

Fake News Detection & Sentiment Analysis on Twitter Data Using NLP

Pooja G.Dolai¹, Akansha T.Hole², Prof.Dnyaneshwar V.Kudande³

¹Pooja G.Dolai, BE Computer Engineering, SRTTC college, Maharashtra, India

²Akansha T.Hole, BE Computer Engineering, SRTTC college, Maharashtra, India

³Prof. Dnyaneshwar V. Kudande, Dept. of Computer Engineering, SRTTC college, Maharashtra, India

Abstract - Messages posted to online social networks (OSN) have recently caused a stir due to the deliberate spread of fake news or rumour. The goal of this research is to understand and analyse the characteristics of fake news, particularly in relation to sentiments, in order to automate the detection of fake news and rumours. We offer a notion that there is a relation between bogus communications or rumours and the sentiments of texts submitted online, based on actual evidence.. We validate our theory by comparing it to cutting-edge baseline text-only fake news detection methods that do not consider sentiments. We ran tests on a standard Twitter fake news dataset and found significant improvements.

Key Words: Fake News Detection, Machine Learning, Natural Language Processing, Sentiment Analysis, Twitter Data.

1. INTRODUCTION

As one of the primary platforms for disseminating news, social media has largely replaced traditional media. This is due to two factors: first, it is less expensive to get news from social media; and second, it is easier to share, comment, and discuss with other readers on social media. According to a survey conducted by [1,] 62 percent of U.S. adults consume news on social media, up from 49 percent in 2012. As the number of people who use social media grows, news on social media spreads more quickly and easily than traditional news sources. However, not all of the information shared on social media is correct, and some of it came from an untrustworthy source.

A large amount of content shared by people contributes to various public opinions, which can sometimes have a significant impact on the way people think in society, even if it is distorted information from fake news created on purpose with the wrong commercial and political intent. As a result of the rapid advancement of various media and communication technology, detecting fake news has recently become a very important and challenging issue. We are interested in a fake detection model for determining the truth of a question from a Korean article using sentence matching based on key sentence retrieval in this paper. Sentence matching is a fundamental natural language processing (NLP) technique. Deep learning research has recently been energised by advances in hardware such as graphics processing units. Various attempts at sentence matching have resulted in the development of NLP techniques based on deep learning..

2. LITERATURE REVIEW

The author embraced information mining measure approach and focussed on the periods of information understanding pre-processing/transformation, information displaying and assessment. They started by looking at the fake news net storehouse, which has only four attributes: id, URL, Title, and tweet ids (for example a rundown of tweet ids of tweets sharing the news what isolated by tabs) [1] By utilising the relevant Twitter API for data slithering, all four qualities can be derived from the properties of a Tweet. The author added more Twitter credits because we believe the dataset can be improved by removing and re-adding more attributions related to tweet ids into the joined dataset. Some of the attributes that can be recovered include a complete check of tweet ids, a total top choice (t fav) tally of the tweet ids, and a total retweet (t retweet) tally of the tweet ids. During the pre- preparing and change measure, we built up an upgraded the dataset to catch Tweets' credits [1]. Proposed Framework :- In this proposed system, they are developing the current writing by presenting troupe procedures with different semantic capabilities to order news stories from various areas as true or fake ensemble strategies alongside Linguistic Inquiry and Word Count (LIWC) set utilized in this examination are the oddity of our proposed approach. [2] Algorithms :- Author used the following learning algorithms in conjunction with our proposed methodology to evaluate the performance : Logistic Regression, Support Vector Machine, Multilayer Perceptron, K-Nearest Neighbors. [2] Due to the multi- dimensional nature of fake news, the recognizing the classification of information isn't so natural. To properly address the issue, a practical approach must clearly include a variety of perspectives. As a result, the proposed methodology combines the Nave Bayes classifier, Support Vector Machines, and semantic analysis. The proposed method is entirely based on Artificial Intelligence moves close, which are required to accurately distinguish between the genuine and the fake, rather than using estimations that cannot reflect abstract limits.. The three-part technique combines Machine Learning figures that are divided into managed learning methodology and traditional language preparation Techniques. (3rd) THE FRAMEWORK'S FIRST FEATURE – NEWS GATHERING The author gathered random news from various articles on various subjects to prepare our prototype. The Accurate and Best performing model is chosen for our forecasts. The pre-marked information that we gathered is organised in a reliable resource like

HANDLING Framework will examine complex news which can be hard for customary model. Dealing with mind-boggling news necessitates the following technological advancements: Tokenization, padding, encoding, embedding grid development, model formation, model training, and forecasting the model are all steps in the process. [4] Formal paraphrase formalised Data collection, data pre-processing, result ordering, and result analysis are all significant advancements in the proposed system. To begin, we take key phrases from the news event and use them as information that the individual must validate. The Twitter streaming API is then used to obtain live data. The filtered data is saved as part of the data set (mongo db).

[5] Formal paraphrase Formal phrasing The unit in charge of data preprocessing is in charge of preparing data for further processing. The arrangement will be based on various news highlights, Twitter surveys such as Sentiment Score, Number of Tweets, Number of Followers, Number of Hashtags, Is Confirmed User, Number of Retweets, and Natural Language Processing (NLP) methods. We are going to describe fake news detection method based on one artificial intelligence algorithm –Naïve Bayes .[5]Text Vectorization and NLTK will be used to compute the TClassifier Sentiment Score (Natural Language Tool compartment). By analysing the effects of arrangement and examination, we can determine whether a piece of news is fake or real. [6] Formalized paraphrase Paraphrase that has been formalised Using the two aforementioned algorithms, the Nave Bayes Classifier and the Support Vector Machine, The following accuracy has been attained.[6]The maximum accuracy of 83 percent on the given training Set was attained by using Naïve Bayes classifier with lidstone Smoothing. Whereas in the previous models which consisted Of only Naïve Bayes (without lidstone smoothing) attained an Accuracy of 74 percent [6]Overview of our method :- High performing NLI models are independently trained and composite with a fine- tuned BERT model to determine soft labels, which are then used to fine-tune the original NLI models, BERT, and the Decomposable [7]Attention model. These are then compoand combined with predictions made via observing transitivity relations After the text edit has been completed, the paper is up for it.

3.SYSTEM OVERVIEW

3.1 PROPOSED SYSTEM

In Existing System, to analyses the behavior of news required maximum resources. To analyze the fake news, we required man power to deep down into it and check the authentication of news. We have to check all possible connection with news manually. It is time consuming and costly approach. Limitation of existing system:-

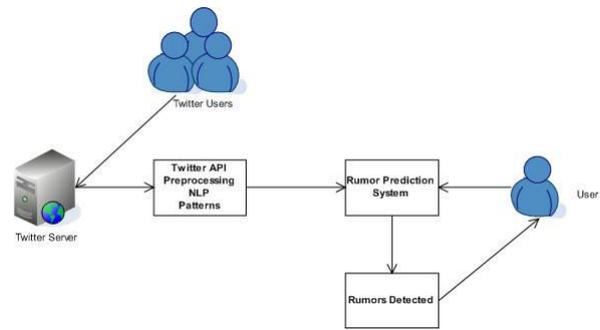


Fig -1: SYSTEM ARCHITECTURE

3.2 EXISTING SYSTEM

In the proposed system, we will fetch tweets from twitter using twitter API based on the query. The collected tweets will be subjected to pre-processing. We will then apply the different patterns and strategic algorithms including some of machine learning algorithms for NLP to supervise the data. The results of the algorithms i.e. the sentiment and influence will be signified in graphical manner (pie charts/bar charts). The proposed system is more essential than the existing one. This is because we will be able to know how the statistics resolved from the representation of the result can have an impact in a particular field as well as influence of negativity spread by rumors.

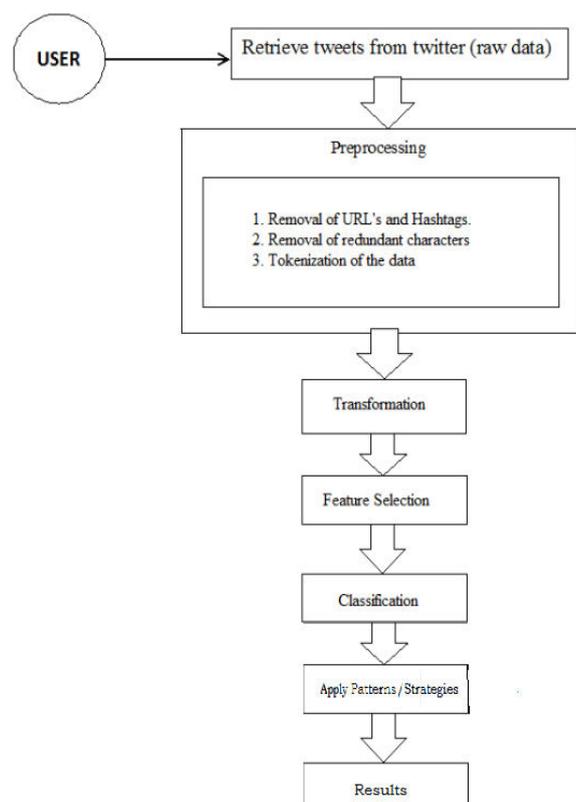


Fig -1: SYSTEM FLOW DIAGRAM

4. MATHEMATICAL ORDER

Let 'S' be the system

Where,

$S = \{I, O, P, Fs, Ss\}$

Where,

I = Set of input

O = Set of output

P = Set of technical processes

Fs = Set of Failure state

Ss = Set of Success state

Identify the input data **I1, I2, , In**

I = {(Twitter Data)}

Identify the output applications as **O1, O2,,On**

O = {(Rumors Detection, Fake News Detection, Sentiment Detection)}

Identify the Process as **P**

P = {(Data Processing, Natural Language Processing, Sentiment Analysis, Pattern Recognition)}

Identify the Failure state as **Fs**

Fs = {(If fake news not predicted)}

Identify the Success state as **Ss**

P = {(Fake news detected successfully)}

5. DISCUSSION

In proposed system we have created one web based application using Python's Flask framework which is light weight. In proposed application we are fetching real time tweets from twitter data and applying algorithm on them to get result out of that. To access data from twitter, you need to have authenticate twitter developer account which allows you to access the data. After accessing the data we are also storing that data into SQL database. Then we applied algorithms on that data. For sentiment analysis, we are using text blob and NLTK libraries. And for fake news detection, we have used TFIDF algorithm. It's taking approximately two to five minutes for execution.

5. CONCLUSIONS

The project embark to solve a real problem of sentiment analysis and genuinely check of Twitter posts. We proposed a technique using knowledge base patterns, strategies and machine learning approaches. These techniques are proposed to increase the perfection of sentiment check for tweets. Patterns can be used to assess of sentiment check for tweets. Patterns can be used to assess if the tweets was a influenced rumor or a original post by any user. By using API of twitter it is feasible to work on live tweets than to work on offline data. Querying and appealing of specific tweets from twitter is possible by using its API. Finding influence or hostility spread by users can be useful in various analytical tasks.

6. RESULT

In proposed system we have created one web based application using Python's Flask framework which is light

weight. In proposed application we are fetching real time tweets from twitter data and applying algorithm on them to get result out of that. To access data from twitter, you need to have authenticate twitter developer account which allows you to access the data. After accessing the data we are also storing that data into SQL database. Then we applied algorithms on that data. For sentiment analysis, we are using text blob and NLTK libraries. And for fake news detection, we have used TFIDF algorithm. It's taking approximately two to five minutes for execution.

According to network complexity it fetch tweets from server and process them. Following screenshot shows the final output of project which has interpreted COVID19 keyword on twitter and which is accurate and efficient..

Image 1: LOGIN PAGE



Image 2: SEARCH FOR TWEETS USING KEYWORDS

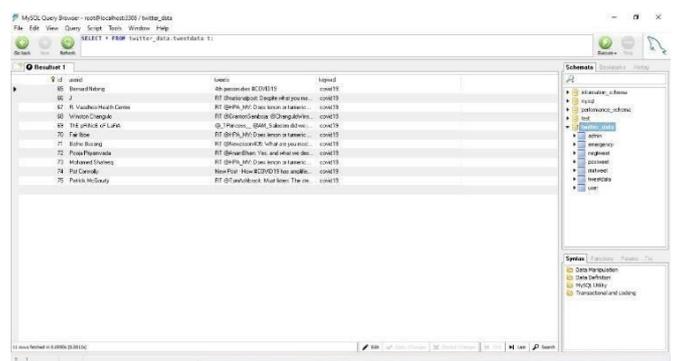


Image 3: DATASET BASED ON SEARCHED KEYWORDS

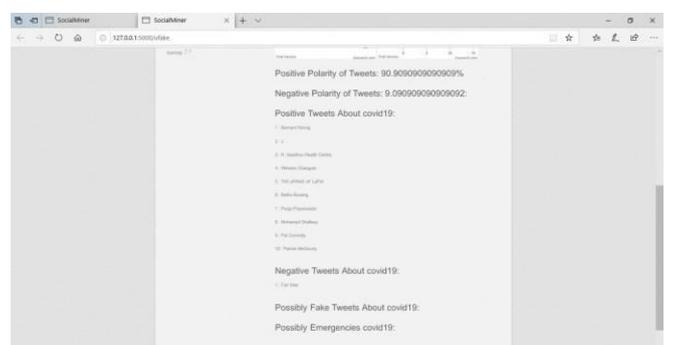


Image 4: NUMBER OF TWEETS ON GIVEN DATA IS POSITIVE OR NEGATIVE



Image 5:FINAL RESULT DISPLAYED

ACKNOWLEDGEMENT

The authors would like to thank the college authorities for their cordial support.

REFERENCES

- [1] 1. Hui Na Chua , Ning Xin Nyow “Detecting Fake News with Tweets’ Properties”978-1-7281-3306-5/19/\$31.00 ©2019 IEEE
- [2] Iftikhar Ahmad , Muhammad Yousaf, Suhail Yousaf , and Muhammad Ovais Ahmad R. Nicole,” Fake News Detection Using Machine Learning Ensemble Methods” 2020
- [3] Harsh Katter”A Smart System For Fake News Detection Using Machine Learning”2019 Sept
- [4] Nikhil Sharma "Fake News Detection using Machine Learning" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456- 6470, Volume-4 | Issue-4, June 2020, pp.1317-1320
- [5] Subhadra Gurav, Swati Sase, Supriya Shinde, Prachi Wabale, Sumit Hirve” Survey on Automated System for Fake News Detection using NLP & Machine Learning Approach”Ijret Jan 2019
- [6] Kushal Agarwalla, Shubham Nandan, Varun Anil Nair,D. Deva Hema “Fake News Detection using Machine Learning and Natural Language Processing” International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7,Issue-6, March 2019
- [7] Timothy Niven, Kai-Chou Yang, Hung-Yu Kao” Fake News Detection as Natural Language Inference” WSDM ’19, February 11–15, 2019, Melbourne,Australia
- [8] Yendhe Rutuja , Kasturi Anusha”Fake News Detection And Sentiment Analysis” September 2020 ISSN (Online) 2456-0774