Classifying Product Reviews

Technical Project Overview

By: Nicole Quevillon

March 2025

Project Goals

Terraria is one of the best-selling video games of all time with over 58 million copies sold since 2011. While the reviews are overwhelmingly positive, the game's features have become increasingly complex as the developer has continued to release major updates.

Business Goal: Develop an understanding of the spread and frequency of topics among reviews of the game to help the team better understand feature importance.

Technical Goal: Develop an unsupervised classification system for Terraria's reviews.

Data Source

The data was obtained from Kaggle

[Source]

Date of Dataset: 2021
Size: 8kb
Format: csv

The reviews are all from **Steam**, which means this sample represents the **PC version** of the game. Reviews for the same game on other platforms such as Xbox or Switch are not included.

The Data

Pre-Processing:

- Only English reviews were kept. The first filter was done using the 'language' ID in the data. There were some reviews incorrectly labeled as English, so a secondary language detection and filtering was applied to all reviews
- Reviews with a sentimentality score of zero were found to be nonsense most of the time, so these were removed
- Some reviews were very short and lacked meaningful content, so only those that were at least 20 tokens long were kept for clustering.

app_name	review_id	language	review
Terraria	85105575	english	Very good and very addicting game! I recommend...
Terraria	85105565	english	i like everything about this game but i'm not ...
Terraria	85105421	english	Its Gucci to all players seaking a god damn ch...

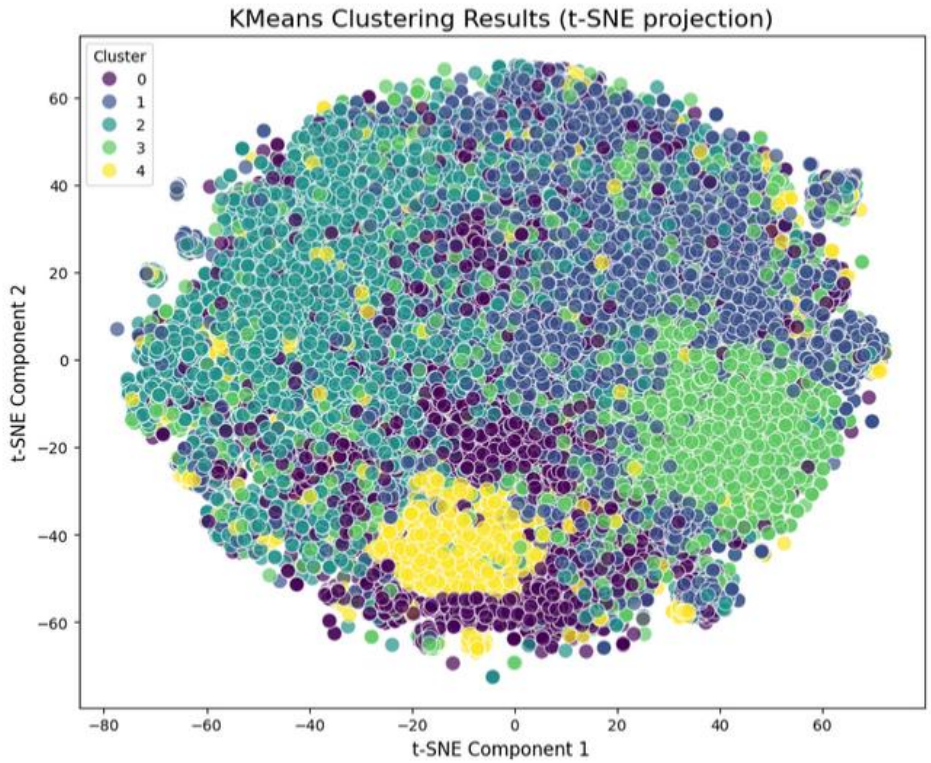
Final Review Count: 45,075 reviews for clustering

Simple Clustering

The reviews were tokenized, lemmatized, and stop words removed. The following clustering models were tested on this data:

- BOW KMeans – silhouette score of 0.02
- BOW Mean Shift
- Word2Vec KMeans – silhouette score of 0.12
- TF-IDF KMeans – silhouette score of 0.02
- DBSCAN – silhouette score of -0.24

These clustering methods performed extremely poorly and did not pick up on coherent clusters.



Topic Modeling

Since we were unable to identify meaningful clusters with simple unsupervised models, we tried **topic modeling with BERTopic.** This was applied to the **raw review text** rather than the tokenized reviews.

After testing various parameters, the following were decided on:

- **ClassTfidfTransformer** was used to help eliminate stop words. This approach worked well because it is not based on a predefined stop word list, so it was able to adjust to the niche topic matter.
- **min_topic_size = 10** was found to give the best results. This meant that our topics could have as few as 10 reviews in them.
- **ngram** was not specified. The topics were more intuitive without including bigrams or trigrams

Topics

The topics were reduced to the top 50. Intuitively the groupings look to make sense. The most significant topics were:

Topic Name	Top Words	Review Count
Minecraft Comparison	terraria, minecraft, as, has, your, are, will…	17,714
2D Sandbox Like Minecraft	minecraft, 2d, sandbox, like, than, more, bet..	6,244
Modding	mods, mod, tmodloader, vanilla, modding, hour…	2,639
Journeys End Update	journey, end, journeys, pc, update, steam, xb..	944
Worth the Price	10, worth, price, would, buy, sale, rate, 99,..	898
Expert Mode	mode, expert, lord, moon, master, moonlord, n..	619
Killed or Died	killed, die, died, slime, he, dies, again, qu…	420
Fix Corrupted Files	fix, cloud, files, corrupted, save, deleted, …	349
Game Updates	update, came, wait, new, now, years, was, ver…	335
Building and Digging	dig, digging, build, house, hell, your, kill,…	186

Issues with the Topic Modeling Results

There are some major issues that remain with our clustering:

- **12k reviews are classified as outliers.** The model was reduced to the top 50 clusters after reviewing the original model hierarchy, which inflated this number significantly. There were a lot of nonsensical clusters before reduction.
- Related to the first point, **the differences in reviews is too fine to be picked up by topic modeling.** Most reviews are extremely generic, leading the topic modeling to latch onto one or two keywords to classify an entire review.
- Our top **topic includes stop words** like has, like, you etc
- **Our clusters are generally very small.** Only the top 9 topics contain at least 1% of the reviews in the dataset.

Conclusion: We have an ok start with some very intuitive and explainable topics, but there is an issue with the topic modeling being unable to extract meaningful results from a large portion of the reviews.

Next Steps

There are a few next steps for this project:

1. We can continue to tweak the **stop-words parameters** for this model. We have used TF-IDF to try and account for these, but there are still some stop words coming through in our topics. That said, stop word removal may impact the interpretability by BERTopic. We would also likely need to create a custom stop words list

2. Given the very generic and jargon-heavy nature of these reviews, **use of an LLM** for clustering would be a good approach to test. It may perform better than our topic modeling.