SCTR's Pune Institute of Computer Technology Dhankawadi, Pune

A PROJECT REPORT ON
Fake News Detection System

SUBMITTED BY
Harsh Dhawale (41124)
Akash Kalme (41141)

Under the guidance of
Prof. U. S. Pawar

DEPARTMENT OF COMPUTER ENGINEERING
Academic Year 2023-24

DEPARTMENT OF COMPUTER ENGINEERING

SCTR's Pune Institute of Computer Technology Dhankawadi, Pune, Maharashtra - 411043

CERTIFICATE

This is to certify that the SPPU Curriculum-based Mini Project
- Develop Fake News Detection System.

Submitted by
Harsh Dhawale (41124)
Akash Kalme (41141)

has satisfactorily completed the curriculum-based Mini Project under the guidance of
Prof. U. S. Pawar towards the partial fulfilment of the final year
Computer Engineering Semester VII,
Academic Year 2023-24 of Savitribai Phule Pune University.

Date:
Place: PICT, Pune   **Name & Sign of Project Guide**

Acknowledgment

It gives me great pleasure to present the mini project on - Develop Fake News Detection System.

First, I would like to take this opportunity to thank my guide Prof. U. S. Pawar for giving me all the help and guidance needed. I am grateful for her kind support and valuable suggestions that proved to be beneficial in the overall completion of this project.

I am thankful to our Head of the Computer Engineering Department, Dr. G. V. Kale, for her indispensable support and suggestions throughout the project. I would also genuinely like to express my gratitude to the CC Prof. Samadhan Jadhav, for his constant guidance.

Finally, I would like to thank my mentor, Prof. U. S. Pawar for her constant help and support during the overall process.

Place: Pune   **Name of Student & Signature**

❖ **Title:**
Develop Fake News Detection System.

❖ **Problem Statement:**
 Develop a reliable and scalable Fake News Detection System to combat the proliferation of misinformation and fake news in online content, ensuring accuracy, real-time operation, and ethical considerations are prioritized.

❖ **Software Requirements:**
- Operating System: Open-Source Linux / Windows
- Python 3.11
- Pandas
- Numpy
- Matplotlib
- Scikit-Learn
- Keras

❖ **Objectives:**
- **Accurate Classification:** Develop ML models that can accurately differentiate between real and fake news articles or content.
- **Model Selection:** Utilize appropriate ML algorithms, including deep learning models like neural networks, to train and deploy for classification.
- **Evaluation Metrics:** Establish clear evaluation metrics (e.g., accuracy, precision, recall) to assess the system's performance and effectiveness.

❖ **Theory:**
 Detecting fake news using Machine Learning (ML) involves applying various ML techniques to analyze and classify news articles or content as either genuine or fake. Here's a theoretical framework for building a Fake News Detection System using ML:
- **Data Collection and Preprocessing:**
  - ○ Gather a diverse dataset of news articles, including both real and fake news, preferably with labelled examples.
  - ○ Preprocess the data by removing noise, such as HTML tags, special characters, and stop words.
  - ○ Tokenize the text and convert it into numerical features suitable for ML algorithms.
- **Feature Extraction:**

- Extract relevant features from the text, such as TF-IDF (Term Frequency-Inverse Document Frequency) vectors, word embeddings (e.g., Word2Vec, GloVe), or character n-grams.
- For multimodal content (e.g., images or videos), use techniques like CNNs or feature extraction from visual or audio content.

- **Model Selection:**

  Choose appropriate ML algorithms for classification, such as:
  - Logistic Regression
  - Naive Bayes
  - Decision Trees
  - Random Forest
  - Support Vector Machines (SVM)
  - Neural Networks (e.g., LSTM, CNN, or transformer-based models like BERT)

- **Training and Validation:**
  - Split the dataset into training, validation, and testing sets.
  - Train the selected model(s) on the training data and optimize hyperparameters.
  - Use the validation set to fine-tune the model and prevent overfitting.

- **Evaluation Metrics:**
  - Evaluate the model's performance using appropriate metrics, such as accuracy, precision, recall, F1-score, and AUC-ROC (Receiver Operating Characteristic - Area Under the Curve).

- **Ensemble Methods:**
  - Experiment with ensemble methods like stacking or bagging to combine the predictions of multiple models for improved accuracy.

- **Continuous Learning and Feedback:**
  - Integrate mechanisms for continuous learning and model updates based on user feedback and emerging fake news tactics.

- **Scalability:**
  - Design the system to handle large volumes of data and adapt to various content sources.

- **Monitoring and Updates:**
  - Continuously monitor fake news trends and adapt the system to evolving tactics employed by malicious actors.

❖ **Outcome**:

1. **Accurate News Classification:** The system is capable of accurately classifying news articles, distinguishing between genuine and fake news with a high degree of precision. This outcome is critical in reducing the dissemination of misinformation and enhancing information integrity.

2. **Real-time Detection:** The system operates in real-time, enabling swift identification of fake news as it emerges. This capability is essential for timely interventions and corrections.

3. **Enhanced Information Integrity:** By reducing the impact of fake news, the system contributes to maintaining the integrity of information in the digital landscape. Users can have more confidence in the news they consume.

4. **Scalability:** The system is designed to handle large volumes of data and various content types, making it adaptable to different platforms and content sources. This scalability ensures its widespread applicability.

5. **Transparency and Explainability:** With integrated explainability techniques, the system provides insights into its decision-making process, increasing transparency and user trust.

6. **Continuous Learning and Adaptation:** The incorporation of mechanisms for continuous learning and model updates ensures the system can adapt to evolving tactics employed by malicious actors. It remains effective over time.

7. **Ethical Considerations:** Addressing ethical concerns, such as privacy, bias, and freedom of speech, in the development and deployment of the system ensures it respects ethical principles and user rights.

8. **User-friendly Interface:** A user-friendly interface allows fact-checkers and news consumers to access the system's output, making it a practical tool for various stakeholders.

9. **Collaboration Opportunities:** The system opens the door to collaboration with news organizations, social media platforms, and fact-checking entities to integrate its capabilities into their workflows.

10. **Reduced Harm from Fake News:** Ultimately, the primary outcome is a reduction in the harm caused by the spread of fake news, leading to a more informed and resilient society.

11. **and Performance:** The system's performance is quantifiable through metrics such as accuracy, precision, recall, F1-score, and AUC-ROC, allowing for continuous evaluation and improvement.

❖ **Result:**

▪ Real News Detection:

```
news2 = """BRUSSELS (Reuters) - NATO allies on Tuesday welcomed President Donald Trump s decision to commit more forces
to Afghanistan, as part of a new U.S. strategy he said would require more troops and funding from America s partners.
Having run for the White House last year on a pledge to withdraw swiftly from Afghanistan, Trump reversed course on
Monday and promised a stepped-up military campaign against  Taliban insurgents, saying:  Our troops will fight to win.
U.S. officials said he had signed off on plans to send about 4,000 more U.S. troops to add to the roughly 8,400 now
in Afghanistan. But his speech did not define benchmarks for successfully ending the war that began with the U.S.-led
invasion of Afghanistan in 2001, and which he acknowledged had required an   extraordinary sacrifice of blood and treasure.
We will ask our NATO allies and global partners to support our new strategy, with additional troops and funding increases
in line with our own. We are confident they will,  Trump said. That comment signaled he would further increase pressure
on U.S. partners who have already been jolted by his repeated demands to step up their contributions to NATO and his
description of the alliance as  obsolete  - even though, since taking office, he has said this is no longer the case.
NATO Secretary General Jens Stoltenberg said in a statement:  NATO remains fully committed to Afghanistan and I am
looking forward to discussing the way ahead with (Defense) Secretary (James) Mattis and our Allies and international
partners. NATO has 12,000 troops in Afghanistan, and 15 countries have pledged more, Stoltenberg said. Britain, a
leading NATO member, called the U.S. commitment  very welcome ."""
manual_testing(news2)
```

✓ 0.0s   Python

```
Linear Regression Prediction: Not A Fake News
Decision Tree Classifier Prediction: Not A Fake News
Gradient Boosting Classifier Prediction: Not A Fake News
Random Forest Classifier Prediction: Not A Fake News
```

▪ Fake News Detection:

```
news = """A week after Supreme Court-appointed panel of experts examining India's regulatory mechanism in an investigation
linked to the Hindenburg allegations has given a clean chit to the Adani Group, the short-seller Hindenburg has announced
that it will now expose Supreme Court who has given clean chit to Adani.Speaking to The Fauxy, Hindenburg said that
Hindenburg is the biggest authority and SC of India has done a contempt by not aligning itself with Hindenburg report.
"We will soon release a report on Supreme Court of India" said the Hindenburg Research CEO Nathan Anderson.
Hindenburg is currently finding SC shares to take short position and make profits when the share prices drop after its
report. Reportedly, Hindenburg is likely to raise questions on Supreme Court's collegium system."""
manual_testing(news)
```

✓ 0.0s   Python

```
Linear Regression Prediction: Fake News
Decision Tree Classifier Prediction: Fake News
Gradient Boosting Classifier Prediction: Fake News
Random Forest Classifier Prediction: Fake News
```

❖ **Conclusion:**

The Fake News Detection System, built using ML, presents a powerful tool to combat fake news and misinformation in the digital landscape. With high accuracy, precision, and recall, the system can effectively differentiate between genuine and fake news articles. Continuous learning and user feedback mechanisms ensure its adaptability to evolving tactics used by malicious actors.