Sentiment Analysis Using Naive Bayes Classifier

Objective

This project aims to analyze sentiment from Twitter data and predict sentiments for news headlines using Natural Language Processing (NLP) techniques.

Dataset Overview

- **Twitter Data: Contains phrases and their corresponding sentiments (Positive, Negative, Neutral).**
- **News Data: Contains news headlines for which sentiment needs to be predicted.**

Data Preprocessing

- **Text Cleaning: Convert text to lowercase and remove unwanted characters.**
- **Vectorization: Convert text into numerical form using TF-IDF (Term Frequency-Inverse Document Frequency) to extract important features.**

Building the Sentiment Analysis Model

- **The dataset is split into training and testing sets (80-20 split).**
- **A Multinomial Naive Bayes model is trained on the Twitter sentiment dataset.**
- **Model performance is evaluated using accuracy score and classification report.**

Sentiment Prediction for News Headlines

- **The trained model is used to predict sentiments for news headlines.**
- **The predicted sentiments are saved in news_with_sentiments.csv for further analysis.**

Results & Insights

- **The model successfully classifies sentiments in the Twitter dataset with reasonable accuracy.**
- **It can effectively predict the sentiment of unseen news headlines.**
- **The system can be expanded by using deep learning models like LSTMs or Transformer-based models (BERT) for better accuracy.**