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

FAKE NEWS DETECTION

Ishan Upadhyay[1], Harmeet Singh[2]

[1-4]*Student, Dept. of Computer Science and Engineering, MIT School of Engineering, MIT ADT University, Pune Maharashtra, India*

Abstract - *In previous years, we have seen the prosper development of fake news for commercial and political purposes in great numbers and is spread in the online world. By using misleading words, online social network users are infected or get caught in the scandal of fake news easily which has caused enormous effects on the offline community. So, it's really important to improve the responsible and dependable data in online community networks that check the fake news appropriately.*

Thus, it is very important to identify the fake news. There are different machine learning techniques that help to solve the above problem of fake news detection.

In this research paper, we have proposed the system which detects the fake news using different NLP APIs from TensorFlow and we have implemented it on different datasets and the system gave positive results. The APIs made the system more efficient and simpler. We have the build a system to detect fake news with a 90% accuracy approximately.

Key Words: **Fake News, Machine Learning, NLP, APIs, TensorFlow.**

1. INTRODUCTION

Fake news could be a type of news that consists of sensitive false information or scam. Since the ideology has existed even before the popularity of the Internet as different publishers that implemented false and deceptive information for their advantage.

Succeeding the arrival of the web, more and more users began abandoning the traditional news channel used to propagate information for the online world. The problem is not only that the alternative medium allows consumers to make a variety of proclamations, but it is also very convenient as running fingers on the screen. The outrage started when the authors started using the news thumbnails for click-baiting. Click baits are expressions that are used to gain attention and after getting caught by the click baiter and after seeing the news their expectations are down. Their only aim is to get to their sites for a really short time.

In this project, we have used different news datasets to build a fakes news detector. We have used natural language processing techniques and machine learning algorithms to classify fake news articles using different NLP APIs from TensorFlow.

2. RELATED WORK

Sarcasm is considered as the most desired form to convey views and feelings on social media platforms. The capacity of different content such as sarcastic or funny content has been increasing day by day and it can be seen on some popular social networking sites such as Twitter, Facebook, etc. A lot of different models have been made to detect sarcasm and side by side research is also going on.

Ellen et al reviewed Sarcasm as a dissimilarity between positive and negative situations. This model uses different algorithms such as bootstrapping for the creation of positive and negative phases. For the training purpose, different tweets were considered which contain #sarcasm, and later it is trained on different algorithms such as SVM and Naive Bayes. [1]

Joshi et al considered a model that was based on the direct and indirect inappropriateness of sentiments which were revealed through tweets.

From the various tweets, the text was split into various grams to detect sarcasm, and the quality of the grams was tested based on the existing collection of the positive and negative words. The model worked using different features of the tweet such that it follows SVM rules. It helped them to beat the present system and the performance was also improved by 10%. [2]

The (SCUBA) system aims to point out the task of sarcasm detection on Twitter with the help of the different behavioural features. They classified patterns using the previous tweets and theories from behavioural and psychological studies to build a structured protocol that is used for detecting sarcasm. It gives a set of ideas such as the main forms of sarcasm which contains the established psychological and behavioural studies. After that, it produced computational features that are used to capture different kinds of sarcasm using the current and previous user tweets. Finally, it combines all these features to train Naive Bayes and SVM such that these can train classifiers. [3]

Bamman et al introduce a model for sarcasm detection which was based on the basis of inferability -such that speakers can only use sarcasm if they are sure that it will be understood. Sarcasm may be between persons who know each other. Normally tweet is classified as sarcastic if there is "#sarcasm" and which may contain a minimum of three words. All these tweets are taken and they are used as a

training set. Basically, it contains three types of features that are tweet, author, and audience. [4]

Go, Alec work was based on substantial pre-processing such that data should become logical with the help of the necessary method. After that, the data were categorized into positive, negative and sarcastic. For feature extraction, all these things are used, and later on, it is used to train SVM classifiers. [6]

3. METHODOLOGY

The Fake News Detector which has been implemented using the Machine Learning approach, by employing Natural language processing.

3.1 Fake News Detection:

[1] Data-oriented –

In computing Data orientation means focusing on the data i.e., no hiding. Data is all the facts about it. In Data-oriented programming, you may or may not have to do a lot of data transformations. Your system is inputting, storing, transmitting, and processing data. Data orientation in Fake news detection is Temporal, Psychology, And Dataset.

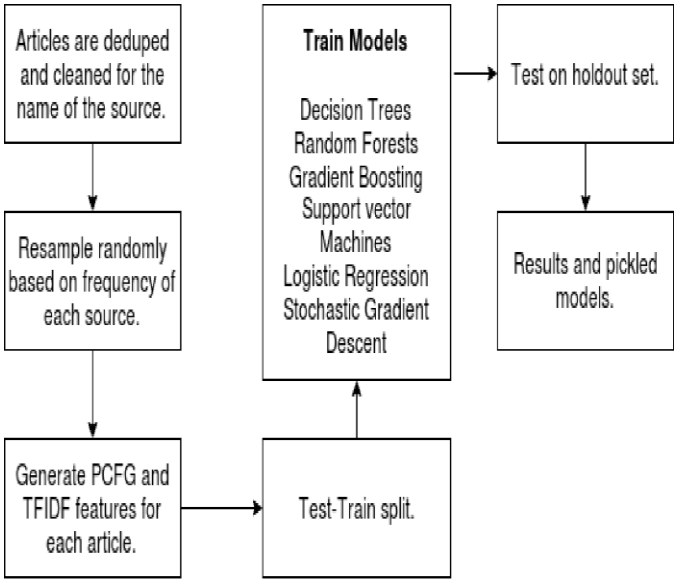


Fig. 1: Pipeline representation.

Fig.1: Pipeline representation

[2] Feature oriented –

In the field of computer programming, feature-oriented programming (FOP) or feature-oriented software development (FOSD) is considered as a prototype for the program generation especially in software product lines (SPLs) and it is also used as incremental growth of programs. News Content, Social Context

[3] Model-oriented-

Model oriented or model-driven approach is a software design approach means how which model or learning you have used to implement your project. Like Semi-Supervised, supervised, Unsupervised learning.

[4] Application-oriented –

Your project must be application-oriented i.e., it is most useful to the outside world. Like our project with some modifications can be implemented for on. Fake news Diffusion and Fake news Intervention.

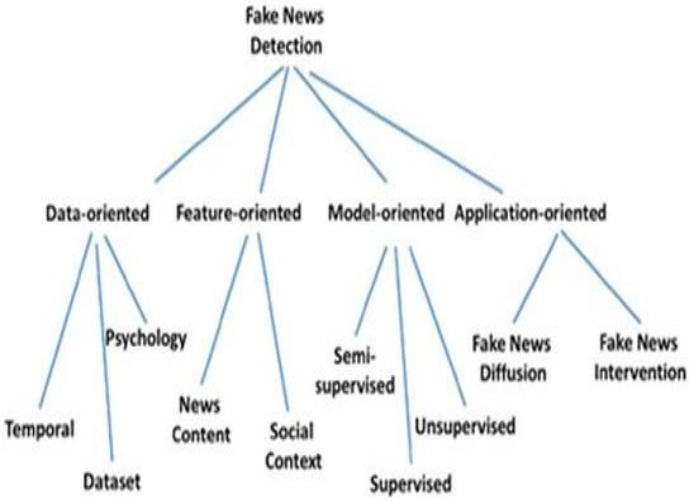


Fig.2: Fake News Detection Types

[5] NLP –

Natural Language Processing is a subpart of AI, which allocates with the interaction between machines and human natural languages. NLP's help the computer to analyse and understand the user's input. Natural language Processing computers can perform various tasks like knowledge extraction, sentiment analysis, speech recognition, fake news detection.

[6] Dataset –

A dataset is considered as a cluster of information. Dataset records the values for every variable, like the height and weight of an object and every member of the information set. The training set to develop this Fake news detector was collected from the data available on Kaggle.

[7] Data Pre-processing –

Data pre-processing is considered as one of the most essential steps in ML. Where the data is transformed into a clear format. Basically, the Real-world data is not complete, it may not be consistent or it may lack specific functioning and it may contain many errors. So, Data pre-processing is used for solving these issues.

3.2 APIs:

[1] Tokenizer by TensorFlow –

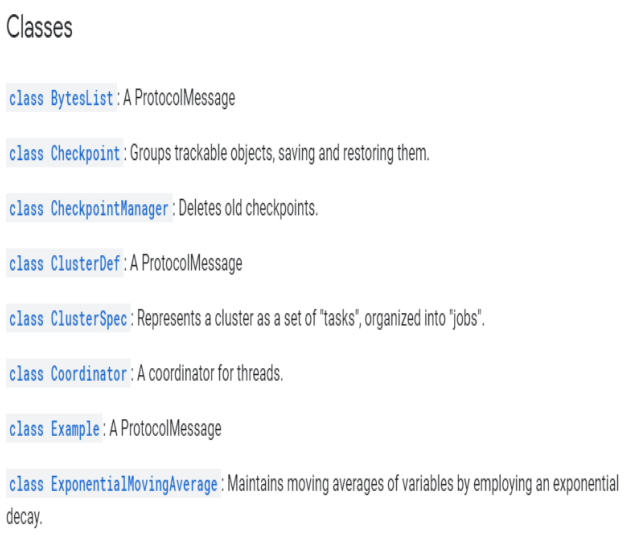
where

Tokenizer allows us to transform a text by changing the text into a sequence of integer where the index of every token will be present in the dictionary or it changes the text into a vector and the coefficient of each token might be binary and it supports word count and it is based on tf-idf.

[2] pad sequence by TensorFlow -

It is the kind of a function which alters a list of sequences into a 2D array. The sequences which are smaller than num_timesteps are filled with values till they become num_timesteps bigger. The sequences which are longer than num_timesteps are shortened such that they can fit in the specified length.

[3] tf.train by TensorFlow –


Classes

class BytesList: A ProtocolMessage

class Checkpoint: Groups trackable objects, saving and restoring them.

class CheckpointManager: Deletes old checkpoints.

class ClusterDef: A ProtocolMessage

class ClusterSpec: Represents a cluster as a set of "tasks", organized into "jobs".

class Coordinator: A coordinator for threads.

class Example: A ProtocolMessage

class ExponentialMovingAverage: Maintains moving averages of variables by employing an exponential decay.

Fig.3: Classes of tf. train

4. RESULTS

4.1 Accuracy graph

This is the accuracy graph of the model, where it is showing the accuracy of the test and train data.

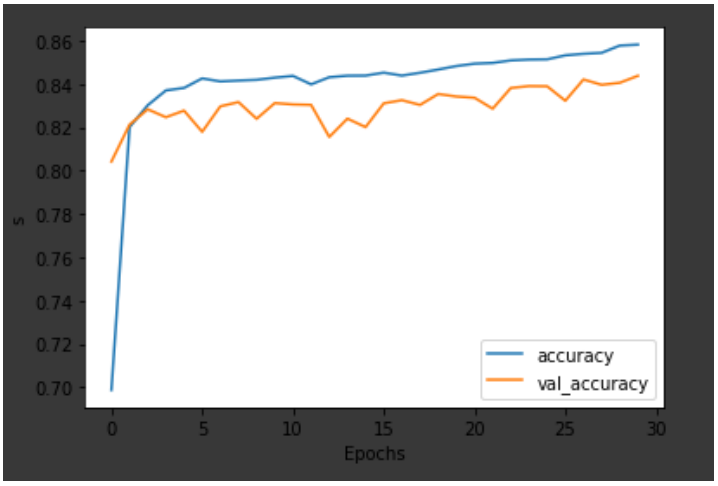


Fig.4: Accuracy graph

4.2 Loss graph

This is the loss graph of the model, where it is showing the accuracy of the test and train data.

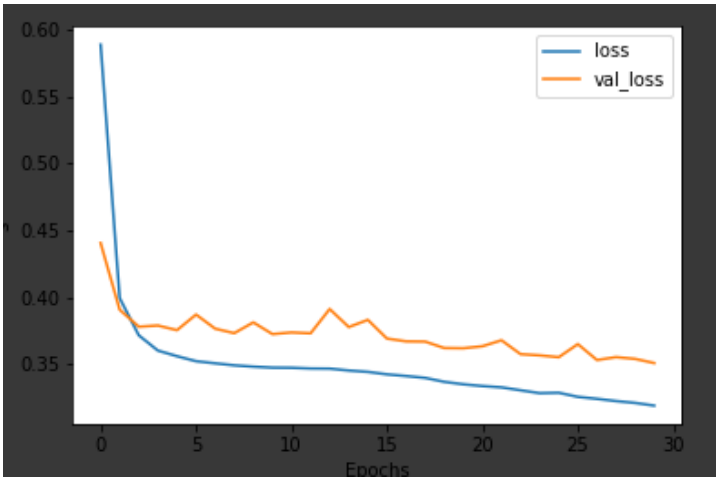


Fig.4: Loss graph

5. FUTURE WORK

The future scope of this project is that fake news detectors can help to filter different websites that contain fake news and the motive is to help users such that they can't get attracted by clickbait. The project can also be used on many social media platforms where there is a massive amount of fake data which can cause damage to the society, with some modifications to remove the same.

6. CONCLUSION

The primary objective of the research was to develop a Fake news detector using ML. The training set to initiate this Fake news detector was devised from the data available on Kaggle. In this system, we have used different APIs of NLP which helped us to obtain the result.

7. REFRENCES

[1] Riloff, Ellen, et al. "Sarcasm as Contrast between a Positive Sentiment and Negative Situation." EMNLP. Vol. 13. 2013.

[2] Joshi, Aditya, Vinita Sharma, and Pushpak Bhattacharyya. "Harnessing Context Incongruity for Sarcasm Detection." ACL (2). 2015.

[3] Rajadesingan, Ashwin, Reza Zafarani, and Huan Liu. "Sarcasm detection on twitter: A behavioral modeling approach." Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. ACM, 2015.

[4] Bamman, David, and Noah A. Smith. "Contextualized Sarcasm Detection on Twitter." ICWSM. 2015.

[5] Maynard, Diana, and Mark A. Greenwood. "Who cares about Sarcastic Tweets? Investigating the Impact of Sarcasm on Sentiment Analysis." LREC. 2014.

[6] Go, Alec, Richa Bhayani, and Lei Huang. "Twitter sentiment classification using distant supervision." CS224N Project Report, Stanford 1.12 (2009).

[7] Neri, Federico, et al. "Sentiment analysis on social media." Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012). IEEE Computer Society, 2012.

[8] Zhao, Jichang, Li Dong, Junjie Wu, and Ke Xu. "Moodlens: an emoticon-based sentiment analysis system for Chinese tweets." In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1528-1531. ACM,2012.

[9] Kouloumpis, Efthymios, Theresa Wilson, and Johanna D. Moore. "Twitter sentiment analysis: The good, the bad, and the omg!." Icwsm 11, no. 538-541 (2011): 164.

[10] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, pp. 1–4, 2015.