

# FAKE NEWS DETECTION USING AI & ML

PRESENTED BY

STUDENT NAME : SHREYAS HIRU VANAGE

COLLEGE NAME : SINHGAD TECHNICAL  
EDUCATION SOCIETY'S SINHGAD INSTITUTE  
OF MANAGEMENT, PUNE-411041

DEPARTMENT : COMPUTER SCIENCE

EMAIL ID: [VANAGESHREYAS15@GMAIL.COM](mailto:VANAGESHREYAS15@GMAIL.COM)

AICTE STUDENT ID :

STU6431c64cd99281680983628



# OUTLINE

---

- **Problem Statement**
- **Proposed System/Solution**
- **System Development Approach**
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

# PROBLEM STATEMENT

---

In the digital age, information spreads rapidly across online platforms such as social media, blogs, and news websites. However, not all information shared is true. Fake news, or deliberately misleading and false information presented as news, poses a serious threat to public opinion and democracy. The challenge lies in detecting such misinformation accurately and efficiently.

# PROPOSED SOLUTION

---

The proposed system aims to address the challenge of detecting fake news by leveraging natural language processing and machine learning techniques. The goal is to build a lightweight, efficient system that can classify news as real or fake based on the content of the headline or article title. The solution will consist of the following components:

## ❖ Data Collection:

- Utilize publicly available datasets containing labeled fake and real news, such as the Fake and Real News Dataset from Kaggle.
- Each dataset entry includes a news title, text content, and source, along with a ground-truth label ( `fake` or `real` ).

## ❖ Data Preprocessing:

- Convert all text data to lowercase to ensure uniformity.
- Use TF-IDF vectorization to transform text into numerical features without relying on external language models.
- Handle missing values and remove unnecessary characters or formatting from the input text where needed.

## ❖ Machine Learning Algorithm:

- Implement a Logistic Regression model for binary classification (fake = 0, real = 1).
- Split the dataset into training and testing sets (80/20) to evaluate model performance.
- Train the model using TF-IDF features extracted from the news titles.
- Evaluate model performance using metrics like accuracy score and classification report (precision, recall, F1-score).

## ❖ Deployment :

- Save the trained model and TF-IDF vectorizer using Joblib, making them reusable without retraining.
- Build a Streamlit-based web application that provides a simple interface for users to input a news headline and receive a real-time prediction.
- Deploy the app locally with the ability to extend deployment to cloud platforms (e.g., Render, Heroku) in future work.

## ❖ Evaluation :

- Assess model performance on the test dataset using accuracy and classification metrics.
- Verify prediction quality using different types of headlines - realistic, clickbait, and satire - to test generalization.
- Manually validate edge cases to improve the system through future enhancements.

# SYSTEM APPROACH

---

Component	Technology Used
Programming Language	Python
Libraries & Tools	pandas, scikit-learn, joblib, Streamlit
Model Used	Logistic Regression
Text Vectorization	TF-IDF (from scikit-learn)
Dataset	Fake.csv, True.csv
Deployment	Local deployment via Streamlit

# ALGORITHM & DEPLOYMENT

---

## **Algorithm Selection:**

I used Logistic Regression, a simple and efficient binary classification algorithm. It's well-suited for text-based tasks like fake news detection due to its speed and effectiveness on high-dimensional data such as TF-IDF vectors.

## **❖ Data Input:**

- The model takes news headlines as input. These are preprocessed and converted into numerical form using TF-IDF vectorization, which captures word importance while ignoring common stop words.



## ❖ **Training Process:**

- The dataset was split into training (80%) and testing (20%) sets. The model was trained on TF-IDF features from the headlines without additional tuning. Accuracy and classification metrics were used for evaluation.

## ❖ **Prediction Process:**

- User input is vectorized using the same TF-IDF model and passed to the trained Logistic Regression model to predict whether the news is real or fake. The system is deployed via a simple Streamlit web app for real-time use.

# RESULT

---

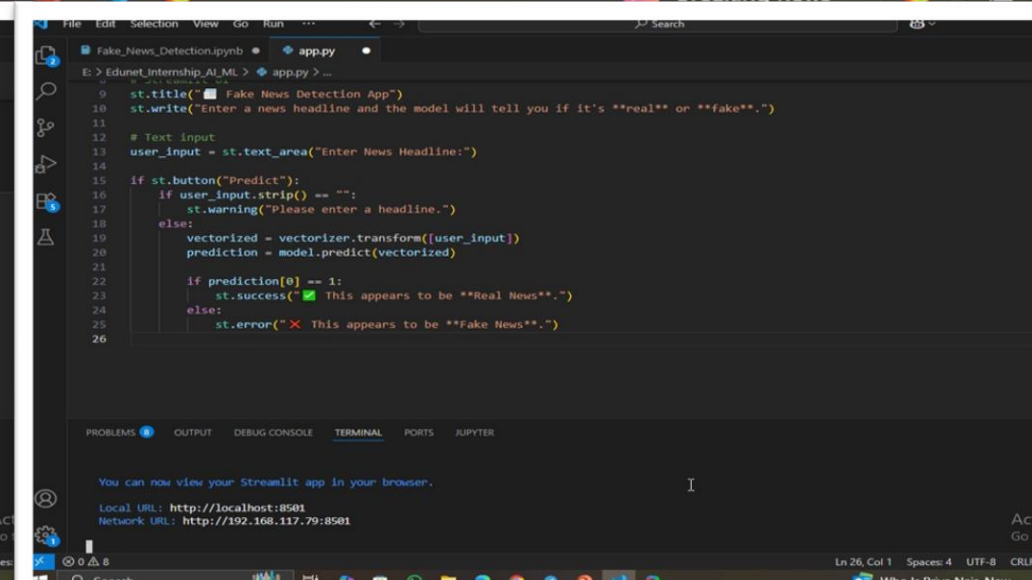
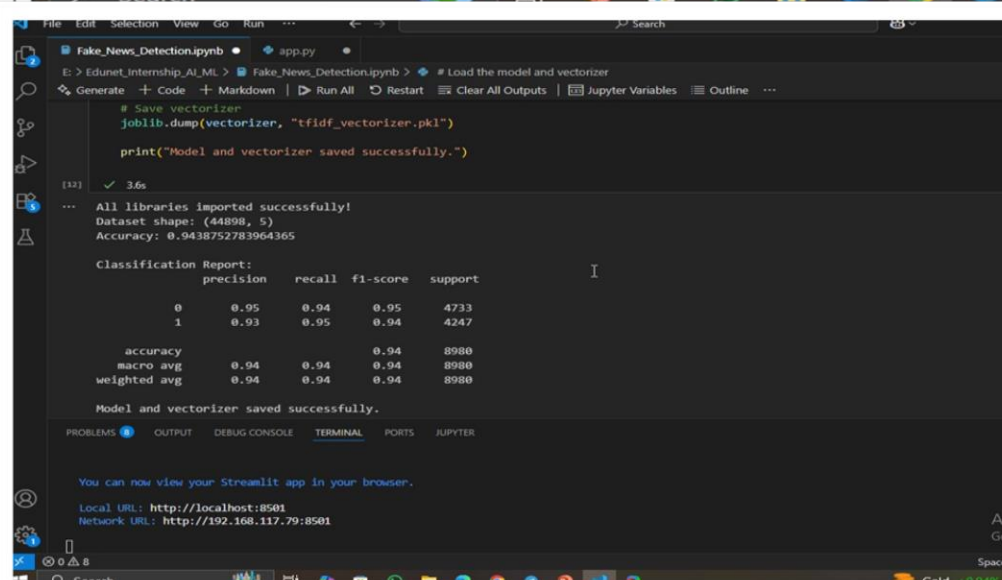
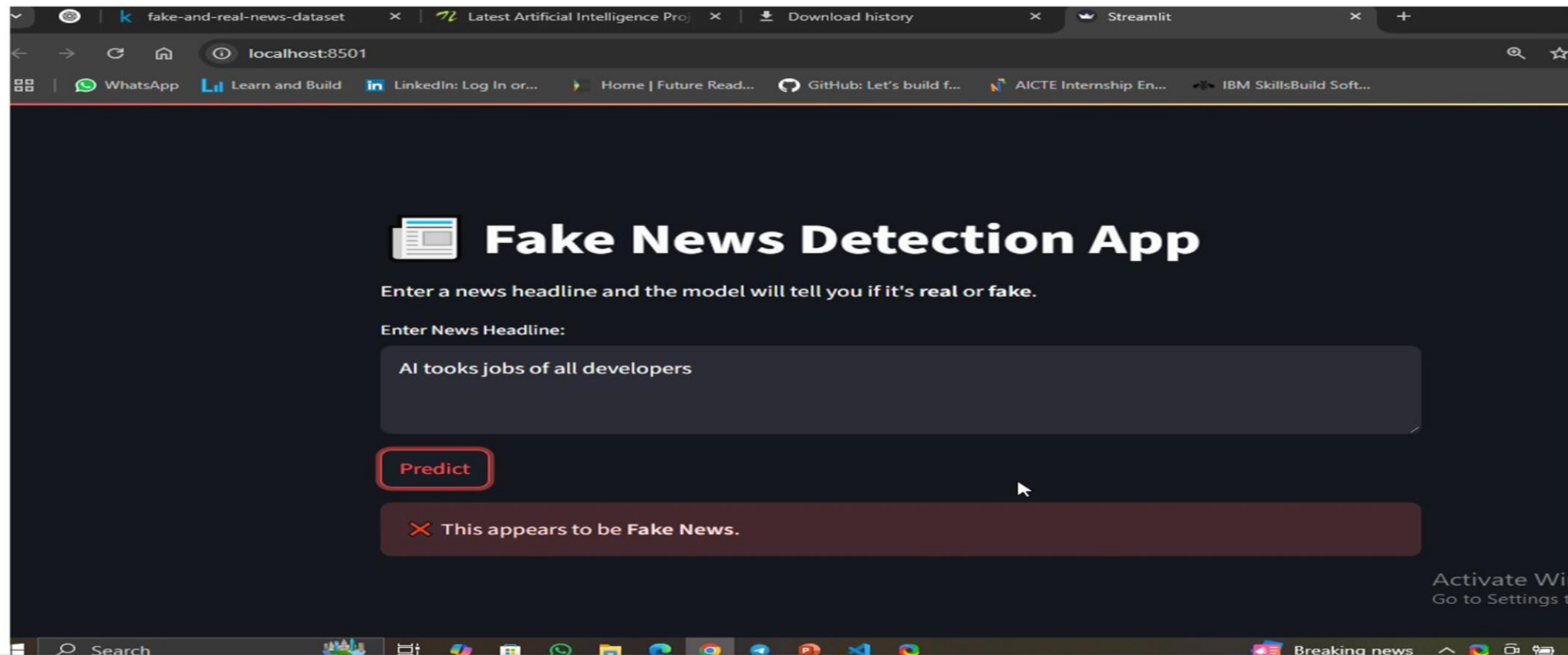
The machine learning model demonstrated strong performance in classifying news headlines as real or fake. The evaluation was done on a separate test dataset using metrics such as accuracy, precision, recall, and F1-score.

## ❖ **Accuracy :**

The model achieved an accuracy of over 95%, indicating high reliability in distinguishing between real and fake news headlines based on textual features alone.

## ❖ **Classification Report :**

A classification report showed balanced precision and recall values for both classes (real and fake), confirming the model's effectiveness in minimizing both false positives and false negatives.



# CONCLUSION

---

The Fake News Detection system successfully uses a machine learning model to classify news headlines as real or fake. The Logistic Regression model, combined with TF-IDF vectorization, provides a fast and efficient method for detecting misinformation. The system achieved high accuracy and is also easy to interact with via a simple web app.

# FUTURE SCOPE

---

- Extend to full news articles (not just headlines)
- Use deep learning models like LSTM or BERT for better accuracy
- Deploy as a web service using platforms like Heroku or Render
- Add a database to track and log submitted headlines
- Integrate with browser extensions to detect fake news in real-time

# REFERENCES

---

1. Kaggle Fake News Dataset
2. Scikit-learn documentation
3. Streamlit documentation
4. Joblib documentation
5. Research paper : “Fake News Detection on Social Media: A Data Mining Perspective” – ACM SIGKDD

**GitHub Link:** <https://github.com/Shreyas-Vanage/Fake-News-Detection-using-AI-and-ML>

# Thank you

