

DSO 560 NLP

Google Play Store Apps Reviews Analysis

Feiran Zhang, Ge Zeng, Ranyue Yang, Xinyue Yu, Xiaoying Zou

Overview

Business Problem

Dataset Review

Preprocessing

Classification Modeling

Deep Learning Models

Feature Importance

Problematic Reviews

Topic Modeling

Business Insights from Topic Modeling

Future Improvement

Contents

1

Overview

Business Problem - Dataset Review - Preprocessing



Overview – Business Problem

1

How to classify a review as good/poor only based on the text?
Which features matter?

Classification
Modeling

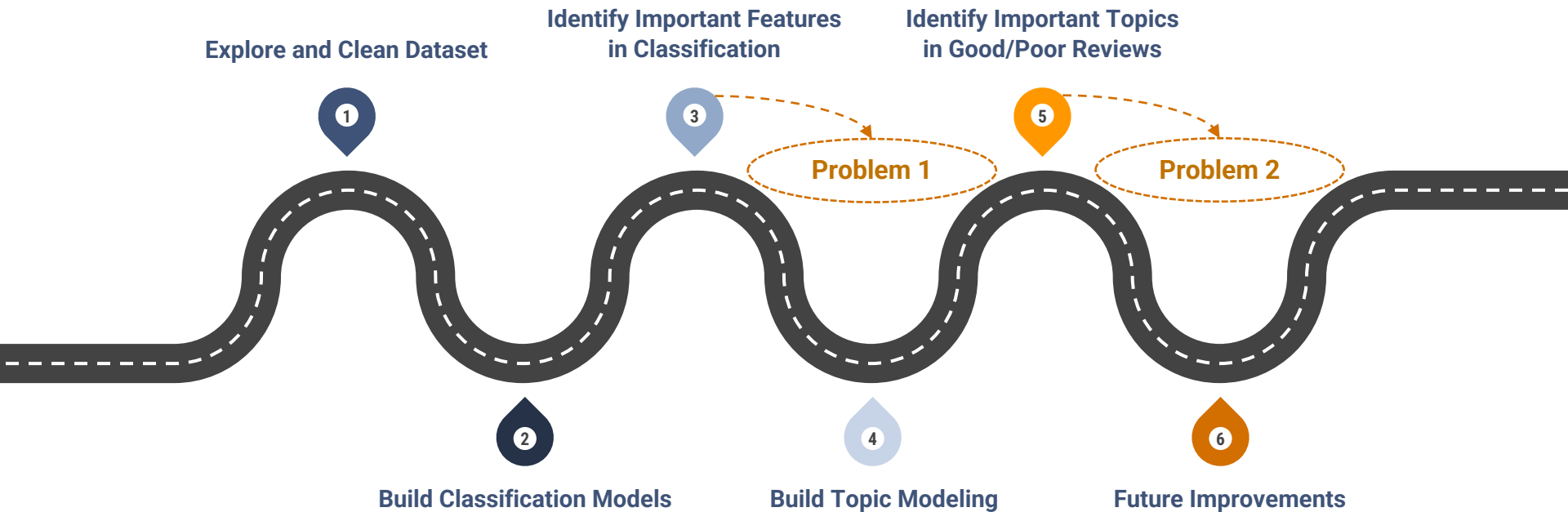
2

Why do users give good/poor reviews?
How can the Apps improve?

Topic
Modeling



Overview – Roadmap





Overview – Dataset Review



Dataset Introduction:

This dataset contains more than 50k reviews of Apps from Google Play Store which are categorized into Browsers, Video Players, File Managers, Mobile Payment, and Communication.



Important Features

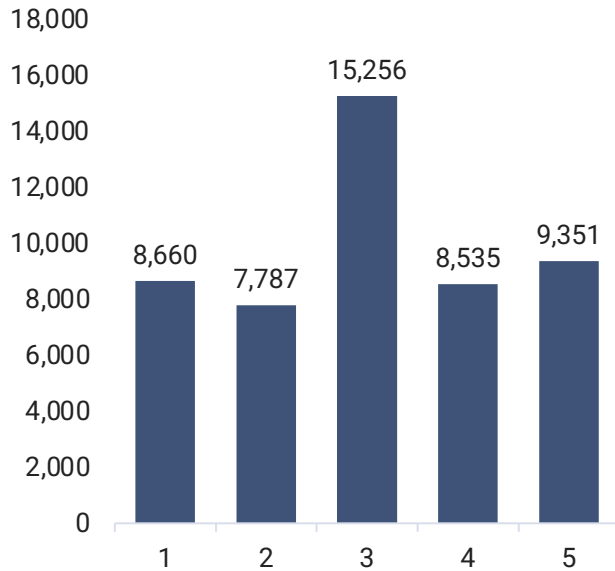
- ☐ **Content:** the reviews of the Apps.
- ☐ **Score:** the corresponding rating of the review. (Values: 1 - 5, 1 is the lowest, 5 is the highest)
- ☐ **Category:** Apps belonging to the categories. (Values: browsers, videoplayer, filemanager, payment, and communication)



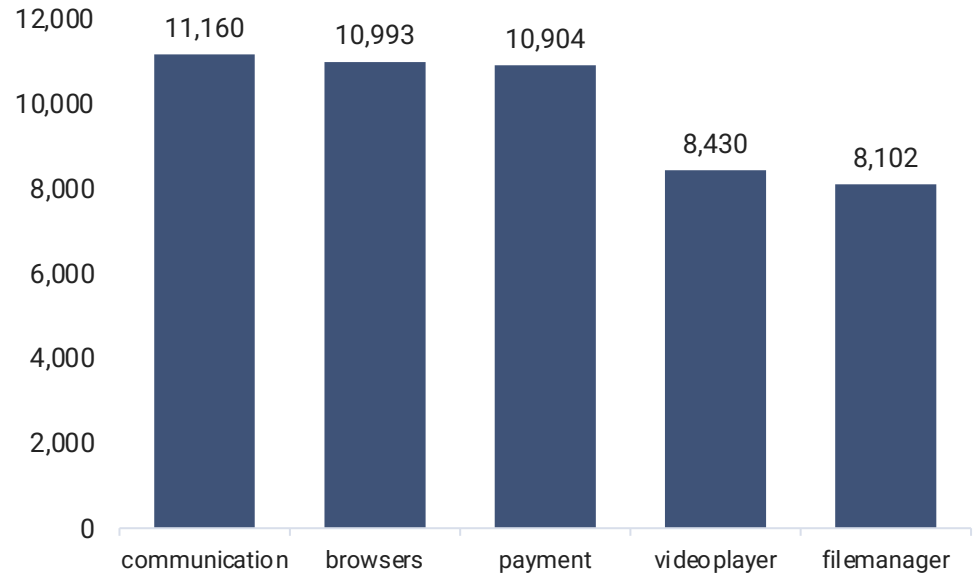
Overview – Dataset Review

The distributions of categories are similar, hence we can analyze each category respectively;

Distribution of Score

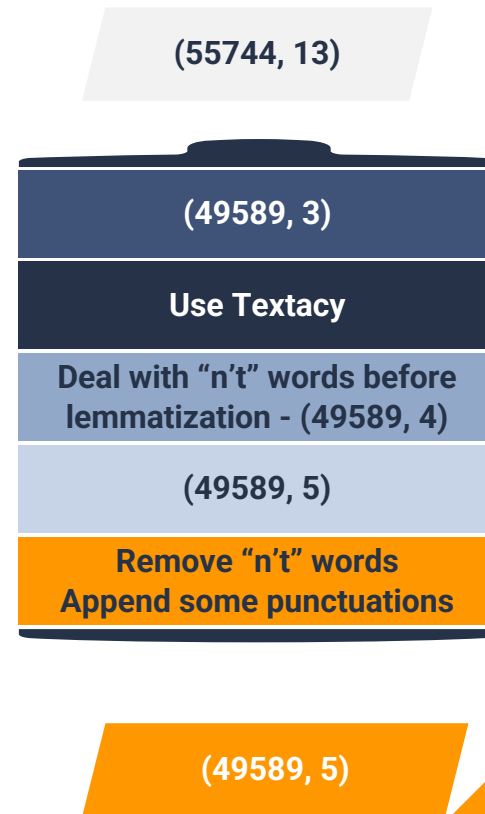
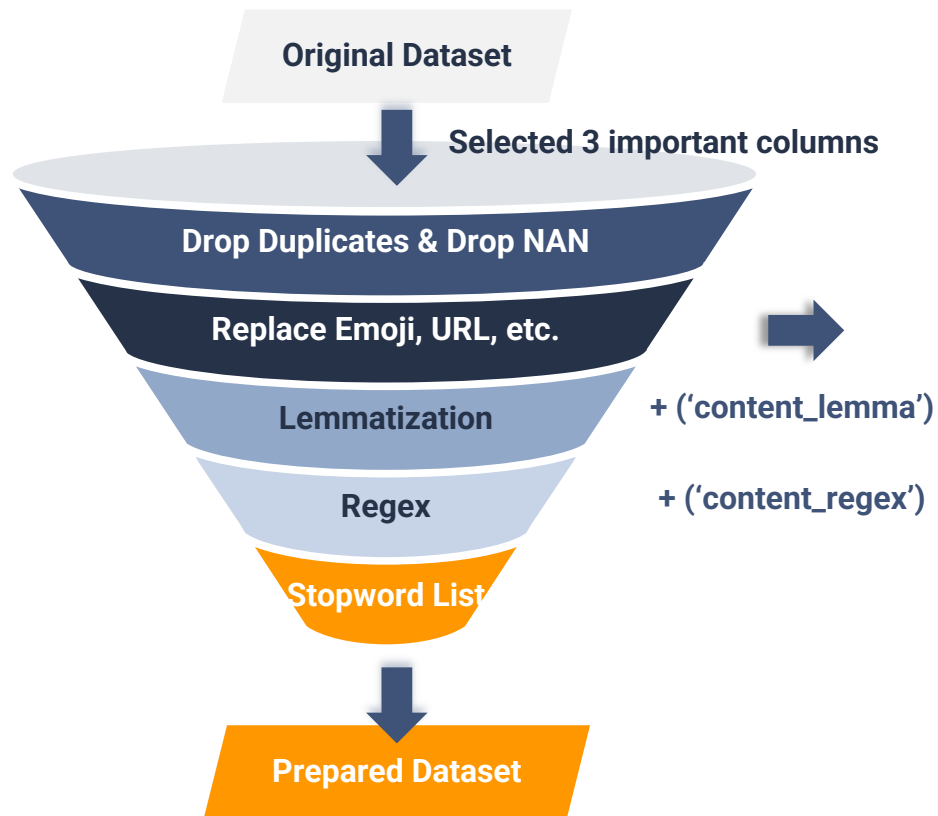


Distribution of Category





Overview – Preprocessing



 **Why do we keep "n't" words?**

The "n't" words will change the sentiment of a review.

2

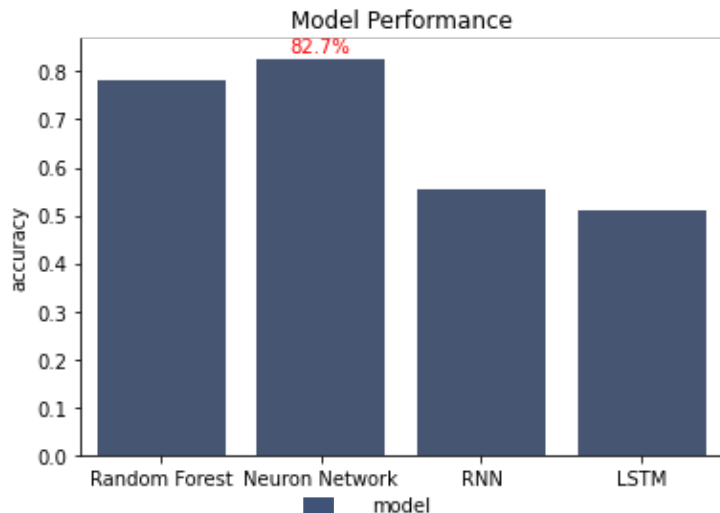
Classification Modeling

Deep Learning Models – Feature Importance

– Problematic Reviews



Classification Modeling – Deep Learning Models



Confusion Matrix	Predicted Positives	Predicted Negatives
Actual Positives	1,381	351
Actual Negatives	273	1,598

Best Result:

A simpler deep learning architecture has the best performance with 82.7% accuracy, **outperforming** a brute random guess significantly by **32.7%**.

Models Comparison:

Random Forest and **Neural Network** perform much better than deep learning architectures with more complexity, i.e. RNN and LSTM, which do not perform so well in our sentiment analysis of app reviews.

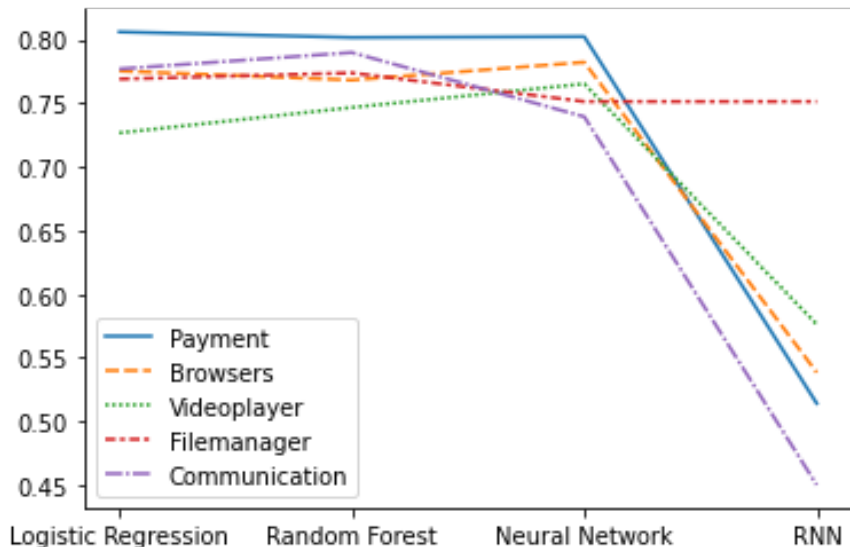
Precision-Recall Balance:

The confusion matrix returned by this model indicates a fairly good balance between precision (79%) and recall (78%).



Classification Modeling – Neural Network

Model Performance



Model Architecture of Neural Network

This model was established using our own trained TF-IDF vectorization embedding with **N-gram = (1,3)**.

The model uses a sigmoid function to output the probability of a positive/negative review (score of 5 being **positive** and score of 1 being **negative**).

Model Hyperparameters:

Vocab Size = 5,000, **Max Length** = 493, **Embedding Size**=50

Layers: Embedding (Our Own Trained TF-IDF Vectorization), **Flatten()**, **Dense** (1 Unit Sigmoid)



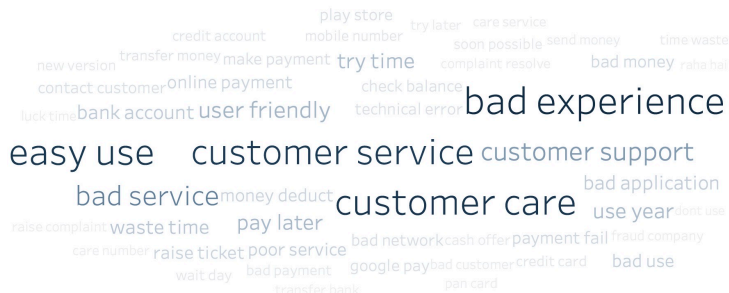
Classification Modeling - Feature Importance

Overall



- Using the whole dataset, the most important features are "easy use", "please fix", "bad experience", "not work" and so on.
- In general, the success of an App is highly related to whether it is easy to use and whether there are some bugs in it.

Payment



- Using the data of payment Apps, the most important features are "customer service", "easy use" and so on.
- For payment Apps, the company should focus more on the customer service and easiness



A word cloud of user feedback for WhatsApp. The words are arranged in a circular pattern, with some words appearing more frequently than others. The most prominent words are "easy use", "please fix", "not work", "can not even", "bad experience", "not able send message", "work properly", "something go", "can not send", "make call", "wo n", "work fine", "create account", "do not know", "internet connection no longer", "good application", "one star", "like whatsapp", "good like", "last update", "phone number", "is not work", "does not work", "verification code", "open chat", "please help", "fix problem", "get notification", "work good not good", "please add not even", "no problem", "bad ever", "many time even though", "do not want", "good use", "can not log", "new version", "work properly", "something go", "can not send", "make call", "wo n", "work fine", "create account", "do not know", "internet connection no longer", "good application", "one star", "like whatsapp", "good like", "last update", "phone number", "is not work", "does not work", "verification code", "open chat", "please help", "fix problem", "get notification", "work good not good", "please add not even", "no problem", "bad ever", "many time even though", "do not want", "good use", "can not log", "new version".

- Using the data of communication Apps, the most important features are "please fix", "good", "easy use" and so on.
- For communication Apps, the companies should focus more on fixing bugs and easiness of their Apps.

is_not work not work no ad share file not able video file copy paste
not show file explorer _good_ use did_not work not _good_ can_not see
waste time _good_ file explorer no longer google account solve _problem_ file delete
nearby sharedoes_not work file manager file management
good _good_
fix _problem_ _last update_ _good_ file manager
wo n easy use work _good_ please fix _good_ file
good work internal storage manage file _good_ application
copy file _one_ _good_ sd card work properly transfer file usb storage thank much
dark mode restore purchase user friendly file not keep crash no way
move file

- Using the data of file managers, the most important features are "last update", "good file manager", "nearby share" and so on.
- For file manager Apps, the company should focus more on the update of their Apps and the function of nearby sharing.



easy use

search result

not work

web browser

keep work

brave browser

grid view

waste time

always use

search engine

not open

try update

not use

give star

slow browser

want use

even though

browser see

keep crash

use browser

download update

user friendly

clear cache

browser use

bad browser

fast browser

please fix

dark mode

tab open

data usage

blocker work

stop use

adblock work

free speech

use firefox

work properly

close tab

mobile browser

address bar

every time

browser fast

not able

not show

google chrome

opera mini

web page

- Using the data of video players, the most important features are "many ad", "not work", "not support" and so on.
- For video player Apps, the company should focus more on the support for varies format of videos and find the balance between user experience and the number of ads.

A word cloud representing user feedback for the 'Download' button. The words are arranged in a circular pattern, with the most frequent words being the largest. The words include: 'download', 'not work', 'does not work', 'does not play', 'does not support', 'download subtitle', 'bad experience', 'no option', 'unable play', 'not download', 'even though', 'android tv', 'not able', 'watch_ad_', 'pro version', 'can not get', 'wo n clear data', 'n entry play', 'soon possible', 'n play', 'waste time', 'download movie', 'does_not play', 'not support', 'download subtitle', 'bad experience', 'does_not work', '_many_', '_ad_', 'error message', 'is_not support', 'not work', '_ad_ free', 'download speed', 'can_not play', '_last update_', 'every time', '_many_ add', 'do_not download', 'no way', 'android_phone_', 'is_not work', 'watch movie', 'play audio', 'can', 'can_not download', 'not play', 'go back', 'not even', '_ad_ every', 'no _ad_', 'does_not show', 'sd card', 'no sound'.



Classification Modeling - Problematic Reviews

Spurious **Negative**

"Why the search page isn't opening in the first go after the last update? User have to search the same thing twice unnecessarily."

Score: 5.0 (Positive)

Prediction: Negative

Analysis - Wrong Given Score

Our model **successfully** captures the negative sentiment of this review which has a spurious 5.0 score. The true sentiment is inconsistent with the actual score given. **Thus, our model can be applied as a complementary testing tool to the existing rating system.**

Spurious **Positive** Prediction

*"Generally speaking, Whatsapp is very **helpful and user friendly**... But there is a major drawback of Whatsapp..... There is no "Edit " option... So once you post your message in Whatsapp, the content of the message becomes permanent... Even if there is some spelling mistakes, you cannot correct it anymore But Telegram has edit option. So Telegram has an edge over Whatsapp... So if Whatsapp wants to retain its users, Edit option should be there in future upgrade of Whatsapp."*

Score: 1.0 (Negative)

Prediction: Positive

Analysis - Wrong Prediction

There are important tokens such as '**helpful**' and '**user friendly**' in the review, which show frequently in our good reviews and make our model predict it as positive. But the overall signal of the review is negative since the user is much more concerned about the drawbacks of Whatsapp.

3

Topic Modeling

Business Insights from Topic Modeling

A person with short dark hair, seen from the back, is looking at a wall covered in various business-related documents, diagrams, and photos. The documents include flowcharts, charts, and text. The person is wearing a grey and black striped sweater. The text "What is the Business Insights?" is overlaid in the center in a large, bold, orange-outlined font. A blue arrow points from the left towards the top of the wall. An orange arrow points from the bottom right towards the center.

What is the Business Insights?



Topic Modeling – Payment (Positive Reviews)

Bank Money Transfer

- ❑ Improve automatic money transfer features and remind customers to set up regular frequent transactions
- ❑ Cooperate with more banks for wider coverage of the banking system
- ❑ Decrease money processing time by enhancing telecommunications with banks

Easy Features & User Friendly Interface

- ❑ Stick to intuitive and simple design styles and user-friendly interfaces



**Positive
Aspects
&
Reasons**

Sending Money

- ❑ Use fraud analytics algorithms to detect transaction anomaly and ensure transaction safety
- ❑ Expand identity verification methods including username, phone numbers, email, etc.



Topic Modeling – Payment (Negative Reviews)

Customer Service

- ❑ Ensure the accessibility of customer service by increasing customer hotlines and online receptionists
- ❑ Establish a formal procedures to handle customer complains and process refund requests within a specific timeframes (such as in 3 days)



Negative Aspects & Solutions

Pay Later Feature Failure

- ❑ Conduct credit scoring for profile screening and use credit metrics as reference of transaction approvals
- ❑ Increase the options of payment dates so customers can choose to repay on different dates

Bank Account Update

- ❑ Adjust transaction fee according to the estimated time between payment applications and bank accounts
- ❑ Add email and text notification features to notify customers of the current stage of transactions

Technical Problems

- ❑ Monitor technical issues on a daily and weekly basis,
- ❑ Analyze the cause of these issues through daily transaction data and customer complaints and debug



Topic Modeling – Browser (Positive Reviews)

Easy To Use Features

- ❑ Streamline features that are less likely to be used and keep user interface neat and clear
- ❑ Highlight no tracking and automatic deleting browsing history options to ensure users of data privacy and security

Extension Support

- ❑ Test the compatibility with different browser extensions on a regular basis
- ❑ Improve the functionality of high-frequency browser extensions



Dark Mode & Data Saving Mode

- ❑ Design more dark mode options and improve dark mode layout for dark modes to increase visibility in different environments
- ❑ Set up an automatic reminder feature for users to transfer to data saving modes when the battery is below a certain threshold

Desktop Versions

- ❑ Adjust browser features according to different platforms to customize different platforms
- ❑ Improve the automatic syncing features of bookmarks and browsing histories



Topic Modeling – Browser (Negative Reviews)

Update Failure

- ❑ Remember user preference of automatic software update
- ❑ Avoid automatic update when users reject the options
- ❑ Report update failure cases, analyze the cause of failures and debug accordingly



AdBlock Failure

- ❑ Monitor and test the AdBlock features to check AdBlock performance on different websites
- ❑ Customize AdBlock features according to different websites and dynamically adjust the features

Forced Grid View

- ❑ Increase view options for users to choose between cascade view and grid view
- ❑ Survey and research user experience preference before conducting software updates



Topic Modeling – Video Player (Positive Reviews)

No ad

- ❑ Provide video players without intrusive ads
- ❑ Create a non-ad version for users to choose from

Many Feature

- ❑ Provide option to switch between languages for dual audio vids
- ❑ Provide features similar to PC version to increase the loyalty of existing PC users
- ❑ Add feature of pan zoom which users can freely zoom video
- ❑ Customize number of seconds to skip
- ❑ Customize audio



Support Multiple Format

- ❑ Support different formats of video such as MP4, MOV, FLV, AVI
- ❑ Support Android TV and other smart TV

Good for Movie Watching

- ❑ Ensure movie can load in quickly
- ❑ Provide automatic recognition subtitles in different languages
- ❑ Provide 4K version for video to improve the user experience in movie watching



Topic Modeling – Video Player (Negative Reviews)

Crashing in Latest update

- ❑ Ensure new version work properly before launching new version
- ❑ Fix issue like (1) when forward the video, audio doesn't stop (2) screen turn off when video is playing (3) OTG connection problem

Too many ad

- ❑ Reduce the length of ads or even provide a non-ad version to increase the user retention



**Positive
Aspects
&
Solutions**

Format Not Support

- ❑ Support more video format such as MPEG4, HEV, OPUS, HEVC
- ❑ Support 4K video which is an increasing need
- ❑ Add more audio format like AAC, EAC3, AC3



Topic Modeling – Communication (Positive Reviews)

Video Call

- ❑ Support video calls in communication app, which is an increasing need
- ❑ Support group video call for a larger number of users instead of only 4
- ❑ Available across all platforms such as Android, IOS, windows, etc.
- ❑ Ensure the video call is clear, smooth, and fast



**Positive
Aspects
&
Reasons**

Status Option

- ❑ Provide status options for users to show their real-time status

Easy to Use

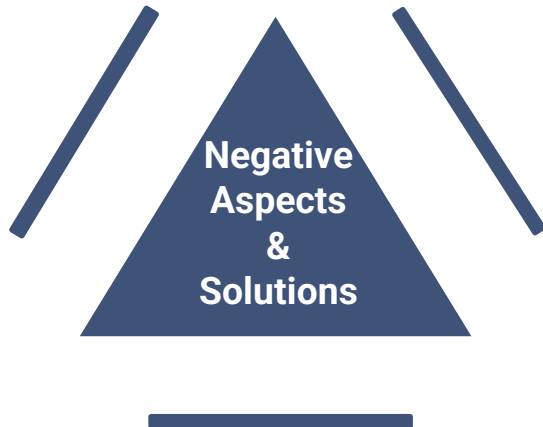
- ❑ Put common features in a prominent place
- ❑ Provide user tutorials to new users
- ❑ Keep consistence when updating to new version



Topic Modeling – Communication (Negative Reviews)

Video Call Problem

- ❑ Provide minimize square frame for video while in chatting page
- ❑ Fix video call auto rotate issue, provide manual rotate option
- ❑ Ensure smooth switching to cellular data when WIFI is disconnected
- ❑ Reduce battery consumption for video calls



**Negative
Aspects
&
Solutions**

Phone Number

- ❑ Improve security and privacy by owning users' phone number
- ❑ Allow users to sign up using email if the PC version does not require a phone number
- ❑ Provide automatic identification of phone numbers to enable users copy & paste phone numbers easily
- ❑ Provide an appropriate way to solve the "phone number already exists" problem

Message Sending Issue

- ❑ Provide option to send message to multiple users at one time
- ❑ Improve message sending speed
- ❑ Show specific reason for the issue when users couldn't send the message successfully



Topic Modeling – File Managers (Positive Reviews)

Easy to use

- ❑ Design a user-friendly and easy-to-use interface
- ❑ Highlight the most important features for users' daily use
- ❑ Improve constantly the user interface design

No ad

- ❑ Ensure app service quality without intrusive ads
- ❑ Find the balance between user experience and the number of ads to increase the profit

**Positive
Aspects
&
Reasons**

Free of Charge

- ❑ Continue to provide free file manager service to users



Topic Modeling – File Manager (Negative Reviews)

Run Slowly

- ❑ Ensure the file preprocessing performance after users use it for a long time
- ❑ Fix the problem of crashing after long-time use
- ❑ Optimize the Apps to enable user use less memory and run faster

Update Problem

- ❑ Combine the adjustment in features in several updates to reduce the number of updates



Negative Aspects & Solutions

SD Card Issue

- ❑ Fix the problem of reading and writing the file from the SD card
- ❑ Ensure that the apps can detect various types of SD cards

4

Future Improvement



Future Improvement

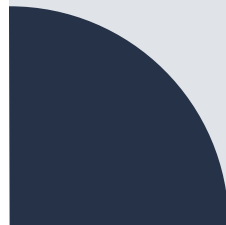
Data Cleaning

Clean reviews more thoroughly and reasonably to achieve higher accuracy



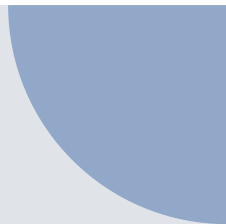
Modeling

Try more models such as transformer and BERT to seek better classification results



Do oversampling based on the 'thumbupcount' to consider the impact of thumb-ups in each review

Feature Engineering



Explore how to Choose a better topic size; try to avoid the same words appearing in different topics

Topic Modelling





THANKS!

Any questions?
You can find us at
zoux@usc.edu/