Department: Information and Communication Technology

Academic year: (2025-2026)

Subject : Capston Project

Semester: 7

Project Title: News Headline Detector

Name: Vatsal Parmar

Roll no: 92310133010

Project Definition and Scope

1. Introduction

With the age of the internet, news and information are disseminated faster than ever via social media and online portals. As convenient as this has been, it has also resulted in the unbridled transmission of false information and misinformation. This erodes public credibility and has a large impact on personal decisions and collective choices. Manually authenticating each piece of information is neither feasible nor viable. Hence, there is an increased need for systems that are automated to recognize false or misleading content more efficiently. To solve this issue, the project "News Headline Detector" has been conceived. It is based on machine learning algorithms and comes as a web application, where the user can input a headline or a link and obtain real-time feedback regarding the authenticity of the news.

2. Problem Statement

The sheer pace of growth in fake news is a grave challenge in present-day digital media. Misleading or provocative headlines tend to misinform the readers, manipulate the public perception, and help create misinformation with long-term social impacts. Manual fact-checking lags far behind the immense volume of online information. Therefore, there is an evident necessity for an automated mechanism capable of efficiently distinguishing between authentic and fictitious headlines with improved precision and consistency.

3. Objectives

• Develop and train a machine learning algorithm that is capable of distinguishing between authentic and false news headlines .

• Implement an easy-to-use web interface wherein users can enter headlines or URLs and receive immediate analysis.

• Offer prediction output along with confidence levels to enhance clarity and earn user trust.

• Validate and test the system with publicly available datasets (Fake.csv, True.csv) to provide reliable performance.

• Deploy the application on an appropriate cloud or deployment platform to make the application easily accessible and scalable.

4. Relevance to ICT Domain

The project is directly related to the ICT areas of Artificial Intelligence (AI), Natural Language Processing (NLP), and web application development.

It employs machine learning methods, specifically the Passive Aggressive Classifier with TF-IDF vectorization, to demonstrate how AI can assist in addressing the increasing problem of misinformation. From the technical perspective, the system marries a ReactJS-driven user interface with a Flask-based backend, demonstrating full-stack development capabilities. Generally, this work is representative of current industry tendencies since AI-based methods are being used more frequently to guarantee the credibility of internet information. 5. Feasibility Analysis.

- **Technical Feasibility:**

The system as proposed is built using Python as the frontend and backend framework and libraries such as scikit-learn and joblib as backend for the training and deployment of the machine learning model. Newspaper3k is utilized for news text extraction and processing, and ReactJS is for an interactive and responsive user interface. All selected technologies are open-source, actively supported, and well-suited for academic and practical applications.

- **Economic Feasibility:**

Because the implementation depends solely on open-source tools, the overall cost of development is very minimal. Deployment can be done using free or low-cost platforms like Vercel or Render, which obviates the cost of expensive infrastructure or commercial licenses. This way, the project is not only financially viable but also cheap to upkeep.

- **Ethical Considerations**

Although the system provides obvious advantages, there are ethical concerns such as biased forecasts or the potential for abuse to restrict access to specific information. For addressing these issues, the system emphasizes transparency through reasons for forecasts and confidence ratings. It is designed to utilize only for education and awareness, refraining from censorship or propaganda campaigns.

6. Market/User Needs Analysis

The News Headline Detector is made for the public audience, such as students, researchers, journalists, and general internet users who are interested in fact-checking the authenticity of news articles.

Studies disseminated in IJCRT and EJTAS report that spreading disinformation has greatly risen over the past decade, emphasizing the urgent need for automated fact-checking tools.

News from Statista and Pew Research Center also indicates that individuals are resorting to fact-checking tools more and more. The project addresses these requirements through a friendly, effective, and clear interface for assessing news headlines.

References:

IJCRT.ORG - FAKE NEWS DETECTION USING MACHINE LEARNING

EJTAS - A Comprehensive Review of Fake News Detection Approaches Using Machine Learning

JRTI - Artificial Intelligence-Based Fake News Detection: A Performance Comparison

7. Literature Review

Several researchers have discussed various approaches to detecting fake news through newer deep learning methods like LSTMs, BERT, and transformer-based models.

Although these models perform well in terms of accuracy, they are usually computationally intensive and resource-hungry.

This work employs a Passive Aggressive Classifier with TF-IDF vectorization in a lighter framework, thus making it efficient and feasible for real-time applications.

The novelty of this work is the analysis of news headlines with reduced processing overhead while maintaining reliable predictions. This approach offers real-time results along with confidence scores compared to conventional fact-checking systems that lack immediate results and confidence scores, thus increasing usability and user trust.

8. Conclusion

The News Headline Detector initiative meets the expanding issue of spotting and tracking false news on the web. Through a simple web interface and machine learning algorithms, the system provides an accessible, affordable, and ethically sound solution. Its greatest innovation is the capacity to perform rapid, real-time headline verification and supply confidence scores to further promote user trust.

Ideation and Stakeholder Needs Analysis

The spread of false information and fabricated news has become a serious problem in the current digital age. Social media websites, websites that offer news online, and blogs tend to forward news headlines without real verification, confusing the masses and influencing decision-making processes. The News Headlines Detector project seeks to tackle this problem by implementing an automated AI-based system for detecting the veracity of news headlines.

This section states the key stakeholders, examines their needs, develops a problem statement, suggests innovative solutions, and establishes relevance to the ICT sector.

1. Stakeholder Identification

Primary Stakeholders:

1.News Readers / General Public:

Require quick and reliable verification of news titles in order not to be misled.

Tend to be exposed to fake news on social media and messaging platforms.

2.Journalists and Media Houses:

Require tools for verifying content prior to publishing.

Minimize manual work in fact-checking.

3.Educational Institutions / Researchers:

Need datasets and analytics for misinformation trend studies.

Utilize automated tools in analyzing fake news dissemination patterns.

Evidence Support:

Pew Research Center, News Consumption in the Digital Age, 2023 – reveals that more than 64% of adults come across misinformation online every week.

McKinsey & Company, Misinformation and Media Trust, 2021 – emphasizes the need for automated verification tools in journalism.

2. Stakeholder Needs Analysis

- News readers need simplicity and speed.
- Journalists require accuracy and reliability.
- Researchers require structured data for insights and reporting.

Stakeholder	Specific Needs	Example Use Case
News Readers	Instant verification of headlines	User pastes headline in tool → receives real/fake result in seconds
Journalists	Automated fact-checking to reduce errors	Journalist checks multiple headlines before publishing
Researchers	Access to labeled datasets and analytics	Research on misinformation trends, visualization of fake news patterns

1.1 Table : Stake Holders relation with their needs and its use cases

3. Problem Statement

The emergence of false news poses a major threat to the propagation of reliable information. News verification currently relies mostly on manual processes, which are both time-consuming and unavailable to common readers. The general public, journalists, and researchers do not have an automated, robust, and user-friendly system to identify legitimate news from fabricated headlines. The problem is to create an system that utilizes artificial intelligence for rapid, precise, and scalable news verification.

4. Ideation of Solutions

AI-powered News Headline Verification Tool

Leverages machine learning (PassiveAggressiveClassifier) and natural language processing (TF-IDF vectorization) to label headlines as authentic or fabricated.

Benefit to stakeholders: Rapid verification for users and journalists; precise detection minimizes misinformation.

Browser Plugin / API Integration

Interacts with social media sites and news platforms to identify suspicious headlines in real-time.

Benefit to stakeholders: Informs users immediately; enables journalists to pre-screen content effectively.

Researchers' Analytics Dashboard

Offers visualizations, statistics, and reports of fake vs real news trends.

benefits stakeholders: Facilitates misinformation research and policymaking allows for dataset export for academic research.

5. Relevance to ICT Domain

The project aligns with existing ICT , including:

Machine Learning & NLP: Computerized news headline analysis.

Web Development: Frontend (React/HTML) and backend (Flask API) integration for engaging user experience.

Cloud / Edge Deployment Potential: Scalable deployment to all users.

Potential Impact:

Minimizes exposure to misinformation for news readers.

Assists journalists in ethical reporting.

Facilitates research on misinformation trends based on real datasets.

ICT Feasibility:

ML model based on Python, Flask API, and React frontend ensure seamless integration and ease of maintenance.

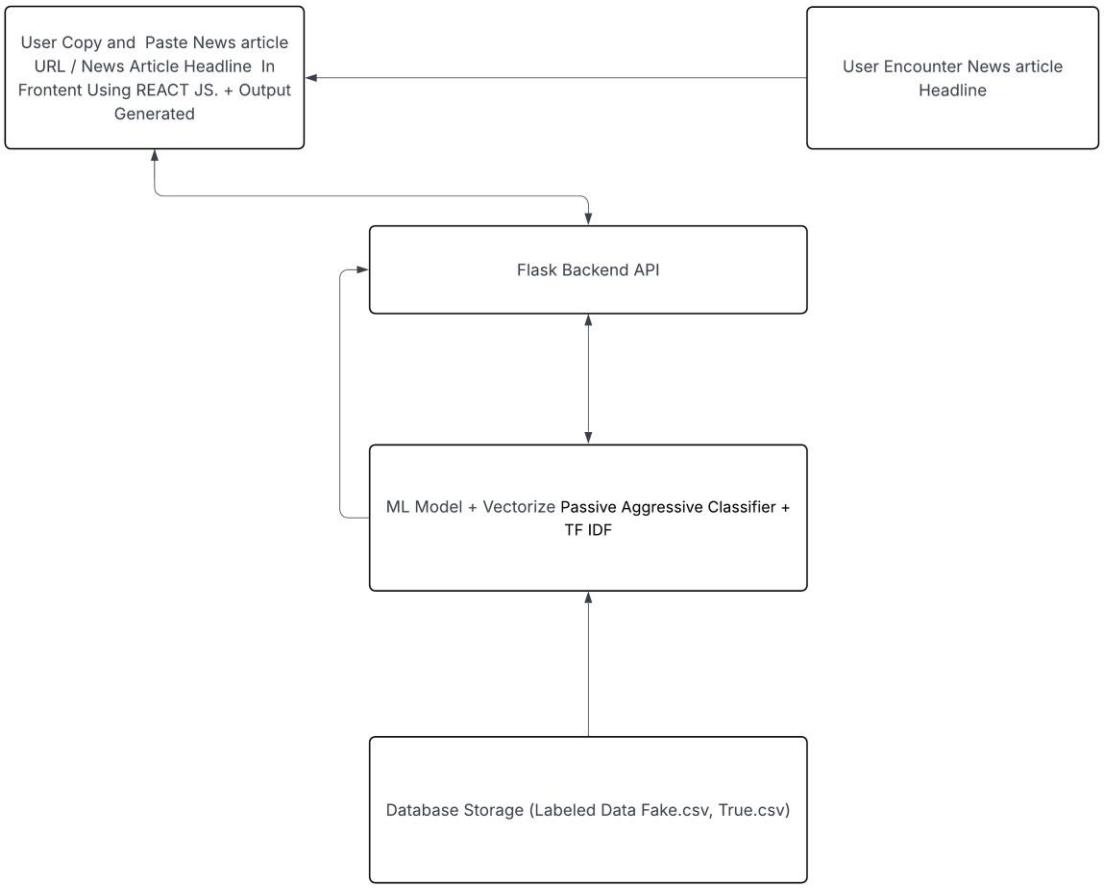
System Design and Architecture

The Fake News Detector system is designed to help users validate whether a given news headline or article is real or fake. The solution uses a machine learning model (Passive Aggressive Classifier with TF-IDF vectorization) trained on real and fake news datasets. The architecture integrates a Python/Flask backend, a ReactJS frontend, and a model persistence layer for classification tasks.

Modular Design

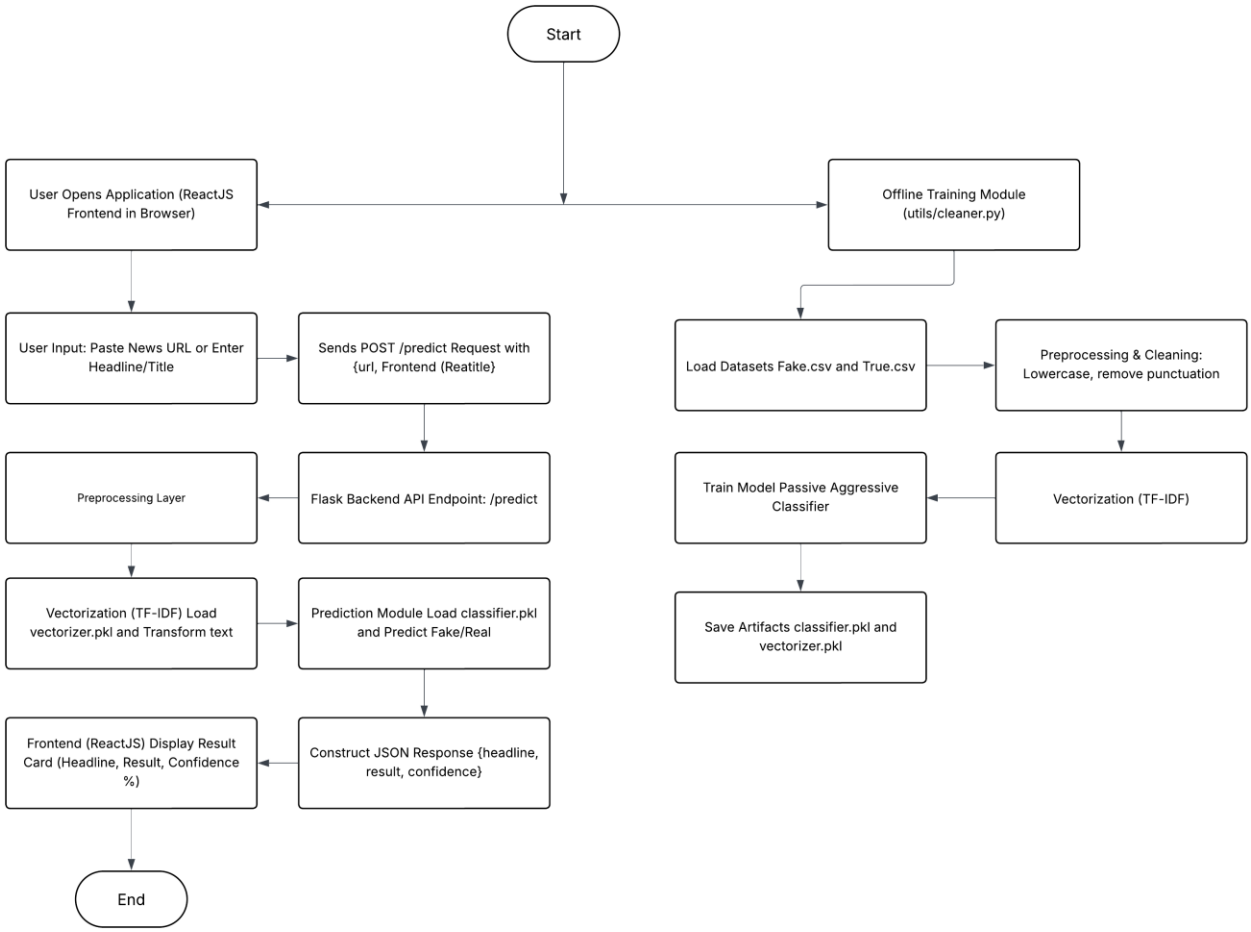
- **Frontend Module (React Js)**
    - Provides user interaction through forms (URL or headline input).
    - Displays prediction results with headline, classification, and confidence.
    - Handles errors gracefully.
- **Backend API Module (Flask App)**
    - Routes (/predict) for handling POST requests.
    - Extracts text from URLs (via newspaper3k) or direct headline input.
    - Passes input through trained ML model for classification.
    - Returns JSON response to frontend.
- **ML Model & Preprocessing Module (utils/cleaner.py)**
    - Data loading and cleaning (from Fake.csv and True.csv).
    - TF-IDF vectorization of text.
    - Model training using Passive Aggressive Classifier.
    - Model persistence via Joblib (classifier.pkl, vectorizer.pkl).
- **Data Storage Module**
    - Dataset files (Fake.csv, True.csv).
    - Model storage (./model/classifier.pkl, ./model/vectorizer.pkl).

System Archicture :-



1.1 System Archicture

Flow Chart :-



1.2 Flow chart of News Headline Detector

Technology Stack

- **Frontend**: ReactJS (user-friendly, modular components like NewsForm.js and Result.js).
- **Backend**: Flask (lightweight Python framework, REST API support).
- **Machine Learning**:
    - TF-IDF Vectorizer for feature extraction.
    - Passive Aggressive Classifier for fast binary classification of fake vs real.
- **Libraries**:
    - joblib for model persistence.
    - pandas & scikit-learn for preprocessing and ML.
    - newspaper3k for news extraction from URLs.
    - flask-cors for cross-origin requests from React frontend.
- **Deployment**:
    - Render for backend hosting (scales easily).
    - Vercel/GitHub Pages for frontend hosting.

Scalability Plan

- **Database Optimization**:

  The current implementation relies on CSV datasets, which limit scalability and query efficiency. Migration to a NoSQL database (MongoDB) will enable:

  - Efficient storage and retrieval of large volumes of articles and prediction logs. Schema flexibility to handle varied news metadata.

- **Caching**:

  We can use Redis or in-memory caching at the backend to:

  - Store results of frequently queried headlines/URLs.
  - Minimize redundant model inference requests.
  - Reduce average response time and server load.

- **Model Improvement**:

  As Dataset grows we require higher accuracy requirements:

  - Upgrade from a traditional ML model (Passive Aggressive Classifier) to a deep learning model such as BERT or RoBERTa.
  - Host models on GPUs for faster inference.
  - Serve models using TensorFlow Serving or TorchServe for production-grade scalability.

- **Cost Consideration**:

  Currently deployed on Render (backend) and Vercel (frontend) free tiers, suitable for small-scale usage. Plan for auto-scaling infrastructure as user load increases:

  - Horizontal scaling of backend containers with load balancers.We can use Cloude services such as AWS for it.
  - Vertical scaling (upgrading compute resources) for model-serving nodes.
  - Implement usage monitoring and alerts to optimize resource allocation and control costs.