FAKE NEWS DETECTION USING NAIVE BAYES ALGORITHM WITH MACHINE LEARNING

Dr.M.Rajeswari[1],
Assistant Professor,
Department of B.Com(Business Analytics),
PSGR Krishnammal College for Women, Coimbatore, India.
rajeshwarim@psgrkcw.ac.in

Rithisha Catharine C[2],
UG Scholar,
Department of B.Com(Business Analytics),
PSGR Krishnammal College for Women, Coimbatore, India.
rithishacatharine@gmail.com

ABSTRACT

It has been called one of the most dangerous developments in modern history. Fake news, made-up stories that have been reported as real events, has become a new form of misinformation. These days a lot of information is being shared over social media and we are not 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 come across a post, without verifying its authenticity. This further results in spreading of it. Fake news leads to convince the reader to believe fake information which proves these articles difficult to read. Fake news detection problem can be solved by the help of Artificial Intelligence algorithms which includes machine learning Machine Learning algorithms.

Keywords: dissemination, Machine learning, fake news, social media.

I. INTRODUCTION

Fake News contains misleading information that could be checked. World is changing rapidly. No doubt we have a number of advantages of this digital world but it has some disadvantages too. There are different issues related to digital world. One among it is 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 problem is it gives an wide opportunity to the cyber criminals to spread a fake news through these platforms easily. Readers read the news and start believing without verifying it. Detecting the fake news is a big challenge because it is a difficult task. If the fake news is not deducted earlier it can create a huge impact in the society. Individuals, organizations, or political parties can be affected through these fake news. Different researchers are working for the deduction of fake news. The use of machine learning is providing helpful in this regard. Machine learning algorithms will detect the fake news automatically once they are trained. There are different algorithms to detect fake news. Naive bayes is one among the algorithms which is used to detect fake news. It is a probabilistic classifier. It is based on probability models that incorporate strong independence assumptions. It is a supervised machine learning algorithm that is widely used in classification and regression problems. The papers aim is to collect the fake news dataset, apply naive bayes algorithm and find its accuracy rate in deducting fake news. The main objective of this paper is to detect fake news with the dataset by using the Naive bayes 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.

II. OBJECTIVE

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

III. RELEATED WORKS

Pointed out various sources of media and made the suitable studies whether the submitted article is reliable or fake. The paper utilizes models based on speech characteristics and predictive models that do not fit with the other current models [5]. used naïve Bayes classifier to detect fake news by Naive Bayes. This method was performed as a software framework and experimented it with various records from the Facebook, etc., resulting in an accuracy of 74%. The paper neglected the punctuation errors, resulting in poor accuracy [3].

Estimated various ML algorithms and made the researches on the percentage of the prediction. The accuracy of various predictive patterns included bounded decision trees, gradient enhancement, and support vector machine were assorted. The patterns are estimated based on an unreliable probability threshold with 85-91% accuracy [1].

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 [4]. utilized the Naive Bayes classifier, discuss how to implement fake news discovery to different social media sites. They used Facebook, Twitter and other social media applications as a data sources for news. Accuracy is very low because the information on this site is not 100% credible [10].

Discuss misleading and discovering rumors in real time. It utilizes a novelty-based characteristic and derives its data source from Kaggle. The accuracy average of this pattern is 74.5%. Clickbait and sources do not consider unreliable, resulting in a lower resolution[2]. Used to distinguish Twitter spam senders. Among the various models used are the naive Bayes algorithms, the clustering, and the decision tree. The accuracy average of detecting spammers is 70% and fraudsters 71.2%. The models used have achieved a low level of intermediate precision to separate spammers from non-spam [9].

Identified fake news in different ways. The accuracy is limited to 76% as a language model. Greater accuracy can be achieved if a predictive model is used [8]. 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[6].

Aimed to utilize machine learning methods to detect fake news. Three common methods are utilized through their researches: Naïve Bayes, Neural Network and Support Vector Machine (SVM). Normalization technique is an essential stage in data cleansing prior machine learning is used to categorizing the data. The output proved that that Naive Bayes has an accuracy of 96.08% for detecting fake messages. Two more advanced methods, the neural network and the machine vector (SVM) reached an accuracy of 99.90%[7].

IV. METHODOLOGY

A. WORKFLOW
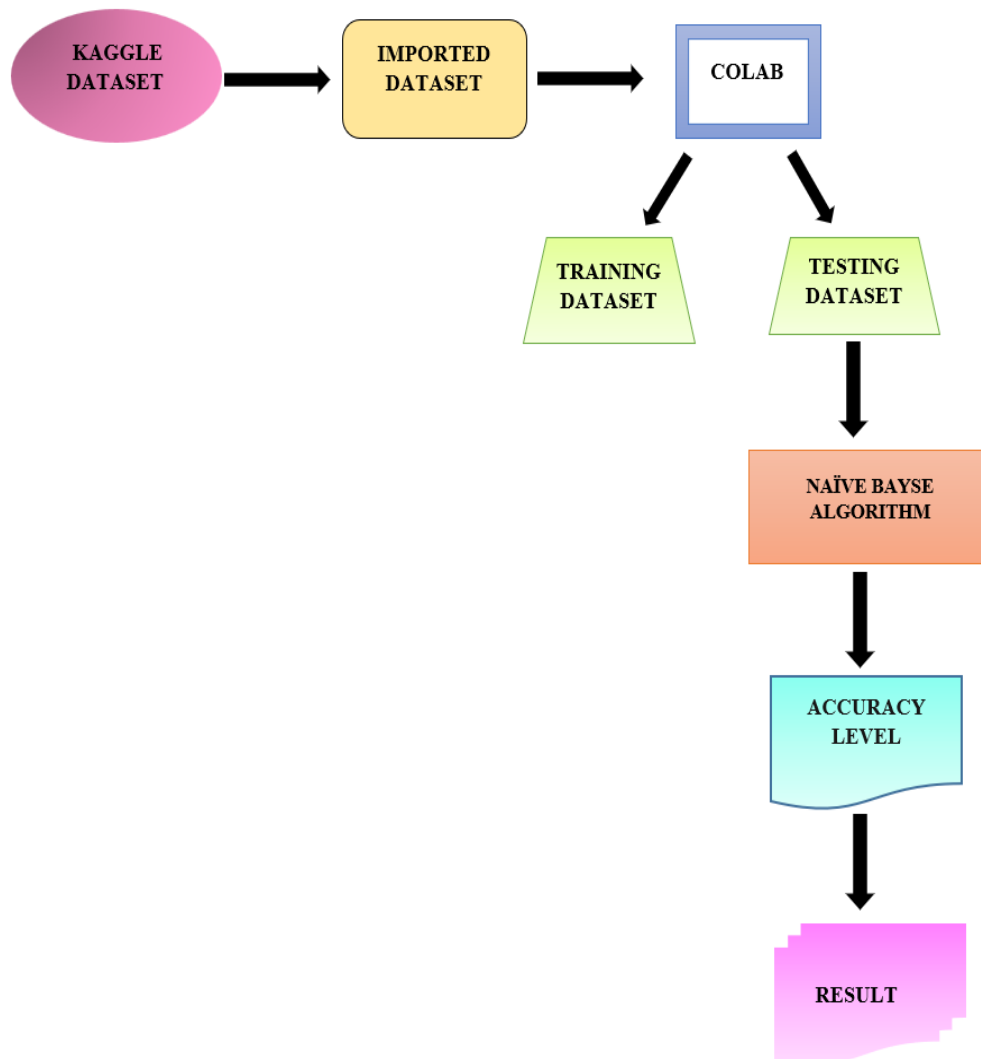
B. PROPOSED SYSTEM

STEP 1: Imported the dataset from Kaggle dataset, modified the dataset and saved in Excel.csv format.

STEP 2: Used 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.

C. NAIVE BAYES ALGORITHM

Naive Bayes This algorithm works on Bayes theory under the assuming that its free from predictors and is used in multiple machine learning problems. Simply put, Naive Bayes assumes that one function in the category has nothing to do with another. For example, the fruit will be classified as an apple when it's of red colour, swirls, and the diameter is close to 3 inches. Regardless of whether these functions depend on each other or on different functions, and even if these functions depend on each other or on other functions, Naive Bayes assumes that all these functions share a separate proof of the apples.

V. IMPLEMENTATION

```
[ ] dct = dict()

    from sklearn.naive_bayes import MultinomialNB

    NB_classifier = MultinomialNB()
    pipe = Pipeline([('vect', CountVectorizer()),
                     ('tfidf', TfidfTransformer()),
                     ('model', NB_classifier)])

    model = pipe.fit(X_train, y_train)
    prediction = model.predict(X_test)
    print("accuracy: {}%".format(round(accuracy_score(y_test, prediction)*100,2)))

    dct['Naive Bayes'] = round(accuracy_score(y_test, prediction)*100,2)

    accuracy: 81.43%
```

Fig5.1

The fig5.1 shows the accuracy level of fake news deduction using naive bayes algorithm.

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

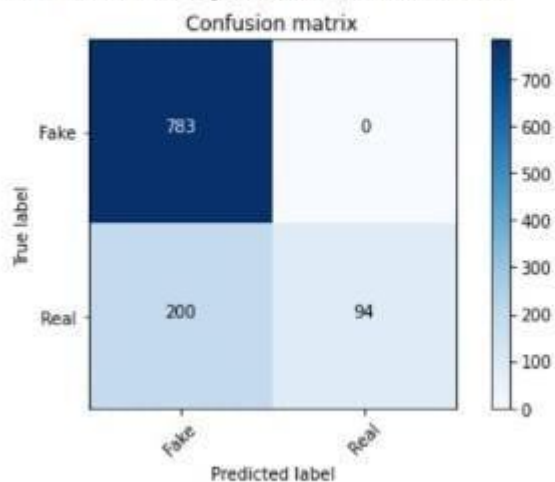Confusion matrix, without normalization

Fig5.2

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

VI. CONCULISION

In this paper, importing data set, executing coding and visualization are done in Google colab. Naive bayes 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.

Naive bayes algorithm has been used in this paper. In further works other algorithms may be utilized in different tools and accuracy rate is predicted to search out which tool and algorithm has high accuracy in predicting the fake news.

REFERENCES

[1]. Akshay Jain and AmeyKasbe. "Fake News Detection." 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). Bhopal, India: IEEE. 2018.

[2]. Cardoso Durier da Silva, F., Vieira, R., & Garcia, A. C. (2019, January). Can machines learn to detect fake news? a survey focused on social media. In Proceedings of the 52nd Hawaii International Conference on System Sciences.

[3]. Gilda, S. "Evaluating machine learning algorithms for fake news detection." 15th Student Conference on Research and Development (SCOReD) (pp. 110-115). IEEE. 2017.

[4]. 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).

[5]. MykhailoGranik and VolodymyrMesyura. "Fake news detection using naive Bayes classifier." First Ukraine Conference on Electrical and Computer Engineering (UKRCON). Ukraine : IEEE. 2017.

[6]. 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.

[7]. Prabhjot Kaur et al. "Hybrid Text Classification Method for Fake News Detection." International Journal of Engineering and Advanced Technology (IJEAT), 2388-2392. 2019.

[8]. Supanya Aphiwongsophon et al. "Detecting Fake News with Machine Learning Method." 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). Chiang Rai, Thailand, Thailand: IEEE. 2018.

[9].Veronica Perez-Rosas et al. Available at:
https://www.researchgate.net/publication/319255985_Automatic_Detection_of_Fake_News August, 2017.

[10]. Yumeng Qin et al. "Predicting Future Rumours." Chinese Journal of Electronics Volume: 27 , Issue: 3 , 5 2018, 514 - 520.