Fake News Detection Using Natural Language Processiong(NLP)

Abstract:

The proliferation of fake news has become a significant concern in today's digital age. The ease with which information can be created and disseminated has made it increasingly difficult to distinguish fact from fiction. Fake news can have serious consequences, ranging from damaging individual reputations to influencing political decisions. In this project, we propose a novel approach for detecting fake news using Natural Language Processing (NLP) techniques. Our approach leverages machine learning algorithms to analyze various linguistic features that are known to differentiate real and fake news articles. We evaluate the effectiveness of our approach using a dataset of real and fake news articles and demonstrate its accuracy in detecting fake news.

Module:

Data Collection:

Collect a large dataset of real and fake news articles from various sources . Ensure that the dataset is balanced, i.e., equal number of real and fake news articles.

Preprocessing:

Remove stop words, punctuation, and special characters from each article. Convert all text to lowercase. Tokenize each article into individual words or phrases.

Feature Extraction:

Extract various linguistic features that are known to differentiate real and fake news articles. These may include, Lexical complexity: Measure the complexity of language used in the article.

Sentiment analysis:

Analyze the emotional tone of the article. Part-of-speech tagging: Identify the parts of speech (nouns, verbs, adjectives, etc.) used in the

article. Named entity recognition: Identify named entities (people, organizations, locations, etc.) mentioned in the article.

Machine Learning:

Train a machine learning model using the extracted features and labeled data (real or fake).Use supervised learning techniques, such as classification, to train the model. Evaluate the performance of the model using metrics like accuracy, precision, recall, and F1 score.

Model Evaluation:

Test the trained model on new data to evaluate its performance. Compare the results with human annotation or other existing approaches.

Results and Discussion:

Present the results of the evaluation, highlighting the strengths and limitations of the proposed approach. Discuss potential improvements and future directions for research.

Conclusion:

Summarize the key findings of the study . Emphasize the importance of developing effective fake news detection systems to mitigate the negative impact of misinformation.

DATASET LINK:

https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset

R.DHIWAKAR

DATE:3.10.2023

NANDHA COLLEGE OF TECHNOLOGY