A Sentimental Analysis Report of Social Media Using Fine Grained NLP

Dr.P.A.Selvaraj, Dr.M.Jagadeeshan, Ramesh Arunachalam

Department of computer applications Kongu Engineering College,

Perundurai-638060,Erode,

Tamil Nadu, India

ABSTRACT:

Social media such as Facebook, Twitter, Instagram, News and reddit have recently attracted millions of users. Now-a-days social networking provides services like message posting, information sharing and opinions sharing to the world. So the government organizations and private organizations have started to consider the social media feeds and posts to know the comments and review about their organizations. But they felt difficult to understand the user opinions from various social media.

To make an ease of access to all social media, we are analysing a Web-Based application (Sentimental Analysis Report Of Social Media Using Fine Grained NLP) which will have the ability to integrate most of the information of popular social media and it will analysed in our server and send back the analysed report to our web app on a specific topic.

Our approach incorporates extraction of feeds from the major social media using API and Scrapping and by using the machine learning technology, we can separate the news feed into two different categories such as positive feeds and negative feeds in a form of pie chart, so that organization can understand the perspectives of different peoples directly from different social media easily through our app and can improve their organizations.

INTRODUCTION:

It is a web-based application where we can be able to search a keyword in most of the popular social media such as Facebook, Twitter, Reddit, LinkedIn, and News. From the extracted feed from different social media is sent to the next module of our application that is to analyse it using Machine Learning Algorithm which is used to separate the extracted feeds into some categories such as positive feeds, negative feeds, and neutral feeds. By this we can understand the public's view on the keyword in the social media. Then the pie-chart is used to display the public's view on the keyword which is from the social media. In the pie-chart positive, weakly positive, strongly positive, neutral, negative, weakly negative, and strongly negative are present. By the analysis, the result will be displayed on the basis of the user's post in the social media.

A module is a software and hardware item that is distinct from the rest of the system. Modular components are known for their portability, which allows them to be used in a number of systems, as well as their interoperability.

• Extraction

• Preparation of dataset

• Report generation using NLP

EXTRACTION:

In this module, we will be extracting the data using the public API and scraping tool. The extraction of data from each social media has different kinds of techniques to obtain the data. Some of the data that are on Facebook, Twitter, reddit, Instagram and news are extracted using the developer API.

PREPRATION OF DATASET:

In this module, the data, which is extracted from Facebook, Twitter, Reddit, Instagram, and news18 are cleaned using BeautifulSoap library. The cleaned data is sent to the TextBlob library, where the tweets are converted into datasets using inbuilt function of TextBlob library.

REPORT GENERATING USING NLP:

The obtained or extracted data from social media is then analysed and processed using the sentiment analysis algorithm of Natural language processing which is a field of machine learning technology. By processing the feeds based on the rating of the each post a cumulative report and pie chart will be generated and displayed to the user.

SENTIMENTAL ANALYSIS:

People frequently see emotion (whether good or negative) as the most essential value of a user's social media view. In reality, emotions provide a more comprehensive set of facts that influences and, in some cases, dictates customer decisions. As a result, Natural Language Processing is extremely useful for emotion-based sentiment analysis. Organizations can use NLP for speech analysis combined with a comprehensive social media monitoring strategy to improve customer experience, promptly manage customer complaints, and adjust their market position by analysing consumer reactions and responding accordingly. Natural Language Processing is being used by a variety of businesses to extract information from social data and recognise the impact of social media on brands and products.

SENTIMENT ANALYSIS WITH FINE-GRAINING

If polarity precision is crucial to your business, you should consider adding the following polarity categories to your polarity categories:

• Extremely positive

• Affirmative

• Undecided

• Disappointing

• Extremely negative

Fine-grained sentiment analysis is what this is known as, and it can be used to analyse 5-star ratings in a review, for example:

• 5 stars for Good review

• 1 star for Bad review

EMOTION DETECTION:

Emotion detection is the process of identifying feelings such as joy, frustration, anger, sadness, and so on. Many emotion recognition systems rely on lexicons (lists of words and the feelings they evoke) or sophisticated machine learning techniques. People communicate emotions in a variety of ways, which is one of the drawbacks of employing lexicons. Some words that are commonly used to communicate anger, such as bad or kill (for example, your product is so horrible or your customer service is killing me), can also be used to express satisfaction (e.g. this is bad ass or you are killing it).

ASPECT-BASED SENTIMENT ANALYSIS:

When analysing text sentiments, such as product evaluations, you'll typically want to figure out the individual aspects or attributes people are praising, neutralising, or criticising. In this text, an aspect-based classifier could figure out that the statement expresses a negative opinion regarding the feature battery life: "The battery life of this camera is too short."

ANALYSIS OF MULTILINGUAL SENTIMENTS:

It can be difficult to analyse bilingual sentiment. It needs a substantial amount of preparation and resources. The bulk of these resources may be found online (for example, emotion lexicons), while others must be created (for example, translated corpora or noise detection methods), but you'll need to know how to code to use them.

ARCHITECTURE DIAGRAM:

In the architecture diagram the user is giving a keyword that to be extracted from the social media and now the feed extraction is done in the social media that is like the keyword. The feed extraction will be done in Facebook, twitter, Instagram , news and Reddit. And now the extracted feeds will be going through sentiment analysis in machine learning. It will separate the positive feeds and negative feeds by using the sentiment analysis. The separated positive and negative reviews and reports will be displayed on the pie-chart and report which is the output send to the user
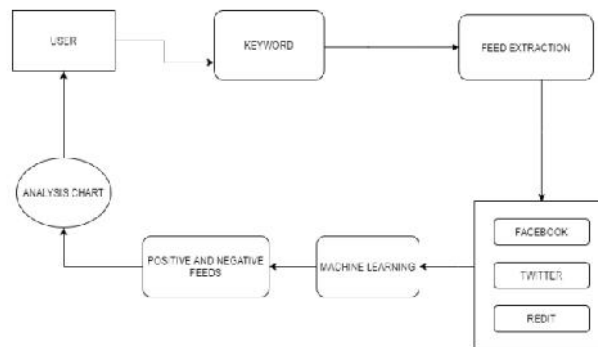
FRAMEWORKS AND LIBRARIES:

TEXTBLOB

Textblob is a text processing package for Python (2 and 3). It provides a straightforward API for doing standard natural language processing (NLP) activities like part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

Text Blob does NLP functions such as tokenization, sentiment analysis, and POS tagging, among other things. Text Blob or Vader are used to calculate sentiment. You evaluate if the text is favourable, negative, or neutral based on polarity and subjectivity. The polarity of a Text Blob is regarded positive if it is >0, negative if it is 0 and neutral if it is ==0. This is based on the compound score for Vader feelings. The classifier is then trained based on your emotions (positive, negative, or neutral) and predictions are made.

Flask:

Flask is a lightweight WSGI web application framework. It's designed to make getting started easy and quick, while also allowing you to grow up to more complex projects. It began as a simple wrapper for Werkzeug and Jinja and has since grown to become one of the most widely used Python web application frameworks.

BEAUTIFUL SOUP :

Beautiful Soup is a Python module for parsing HTML and XML files. It creates a parse tree for parsed sites, which can be used to scrape data from HTML. In our project, this software is utilised to extract feeds from social networking sites**.**

IMPLEMENTATION:

EXTRACTION OF DATA:

The data is extracted from twitter, Facebook, Instagram, Reddit, News18 using APIs namely praw, tweepy, request.

 • Praw is an API which is used for accessing the Reddit API and getting the data from the Reddit website.

 • Tweepy is an API which is used for accessing the Twitter API and getting the data from the Twitter website.

 • Request API is used to get data from any website. Since we cannot access data from Facebook and Instagram, so we are using this request API to get the same through a different website that is Hash edit.

After getting the authorization we must use that to obtain the data from the social media sites. We have done testing the API by checking whether the data is correctly extracted from the social media. The below figure data is belonging to the keyword "Indian Railway" and the data were fetched correctly.

ANALYSIS OF DATA:

To analyse the data using the sentiment analysis algorithm we have test it on the provided dummy data. In that case we will be sending the data one by one for the analysis processing using the textblob which is library file in python packages which is use concept of natural language processing. By using the textblob we can analysis the data word by word any filter out the unwanted word and results were merged based upon the polarity level of the word.
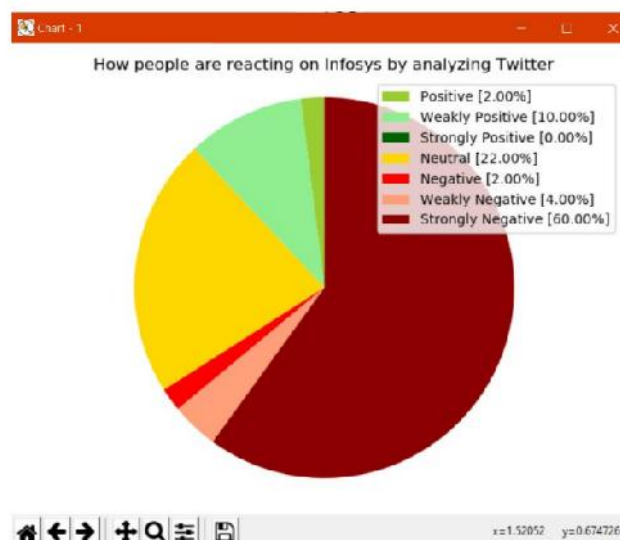
DISPLAYING CHART :

Next, we must test the pie chart also. After doing the sentiment analysis for the data. We must display the result in a pie chart so that the end user can understand the report easily. In order for testing, it we have used the sample results and transformed it into a pie chart

SOFTWARE TESTING

Software testing is a way of analysing the functionality of a software application with the purpose of detecting whether the generated software fits the given requirements and identifying faults in order to produce a high-quality output that is defect-free.

LEVELS OF TESTING:

- System testing
- integration testing
- unit testing

UNIT TESTING:

Unit testing is used to see if the source code's constituent modules are working properly. i.e. the developer testing each and every unit of the application individually in the developer's environment. It's also known as Component Testing or Module Testing.

SOCIAL MEDIA API TESTING:

To test whether the application program interface of the social media we may use Postman application to do the unit testing.
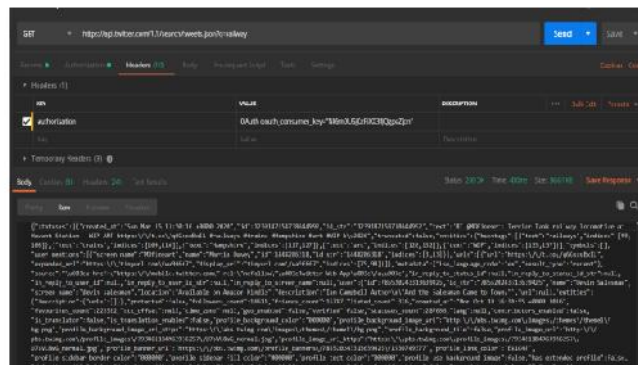
To test the application program interface, we may need to gain the Authorization from the social media such as Access token, Token Secret, Consumer Key, Consumer Secret
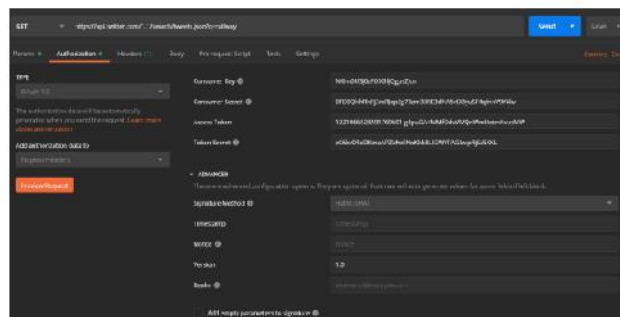
SYSTEM TESTING (END TO END TESTING):

It's a closed-loop test. End-to-end scenario testing is the process of testing a fully integrated application. To guarantee that the programme is compatible with all of the target systems. Check for desired outputs by thoroughly checking each input in the programme. Experiences of users with the application are being tested.. To check the user's experience, we have executed the application completely and checked whether the data input and output were as per the plan. All the modules were integrated, and application is executed. This application is a platform independent it can work on any operating system. When the application is being

launched, we can see the user interface were displayed to interact with the user. By getting the input from the user such as keyword and media Do the analysis button is clicked.

After clicking the analysis button we can see the process is that has been executed in the terminal and it started to collect data from the Facebook, then the data from the Facebook is being analysed and then report were stored. Then the data from the Twitter is being analysed and then report were stored. Finally, the data from the reddit is being analysed and then report were stored.

From the stored result data report the pie chart has been created for each social media individually. And it has displayed to the user. From the stored result data report the pie chart has been created for each social media individually. And it has displayed to the user.
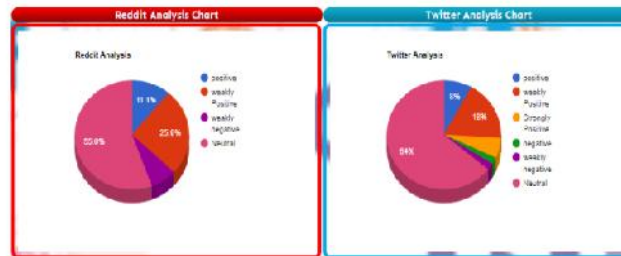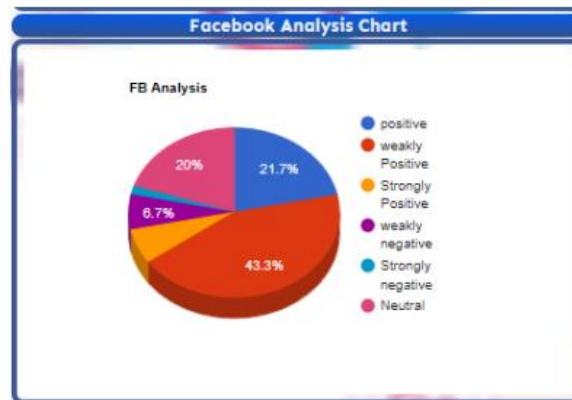
The confusion matrix considering the 5 factors that is:

 • Strongly Positive

• Positive

• Strongly Negative

• Negative

• Neutral

	0	1	2	3	4
0	26	6	309	80	171
1	2	47	80	64	406
2	28	9	359	256	389
3	0	4	134	300	181
4	3	34	206	179	525

According to the reading, in the above confusion matrix, the calculated results of Accuracy, Precision and recall is given below:

 Accuracy = 33%

 Precision = 55%

 Recall = 31%

Which is very less Similarly if we see closely Strongly Positive and Positive can be taken as Positive andsimilarly the Strongly Negative and Negative can be taken as Negative, So taking that as consideration, the confusion matrix considering the 3 factors given below is:

• Positive

• Negative

• Neutral

	0	1	2
0	722	336	575
1	134	300	185
2	291	243	1012

According to the confusion matrix values the calculated results of Accuracy, Precision and recall is given below:

Accuracy = 53%

Precision = 65%

Recall = 57%

We can see the increase in the reading when it is tested on a different dataset having only Positive and Negative values the following was the confusion matrix obtained:

	0	1
0	380	120
1	106	394

According to the confusion matrix values the calculated results of Accuracy, Precision and recall is given below:

 Accuracy = 77%

Precision = 78%

Recall = 76%

 So, from this we can say that our machine gives an average accuracy of 54% and it ranges from 54% to 77%. In best case the machine gives 77% accuracy, 78% precision and 76% recall.

CONCLUSION:

Thus, the project helps to analyse the data from the popular social media. For sentiment analysis data is taken from blogs, social media website like Facebook, Twitter, Instagram, NEWS18. People freely express their view on these media about certain brands. By analysing these reviews one can extract the information about their area and can-do improvement and it display a pie chart as an output to the user. Which helps the user to understand about the organization and public opinions clearly.

REFERENCE:

1. Behrooz Noori (2021) "**Classification of Customer Reviews Using Machine Learning Algorithms**", Applied Artificial Intelligence, 35:8, 567-588, DOI: 10.1080/08839514.2021.1922843
2. Yonghui Dai & Tao Wang (2021) "**Prediction of customer engagement behaviour response to marketing posts based on machine learning**", Connection Science, 33:4, 891-910, DOI: 10.1080/09540091.2021.1912710
3. Raghavendra Vijay Bhasker Vangara (2020) "**Opinion Mining Classification using Naive Bayes Algorithm**", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-5
4. Kiran Chaudhary1 ,MansafAlam(2021), "**Machine learning-based mathematical modelling for prediction of social media consumer behaviour using big data analytics**",HTTPS://DOI.ORG/10.1186/S40537-021-00466-2
5. Mazhar Javed Awan1,2,*, MohdShafryMohd Rahim1 **(2020)"Social Media and Stock Market Prediction:A Big Data Approach**", Computers, Materials & Continua DOI:10.32604/cmc.2021.014253

.