November 2022

Operation Optimization with a Dive into Genshin Impact’s Comments

Group 14

Shen Yi, Lin Fangzhou,Wu Yangyi, Li Kongwen, Chi Yijin

1.1 Genshin Impact in the Game Industry

Open-Word
Action Role-Playing Game

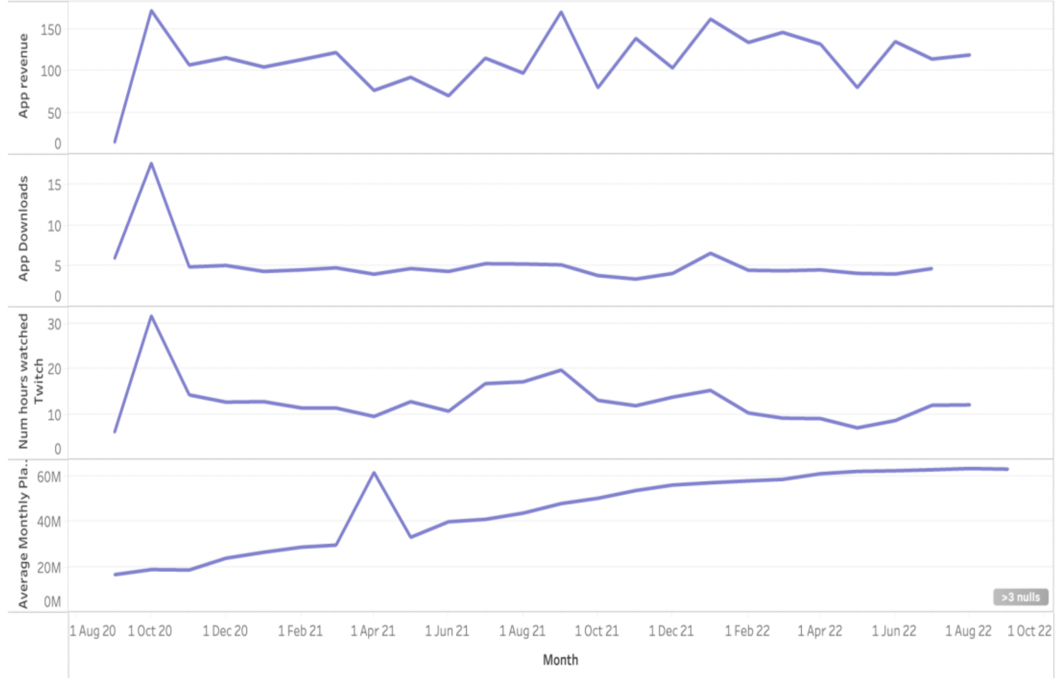
Launch time: Sep 2020

Award & Ranking:
Best Mobile Game (2021)
Top Grossing Mobile Games Worldwide (2020-2022)

Another two games being developed by miHoYo:
Honkai: Star Rail
Zenless Zone Zero

Performance of Genshin Impact

Top to down: App revenue(↑), App downloads (↑),
Hours watched on Twitch, Average monthly players (↑)

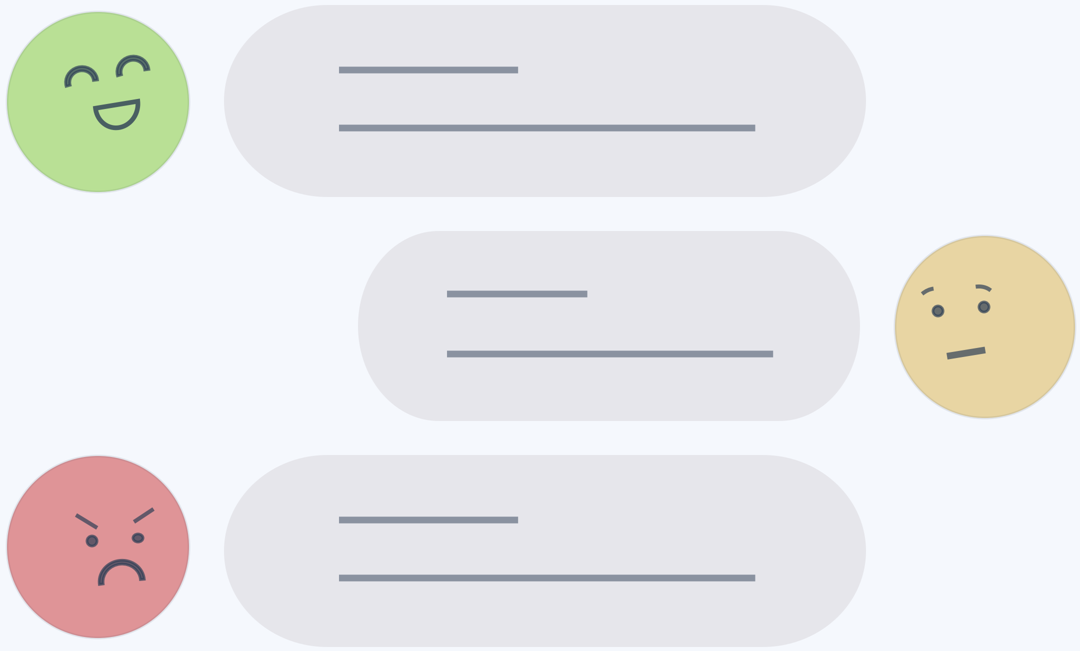


Introduction | Project Design | Data Analysis | Deliverables

1.2 Business Problem

Genshin Impact's **player pool keeps growing larger** and results in an increasing number of comments in the application stores, including Google Play Store. The **variety in comments also keeps growing**. As popular as today, a potential **reputation drop would be costly**.

And meanwhile, what **references** can be derived for the two developing games targeting similar players?



1.3 Objectives

To Prevent Reputation Drop

Update the auto-reply corpus

Detect what is the timing to take some maintenance actions in advance

What is complained by the players?

To Derive Reference for The Two Developing Games

What features are appreciated by the players?

2.1 Stakeholder Strategy

Stakeholder

Operation Team

Key Concerns

1. Find out major topics involved in players’ comments
2. Score isn’t an indicator as effective as expected, sentiment of comments maybe a good surrogate
3. Penalty should be higher on misclassification of a timing when intervention are necessary

Response Action

1. Apply LDA model to find out the drawbacks and strengths of the product
2. Weighted average of score and sentiment to create a new indicator of reputation
3. Use cost function to determine the model

2.2 Data Preparation

External Sources

Google Trends | Google Play

Data Preprocessing

- Data Cleaning
- Pre-Sentiment Analysis
- Set Up New Attributes

- Polarity (-1 - 1)
- Subjectivity (0 - 1)
- Length = length of comment
- ABS = |Score – average Score|
- Days of Comment
- **Usefulness** = 1 - noteworthy
  0 - normal

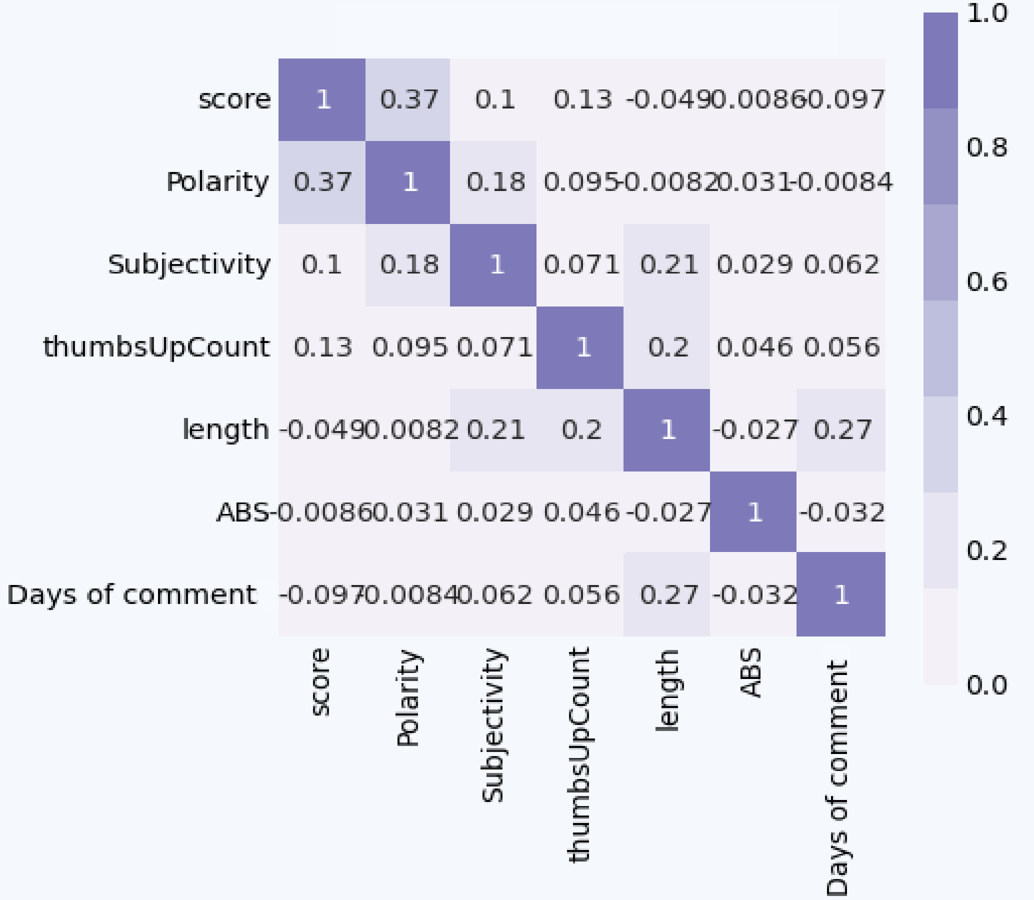
Quntile 1 of 'ThumbsUpCount' (6) as the split benchmark

2.3 Data Set

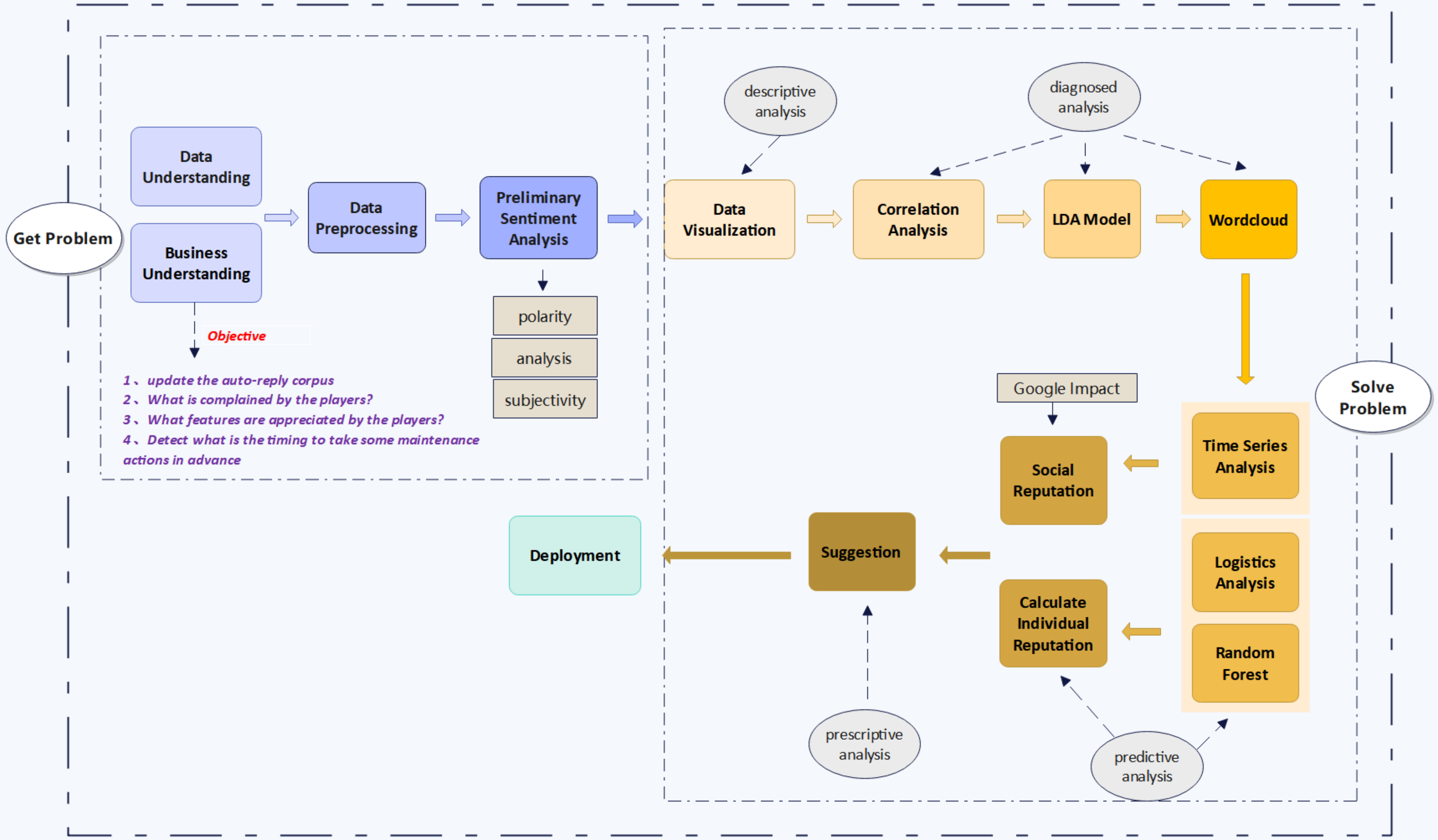
15 Variables	Type		Description
reviewID	Categorical	Char	ID of the comment
content	Categorical	Char	comment text
score	Numerical	Integer	players' rating of the game
thumbsUpCount	Numerical	Integer	number of likes from other players for each comment, quantile 1(25%) is 6
reviewCreatedVersion	Categorical	Char	game version
at	Numerical	Decimal	time of the review created
replyContent	Categorical	Char	game official's reply to player' comments
repliedAt	Numerical	Decimal	time of the game official's reply
polarity	Numerical	Decimal	polarity of players' comment based on sentiment scoring
subjectivity	Numerical	Decimal	subjectivity of players' comment based on sentiment scoring
ABS	Numerical	Integer	=|Score – average Score|
reputation	Numerical	Decimal	calculated from score and sentiment
review_length	Numerical	Integer	the length of the review (count of words)
usefulness	Categorical	Char	binary variable, 1 if noteworthy comment, and 0 if not noteworthy (just normal)
days of comment	Numerical	Integer	How many days the review has been published

Training+Testing: 1273 observations　Deployment: 642 fresh comments

Correlation Heatmap



2.4 Analytical Pathway



Objective 1

To deal with the growing variety in comments,
the operation team what to update the auto-reply corpus.

1. The type and distribution of topics
2. The distribution of sentiment

3.1 LDA

Pre-Processing

Tokenization

↓

Case lowering

↓

Lemmatization

↓

Remove stopwords

↓

Remove punctuation

↓

Remove token with length smaller than 3

Additional Stopwords

'game’, 'even', 'really', 'like', 'still', 'get', 'also', 'one'

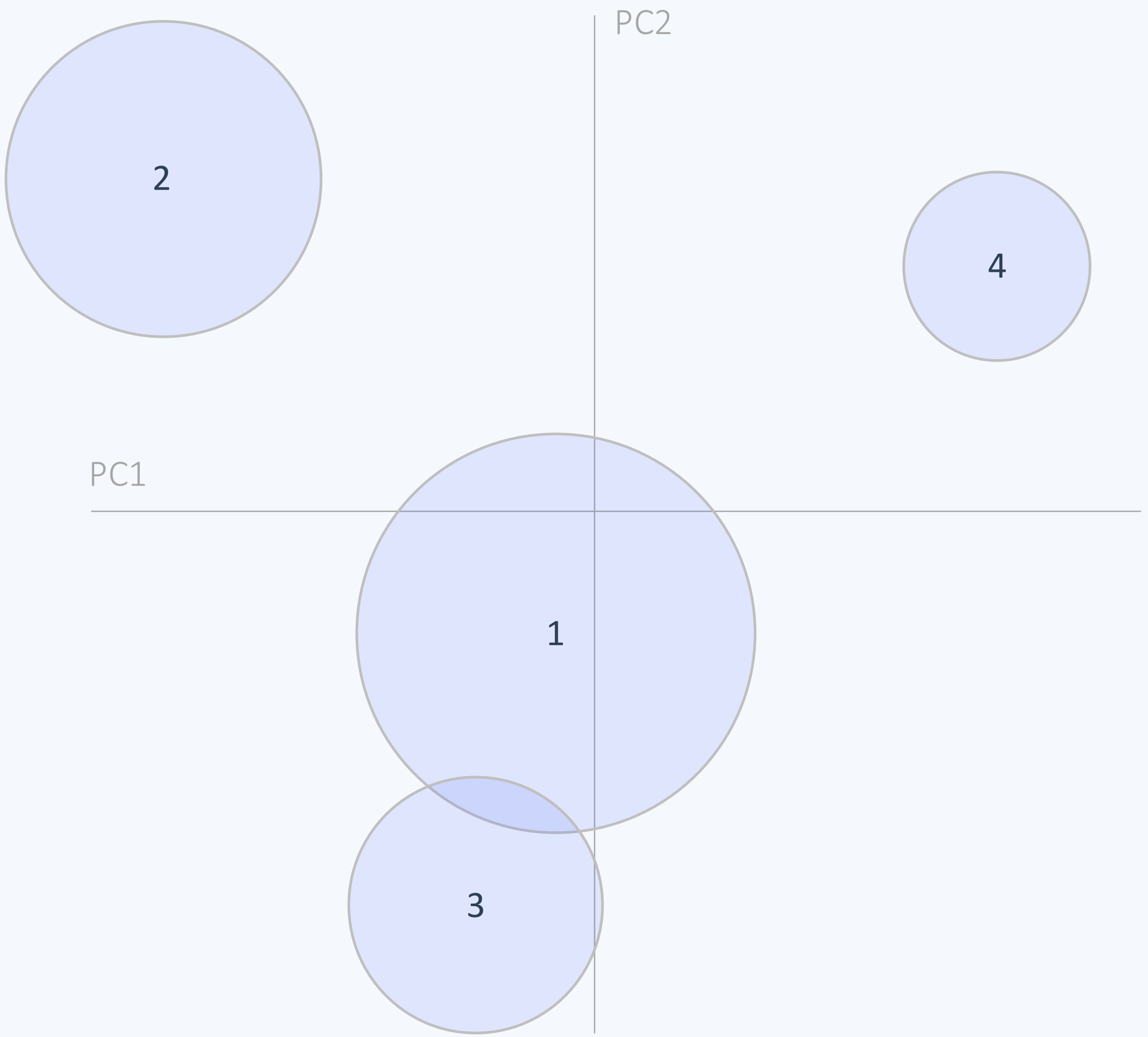
3.1 LDA

LDA Model Parameters
num_topics = 4
Passes = 10
Chunksize = 128

4 Topics
Topic 1: 43.7% of tokens
Topic 2: 23.4% of tokens
Topic 3: 18.8% of tokens
Topic 4: 14.0% of tokens

Coherence
Umass: -2.5639
CV: 0.4093

Intertopic Distance Map (via multidimensional scaling)



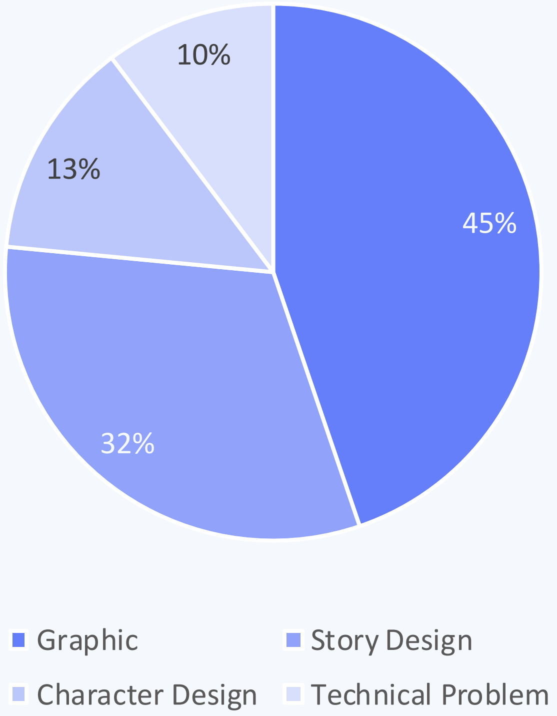
3.2 Topics

Most Relevant Terms for the 4 Topics (Sorted by Estimated Frequency)	
Story Design character, story, great, player, time, good, fun	**Mobile Technical Problems** phone, play, mobile, problem, storage, controller, space, fix
Graphic good, update, graphic, character, play, new	**Character Design** character, best, amazing, look, love, fighting

3.2 Topics

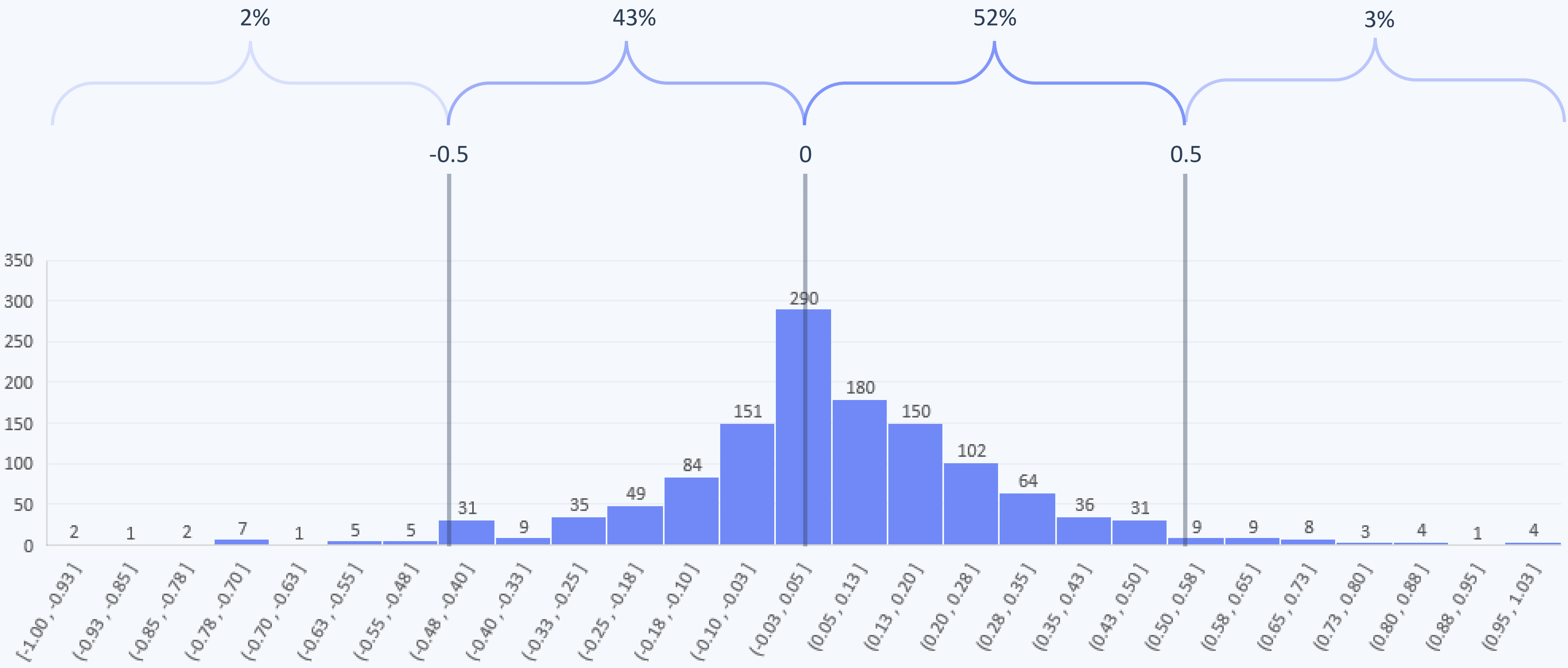
Topic		Review Number
1	Story Design	404
2	Technical Problem	131
3	Graphic	570
4	Character Design	168
Total		1273

Suggested Proportions of Topics for the Redesigned Auto-Reply Corpus



3.3 Sentiment Distribution (Polarity)

55% comments are positive, 45% comments are negative



Objective 2&3

Improvement should focus on what is doing bad and references can be taken from what is doing well.

1. What is complained by the players?
2. What is appreciated by the players?

3.4 Diagnosis on comments based on sentiment

Negative Comments	Positive Comments

Negative Comments

Potential problems

System bug, time(service response and some quests are too long), occupy a large amount of storage

Positive Comments

Advantages

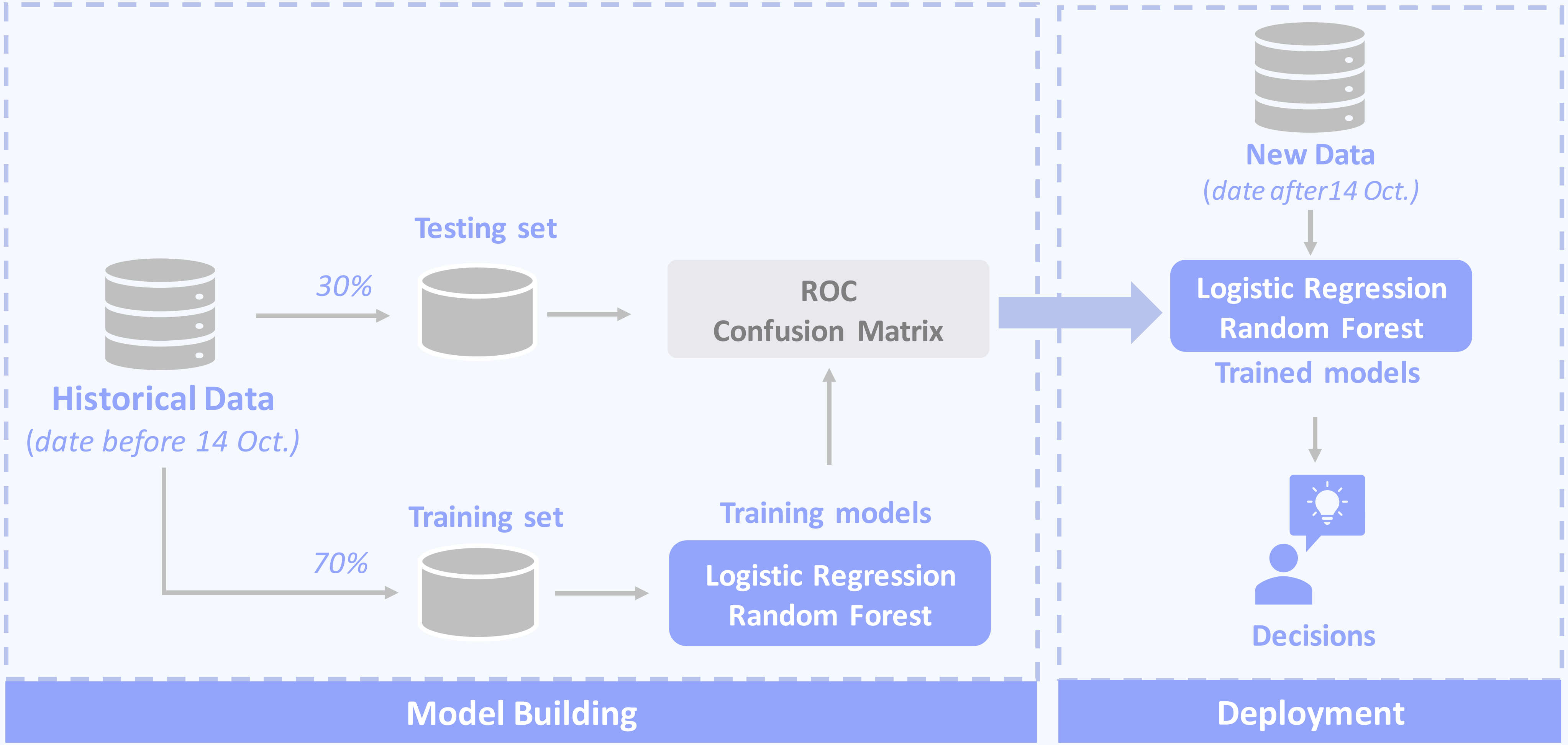
Elaborate design of characters and storylines

Objective 4

In preventing the reputation decline, the operation team want to detect what is the timing to take some maintenance actions in advance.

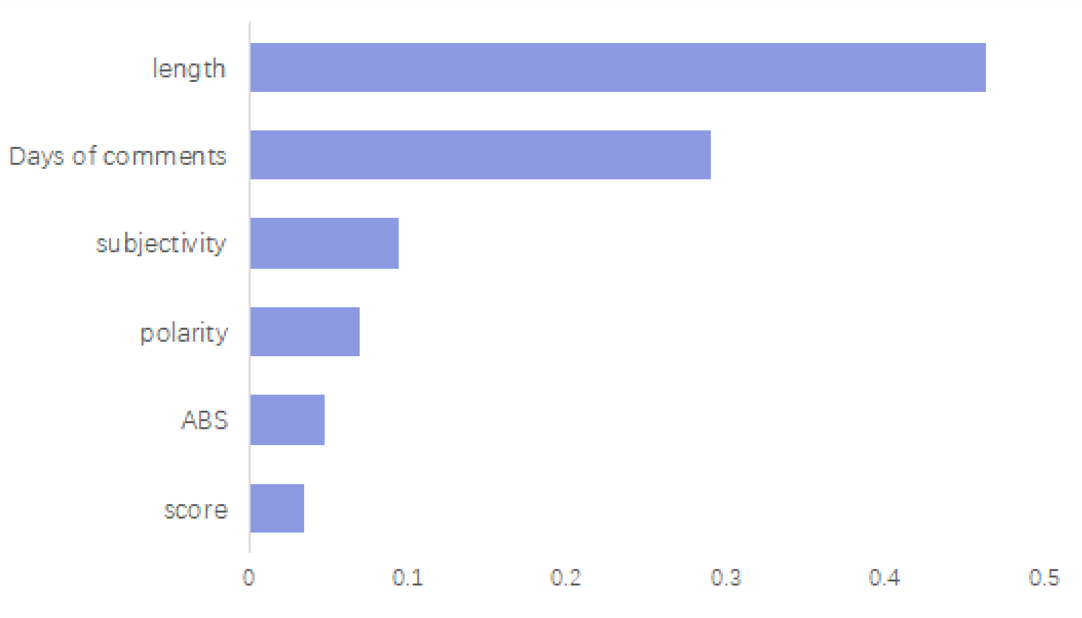
1. Develop a classification model to predict noteworthy comments
2. Construct an indicator for player reputation by leveraging score and sentiment.
3. Determine a pre-warning model to decide whether the comments are showing an alert so that actions need to be taken to appease the players.

3.5 Classification model building pathway

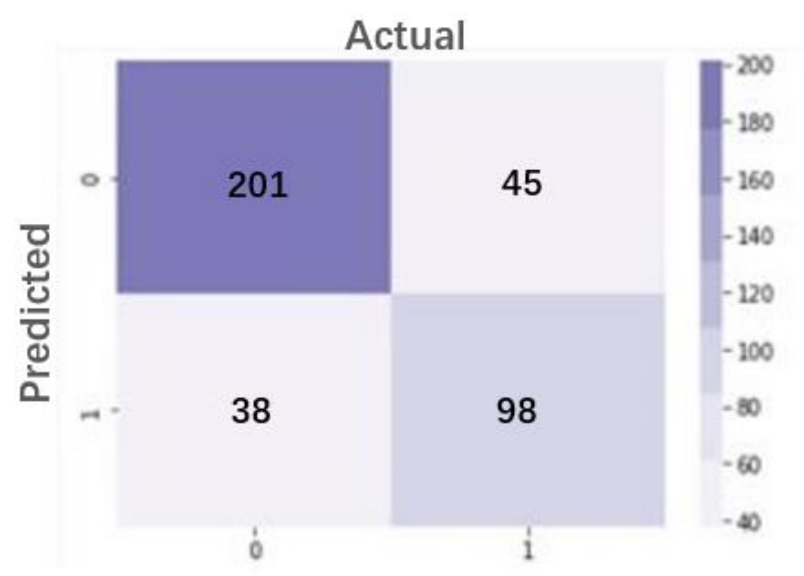


3.6 Random Forest

Importance of variables



Confusion matrix



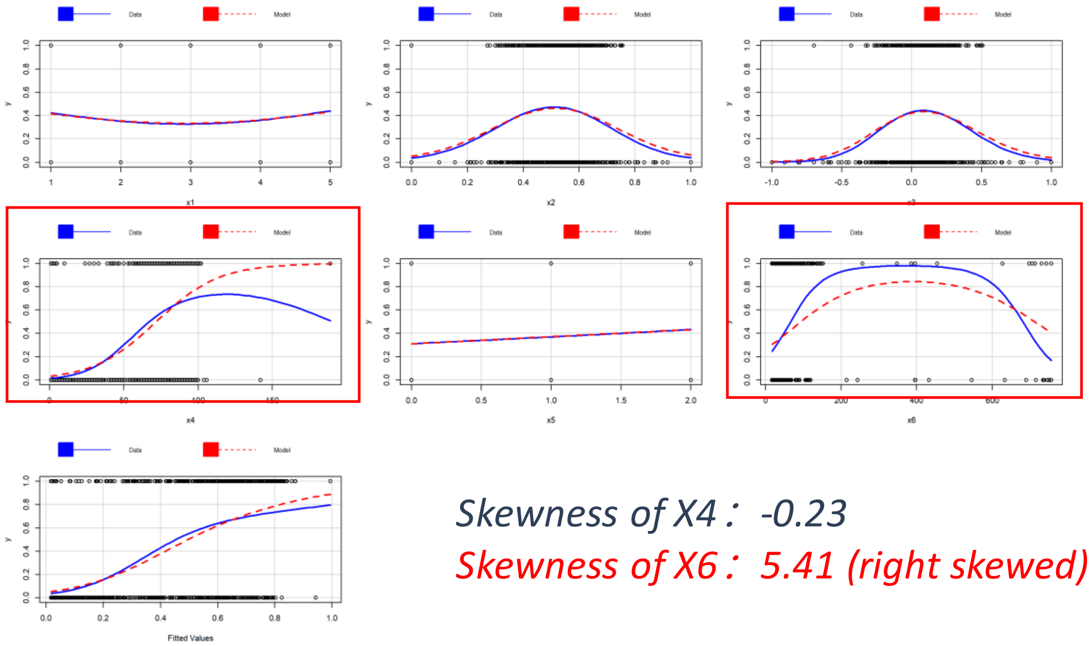
Accuracy: 0.783
Sensitivity: 0.685
Specificity: 0.841

3.7 Logistics Regression – Feature Selection

Model 1

$$\log\left(\frac{p}{1-p}\right) = -4.227 + 0.05 \times score + 0.4 \times subjectivity + 0.354 \times polarity + 0.048 \times length + 0.41 \times ABS - 0.001 \times Days\ of\ Comment$$

Variable	P-value
X1-score	0.463824
X2-subjectivity	0.527310
X3-polarity	0.475314
X4-length	< 2e-16 ***
X5-ABS	0.000102 ***
X6-Days of comment	0.292379



Skewness of X4： -0.23

Skewness of X6： 5.41 (right skewed)

Cook, R. D., & Weisberg, S. (1997). Graphics for assessing the adequacy of regression models. Journal of the American Statistical Association, 92(438), 490-499.

Fox, J. and Weisberg, S. (2019) An R Companion to Applied Regression, Third Edition. Sage.

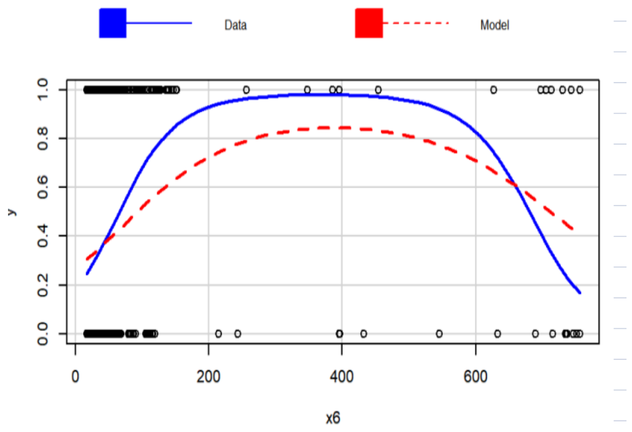
3.7 Logistics Regression – Feature Selection

Model 2

$$\log\left(\frac{p}{1-p}\right) = -9.092 - 0.012 \times score + 0.281 \times subjectivity + 0.466 \times polarity + 0.044 \times length + 0.521 \times ABS - 0.009 \times Days\ of\ Comment + 1.547 \times \log(Days\ of\ Comment)$$

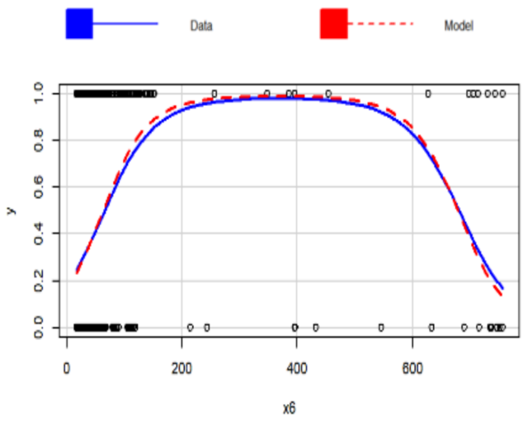
Variable	P-value
X1-score	0.862
X2-subjectivity	0.659
X3-polarity	0.363
X4-length	< 2e-16 ***
X5-ABS	2.54e-06 ***
X6-Days of comment	1.55e-08 ***
X7-Log(Days of comment)	1.32e-09 ***

Model 1



AIC：872.15

Model 2



AIC：834.84

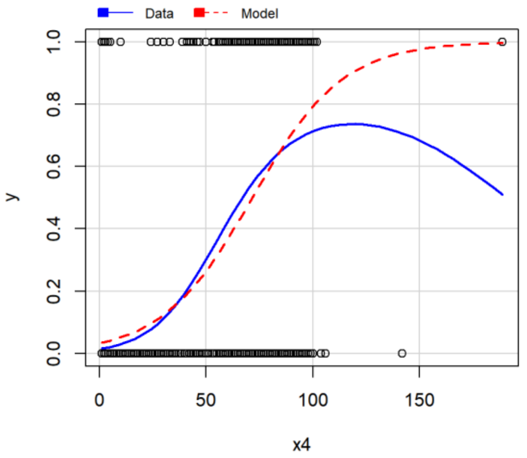
3.7 Logistics Regression – Feature Selection

Model 3

$$\log\left(\frac{p}{1-p}\right) = -9.386 + 0.082 \times score + 0.082 \times subjectivity + 0.504 \times polarity + 0.072 \times length + 0.519 \times ABS - 0.008 \times Days\ of\ Comment + 1.461 \times \log(Days\ of\ Comment) - 0.0002 \times length^2$$

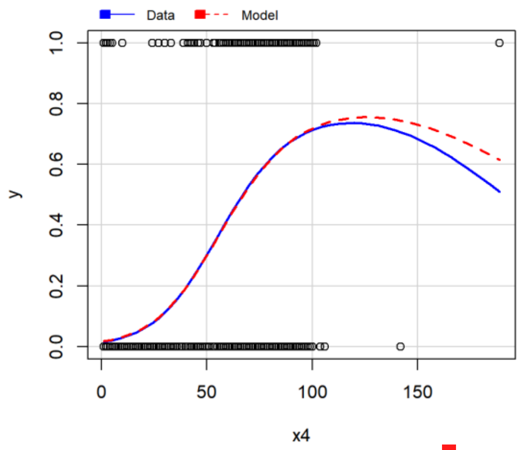
Variable	P-value
X1-score	0.88835
X2-subjectivity	0.90407
X3-polarity	0.34928
X4-length	1.38e-09 ***
X5-ABS	2.77e-06 ***
X6-Days of comment	9.07e-08 ***
X7-Log(Days of comment)	7.43e-09 ***
X8-length^2	0.00869 **

Model 2



AIC： 834.84

Model 3



AIC： 830.39

3.7 Logistics Regression – Feature Selection

Model 4

$$\log\left(\frac{p}{1-p}\right) = -9.368 + 0.072 \times length + 0.522 \times ABS - 0.008 \times Days\ of\ Comment + 1.464 \times \log(Days\ of\ Comment) - 0.0002 \times length^2$$

Variable	P-value
length	9.68e-10 ***
ABS	2.23e-06 ***
Days of comment	5.70e-08 ***
Log(Days of comment)	4.82e-09 ***
length^2Length^2	0.00757 **

Remove all non-significant variables in Model 3 with the step function to get Model 4

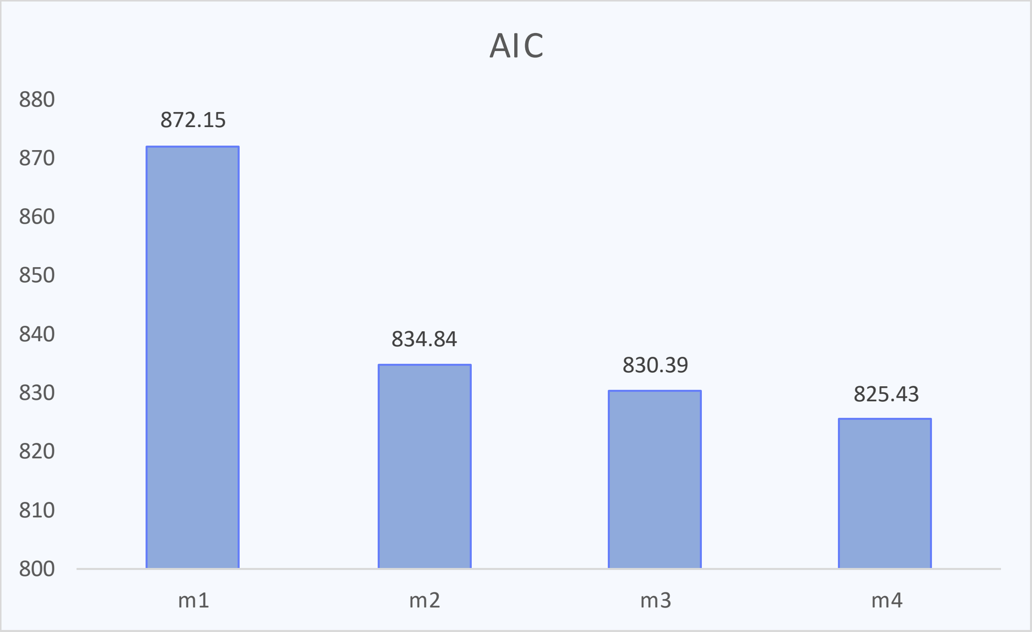
Model 3

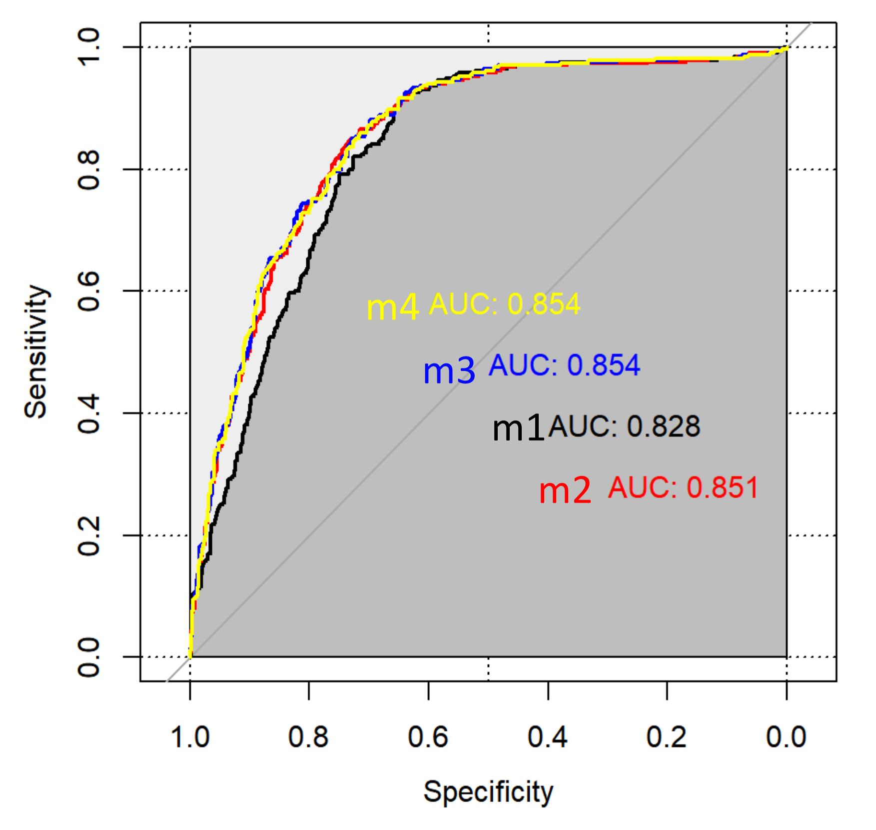
AIC： 830.39

Model 4

AIC： 825.43

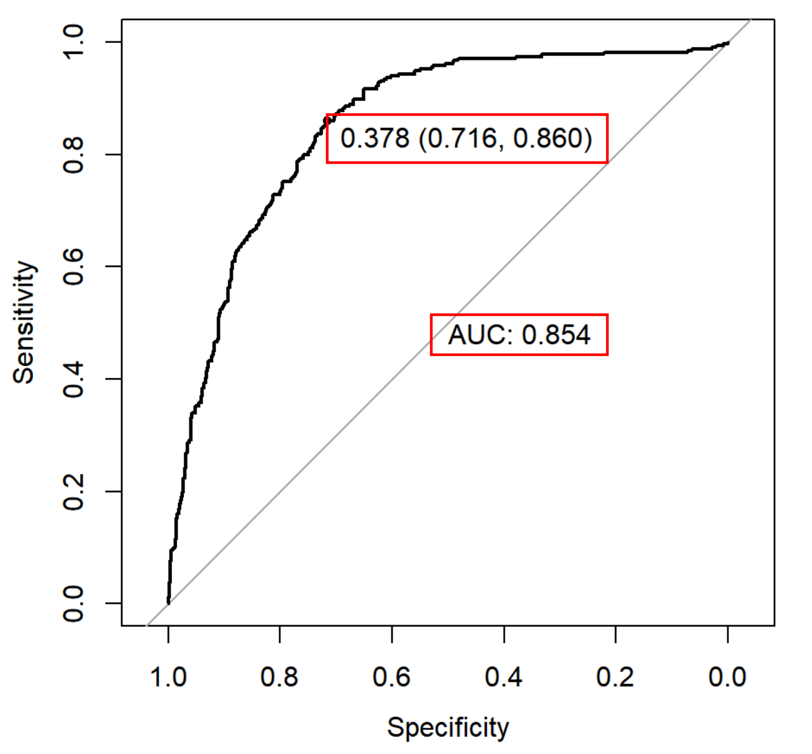
3.7 Logistics Regression – Model Comparison



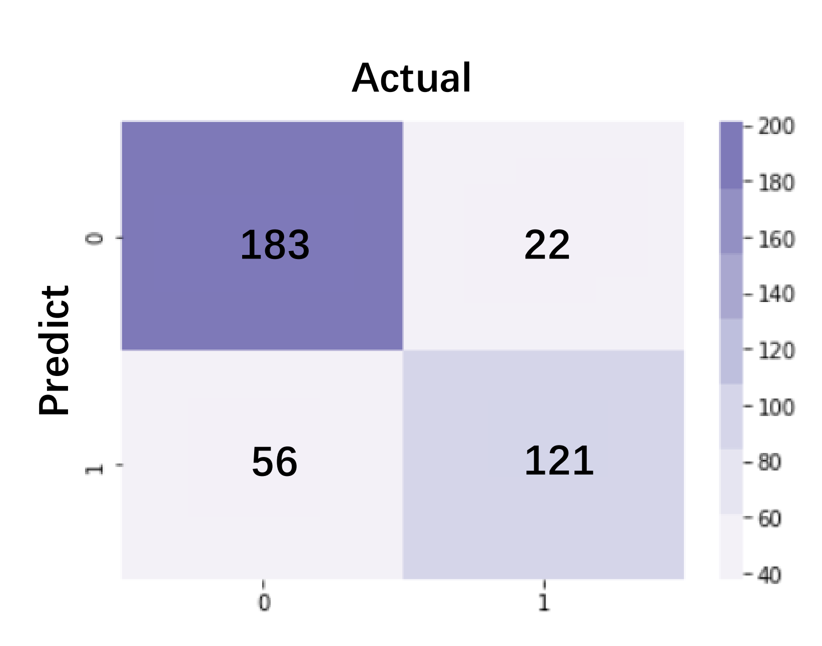


3.7 Logistics Regression – Model 4 Accuracy

ROC(train)



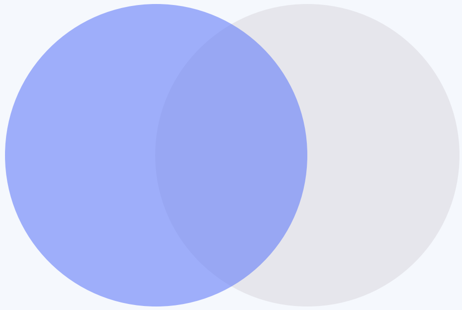
Confusion matrix (test)



Accuracy score: 0.796
Sensitivity: 0.846
Specificity: 0.766

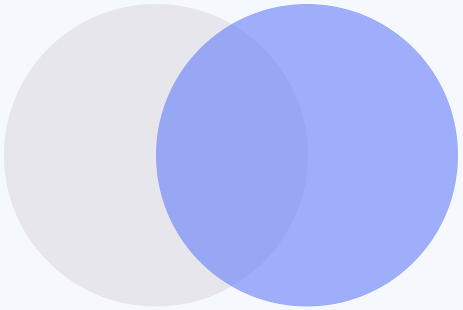
3.8 Aggregate the Classification Results

Logistic Regression



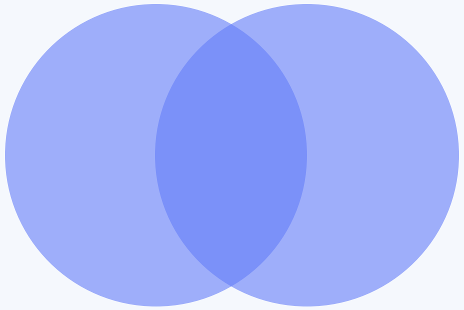
1-Recall = 22/(22+121) = 15.38%

Random Forest



1-Recall = 45/(45+98) = 31.47%

Union



1-Recall = 20/144 = 13.89%

The operation team regard missing the timing to intervene as more costly.

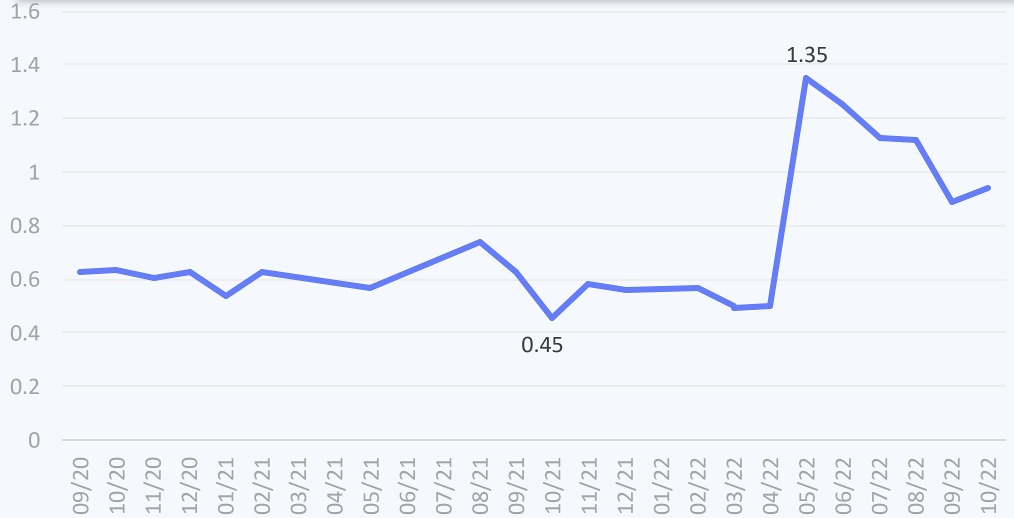
Cost function = (-50) * FN + (-10)* FP

3.9 Reputation Construction

$$Reputation = 0.3 * Score + 0.7 * Polarity$$

Variable	Range	Type
Reputation	**[-0.4, 2.2]**	**Decimal**
Polarity	[-1, 1]	Decimal
Score	[1, 5]	Integer

Monthly Reputation



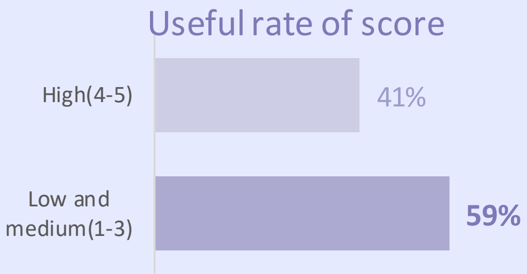
1. Why to use polarity?

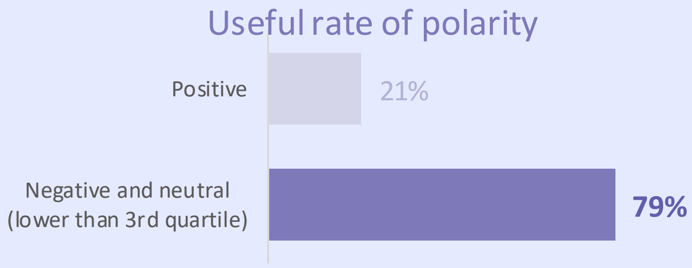
Reviews with high(low) score but negative(positive) sentiment

Score	Polarity	Content
5	-0.4	I was gonna play this game **but** now the download resource data is 18 GB **man**
5	-0.15	This game is very fun, has been playing it ever since the game release. The few **problem** I had is when I play in mobile it lag all the time. There's also an **issue** in the cut scenes. Hope mihoyo can fix these problems for all of us mobile user.
5	-0.13	**Update:** Issue resolved! **Sad** that they didn't address the mobile **issue** in their recent development notes. **Stuck in** the white screen of death.
1	0.78	Very **nice** size.
1	0.42	Happy 2nd anniversary, great rewards, exciting events especially the 10% only win welkin, **LOVE IT**!

2. Why for this weight?

Players tend to pay more attention to reviews with negative and neutral sentiment polarity.





3.10 Pre-Warning Model

Historical data of thumbsUpCount > 6

1. Calculate **R1** : the first quartile of reputations (unchanged for a version)

2. Set every two-week data as a window

3. Calculate **Bn** : the number of reviews whose reputation is lower than **R1** in **Window n**

4. Calculate **An** : the number of reviews in **Window n**

5. Calculate **Pn = Bn/An** : the percentage of the review number with low reputation in Windown

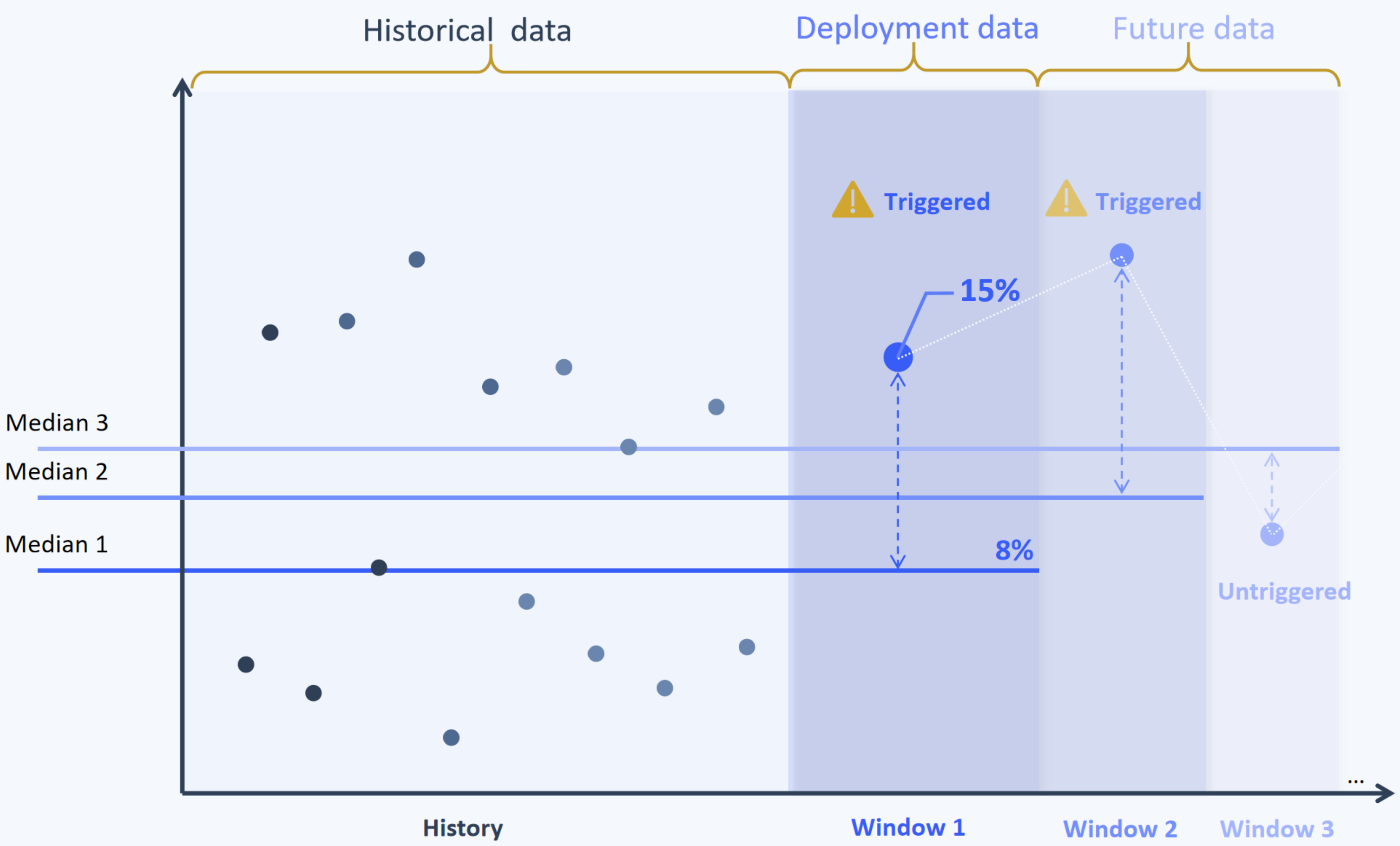
6. Set the median **M** of **P1, …, Pn** as the triggered line

Deployment data of model predicted

7. Calculate **P = B/A** in deployment data set(a window)

8. If **P > M**, a warning is triggered. Actions are required to prevent a reputation drop.

Comparison on percentage of the useful reviews with low reputation



Beyond the Players: Social Reputation

Genshin Game operation recommendations based on Time Series analysis of Google Trend

1. Related Variables
2. Google Trend Time Series Analysis, Modelling, and Forecast

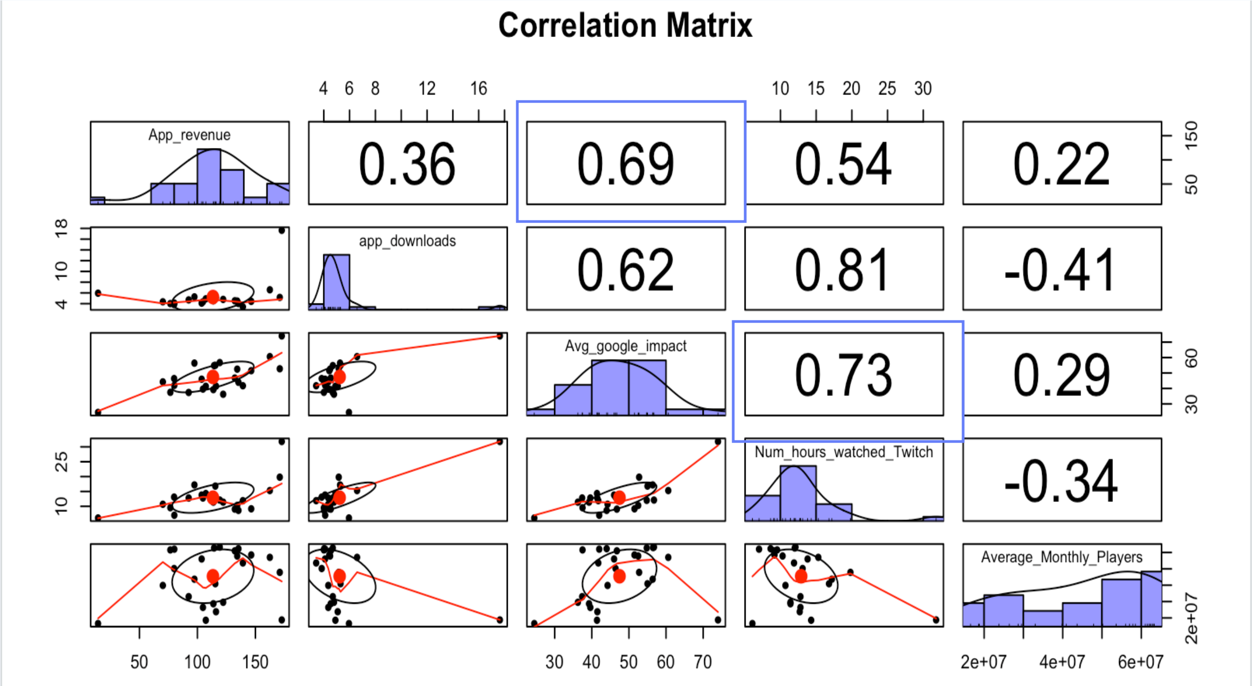
3.11 Correlation Matrix

Why use Google Trend?

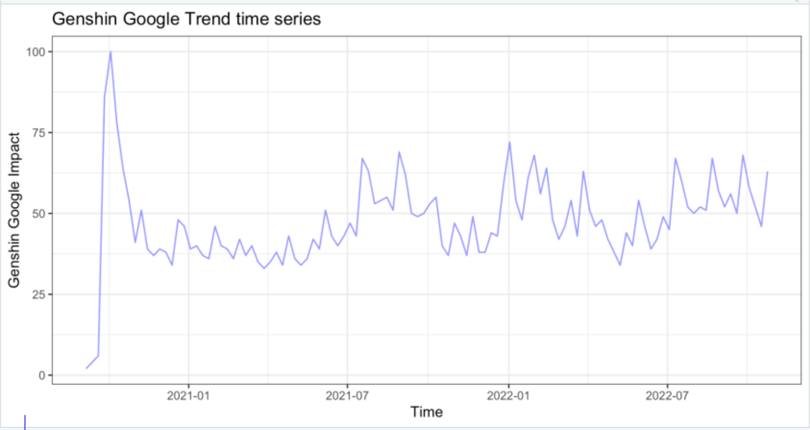
We use Google trends to figure out Genshin's popularity in society.

What may be related to Google trend?

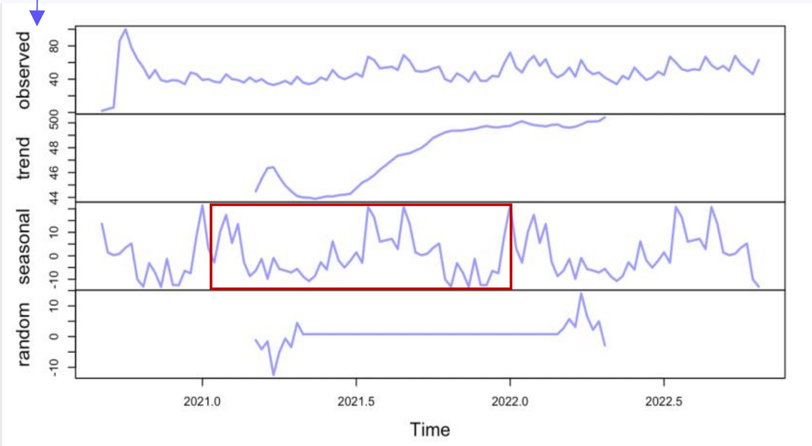
From correlation matrix, we can find Google trend has a relatively strong and positive correlation with App Revenue and the number of hours players watch Twitch.



3.12 Time Series Analysis of Google Trend



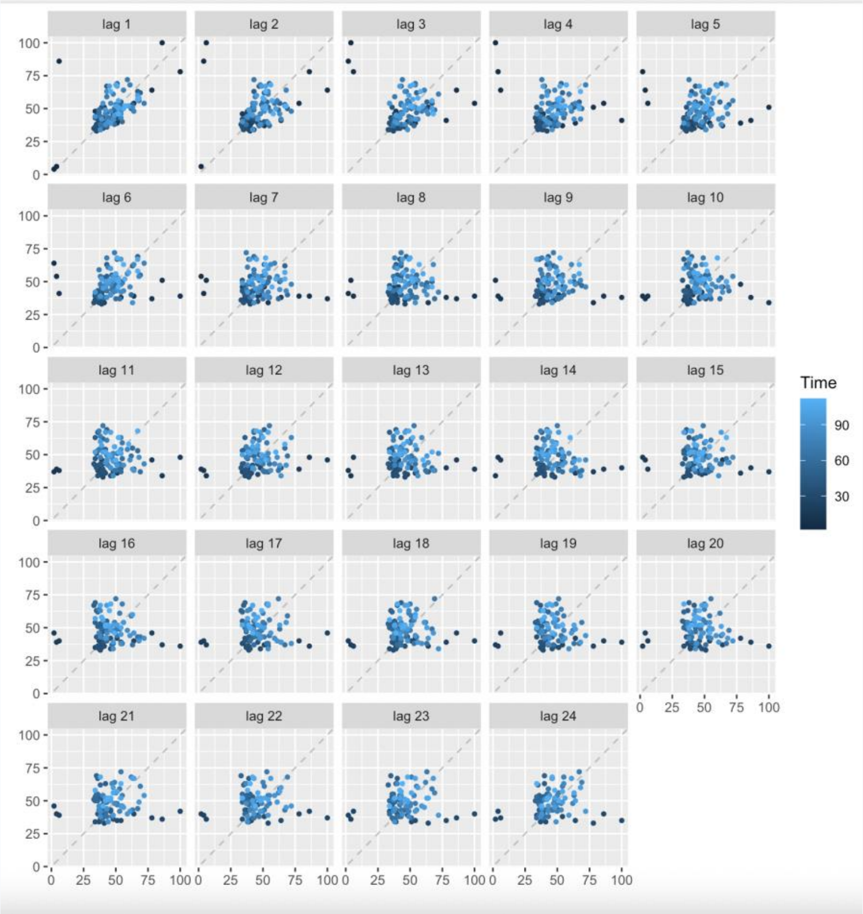
Decomposition: addictive



- Increasing trend
- Seems to be some seasonality in terms of years

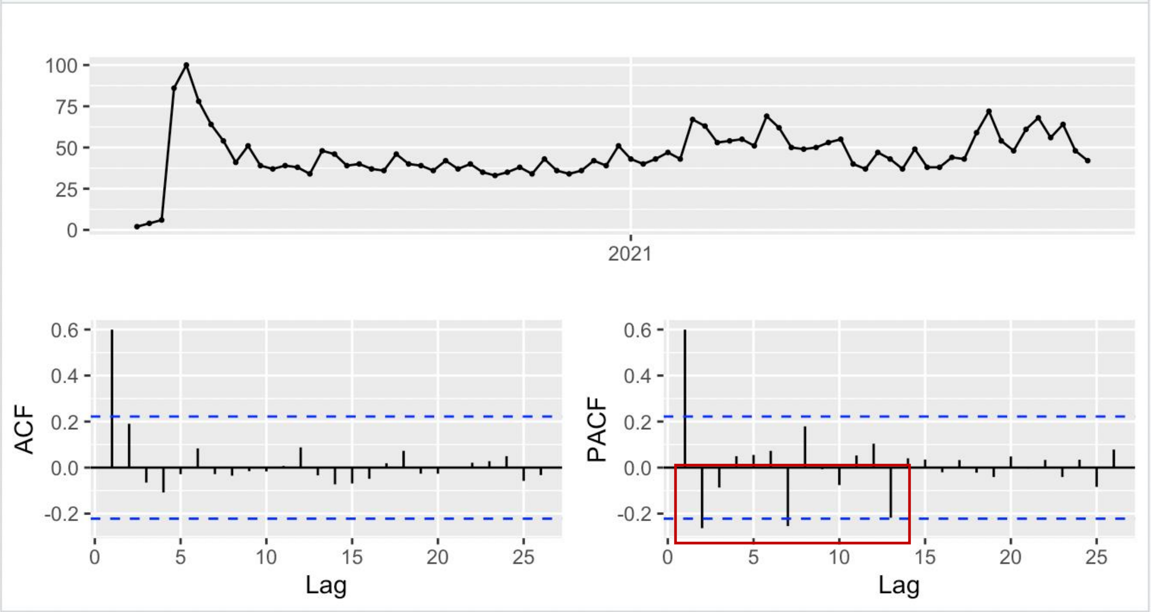
Lag plot

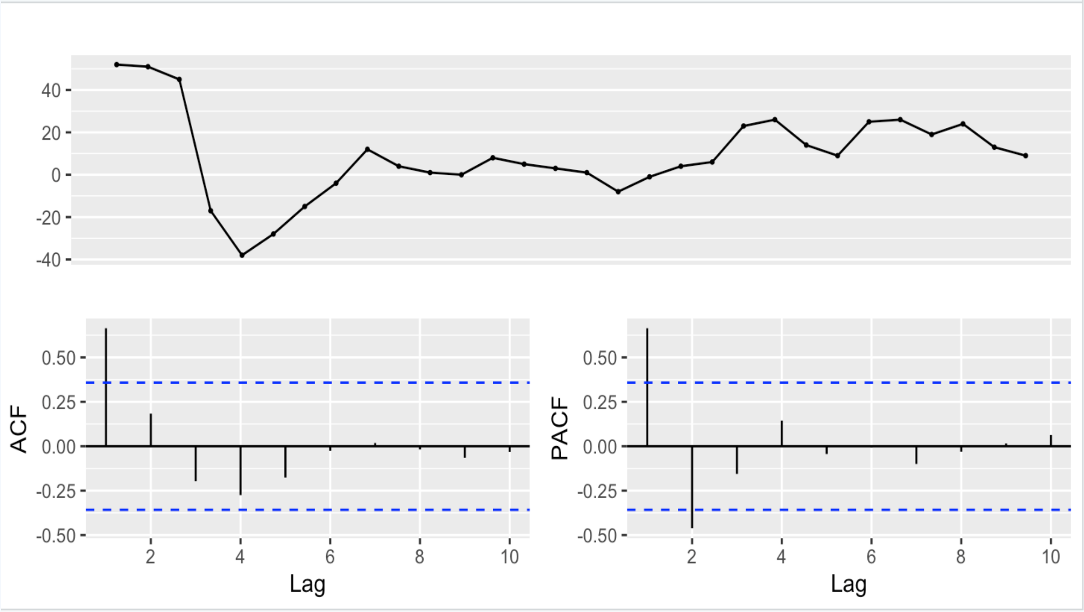
- shows the time series is random
- Small autocorrelation present in the data



3.12 Time Series Analysis of Google Trend

Here, we divide the data into training and testing and plot a time series for Google Impact along with its ACF and PACF,
All data-- training (70%) : testing (30%)





It shows a little seasonality, but it's not obvious. It might become obvious with more data. So, a seasonal difference analysis is added for further analysis.

3.13 ARIMA Model for Google Trend

Small

Model complexity

Large

Complexity	ARIMA Model	AIC	RMSE (training)	RMSE (testing)	p-value (Ljung-Box test)
1 variable	(1,0,0)(0,0,0)	610.65	11.626	8.701	0.0962
1 variable	(0,0,1)(0,0,0)	614.93	11.96	9.647	0.0385
2 variables	(1,0,0)(1,0,0)	608.69	10.98	**8.333**	0.1978
2 variables	(0,0,1)(1,0,0)	614.97	11.622	9.494	0.0387
2 variables	(1,0,1)(0,0,0)	608.17	11.290	9.466	0.0360
2 variables	(1,0,1)(0,1,0)	**245.3**	**7.906**	NA	0.1123
2 variables	(2,0,0)(0,0,0)	607.41	11.235	9.484	0.0361
3 variables	(1,0,1)(1,0,0)	605.28	10.480	9.063	0.1807
3 variables	(2,0,1)(0,0,0)	608.82	11.190	9.447	0.0280

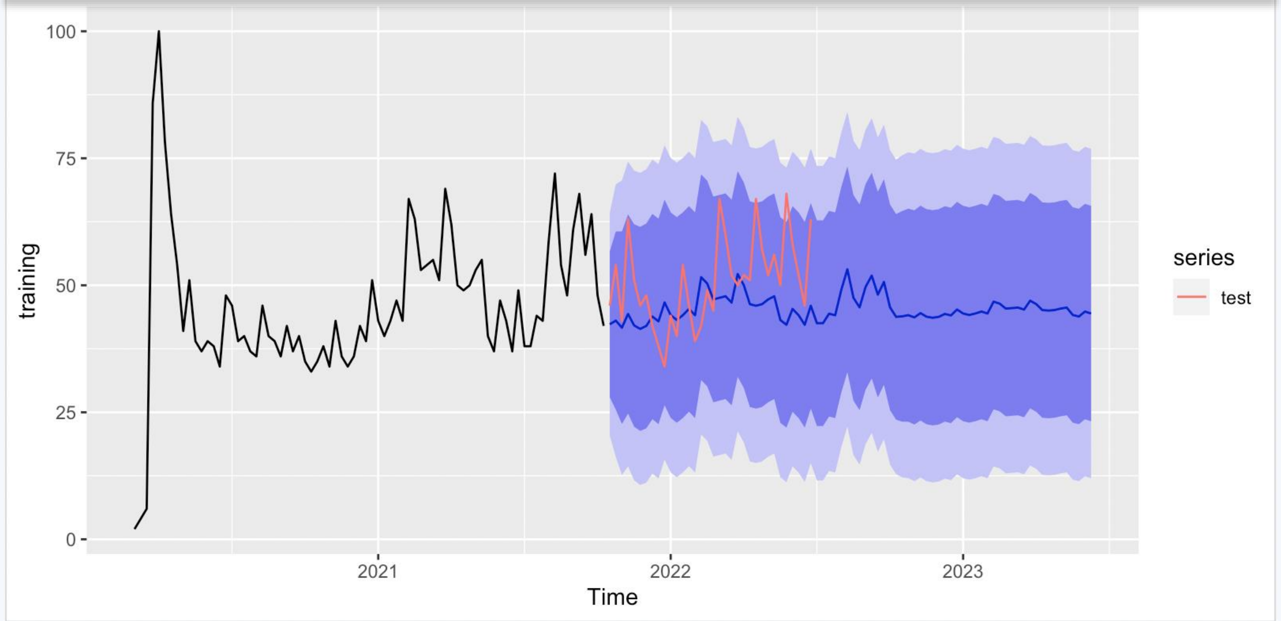
Principles of model selection
1. Small AIC and test RMSE
2. Less complexity model

1. Conducted seasonal difference, but data were too small to be tested
2. No need to do seasonal difference.

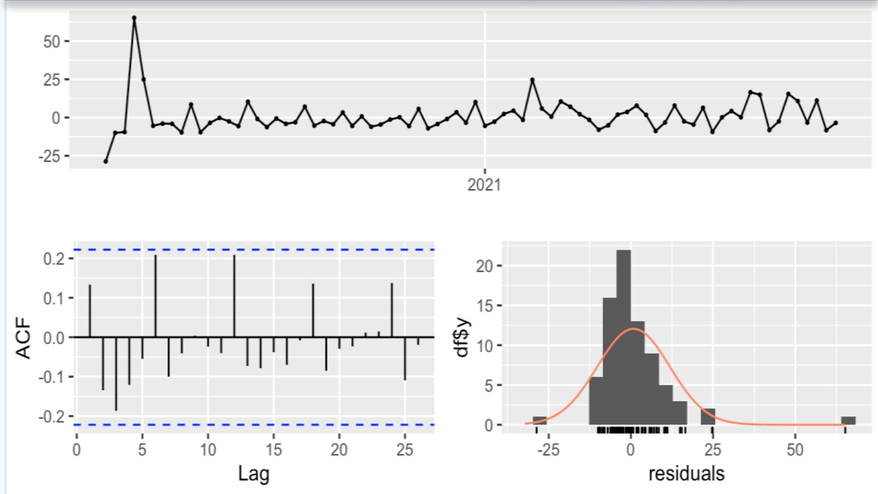
Conclusion: Although ARIMA(1,0,0)(1,0,0) result is not the best in training data, it is the best in testing and the model complexity is small.

3.14 Google Trends Forecast Results

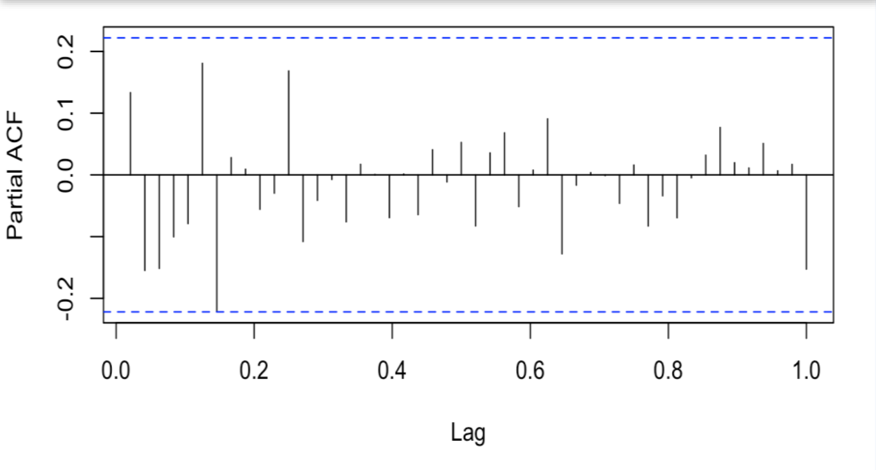
Forecasts from ARIMA(1,0,0)(1,0,0)



Residuals from ARIMA(1,0,0)(1,0,0)



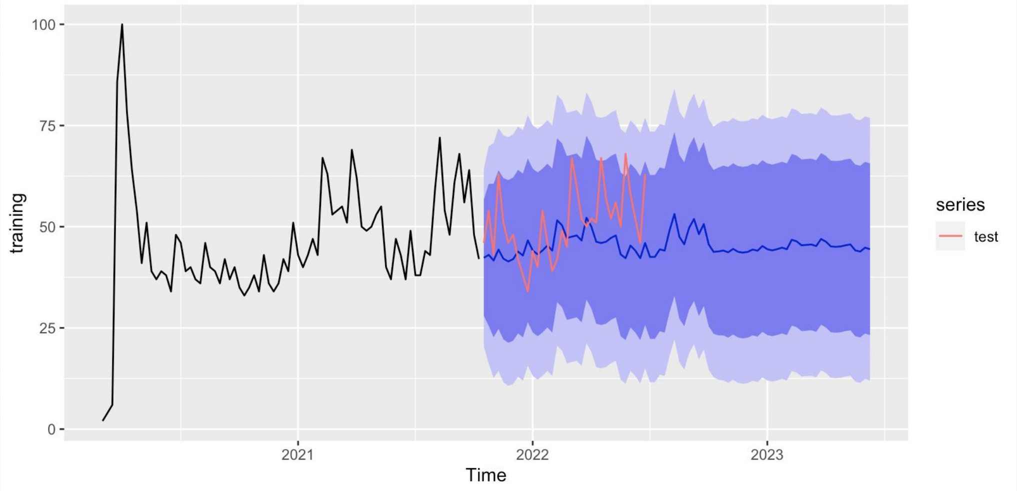
PACF of residuals



- The residuals distributed evenly around 0.
- From the ACF and PACF, confirming that the data are white noise.
- Residual plots is relatively normal distributed.

3.14 Google Trends Forecast Results

Forecasts from ARIMA(1,0,0)(1,0,0)



Google Trend

- Definition: Searching popularity
- When being widely discussed/searched
  —> the game is regarded as more popular

Recommendations

When the operation team find that the whole society is paying high attention to the game, it should be a good time for the game to attract new players.

4.1 Suggestion Summary

Prevent Reputation Drop

What is complained by the players?

System bug, time (service response and some quests are too long), occupy a large amount of storage

Word Cloud

Derive Reference for Developing Games

What features are appreciated by the players?

Elaborate design of characters and storylines

Word Cloud

Player reputation

Update the auto-reply corpus

Topics about Graphic(45%), story design(32%), character design(13%), technical problem(10%);

Positive(55%) and Negative(45%) Sentiment

LDA

When to take maintenance actions?

The percentage of useful review with low reputation within a window is higher than that of the median level in history

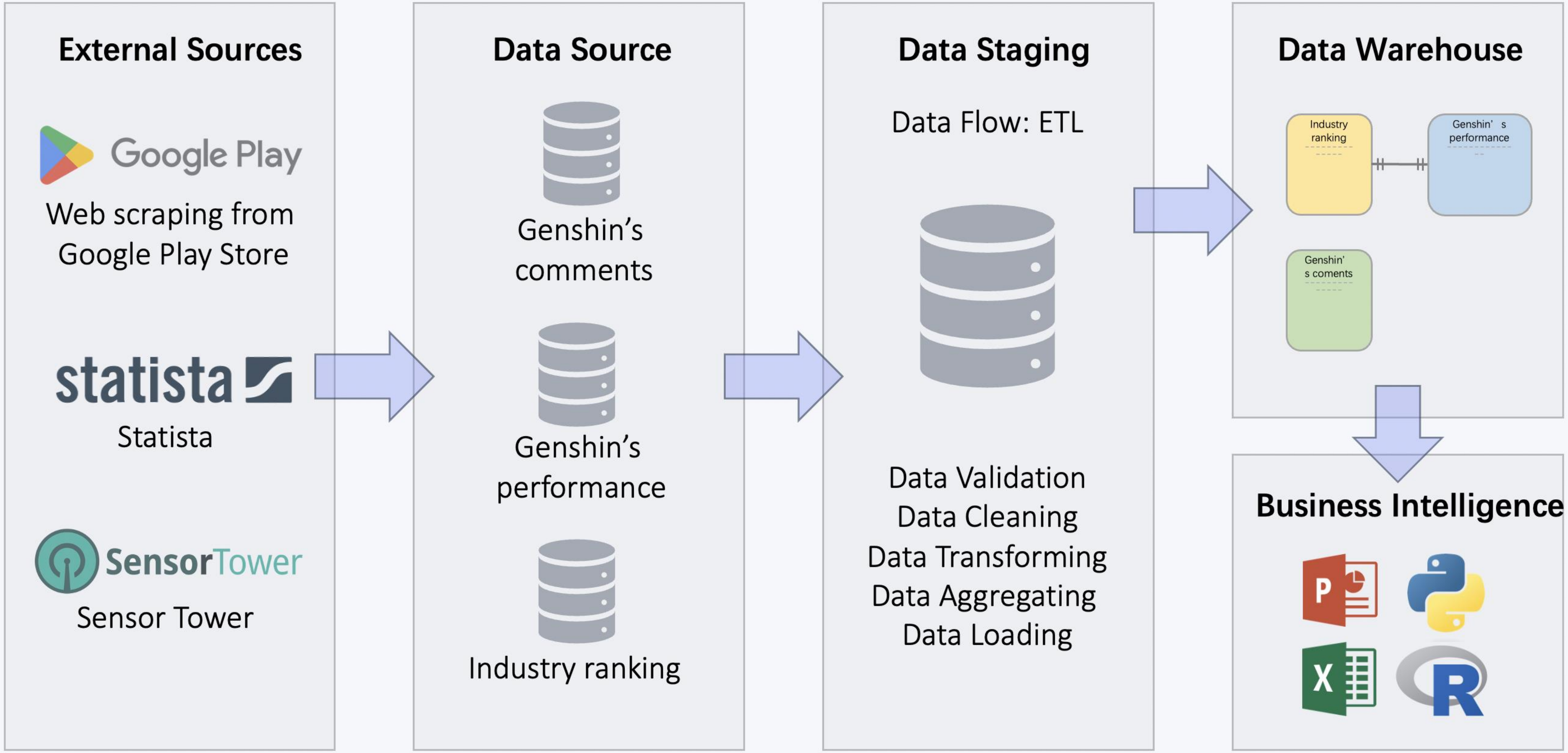
Pre-warning model

Social reputation

When the whole society is paying high attention to the game, it should be a good time for the game to attract new players

Google trend

4.2 Data Management



References

Cook, R. D., & Weisberg, S. (1997). Graphics for assessing the adequacy of regression models. Journal of the American Statistical Association, 92(438), 490-499.

Fox, J. and Weisberg, S. (2019) An R Companion to Applied Regression, Third Edition. Sage.

Weisberg, S. (2005) Applied Linear Regression, Third Edition, Wiley, Section 8.4.

Thank you.