



Classifying Customer Feedback (Delivery and Ride Hailing)

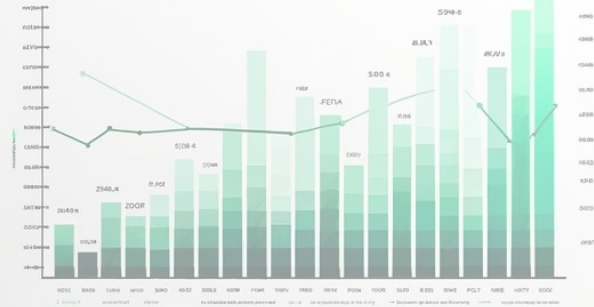
Irfan, Zheng Gang, Ng Wei

Content

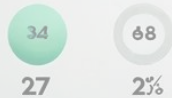
- **Sentiments for Grab Applications**
- **Introducing Greg and our Problem Statement**
- **Reddit Data & Initial Findings**
- **Modelling & Evaluation**
- **App Demo**
- **Conclusion & Recommendations**



High-stick
of minomafic



Ride Frequency



U2D GRAB DLE	FEEDBACK
1 5%	1 10%
2 10%	2 20%
3 15%	3 30%

Ride



2020/1	2020/2
1 10%	1 10%
2 20%	2 20%
3 30%	3 30%

Ride Frequency



FEEDBACK	FEEDBACK
1 10%	1 10%
2 20%	2 20%
3 30%	3 30%

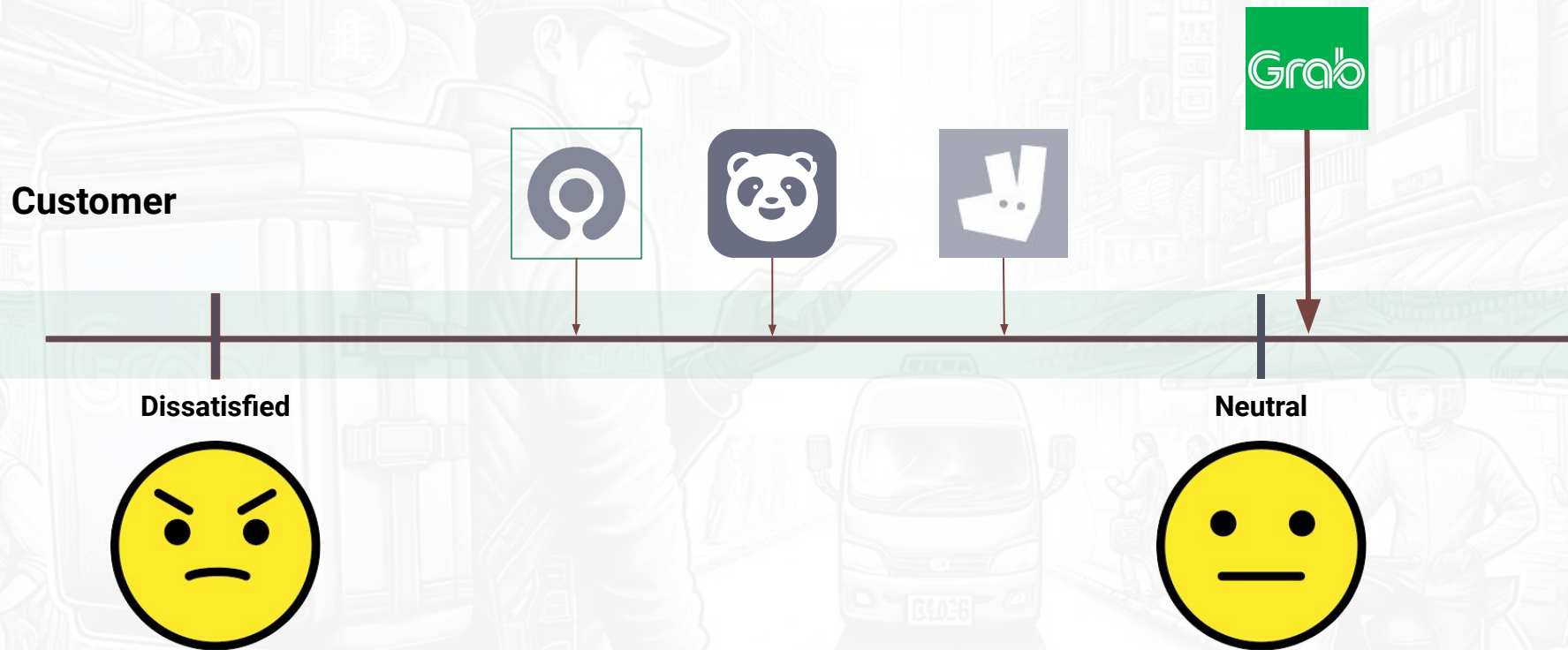
Sentiments of Grab applications

Sentiment Analysis of App Store Reviews*



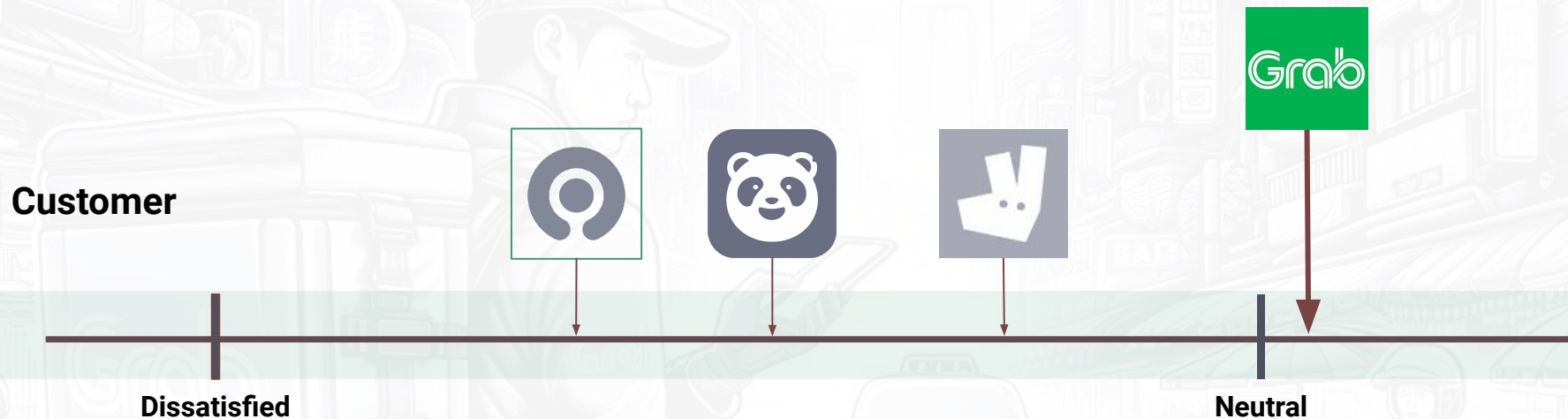
*Based on Google play store comment reviews using Hugging Face BERT Sentiment Analysis. Not drawn to scale.

Sentiment Analysis of App Store Reviews*



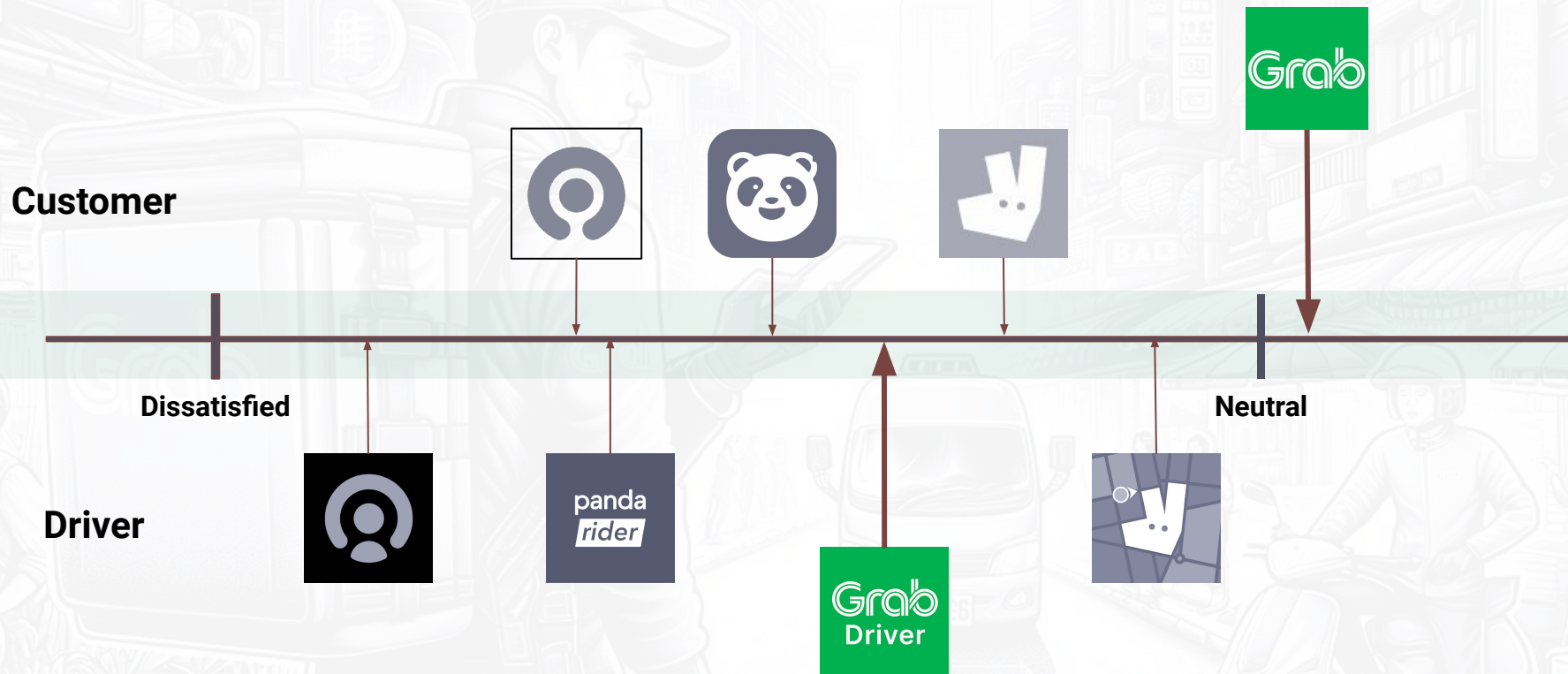
*Based on Google play store comment reviews using Hugging Face BERT Sentiment Analysis. Not drawn to scale.

Sentiment Analysis of App Store Reviews*



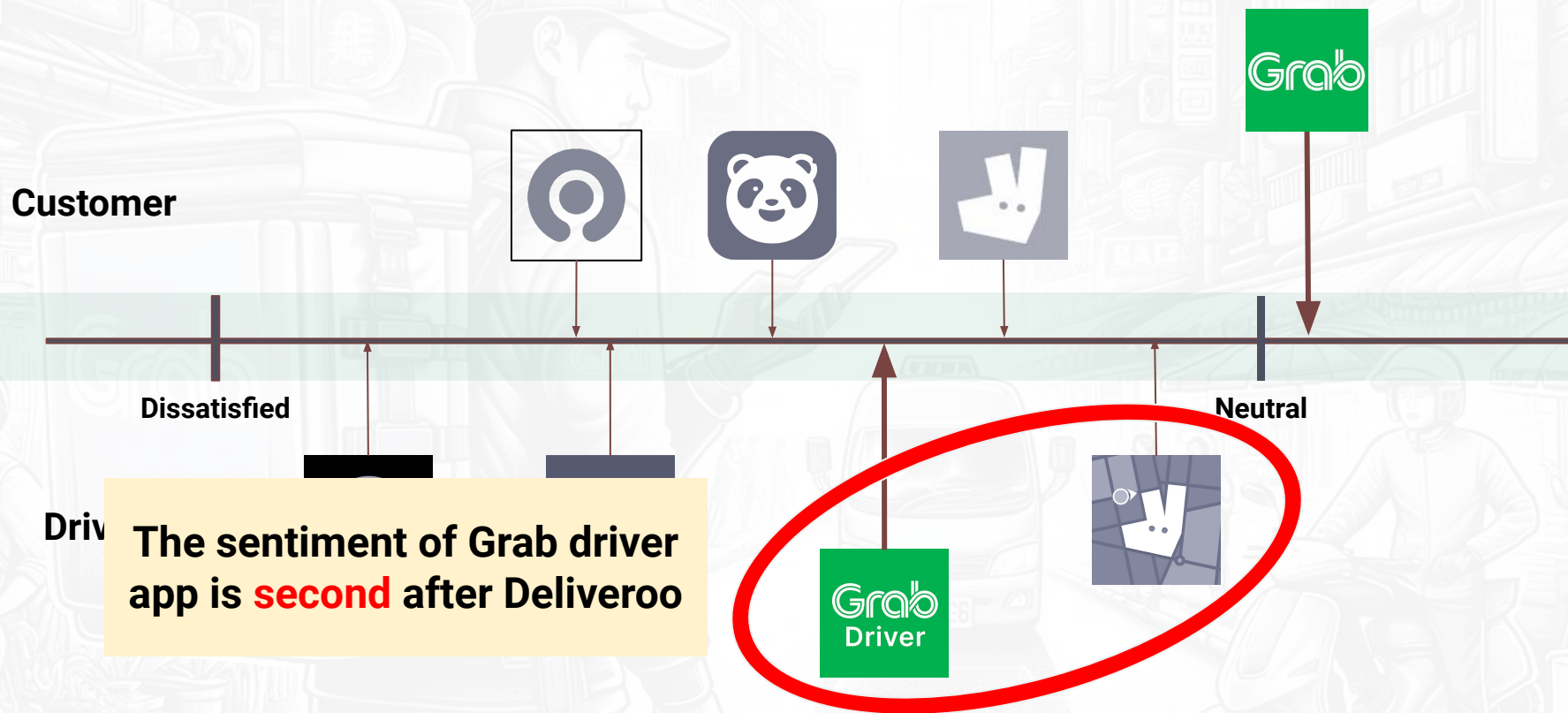
Driver

Sentiment Analysis of App Store Reviews*



*Based on Google play store comment reviews using Hugging Face BERT Sentiment Analysis. Not drawn to scale.

Sentiment Analysis of App Store Reviews*



*Based on Google play store comment reviews using Hugging Face BERT Sentiment Analysis. Not drawn to scale.

Same Ratings



4.1★

1.61M reviews



4.1★

38K reviews

Same Ratings

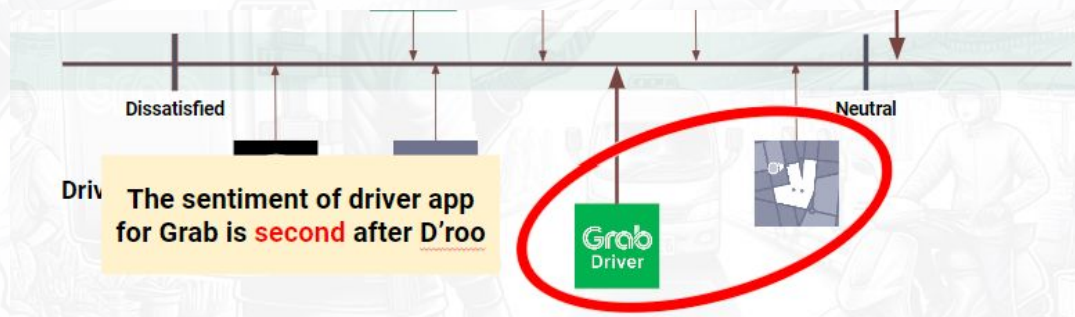
Sentiment Analysis shows otherwise



4.1★
1.61M reviews



4.1★
38K reviews



App Ratings **do not show
the full picture.**



4.1★

1.61M reviews



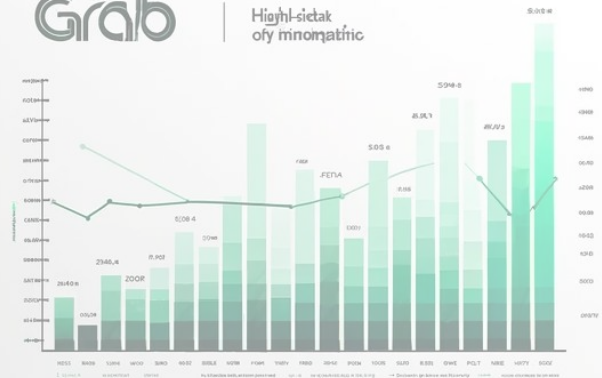
4.1★

38K reviews

Use **app store reviews to
gain insights on our
users' sentiments.**



High-stick
of minomafic



Ride Frequency



27

2020

2020



2%

2020

2020

Ribde



2020

2020

Ride Frequency



27

2020

2020



2%

2020

2020

Introducing Greg

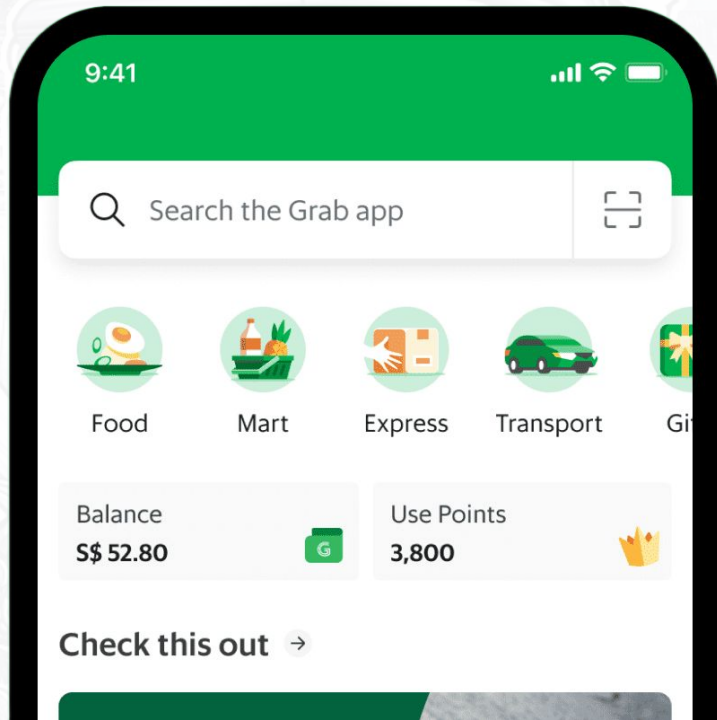
Greg

Profile: 30-year-old, Product Manager


- Oversees the product feature pipeline.
- Need to prioritize features between **ride hailing** and **delivery** due to limited tech resources.
- Aim to extract actionable insights via **reviewing customer feedback** on app store.



Multiple services on the same app



Reviews...

 **Grab - Taxi & Food Delivery**
Ratings and reviews

g greeku o

★★★★★ February 3, 2024

What a terrible app. Order a taxi from the airport and by the time the driver finished putting our bags in his trunk extremely slowly (he wouldn't accept any help) I already had a 3 dollar waiting fee added to the booking. Then it said I couldn't use my credit card for the trip so I had to cancel for an extra 4 dollars (which it took off the same credit card). Clearly the app is a scam and should be avoided at all cost.

52 people found this review helpful

Did you find this helpful? ☐ Yes ☐ No

H Hitesh Bansal

★★★★★ March 9, 2024

Pathetic service by grab. Remember this app loots you of your hard earned money. If the merchant changes the order because they don't have any item available and your order amount will fall lower than the amount on which you applied the promo, then grab will cancel that promo and you will not get any discount! For the mistake of grab and the merchant the customer ends up paying higher amount. What's the point of promising a promo on app if you are not going to honor it?

24 people found this review helpful

Did you find this helpful? ☐ Yes ☐ No

Ride Hailing



Delivery



[Google Play Store](#)



Greg

Role: Product Manager

4.8★

12.9M reviews

Challenge:

**Overwhelmed by vast user reviews;
struggles to classify reviews
between delivery and ride hailing.**



Problem Statement

How can we **distinguish** between customer feedback related to Grab's **ride-hailing** and **delivery** **fast** and **accurately**?



High-stick
of minomafic



Ride Frequency



27

2020

2020



2%

2020

2020

Ribde



2020

2020

Ride Frequency



27

2020

2020



2%

2020

2020

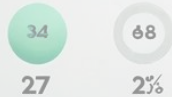
Research

Grab

High-stick
of minomafic



Ride Frequency



USD GRAB DLE
570
570
570

Ride



USD GRAB DLE
570
570
570

Ride Frequency



USD GRAB DLE
570
570
570

Research

Reddit



Why Reddit ?

Easy to use Public API

Why Reddit ?

Easy to use Public API

Service Types Specific Threads (Subreddit)

Why Reddit ?

Easy to use Public API

Service Types Specific Threads (Subreddit)

Substantial Data Volume

Why Reddit ?

Easy to use Public API

Service Types Specific Threads (Subreddit)

Substantial Data Volume

Negative Sentiment

Our Data.....

Ride-Hailing



Uber

Grab equivalent

Subreddit with **53k members**

Our Data.....

Ride-Hailing



Uber

Grab equivalent

Subreddit with **53k members**

Delivery



UberEats

Grabfood equivalent

Subreddit with **143k members**



Initial Findings

Key Term Differences

Delivery

Order

Customer



Ride Hailing

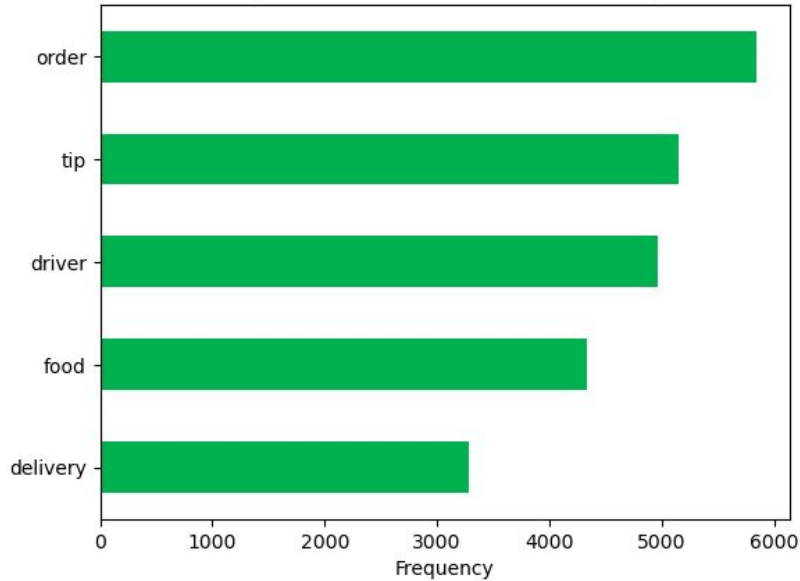
Ride

Passenger

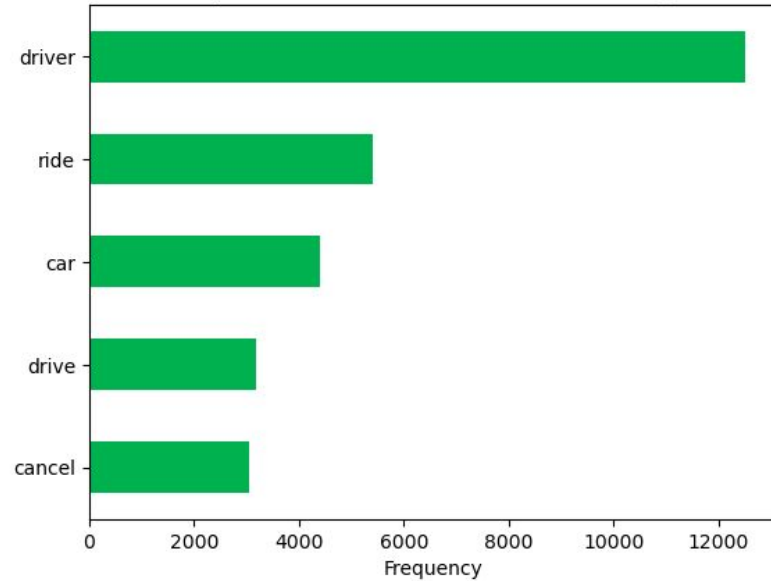


Frequency analysis

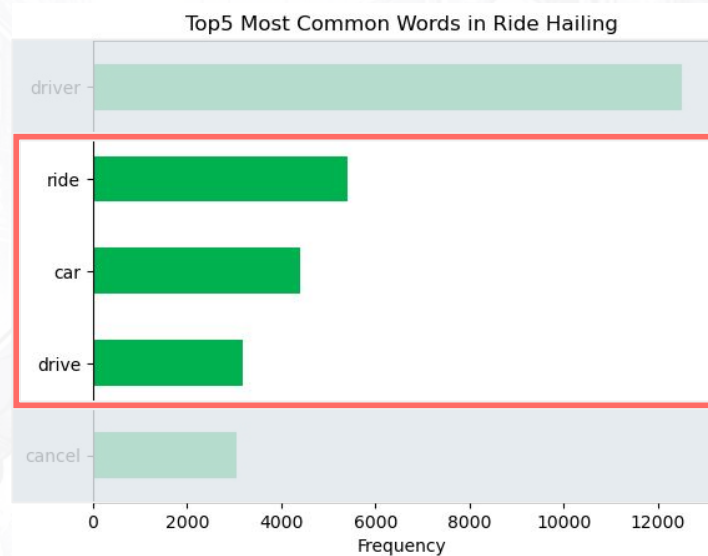
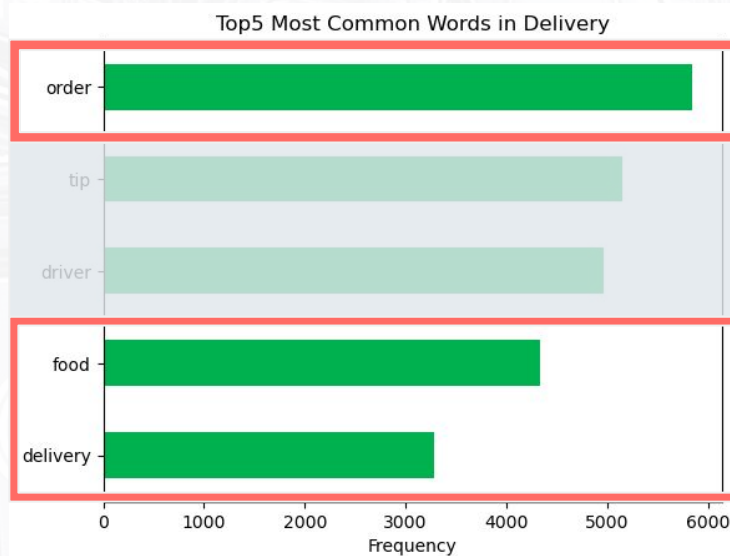
Top5 Most Common Words in Delivery



Top5 Most Common Words in Ride Hailing



Frequency analysis

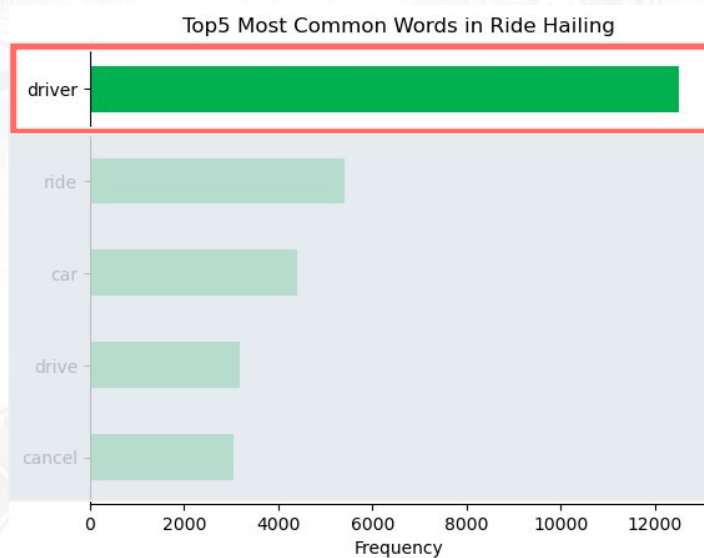
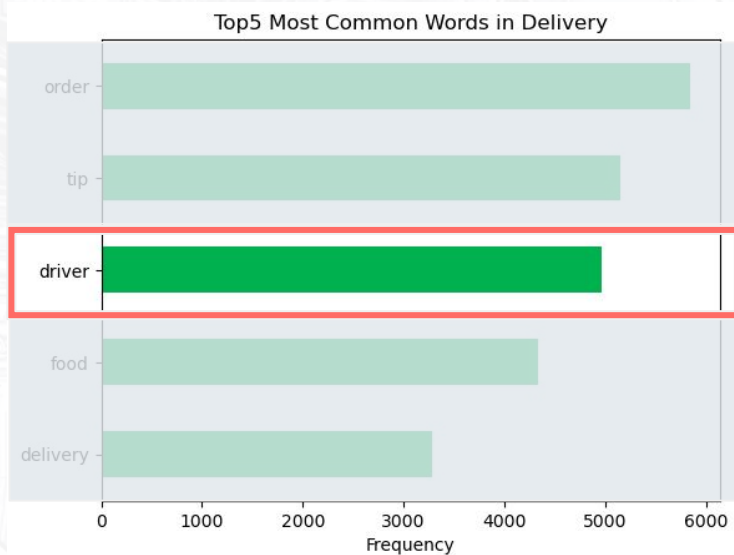


As expected, we see service type specific words rank high in our frequency analysis:

Delivery: **Order, Food, Delivery**

Ride Hailing: **Ride, Car, Drive**

Frequency analysis



Word that rank high in frequency for both.
Example: **Driver**

Large difference in absolute frequencies
Example: **Driver appears much more in ride hailing than in delivery.**



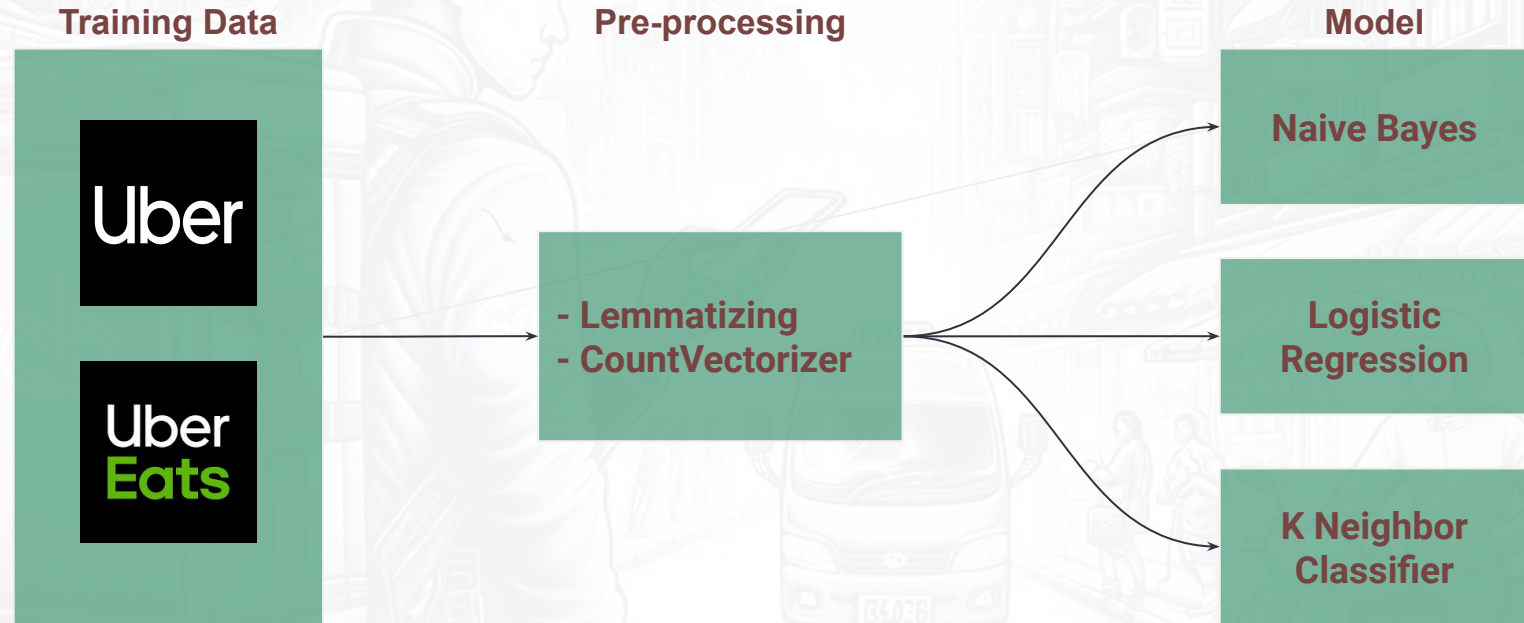
Problem Statement

How can we **distinguish** between customer feedback related to Grab's **ride-hailing** and **delivery** **fast** and **accurately**?



Modelling Process

Modelling Process





Model Selection

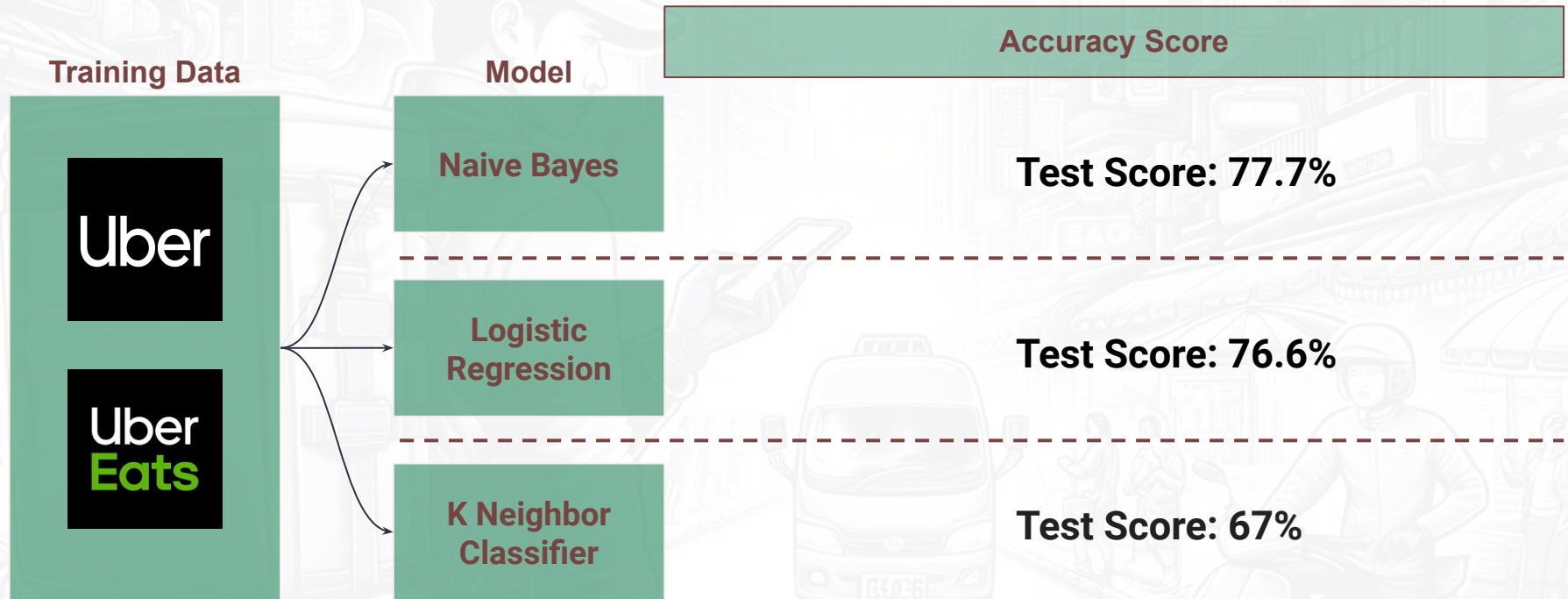
Maximizing Accuracy Score

- **Select model with highest accuracy score**
- **Accuracy score measures the proportion of posts correctly predicted to the total number of posts**

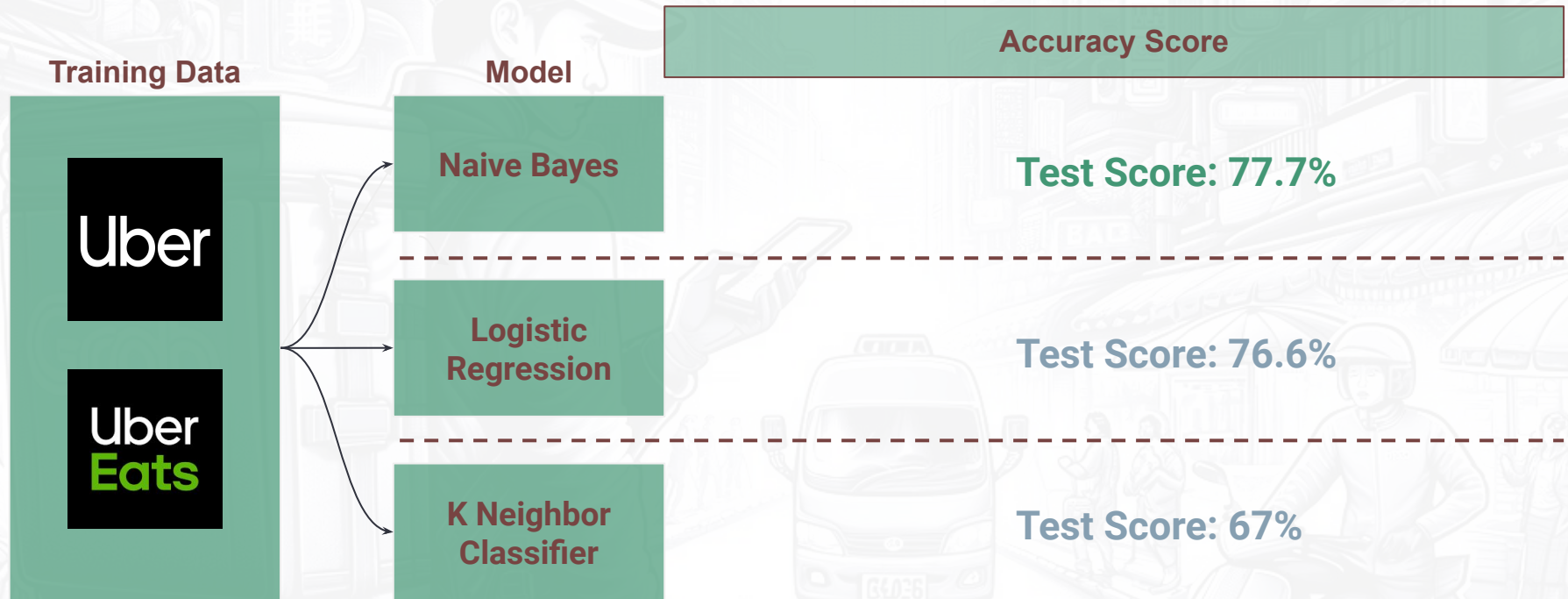
Why?

- **Predicting a ride-hailing post as delivery has the same negative impact as predicting a delivery post as ride-hailing.**

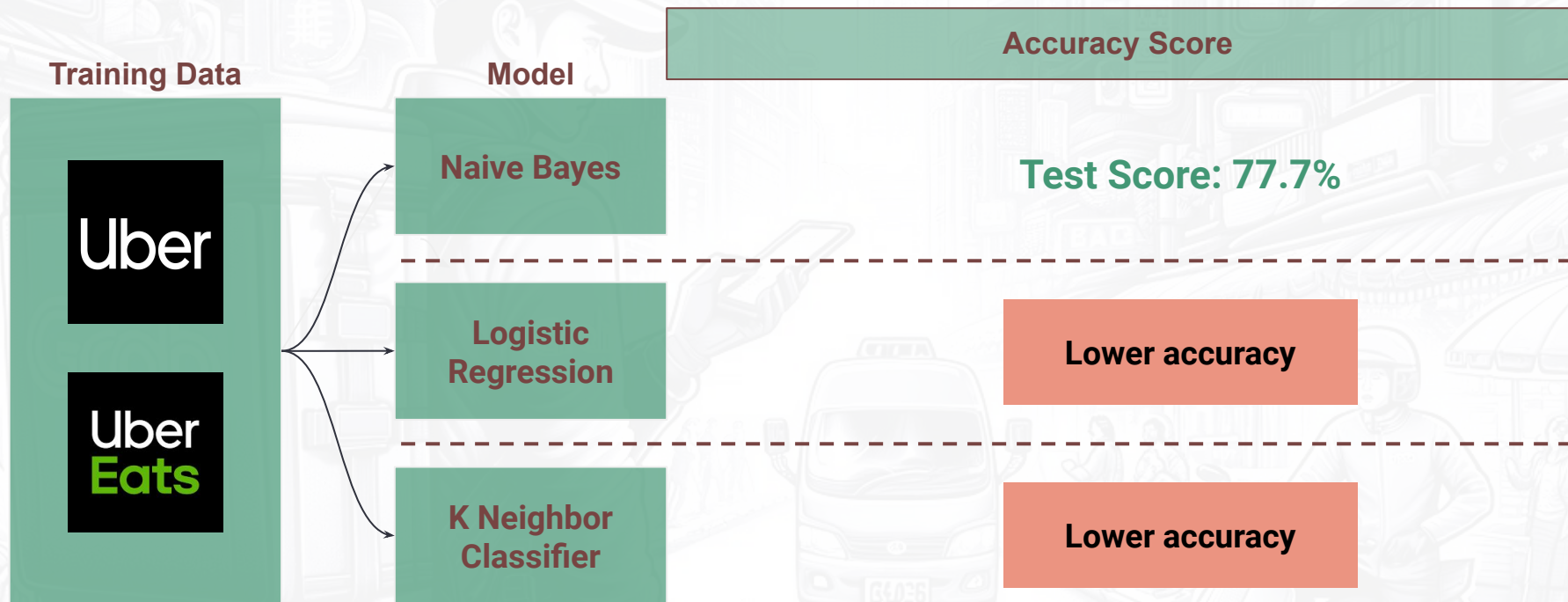
Subject training data to different models...



Subject training data to different models...



Subject training data to different models...



Comparing computational intensity...

Model	Time taken to classify test set
Naive Bayes	1.2 seconds (fastest)
Logistic Regression	1.5 seconds
K Nearest Neighbour	53 seconds

Efficiency: Naive Bayes classifiers are incredibly fast compared to more sophisticated methods.

[Classification Algorithms: KNN, Naive Bayes, and Logistic Regression](#) by Brandon Wohlwend

Comparing computational intensity...

Model	Time taken to classify test set
Naive Bayes	1.2 seconds (fastest)
Logistic Regression	1.5 seconds
K Nearest Neighbour	53 seconds

Efficiency: Naive Bayes classifiers are incredibly fast compared to more sophisticated methods.

[Classification Algorithms: KNN, Naive Bayes, and Logistic Regression](#) by Brandon Wohlwend

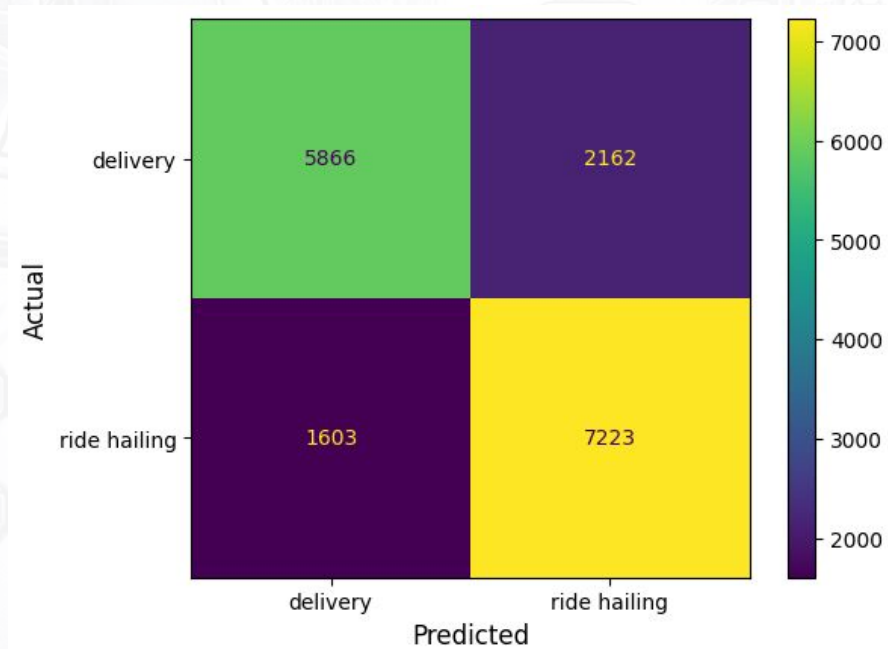
Model Selected - Naives Bayes Model due to:

1. Higher **Accuracy**
2. Least computational Intensive - **Fastest**



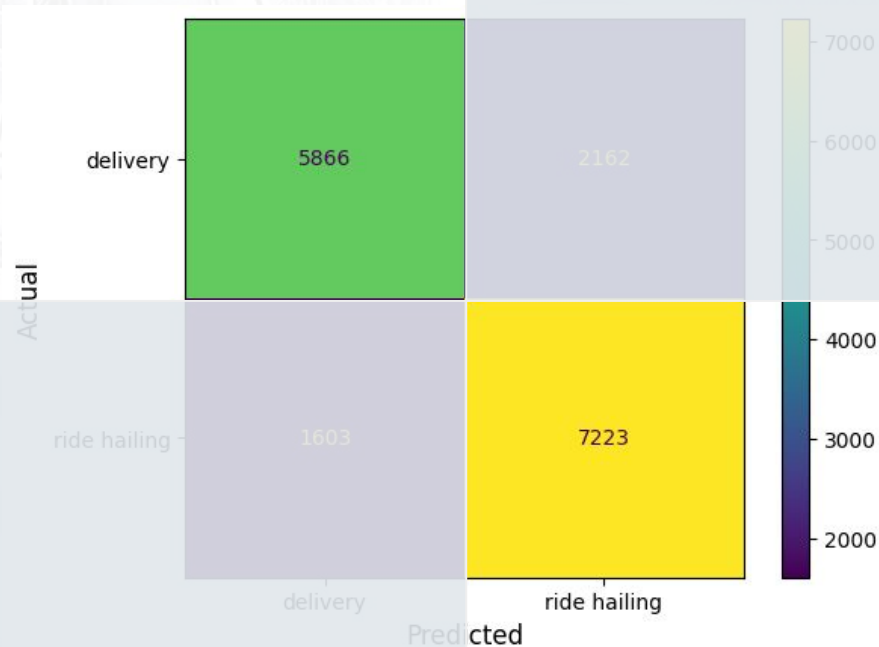
Model Evaluation

Confusion Matrix



Confusion Matrix

**Correctly Predicted
73% of delivery
comments**



**Correctly Predicted
81% of ride hailing
comments**

Confusion Matrix

Wrong
Predicted: Delivery
Actual: Ride Hailing

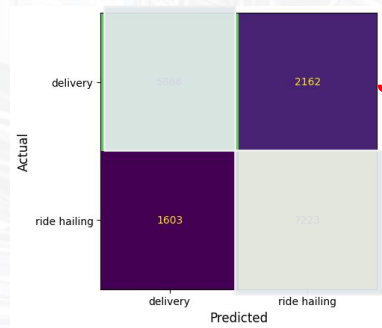


Wrong
Predicted: Ride Hailing
Actual: Delivery

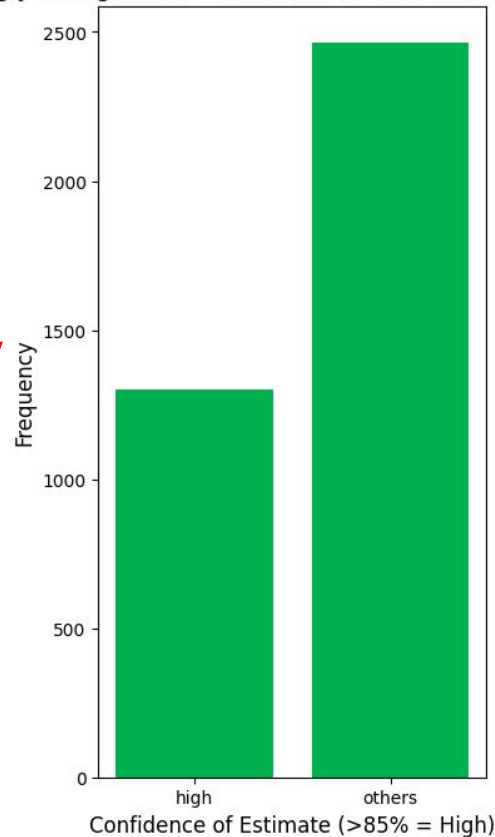
Among those that was wrongly classified....

Actual \ Predicted	delivery	ride hailing
delivery	2088	2162
ride hailing	1603	7223

Among those that was wrongly classified....

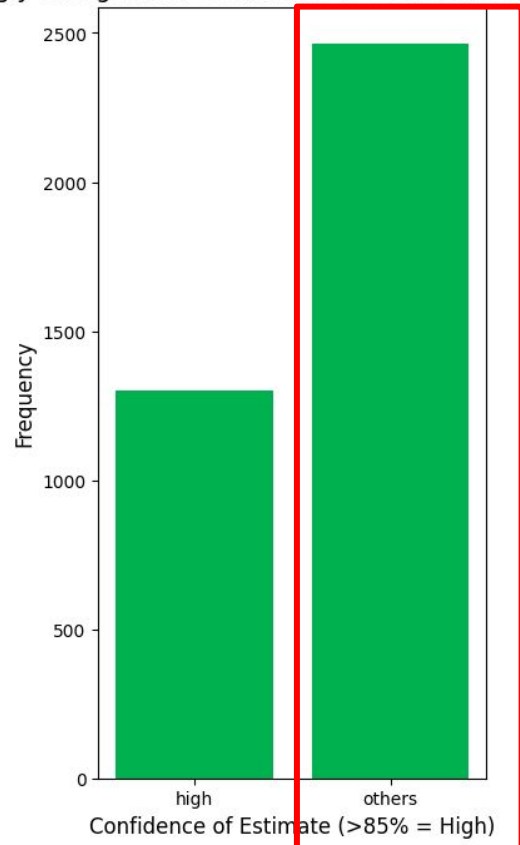


Wrongly Categorized Comments vs Confidence of Estimate



Among those that was wrongly classified....

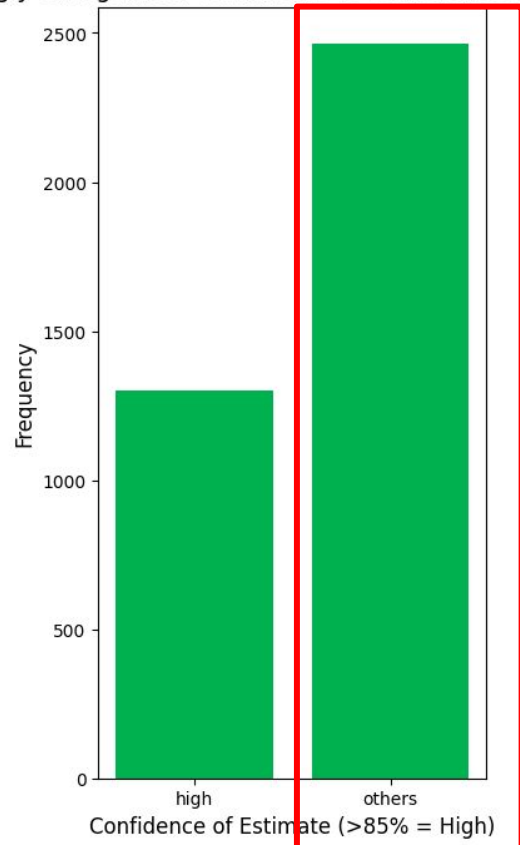
Wrongly Categorized Comments vs Confidence of Estimate



Comments classified with low level of confidence

Among those that was wrongly classified....

Wrongly Categorized Comments vs Confidence of Estimate

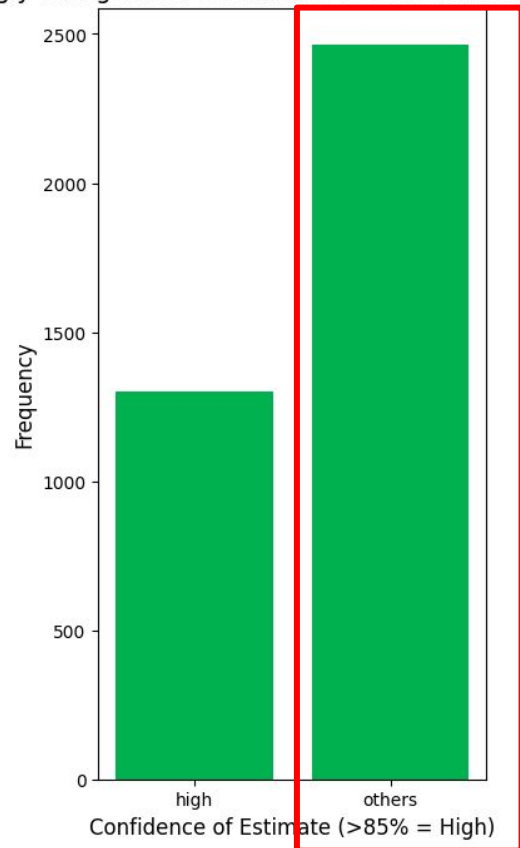


Comments classified with low level of confidence

Model is unsure of which classification comment belongs to.

Among those that was wrongly classified....

Wrongly Categorized Comments vs Confidence of Estimate



Comments classified with low level of confidence

Model is unsure of which classification comment belongs to.

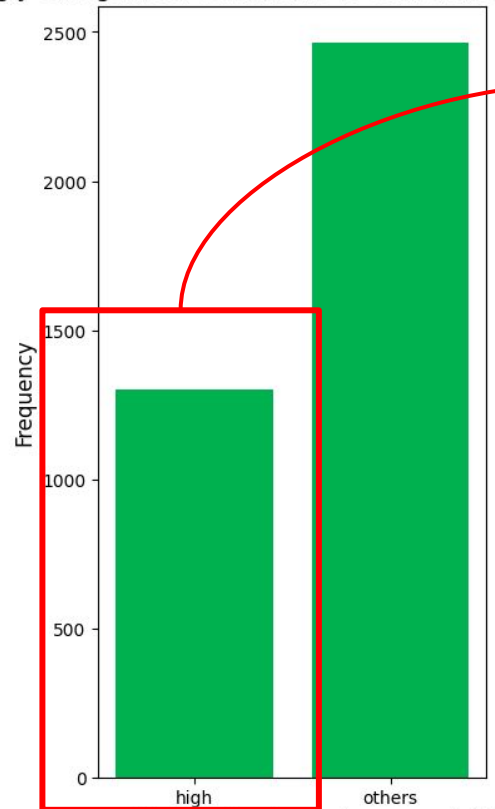
Due to:

General comments

E.g. “glad you could get that off your chest”

Among those that was wrongly classified....

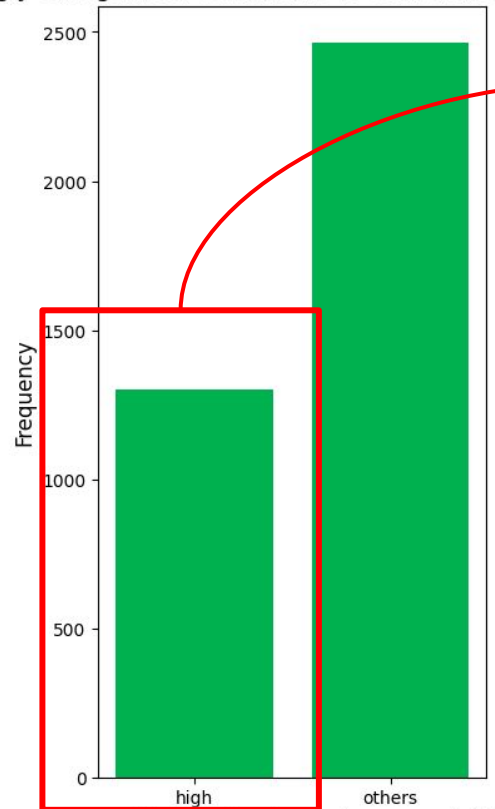
Wrongly Categorized Comments vs Confidence of Estimate



Comments wrongly classified with high level of confidence

Among those that was wrongly classified....

Wrongly Categorized Comments vs Confidence of Estimate

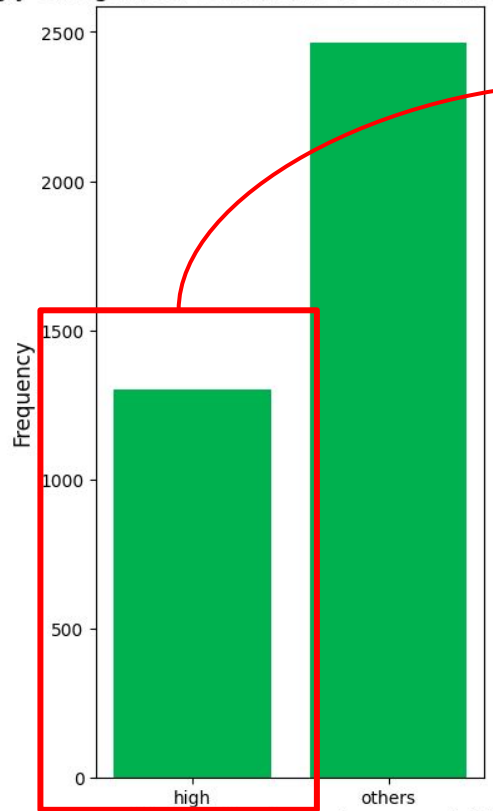


Comments wrongly classified with high level of confidence

Model is very sure, but prediction differs from the actual.

Among those that was wrongly classified....

Wrongly Categorized Comments vs Confidence of Estimate



Comments wrongly classified with high level of confidence

Model is very sure, but prediction differs from the actual.

Due to:

Ride hailing comments that were posted in delivery thread (vice versa)

E.g. "I wouldn't take him & his dog to his destination" in delivery subreddit.



Our classifier in action - App Demo

Our classifier in action - App Demo



Test:
414 Grab App Reviews

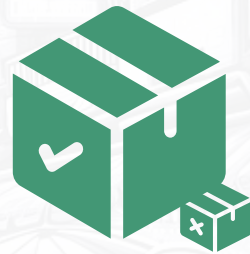
Our classifier in action - App Demo



Test:
414 Grab App Reviews

Model

Our classifier in action - App Demo



Test:
414 Grab App Reviews

Model

Accuracy:
82.37%



Conclusion and Recommendation



Problem Statement:

How can we distinguish between customer feedback related to Grab's ride-hailing and delivery fast and accurately?





Problem Statement:

How can we distinguish between customer feedback related to Grab's ride-hailing and delivery fast and accurately?

Accurate

- 77.7% on sub-reddit test data set
- **82.37%** on Grab app reviews



Problem Statement:

How can we distinguish between customer feedback related to Grab's ride-hailing and delivery fast and accurately?

Accurate

- 77.7% on sub-reddit test data set
- **82.37%** on Grab app reviews

Fast

- Naive Bayes → **computationally inexpensive**



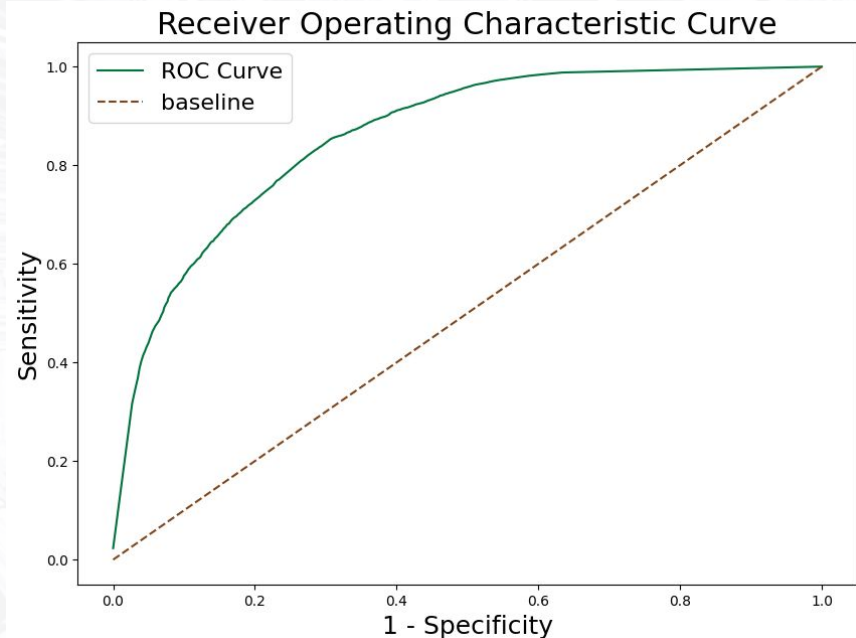
“I saved 2 hours just by using the classifier. I can now focus on more value-adding tasks!”

Recommendation

1. Extend training data to include **Southeast Asia languages** like those from hardware zone forums to address **linguistic nuances** (e.g., Grab use "rider" vs. Uber use "driver").
2. Consider a **multi-class classifier** instead of binary to cover **more categories** (e.g., GrabPay, app performance).



Thank you





Sensitivity \rightarrow True Positive Rate

1 - Specificity \rightarrow False Positive Rate

We can see that our model has a moderate performance.

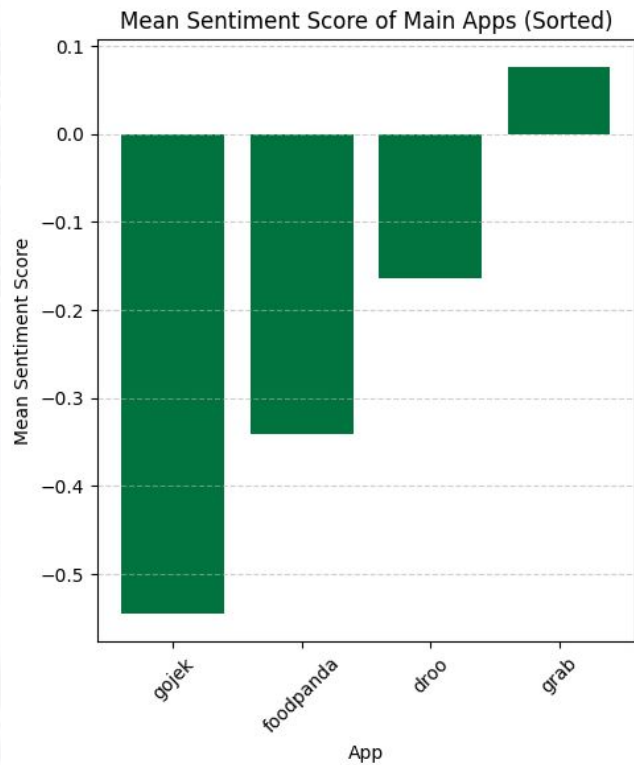
A probability distribution graph will show it better.

Subject training data to different models

Accuracy Score			
Training Data	Model	CountVectorizer	TF-IDF Vectorizer
 	Naive Bayes	Train: 84.5% Test: 77.7%	Train: 93.2% Test: 78.18%
	Logistics Regression	Train: 84.2% Test: 76.6%	Train: 88.5% Test: 77.3%
	K Neighbor Classifier	Train: 69% Test: 67%	Train: 94.2% Test: 58.2%

Sentiment Analysis (Consumer)*

Backup



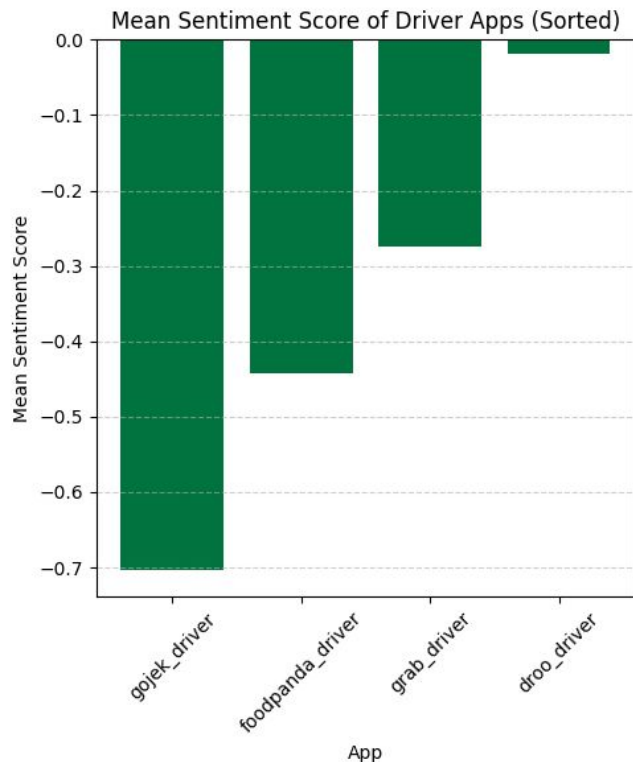
The comments on consumer app is generally negative on all competitors

Grab is the only company that more neutral vs negative.

Grab has the best sentiment score when compared to local competitors

Sentiment Analysis (Driver)*

Backup



Driver reviews have a negative sentiment across all apps.

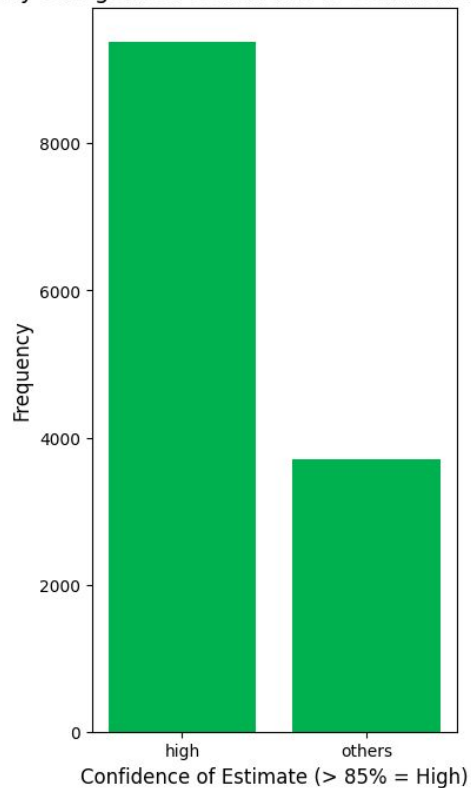
D'roo has the best sentiment score when compared to local competitors for driver app.

There is room for improvement on driver experience in order to better sentiment for Grab

Correctly Classified Comments

Backup

Correctly Categorized Comments vs Confidence of Estimate



Confidence of prediction

Backup

