IMPERIAL

App Store Applications Reviews Analysis

How Reviews Help Developers Improve Their Products

Baoqi Zhang, Runyu Cai, Yifan Jiang, Jiahui Chen, Yaoguo Zhong
March 7, 2025

Contents

Population of Interest:

App users who leave reviews.

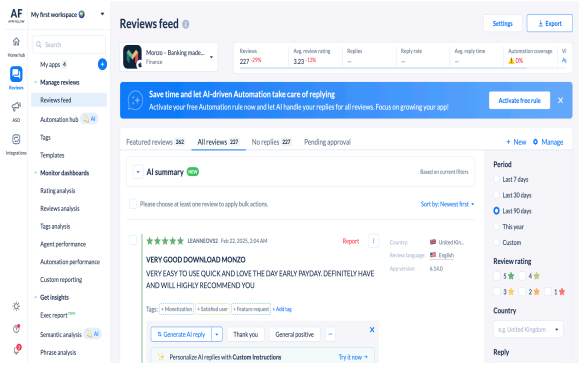
App developers who regularly check reviews.

Quantities of Interest:

- App rating and Categories Predictions
- Common topics in user reviews

Data Introduction

- Data source website: AppFollow
- We explore 10 different categories of apps: Finance, Games, Food Delivery, Dating, Entertainment, Shopping, Productivity, Social Media, Travel, and Music
- Each category contains information on 5–10 different apps
- Each document has about 30,000–50,000 data points, forming a powerful dataset for analysis

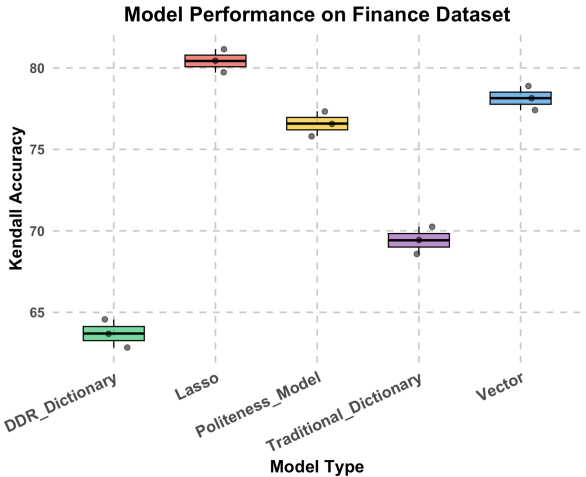


Lasso and Benchmarks Comparison

Financial Apps Dataset Only

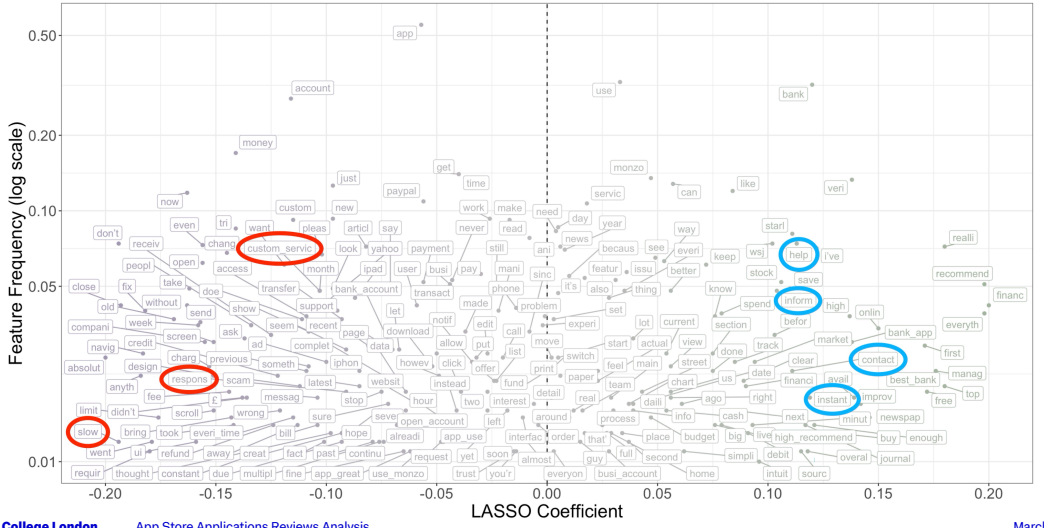
Key Insights:

- Use LASSO regression as the main model and compare it to four benchmark models
- LASSO regression achieves the best performance
- LASSO-based sentiment analysis will be solid in our future research



Lasso Model

Coefficient Plot



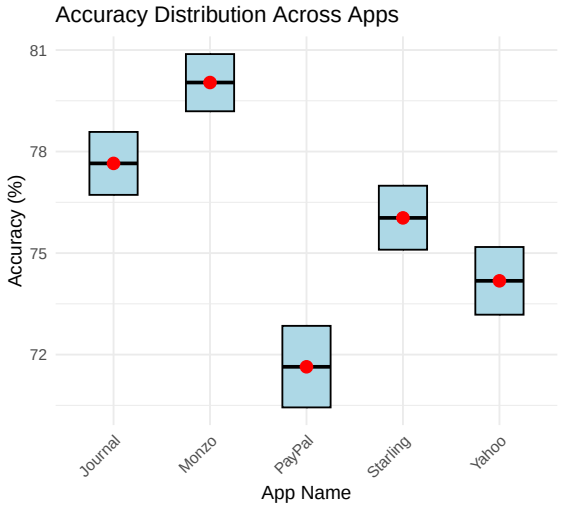
Transfer Learning:
Rolling Predictions

Key Insights:

- **App-Level Predictions:** prediction accuracy differs across apps
- **Accuracy:** Generally high at the app level

Conclusion:

- Transfer learning helps predict app ratings
- Providing a solid logical foundation for further research

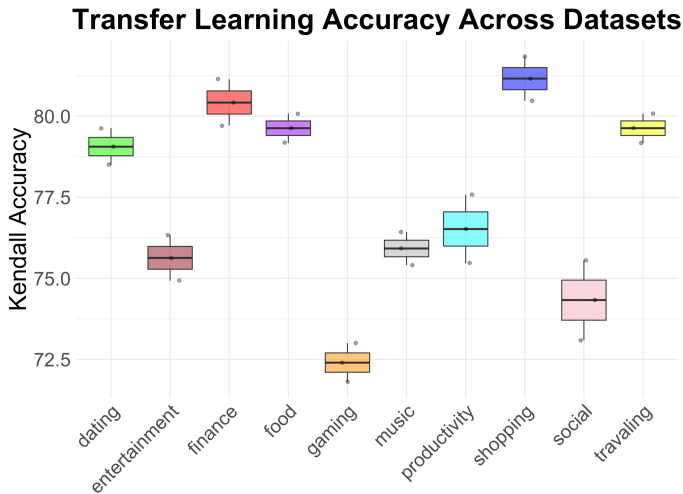


Transfer Learning Performance on App Ratings

Transfer Learning:
Use financial Reviews to Predict Other Categories

Key Insights

- **Categorical Level:** 10 categories analyzed
- **Target Categories** Dating, Entertainment, Finance, Food, Gaming, Music, Productivity, Shopping, Social, Traveling.
- **Accuracy:**Overall acceptable, but variations exist.



(Above) Topic Correlation with Star Ratings

Transfer Learning:
Improving with Game Data

Key Insights

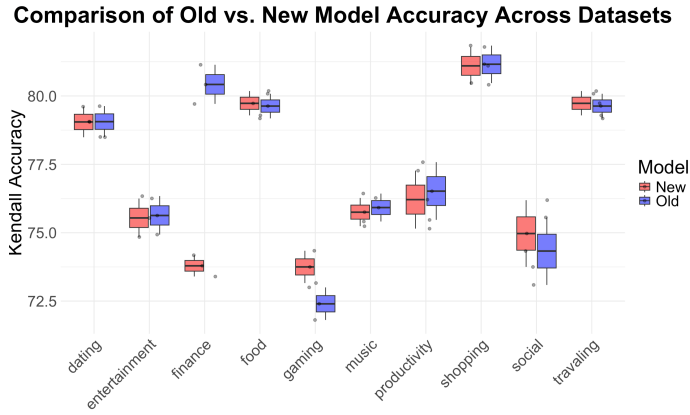
- Game accuracy particularly low
- Improve by add more reviews to training setSocial, Traveling.
- More reviews added, Higher the accuracy

(Above) Topic Correlation with Star Ratings

Transfer Learning:
Improved Model

Key Insights

- Trade-off in Transfer Learning
- Domain-Specific Knowledge Gets Weakened
- Structured Approach Required

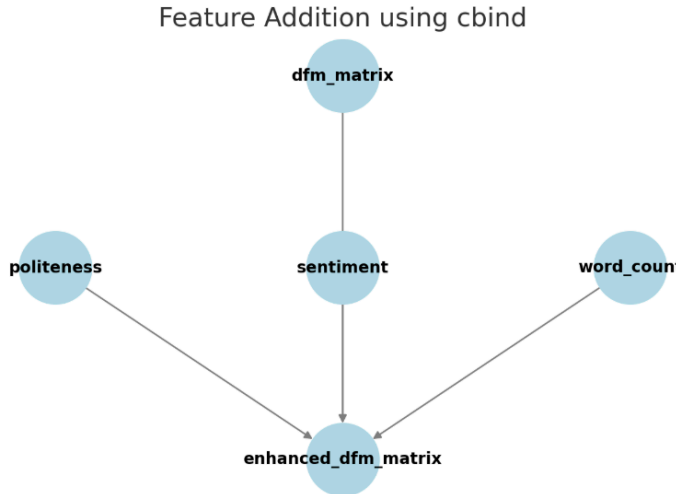


(Above) Topic Correlation with Star Ratings

Pre-processing
Detailed Steps

Pre-processing Steps:

- Reviews often contain long sentences; we use trigrams, stopwords, tokenization, and word stemming.
- A minimum token proportion of 0.001 is set to avoid dropping too many words.
- For enhanced sentiment analysis, we apply negation review and lemmatisation (treating words following negations as having opposite sentiment).
- Additional features (dictionaries, politeness scores, and sentiment scores) are added using `cbind()` to boost model accuracy.



(Above) Feature Addition using `cbind()`

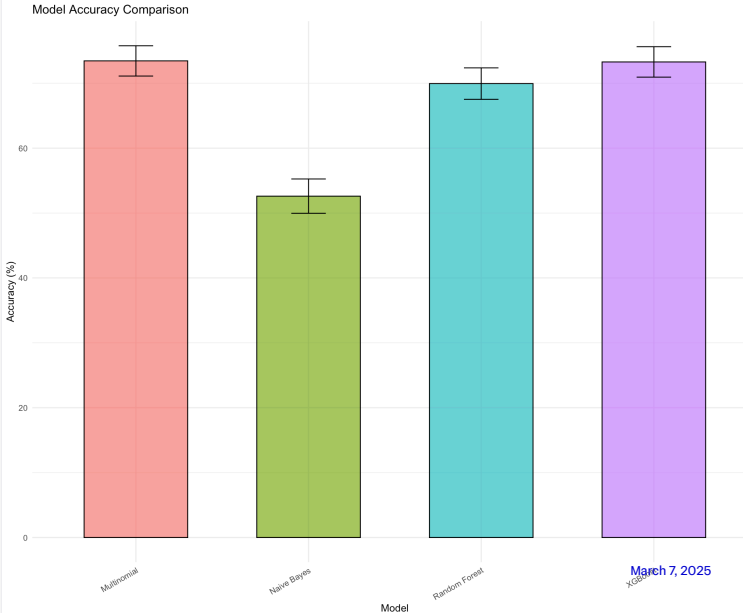
Multinomial Model

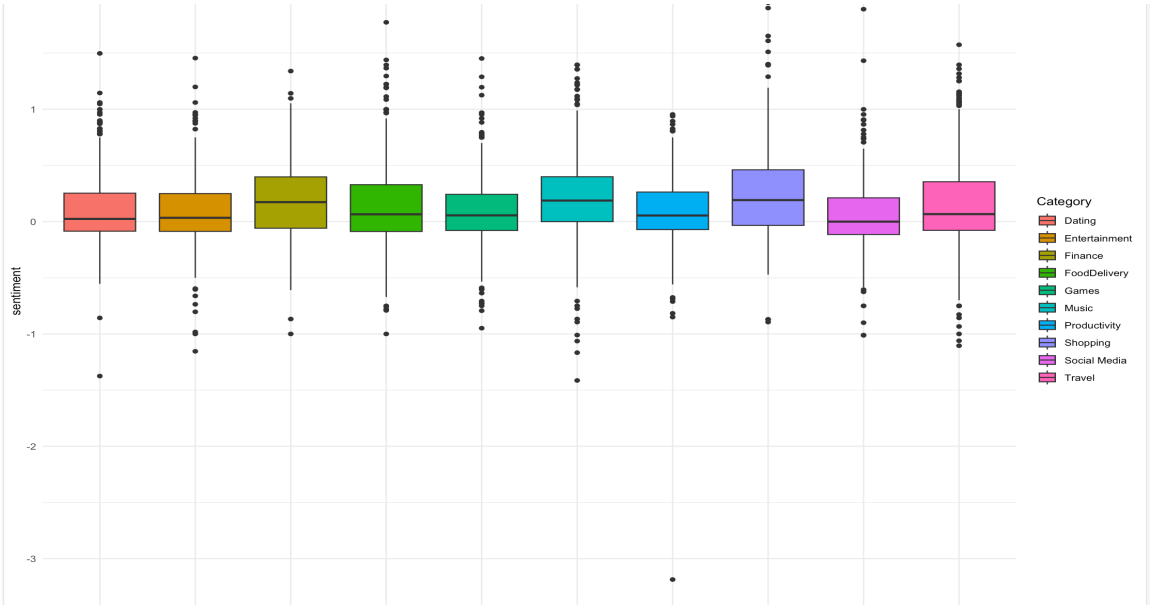
10 Categories Model

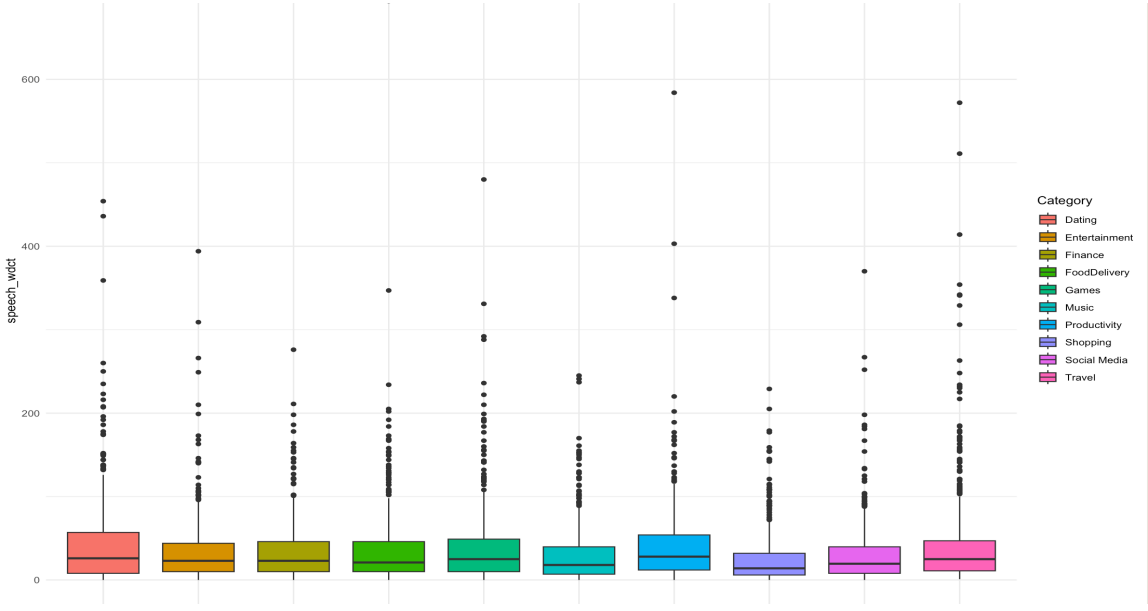
Model Comparison: Multinomial vs. Other Classifiers

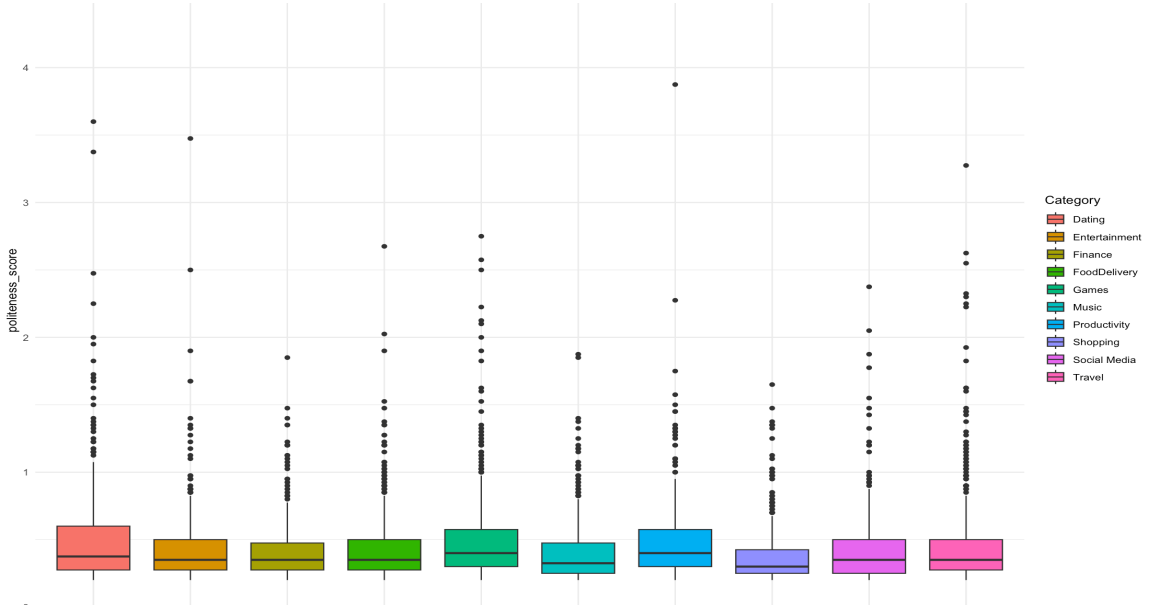
Key Findings:

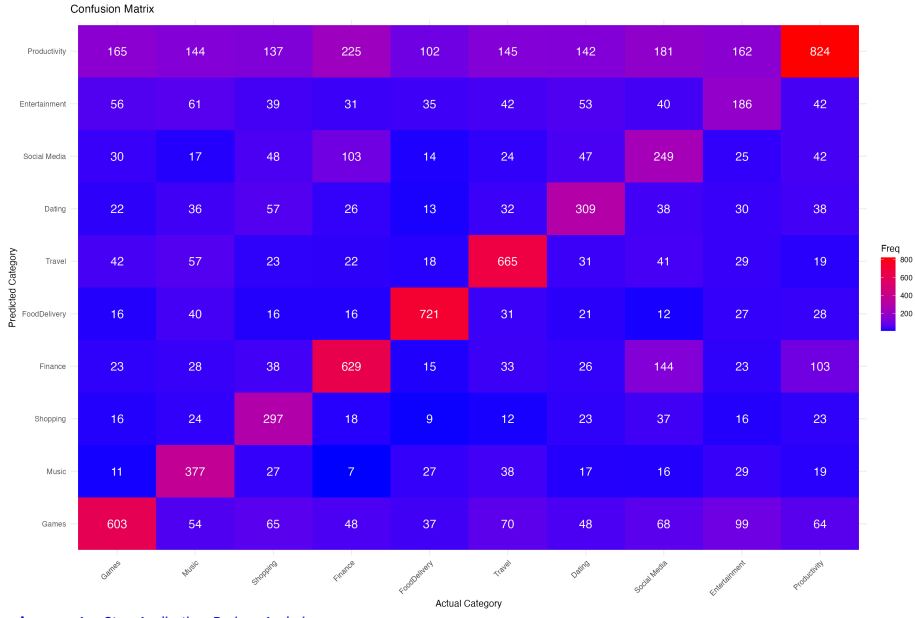
- **Multinomial Model** achieved the highest accuracy, making it effective for review classification.
- **XGBoost** performed similarly well, benefiting from its ability to capture non-linear patterns.
- **Random Forest** showed moderate performance, outperforming Naïve Bayes.
- **Naïve Bayes** had the lowest accuracy due to its assumption of word independence.




sentiment
1
0
-1
-2
-3
Category
Dating
Entertainment
Finance
FoodDelivery
Games
Music
Productivity
Shopping
Social Media
Travel


600
400
200
0
speech_wdct
Category
Dating
Entertainment
Finance
FoodDelivery
Games
Music
Productivity
Shopping
Social Media
Travel


politeness_score
4
3
2
1
Category
Dating
Entertainment
Finance
FoodDelivery
Games
Music
Productivity
Shopping
Social Media
Travel


Confusion Matrix
Productivity 165 144 137 225 102 145 142 181 162 824
Entertainment 56 61 39 31 35 42 53 40 186 42
Social Media 30 17 48 103 14 24 47 249 25 42
Dating 22 36 57 26 13 32 309 38 30 38
Travel 42 57 23 22 18 665 31 41 29 19
FoodDelivery 16 40 16 16 721 31 21 12 27 28
Finance 23 28 38 629 15 33 26 144 23 103
Shopping 16 24 297 18 9 12 23 37 16 23
Music 11 377 27 7 27 38 17 16 29 19
Games 603 54 65 48 37 70 48 68 99 64
Predicted Category
Games
Music
Shopping
Finance
FoodDelivery
Travel
Dating
Social Media
Entertainment
Productivity
Actual Category
Freq
800
600
400
200

Topic Modelling
Refining Topic Labels

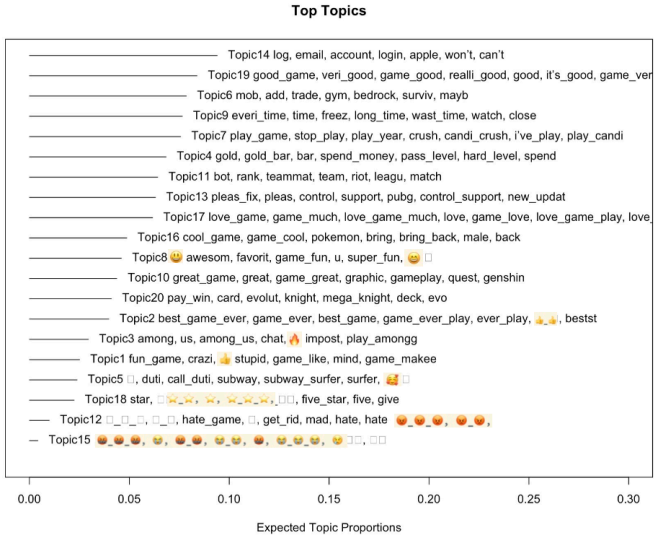
Key Insights

Topic Identification:

- Extracted key themes from app reviews.
- Topics highlight user experiences and concerns.

Redefining Topic Names:

- Original labels lacked clarity.
- Renamed topics based on keyword patterns.
- Some topics included **emoji artifacts**, Exaggerate the emotions of reviews.



(Above) Topic Correlation with Star Ratings


made
game
surfer
life
call
subway
subway_surfer
happi
run
call_duti
like
real
duti
hope
die
amaze
feel
gotta
love_subway
wonder_game
bye
amazing
lol
thank
made_game
mobil
train
kid
christma
love_candi_crush
love_candi
love_subway_surfer
call_duti_mobil
duti_mobil
game_made
real_life
game_hope
amaze_game
play_subway_surfer
battl_royal
wonder
better
multiplay
wow
jump
fortnit
guy
i'm
action
man
never
gonna
depress
develop
season
joy
track
dream
becaus
polic
theme
teach
peac
feel_like
true
omg
son
you'r
know
also
i'm_happi
keep
eat
gamer
boy
movi
just
play_subway
glad


us
among
among_us
chat
say
game
impost
like
peopl
quick
free
name
want
pls
just
ban
fire
bean
becaus
lobbi
kill
hack
role
play
someon
sorri
got
guy
report
quick_chat
give_us
word
also
said
wanna
everyon
crewmat
talk
thing
make
see
use
child
crew
room
task
hide
map
seek
get
ask
onli
love_among
love_among_us
read
think
person
mean
gonna
thank
cuz
tell
play_among_us
type
inappropri
can
random
hacker
yall
stuff
impostor
lie
swear
innersloth
better
free_chat
one
play_among
peopl_say
don't
now
instead
i'm
vote
yeah
someth
parent


pokemon
game
pokémon
catch
right
just
event
bring
niantic
come
back
charact
now
go
cool
new
cool_game
bring_back
game_cool
one
day
old
like
even
sinc
actual
year
player
right_now
natlan
femal
away
becaus
feel
took
want
start
rare
yet
chang
come_back
use
far
make
servic
around
batl
male_charact
origin
now_it'
care
ticket
complet
shini
everyth
month
went
white
region
sad
alreadi
onli
walk
compani
half
also
first
get
look
see
sound
disappoint
behind
girl
came
feel_like
fan_servic
turn
still
releas
pokemon_go
got
back_game
male
communiti
i'm
back_old
fan
dynamax
gone
go_back
music
miss
listen
effect
playabl

Topic Modelling

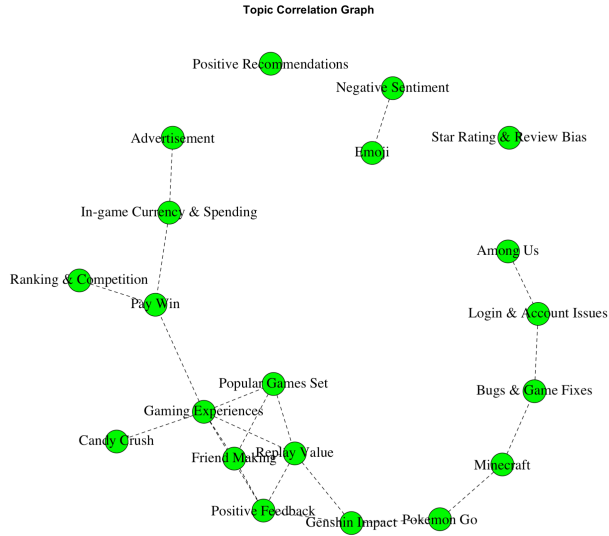
Topic Relationships and Correlations

Topic Clusters:

- **Game Reviews, Replay Value, Friend Making** are highly interconnected.
- **Pay-to-Win** links to **Ranking**, showing fairness concerns.

Isolated Topics:

- Positive Recommendation Picks and Star Ratings are standalone.
- Negative Sentiment interacts with Emoji.



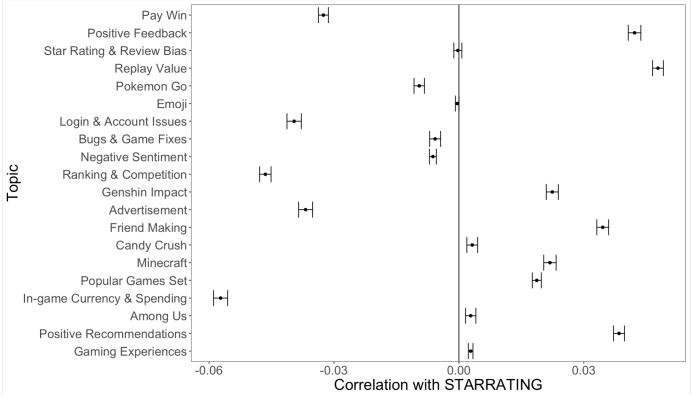
(Above) Topic Correlation Network

Topic Modelling

Topic Correlations with Ratings

Key Insights

- Some topics strongly correlate with **Star Ratings**.
- **Pay-to-Win, Good Game Reviews** show notable effects.
- **Negative impact:** Bug Fixes, Login Issues.
- **Positive impact:** Replay Value, Graphics.
- Some topics show weak influence on ratings.



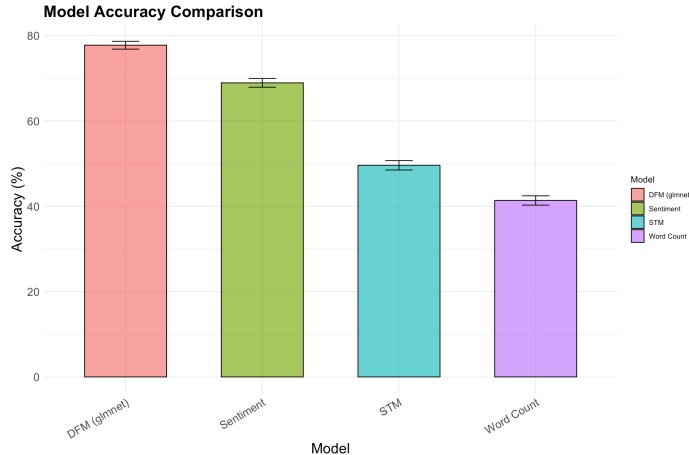
(Above) Topic Correlation with Star Ratings

Topic Modelling

Model Accuracy Comparison

Key Insights

- **DFM (glmnet)** achieves the highest accuracy.
- **Sentiment model** performs well but slightly lower.
- **STM shows moderate accuracy**, indicating topic structure alone is less predictive.
- **Word Count** is the weakest predictor.



(Above) Topic Correlation with Star Ratings

Model Limitations

Challenges and Constraints

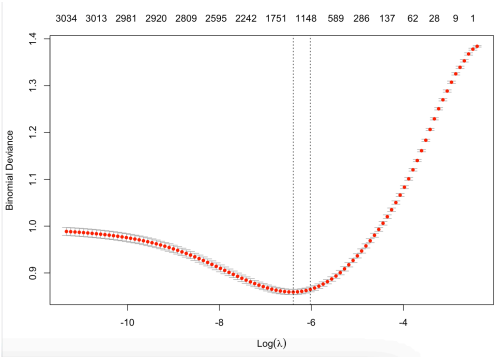
Models Limitations:

- **Lacks interpretability** – Lasso selects words but does not explain their impact.
- **Ignores context** – Lasso fails to capture negations or word dependencies.
- **Feature dropping** – Penalization removes useful correlated features.
- **Lower Predictive Accuracy** – Topic models prioritize theme extraction over rating prediction.

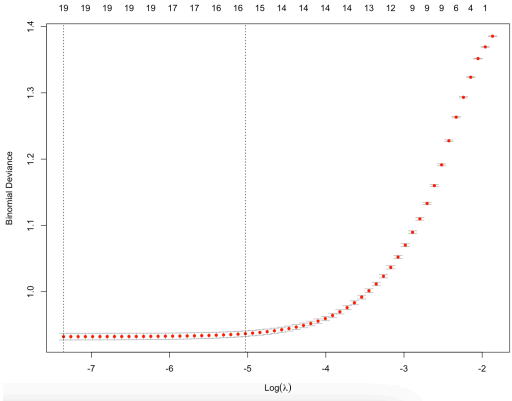
General Limitations:

- **Noisy data** – mixed positive and negative feedback in reviews.
- **Rating biases** – external factors (e.g., updates, promotions) influence scores.
- **Temporal shifts** – app changes over time but model relies on historical data.
- **Sentiment complexity** – sarcasm and nuanced expressions mislead models.
- **Data** – Only app store reviews may not reflect all aspects of users. Should combine with other platforms such as google store and etc.

Appendix



N-Grams Model



Topic Model

IMPERIAL

Thank you.
Questions?

App Store Applications Reviews Analysis
March 7, 2025