Twitter Rumour Detection in the Health Domain

Anjana Krishna Sajith[1], Manika Tyagi[2], Riddhi Maniktalia[3], Muskan Bansal[4]

[1]SCOPE, *Vellore Institute of Technology*, Vellore, India
Email: *anjanakrishna. sajith2020[at]vitstudent.ac.in*

[2]SCOPE, *Vellore Institute of Technology*, Vellore, India
Email: *manika. tyagi2020[at]itstudent.ac.in*

[3]SCOPE, *Vellore Institute of Technology*, Vellore, India
Email: *riddhi. maniktalia2020[at]vitstudent.ac.in*

[4]SCOPE, *Vellore Institute of Technology*, Vellore, India
Email: *muskan. bansal2020[at]vitstudent.ac.in*

Abstract: ***Due to the widespread usage of the internet now a-days, a huge quantity of data is produced every day as a result of our daily activity on many platforms including social media, news articles, forums, emails, and websites. Twitter is a microblogging website among social networking sites where users may express their ideas, opinions, and knowledge in a condensed amount of words. Yet, the platform is also susceptible to misinformation that could negatively impact people's lives, particularly in situations involving health. This study focuses on finding health-related rumours on Twitter and figuring out how they spread in order to address this problem. Data gathering, identification, verification, and the study of networks that spread rumours are the four mod-ules that make up the project. These modules make use of data mining, manual data processing, supervised machine learning, and social network theories. The implementation is carried out using the Python programming language and pertinent libraries like Pandas and Numpy. Several researchers are investigating various approaches to identify whether a subject is a rumour or not in order to tackle the issue of fake news. This study suggests a brand-new rumour detection technique that includes recently developed features like impact potential and measurements of network parameters.***

Keywords: Rumor Detection and Classification; Machine Learning; Twitter; Online Social media

1. Introduction

Social media is becoming increasingly popular for news consumption due to its easy access, fast dissemination, and low cost. Social media, on the other hand, also makes it possible for "fake news, " or news that contains purposefully misleading information, to spread widely. Currently, a lot of data is being regularly generated in social media. Modern breaking news first appears on microblogs before reaching traditional media channels. As belief, opinion, and sentiment in online communication are sociopragmatic phenomena, microblogging websites are rich sources of information that have been successfully tapped for this purpose. One of the most well-known microblogging services is Twitter.

Yet, the unprecedented availability of such vast amounts of data has a double-edged effect; it is simple to obtain erroneous information from these sources, and it is difficult to prevent the dissemination of misleading information, whether on purpose or unintentionally. An inundation of data is presented to the information seeker. Social media fake news has a serious potential to harm society. Fake news identification and mitigation also pose special difficulties. Many current research projects make use of different data aspects, including network features, to address these problems. In recent years, social networks have developed as a vital means of disseminating information, bringing with them a number of benefits. Simultaneously, rumours, which are unconfirmed and instrumentally relevant information claims in circulation, are becoming a possible threat to society. As a result, despite the fact that determining whether a topic is a rumour in social microblogs has been explored in numerous ways, There is a need to determine if a post is a rumour or not. This is the subject of this study.

The aim is to create a new rumour detection system that takes advantage of newly created features such as impact potential and network characteristics measurements. We put our method to the test on a real-world dataset. A collection of health-related tweets culled from the Twitter microblog. We see positive outcomes, as evidenced by the reasonable levels of precision; the algorithm can properly detect around 90% of rumours.

Because rumours have the potential to be harmful, especially in the area of health, it is crucial to create machine learning models for handling them. By employing data mining and processing methods to aggregate data from Twitter and create a model for rumour identification, this study seeks to address this difficulty. The creation of datasets as well as methods for gathering and annotating data from social media are explained. Also, the elements of a rumour categorization system are described, and its precision is examined. The objective is to develop a machine learning model that can identify fraudulent health-related tweets with high accuracy and minimal complexity. To discover social network linkages at various degrees of rumour diffusion, the processed dataset will be used. We will conclude with a brief discussion of the social network that rumour spreads on, emphasizing the value of social network analysis in comprehending the dynamics of rumour on social media.

2. Related Work

Researchers have studied several aspects of credibility of online information. For example, Sardar [1] uses preprocessed data for rumour detection. Here the output with preprocessed data gives lower accuracy in comparison to the output without preprocessing data. Another relevant work is by Sahana [2] whose model has accuracy of recall rate, but the model is only focused on one rumour topic. We also went through Zhiwei [3] work on detection and analysis of the US presidential election related rumours on Twitter. They achieved an accuracy of 82%, and stated the efficiency of their algorithm can be worked on as a future scope.

Qiao [4] used content and user based features to detect rumours on social networks. They realised that the combined method of content and user features is better than others with a 7.1% improvement in precision. Here, user credibility is a limitation. Another method was by Yan [5], they used some threshold based on their original dataset and among those thresholds they achieved an accuracy of 88%. Carlos [6] focused on a broader dataset in comparison to Qiao [4]. They focused on automatically assessing the credibility of a set of tweets. They classified tweets related to trending topics as credible or not by using a supervised learning method.

Aditi [7] et al worked on the credibility and ranking of tweets during high impact events. They adopted a supervised machine learning and relevance feedback approach using var-ious features to rank tweets according to their credibility score. Moreover, they performed logistic linear regression analysis. With respect to few features to estimate the good predictors for credibility of tweets. A few more researchers worked on the similar lines like P. Kumaraguru [8]. They performed a characterization analysis to understand the influence patters for the spread of fake images. Methods like retweet count, network relations and decision trees were used to analyze. Statistical and empirical approaches were also adopted by Amanda [9] et al.

A method which had a limited focus on retweets was by Marcelo [10]. They analyzed and characterized the social network of the community surrounding the topic and con-firmed that network topology characteristics remained un-changed regarding studies performed under normal circum-stances. Fan [11] et al used a completely different set of features for their model. They introduced the client program used features and the event location feature into the features used for classification respectively to study their effectiveness. KP Krishna [12] used the concepts of cognitive psychology for de-tecting misinformation in online social networks. He evaluated the spread of misinformation in online social networks by an algorithm which used psychology.

Bruno [13] and his work lacked on implicit features of content and its users. Ke Wu [14] detected false rumours on Sina Weibo by introducing propagation tree and used a random walk kernel algorithm to build rumour detection classifiers.

V. Janakiraman [15] worked on a completely different dataset. They worked on aviation data using extreme learning machine by adapting unsupervised ELM algorithms based on the au-to encoder to detect anomalies in large aviation data sets.

A method that is based on user behaviour to detect rumours was formed by Gang [16] et al. However, their detection method is based on the aggregate of all the users and thus has much room for improvement. Ke Qu [17] studied message propagation patterns on Sina Weibo and used them as high order features to construct a graph-kernel-based SVM classifier for rumour identification. Weiling Chen [18] et al work examined the strengths and limitations of existing techniques, and proposed future directions in this area.

Joao [19] et al paper stated that in future they will investigate the use of more sophisticated deep learning models, such as attention-based and transformer-based models, to capture the complex patterns and relationships in social media data. Tahir [20] also stated that they will work of the use of ex-plainable AI techniques to improve the interpretability and transparency of the detection models. The use of more diverse features such as network and content features in coming times will be implemented in the model was stated by Faisal [21].

The objective of Zhiyuan [22] et al literature survey paper is to review and analyze the existing research on rumour detection using propagation graph neural networks with attention mechanism. Na Bai [23] paper highlights that GCNNs have the ability to capture the complex relationships and patterns of rumour propagation within a social network. Ash Mohammed [24] paper showed promising results in a varity of social network analysis applications, such as link prediction community detection, and node classification.

The model by Adyan [25] captured the complex relation-ships between words and phrases in tweets. Han [26] et al work proposed a novel approach that leveraged the hierar-chical structure of social networks. Whereas Jing [27] et al work proposed an approach that leveraged the tree-structured recursive neural networks.

Shouzhi [28] model incorporated user and content-level fea-tures to improve its accuracy in detecting rumors. Sarah [29] future scope was something out of the box, he use of more diverse and multimodal data sources, such as images, videos, and audio, to improve the accuracy of rumor detection models. Dinesh [30] needed a better preprocessing process.

Rosa [31] paper limitation was that it collectively tested with multiple domains and its comparison with each other resulted in ambiguity outputs. The paper by Hardeo [32] used a combination of methods-user based, content user based in order to get results which led tto better precision. With the help of two hidden layers of auto encoding by Oluwaseun [33] gave improved results.

This author-Suchita [34] used some examples of rumours and according to verified news channels and general public tweets ratio detected rumours by using sentiment and semantic analysis. Suthendran [35] et al used an API, which had some constraint features as the user will be limited to send only 180 requests for every 15 minutes. BET based classificated for detecting rumours by Rini [36] was a novel approach which explored the impact of different rumours too.

Sarah [37] approach was effective in detecting high-engaging breaking news rumors with an accuracy of 89.7%. The authors also found that social context features, such as user credibility and social relationships, play an important role in rumor detection. Wanlun [38] found that incorporating user feedback in the DNN model improves its performance and reduces the impact of noisy data. C Castillo [39] shows that the proposed model is effective in predicting the credibility of information on Twitter with an accuracy of 86%. A Friggeri [40] findings of the paper show that rumors tend to spread quickly and widely on social media, with a median cascade size of 13 users. The authors found that the early spreaders of rumors have a significant impact on their propagation and that rumors are more likely to be retweeted if they are novel and generate strong emotions.

3. Methodology

We go into great detail about our study process in this section. We first go over the procedure for gathering data from Twitter, then we go over the various analytical methods used in this research.

a) Data

William [41] published an annotated Twitter data set on Kaggle in 2018. The dataset description is as follows:

- id: unique id for a news article
- title: the title of a news article
- author: author of the news article
- text: the text of the article; could be incomplete
- label: a label that marks the article as potentially unreli-able
  - 0: reliable
  - 1: unreliable

The size of our data set in total 20, 799 tweets in total for ourexperimentation.

1) Scraping data: Further during analysis, we scraped Twitter data using the snscrape module for tweets containing the term ”covid-19” between two dates. The scraped data includes the tweet ID, date, user display name, tweet content, user location, number of retweets, and the ID of the original tweet. The stopwords were also removed from the data with the help of the stop words module from Natural Language Toolkit library. The next step was text preprocessing, along with stemming.

b) Packages used

The project implementation has been achieved by using a number of modules and packages. Pandas, an open-source, BSD-licensed Python library that offers high-performance, user-friendly data structures and data analysis tools for the Python programming language, is one of the key packages utilised in this strategy. NumPy, which stands for Numerical Python, is the next significant package utilised. It is a library made up of multidimensional array objects and a selection of functions for processing those. Using NumPy, mathematical.

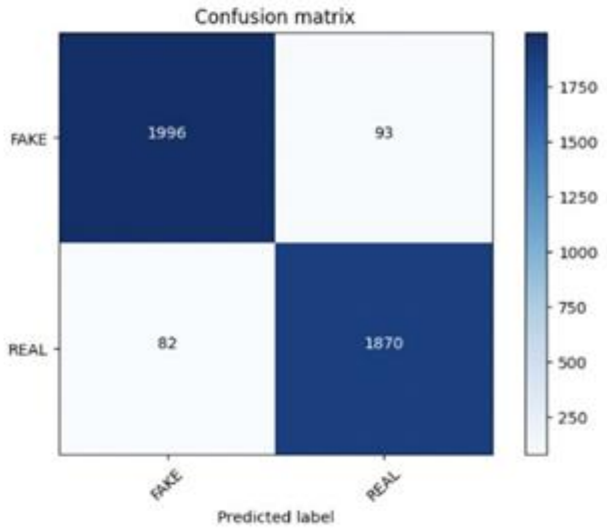

Figure 1: Confusion matrix using Decision Tree Classifier

and logical operations on arrays can be performed. It also has functions for working in the domain of linear algebra, fourier transform, and matrices.

c) Train and test data

The first step being used to build the model is reading the CSV data file, using the Pandas package. The dataset being used has 20, 799 entries of tweets in the health domain. The dataset consisted of the tweet title, the author, actual tweet along with the label which is 0 for rumour and 1 for not a rumour. After this the cleaning of the dataset is done where the NA (not applicable) and authors column is removed to get a clear picture of the dataset. Here 80% of the dataset is used for training the model and the rest 20% for testing the model. Now the dataset is divided into dependent and independent, the independent variable x comprises the title and the tweet whereas the dependent variable y has the label. Next the Count Vectorizer module is used which converts text to numerics, these numerics are preserved in a matrix. Now the NumPy package is used to stack head matrix over body matrix. Now the dataset is split into a training and testing set.

d) Model Implementation

Another module being used at this point is Decision Tree Classifier, which is the main algorithm for the project. A supervised learning method called a decision tree can be used to solve classification and regression problems, but it is typically favoured for doing so. It is a tree-structured classifier, where internal nodes stand in for a dataset's features, branches for the decision-making process, and each leaf node for the classification result. The Decision Node and Leaf Node are the two nodes of a decision tree. Whereas Leaf nodes are the results of decisions and do not have any more branches, Decision nodes are used to create decisions and have numerousbranches.

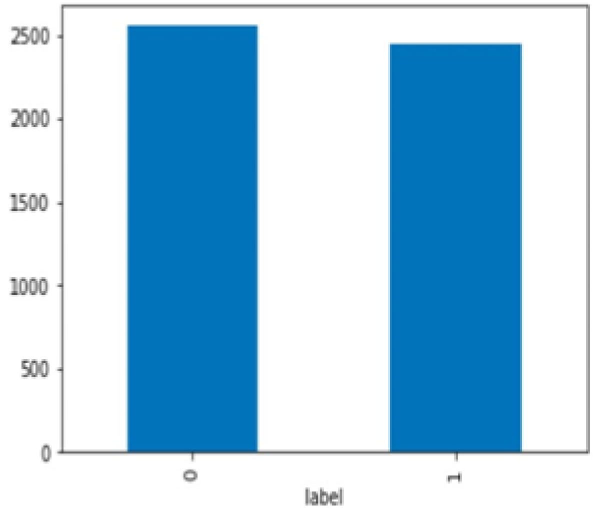

Figure 2: Analyzing data through a histogram

The given dataset's features are used to execute the test or make the decisions. It is a graphical depiction for obtaining all feasible answers to a choice or problem based on pre-determined conditions. Because it starts with the root node and expands on succeeding branches to build a structure resembling a tree, it is known as a decision tree. Using the decision tree we have classified the tweet as being a rumouror not.

We have used other algorithms in order to search for the best accuracy. A supervised learning technique for classification tasks is K-Nearest Neighbors (KNN). It works by determining the K training instances closest to the new data point being classified in the feature space, and then classifying the new data point using the majority class among those K neighbours. KNN is used to categorise Twitter rumours by first represent-ing each tweet as a feature vector. The frequency of each word in the tweet is taken into account while creating the feature vectors, and it is contrasted with the frequency of the same phrase over the full corpus of tweets. The KNN method is then used to classify the tweets as rumour or non-rumor based on the features after the feature vectors have been created. The accuracy of this approach was 84.18%.

Because it can handle high-dimensional data and effec-tively separate classes even when the data cannot be linearly separated, SVM is a popular choice for classifying Twitter rumours. A regularisation parameter in SVM can also be adjusted to avoid overfitting. Overall, SVM has been found to perform well in trials and is a promising algorithm for classifying Twitter rumours. The accuracy discovered with thistechnique was 97.07%.

e) Evaluation Metrics

Now after the modelling and testing, to find the accuracy of the model the confusion matrix is prepared using the con-fusion matrix module. This confusion matrix is then plotted using plot confusion matrix module as shown in Fig. 1, Fig

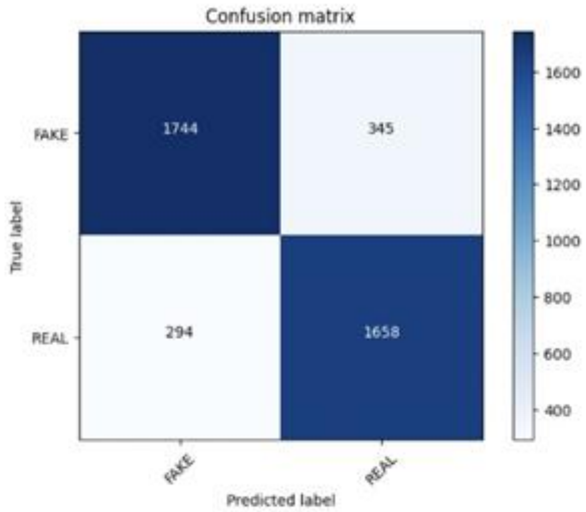

Figure 3: Confusion matrix using KNN

3 and Fig 4 for Decision tree classifier, KNN and SVM algorithms respectively.

Moreover, after the confusion matrix is plotted the accu-racy is found to be around 94.5%, using basic mathematics techniques.

f) Analyzing the Data

For the purpose of locating and dispelling rumours, Twitter data analysis is essential. Twitter is a platform that allows for the rapid and immediate dissemination of information. We can immediately spot new rumours by analysing Twitter data, and we can then act quickly to stop them from spreading further. The source of a rumour, the substance of tweets, and the attitudes surrounding it may all be determined with the aid of Twitter data analysis. This knowledge can help in the creation of efficient mitigating tactics to deal with the rumour

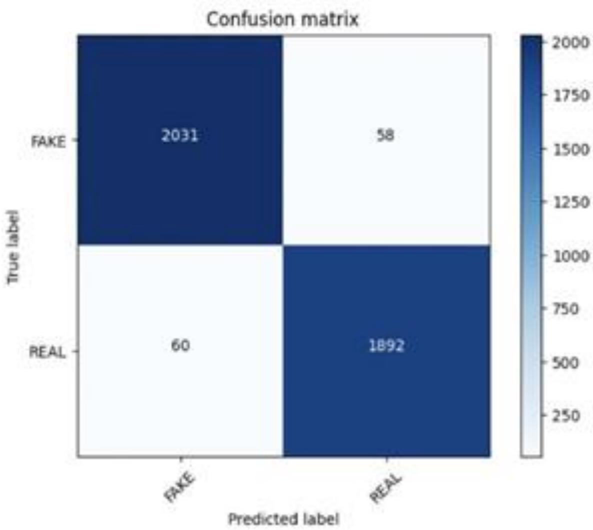

Figure 4: Confusion matrix using SVM

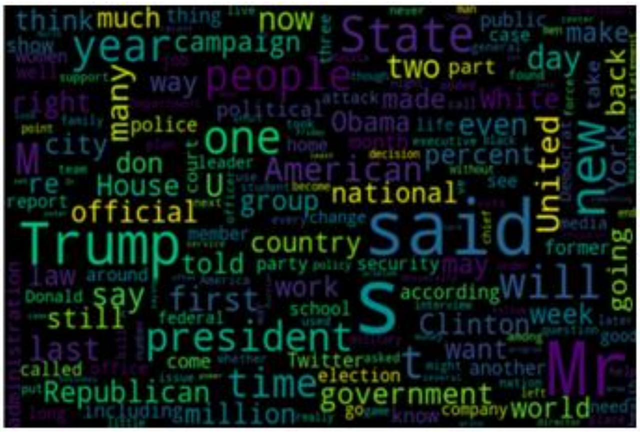

Figure 5: Analyzing data through a WordCloud

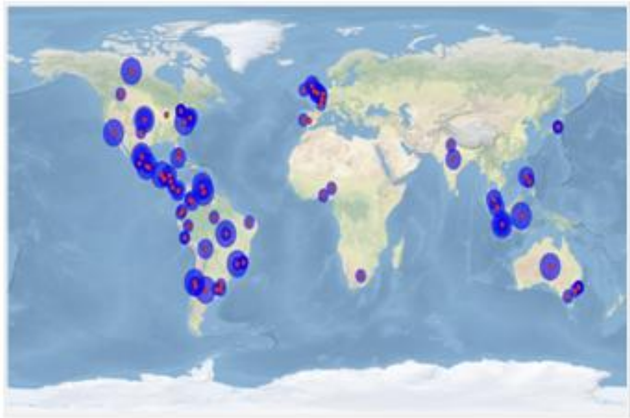

Figure 6: Analyzing data through a GeoMap

and stop it from escalating. Overall, Twitter data analysis is a crucial tool for spreading factual information and eliminating harmful falsehoods.

1) *Through Histogram:* To analyse the dataset a histogram is made i.e. Fig 2, which shows the consistency of data which is fake and real. With matplotlib, the histogram is plotted. For 2D displays of arrays, the pyplot library and Matplotlib are fantastic Python visualisation libraries. NumPy arrays are the foundation of the multi-platform data visualisation software Matplotlib. One of visualization's biggest advantages is that it gives us visual access to vast volumes of data in forms that are simple to understand. There are numerous plots in Matplotlib, including line, bar, scatter, histogram, etc.
2) *Through WordCloud:* Using WordCloud library the word cloud of rumours as well as real tweets, this gives a clear idea of the key words being used in every fake and real tweets. Word Cloud is a great way to represent text data. The size and color of each word that appears in the Word cloud indicates it's frequency or importance. The word cloud so formed will help in social network analysis by getting the most used data and then the topic of fake news can be easily found or detected. It is shown in Fig 5.
3) *Through GeoMap:* This GeoMap is plotted based on the popular Five Thirty Eight journal method. The size of the blue circles in Fig 6 represent the frequency of tweets from each location. If a location has less than or equal to 10 tweets, the circle radius is set to 1.5. If a location has less than 20 tweets, the radius is set to 3.0. If a location has less than 35 tweets, the radius is set to 4.5. If a location has more than 35 tweets, the radius is set to 6.0.
4) *Through Interaction Network:* The centre nodes in this Fig. 7—which shows the graph's greatest linked component—are highlighted. The analysis aims to provide insights into the structure and behavior of the network of Twitter usersand their interactions.

4. Results

In this section, we summarize the results obtained for the characterization and classification analysis performed. In respect of social information network, the analysis is as follow,

A. Clustering

The cluster so formed as shown in Fig 8 is dimensionally reduced, it uses blue colour for not rumour and fake for rumour. This gives clarity of the relation between tweets, users, re-tweets, reactions etc. t-SNE has been used here to have a scatter plot, which is to get a cluster by dimensionally reducingthe data figure.

B. Barabasi Albert Graph

A Barabasi Albert Graph is drawn from the cluster, shown in Fig 9, which tells about how the relation between the users spread the rumour, this is being done by NetworkX package of python language, Network X is a Python language software package for the creation, manipulation, and study of the structure, dynamics, and function of complex networks.

Figure 7: Analyzing data through a Interaction Network

.

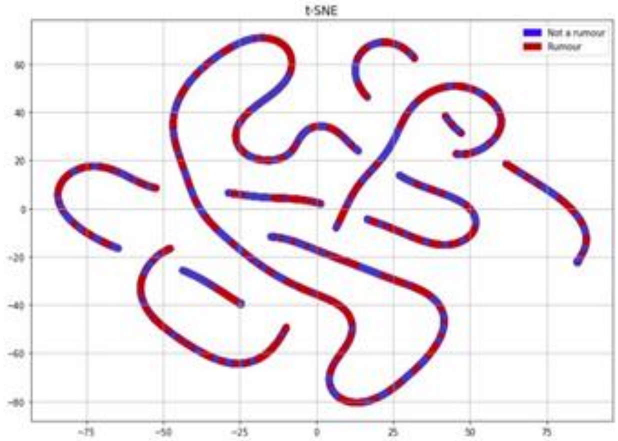

Figure 8: Clusters using Dimensionality Reduction

C. Frequency Distribution Graph

Last but not least, the social network analysis is completed by determining the frequency distribution of fake and actual tweets, which provides details on the intensity of both types of tweets. It explains how bogus tweets propagate and how they ultimately disappear. In Fig 10, the green line represents the graph of actual tweets, while the blue line depicts how rumours begin at their peak and then abruptly begin to slope downward. Additionally, they last longer, even when rumours fade.

5. Conclusion

As the use of social media has grown, research into the creation of rumour detection and verification tools has gained popularity. These tools allow both average users and professional practitioners to gather news and facts in real-time, but they also have the unfavourable side effect of spreading unverified information. The development of rumour detection systems has been the subject of studies reported in the scientific literature. This project has summarised those studies and

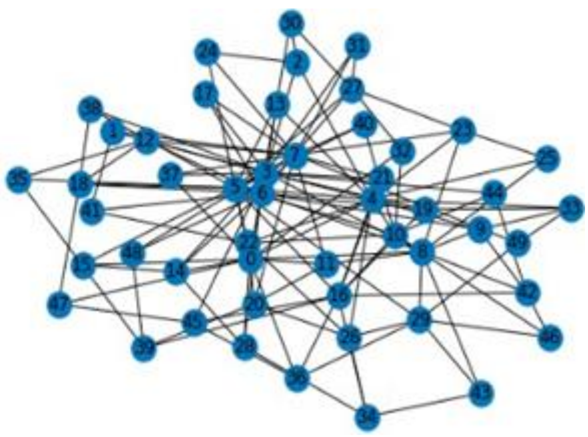

Figure 9: Barabasi graph

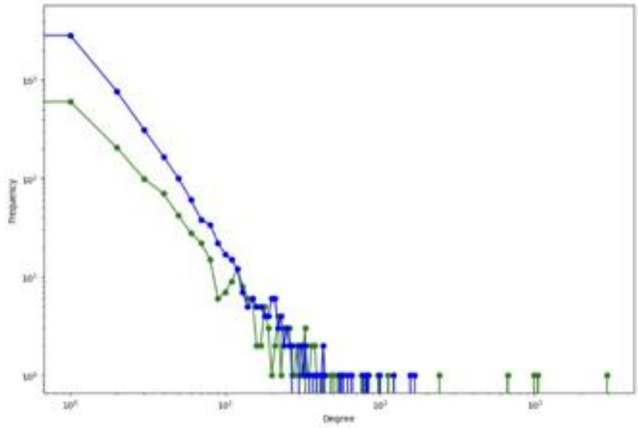

Figure 10: Frequency Distribution graph

described an alternative method for the development of the two key components, rumour detection and rumour verification, as well as social network analysis. Real-time data combined with the rumour detection model and veracity verification model leads to a considerable improvement over previously created rumour identification techniques. This web application will act as a cutting-edge platform for various health sectors to connect with the trending tweets about the relevant subject and identify rumours, if any, are there. They could use this to stop the spread of untrue information, misinformation, mis-leading instructions, malicious defamation, image destruction, the dissemination of private information, etc.

References

[1] Sardar Hamidian et al, Rumor Detection and classifica-tion for Twitter Data
[2] Sahana V P et al, Automatic Detection of Rumored Tweets and finding its Origin
[3] Zhiwei Jin et al, Detection and Analysis of 2016 USPresidential Election Related Rumors on Twitter
[4] Qiao Zhang et al, Automatic Detection of Rumor on Social Network
[5] Yan Zhang et al, Detecting Rumors on Online SocialNetworks Using Multi-layer Autoencoder
[6] Carlos Castillo et al, Information credibility on Twitter
[7] Aditi Gupta et al, Credibility ranking of tweets duringhigh impact events
[8] Aditi Gupta et al, Characterizing and Identifying FakeImages on Twitter during Hurricane Sand
[9] Amanda Lee Hughes et al, Twitter adoption and use inmass convergence and emergency events
[10] Marcelo Mendoza et al, Twitter Under crisis: Can wetrust what we RT?
[11] Fan Yang et al, Automatic Detection of rumour on SinaWeibo
[12] KP Krishna Kumar et al, Detecting Misinformation in Online Social Networks using Cognitive Psychology
[13] Bruno Goncalves et al, Detecting and tracking politicalabuse in social media
[14] Ke Wu et al, False rumors detection on sina weibo bypropagation structures
[15] V. Janakiraman et al, Anomaly detection in aviation datausing extreme learning machines

[16] Gang Liang et al, Rumor identification in microbloggingsystems based on User's behavior
[17] Ke Wu et al, False rumors detection by propagationstructures
[18] Weiling Chen, Unsupervised rumor detection based onusers' behaviors using neural networks
[19] Joao, A systematic mapping on automatic classificationof fake news in social media
[20] Tahir Ahmad et al, Efficient Fake News Detection Mechanism Using Enhanced Deep Learning Model
[21] Faisal Saeed et al, Detecting Health-Related Rumors onTwitter using Machine Learning Methods
[22] Zhiyuan Wu et al, Rumor detection based on propaga-tion graph neural network with attention mechanism
[23] Na Bai et al, Rumour Detection Based on Graph Con-volutional Neural Net
[24] Ash Mohammad Abbas, Social network analysis usingdeep learning: applications and schemes
[25] Adyan Marendra Ramadhani et al, Twitter sentimentanalysis using deep learning methods
[26] Han Guo, Rumor Detection with Hierarchical SocialAttention Network
[27] Jing Ma, An Attention-based Rumor Detection Model with Tree-structured Recursive Neural Networks
[28] Shouzhi Xu et al, Rumor detection on social media us-ing hierarchically aggregated feature via graph neural networks
[29] Sarah A. Alkhodair et al, Detecting breaking news rumors of emerging topics in social media by
[30] Dr. Dinesh B. et al, Rumour Detection with Twitter and News Channel Data Using Sentiment Analysis and Classifica-tion
[31] Rosa Sicilia et al, Twitter Rumour Detection in the Health Domain
[32] Hardeo Kumar Thakur et al, Rumour Detection on Twit-ter Using a Supervised Machine Learning Framework
[33] Oluwaseun Ajao et al, Sentiment Aware Fake News Detection on Online Social Networks
[34] Suchita Jain et al, Towards Automated Real-Time De-tection of Misinformation on Twitter
[35] Suthendran Kannan et al, Isolating Rumors Using Sen-timent Analysis
[36] Rini Anggrainingsih et al, BERT based classification system for detecting rumours on Twitter
[37] Sarah A. Alkhodair et al, Detecting High-Engaging Breaking News Rumors in Social Media
[38] Wanlun Ma et al, Social Media Event Prediction using DNN with Feedback Mechanism
[39] C Castillo et al, Information credibility on twitter
[40] A Friggeri et al, Rumor cascades
[41] Dataset link: https://www. kaggle.com/competitions/ fake-news/data