Genre Based Sentiment Analysis on Popular PC Games

Alex Barnes, Anjun Luo, Rishabh Aggarwal

BAX 452 - Machine Learning

Table of Contents

Executive summary

This project aims to conduct in depth analysis on the user reviews of popular personal computer video games segmented by genres. A list of the most popular PC games with different genres and published between 2020 and 2021 is obtained from the RAWG database. All of the text and ratings from critic and user reviews on each of these games are then web-scrapped from MetaCritic.com and stored in mongoDB.

We were able to quickly identify the top keywords mentioned by users within a gaming genre by utilizing various natural language processing techniques. We constructed the frequency distribution of keywords based on sentiment and outputted our findings in the form of word clouds. We also implemented machine learning classifiers such as random forest and weighted logistic regression in attempts to predict user ratings based on the sentiment analysis for player reviews. Our model was able to achieve 80% accuracy in successfully classifying the sentiments as positive, negative or neutral.

Through this exercise we came to the conclusion that there is tremendous potential in applying machine learning techniques to the review texts of video games. Game developers will be able to obtain genre-specific insights on their customers. After further refinement, our analyses and models can be applied to a broader range of games.

Introduction & Background

The video game industry started in the 1950s and has been rapidly expanding on a global scale ever since. According to a report by Fortune Business Insights, global gaming market size is expected to reach around $500 billion by 2028 while a CAGR of 13.2% is expected between 2021 and 2028. It is estimated that around 2.8 billion people worldwide play video games, that is roughly one third of the world's population. It is clear that the gaming industry has become one of the larger sectors under the entertainment industry, and within the gaming industry there exists massive growth and profitable opportunities.

The structure of the gaming industry is rather complex. Video games can be developed and published on various platforms such as mobile, personal computers, consoles, virtual reality, handhelds as well as acardae. Each game can be published on one or more platforms and the game could be played either online or local. Traditionally, consoles and personal computers have been the stable platform in the gaming community, while mobile gaming has seen the largest growth in the past 10 years. In addition to the different platforms they are on, video games themselves can be categorized into over 30 different genres.

There are many gaming companies operating in different platforms and specializing in different genres, however, they are all after one thing – the players. The players are the ones to ultimately decide whether or not they want to purchase a game and continuously pay for gaming services (including monthly subscription or in-game cosmetics). The one-time purchase and the recurring payments are the two major forms of revenue stream for gaming companies. It is crucial for a game developer to understand what players are looking for especially in this highly segmented market. What a mobile puzzle gamer is looking for could be completely different than a console gamer that plays action games. By analyzing the needs and feedback from gamers that play a specific genre of the game on a specific platform can translate to valuable insights for gaming companies that specialize in that field. These insights can contribute to the refinement of current games as well as the development of new games. Both are essential for gaming companies to retain players and attract new players which directly correlates to the revenue generated.

Strategies & Business model

When it comes to the data of their players, gaming companies are definitely not in shortage of that. In general, almost any action a player does in the game is tracked and recorded. For example, the average amount of time per player spent per session of the game, the total number of players by month, the most popular or purchased in-game items in the last year, and the most common action a player takes before exiting the game. Almost all of these metrics provide a way for the company to gauge the performance of their game or to understand their customers. Some of these metrics are extremely valuable for companies to analyze the profitability of their games such as what items are being purchased by whom. However, these data are rather “passive”, in the sense that they only represent what the users do in game, and fail to answer other business questions such as “why are the average number of active users declining” or “how can we attract more new players”.

This type of question that highly relates to an individual player’s gaming experience requires data that are more “active”, for example, written feedback by the players. Traditionally, gaming companies have online forums for their gaming communities. These forums are a media platform in which players can communicate with other players as well as game developers. Game developers can judge whether a topic is important or widely recognized by the amount of replies in the tread and take proper action to respond to the players. There are a couple drawbacks to the forum, first thing is that most of the topics are related to “what needs to be fixed” in the game, and the second thing is forums are usually game-specific. That is to say, if one would want to know what features the community like about the game, it might not be easily found on a forum. Furthermore, if a gaming developer wants to find out information about a specific genre, such as what the users are looking for and avoiding in regards to a specific genre, that information is hard to obtain even from a collection of forums.

Our project aims to address direct player feedback by analyzing critic and user reviews on MetaCritic.com. Given the complex structure of the gaming industry, we will narrow our scope to just personal computer games that were published in the last two years. This would allow us to test our theories and models before applying our findings to a broader environment. Our plan is to aggregate all of the reviews on the most popular games under a certain genre and then utilize sentiment analysis and word cloud to highlight any frequently mentioned paired-words in both positive and negative sentiment. This will provide game developers insight on direct player feedback in different sentiment with regards to specific genre on a specific platform. In addition to the review text, the actual review rating will also be analyzed. The overall user and critic rating plays a significant role in when a new player is deciding to purchase. This is even more important for gaming companies that rely on the one time purchase revenue as their main revenue stream. We will examine the relationship between critic reviews and user reviews as well as building models to classify and predict the rating of a review based on the review texts.

Analysis

In this project, we segmented our analysis based on genre of the game using the actual critic reviews and user reviews collected in the database, some of the games have tens of thousands of reviews from players. Because desired gameplay characteristics vary by genre, this dataset offers in-depth genre-based research, allowing game developers to gain a better understanding of a subset of their gamers. The overall objective of this project has been categorized into two phases.

1. **Determine what users actually look for in different gaming genres** : Implemented by analyzing the frequency distribution of words and plotting a word cloud using Matplotlib

2. **Predict user ratings based on sentiment analysis for player reviews** : Implemented by using various robust ML classifiers such as Random Forest, Weighted Logistic regression.

Fig1 shows the sequence of methodology being used to achieve the overall project scope. Apart from these methods, the following mentioned technologies are used:

1. **Python -** Both the web scraper and classifier/regression models were made using Python.
2. **Numpy/ Pandas -** These modules are included to perform all the data preprocessing steps.
3. **Seaborn/Matplotlib** - These modules are used for plotting graphs and visualizations.
4. **Scikit-learn** - This is used for developing and training the classifier/regressor.
5. **Word Cloud** - Graphical representation of the frequency distribution of popular words.
6. **NLTK**- Natural language processing model being used to clean,transform, label the data.

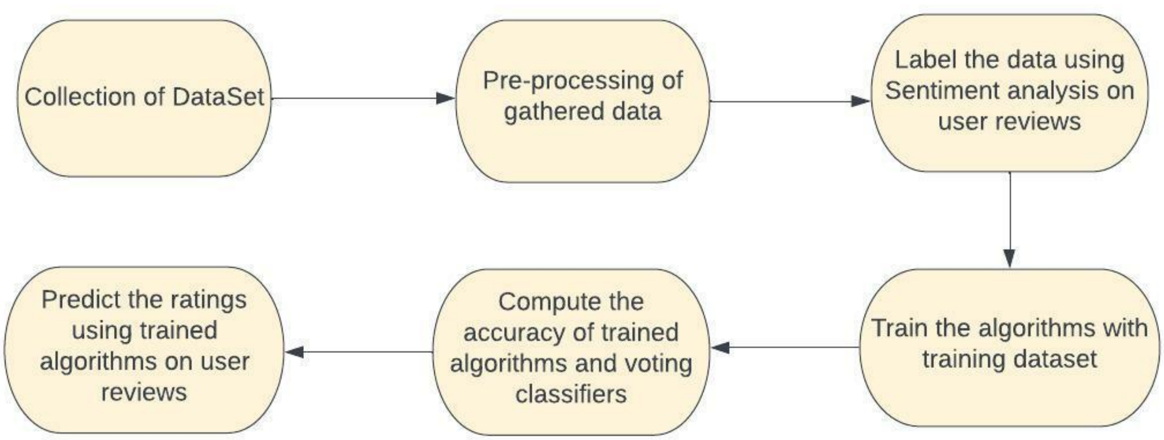

Fig1. Methodology

After collecting the data containing both the actual player reviews and critic reviews through web scraping from Metacritic.com and RAWG API for top 27 PC games, we stored the dataset into MongoDB collection. Using the Pymongo module , we queried the database for sentiment analysis of user reviews based on genre of the game using the methodology mentioned shown in Fig 1.

1. **Data Preprocessing** : User review contains various irrelevant words such as numerical digits, special characters, verbs , punctuations, symbols etc. These words are considered as noise and are therefore being removed before performing sentiment analysis. We used the NLTK library in python to perform data preprocessing /cleaning. Stop words such as “the”, “a”,”or”,”what” are also removed and stored in pandas data frame along with user ratings.

2. **Labeling the dataset**: Based on user ratings on the scale 1-10, labels have been assigned to each user review for all games belonging to a particular genre. User ratings above 7 have been labeled as positive, user ratings between 5-7 have been labeled as Neutral and below 4, user ratings are classified as Negative.

3. **WordCloud**: WordCloud is then constructed using a word cloud library in Python for each Positive and Negative reviews by players for every genre of the game to determine which things users actively look for in a game by genre.

4. **Bag of words/Vectorization**: In order to draw the relevant meaning from the text , a vocabulary needs to be formed based on which ML algorithms can be trained to accurately predict the sentiments of text reviews.To achieve this , text reviews are converted into vectors and we have used Bag of Words feature extraction technique to represent text based on occurrence of known words.

5. **Training and Testing Set**: In order to measure the accuracy of ML models, We splitted the dataset into two parts : training and testing data. For the training set, we took 80% of the dataset and the remaining 20% was used at testing data. The algorithms are trained using labeled data and then tested on a test set.

6. **Machine Learning Algorithm Model**: We have used an improvised Weighted Logistic Regression (for handling imbalanced datasets), Random Forest classifier to predict the sentiments/tone of user reviews. Based on the sentiment label of the user, overall user rating is then being computed using Random Forest regression due to its higher accuracy.

Results and Performance Evaluation:

With respect to the scope of the first objective of this project, we obtained some really interesting insights from player text reviews and did an overhaul comparison on what users are looking for in each genre. Fig 2 and 3 shows the word cloud comparison for positive sentiment reviews of genre **Sports Vs Adventure.** We can clearly see from the word cloud, In Sports genre, players are most concerned about mission, cars,variety, horizon,racing, gorgeousness of the games whereas in Adventure genre, people are most concerned about story, the beginning, ending, characters,industry, reality etc which can be really helpful for gaming companies to focus upon while improving the games in those genres. (See appendix 1 for more outputs)

Fig 2 : WordCloud on reviews for Sports genre

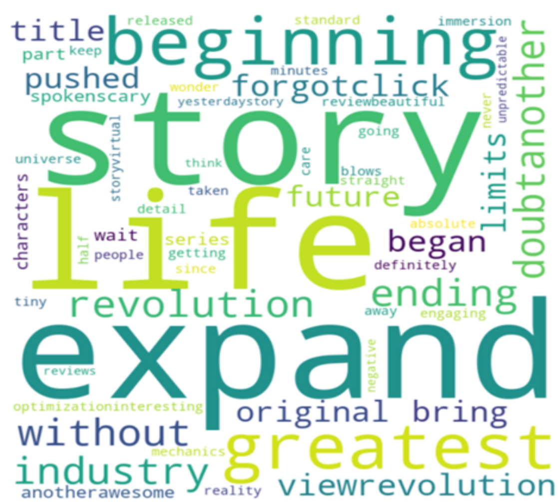

Fig 3: WordCloud on reviews for Adventure Genre

Towards our second objective, we used Random Forest classifier and Logistic Regression classification algorithm but due to highly imbalanced dataset , both these models were not able to accurately predict minority label classes (for eg Neutral) as shown by the confusion matrix obtained in fig 5. We can see that for Neutral class(label 1) ,the random forest classifier wasn't able to classify any of the reviews to Neutral class. On the other hand, when we did weighted logistic regression using actual weights for each class label **(Positive , Negative and Neutral)** in the conventional Logistic regression, the model was able to predict even minority label class (as can be seen from the figure on the right), and hence within class accuracy also improved in classifying user sentiments based on reviews .

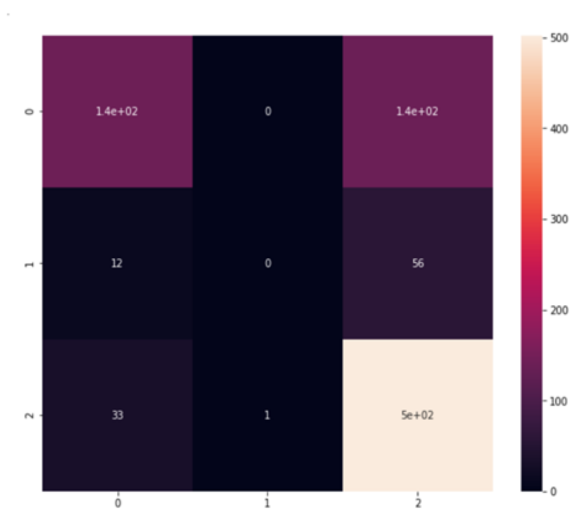

Fig: Confusion Matrix for Random Forest Classifier

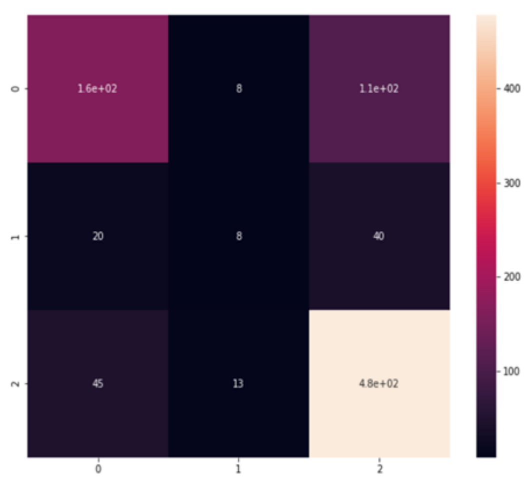

Fig: Confusion Matrix for Weighted Logistic Regression

The **Weighted Logistic Regression Model** is able to fix the inaccuracy in classifying **minority** class labels and we obtained **accuracy of over 80%** on the out of sample test data as shown in Figure below The reason for low recall of Neutral and negative classes is due to them being minority label class with significantly less data as compared to majority label class positive. As a result, we get high precision for positive sentiments.

	precision	recall	f1-score	support
Neutral	0.67	0.11	0.19	18
Negative	0.50	0.04	0.08	23
Positive	0.82	0.99	0.90	166
accuracy			0.81	207
macro avg	0.66	0.38	0.39	207
weighted avg	0.77	0.81	0.74	207

```
Training set score: 0.981
Test set score: 0.812
```

Fig: Classification Report of Weighted Logistic regression

The Random Forest Regression was able to predict the user ratings based on user sentiments classified previously with accuracy of 77% on train data set and over 32% accuracy was achieved on test data set as shown in the figure below:

```
Training set score: 0.776
Test set score: 0.324
```

Fig. Accuracy Results for Random forest Regression to predict user ratings

Recommendation & Business Values

We are successfully able to present an insightful comparison of word cloud on player reviews for different genres filtered on positive and negative sentiments. With both review text data, numerical score data, and game metadata we are able to dive deeper into what people are saying about games and things they are the most concerned about within each genre of the games.This can be used by gaming companies to develop a robust understanding on the user demands much quickly through a wordcloud and further improvements can be made in the future game development according the the players' preferences.

Using Weighted logistic regression we are successfully able to separate the sentiment of games for different genres. Based on different sentiments for user text reviews , we are able to successfully predict the user ratings reflecting overall public opinion and tone about the games for different genres. When a new player is contemplating whether or not to buy, the total user rating is quite important. This is especially true for gaming companies that rely on one-time purchase income as their primary source of revenue. In future we are looking to further improve the accuracy of the regression model and also include aspect based sentiment analysis where classification is done on various different aspects of games such as Gameplay, Graphics and Audio. Sentiment analysis across these aspects will provide gaming companies a much deeper understanding of user expectations and in turn allowing them to enhance user experience by much larger scale for their end customers.

Summary and Conclusions

In this project, our team has implemented various machine learning techniques from bag of words, sentiment analysis to classification on the user reviews of video games. For the purpose of this project, we have limited our scope to just the popular PC games that were published between the last two years. After processing the review texts, we were able to extract meaningful, genre-specific insights from the reviews. Using weighted logistic regression in our classification exercise, our team achieved a model with over 80% accuracy with test data. To conclude, we believe that our theories and approaches in attempt to understand the genre-specific video game player behavior are proven effective, and with further tuning and testing, our models can be applied to different game genres under different platforms

Appendix

We can see overhaul word cloud comparison for different genres in both positive and negative sentiments in the figures below:

1. Genre - RPG

Fig. Positive Sentiment WordCloud

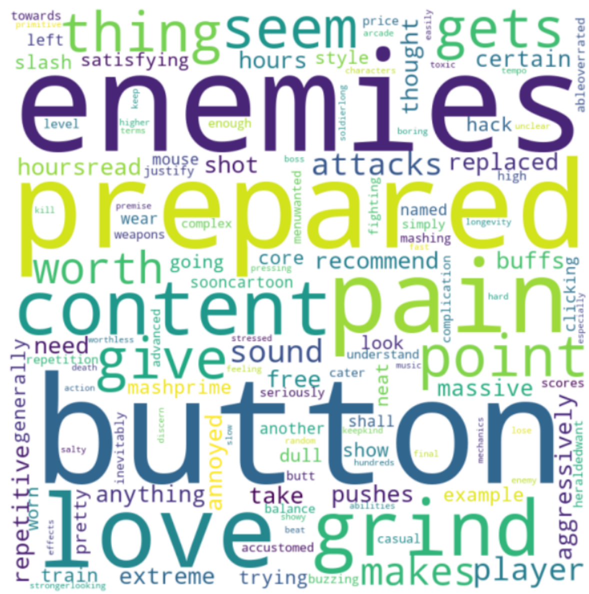

Fig. Negative Sentiment WordCloud

2. Genre - Indie

Fig. Positive Sentiment WordCloud

Fig. Negative Sentiment WordCloud

3. Genre Action :

Fig. Positive Sentiment WordCloud

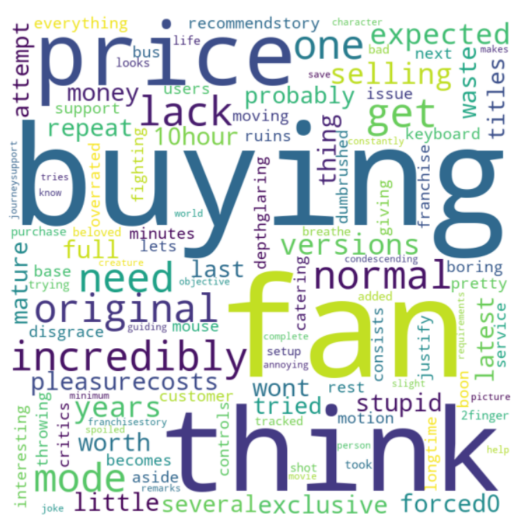

Fig. Negative Sentiment WordCloud

4. Genre Sports

Fig. Positive Sentiment WordCloud

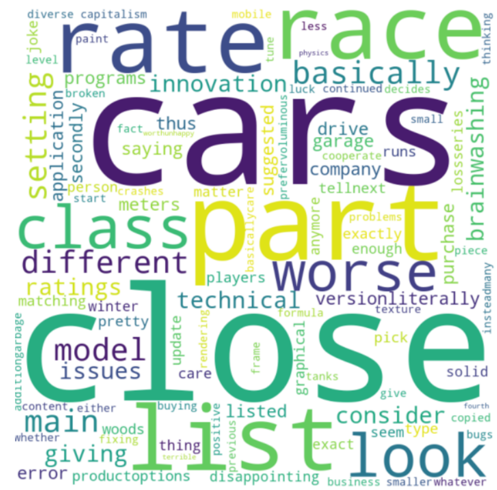

Fig. Negative Sentiment WordCloud

5. Genre Adventure

Fig. Positive Sentiment WordCloud

Fig. Negative Sentiment WordCloud

6. Genre Casual

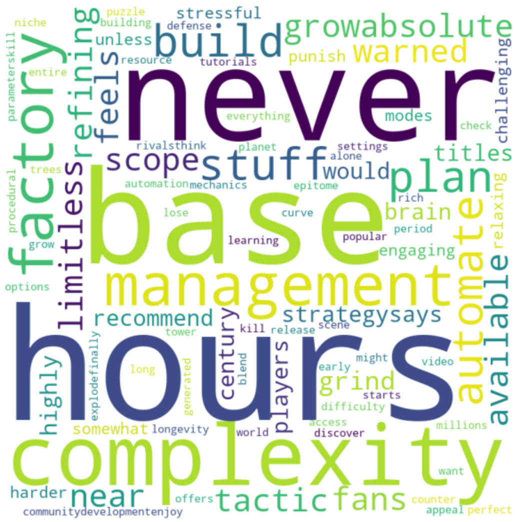

Fig. Positive Sentiment WordCloud

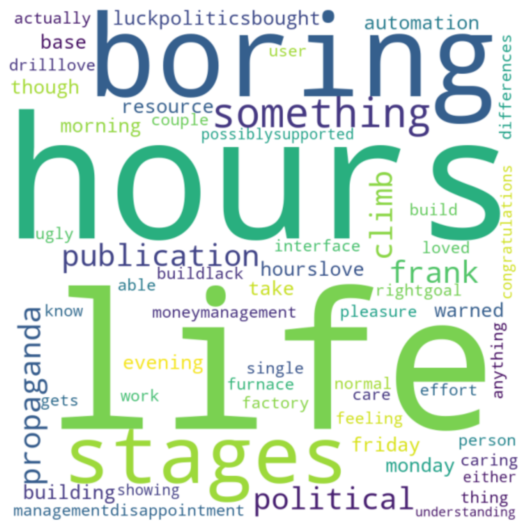

Fig. Negative Sentiment WordCloud

7. Genre Strategy

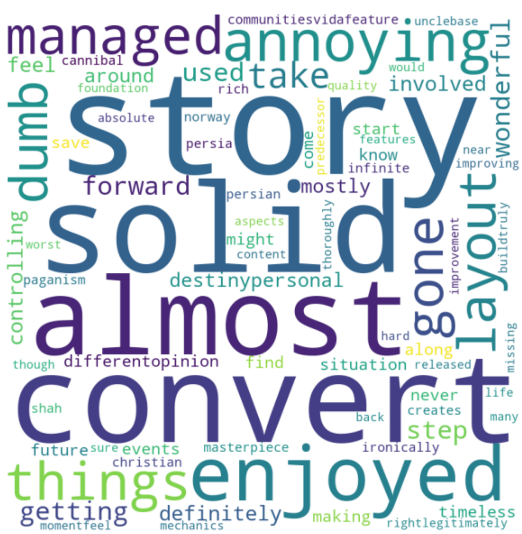

Fig. Positive Sentiment WordCloud

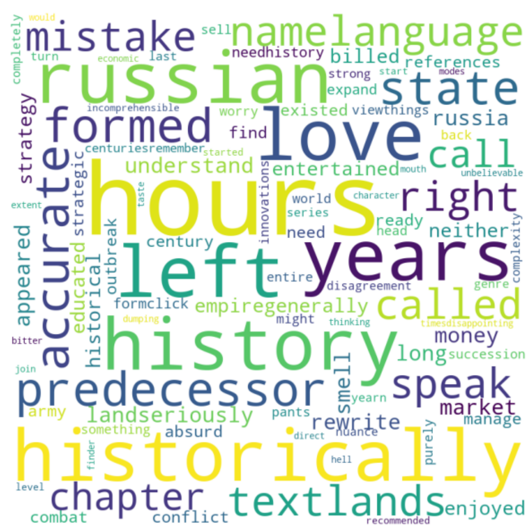

Fig. Negative Sentiment WordCloud

References

- https://www.globenewswire.com/news-release/2022/01/24/2371836/0/en/Gaming-Market-Size-Moving-Upwards-to-Hit-USD-545-98-Billion-by-2028-Global-Gaming-Industry-Share-Heightening-Vision-in-Games-Sector-Fortune-Business-Insights.html#:~:text=Fortune%20Business%20Insight%2C%20in%20its,users%20with%20premium%20quality%20games.
- https://financesonline.com/number-of-gamers-worldwide/#:~:text=Statistics%2C%20and%20Predictions-,Number%20of%20Gamers%20Worldwide%202022%2F2023%3A%20Demographics%2C%20Statistics%2C,%2Don%2Dyear%20growth%20forecast.