FAKE NEWS DETECTION USING MACHINE LEARNING EMPLOYING RANDOM FOREST ALGORITHM

Dr.M.Rajeswari,

Assistant professor,

Department of B.com (Business Analytics),

PSGR Krishnammal College for women Coimbatore, India.

rajeswarim@psgrkcw.ac.in

K.Dhana Priya,

UG Scholar,

Department of B.com (Business Analytics),

PSGR Krishnammal College for women Coimbatore, India.

dhanapriya1107@gmail.com

ABSTRACT

The advent of World Wide Web and therefore the rapid adoption of social media platforms paved the way for information dissemination. Nowadays lots of data is being shared over social media and that we aren't able to differentiate between which information is fake and which is real. People immediately start expressing their concern or sharing their opinion as soon as they are available across a post, without verifying its authenticity. This further ends up in spreading of it. Fake news and rumours are the foremost popular styles of false and unauthenticated information and may be detected as soon as possible for avoiding their dramatic consequences.

Keywords: dissemination, fake news, social media

1. INTRODUCTION

World is changing rapidly.no doubt we've variety of benefits of this digital world but it's some disadvantages too. There are different issues associated with digital world.one among its fake news.one can easily spread a fake news without knowing its impact on the society. There are different online platforms to spread fake news .it includes Facebook, twitter etc., Online platforms are useful to the users to access the news easily but the matter is it gives an wide opportunity to the cyber criminals to spread a fake news through these platforms easily. Readers read the news and begin believing without verifying it. Detecting the fake news could be a big challenge because it's a difficult task. If the fake news isn't deducted earlier it can create a large impact within the society. Individuals, organizations, or political parties will be affected through these fake news. Different researchers are working for the deduction of faux news. The employment of machine learning is providing helpful during this regard. Machine learning algorithms will detect the fake news automatically once they're trained. There are different algorithms to detect fake news. Random forest is one amongst the algorithms which is employed to detect fake news.it is a supervised machine learning algorithm that's widely utilized in classification and regression problems. The papers aim is to gather the fake news dataset, apply random forest algorithm and find its accuracy rate in deducting fake news.

II OBJECTIVE

The main objective of this paper is to detect fake news with the dataset by using the random forest classifier algorithm and then predicting the accuracy level. The random forest algorithm is used in this paper to analyse and visualize the values of the predicting data.

III RELEATED WORKS

Social networks are stepping up in using digital fake news detection tools and educating the public towards spotting fake news. At the time of writing, Facebook uses machine learning algorithms to identify false or sensational claims used in advertising for alternative cures, they place potential fake news articles lower in the news feed, and they provide users with tips on how to identify fake news themselves[1]. Some approaches detect fake news by using metadata such as a comparison of release time of the article and timelines of spreading the article as well where the story spread [2].The spreading of false political information have increased due to the emergence of streamline media environments. In a recent study it was found that 43% (13 of 30) false news stories were shared on social media platforms, like Twitter, with links to non-credible news websites [3]

There has been a rapid increase in the spread of fake news in the last decade, most prominently observed in the 2016 US elections [5]. Such proliferation of sharing articles online that do not conform to facts has led to many problems not just limited to politics but covering various other domains such as sports, health, and also science [3]. One such area affected by fake news is the financial markets [6], where a rumor can have disastrous consequences and may bring the market to a halt. A number of studies have primarily focused on detection and classification of fake news on social media platforms such as Facebook and Twitter [6, 7].

At conceptual level, fake news has been classified into different types; the knowledge is then expanded to generalize machine learning (ML) models for multiple domains [4, 8, 9]. The study by Ahmed et al. [10] included extracting linguistic features such as n-grams from textual articles and training multiple ML models including K-nearest neighbour (KNN), support vector machine (SVM), logistic regression (LR), linear support vector machine (LSVM), decision tree (DT), and stochastic gradient descent (SGD), achieving the highest accuracy (92%) with SVM and logistic regression. According to the research, as the number of increased in -grams calculated for a particular article, the overall accuracy decreased. The phenomenon has been observed for learning models that are used for classification.

Shu et al, He achieved better accuracies with different models by combining textual features with auxiliary information such as user social engagements on social media. The authors also discussed the social and psychological theories and how they can be used to detect false information online. Further, the authors discussed different data mining algorithms for model constructions and techniques shared for features extraction. These models are based on knowledge such as writing style, and social context such as stance and propagation. Different researchers are working for the detection of fake news. The use of Machine learning is proving helpful in this regard. Researchers are using different algorithms to detect the false news. Researchers in (Wang, 2017) said that fake news detection is big challenge. They have used the machine learning for detecting fake news

IV METHODOLOGY

The method used here is random forest algorithm. It provides high accuracy for large proportion of data.it is a popular machine learning that belongs to the supervised learning technique made up of decision tree. The main aim of this paper is to predict the accuracy level.

A. PROPOSED SYSTEM

STEP 1: Import the dataset from kaggle dataset, modify the dataset and save in Excel.csv.fornat.
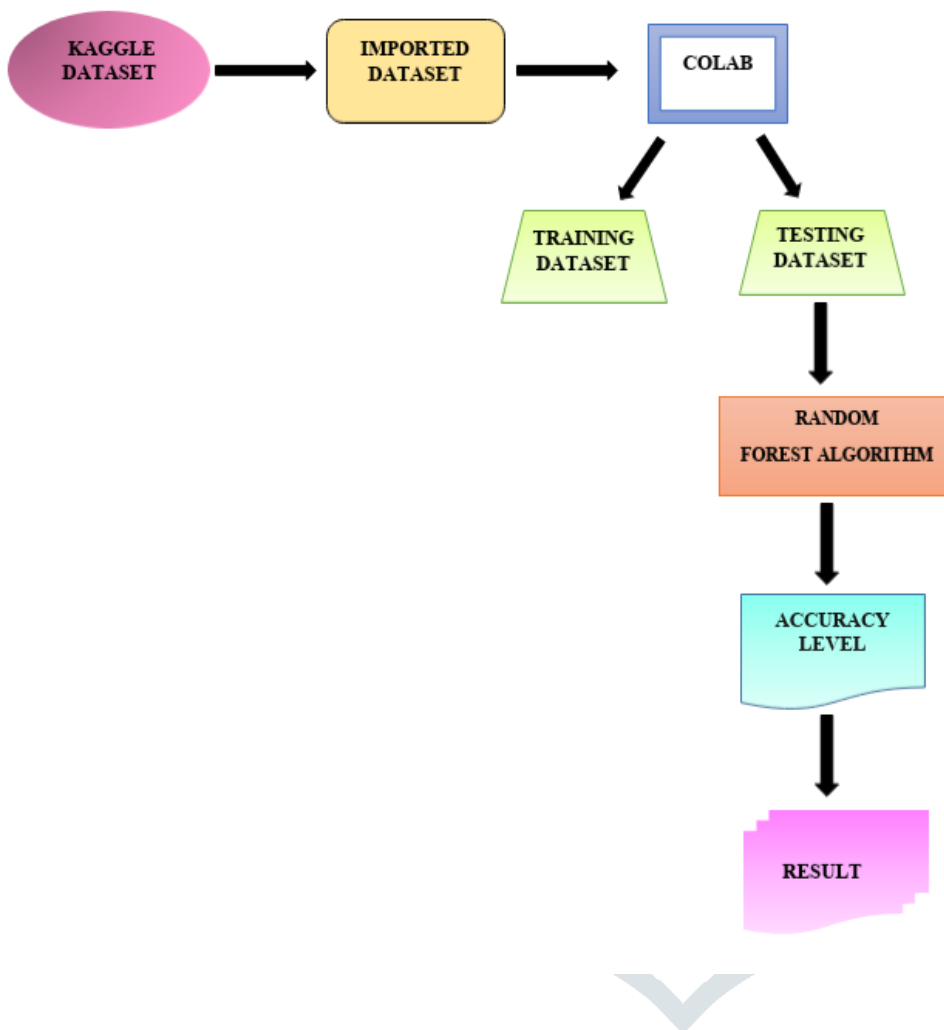
STEP 2: Use Google colab for executing python coding and removed all unwanted data from dataset.

STEP 3: Then dataset is separated into training dataset and testing dataset.

STEP 4: Visualization are made in Google colab for better understanding of dataset.

STEP 5: Finding accuracy with Confusion matrix, without normalization.

A.WORKFLOW

B. RANDOM FOREST ALGORITHM

Random forests or random decision forests is an ensemble learning methods for classification, regression and other tasks that operates by constructing a multitude of decision tree at training time and outputting the class that is the mode of the classification or mean/average prediction. It is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

V IMPLEMENTATION AND RESULT

```
from sklearn.ensemble import RandomForestClassifier

pipe = Pipeline([('vect', CountVectorizer()),
                 ('tfidf', TfidfTransformer()),
                 ('model', RandomForestClassifier(n_estimators=50, criterion="entropy"))])

model = pipe.fit(X_train, y_train)
prediction = model.predict(X_test)
print("accuracy: {}%".format(round(accuracy_score(y_test, prediction)*100,2)))
dct['Random Forest'] = round(accuracy_score(y_test, prediction)*100,2)
```

accuracy: 98.61%

Fig5.1

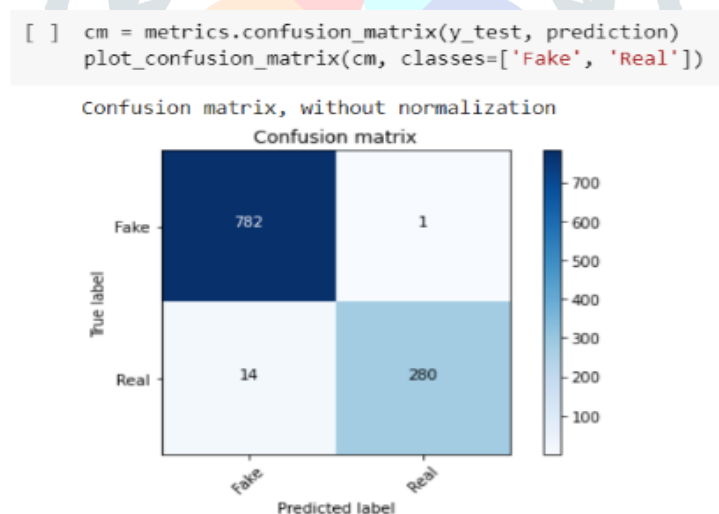The fig5.1 shows the accuracy level of fake news deduction using random forest algorithm.

```
[ ] cm = metrics.confusion_matrix(y_test, prediction)
    plot_confusion_matrix(cm, classes=['Fake', 'Real'])
```

Confusion matrix, without normalization

Fig5.2

The fig5.2 shows the visualization of fake and real news in confusion matrix.

VI CONCULISION AND FURTHER WORK

In this paper, importing data set, executing coding and visualization are done in Google colab. Random forest algorithm is used to predict the accuracy level. Fake news play a vital role in decision of individuals. Spread of fake news should be considered as a serious issue and proper steps should be taken to control them. In controlling spread of fake news machine learning helps to detect the news easier. Random forest algorithm has been used in this paper. In further works other algorithms can be used in

different tools and accuracy rate can be predicted to find which tool and algorithm has high accuracy in predicting the fake news.

REFERENCES

[1] Agarwal, Arush, and Akhil Dixit. "Fake News Detection: An Ensemble Learning Approach." 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2020.

[2]Ahmed, Hadeer, Issa Traore, Sherif Saad. "Detection of online fake news using n-gram analysis and machine learning techniques." International conference on intelligent, secure, dependable systems in distributed & cloud environments. Springer, Cham, 2017

[3]Kanoh, H.: Why do people believe in fake news over the Internet? An understanding from the perspective of existence of the habit of eating and drinking. Proc. Comput. Sci. 126, 1704–1709 (2018).

[4]N. K. 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.

[5]K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.

[6]S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146–1151, 2018.

[7]H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," Journal of Economic Perspectives, vol. 31, no. 2, pp. 211–236, 2017.

[8] V. L. Rubin, N. Conroy, Y. Chen, and S. Cornwell, "Fake news or truth? using satirical cues to detect potentially misleading news," in Proceedings of the Second Workshop on Computational Approaches to Deception Detection, pp. 7–17, San Diego, CA, USA, 2016.

[9]H. Jwa, D. Oh, K. Park, J. M. Kang, and H. Lim, "exBAKE: automatic fake news detection model based on bidirectional encoder representations from transformers (bert)," Applied Sciences, vol. 9, no. 19, 2019.

[10]H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using n-gram analysis and machine learning techniques," in Proceedings of the International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, pp. 127–138, Springer, Vancouver, Canada, 2017.