



Fake News Detection Using Machine Learning

This project report from Bundelkhand University explores an intelligent system to detect fake news automatically using Machine Learning and Natural Language Processing. The system analyses news text to classify articles as Real or Fake, helping combat misinformation in the digital era. It demonstrates practical data science applications to maintain information integrity on online platforms.

Project Objectives and System Analysis

Project Objectives

- Develop ML model to classify news as fake or real
- Preprocess data using NLP techniques
- Evaluate algorithms like Logistic Regression and Naive Bayes
- Implement user-friendly Streamlit interface

System Analysis

The system accepts news text, preprocesses it, converts it to numerical vectors via TF-IDF, and predicts authenticity using ML models. It aims for quick, accurate predictions with an intuitive UI, scalable for large datasets.



Feasibility Study

Technical Feasibility

Uses open-source Python libraries like scikit-learn, NLTK, and Streamlit; no special hardware needed.

○ Operational Feasibility

Simple web interface allows easy use by non-technical users with fast real-time predictions.

○ Economic Feasibility

No licensing costs; development on personal or university machines; minimal expenses.

Time Feasibility

Project completed within 4–6 weeks, covering data prep, model training, UI design, and testing.



Software and Hardware Requirements

Software

- Windows 11 OS
- Python 3.12, Jupyter Notebook
- Libraries: scikit-learn, pandas, numpy, nltk, streamlit
- Browsers: Chrome, Bing

Hardware

- Intel Core i3 or equivalent processor
- Minimum 4 GB RAM (8 GB recommended)
- 5 GB free disk space
- 13" or larger display recommended

System Design and Architecture



Data Preprocessing

Cleaning news text: removing special characters, stopwords, and stemming.

Feature Extraction

Transforming cleaned text into TF-IDF numerical vectors.

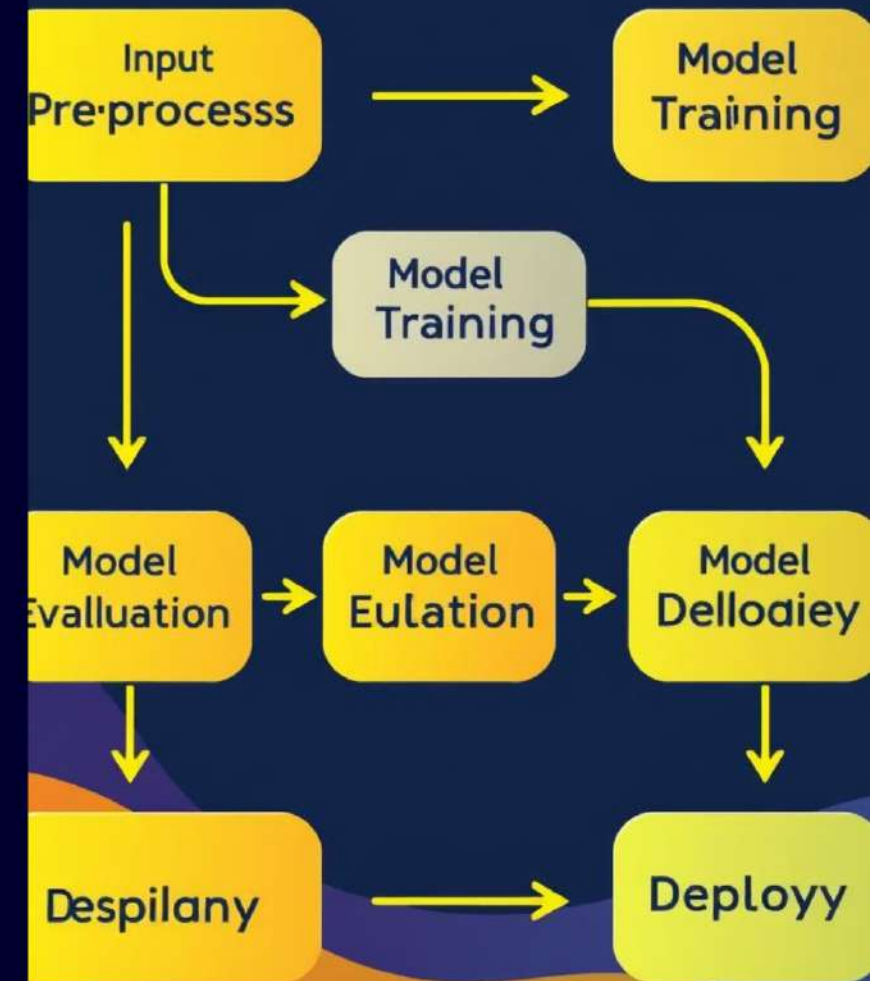
Model Training & Prediction

Using Logistic Regression and Naive Bayes to classify news articles.

User Interface

Streamlit-based web app for input and displaying classification results.

Machine Learning



Coding and Implementation Highlights

Preprocessing Code

Uses NLTK for stopword removal and Porter Stemmer for word normalization.

Regular expressions clean non-alphabetic characters.

Model Training

Splits data into training and test sets; trains Logistic Regression and Naive Bayes models.

Evaluates models using accuracy, precision, recall, and F1-score.



Testing, Validation, and Model Evaluation

Validation Checks

Ensures input is non-empty, cleansed, and properly formatted before prediction.

User warnings for empty or invalid inputs improve reliability.

Model Performance

- Logistic Regression Accuracy: 98.6%
- Naive Bayes Accuracy: 92.75%
- Logistic Regression chosen for deployment due to superior metrics

Conclusion and Future Scope

Conclusion

The project successfully built an accurate fake news detection system using ML and NLP, with a user-friendly interface.

It aids in combating misinformation and demonstrates practical application of data science skills.

Future Enhancements

- Support for multiple languages and deep learning models
- Real-time news API integration and cloud deployment
- Mobile apps, browser extensions, and fact-checking integration
- User authentication, feedback system, and model retraining features