

Twitter Sentiment Analysis Of Movie Review with ML Techniques

Prashant Pathak
MT19051
Mtech-CSE

Project Overview and Data Acquisition

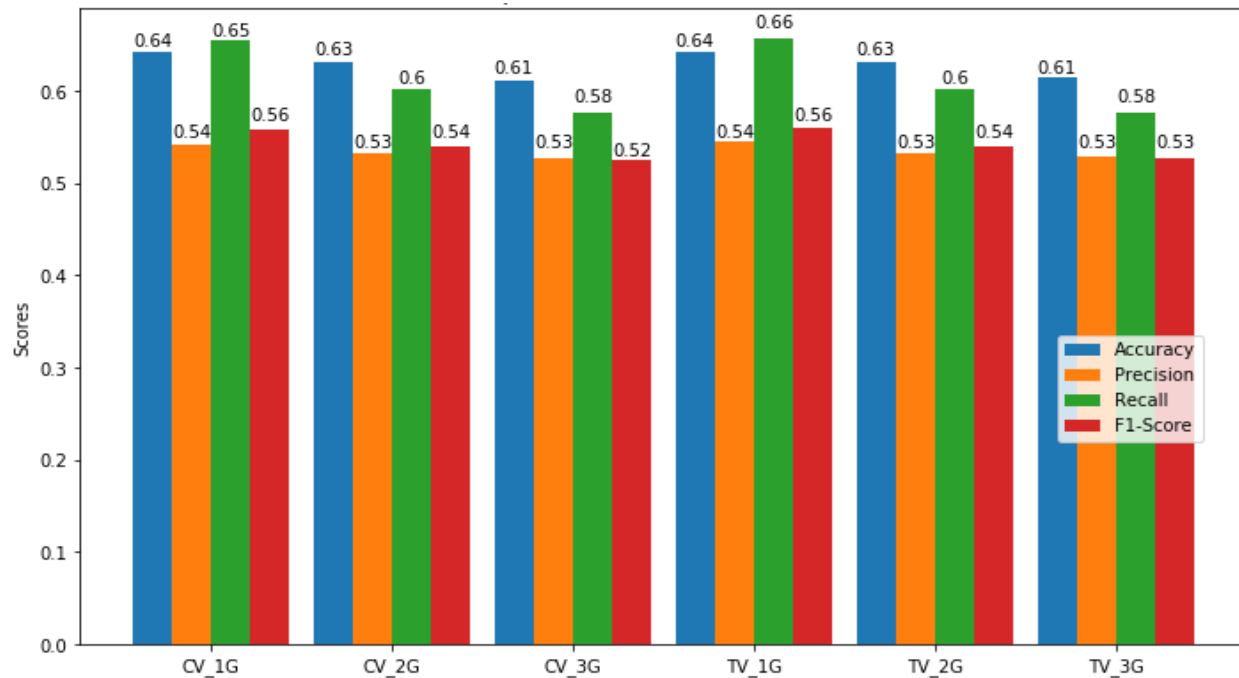
- Sentiment analysis is extracting and analyzing someone's views or opinions from textual data.
- Need for Sentiment Analysis.
- Why Twitter?
 - Twitter has 321 million active users, which makes it useful platform for performing community or market research.
- Movie Review Sentiment analysis.
- Types of Sentiment Analysis
- Data Gathering and its difficulties.

Data Preprocessing and Labeling

- Removing duplicate rows
- Remove neutral words ex Hashtags
- De-encode Emojis
- Tokenization and POS tagging
- Removing Stop word
- Lemmatization and Stemming
- Use Lexicon Sentiment analyzer to label the Data for further processing.

Feature Selection

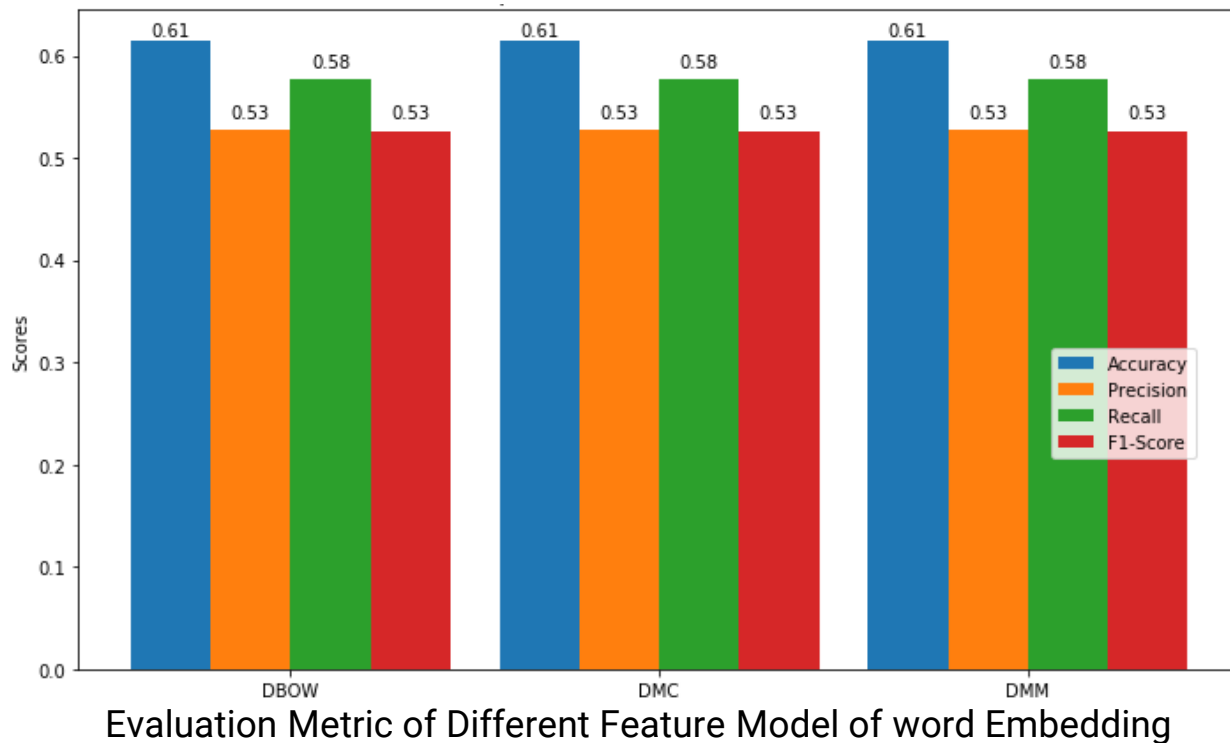
- Bag of Words
 - Count Vectorization
 - TFIDF



Evaluation Metric of Different Feature Model of Bag of Words

Feature Selection

- Word Embedding



Results

- In this project we explored the various feature extractions and found that the 1-Gram performs the best for this data
- Balancing the data does no good for our classifier as f1 score of Naive Bayes reduces after balancing the data.
- Linear SVM performs best for this tweet data.
- More Data can improve the accuracy and f1 score..

Model name	Accuracy	F1-score
Naive Bayes(On unbalances classes)	64.26	56
Naive Bayes	61.8	52.7
SVM	72.42	59.0
Logistic Regression	72.18	55.94