Fake News Detection using Machine Learning Algorithms

Siddhi Sachin Pande, Nisha Shekhar Dhotre, Nandini Maruti Panchal, Vaishnavi Prashant Pardeshi, Prof. Suparna Shirish Naik

Student, Department of Computer Technology, Bharati Vidyapeeth's Jawaharlal Nehru Institute of Technology (Poly), Pune, Maharashtra, India.
Student, Department of Computer Technology, Bharati Vidyapeeth's Jawaharlal Nehru Institute of Technology (Poly), Pune, Maharashtra, India.
Student, Department of Computer Technology, Bharati Vidyapeeth's Jawaharlal Nehru Institute of Technology (Poly), Pune, Maharashtra, India.
Student, Department of Computer Technology, Bharati Vidyapeeth's Jawaharlal Nehru Institute of Technology (Poly), Pune, Maharashtra, India.
Professor, Department of Computer Technology, Bharati Vidyapeeth's Jawaharlal Nehru Institute of Technology(Poly), Pune, Maharashtra, India.

Abstract: We all use multiple platforms of social media such as Facebook, WhatsApp, Instagram, Twitter, etc, for sharing information and media. Due to this sharing of media and information in large numbers there are usually many fake news or rumors which come along with it and we cannot tell the difference between the fake news and the real one. This project focuses on detecting the fake news from real and methods we have used to do so. In this project, we will implement few of machine learning techniques like Random Forest classifier, Logistic regression, Decision tree. These techniques will help in providing good and accurate result.
Keywords - Internet, social media, Artificial Intelligence, Machine Learning, Rumours.

I. INTRODUCTION

Social media is a digital platform that allows us users to create and share the content or information with everyone. Social media has large influence on a wide range of websites and apps. It has made its importance in our lives over a very short time. We all check our social media apps daily to stay updated with current world affairs and news. Information and news are shared within seconds around the world by so many users that sometimes there are misinformation. In previous days, people would not have known so much information about one another as they did not have any means to communicate quickly. Now, everyone is able to share news and information with everyone and some of them are fake news as well. Fake news can affect people's lives. Social media and the internet have had a huge impact on our lives.

II. IMPLEMENTATION

- Data Pre-Processing

  To pre-process your text is to make text in the simplest form that is predictable and analysable. The goal of pre-processing is to remove unnecessary text. By removing unnecessary features from our text, we can reduce complexity and increase predictability to make model faster. Removing punctuation, special characters, comma's, etc. doesn't drastically change the meaning of a text. Types of text pre-processing will include lowercase text, URL removal, etc.

- Feature Extraction

  This type of models is used in natural language processing domain to obtain features from the input text data. The extracted features are mainly used to train a machine learning classifier. Their main purpose is to characterize the input text and make it suitable to be processed. The number of rows represents the number of all input documents. The number of columns represents the number of all features extracted from all input text documents.

- Models

  1. Logistic Regression –

     Logistic regression is another Supervised Machine learning algorithm which is used for data that is co-dependent on each other. It is used to capture the relation between the binary valued data event occurring. The best example of it is to determine, if a person is infected by covid-19 or not.

  2. Random Forest Classifier –

     It is a classifier which takes the average to improve the predictive accuracy of the dataset and control over-fitting by fitting a number of decision tree classifiers on various dataset.

  3. Decision Tree Classifier –

     A decision tree is a very specific type of probabilistic classifier that enables you to make a decision on availability of resources. Here internal node represents a decision that has been taken and leaf node shows the outcome of decision and the output in predicted by traversing through the tree.

IV. LITERATURE SURVEY

[1] Paper Name - Fake News detection Using Machine Learning
Author Name - Nihel Baarir, Abdelhamid Djeffal
In this work, author propose a system for Fake news detection that uses machine learning techniques. We used term frequencyinverse document frequency (TF-IDF) of bag of words and n-grams as feature extraction technique, and Support Vector Machine (SVM) as a classifier. We propose also a dataset of fake and true news to train the proposed system. Obtained results show the efficiency of the system.

[2] Paper Name - Fake News Detection by Decision Tree
Author Name - Shikun Lyu, Dan Chia-Tien Lo
Authors propose is to analyse the performance of several machine learning algorithms such as Fake news tracker, SVM, decision trees. The results indicate that the SVM and the decision trees are suitable.

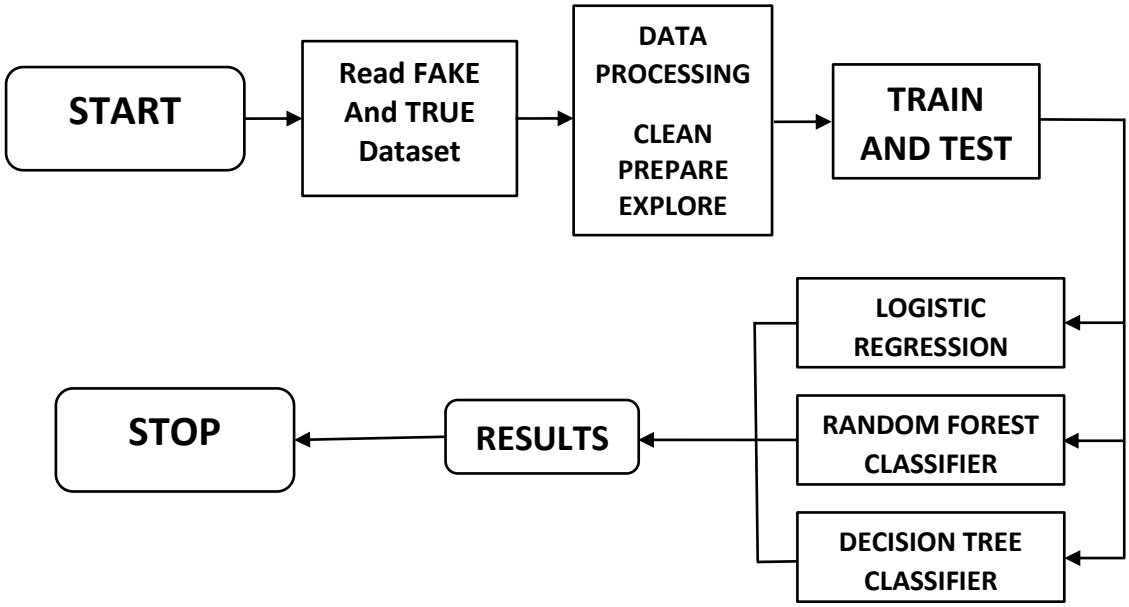
[3] Paper Name - Study of Fake News Detection using Machine Learning and Deep Learning Classification Methods
Author Name - Keshav Nath, Priyansh Soni, Aman Ahuja
Author have compared the performance of various Machine Learning and Deep Learning models. From their experimentation, it is realized that Random Forest and Bag of Words outperformed the rest with an accuracy of 98.8%.

[4] Paper Name - Fake News Detection
Author Name - K. Harshitha, Aditya V, Dr. P. Lakshmi Harika, P. Veejendhiran
The primary objective of the author is to detect the fake news to ensure creditability, benefits of the real news, to deserve the truth by using Machine learning. They ensure that decision trees and random forests exhibit superior accuracy when compared to other models, with respective values of 99.6 and 99.1

[5] Paper Name - A RESEARCH PAPER ON FAKE NEWS DETECTION
Author Name - Mayur Bhogade, Bhushan Deore, Abhishek Sharma, Omkar Sonawane
In this paper, intend is to perform a parallel grouping of different news stories accessible online with the help of thoughts identifying with Artificial Intelligence, Natural Language Processing, and Machine Learning. The result of the project determines the fake news detection for social networks using machine learning and also checks the authenticity of the publishing news website.

V. SYSTEM ARCHITECTURE

Architecture of the proposed fake news detection system.



VI. CONCLUSION

Manual classification of news articles requires in-depth knowledge and expertise in identifying anomalies in the text. It takes a lot of time to verify a single article manually that's why We discussed the use of machine learning models and ensemble methods to classify fake news articles. It is important that we have a software or mechanism that tells the fake news apart from the real vice versa, or at least an awareness that not everything we read or see on social media may be true. This way, we can help people to make more informed decisions.

VII. References

[1] Shailendra Kumar, Subodh Kumar, Pooja Yadav, Meghna Bagri" A Survey on Analysis of Fake News Detection Technique".
[2] Shikun Lyu, Dan Chia-Tien Lo" Fake News Detection by Decision Tree" IEEE Southeast Conference 2020.
[3] Nihel Fatima Baarir, Abdelhamid Djeffal "Fake News detection Using Machine Learning" 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH)
[4] Keshav Nath, Priyansh Soni, Aman Ahuja" Study of Fake News Detection using Machine Learning and Deep Learning Classification Methods"2021 IEEE
[5] Uma Sharma, Siddarth Saran, Shankar M. Patil, 2021, Fake News Detection using Machine Learning Algorithms, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NTASU – 2020 (Volume 09 – Issue 03)
[6] De Beer, Dylan, and Machdel Matthee. "Approaches to Identify Fake News: A Systematic Literature Review." Integrated Science in Digital Age 2020 vol. 136 13–22. 5 May. 2020, doi:10.1007/978-3-030-49264-9_2
[7] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, "Fake News Detection Using Machine Learning Ensemble Methods," Complexity, 17-Oct-2020. [Online]. Available: