

OpinionPoint - A Flask Based Web Application For Sentiment Analysis

Chirag Vijay

Student

School of CSE

REVA University

Bengaluru, India

R18CS101@cit.reva.edu.in

Clinton Jude Noronha

Student

School of CSE

REVA University

Bengaluru, India

R18CS260@cit.reva.edu.in

Devika Vijayan

Student

School of CSE

REVA University

Bengaluru, India

R18CS114@cit.reva.edu.in

Chatakonda Venkata Karthik

Student

School of CSE

REVA University

Bengaluru, India

R18CS092@cit.reva.edu.in

Vinay Kumar M

Student

School of CSE

REVA University

Bengaluru, India

vinaykumar.m@reva.edu.in

Abstract—In the present era, many people tend to voice their opinions via different platforms such as Facebook, Twitter, Reddit etc, and the data posted by these people (i.e. the users) is growing exponentially. This growth has attracted various clients relying on data on the web to gather information on various stuff (such as products, services, events and so on) in an attempt to forecast opinions and attitudes. This task of attempting to forecast opinions and attitudes by studying and analyzing people’s opinions, emotions, sentiments, evaluations, and attitudes is known as opinion mining, or sentiment analysis [1]. In this paper, we explain the process in developing a web based sentiment analysis application with Flask (a micro web framework for Python) called “OpinionPoint” where opinions are gathered through tweets posted on Twitter. We have chosen Twitter because we can easily extract tweets with the help of Twitter API which is useful in generating a huge amount of sentiment data after processing. Twitter sentiment analysis allows one to keep track of what’s being said about a certain product or service in real-time or over a specific period of time, and it also helps one detect angry/dissatisfied customers or negative mentions before they can escalate.

Keywords— *OpinionPoint, Sentiment Analysis, Opinion Mining, Python, Flask, Twitter*

I. INTRODUCTION

Despite the long history of linguistics and natural language processing (NLP), according to the author in [1], not much research on people's thoughts or sentiments had taken place before the year 2000; however, since then, this discipline has become an active area of focus. Three explanations are given by the author for the sudden interest in this topic. For starters, it has various applications in numerous fields. Second, it presents a variety of difficult research questions that have never been investigated previously. The third area on which we concentrate is the vast amount of opinionated material available on the web through social media platforms. All this information was vital, and it would not have been possible to conduct study without it.

With the massive rise of social media on the internet, various services such as Twitter have developed as a plethora of data where individuals from all over the world express their opinions via tweets on a variety of topics. According to the data available, Twitter has approximately 396.5 million users, with approximately 206 million daily active users and approximately 500 million tweets being generated daily [2]. Through these tweets, one can easily extract huge amounts of data for analysis. We leverage this opinionated data available in the form of tweets and simplify the task of performing sentiment analysis through OpinionPoint, which is a web application built with Flask framework. It performs sentiment analysis based on a keyword inputted by the user and delivers clean, straightforward and neat results with supporting graphical visuals for clear understanding.

II. LITERATURE SURVEY

Doxa, an Ancient Greek Term which refers to an Opinion [3], is used to introduce the concept of sentiment analysis. On Lumen Learning website [4], in Chapter 6 under American Government section named as “Polling the public”, it is stated that During the 1824 United States Presidential election Harrisburg Pennsylvanian performed a “straw polling” in order to predict the results of the election, however the prediction was erroneous, presumably as a result of an improper sample. In 1997 the term semantic orientation was used by Hatzivassiloglou et al., in their paper which was based on predicting the results based on certain adjectives [5]. They took the corpus-based approach. It could explain that the terms 'bull' and 'bear' are contradictory in the reports of the stock market. The model achieved 90% precision using only adjectives.

Machine Learning was applied to sentiment analysis by Pang et al., in 2004. They suggested a detector based on the subjectivity which picks out various subjective sentences and the utilize various techniques to do the text categorization on those sentences. They achieved 86.4% accuracy. [6].

In 2005, Daniel et al., published their paper titled Predictive power of Online chatter [7]. They conducted a study to assess whether the comments which are present online have an impact on product’s sales figures or not. They have used Amazon sales data of books to compile this study and discovered that good comments led to greater sales of books.

In 2012, Onur Kucuktunc et al., published a paper which was based on Sentiment analysis of Yahoo! [8]. Within the context of an online question and answer site, they employed a sentiment extraction technique to study the influence of numerous parameters such as gender, age, education level, etc. They then discovered that the replies varied depending on the users attributes, for example, the answer which was best rated was having a neutral effect on the users. Eventually, their findings started being used in marketing and product recommendations.

In 2014, Cambridge Analytica gathered the personal information of around 87 million Facebook users, which was then used to create 30 million "psychographic" profiles of potential voters [9]. This data is believed to have been utilised to influence voter sentiment in favour of politicians who hired them.

III. MOTIVATION

The primary reason for conducting any type of Sentiment Analysis is to "know" and "understand." We concentrated on the same thing, which was to understand or get to know the user. It is a natural curiosity to want to discover what the majority of people believe or which point of view is currently dominating the realm of opinions.

It is the desire to keep up with the constantly changing world of social media and to always be aware of what information is now trending and why it is popular. A piqued interest in learning what the general public thinks about a particular political party, football team, or film in order to gain a better knowledge of how trends flow.

Knowing how wonderful your products are or how popular your brand is can be significant in the case of business or industries, and knowing how customers react to new offers that are launched can be important as well.

IV. FLOW OF DATA

All the data used for analysis is in the form of the tweets that were being posted on the Twitter. When we use the Twitter API, it queries our request to the database, which is initially processed through Tweepy., which is a python module that allows developers to easily access the endpoints of the Twitter (Fig. 1). Tweepy takes care of various low-level aspects such as HTTP requests, rate restriction, authentication, etc on behalf of the users.

Without Tweepy, this task would be lengthy and quite erroneous. It allows one to emphasize more on the functionality rather than low-level details [10].

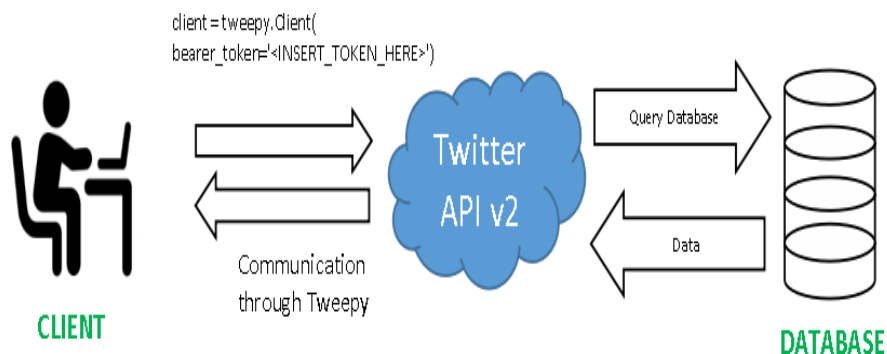


Fig. 1: Tweepy Python Library for easy communication with Twitter API v2

V. SOLUTION: OPINIONPOINT

After analyzing the current challenges faced, we propose our solution which is a Web Application based on Flask for Sentiment Analysis.

A. Features

- Free and easy-to-use web application for real time analysis of Twitter data to gather opinions.
- Generates a detailed report of the results after analysis is completed.
- Generates meaningful graphical representation of the detailed report.

B. Tools and Implementations

- 1) Visual Studio Code: All frontend (i.e. html, css, js) and backend code (i.e. python) was written on Visual Studio Code.
- 2) Flask: It is a micro-framework for web applications that is built on Python, built by Armin Ronacher. Flask is built with the Werkzeug WSGI toolkit and the Jinja2 template engine. The term "microframework" refers to Flask's goal of keeping the core basic but extensible [11].
- 3) Twitter API v2: Twitter API v2 is the most recent version of the Twitter API that all Twitter developers can utilise. Currently, it offers three levels of access (i.e. Essential, Elevated, Academic Research) and we have used the Essential level for this project [12].

Python Libraries Used:-

- a) Tweepy: Tweepy is a Python library which is Open source that provides a very simple way for accessing the Twitter API. Twitter's models and API endpoints are represented by a set of classes and methods. Also it includes various details such as: HTTP request data encoding and decoding, pagination of results, Authentication using OAuth, Rate Caps.

- b) TextBlob: TextBlob is a Python based text processing based library. It provides an API for analyzing into various (NLP) activities including noun phrase extraction, sentiment analysis, etc. [13].
- c) Matplotlib: Matplotlib, a Python package which is used to create different types of visualizations. It is built on NumPy arrays and can be used with SciPy. One of the most significant benefits of visualization is that it provides visual access to massive amounts of data in consumable formats.

C. Working

a) Front-end

Our web based application has been developed on Flask which is a lightweight python micro framework.

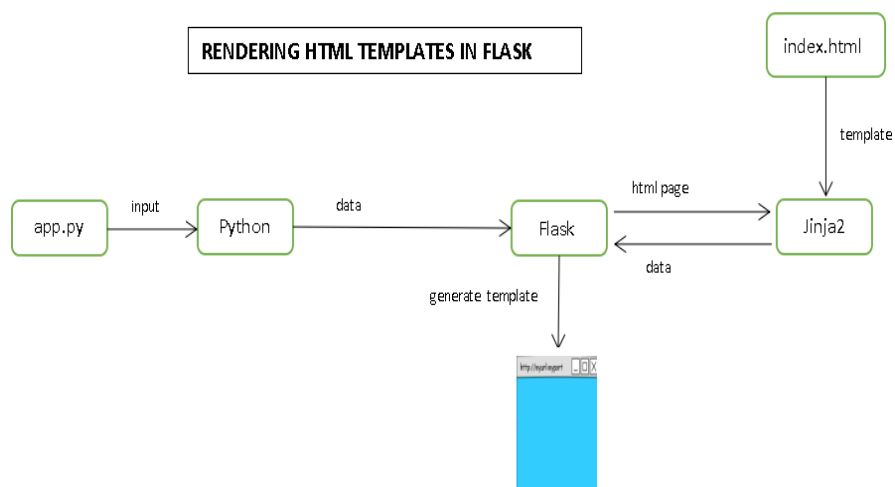


Fig. 2: Rendering of HTML Templates in Flask

Flask configures Jinja2 template engine for rendering HTML files (Fig. 2). Our application consists of three HTML files: the home page (Fig. 3), the sentiment analysis report page (Fig. 4) and the graphical visualization page (Fig. 5).

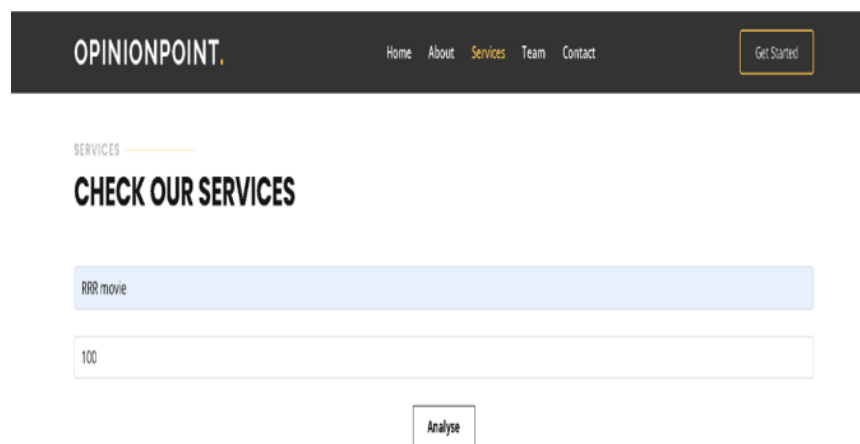


Fig. 3: Home Page



Fig. 4: Sentiment Analysis Report Page

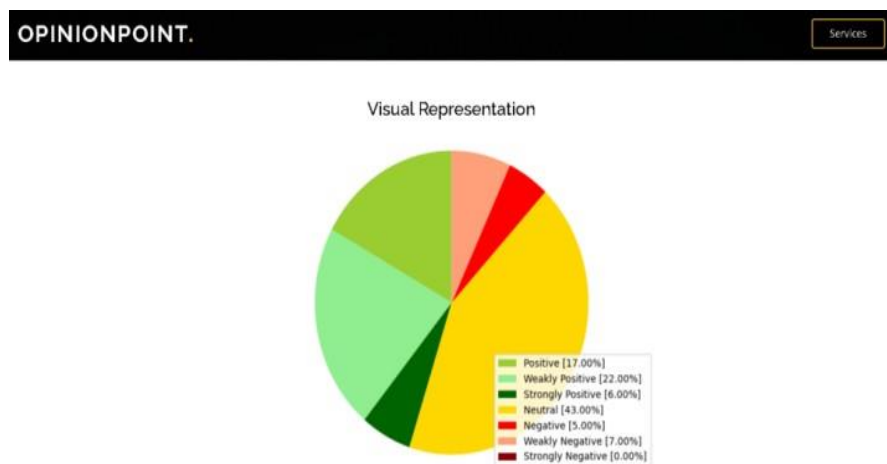


Fig. 5: Graphical Visualization Page

b) Back-end

In the figure below (Fig. 6), we can see the basic implementation of any sentiment analysis procedure.

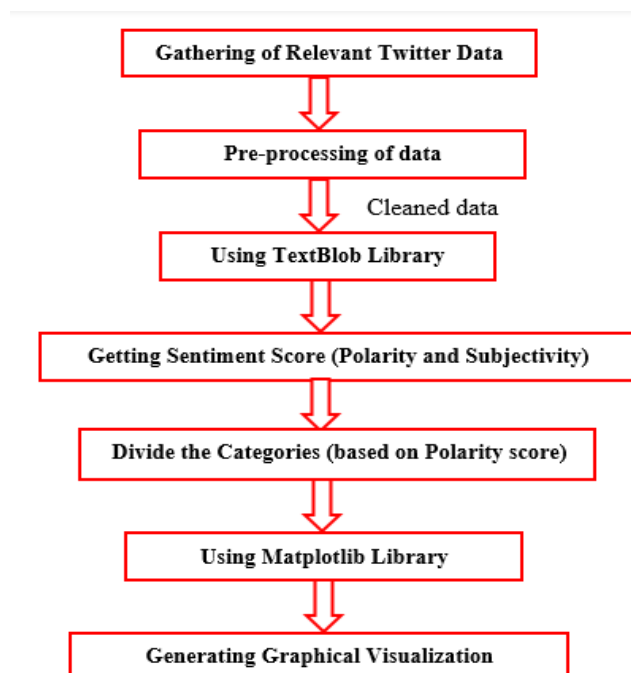


Fig. 6: Procedure to perform sentiment analysis

We will now go through each of the steps mentioned in the above figure (Fig. 6). The first step begins with collection of relevant data from twitter. This is done by querying the keyword inputted by the user through tweepy which in-turn communicates with Twitter API v2 (Fig. 1). Twitter API fetches for us the tweets related to the keyword and the fetched tweets are then stored in an excel sheet.

The next step involves the pre-processing performed on the tweets gathered. In this step, all of the tweets gathered need to be cleaned. Since the tweets are not always clean, the result could be affected (i.e. the accuracy of the results are proportional to how accurate the data is). The 5 main cleaning challenges that faced are :

1. Emotes are often used in tweets. However, when you pull the tweets, they are turned into unicodes. As a result, we eliminate them since we examine the sentiment of words rather than emotes.
2. Mentions, such as "@mention," aren't required to evaluate the tweet. Well, not all remarks can have an effect; only some usernames that give sentiments can have an effect on the outcome.
3. Hashtags can convey meaning in some circumstances, such as when a person posts a tweet with "#Happy," indicating that the user is pleased with the tweet he or she wrote. As a result, we remove only the hash, i.e., "#," while keeping the word as it is.
4. Remove stopwords such as is, to, as, the, and so on that are unnecessary because they do not convey any mood.
5. Tokenising the words will help us to make the tweet easy to extract the sentiment from it. We use a module named "tokenize" from the "nltk" library which tokenizes the words in the sentence. For example, the word "Crazyyyy" will become "Crazy" which gives a proper sentiment.

After the pre-processing of the data, in the next step all the cleaned tweets are inputted to a python library called Textblob. By using this library we can find the subjectivity and polarity of a tweet, these are the main factors which decide the sentiment of the tweet. The polarity and subjectivity is being calculated by the Textblob's sentiment function. Polarity is a number whose data type is float and range is from [-1, 1] where a positive result indicates 1 and a negative result indicates -1. Subjectivity refers to the degree in which the facts and opinion are present in the text. If the subjectivity is higher, it means that the text contains opinion rather than the facts. It is also a float which lies in the range of [0, 1]. The textblob.sentiments module contains two sentiment analysis implementations, PatternAnalyzer and NaiveBayesAnalyzer. For our project, we have used PatternAnalyzer.

Next we find the polarity score of each and every tweet and then categorize each tweet into different categories such as strongly positive, weakly positive, strongly negative, neutral and so on. The categorization is based on the polarity score. In Fig. 4, we can see this categorization in the detailed report page.

Finally, to represent this result visually, we use Matplotlib library as this library can help to visualize easily with clean and simple graphs and charts as seen in Fig. 5.

VI. LIMITATIONS

1. Given that the tweets will represent our views on the topic, there is a potential that the tweets' emotion will be distorted. For instance, sarcasm is a significant root cause. Algorithms aren't perfect at figuring out when people are being honest about their feelings or not. Which is why it's so difficult to comprehend what people are saying, even when they're telling the truth.
2. There are three versions of the Twitter API and three different API access levels; we are currently

using version Twitter API v2 and the essential level, which imposes constraints on how we utilize the API. Among the restrictions that influence our project are the following:

- The number of tweets that we can collect is restricted to 100.
- Because the twitter api only returns tweets that were recently posted, there is no way to retrieve tweets from several timelines (Since there is no access to advanced filter operators).
- Each month, we can retrieve a maximum of 500k Tweets.

The restriction issue can be resolved by applying for the third level of Twitter API v2 access i.e. “Academic Research” on the developer account.

VII. CONCLUSION

Presently, sentiment analysis has become a critical component of various huge enterprises and organisations and its use for data mining or obtaining competitive intelligence will only increase in the future.

"OpinionPoint" is a project that analyses millions of tweets and visually represents what people believe. To accomplish our goal of establishing a sophisticated website for our visitors, we incorporated a variety of technologies, including the Twitter API, the Python library "Textblob," Flask, and Bootstrap. OpinionPoint's user interface is straightforward and clear, and our users receive neat results. Our users may be non-programmers, business owners, industrialists, or anybody else interested in staying current with trends.

- The website can be used to determine whether or not a product or service is popular.
- It can also be used to determine whether a policy is popular or well-accepted in society.
- Additionally, to get a sense of how people feel about any generic idea or issue.
- The best feature of the initiative is that everything is real-time; every opinion you receive is immediate.

As we live in a world where technology are constantly developing, we will continue to work on discovering better APIs and algorithms to optimize our project for the benefit of our users.

ACKNOWLEDGMENT

The success of the completion of this project is an effort of strategic teamwork and a few people who were very helpful to us in this project. We would like to express our profound gratitude to our mentor, Prof. Vinay Kumar M, for his assistance and valuable inputs during the process of developing our project and conducting our research. Additionally, we would also like to express our sincere gratitude to our institution REVA University for providing us with this incredible opportunity and unwavering assistance in completing our project.

REFERENCES

- [1] Liu B. Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012.
- [2] <https://thesocialshepherd.com/blog/twitter-statistics>
- [3] Liddell, Henry George, and Robert Scott. 1940. "δοκέω." In A Greek-English Lexicon, edited by H. S. Jones and R. McKenzie. Oxford. Clarendon Press. – via Perseus Project.
- [4] Lumen Learning. 2020. "Polling the Public." Chapter 6 in: American Government, Lumen Learning.
- [5] Hatzivassiloglou, Vasileios, and Kathleen R. McKeown. 1997. "Predicting the Semantic Orientation of Adjectives." 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics, July, pp. 174-

- [6] Pang, Bo, and Lillian Lee. 2004. "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts." Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04), pp. 271-278, July.
- [7] Gruhl, D., R. Guha, Ravi Kumar, Jasmine Novak, and Andrew Tomkins. 2005. "The predictive power of online chatter." KDD '05: Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, pp. 78-87, August.
- [8] Kucuktunc, O., B. Barla Cambazoglu, Ingmar Weber, and Hakan Ferhatosmanoglu. 2012. "A large-scale sentiment analysis for Yahoo! Answers, Proceedings of the 5th ACM International Conference on Web Search and Data Mining." WSDM'12, ACM, pp. 633-642, February 8-12.
- [9] Chan, Rosalie. "The Cambridge Analytica whistleblower explains how the firm used Facebook data to sway elections". Business Insider. Retrieved May 7, 2020.
- [10] Sreedhar, K. S., Ahmed, S. T., & Sreejesh, G. (2022, June). An Improved Technique to Identify Fake News on Social Media Network using Supervised Machine Learning Concepts. In 2022 IEEE World Conference on Applied Intelligence and Computing (AIC) (pp. 652-658). IEEE.
- [11] Basha, S. M., Poluru, R. K., & Ahmed, S. T. (2022, April). A Comprehensive Study on Learning Strategies of Optimization Algorithms and its Applications. In 2022 8th International Conference on Smart Structures and Systems (ICSSS) (pp. 1-4). IEEE.
- [12] <https://developer.twitter.com/en/support/twitter-api/v2>
- [13] <https://textblob.readthedocs.io/en/dev/>