MSc in Business Administration and Data Science
Final Project Natural Language Processing and Text Analytics

From Reviews to Insights:
Advanced Topic Modeling in Gaming

Álvaro Díaz-Dávila Álvarez; 167299

Jan Bendix Portius; 167605

Jialing Zhou; 166673

Long Chen; 166647

Abstract

This study analyzes Steam video game reviews to identify players' needs and preferences using three topic modeling methods: Latent Dirichlet Allocation (LDA), Top2Vec, and BERTopic. The objective is to determine which method provides the most accurate and insightful analysis of these reviews. The findings indicate that BERTopic outperforms the other models in terms of coherence, relevance, interpretability, and topic diversity, offering actionable insights for game developers. Additionally, augmenting the results with OpenAI's GPT-4o proves to be useful in enhancing the interpretability of the identified topics. The study concludes by recommending BERTopic for analyzing video game reviews and suggests further research to enhance topic modeling evaluation metrics.

Keywords: Topic Modeling, Video Game Reviews, BERTopic, LDA, Top2Vec, Interpretability, GPT-4o.

1 Introduction

Video game reviews provide a valuable source of feedback for developers, revealing players common needs and preferences (Tong, 2021). Identifying and addressing these concerns can lead to improved game design, enhanced user satisfaction, and ultimately, increased sales. Topic modeling can help uncover these relevant needs and preferences, offering actionable insights for game development and marketing strategies.

To identify the most effective topic modeling approach for analyzing video game reviews, this study compares three methods: Latent Dirichlet Allocation (LDA), Top2Vec, and BERTopic. By evaluating these models through both quantitative and qualitative metrics, the study aims to determine which method performs best in extracting meaningful insights from the reviews.

Additionally, the output of the best performing model is augmented with OpenAI GPT-4o to enhance the interpretation of the identified topics. The integration of GPT-4o enables a more robust and nuanced understanding of the results, facilitating companies to better understand and respond to their customers' preferences for video games.

All review data used in this study is sourced from Steam, which is the largest digital distribution platform for PC games. It has a strong global presence, holding an estimated 75 percent market share in the U.S. and 80 percent in the EU (Osum, 2024), which makes the collected game review data highly representative.

To incorporate potential differences across games, this study applies topic modeling to reviews from three different games: PlayerUnknown's Battlegrounds, Grand Theft Auto V, and Rust. They are the three most popular games in the dataset, accounting for 73 percent of the total number of reviews of this dataset.

The remainder of this paper is organized as follows: Section 2 reviews the relevant literature. Section 3 covers the methodology, including dataset description, exploratory data analysis (EDA), model overview and metrics. Section 4 presents the results of the study and Section 5 discusses these results. Section 6 outlines the limitations of the study and Section 7 provides suggestions for future research. Finally, Section 8 concludes the paper.

2   Literature Review

Busurkina et al. (2020) conducted a study on the important features of game reviews by applying the Structural Topic Model (STM) to Steam review data. They identified 120 topics and grouped them into seven meta-topics: Battle and Achievements, Value, Social Interactions, Game Performance and Visuals, Experience, Social Influence, and Gaming Progress. Their findings support the idea that reviews are valuable for product improvements, as they provide detailed insights into various aspects of the gaming experience. However, the study did not differentiate between individual games; instead, it analyzed a single sample containing all games. As a result, potential differences between games remain unclear. Understanding these differences could further enhance the value of reviews for developers by highlighting specific areas for improvement tailored to each game.

When considering the models that can be used for such analyses, recent advancements in topic modeling have highlighted the differences in performance between BERTopic, Top2Vec, and LDA. Gan et al.'s recent study (2024) conducted a comparative analysis of these methods using data from the #ChatGPT topic on Twitter and Weibo. The analysis revealed that BERTopic demonstrated superior performance in cluster separation and topic independence, outperforming LDA and Top2Vec by at least 34.2 percent in both Chinese and English datasets. While traditional evaluation metrics like the silhouette coefficient were used, the study noted their limitations, suggesting a need for further exploration of alternative evaluation indicators and advanced modeling approaches.

Further supporting these findings, Chen et al. (2023) conducted a comparative study on the application of state-of-the-art topic modeling techniques - LDA, Top2Vec, and BERTopic - specifically for analyzing the impact of news on financial markets. They developed a framework called "News Impact Analysis" (NIA) to automate the correlation of news topics with financial market movements. Their results show that BERTopic outperforms LDA and Top2Vec in terms of coherence, interpretability, and computational efficiency, making it particularly suitable for large-scale financial news datasets. The study emphasizes that there is not a one-size-fits-all metric for evaluating topic models, highlighting the limitations of coherence scores and the need for complementing them with human judgment. However, as their study focused on long newspaper articles, its findings might not directly apply to the analysis of short game reviews in the current use case.

Aside from the advancement of models applied in topic modeling, the interpretation of the output is also of great interest and importance to researchers. Gillings and Hardie (2023) compared two interpretation approaches on the output of the LDA model for a collection of articles regarding climate change. The result indicates that interpretation based on scrutinizing the most representative documents for each topic outperforms the "eyeballing" method, which involves merely examining the most associated words for each topic. This study provides valuable insight for the interpretation process by emphasizing the importance of considering the context provided by representative documents in addition to keywords.

Considering the reviewed literature it is clear that

the three main algorithms used in the field are LDA, Top2Vec, and BERTopic, justifying the decision to choose them as well. In terms of metrics, there is no single quantitative metric that fully captures the quality of the topics, highlighting the necessity to complement them with human evaluation.

3 Methodology

3.1 Data Analysis Process

Figure 1 provides an overview of the underlying workflow:

3.2 Data Description

The dataset used in this project is a collection of reviews for the most downloaded games on Steam in 2019. It is obtained from Kaggle, and the reviews range from as early as 2010 to 2019, with the bulk of the reviews coming from the last two years. This dataset contains approximately 435,000 reviews and includes 8 different features related to each comment. These features include the game the review was written about, the number of hours the user played before writing the review, and the number of upvotes the review received, among others.

There are no duplicated rows, and the only missing values are in the reviews column, accounting for 0.03 percent of the dataset.

The reviews cover 48 different games; however, the three most popular games account for nearly 73 percent of the reviews, and the top five games represent 94 percent of the reviews, as illustrated in Table 1. Given this distribution, the decision is made to focus the analysis on the top three games, as they have the highest number of reviews, allow-
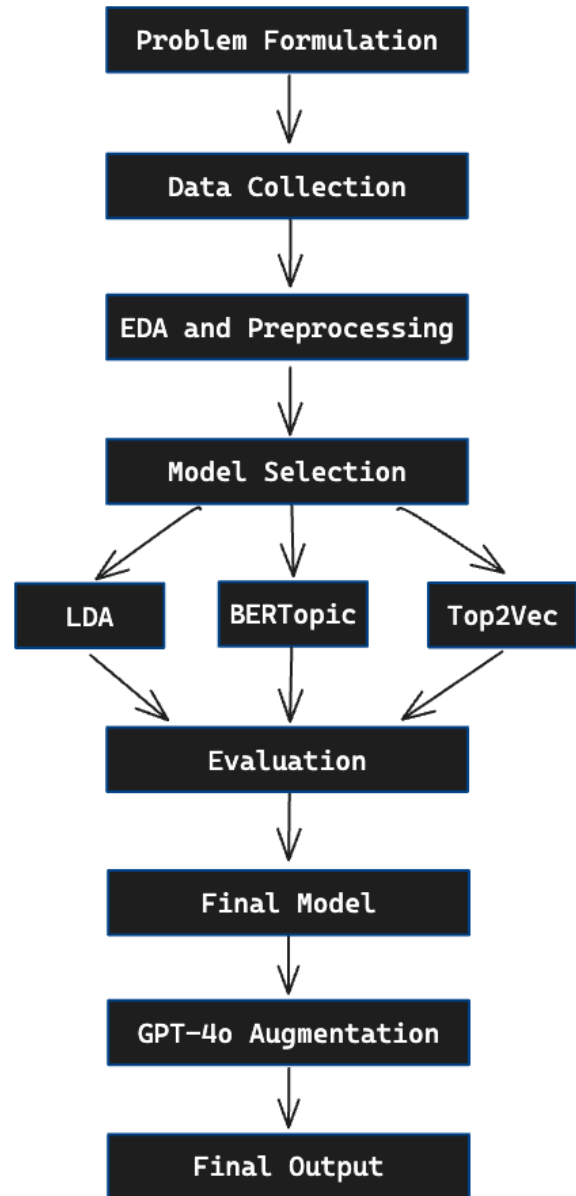

Figure 1: Data analysis process.

ing for a more extensive and detailed examination of each game.

The average length of a review is 40 words, with a highly skewed distribution. The median review has only 13 words, while some reviews exceed 1,000 words. This skewness is not a concern because the models used in this report are robust to outliers, so no transformations are applied to maintain an accurate representation of the reviews.

3.3 Preprocessing

Based on the exploratory data analysis, no duplicate rows exist and the missing values in the review feature are removed from the dataset. Then the selected three games' reviews are filtered out of the dataset for further training process.

As Top2Vec and BERTopic leverage pre-trained embeddings and do not require extensive preprocessing, the following steps are exclusively for preprocessing data for training the LDA model. The reviews extracted from the dataset are converted to lowercase and tokenized by applying *simple_preprocess* from Gensim. All the stop words have been removed as these data are considered containing too little useful information for formulating a specific topic. Then the remaining tokens are lemmatized to base format to improve token consistency and enhance the topic formulation coherence. The selected and preprocessed dataset of the three games' reviews are utilized to create separate dictionaries of bag-of-words representation for the LDA training. Further, the tokens that appear in less than 10 documents or appear in more than 50 percent of all documents are filtered out to reduce noise and improve the quality of topics generated.

3.4 Models

The study applies three models - LDA, Top2Vec and BERTopic - to independently process and analyze the reviews of each of the top three chosen games. An overview of each model is provided subsequently.

LDA: Latent Dirichlet allocation (LDA) is a generative probabilistic model of a corpus introduced by Blei, Ng and Jordan in 2003, and has become one of the most commonly used topic modeling methods. The basic idea is that documents can be represented as random mixtures over latent topics, where each topic is characterized by a distribution over words (Blei et al., 2003). A common statistical inference method used for parameter estimation is Gibbs Sampling, which is an iterative Monte Carlo algorithm to generate a sample from joint distribution when only conditional distributions of each variable is efficiently computed (Jelodar et al., 2019). Although LDA has the advantage of revealing latent semantic connections between words and is capable of generating multiple topics for one document compared with some previous models, it also has some constraints (Maier et al., 2021). As LDA relies on bag-of-words methods and ignores ordering or semantics of words, specific context information regarding semantic relations among words is omitted (Lenci, 2008). Moreover, LDA assumes that topics are independent, making it fail to model correlation between topics (Blei et al., 2007).

Top2Vec: Top2Vec is an advanced topic modeling algorithm introduced by Pietsch and Rocktäschel in 2020, designed to better capture semantic information in text corpora. Unlike traditional models like LDA, which rely on bag-of-words representations, Top2Vec integrates word embeddings

Game	Number of Reviews	Cumulative Percentage
Playerunknown's Battlegrounds	144846	33.42
Grand Theft Auto V	99677	56.42
Rust	70907	72.78
Rocket League®	67765	88.42
Dead by Daylight	22178	93.54
Monster Hunter: World	18390	97.78
Astroneer	2658	98.40
The Elder Scrolls V: Skyrim Special Edition	1471	98.73
Resident Evil 2 / Biohazard Re:2	1384	99.05
Sid Meier's Civilization®VI	522	99.17

Table 1: Number of Game reviews and their cumulative percentages.

with the clustering technique HDBSCAN. The algorithm works by embedding documents in a high-dimensional vector space and then clustering these vectors to identify topics. Notably, the topic vectors in Top2Vec are calculated from dense areas of highly similar documents, ensuring that the topics are both informative and representative of the corpus. This approach allows for the automatic determination of the number of topics without requiring stop-word lists, stemming, or lemmatization, and facilitates a more nuanced representation of document content. Additionally, Top2Vec internally leverages a novel evaluation method using mutual information to measure the informativeness of topics, consistently outperforming traditional models in producing more coherent and meaningful topics. Despite the computational demands of the embedding process, Top2Vec is particularly effective for diverse and large-scale text datasets (Angelov, 2020).

BERTopic: BERTopic is a neural topic modeling technique that uses pre-trained language models to create document embeddings, clusters these embeddings, and generates topic representations using a class-based variation of TF-IDF (Grootendorst, 2022). Similar to Top2Vec, BERTopic leverages embeddings to capture semantic information and identify topics. Egger and Yu (2022) highlight several advantages of BERTopic. It is highly versatile and stable across various domains, requires no preprocessing due to its use of embeddings, and automatically identifies the number of topics. However, they mention some drawbacks to this embedding approach. It can generate an excessive number of topics, necessitating labor-intensive inspection of each one. Additionally, each document is assigned to a single topic, even though documents often encompass multiple topics. Moreover, objective evaluation of the results is challenging.

3.5 GPT-4o Interpretability Augmentation

ChatGPT, released by OpenAI in 2022 and subsequently updated, offers a sophisticated chatbot interface powered by Generative Pre-trained Transformers (GPT), which are advanced language models known for their exceptional text generation capabilities. These models excel in various applications, in-

cluding content creation, language translation, and contextual understanding.

In this report, GPT-4o is utilized not as a topic modeling algorithm, but as a support model to help overcome a common limitation of these three models: the lack of interpretability of results. Typically, the output of such models is limited to the most representative words and reviews for each topic, necessitating manual effort to extract the main ideas and understand the underlying meaning. By using the OpenAI API interface, the most representative words and reviews for each topic are sent with a custom prompt to streamline and facilitate the process of creating detailed summaries for each topic.

With the best performing model, OpenAI's GPT-4o model is used to interpret the topics and provide a detailed summary. This facilitates and enhances the interpretability of the output and assigns a score from 1 to 10 for the overall sentiment, indicating the negativity or positivity of each topic. For each topic, the OpenAI API is called with the following structure:

1. **System Prompt**: A set of instructions tailored to our use case. These instructions include an introduction to the data, the specified game, and how the output is supposed to be structured. Additionally, they provide guidelines on tone and formatting:

   You are an assistant whose job is to write an informative summary based on the content of Steam videogame reviews provided by the user.
   The summary should be descriptive and informative, allowing someone

to fully understand the overall sentiment and key points discussed in the reviews.
Additionally, include a score from 1 to 10, where 1 indicates highly negative reviews and 10 indicates highly positive reviews. Do not explain what the game is; assume the reader is already familiar with it.

2. **User Prompt**: This prompt contains data obtained for each topic from the best performing model, including the topic name, the most common words, and the most representative reviews. The data is formatted as a string with delimiters and title headings to clearly signal the different sections to the GPT-4o model as suggested by OpenAI (n.d.).

   Topic Name: {topic_name}
   —————
   Most common words:
   {most_common_words}
   —————
   Most representative reviews:
   {most_common_reviews}

3.6 Metrics

To compare the performance across all three models, both quantitative and qualitative metrics are used. As a quantitative metric, topic diversity measures the distinctiveness of the topics. As indicated in the relevant literature, there is no one-size-fits-all quantitative metric for assessing performance. Therefore, qualitative metrics are also employed, as they can provide a more comprehensive picture of the models' performance. These qualitative metrics

are obtained by human evaluators who manually rate the coherence, relevancy, and interpretability of the different topics.

3.6.1 Quantitative Metrics

Topic diversity measures the distinctiveness of the topics produced by a model. High topic diversity indicates that the topics are varied and cover different aspects of the corpus, while low diversity suggests that topics may be redundant. To evaluate the diversity of topics generated by a topic modeling algorithm, several steps are taken. Firstly, after executing the algorithm, the top N words for each topic are extracted based on their highest probability within the respective topic. Subsequently, pairwise overlap between topics is calculated. This involves measuring the similarity of the top words between each pair of topics using the metric - Jaccard similarity coefficient, which quantifies the similarity between finite sample sets. Specifically, the Jaccard similarity is computed as the intersection of the sets divided by their union:

$$\text{Jaccard Similarity} = \frac{|A \cap B|}{|A \cup B|}$$

Here, A and B represent the sets of top words from two different topics. Once pairwise similarities are obtained, an overall diversity measure is derived by aggregating these overlaps. A common approach that involves computing the average pairwise Jaccard similarity and subtracting it from one is the diversity score:

$$\text{Topic Diversity} = 1 - \frac{1}{\binom{T}{2}} \sum_{i<j} \text{Jaccard Similarity}(T_i, T_j)$$

Here, T is the total number of topics, and $\binom{T}{2}$ represents the number of unique topic pairs considered for the calculation. This score shows the extent to which topics are distinct from each other across the corpus. The process considers every unique pair of topics to ensure a comprehensive evaluation of topic diversity. The higher the topic diversity score (approaching 1), the more distinct the topics are, covering different areas of the text corpus, which is typically desirable. Conversely, the lower the topic diversity score (approaching 0), the more topics are similar, suggesting redundancy in the model.

3.6.2 Quantitative Metrics

To complement the quantitative metric, a human evaluation is conducted to assess the performance of the topic modeling approaches from a qualitative perspective. This evaluation focuses on three key criteria: coherence, relevance, and interpretability of the topics generated by each model.

Evaluation Criteria

1. **Coherence**: This criterion examines whether the terms within each topic logically fit together. A high score indicates that the words in a topic are strongly related, suggesting that the topic is well-defined and coherent.

2. **Relevance**: This assesses whether the identified topics are relevant to the domain, in this case, video games. A relevant topic provides useful insights specific to the video game industry.

3. **Interpretability**: This measures how easily one can understand what each topic is about.

High interpretability means that the topic is clear and understandable, making it easier for developers and analysts to utilize the insights provided.

Each of these criteria is evaluated using a Likert scale from 1 to 7, where 1 indicates strong disagreement and 7 indicates strong agreement with the criterion being met.

For the human evaluation, the top five most frequent topics for each game from each model are selected, excluding outliers if applicable, resulting in 15 topics per model (5 topics from each of the three games: PlayerUnknown's Battlegrounds, Grand Theft Auto V, and Rust). This ensures that a diverse range of topics is included in the evaluation, providing a comprehensive assessment of each model's performance.

The study engages 13 evaluators to ensure reliable feedback. The models are anonymized to minimize potential biases, ensuring that evaluators are not influenced by preconceived notions. Furthermore, the order of the topics is randomized for each evaluator. This process, facilitated by the Qualtrics survey platform, ensures an unbiased evaluation by preventing order effects and maintaining consistency in the evaluation process.

The evaluators rate each topic based on the three criteria mentioned above. The results are then summarized by calculating the mean score for each criterion across all games, leading to a concise representation of each model's performance. This aggregate data allows to compare the models directly and make an informed recommendation on which method provides the most consistent, relevant, and interpretable topics.

By incorporating human evaluation, a qualitative layer is added to the analysis, enhancing the robustness of the study. This holistic approach ensures that the findings are practically meaningful for understanding and responding to video game reviews.

4   Results

To ensure consistency across conditions, the topic diversity scores are calculated for the top 20 topics generated by each model. The results are presented in Table 2. The BERTopic model performs the best, followed by LDA, and then Top2Vec. However, their average scores all exceed 0.8. Even the best-performing BERTopic and the least-performing Top2Vec differ by only 16.5 percent. This indicates that the topics generated by these three models are highly distinct overall and able to cover different areas of the text corpus.

Topic diversity Score	LDA	Top2Vec	BERTopic
PUBG	0.971	0.877	0.994
GTA	0.960	0.799	0.973
RUST	0.957	0.794	0.989
Average	0.963	0.823	0.985

Table 2: Topic diversity scores across games and models.

The results of the human evaluation are summarized in Table 3. This table presents the mean scores for each criterion across all models. Based on the human evaluations, BERTopic is the best-performing topic modeling model, while LDA performs the worst. In terms of coherence and interpretability, BERTopic achieved the highest scores, with 6.20 and 5.87 respectively. For coherence, BERTopic scored 19 percent higher than Top2Vec

(5.22) and 98 percent higher than LDA (3.13). For interpretability, BERTopic scored 16 percent higher than Top2Vec (5.05) and 83 percent higher than LDA (3.20). In terms of relevance scores, the differences among the three models are not as significant: the highest score, achieved by Top2Vec, is only slightly higher than the second highest, BERTopic, by 0.05 (5.65 vs. 5.6). LDA also scored above 4.5, only about 20 percent lower than the other two models. This indicates that regardless of the model used, the topics generated are still relevant to the documents.

Metric	LDA	Top2Vec	BERTopic
Coherence	3.13	5.22	6.20
Relevance	4.53	5.65	5.60
Interpretability	3.20	5.05	5.87

Table 3: Mean topic modeling scores based on human judgement.

5 Discussion

5.1 Performance Evaluation

The models show significant differences in their performance for most evaluated criteria. These differences can be attributed to various factors inherent to each modeling approach.

Considering the topic diversity score, when generating topics, LDA models the distribution of topics within documents and words within topics explicitly. This probabilistic approach tends to ensure that topics are distinct because it balances the likelihood of words belonging to different topics. Its generative process makes topics to be more separated in the word space, making it obtain a high topic diversity score. Top2Vec uses embeddings to group similar documents and then finds dense areas in the embedding space to define topics. While this approach is powerful for finding semantically similar topics, it might not enforce the same level of distinctiveness between topics as LDA does. Furthermore, Top2Vec automatically selects the number of topics, and for the dataset used, the number of topics formed by individual games is much larger than 20. After taking all the topics it generates into account, it can reach the same level of topic diversity as the other two models do.

To further evaluate the three models, a more in-depth qualitative analysis is necessary. This allows for a more thorough analysis of the performances. LDA receives the lowest scores in all three criteria in human evaluation, with coherence at 3.13, relevance at 4.53, and interpretability at 3.20. The lower scores for coherence and interpretability may be due to LDA's reliance on the bag-of-words approach, which ignores the order and context of words. This can result in topics that are less cohesive and harder to interpret. Some of the topics include a mix of unrelated terms, for example the following topic: "great", "suck", "like", "dont", "winner", "optimize", "boring", "cheat", "play", and "perfect". This makes it difficult for evaluators to discern a clear theme, leading to more fragmented topics, impacting both their coherence and interpretability, but also the perceived relevance.

Both Top2Vec and BERTopic perform significantly better than LDA, with Top2Vec achieving coherence at 5.22, relevance at 5.65, and interpretability at 5.05, while BERTopic achieved coherence at 6.20, relevance at 5.60, and interpretability at 5.87. These improvements can be attributed to their

use of advanced embeddings and clustering techniques, which capture semantic information more effectively. The results indicate that Top2Vec and BERTopic, which leverage advanced embeddings and clustering techniques, are better suited for generating pertinent and coherent topics from video game reviews in the qualitative metric perspective.

Incorporating both quantitative and qualitative evaluation, BERTopic demonstrates superior overall performance compared to the other methods. It produces more contextually rich topics that are distinctive, coherent, relevant and easier to interpret. Therefore, BERTopic is chosen to be the final model with most effective performance for understanding and addressing video game reviews, and its output undergoes further analysis in the following sections.

5.2 Analysis of the Final Model

The output of the final chosen BERTopic model includes associated keywords as well as representative documents of each topic, which provides the users with a concise summary as well as a comprehensive understanding of the main topics covered by the model. Users can grasp the common themes or recurring concepts presented in keywords as a guide to dive deeper into the specific area of interest (Samsir et al., 2023).

To illustrate the interpretation process with human cognition, the output of BERTopic for PlayerUnknown's Battlegrounds is taken as an example, as this is the game with the highest number of reviews. The result shown in Figure 2 provides the topic word scores for the highest scoring tokens in each of the top eight topics for the game. These keywords are further analyzed in conjunction with representative documents to provide context. For

instance, Topic 1 is represented by several Chinese keywords with top-ranking scores, which includes "Trash game", "Negative review" and "Poor gaming experience". Combined with the result of representative documents, which indicate that a large amount of cheaters significantly influence users' gaming experience by outperforming the legitimate players who should have performed well and cause the users to have extremely negative reviews, the topic suggests that users in the Chinese market hold strong negative emotion towards the cheaters in the game. This topic offers valuable insight for the developers of the game, highlighting the need to take action to prevent the interference of cheaters and improve the gaming experience and user satisfaction for legitimate players. Another implication of this topic is that the output of topic modeling may include words in different languages, which complicates the interpretation process as users need to possess professional knowledge not only in the gaming field, but also in different linguistic domains to better understand and interpret the latent insight from the output.

5.3 GPT-4o Augmentation

Aside from the human evaluation of the output, utilizing the OpenAI GPT-4o model enhances the interpretation by providing an overall score, a summary of the information, and an analysis of the overall sentiment of the text. An example is provided by the following description for Topic 1:

> *"The reviews for PLAYERUNKNOWN'S BATTLEGROUNDS on Steam are overwhelmingly negative. The most common complaints center around the rampant*

10

Figure 2: Distribution of top words in each topic identified by BERTopic.

presence of cheaters ('外挂太多'), which significantly impacts the gaming experience. Many users describe the game as '垃圾游戏' (trash game) and express frustration with the developers, referred to as '垃圾蓝洞' (trash Bluehole). The sentiment is further emphasized by phrases like '游戏体验极差' (extremely poor gaming experience) and '垃圾服务器' (trash servers). A recurring sarcastic comment, 'hahahahahahahahahahahahah ahahahahahahahahahahahahahahahahaha,' highlights the mockery and dissatisfaction among players. Overall, the sentiment score is a 2 out of 10."

The augmentation of the output of BERTopic using GPT-4o reveals three main benefits:

1. **Streamlined Pipeline**: The model eliminates the need for human intervention, significantly accelerating the process and ensuring timely and efficient outputs. This automation leads to more consistent and reliable results.

2. **Inference Capabilities**: With its extensive general knowledge, the model can understand complex and abstract patterns that require contextual awareness. For example, it can grasp user discontent in GTA reviews regarding Rockstar's changes to modding policies, highlighting its nuanced understanding of user sentiment. Additionally, it assigns a quantitative score to the overall sentiment of the text, which can be used for filtering reviews by positive and negative sentiments.

3. **Language Barrier**: The GPT-4o model's multilingual capabilities allow it to interpret reviews in various languages and provide accurate and comprehensive summaries in English. This is particularly useful for understanding Chinese reviews on PUBG, where the model effectively identifies and conveys

concerns about widespread hacking issues.

One drawback of the model is the tendency of large language models (LLMs) to hallucinate and generate stories or summaries that do not accurately reflect the actual content of the comments. However, this effect can be mitigated by manually checking a representative sample of the reviews and introducing specific prompts to limit hallucinations. Additionally, the frequency of these hallucinations decreases with the introduction of updated LLMs, which offer improved accuracy and reliability in interpreting and summarizing user reviews.

6 Limitations

The study has certain limitations. First, traditional topic modeling is often applied to automatically categorize news articles, research papers, or any large collection of documents, which utilize standard language. However, upon examining the original data, many reviews revealed ambiguity with informal expressions, such as "reeeeeeeee". To some extent, this introduces additional noise within the data. Second, due to the current lack of universally accepted evaluation standards for topic modeling results, the metrics are largely based on human evaluation. This reliance on subjective assessment may introduce a degree of bias into the evaluation of model performance. Without standardized benchmarks it is challenging to objectively compare and validate the effectiveness of different models. Finally, the human evaluation is conducted with a relatively small sample size of 13 evaluators. This limited number of evaluators increases the potential for bias, particularly given that the evaluators share similar backgrounds and perspectives. Consequently, the generalizability of the findings may be restricted, and the results may not fully represent the diversity of potential viewpoints.

7 Outlook

Looking ahead, the findings highlight several promising avenues for future work in video game review analysis. Expanding data sources to include platforms like PlayStation Network and Xbox Live may provide a more comprehensive understanding of relevant topics. Longitudinal studies could track how these topics evolve over time, offering valuable insights into the changing preferences and concerns of players. Additionally, combining quantitative metrics with qualitative analysis would uncover deeper, more nuanced insights, enhancing the understanding of player experiences and the effectiveness of the models. These steps will be crucial in advancing the field and providing more robust feedback for game developers and marketers.

8 Conclusion

Given that video game reviews possess many valuable insights for game developers, this study examines three topic modeling methods - LDA, Top2Vec and BERTopic, to determine which method performs most effectively to extract diverse, relevant, coherent and interpretable topics from these reviews. The output of these three models are evaluated using a quantitative metric called topic diversity, and complemented by a qualitative human evaluation through anonymous questionnaires. The assessment suggests that BERTopic is the most suitable model to conduct topic modeling on video

12

game reviews. The output of BERTopic is then interpreted and analyzed by both human evaluators and GPT-4o, illustrating the advancement of integrating GPT-4o into the topic modeling process. This integration reveals valuable insights for developers. Despite limitations including noise from non-informative words and lack of unified evaluation standard, applying BERTopic for topic modeling on video game reviews remains promising for the future with several potential directions, such as expanding the dataset scope, conducting longitudinal studies or combining quantitative evaluation with qualitative analysis. By revealing more valuable and nuanced insights from the output of topic modeling on video game reviews, this technique will facilitate game developers to enhance business decisions and advance the field in multiple aspects.

References

Angelov, D. (2020). Top2Vec: Distributed representations of topics (arXiv:2008.09470). arXiv. https://doi.org/10.48550/arXiv.2008.09470

Blei, D. M., Ng, A. Y., Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993-1022.

Blei, D. M., Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, 1(1), 17-35.

Busurkina, I., Karpenko, V., Tulubenskaya, E., Bulygin, D. (2020). Game experience evaluation. A study of game reviews on the Steam platform. In D. A. Alexandrov, A. V. Boukhanovsky, A. V. Chugunov, Y. Kabanov, O. Koltsova, I. Musabirov (Eds.), *Digital Transformation and Global Society*. DTGS 2020. Communications in Computer and Information Science (Vol. 1242). Springer, Cham. https://doi.org/10.1007/978-3-030-65218-0_9

Chen, Z., Wang, Y., Zhang, X., Li, Y. (2023). Leveraging state-of-the-art topic modeling for news impact analysis on financial markets: A comparative study. *Electronics*, 12(12), 2605. https://doi.org/10.3390/electronics12122605

Egger, R., Yu, J. (2022). A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. *Frontiers in Sociology*, 7, 886498. https://doi.org/10.3389/fsoc.2022.886498

Gan, L., et al. (2024). Experimental comparison of three topic modeling methods with LDA, Top2Vec and BERTopic. In H. Lu J. Cai (Eds.), *Artificial Intelligence and Robotics*. ISAIR 2023. Communications in Computer and Information Science (Vol. 1998). Springer, Singapore. https://doi.org/10.1007/978-981-99-9109-9_37

Gillings, M., Hardie, A. (2023). The interpretation of topic models for scholarly analysis: An evaluation and critique of current practice. *Digital Scholarship in the Humanities*, 38(2), 530-543.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv. https://doi.org/10.48550/arXiv.2203.05794

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications*, 78, 15169-15211.

Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Rivista Di Linguistica*, 20(1), 1-31.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., ... Adam, S. (2021). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. In *Computational Methods for Communication Science* (pp. 13-38). Routledge.

OpenAI. (n.d.). Strategy: Test changes systematically. OpenAI. Retrieved May 23, 2024, from `https://platform.openai.com/docs/guides/prompt-engineering/strategy-test-changes-systematically`

Osum. (2024). Crushing the competition: Steams dominant market share. Retrieved May 19, 2024, from `https://blog.osum.com/steam-market-share/`

Samsir, S., Saragih, R. S., Subagio, S., Aditiya, R., Watrianthos, R. (2023). BERTopic modeling of natural language processing abstracts: Thematic structure and trajectory. *Jurnal Media Informatika Budidarma*, 7(3), 1514-1520.

Tong, X. (2021). Positioning game review as a crucial element of game user feedback in the ongoing development of independent video games. *Computers in Human Behavior Reports*, 3, 100077. `https://doi.org/10.1016/j.chbr.2021.100077`

Annexes

GPT-4o Interpretation of BERTopic Results - PUBG

Topic Number	ChatGPT Response	Sentiment Score
1	The reviews for PLAYERUNKNOWN'S BATTLEGROUNDS (PUBG) on Steam are overwhelmingly positive, with many players praising it as the best battle royale game in the genre. Common sentiments highlight the game's realistic style, variety of weapons, and engaging gameplay. Phrases like 'best battle royale game,' 'This is BATTLE ROYALE,' and 'battle royale!' are frequently mentioned, underscoring the game's strong reception among fans of the genre.	9.0
2	The reviews for PLAYERUNKNOWN'S BATTLEGROUNDS on Steam are overwhelmingly negative. The most common complaints center around the rampant presence of cheaters ('外挂太多'), which significantly impacts the gaming experience. Many users describe the game as '垃圾游戏' (trash game) and express frustration with the developers, referred to as '垃圾蓝洞' (trash Bluehole). The sentiment is further emphasized by phrases like '游戏体验极差' (extremely poor gaming experience) and '垃圾服务器' (trash servers). A recurring sarcastic comment, 'hahahahaha-hahahahahahahahahahahahahahahahahahahahahahahaha,' highlights the mockery and dissatisfaction among players. Overall, the sentiment score is a 2 out of 10.	2.0
3	The reviews for PLAYERUNKNOWN'S BATTLEGROUNDS on Steam frequently mention the pervasive issue of hackers in the game. Common words in the reviews include terms like 'hackers', 'hacker', 'hacking', and 'hacks', indicating that cheating is a significant concern among players. Phrases such as 'Fun when no hackers' and 'so many hackers in the game...' suggest that while the game itself is enjoyable, the presence of hackers severely detracts from the experience. Overall, the sentiment is largely negative due to the frustration caused by the hacking problem.	3.0
	Continued on next page	

Topic Number	ChatGPT Response	Sentiment Score
4	The reviews for PLAYERUNKNOWN'S BATTLEGROUNDS (PUBG) highlight a generally positive sentiment towards the game, particularly emphasizing the enjoyment of playing with friends. Commonly mentioned aspects include the fun and engaging gameplay experience, with some users expressing their love for the game despite occasional issues. The game is frequently described as a good shooter with a satisfying campaign, providing an enjoyable time for players. Overall, the sentiment is favorable, with most users appreciating the game's quality and the social aspect of playing with friends.	8.0
5	The reviews for PLAYERUNKNOWN'S BATTLEGROUNDS on Steam are generally positive, with a common theme revolving around its early access status. Many users acknowledge that the game, while still in early access, has met or exceeded their expectations. Common sentiments include appreciation for the game's current state despite its early access tag and a recognition that bugs and issues are to be expected at this stage. Some users suggest waiting for the full release if one prefers a more polished experience. Overall, the sentiment is that the game is promising and enjoyable, even in its early access phase.	7.0

GPT-4o Interpretation of BERTopic Results - Grand Theft Auto V

Topic Number	ChatGPT Response	Sentiment Score
1	The reviews for Grand Theft Auto V on Steam are overwhelmingly negative, primarily due to the actions of Rockstar and Take-Two Interactive towards the modding community. Users express significant frustration over the company's hostile stance towards mods, particularly highlighting the cease and desist orders issued to OpenIV, a popular modding tool. This move is seen as anti-consumer and detrimental to the game's longevity and appeal on PC, where modding is a significant draw for many players. Additionally, the in-game economy of GTA Online is criticized for being a grind fest designed to push players towards microtransactions, further souring the overall experience. The sentiment is that the developers prioritize monetary gain over player satisfaction and community support. Overall, the sentiment score is 2 out of 10, indicating highly negative reviews.	2.0
2	The reviews for Grand Theft Auto V on Steam indicate a strong sentiment towards modding as a primary motivation for purchase. Many users explicitly mention that they bought the game specifically for the mods or modding capabilities. Common words in the reviews include 'mods,' 'modding,' and 'bought,' highlighting the importance of these features to the players. There is also a sense of disappointment expressed by some users, indicated by words like 'removed' and 'anymore,' suggesting that some modding capabilities may have been restricted or removed, leading to dissatisfaction.	6.0

Topic Number	ChatGPT Response	Sentiment Score
3	The sentiment in the reviews regarding the cease and desist order for OpenIV is overwhelmingly negative. Users are highly dissatisfied with the removal of OpenIV, a popular modding tool, expressing that without it, the game loses significant value. The reviews are dominated by repetitive and emphatic calls for the return of OpenIV, with many users stating that no mods mean no game for them. The community sentiment is one of frustration and disappointment towards the decision made on behalf of Take-Two Interactive.	2.0
4	The reviews for Grand Theft Auto V on Steam reveal a significant dissatisfaction among players regarding Rockstar and Take-Two's stance on modding. Many users express frustration and disappointment over the companies' actions to shut down mods, which they feel undermines the modding community that adds value to the game. Common sentiments include a lack of support for future purchases from these companies and a strong recommendation against the game due to these anti-modding policies. The overall sentiment is highly negative, reflecting a deep-seated disapproval of Rockstar and Take-Two's approach to modding.	2.0

Topic Number	ChatGPT Response	Sentiment Score
5	The reviews for Grand Theft Auto V on Steam are heavily influenced by the controversy surrounding Take Two Interactive's cease and desist order against OpenIV, a popular modding tool. The sentiment is predominantly negative, with many users expressing their frustration and disappointment. They highlight that the single-player mode is enjoyable and that mods significantly enhance the gaming experience. However, the forced shutdown of OpenIV is seen as a move to push players towards the online mode, which is criticized for its microtransactions and hacker issues. Reviewers feel that this action undermines the modding community and diminishes the game's value. Some updates mention that Take Two has retracted the cease and desist, but skepticism remains about the company's future actions. Overall, the reviews reflect a strong disapproval of Take Two's handling of the situation, despite the game's inherent qualities.	3.0

GPT-4o Interpretation of BERTopic Results - Rust

Topic Number	ChatGPT Response	Sentiment Score
1	The reviews for Rust paint a picture of a complex, evolving game that has both its strengths and weaknesses. On the positive side, players appreciate the game's constant evolution and the developers' responsiveness to community feedback. The ability to host and mod servers adds a layer of customization that many find appealing. The crafting system is praised for its depth and flexibility, allowing for a wide range of play styles and strategies. Additionally, the game's graphics and immersive environment receive high marks, contributing to a compelling survival experience. However, the reviews also highlight several significant drawbacks. Performance issues, such as lag and long loading times, are common complaints, even among those with high-end gaming setups. The game's punishing nature, where players can lose weeks of progress in a single raid, is a double-edged sword—adding to the challenge for some, but causing frustration for others. The community is often described as toxic, with new players frequently targeted and harassed. This, combined with the presence of hackers, can make the game unwelcoming for newcomers. Overall, Rust is seen as a game with great potential and many unique features, but it also has areas that need improvement to enhance the player experience.	7.0
2	The reviews for Rust on Steam are overwhelmingly positive, with users frequently describing it as the best survival game they have ever played. Commonly used words in the reviews include 'survival,' 'best,' 'games,' 'ever,' 'multiplayer,' 'played,' 'one,' 'game,' 'great,' and 'amazing.' The sentiment reflects a strong appreciation for the game's multiplayer experience and overall gameplay quality. Phrases like 'Best survival game!' and 'The best survival game' are recurrent, highlighting the high regard players have for Rust.	9.0
		Continued on next page

21

Topic Number	ChatGPT Response	Sentiment Score
3	The reviews for Rust on Steam indicate a generally positive sentiment, with many users acknowledging that the game is still in its alpha stage. Common words in the reviews include 'alpha', 'still', 'stage', and 'wait', suggesting that players are aware of the game's ongoing development and potential. Despite the presence of bugs and the game not yet being in beta, users express enjoyment and optimism about its future. Phrases like 'great game and its still in alpha!' and 'Game is still in Alpha but I like it.' are indicative of this sentiment.	7.0
4	The reviews for Rust on Steam are generally positive, particularly emphasizing the game's PvP and survival aspects. Players appreciate the blend of PvP and PvE elements, highlighting the excitement and challenges these bring. The game's building mechanics and server variety are also frequently mentioned, with many players enjoying the different server experiences available. Overall, the sentiment is that Rust provides a thrilling and engaging survival experience, with a strong focus on player interactions and community. The most common reviews reflect a high level of satisfaction with the game's core elements.	8.0
5	The reviews for Rust on Steam are overwhelmingly positive, especially highlighting the enjoyment of playing with friends. Common words in the reviews include 'friends', 'play', 'fun', 'great', and 'good', indicating a strong social and enjoyable aspect to the game. Phrases like 'Great Game Fun To play with friends!' and 'Fun game to play with friends!' are frequently mentioned, emphasizing the game's appeal in a multiplayer setting. Overall, the sentiment is very positive.	9.0

LDA Results

Topic Name	Count of Documents	Representation	Probability
15	90232	['fun', 'friend', 'lot', 'play', 'great', 'bug', 'time', 'buggy', 'need', 'recommend']	['0.1437', '0.0365', '0.0354', '0.0336', '0.0285', '0.0283', '0.0233', '0.0212', '0.0190', '0.0148']
0	89320	['early', 'access', 'review', 'issue', 've', 'like', 'release', 'people', 'time', 'great']	['0.0365', '0.0301', '0.0233', '0.0176', '0.0131', '0.0118', '0.0097', '0.0094', '0.0083', '0.0075']
9	88044	['fix', 'lag', 'hacker', 'bug', 'server', 'issue', 'update', 'problem', 'fixed', 'awesome']	['0.0909', '0.0905', '0.0735', '0.0466', '0.0385', '0.0284', '0.0280', '0.0256', '0.0195', '0.0166']
2	87716	['play', 'fun', 'friend', 'playing', 'time', 'people', 'squad', 'like', 'solo', 'win']	['0.0529', '0.0358', '0.0295', '0.0217', '0.0211', '0.0196', '0.0190', '0.0173', '0.0164', '0.0148']
6	87168	['play', 'got', 'like', 'hour', 'im', 'pubg', 'playing', 'know', 'banned', 'played']	['0.0399', '0.0290', '0.0248', '0.0191', '0.0189', '0.0183', '0.0160', '0.0158', '0.0154', '0.0139']
11	86863	['gun', 'shot', 'like', 'yes', 'weapon', 'player', 'time', 'map', 'people', 'kill']	['0.0141', '0.0130', '0.0128', '0.0120', '0.0117', '0.0115', '0.0108', '0.0083', '0.0083', '0.0078']
19	86484	['money', 'crate', 'devs', 'developer', 'buy', 'key', 'pay', 'instead', 'skin', 'like']	['0.0315', '0.0217', '0.0142', '0.0134', '0.0120', '0.0104', '0.0104', '0.0102', '0.0098', '0.0094']
7	84833	['pubg', 'like', 'battle', 'better', 'royale', 'fortnite', 'feel', 'gameplay', 'player', 'experience']	['0.0501', '0.0216', '0.0194', '0.0157', '0.0140', '0.0130', '0.0125', '0.0123', '0.0118', '0.0101']
18	84413	['good', 'great', 'pretty', 'simulator', 'need', 'job', 'graphic', 'tho', 'yeet', 'like']	['0.5241', '0.1036', '0.0438', '0.0209', '0.0139', '0.0107', '0.0104', '0.0093', '0.0090', '0.0088']
3	84171	['buy', 'dont', 'bad', 'money', 'worth', 'garbage', 'waste', 'bug', 'hour', 'trash']	['0.1142', '0.0823', '0.0643', '0.0605', '0.0337', '0.0261', '0.0253', '0.0211', '0.0211', '0.0177']

Top2Vec Results

Topic	Count	Representation	Scores
0	11508	['kotk' 'fortnite' 'pubg' 'desync' 'battleye' 'netcode' 'dayz' 'fpp' 'hitreg' 'arcadey' 'anticheat' 'playerbase' 'gamemodes' 'aimbots' 'aimbotting' 'rubberbanding' 'lag' 'battleye' 'gameplay' 'game breaking' 'battlegrounds' 'tickrate' 'firefights' 'unplayable' 'replayability' 'csgo' 'laggy' 'wallhack' 'teamkilling' 'severs' 'arma' 'battlefield' 'multiplayer' 'gunplay' 'daybreak' 'akm' 'gaem' 'playable' 'aimbot' 'mmo' 'overwatch' 'devs' 'killcam' 'headshotted' 'deathmatch' 'teamkill' 'pvp' 'game' 'vss' 'singleplayer']	[0.49168184 0.48014042 0.46266407 0.4305604 0.41654396 0.41653085 0.4071164 0.39269513 0.39227673 0.39210242 0.386366 0.3777981 0.3746913 0.37269795 0.37235016 0.3680516 0.35819936 0.35332203 0.35300237 0.3488353 0.3427087 0.34186256 0.34073558 0.34036624 0.3365125 0.331802 0.32885748 0.32759607 0.32689765 0.32317904 0.3216731 0.3178348 0.31515887 0.3131408 0.31283408 0.3126216 0.30927825 0.30919307 0.30883285 0.30874044 0.30823 0.305264 0.3043666 0.30393994 0.3017086 0.30125028 0.3010005 0.30079308 0.2932008 0.29308003]
1	4560	['framerates' 'lagg' 'optimised' 'framerate' 'unoptimized' 'laggy' 'fps' 'gpu' 'pubg' 'optimized' 'lagging' 'lags' 'reshade' 'optimizations' 'vram' 'radeon' 'gaming' 'playable' 'stutters' 'processor' 'optimisation' 'unplayable' 'geforce' 'arma' 'csgo' 'overclocked' 'battlegrounds' 'laggs' 'fortnite' 'dayz' 'lag' 'cpu' 'rubberbanding' 'specs' 'kotk' 'tickrate' 'netcode' 'ghz' 'desync' 'pc' 'graphics' 'choppy' 'battleye' 'gtx' 'sli' 'rig' 'optimization' 'arcadey' 'gameplay' 'battlefield']	[0.545182 0.51620364 0.5155623 0.51004386 0.49722642 0.4735311 0.46534294 0.4301517 0.42901444 0.40795827 0.3939476 0.39120483 0.38854545 0.3867249 0.3834663 0.37761366 0.3772636 0.3731075 0.3663208 0.36372817 0.36313945 0.36242437 0.36204237 0.36046186 0.36029872 0.36012405 0.35889083 0.3580446 0.3558361 0.35372418 0.35280353 0.35199583 0.3504722 0.34927034 0.34690094 0.34582928 0.3440609 0.34285724 0.34226462 0.34113127 0.33797947 0.33638683 0.33630878 0.33559102 0.33476895 0.3332423 0.32722807 0.3210858 0.31941748 0.31699973]
2	3232	['replayability' 'gameplay' 'game' 'fortnite' 'playable' 'playerbase' 'arcadey' 'unplayable' 'kotk' 'gamebreaking' 'games' 'pubg' 'juego' 'netcode' 'mmo' 'gaem' 'gamemodes' 'playability' 'playtime' 'lagg' 'gaming' 'ingame' 'multiplayer' 'gamei' 'laggy' 'goty' 'rubberbanding' 'dayz' 'playstyles' 'desync' 'played' 'sux' 'simulator' 'dota' 'patching' 'sucks' 'battlegrounds' 'patched' 'overwatch' 'csgo' 'play' 'severs' 'playstyle' 'perma' 'addicting' 'dev' 'janky' 'matchmaking' 'enjoyable' 'battleye']	[0.38278255 0.3783844 0.35409987 0.3487621 0.3364457 0.33632544 0.32313284 0.32202345 0.3201272 0.3108546 0.30795613 0.30652177 0.30151975 0.3003197 0.29786223 0.2955274 0.29505193 0.27790147 0.27556112 0.2753896 0.27392048 0.2738885 0.27199566 0.26846805 0.26703858 0.26514143 0.26498428 0.26128644 0.26042825 0.25769407 0.2566529 0.24963276 0.24929422 0.2491929 0.24879271 0.24764374 0.24728236 0.24661173 0.24486893 0.24467981 0.24124102 0.23831117 0.2376617 0.23717284 0.23684858 0.23560722 0.23511684 0.23474546

24

			0.23362109 0.23306778]
3	3055	['hackers' 'hacker' 'aimbots' 'aimbot' 'battleye' 'hacks' 'hacking' 'anticheat' 'aimbotting' 'wallhack' 'hack' 'battleeye' 'hacked' 'desync' 'hitreg' 'killcam' 'netcode' 'hax' 'wallhacks' 'cheaters' 'aimbotters' 'kotk' 'cheats' 'exploits' 'teamkilling' 'leaderboards' 'headshotted' 'patched' 'cheater' 'dayz' 'fpp' 'rubberbanding' 'teamkill' 'fortnite' 'glitching' 'severs' 'unplayable' 'pubg' 'lagg' 'deathcam' 'firefights' 'arcadey' 'vac' 'hitboxes' 'laggy' 'csgo' 'dmg' 'gamebreaking' 'lobbies' 'pvp']	[0.62787056 0.608354 0.5737151 0.541691 0.52523625 0.52092564 0.5199369 0.512089 0.5076853 0.48341262 0.48005304 0.44003126 0.4257657 0.40371722 0.40225205 0.3926791 0.3887965 0.38713995 0.3833663 0.38034168 0.36925304 0.36768258 0.3569006 0.35151386 0.345284 0.3442431 0.3405751 0.33905548 0.33819997 0.33642632 0.33315635 0.33187455 0.33047378 0.32649747 0.32447934 0.32126883 0.32082796 0.31647986 0.3154951 0.31526697 0.31425172 0.3125828 0.31250972 0.31234717 0.307517 0.30723488 0.30700547 0.29999596 0.29922408 0.29896232]
4	2822	['game' 'brilliant' 'superb' 'nice' 'good' 'fantastic' 'awsome' 'goood' 'enjoyable' 'awesome' 'great' 'excellent' 'fun' 'dope' 'amazing' 'cool' 'hilarious' 'decent' 'beautiful' 'nicely' 'classic' 'gud' 'wonderful' 'super' 'epic' 'gg' 'lovely' 'games' 'funny' 'neat' 'noice' 'phenomenal' 'entertaining' 'impressive' 'perfect' 'sweet' 'pretty' 'perfectly' 'happy' 'omg' 'enjoying' 'lol' 'wow' 'replayability' 'fine' 'immense' 'enjoyed' 'enjoy' 'legit' 'cheers']	[0.5743289 0.5332968 0.5238197 0.52243745 0.522409 0.51258683 0.50410056 0.5006144 0.5004802 0.49910486 0.49906075 0.4947645 0.49395737 0.48707497 0.4837019 0.47934252 0.477941 0.4760693 0.47604334 0.4716461 0.4706331 0.4700805 0.46969086 0.4688166 0.46477073 0.45583093 0.4534405 0.4523803 0.4514814 0.4359731 0.43465388 0.43394944 0.43303835 0.4277826 0.4271186 0.42383164 0.41779023 0.41462544 0.41407174 0.40917838 0.40724707 0.4049526 0.40468875 0.404248 0.40394413 0.4021597 0.40208024 0.39858148 0.39762104 0.39745322]
5	2616	['gamebreaking' 'unplayable' 'kotk' 'fixes' 'devs' 'buggy' 'desync' 'fortnite' 'patching' 'patched' 'gaem' 'glitching' 'rubberbanding' 'bugged' 'playable' 'battleye' 'netcode' 'bugs' 'pubg' 'dayz' 'glitchy' 'playerbase' 'glitches' 'ubisoft' 'arcadey' 'fix' 'lagg' 'hitreg' 'dev' 'bug' 'tickrate' 'game' 'anticheat' 'refunding' 'gameplay' 'replayability' 'battleeye' 'severs' 'janky' 'gamemodes' 'developers' 'fpp' 'nerf' 'cashgrab' 'daybreak' 'glitched' 'developer' 'laggy' 'refunded' 'simulator']	[0.44086945 0.4079635 0.40578192 0.3756258 0.37396035 0.365472 0.36374146 0.36267453 0.33855072 0.33849326 0.33673155 0.33629137 0.32929912 0.32922143 0.32920873 0.3284005 0.32740283 0.32709044 0.32689816 0.3226059 0.32074627 0.31514102 0.31342453 0.3130166 0.3093692 0.30826887 0.30691022 0.30553454 0.29684877 0.29648814 0.28988487 0.2876064 0.28678638 0.28349984 0.28242114 0.27666813 0.27425396 0.274136 0.27397004 0.27034366 0.2703151 0.27013353 0.26731977 0.26457328 0.25942445 0.2591933 0.2567072 0.25644132

BERTopic Results

Topic	Count	Name	Representation	Representative Docs
0	2086	0_royale_battle_best_genre	['royale', 'battle', 'best', 'genre', 'royales', 'style', 'realistic', 'games', 'type', 'weapons']	['best battle royale game', 'This is BATTLE ROYALE.', 'battle royale!']
1	1519	1_中国第一_垃圾游戏_差评_垃圾	['中国第一', '垃圾游戏', '差评', '垃圾', '游戏体验极差', '外挂太多', '垃圾服务器', '垃圾蓝洞', 'hahahahahahahahahahahahahahahahahahahahahahahahahahahahahaha', '外挂太多了']	['外挂太多 影响游戏体验...差评', '中国第一！！！', '垃圾游戏 不打挂 垃圾蓝洞']
2	1493	2_hackers_hacker_hacking_hacks	['hackers', 'hacker', 'hacking', 'hacks', 'hack', 'many', 'ruined', 'ruining', 'everywhere', 'full']	['Fun when no hackers', 'so many hackers in the game...', 'Good game many hackers']
3	1126	3_pubg_playing_has_friends	['pubg', 'playing', 'has', 'friends', 'campaign', 'games', 'is', 'time', 'shooter', 'experience']	['I love playing PUBG but then....', 'Pubg is a very good game.', 'Pubg has a fun play it.']
4	1009	4_access_early_an_expected	['access', 'early', 'an', 'expected', 'bugs', 'expect', 'release', 'still', 'wait', 'released']	['Good game for early access.', '"EARLY ACCESS"', 'Early access 1.0']
5	1002	5_good_game_belaive_jajajaja	['good', 'game', 'belaive', 'jajajaja', 'yar', 'maan', 'photos', 'gamevery', 'yeh', 'tis']	['good game', 'Good Game', 'Good Game']