

News Sentiment Analysis System

Project Description

News sentiment analysis is a system that automatically identifies the sentiment (positive, neutral, or negative) of news articles or headlines. The purpose of this project is to develop a machine learning algorithm for news sentiment analysis using a given dataset. The dataset consists of news articles and their associated sentiment labels, and the goal is to train a classifier that can accurately predict the sentiment of new articles.

- The number of classes in the dataset is three (positive, neutral, and negative).
- The dataset consists of 4846 news articles
- 70% of the data for training and 30% for testing
- [ThisGameZR/News-sentiment-analysis \(github.com\)](https://github.com/ThisGameZR/News-sentiment-analysis)

Methodology

- Data-preprocessing
 - Giving dataset column heads -> sentiment, text
 - removing URLs, special characters, and stop words.
 - stemming the words using the PorterStemmer algorithm
- split the preprocessed data into training and testing sets using a 70/30 split.
- used the TF-IDF vectorizer to convert the preprocessed text data into numerical feature vectors
- SVM Linear Classifier
 - I try with various classifiers and SVM Linear seems to perform the best

Results

The classification reports for the Multinomial Naive Bayes and Gaussian Naive Bayes classifiers are as follows:

Classification report:

Class	Precision	Recall	F1-score	Support
negative	0.71	0.54	0.61	179
neutral	0.76	0.91	0.83	847
positive	0.75	0.52	0.62	428
Accuracy			0.75	1454
Macro avg	0.74	0.66	0.68	1454
Weighted avg	0.75	0.75	0.74	1454

Based on the classification reports, we can see that SVM-Linear classifier achieved an overall accuracy of 0.75, the classifier is very accurate on predicting 'neutral'. However, the classifier performed poorly for the 'negative' and 'positive' classes, with recall scores below 0.6

One potential problem with making false classifications is that it can lead to misleading information. For example, if a positive news article is misclassified as negative, it could cause unnecessary concern among readers. Similarly, if a negative news article is misclassified as positive, it could lead to lack of caution among readers.

Another potential problem is that misclassifications can impact the miscalculations of events that rely on the news sentiment. For example, if a stock trading algorithm relies on sentiment analysis to make buy or sell decisions, a misclassification could lead to poor investment decisions and financial losses.