Multi-Label Text Classification for News Categorization

Abstract:
News articles often cover multiple topics, requiring a multi-label classification approach to categorize them accurately. This project focuses on developing a **multi-label text classification system** using **Natural Language Processing (NLP)** to classify news articles into multiple categories, such as politics, economy, sports, technology, etc.

Methodology:
The system will preprocess news articles using standard **NLP techniques** like **tokenization**, **lemmatization**, and **stopword removal**. **Word embeddings** (e.g., **Word2Vec** or **GloVe**) will be used to represent text data in a vectorized form. A **multi-label classification model** will be trained using algorithms like **Logistic Regression**, **Random Forest**, and deep learning models such as **CNNs** or **BERT** for better contextual understanding. The system will use techniques such as **sigmoid activation functions** and **binary cross-entropy loss** to handle multiple labels for a single article.

Outcome:
The system will accurately categorize news articles into multiple relevant categories, improving the organization of news content. It will help in content recommendation, targeted advertising, and news aggregation, providing users with personalized, multi-topic news feeds.