

Genshin Impact Reviews Analysis

Topic Modeling & Sentiment Analysis

Luan Yuanfeng

Contents

1

Introduction & EDA

- What is this project about?
- Findings from EDA

2

Topic Modeling

What topics can be inferred from the reviews?

3

Sentiment Analysis & Classification

Can the sentiment be predicted?



Introduction

Project Aim

1. Train a model to assign reviews into different topics/categories
2. Train a model to classify reviews based on sentiment



GENSHIN IMPACT



Sep 28

Released on PC(Windows), PlayStation 4, Android and iOS

\$400M

In first 2 months

2 Best Game of 2020

From Google Play Store and App Store

Isn't Genshin Impact a Free Game?

- Gacha Game: Random Loot box
- 1 Wish ~ \$2 ⇒ 1 Premium Character approx. \$200
- F2P players gets free wishes as they play





GENSHIN IMPACT





Data & Exploratory Data Analysis

Data Scrapped From

1. Google Play Store
[102,582 reviews](#)
2. App Store
[11,617 reviews](#)

All reviews taken from [US stores](#) for
[English](#) reviews

Data Dictionary

- [Content](#): Text of reviews
- [Score](#): Ratings of reviews
- [Source](#): Either Google Play Store or App Store
- [Update](#): The update version of game when review was posted
(feature engineered from date)

Data Preparation Process

1. Review length

Kept only reviews with length between **3 to 100** words.

2. Duplicates

Removed several reviews with the **same content** and posted at the **same time**.

3. Non-English Reviews

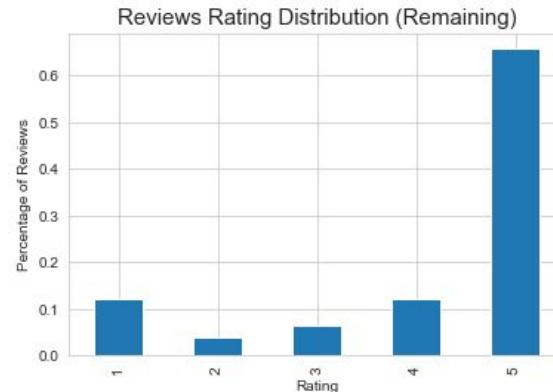
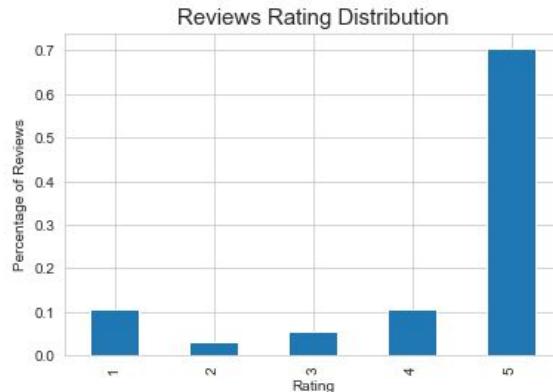
There were some reviews that were **not in English**. These reviews were removed.

4. Update Version

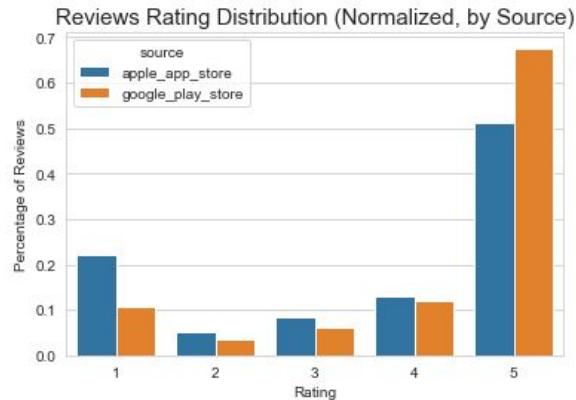
Created the **Update** feature using the *date and time* the review was posted.

Remaining Dataset

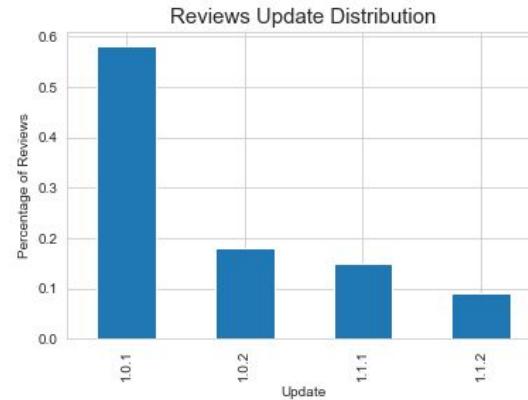
- **88,199 Reviews** (77% Original Dataset, Original: 114,199 Reviews)
- Rating distribution imbalance **slightly improved**



Exploratory Data Analysis



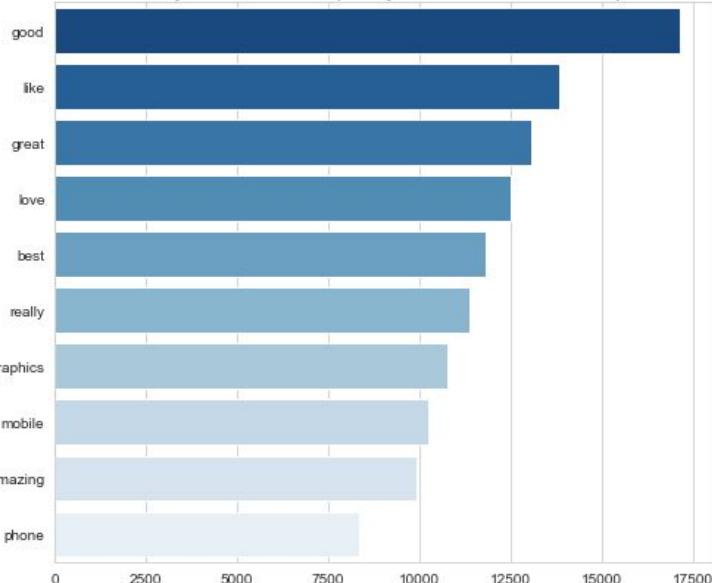
There are more critical reviews from
App Store



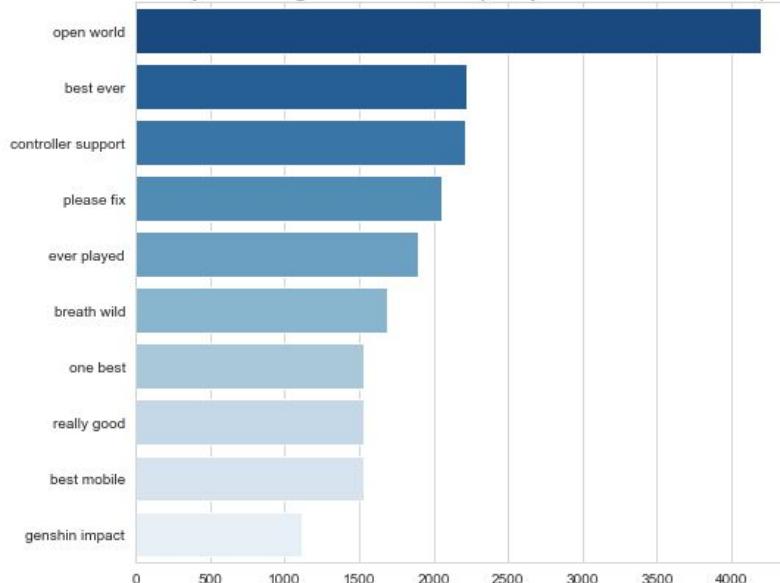
Huge proportion of the reviews are
from the first update

Top 10 N-grams

Top 10 Words (Stopwords Removed)

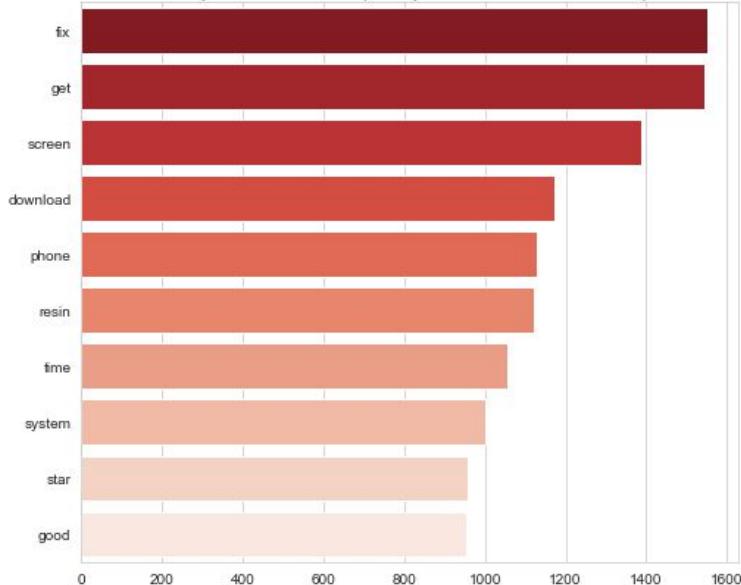


Top 10 Bi-grams Words (Stopwords Removed)

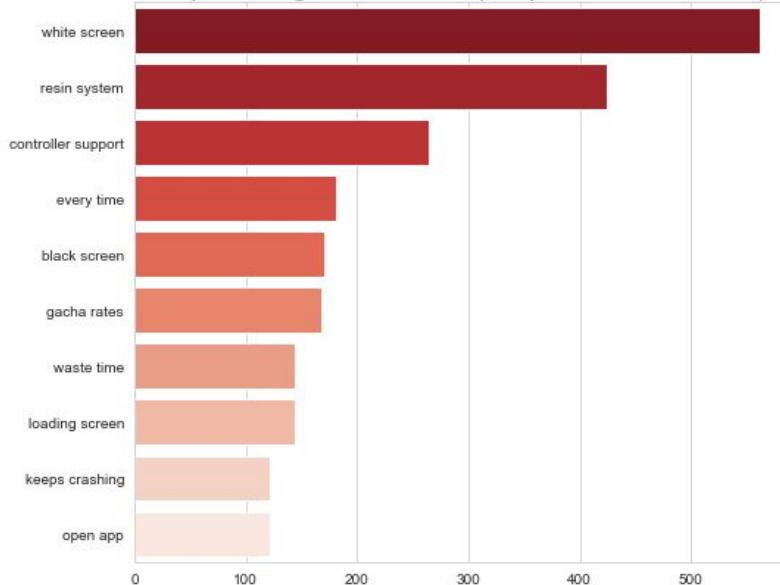


Top 10 N-grams (Negative Reviews*)

Top 10 Words (Stopwords Removed)



Top 10 Bi-grams Words (Stopwords Removed)



*Reviews with rating = 1



Topic Modeling

Topic Modeling (Latent Dirichlet Allocation)

- Using [LDA Model](#) from the Gensim Package
- [Unsupervised](#) learning model:
Do not know the topics or the optimal number of topics
- How to choose [optimal number](#) of topics?
 - Test with a range of number of topics
 - Check [coherence score](#) and [representative texts](#) to choose optimal number of topics

Topic Modeling (Latent Dirichlet Allocation)

- Using [LDA Model](#) from the Gensim Package
- [Unsupervised](#) learning model:
Do not know the topics or the optimal number of topics
- How to choose [optimal number](#) of topics?
 - Test with a range of number of topics
 - Check [coherence score](#) and [representative texts](#) to choose optimal number of topics

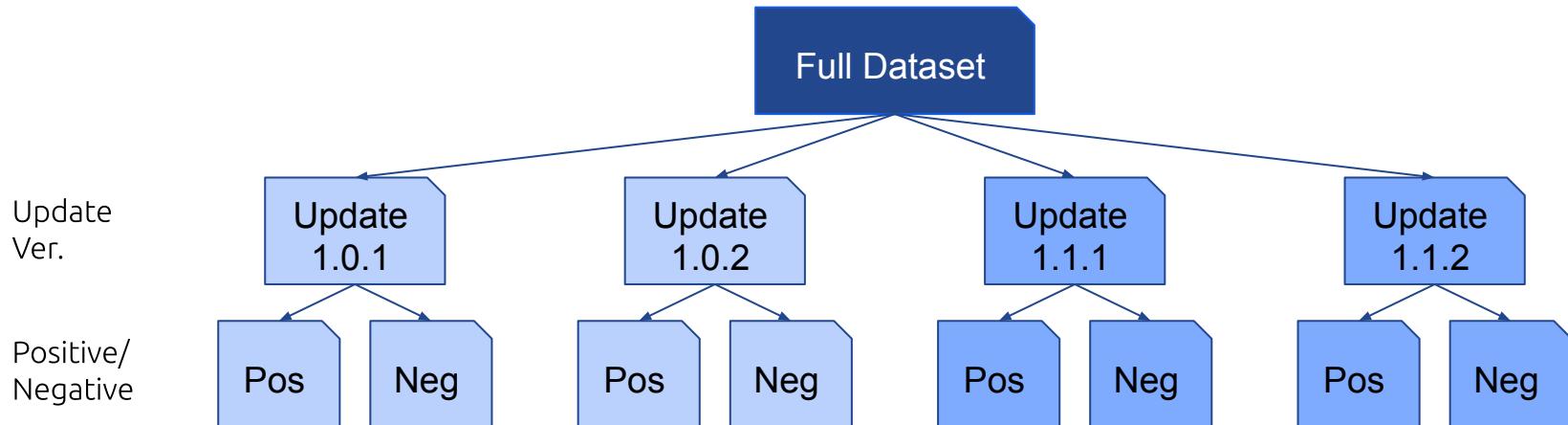
Coherence Score

Metrics for LDA model evaluation.

Measures how [semantically similar](#) are the keywords for each topic.

Splitting the Dataset

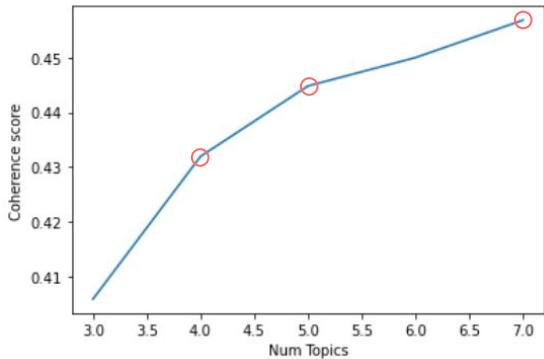
Split the dataset as the mains topics in the reviews might be different.



Example Process

Update 1.1.2, Positive Reviews

Coherence Score Plot



Look at the coherence score plot and **choose 2 (or 3) points**

Representative Texts

Top 7 texts for Topic 3

Keywords: phone, good, say, fix, really, great, need, get, lag, problem

No 1. A great game with minor problems. A audio issues keeps coming up on multiple android devices including Samsung galaxy A51 and my Samsung galaxy tab s6 lite. Very scratchy audio performance even on low settings. To fix try screen recorder pased. Please fix this , other then this sound problem game runs perfect.

No 2. Love it but... I honestly love this game. However when I was trying to link my account, I ended up creating a new accout and now everytime I log in with Game Center, my process would be gone because it's a new account. I have no idea how to get my process back and I'm honestly upset since I worked hard on leveling up and saving up. Please fix this.

No 3. Playing on a galaxy tab s6 256gb 8gb ram, and the game audio is really bad and choppy. A friend's galaxy note 10 was running fine, but after the update 1.1 the game's audio became choppy too. Please fix this issue asap because the audio experience is really horrendous.

No 4. I feel like am ina different world now wwwwbut i have 1 thing to ask plzz can u stop deleting the 8 gb whenever i uninstall the game coz it took long to download and wwwrytime i download or delete i have to download 8 gb is impossible so plz do the thing

No 5. I was really happy playing this game until I update it this morning and when I open it, it wanted me to download the resources again? Like why? I check my files and the games file resources are still here.. Please fix this, km really disappointed because I just downloaded this game yesterday I only played it once, and I don't want to waste another 5gb. Please.

No 6. Game is great, 1 main thing to improve is scenes preformance, they lag even on great high end phones, I just bought t his phone 2 days ago (galaxy s10+, has some good specs) and it lags. on slightly older phones the scenes straight up freez e.

No 7. There's so much to say about the reason why I love this game! BUT the only issue I have is that I have two accounts now and would like to delete one so I can link my email to the account I want to use. That's why I hope there will be a possibility in the future that we can delete our accounts on our own as fast as possible or maybe another way to solve this.

Examine the **topic keywords** and **most representative texts** to distinguish topics

Example Process

Update 1.1.2, Positive Reviews

Suggestions

player
server fun make
time day
resin system
friend co_op

Characters System

character
nice love
get really
give much good
star wish

Technical Issues

problem lag
fix need say
good phone
great really get

Love for the Game

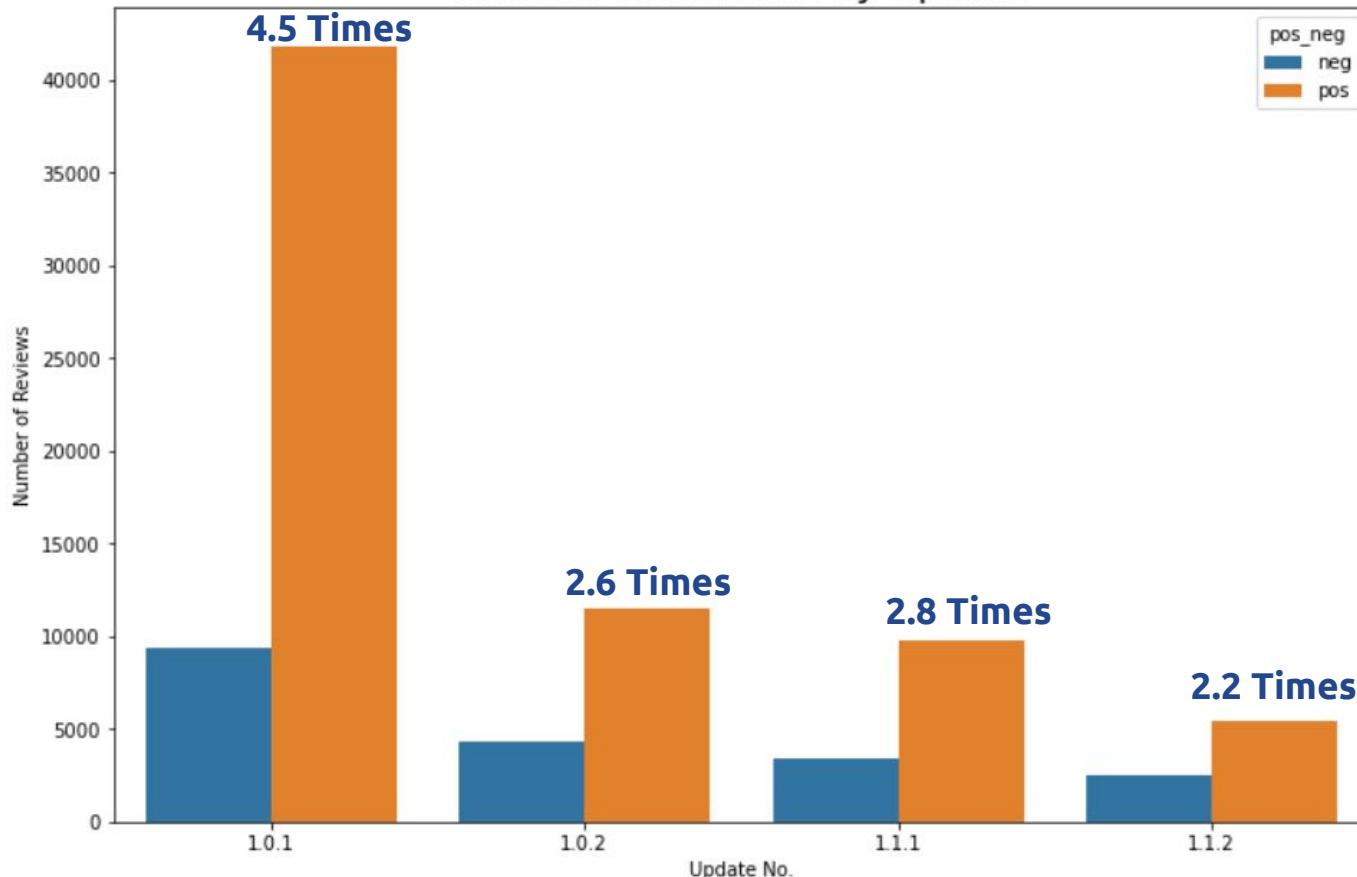
great love mobile
ever good really amaze
fun graphic best



Trends from Topic Modeling

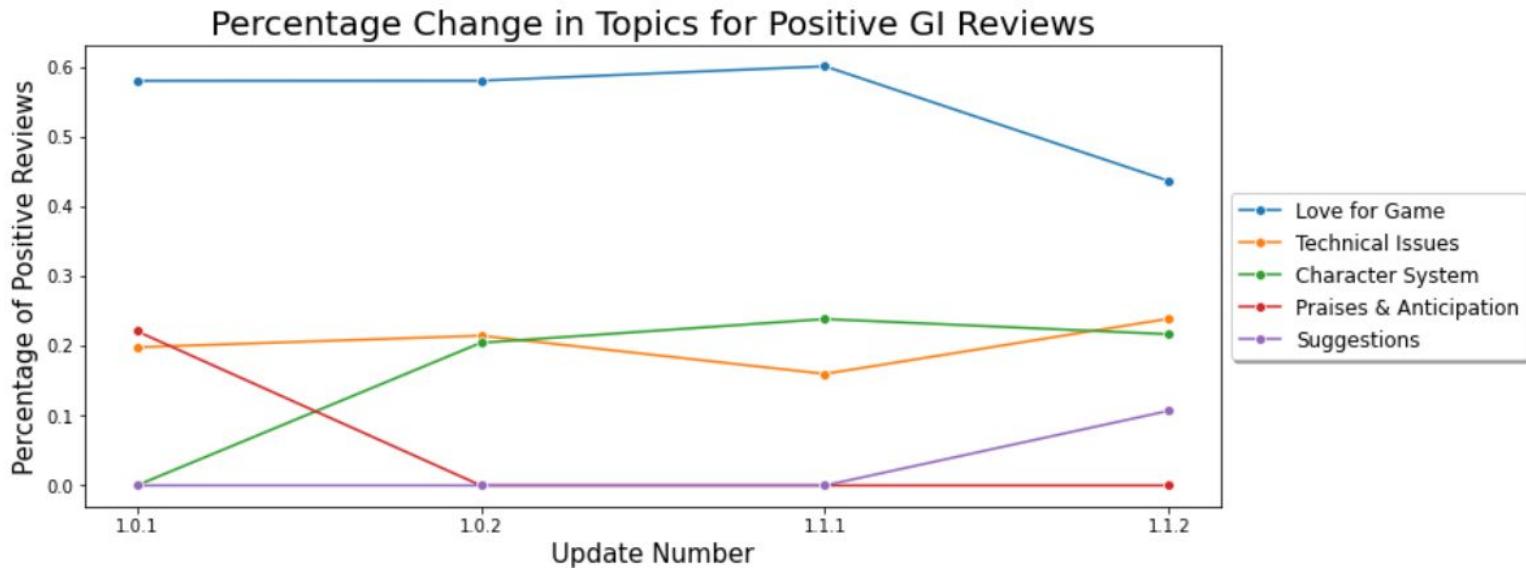


Number of Reviews by Update

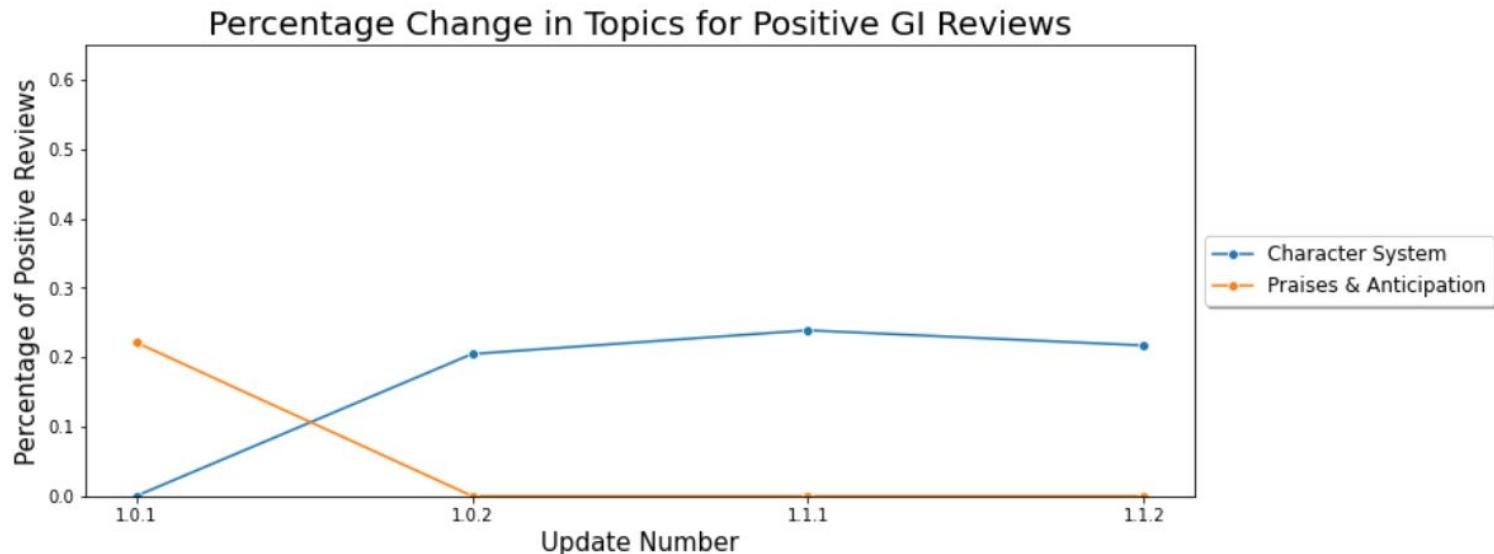


Trends for Positive Reviews

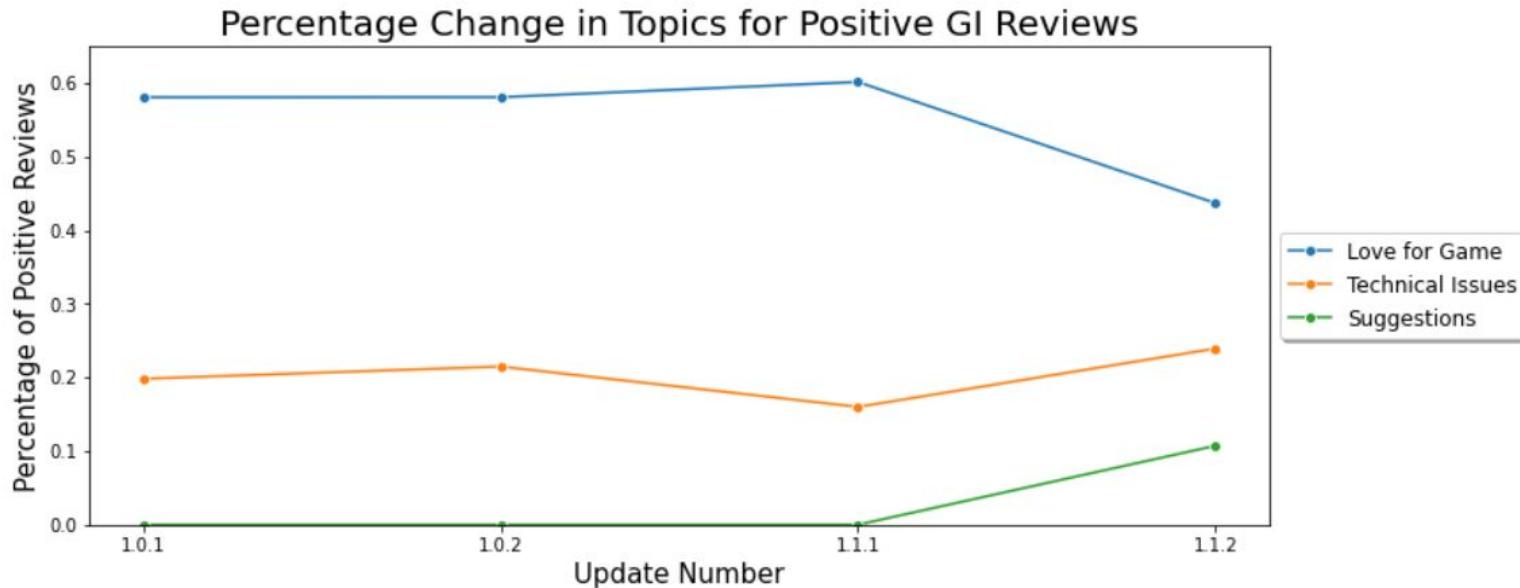
11



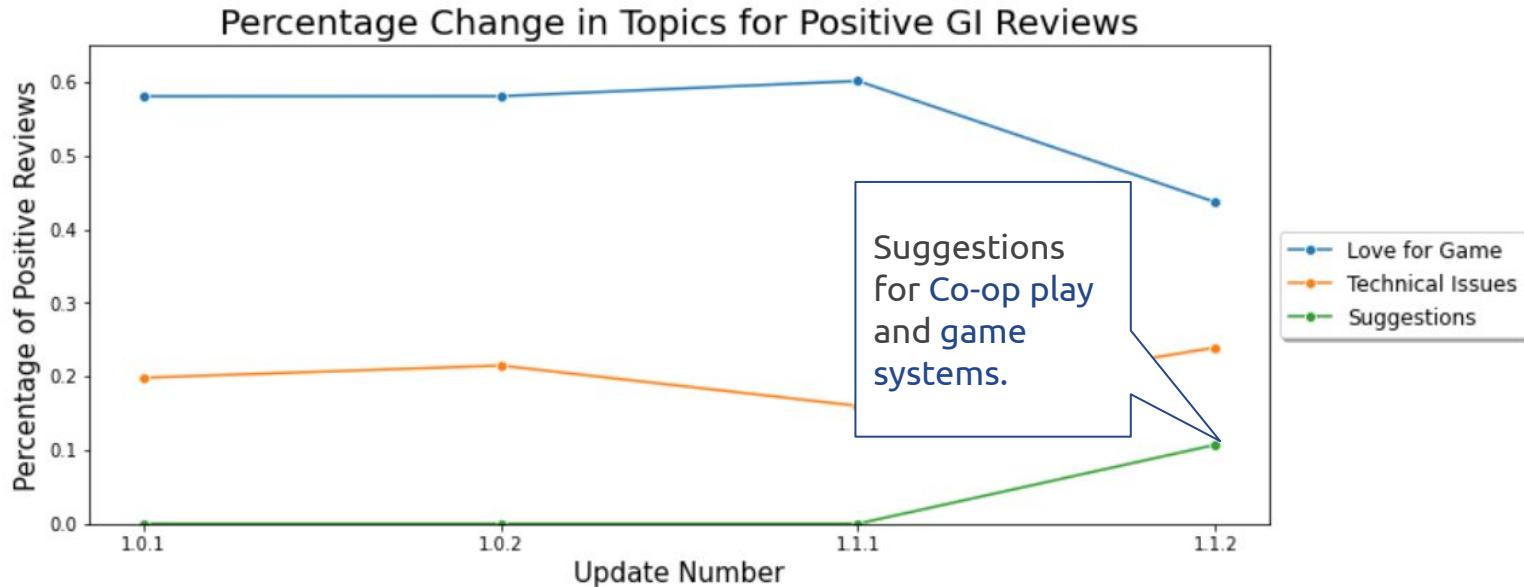
Trends for Positive Reviews



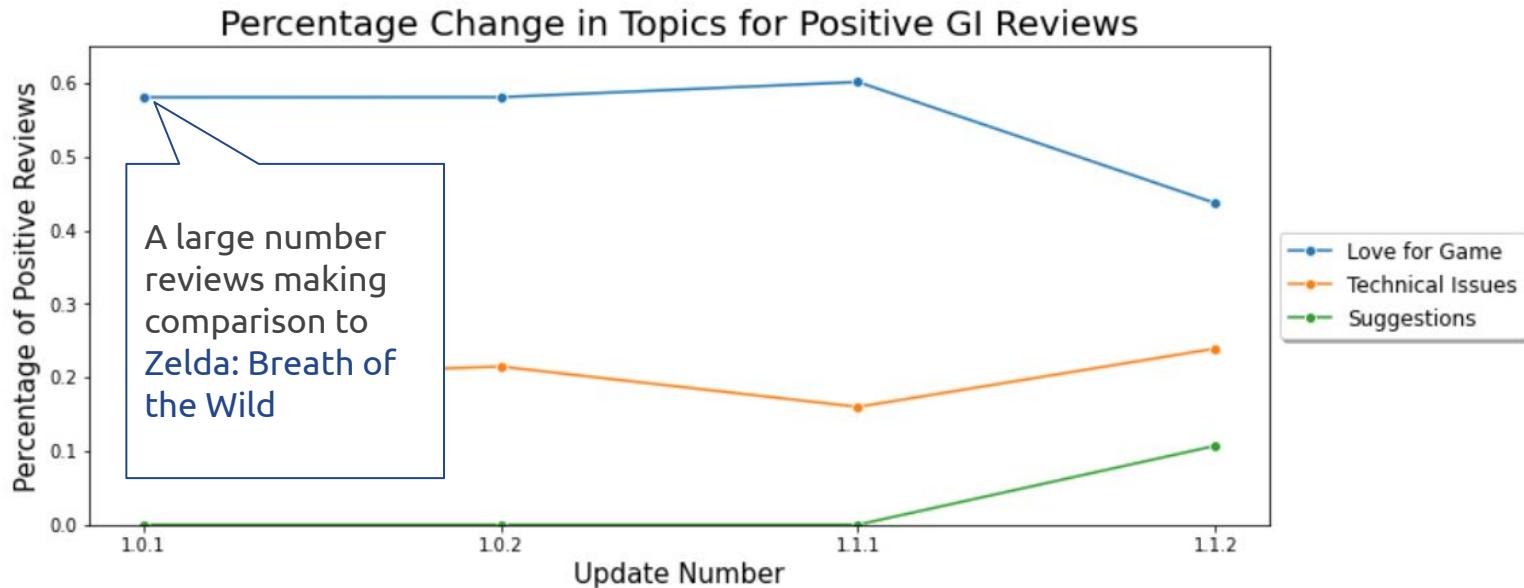
Trends for Positive Reviews



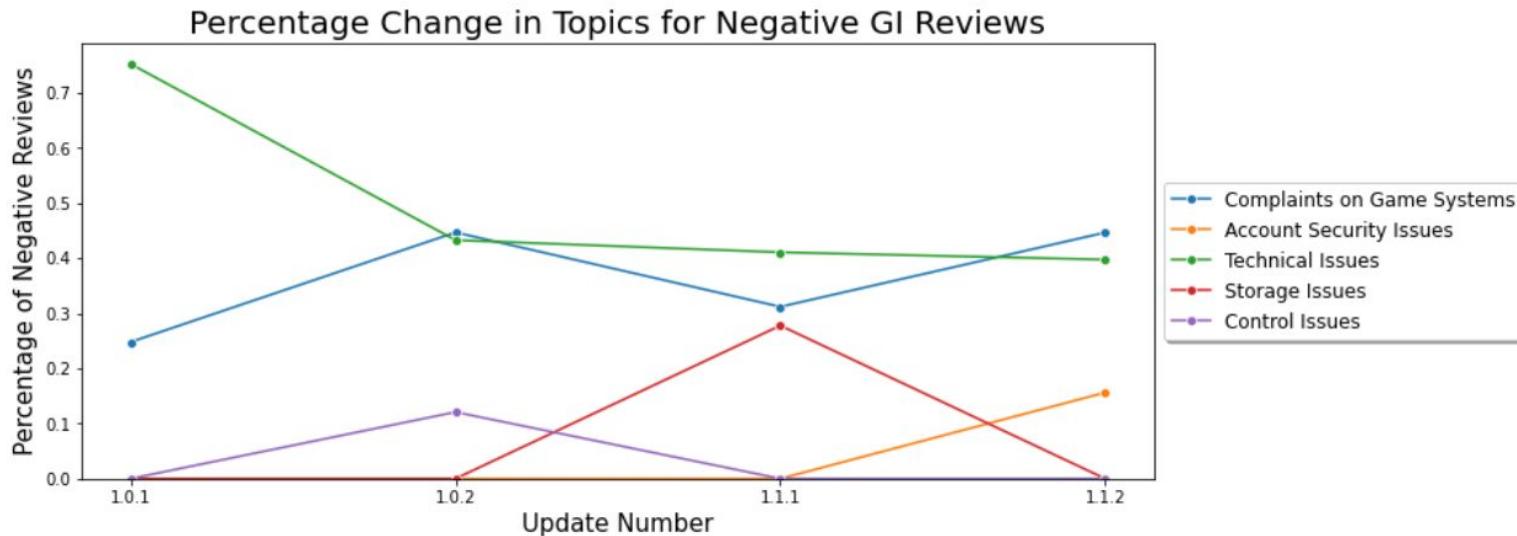
Trends for Positive Reviews



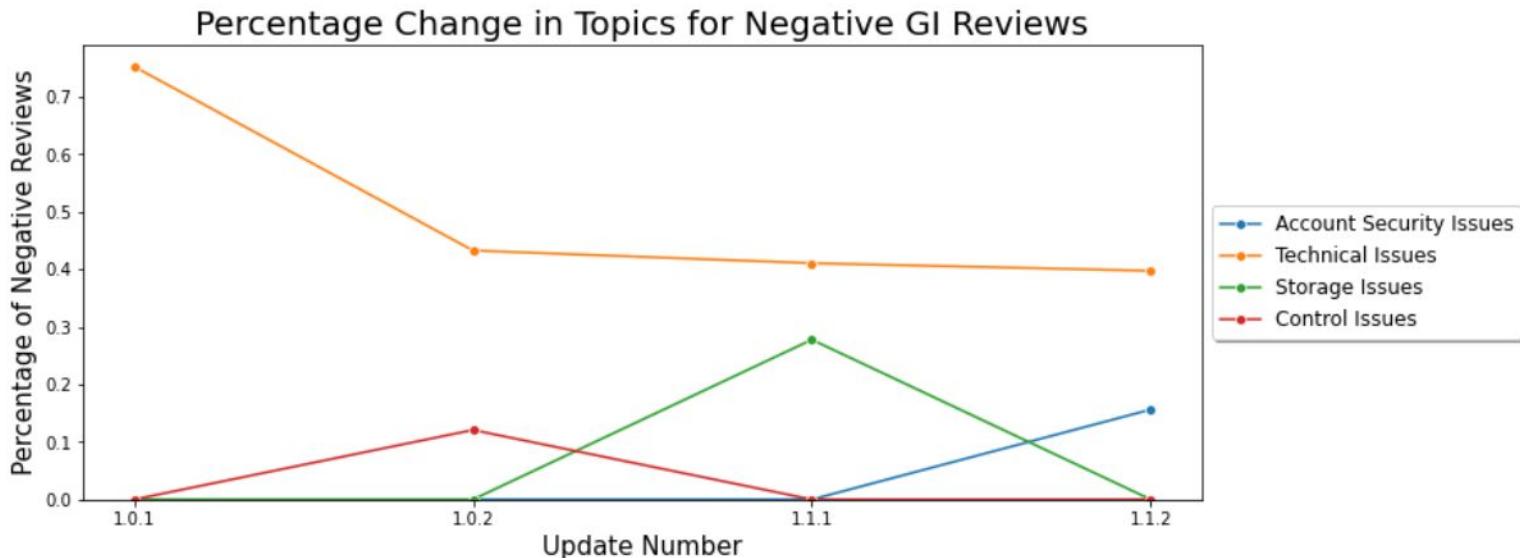
Trends for Positive Reviews



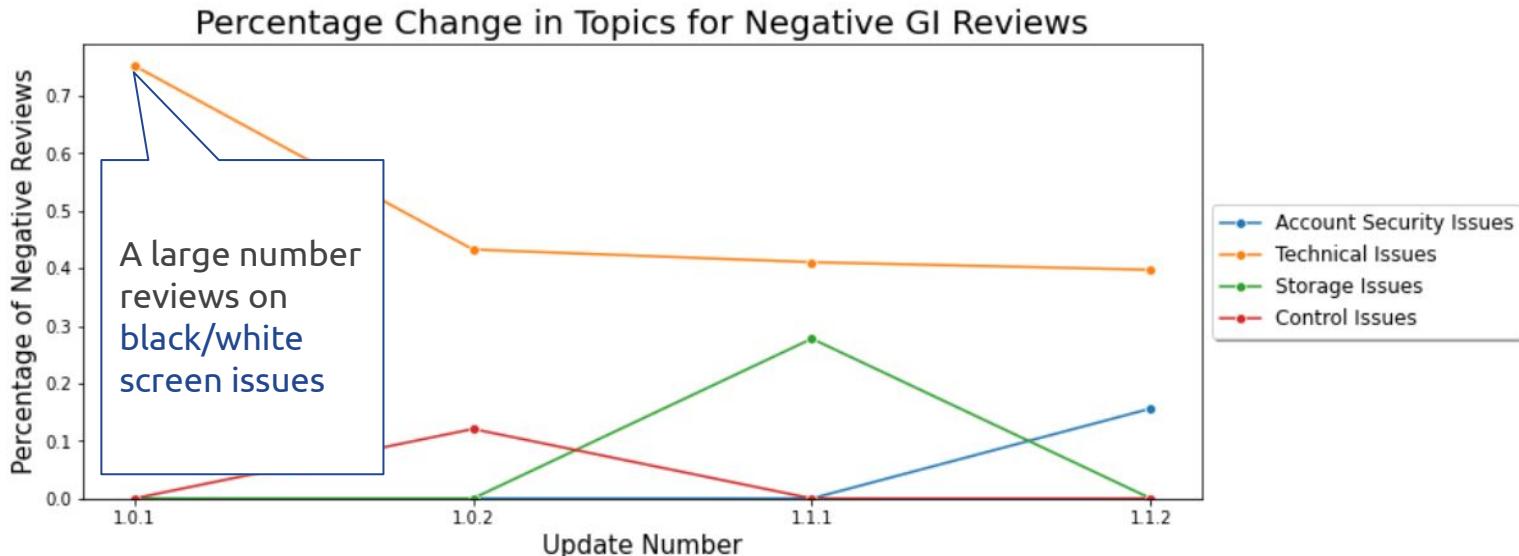
Trends for Negative Reviews



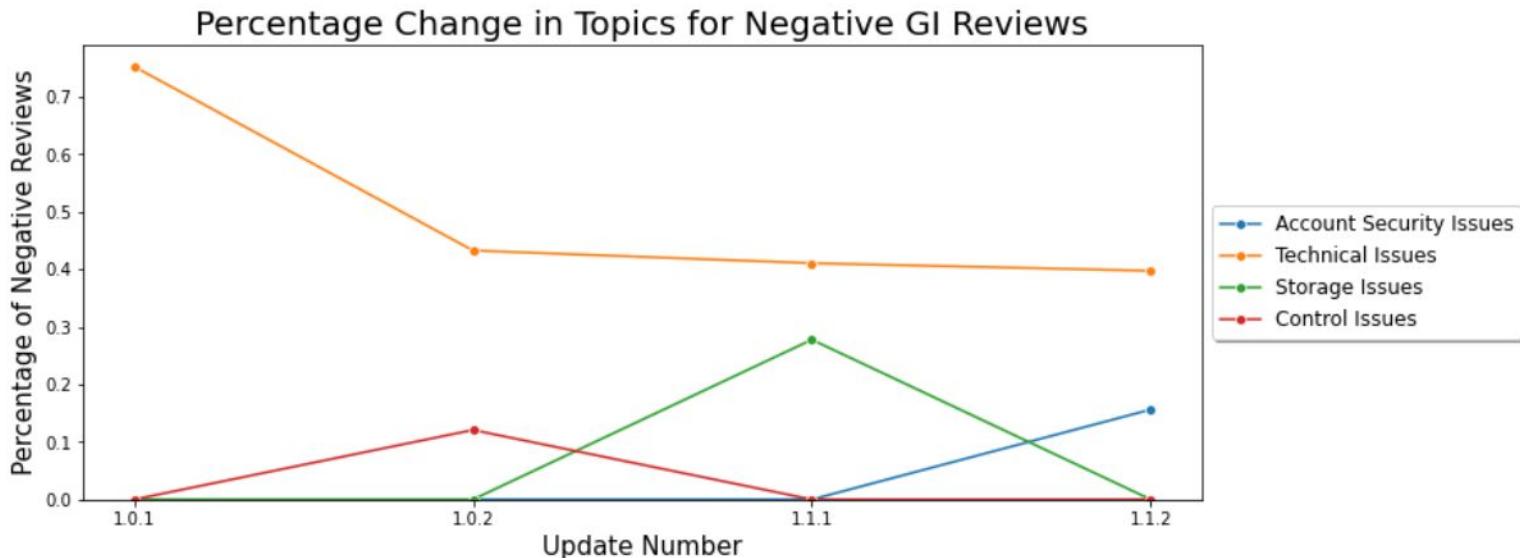
Trends for Negative Reviews



Trends for Negative Reviews



Trends for Negative Reviews



Account Security Issue

Regarding Account Security and Account Theft

Date Posted Nov 18, 2020

Dear Travelers,

Recently, we have received feedback from our social media and customer service channels regarding issues of account security, including account theft. We take all feedback of this nature extremely seriously and promptly investigate each case based on the information provided by Travelers.

Following a careful examination, we can confirm that the miHoYo Account system has not been subject to any malicious attacks.

As for the feedback on potential security vulnerabilities with the verification of miHoYo Account information, we promptly contacted Travelers who initially reported this to us and fixed the issue. On further examination, no user account information has been impacted by this issue.

Account Security Issue

Websites Selling Accounts

AI America Europe ASIA Asia China TW,HK,MO

All Kqing Diluc Qiqi Mona Jean Albedo Zhongli Ganyu Product Enter Items Name Items Search

Server: ASIA Category: Multiple Five-star
Diluc & Mona & Qiqi+Lost prayer to the Sacred winds | 3 Five-star Character +1Five star weapon | not real-named, Can be Bound to Mobile Phones, Can Change Passwords, 100% Safe Guaranteed!Fully leveled character and weapons. Best quality account.

★★★★★

Delivery Instantly In Stock(1)

\$ 119.00

Buy Now

Server: America Category: Multiple Five-star
Diluc & Mona & Qiqi | 3 Five-star Character + Random 4-star | not real-named, Can be Bound to Mobile Phones, Can Change Passwords, 100% Safe Guaranteed!

★★★★★

Delivery In 1 - 24 Hours In Stock(1)

\$ 65.99

Buy Now

Server: America Category: Multiple Five-star
Kqing & Diluc & Qiqi| 3 Five-star Character + Random 4-star | not real-named, Can be Bound to Mobile Phones, Can Change Passwords, 100% Safe Guaranteed!

More Info ▾

★★★★★

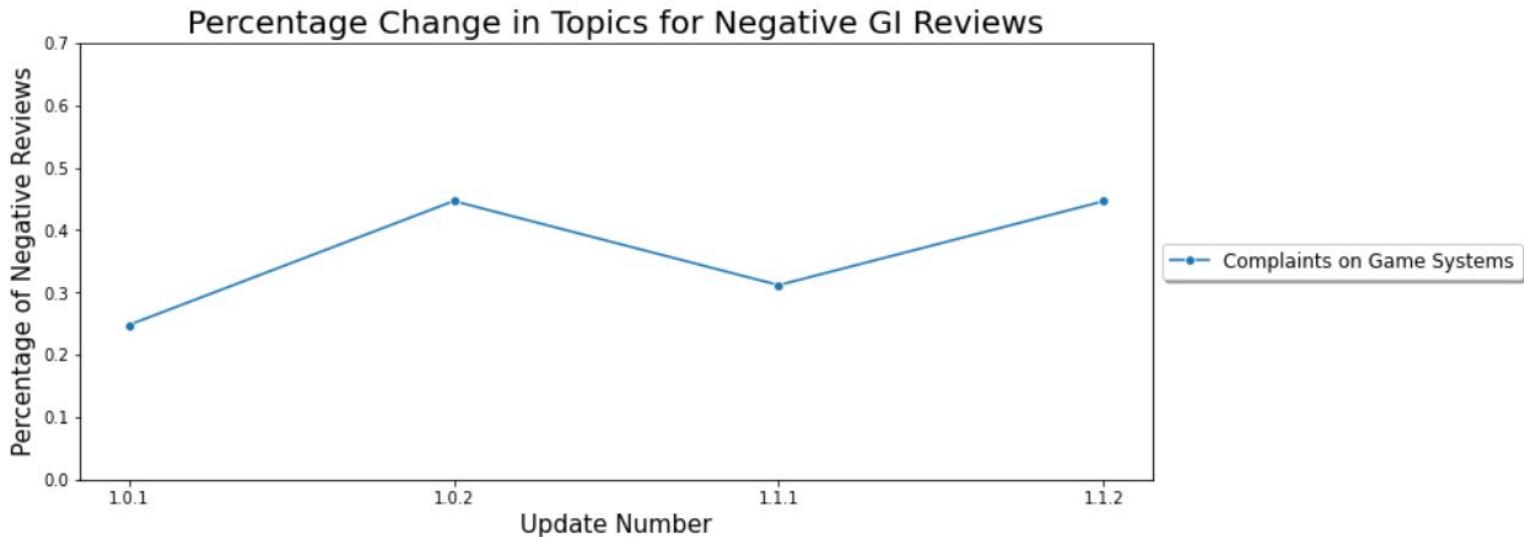
Delivery In 1 - 24 Hours In Stock(1)

\$ 59.99

Buy Now

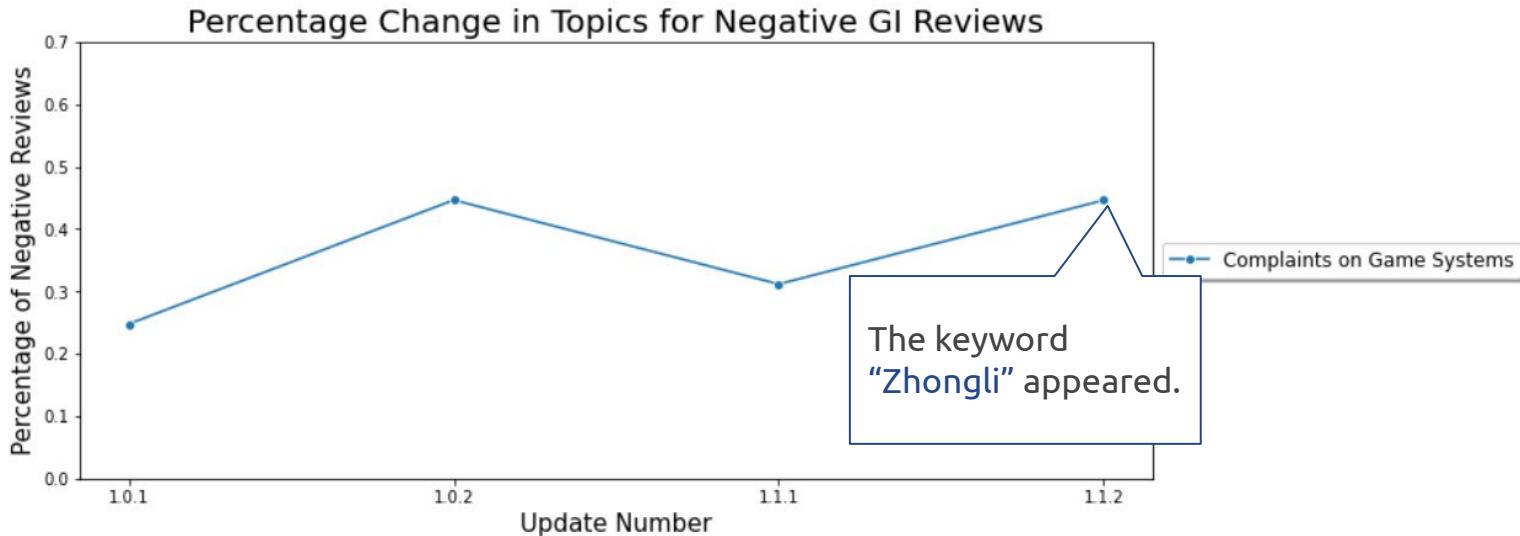
Trends for Negative Reviews

11.1



Trends for Negative Reviews

11.1



Zhongli - “God” of Earth Element



- Perceived by community as **very weak**
 - A lot of backlash in feedbacks and reviews
- ⇒ Mihoyo decided to **increase the strength of the character**

The False Geo Archon	The One True Geo Archon
Weak Physical Body	Bulky Physical Body
Typical DMG output	Explosive DMG output
Gewgaw 猴动作	Real ground splitting aggression
Only shield and CC	Berserk like, undeterrable

米哈游·原神

Sentiment Analysis & Classification



Models Tested

Models	Accuracy
1. Textblob + Binning	13.98%
2. Vader + Binning	51.97%
3. Logistic Regression	70.86%
4. Multinomial NB	72.11%
5. Support Vector Machine	69.98%
6. LSTM (Neural Network)	66.84%

Due to the [imbalanced dataset](#), [class weight](#) attributes were used whenever possible.

However, even with LSTM, the accuracy is very low. This could be due to:

1. [Highly imbalanced dataset](#)
2. [Insignificant difference](#) between some ratings

Try Different Number of Classes (Using LSTM)

Models	Accuracy
3 Classes (Pos, Neu, Neg)	77.00%
2 Classes (Pos, Neg) - Ratings = 3 → Pos	84.37%
2 Classes (Pos, Neg) - Ratings = 3 → Neg	85.35%
2 Classes (Pos, Neg) - Ratings = 3 excluded	88.51%

This is reasonable as a review with rating of 3 might mean both positive or negative for different people.

⇒ Noise for the model.

Aim of project: Classify new feedbacks.

Ability to classify them into positive and negative classes is sufficient.

Optimising the Model

Hyperparameter Optimized	Accuracy	F1-Score
Benchmark	88.51%	71.51
Learning Rate	89.57%	72.52
Drop Out Rate	89.12%	72.32
Three Layer LSTM	90.49%	72.97

Testing the Trained Models



Trying Out the Models on New Reviews

"The game is truly **one of the best in its genre**. the **world is fun to explore**, the **combat is quite engaging!**

The **biggest issue**, I have with this game is the resin system! It **feels too restrictive**, and **heavily halts the players progression** in becoming stronger and upgrading their characters. All the fun content requires you to have resin, and there's **not enough ways to refill resin reasonably**. I often **end my play sessions unsatisfied** because I just wanted to play and farm a bit more"

Class Prediction	Topic Prediction
Neutral (51.67%)	Character System, Suggestions

Trying Out the Models on New Reviews

"The game is truly **one of the best in its genre**. the **world is fun to explore**, the **combat is quite engaging!**"

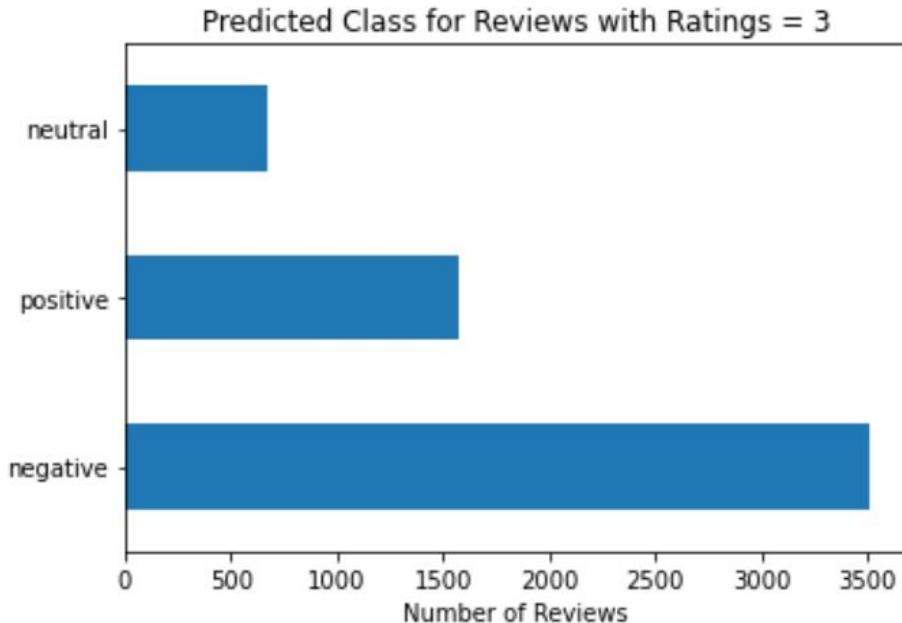
Class Prediction	Topic Prediction
Positive (99.94%)	Love for Game

Trying Out the Models on New Reviews

"The **biggest issue**, I have with this game is the resin system! It **feels too restrictive**, and **heavily halts the players progression** in becoming stronger and upgrading their characters. All the fun content requires you to have resin, and there's **not enough ways to refill resin reasonably**. I often **end my play sessions unsatisfied** because I just wanted to play and farm a bit more"

Class Prediction	Topic Prediction
Negative (10.33%)	Complaints on Game Systems

Trying the Model on Reviews with Rating = 3



Approx. Distribution (in%):

61% → Negative

27% → Positive

12% → Neutral

Conclusion

Accuracy of Sentiment Analysis model could be affected by the **subjectivity of the ratings** from the reviewer

- A review with rating of 4 could have similar content as a review with rating of 2 or even 1

Ways to overcome:

- Gather more data
→ More generalized
- Use more complex models
(i.e. BERT)

Possible Improvements

LDA Model could be improved by using the positive/negative class predicted

Additional techniques for Sentiment Analysis:

- GloVe embedding
- BERT classification

Thank you! :)



Credits

This is where you give credit to the ones who are part of this project.

Did you like the resources on this template? Get them for **free** at our other websites.

Presentation template by [Slidesgo](#)

Icons by [Flaticon](#)

Images & infographics by [Freepik](#)

Author introduction slide photo created by **katemangostar** - Freepik.com

Big image slide photo created by **jcomp** - Freepik.com

Text & Image slide photo created by **rawpixel.com** - Freepik.com

Text & Image slide photo created by **Freepik**