Fake News Classifier – Internship Project Report

Intern Name: Muhammad Awuf A.J

Department: B.Tech – Artificial Intelligence and Data Science

Internship Organization: Elevate Labs

Submission Date: July, 2025

1. Introduction

With the rapid spread of information on digital platforms, distinguishing between real and fake news has become a critical challenge. Misinformation not only affects public opinion but also has serious consequences on society. The aim of this project is to build a machine learning-based classifier that can accurately detect whether a news article is fake or real by analyzing its content.

2. Abstract

In this project, I developed a Fake News Classifier using Natural Language Processing (NLP) and machine learning techniques. The project pipeline includes text preprocessing using NLTK, feature extraction using TF-IDF Vectorizer, and model training using Logistic Regression and Naive Bayes. A Streamlit-based user interface was developed to allow users to input news content and instantly receive a classification result. Logistic Regression achieved 99% accuracy, making it the preferred model for deployment.

3. Tools Used

- Python (Programming Language)

- Jupyter Notebook (for development)

- Libraries:

  - pandas, numpy (Data handling)

  - NLTK (Text preprocessing)

  - scikit-learn (Model building)

  - matplotlib, seaborn (Visualization)

  - Streamlit (Web-based UI)

- pickle (for model saving)

4. Steps Involved in Building the Project

- Step 1: Dataset Collection

Two datasets (Real.csv and Fake.csv) were downloaded from Kaggle and combined with appropriate labels.

- Step 2: Text Preprocessing

Used NLTK to tokenize the text, convert to lowercase, remove punctuation and stopwords.

- Step 3: Feature Extraction

Applied TF-IDF Vectorizer (max 5000 features) to convert cleaned text into numerical vectors.

- Step 4: Model Training

Trained Logistic Regression and Naive Bayes classifiers using scikit-learn. Logistic Regression performed best with 99% accuracy.

- Step 5: Evaluation

Used accuracy score, classification report, and confusion matrix to evaluate both models.

- Step 6: Model Saving

Saved the trained model and vectorizer using pickle for deployment.

- Step 7: Streamlit UI

Developed a web interface using Streamlit to allow users to paste a news article and check whether it is fake or real.

5. Conclusion

The Fake News Classifier project successfully demonstrates how machine learning and NLP can be applied to detect misinformation. Logistic Regression was the most effective model for this dataset. The project strengthened my skills in data preprocessing, model evaluation, and deploying ML models through a web interface. This application can be further expanded by including more complex models and larger datasets in future iterations.