DNN Based Fake News Identification and Analysis

[1]Nagumalla Lalitha Naga Lakshmi Alekya, [2]S Surya Godha Devi
[1,2]Dept. of CSE, KIET, Kakinada, AP, India

Abstract
As the world is growing towards Internet-based digitalization over the social media, millions of articles are uploaded every second, and the Internet users are more keep on social media applications such as WhatsApp, Meta, Instagram, Twitter and other online applications. The aim of the project is to detect the Fake or Real News in online available articles. There are many news channels online, who are publishing the daily news and more people read online media as it has many benefits such as paperless, fingertip availability of news, Attractive headlines and colorful titles. The recent advancement of the online social media platforms has impacted the daily life of civilization. The people who are using the above online news platform should not get the fake news regarding, we need to identify the challenging solution that can detect the fake news. The application of the Project is detecting the fake news using Machine Learning (ML) and Deep learning (DL). The project uses two major algorithms namely Passive Aggressive Classifier and Adaptive Boosting Classifier (ABC) and Long Short Term Memory (LSTM). We have made the strong analysis and visualization on the dataset supplied. Our models are providing 95% & above accuracy. Finally we have developed the Web Application for the prediction of Forged and Actual news.

Index Terms
Passive Aggressive Classifier, LSTM, TF-IDF, BOW, Flask-Web Application.

I. Introduction

Today's world information is growing every day, so the data is the important asset for every organization. The evaluation of genuine information is very difficult task to perform as many cases are noticed with forged entities. If we talk about the news, the information shared in the news article, knowing the genuine news is very important for everyone to aware about the daily events happening around the globe. The news may be genuine or fake; the fake information may impact on our daily activities, culture and outside environment. It also affects the normal life of people who are more prone to the newspaper reading every day.

The project basically concentrate on developing the system to classify the fake and real news, as we all know that the false information spread very fast like wild fire and it spread to all in a very fast manner. Here we are using the concepts of ML and DL to identify and predict the fake news using the models. We have used the dataset from the different sources of news channels. Finally we have developed the web-based application to show the live demonstration of predicting the news either Real or Fake. In the project, we have used many feature engineering techniques required to classify the news articles or the text. In this context we have used the classification methodology as TF-IDF (Term Frequency- Inverse Data Frequency) and Bag or Words (BOW) for Feature selection called as Feature Engineering on the dataset. Here the textual data is converted into feature vectors then it can be used as the input to the machine learning models to classify or group the news as Fake or Real. Also the prediction part can be shown by using Flash Web Application.

Here in this project we are rejecting some of the features which are not part of our context of application. We ignore some of the attributes such as from which sources which are taking the news contents etc. Only the important attributes are selected for the study. We are also aim to represent several ML and DL models to regulate the finest way to classify news.

The news may be from the online media or offline media. Fake news is, which a part of publicity is or yellow reporting that consists of cautious falsehood or swindles spread, which can be present to the public. Most of the news channels are showing fake news just for the TRP achievements that leads to misinformation among the communities and can be spread across the various platforms.

We tried to develop the system where we can predict the purity of news contents and help the user to identify the falsity of the news. In this project we are checking with the static and dynamic data available of Internet sources. The list of news contents are collected and converted into (Comma Separated Value) .csv file and further can be used as a Dataset in our project.

The dataset is not in proper format, so we further store it into the local machine, preprocessing of data is applied. ML & DL techniques are the best solution to extract the information available for the data classification and labelling and further we can use this data for the purpose of prediction.

The figure 1.1 shows a simple representation of the project by using the Web Application using Flask Framework. In this application User open the index page, check the sources from the internet and enters the news title from any one of the Online News Channels. The model works behind the scene and give us the prediction after entering the title or the caption news and alert as whether the input information or news is fake or real.

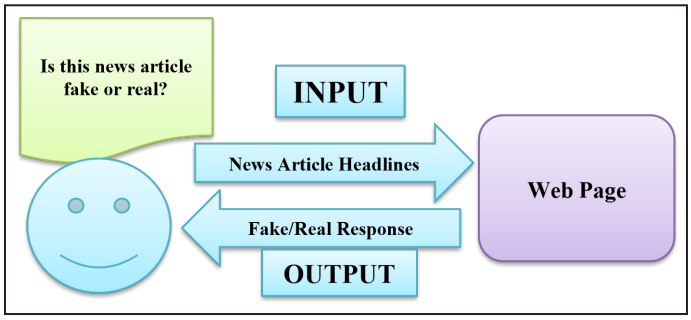

Fig. 1: Basic Layout of Project

II. Literature Review

The above problem can be designated by using the ML & DL approaches, the objective can be solved and build the modelling and finding the predictions. Many authors across the globe have contributed towards the research methodologies and contribute their research work on the trending global issue. The literature survey made helped in increasing the awareness among the civilians and finding the appropriate solutions. The basic survey is more useful to grow towards research directions, awareness and challenges in today's issues. Everyone in this world can post any content over the Internet and there are no restrictions to share the contents over the various social media platforms.

News comes from different websites, majorly from the website-based networking social Medias. The spread of fake news is not good for the civilians as it leads to many problems in terms of communal rights, fights among the social life and negativity among the specific category of peoples. As the technology advancements and powerful media resources, on the same pace the precautionary measures are required to stop such activities. There are many websites which gives false information and attempt to exploit the people's mind-set.

Fake news detection and classification is made to halt the rumors that are being feast through the different social media platforms and messaging applications. We are seeing various mob lynching activities that leads to attempt of murders and other activities, so we need to protect the social life thereby protecting the society from these kinds of unwanted activities.

A. Survey on Text Classification for Identifying Fake News

Text classification is very important issue in identification of fake or real news. There are different methods to classify the content of news. The primarily used approaches are TF-TDF, Bow and Word2vec (Word to Vector). By using the Gensim Module we can develop the classification models that use the average of all words to represent the complete document or news content. The classification report by calculating the Precision, Recall, Support and F1-Measure can further clarify the actual representation of data.

B. Fake News Detection Using Deep Learning Techniques

There are many algorithms available to solve the fake news issues using deep learning. Among that we can develop the models using Convolutional Neural Network (CNN), Deep Neural Network (DNN), Recurrent Neural Network (RNN) and many more. Among that some of the algorithms are giving better accuracy and less loss. The algorithms which are most used in all the literature survey are Shallow-and-Wide CNN, Bi-Directional LSTM, Bidirectional RNN and combinations of (Bi-)LSTM + BCNN.

C. Fake Information and News Detections using Machine Learning technique

Many Machine Learning algorithms (Either Supervised learning or Unsupervised learning) are used in solving and classifying the issue and to distinguish between real and fake news articles. The most common algorithms used are: Support Vector Machine, Naïve Bayes, Logistic Regression, K-Nearest Neighbor, Decision Tree and Ensemble algorithms.

III. Objectives

In all parts of world, lot of research is going on to address the different issues pertaining to the Threat from the insiders. Here in our proposed system, we have mentioned the various methods to resolve the issue. The process of detecting the insider and misbehaviour der threat traces the fake alarms to the administrator whenever it finds the malicious activity of many employees in the organization.

To confrontation the aforesaid explained, we have tried to focus on the detection of Fake news using Machine Learning and Deep Learning model and Standard reviews Dataset and predicting the possibility of Real or Fake .

As per the above mentioned aim of the project, herewith we have framed the objectives as:
The main aim of this project is:

- **Objective 1:** Exploratory Data Analysis and Visualization on Dataset
- **Objective 2:** ML Models: Passive Aggressive Classifier and Adaptive Boosting (ADABOOST) Algorithm.
- **Objective 3:** Deep Learning Model Deployment using CNN Model and Comparison Study of above all model.
- **Objective 4:** Web Application to predict Fake or Real News.

IV. Proposed Architecture

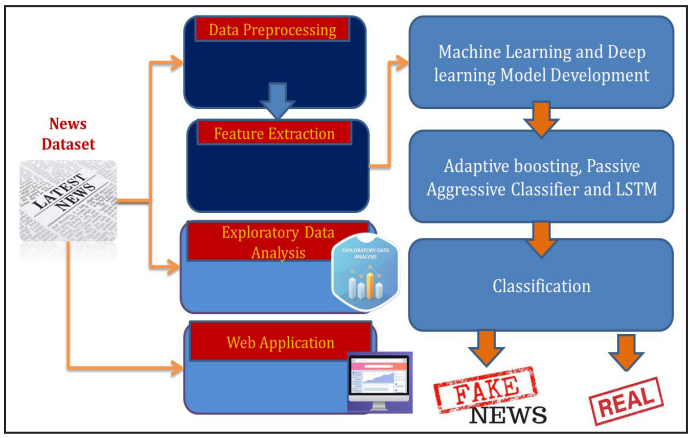

Fig. 2: Overall Architecture Diagram

The proposed system consists of the overall design of the system for Forged and actual News predictions. The information is collected from genuine sources and is used as Fake review dataset. The initial stage is Data Collection, we acquired the data.

The Second Stage of the system is Data Preporcessing, obviously whenever we are using any data , the data should be very much informative and clean , so that where will not be difficulty in model training.
Another feature to the system is Feature extraction of attributes. The important attributes for the purpose of study are used here. Once the above process of data acquisition till extraction, we use the concept of Data Engineering called as Exploratory Data Analysis, Which gives lot of weightage to the system in terms of Analysis and Visualization of data in various directions.

Numerous ML Algorithms are used as Models to classify the text. Along with ML , we have used deep learning algorithm based CNN model .Finally we have developed the web-based application to take the input from the user, deploy the model using machine learning and application has to predict the scenario as Fake or Real news.

A. Data Collection

This is the first stage of our study and model development. The data is collected from numerous sources. The dataset contains almost 8,000 entries (news) with features as Title, Text and label. The dataset is useful in model evaluation using numerous models. The dataset contains the information about for real and fake.

Table 1: Dataset Information

Dataset Name	Attribute List			
Fake.csv	Title	Text	Subject	Date
Real.csv	Title	Text	Subject	Date

B. Data Preporcessing

Before feeding data to the model, the information or the input need to be preprocessed, we are applying various schemes of preprocessing like cleaning the data, removing irrelevant rows and columns, data abstraction and final data aggregation.

C. Feature Extraction

Feature extraction is a process of renovating the fresh input data into the integral attributes and the main thing is we must preserve the data during the extraction phase. The data integrity, data confidentiality and storage are important and retain the original information of dataset. The raw data is processed at each stage and cleaned attributes which required for the study are extracted.
The textual information (text) is converted to another form to numerical keywords to input as a training data to the classifiers. The input to the system is text data and output as class label (Fake or real). The input is again converted into series of vectors, later we can use any of the encoding techniques to convert it into the vectors.

D. Machine Learning Algorithms

1. Passive Aggressive Classifier

In this type of text classification algorithm, it takes single response at one stage. Every time the weight on the model is adjusted and deployed in the model implementation. In each iteration, it checks for the prediction, if the predicted value is correct, the model behaves same. It acts as passive, if the model build is incorrect. After every iteration, it adjusts the weight, till it gets the correct predictive outcome. Then model classifying news as either "Real" or "Fake".

2. AdaBoost –Adaptive Boosting Algorithm

In Adaptive Boosting algorithm, which is also a part of ensemble technique are used to increase the output or accuracy of the model developed. Most likely this algorithm is used in binary classification of models. The working of algorithm depends on the input applied. It follows the methodology of repetitive approach in gaining the knowledge from the mistakes of weak classifiers, and makes them into strong classifier. The feature vectors are given as the input to the model to classify the fake and real news prediction.

3. Deep Learning Model-LSTM

The deep learning models we are implementing here are, Namely LSTM. We have used CNN layers for feature extraction and LSTM module for the prediction. CNN LSTM is a powerful category of deep learning model that is used for deep data learning for both spatially and temporally dataset and performs all the input and output. CNN uses a process called Convolution in finding the relationship between the two variable or functions. The function shape can be modified.

V. Experimental Results

The implementation can classified into different modules of project and are listed as: Data Acquisition

- Initial Data Pre-Processing
- Feature Engineering
- Exploratory Data Analysis (EDA)
- Model Training And Testing
- Model Deployment
- Model Performance Estimation
- Comparison Study of The Model
- Fake news Prediction Web Application using Flask Framework

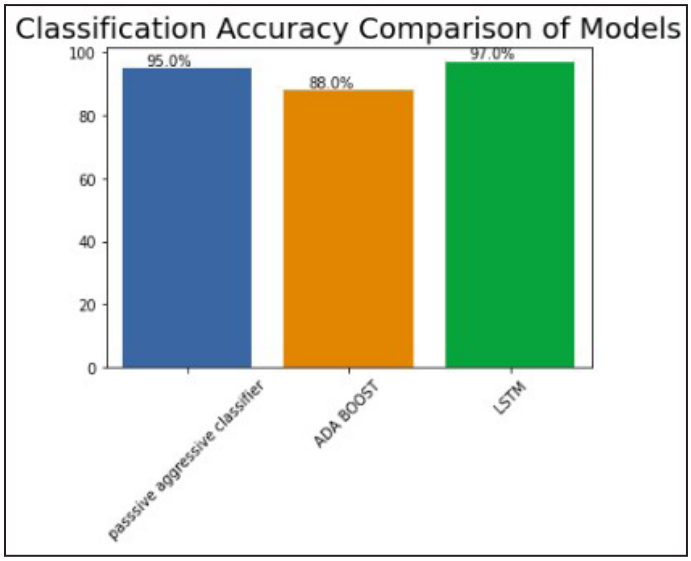

Fig. 3: Comparison Study of ML & DL Algorithms

VI. Implementation of Web - Based Application For Fake News Predictions

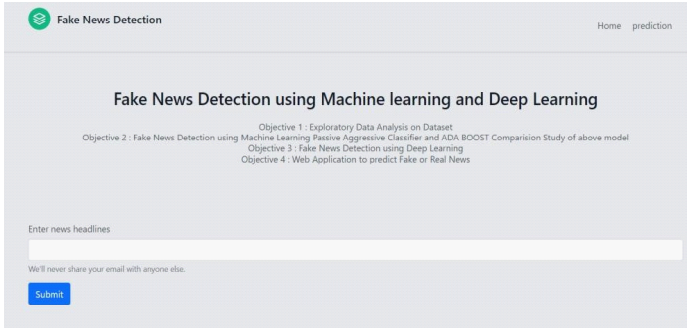

Fig. 4: Front End Design for prediction

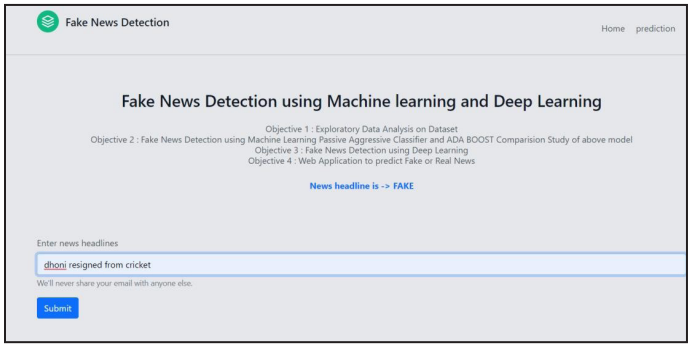

Fig. 5: Fake New Prediction Screen

In this project we have tried to implement a system which can predict the Fake and real News using standard Dataset. We applied

two machine learning algorithm namely Passive Aggressive Classifier and Adaptive Boosting Algorithm , and deep learning algorithm LSTM is used and the accuracy are 95%, 88% and 97% respectively. Among all LSTM provides highest accuracy score. Apart from the above results we have used pickle file to deploy the model. We have created the web application to predict the fake and real news.

References

[1] Abdullah Al MamunSardar, ShahaluAkter Salma, Md. Sanzidul Islam, Md. Arid Hasanand Touhid Bhuiyan.2021. Team Sigmoid at CheckThat!2021 Task 3a: Multiclass fake news detection with Machine Learning. CLEF 2021 – Conference and Labs of the Evaluation Forum. 2936.

[2] Dr. K Jagan Mohan,,AnjuriAnusha. 2021. International Journal of Computer Science & Communication, 12(2): 38-44.

[3] Youssef Taher. 2021. Automatic Fake News Detection based on Deep Learning, FastText and NewsTitle. International Journal of Advanced Computer Science and Applications,13(1): 146-158.

[4] Md. RakibHasan, IsmathAraItu. 2021. A Distinctive Approach for Detecting Fake News using machine learning. International Journal of Innovative Technology and Exploring Engineering, 11(3): 29-35.

[5] Alim Al Ayub Ahmed. 2021. Detecting Fake News using Machine Learning: A Systematic Literature Review. Biomedical Signal Processing. 36-41.

[6] PoojaDeshmukh, Dr. R. R. Shelke. Prediction Of Fake News Using MachineLearning On ISOT Dataset. Science, Technology and Development, 10(6): 547-558.

[7] Akash Dnyandeo Waghmare. 2021. Fake News Detection Of Social Media News In Blockchain Framework. Indian Journal of Computer Science and Engineering. 12(4): 972-980.