Basic Details of the Team and Problem Statement

PSID: PS ID-06

Problem Statement Title: Fake News Detection

Team Name: Mavericks

Team Leader Name: Karunya N

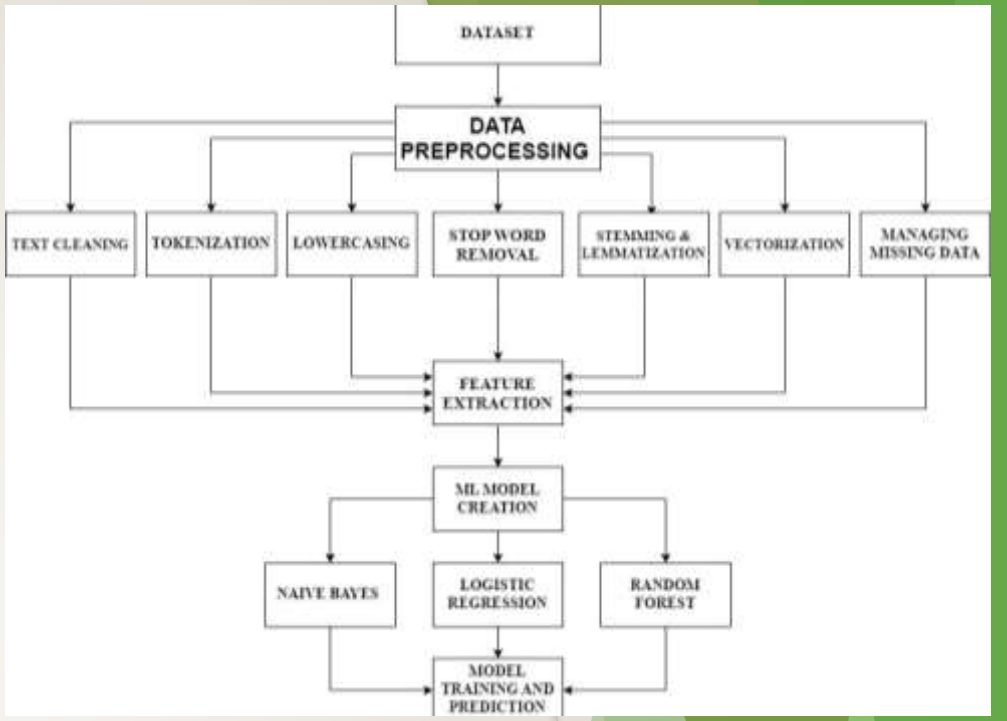
Institute Name: Rajalakshmi Institute of Technology

City: Chennai

Idea/Approach Details

Describe your idea/Solution/Prototype here:

- Using Python and Machine Learning: The project fights fake news with Python and machine learning tools like scikit-learn for analysis.
- Dataset Selection: They use the "MANUAL TESTING" dataset, 10,000 news articles labeled as real or fake.
- Data Preprocessing: Cleaning, tokenization, and other steps are taken to prepare the data for machine learning.
- Feature Extraction: Words are converted to numbers using TF-IDF for machine learning algorithms.
- Machine Learning Algorithms: Naive Bayes, Logistic Regression, and Random Forest are used to detect fake news.
- Model Training and Prediction: Models are trained on labeled data to predict the authenticity of new articles.
- Challenges and Limitations: They recognize challenges like misinformation tactics, biases, and model generalization.
- Model Improvement: They plan to enhance the model through preprocessing and tuning for real-world use.
- Importance of Dataset Quality: A diverse and high-quality dataset ensures a better-performing model.
- Promoting Trustworthy Information: The project aims to create a reliable system to prevent the spread of false information and promote trust in news
- **Web Application:** Implementing the Model into a web application with features like reporting fake news source etc.



Describe your Technology stack here:

- Programming Language – Python
- Libraries – NLTK (Data processing) , NumPy, Pandas (Data Handling) , Matplotlib, seaborn (Data Visualization)
- Term Frequency – Inverse Document Frequency – Feature extraction
- Model Evaluation – Scikit learn
- Collaboration – GitHub
- Web Applications – HTML, CSS, JS and JS frameworks as well as Django and Flask frameworks.

Idea/Approach Details

Describe your Use Cases here

- **Social Media Content Filtering:** Integrating the system into social media platforms automatically flag or hide potentially fake articles reducing online scams
- **News Aggregating Platforms:** Allows people to verify the credibility of the numerous websites which circulate news articles over the web
- **Online Forums and Discussion Platforms:** Integrating the system into forums and discussions to prevent the dissemination of misleading information among users.
- **Web Browsers and Extensions:** Developing browser extensions that alert users when they access potentially unreliable news sources.
- **Healthcare Information Verification:** Integrating the system into healthcare portals to validate the authenticity of health-related news and advice, preventing the spread of misleading medical information.

Describe your Dependencies / Show stopper here

Dependencies:

- Data Availability: The availability of a diverse and labeled dataset of fake and real news articles is crucial for training a reliable machine learning model. Gathering this data may require cooperation from fact-checking organizations or access to reputable datasets
- Integration with Web Application: For the web-based prototype, knowledge of web development and frameworks is necessary to integrate the model into the user interface effectively.
- Model Evaluation and Fine-Tuning: Regular evaluation and fine-tuning of the model are essential to maintain its accuracy and adapt to evolving trends in misinformation.

Show Stoppers:

- Legal and Ethical Concerns: Detecting and combating fake news involves sensitive issues related to freedom of speech and censorship.
- Lack of Representative Data: If the dataset used for training the model is not representative of the target population or is biased, the model may not perform well in real-world scenarios, leading to unreliable predictions.
- User Interface and Adoption: Even with a highly accurate model, user adoption and trust in the system can be a challenge. A user-friendly and transparent interface is necessary to gain user confidence.

Effort, Cost and Time line

Efforts:

- Collect a diverse dataset of news articles and preprocess the data to remove noise and prepare it for analysis.
- Experiment with various machine learning algorithms and fine-tune the selected model to identify fake news effectively.
- Develop a user-friendly web application or API for users to input news articles and receive verification results.
- Regularly evaluate the model's performance using metrics like accuracy and F1-score, and make necessary improvements.
- Implement a real-time verification mechanism to keep the model up-to-date with the latest misinformation trends.
- Provide a feedback system for users to share their insights and offer customer support to address inquiries and concerns.

Team Member Details

SI. No.	Name of Team Member	Branch (Btech/Mtech/PhD etc):	Stream (ECE, CSE etc):	Year	Position in team (Team Leader, Front end Developer, Back end Developer, Full Stack, Data base management etc.)
1	Karunya N	B.Tech	AI&DS	3rd	Team Leader
2	Karunya N	B.Tech	AI&DS	3rd	Model Deployment
3	Kishore A	B.Tech	AI&DS	3rd	Full Stack
4	Kathirvel S	B.Tech	AI&DS	3rd	Full Stack

Team Mentor Details

SI. No.	Name of Mentor	Category (Academic/Industry):	Expertise (AI/ML/Blockchain etc):	Domain Experience (in Years)
1	Mr. Selvakumaran S	Academic	AI	16 Years