

Dataset3 : Amazon Review Sentiment Analysis

Problem: To perform sentiment analysis on Amazon reviews to determine whether a review is positive or negative.

Data Set :

- Amazon Reviews Kaggle competition. 400000 rows, each row in the format {sentiment(text), review(text)}
- Test set is every 5th sample from the dataset. Remaining goes to Train set.

Pre-processing steps:

- To Lower Case, Remove Punctuations and Stop Words
- Negation Rule and Conjunction Rule

Machine Learning Algorithms

▶ Neural Networks

1. Good to model data with large number of input features.
2. Converts the data into vectors, which provides better results than textual data.

▶ Decision Trees

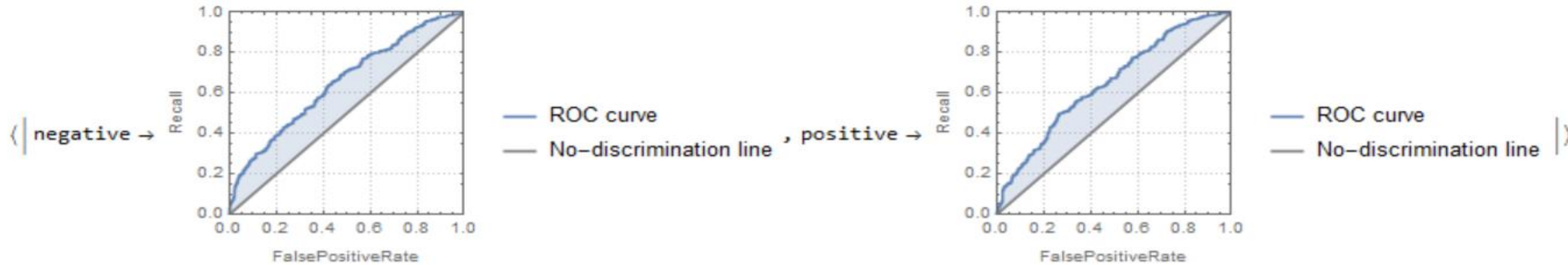
1. Does not require domain knowledge or parameter setting, hence can be used for classifying any dataset.
2. Performs variable screening implicitly.

▶ Support Vector Machines

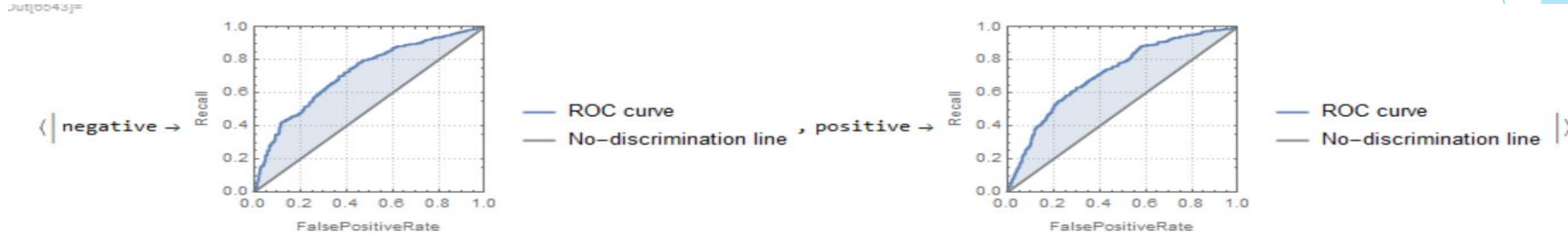
1. Most widely used classification algorithm for sentiment analysis .
2. Generate better result

Results:

- Random Forest[Decision Trees]:Accuracy :0.586587 , Precision: 0.603563, 0.572727 , Recall: 0.535573, 0.638945



- Neural Networks: Accuracy : 0.658659, Precision: 0.666667, 0.650794 , Recall:0.652174, 0.665314



- SVM: Accuracy :0.674675 , Precision: 0.696312, 0.656134, Recall: 0.634387, 0.716024

