Final Report

—— Extracting product performance from user reviews of Steam to improve game experience

1. Executive summary

Using data exploration, data preprocessing and machine learning methods; we draw charts to help data visualization, create labels for comments through sentiment analysis, and help understand text through topic modeling. We find that many negative reviews related to adventure-action games are *weapon*, *combat* and *enemy*. The game's storyline is a concern for many players. Some players encountered technical issues like crashes and bugs. Combined with the current game trends, we recommend game developing companies pay more attention to the weapon design, which means the weapon types in the game should have enough variety and playability. In terms of combat type and enemy design, designers should consider the style characteristics of their own games, but still need to ensure the high quality of these two aspects, and more funds can be invested in the improvement of combat quality, such as a sufficiently good physics engine.

2. Project objectives

Our goal is to find out which type of the games and which kind of characteristics of those games on steam that are most likely to dissatisfy the players. There might be some common topics of the poor reviews like ‘Poor setting’, ‘Violence’,’ Pornography’, ’Tedious’ and the stuff like that. By acknowledging those games with high frequency negative comment words, it can serve as a warning sign to subsequent game design companies so that they can make modifications and improvements while designing new games.

3. Data Description

【Data Source: https://www.kaggle.com/datasets/andrewmvd/steam-reviews】

Steam is the leading PC platform for games. Throughout its many years of online service, Steam has amassed a large number of reviews around its games. These reviews represent a great opportunity to break down the satisfaction and dissatisfaction factors around games as well as sentiment over time.

The dataset contains over 6.4 million publicly available reviews in English from the Steam store run by Valve. The dataset is streamlined and complete, but unfortunately there is no indicator for the game genres. We are unable to analyze each game type in detail. Besides, since the file was too large for python to handle the data, we chose to sample a random portion for our study.

Column Name	Description
app_id	Game ID. Each game has a unique ID.
app_name	Game name
review_text	Review content submitted by users.
review_score	Whether the review recommends the game or not. 1 for positive reviews and -1 for negative reviews.
review_votes	Whether any other users found this review helpful 1 means yes and 0 means no.

The overall distribution of the data (*Figure 1*) shows that there are far fewer negative reviews than positive reviews, roughly one-fifth as many as positive ones. And from *Figure 2*, we can see that the number of reviews received varied widely from game to game, with some of them significantly hotter than others.

4. Methodology

A. Explore the dataset
   * Summarize the distribution of positive/negative comments
   * Summarize the number of reviews per game

B. Data Preprocessing
   * Tokenize / Laminate the text data
   * Remove the stop words/emoji/meaningless signs
   * [While supervised learning], Use Tf-IDF to vectorize the words

C. [Main Methods] - Data Analysis
   * Sentiment Analysis
     -> Use the unsupervised machine learning method - VADER Lexicon to set the positive/negative score for each sentence
     -> After trying different thresholds, we set the optimal positive/negative classification score for each row of text
     -> Using group by method, we compute the negative score for each game and select the top 3 games with the most negative score
     -> By computing the word frequency, we draw word clouds for positive/negative reviews for later analysis
     ->Use the supervised machine learning method - Support Vector Machine (SVM) to compare the result with the unsupervised- VADER Lexicon method above.

D. Topic Modeling
     -> Normalizing the review text column focused on games with negative scores.
     -> Split the data into 4 topics and show the most exclusive terms (λ =0.01).
     -> Visualize the topic distribution and label those topics by game type.

5. Results and their Discussion

(1) Sentiment Analysis

Since we have human-created labels for sentiment polarity in our data (positive/negative), we used both the unsupervised lexicon-based method and the supervised machine learning method for our sentiment analysis.

A. VADER Lexicon-based analysis

Through exploration, we found that in order to maximize the overall accuracy, the threshold parameter for defining positive polarity should be set at -0.8788 *(See Figure 3)*. The corresponding precision is about 82.89% and the recall rate is about 98.57%. Therefore, in this case, the VADER lexicon-based tool is better at identifying negative reviews than positive reviews.

By grouping the game names, we calculated the average VADER score, the number of negative reviews received, and the percentage of negative reviews to the total reviews for each game. The 3 games with the lowest average scores are *Run The Gamut*, *Snow Light* and *Hard to Be a God*. At the same time, however, we found that the games with high rates of negative reviews are those that originally had very few reviews, which is reasonable since poor games cannot attract more players to try them. The only game with more than 100 total reviews and a negative review rate in the top 10% is *Condemned: Criminal Origins*.

After getting the predicted polarities for each record, we paid particular attention to the high frequency words that appear in those negative comments, as we may be able to detect the points of dissatisfaction of users from them. Ignoring some intonation words and adverbs in the word cloud *(Figure 4)*, we can broadly identify the followings:

- Many of the games to which these negative reviews belong are adventure-action games. Many words such as *weapon*, *combat*, *enemies*, *fight* and *battle* all point to games related to confrontation and competition.
- The game's storyline is a concern for many players, which is reflected in the words such as *story*, *character* and *mission.* For example, some people felt bad because the game story ended too abruptly.
- Some players encountered technical issues like *crash*es and *bug*s.

B. SVM-based analysis

We split the data into a training set and a test set at the ratio of 7 to 3. The accuracy of the final obtained model is 83.1%, a little higher than that of the lexicon-based sentiment classifier, but not too much different.

(2) Topic Modeling

After preprocessing the data, we created visualization for the LDA models with 4 topics and λ equals to 0.01 which indicates the sense of exclusiveness. The exclusive term is particularly informative of the contents of the topic as these terms almost do not appear in any other topic. We will exclude those adjunct words while analyzing the topic. In principle, the largest bubble of a topic in the visualization usually stands for the largest amount of text devoted to that topic. In the biggest selected topic, the 5 most exclusive terms are weapon, map, gun, war and team. Indeed, the shooting games based on team works such as *Playerunknown's Battlegrounds* and *Counter Strike* are quite popular among players and there are some common flaws including 'Uneven fire power of weapons and guns', 'Lack of diversity in the map' and 'Imbalance level of teammates. Therefore, we decided to label this topic as "Shooting games". In the second largest selected topic, the 5 most exclusive words are story, puzzle, music, style and RPG. Besides the adventure-action games, another hot game type is role-playing game which includes a whole storyline for the player to explore. Similarly, RPG games will always encounter some negative reviews like 'A tiger's head, a snake's tail' or 'Puzzling and bewildering plots'. As a matter of fact, we decided to label this topic as "RPG games". In the third selected topic, the 5 most exclusive items are awesome, friend, buy, download and recommend. At this part, rather than focusing on the game types, the negative reviews are mainly complaining about the poor game setting systems. For lacking the regular maintenance, players are sometimes suffering from crashed communication channel via online friends or tough game top-ups issues. Thus, we decided to label the third topic as "Game setting". We decided to not consider the last topic since the result is too diverse to form a general topic.

6. Conclusion

(1) Recommendations

Because open-world action-adventure games are the game trend in recent years, which means many companies will try to make similar games to occupy the market share, and many reviews we analyzed are related to this type of game, we try to make recommendations to this type of game designing. According to our analysis results above, we know that for this type of game weapon design, combat, and enemy design are the three parts that require the designer's attention the most. Combined with the current most popular and most acclaimed open world adventure action game Elden Ring for analysis. Elden Ring is undoubtedly a perfect solution to the most common problems with bad reviews for this type of game. Compared with the previous game works, this game has added many weapon types and weapon special effects, which greatly enhances the player's game experience. In terms of combat design, Elden Ring continues the consistent combat system of Dark Soul-like games, which has been loved by players for many years. In terms of enemy design, since the game type of Elden Ring itself tends to be dark, bloody and violent, the enemy design is also relatively bloody. For the new game design, the diversified design of weapons can be borrowed from the Elden Ring, but for the combat design and enemy design, I think each game has its own suitable type, and the designer should pay attention to the quality. rather than type. No matter what type of combat mode it is, it needs to ensure high quality, such as excellent physics engine, model collision volume, etc. The enemy type needs to be designed according to its own game type and style.

(2) Shortcomings

The main disadvantage of our analysis process is that the text data itself is too cluttered and general. Most of the words in the word cloud we made through sentiment analysis have no guiding value, and we can only infer and analyze through a small number of high-frequency words. I think we need more advanced ways to filter words to ensure that the high-frequency words that appear in the results are more representative.

Appendix

Figure 1: Positive/Negative reviews count

Figure 2: Number of reviews per game

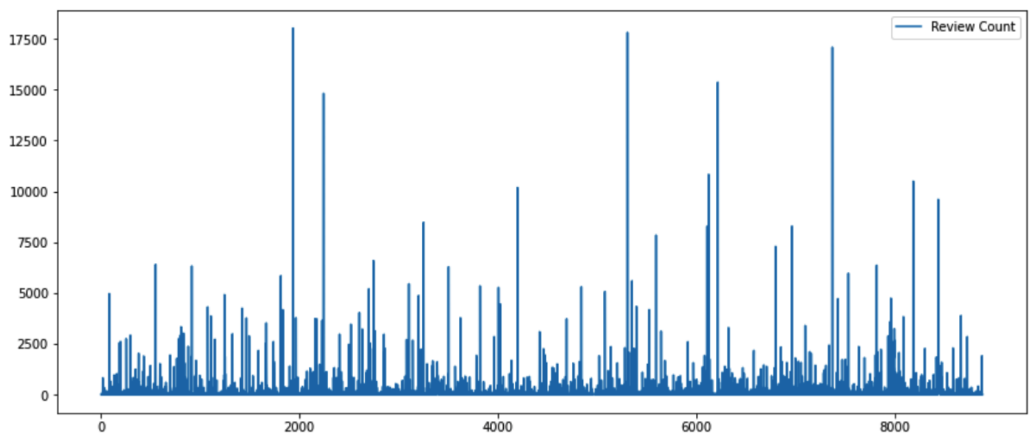

Figure 3: Changes in accuracy rate for a range of thresholds from -1 to 1 (VADER score)

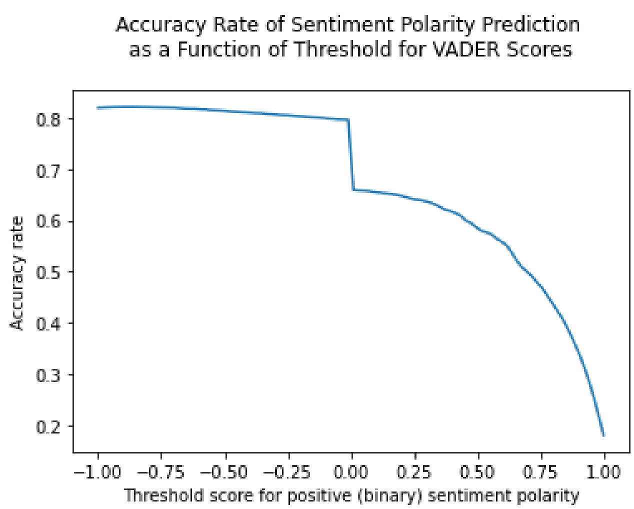

Figure 4: Word Cloud of negative reviews based on VADER-Lexicon analysis