

Synopsis on

Fake News Detection Using Machine Learning

Submitted in partial fulfilment of the requirement for
the award of the degree of

MASTER OF COMPUTER APPLICATION

In The Discipline of Computer Science & Engineering

SACHIN MAURYA

O23MCA110402

Department of Computer Science and Engineering

Chandigarh University

Mohali, Punjab

2023-2025

Objective

The main objective of this project is to **develop an automated system** that can intelligently detect **fake news articles** using the power of **Machine Learning (ML)** and **Natural Language Processing (NLP)**. With the increasing spread of false or misleading news across digital platforms, especially social media, it has become crucial to create systems that can help identify and filter out such content to protect public opinion and promote authentic information sharing.

This system will **analyse the text** of news articles (title, author, content) and classify whether the news is **real or fake**. The classification will be based on patterns learned from previously labelled datasets using **supervised learning algorithms**.

Key goals include:

1. Automating Fake News Detection

The primary objective of this project is to **automate the process of identifying fake news** using modern technologies such as **Machine Learning (ML)** and **Natural Language Processing (NLP)**. With the massive growth of digital content and the viral nature of social media, **fake news has become a serious threat** to public trust, democracy, and societal stability. Manual verification of every piece of news is practically impossible due to the volume and speed at which content spreads. Hence, there is a clear need for an **automated solution**.

This project aims to:

- **Develop a system** that can automatically classify news as **real or fake** based on the text content (title, author, and full article).
- **Utilize supervised machine learning algorithms** like Logistic Regression, SVM, or Random Forest, trained on labelled datasets to learn patterns commonly found in fake and real news.
- **Apply NLP techniques** to clean and process the raw text data, including removing stop words, stemming, and converting text into numerical features using methods like TF-IDF.
- **Evaluate the model** using standard metrics such as **accuracy, precision, recall, F1-score**, and **confusion matrix** to ensure the reliability of the system.
- **Design a user-friendly interface** (using Flask or similar tools) where users can input a news headline or article and receive an instant prediction about its authenticity.

- **Provide real-time, scalable solutions** that can assist individuals, journalists, media organizations, and fact-checkers in filtering out misleading content.

2. Enhancing Accuracy through NLP Preprocessing and ML Models

The core objective of this project is to **increase the accuracy and reliability of fake news detection systems** by leveraging advanced **Natural Language Processing (NLP)** techniques and **Machine Learning (ML) algorithms**. Traditional rule-based systems often fail to generalize across diverse types of news content. Therefore, this project focuses on **automating the detection process** with a higher degree of **precision, recall, and overall model performance**.

The goal is to:

- **Use NLP preprocessing steps** such as text cleaning, stop word removal, stemming/lemmatization, and vectorization (e.g., TF-IDF) to **transform raw news data** into structured, meaningful representations.
- **Build and train multiple ML models** such as Logistic Regression, Support Vector Machines (SVM), and others to find the most **accurate and generalizable model** for classifying fake and real news.
- **Experiment with various combinations** of preprocessing techniques and ML algorithms to **identify the best pipeline** that maximizes detection accuracy.
- **Evaluate model performance** using key metrics (accuracy, precision, recall, F1-score) and techniques (e.g., confusion matrix, cross-validation) to ensure the system is not only accurate but also robust against misleading patterns.
- **Deploy the final system** via a user-friendly interface where users can input news articles or headlines and receive predictions instantly.

By **enhancing the preprocessing pipeline and model selection process**, this project aims to deliver a **highly accurate and efficient fake news detection system**, addressing one of the most critical challenges in today's digital information era.

3. Building a User-Friendly Interface to Test News Content

An essential objective of this project is to design and develop a **user-friendly, interactive interface** that allows end users to test and verify the authenticity of news content efficiently. While the backend system involves complex Natural Language Processing (NLP) and Machine Learning (ML) operations, the frontend aims to present this functionality in a **simple, accessible, and intuitive manner**, even for non-technical users.

The specific goals of this objective include:

- **Creating a clean and responsive UI** using tools like HTML, CSS, and frameworks like Flask or Streamlet that can work across devices.
- Allowing users to **input news headlines or full article content** through text fields or file uploads.
- Providing **instant predictions** on whether the news is real or fake, based on trained ML models, with clearly marked labels (e.g., “Fake” or “Real”).
- Displaying **confidence scores or probability levels** to help users understand the certainty of the model's prediction.
- Ensuring the interface is **lightweight, fast, and visually appealing** to encourage real-time interaction and adoption.
- Including additional features such as **text summary, source link verification, or sentiment analysis** (optional for extended functionality).

This user interface bridges the gap between sophisticated backend intelligence and practical real-world usability. It ensures that the system can be **effectively utilized by journalists, educators, researchers, and the general public** for immediate fake news detection and analysis.

4. Deploying the Model Using Python and Flask for Real-Time Classification

Another critical objective of this project is the **deployment of the trained machine learning model** using **Python and the Flask framework** to enable **real-time fake news classification**. While model training and evaluation are essential stages, deployment ensures that the solution is **accessible and operational for end users beyond the development environment**.

The key elements of this objective include:

- **Utilizing Flask**, a lightweight web application framework in Python, to **create APIs and route the model's predictions** based on user input.
- Developing a **real-time prediction pipeline** where users can enter news headlines or full articles and get **instant classification results**.
- Ensuring **seamless integration** between the front-end user interface and the backend ML model through Flask endpoints.

- Implementing **efficient data preprocessing** and **model inference logic** within the deployed application to maintain speed and performance.
- Hosting the Flask application either **locally or on cloud platforms** such as Heroku, AWS, or Render to make it publicly accessible.
- Incorporating necessary **security, logging, and error handling mechanisms** to ensure a reliable and user-friendly experience.

2.Resources Required

Software Requirements:

- Programming Language: **Python 3.x**
- Libraries: **pandas, numpy, sklearn, nltk, re**
- Machine Learning Algorithms: **Logistic Regression, SVM, etc.**
- NLP Tools: **TF-IDF Vectorizer, Stopword Removal, Stemming**
- IDE: **Jupyter Notebook / VS Code**
- Web Framework: **Flask**
- Dataset: **Fake News dataset (from Kaggle or other sources)**

Hardware Requirements:

- Processor: **Intel Core i5 or higher**
- RAM: **8 GB or more**
- Storage: **At least 1 GB free space for datasets and models**

Others:

- Internet connection (for downloading libraries/datasets)
- Google Colab or Anaconda (optional cloud-based execution)