Sentiments Analysis on SARS-Cov-2 Lockdown Strategy Using Machine Learning Techniques

Amol Nimbalkar[1], Ajay Sahane[2], Rushikesh Shitole[3], Ankush Motipwar[4]

[1,2,3,4]*Student, Dept. of Computer Engineering, G H Raisoni College of Engineering and Management, Pune, Maharashtra, India*

[5]*Asst. Professor, Dept. of Computer Engineering, G H Raisoni College of Engineering and Management, Pune, Maharashtra, India*

Abstract - *Twitter, as a social network, is a very common medium to share thoughts and communicate with other people in the online world. Tweets collected in aggregate will represent public opinion about events. This paper gives an optimistic or negative feeling to Twitter posts using a well-known machine learning approach for text categorization. In addition, we use manually labeled (positive/negative) tweets to create a qualified system to perform a task. The task is looking for a correlation between twitter sentiment and events that have occurred. The qualified model is built on the classification system of Naive Bayes and Support Vector Machine(SVM).*

We also used external lexicons to detect arbitrary or objective tweets, added Unigram and Bigram features, and used TF-IDF (Term Frequency-Inverse Document Frequency) to filter out the features. We used the Twitter Streaming API and some of the official hash tags for mine, filter and process tweets to examine the public's view of unusual incidents. The same method can be used as a basis for forecasting future events. In the form of the twitter sentiment analysis, the most basic sentiment analysis quantifies the mood of a tweet or message by counting the number of positive and negative terms.

Key Words: **Twitter Streaming API, Opinion Mining, NLP, Sentiment Analysis, Naive Bayes, SVM, BLR.**

1. INTRODUCTION

Twitter, one of the most prominent online social networking and micro-blogging services, is a very popular way of sharing thoughts and communicating with other people in the online world. Twitter tweets offer true raw data in the medium of brief texts that convey thoughts, concepts and events that have been caught at the moment. Tweets (Twitter posts) are well-suited streaming data outlets for opinion mining and sentiment polarity detection. Opinions, assessments, feelings and speculations also represent the states of individuals; they consist of opinions articulated in a vocabulary made up of abstract terms.

Social media are emerging exponentially on the Internet. This media expertise allows individuals, companies and organisations to interpret facts for important decision-making. Opinion mining is also referred to as sentiment analysis, which entails developing a framework to collect and analyse thoughts on the product made in ratings or tweets, articles, blog posts on the internet. Sentiment is automatically graded for critical applications such as opinion mining and summarization.

Making useful choices in the study of ads where the classification of emotion is implemented effectively. Reviews include feelings that are articulated differently in different domains, and it is expensive to annotate data for each new domain. An study of online user feedback in which companies are unable to find out just what people loved and did not like in text level and sentence level opinion mining. So now opinion mining on-going analysis is in phrase-level opinion mining. It does a fine-grained analysis and discusses the viewpoint explicitly in the online feedback. The suggested framework is built on a phrase level for reviewing consumer feedback. Phrase-level opinion mining is also well-known as opinion-based mining. It is used to extract the most relevant features of the item and to forecast the orientation of each element of the item analysis. The projected framework incorporates element extraction using periodic item mining in consumer product feedback and mining thoughts, whether positive or negative. It identifies the sentimental orientation of each aspect by supervised consumer review learning algorithms.

Data mining technology has effectively shaped a wide range of approaches, techniques and algorithms for managing large volumes of data to solve real-world problems. The main goals of the data mining process are to successfully manipulate large-scale data, mine actionable laws, trends and acquire informative information. The proliferation of social media has provided extraordinary platforms for people to express their views online. Since social media is commonly used for a range of uses, massive content of usercreated data exists and can be made usable for data mining. Recent data mining analysis focuses on mining exploration.

1.1 Literature Review

New modes of networking, such as micro blogging and instant messaging, have arisen and been commonplace over the last decade. Although there is no limit to the variety of information communicated by tweets and emails, these short messages are mainly used to express the thoughts and emotions that people have about what's going on in the world around them.

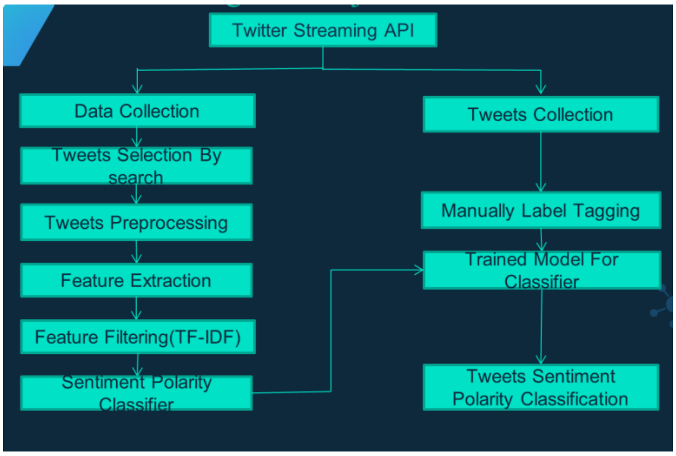
The algorithms and methods used in the current method were complicated and difficult to interpret, and the results were not quite reliable and consistent. The suggested framework consists of approaches and procedures that have reliable and stable effects and are therefore simple to explain. Thanks to the consideration of the proposed method, the difficulty of solving the problems is minimized.

Tweets and texts are short: a phrase or a headline rather than a paragraph. The vocabulary used is very casual, with inventive spelling and punctuation, mispellings, slang, new terms, URLs, and genre-specific terminology and abbreviations, such as RT for "re-tweet" and hash tags, which are a kind of twitter message tag.

Table-1

Paper Title	Part Used
Sentiment Analysis on Twitter using Streaming API	NLTK, Twitter API, Different Phases
Sentiment Analysis of Tweets using Machine Learning Approach	sentiment analysis on social media
Sentiment Analysis on Twitter Data-set using Naive Bayes Algorithm	Framework Implementation of Naïve Bayes Algorithm
Comparative Analysis of Sentiment Orientation Using SVM and Naive Bayes Techniques	Methodology: Dataset, Text Processing, Porter Algorithm.
Comparative Study of Classification Algorithms used in Sentiment Analysis	Classifications: Naïve Bayes, Max Entropy, Boosted Trees, Random Fores

2. DESIGN AND ARCHITECTURE



2.1Tweets Collection:

The first module consists of collecting data for the creation of a training set and then collecting tweets from a real case. Data processing consists of two phases using the Twitter streaming API: The first is the array of data to be used as a training set to create a model. This consisted of a collection of Tweets manually labelled as "Positive" or "Negative." The second move is to gather Tweets during some given occurrence and mark them according to any of the official Hash tags. In addition, the twitter usernames of the celebrities involved associated with the incident are used to retrieve tweets pertaining to the event. The data is as a series of documents in JSON format

2.2Tweets Text Pre-processing:

As a first step in seeking a feeling for tweets, and in order to get a detailed classification of feelings, we wanted to filter out noise and senseless signals that do not lead to a feeling for tweets from the original document.

2.3Feature Extraction:

Selecting a useful list of terms as features of a text and deleting a large number of words that do not add to text sentiment is known as an extraction function.

2.4Feature Filtering:

Term Frequency-Inverse Document Frequency (TFIDF) is a numerical statistical tool for filtering features by weighting and marking of unigram and N-gram using the frequency of terms in the file.

2.5 Dual Sentiment Analysis:

Next, we reinforce the DSA algorithm by introducing a limited data expansion process. Second, we expand the DSA system from the classification of emotion polarity to the classification of positive-negative-neutral sentiment. Third, we suggest a corpus-based approach to create a pseudo-antonym dictionary that could eliminate the reliance of DSAs on an external antonym dictionary.

2.6 Sentiment Classifier:

This process picks 90 per cent for the training set and 10 per cent for the evaluation set, which is repeated in 10 separate parts of the data set.

2.7 Building a Trained model:

Labeling an opinionated text and categorizing it as a positive or negative class overall is called emotion polarity labeling. The neutral label is used for more impartial objects which have no opinion in the text or where there is a combination of positive and negative views in the text. Both subjective tweets, like optimistic or pessimistic feelings, need to be included. There are approaches to retrieve valuable words to detect the emotions in tweets.

3. REQUIREMENT ANALYSIS

This segment outlines the conditions. It defines the hardware and software specifications that are needed to operate the application properly.

3.1 System requirement specification:

Structured compilation of knowledge that reflects the specifications of the method. Company Analysts, also referred to as machine analysts, are responsible for analysing the business demands of their customers and partners to better determine business issues and recommend solutions. In the field of life-cycle production of applications, BA usually plays a coordination role between the corporate side of the company and the IT department or external service providers.

3.2Feasibility:

The technology will be theoretically built and can be used if it always needs to be a reasonable investment for the company. In terms of economic viability, the risk of designing the technology is measured against the ultimate value gained from the proposed technologies. Financial gains must be equal to or higher than losses. The device is commercially viable. No additional hardware or software is needed. As the interface for this device is designed using the current tools and technology available at NIC, there are negligible costs and economic viability for some of them.

3.3Functional Requirements:

Technical Requirement determines the role of the software system and how the system must operate when it is faced with particular inputs or conditions. This may involve measurements, handling and analysis of data and other unique functions.

3.4Non-Functional Requirements:

Non-functional specifications, as the name implies, are certain requirements which are not strictly related to the basic functionality of the device. They can refer to emerging device resources, such as reliability of response time and storage occupancy. Alternatively, system restrictions can be established, such as the functionality of input output devices and the data representations used in system interfaces. Many non-functional criteria apply to the system as a whole rather than to specific aspects of the system.

3.5 SoftwareRequirement:

[1]Operating System: Windows
[2]Technology: Python(3.6)
[3]IDE: NLTK(JetBrains PyCharm 2020)

3.6 Hardware Requirement:

[1]Hardware: Pentium

[2]Speed: 1.1 GHz

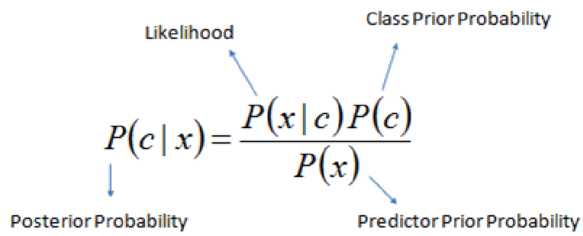
[3]RAM: 1 GB

[4]Hard Disk: 20GB

4. ALGORITHMS OF PROPOSED SYSTEM

4.1 Introduction to Naive Bayes Algorithm:

Naive Bayes is essentially a probabilistic approach to classifying tweets on the basis of their polarity in terms of either a positive opinion or a negative opinion.



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Figure 4.1: Naive Bayes Classification

The Multinomial Case Model is used to count term occurrences. However, tweets are such short texts that seldom include several occurrences of the same expression, so Bernoulli was adequate for sentiment analysis. The downside will be that conditional independence sometimes does not keep the text in practice, but this model has done reasonably well for this principle.

In this part, we present the implementation of our Hadoop architecture for the efficient execution of the Naive Bayes algorithm. We need a qualified SentiWordNet dictionary that is accessible online to apply the Naive Bayes algorithm. It consists of a series of various words with their synonym and their polarity. The synonym is the same term definition it would have the same polarity. Polarity is the positivity of the term in the context of the sentence.

4.2 Introduction to SVM Algorithm:

We can see that SVM is one of the most reliable classifiers. As far as sentiment analysis is concerned, a multiclass SVM with one vs. one scheme is considered, and provides a binary classifier for each particular class pair, and then uses an opinion mechanism to eventually select one class. One downside of SVM is that if the scale of the vector function is greater than the number of training measurements, it appears to be over-adjusted to the training data, which reduces the precision of the test data. Another downside to SVM arises with a neutral view, which may pose a problem to our project due to the limited size of our dataset.

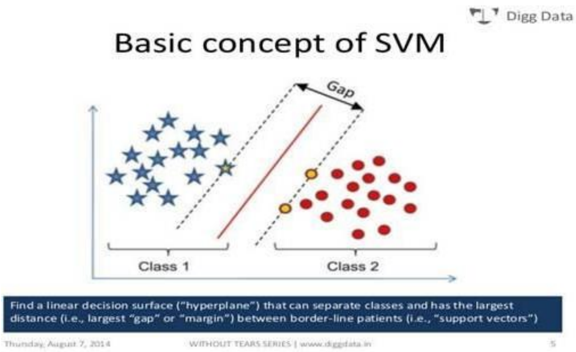


Figure 4.2: Basic Concept of SVM.

4.3 Introduction to BLR Algorithm:

Bayesian Logistic regression brings a positive or negative feeling to Twitter posts as it is a well-known machine learning tool for text categorization. In addition, we use manually labelled (positive/negative) tweets to create a qualified system to execute a task. The task is to find a connection between the twitter emotions and the events that have happened. The trained model is built on the Bayesian Logistics Regression (BLR) classification system.

In mathematics, the logistic model is used to model the likelihood of an actual class or occurrence such as pass/fail, win/loss, live/dead or healthy/sick. This can be generalised to model other types of incidents, such as if the picture includes a cat, dog, lion, etc.

A probability of 0 and 1 would be allocated to each target being observed in the image and the number would be added to one.

5. SMILEY DETECTION

Stuff that we are applying to the suggested technique: now, in most tweets, people are using smileys to express their thoughts. So, in order to do an emotion analysis, we should be in a position to characterize these tweets on the basis of an impression polarity consisting of a variety of smileys or emoticons.

Figure 5: Smily Detection.

6. RESULT

This work is of great value to individuals and businesses who are focused on emotion analysis. Sales Promotion, Product Promotion, etc. The main features of this system are the teaching module on the Nave Bayes SVM Classification, the BLR Time Variant Analytics and the Continuous Learning System.

The fact that the research is carried out in real time is the main highlight of this article. Several current programmes store old tweets and conduct sentiment analysis on them, giving results on old data and saving a lot of space. But this device does not store tweets that are cost efficient as no storage space is required. All research is also done on real-time tweets. The customer is then guaranteed that fresh and valid findings will be obtained.

6. CONCLUSION

Sentiment polarity tests for different individuals and incidents and see how positively or negatively people respond or speak about them. The Dual Sentiment Analysis (DSA) model is very good for polarity classification and greatly outperforms most alternative methods of polarity change consideration. Creating flipped feedback to aid in the classification of supervised sentiment.

In addition, we improve the DSA algorithm by introducing a selective data expansion strategy that selects higher-feeling training feedback for data expansion. Experimental findings suggest that using the chosen portion of the data expansion training feedback will provide higher value than using all reviews.

REFERENCES

[1] ”Sentiment Analysis on Social Media” IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining-Carlo Aliprandi, Federico Capeci, Montserrat Cuadros, Tomas.

[2] ”MapReduce programming with apache Hadoop”M. Bhandarkar-International Symposium on Parallel Distributed Processing (IPDPS).

[3] “Comparative Analysis of Sentiment Orientation Using SVM and Naïve Bayes Techniques.” Shweta Rana, Archana Singh.

[4] "Review on Developing Corpora for Sentiment Analysis Using Plutchik’s Wheel of Emotions with Fuzzy Logic. “International Journal of Computer Sciences and Engineering (IJCSE) 2 (2013): 14-18. Chafale, Dhanashri, and Amit Pimpalkar.

[5] "Tokenization and Filtering Process in RapidMiner." International Journal of Applied Information Systems (IJAIS)–ISSN (2014): 2249-0868. Verma, Tanu, and Deepti Gaur Renu.

[6] https://www2.bui.hawhamburg.de/pers/ursula.schulz/astep/porter.pdf

[7] "Sentiment analysis and opinion mining: a survey." International Journal 2.6 (2012). Vinodhini, G., and R.M. Chandrasekaran.

[8] “Sentiment Analysis on Twitter Data-set using Naïve Bayes Algorith” Huma Parveen, Prof. Shikha Pandey.

[9] "Sentiment Analysis on Social Media", 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) Federico Neri, Carlo Aliprandi, Federico Capeci, Montserrat Cuadros, Tomas.

[10] https://en.wikipedia.org/wiki/Sentiment_analysis

[11] https://en.wikipedia.org/wiki/Text_processing

[12] https://en.wikipedia.org/wiki/Naive_Bayes_classifi er

[13] https://en.wikipedia.org/wiki/Support_vector_mac hine

BIOGRAPHIES

Amol Nimbalkar
Student at Dept. of Computer Engineering, G H Raisoni College Of Engineering and Management, Pune, Maharashtra, India.

Ajay Sahane
Student at Dept. of Computer Engineering, G H Raisoni College Of Engineering and Management, Pune, Maharashtra, India

Rushikesh Shitole
Student at Dept. of Computer Engineering, G H Raisoni College Of Engineering and Management, Pune, Maharashtra, India

Ankush Motipwar
Student at Dept. of Computer Engineering, G H Raisoni College Of Engineering and Management, Pune, Maharashtra, India