****

**A REPORT ON SENTIMENT ANALYSIS 1**

**DEPARTMENT OF COMPUTER SCIENCE,**

**COLLEGE OF COMPUTING,**

**MCPHERSON UNIVERSITY, SERIKI SOTAYO, OGUN STATE.**

**BY: GROUP 4**

**TABLE OF CONTENTS**

[1.0 INTRODUCTION 3](#_xok1qk3yvp5y)

[1.1 IMPORTANCE OF SENTIMENT ANALYSIS 3](#_kimtuet5kibw)

[2.0 METHODOLOGY 4](#_c3vgqmjwfc9a)

[2.1 STEPS FOLLOWED 4](#_z1wxyxtbvbg3)

[2.2 METHODOLOGY USED 5](#_li3m0ln0sr2m)

[3.0 CHALLENGES FACED 6](#_zyoqwb27cwt)

[4.0 CONCLUSION 6](#_etn0qe3dauvr)

# **1.0 INTRODUCTION**

Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) technique that involves analyzing and determining the emotional tone or sentiment expressed in written or spoken language. It categorizes text as positive, negative, or neutral, providing valuable insights into the opinions and feelings of individuals. This analysis is the classification of textual data either from CSVs or live data on social media posts(basically any type of textual data that can be classified) into different sentiment categories such as positive, negative and neutral. Some models go as far as detecting emotions like sarcasm, anger, joy or sadness.

**How Does Sentiment Analysis Work?**

Sentiment analysis employs a combination of linguistic rules to assess the sentiment of text. It identifies keywords, phrases, and linguistic cues that indicate positive or negative emotions. These algorithms then assign a sentiment score to the text, which can range from highly positive to strongly negative

## **1.1 IMPORTANCE OF SENTIMENT ANALYSIS**

Understanding public sentiment is important for businesses, governments and organizations. Opinion mining enables these authoritative figureheads to extract meaningful insights from large or small amounts of textual data, which helps in decision-making.

The following listed include how and where sentiment analysis can be used:

A. **Business Insights**

In the business world, understanding customer sentiment is critical. By analyzing customer feedback, reviews, and social media mentions, companies can gain valuable insights into customer satisfaction and identify areas for improvement. This data can inform product development, marketing strategies, and customer service enhancements.

B. **Brand Reputation Management**

A positive online reputation is invaluable. Sentiment analysis allows brands to monitor their online presence, track mentions, and address negative sentiments promptly. By doing so, they can protect their brand’s image and maintain customer trust.

C. **Market Research**

Sentiment analysis is a game-changer in market research. It enables companies to gauge public opinion about their products or services and those of their competitors. This data helps in making informed decisions, launching new products, and staying ahead of market trends.

D. **Political Analysis**

In the political arena, sentiment analysis is used to assess public opinion, especially during election campaigns. It helps politicians understand their constituents’ concerns and tailor their messages accordingly.

E. **Customer Feedback Analysis**

By analyzing customer feedback, companies can identify recurring issues and address them promptly. This not only improves customer satisfaction but also fosters loyalty.

F. **Stock Market Predictions**

Sentiment analysis is increasingly used in the financial sector to predict stock market trends. It analyzes news articles, social media posts, and financial reports to assess market sentiment, helping investors make more informed decisions.

G. **Crisis Management**

During a crisis, understanding public sentiment is crucial for effective crisis management. Sentiment analytics can help organizations gauge the severity of public sentiment and respond accordingly.

H. **Social Media Monitoring**

With the proliferation of social media, sentiment analytics is essential for monitoring brand mentions, tracking trends, and managing online reputation.

# **2.0 METHODOLOGY**

To implement a functional sentiment analysis the social media app: REDDIT was used. A post on Reddit: “Trump taps Rep. Matt Gaetz as attorney general,” was picked to analyze. The steps and methods used in the process of analyzing this post will be highlighted and explained extensively below:

## **2.1 STEPS FOLLOWED**

1. Authenticate with Reddit: this involved getting the client\_id and client\_secret which was done with the help of ChatGPT to guide the developer on how to gain access to reddit as it is an open-source platform.
2. Get The Reddit post: a post was chosen and the comment section was analyzed.
3. Extract the topic and all comments
4. Organize data into a dataframe
5. Save to a csv file if needed
6. Display first few comments
7. Data Cleaning Process:
8. Function to clean text
9. Apply function to each text

VIII. Write a function to get the sentiment

IX. Apply the sentiment Analysis

X. Inspect the results

XI. Classify the sentiments

## **2.2 METHODOLOGY USED**

The methodology used follows these key steps:

1. **Importing Libraries**: The code begins by importing necessary libraries:

import praw

import pandas as pd

**praw (Python Reddit API Wrapper)**: This library is used to interact with Reddit's API to fetch posts and comments.

**pandas**: Used to organize extracted comments into a structured DataFrame for further processing.

1. **Data Collection (Reddit Scraping)**
2. Uses praw (Python Reddit API Wrapper) to authenticate and extract comments from a specific Reddit post.
3. The script retrieves the post title and all associated comments, including nested replies.
4. Stores the extracted comments in a Pandas DataFrame and optionally saves them to a CSV file

3.  **Data Cleaning:** To clean the text regular expressions (re module) were used to :

1. Removes URLs.
2. Removes mentions (@username).
3. Eliminates special characters, keeping only letters and spaces.
4. Converts text to lowercase and removes extra whitespace.

A new column (cleaned comment) is created in the DataFrame with the processed text.

4. **Sentiment Analysis:** TextBlob was used for sentiment analysis.

1. A function was defined to compute sentiment polarity (ranging from -1 to +1). The scale was used to categorize each comment into different classes.
2. This function was applied to the cleaned comments, storing the sentiment scores in a new column (sentiment).

# **3.0 CHALLENGES FACED**

1. **Limitations of the Reddit API**: The Reddit API has rate limits on the number of requests in a given time frame, which might impact large-scale data extraction.
2. **Data Imbalance:** Unequal distribution of sentiment classes in datasets.
3. **Differentiating Clime specific words:** Sentiment varies across contextsand domains (e.g., “sick” can be positive in slang but negative in healthcare)
4. **Data Preparation:** Comments on Reddit may include spam, responses of bots, and other irrelevant texts that will need filtering.
5. **Ambiguous Sentiment**: Some comments may contain sarcasm, mixed sentiments, or ambiguous language that makes sentiment analysis difficult. Security Risks: Hardcoded API credentials are a security risk and must be handled via environment variables or configuration files.
6. **Internet Dependency:** The script requires an active internet connection; hence, it may fail to fetch the data due to API downtime or network failure.
7. **Detecting Sarcasm and Irony**: Difficulty in identifying and differentiating non-literal language.

# **4.0 CONCLUSION**

In this project, the script can fetch comments in a targeted post on the social network website: Reddit using a library called PRAW. Finally, we paid close attention to the information collected and they were given structure so that the sentiment analysis can follow easily. While the initial code block collects data, the other processes would further implement a Sentiment Analysis using TextBlob.