Journal of Applied Computer Science & Artificial Intelligence

Research Paper Vol.-2, Issue- 1, June 2022 ISSN: 2582-8673

TWACH: THE BIAS PLATFORM
Twitter Sentiment Analysis

Anjan Kumar Dey, Ashish Abhishek Minz, Ayan Chatterjee, Debajyoti Ghosh, Debraj Karmakar, Dr. Dharampal Singh

[1] JIS College of Engineering, Kalyani, India, 741235, debrajkarmakar010@gmail.com

Available online at: http://jacsai.org/

Received: ………./2021, Revised: ………..2021, Accepted: ………2021, Published: 30/June/2021

***Abstract*—** This project aims to solve the problem of sentiment analysis on Twitter. Classify tweets according to the sentiment they express: positive, negative, or neutral. We may collect data and analyze tweets by providing a username or any hashtag. The user enters the number of tweets to extract and the extraction date with the hashtag or username. If the input is not empty, using the Twitter API we pull tweets from Twitter in real-time. The NLP technique is then used to clean them. The tweets are analyzed and classified according to their polarity and subjectivity scores in positive, negative, or neutral lists and finally displayed. Analyzing public sentiment or opinion about something is important for many applications such as companies trying to find out the public sentiment of their products in the market, predicting political elections and socio-economic phenomena such as the stock market. The goal of this project is to develop a functional classifier and automatic sentiment classification of recent tweets.

Keywords— Sentiment, Tweets, NLP, Hashtag, Username, Polarity, Subjectivity

I. INTRODUCTION

Data analysis is the process of applying organized and systematic statistical techniques to data to describe, summarize, control and condense it. It applies these techniques to modify the data according to our needs. Data mining is a multi-step process that involves collecting, organizing, and cleaning data. It is necessary because different sources like social media, company data, public data, transactions, etc**.** generate more and more data every day and it is important to manage, analyze and organize this big data. Social media has become a turn of the 21st century, something we live by. It has changed the way we and businesses look at advertising, marketing, globalization or politics. It is estimated that the data we generate will increase faster than ever and in 2020, approximately 1.7 megabytes of additional data were generated by each person.5 Bytes of data quintilies are produced by humans every day.

II. MOTIVATION

The database we chosen for the proposed study is Twitter.It is a very popular microblogging service for very popular microblogging items and conventional Internet items and web drop we believe that Twitter is aA study was conducted by Tumasjan and Al.in in Germany to predict the results of the federal elections, which concluded that Twitter is a good reflection of offline feeling**.**

III. SCOPE

This project hopes to be useful to businesses, politicians and ordinary people**.** It will be useful to examine public opinion on their programs. Businesses can get reviews of their new product or the latest hardware or software. Additionally, filmmakers and viewers can see reviews about the movie while streaming. By analyzing the tweets, the analyzer can get results on how positive, negative or neutral people are about it.

IV. OVERVIEW OF THE PROJECT

This project is a web application which is used to analyze tweets on a certain topic using a hashtag or a certain person using their username. We'll build a classifier to perform sentiment analysis on these tweets to determine if they're positive, negative, or neutral. This application can be used by any office in the organization to review their work or by any other business to review their products or brands. The main feature of our app is that it helps to determine public opinion on products, government work, politics or any other by analyzing tweets.

V. DOMAIN INTRODUCTION

This tweet feelings analysis project **is** under the field of "data minification" and "classification of models". The vocal marking pieces are a syntactic approach, automatically identifies the part of the speech to each phrase belongs to: name, pronoun, adjective, verb, adverb, etc. Social networking sites or microblogging sites are great sources of information because people freely share and discuss their views on various topics and can be used as data sources in the sentiment analysis process. Sentiment classification techniques can be divided into machine learning approach, vocabulary-based approach, and hybrid approach. The Machine Learning (ML) approach applies famous ML algorithms and uses language features.

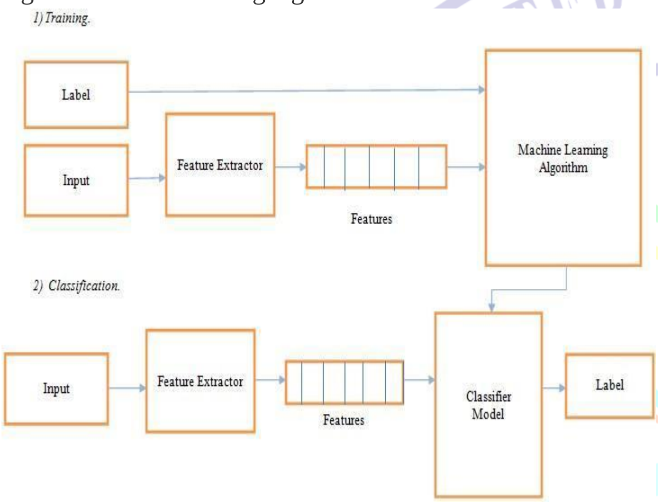


Figure1: Machine Learning Approach

The Lexicon Based approach is based on a sense of lexicon, a collection of known and precompiled feeling terms.

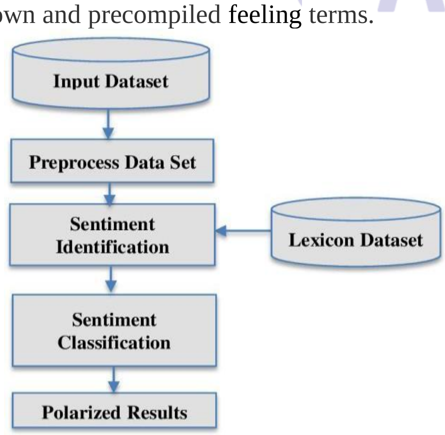
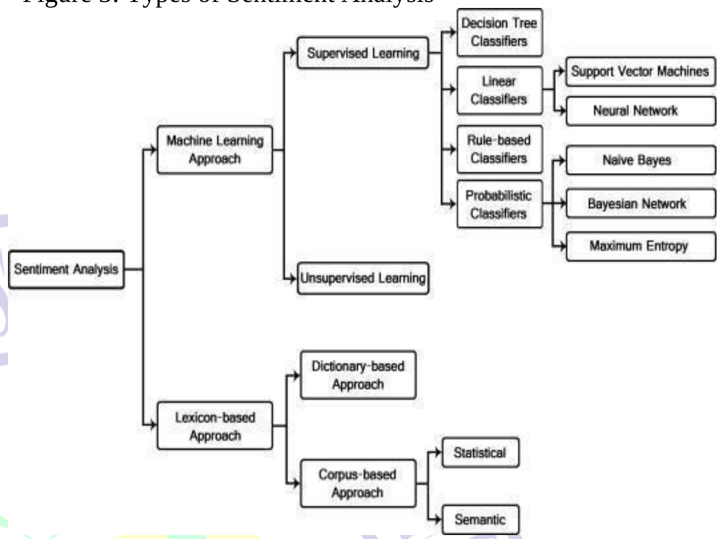


Figure 2: Lexicon-based Approach

The hybrid approach combines the two approaches and is very common, with sentiment lexicons playing a key role in most methods. The different approaches and the most popular algorithms:

Figure 3: Types of Sentiment Analysis



The text classification methods using the ML approach may be approximately divided into supervised and unopened learning methods.

VI. APPLICATION

The application of the feeling analysis varies from commercial activities and marketing, health, policy for the analysis of public actions.

1. **Applications that use website reviews:** The internet now has a huge collection of reviews and opinions on almost everything. This includes product issue feedback, service feedback, product reviews, etc. So, there is a need for a sentiment analysis system that can extract sentiments about certain services or products that can help us automate the analysis of ratings or reviews for a certain article, program, product, etc. This would meet the needs of providers and users. This allows us to keep track of what is being said about a service or product on social media and can help us catch angry and dissatisfied customers or negative mentions before they turn into a major crisis or n negatively impact an organization's brand image.

2. **Business Intelligence Applications:** People now tend to search for product reviews online before buying them. Thus, for many companies, online opinion becomes a decisive factor in the success or failure of their product or service. Sentiment analysis thus plays an important role in business. It can help businesses and organizations take the right steps to improve their products or services and their

business strategy. Sentiment analysis creates benefits for business owners by identifying their popularity with customers and how customers feel about their product or service and evaluating their share price through social media and rating the effectiveness and capability of corporate brand communication.

3. **Applications such as subcomponent technology:** A sentiment prediction system can be useful in recommender systems. These systems do not recommend articles that receive a lot of negative comments or low ratings. Communication on the Internet can be dangerous. We may encounter abusive language and other negative elements. These can be detected by identifying a very negative feeling and acting on it accordingly.

4. **Applications in all fields:** Recent research in sociology and other fields such as medicine and sports has also benefited from sentiment analysis showing trends in human emotions, especially on social media.

5. **Policy Applications:** Predict the results of elections and popular surveys is also an emerging application of feeling analysis.

Applications in the healthcare sector: We can see the application of sentiment analysis in the healthcare sector and where the study uses sentiment analysis as a service framework is offered and uses spatio -temporal to identify outbreak locations. Sentiment analysis can be used to determine a person's level of depression by observing and analyzing textual emotions in their social media. Additionally, sentiment analysis can identify the sentiment needs of people during a disaster and prepare an appropriate relief response.

VII. LITERATURE REVIEW

The feeling analysis has been managed as a task of transforming the natural language to many levels of granularity.(2009), Pak and Paroubek (2010) and (Bermingham and Smeaton, 2010). Go et al. (2009) used distance learning to acquire sentiment data. They build models using Naive Bayes, MaxEnt and Support Vector Machines (SVM) and report that SVM outperforms other classifiers.

In terms of feature space, they are trying a Unigram, Bigram model in combination with POS (parts of speech) features. Point out that the unigram model outperforms all other models and the bigrams and POS features don't help. (Das and Chen, 2001) reported that Naive Bayes and SVM classifiers achieve higher accuracy without applying negation rules. (Joachims, 1998) reported that both of these classifiers achieved accuracy better than 0.9 using unigram features in traditional argument-based classification. used by YungMing Li and TsungYing Li (2013) as a feeling polarity classifier.Unlike the binary classification problem, they argued that the subjectivity of the opinion and the credibility of the expressor should also be taken into account for the accuracy of the model. They proposed a framework that provides a compact numerical synthesis of views on microblogging platforms. Hanhoon Kang, Seong Joon Yoo, and Dongil Han (2012) proposed an improved NB classifier to correct the tendency of positive classification accuracy to appear up to about 10% higher than negative classification accuracy. This decreases the average precision when the precisions of the two classes are expressed as an average value. They reported that using this algorithm with restaurant reviews narrowed the gap between positive and negative accuracy compared to NB and SVM. Accuracy is improved in recall and precision compared to both NB and SVM. After analyzing these articles, it is clear that improvements in sentiment classification algorithms are still an open field of research.

Naïve Bayes and Support Vector Machines are the most widely used ML algorithms for solving sentiment classification problems. They are often considered a reference model with which to compare many proposed algorithms.SVM is known to be the model that gives the best results, but in this project, we only focus on the probabilistic model which is a naive berry that has been widely used in this area.

1. Naïve Bayes Classifier (NB)

It is a probabilistic classifier that uses mixing models for classification.It uses Bayes' theorem to predict the probability that a given set of features belongs to a particular label.

$$P(label \mid features) = \frac{P(label) * P(features \mid label)}{P(features)}$$

P (label) is the previous probability of a label or the likelihood that a random function defines the label Leban.P (feature | Label) is the previous probability that a certain set of functions is classified as label.

$$P(label \mid features) = \frac{P(label) * P(f1 \mid label) * \ldots\ldots * P(fn \mid label)}{P(features)}$$

Rather than p (characteristics) explicitly, our algorithm calculates the numerator for each label and standardize them so that they have a sum to a:

$$P(label|features) = \frac{P(label) * P(f1|label) * ... * P(fn|label)}{SUM[l](P(l) * P(f1|l) * ... * P(fn|l))}$$

Although the assumption we made above is not always contained in the true word, the Naive Bayes model works surprisingly well. (Domingos and Pazzani, 1997) show that even with some dependent characteristics, Naive Bayes is optimal for some problems. Naïve Bayes is one of the quick and easy ML algorithms to predict a class of dataset. It can be used for binary and multiclass classifications. This is the most popular choice for text classification problems

Advantages of Naïve Bayes Classifier:

- It is easy and simple to implement
- It performs well even without much training data
- It handles both continuous and discrete data
- It is highly scalable with the number of predictors and data points
- It is fast and can be used to make real-time predictions
- It is not sensitive to irrelevant features

2. TextBlob

Textblob is a python library that provides text mining, text analysis and text processing modules for python developers.It reuses NLTK corpora.Textblob is a sentence level analysis tool.First, it takes a dataset as the input then it splits the review into sentences.A common way of determining polarity for an entire dataset is to count the number of positive and negative sentences/reviews and decide whether the response is positive and negative based on total number of positive and negative reviews.Polarity and subjectivity of a given review can be obtained using the sentiment module.It returns a named tuple with two parameters called polarity and subjectivity.The default sentiment analyzer algorithm uses the implementation PatternAnalyzer which is based on a library called pattern.The sentiment analysis lexicon bundled in Pattern focuses on adjectives.It contains adjectives that occur frequently in customer reviews, handtagged with values for polarity and subjectivity.The polarity score goes up from 1 to 1 and the subjectivity ranges are 0 to 1 where 0 is most objectives and 1 is the most subjective.

VIII. SYSTEM REQUIREMENTS STUDY

- USER CHARACTERISTICS

Sentiment evaluation is a system that automates mining of attitudes, views, evaluations and feelings from textual content, tweets, speech and database reassets thru Natural Language Processing (NLP). It entails classifying evaluations in textual content into classes like "positive" or "negative" or "neutral". Also referred to as subjectivity evaluation, appraisal extraction or opinion mining.

The phrases sentiment, view, opinion and perception are used interchangeably however they're extremely different.

Opinion:

A end open to dispute (due to the fact distinct professionals have distinct evaluations)

View: subjective opinion

Belief: planned recognition and highbrow assent

Sentiment: opinion representing one’s feelings

Sentiment Analysis consists of many responsibilities consisting of sentiment extraction, subjectivity classification, sentiment classification, opinion unsolicited mail detection, summarization of evaluations and lots of more. It ambitions to research people’s sentiments, evaluations, attitudes, emotions, etc. toward factors consisting of individuals, organizations, products, subjects and services.

The main components of this application are:

- **UI:** The graphical interface of this net software has been evolved the usage of html, CSS, and streamlit that is an open-supply Python library.

  It is straightforward and includes a textual content enter area for a hashtag to be searched, sliders at the sidebar to enter different fields and buttons: one to examine the hashtag and the alternative to show the uncooked twitter information that has been received from information mining via the twitter API.

  .

- **Backend:** The backend of this internet utility has been advanced the use of Python3 and its libraries.

In this project some libraries used:

- **Tweepy,** A Python library for getting access to the Twitter API for scraping information or information mining.

- **Natural Language Toolkit (NLTK),** A NLP library in Python, to smooth the tweets.
- **TextBlob,** A Python (2 and 3) library for processing textual data. It affords a easy API for diving into not unusualplace herbal language processing (NLP) tasks. It is used for sentiment analysis.
- **Scikit-learn (Sklearn)** Is the maximum beneficial and sturdy library for system getting to know in Python.It gives a choice of green gear for system getting to know and statistical modeling which include numerous classification, regression and clustering algorithms.
- **Jupyter Notebook,** An interactive computational environment, wherein you could integrate code execution, wealthy text, mathematics, plots and wealthy media. It is used for quicker implementation of the experiments.

  The net software is hosted the usage of heroku, a cloud platform.

HARDWARE REQUIREMENTS:

- x86 64-bit CPU (Intel / AMD architecture)
- four GB RAM
- five GB loose disk space

SOFTWARE REQUIREMENTS:

Windows 7 or higher

A net browser ideally Chrome, Firefox

Python-3

Twitter developer account

IX. SYSTEM ANALYSIS

1) Feasibility study

- **Technical Feasibility:**
  Our system is technically feasible because all the required tools are readily available.
- **Operational Feasibility:**
  In this project, we have done a simplified web application to analyze tweets.
- **Economic Feasibility:**
  It is a web application hosted on a free cloud services platform. heroku Creating the application is not expensive.
- **Schedule Feasibility:**
  This application development is doable in terms of time.

2) Requirement Definition

After a thorough analysis of the system problems, we familiarized ourselves with the requirements required by the current system. The requirement that the system needs is categorized into functional and non-functional requirements. These requirements are listed below:

1) Functional Requirements:

Functional requirements are product features or functions that developers must implement to enable users to perform their tasks. The functional requirements that this system must require are as follows:

- The system should be able to set Twitter authentication keys.
- The system should be able to authorize tweepy and access the Twitter API.
- The system must be able to analyze the data and classify each polarity of the tweets.

2) Non-Functional Requirements:

Non-functional requirements define system attributes such as reliability, maintainability, security, performance, scalability, and user-friendliness.

- Easy to use
- The system must provide better accuracy
- To operate with efficient productivity and response time
- Easy to maintain and reliable

X. SYSTEM DESIGN

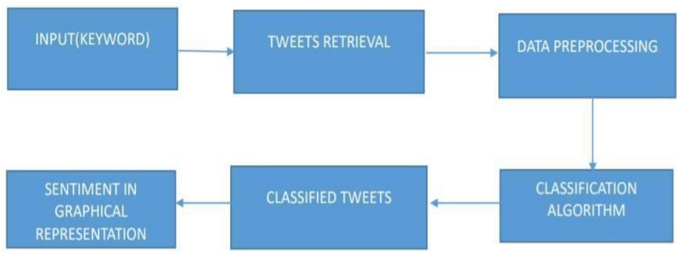


Figure 4: System Design

XI. METHODOLOGY

For the development a dataset containing labelled tweets is extracted from Kaggle.Preprocessing takes place on the data obtained.Then by using Naive Bayes algorithm a classifier will be trained using training data.Several methodologies will be used to extract features from the source text.After features are added to the feature vector, each tweet in the training data is associated with the class label and passed to the classifier.Test tweets will be given to the model and classification will be performed with the help of these trained classifiers.We get the tweets classified into the positive and negative.We then extend the classification to include the neutral class.The model can now be saved and loaded when a prediction is to be made.It is ready to be implemented in our web application.

I. Loading sentiment data

The dataset for this project is pulled from Kaggle. This dataset contains over one million tweets extracted using the Twitter API. Tweets have been annotated (0=negative, 4=positive) and can be used to detect sentiment.

It contains the following 6 fields:

1. target: the polarity of the tweet
2. ids: The id of the tweet
3. date: the date of the tweet
4. flag: The query. If there is no query, then this value is NO_QUERY.
5. user: the user that tweeted
6. text: the text of the tweet

The files contain positively and negatively tagged tweets. First, the dataset is loaded. We use a fraction of this to train our model for better time efficiency. 200,000 tweets have been downloaded. We changed the positive polarity value to: 4 → 1. Next, we check the number of tweets with positive and negative tags.

No. of positive tagged sentences: 99956

No. of negative tagged sentences: 100044

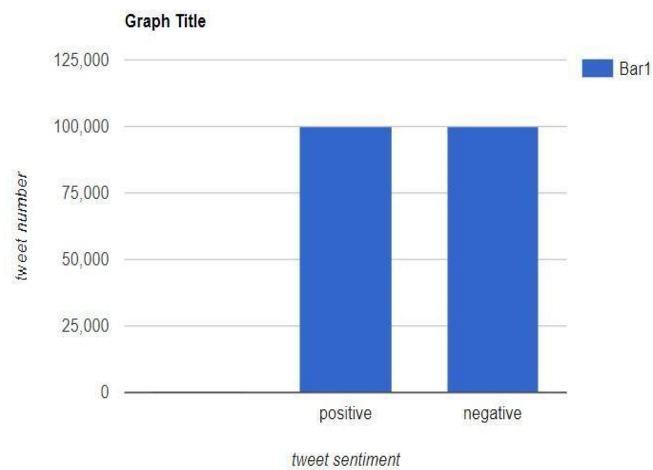


Figure 6: Dataset bias

We see that the dataset is balanced. The base probability is therefore 50%.

We eliminate unnecessary columns leaving only polarity and text in the data.

II. Pre-processing Data

Once the data is loaded, pre-processing is performed. The pre-processing steps performed are as follows:

- Lowercase: Each text is converted to lowercase.
- URL Removal: Links beginning with "http" or "https" or "www" are replaced with "".
- Removing Usernames: Replace @Usernames with the word "".
- Deletion of short words: Words with a length less than 2 are deleted.
- Noise removal: Noise words are English words that don't add much meaning to a sentence. They can be safely ignored without sacrificing the meaning of the sentence. (Ex: "the", "she", "have")
- Lemmatization: Lemmatization is the process of converting a word into its base form.(For example:

ISSN: 2582-8673

Running, running, running are all forms of the word travel, so it's the lemma of all these words).

III. Training Naïve Bayes Classifier

1) Convert text to word frequency vectors:
TFIDF: This is an acronym that stands for Term Frequency – Inverse Document Frequency which are the components of the resulting scores assigned to each word.

   a. Term Frequency: This summarizes how often a given word appears within a document.
   b. Inverse Document Frequency: This downscales words that appear a lot across documents.

2) Split Train and Test -
The preprocessed data is split into 2 datasets:

   a. Training Data - The dataset on which the model will be trained. Contains 80 tons.

   b. Test Data: Dataset against which the model would be tested. Contains 20 tons.

3) Model Building:
We build the model of Bernoulli Naive Bayes of the Sklearn Library that fueled the training data.

IV. Implementation of evaluation metric

Finally, to test the accuracy of the model, the F1 score is evaluated. In the statistical analysis of binary classification, the F1 score or Fmeasure is a measure of the accuracy of a test. The confusion matrix is used for data evaluation. A confusion matrix is a technique for summarizing the performance of a classification algorithm. Computing a confusion matrix can give us a better idea of what our classification model is doing well and what kinds of errors it is making. The number of correct and incorrect predictions is summarized with count values and decompositions for each class.

There are 4 important terms in the confusion metric:

1. True positives: the cases where we predicted YES and the actual output was also YES.
2. True Negatives: Cases where we predicted NO and the actual output was NO. False Positives: Cases where we predicted YES and the actual output was NO.
3. False negatives: the cases where we predicted NO and the actual output was YES.

To evaluate our model, we found:
model accuracy on training data: 86.868749999
Test data model accuracy: 75.69749999
Score F1: 0.76

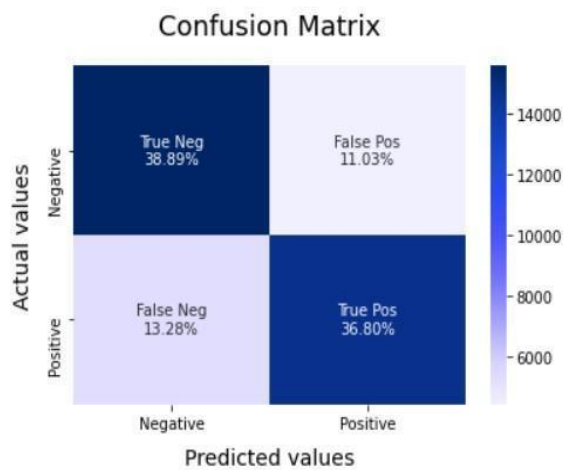


Figure 7: ***Confusion Matrix of Model***

Now let's save this model and load it during a prediction.

V. Handle the neutral class

In previous sections, neutral sentiment has been ignored. The training and test data only contained text with positive and negative sentiments. In this section, we explore what happens when neutral sentiment is introduced.

Naive Bayes with Three Classes

Our model returns its results as a named tuple of the form: Sentiment (classification, p_pos, p_neg) eg: Sentiment (classification = 'pos', p_pos = 0.5057908299783777, p_neg = 0.49420917002162196)

We have extended the Naive Bayes classifier to manage 3 classes: positive, neutral and negative. Taking the difference of p_pos and p_neg as the polarity score (polarity = p_pos p_neg), we have classified the three classes as follows: Positive: if the polarity score is greater than or equal to 0.

Negative: if the polarity score is lower or equal to 0. Neutral: if the polarity score is between 0.2 and 0. Although rudimentary, it works quite well for this type of classification.

Finally, after classifying the tweets, you can view the count using graphs. We also visualize the polarity and subjectivity achieved through the text blob library in a scatter plot.

XII. IMPLEMENTATION

In this document, the Tweepy Python library was used to extract data from .more, Tweepy Application Programming Interface) allows proper data retrieval via keywords.

Implementation In Web Application:

For example: give the entry of "Uber" into the Hashtag option, we can analyze the feelings (positive, negative or neutral), the polarity and the subjectivity of tweets labeled with #Uber as shown below.

SNAPSHOTS:

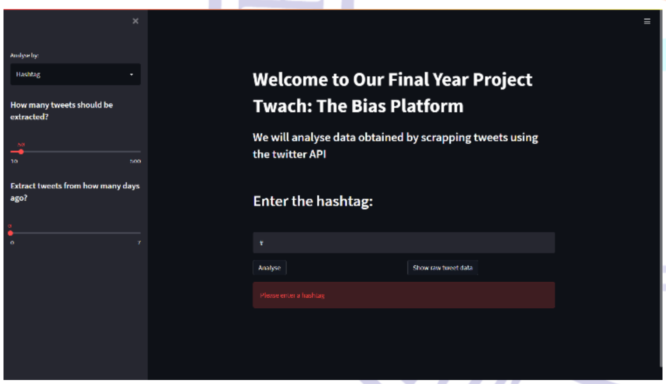


Figure 9: ***Snapshot of Web Application #1***

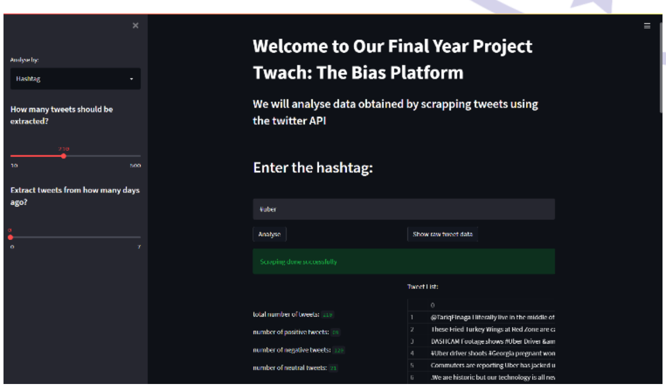


Figure 10: ***Snapshot of Web Application #2***

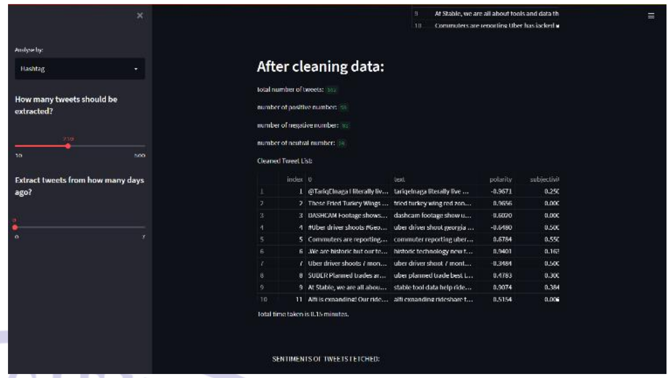


Figure 11: ***Snapshot of Web Application #3***

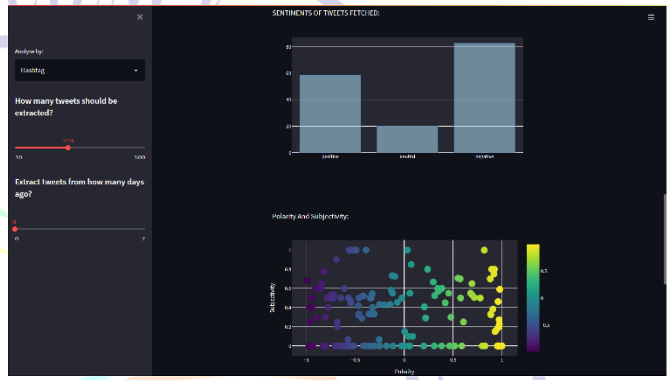


Figure 12 : ***Snapshot of Web Application #4***

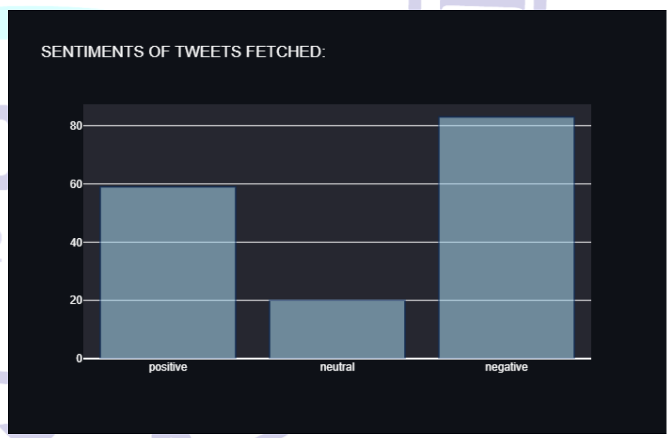


Figure 13 : **Graphical Representation of Tweet Sentiments by Bar Graph**

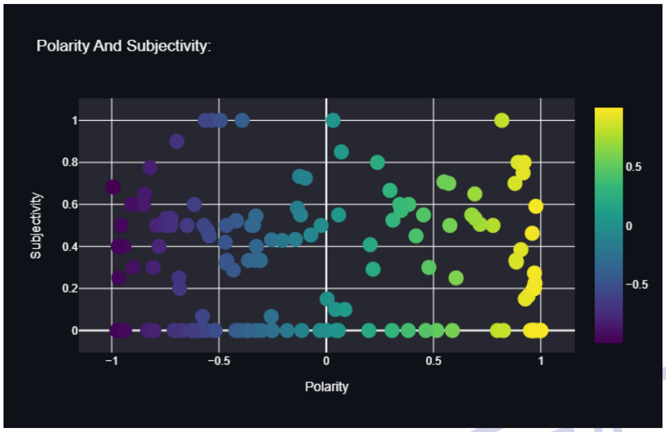


Figure 14: ***Graphical Representation of Tweet Sentiments by Scatter Plot***

XIII. LIMITATIONS

The analysis of feeling is a very difficult task.Example: The word "unpredictable" is positive in the realm of movies, dramas, etc., but if the same word is used in the context of driving a vehicle, then it has a negative opinion. Sarcasm detection: Sarcasm phrases express negative opinions about a target who uses positive words in a unique way. Example: "Wow, that's great!" ; Very well ! The sentence contains only positive words but actually expresses a negative feeling.

1. **Contrasting Expressions:** In some sentences, only part of the text determines the overall polarity of the document.

Example: “This movie should have been fantastic. fantastic plot, popular actors and supporting actors are equally talented”.In this case, a simple bag of words approach will define it as a positive feeling, but the ultimate feeling is negative.

2. **Explicit denial of the feeling:** the feeling can be denied in many ways instead of using the simple no, no, never, etc. It is very difficult to identify such negatives. Example:

"Avoid all the clichés and predictability found in Hollywood movies."

Here, the words cliche and predictable convey a negative feeling, but the use of the word "avoid" negates the respective feelings.

3. **Order dependence:** Analysis of speech structure is essential for sentiment analysis and opinion extraction. Example:

Dan is older than John, expresses the exact opposite opinion that John is older than Dan. Building a classifier for subjective vs. objective tweets: Our current research mainly focuses on the correct classification of positive, negative, and neutral tweets. You need to consider ranking tweets with sentiment vs.no feeling up close.

4. **Handling comparisons:**

The bag-of-words model doesn't handle comparisons very well.

Example: "IITs are better than most private colleges", the tweet would be considered good for both IITs and private colleges that use the pattern word bundle because it doesn't take into account the relationship with the " better".

5. **Applying Sentiment Analysis to Facebook Posts:** There hasn't been much work done on sentiment analysis on Facebook data, mainly due to various Facebook Graph API restrictions and security policies data access.

6. **Internationalization:** Current research mainly focuses on English content, but Twitter has many different users from all over the world with different languages

7. **Entity recognition:** It is necessary to separate the text on a specific entity and then analyze sentiment in his comparisons. Example:

"I hate Microsoft, but I love Linux". A simple bag-of-words approach will qualify this as neutral, however, a specific sentiment for both entities is present in the statement.

XIV. CONCLUSION AND FUTURE SCOPE

Therefore, via way of means of coming into a keyword, we get the visible illustration of the tweets analyzed (positive, negative, neutral) in addition to the polarity and subjectivity thru the internet app.

The carried out systematic literature evaluate offers facts on research on sentiment evaluation in social media. There are numerous techniques added via way of means of researchers. Choosing the maximum green approach of sentiment evaluation relies upon at the facts itself. Both techniques validated a comparable accuracy. The matters that we want to consider are the shape of the text, time and quantity of facts. If the facts shape is messy, there's confined time to be had to examine and a small quantity of facts, lexicon-primarily based totally techniques are extra appropriate. Whereas, device getting to know primarily based totally techniques are extra appropriate for larger facts because it calls for extra facts and time to train. In order to enhance the accuracy and first-rate of the result, combining each lexicon and device getting to know techniques could be recommended.

We recognized the maximum not unusualplace kind of social media web website online to extract facts from for sentiment evaluation, that is Twitter. Most of the reviewed papers use twitter as their social media context. This is because of the accessibility, availability and richness of Twitter content. There are tens of thousands and thousands of tweets each day on nearly any subject matter. This suggests that social media is turning into a treasured supply of facts.
We show the software of sentiment evaluation in social media thru making an internet app. Sentiment evaluation has a extensive software and may be applied in exceptional regions along with political forecasting an election result, enhancing method and first-rate of merchandise in business, tracking ailment outbreak, notion in the direction of a selected subject matter or individual and enhancing reaction to a disaster. This suggests that sentiment evaluation performs a big position in expertise people's notion and allows in choice making.

Sentiment Analysis, specifically withinside the area of micro-running a blog remains withinside the growing degree and some distance from being complete. We endorse a few thoughts which we experience are really well worth exploring withinside the destiny and might bring about advanced performance.

1. We can also additionally discover extra dependable reassets of records on different social media systems like Facebook, Instagram, LinkedIn, Youtube, etc. It is vital to discover different social media systems to carry out and examine sentiment evaluation outcomes.

2. In our contribution, we selected to apply the Naïve Bayes version. But there are different fashions that can offer thrilling outcomes along with lexicon-primarily based totally algorithms. In this task we attempted to reveal the primary manner of classifying tweets into positive, poor or impartial classes the use of our changed Naive Bayes. We ought to in addition enhance our classifier with the aid of using seeking to extract extra capabilities from the tweets, tuning the parameters of the Naïve Bayes classifier, attempting exclusive styles of capabilities, or attempting any other classifier all together.

3. We are specializing in popular sentiment evaluation. There is capacity for studies withinside the area of sentiment evaluation with partly recognised context. For example, customers usually use our internet site to go looking with precise varieties of key phrases which may be divided into awesome classes: celebrities,politics/politicians, sports/sportsmen, brands/products, movies/media and music. So we will try to carry out separate sentiment evaluation on tweets that simplest belong to this sort of classes (i.e. the schooling records could now no longer be popular however precise to this sort of classes) and examine the outcomes with those who we'd get if we follow popular sentiment evaluation on it instead.

4. One extra characteristic that we sense is really well worth exploring is whether or not the records approximately relative function of phrases in a tweet has any impact at the overall performance of the classifier. Although Pang et al. explored a comparable characteristic and suggested poor outcomes, they labored on an exceptionally easy version and their outcomes had been primarily based totally on critiques which can be very exclusive from tweets.

XV. ACKNOWLEDGEMENT

We would really like to explicit our honest gratitude to Dr. Dharampal Singh of the branch of Computer Science and Engineering, whose function as a challenge manual changed into priceless for the challenge. We are extraordinarily grateful for the eager hobby he took in advising us, for the books and reference substances supplied and for the ethical guide prolonged to us.

Last however now no longer the least we bring our gratitude to all the academics for supplying us the technical talent to be able to continually stay as an asset at some point of our lives and to all non-coaching group of workers for the gracious hospitality they supplied us.

Place: JIS College of Engineering, Kalyani
Date: 23rd February, 2022
Anjan Kumar Dey
Ashish Abhishek Minz
Ayan Chatterjee
Debajyoti Ghosh
Debraj Karmakar
department of Computer Science and Engineering,
JIS College of Engineering, Kalyani,
West Bengal, India.

REFERENCES:

- "DocumentationTweepy"- https://docs.tweepy.org/en/latest/api.html"documentation textblob".https://textblob.readthedocs.io/en/dev/ "documentation simplifiée" https: //docs.streamlit.io/fr/stable/com/nlpinpythondatacleaning 6313a404a470 https://python.
- plainenglish.io/scrapingtweetswithtweepypython5941 3046e788

- Alexander Pak and Patrick Paroubek.Twitter as a Corpus for Sentiment Analysis and Opinion Mining.In Proceedings of international conference on Language Resources and Evaluation (LREC), 2010.
- Pang and L.Lee."A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts" in Proceedings of ACL, 2004.
- Peter D.Turney.Thumbs Up or Thumbs Down?Semantic Orientation Applied to Unsupervised Classification of Reviews.In Proceedings of the Annual Meeting of the Association of Computational Linguistics (ACL), 2002.
- Hu, M., and Liu, B.(2004)."Mining and Summarizing Customer Reviews." Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD '04 (2004).
- Kang, Hanho, Seong Joon Yoo et dongil Han.Twitter Feeling Classification Using Final Supervision.Project Technical Report, Stanford University, 2009.
- R Kusumawati et al 2019 J.Phys.: Conf.ser.1320 012016 Conf.ser.1320 012016 Performance Performance Performancecom/280203
- Rahman, S.El, F.AlOtaibi, and W.AlShehri.(2019, 34 April 2019).“Sentiment Analysis of Twitter Data”, in the 2019 International Conference on Computer and Information Sciences (ICCIS).
- Joyce, Brandon, and Jing Deng.(2017) “Sentiment Analysis of Tweets for the 2016 US Presidential Election”, in IEEE MIT Undergraduate Research Technology Conference (URTC), Cambridge, MA, USA: IEEE.
- Poecze, Flora, Claus Ebster, and Christine Strauss.(2018) "Social Media Metrics and Sentiment Analysis to Evaluate the Effectiveness of Social Media Posts". Procedia Computer Science 130:660666.
- Suman, N., P. Gupta and P.Sharma. (2017, 1112 December 2017). "Analysis of Stock Price Flow Based on Social Media Sentiments", at the 2017 International Conference on Next Generazione Informatica and Information Systems (ICNGCIS). Ali, Kashif, Hai Dong, Athman Bouguettaya, Abdelkarim Erradi and Rachid Hadjidj.(2017) "Analysis of the Sentiment of Service: A Framework for Analytical Analysis Based on Social Media", in the IEEE International Conference on Web Services (ICWS), Honolulu, HI, USA: IEEE.HASSAN,