Project Title: News Article Classification (Fake/Real)
Name: Hadeel Ismail
Internship Duration: 2 Weeks
Tools Used: Python, NLTK, Sklearn, Pandas, Streamlit

1. Introduction

With the rise of misinformation, detecting fake news has become critical. This project aims to classify news articles as **real** or **fake** using natural language processing and machine learning techniques. A Streamlit web app was built to allow real-time predictions.

2. Abstract

A Naive Bayes classifier was trained on a dataset of labeled news articles using the TF-IDF vectorization method. The model was deployed using a simple Streamlit interface. It enables users to paste any news content and instantly check if it's real or fake based on the model's prediction.

3. Tools & Technologies Used

- **Python** for scripting
- **NLTK** for text preprocessing
- **Sklearn** for TF-IDF and model training
- **Pandas** for data handling
- **Streamlit** for building the web app interface

4. Steps Involved in Building the Project

1. Collected and merged real/fake news datasets
2. Cleaned the text using NLTK (stopwords, punctuation)
3. Applied TF-IDF vectorization
4. Trained a Naive Bayes classifier
5. Evaluated model using accuracy and F1-score (achieved ~94.6% accuracy)
6. Created a Streamlit UI to input and classify news text
7. Tested and validated with real-time examples

5. Conclusion

This project demonstrated how machine learning and NLP can be used to fight misinformation. The model shows high accuracy and the Streamlit app provides a user-friendly interface for real-time prediction. This project enhanced my understanding of ML pipelines and real-world app deployment.