

Game Sense: Understanding Player Sentiment Through NLP

Exploring models to predict player
sentiment



Bio

- Bachelor of Computing and Information Sciences
- Experience in Customer Service, Customer Relations and Logistics
- Aspiring Data Analyst



Index

Overview	Page 4
Dataset	Page 5
Methods	Page 6
Results	Page 7/8
Findings	Page 9
Feature Importance	Page 10/11
Conclusion	Page 12
Future Improvements	Page 13
EDA	Page 14/15

Overview

Business Statement

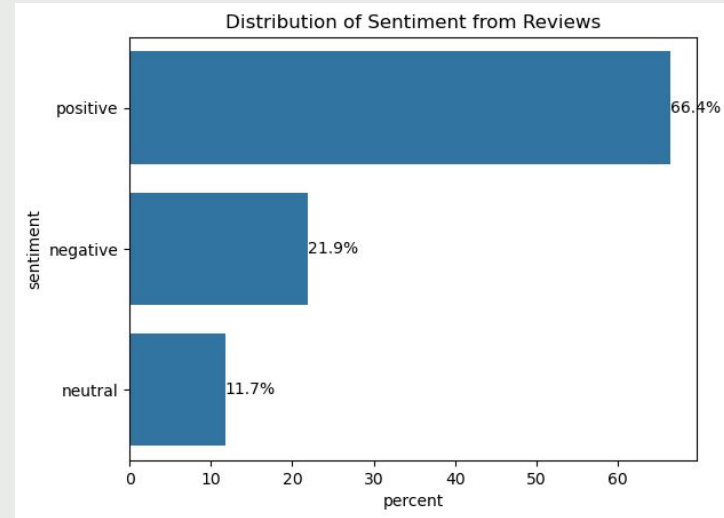
- Reviews are relied on when deciding where to spend
- Misleading reviews can damage trust and game reputation
- Benefits players and developers

Data Science Statement

- Preprocessed textual review
- Testing various models
- Achieve an acceptable score between 70/80%

Dataset Overview

- Source: Kaggle
- Scrapped from backloggd.com
- 1512 Samples, 13 Features
- Features:
 - Title
 - Release Date
 - Team
 - Rating
 - Times Listed
 - Number of Reviews
 - Genres
 - Summary
 - Reviews
 - Plays
 - Playing
 - Backlogs
 - Wishlist
- Feature Engineering:
 - Vader Sentiment Analysis
 - Count Vectors
 - TF-IDF Word Level
 - TF-IDF N-gram level
 - TF-IDF Character Level



Methods

Train / Test Split

70% Training Set / 30% Test Set

Logistic Regression

A simple, interpretable baseline model

Naive Bayes

A simple, easy to implement and scalable model

Support Vector Machine

Effective in high dimensional spaces while robust to noise and outliers.

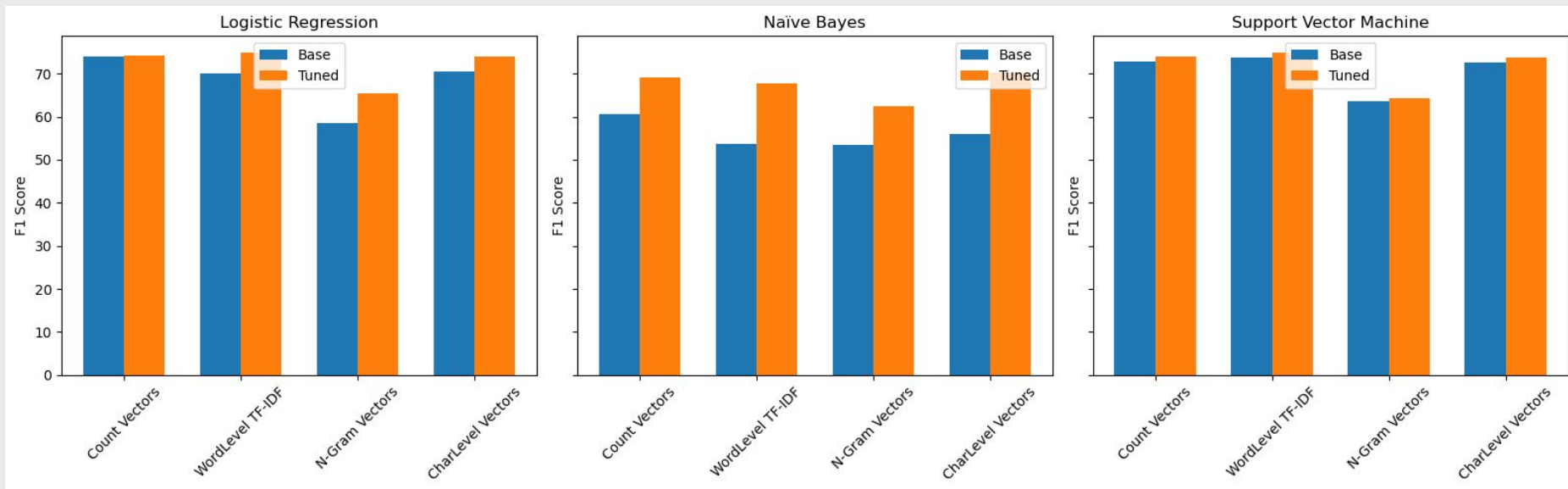
Models - Baseline (without tuning)

<u>F1-Score</u>	<u>Count Vectors</u>	<u>WordLevel TF-IDF</u>	<u>N-Gram Vectors</u>	<u>CharLevel Vectors</u>
<u>Logistic Regression</u>	74.02	69.93	58.55	70.43
<u>Naive Bayes</u>	60.63	53.68	53.47	55.87
<u>SVM</u>	72.7	73.72	63.57	72.6

Models - Baseline (with tuning)

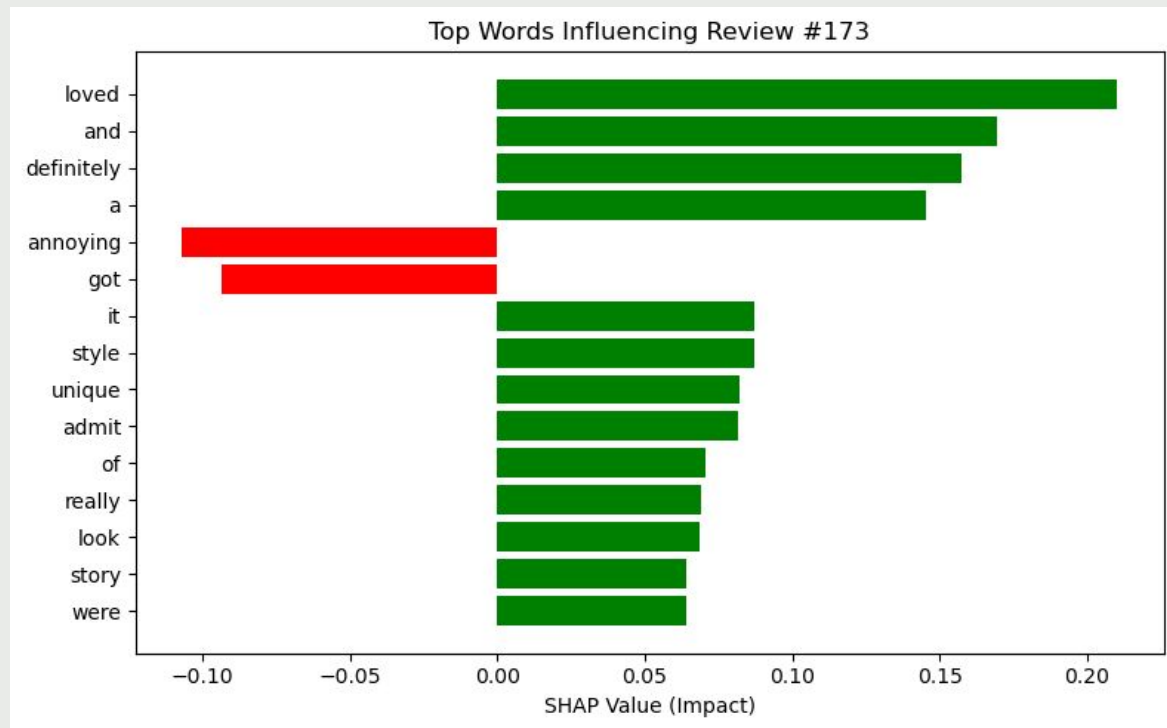
<u>F1-Score</u>	<u>Count Vectors</u>	<u>WordLevel TF-IDF</u>	<u>N-Gram Vectors</u>	<u>CharLevel Vectors</u>
<u>Logistic Regression</u>	74.2	74.98	65.5	73.9
<u>Naive Bayes</u>	69.22	67.75	62.36	70.29
<u>SVM</u>	74.07	74.96	64.2	73.73

Findings



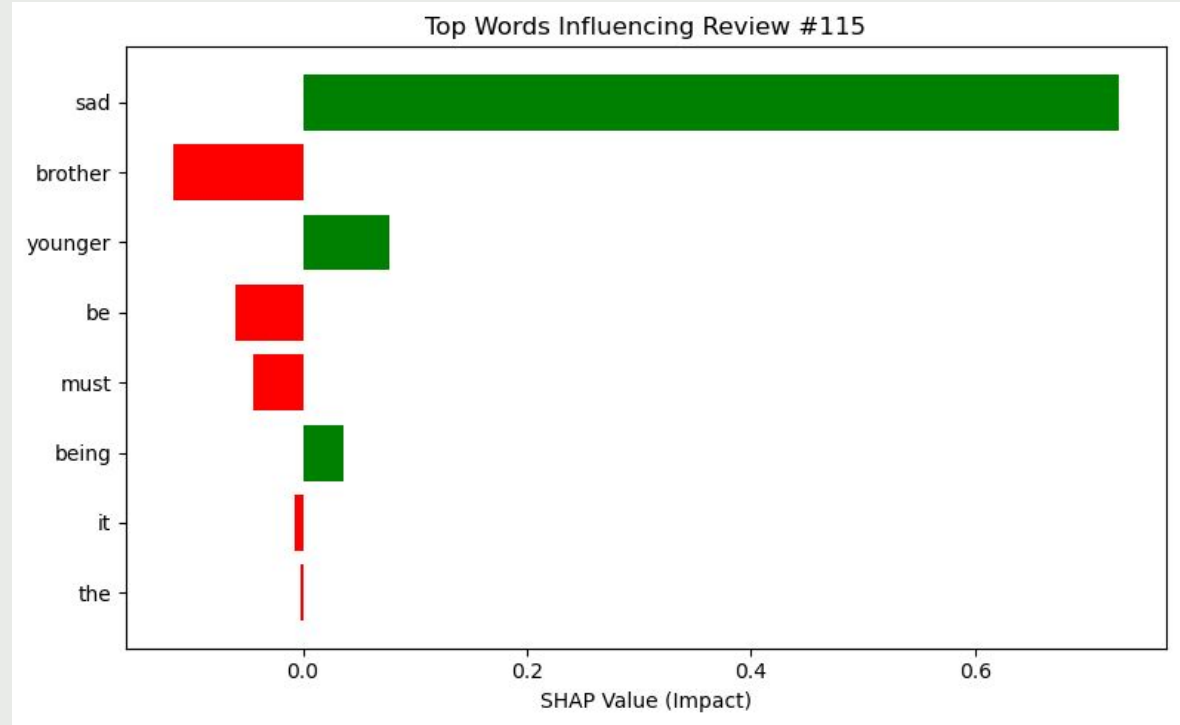
Feature importance per Review

Loved the art style and the story that it told, but a lot of the puzzles were extremely confusing and took a lot of time to figure out. I'll admit, there were a few that I had to look up a guide for, because it got really annoying at a few sections, but it was decent overall and definitely a unique puzzle game.
Predicted Sentiment: positive



Feature importance per Review

It must be sad being the overlooked
younger brother.
Predicted Sentiment: negative



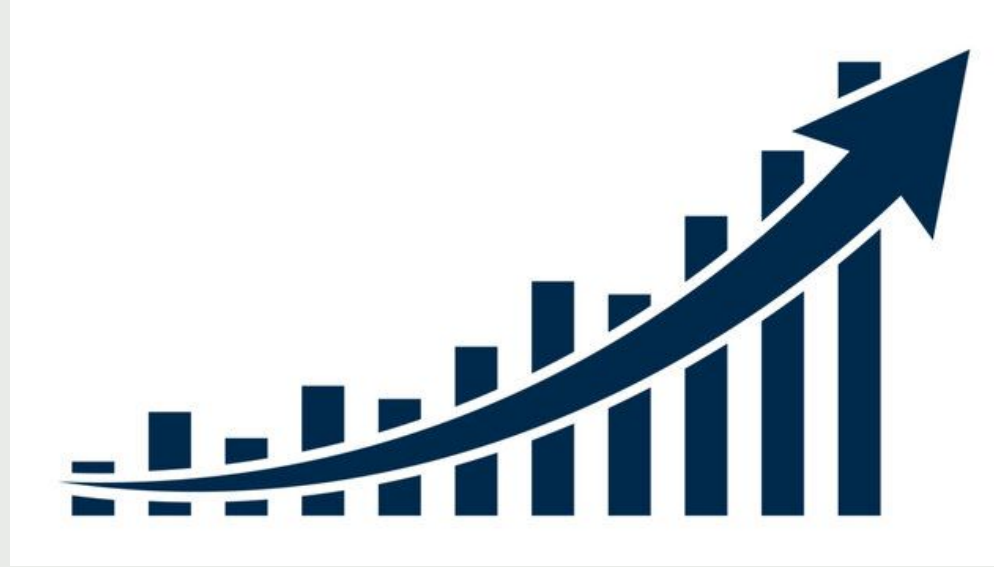
Conclusion

- Best performing model overall is SVM
- Best performing model with vectorizer is Logistic Regression with Word-Level TF-IDF
- Models meets the sentiment analysis requirement at 74.98%
- Models improvement after hyperparameter tuning

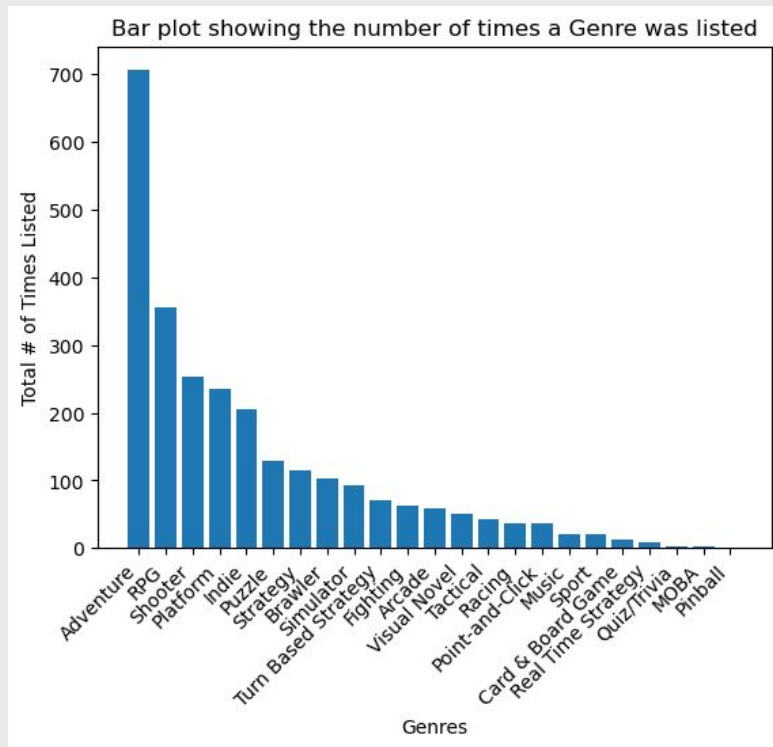


Future Improvements

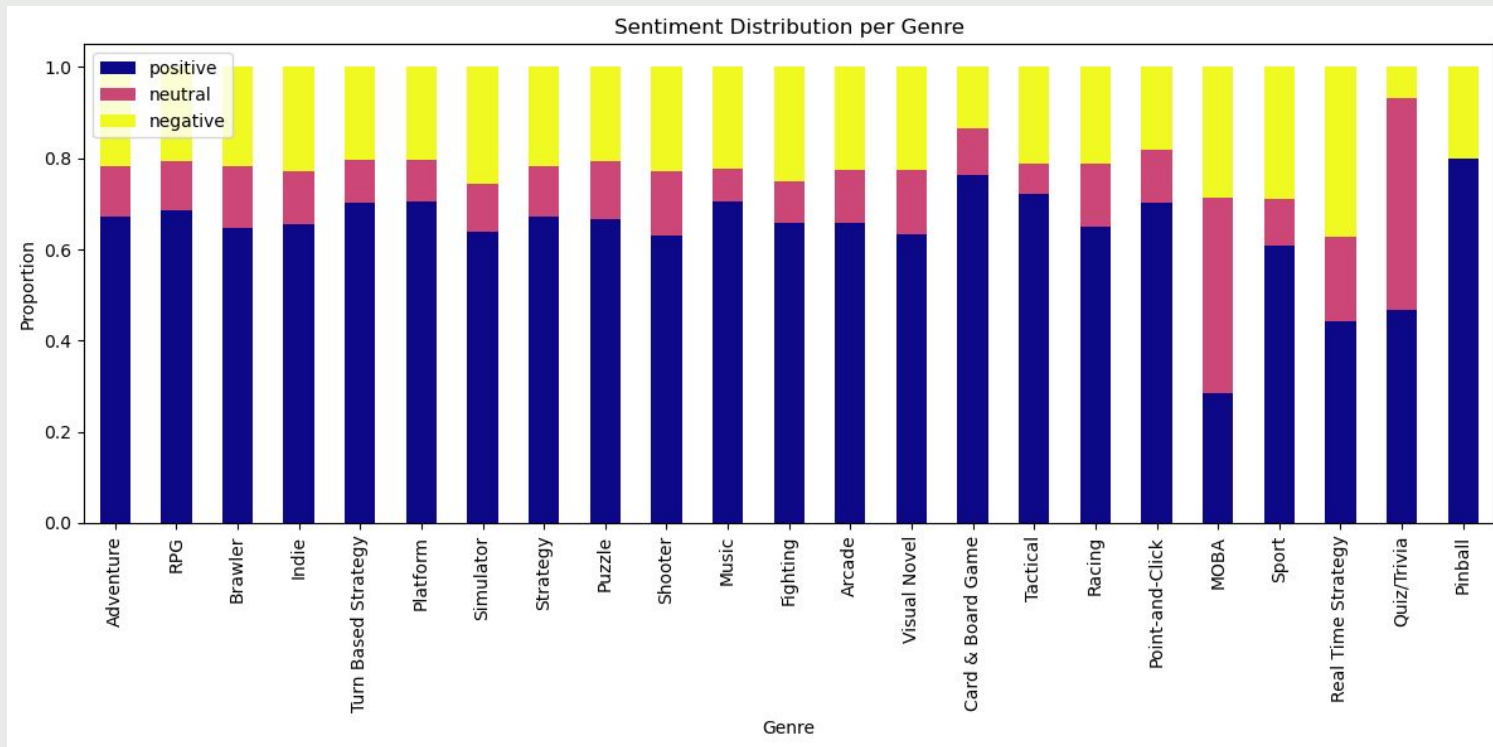
- Deploy an app to predict sentiments
- Improve score
- Bigger dataset
- Deep learning or transformer-based models



Exploratory Data Analysis



Exploratory Data Analysis



Thank You!
Q/A?

