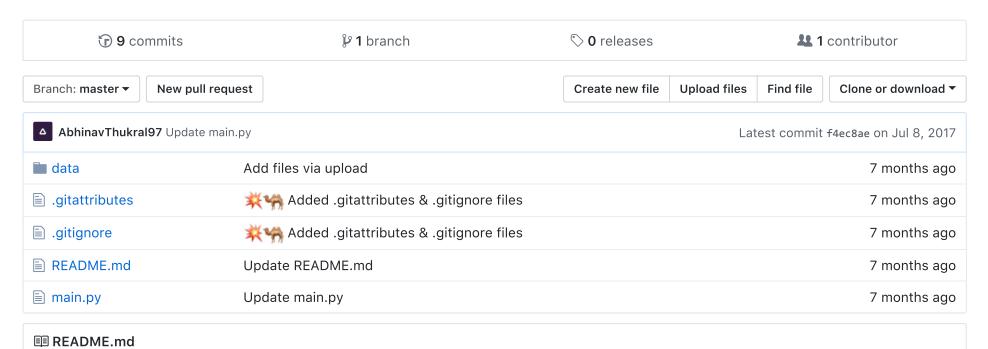
AbhinavThukral97 / SentimentAnalysis

Implemented text analysis using machine learning models to classify movie review sentiments as positive or negative. Built using Python 3.6.1.

#machinelearning #scikitlearn-machine-learning #sentiment-analysis #text-analysis #naive-bayes #logistic-regression #svm #knn



Movie Reviews Sentiment Analysis using machine learning

Implemented text analysis using machine learning models to classify movie review sentiments as positive or negative. Built using Python 3.6.1.

1. Tuned CountVectorizer (1_gram) to get appropriate features/tokens and then transformed to obtain input variable (document term matrix).

- 2. Splitted training test with test size of 20%
- 3. Used the following models to train on training data.
 - Naive Bayes
 - Logistic Regression
 - SVM (Support Vector Machine)
 - KNN (K Nearest Neighbors)
- 4. Tested models on test data and calculated accuracy of predictions
- 5. The results were as follows:

Naive Bayes: 98.9161849711%

Logistic Regression: 99.3497109827%

SVM: 99.0606936416%KNN: 98.6994219653%

- 6. Analysed further by observing confusion matrix
- 7. Used Naive Bayes model to observe the number of tokens (words) and the positivity/negativity associated with that word.
- 8. Implemented searching of selected words in the pandas dataframe to analyse specific words in the feature sets
- 9. Used the most accurate model (Logistic Regression) to train on the entire dataset. df
- 10. Took custom review inputs and predicted Positive/Negative review.

Dataset Source (from Kaggle): https://inclass.kaggle.com/c/si650winter11/data