Text Classification of Steam User Reviews

Classifying Reviews Using Steam Data Set

Executive Summary

- Steam library is one of the largest PC digital game distributors in the market.
- Steam reviews data set consists of 9.6 million English reviews of games ranging from triple A games to smaller indie games.
- Each review contains the user opinion, positive or negative sentiment and number of users who found the opinion helpful. After data wrangling, each review contains tokenized text of the user generated reviews.
- The best model created was a Random Forest Classifier, which is able to predict a sentiment score per user review. This model can be used to evaluate newly developed game reviews to predict user sentiments.
- Our model has an AUC of 0.872, which is a 87.2% probability of classifying our user reviews correctly.