Sentiment Analysis Detecting Polarity in Text

Pragathi R Gowda
Dept of CSE,
GSSSIETW, Mysuru

Vinutha N
Dept of CSE,
GSSSIETW, Mysuru

Prathiksha M
Dept of CSE,
GSSSIETW, Mysuru

Harshitha M S
Dept of CSE,
GSSSIETW, Mysuru

Shyleshwari M Shetty
Assistance Professor, Dept of CSE
GSSSIETW, Mysuru

Abstract: we can see that now a days there is raise of social media, which in turn have changed the perspective in every field such as networking, socialization and personalization. People are using these data for different purpose like predicting the results of the election, communication, business, and marketing and in many more fields. Here extracting the valuable information that is from twitter which is a social media platform and then analyzing the tweets and then classifying those tweets into different polarity like positive, negative and neutral. And yes the data extracted can be from any of the domain which is specified. The application proposed here facilitates the extraction of some keywords, entities, synonyms and parts of speech that will be used in tweets and this will be used to segregate or classify the tweets that are data and perform sentiment analysis. This project is mainly on twitter data abstracting but we can also do this polarity check for other applications also like for example Facebook , Instagram.

I.INTRODUCTION

The project is mainly about extracting the data from the twitter application and then detecting the polarity for the extracted data. This can mainly help in stopping the hatred speech that is text content in the twitter. Here in our project we are detecting the polarity of the text and classifying then as positive, weakly positive, strongly positive, negative, strongly negative, weakly negative and neutral. This project can be useful for such a way that it reduces the time for reading the each and every comment from the twitter here instead we have the final result automatically. For example consider the review of any product instead of reading each and every comments in the twitter we can have the final result as a graph and can easily come to conclusion that how many people are liking the product how many are not. We here have also added the extra feature that is we also save the each comment specifying its polarity.

II. LITERATURE SURVEY

Apoorv Agarwa, et.al [1] Improve pre-processing techniques of tweets and use baseline machine learning methods. "Using many emoji occurrences to find out any-domain representations for detecting sentiment, emotion and sarcasm",

Bjarke Felbo, et.al [2] The goal is to predict emojis for a tweet Develop and train a deep learning model using LSTMs Use transfer learning to infer on other datasets with finetuning. Improve pre-processing techniques of tweets and use baseline machine learning methods.

Moumeen, et.al [3] Given an opinionated document d, Discover all opinion in Identify all synonyms in Determines whether there's opinion on a feature during a sentence, and if so,whether it'spositive or negative.Garin Kilpatrick [4] Introduced list of all twitter tools to collect and analyze Twitter data. He divided all Twitter tools into 53 categories. These tools provide facility in backup tweets, analysis, tweets translation, voice   tweet, and Twitter statistics.

IlknurCelik et al. [5] studied semantic relationship between entities in Twitter to provide a medium where users can easily access relevant content, they are interested in.

MorNaamen et al. [6] studied the user's behavior on Twitter. They applied human coding and chemical analysis of tweets to know users activities on Twitter. They analyzed that majority of users specializein self(memo formers) while small portion of users share information withothers (informers).

III DESIGN

System Requirements:
Hardware Requirements
Core i3,5th gen +processor
4GB +RAM
80+ Harddisk
SoftwareRequirements:
Programming language: Python
Framework:Tweepy (Twitter),Flask(Web)
IDE : Jupiter Lab, VSCode
Operating System : Windows 7+ or Mac ,Linux

Existing System
- Apriori algorithm is used this algorithm fails to handle large data set.
- Data retrieval is not allowed based on the domain or query entered by the user.
- The database for the searched domain will not be stored.

Proposed System
- We are extracting the data from Twitter by using Twitter API
- After extracting the tweets preprocessing of tweets is done.
- Naive Bayes algorithm is used here to classify the data
- This is a web application.

- Here the result is obtained for the particular domain we search for ,that is hashtag (eg: #coronaSystem Architecture
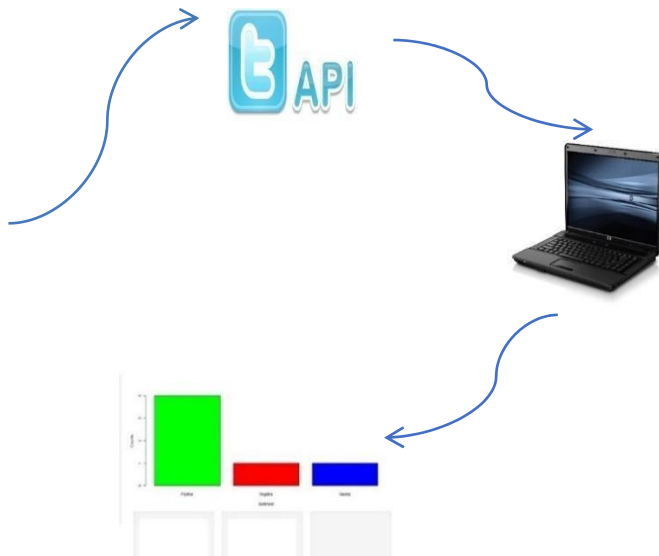

Fig. 1. System Architecture of Sentiment Analysis

Twitter:

This is an application which we are using to read the data and are giving the output based on the data provided(here data means the comments). We can find the results for any of the data which we want that can be of any domain.

Twitter API:

Here in this module when the user creates his/her personal account they will be able to access to the data from the twitter. This application is used to extract the data from the twitter. Here data that is comments can be extracted by specifying the topic for which we want to know the results.

Web Application:

Here we code the program in python using different algorithms so that it works accordingly. We have used naïve Bayes and multinomial NB algorithms as these algorithms main purpose is to classify the text. As we have coded the program in python we are developing this as an web application (As python does not support android).

Output:

As shown in the figure our output will be displayed in the form of graph which makes us easy to understand. Output is shown both in graph and percentage.

Disadvantages

- Here this dataset works only regarding to a particular domain or topic.
- We use Naive Bayes algorithm and this may take more time to extract large data set
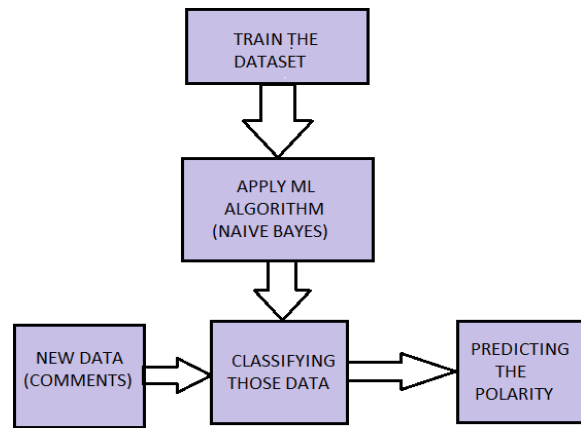- We can not make this as an Android application as we use python language.

System flow diagram

Fig 2 : System Flow Diagram of Sentiment Analysis

In figure 2 it shows the steps used for creating the application.

VI.IMPLEMENTATION

Feasibility Study: A feasibility study is analysis that takes all of a project's relevant factors into account- including economic, technical, legal, and scheduling considerations-to ascertain the likelihood of completing the project successfully.

Technical Feasibility: A Project is taken into account technically feasible if it's the required expertise, infrastructure and capital to develop, install, operate and maintain the proposed system. Operational Feasibility: Operational Feasibility will measure how well the proposed new system solve the problem and how it satisfies the Requirements identifies in system phase development.

Naïve Bayes Classifier (NB): Naïve Bayes classifiers are among the simplest Bayesian network models. But they might be including kernel density estimation and achieve higher accuracy levels. Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables in learning problem. The model works with the BOWs feature extraction which ignores the position of the world in the document.

$$(1), P(label|features) = \frac{P(label)*P(features|label)}{p(features)}$$

Equation 1 is,p(label) :is the prior probability of a label or the likelihood that a random feature set the label. P(features|label) :is the prior probability that a given feature set is being classified as a label. P(features) :is the prior probability that a given features set is occurred. Given the Naive assumption Which states that each one features are independent, the equation 2 might be rewritten as follows:

$$(2), (label|features) = \frac{P(label)*P(f1|label)*\ldots\ldots*P(fn|labe)}{P(features)}$$

Special Issue - 2020

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
IETE – 2020 Conference Proceedings

Algorithm:

Dictionary generation Counts occurrence of all word in our whole data set and make a dictionary of some frequent words.

Feature set generation all document is represented as a feature vector over the space of dictionary words.

For each and every document, keep track of dictionary words along with their number of occurrence in that document.

White Box Testing: White Box Testing may be a testing technique, that examines the program structure and derives test data from the program logic/code. In order perform white box testing on an application, the tester needs to possess knowledge of the internal working of the code.

Black Box Testing: It is a method of software

testing that examines the functionality of an application based on the specifications.

Verification and Validation: Verification and Validation are the processes to check whether a software system meets the specification and that it fulfills its intended purpose or not. Verification and Validation are two different things
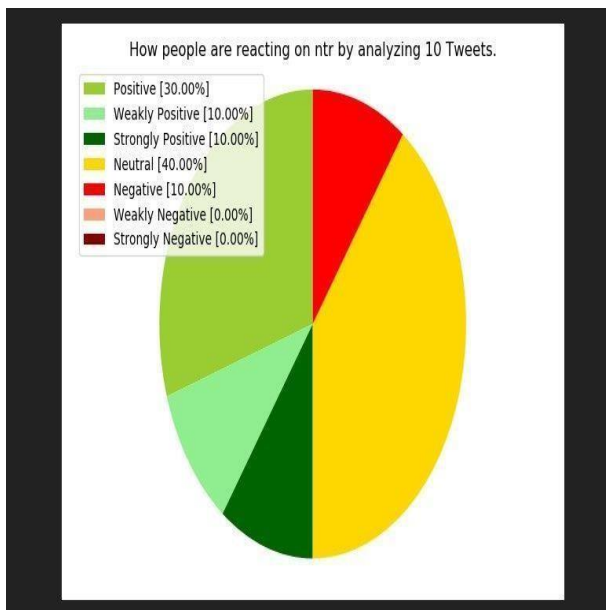
FINAL RESULT:

Fig3.Showing results as pie chart with percentage

Figure 3 showing the final result , that is as a pie chart . Displaying the percentage in different categories.

1.26E+18 RT @BinduChowdary99:	neutral	
1.26E+18 RT @NarikesthaOba:	Positive	
1.26E+18 Happy birthday to you nt	Strongly Positive	
1.26E+18 RT @vamsikaka:	Weakly Positive	
1.26E+18 RT @BinduChowdary99:	neutral	
1.26E+18 Love You NTR	Positive	
1.26E+18 Happy birthday to you nt	Strongly Positive	
1.26E+18 Kannada Superstar	Weakly Positive	

Fig4.Saving each comments in a file defining its polarity

Figure 4 is the data that will be stored in a csv file for each   search we do.

Applications of sentiment analysis:
- Sentiment analysis can be applied in many fields such as Movies, Products, etc.
- In Business can analyze brands and new product perceptions.
- Application in Business Intelligence.
- In Social media's can find the like minded individuals and reviews from feedback shows the percentage of positive and negative which can be also used as public sentiments.
- Applications in blogs, articles, and any posts and tweets in twitter.
- In the Politics it gives the insights about what does people think about the issues and the candidates.

VII. FUTURE ENHANCEMENT

Sentiment analysis is uniquely powerful tool for the Business that are looking to measure attitudes, Feelings, and emotions regarding their brands. By Investigating and analyzing customer sentiments, Those are able to get an inside look at consumer Behavior and ultimately better serve their audience With the products, services and experiences offer.

The future of sentiment analysis is going to continue a dig deeper and far past the surface of number of number of likes , comments and shares and it will aim to reach and truly understand the significance Of social media interactions and what they tell us About the consumers behind the screens. This also Predicts broader applications for sentiment analysis So brands will continue to leverage this tool but so Will individuals in the public eye, government of Education centers and many other organizations.

CONCLUSION

This approach provides improvement in the accuracy by using simplest features and small amount of data sets. It helps user to get conclusion about polarity of comments , whether it is positive or negative or the neutral so that not needed to read all comments. So using this one can save their time and within instant they can get the final results. The important part of gathering is to know what other people think. The sentimental analysis done here is obtaining opinions of people either the positive or the negative. Depends upon the views which are extracted from the tweets. Analysis will helps in more places like for News, Political issues, about current issues happening in the country and how people views about it will be get to know and utilized properly. It is also useful in businessdevelopments like improving customer services and developing quality products. Obtains opinions of people depends upon the views extracted from the tweets.

REFERENCES

[1] Corrêa EA Jr, Marinho VQ, Santos LB (2017) Nilc-usp at semeval-2017 task 4: a multi-view ensemble for twitter sentiment analysis.

[2] Dovdon E, Saias J (2017) ej-sa-2017 at SemEval-2017 Task 4: Experiments for Target oriented Sentiment Analysis in Twitter.

[3] Jabreel M, Moreno A (2017) SiTAKA at SemEval-2017 task 4: sentiment analysis in twitter based on a rich set of features. In: Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017).

[4] Agarwal B, Ramampiaro H, Langseth H, Ruocco M (2018) A deep network model for paraphrase detection in short text messages. Inf Process Manag 54(6):922–937

[5] Allahyari M, Pouriyeh S, Assefi M, Safaei S, Trippe ED, Gutierrez JB, Kochut K (2017) A brief survey of text mining: classification, clustering and extraction techniques.

[6] Asghar MZ, Kundi FM, Ahmad S, Khan A, Khan F (2018) T-SAF: Twitter sentiment analysis framework using a hybrid classification scheme.

[7] Baecchi C, Uricchio T, Bertini M, Del Bimbo A (2016) A multimodal feature learning approach for sentiment analysis of social network multimedia.

[8] Baziotis C, Pelekis N, Doulkeridis C (2017) Datastories at semeval-2017 task 4: deep lstm with attention for message-level and topic-based sentiment analysis. In: Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017), pp.

[9] Chen P, Sun Z, Bing L, Yang W (2017) Recurrent attention network on memory for aspect sentiment analysis. In: Proceedings of the 2017 conference on empirical methods in natural language processing, pp.

[10] Abulaish et al., 2009 M. Abulaish, M.N. Doja, T. Ahmad Feature and opinion mining for customer review summarization Pattern Recognition and Machine Intelligence.

[11] Balahur and Montoyo, 2008 Balahur, A., Montoyo, A., 2008. A feature dependent method for opinion mining and classification. Paper presented at the International Conference on Natural Language Processing and Knowledge Engineering, NLP-KE 08.

[12] Dragoni M, Federici M, Rexha A (2018) An unsupervised aspect extraction strategy for monitoring real-time reviews stream.

[13] Hanafiah N, Kevin A, Sutanto C, Arifin Y, Hartanto J (2017) Text normalization algorithm on Twitter in complaint category.

[14] Vechtomova O (2017) Disambiguating context-dependent polarity of words: An information retrieval approach.

[15] Yang Z, Hu Z, Salakhutdinov R, Berg-Kirkpatrick T (2017) Improved variational autoencoders for text modeling using dilated convolutions. In: Proceedings of the 34th international conference on machine learning.