Sentiment Analysis for Shopee

Grace Chia

Business problem

Shopee's product managers wish to **improve the online shopping experience** for customers
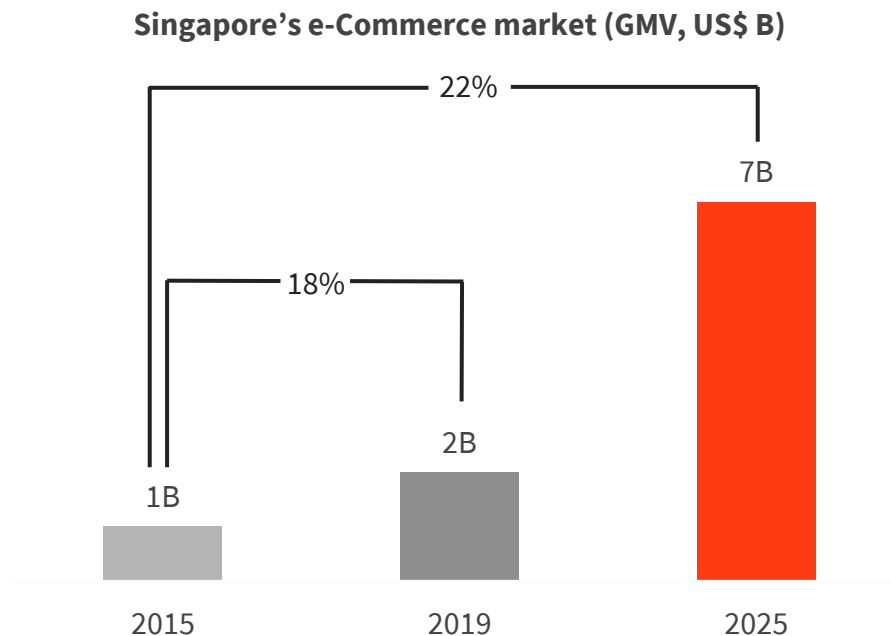
Classify positive and negative reviews on Shopee Singapore's Google Play using NLP

The model that achieves the highest **accuracy and recall** on the validation set will be selected for production

Identify **pain points** among dissatisfied customers using topic modeling

Why does it matter?

The e-Commerce market presents opportunities, expecting to grow 3.5x in size by 2025

Singapore's e-Commerce market (GMV, US$ B)

22%

18%

7B

2B

1B

2015   2019   2025

Source: Google, Temasek, Bain & Company e-Conomy SEA 2019

Data Collection

Scraped Google Play and collected 4,461 user reviews on Shopee

REVIEWS

Review Policy

4.7

★ ★ ★ ★ ★

👤 51,917 total

5
4
3
2
1
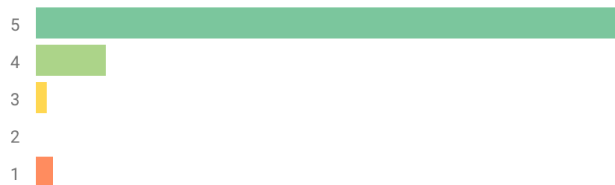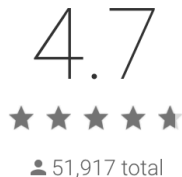
Yi Lin Lee

★ ★ ★ ★ ★ 3 June 2020

👍
38

App is really user-friendly and easy to use! The interface is not cluttered at all unlike the other shopping apps I've tried. Additionally, I usually like to do price comparisons before making purchases, and I almost always find Shopee to have the best deals. You can also be reassured that you're bu

1-3 stars: Negative sentiment, class 1

4-5 stars: Positive sentiment, class 0

Goal
Can we predict sentiment based on user reviews?

Data Cleaning

Key data cleaning steps that helped to improve accuracy

Removed reviews that were
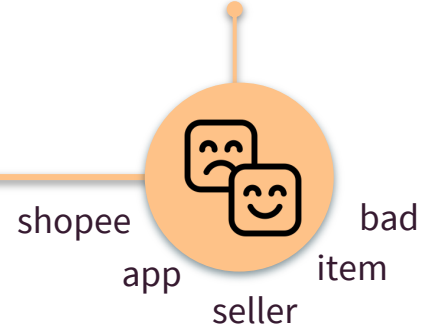**non-English** or **gibberish**

Apa punya
app ni?

Used NLTK to remove
stopwords and **stemmed**
words to root form

Removed frequently occurring
words that **appeared in both
positive and negative reviews**

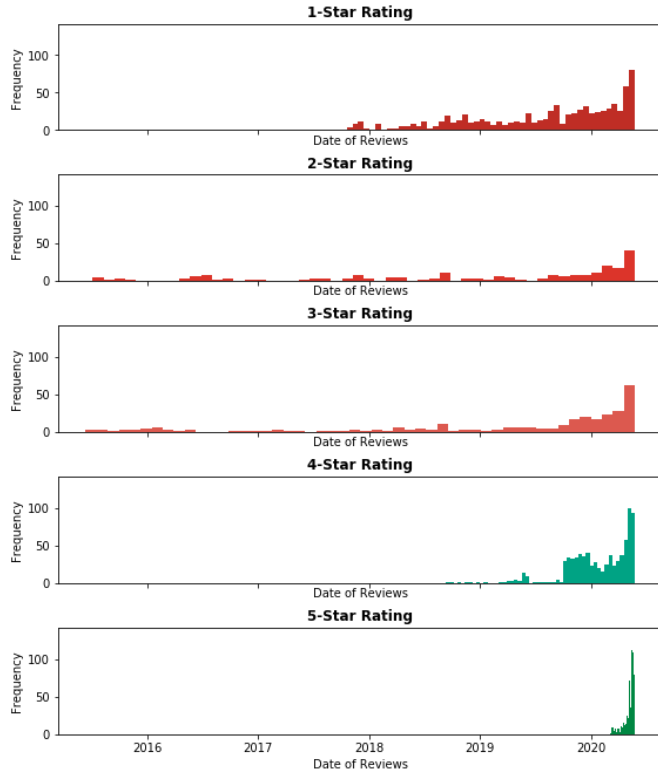shopee
app
seller
bad
item

Shopee

EDA

Growth in the no. of reviews due to increased e-commerce activity during COVID-19 lockdown
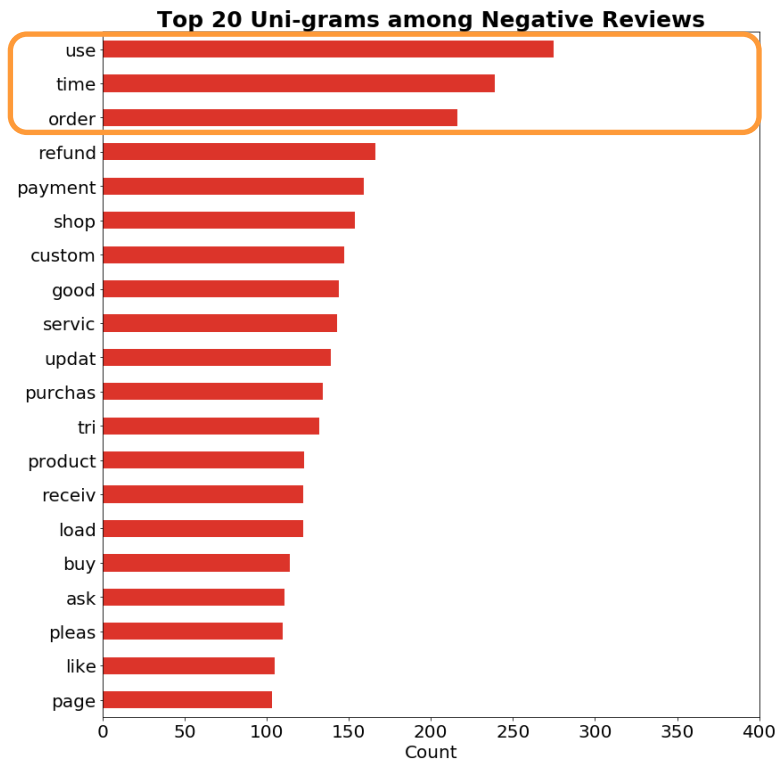
Distribution of Review Dates by Users' Ratings

"Good experience! Have been **shopping more during circuit breaker** period. Playing games and collecting some coins to deduct off purchase."

'Use', 'time' and 'order' are the top 3 words seen in negative reviews
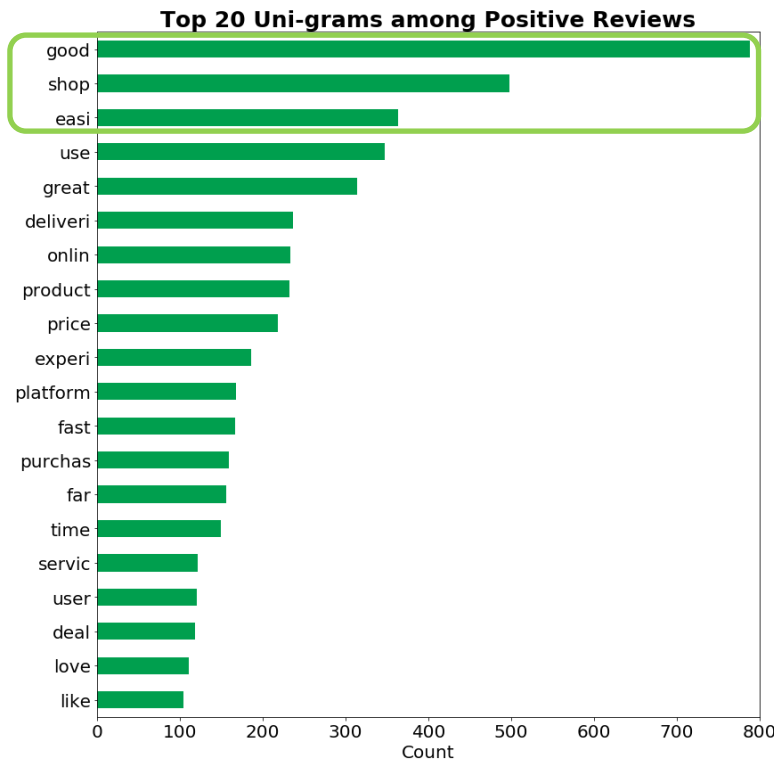
Top 20 Uni-grams among Negative Reviews

"Terrible payment.
For a long **time** I have been
using credit/debit payment
without any glitch. Today, I kept
being denied of my payment.."

User rated Shopee 2-stars

'Good', 'shop' and 'easi' are the top 3 words in positive reviews

Top 20 Uni-grams among Positive Reviews

"Quality of goods matched the review and picture advertised. Purchase status and whereabouts can be tracked from the app and delivery service providers. Had a **good shopping** experience with Shopee!"
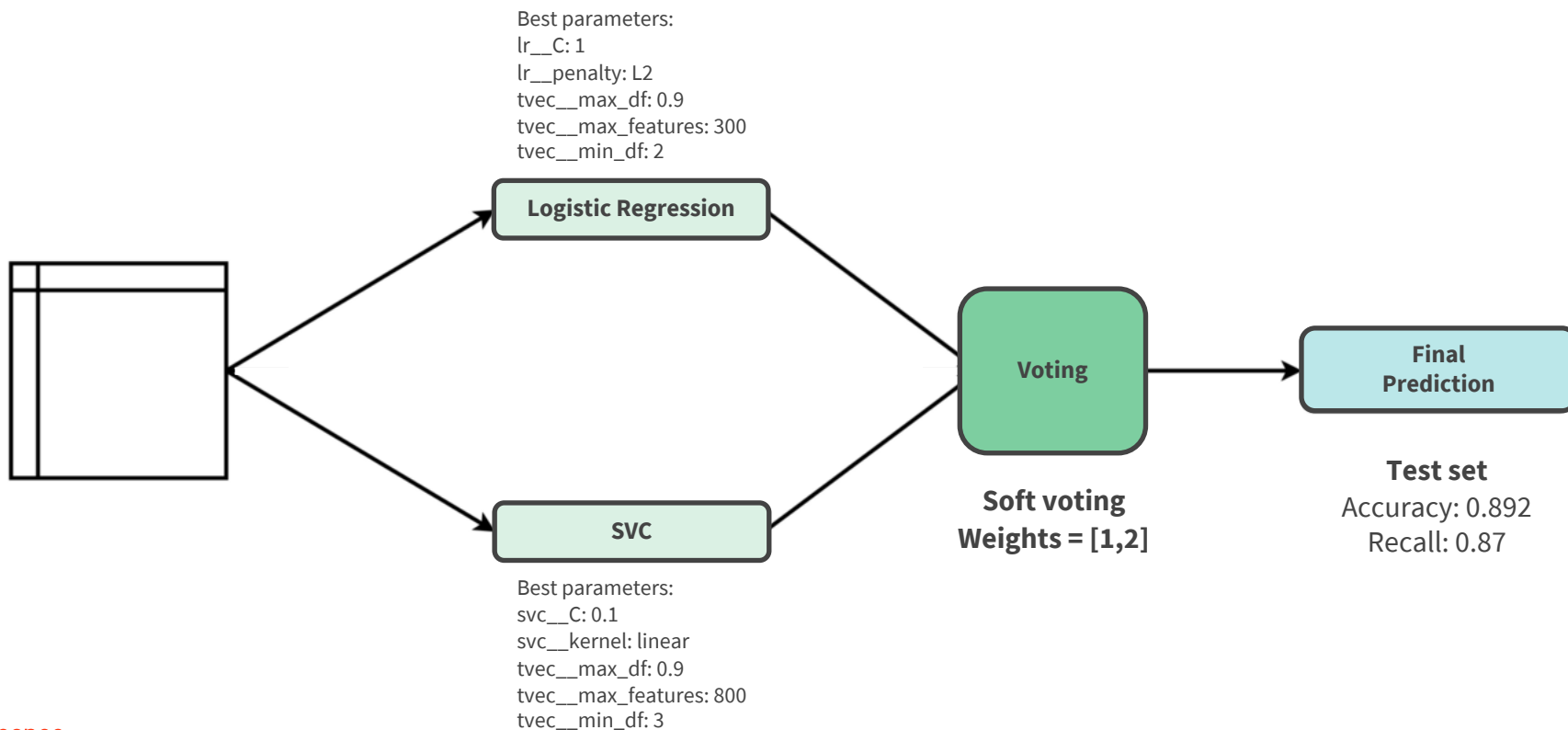
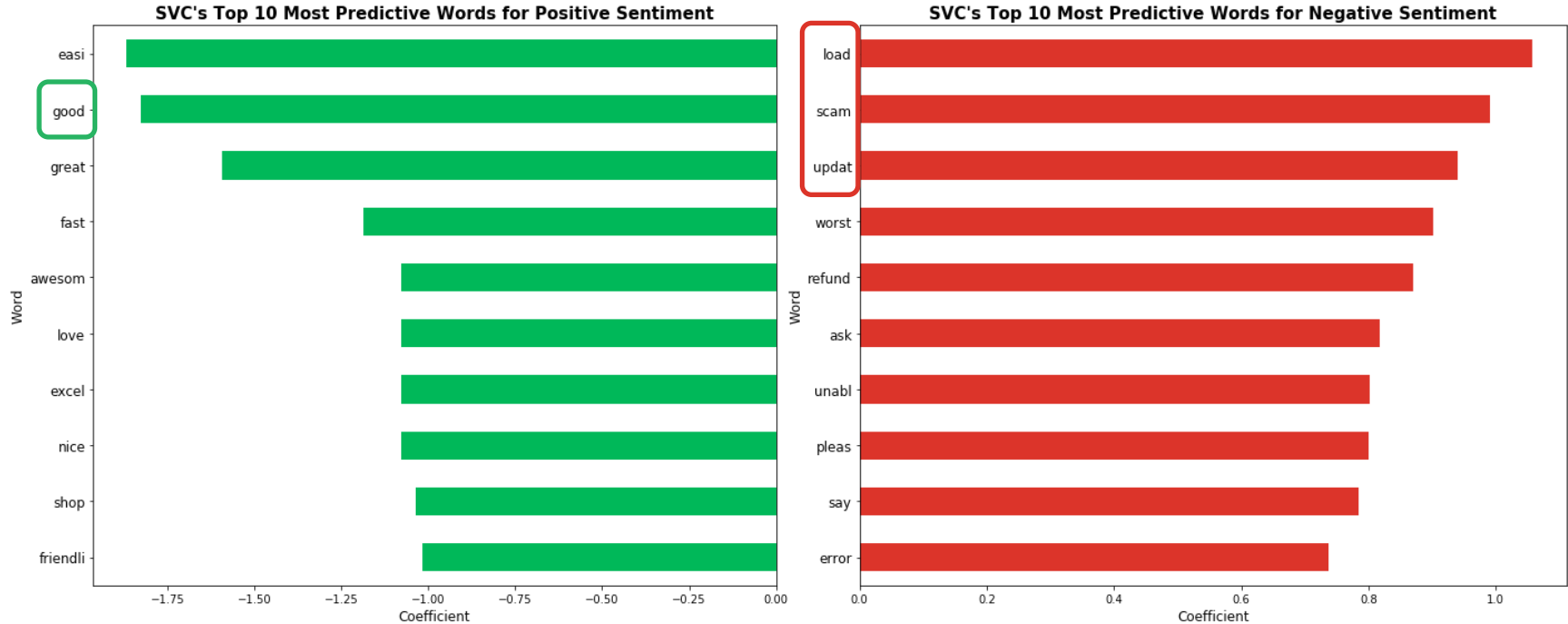User rated Shopee 4-stars

Modeling

Voting Classifier selected for its high accuracy and recall

	Accuracy on Training Set	Accuracy on Validation Set	Recall on Validation Set
Voting Classifier (TF-IDF Logistic Regression & TF-IDF SVC)	0.902	0.904	0.87
TF-IDF & SVC	0.901	0.899	0.87
Count Vectorizer & Naïve Bayes	0.898	0.898	0.83
Bidirectional LSTM	0.916	0.896	0.82
TF-IDF & Logistic Regression	0.906	0.892	0.81
TF-IDF & Naïve Bayes	0.904	0.882	0.79
Count Vectorizer & SVC	0.900	0.872	0.75
Count Vectorizer & Logistic Regression	0.882	0.861	0.71
Baseline (Majority class)	0.61		-

Assigned a higher voting weight to SVC than LR as the SVC had a better recall than LR

Best parameters:
lr__C: 1
lr__penalty: L2
tvec__max_df: 0.9
tvec__max_features: 300
tvec__min_df: 2

Logistic Regression

Voting

Final Prediction

Best parameters:
svc__C: 0.1
svc__kernel: linear
tvec__max_df: 0.9
tvec__max_features: 800
tvec__min_df: 3

SVC

Soft voting
Weights = [1,2]

Test set
Accuracy: 0.892
Recall: 0.87

Shopee

14

Most predictive words for each sentiment

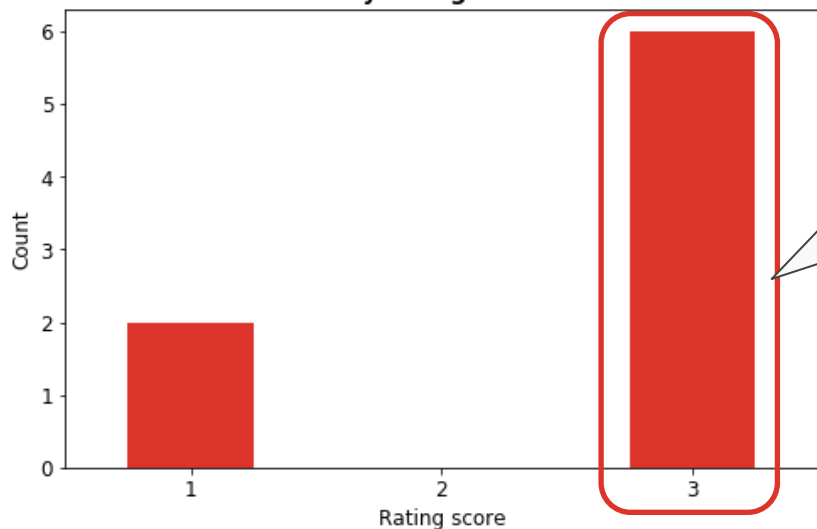SVC's Top 10 Most Predictive Words for Positive Sentiment

SVC's Top 10 Most Predictive Words for Negative Sentiment

Limitations: Misclassifications tend to occur when users write mixed reviews

No. of times the word 'good' appeared in False Negative reviews
By rating score

"**Good** platform for products and purchases **but the app lags and hangs** very often. Hopefully the app can be more stable so that consumers can have a better experience shopping."
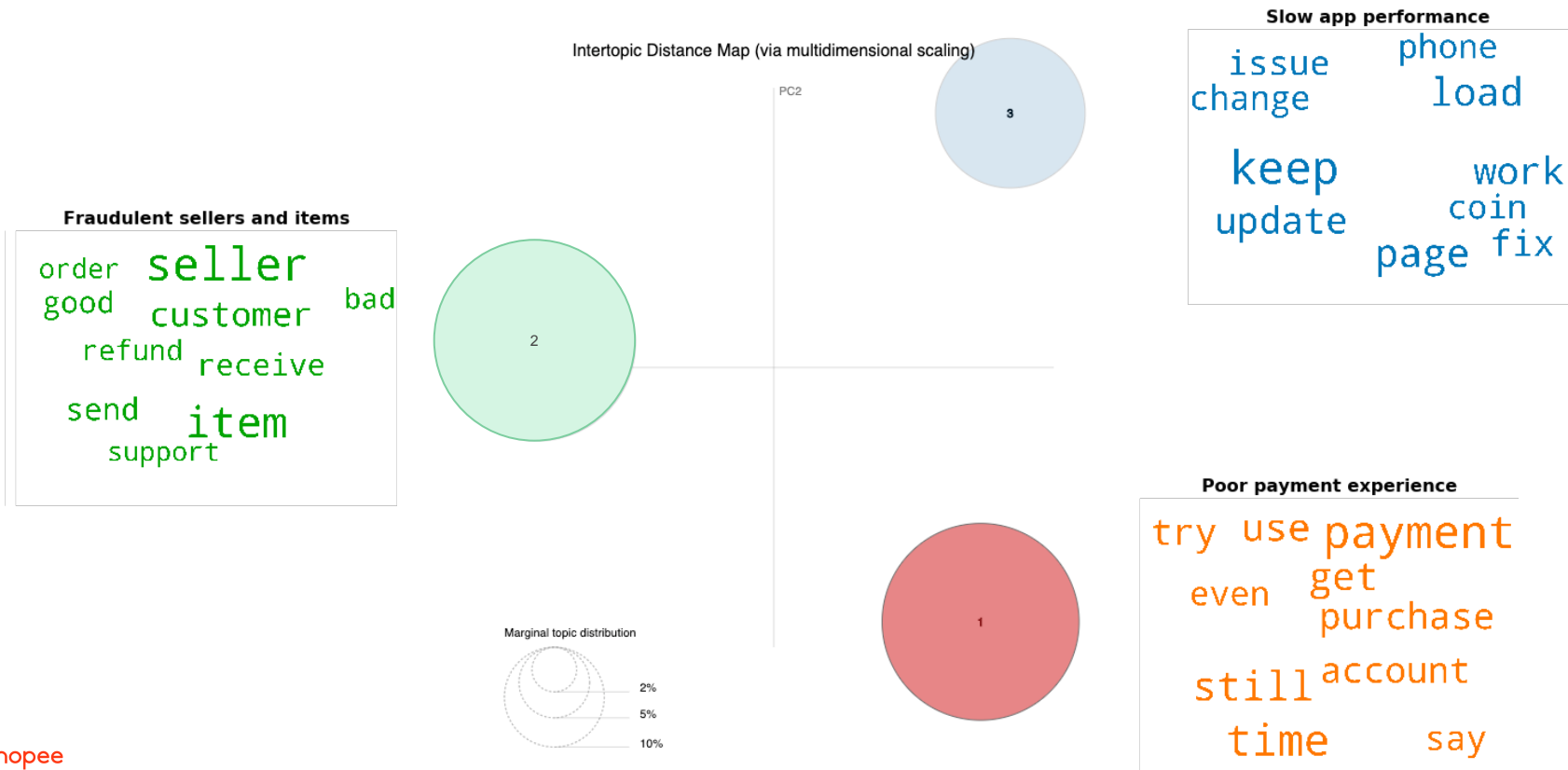
Demo time!

Write a Review on Shopee

Review | Predict my sentiment!

https://shopee-sentiment-analysis.herokuapp.com/

Shopee

Topic modeling for negative sentiment

Recommendations

Customer Pain Points	Recommendations
Poor payment experience	• Software engineers need to enhance the stability of its payment gateway service
Fraudulent sellers and items	• Identify and block fraudulent sellers by tracking behavioural pattern • Improve buyers' protection scheme to regain trust
Slow app performance	• Optimise Shopee's traffic scheduling platform latency, especially during big sale events • Perform rigorous stress testing before releasing new app updates

Conclusion

Classification model

- Production model generalises well on unseen data, achieving a relatively high accuracy (0.892) and recall (0.87) on the test set
- This model serves as a good start for Shopee's product managers to classify sentiments and understand customers' pain points through Google Play reviews

Topic modeling

- Identified 3 pain points among dissatisfied customers for product managers to act on:
    - Poor payment experience
    - Fraudulent sellers and items
    - Slow app performance

Next steps

- Train model on reviews from the Apple App Store
- Include the 3 topics found as features in our classification model

Thank you!