

News article classification (fake/real)

1. Introduction :

- This project is based on Fake News Detection using NLP and Machine Learning.
- In the digital era, misinformation spreads rapidly through online platforms. To combat this, our project focuses on building a machine learning model that classifies news articles as either fake or real using natural language processing techniques.

2. Abstract :

- The project involves the collection and cleaning of a labeled news dataset containing real and fake news articles.
- We used NLP preprocessing like lemmatization and stopwords removal, followed by TF-IDF vectorization. A Logistic Regression classifier was trained on the processed data.
- The final model achieves high accuracy on unseen news samples. The model is deployed using Streamlit, allowing users to paste a news article and receive a prediction.

3. Tools Used :

- Python
- Pandas, NLTK
- Scikit-learn
- Streamlit
- Pickle (for saving model)
- Jupyter Notebook (for development)

4. Steps Involved in Building the Project :

1. Loaded and merged real and fake news datasets from Kaggle.
2. Preprocessed text: lowercasing, punctuation removal, lemmatization, and stopwords filtering.
3. Converted text into numerical features using TF-IDF vectorization (with n-grams).
4. Trained a Logistic Regression classifier and evaluated it using accuracy, F1-score, and confusion matrix.
5. Saved the trained model and vectorizer using pickle.
6. Built a web interface using Streamlit for real-time news classification.
7. Deployed the app on Streamlit Cloud to make it accessible online.

5. Conclusion:

This project demonstrates the effectiveness of natural language processing and machine learning in detecting fake news. The Logistic Regression model performs well and serves as a valuable tool to validate the authenticity of online content. Future improvements can include trying advanced models, multilingual support, and integration with fact-checking APIs.