Board Game Sentiment Analysis Predictor

Winnie Zhang | BrainStation Data Science Bootcamp

Problem Statement

The goal of this project is to better understand the ratings people give board games, such as which words used to describe board games end up making it more popular or which categories are more likely associated with a higher rated board game. Using numerical and categorical data along with text data processed using Natural Language Processing (NLP) with machine learning techniques, I built a classification model that can predict if the sentiment of a board game review is positive or negative). This will give companies making board games, and especially those who are trying to raise the funds to release their own board game better awareness of what they should focus on when creating their game.

Background

Every year, more and more board games are being published. It was in the year 1998 that the number of board games released in a year was over 1000 and in 2019, over 4500 board games were released[1]. With the vast number of board games being released every year, it's going to become more and more difficult for board games to stand out. Furthermore, over the past few years, Kickstarter, a global crowdfunding platform, has helped more and more board game launch into the market[2]. However, this doesn't mean that these board game necessarily perform well or rate well. With increasing competition, it's essential for companies, or indie tabletop game makers to have their board games stand out. Furthermore, with the pandemic causing digital fatigue, it's been causing a rapid growth of the tabletop game industry, with it expected to exceed $12.4 billion in 2021, and it's expected to grow another $2.5 billion in the next 3 years[3].

Data Source

The data for this project was scraped from the website BoardGameGeek. This was done by a user on Kaggle, where they also uploaded and shared this dataset. This consists of 5 different csv files. 2 of the files are from 2020, which I didn't use. I used the 3 most recent files from January 2022. One file is a dataframe containing all the user comments and ratings for each game. This file consists of 19 million rows and is around 1.5 GB in size. Another file contains some basic information about the game, such as the game description, the average rating, and the number of users who rated the game. I didn't use this file in the end because it contained the same information as the next file I will discuss. The third file contains detailed information about the board game, such as its description along with the minimum age of the player, the minimum and maximum play time.

Data Preprocessing and Feature Engineering

Since I am interested in natural language processing to predict reviews, I dropped all the rows that didn't contain a comment from the user. For the user reviews dataset, I also engineered 3 features: review length, number of reviews the user gave in total, and the average score of their reviews. I also removed all the non-English reviews as the majority of the reviews was English. For the games table, I dropped most of the columns as they were missing a lot of values. For many of the columns, they were also a list of items. I further separated these lists into separate columns and binarized them so that this information could be extracted. Specifically, I chose to do this for board game category, mechanic, and family I chose the most frequent ones.

Exploratory Data Analysis (EDA)

Figure 1 shows the distribution of the dependent variable, the rating of the board game, which is rated out of 10. The majority of the ratings are around 7/10 or 8/10. Since I'm interested in learning more about the sentiment of the reviews, I binarized the ratings so that ratings 8 or higher are considered positive (1) and ratings lower than 8 are considered negative (0) (Figure 2). During EDA, I also saw that reviews with the word "boring" had a much lower rating than reviews without it. Furthermore, some categories of board games such as Party Games and Children's games had a lower rating than those not under those categories. Some categories had higher ratings, such as War Games.

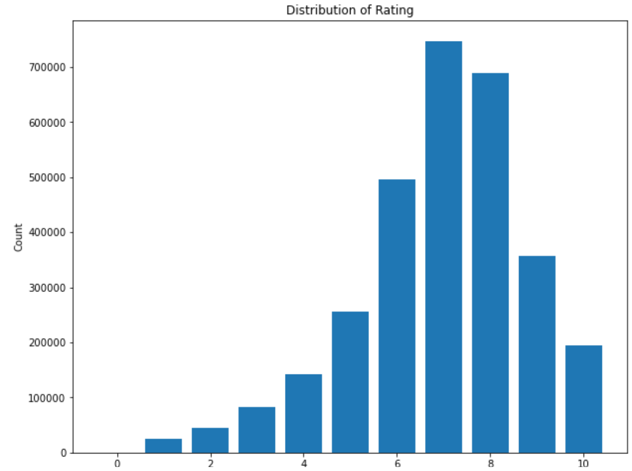

Figure 1 Distribution of table top review ratings

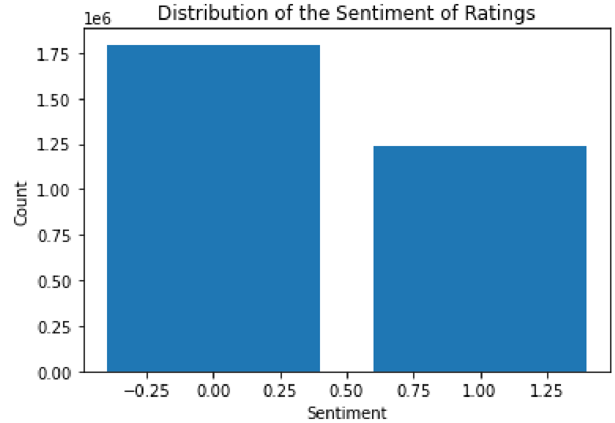

Figure 2 Distribution of the sentiment of reviews

Natural Language Processing (NLP) and Modelling

To help encode the text seen in the both the reviews and the descriptions of the game, I used both Bag of Words (BoW) and TF-IDF vectorizer. For both vectorizers, I used stemming. To ensure that the dataframe didn't get too big, I ensured that words occurred a minimum of 15 or 20 times, and that the max number of tokens was limited to 3000.

The candidate machine learning models evaluated included Linear Regression (L1 and L2), Decision Tree, and Support Vector Machine (SVM). Grid search cross validation was used to select the best hyperparameters for each model and ensure the best train and test scores. Table 1 shows the test set accuracy results.

Model	Vectorizer	Accuracy
Logistic regression (L1)	Bag of Words	0.766
Logistic regression (L2)	Bag of Words	0.769
Decision Tree	Bag of Words	0.738
Logistic Regression (L1)	TF-IDF	0.775
Logistic Regression (L2)	TF-IDF	0.772
Decision Tree	TF-IDF	0.748
SVM	TF-IDF	0.755

Table 1 Classification model scores on the test set. Hyperparameters were optimized were selected via grid search cross validation.

The best performing model is the TF-IDF Logistic Regression with L1 regularization. This model had an accuracy of 77.5%, a precision of 74.9%, and a recall of 69.2%.

Findings

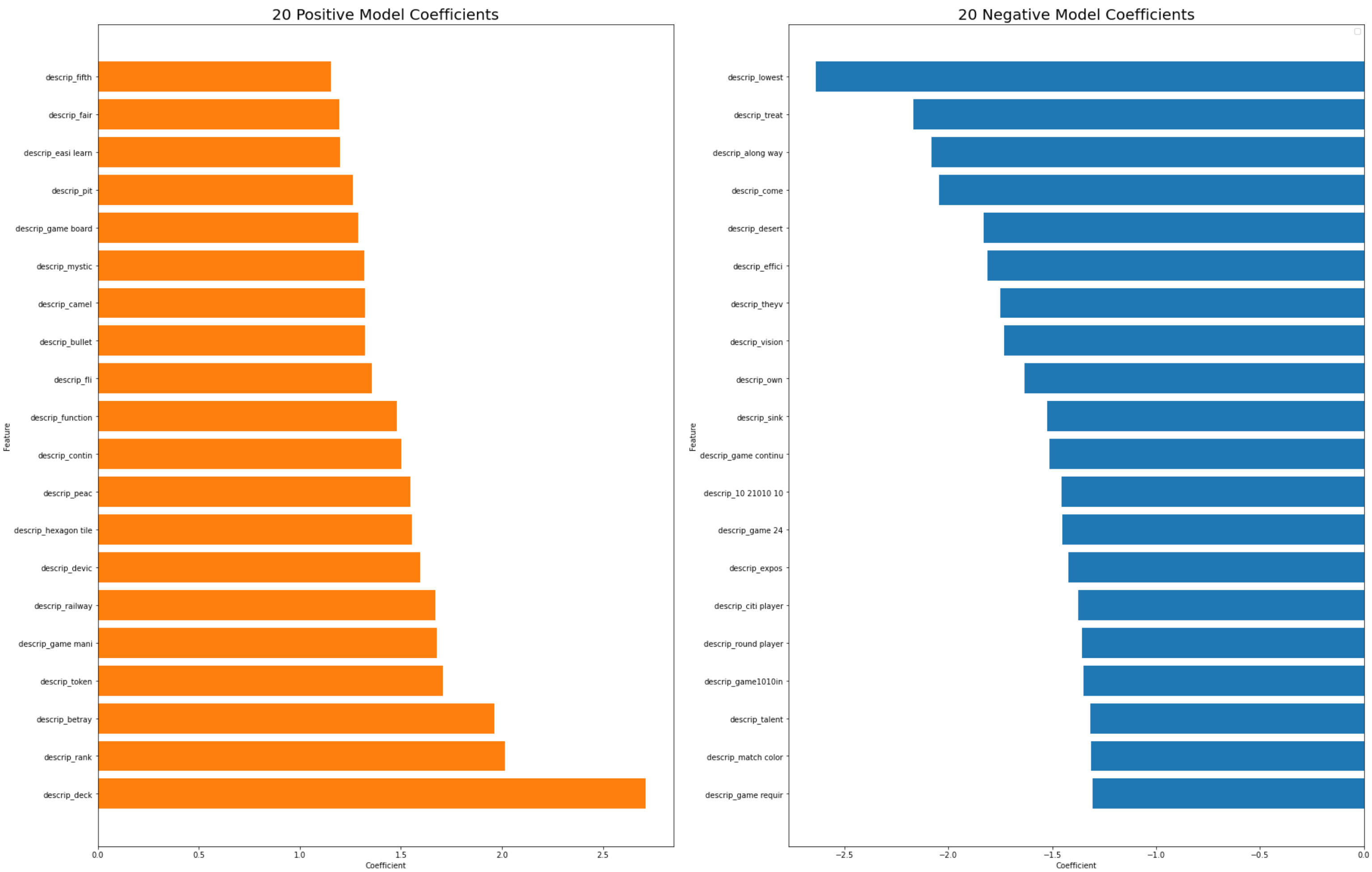

Figure 3: Top 20 positive ad negative tokens by classification coefficient value for description.

The key findings for Figure 3 are:

- Betrayal games and games involving decks are rated well
- Colour matching games don't rate well
- Games that involve game manipulation also tend to have a positive sentiment

It's much more difficult to analyze the 20 negative coefficients as most of the words don't make sense in the context. Though based on this, they shouldn't focus on boring concepts like colour matching, as people tend to prefer games that allow them to betray others.

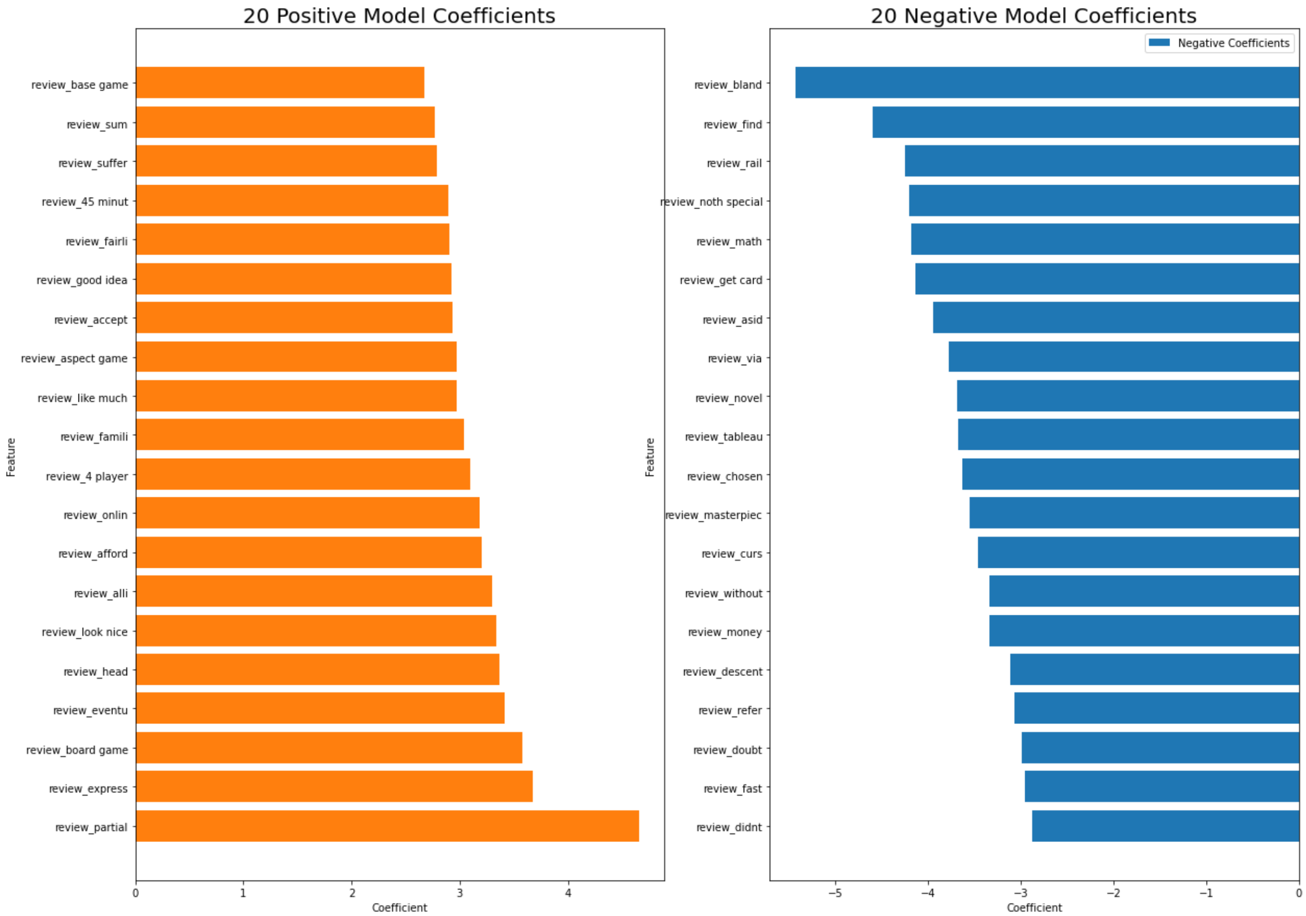

Figure 4 Top 20 and positive and negative tokens for reviews

The key findings for Figure 4 are:

- Board games that are aesthetically pleasing result in a positive sentiment
- Affordable and family friendly board games rate better
- Board games that are bland and include math rate lower

An important part of board games is making sure that it looks as good as it plays, so companies should also focus on how good the board game looks. The board games should focus on catering towards families or groups of friends and avoid subjects that require math.

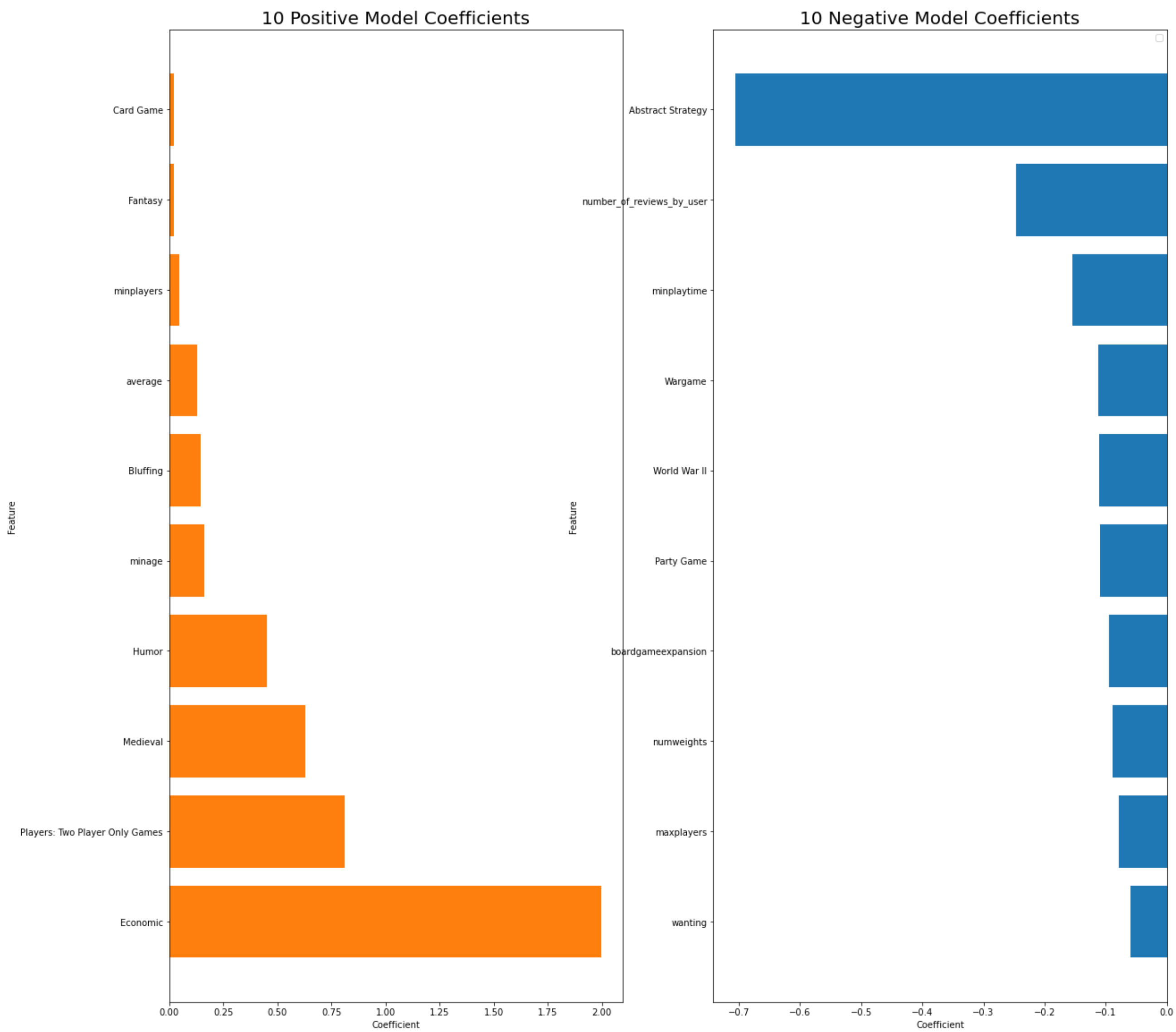

Figure 5: Top 10 positive and negative features that aren't text based.

The key findings for Figure 5 are:

- Abstract strategy games are more likely to cause a negative sentiment in a review
- Games with economic theme have positive sentiment

What's also surprising is that 2 player games have a positive coefficient as we saw earlier, 4 player games had a positive coefficient. However, this was a category we extracted and added, and 4 player games wasn't an option so this may be the reason why it doesn't appear here. All in all, people have a more positive sentiment towards games they don't play on their own.

Conclusion

In this study, we trained a model that can predict whether the review a user gives a board game is positive or negative in sentiment with an accuracy of 77.5%. We have also identified the features that have the biggest influence on rating score. Creating a game that's easy to learn and is good looking is essential for a positive sentiment, without too much math or boring concepts like color matching. The board game should be catered towards families or groups of 4.

Next Step

The next steps I would take are:

- Use more hyperparameters for the SVM. This is a computationally expensive model, so I didn't try out a lot of hyperparameters when optimizing it, which is why the accuracy may not be higher. I also only ran the SVM model using my TF-IDF vectorized data. I would run it again using my Bag of Words data.
- Use XGBoost to further improve the model for better prediction and interpret this model. XGBoost is an ensemble learning method, where it trains new models based on the errors of the previous models.
- Use neural networks and Word2Vec to see if this improves the modelling.
- Use lemmatization instead of stemming. Stemming has cut off some of the words so they're difficult to understand. Lemmatization would help preserve the words so they're easier to understand

References

1. Number of Games Published by Year | BoardGameGeek
2. How Kickstarter has CHANGED Board Games | Board Game Atlas
3. "Digital fatigue" is fuelling board game sales among adults | Dicebreaker
4. BoardGameGeek Reviews | Kaggle