International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

Twitter Rumour Identification and Verification

Neha Sawant[1], Sameer Thakare[2], Saloni Wandile[3], Ojas Patil[4]

[1,2,3,4]*Final Year Student, Department of Information Technology, MAEER's MIT Pune, Maharashtra, India*

Abstract - *Social media has emerged as one of the main sources of news and information. An important characteristic of social media is rapid emergence and spread of new information. Twitter is one such platform of micro-blogging on social media for sharing thoughts, opinions, news and information in limited words. This information may contain rumours too. False information and rumours create panic in the society, heightens the bad effects on society, government, organizations, businesses and individuals as well. Considering the harmful consequences of fake news, there is a profound need to detect rumours early, verify their veracity and prevent it from spreading. Our hypothesis states that the content having a negative sentiment is more likely to have rumour. In this paper, we present our rumour detection approach to verify our hypothesis. The project is divided into three modules: Classification, Detection and Verification carried out using natural language processing and supervised machine learning. The tweets are classified based on sentiment analysis. The rumours are detected using the sentiment score and some other features. The rumour's veracity is checked using an external module for obtaining verified news.*

Key Words: Machine Learning, Rumour Detection, Sentiment analysis, Feature Selection, Pattern Matching, Data Set, Fake news, Keyword Extraction.

INTRODUCTION

With the proliferation of Social media, it is easy to access information for all users. Micro-blogging sites such as Twitter are a rich source of news. According to a research by Pew Research Center, 57% people say they expect the news they see on social media to be largely inaccurate. Fake news related to any government, political or business organization, celebrity or natural phenomena may lead to chaos in people. Under the broad and rapid deployment of information and the absence of strategies to ensure the reliability of such information, the need of this project emerged. A rumour is an unverified and instrumentally relevant information statement in circulation. Rumors are ubiquitous and with vast public involvement, they have the capability to impose real damage to individuals, organizations, and the government. Viral rumors about individuals that condemn them for their actions may lead to hate campaigns and eventually harm their reputation.

This may affect individual's self-esteem and confidence level. Rumors accelerate the dynamic nature of share markets and consequently elaborate their effect on organizations. Sometimes misinformation about the outburst of a disease might affect the tourism of a country and likewise other government sectors. Analysis of rumors led to its aspect of public participation through various perspectives, for example, political, influence on markets. Nevertheless, the existence of invalid information over the social network makes the users unhappy, may lead to bad-mouthing of any business or political organization and may also lead to chaos in the real world, particularly in the crisis. For instance, consider the current crisis faced by the world – The COVID-19 pandemic caused by corona virus. Following are some rumours being spread on Twitter:

"Corona virus is not heat-resistant. Walking outside is enough to disinfect you"
"Social distancing is not effective"
"Drinking bleach and ingesting colloidal silver will cure COVID-19."
"The National Guard just announced that no more shipments of food will be arriving for 2 months - run to the grocery store ASAP and buy everything!"

These rumours may mislead people to do harmful or fatal things as remedies or make them careless about the necessary precautions or create chaos. To overcome these problems and other possible issues, automatic and early detection of rumours on Twitter can be carried out using this rumour detection system. The primary goal of the rumour detection system is to stop the spread of false news and misinformation about an event, person or organization on a social media platform such as Twitter. The project not only detects the potential rumours spreading on Twitter but also checks their veracity by verifying them with reliable news sources at real time.

RELATED WORKS

A survey of the literature suggests detecting rumour on social media has gained lots of attention lately.

[1] Dr. Dinesh B. Vaghela1, Divya M. Patel2 , "Rumour Detection with Twitter and News Channel Data Using Sentiment Analysis and Classification" in International Journal of Advance Engineering and Research Development Volume 5, Issue 02, February -2018.

- In this paper, detection approach is based on the sentiment classification. The paper has results of comparison between different supervised learning techniques for detection of rumours. The limitation is that only one feature, sentiment polarity, is used to

detect a rumour. The accuracy of their model ranges from 60-70% using different algorithms.

[2] Rosa Sicilia, Stella Lo Giudice, Yulong Pei, Mykola Pechenizkiy, Paolo Soda, "Twitter Rumour Detection in the Health Domain" in Expert Systems With Applications (2018).

- The paper explores the selection of different features using different classification methods for rumour detection which is beneficial in selecting effective combination of classifiers and features. It aims at analyzing which features are the most informative and whether the newly introduced ones are meaningful or not for the classification purpose.

[3] Hardeo Kumar Thakur, Anand Gupta, Ayushi Bhardwaj and Devanshi Verma, "Rumour Detection on Twitter Using a Supervised Machine Learning Framework" in International Journal of Information Retrieval Research, Volume 8 Issue 3, July 2018.

- In this article, a two-fold supervised machine-learning framework is proposed that detects rumours by filtering and then analyzing their linguistic properties. This method attempts to automate filtering by training multiple classification algorithms with accuracy higher than 81.079%. Finally, using textual characteristics on the filtered data, rumours are detected.

[4] Suchita Jain, Vanya Sharma, Rishabh Kaushal, "Towards Automated Real-Time Detection of Misinformation on Twitter" in Intl. Conference on Advances in Computing Communications and Informatics (ICACCI), pp. 2025-2020, IEEE 2016.

- This paper introduces an algorithm that detects rumors on Twitter using tweets from the verified news channels as base. The algorithm is based on the premise that Verified News Channel accounts on Twitter furnish credible information as compared to the naive unverified account of user. The paper proposes an approach to detect rumors from the data using sentiment and semantic analysis. Accuracy of the model ranges from 60-76%.

[5] V. Sivasangari, Ashok Kumar Mohan, K. Suthendran, M. Sethumadhavan, "Isolating Rumors Using Sentiment Analysis" in Journal of Cyber Security and Mobility, Vol. 7 1, 181–200. River Publishers, 12 June 2018.

- The model retrieves data using Twitter Scraper instead of Twitter API. Sentiment analysis is used to differentiate the true and false text for that large volume of scraped information. The hashtag input from the text is separated based upon positive and negative tags.

[6] Oluwaseun Ajao1, Deepayan Bhowmik2 and Shahrzad Zargari, "Sentiment Aware Fake News Detection on Online Social Networks" in ICASSP, IEEE 2019.

- The paper hypothesize that there exists a relationship between a rumour and the sentiment of the texts posted online. In the determination of the word relevance and usage within the corpus, they consider sentiments for the terms and words. Topic models enable the identification of most relevant words and concepts within a text corpus. The emotion ratio of negative to positive words is computed.

PROPOSED SOLUTION

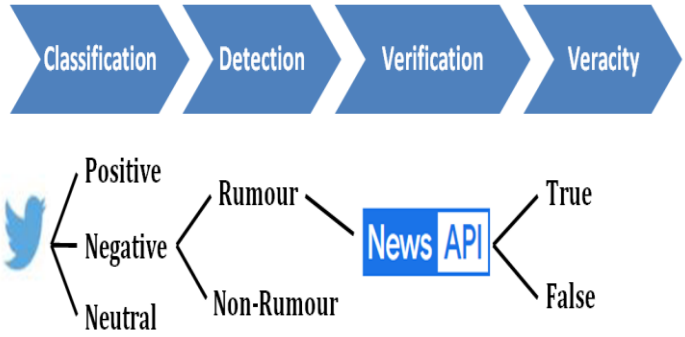

Fig -1: Proposed System

The rumour detection system facilitates a Python based Application for early detection and verification of rumours on Twitter. The procedure is carried out by finding out new trending topics in the different sectors to recommend keywords to the users which then will be used to classify the tweets as positive, negative or neutral by performing Sentiment Analysis using Naïve Bayes' probabilistic model in order to detect rumours related to that field on Twitter. Then, confirming the veracity of the rumours by using the news API contains data retrieved from verified news sources. The system proactively informs the user about currently trending topics on twitter. The user may also enter any topic of interest to check for presence of rumours. The machine learning model is trained to classify words into categories- positive and negative and analyse the sentiment polarity of the tweets for presence of potential rumours. Since a rumour is of negative sentiment, the action of rumour detection is carried out in the negative tweets. The rumour is detected in the negative tweets using various features responsible. The system then verifies the veracity of the potential rumour by checking its presence in reliable news sources. The system receives real-time trending keywords, tweets and news too. Thus, serves the purpose of early detection and verification of rumours on Twitter. The previous rumour detection projects do not give complete assurance of the veracity of rumours. We fill in this gap by finding the veracity of rumours by matching them against a real-time, verified news source. We are using News API as a

source of verified information because the News available on the platform has verified information. It holds accountability for the information posted. It strives to maintain the reputation to post the correct news as fast as possible and consider the fact that the news would affect a large user base. Thus, information from a News API can be considered trustworthy.

IMPLEMENTATION

- Packages and Technologies Used

The software platform used is Microsoft Visual Studio Code, language used for development is Python and libraries such as Python, NodeJs, npm are installed which are free and open source. The algorithm was implemented in python programming language with help of other APIs and packages. We harvested trending keywords from Twitter API. The tweet data was collected from Twitter API. TextBlob package was used for sentiment mining. News API was used for rumour verification module. The best way to access, collect, and store data from social media platforms is generally through application programming interfaces (APIs) APIs are easy-to-use interfaces that are usually accompanied by documentation that describes how to request the data of interest. They are designed to be accessed by other applications as opposed to web interfaces, which are designed for people; APIs provide a set of well-defined methods that an application can invoke to request data. For instance, in a social media platform, it may be desirable to retrieve all data posted by a specific user or all the posts containing a certain keyword. Twitter provides a range of metadata with each tweet collected, including tweet language, location, and so on, as well as details of the user posting the tweet.

- Algorithm Used

Multinomial Naive Bayes Classifier
Naive Bayes Classifier is a classification algorithm that relies on Bayes' Theorem. This theorem provides a way of calculating a type or probability called posterior probability, in which the probability of an event. An occurring is reliant on probabilistic known background. The Naive Bayes' algorithm assumes that the features are independent of other features.

$$P(A|B) = P(B|A) * P(A)/P(B)$$

The algorithm first computes the probability of a word to be present in a positive or negative tweet. This is computed from the training data.

Probability (word in positive tweet) = frequency of occurrence of this word in positive tweets, for instance, what fraction of all tweets containing this word are positive. A tweet contains many words. The probability of a set of words to be in a positive tweet is defined as the product of the probabilities for each word. This is the Naïve assumption. Using the pre-estimated values of these probabilities, you can compute the probability of a tweet to be positive or negative using Bayes theorem. Whenever a new tweet is fed to the classifier, it will predict the polarity of the tweet based on the probability of it having the polarity.

- TRIV : The Web Application

TRIV i.e. Twitter Rumour Identification and Verification is created using Microsoft Visual Studio Code. It showcases the implementation and results of the proposed rumour detection and verification model. It provides an interactive portal to the Twitter-users to obtain the veracity of tweets and also contribute to the rumor detection process. The application displays multiple findings related to the rumour topic such as the sentiment analysis report and the list of classified tweets. Along with this functionality, it also displays the detected rumour and the closest news headline match if in case its veracity is true. If a rumour is classified as misinformation, it is recorded in a database and displayed on the application portal. This interactivity helps in maintaining the integrity of news and information available on Twitter and making the users a part of the process.

SYSTEM DESIGN

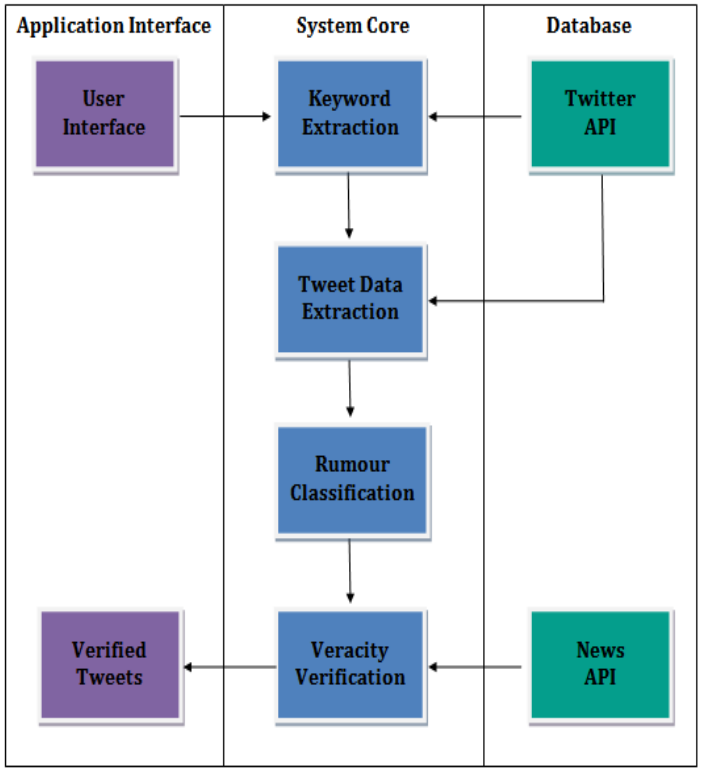

Fig -2: System Block Diagram

1] Trending Keywords Extractor

One of Twitter API's contents is retrieval of trends near a location. The API returns 50 trending topics for a specific

location or the whole global community as requested. The response is an array of trending objects along with name of the topic, query parameter to search for the topic on Twitter, its URL and the tweet volume for the last 24 hours related to the topic. The application interface recommends the trending keywords which are extracted from the Twitter API. The user may select one of these recommended topics or type a topic name of his interest as a keyword for searching the Tweets. The application, when not in use, proactively notifies the user about the currently trending topics on Twitter which the user can choose to analyze.

[2] Data Retrieval for the Application

Initially we have to log in into our Twitter account. Create new application on Twitter and fill out the given form to register our Twitter application and get our own credentials. Then authenticate our Python script with the API using the credentials obtained from Twitter which include API key, API secret, Access token, Access token secret, etc. We a python library called Tweepy to connect the API and download the data. Create a function in our python program to download tweets based on a search keyword. The search keyword is either entered by the user or chosen from the trending keywords retrieved from Twitter. Data related to Tweets, User and Network can be retrieved. The data is in JSON format. Thus we use json library for parsing the data. Some more libraries are used in the program such as pandas for data manipulation, re for regular expressions, matplotlib for data representation, etc. Then we structured the Tweets data into a data frame for easy data operations.

[3] Rumour Detection in Real-time Tweets

The system takes the entered keyword as input and extracts tweets containing those keywords from the Twitter API in real time. The tweets retrieved are cleaned and fed to Textblob API for sentiment analysis. The tweets are then fed to the Naïve Bayes' classifier to classify as positive or negative based on their sentiment score. The tweets having negative sentiment greater than the threshold value are further classified as rumour or non-rumour based on the selected features. Once a tweet is marked as a potential rumour, its veracity is checked using a keyword matching algorithm.

[4] Verifying the veracity of Rumour

The considered rumour will be then matched against real time news articles retrieved from a verified news source. News API is a verified and simple API that returns JSON metadata of news headlines and articles live all over the web in real time. The data is parsed using json library. Recurring words from the rumour text are extracted. This list of keywords is matched against the data retrieved from the News API. According to the result of the match, the application will show the result whether the potential rumour is true or false. If the match for keywords from the tweets is found in the news data, the tweet's veracity is true. This means that it is not a rumour. If there is no match in the news database, that tweet is a rumour.

MACHINE LEARNING MODEL

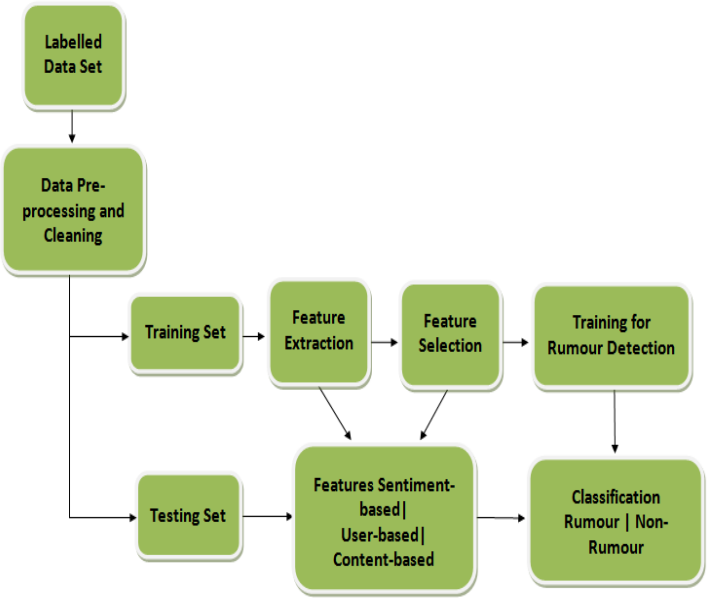

Fig -3: Machine Learning Model

[1] Preparing the Training and Testing Set for the Model

Labelled Twitter datasets are available free of cost online. We downloaded our dataset from Kaggle.com which consists of metadata such as Tweet ID, content, timestamp, re-tweet count, user profile picture details, account verification status, location enabled/disabled status, follower count, and rumour label (rumour or non-rumour), etc. We divided the dataset into 90% training set and 10% testing set for effective training and testing of the supervised machine learning model.

[2] Pre-processing Tweets in the Data Sets

Cleaning the data is very important because it improves the data quality and increases the productivity of the model. Removal of unnecessary strings of words or characters improves decision-making ability. This implies to both the training and testing sets. The process includes:

- Removal of twitter handles (@user).
- Removal of punctuations, numbers, special characters.
- Removal of short words.

Words having two or less letters are removed.

- Tokenization

Sentences are converted into lists of words.

- Stemming

Transforming any form of a word to its root word. It is a process of stripping the suffixes ("ing", "ly", "es", "s" etc) from a word.

- Extraction of hashtags

[3] Feature Extraction from Cleaned Tweets

[3.1] Sentiment-based features
These involve features related to sentiments and meaning of words.

- Sentiment score

Experiments prove that there exists a relation between sentiment polarity of words and false news. False news is spread with a negative intention and thus tends to have a negative sentiment. Sentiment score is obtained using the Textblob API.

- Denial term and Question Existence

The re-tweets of a potential rumour sometimes contain denial, doubt or a question about its veracity.

[3.2] Content-based features
Content features are directly extracted from Tweet texts and other characteristics of the Tweets.

- Retweets number

The Tweets containing rumours have a lesser re-tweet count since people tend to doubt its truthfulness and decide to not further spread it.

- WordCloud

Collecting the common words used in the tweets labelled as positive or negative in the dataset and making two separate feature vectors for positive and negative tweets.

- URL

Existence of a URL is a very informative feature because most of the genuine users have a strong habit of supporting a statement with a link to a web page containing verified information. Rumours usually indicate news without reference, so that can't be verified. Typically it shows an instantaneous and huge spread in the network, in other words an infectious behavior. An example retrieved from the dataset is: "Corona disease can be cured by drinking alcohol!!! #Coronavirus."
The non-rumour class usually identifies all referenced news with at least one link to a certified and official webpage, such as news papers, hospitals, universities, etc.

[3.3] User-based features
These features consist of information related to the users and their accounts.

- Followers count

Verified accounts and renowned users have a high follower count. Such users do not usually spread rumours.

- Profile Photo and Location

Rumour-mongers usually keep no profile picture so as to hide their identity and they also disable their location setting to make themselves untraceable.

- Account Verification Status

It is evident that verified accounts do not spread unverified misinformation since they have a reputation to maintain.

[4] Training the model

The Naïve Bayes' algorithm is used to prepare a machine learning model. The model is fit on the training set. The model is trained by plugging our feature vector into the Naive Bayes Classifier.

[5] Testing the Model

The fitted model is used to detect the results for the content in the testing model in order to obtain an unbiased evaluation of how accurately the model works. We now have a list of rumours along with their predicted labels (rumour or non-rumour) and their actual labels. We used this information to evaluate the accuracy of our model.

GUI

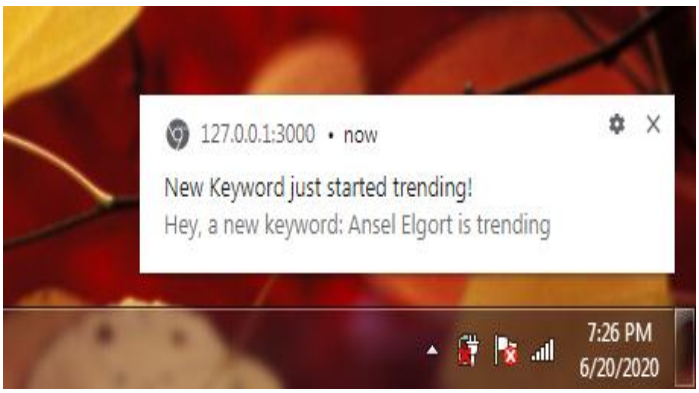

Fig -4: Notification from Application

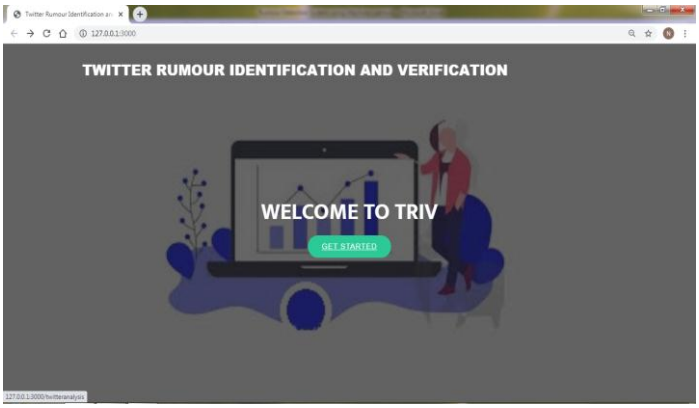

Fig -5: Splash Screen

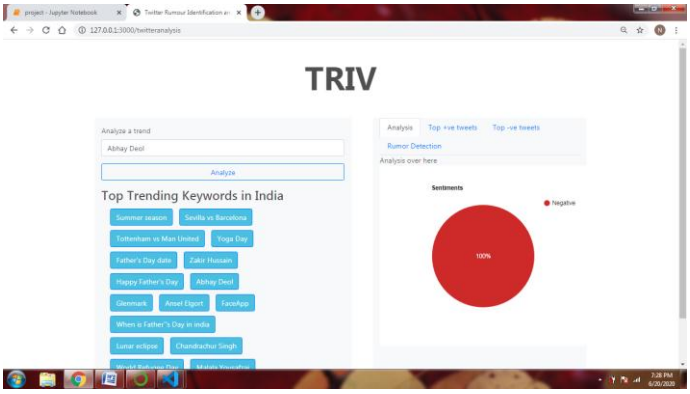

Fig -6: Sentiment Analysis Report

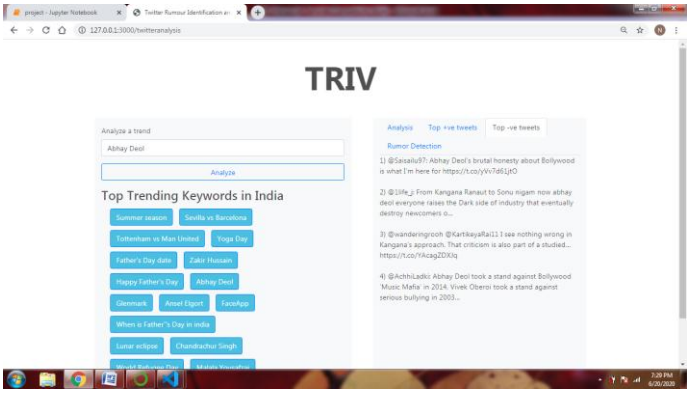

Fig -7: Potential Rumours Detected

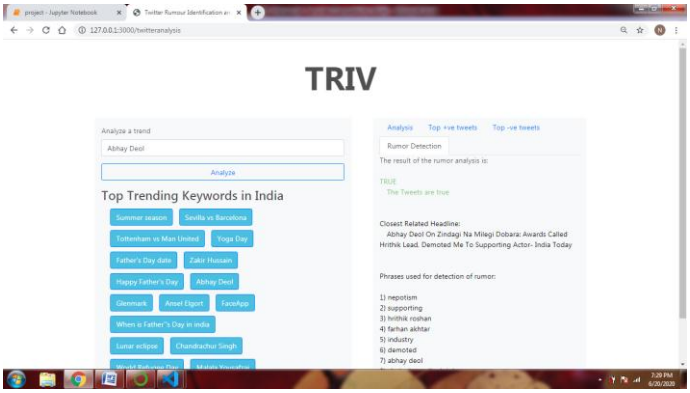

Fig -8: Example of a Tweet being True Information

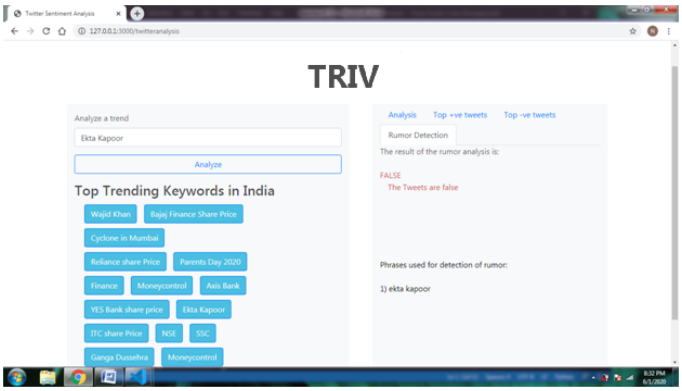

Fig -9: Example of a Tweet being False Information

EVALUATION METRICS

The evaluation metric used is accuracy, precision, recall and F-measure. While predicting values against labeled, we get four bins which are True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). TP is rumoured event is predicted as rumour, TN is non-rumoured event is predicted as non-rumour, FP is non-rumoured event is predicted as rumour, and FN is rumoured event is predicted as non-rumour. Precision is the fraction of the correctly predicted rumour micro-blogs to all rumor micro-blogs identified. Recall is the proportion of correctly predicted rumor micro- blogs to all the rumor micro-blogs. Whereas F-measure can be considered as the harmonic mean of recall and precision.

Evaluation Metric	Formula	Value for Rumour	Value for Non-Rumour
Precision	TP / (TP + FP)	0.74	0.73
Recall	TP / (TP + FN)	0.73	0.75
F-measure	2 * (Recall * Precision) / (Recall + Precision)	0.73	0.75
Accuracy	(TP + TN) / Total events	0.74	

Table -1: Evaluation Metrics

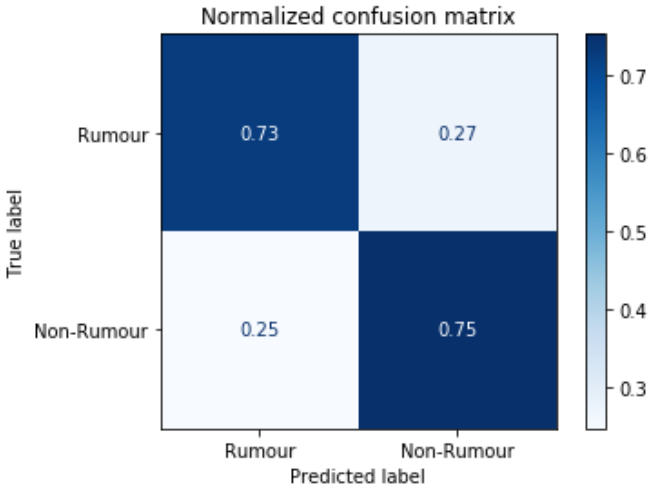

Fig -10: Confusion Matrix

CONCLUSIONS

Research on the development of rumour detection and verification tools has become increasingly popular as social media penetration has increased, enabling both ordinary users and professional practitioners to gather news and facts in a real-time fashion, but with the problematic side effect of the diffusion of information of un-verified nature. This

project has summarized studies reported in the scientific literature toward the development of rumour detection systems and has described a different approach to the development of the two main components, rumour detection and rumour veracity verification. Coupling the rumour detection model with the veracity verification model using real-time data achieves significant improvement over previously developed rumour detection methods. This web application would serve as an innovative platform for different sectors such as politics, government, business, health, film industry, and many more to reach the trending tweets related to the respective subject, perform the sentiment analysis and detect rumours if present. This would help them to prevent the spread of false news, misinformation, misleading instructions, intentional bad-mouthing, image destruction, spread of confidential information, etc.

FUTURE WORK

Future work mainly includes additional sources of sentiment extraction could be from images, embedded text in the image and other visual media such as animations (GIFs) and videos may enhance model performance.

Also, integration of this project with Twitter would help to detect rumours but also stop their spread immediately. The output of the project could help twitter to alarm the users about false information and also help people to report the spread of misinformation to Twitter. An important limitation toward the development of rumour classification systems has been the lack of publicly available datasets. In recent years, research in rumour detection has largely focused only on the data available on public platforms like twitter, etc. Future work should focus on rumour detection on other social media platforms such as Facebook, Whatsapp, Instagram, etc. The project scope must be extended to these platforms as these play an important and huge role in the spread of information on social media.

ACKNOWLEDGEMENT

We would like to thank Prof. Priti Chakurkar for her valuable guidance, time, support, comments, suggestions, persuasion, for providing all the necessary facilities and the necessary background knowledge about the topic, all of which was indispensable in the completion of this project.

REFERENCES

[1] Dr. Dinesh B. Vaghela1, Divya M. Patel2 , "Rumour Detection with Twitter and News Channel Data Using Sentiment Analysis and Classification" in International Journal of Advance Engineering and Research Development Volume 5, Issue 02, February -2018.

[2] Rosa Sicilia, Stella Lo Giudice, Yulong Pei, Mykola Pechenizkiy, Paolo Soda, "Twitter Rumour Detection in the Health Domain" in the Journal of Expert Systems With Applications-2018.

[3] Hardeo Kumar Thakur, Anand Gupta, Ayushi Bhardwaj and Devanshi Verma, "Rumour Detection on Twitter Using a Supervised Machine Learning Framework" in International Journal of Information Retrieval Research, Volume 8 Issue 3, July 2018.

[4] Suchita Jain, Vanya Sharma, Rishabh Kaushal, "Towards Automated Real-Time Detection of Misinformation on Twitter" in Intl. Conference on Advances in Computing Communications and Informatics (ICACCI), pp. 2025-2020, IEEE 2016.

[5] V. Sivasangari, Ashok Kumar Mohan, K. Suthendran, M. Sethumadhavan, "Isolating Rumors Using Sentiment Analysis" in Journal of Cyber Security and Mobility, Vol. 7 1, 181–200. River Publishers, 12 June 2018.

[6] Oluwaseun Ajao1, Deepayan Bhowmik2 and Shahrzad Zargari, "Sentiment Aware Fake News Detection on Online Social Networks" in ICASSP, IEEE 2019.

BIOGRAPHIES

Neha Sawant
Pursuing B.E. in Information Technology
MIT Pune

Sameer Thakare
Pursuing B.E. in Information Technology
MIT Pune

Saloni Wandile
Pursuing B.E. in Information Technology
MIT Pune

Ojas Patil
Pursuing B.E. in Information Technology
MIT Pune