News Article Classification (Fake/Real)

1. Abstract

This project aims to classify news articles as either Fake or Real using machine learning models. By leveraging natural language processing techniques and classification algorithms, the model can analyze textual data and help in the identification of misinformation. This system can be useful in today's world where fake news can have significant consequences.

2. Introduction

With the rise of social media and digital platforms, the spread of fake news has become a serious issue.

Automated systems capable of distinguishing between real and fake news can help combat misinformation. This project utilizes a dataset of labeled news articles and applies various machine learning techniques to build an effective classifier.

3. Tools Used

- Python 3
- Jupyter Notebook
- Pandas, NumPy
- Matplotlib, Seaborn
- Scikit-learn
- NLTK (Natural Language Toolkit)
- WordCloud

4. Steps Involved in Building the Project

1. **Data Collection**: Merged `Fake.csv` and `True.csv` datasets.
2. **Preprocessing**: Removed punctuation, stopwords, and normalized text.
3. **EDA**: Visualized word distributions and word clouds.
4. **Feature Extraction**: Applied TF-IDF vectorization.
5. **Model Training**: Trained models like Logistic Regression, Naive Bayes, Random Forest, and SVM.
6. **Evaluation**: Assessed performance using accuracy, confusion matrix, and classification report.

5. Conclusion

The fake news detection system built using machine learning models proved to be effective in distinguishing between real and fake news. With proper text preprocessing and evaluation metrics, the models demonstrated promising results. This project lays the foundation for future advancements such as using deep learning models or deploying it as a web application.