Fake News Detection Using NLP

Problem Statement:

The proliferation of fake news in the digital age poses a significant threat to information integrity and public trust. In this era of information overload, distinguishing between genuine news articles and fabricated ones has become a pressing concern. To address this challenge, our project aims to develop a robust and accurate model for automated fake news detection.

Technique:

Our project leverages Natural Language Processing (NLP) techniques to build a sophisticated machine learning model. We will analyze a comprehensive dataset containing both genuine and fake news articles' titles and text. By applying advanced NLP algorithms and feature engineering, we will extract meaningful patterns, linguistic cues, and semantic relationships from the textual content.

The key techniques and components of our approach include:

Text Preprocessing: Cleaning and preprocessing the text data by removing stopwords, punctuation, and special characters. Tokenization and stemming/lemmatization will also be applied to standardize the text.

Feature Extraction: Utilizing various NLP techniques, such as TF-IDF (Term Frequency-Inverse Document Frequency) and Word Embeddings (e.g., Word2Vec or GloVe) to represent words and phrases in a numerical format.

Model Selection: Evaluating different machine learning and deep learning algorithms for classification tasks, such as Logistic Regression, Random Forest, Support Vector Machines, and Neural Networks (e.g., LSTM or Transformer-based models).

Model Training: Splitting the dataset into training and testing sets to train the chosen model. We will also explore techniques like cross-validation for hyperparameter tuning to optimize model performance.

Evaluation Metrics: Assessing the model's performance using evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to ensure its effectiveness in distinguishing fake from real news.

Explainability: Implementing techniques to interpret and explain the model's predictions, such as feature importance analysis and attention mechanisms.

Deployment: Developing a user-friendly interface or API that allows users to input news articles and receive real-time predictions about their authenticity.

By combining these NLP techniques and machine learning methods, our project aims to contribute to the ongoing efforts to combat the spread of misinformation by providing a reliable tool for fake news detection.

Team Members

Mohamed Abdul Rahman S

Sam Selin D

Yabesh Jesilen V

Naga Mari Muthu A

Rahul B

JP College Of Engineering