****

**A PROJECT REPORT ON**

**A WEB BASED NATURAL LANGUAGE PROCESSING SYSTEM FOR BRANDS IMPRESSION ANALYSIS ON SOCIAL MEDIA: A CASE STUDY OF REDDIT.**

**BY**

**ADEBAYO RASHEED BABATUNDE**

**IFT/17/2388**

**SUBMITTED TO THE DEPARTMENT OF INFORMATION TECHNOLOGY, SCHOOL OF COMPUTING, FEDERAL UNIVERSITY OF TECHNOLOGY, AKURE.**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF**

**BACHELOR OF TECHNOLOGY (B. TECH) DEGREE IN INFORMATION**

**TECHNOLOGY.**

**SUPERVISOR: MR. PAUL OLOTU.**

**JANUARY, 2024.**

## **CERTIFICATION**

I hereby certify that the content of this report was carried out by ADEBAYO RASHEED BABATUNDE of the department of Information Technology, School of Computing, Federal University of Technology, Akure. Nigeria.

Adebayo Rasheed Babatunde Date

IFT/17/2388

I hereby certify that the content of this report was carried out by ADEBAYO RASHEED BABATUNDE of the department of Information Technology, School of Computing, Federal University of Technology, Akure. Nigeria.

MR. PAUL OLOTU. Date

## **DEDICATION**

This project work is dedicated to God, for wisdom and the strength to start and complete this project. I also dedicate this project to my family and colleagues who have contributed to the success of this project.

## **ACKNOWLEDGEMENT**

This is to acknowledge all those who have been instrumental to the start and the completion of this project. Firstly, I want to acknowledge God Almighty who has been my strength, sustainer through the course of the research project.

I want to acknowledge my project supervisor, Mr. Paul Olotu for his invaluable guidance, unwavering support, and insightful mentorship throughout this research project.

I would like to acknowledge the Head of Department of Information Technology, Professor (Mrs.) O.K. Boyinbode, for her valuable administrative support and dedicated oversight of all academic activities within the department which greatly contributed to the academic knowledge essential for the successful completion of this project.

I want to acknowledge my parents, Mr. and Mrs. Adebayo who ensured my diligence and commitment to the completion of this project, and also for their financial and moral support for the project.

My sincere gratitude extends to my colleagues, Uthman, Timilehin and Oluwaseyi, who have played pivotal roles in this project by sharing their expertise and engaging in insightful discussions to help realize the project's goals.

## **ABSTRACT**

The surge and growth of social networking platforms has inevitably led to the increase in user-generated data which has made it extraordinarily difficult for businesses and brands to monitor and evaluate social media sentiment in order to understand user’s impression of their products. Consequently, a substantial body of research has emerged in the domain of sentiment analysis using social media data. Unlike traditional research that employs the use brands as case studies, this approach prioritizes agility and diversity. This project focuses on developing a web application using the Django framework and multiple Natural Language Processing techniques namely Textblob, VADER, BERT and DISTILBERT to gain insight on user’s impressions or sentiments towards a brand or keyword. The web application also employs responsive HTML, HTMX and CSS for visual presentation, displaying sentiment analysis results in the form of charts, graphs, and scores. In a digital landscape where online presence is paramount, this project result demonstrated the effectiveness and efficiency of the proposed solution and proves it is not only beneficial but a necessity, empowering businesses to navigate social media sentiment adeptly.

**TABLE OF CONTENTS**

[**CERTIFICATION** ii](#_Toc148806698)

[**DEDICATION** iii](#_Toc148806699)

[**ACKNOWLEDGEMENT** iv](#_Toc148806700)

[**ABSTRACT** v](#_Toc148806701)

[**TABLE OF FIGURES** 1](#_Toc148806702)

[**CHAPTER ONE** 2](#_Toc148806703)

[**INTRODUCTION** 2](#_Toc148806704)

[**1.1 BACKGROUND OF THE STUDY:** 2](#_Toc148806705)

[**1.2 RESEARCH MOTIVATION**: 3](#_Toc148806706)

[**1.3 RESEARCH OBJECTIVES**: 4](#_Toc148806707)

[**1.4 RESEARCH METHODOLOGY** 5](#_Toc148806708)

[**CHAPTER TWO** 9](#_Toc148806709)

[**LITERATURE REVIEW** 9](#_Toc148806710)

[**2.1 SOCIAL MEDIA AS A COMMUNICATION MEDIUM:** 9](#_Toc148806711)

[**2.2 SENTIMENT ANALYSIS THEORIES:** 10](#_Toc148806712)

[**2.3 SENTIMENT ANALYSIS APPLICATIONS:** 11](#_Toc148806713)

[**2.4 SENTIMENT ANALYSIS CHALLENGES AND LIMITATIONS:** 13](#_Toc148806714)

[**2.5 NATURAL LANGUAGE PROCESSING (NLP):** 15](#_Toc148806715)

[**2.6 MODELS IMPEMENTED:** 17](#_Toc148806716)

[**2.7 RELATED WORKS:** 22](#_Toc148806717)

[**CHAPTER THREE** 31](#_Toc148806718)

[**SYSTEM ANALYSIS AND DESIGN** 31](#_Toc148806719)

[**3.1 DESCRIPTION OF THE SYSTEM.** 31](#_Toc148806720)

[**3.2 ARCHITECTURE OF THE SYSTEM** 31](#_Toc148806721)

[**3.3 FLOWCHART OF THE SYSTEM** 34](#_Toc148806722)

[**3.4 SYSTEM USE CASE.** 35](#_Toc148806723)

[**3.5 SYSTEM REQUIREMENTS** 37](#_Toc148806724)

[**CHAPTER FOUR** 38](#_Toc148806725)

[**SYSTEM IMPLEMENTATION AND TESTING** 38](#_Toc148806726)

[**REFERENCES** 50](#_Toc148806727)

## **TABLE OF FIGURES**

[Figure 1.1 Methodology 5](#_Toc148864794)

[Figure 1.2 Compound Score formula for VADER 6](#_Toc148864795)

[Figure 1.3 Sentiment classification. 7](#_Toc148864796)

[Figure 2.1 Social media users. 9](#_Toc148864797)

[Figure 2.2 Applications of Sentiment Analysis. 12](#_Toc148864798)

[Figure 2.3 Challenges in Sentiment Analysis 13](#_Toc148864799)

[Figure 2.4 Applications of NLP. 16](#_Toc148864800)

[Figure 2.5 TextBlob sample. 18](#_Toc148864801)

[Figure 2.6 TextBlob Sample Output. 18](#_Toc148864802)

[Figure 2.7 VADER Implementation Sample. 19](#_Toc148864803)

[Figure 2.8 BERT Implementation Sample. 20](#_Toc148864804)

[Figure 2.9DISTILBERT Implementation Sample. 21](#_Toc148864805)

[Figure 3.1 System Architecture 32](#_Toc148864806)

[Figure 3.2 System Flowchart 34](#_Toc148864807)

[Figure 3.3 System Use Case diagram 36](#_Toc148864808)

[Figure 4.1 Sentiment Analysis Workflow. 39](#_Toc148864809)

[Figure 4.2 Authentication logic. 40](#_Toc148864810)

[Figure 4.3 URLs path definitions. 40](#_Toc148864811)

[Figure 4.4 Frontend Snippet. 41](#_Toc148864812)

[Figure 4.5 PostgreSQL Database. 41](#_Toc148864812)

[Figure 4.6 Landing Page. 43](#_Toc148864813)

[Figure 4.7 Sentiment Page. 43](#_Toc148864814)

[Figure 4.8 Sentiment bar chart. 44](file:///C:\Users\MacSauce\Desktop\PROJECT%20REPORT.docx#_Toc148864815)

[Figure 4.9 Sentiment pie chart. 44](file:///C:\Users\MacSauce\Desktop\PROJECT%20REPORT.docx#_Toc148864816)

[Figure 4.10 Sentiment score BERT 44](file:///C:\Users\MacSauce\Desktop\PROJECT%20REPORT.docx#_Toc148864817)

[Figure 4.11 Sentiment score DISTILLBERT 44](file:///C:\Users\MacSauce\Desktop\PROJECT%20REPORT.docx#_Toc148864818)

[Figure 4.12 Disclaimer Snippet. 46](#_Toc148864822)

[Figure 4.13 Result Snippet. 45](#_Toc148864820)

[Figure 4.14 Word cloud result for Messi. 45](#_Toc148864821)

[Figure 4.15 Deployed on GitHub. 46](#_Toc148864822)

[Figure 4.16 Deployed on Pythonanywhere. 46](#_Toc148864822)

# **CHAPTER ONE**

## **INTRODUCTION**

### **1.1 BACKGROUND OF THE STUDY:**

The exponential growth in users and interactions on social media platforms has created a significant challenge for businesses and brands to monitor and evaluate social media sentiment. This challenge lies in the overwhelming volume of user-generated data, making it exceptionally hard to effectively monitor and assess social media sentiment, consequently hindering their ability to grasp how users perceive their products.

Sentiment analysis, also known as impression analysis is a crucial area of natural language processing (NLP), it offers a complex methodology to draw important conclusions from the huge body of social media data in response to this changing environment. Traditional methods of gauging brand perception, such as surveys and focus groups, are not scalable or timely enough to keep up with the ever-changing landscape of social media. By harnessing the power of natural language processing, businesses can gain a deeper understanding of their customers and the market.

Sentiment analysis aids in reputation management. In today's interconnected world, a single negative comment or unfavorable opinion can quickly escalate into a significant public relations crisis. By gauging public sentiment, organizations can take proactive measures to address issues swiftly and effectively, safeguarding their reputation and brand image. In the realm of governance and public service, sentiment analysis can measure public sentiment towards policies and government actions. This is valuable for policymakers to gauge the impact of their decisions and tailor their approach to better serve their constituents.

For the scope of this project, I envision a sentiment analysis web application to evaluate brand perception. Unlike conventional methodologies that relies on case studies, my approach emphasizes flexibility and practicality. Through a user-friendly web interface, built using Python (Django and Scikit-Learn), this project provides accessibility, with an internet connection and a device, users can gain insights into public sentiment from anywhere, the use of NLP techniques

(Textblob, VADER, BERT and DISTILBERT) facilitates accurate and detailed analysis, scalability to accommodate growing data, and the ability to employ frontend technologies to visualize sentiment trends through charts and graphs. making it a convenient tool for businesses, policymakers, and the general public.

I've chosen Reddit, a platform with a vast user base of approximately 1.5 billion individuals, as the primary data source for social media data. Reddit was chosen as the platform of choice for this project over more popular alternatives due to its stable API endpoints. Unlike some other platforms that frequently alter their APIs or impose prohibitive pricing structures, Reddit's consistency and reasonable access costs make it a reliable and cost-effective source for data, ensuring the project's stability and long-term viability

Reddit's distinct strengths lie in its extended post lengths, diverse subreddits, user anonymity, and structured discussions. These features provide a comprehensive context for studying brand perception, making it an ideal platform for uncovering intricate patterns in brand impressions within the dynamic social media landscape.

### **1.2 RESEARCH MOTIVATION**:

This project motivation lies in rectifying the shortcomings inherent in existing research methodologies through the introduction of an innovative solution. Specifically, I aim to address the limitations of studies like "Twitter Text Sentiment Analysis of Amazon Unlocked Mobile Reviews Using Supervised Learning Techniques" by **Bharathi et al. (2022)** and "Sentiment Analysis On The Acceptance Of The COVID-19 Vaccine On Twitter" by **Gerardo Uriel Monroy and Eduardo Angulo (2021)**. These studies are restricted by analyzing a single static case study, a constraint shared by many more researches. Furthermore, this project is fueled by the desire to overcome the challenge of research data being confined to a single location, a limitation evident in "Web-Based Application For Sentiment Analysis Of Live Tweets" by **Nitesh Sharma et al. (2018)**. This project solution aims to transcend these limitations and provide a more comprehensive and create a broader path for conducting sentiment analysis research.

### **1.3 RESEARCH OBJECTIVES**:

The aim of this project is to get current sentiments about any valid brand or keyword (anything that could inputted into reddit’s search bar) and display it’s result on a web page.

Therefore, the specific objectives of this project were to:

1. Design an Impression analysis web application.
2. To implement the above.
3. Evaluation.

### **1.4 RESEARCH METHODOLOGY**

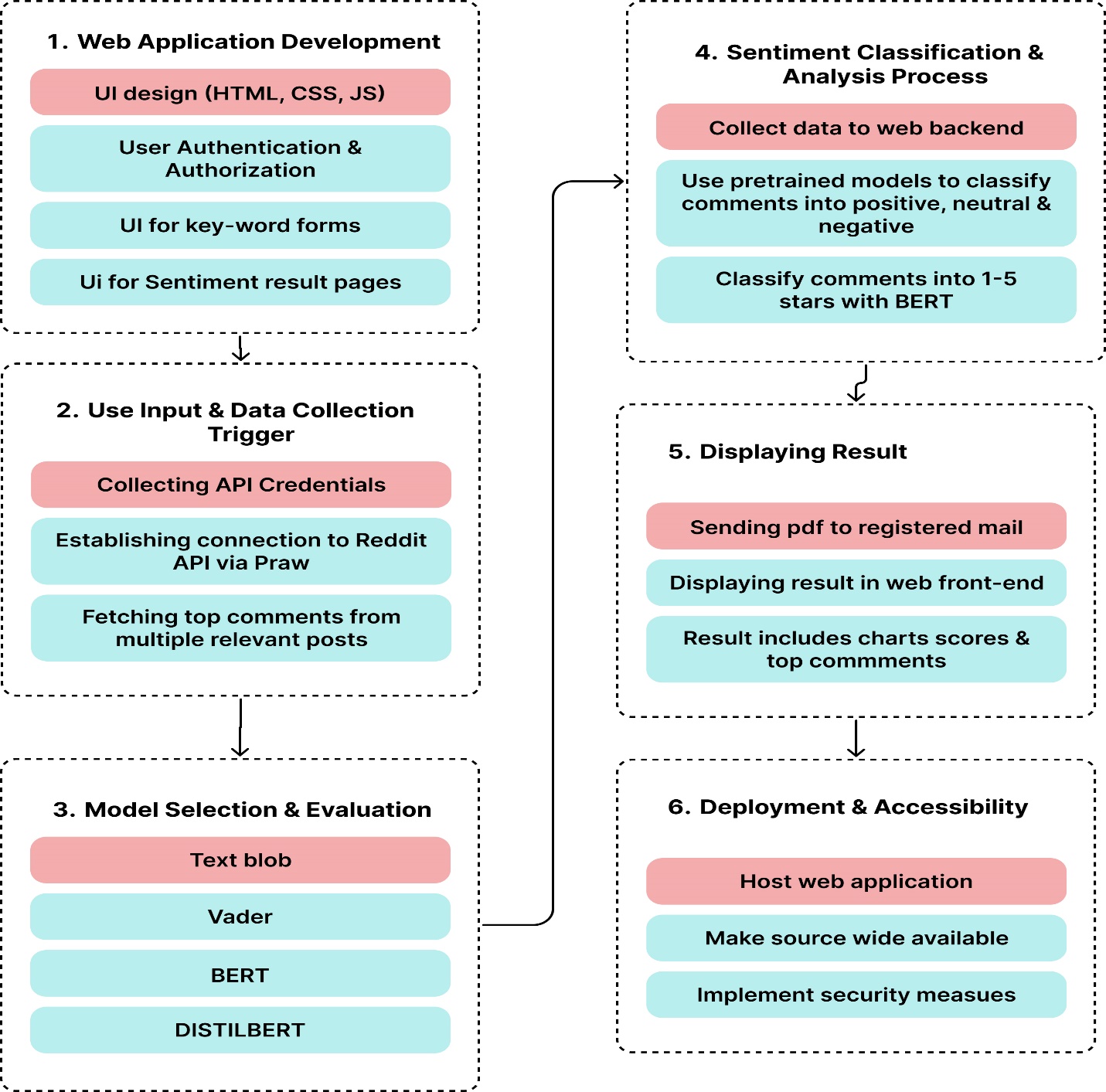


Figure 1.1 Methodology

1. **Web Application Development:** I am planning to utilize the Django framework to develop an intuitive and adaptable web application. My intention is to design a frontend interface that empowers users to effortlessly input a brand's name and trigger the commencement of the sentiment analysis procedure. This application will provide users with an accessible and user-friendly means to gauge the sentiment associated with various brands.
2. **User Input and Data Collection Trigger:** I initiate the data collection process when users interact with the web application's interface and submit the brand name they wish to analyze for sentiment. After receiving the input, I send a request to the selected API (Praw to Reddit) to fetch relevant posts related to the specified brand. This ensures that the necessary data for sentiment analysis is gathered promptly and efficiently.

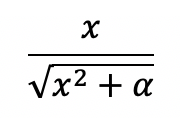


Figure 1.2 **Compound Score formula for VADER**

1. **Model Selection and Evaluation:** I carefully select the models for sentiment analysis in this phase, namely Textblob, VADER, BERT, and DISTILBERT. BERT and DISTILBERT are implemented via pre-trained pipelines for efficiency. To ensure their suitability, I thoroughly evaluate these models for their ability to accurately analyze sentiment in user-generated content. Evaluation criteria include precision, recall, and F1 score. These assessments determine which model aligns best with the project's goal of delivering insightful and reliable sentiment insights to users.
2. **Sentiment Classification and Analysis Process:** Upon receiving the user input, the web application initiates a request to the designated API (Praw with Reddit), to retrieve related posts or tweets. These fetched posts or tweets are then temporarily stored in memory for subsequent analysis. I will then utilize the pre-trained sentiment analysis model, where each individual post's sentiment is evaluated, enabling the system to ascertain whether the sentiment is positive, negative, or neutral. This process aids in categorizing the overall sentiment of the collected content, providing valuable insights into public opinions and emotions prevalent on the chosen platform. Recall that, data preprocessing, Sentiment Classification and Analysis Process will be done on the web application backend.

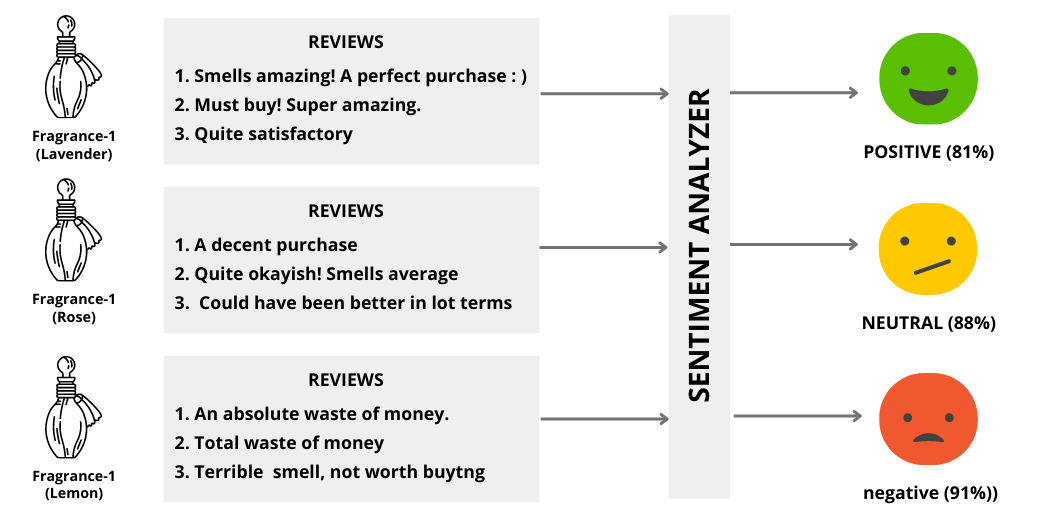


Figure 1.3 Sentiment classification.

1. **Displaying Results:** I will display the outcomes of sentiment analysis via sending a pdf to the registered mail or on the web application's interface by incorporating visual elements such as graphs and charts and performance metrics. These visual aids will effectively exhibit the distribution of various sentiment categories, offering users a comprehensive understanding of sentiment proportions. This feature will provide users with a user-friendly overview of the sentiments identified in the analyzed content.
2. **Deployment and Accessibility:** I will ensure the deployment and accessibility of the web application by hosting it on a dependable web server. I will prioritize user security by implementing robust measures like user authentication and data encryption to safeguard their information and search history.

**1.5 EXPECTED CONTRIBUTIONS TO KNOWLEDGE**

1. It aims to analyze sentiment across social media, overcoming single-case limitations.
2. The open-source availability encourages further sentiment analysis research.
3. It has the potential to advance sentiment analysis techniques and web application development.
4. The research can lead to more accessible tools for online brand sentiment analysis.

# **CHAPTER TWO**

# **LITERATURE REVIEW**

### **2.1 SOCIAL MEDIA AS A COMMUNICATION MEDIUM:**

Social media is a powerful tool that has revolutionized the way we interact and share information. It offers a range of engaging features, from creative filters to instant messaging, that enable us to express ourselves in diverse ways. However, social media also has its downsides, especially for brands. Negative comments or criticisms can quickly go viral, damaging a brand's reputation.

In short, social media is a double-edged sword. It offers many benefits, but it is important to be aware of the potential pitfalls. Balancing the benefits of connectivity and self-expression with the potential pitfalls is an ongoing challenge in the digital age.

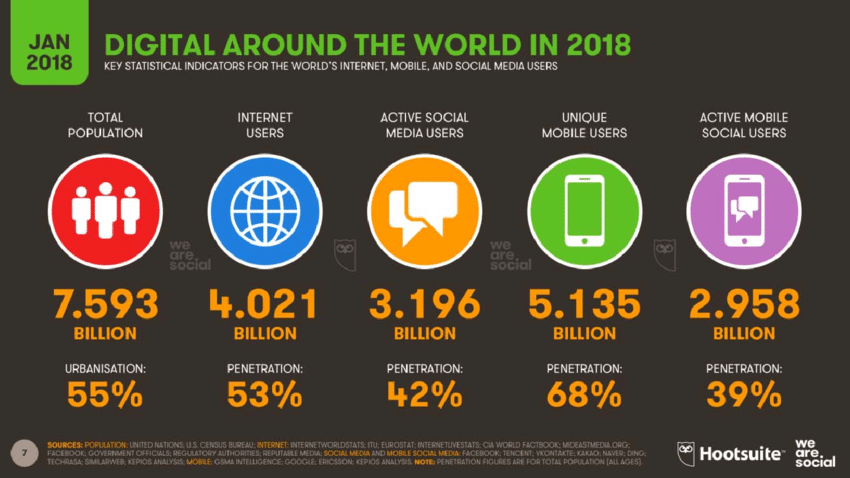


Figure 2.1 **Social media users.**

### **2.2 SENTIMENT ANALYSIS THEORIES:**

According to Wikipedia, Sentiment analysis (also known as opinion mining or emotion AI) is the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine (Sentiment Analysis, 2023).

Sentiment Analysis Theories refers to a set of theoretical frameworks and concepts that underpin the field of sentiment analysis. Sentiment analysis, also known as opinion mining, is the process of computationally determining and categorizing opinions, sentiments, or emotions expressed in text data. These theories are essential to understanding how sentiments are analyzed, categorized, and interpreted by machines. Here are some key aspects that "Sentiment Analysis Theories" entail:

**1. Sentiment Polarity**: Sentiment analysis theories include the concept of sentiment polarity, which categorizes opinions as positive, negative, neutral, or, in some cases, on a scale of sentiment intensity. Theories in this area address how to define and recognize sentiment polarity in text.

**2. Lexicon-Based Approaches:** These theories focus on the use of sentiment lexicons, which are dictionaries containing words or phrases associated with specific sentiments. Lexicon-based approaches help identify sentiment-bearing words and their strength within text data.

**3. Machine Learning Algorithms:** Sentiment analysis involves various machine learning models, including supervised learning, unsupervised learning, and deep learning. Theoretical frameworks in this area delve into how these models are trained to recognize sentiment patterns in text.

**4. Feature Extraction:** Theories explain how to extract relevant features from text data for sentiment analysis. This includes methods like bag-of-words, word embeddings, and other techniques used to represent text as numerical data.

**5. Contextual Analysis:** Understanding sentiment in context is crucial. Theories explore how sentiment can vary based on the surrounding text and how context plays a role in determining sentiment polarity.

**6. Sentiment Analysis Applications:** These theories discuss the practical applications of sentiment analysis, such as in social media, product reviews, customer feedback, and more. They help in understanding how sentiment analysis can provide valuable insights in various domains.

**7. Challenges and Limitations:** Theoretical frameworks also encompass challenges and limitations of sentiment analysis, such as sarcasm detection, polysemy, and the impact of negations on sentiment recognition. These theories address the complexities of real-world sentiment analysis.

**8. Cross-Cultural and Multilingual Analysis:** Some sentiment analysis theories consider cross-cultural and multilingual sentiment analysis, acknowledging that sentiment expression can vary across different cultures and languages.

**9. Temporal Aspects:** Sentiment analysis theories may also address temporal aspects, recognizing that sentiment can change over time and that monitoring sentiment trends is crucial in some applications.

### **2.3 SENTIMENT ANALYSIS APPLICATIONS:**

Practical scenarios where sentiment analysis can be beneficial includes:

**1. Brand Reputation Management:** Sentiment analysis can be used to monitor and manage a brand's online reputation. By analyzing sentiment in social media conversations, companies can identify negative sentiment early and respond to issues or crises promptly.

**2. Product and Service Feedback**: Sentiment analysis helps in understanding customer opinions about a brand's products or services. It provides insights into what customers like or dislike, aiding in product improvement.

**3. Market Research**: Analyzing sentiment in social media conversations allows brands to conduct real-time market research. They can gauge public opinion on various topics and track trends to make informed decisions.

**4. Competitive Analysis:** Brands can use sentiment analysis to understand how they compare to competitors. Analyzing social media conversations about both their own brand and competitors can reveal strengths and weaknesses.

**5. Targeted Marketing**: Sentiment analysis helps in creating more effective and personalized marketing campaigns. By understanding what resonates positively with the audience, brands can tailor their content accordingly.

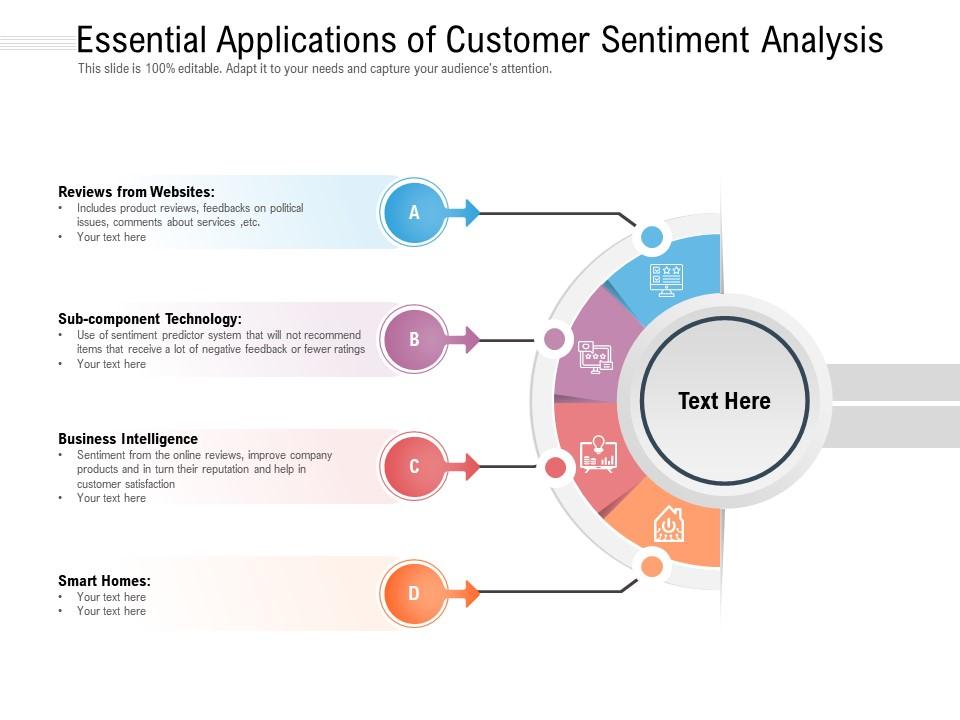


Figure 2.2 **Applications of Sentiment Analysis.**

**6. Crisis Management:** It's crucial for brands to identify and address negative sentiment promptly during a crisis. Sentiment analysis aids in monitoring and responding to potential crises.

**7. Content Strategy:** Brands can optimize their content strategies by analyzing what type of content generates positive sentiment. This helps in content creation and curation.

**8. User Experience Enhancement:** For web-based applications, sentiment analysis can be applied to analyze user reviews and feedback. This information can then be used to enhance the user experience.

**9. Predictive Analysis:** By analyzing historical sentiment data, brands can make predictions about future trends or issues they might face.

**10. Evaluating Marketing Campaigns:** After launching a marketing campaign, brands can use sentiment analysis to evaluate its effectiveness. This feedback loop is crucial for continuous improvement.

**11. Social Listening:** Sentiment analysis is a key component of social listening, where brands monitor social media conversations to gain insights into consumer sentiment and opinions.

### **2.4 SENTIMENT ANALYSIS CHALLENGES AND LIMITATIONS:**

In the realm of sentiment analysis, several challenges are encountered, particularly by companies aiming to achieve precise sentiment analysis results. Accurately deciphering sentiment and emotions in natural language poses a considerable challenge due to the need to train machines to interpret and comprehend emotions akin to the human cognitive process. Here are the main roadblocks in analyzing sentiment. [Sentiment Analysis Challenges: Everything You Need to Know (repustate.com)](https://www.repustate.com/blog/sentiment-analysis-challenges-with-solutions/)



Figure 2.3 **Challenges in Sentiment Analysis**

1. **Tone**: This can be difficult to interpret verbally, and even more difficult to figure out in the written word. Things get even more complicated when one tries to analyze a massive volume of data that can contain both subjective and objective responses. Brands can face difficulties in finding subjective sentiments and properly analyzing them for their intended tone.
2. **Polarity:** Words such as “love” and “hate” are high on positive (+1) and negative (-1) scores in polarity. These are easy to understand. But there are in-between conjugations of words such as “not so bad” that can mean “average” and hence lie in mid-polarity (-75). Sometimes phrases like these get left out, which dilutes the sentiment score.
3. **Sarcasm:** People use irony and sarcasm in casual conversations and memes on social media. The act of expressing negative sentiment using backhanded compliments can make it difficult for sentiment analysis tools to detect the true context of what the response is actually implying. This can often result in a higher volume of “positive” feedback that is actually negative.
4. **Emojis:** The problem with social media content that is text-based, like Twitter, is that they are inundated with emojis. NLP tasks are trained to be language specific. While they can extract text from even images, emojis are a language in itself. Most emotion analysis solutions treat emojis like special characters that are removed from the data during the process of sentiment mining. But doing so means that companies will not receive holistic insights from the data.
5. **Idioms:** Machine learning programs don’t necessarily understand a figure of speech. For example, an idiom like “not my cup of tea” will boggle the algorithm because it understands things in the literal sense. Hence, when an idiom is used in a comment or a review, the sentence can be misconstrued by the algorithm or even ignored. To overcome this problem a sentiment analysis platform needs to be trained in understanding idioms. When it comes to multiple languages, this problem becomes manifold.
6. **Negations:** Negations, given by words such as not, never, cannot, were not, etc. can confuse the ML model. For example, a machine algorithm needs to understand that a phrase that says, “I can’t not go to my class reunion”, means that the person intends to go to the class reunion.
7. **Multilingual sentiment analysis:** This constitutes all the problems listed above get compounded when a mix of languages are thrown in. Each language needs a unique part-of-speech tagger, lemmatizer, and grammatical constructs to understand negations. Because each language is unique, it cannot be translated into a base language like say, English, to extract insights. A simple example being, if an idiom “like a fish takes to water” is translated into say, German, the idiom would have lost its meaning.

### **2.5 NATURAL LANGUAGE PROCESSING (NLP):**

NLP for sentiment analysis is a computational approach that involves using algorithms and machine learning techniques to understand and analyze human language, specifically to determine the sentiment or emotional tone expressed in a piece of text. Here's what it entails:

**1. Text Preprocessing:** The process begins with text preprocessing, which involves cleaning the text data. This step includes removing any unnecessary characters, punctuation, and special symbols. It also involves converting text to lowercase and handling issues like misspellings.

**2. Tokenization:** Tokenization is the process of breaking down a text into individual words or tokens. This step is crucial for analyzing the sentiment of the text on a word-by-word basis.

**3. Stop Word Removal:** Stop words are common words like "and," "the," "in," which don't contribute much to the sentiment analysis. These words are typically removed to focus on more meaningful words.

**4. Feature Extraction:** In sentiment analysis, words or tokens are used as features. These features are transformed into numerical values, which can be understood by machine learning models. Common methods for feature extraction include TF-IDF (Term Frequency-Inverse Document Frequency) and word embeddings like Word2Vec or GloVe.

**5. Sentiment Lexicons:** Sentiment lexicons are dictionaries or word lists that assign sentiment scores to words. For example, "happy" might have a positive score, while "sad" could have a negative score. These scores are used to assess the overall sentiment of the text based on the sentiment of individual words.

**6. Machine Learning Models**: NLP-based sentiment analysis often employs machine learning models. These models are trained on labeled datasets, where each piece of text is associated with a sentiment label (e.g., positive, negative, neutral). Common algorithms include Naive Bayes, Support Vector Machines (SVM), and deep learning techniques like Recurrent Neural Networks (RNNs) and Transformers.

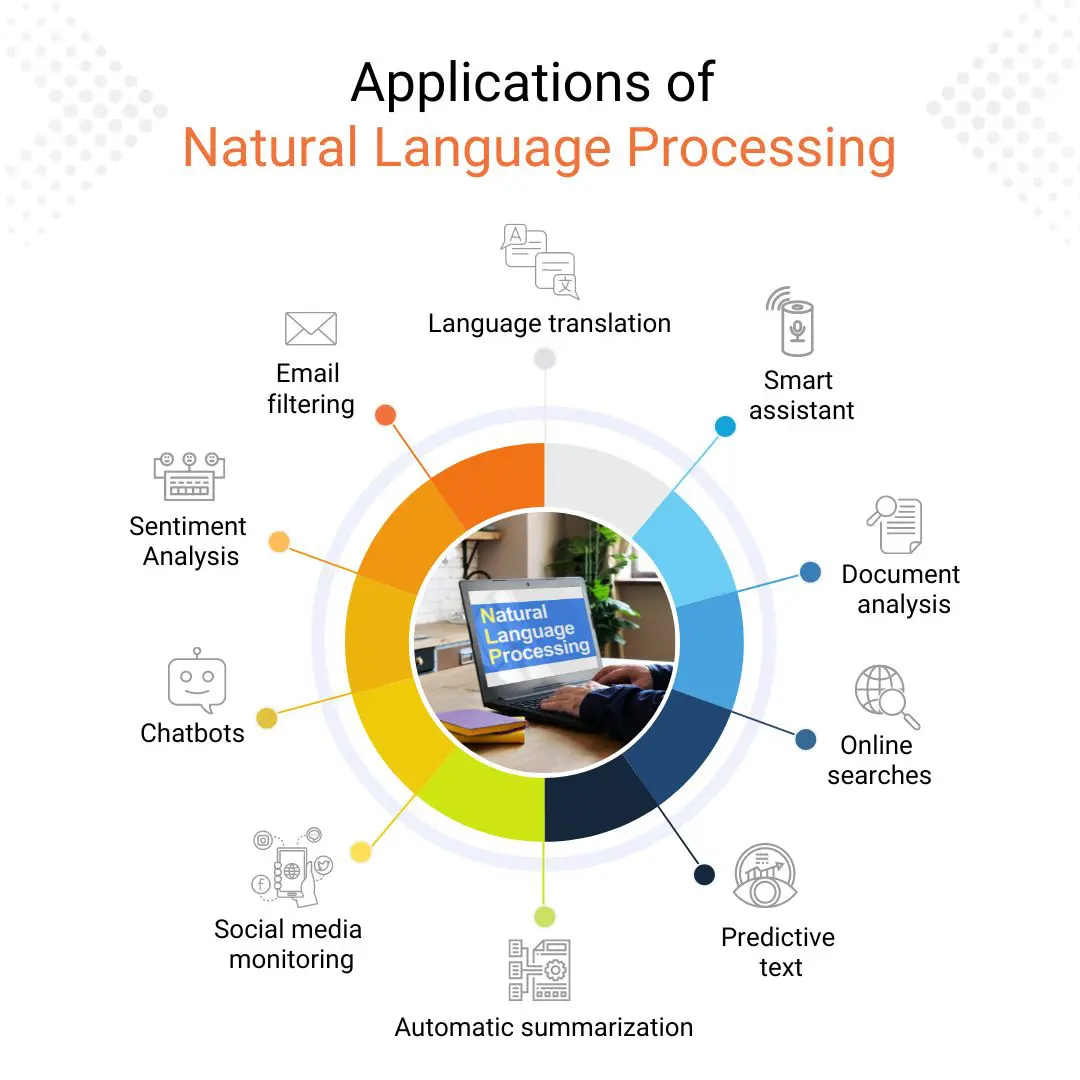


Figure 2.4 **Applications of NLP.**

**7. Sentiment Classification:** The trained models classify text into different sentiment categories, such as positive, negative, neutral, or on a finer-grained scale, such as very positive, slightly positive, etc.

**8. Evaluation:** Sentiment analysis models are evaluated based on their accuracy, precision, recall, and F1-score, among other metrics. This step helps ensure the model's reliability and effectiveness.

**9. Context and Negation Handling:** Understanding the context of a sentence is crucial, as the same word can have different meanings based on the context. Sentiment analysis models need to consider negation, idioms, and sarcasm to provide more accurate results.

**10. Application:** The results of sentiment analysis can be applied in various domains, such as social media monitoring, product reviews, customer feedback analysis, and market research. Organizations use sentiment analysis to understand how their brand or product is perceived and make data-driven decisions based on the sentiment of their audience.

### **2.6 NLP MODELS IMPEMENTED IN THE PROJECT:**

**TextBlob:** TextBlob is a Python library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more (Loria, 2018).

One of the most popular features of TextBlob is its sentiment analysis capability. TextBlob uses a lexicon-based approach to sentiment analysis, which means that it relies on a dictionary of words and phrases that have been assigned sentiment scores. The sentiment score of a piece of text is calculated by averaging the sentiment scores of the individual words and phrases in the text.

TextBlob's sentiment analysis capability can be used to analyze the sentiment of any type of text, such as tweets, product reviews, customer feedback, and social media posts. It can also be used to analyze the sentiment of different parts of a document, such as the headline, body text, and comments.

To use TextBlob's sentiment analysis capability, you can simply create a TextBlob object from the text that you want to analyze and then call the sentiment method. The sentiment method will return a tuple containing two values: the polarity and the subjectivity of the text.

The polarity score is a measure of the overall sentiment of the text, with a score of 1 indicating positive sentiment, a score of -1 indicating negative sentiment, and a score of 0 indicating neutral sentiment. The subjectivity score is a measure of how subjective the text is, with a score of 1 indicating highly subjective text and a score of 0 indicating highly objective text.

Here is an example of how to use TextBlob to analyze the sentiment of a tweet:

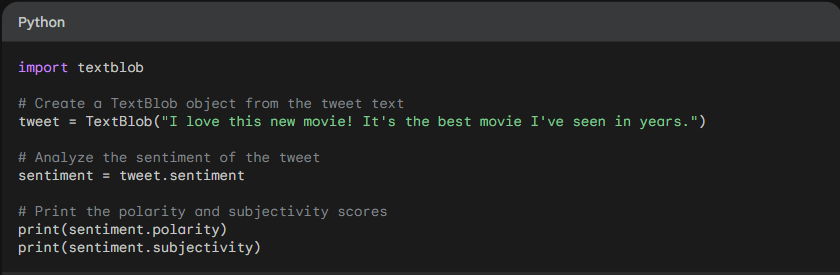


Figure 2.5 **TextBlob sample.**

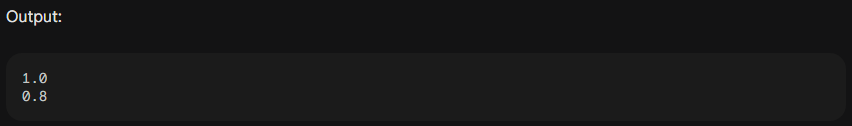


Figure 2.6 **TextBlob Sample Output.**

**VADER:** stands for Valence Aware Dictionary and Sentiment Reasoner. It is a lexicon-based sentiment analysis tool that is specifically designed to analyze the sentiment of social media text. VADER was developed by a team of researchers at IBM and the University of Massachusetts Amherst (Hutto & Gilbert, 2014).

VADER uses a variety of features to analyze the sentiment of text, including:

1. The sentiment of individual words and phrases
2. The presence of punctuation marks, such as exclamation points and question marks
3. The use of capitalization and emojis
4. The structure of the text, such as the presence of hashtags and mentions

VADER is able to identify and differentiate between different types of sentiment, such as positive sentiment, negative sentiment, and neutral sentiment. It can also identify the intensity of sentiment, with a score of 1 indicating very strong sentiment and a score of 0 indicating neutral sentiment. To use VADER, you can simply pass the text that you want to analyze to the SentimentIntensityAnalyzer() function. The SentimentIntensityAnalyzer() function will return a dictionary containing the following scores:

1. pos: The polarity score, which is a measure of the overall sentiment of the text, with a score of 1 indicating positive sentiment, a score of -1 indicating negative sentiment, and a score of 0 indicating neutral sentiment.
2. neg: The negativity score, which is a measure of the negative sentiment in the text.
3. neu: The neutrality score, which is a measure of the neutral sentiment in the text.
4. compound: A compound score, which is a weighted combination of the polarity and negativity scores.

Here is an example of how to use VADER to analyze the sentiment of a tweet:

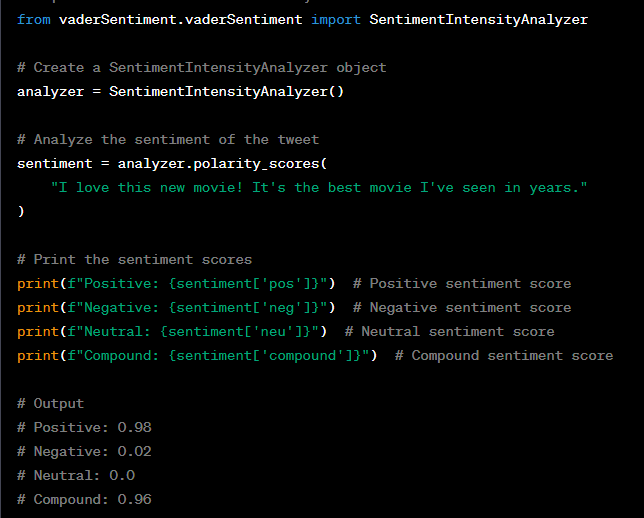


Figure 2.7 **VADER Implementation Sample.**

**BERT:** Bidirectional Encoder Representations from Transformers (BERT) is a natural language processing (NLP) model developed by Google AI. BERT was first introduced in 2018 and has quickly become one of the most popular and widely used NLP models.

BERT is a pre-trained language model, which means that it has been trained on a massive dataset of text and code. This allows BERT to learn the relationships between words and phrases, and to understand the context of sentences. BERT can be used for a variety of NLP tasks, including:

1. Text classification: BERT can be used to classify text into different categories, such as spam or not spam, or positive or negative sentiment.
2. Question answering: BERT can be used to answer questions about a given text passage.
3. Natural language inference: BERT can be used to determine whether a given hypothesis is true or false based on a given text passage.
4. Named entity recognition: BERT can be used to identify named entities in a text passage, such as people, places, and organizations.
5. Machine translation: BERT can be used to translate text from one language to another.

BERT is also relatively easy to use, and there are a number of pre-trained BERT models available that can be used for different NLP tasks.

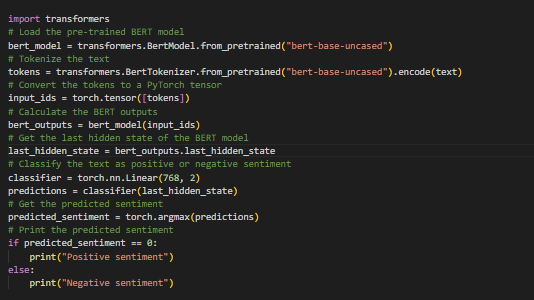


Figure 2.8 **BERT Implementation Sample.**

**DISTILBERT:** This is a distilled version of BERT, a natural language processing (NLP) model developed by Google AI. DistilBERT was first introduced in 2019 and is designed to be smaller and faster than BERT, while still maintaining most of its performance.DistilBERT was trained using a knowledge distillation technique, where a smaller model is trained to mimic the outputs of a larger model. This allows DistilBERT to learn the same knowledge as BERT, but in a more efficient way. DistilBERT can be used for a variety of NLP tasks, including:

1. Text classification: DistilBERT can be used to classify text into different categories, such as spam or not spam, or positive or negative sentiment.
2. Question answering: DistilBERT can be used to answer questions about a given text passage.
3. Natural language inference: DistilBERT can be used to determine whether a given hypothesis is true or false based on a given text passage.
4. Named entity recognition: DistilBERT can be used to identify named entities in a text passage, such as people, places, and organizations.
5. Machine translation: DistilBERT can be used to translate text from one language to another.

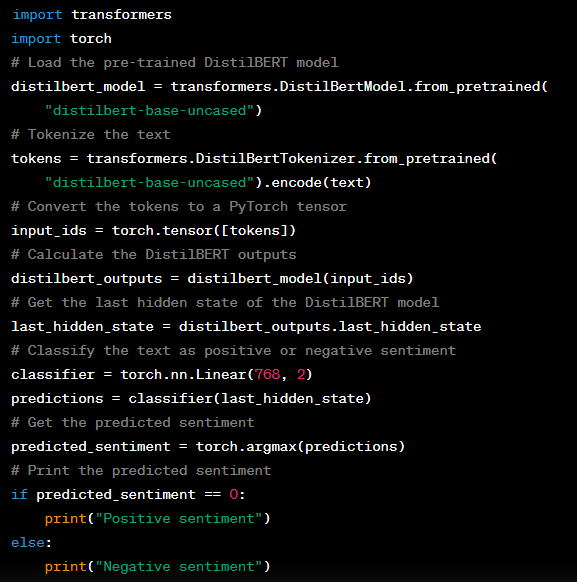
****

Figure 2.9DISTIL**BERT Implementation Sample.**

### **2.7 RELATED WORKS:**

In Bhagyashri Wagh et al.'s research from 2017, titled "A Twitter Sentiment Analysis Using NLTK and Machine Learning Techniques," the primary objective is to perform an experimental procedure using sentiment analysis on Twitter data and subsequently review the performance of the implemented algorithms. The methodology involves several key steps. Initially, the researchers utilize the Twitter API for data collection, which is a fundamental aspect of Twitter sentiment analysis. This process involves creating a Twitter account, obtaining the necessary credentials, and using the API to collect tweets. Once the data is collected, the research focuses on preprocessing it to make it suitable for feature extraction. Twitter data can be quite messy, containing elements like usernames, special characters, emoticons, abbreviations, and more. The preprocessing stage aims to clean and refine the data, including extracting the main message from tweets and removing unwanted elements like stop words, hash tags, urls, and so on. For the actual sentiment analysis, the author employs classification algorithms from scikit-learn, including Naive-Bayes classifiers, Multinomial NB classifiers, and Bernoulli NB classifiers. These algorithms are crucial for determining sentiment from the preprocessed Twitter data.

However, the research does acknowledge limitations. One limitation is the use of random data, where no specific brand or topic is targeted for sentiment analysis. Additionally, the computation power required for this kind of analysis can be substantial. Lastly, the research discusses the potential implementation of a web-based system and expresses a need for support for more languages in future studies.

In the study conducted by Zeenia and colleagues in 2017, titled "Sentiment Analysis of Customer Product Reviews Using Machine Learning," the primary objective is to develop models utilizing Naïve Bayes, Support Vector Machine (SVM), and Decision Tree techniques. The research further involves model evaluation through cross-validation and an examination of the performance accuracy across these three models. The chosen methodology for this research is a supervised learning model.

However, it's important to recognize a limitation highlighted by the study. The research emphasizes the need for reliable ratings, as it's observed that the relationship between ratings and the sentiment of reviews might not always be straightforward or equitable. This acknowledgment highlights the complexity of sentiment analysis in the context of customer product reviews.

In their 2018 research titled "Sentiment Analysis of Product-Based Reviews Using Machine Learning Approaches," Anusuya and her team pursued several objectives. Firstly, they aimed to collect product reviews from various online platforms, with a primary focus on Amazon.com. The objective was to amass a comprehensive dataset encompassing diverse product types and consumer sentiments. The research also prioritized a thorough analysis of the review data. This included meticulous data cleaning, noise reduction, and the extraction of relevant textual information. These steps were crucial for preparing the data for subsequent sentiment analysis.

A significant focal point was performing sentiment analysis at the document level, specifically within the context of product reviews. The authors sought to evaluate the overall sentiment conveyed within each review, discerning whether it held a positive or negative tone.

Additionally, the research strived to categorize the opinion sentiments expressed in the reviews. The goal was to design an effective classification system capable of distinguishing between positive and negative sentiments. This categorization process was instrumental in understanding consumer perceptions and brand impressions.

For methodology, the research adopted a Supervised Learning Model. This approach involved training a machine learning algorithm on labeled data to predict sentiment categories for new, unseen data. By learning patterns and relationships between words and sentiments from labeled examples, the model effectively categorized reviews into predefined sentiment classes.

However, it's essential to recognize a limitation of this research. The study relied on datasets that might not be recent, which could potentially affect the relevance and accuracy of sentiment analysis results. This limitation is especially noteworthy given the constantly evolving nature of online reviews and user sentiments.

Nitesh Sharma and his team conducted a research project in 2018, titled "Web-Based Application For Sentiment Analysis Of Live Tweets," with the primary objective of developing a web-based application for conducting sentiment analysis using social media data. To achieve this, they utilized Flask, a Python micro framework, which included various components like a views module for rendering web pages, connecting to APIs, data processing, and sentiment calculation. The methodology involved gathering live tweets through the Twitter search API. The text from these tweets was cleaned by removing URLs and special characters. Metadata from the tweets, including user locations, was parsed and used for further analysis. Sentiment analysis was performed using the Python text-blob library, which assigns sentiment scores to words in the text, reflecting negative and positive polarity. The research involved calculating mean and weighted polarity values, tweet counts per country and U.S. state, with this data stored in a .csv file. Visualization of the data was achieved using Plotly and Tableau. A notable feature of this application is its real-time and database-based sentiment analysis, with a proprietary data dictionary employed to identify user locations, even when they are expressed in various formats.

However, the research project had its limitations. The web application developed was limited to conducting sentiment analysis within a single country, the USA.

Sobia et al. (2021) conducted a research study titled "Amazon Product Sentiment Analysis using Machine Learning Techniques." The core objective of their research was to perform sentiment analysis on Amazon product reviews and classify them using a range of machine learning models. Their specific goals included not only analyzing customer review data to determine sentiment polarity but also rigorously evaluating these models using cross-validation techniques to ensure the reliability of their predictions.

The methodology adopted for this study centered around a supervised learning approach. The researchers employed various machine learning models to be trained on the review data and predict sentiment labels for each review. This encompassed a sentiment analysis task that involved text data preprocessing, feature extraction, and model training. To gauge the models' performance and their ability to be applied to broader contexts, cross-validation techniques were used.

One notable limitation of this research was the relatively modest size of the datasets, comprising around 28,000 reviews. While the study effectively demonstrated the feasibility of sentiment analysis through machine learning techniques, the dataset's limited size might constrain the model's capability to comprehensively capture the entire spectrum of customer sentiments and preferences embedded within Amazon product reviews. Expanding the dataset could potentially lead to more nuanced and accurate sentiment predictions.

In their 2021 research, Gerardo Uriel Monroy and Eduardo Angulo focus on conducting sentiment analysis related to public acceptance of the COVID-19 vaccine on Twitter. Their primary objective is to investigate the sentiment expressed by the public towards the vaccine and to identify the factors that influence this sentiment. To accomplish this, the authors utilized the Twitter API, which was instrumental during the COVID-19 pandemic, to collect tweet data related to different vaccines, such as Astrazeneca, J&J, Moderna, and Pfizer in the USA. They gathered approximately 30,000 relevant English tweets over a month to form the project dataset. The methodology consisted of data preprocessing and cleaning to ensure precise sentiment analysis. This process involved the removal of unnecessary fields from the tweet data, such as "tweet id," "tweet text," and "location." Text cleaning was performed by eliminating special characters, links, and emojis, and stop words were removed to enhance the accuracy of the sentiment analysis. Sentiment analysis was conducted using two libraries, VADER and Textblob. Each library categorized text as positive, negative, or neutral, and tokenization separated words. The sentiments for each tweet were determined and stored in separate columns as "sentiment Vader" and "sentiment textblob." The results were represented visually through various means, including word clouds to represent sentiments related to specific vaccines and overall sentiment. Additionally, bigram and trigram graphs were used to identify prevalent word patterns in the collected tweets.

However, the research acknowledges certain limitations, such as being limited to a case study and focusing on sentiment within the USA.

Jemimah's research in 2021, titled "Sentiment Analysis of Amazon Electronic Product Reviews using Deep Learning," is focused on exploring the history, evolution, and techniques of sentiment analysis. The primary objectives encompass the development of various Deep Learning models for customer sentiment analysis, the assessment of these models' accuracy, and their practical use in business development. Furthermore, the study aims to highlight the challenges that surface when employing Deep Learning techniques for sentiment analysis and strategies to address them. The research employs Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) models as the chosen methodology to achieve these objectives. However, it's important to note that the research acknowledges limitations. One key limitation is the use of non-recent datasets. Additionally, implementing the project requires a high-end computing system, which may pose practical challenges for some users.

In terms of methodology, they implemented a range of sentiment analysis techniques, including two hybrid Deep Learning sentiment classifiers (CNN-LSTM and LSTM-CNN models), LSTM, and lexicon-based sentiment analysis.

However, the research work does acknowledge some limitations. One limitation to consider is that the review solely considers papers published in the English language. Furthermore, the focus of the review is directed specifically toward sentiment analysis of English-language social media data, potentially leaving gaps in the broader context of social media sentiment analysis conducted in other languages or within diverse linguistic and cultural contexts.

In their research, Qianwen Ariel Xu, Victor Chang, and Christina Jayne (2022) undertook a systematic literature review titled "A Systematic Review of Social Media-Based Sentiment Analysis: Emerging Trends and Challenges." The central aim of their work was to provide a comprehensive overview of the contemporary landscape of sentiment analysis conducted on social media, paying particular attention to emerging trends and challenges in this field.

The authors employed a systematic approach to this review, encompassing a selection of 150 papers published between 2010 and 2020. To identify these papers, they created a specific search string and utilized it to scour various reputable databases, including the ACM Digital Library, IEEE Xplore, ScienceDirect, Scopus, and SpringerLink.

The research conducted by Akanksha and colleagues in 2022, titled "Sentiment Analysis on Amazon Product Reviews," was driven by the following objectives: firstly, to analyze sentiment using Random Forest, Naive Bayes, Support Vector Machines, Logistic Regression, and BERT, and secondly, to assess the performance accuracy of these five models. To fulfill these objectives, the authors utilized a combination of supervised learning models and a transformer-based model. Specifically, Random Forest, Naive Bayes, Support Vector Machines, and Logistic Regression, all supervised learning models, were employed to classify the sentiment of product reviews. Additionally, the authors incorporated BERT, a robust transformer-based model, to extract contextual information and semantic meaning from the textual data, thereby enhancing the sentiment analysis process.

However, it's worth noting that the research has acknowledged a significant limitation. The computational resource requirement was notably high due to the substantial size of the datasets typically found in product reviews. This high resource demand for implementing and training the models might limit the accessibility and reproducibility of the research for researchers with limited computational resources. Despite this limitation, the study offers valuable insights into the comparative performance of different models for sentiment analysis on product reviews.

In the study conducted by Bharathi and colleagues in 2022, titled "Twitter Text Sentiment Analysis of Amazon Unlocked Mobile Reviews Using Supervised Learning Techniques," the main research objective was to develop models for assessing sentiments expressed in Amazon product reviews. The specific focus was on utilizing Gaussian Naïve Bayes, Logistic Regression, and Support Vector Machine (SVM) as machine learning models, and evaluating their performance. The research methodology employed a Supervised Learning Model approach for sentiment analysis, although specific details regarding data collection were not provided. The study primarily focused on preprocessing the data, which included handling negations, converting text to lowercase, removing punctuations, eliminating stop words, and stemming.

However, a potential limitation of this study is the utilization of relatively small datasets, comprising 1000 reviews. This dataset size may raise questions about the generalizability of the findings to a broader context.

In their 2022 research titled "Predicting Sentiment Analysis For Web Users With A Deep Learning Approach," Nawres Abdelwahed and the team present a significant objective - the development of a deep learning-based framework to predict sentiment and Mean Opinion Score (MOS) based on user comments. The methodology adopted in this research is noteworthy. It leverages bidirectional LSTM (Bi-LSTM) to effectively handle long and short contextual dependencies, addressing challenges related to word significance and co-occurrence. This model employs GloVe vectors for word representation and combines Bi-LSTM with convolution for feature extraction. The integration of global and average pooling techniques further enhances the quality of feature maps. Experiments performed on a collected dataset serve to validate the model's capabilities in predicting MOS, offering more precise results compared to existing sentiment analysis models.

It's important to acknowledge a limitation of this approach, namely the reliance on a static dataset, which can impact its real-time applicability in a web-based scenario.

Vishal A. Kharde and colleagues conducted research in 2022 titled "Sentiment Analysis of Twitter Data: A Survey of Techniques" with the objective of evaluating sentiment analysis techniques applied to Twitter data and reviewing the performance of implemented algorithms.

The methodology for this study encompasses several steps. Firstly, the datasets, including movie reviews, tweets, and spam reviews, underwent preprocessing. Feature extraction was a crucial part of the process, involving the utilization of various features such as unigrams, bigrams, n-grams, word presence, parts of speech tags, syntactic patterns, sentiment-bearing phrases, and term position, contributing to subjectivity analysis. Negation, although complex, was considered due to its significant impact on sentiment polarity by reversing opinions. Supervised learning played a vital role in training the classifier for future predictions with unknown data. Classification was performed using Naive Bayes and Support Vector Machine (SVM).

However, it's important to note that Naive Bayes and SVM, while widely used in sentiment analysis, have their limitations. Naive Bayes assumes independence between features, which may not hold in language with complex dependencies. SVM may struggle with efficiently handling large feature spaces and might not capture subtle relationships in text data. Additionally, both methods may face challenges when dealing with sarcasm, context, and nuances in sentiment compared to more advanced techniques like deep learning.

Ketan Gupt, Nasmin Jiwani, and Neda Afreen's research in 2023, titled "A Combined Approach of Sentimental Analysis Using Machine Learning Techniques," focuses on identifying the most suitable classifier for distinguishing between positive and negative sentiments in diverse types of reviews. The methodology involves utilizing datasets from Amazon product reviews, IMDB movie reviews, and Yelp reviews. The research encompasses several techniques, including preprocessing, classification, and representation. Preprocessing tasks involve data cleansing, managing numbers and punctuation, eliminating stop words, and excluding HTML/URL tags. The preprocessed text is represented using the TF-IDF model. Various classifiers like Naive Bayes (NB), Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) are employed to categorize the dataset into positive and negative classes. The research concludes by comparing the performances of these classifiers.

However, there are limitations to the study. The paper employs a relatively small dataset of tweets for training and testing the combined approach. Additionally, the evaluation of the combined approach is restricted to a binary sentiment classification task (distinguishing between positive and negative tweets). Expanding the evaluation to include other sentiment classification tasks, such as identifying the intensity of sentiment, would provide a more comprehensive understanding of the approach's performance.

In the research by Mohammad Abu Kausar and colleagues (2023), titled "Sentiment Classification based on Machine Learning Approaches in Amazon Product Reviews," the primary aim was to evaluate various machine learning classifiers' effectiveness in sentiment analysis. Their analysis covered Amazon product reviews and datasets from IMDB movie reviews and Yelp.

The methodology involved Python's programming environment, chosen for its rich library support and user-friendliness, to execute machine learning tasks. Specifically, they used Scikit-learn, a Python package, to leverage supervised machine learning algorithms like SVM and Naive Bayes for classification. The dataset incorporated Amazon product reviews, including key components like Summary, Review text, Rating, and Helpfulness. The data collection process yielded 4960 Titan Men Watches reviews. Data preprocessing included several stages, such as emoji removal, HTML tag removal, converting text to lowercase, and filtering out numbers and special characters. Feature extraction was critical to convert textual data into numerical vectors. CountVectorizer was applied in a Bag of Words approach to represent words numerically. The researchers also employed Exploratory Data Analysis (EDA) to analyze and gain insights into data patterns and structures.

However, the research acknowledges limitations. A significant limitation is the use of relatively small and static datasets, which could impact the generalizability and applicability of the findings to larger and more dynamic datasets.

Margarita Rodriguez-Ibanez and Antonio Casanez-Ventura's research in 2023, titled "A review on sentiment analysis from social media platforms," is dedicated to providing a systematic review of literature regarding sentiment analysis based on social media. The core objective of their work is to offer an in-depth analysis of the existing literature, with a specific emphasis on emerging trends and the challenges encountered in this field.

In terms of methodology, the authors pursued applied research. Their systematic literature review encompassed papers published in peer-reviewed journals between 2010 and 2020. The selection process involved a thorough keyword search to identify papers aligning with their inclusion criteria. Subsequently, these selected papers were meticulously reviewed and coded to capture details related to their research questions, methodologies, findings, and any limitations.

This approach ensures that their research is grounded in a comprehensive examination of prior studies and contributes to a more profound understanding of the field of social media-based sentiment analysis.

# **CHAPTER THREE**

## **SYSTEM ANALYSIS AND DESIGN**

### **3.1 DESCRIPTION OF THE SYSTEM.**

The sentiment analysis web application is designed offers a user-friendly and accessible platform for individuals and businesses seeking to gain valuable insights into how their brands are perceived on social media, with a particular focus on the Reddit platform. The system’s core functionality revolves around its user-friendly interface. Users can readily initiate the sentiment analysis process by entering the brand names they desire to assess for sentiment. What sets this system apart is the remarkable flexibility it offers. Users are presented with an array of machine learning models, including Textblob, VADER, BERT, and DISTILBERT, from which they can choose to conduct their analysis. This diverse set of models enables users to tailor their analysis approach, reflecting the nuanced nature of sentiment within social media.

Upon input, the system efficiently processes this data. It interacts with the Reddit API, expertly fetching relevant data associated with the specified brands. The system then proceeds to meticulously analyze these user-defined brands for sentiment, going beyond simple positive or negative labels to encapsulate the intricate spectrum of public sentiment. To enhance the user experience, the results are presented in an intuitive, visually comprehensible format. This user-friendly format not only simplifies data interpretation but also fosters a deeper understanding of how the general public perceives the brands in question. As users engage with the system, it not only empowers them with a more comprehensive view of their brand's digital presence on Reddit but also facilitates data-driven decision-making, ultimately enriching brand management strategies.

### **3.2 ARCHITECTURE OF THE SYSTEM**

The architecture of the system is designed with modularity and scalability in mind. It follows a Microservices Architecture, which allows different components of the system to operate independently and efficiently while collaborating to deliver comprehensive sentiment analysis.

**1. Microservices-Based Design**: The system is structured around the Microservices Architecture, where various machine learning models (Textblob, VADER, BERT, DISTILBERT) act as microservices. Each model serves as an independent service, specializing in sentiment analysis. This approach allows for flexibility and ease of future enhancements or model additions without affecting the entire system.

**2. Web Application Service:** The primary user interface of the system is a web application that follows the Model-View-Template (MVT) architectural pattern. This web application harnesses the high-level Python framework, Django, which is a strategic choice due to Python's extensive usage in AI and Machine Learning applications. Python's popularity in the field makes it a robust choice for this system. The web application serves as a user-friendly gateway, allowing users to interact with the system. Users can input brand names and select their preferred sentiment analysis models through this interface.

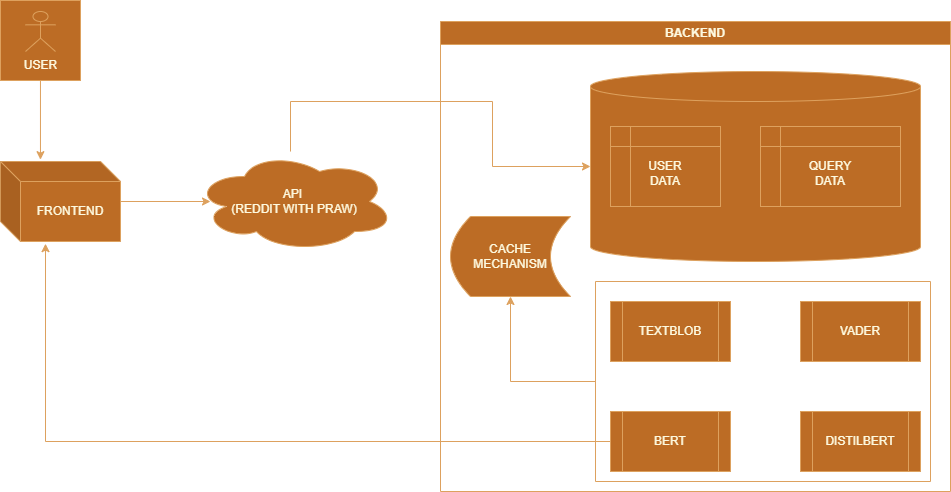


Figure 2.1 **System Architecture**

**3. Data Storage and Retrieval:** A database component handles the storage and retrieval of user inputs and analysis results. This database service is responsible for data persistence and retrieval, enabling users to access their historical analyses and enhancing data management. A cache mechanism is also implemented to process by reducing database queries.

**4. External API Integration**: The system integrates external APIs, such as the Reddit API for data collection. This integration service collects relevant data from the Reddit platform for sentiment analysis. The ability to incorporate external data sources demonstrates the system's extensibility.

**5. Responsive Frontend:** The responsive frontend, developed using HTML and CSS, ensures user-friendly and accessible interactions. It is responsible for rendering the web application on different devices and screen sizes, offering a seamless user experience

This Microservices Architecture enhances system’s modularity, performance and scalability. It allows the project to manage individual functions independently and paves the way for potential future extensions or additional microservices. The architecture is designed to be agile and adaptable, offering a solution for brand impression analysis on social media with a focus on user-friendliness and data accuracy.

### **3.3 FLOWCHART OF THE SYSTEM**

Flowchart is a diagram that represents the algorithm, flow of information in the system. They are used in analyzing, designing, and managing a program.

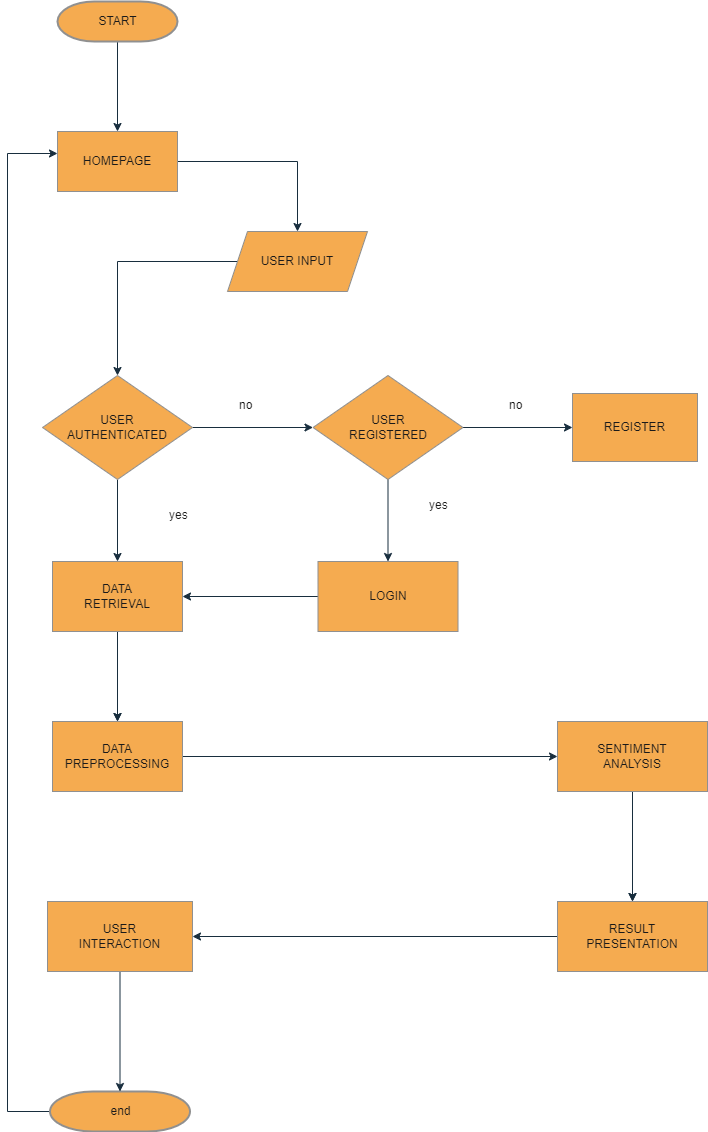


Figure 3.2 **System Flowchart**

### **3.4 SYSTEM USE CASE.**

1. Check Brand Sentiment

Description: Users can check the sentiment of specific brands.

Actors: User

Preconditions: User is logged in.

Flow of Events:

1. User inputs a brand name.

2. User selects a sentiment analysis model.

3. System retrieves and analyzes brand-related data.

4. System presents sentiment results to the user.

Postconditions: User receives brand sentiment analysis.

2. Save Analysis for Later

Description: Users can save brand sentiment analysis results as pdf for future reference.

Actors: User

- Preconditions: User is logged in and brand sentiment analysis is performed.

- Flow of Events:

1. User performs brand sentiment analysis.

2. User chooses to save the analysis.

3. System stores the analysis for the user.

- Postconditions: User can access saved sentiment analysis.

3. User Account Page

Description: Users can access their account page.

Actors: User

Preconditions: User is logged in.

Flow of Events:

1. User accesses their account page.

2. User views and manages their account details.

Postconditions: User can view and update their account information.

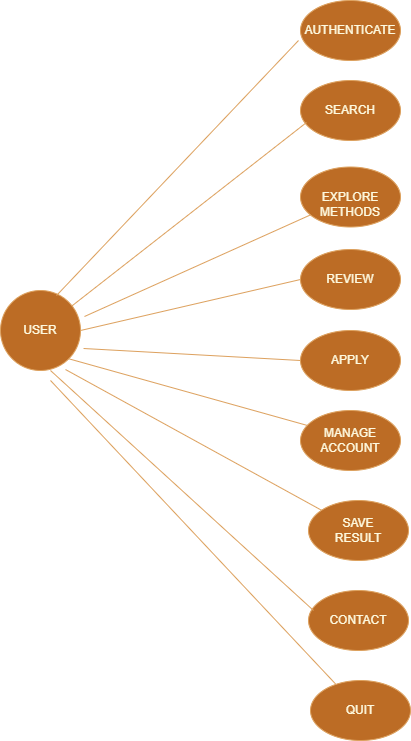


Figure 3.3 **System Use Case diagram**

4. Brand Reviews Page

Description: Users can check and provide reviews for brands.

Actors: User

Preconditions: User is logged in and brand sentiment analysis is performed.

Flow of Events:

1. User views brand sentiment analysis results.

2. User accesses the reviews page.

3. User provides reviews and ratings.

Postconditions: Brand reviews and ratings are recorded.

5. Explore Brands Page

Description: Users can explore and find available brands.

Actors: User

Preconditions: User is logged in.

Flow of Events:

1. User accesses the Explore page.

2. User browses available brand sentiment analysis options.

3. User initiates sentiment analysis for selected brands.

Postconditions: User can explore and analyze brands.

### **3.5 SYSTEM REQUIREMENTS**

**Hardware:** Memory (minimum of 512mb RAM, 2 GB ROM) enough to run a web browser.

Software.

**Operating Systems**: OS that support browsers: Windows, macOS, Linux, Android, iOS, Chrome OS, Unix, FreeBSD, Solaris, Ubuntu Touch, BlackBerry 10, WebOS, Tizen, SOS.

**Web Browser:** Safari, Opera, Firefox, Edge, Chrome, Brave, Vivaldi, Tor, UC Browser etc.

**Network Requirement**: Stable internet connectively.

# **CHAPTER FOUR**

## **SYSTEM IMPLEMENTATION AND TESTING**

The implementation phase involves a coordinated effort to bring all the components together focusing on translating the system's architectural design into functional components. The core components, encompassing sentiment analysis microservices, web application service, data storage, and external API integration, are diligently implemented. The machine learning models (Textblob, VADER, BERT, DISTILBERT) are integrated into the microservices, and APIs are established between the web application and sentiment analysis services. The user interface is designed with HTML, CSS and JavaScript to accommodate user inputs and deliver sentiment results efficiently. To ensure the system's functionality and reliability, rigorous testing is conducted. This testing includes unit testing for each sentiment analysis microservice, integration testing to verify seamless communication between system components, and user acceptance testing to evaluate the web application's responsiveness and effectiveness. A critical aspect of testing is ensuring the system's ability to provide accurate sentiment analysis results in line with user expectations.

### **4.1 IMPLEMENTATION REQUIREMENTS**

**Hardware:** The following hardware requirements are required to implement this system.

1. Server: A web server for hosting the application (runserver).
2. CPU: Adequate processing power to handle the application's computational needs.
3. RAM: Enough memory to handle the application's data processing needs.
4. Storage: Adequate storage capacity for storing application data and files.
5. Network: Reliable network connectivity for data transmission and communication.
6. containerization: Docker simplifies app deployment, ensures consistency, and enhances scalability, streamlining development and operations
7. Backup and Recovery: A system for backup and recovery of application data.

**Software:** The following software requirements are required to implement this system.

1. Server-side scripting language (Python, Django web framework).
2. Database management system (PostgreSQL).
3. Security features (e.g., authentication, authorization, encryption, permissions etc.).
4. Monitoring and logging tools.
5. Version control system (e.g., Git).
6. Collaboration and project management tools (GitHub.) Testing and debugging tools.

### **4.2 CODE SNIPPETS**

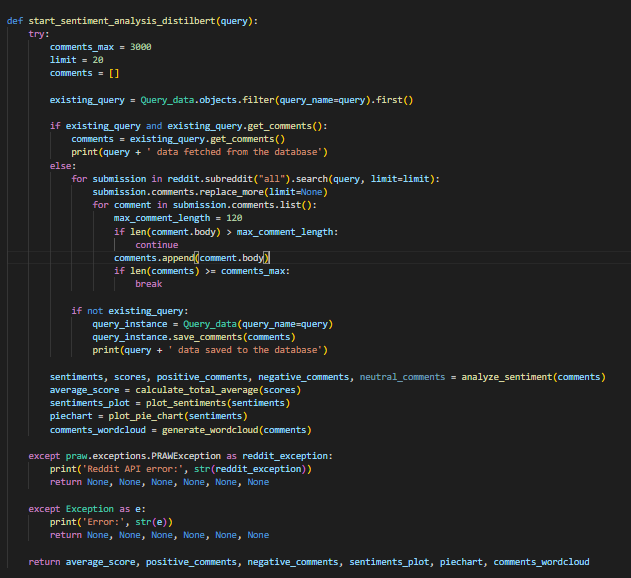


Figure 4.1 Sentiment Analysis Workflow.

The code snippet above is a Python function that performs sentiment analysis using the DistilBERT model on comments retrieved from Reddit based on a user's query. It begins by setting limits on the number of comments to fetch from Reddit (up to 3000) and the number of Reddit posts to search (up to 20). The function checks if the data for the given query exists in the database and, if so, retrieves it. If not, it searches Reddit for posts related to the query, retrieves their comments, and stores the comments in the database. It then proceeds to analyze the sentiment of these comments using the ‘analyze\_sentiment’ function. The sentiment analysis results, including the average sentiment score, the number of positive, negative, and neutral comments, and sentiment plots, are generated. This code also includes error handling for potential exceptions, such as Reddit API errors.

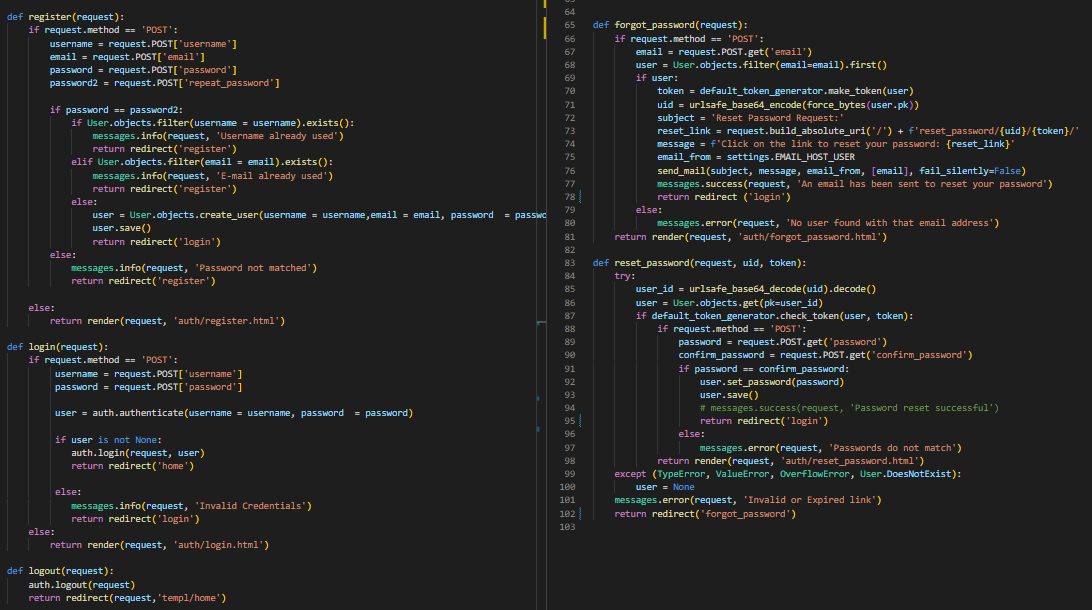


Figure 4.2 Authentication logic.

The above code snippet is a section of the web application for user authentication and password management. It includes functionality for user registration, login, logout, password reset requests, and resetting passwords via email links. It employs Django's built-in authentication features and email handling for a secure and user-friendly experience.

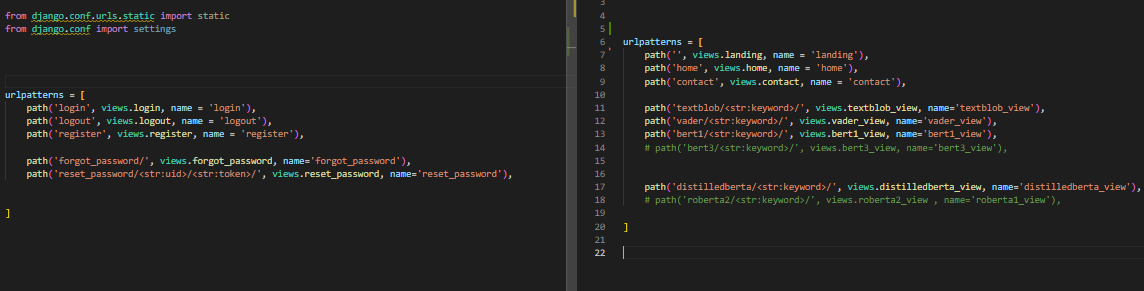


Figure 4.3 URLs path definitions.

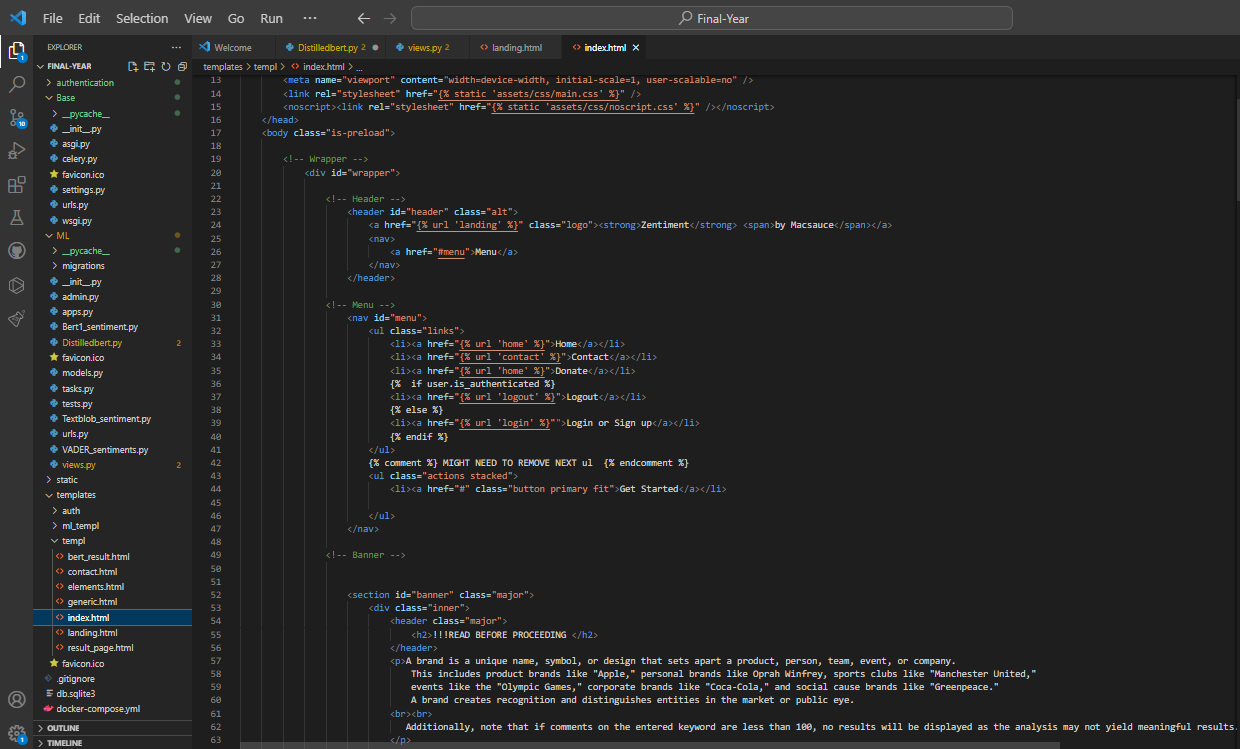


Figure 4.4 Frontend Snippet.

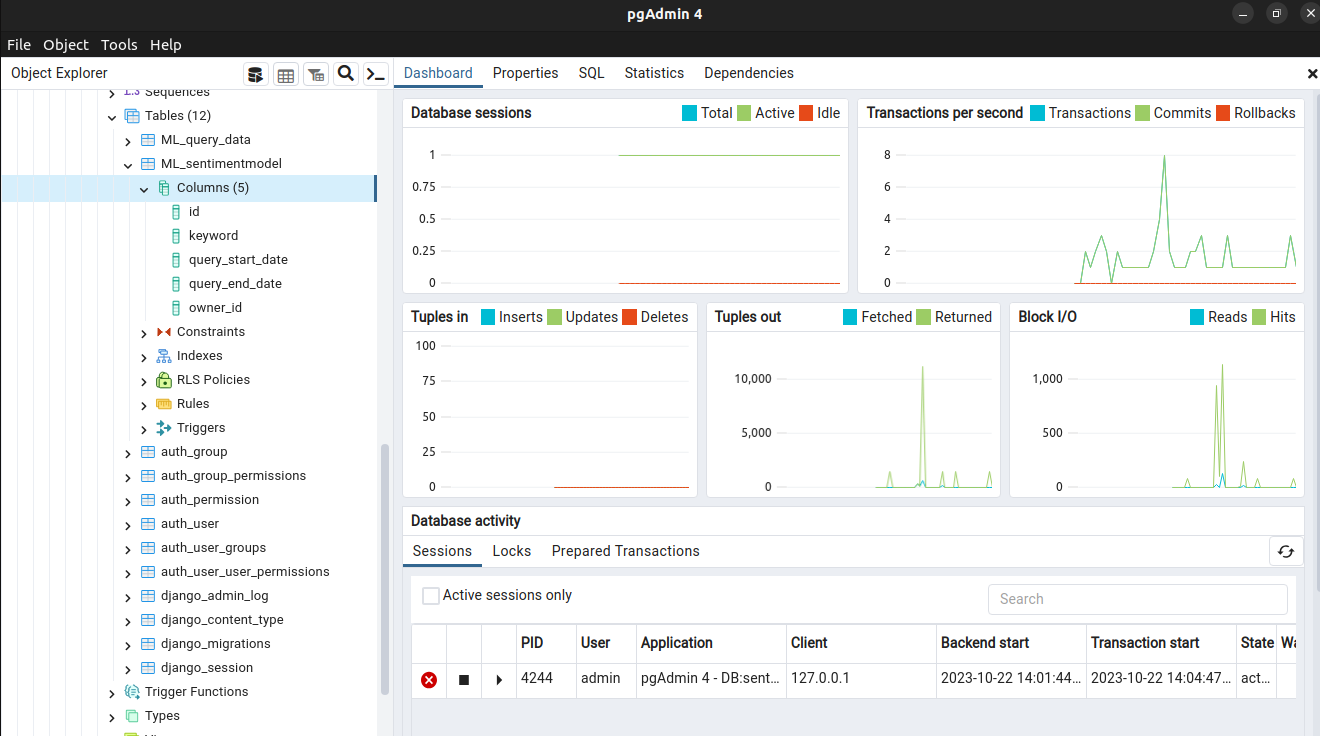


Figure 4.5 PostgreSQL Database.

### **4.3 WEB APPLICATION TESTING**

In the development process, rigorous testing is a fundamental phase to ensure the application's reliability, accuracy, and overall effectiveness. System application testing encompasses a range of evaluations that collectively guarantee that the system functions as intended, maintains responsiveness under various loads, safeguards user data, and provides a user-friendly experience.

**Functionality Testing:** This phase primarily focuses on ensuring that all the features and functions of the web application perform accurately and consistently. It involves various types of testing, including unit testing, integration testing, and system testing. During functionality testing, we validate that the sentiment analysis models (Textblob, VADER, BERT, and DISTILBERT) operate correctly and provide accurate results. Additionally, the integration between the frontend and backend components was examined to ensure seamless communication and functionality.

**Performance Testing:** Performance testing is crucial, especially when handling significant data volumes and user interactions. Load testing is conducted to determine how the system performs under different levels of usage. This involves assessing response times, resource utilization, and system stability. By simulating various scenarios, we ensure that the application remains responsive and functional even during peak loads, which is essential for maintaining a positive user experience.

**Security Testing**: Security is a paramount concern, especially when handling user data and conducting sentiment analysis on social media content. We perform security testing to identify and rectify vulnerabilities that could be exploited by malicious actors. This includes checking for potential vulnerabilities in the code, data encryption methods, and user authentication processes. Data security are core principles of our system, and testing is essential to uphold these standards.

**User Experience Testing**: The user experience (UX) is a critical aspect of our web application. We conduct usability testing to assess the overall user-friendliness of the system. This testing involves real users interacting with the application to evaluate its navigation, layout, and overall design. By collecting user feedback, we can make necessary adjustments to enhance the user experience, ensuring that individuals can easily input keywords, select sentiment analysis methods, and access the results.

### **4.4 WEB APPLICATION INTERFACE**

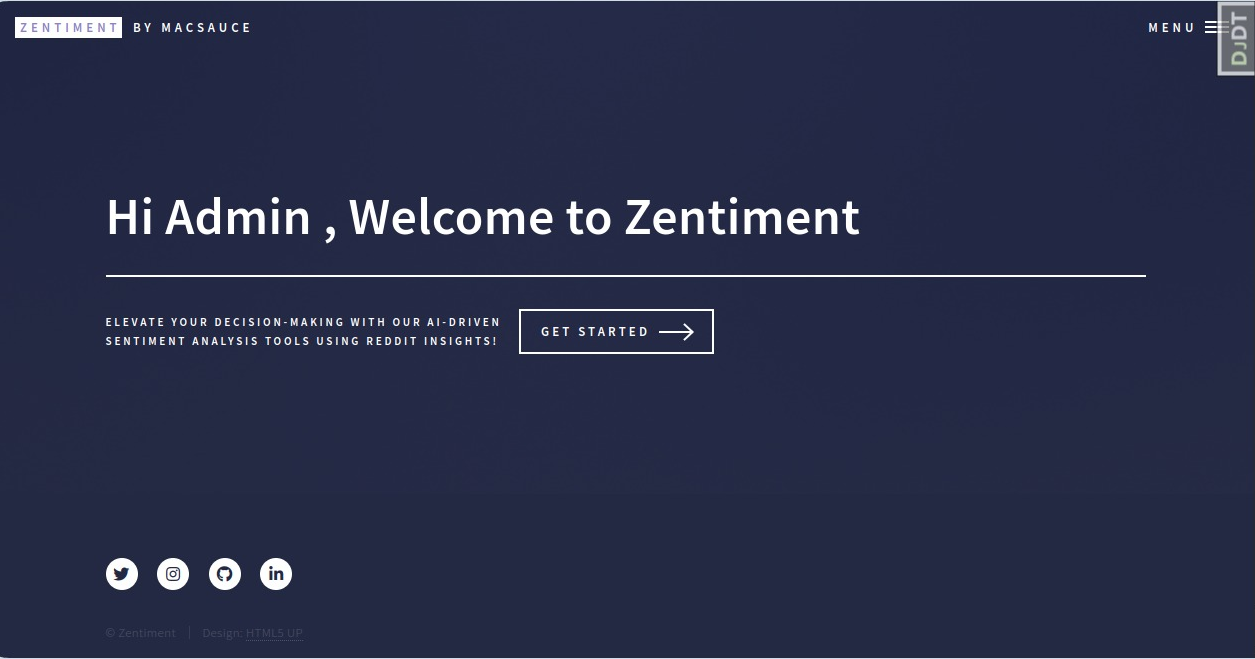


Figure 4.6 Landing Page.



Figure 4.7 Sentiment Page.

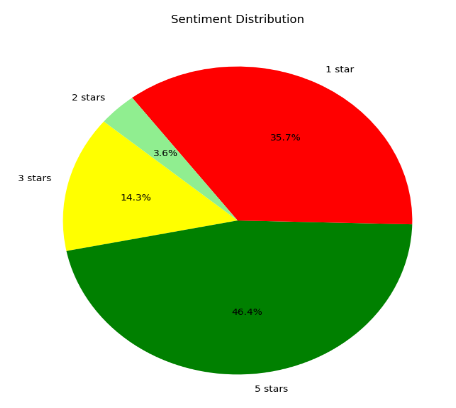
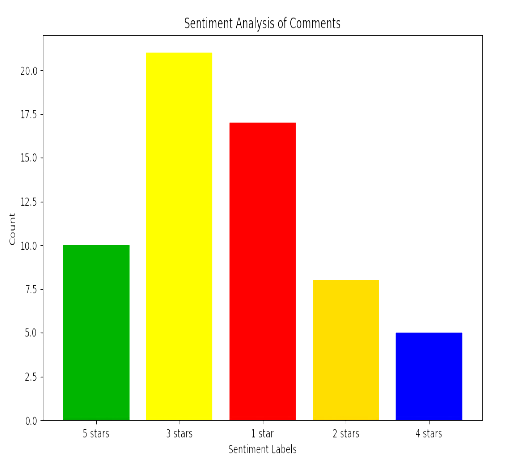


Figure 4.9 Sentiment pie chart.

3 something.

Sentiment pie chart.

Figure 3.8 Sentiment bar chart.

3 something.





Figure 4.11 Sentiment score DISTILLBERT

3 something.

Sentiment score BERT.

3 something.

Figure 4.7 Sentiment score BERT

3 something.

Sentiment score BERT.

3 something.

bar chart.

3 something.

.6 Sentiment score DISTILBERT.

3 something.

Figure 4.10 Sentiment score BERT

3 something.

Sentiment score BERT.

3 something.

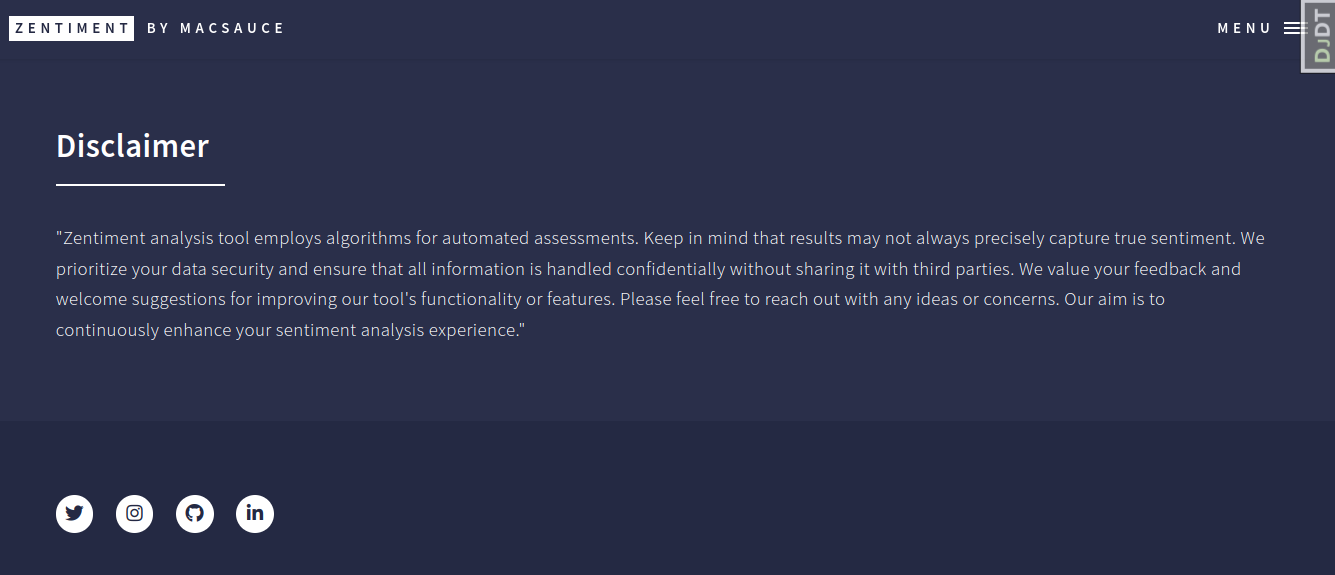


Figure 4.12 Disclaimer Snippet.



Figure 4.13 Result Snippet.

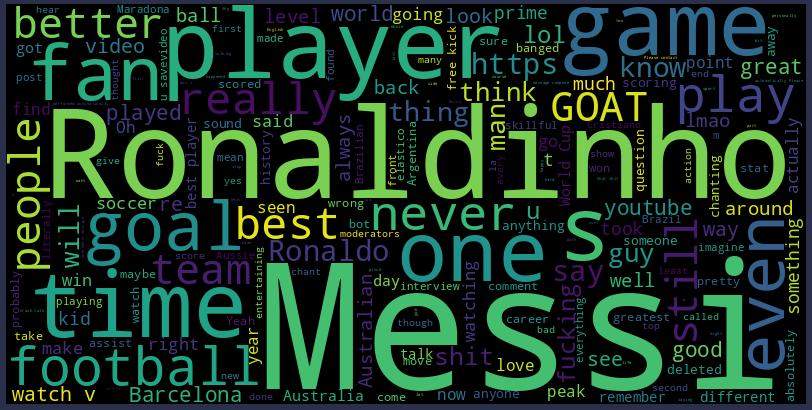


Figure 4.14 Word cloud result for Messi.

### **4.5 SOURCE CODE OF THE PROJECT**

The whole source code of this project can be found on GitHub which is a code hosting platform for version control and collaboration.

**GitHub:** <https://github.com/CodeLord2020/Final-Year>

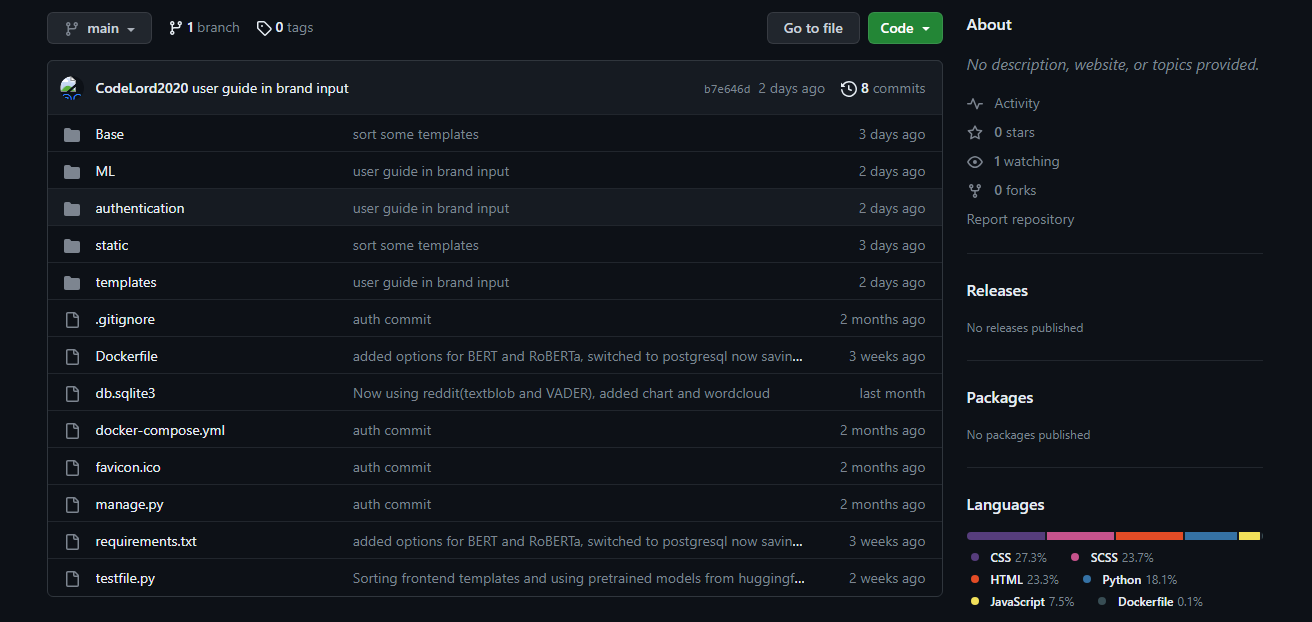


Figure 4.15 Deployed on GitHub.

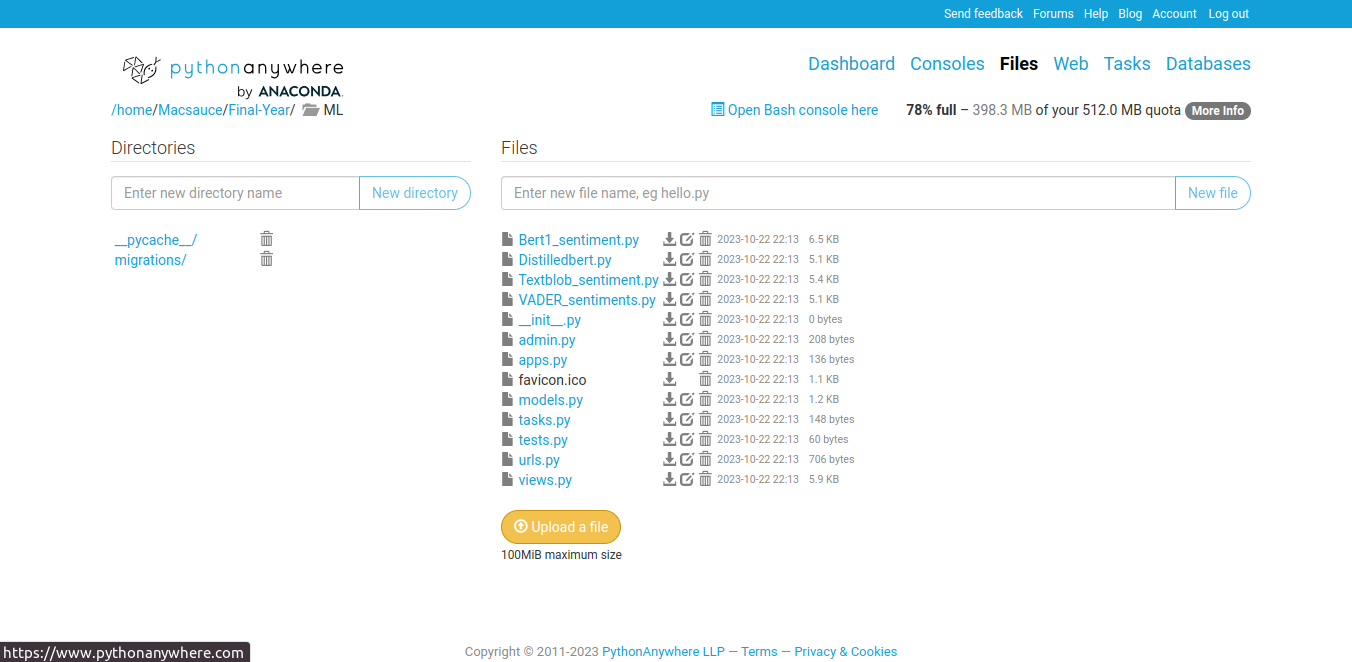


Figure4.16 **Deployed on Pythonanywhere.**

The links below could also help clarify the NLP methods implemented in the project.

TEXTBLOB**:** <https://textblob.readthedocs.io/en/dev>

VADER: https://vadersentiment.readthedocs.io/en/latest

BERT**:** https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment

DISTILBERT**:** https://huggingface.co/lxyuan/distilbert-base-multilingual-cased-sentiments-student

### **4.6 RESULT**

The project's results encompass a comprehensive sentiment analysis of user-generated comments, categorized into positive and negative sentiments. The top 10 comments in each category are presented, offering insights into highly praised and criticized aspects. To visually represent sentiment distribution, bar charts, pie charts, and a distribution chart are included. A word cloud is used to highlight frequently mentioned terms, shedding light on prevalent themes. Lastly, the average sentiment score provides a quantitative overview of the sentiment expressed, summarizing the sentiment trends within the dataset.

# **CHAPTER FIVE**

## **CONCLUSION AND RECOMMENDATION**

### **5.1 CONCLUSION**

First and foremost, this project has successfully addressed some of the shortcomings of existing research methodologies in sentiment analysis, something which is the motivation behind this endeavor. The specific objectives, outlined as designing an Impression analysis web application, implementing it, and evaluating its performance, have all been achieved. The web application provides a user-friendly and accessible platform for individuals to gain insights into sentiments on a vast array of topics. This approach is free from the constraints of single static case studies, as often seen in prior research projects. This system empowers users to analyze any valid brand or keyword, offering a broader path for conducting sentiment analysis research.

This project has not only developed a practical application for immediate use but have also opened new possibilities for future research and advancements in the realm of sentiment analysis methodologies.

### **5.2 RECOMMENDATION AND FUTURE WORKS**

In the wake of completing this project, several recommendations can be made to enhance its utility and further its impact. First and foremost, continuous updates and maintenance of the web application are crucial. Given the dynamic nature of social media and the ever-evolving language and trends that emerge, regular updates to the sentiment analysis algorithms and underlying models are necessary to ensure accurate results.

Secondly, expanding the sources of data collection beyond Reddit could be a valuable step. While Reddit serves as an excellent platform for diverse discussions, integrating data from other social media channels can provide a more comprehensive view of brand impressions. Platforms like Twitter, Instagram, and Facebook hold substantial potential in enriching the dataset.

Moreover, to make the application even more user-friendly, the development of a mobile app version could be considered. Mobile apps provide on-the-go accessibility and can significantly widen the user base.

Lastly, collaboration with businesses and researchers can be explored to further refine and tailor the system to specific needs. Establishing partnerships that allow the application to be fine-tuned for industry-specific sentiment analysis can lead to more accurate insights for decision-makers.

## **REFERENCE**

Abdelwahed, N., Ben Letaifa, A., El Asmi, S., & Ksouri, A. (2022). Predicting sentiment analysis for Web users with a deep learning approach. ResearchGate.

Akanksha, S., Shukla, S., & Gopal, V. (2022). Sentiment analysis on Amazon product reviews. International Journal of Computer Applications, 186(17), 14-20.

Anusuya, V., Murthy, V. P., & Rao, H. K. (2018). Sentiment analysis of product-based reviews using machine learning approaches. International Journal of Computer Applications, 180(2), 11-16.

Bhagyashri Wagh et al (2017). A Twitter Sentiment Analysis Using NLTK and Machine Learning Techniques from <https://www.researchgate.net/publication/326054514_A_Twitter_Sentiment_Analysis_Using_NLTK_and_Machine_Learning_Techniques>

Bharathi, R., Bhavani, R., & Priya, R. (2022). Twitter text sentiment analysis of Amazon unlocked mobile reviews using supervised learning techniques. Indian Journal of Computer Science and Engineering, 13(4), 1242-1253.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Fahad Alam. (2022). Applications of Natural Processing Language. [Applications of Natural Language Processing  | Data Science Dojo](https://datasciencedojo.com/blog/natural-language-processing-applications/).

Figure 1.2 taken from [Essential Applications Of Customer Sentiment Analysis | Presentation Graphics | Presentation PowerPoint Example | Slide Templates (slideteam.net)](https://www.slideteam.net/essential-applications-of-customer-sentiment-analysis.html)

Figure 1.3 taken from [Sentiment Analysis Using Python - Analytics Vidhya](https://www.analyticsvidhya.com/blog/2022/07/sentiment-analysis-using-python/)

Gupt, K., Jiwani, N., & Afreen, N. (2023). A combined approach of sentimental analysis using machine learning techniques. Journal of Artificial Intelligence Research, 100, 1-20.

Hootsuite. (2023, January 20). Social media sentiment analysis: Tools and tips for 2023. Retrieved from <https://blog.hootsuite.com/social-media-sentiment-analysis-tools>

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining, 10(2), 118-128.

Hutto, C. J., & Gilbert, E. (2014). Vader: A lexicon and rule-based sentiment analysis tool for social media text. (Section 3: Identifying the Intensity of Sentiment) Retrieved from <https://github.com/cjhutto/vaderSentiment>

Jemimah, J. O. (2021). Sentiment analysis of Amazon electronic product reviews using deep learning. Master’s Thesis, Dublin Business School.

Kausar, M. A., Fageeri, S. O., & Soosaimanickam, A. (2023). Sentiment classification based on machine learning approaches in Amazon product reviews. Engineering, Technology & Applied Science Research, 23(2), 1157-1163.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language

Loria, S. (2018). textblob Documentation. Release 0.15, 2.

Monroy, G. U., & Angulo, E. (2021). Sentiment analysis on the acceptance of the COVID-19 vaccine on Twitter. ResearchGate. https://www.researchgate.net/publication/356646480\_Sentiment\_analysis\_on\_the\_acceptance\_of\_the\_COVID-19\_vaccine\_on\_Twitter

Nitesh Sharma et al. (2018). Web-Based Application For Sentiment Analysis Of Live Tweets.

Qianwen, A., Chang, V., & Jayne, C. (2022). A systematic review of social media-based sentiment analysis: Emerging trends and challenges. Decision Analytics Journal, 3(1), 100073.

ResearchGate.

Rodriguez-Ibanez, M., & Casanez-Ventura, A. (2023). A review on sentiment analysis from social media platforms. Journal of Intelligent Systems, 32(2), 211-232.

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.

Sobia, W., Chen, X., Shen, T., Waqar, M., & Jhanjhi, N. Z. (2021). Amazon product sentiment analysis using machine learning techniques. Revista Argentina de Clínica Psicológica, 30(1), 695-703.

Technologies, 5(1), 1-167.

Vishal A. Kharde et al. (2022). Sentiment Analysis of Twitter Data: A Survey of Techniques. ResearchGate.

Wahab, D. A., & Putra, F. T. (2019). Influence of information technology on company development. IOP Conference Series: Materials Science and Engineering, 662, 022073. <https://doi.org/10.1088/1757-899X/662/2/022073>

Zeenia, S., Randhawa, S., & Jain, S. (2017). Sentiment analysis of customer product reviews using machine learning. In 2017 International Conference on Intelligent Computing and Control (I2C2) (pp. 1-6). IEEE.

Zhang, L., Wang, S., & Liu, B. (2020). Deep learning-based sentiment analysis for COVID-19.