**COLLABORATIVE PROJECT WITH INTEL**

**PROJECT TITLE : FAKE NEWS DETECTION USING PYTHON AND MACHINE LERANING**

**TEAM NAME :** **TEAM DEBUG THUGS**

**Team Mentor : Dr T V RAJINI KANTH, Professor & Head,**

**Department of CSE-AI&ML**

[**rajinitv@gmail.com**](mailto:rajinitv@gmail.com)

[**rajinikanthtv@sreenidhi.edu.in**](mailto:rajinikanthtv@sreenidhi.edu.in)

**, Ph. No: 9849414375**

**Team Members : CH. Abhinay**  Roll no: 21311A6624 AIML Team lead

**P. Guru Vamshi**  Roll no: 21311A6610 AIML Team member

**M. Sai Krishna** Roll no: 21311A6609 AIML Team member

**Institute Name** **:** **Sreenidhi Institute of Science and Technology**

**Yamnampet, Ghatkesar**

**Hyderabad - 501301**

**Date of submission :** **15 -07- 2023**

**ABSTRACT:** Internet is one of the important inventions and a large number of persons are its users. These persons use this for different purposes. There are different social media platforms that are accessible to these users. Any user can make a post or spread the news through these online platforms. These platforms do not verify the users or their posts. So some of the users try to spread fake news through these platforms. These fake news can be a propaganda against an individual, society, organization or political party. A human being is unable to detect all these fake news. So there is a need for machine learning classifiers that can detect these fake news automatically. Use of machine learning classifiers for detecting the fake news is described in this systematic literature review.

***Keywords:* Online fake news, Machine learning, fake news, Text Classification, social media**

**INTRODUCTION**

The way information is exchanged and consumed has changed dramatically with the advent of the digital age and the explosive expansion of social media platforms. While there are many advantages to this, it has also given rise to the false news epidemic, which poses a serious problem for society. In order to deceive or influence readers, fake news is defined as inaccurate or misleading material that is presented as news that is factual.

False information can spread quickly, harm people's reputations, and even have a negative impact on democratic processes and public opinion. Therefore, there is a growing demand for reliable and potent techniques to identify and counter bogus news.

The goal of this project is to create a system for detecting fake news using machine learning methods. The project tries to identify and categorise news stories as either authentic or fraudulent by utilising the strengths of natural language processing, text classification, and predictive modelling. The system will be able to accurately forecast the veracity of news articles by using sophisticated algorithms and analysing multiple textual elements, including linguistic patterns, source reliability, and contextual data.

The design, implementation, and evaluation of the fake news detection system are all thoroughly covered in this project report. The technique used, the dataset used for training and testing, and the choice of suitable machine learning algorithms are all described. The study also includes performance data and results from the system evaluation, which show how well the system works to identify bogus news.

The study also highlights the difficulties encountered during the development process, including feature engineering, data pretreatment, and model selection. It emphasises how crucial it is to overcome bias, guarantee model fairness, and take ethical considerations into account when designing and deploying such systems.

Overall, this initiative adds to the expanding body of knowledge and real-world applications in the fake news identification sector. The creation of an accurate and trustworthy system is crucial for halting the spread of false information and advancing information integrity in the digital era.

Our Project is Done using NLP Techniques, with Machine learning and Logistic Regression.

**MOTIVATION**

Our project on fake news detection is motivated by the pressing need to combat the proliferation of misinformation. Fake news poses significant threats to information integrity, democratic processes, and social cohesion. By developing reliable detection systems, we aim to preserve information accuracy, protect democratic principles, and limit the negative impacts of false information. Our project also aims to promote media literacy and critical thinking, empowering individuals to discern credible sources. Through these efforts, we strive to contribute to a more informed society and uphold the values of transparency and accuracy in the information ecosystem

**DATA SOURCES**

Data Sources for Fake News Detection

In our project on fake news detection, we have utilized various data sources to develop and evaluate our model. The availability of diverse and reliable data is crucial for training an effective fake news detection system. The following data sources have been instrumental in our research:

**1. News Websites:**

We collected a substantial amount of articles from reputable news websites such as CNN, BBC, The New York Times, and others. These sources provide a vast collection of authentic news articles that serve as a reference for training the model to identify genuine news content.

**2. Research Datasets:**

Academic research papers and studies on fake news detection have provided valuable datasets for our project. These datasets are meticulously curated and labeled, offering a diverse range of fake and real news articles. By utilizing these research datasets, we were able to incorporate state-of-the-art techniques and benchmark our model against existing approaches.

### Some Websites we referred:

### <https://data-flair.training/blogs/advanced-python-project-detecting-fake-news/>

### <https://www.kaggle.com/code/midouazerty/detecting-fake-news-step-by-step>

It is worth noting that throughout our data collection process, we prioritized the authenticity and reliability of the sources used. We ensured rigorous preprocessing and labeling of the datasets to maintain their quality and integrity. Adhering to the data usage policies and terms of service of each source was essential in maintaining ethical practices in data collection.

Pre-processing is the process of transforming or changing data through a sequence of procedures. Before our data is fed to the algorithm, it is transformed. Data processing, especially when done by a computer, is the act of performing It is a way for remodeling unclean records into easy records sets. In different words, every time records is received from numerous sources, its far achieved so in a uncooked way that makes evaluation impossible. After that, it changes the raw file to a readable format (graphs, documents, etc.)

By leveraging these diverse data sources, we aimed to create a robust fake news detection system that can effectively distinguish between genuine and false information. The combination of news articles, fact-checking reports, social media data, and research datasets has enriched our project and enabled us to address the challenges of identifying fake news in today's information landscape.

**PYTHON LIBRARIES USED IN THE PROGRAMS**

**1. Streamlit:** Streamlit is a powerful library that allowed us to build interactive web applications with ease. We utilized it to create a user-friendly interface for our fake news detection system. With Streamlit, we could design the layout, add interactive widgets, and display the output in a visually appealing manner.

**2. Numpy:** NumPy is a fundamental library for scientific computing in Python. It provided support for efficient numerical operations on arrays and matrices. In our project, we used NumPy to handle and manipulate the data in a structured manner, perform mathematical computations, and facilitate data preprocessing tasks.

**3. Pandas** :Pandas is a versatile data manipulation library widely used in data analysis projects. It provides powerful data structures and data analysis tools, such as dataframes, which allowed us to handle and analyze structured data effectively. We utilized Pandas to load, clean, and preprocess the datasets, as well as perform exploratory data analysis.

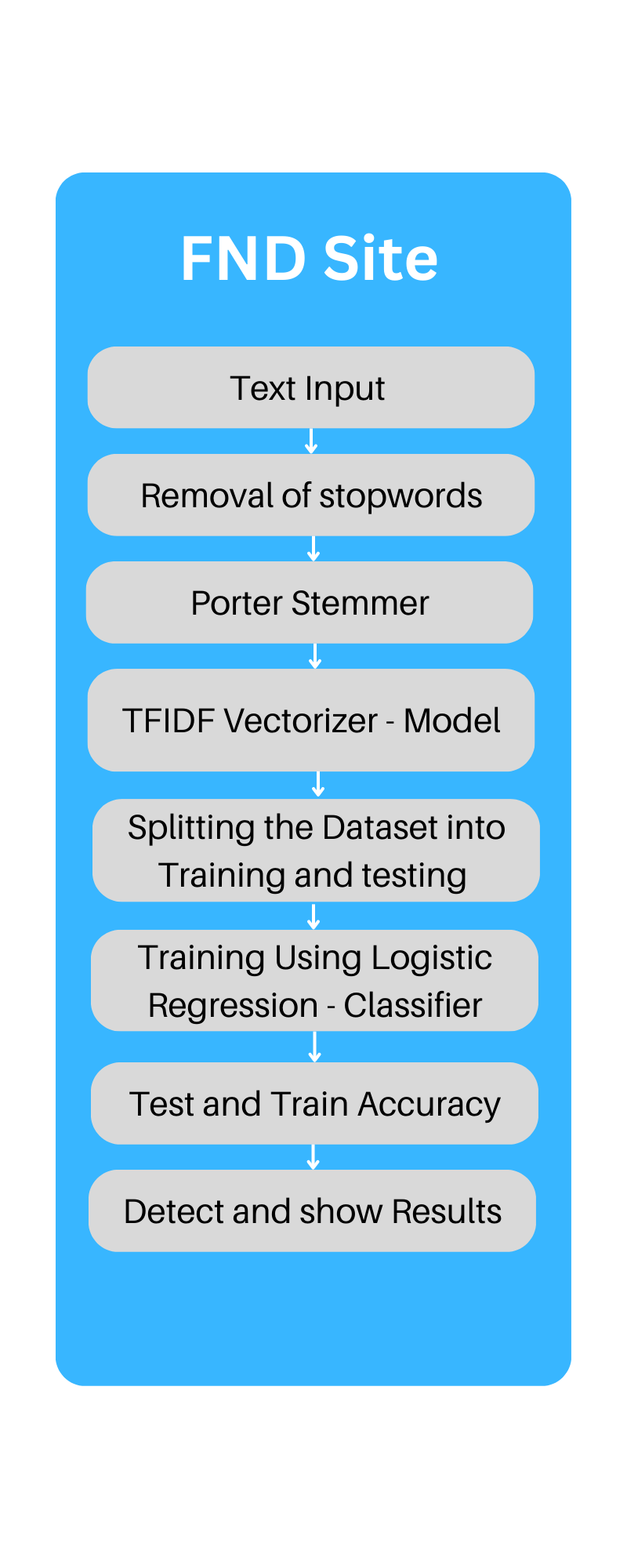
**4. RE (regular expressions):** The `re` module in Python provides functionality for working with regular expressions. We leveraged regular expressions to perform pattern matching and text manipulation operations in our text preprocessing tasks. For example, we used regular expressions to remove special characters, symbols, and patterns from the input text, ensuring cleaner and more consistent data.

**5. NLTK (Natural Language Toolkit):** NLTK is a comprehensive library for natural language processing (NLP) tasks in Python. It offers a wide range of tools, resources, and corpora for tasks such as tokenization, stemming, lemmatization, and stopwords removal. We utilized NLTK to access the stopwords corpus, which allowed us to filter out common words that do not carry significant meaning in the context of fake news detection. Additionally, we used the Porter stemming algorithm from NLTK to reduce words to their base or root form.

**6. Sklearn (scikit-learn):** Scikit-learn is a popular machine learning library that provides a wide range of algorithms and utilities for data modeling and analysis. We leveraged scikit-learn to implement the logistic regression[2] algorithm for binary classification in our fake news detection system. Additionally, we utilized the TF-IDF[3] vectorizer from scikit-learn to convert textual data into numerical feature vectors, enabling the training of our machine learning model.

By harnessing the capabilities of these libraries, we were able to preprocess the data, extract relevant features, train the model, and evaluate its performance, ultimately empowering our fake news detection system with effective and efficient algorithms and tools.

**ARCHITECTURE**

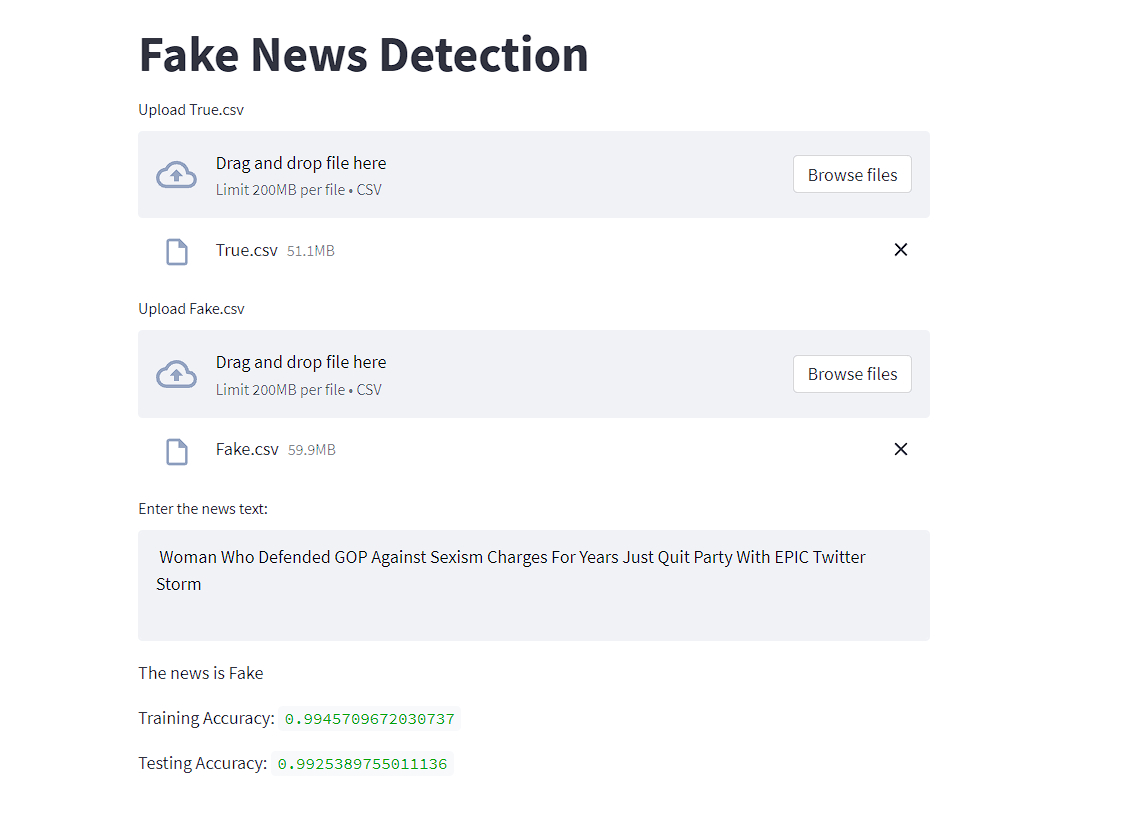
****

**Fig.1: Architecture diagram of Fake News Detection**

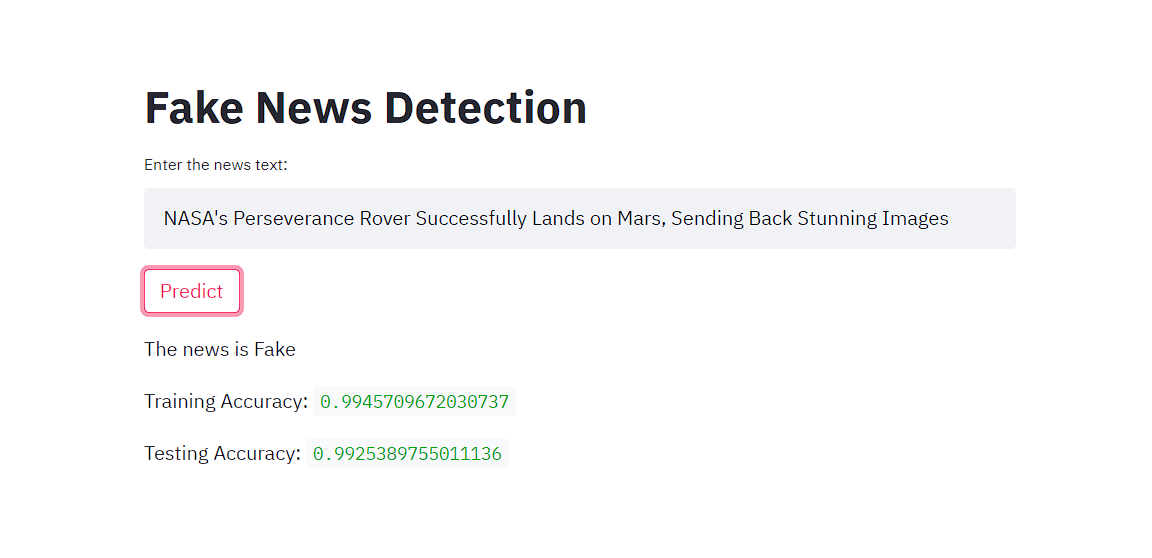
**RESULTS**

**If the news is Fake:**

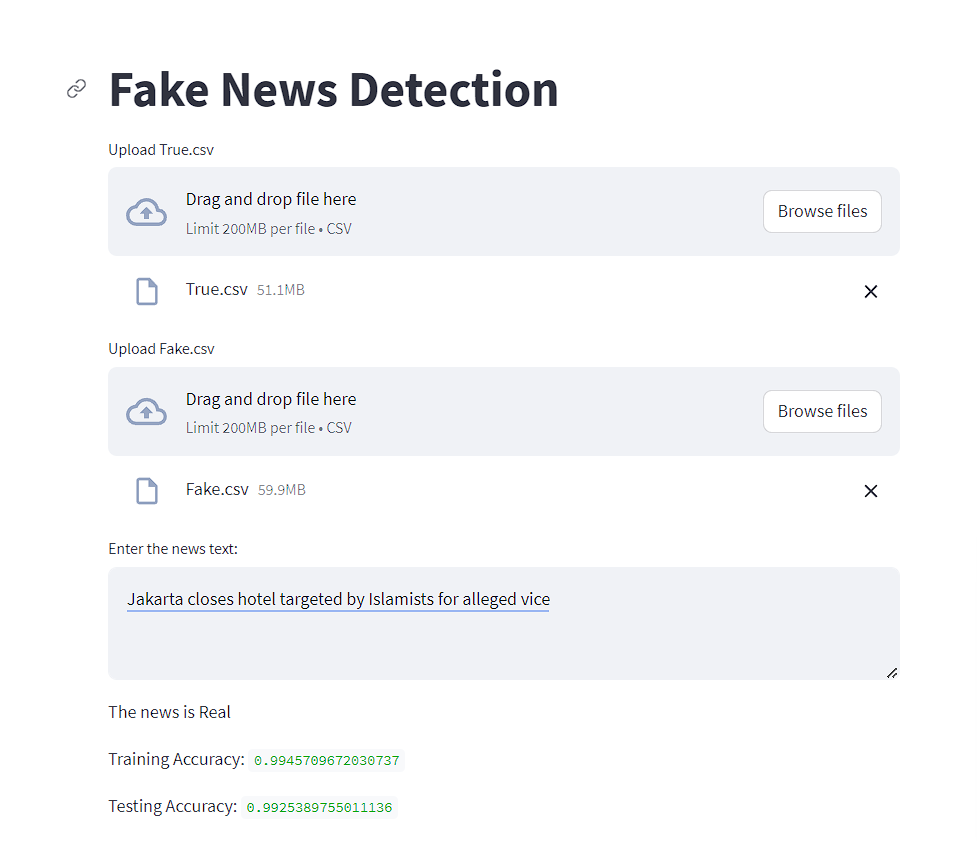
**Dynamic**

****

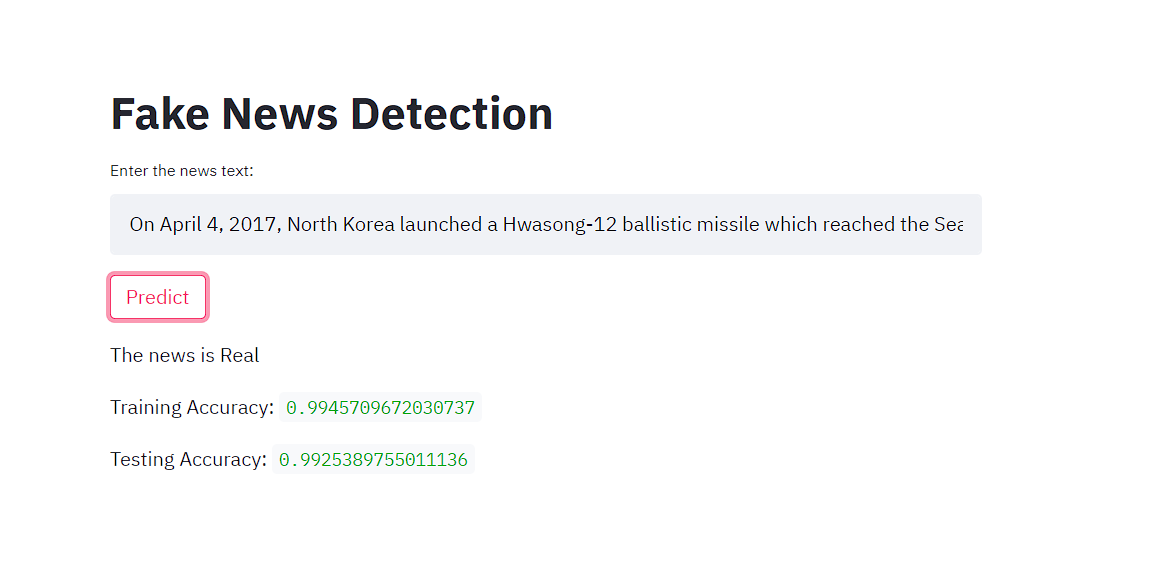
**Static**

****

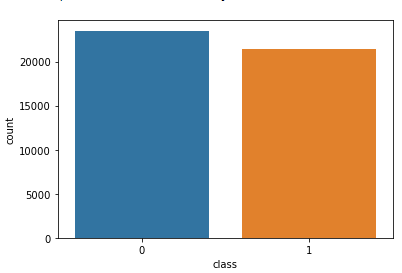
**If the News is Real:**

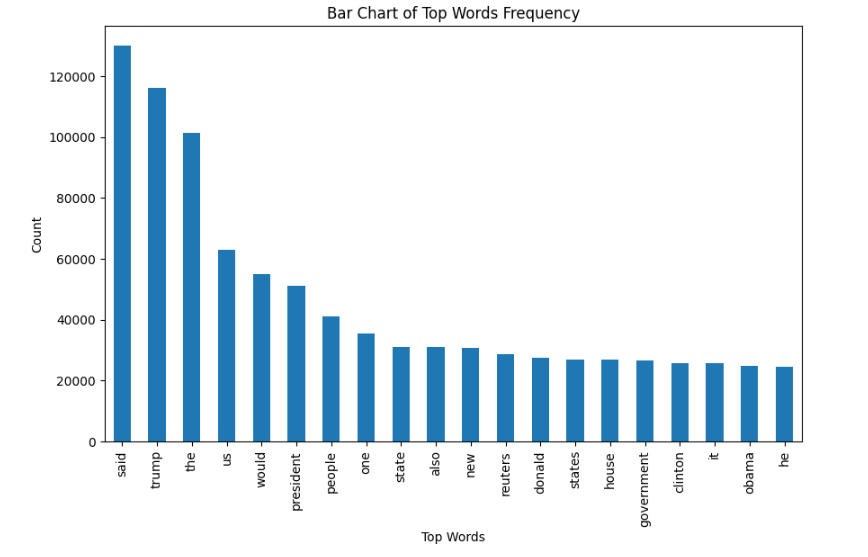
****

**Dynamic**



**Static**





**CONCLUSION**

In conclusion, our project focused on developing a fake news detection system using machine learning techniques. We utilized a dataset consisting of news articles with features such as Title, Text, Subject, and Date.

By preprocessing[1] the dataset and transforming the textual data into numerical features using TF-IDF[3] vectorization, we trained a logistic regression[2] model. We evaluated the model's accuracy on both the training and testing data, which provided insights into its performance.

To enhance usability, we created a user-friendly interface using Streamlit. This interface allows users to input news text and obtain predictions regarding its authenticity, enabling quick verification of news articles.

Our project demonstrated promising results in detecting fake news, indicating the effectiveness of our approach in distinguishing between real and fake information. However, continuous updates and improvements are necessary to address evolving fake news techniques.

In summary, our project highlights the potential of machine learning in combating the proliferation of fake news. By providing a reliable tool for news verification, we contribute to promoting accurate and trustworthy information dissemination in the digital

age.

**REFERENCES**

**[1]** [**Preprocessing**](https://www.ijert.org/fake-news-detection#:~:text=Pre%2Dprocessing%20is,graphs%2C%20documents%2C%20etc.))

**[2]** [**Logistic Regression**](https://www.ijert.org/fake-news-detection#:~:text=Early%20withinside%20the%20twentieth%20century%2C%2C%20the%20biological%20sciences%20began%20to%20employ%20logistic%20regression.%20Then%2C%20it%20was%20put%20to%20much%20different%20social%20science%20uses.%20When%20the%20dependent%20variable%20(target)%20is%20categorical%2C%20logistic%20regression%20is%20utilized.)

**[3]** [**TFIDF**](https://www.geeksforgeeks.org/understanding-tf-idf-term-frequency-inverse-document-frequency/)

**FUTURE SCOPE**

Our Fake News Detection project has exciting future prospects. We can explore advanced techniques like neural networks and natural language processing to improve the accuracy of our model and handle complex language patterns. By incorporating sentiment analysis and fact-checking APIs, we can gain more context for news verification. Expanding our dataset to include a wider range of sources and genres will make our model more versatile. Additionally, integrating social media platforms and real-time news feeds will allow us to detect and debunk fake news in real-time. Collaborating with media organizations and fact-checking agencies will help us make a greater impact in the fight against misinformation.

**SOURCE CODE AND CONFIGURATION FILES**

**Dynamic Upload**

**<https://github.com/Theagentvikram/FakeNewsDetection/blob/dev/FakenewsDet2.py>**

[**https://Fakenews.streamlit.app**](https://Fakenews.streamlit.app)

**Static Upload**

[**https://github.com/Theagentvikram/FakeNewsDetection/blob/main/StT.py**](https://github.com/Theagentvikram/FakeNewsDetection/blob/main/StT.py)

[**https://fakenewsdetection.streamlit.app/**](https://fakenewsdetection.streamlit.app/)