаження набору даних для аналізу тональності SVM	Якщо набір даних не завантажений, неможливо виявити думки; Відображення результатів тональності для кожного відгуку з оцінками.	Відображення результатів тональності для кожного відгуку з оцінками	Високий	Високий
04	Аналіз тональності Наївного Байеса	Перевірка завантаження набору даних для аналізу тональності Наївного Байеса	Якщо набір даних не завантажений, неможливо виявити думки; Відображення результатів тональності з алгоритму Наївного Байеса.	Відображення результатів тональності з алгоритму Наївного Байеса	Високий	Високий
05	Перегляд матриці плутанини	Перевірка вибору рядка для відображення	Якщо рядок не вибраний, неможливо застосувати	Отримання кількості 7 типів	Високий	Високий

		я матриці плутанини	матрицю плутанини; Отримання кількості 7 типів.			
06	Точність SVM	Перевірка вибору рядка для розрахунку точності SVM	Якщо рядок не вибраний, неможливо розрахувати точність; Відображення точності SVM.	Відображення точності SVM	Високий	Високий
07	Точність Наївного Байеса	Перевірка вибору рядка для розрахунку точності Наївного Байеса	Якщо рядок не вибраний, неможливо розрахувати точність; Відображення точності Наївного Байеса.	Відображення точності Наївного Байеса	Високий	Високий
08	Діаграма правильно класифікованих	Перевірка наявності наборів даних для SVM та Наївного Байеса	Якщо набори даних не існують, неможливо отримати діаграму; Відображення діаграми для кількості позитивних відгуків, виявлених кожним алгоритмом.	Відображення діаграми для кількості позитивних відгуків, виявлених кожним алгоритмом	Високий	Високий
09	Вихід	Перевірка завершення всіх операцій перед виходом	Якщо операції не завершені, неможливо вийти; Вихід з екрану.	Вихід з екрану	Високий	Високий

Розроблений програмний продукт має значну практичну цінність, оскільки він дозволяє користувачам ефективно приймати обґрунтовані рішення щодо придбання товарів або послуг, ґрунтуючись на агрегованому та автоматизованому аналізі онлайн-відгуків, що розміщуються на веб-сайтах. Запропонований підхід продемонстрував свою ефективність шляхом успішного тестування на різномірних наборах даних, що характеризуються різними рівнями тональності:

- Відгуки про анімаційний розважальний серіал, що мали переважно низькі рейтинги.

- Відгуки про технічний пристрій, які характеризувалися нейтральною тональністю.

- Відгуки про фільм, що мали високі позитивні рейтинги. Ця різноманітність тестових сценаріїв підтверджує універсальність та надійність системи у обробці відгуків з різним емоційним забарвленням.

Незважаючи на досягнуті результати, існують перспективи для подальшого вдосконалення системи, зокрема у сфері контекстного аналізу тональності. Поточні методи, що ґрунтуються на лексичних ресурсах, таких як SentiWordNet, мають суттєве обмеження: вони не здатні повною мірою враховувати контекст вживання слів у реченні. Це може призводити до неточностей, коли, наприклад, позитивне за змістом речення через наявність заперечення або іронії помилково класифікується як негативне, або навпаки.

Таким чином, підвищення точності поточної системи рейтингу є можливим шляхом інтеграції більш досконалих методів обробки природної мови, які дозволяють розрізняти контекстне значення слів та фраз. Це може включати застосування векторних представлень слів (word embeddings), глибоких нейронних мереж (наприклад, RNN, LSTM, Transformer-моделей) та інших передових технік НЛП, що здатні захоплювати семантичні та синтаксичні залежності у тексті. Розробка та імплементація таких контекстно-обізнаних алгоритмів є пріоритетним напрямком для майбутніх досліджень та подальшого розвитку проекту.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
						71
Змн.	Арк.	№ докум.	Підпис	Дата		

## ВИСНОВКИ

У ході виконання дипломної роботи було проведено комплексне дослідження теоретичних і практичних аспектів застосування методів машинного навчання для обробки та аналізу користувацьких відгуків (user feedback). Результати аналізу підтвердили актуальність проблеми ефективної класифікації текстових даних у контексті зростання обсягів зворотного зв'язку від користувачів у цифрових системах.

У першому розділі роботи здійснено детальний огляд предметної області, пов'язаної з аналізом відгуків, класифікацією їхньої тональності та особливостями впровадження методів машинного навчання. Визначено основні типи та методи збору user feedback, описано ключові підходи до їх аналізу, зокрема з використанням онтологій, таких як SentiWordNet. Здійснено критичний аналіз сучасних інструментів і SaaS-платформ для аналізу відгуків, таких як Brandwatch, Talkwalker та Qualtrics, що дозволило виявити їхні переваги й обмеження у контексті точності класифікації та масштабованості.

У другому розділі розроблено архітектуру програмного забезпечення для автоматизованого аналізу тональності user feedback. Визначено функціональні та нефункціональні вимоги до системи, спроектовано користувацький інтерфейс та побудовано UML-діаграми (діаграми варіантів використання, послідовності, потоків даних, активності), що забезпечують візуалізацію логіки та структури програмного продукту.

У третьому розділі здійснено програмну реалізацію системи, в якій інтегровано два основні алгоритми машинного навчання: метод опорних векторів (SVM) та Наївний Байєс. Проведено експериментальні дослідження з метою оцінки точності та ефективності обраних моделей. Результати класифікації було візуалізовано за допомогою матриць плутанини та відповідних графіків, що дозволило виявити переваги SVM у точності

					БР.ІП – 40.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		72

класифікації тональності відгуків. Проведено тестування системи, включаючи прийняття користувачем (UAT), яке підтвердило працездатність та зручність використання запропонованого рішення.

Таким чином, у роботі доведено доцільність застосування методів машинного навчання для автоматизованої обробки користувацьких відгуків, зокрема для класифікації їхньої тональності. Розроблена система може бути адаптована для використання в бізнес-середовищі, CRM-системах та платформах для аналізу клієнтського досвіду, сприяючи прийняттю обґрунтованих управлінських рішень на основі оброблених текстових даних.

					БР.ІП – 40.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		73

## ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.
2. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Now Publishers Inc.
3. Cambria, E., & White, B. (2014). JumpingNLP: A leap towards a cognitive computing paradigm. IEEE Intelligent Systems, 29(2), 65-72.
4. Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82-89.
5. Turney, P. D. (2002). Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (pp. 417-424).
6. Boiy, E., & Moens, M. F. (2009). A machine learning approach to sentiment analysis in the legal domain. Artificial Intelligence and Law, 17(1), 1-28.
7. Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability from text using machine learning. Proceedings of the 2003 IEEE International Conference on Systems, Man & Cybernetics (Vol. 3, pp. 2650-2655).
8. Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for cross-lingual sentiment classification. ACM Transactions on Information Systems (TOIS), 26(3), 12.
9. Joachims, T. (1999). Transductive Inference for Text Classification using Support Vector Machines. Proceedings of the Sixteenth International Conference on Machine Learning (pp. 200-209).
10. Cortes, C., & Vapnik, V. N. (1995). Support-vector networks. Machine Learning, 20(3), 273-297.

					БР.ІІІ – 40.00.00.000 ІІЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		74

11. McCallum, A., & Nigam, K. (1998). A Comparison of Event Models for Naive Bayes Text Classification. AAI-98 Workshop on Learning for Text Categorization.
12. Lewis, D. D. (1998). Naive (Bayes) At Forty: The Independence Assumption in Information Retrieval. ECML-98.
13. Esuli, A., & Sebastiani, F. (2006). SentiWordNet: A publicly available lexical resource for sentiment analysis. Proceedings of LREC.
14. Ren, X., & Sukthankar, R. (2011). Contextual sentiment analysis for consumer reviews. Proceedings of the IEEE International Conference on Semantic Computing (pp. 222-229).
15. Wiebe, J., Bruce, R., & O'Hara, M. (1999). Development and use of a gold-standard data set for subjectivity classifications. Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics (pp. 308-315).
16. Porter, M. F. (1980). An algorithm for suffix stripping. Program: electronic library and information systems, 14(3), 130-137.
17. Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.
18. Aggarwal, C. C., & Zhai, C. (2012). Mining Text Data. Springer Science & Business Media.
19. Hearst, M. A. (1994). Context and structure in automated text summarization. Proceedings of the AAI Fall Symposium on Text-Based Intelligent Systems.
20. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093-1113.
21. Ghose, S., & Sundararajan, R. (2016). A Comparative Study of Sentiment Analysis using Supervised Machine Learning Algorithms. International Journal of Computer Applications, 144(13).

- 22.Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT.
- 23.Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems, 30.
- 24.Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.
- 25.Liu, B. (2010). Sentiment Analysis and Subjectivity. In N. Indurkha & F. Damerau (Eds.), Handbook of Natural Language Processing (2nd ed., pp. 627-666). CRC Press.
- 26.Plap D. (2017). User Acceptance Testing (UAT) as a quality assurance method for software projects. Proceedings of the International Scientific Conference "Science and Society".
- 27.Lee, S., & Lee, S. K. (2015). Applying user acceptance testing (UAT) in agile software development. Proceedings of the 2015 International Conference on Software Engineering Research, Management and Applications (SERA) (pp. 1-6).
- 28.Al-Omari, M., & Abu-Salem, M. (2012). Software Requirements Specification Document for a Student Information System. International Journal of Modern Education and Computer Science, 4(2), 1-13.
- 29.Al-Shamri, M. Y. H. (2020). GUI design issues and challenges. Journal of Software Engineering and Applications, 13(04), 84-93.
- 30.Niemann, A. (2015). User Interface Design for Sentiment Analysis Tools. Proceedings of the Human-Computer Interaction International Conference.
- 31.Korfhage, R. R. (1997). Information Retrieval in Practice: Applications of Hypertext, Object Orientation, and Other Paradigms. John Wiley & Sons.
- 32.Sarawagi, S. (2017). Information Extraction: Techniques and Challenges. Now Publishers Inc.

					БР.ІІІ – 40.00.00.000 ІІЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		76

33. Cambria, E., Schuller, B., Xia, Y., & Liu, B. (2017). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 32(6), 15-27.
34. Resch, M., & Schreiber, S. (2016). Comparison of different machine learning algorithms for sentiment analysis. *Proceedings of the 2016 IEEE International Conference on Big Data (Big Data)* (pp. 3173-3180).
35. Dave, K., Vaithyanathan, S., & Ghoting, A. (2007). Mining social media for consumer insights: a case study of online product reviews. *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 237-246).
36. Schouten, K., & Frasincar, F. (2016). Survey on sentiment analysis of textual Internet content. *IEEE Transactions on Cybernetics*, 46(5), 1116-1131.
37. Giachanelli, F., & Rosso, P. (2019). Sentiment analysis on social media: A systematic review. *Expert Systems with Applications*, 125, 36-54.
38. Smith, A. (2017). *Understanding Customer Experience: A Practical Guide to CX* (2nd ed.). Kogan Page.
39. Van den Bosch, A. (2013). Text analytics, natural language processing and text mining. *Data Science Journal*, 12, 1-22.

## **ДОДАТКИ**

## Додаток А

### Клас Login

```
public class Login extends JFrame {
    CustomPanel p1; // Панель для інтерфейсу
    JLabel l1, l2; // Мітки для "Ім'я користувача" та "Пароль"
    JTextField tf1, tf2; // Текстові поля для введення імені та пароля
    JButton b1, b2; // Кнопки "Увійти" та "Скинути"
    Font f1; // Шрифт для елементів інтерфейсу

    public Login() {
        super("Вікно входу");
        p1 = new CustomPanel("Екран входу");
        p1.setTitle("Екран входу");
        p1.setLayout(null);
        JPanel main = new HeadingPanel("", Theme.GLOSSY_METALIC_BLUE_THEME); // Основна
        main.setLayout(new BorderLayout());
        f1 = new Font("Microsoft Sanserif", Font.BOLD, 13);

        // Панель для імені користувача
        JPanel pan1 = new HeadingPanel("", Theme.GLOSSY_METALIC_BLUE_THEME);
        l1 = new JLabel("Ім'я користувача");
        l1.setForeground(Color.white);
        l1.setFont(f1);
        pan1.add(l1);
        tf1 = new JTextField(15);
        tf1.setFont(f1);
        pan1.add(tf1);

        // Панель для пароля
        JPanel pan2 = new HeadingPanel("", Theme.GLOSSY_METALIC_BLUE_THEME);
        l2 = new JLabel("Пароль");
        l2.setForeground(Color.white);
        l2.setFont(f1);
        pan2.add(l2);
        tf2 = new JPasswordField(15);
        tf2.setFont(f1);
        pan2.add(tf2);

        // Панель для кнопок
        JPanel pan3 = new HeadingPanel("", Theme.GLOSSY_METALIC_BLUE_THEME);
        b1 = new JButton("Увійти");
        b1.setFont(f1);
        pan3.add(b1);
        b1.addActionListener(new ActionListener() {
            public void actionPerformed(ActionEvent ae) {
                login(); // Виклик методу входу
            }
        });

        b2 = new JButton("Скинути");
        b2.setFont(f1);
        pan3.add(b2);
        b2.addActionListener(new ActionListener() {
            public void actionPerformed(ActionEvent ae) {
                tf1.setText(""); // Очищення полів
                tf2.setText("");
            }
        });

        // Додавання панелей до основної панелі
        main.add(pan1, BorderLayout.NORTH);
        main.add(pan2, BorderLayout.CENTER);
        main.add(pan3, BorderLayout.SOUTH);
        p1.add(main);
        getContentPane().add(p1, BorderLayout.CENTER);
    }
}
```

```

public void login() {
    String user = tf1.getText();
    String pass = tf2.getText();
    if (user == null || user.trim().length() <= 0) {
        JOptionPane.showMessageDialog(this, "Потрібно ввести ім'я користувача");
        tf1.requestFocus();
        return;
    }
    if (pass == null || pass.trim().length() <= 0) {
        JOptionPane.showMessageDialog(this, "Потрібно ввести пароль");
        tf2.requestFocus();
        return;
    }
    try {
        if (user.equals("xxxx") && pass.equals("****")) { // Перевірка логіну та пар
            setVisible(false);
            UploadDocument ud = new UploadDocument(this); // Відкриття вікна для ана
            ud.setVisible(true);
            ud.setExtendedState(JFrame.MAXIMIZED_BOTH);
        } else {
            JOptionPane.showMessageDialog(this, "Невірний користувач");
        }
    } catch (Exception e) {
        e.printStackTrace();
    }
}
}
}

```

## Клас UploadDocument

```

public class UploadDocument extends JFrame {
    JLabel l1;
    JPanel p1, p2, p3, p4;
    Font f1;
    JScrollPane jsp, jsp1;
    Login login;
    JButton b1, b2, b3, b4, b5;
    JFileChooser chooser;
    DefaultTableModel dtm, dtm1;
    JTable table, table1;
    POSModel model;
    boolean loaded = false;
    SentiWordNetDemoCode sentiwordnet;
    SVMAnalysis svm;
    NaiveBayes nb;
    double svm_accuracy;
    double nb_accuracy;
    Porter stemmer = new Porter();
    ArrayList<Count> svmcoun = new ArrayList<Count>();
    ArrayList<Count> nbcount = new ArrayList<Count>();

    public UploadDocument(Login log) {
        super("Аналіз думок");
        login = log;
        p1 = new HeadingPanel("Назва проєкту", Theme.GLOSSY_METALIC_BLUE_THEME);
        p1.setPreferredSize(new Dimension(600, 50));
        l1 = new JLabel("<html><body><center>Новий підхід до використання SentiWordNet д

```

```

l1.setFont(new Font("Times New Roman", Font.BOLD, 18));
l1.setForeground(Color.white);
p1.add(l1);
getContentPane().add(p1, BorderLayout.NORTH);

// Налаштування таблиці для результатів
f1 = new Font("Courier New", Font.BOLD, 14);
p2 = new JPanel();
p2.setLayout(new BorderLayout());
dtm = new DefaultTableModel() {
    public boolean isCellEditable(int r, int c) {
        return false; // Таблиця не редагується
    }
};
table = new JTable(dtm);
table.setRowHeight(30);
table.setAutoResizeMode(JTable.AUTO_RESIZE_OFF);
table.setFont(f1);
dtm.addColumn("Назва файлу");
dtm.addColumn("Відгук");
dtm.addColumn("Думка");
dtm.addColumn("Оцінка");

// Панель для кнопок
p3 = new HeadingPanel("", Theme.GLOSSY_METALIC_BLUE_THEME);
p3.setPreferredSize(new Dimension(150, 80));
chooser = new JFileChooser(new File("."));

// Кнопка "Завантажити словник"
b1 = new JButton("Завантажити словник");
b1.setFont(f1);
p3.add(b1);
b1.addActionListener(new ActionListener() {
    public void actionPerformed(ActionEvent ae) {
        try {
            if (!loaded) {
                setCursor(new Cursor(Cursor.WAIT_CURSOR));
                loadDictionary();
                svm = new SVMAnalysis();
                nb = new NaiveBayes();
                sentiwordnet = new SentiWordNetDemoCode("SentiWordNet_3.0.0_2013
                Stopwords.readWords(stemmer);
                setCursor(new Cursor(Cursor.DEFAULT_CURSOR));
                loaded = true;
                JOptionPane.showMessageDialog(UploadDocument.this, "Словник зава
            }
        } catch (Exception e) {
            e.printStackTrace();
        }
    }
});

// Кнопка "Аналіз настрою SVM"
b2 = new JButton("Аналіз настрою SVM");
b2.setFont(f1);
p3.add(b2);
b2.addActionListener(new ActionListener() {
    public void actionPerformed(ActionEvent ae) {
        clearTable();
        int option = chooser.showOpenDialog(UploadDocument.this);
        if (option == JFileChooser.APPROVE_OPTION) {
            File file = chooser.getSelectedFile();
            setCursor(new Cursor(Cursor.WAIT_CURSOR));
            svm(file);
        }
    }
});

```

```

        setCursor(new Cursor(Cursor.DEFAULT_CURSOR));
    }
}
});

// Кнопка "Точність SVM"
b4 = new JButton("Точність SVM");
b4.setFont(f1);
p3.add(b4);
b4.addActionListener(new ActionListener() {
    public void actionPerformed(ActionEvent ae) {
        double positive = 0;
        double neutral = 0;
        for (int i = 0; i < dtm.getRowCount(); i++) {
            String value = dtm.getValueAt(i, 2).toString().trim();
            if (value.equals("Positive")) {
                positive++;
            }
            if (value.equals("Neutral")) {
                neutral++;
            }
        }
        svm_accuracy = (positive + neutral) / (double) dtm.getRowCount();
        JOptionPane.showMessageDialog(UploadDocument.this, "Точність SVM: " + sv
    }
});
}

public void svm(File file) {
    try {
        svmcount.clear();
        String result = "none";
        File[] file_list = file.listFiles();
        for (int i = 0; i < file_list.length; i++) {
            File fname = file_list[i];
            String review = svm.analyze(fname.getPath(), model, sentiwordnet, stemme
            Count c = new Count(review, svm.neutral, svm.wp, svm.p, svm.sp, svm.n, s
            svmcount.add(c);
            int positive = svm.wp + svm.p + svm.sp;
            int negative = svm.n + svm.wn + svm.sn;
            int neutral = svm.neutral;
            if (positive > negative && positive > neutral)
                result = "Positive";
            else if (negative > positive && negative > neutral)
                result = "Negative";
            else if (neutral > positive && neutral > negative)
                result = "Neutral";
            Object row[] = {fname.getName(), review, result, svm.score};
            dtm.addRow(row);
        }
    } catch (Exception e) {
        e.printStackTrace();
    }
}
}
}
}

```

## БІБЛІОГРАФІЧНА ДОВІДКА

**Тема дипломної роботи:** “ Застосування методів машинного навчання для обробки даних з user feedback ”

Обсяг пояснювальної записки: 77 аркушів.

Дата закінчення роботи: 10 червня 2025 р.

Підпис студента \_\_\_\_\_