CIS 593 Special Topics in CIS Section 51

Course project proposal on Fake News Detection Using Machine Learning

Submitted by

Kumudini Matta - 2888056

Sai Venkata Harshini Rayavarapu - 2894102

INTRODUCTION :

This project focuses on leveraging machine learning techniques to accurately identify fake news articles by examining the linguistic and textual characteristics of their content. In recent years, the spread of misinformation has become increasingly prevalent across social media platforms and digital news outlets, posing significant challenges to public trust and information reliability. As a result, the development of automated systems capable of detecting deceptive content has become an essential task in the field of data mining and natural language processing. The primary objective of this project is to design, implement, and evaluate a range of classification models that can distinguish between real and fake news based on key features extracted from the text. By comparing the performance of different algorithms, we aim to identify not only the most accurate model but also the most interpretable, thereby contributing a practical and transparent solution to combat the dissemination of false information online.

Data Sets:

The dataset used in this project is the “Fake and Real News Dataset” available on Kaggle. It consists of two separate CSV files: 'Fake.csv' containing fake news articles and 'True.csv' containing real news articles. Combined, these files include approximately 44,000 articles.

Each file contains the following columns:
- **title**: Headline of the news article
- **text**: Full body content of the article
- **subject**: Topic category (e.g., politics, world news)
- **date**: Date of publication

We will merge these files and add a binary label column for classification (1 for FAKE, 0 for REAL). The dataset is balanced and well-suited for binary classification tasks. The dataset can be accessed here: https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset

Issues with Research:

1. Our objective is to identify the most suitable machine learning model for detecting fake news based on textual data. This will involve assessing the performance of multiple algorithms, focusing not only on their classification accuracy but also on how easily their predictions can be interpreted and understood

2. Proper text preprocessing is a key factor in enhancing the effectiveness of natural language processing models. Selecting the most appropriate preprocessing techniques—such as removing stop words and transforming text into numerical representations—is vital for improving model accuracy and overall performance.

3. Identifying the specific words or features that significantly influence fake news classification enhances the interpretability of the model and promotes greater transparency in how predictions are made.

Potential Solutions:

A. **Comparing Machine Learning Algorithms for Fake News Detection:**

We plan to develop and compare several classification algorithms, including Logistic Regression, Naive Bayes, and Support Vector Machines. The textual data will undergo preprocessing through standard NLP methods such as tokenization, converting text to lowercase, and removing stopwords. To convert the processed text into numerical form, we will apply the TF-IDF technique. Each model's effectiveness will be assessed using key evaluation metrics such as accuracy, precision, recall, F1-score, and the ROC-AUC score.

B**. Hyperparameter Tuning for Improved Model Performance:**

We will optimize our models by applying hyperparameter tuning methods such as Grid Search and Randomized Search. To ensure the robustness and generalizability of our results, we will incorporate cross-validation during the tuning process. This will involve

adjusting parameters like the regularization factor in Logistic Regression and kernel-related settings in Support Vector Machines to achieve optimal performance.

C. **Feature Interpretability and NLP Analysis:**

To improve the interpretability of our models, we will analyze the key words and features that have the greatest impact on classification outcomes. Visualization tools such as word clouds and feature importance plots will be utilized to highlight and explain the most influential terms associated with both fake and real news predictions.

Evaluations :

To assess the effectiveness of each model, we will employ k-fold cross-validation, ensuring reliable performance across different data splits. The models will be evaluated using a range of metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. Additionally, visual tools such as confusion matrices and comparative performance charts will be used to illustrate the strengths and limitations of each algorithm.

Expected Outcomes :

The primary outcome of this project is to determine the most effective machine learning model for detecting fake news. We anticipate that traditional models like Logistic Regression and Naive Bayes will demonstrate strong performance when paired with TF-IDF-based feature extraction. In addition, we plan to visualize and examine the key textual elements that contribute most to classification decisions. Ultimately, this project aims to deliver a practical, scalable, and interpretable approach to combating misinformation through machine learning.