News Article Classification (Fake/Real)

Abstract

This report details the development of a News Authenticity Checker, a tool designed to classify news headlines as real or fake. Leveraging machine learning techniques, the project aims to combat misinformation by providing users with a simple interface to verify news authenticity. The system was built using Python, with a focus on natural language processing (NLP) and a basic graphical user interface (GUI). Key steps included data collection, model training, and interface development. The final tool successfully identifies fake news with reasonable accuracy.

1. Introduction

The rise of misinformation in digital media has made it critical to develop tools that can verify the authenticity of news. The News Authenticity Checker addresses this challenge by using machine learning to analyze news headlines and predict their validity. This project was motivated by the need to empower users to distinguish between real and fake news, especially during significant events like the COVID-19 pandemic. The tool provides a user-friendly interface where a headline can be input, and a prediction (REAL or FAKE) is generated.

2. Tools Used

The project utilized the following tools and technologies:

Python: Primary programming language for development.

- Scikit-learn: For implementing machine learning algorithms like logistic regression.
- NLTK: For natural language processing tasks such as tokenization and text preprocessing.
- Tkinter: To create a simple graphical user interface.
- Pandas and NumPy: For data manipulation and numerical operations.

3. Steps Involved in Building the Project

The development process consisted of the following steps:

- Data Collection: A dataset of news headlines labeled as real or fake was gathered from publicly available sources.
- Data Preprocessing: Headlines were cleaned using NLTK, including tokenization, stop-word removal, and vectorization using TF-IDF.
- Model Training: A logistic regression model was trained on the preprocessed data to classify headlines.
- Interface Development: A GUI was built using Tkinter, featuring a text box for headline input, a "Check Authenticity" button, and a prediction display.

- Testing and Refinement: The model was tested with various headlines, and adjustments were made to improve accuracy.

4. Conclusion

The News Authenticity Checker successfully demonstrates the potential of machine learning in combating misinformation. The tool accurately classifies news headlines, as seen in examples like identifying a fake claim about micro chipping via vaccines and a plausible tariff-related headline as real. While the project achieves its goal, future improvements could include expanding the dataset, incorporating more advanced NLP models, and enhancing the user interface for broader accessibility.