IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, May 2024

Fake News Detection using Machine Learning

Anusha Amrutkar[1], Jay Pawar[2], Usha Sahal[3], Dr. B. S. Shirole[4]
Department of Computer Engineering
Sanghavi College of Engineering, Nashik, India

Abstract: *In our modern era where the internet is global, everyone relies on various online resources for news. Along with the increase in the use of social media platforms like Facebook, Twitter, etc. news spread rapidly among millions of users within a very short span of time. The spread of fake news has far-reaching consequences like the creation of biased opinions. The project demonstrated for detecting the fake news. The dataset was provided by the company. Here I am performing binary classification of various news articles available online with the help of concepts pertaining to Artificial Intelligence, Natural Language Processing and Machine Learning. Using decision tree classifier provides the ability to classify the news as fake or real.*

In this project different feature engineering methods for text data has been used like Bag of words model and word embedding model which is going to convert the text data into feature vectors which is sent into machine learning algorithms to classify the news as fake or not. With different features and classification algorithms we are going to classify the news as fake or real and the algorithm with the feature which gives us the best result with that feature extraction method and that algorithm we are going to predict the news as fake or real

Keywords: News Identification dataset, Deep Learning, Machine Learning, Classification

I. INTRODUCTION

Data or information is the most valuable asset. The most important problem to be solved is to evaluate whether the data is relevant or irrelevant. Fake data has a huge impact on lot of people and organizations. Since fake news tends to spread fast than the real news there a need to classify news as fake or not. In the project the dataset used is from Kaggle website where real news and fake news are in two separate datasets we combined both the datasets into one and trained with different machine learning classification algorithms to classify the news as fake or not. In this project different feature engineering methods for text data has been used like Bag of words model and word embedding model which is going to convert the text data into feature vectors which is sent into machine learning algorithms to classify the news as fake or not. With different features and classification algorithms we are going to classify the news as fake or real and the algorithm with the feature which gives us the best result with that feature extraction method and that algorithm we are going to predict the news as fake or real. In this project we will be ignoring attributes like the source of the news, whether it was reported online or in print, etc. and instead focus only the content matter being reported. We aim to use different machine learning algorithms and determine the best way to classify news .

II. PURPOSE

A fake news classification system using different feature extraction methods and different classification algorithms like Support Vector Machine, Logistic Regression, Gradient Boosting, XG-BOOST, Decision Tree, Random Forest and the best algorithm we are going to use it in predicting the news as fake or real. In order to create a real time application, the algorithm should be fed with the most recent data. Data is of different sizes so that should be properly cleaned to get better results. So we are using different algorithms and feature extraction methods like Bag of words model and Word embedding model to get the best result.

III. OBJECTIVE OF SYSTEM

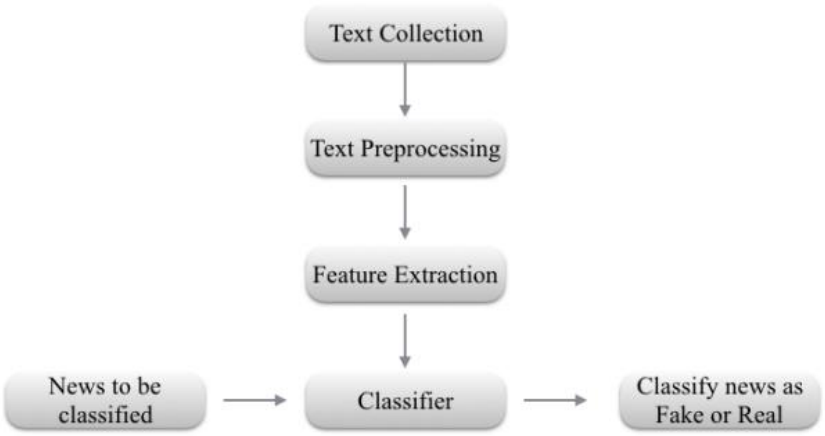
- To achieve our goal of developing machine learning model to classify news as fake or real,
- We need perform following tasks in the same order as stated.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, May 2024

- Data Collection and Analysis Preprocessing the data Text feature extraction Using different classification algorithms
- Taking the best classification algorithm and feature extraction method

IV. SYSTEM ARCHITECTURE

In future works, we intend to use highly sophisticated classifying approach, like deep learning with sentiment analysis also and consider many text features like publisher, urls etc., which may increase the accuracy of the classification of news as fake or real. Automatic fake news detection may be done using the latest news and training the model regularly to get the best results. So this can be used as a filter to upload the news.



User Registration:
Here User has to register with required parameters such as name, mobile, password and hardware ID.

User Login:
After user registration done successfully user can login to the system
Predict News
Feed Back Generation
Algorithm & Mathematical Model
S={I, O, P, S, C, P, Ad, Q, G,H/w, S/w, Failure, Success}
Where
S=System
C= Check News U=User
Ad=Admin
G= Detection.
Procedures {P}= {Pr, Cc, Qid, Amt} Where,
Pr= Check News
Qid= Find News
O is Output of system

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, May 2024

VI. RESULT

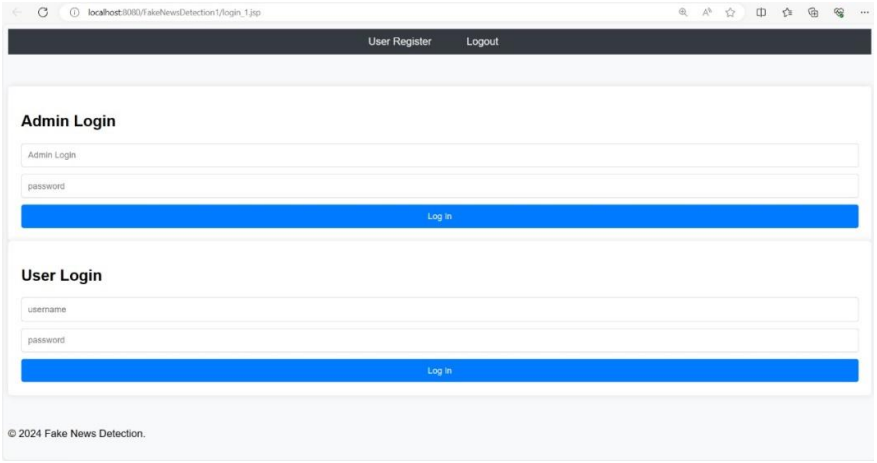


Fig 1. Login Page

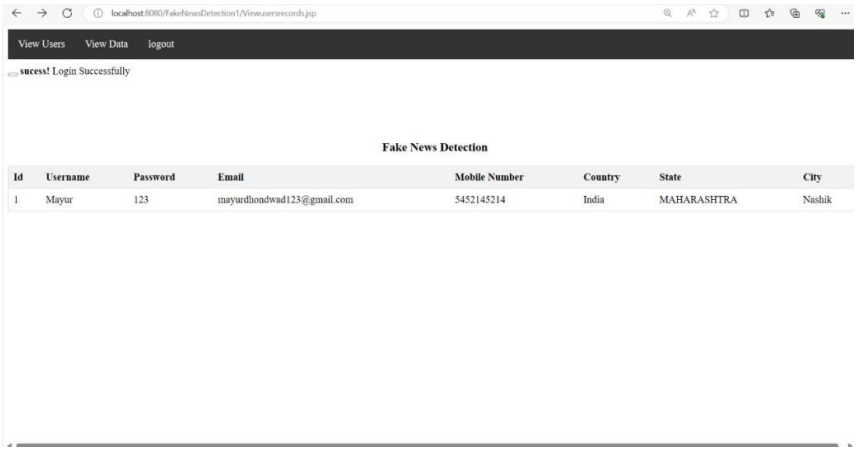


Fig 2.View Users

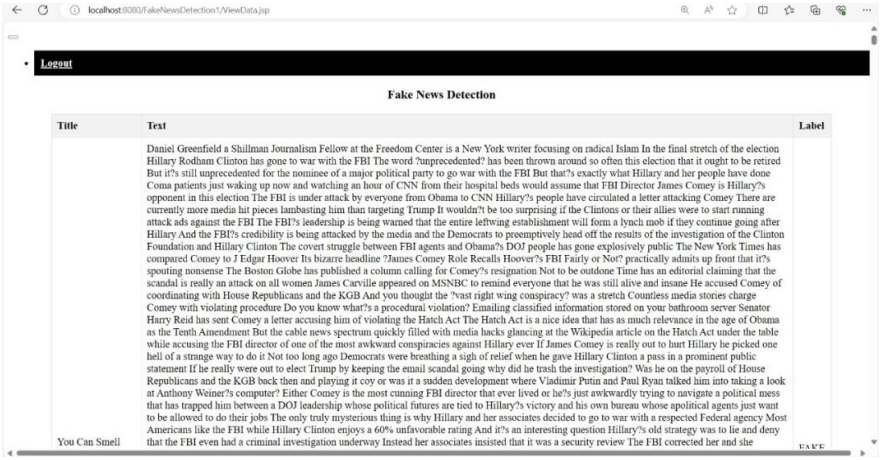


Fig.3 Train Data

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, May 2024

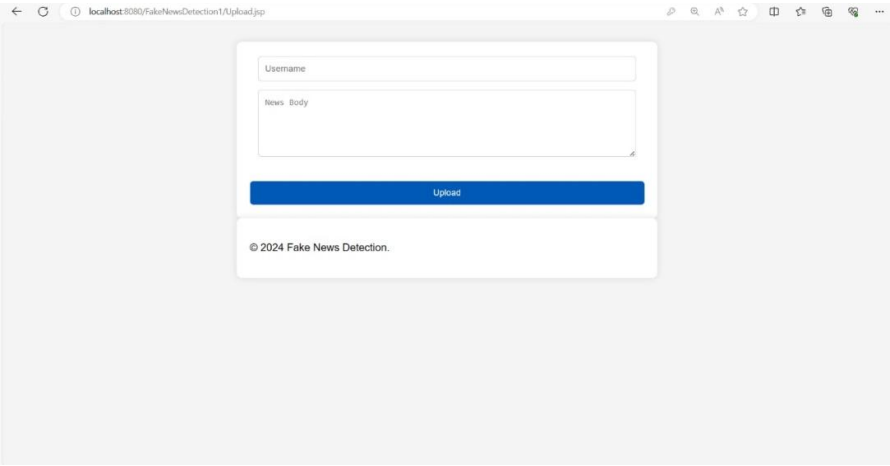


Fig 4. Detection

VII. CONCLUSION

In this paper three different feature extraction methods like Count Vectorizer, TFIDFVectorizer, Word Embedding has been used. And also different classification algorithms like Linear SVC, Logistic Regression Classifier, Decision Tree Classier, Random Forest Classifier, XG-BOOST Classifier, Gradient Boosting Classifier have been used to classify the news as fake or real. By using the classification algorithms we got highest accuracy with SVM Linear classification algorithm and with TF-IDF feature extraction with 0.94 accuracy. Even though we got the same accuracy with Neural Network with Count Vectorizer, Neural Networks and take more time to train and its complex so we used Linear SVC which is not so complex and takes less time to computer.

VIII. ACKNOWLEDGMENT

We express our heartfelt gratitude to our esteemed mentors and professors, especially **Dr. B. S. Shirole+-**, for their invaluable guidance in our academic and project endeavours. We also extend our thanks to the Computer Engineering Department and its staff for their continuous support. Our sincere thanks go to **Dr. B. S. Shirole**, Principal of SANGHAVI COLLEGE OF ENGINEERING, Nashik, for his support and permission to complete this project. We appreciate the assistance of our department's support staff, and we're grateful to our parents, friends, and all those who supported us throughout this project.

REFERENCES

[1] U. A. Butt, R. Amin, H. Aldabbas, S. Mohan, B. Alouffi, and A. Ahmadian, “Cloud-based email phishing attack using machine and deep learning algorithm,” Complex Intell. Syst., pp. 1–28, Jun. 2022.
[2] D. C. Le and A. N. Zincir-Heywood, “Machine learning based insider threat modelling and detection,” in Proc. IFIP/IEEE Symp. Integr.Netw.Service Manag. (IM), Apr. 2019, pp. 1–6.
[3] P. Oberoi, “Survey of various security attacks in clouds based environments,” Int. J. Adv. Res. Comput. Sci., vol. 8, no. 9, pp. 405–410, Sep. 2017.
[4] A. Ajmal, S. Ibrar, and R. Amin, “Cloud computing platform: Performance analysis of prominent cryptographic algorithms,” Concurrency Comput., Pract. Exper., vol. 34, no. 15, p. e6938, Jul. 2022.
[5] U. A. Butt, R. Amin, M. Mehmood, H. Aldabbas, M. T. Alharbi, and N. Albaqami, “Cloud security threats and solutions: A survey,” Wireless Pers. Commun., vol. 128, no. 1, pp. 387–413, Jan. 2023