Executive Summary
Seeking the Truth on the X platform: Detecting Disinformation using Machine Learning Algorithms

Ciobanu Sergiu-Tudor

Introduction

The digital era has fundamentally shifted information consumption from traditional media to social platforms like X, creating an ideal environment for the propagation of fake news. While deceptive content is not a modern invention, exemplified by the 1938 “War of the Worlds” radio broadcast that caused public panic, its nature has evolved from entertainment to a dangerous tool utilized for political and financial manipulation. Unlike traditional journalism, which adheres to editorial standards, social media allows for the unrestricted dissemination of fabricated narrative, posing a critical challenge to information integrity.

Methodology and Exploratory Analysis

This study utilized the CIC TruthSeeker 2023 dataset, a comprehensive repository of real and fake tweets, to analyze and detect disinformation. Exploratory Data Analysis (EDA) revealed distinct behavioral patterns between the two categories. Users disseminating fake news consistently displayed lower credibility scores compared to those sharing factual content. Linguistic analysis showed that real news tweets utilize a diverse vocabulary, whereas fake news content tends to be shorter and less lexically rich, a design choice intended to capture immediate attention and encourage rapid, unverified sharing. Furthermore, it was observed that accounts spreading disinformation typically have substantially fewer followers, suggesting the predominance of "throwaway" or automated profiles created to amplify false narratives.

Data Preprocessing and Model Implementation

To prepare this unstructured data for analysis, a rigorous cleaning pipeline was implemented. Textual data underwent normalization, tokenization, and stop-word removal to eliminate noise such as URLs, special characters, mentions, numerical values, and so on. Feature filtering was applied to remove columns with over 70% null values, specifically sparse Named Entity Recognition (NER) features, to prevent data sparsity from skewing results. The cleaned text was transformed into numerical vectors using TF-IDF, allowing algorithms to weigh the semantic importance of terms. Three machine learning architectures were trained on this processed data: Logistic Regression, Random Forest, and Decision Tree.

Results and Conclusions

The experimental results demonstrated that rigorous preprocessing enables traditional machine learning models to perform with high efficacy. The Logistic Regression model achieved superior performance, recording an Accuracy of 0.9427 and an F1-Score of over 0.94. This notably outperformed the Random Forest classifier, which achieved an accuracy of 0.9329, and the Decision Tree model, which lagged with an accuracy of 0.889. The Logistic Regression model also maintained a low False Positive rate, crucial for avoiding the misclassification of legitimate news.

These findings conclude that fake news detection on current datasets is a linearly separable problem that can be effectively addressed with efficient, interpretable algorithms.