FAKE NEWS DETECTION USING LOGISTIC REGRESSION ALGORITHM WITH MACHINE LEARNING

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

A.SRINIRANJANI[2]
UG Scholar,
Department of B. Com (Business Analytics)
PSGR Krishnammal College for Women Coimbatore, India.
ninijamuna02@gmail.com

ABSTRACT

Most of the smart phone users prefer to read the news via social media over internet. The news websites are publishing the news and provide the source of authentication. The question is how to authenticate the news and articles which are circulated among social media like WhatsApp groups, Facebook Pages, Twitter and other micro blogs & social networking sites. It is harmful for the society to believe on the rumors and pretend to be a news. The need of an hour is to stop the rumors especially in the developing countries like India, and focus on the correct, authenticated news articles. This paper demonstrates a model and the methodology for fake news detection. With the help of Machine learning and natural language processing, it is tried to aggregate the news and later determine whether the news is real or fake using Logistic regression.The proposed model is working well and defining correctness of the results 97.21 % of accuracy .

Key Words: Accuracy level, Analysing, Logistic regression.

I. INTRODUCTION

In the world of rapidly increasing technology ,information sharing has become an easy task. There is no doubt that internet has made our lives easier and access to lots of information. This is an evolution in human history, but at the same time it unfocusses the line between true media and maliciously forged media. Today anyone can publish content – credible or not – that can be consumed by the world wide web. Sadly, fake news accumulates a great deal of attention over the internet, especially on social media. People get deceived and don't think twice before circulating such mis-informative pieces to the world. This kind of news vanishes but not without doing the harm it intended to cause. The social media sites like Facebook, Twitter, Whatsapp play a major role in supplying these false news. Many scientists believe that counterfeited news issue may be addressed by means of machine learning and artificial intelligence.Fake news detection is made to stop the rumors that are being spread through the various platforms whether it be social media or messaging platforms, this is done to stop spreading fake news which leads to activities like mob lynching, this has been a great reason motivating us to work on this project. We have been continuously seeing various news of mob lynching that leads to the murder of an individual; fake news detection works on the objective of detecting this fake news and stopping activities like this thereby protecting the society from these unwanted acts of voilence.

II. OBJECTIVES

The most common algorithms used by fake news detection systems include machine learning algorithms such as Sup port Vector Machines, Random Forests, Decision trees, Sto chastic Gradient Descent, Logistic Regression and so on. In this project have attempted to implement one out of these algorithms to train and test our results to attain the maximum accuracy using logistic algorithm.

III. RELATED WORKS

Because of the multi-dimensional nature of fake news, the recognizing the category of news is not so easy. It is obvious that a practical technique must contain a few perspectives to precisely handle the issue. This is the reason the proposed strategy is a mix of Naïve Bayes classifier, Support Vector Machines, and semantic investigation. The proposed strategy is completely made out of Artificial Intelligence draws near, which is basic to precisely order between the genuine or the fake, rather than utilizing calculations that can't mirror subjective capacities. The three-section strategy is a blend between Machine Learning calculations that subdivide into managed learning procedures, and characteristic language preparing techniques[1].

Facebook and WhatsApp are also working on fake news detection as they wrote in an article. They have been working for almost one year, and it is currently under the alpha phase[3]. Samir Bajaj of Stanford University published a research paper on fake news detection. He detects fake news with the help of NLP perspective and implements some other deep learning algorithm. He took an authentic data set from Signal Media News dataset. Several approaches have been taken to detect the fake news after massive widespread fake news in recent times[2][6].

However, the biggest reason why false information is able to thrive continuously is that humans fall victim to Truth-Bias, Naïve Realism, and Confirmation Bias. When referring to people being naturally "truth-biased" this means that they have "the presumption of truth" in social interactions, and "the tendency to judge an interpersonal message as truthful, and this assumption is possibly revised only if something in the situation evokes suspicion" (Rubin, 2017).

Basically humans are very poor lie detectors and lack the realization that there is the possibility they are being potentially lied to. Users of social media tend to be unaware that there are posts, tweets, articles or other written documents that have the sole purpose of shaping the beliefs of others in order to influence their decisions. Information manipulation is not a well-understood topic and generally not on anyone's mind, especially when fake news is being shared by a friend. Users tend to let their guard down on social media and potentially absorb all the false information as if it were the truth. This is also even more detrimental considering how young users tend to rely on social media to inform them of politics, important events, and breaking news (Rubin, 2017)[4].

Myithalle Granik e al in their paper [2] shows a simple approach for fake news detection using naive Bayes classifier. This approach was implemented as a software system and tested against à data set of Facebook news posts. They were collected from three large Facebook pages each from the right and from the left, as well as three large mainstream political news pages (Politico, CNN, ABC News). They achieved classification accuracy of approximately 74%. Classification accuracy for fake news is slightly worse. This may be caused by the skewness of the dataset only 4.9% of it is fake neves[5].

Himank Gupta et. al gave a framework based on different machine learning approach that deals with various problems including accuracy shortage, time lag (BotMaker) and high processing time to handle thousands of tweets in 1 sec. Firstly, they have collected 400,000 tweets from HSpam14 dataset. Then they further characterize the 150.000 spans tweets and 250.000 non-spam tweets. They alsodectved some lightweight features along with the Top-30 words that are providing highest information gain from Bag-of-Words model. They were able to achieve an accuracy of 91.65% and surpassed the existing solution by approximately18%[4].

IV. METHODOLOGY

A. PROPOSED SYSTEM

STEP 1: Imported the dataset from Kaggle dataset, modified the dataset and saved in Excel.csv format.
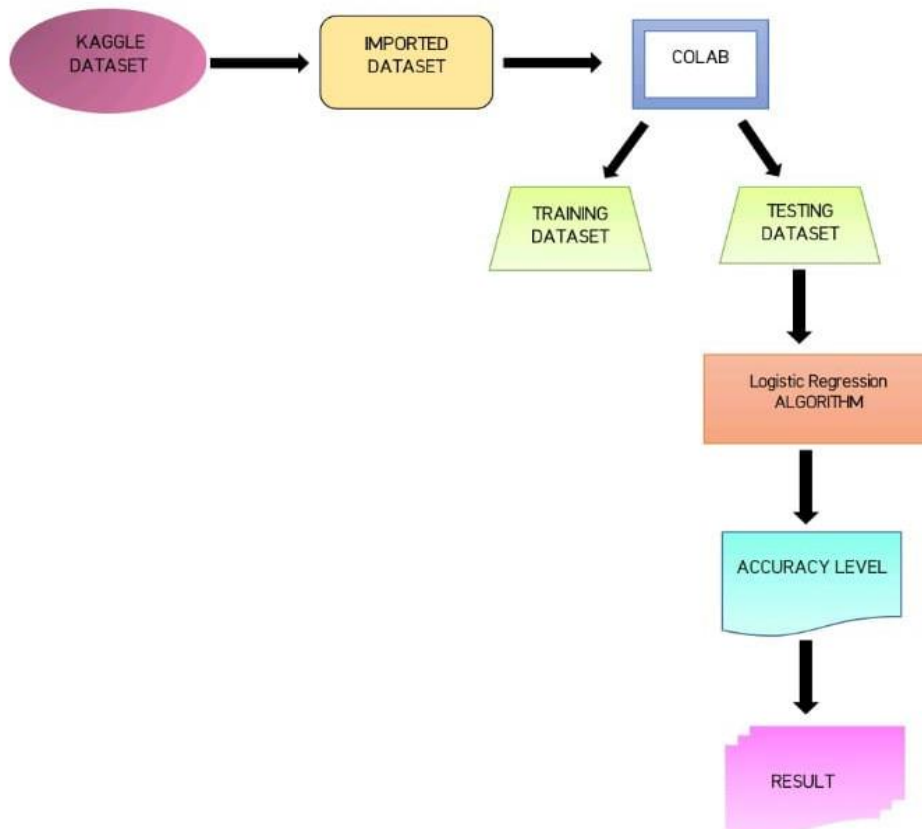
STEP 2: Used Google colab for executing python coding and removed all unwanted data from dataset.

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 using Logistic Regression.

B. WORK FLOW

C.LOGISTIC REGRESSION

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable**.**

- Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.
- Logistic Regression can be used to classify the observations using different types of data and can easily determine the most effective variables used for the classification.
- Logistic regression uses the concept of predictive modeling as regression; therefore, it is called logistic regression, but is used to classify samples; Therefore, it falls under the classification algorithm.

V. RESULT

```
# Vectorizing and applying Tf-IDf
from sklearn.linear_model import LogisticRegression

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

# Fitting the model
model = pipe.fit(X_train, y_train)

# Accuracy
prediction = model.predict(X_test)
print("accuracy: {}%".format(round(accuracy_score(y_test, prediction)*100,2)))
dct['Logistic Regression'] = round(accuracy_score(y_test, prediction)*100,2)
```

```
accuracy: 97.21%
```

FIG 1

From the above FIG 1 train dataset is imported for getting accuracy of fake news and as a result there is 97.21% of accuracy of fake news.
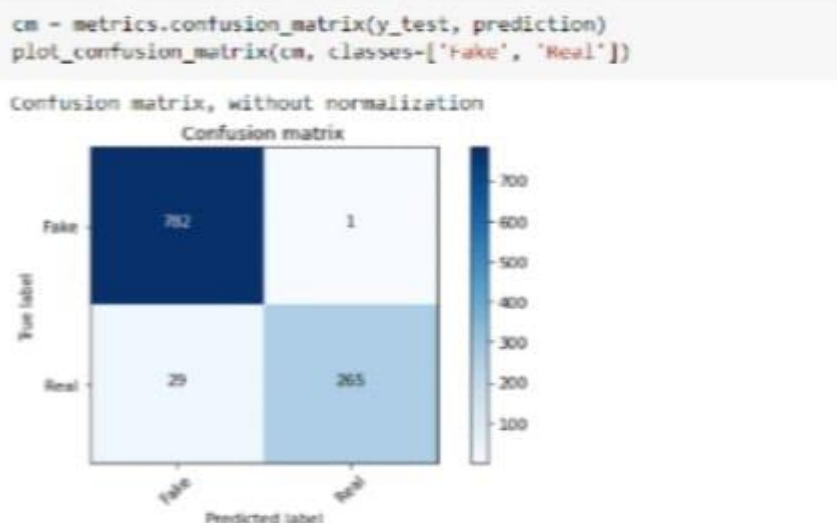
```
cm = metrics.confusion_matrix(y_test, prediction)
plot_confusion_matrix(cm, classes=['Fake', 'Real'])
```

Confusion matrix, without normalization

FIG 2

In the above FIG 2 plotting the fake and real news which has more accurate ,here there is more fake news as compared to true news.

VI. CONCLUSION

Vast spreading of fake news through the net will deliver bad impacts to the society. Fake news will mislead readers and deceive them to the ultimate confusion in believing something that is not true to be true. This is the danger of fake news as people nowadays still are unable to differentiate between fake news and real news in their daily life with their naked eyes. However, this problem can be certainly solved by harnessing the power of machine learning to predict news to be fake or not. Within this capability.Here we predicted the accuracy: 97.21 for the real and fake news using Logistic Regression.

REFERENCE

1.A SMART SYSTEM FOR FAKE NEWS DETECTIONUSING MACHINE LEARNINGhttps://www.researchgate.net/publication/339022255_A_smart_System_for_Fake_News_Detection_Using_Machine_Learning.

2. . A. Martínez -Garcia, S. Morris, M. Tscholl, F. Tracy, and P. Carmichael, "Case-based learning, pedagogical innovation, and semantic web technologies," IEEE Trans. Learn. Technol., vol. 5, no. 2, pp. 104-116, 2012.

3. https://indianexpress.com/article/technology/ social/whatsapp-fight-against-fake-news-top features-to-curb-spread-of-misinformation.

4. . H Gupta, M S Jamal. S. Madsetty and M. S. Desarkar, "A framework for real-time sparn detection in Twitter," 2018 10th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru.

5. . M. Granik and V. Mesyura, Fake news detection using naive Bayes classifier" 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, 2017, pp. 900-903.
.
6.P. R. Humanante-Ramos, F. J. Garcia Penalvo, and M. A. Conde-Gonzalez, "PLES in Mobile Contexts: New Ways to Personalize Learning," Rev. Iberoam. Tecnol. del Aprendiz., vol. 11, no. 4, pp. 220-226, 2016.

7. Rubin, V. (2017). Deception detection and rumor debunking for social media. Handbook of Social Media Research Methods.