Fake News Detection & Sentiment Analysis on Twitter Data Using NLP

Prof .Dnyaneshwar Kudande[1], Pooja Dolai[2], Akansha Hole[3]

[1]Assistant Professor, Dept. of Computer Engineering, SRTTC FOE, MH, India
[2]Student, Dept. of Computer Engineering, SRTTC FOE, MH, India
[3] Student, Dept. of Computer Engineering, SRTTC FOE, MH, India
(Savitribai Phule Pune University)

---***---

Abstract - *Messages posted to online social networks (OSN) causes a recent stir because of the intended spread of faux news or rumor. This work aims to know and analyses the characteristics of faux news especially in reference to sentiments, for the automated detection of faux news and rumors. supported empirical observations, we propose a theory that there exists a relation between fake messages or rumors and sentiments of the texts posted online. We verify our theory by comparing with the state-of-the-art baseline text-only fake news detection methods that don't consider sentiments. We performed assessments on standard Twitter fake news dataset and show good improvements in detecting fake news or rumor posts*

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

1.INTRODUCTION

Social media has replaced the traditional media and become one among the main platforms for spreading news , The reasons for this replacement are due to: i) less expensive to get news from social media; and ii) easier to share, comment and discuss with other readers on social media. According to a survey conducted by [1], 62% of U.S. adults consume news on social media while in 2012, only 49% of consumers get news from social media. As there is a significant increase in social media users, news on social media tends to travel faster and easier than traditional news sources. However, not all the news disseminated on social media are accurate and some of them came from the unverified source.

A huge amount of contents shared by people make up various public opinions which sometimes greatly influence a way of thinking of the people in the society even though they are distorted information from fake news created on purpose with wrong commercial and political intent. Therefore, fake news detection becomes very important and challenging issue recently with the fast advancement of various media and communication technology. In this paper, we are concerned with a fake detection model for finding the truth of the question from a Korean article using sentence matching based on key sentence retrieval. Sentence matching is a fundamental technique of the natural language processing(NLP) which checks whether two sentences are similar or not semantically. Recently, deep learning research has been activated by the advance of hardware such as graphics processing unit. NLP techniques based on deep learning have been developed through various attempts for sentence matching.

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. For information understanding, they firstly analysed the fake news net storehouse that just have 4 attributes which are id, URL, Title, and tweet_ids (for example a rundown of tweet ids of tweets sharing the news what isolated by tabs).[1] All these four qualities can be gotten coordinated from a Tweet's properties utilizing pertinent Twitter API for information slithering. The author further crept more credits from Twitter as we think about the dataset can be improved by removing more ascribes related to the tweet ids and add them into the joined dataset. A few attributes that can be recovered are a complete check of tweet_ids, all out top choice (t_fav) tally of the tweet_ids, and aggregate retweet (t_retweet) 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) highlight 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 of fake news detection classifiers.[2].Algorithms as follow :Logistics 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. Clearly a practical method should contain a couple of viewpoints to correctly deal with the issue. This is the explanation the proposed methodology is a mixture of Naïve Bayes classifier, Support Vector Machines, and semantic examination. The proposed technique is totally made out of Artificial Intelligence moves close, which is essential to accurately arrange between the veritable or the fake, as opposed to using estimations that can't reflect abstract limits. The three-segment technique is a mix between Machine Learning figurings that partition into managed learning methodology, and the characteristic language preparing Techniques.[3]Framework FEATURE 1 – NEWS GATHERING The author assembled random news on different articles with various subjects to prepare our model. By considering these, System distinguishes news goal utilizing AI calculation. Pre Labeled news are utilized to prepare our models. The Accurate and Best performing model is chosen for our forecasts. The pre marked information that we gathered is structure a reliable resource like Kaggle. [4]Framework FEATURE 2 - COMPLEX NEWS HANDLING Framework will examine complex news which can be hard for customary model. Following advances are needed for dealing with of the mind boggling news, which are as per the following Tokenising, padding, encoding, Embedding grid development, Model Formation, Model Training lastly foreseeing the model.[4]The proposed system comprises of four significant advances: data assortment, data pre processing,order and analysis of results. We first take key expressions of the news occasion as an information that the individual need to validate. After that live information is gathered from twitter streaming api. The filtered information is put away in the data set (mongo db).[5] The information preprocessing unit is answerable for setting up an information for additional handling. Arrangement will be founded on different news highlights, twitter surveys like Sentiment Score, Number of Tweets ,Number

of devotees ,Number of hashtags ,is confirmed User ,Number of retweets and NLP methods. We are going to describe fake news detection method based on one artificial intelligence algorithm –Naïve Bayes .[5]Classifier Sentiment Score will be determined utilizing Text Vectorization calculation and NLTK(Natural Language Tool compartment). By doing the assessment of impacts obtained from arrangement and examination, we can choose the portion of news being fake or real. [6]Using the above-mentioned algorithms, i.e. Naïve Bayes Classifier, Support Vector Machine and Logistic Regression, 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. METHODOLOGY

The basic idea of our project is to build a model that can predict the likely of real time news events.
As shown in Fig. 1,
The proposed framework comprise of four major steps: Data collection, Data preprocessing, Classification and Analysis of results.

3.1 EXISTING 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:-
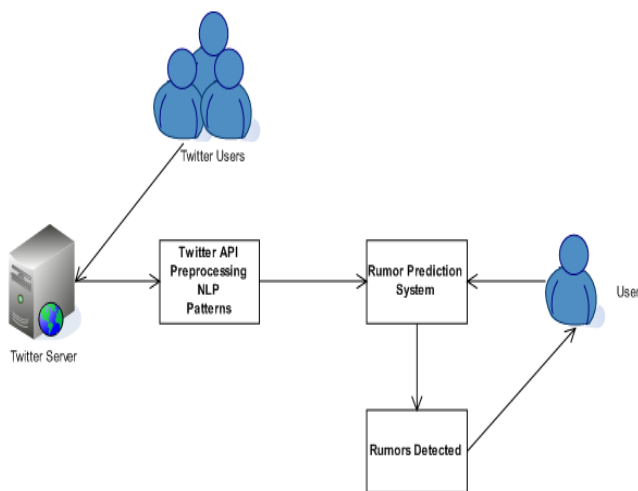
Figure no:-1

3.2 PROPOSED 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.
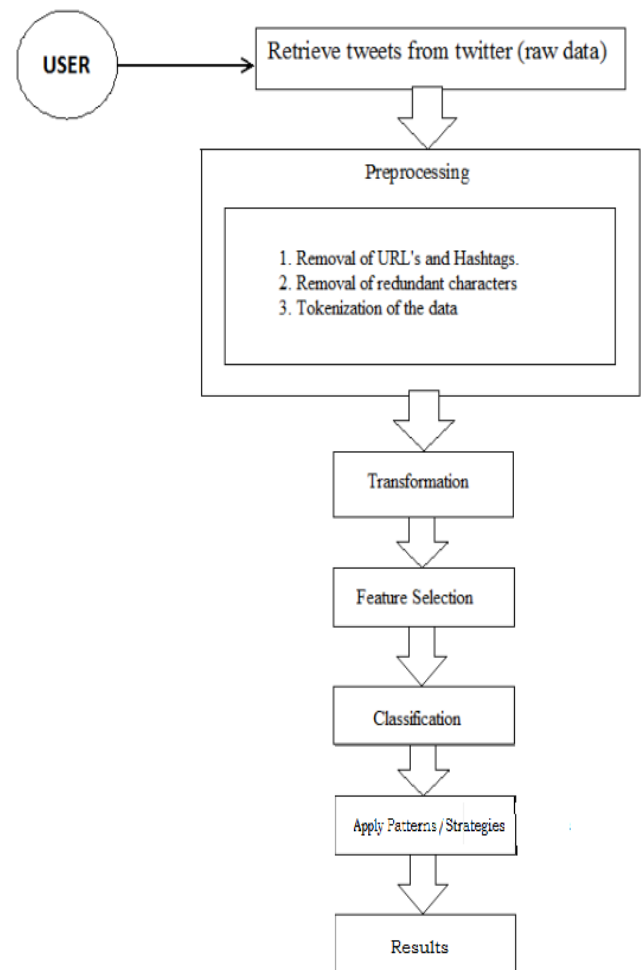

Figure no :-2 Figure no 2 represents the methodology for proposed system.

4.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.CONCLUSION

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 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.

REFERENCES

[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