

LEVERAGING SENTIMENT ANALYSIS OF STEAM REVIEWS FOR GROWTH IN GAME SUCCESS METRICS

**Presented By : Vedant Tomer
Supervisor : Vivek Kshirsagar**

Dublin Business School | 2024

ABSTRACT

The research explores sentiment analysis and machine learning to predict game recommendations and review helpfulness on Steam. It uses lexicon-based (VADER) and deep learning (DistilBERT) methods to extract sentiment scores, which, combined with review metadata, train Linear SVC and Multinomial NB models. Results show that incorporating sentiment analysis improves prediction accuracy, especially in handling class imbalance, highlighting its value in the gaming industry.

OVERVIEW

- Introduction
- Problem
- Literary Review
- Theoretical Framework
- Objectives
- Methodology
- Implementation
- Results and Conclusion
- Thank You

INTRODUCTION

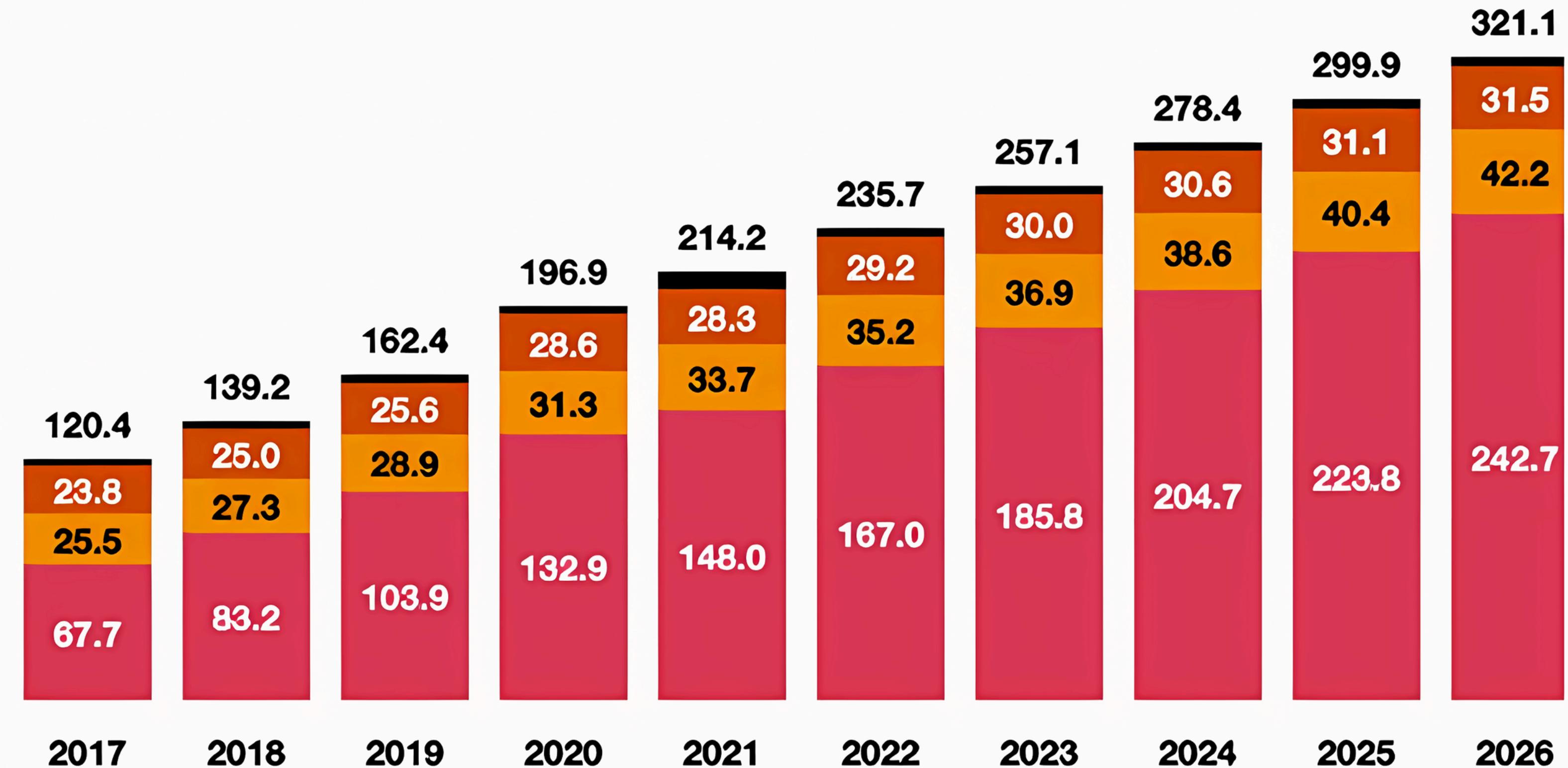
- The video game industry is a dynamic and rapidly growing market, where the success of a game hinges on various factors, including gameplay, graphics, storyline, and player engagement.
- Traditional methods for predicting game success have often relied on sales data and player statistics.
- The rise of digital platforms and social media has provided a new avenue for understanding player sentiment and its impact on game success metrics.



STEAM®

Total global video games revenue, by segment (US\$bn)

Social/casual gaming PC games Console games
Integrated video games advertising



PROBLEM

The abundance of user-generated content on digital platforms like Steam offers a valuable source of insights into player perceptions and experiences.

First Problem

The problem addressed in this research is the need for more effective methods to predict game success in the video game industry.

Second Problem

The traditional methods, which rely heavily on sales data and player statistics, may not fully capture the nuances of player sentiment and its impact on a game's performance.

1

The literature review highlights the growing interest in sentiment analysis and its applications in the gaming industry.

2

Studies have explored the use of machine learning models like BERT, BiLSTM, and CRF to decipher sentiments in game reviews.

3

Research has also addressed challenges such as spam detection, cross-platform analysis, and the extraction of actionable insights beyond sentiment polarity.

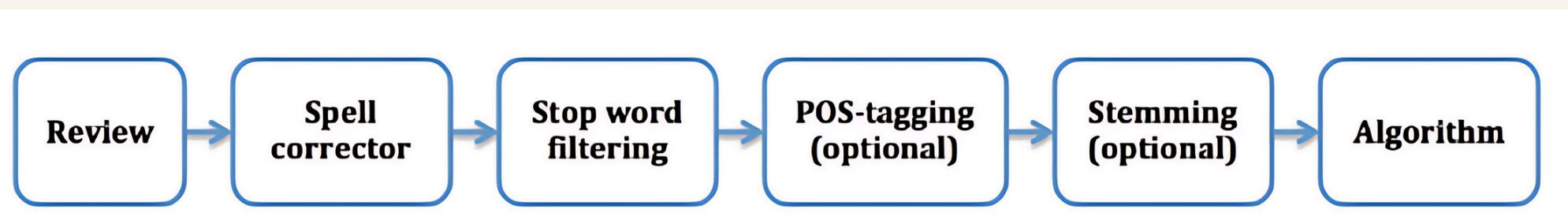
THEORETICAL FRAMEWORK

- The theoretical foundation of this research lies in the field of sentiment analysis, which involves extracting and evaluating subjective information from text.
- The study utilizes both lexicon-based (VADER) and deep learning (DistilBERT) methods to analyze sentiment in Steam reviews.

OBJECTIVES

- The primary objective is to investigate the potential of sentiment analysis and machine learning in predicting game recommendations and review helpfulness on Steam.
- The study evaluates the performance of machine learning models in predicting outcomes by integrating sentiment analysis with review attributes.

METHODOLOGY



The models' accuracy is evaluated using relevant metrics, and the tool is implemented using Python and Streamlit.

The research methodology involves collecting and analyzing user reviews from Steam, categorizing them into positive, neutral, and negative sentiments using sentiment analysis algorithms.

Sentiment scores are calculated to determine the overall sentiment towards each game, then these scores are used to train machine learning models.

IMPLEMENTATION

● Phase 1

Data Acquisition and Cleaning:
Collected and preprocessed Steam game reviews using a custom web scraper, removing noise and inconsistencies.

● Phase 2

Sentiment Quantification: Applied TextBlob, VADER, and a fine-tuned DistilBERT model for multi-model sentiment scoring.

● Phase 3

Feature Transformation: Converted text data into numerical form using TF-IDF with positional weighting for better sentiment capture.

● Phase 4

Predictive Modeling: Trained and evaluated Linear SVC and Multinomial NB models with 10-fold cross-validation and SMOTE for balanced performance.

RESULTS AND CONCLUSION

Steam Review Data

Acquisition

The initial phase of this research involved the acquisition of a substantial corpus of Steam game reviews. This was accomplished through a meticulously designed web scraping process, leveraging the capabilities of the steamreviews library in Python.

578080	PUBG: BATTLEGROUNDS	very good
578080	PUBG: BATTLEGROUNDS	At 69h Yes i will recommend this game for other players, i enjoy this game more in first person ngl, Good game
578080	PUBG: BATTLEGROUNDS	I LOVE THIS GAME I LOVE CHINA THANK YOU CHINESE COMPANY YAHOOOOOOO !!!!!!!!!!!!!!!
578080	PUBG: BATTLEGROUNDS	Cheaters, Dated gameplay...
578080	PUBG: BATTLEGROUNDS	This game is actually pretty bad. I don't know what fps to play but this isn't it.

The Streamlit Sentiment Analysis Tool provides users with both a quick assessment of individual text snippets and a more in-depth analysis of sentiment within a CSV file containing reviews.

Sentiment Analysis Tool

Analyze Individual Text

Enter your text:

good car

Polarity Score: 0.7

Subjectivity Score: 0.6

Text to Clean:

this is a good car

Cleaned Text: good car

Analyze CSV File

Upload CSV File

Drag and drop file here
Limit 200MB per file

Browse files

Sentiment Analysis

The analysis involved three primary approaches: TextBlob for polarity scoring, VADER (Valence Aware Dictionary and sEntiment Reasoner) for lexicon-based sentiment analysis, and a pre-trained DistilBERT model fine-tuned on sentiment classification.

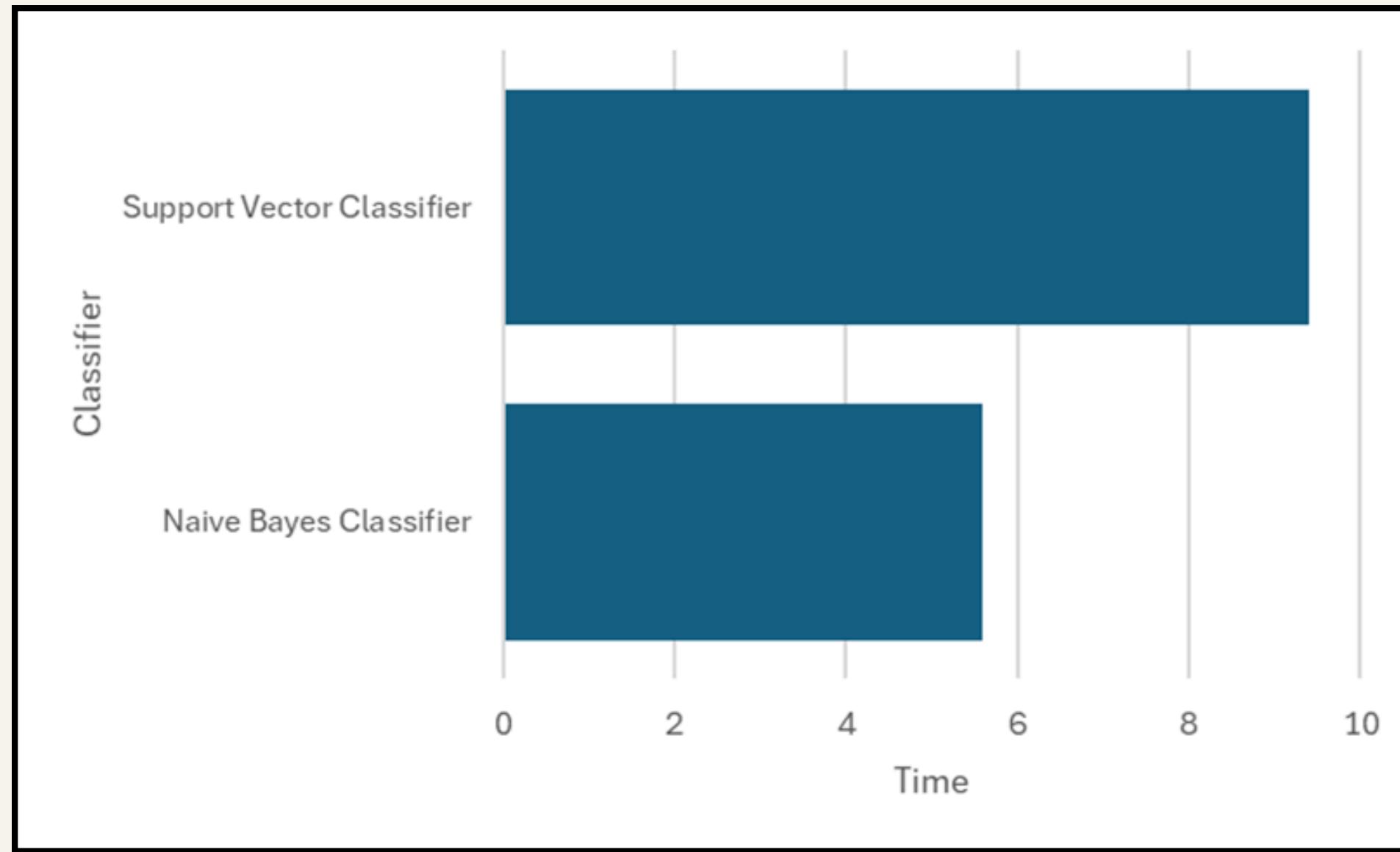
1	app_id	game_name	review	sentiment_textblob	sentiment_vader	sentiment_bert
2	578080	PUBG: BATTLEGROUNDS	good	0.1499999999999999	0.4404	POSITIVE
3	578080	PUBG: BATTLEGROUNDS	Save your	-0.216986423	-0.9913	NEGATIVE
4	578080	PUBG: BATTLEGROUNDS	I'll give cre	-0.17125	-0.296	NEGATIVE
5	578080	PUBG: BATTLEGROUNDS	One of the	1	0.7717	POSITIVE
6	578080	PUBG: BATTLEGROUNDS	I kinda ha	-0.3	-0.5279	NEGATIVE
7	578080	PUBG: BATTLEGROUNDS	Too much	-0.233333333	-0.5574	NEGATIVE
8	578080	PUBG: BATTLEGROUNDS	PUBG is	-0.025	0.296	NEGATIVE

```
Iteration 1  
Cross-validation precison: 0.9701268804539858  
Iteration 2  
Cross-validation precison: 0.969599312493286  
Iteration 3  
Cross-validation precison: 0.9707495429616088  
Iteration 4  
Cross-validation precison: 0.9719782292396401  
Iteration 5  
Cross-validation precison: 0.9699757869249395  
Iteration 6  
Cross-validation precison: 0.971284418442381  
Iteration 7  
Cross-validation precison: 0.9694125262987539  
Iteration 8  
Cross-validation precison: 0.9702374556785215  
Iteration 9  
Cross-validation precison: 0.9704205657739056  
Iteration 10  
Cross-validation precison: 0.9728140676411888  
Mean cross-validation precision: 0.9706598785908209
```

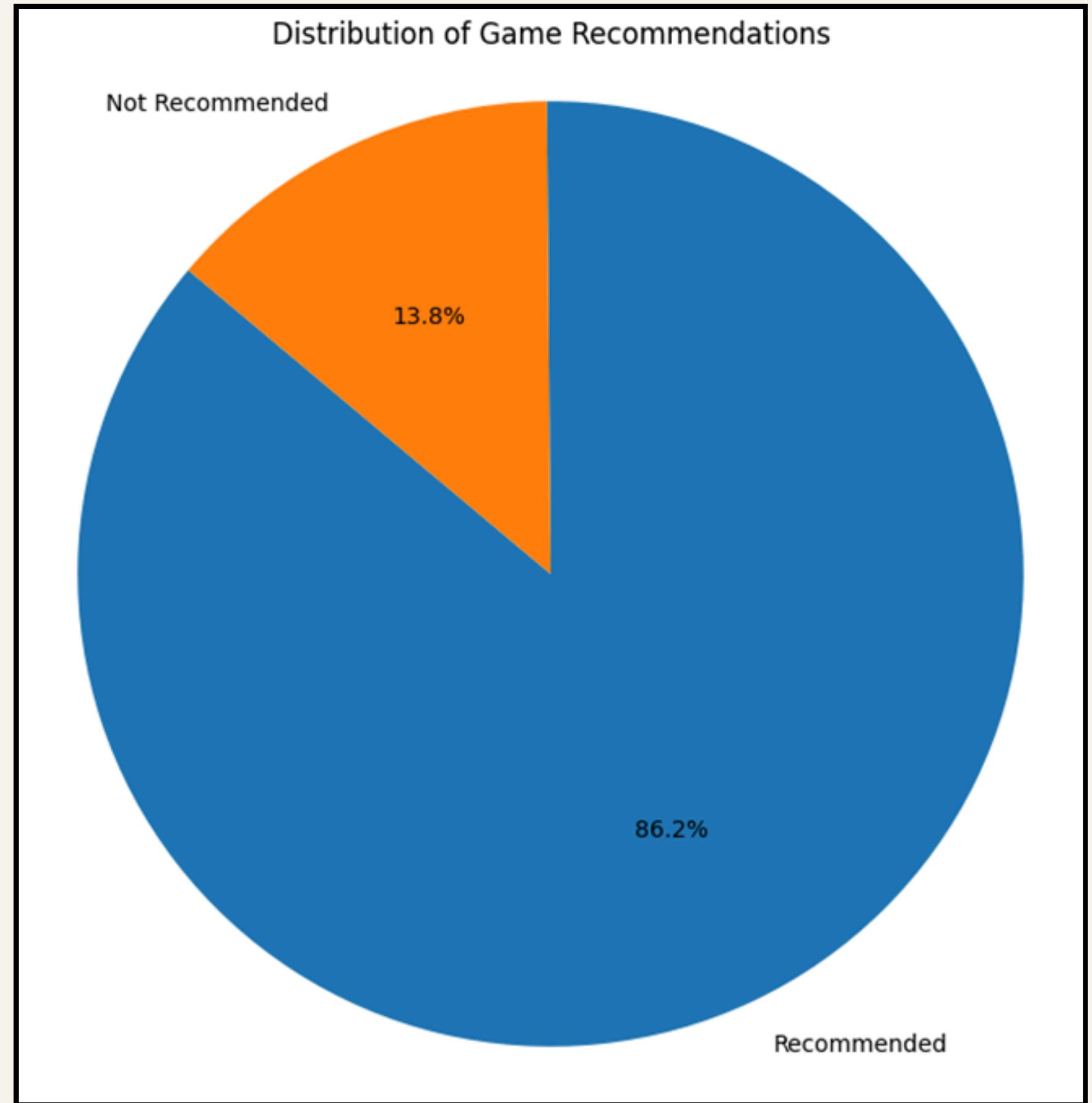
Precision score for
Support Vector Classifier

```
Iteration 1  
Cross-validation precison: 0.9746077507346012  
Iteration 2  
Cross-validation precison: 0.9739193163531435  
Iteration 3  
Cross-validation precison: 0.9749277295975094  
Iteration 4  
Cross-validation precison: 0.975890868596882  
Iteration 5  
Cross-validation precison: 0.9739265505381116  
Iteration 6  
Cross-validation precison: 0.9757854491051144  
Iteration 7  
Cross-validation precison: 0.9730620913206282  
Iteration 8  
Cross-validation precison: 0.9749806265913872  
Iteration 9  
Cross-validation precison: 0.9742620490299627  
Iteration 10  
Cross-validation precison: 0.975867269984917  
Mean cross-validation precision: 0.9747229701852257
```

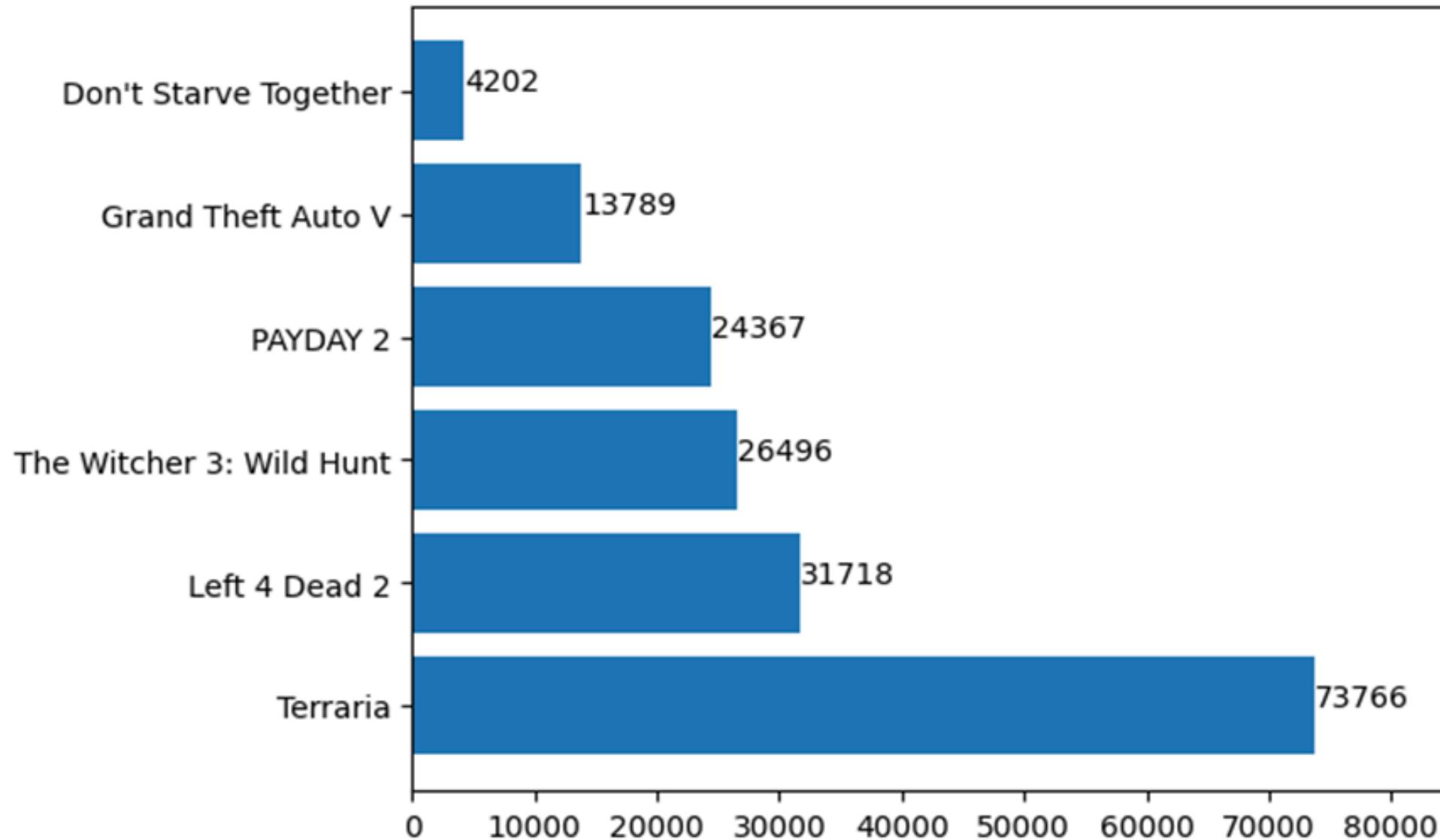
Precision score for Naive
Bayes Classifier



**Time Comparison of both
the Classifiers**

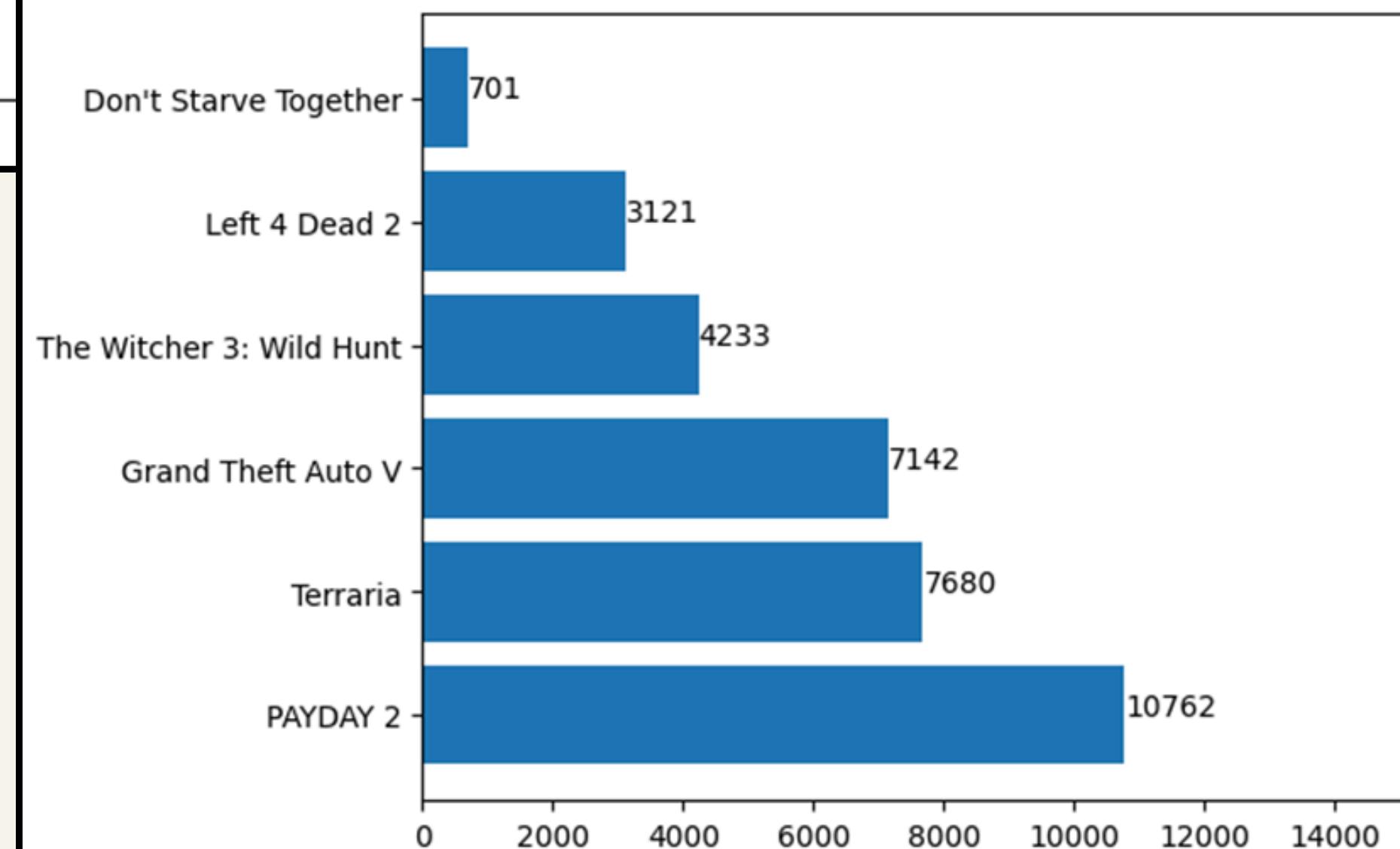


Most Recommended Games



**Comparison of Games to
Quality of Comments**

Most Recommended Comments



Word Cloud for Rating 1



Word cloud for Positive

Reviews

As we can see words like friend, heist, people etc. signify that players prefer multiplayer environment where they can interact and play together with their friends.

Word Cloud for Rating -1



Word Cloud for Negative

Reviews

Words like DLC, buy, pay-to-win, and microtransaction typically indicate player dissatisfaction with games involving useless payment transactions.

Dublin Business School | 2024

THANK YOU