Name:-

Zainab Batool

Sap ID:-

70140186

Section:-

Artificial Intelligence(7E)

Submitted to:-

Sir Asif Ahsan

Assignment topic:-

Project Proposal: AI-Based Fake News Detection System

1. Introduction

The rapid spread of digital information has made news consumption easy but has also increased the risk of **fake news** dissemination. Fake news refers to false or misleading content presented as legitimate news, which can negatively influence public opinion and societal behavior. Manual verification of vast amounts of online news is impractical due to scale and speed. Therefore, automated fake news detection systems using **Artificial Intelligence (AI)** and **Machine Learning (ML)** are becoming essential tools for improving information reliability.

This project aims to build an AI system that classifies a news article as *Fake* or *Real* by analysing its text content. The system uses **Natural Language Processing (NLP)** techniques to convert text into numerical features and trains a machine learning classifier to make predictions.

2. Problem Statement

Fake news has the potential to mislead readers and disrupt decision-making at personal, social, and political levels. Traditional manual fact-checking is time-consuming and often fails to keep up with the volume of online content. Therefore, the need for an automated fake news detection solution is critical. This project addresses the challenge of identifying fake news articles using machine learning by analyzing text patterns rather than relying on manual human review.

3. Objectives

The primary objectives of this project are:

1. To preprocess and merge labeled fake and real news datasets.
2. To extract textual features using TF-IDF (Term Frequency–Inverse Document Frequency).
3. To train a machine learning model that accurately classifies news as fake or real.
4. To prepare a complete software with documentation and GitHub repository for academic submission.

4. Dataset

The dataset used in this project is obtained from **Kaggle**, a publicly available repository for data science resources. It consists of two CSV files:

- **Fake.csv** — news articles labeled as fake
- **True.csv** — news articles labeled as real

Due to file size limitations, the dataset files are referenced from Kaggle rather than directly uploaded to GitHub. This dataset is widely used in similar research and practical projects.

Reference: *Fake and Real News Dataset on Kaggle*

5. Methodology

The proposed fake news detection system follows these steps:

1. **Data Loading & Preparation:** Load the fake and real news articles from CSV files and assign labels.
2. **Text Preprocessing:** Clean text by lowering case, removing stop words, and eliminating noise.
3. **Feature Extraction:** Apply **TF-IDF Vectorization** to convert text into numerical features that machine learning models can interpret.
4. **Model Training:** Train a **Multinomial Naive Bayes** classifier using the TF-IDF features.
5. **Evaluation:** Test the model on a portion of the dataset and calculate performance metrics such as accuracy.

6. Expected Results

The trained model is expected to classify news articles into fake or real categories with reasonable accuracy suitable for a semester project. While simpler models like Naive Bayes may not achieve production-level performance, they are easy to explain, computationally efficient, and effective with textual features such as TF-IDF.

7. Tools and Technologies

- **Programming Language:** Python
- **Libraries:** Pandas, Scikit-learn (TF-IDF, Naive Bayes)
- **Dataset Source:** Kaggle
- **Version Control:** GitHub

8. Limitations

- The model depends solely on text content and cannot consider images or videos.
- Classification quality is limited by how representative the dataset is.
- Models like Naive Bayes assume word independence, which may not hold for all news text.

10. Reference Points

1. Fake News Detection Using Naive Bayes Classifier (Journal Article)
This study uses TF-IDF and Naive Bayes for fake news detection, supporting the methodology used in this project.
🔗 https://jmss.a2zjournals.com/index.php/mss/article/view/22 (jmss.a2zjournals.com)

2. Detection of Hoax News Using TF-IDF and Naive Bayes (Conference/Journal)
This research demonstrates Hoax news detection using TF-IDF and machine learning classifiers.
🔗 https://mjgcs.mase.or.id/index.php/mjgcs/article/view/24 (mjgcs.mase.or.id)

3. Enhancing Fake News Detection with Hybrid NLP & ML (Journal Paper)
This paper discusses combining TF-IDF with machine learning classifiers like Logistic Regression and SVM to improve fake news detection results on a Kaggle dataset.
🔗 https://www.icck.org/article/abs/TIS.2024.461943 (ICCK)