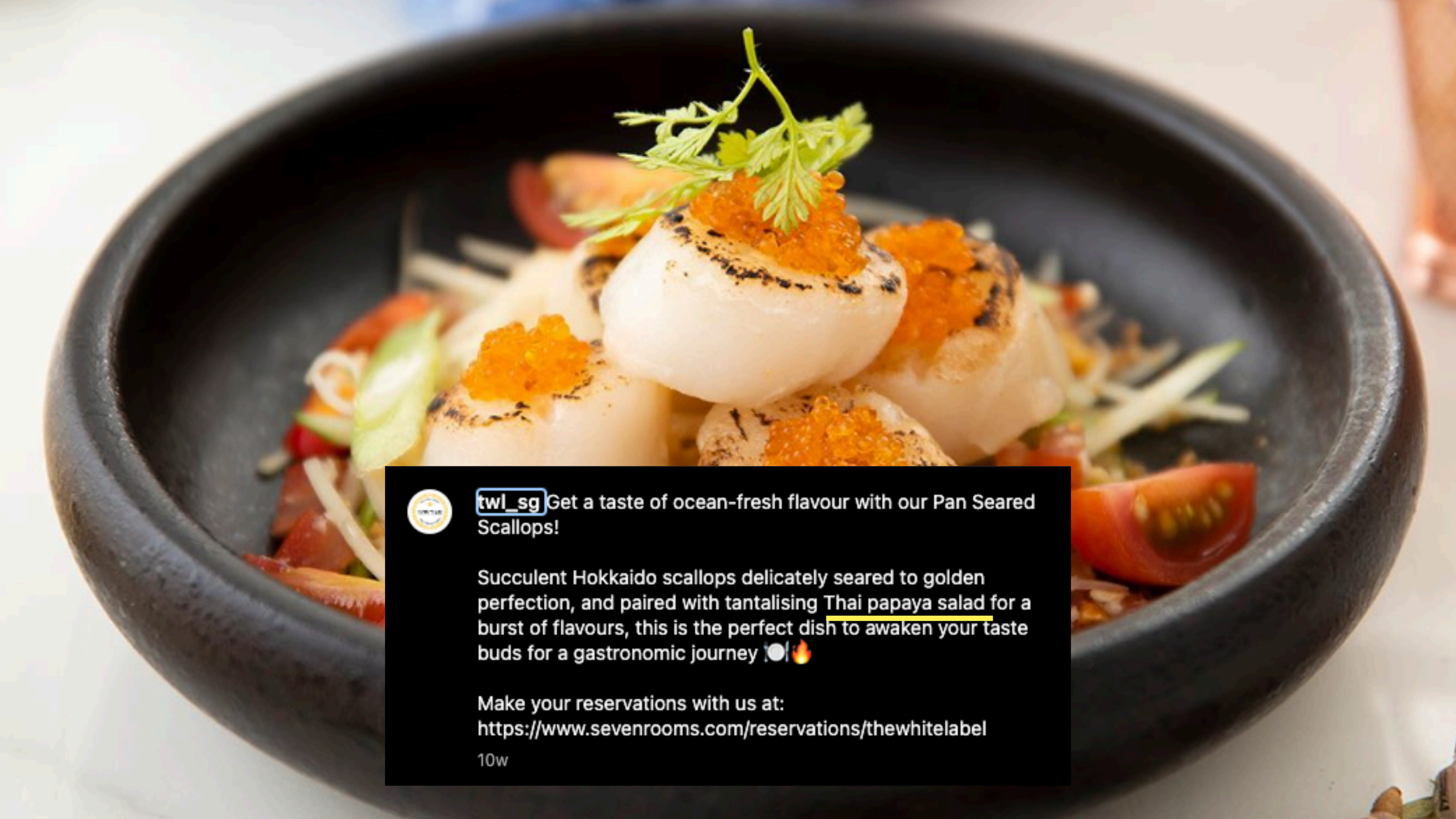
Authentic Food Place

A NLP project that involves topic modelling and classifier models.

Liyena Yusoff



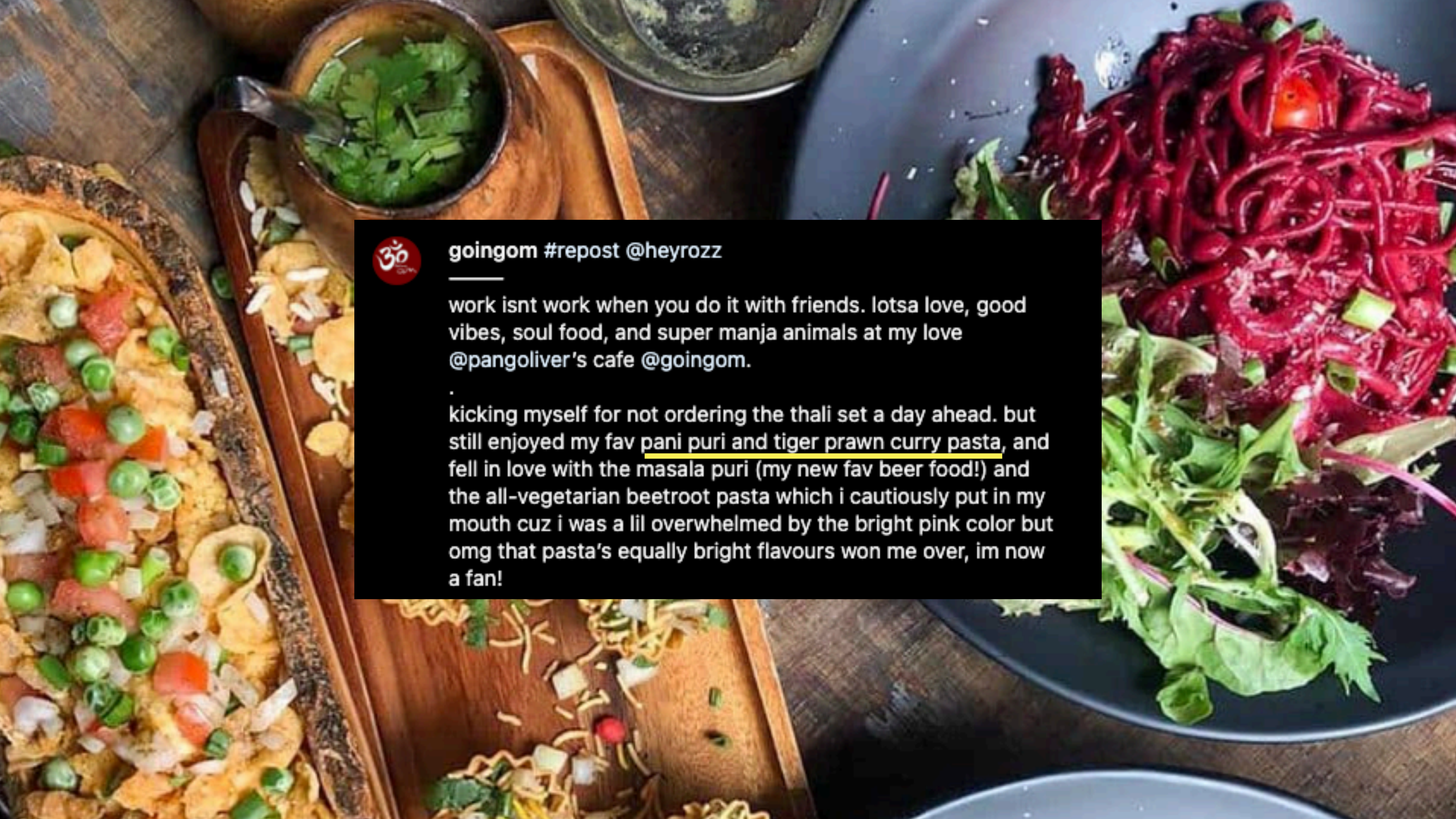
twl_sg Get a taste of ocean-fresh flavour with our Pan Seared Scallops!

Succulent Hokkaido scallops delicately seared to golden perfection, and paired with tantalising Thai papaya salad for a burst of flavours, this is the perfect dish to awaken your taste buds for a gastronomic journey 🍽️🔥

Make your reservations with us at:
https://www.sevenrooms.com/reservations/thewhitelabel

10w

themalayancouncil.sg Come join us to savour our popular Roti Kirai Beef Ribs! If two pieces of Roti Kirai is not enough, you can always opt for extras. 😉
Edited · 3d



goingom #repost @heyrozz

work isnt work when you do it with friends. lotsa love, good vibes, soul food, and super manja animals at my love @pangoliver's cafe @goingom.

.

kicking myself for not ordering the thali set a day ahead. but still enjoyed my fav pani puri and tiger prawn curry pasta, and fell in love with the masala puri (my new fav beer food!) and the all-vegetarian beetroot pasta which i cautiously put in my mouth cuz i was a lil overwhelmed by the bright pink color but omg that pasta's equally bright flavours won me over, im now a fan!

RUMAH MAKAN
MINANG

The
DIM SUM PLACE包

Komala
EST Vilas 1947

Problem Statement

With the growing challenge of finding truly authentic food, this project aims to recommend authentic dining experiences based on consumer preferences and guide restaurants in aligning traditional menus with authentic tastes.

This addresses the difficulty customers face in navigating a culinary landscape where traditional flavors may be diluted or overshadowed by modern interpretations.

Objectives

- Analyze TripAdvisor restaurant reviews to identify their sentiments for the authenticity of the dishes.
- Develop a **NPS-like score** to rank restaurants.
- Develop a model that could **classify** TripAdvisor reviews into detractor or promoter of the restaurant.
- Provide actionable insights for both consumers seeking authentic experiences and restaurants aiming to highlight their authenticity.

Workflow

Web scraping

Restaurant reviews from TripAdvisor

Data cleaning

- restaurant name
- tokenized reviews
- lemmatized words

Exploratory Analysis

- top words (overall)
- ‘authentic’ counts
- sentiment analysis

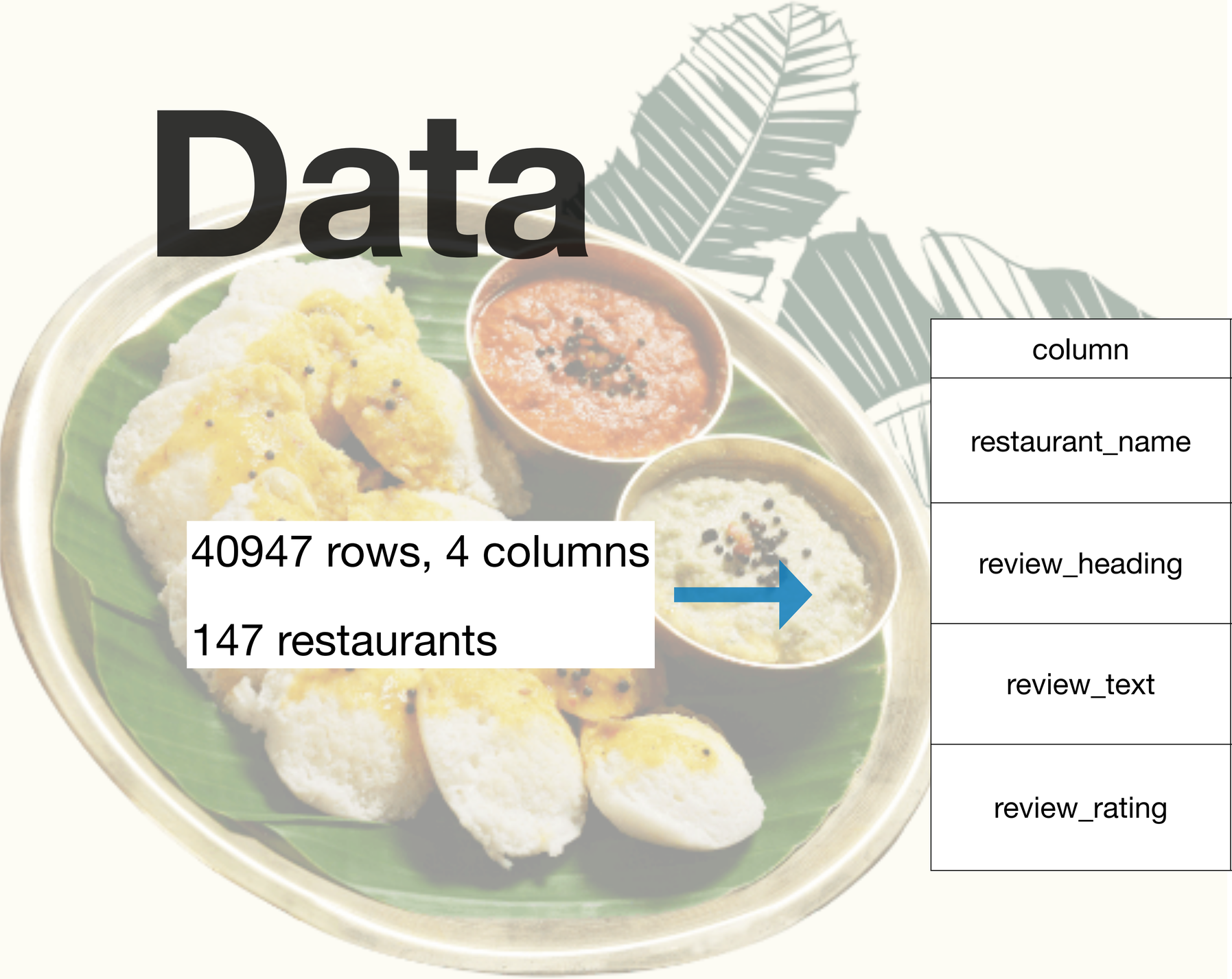
Pre-processing
Modelling

1. Topic Modelling
2. Classification

App

Streamlit

Data



40947 rows, 4 columns

147 restaurants

column	data type / info
restaurant_name	string
review_heading	string
review_text	string
review_rating	float 1.0 to 5.0

Workflow

Web scraping

Restaurant reviews from TripAdvisor

Data cleaning

- restaurant name
- tokenized reviews
- lemmatized words

Exploratory Analysis

- top words (overall)
- ‘authentic’ counts
- sentiment analysis

Pre-processing Modelling

1. Topic Modelling
2. Classification

App

Streamlit

Data cleaning

restaurant	“SEARCH("Reviews-", A2)”	“SEARCH(“-Singapore", A2)”	“ =MID(A2, B2 + LEN("Reviews-"), C2 - B2 - LEN(“Reviews-")) ”	“ =SUBSTITUTE(D2, "_", " “) ”
https://www.tripadvisor.com.sg/Restaurant_Review-g294265-d10460729-Reviews-Pistachio_Middle_Eastern_Mediterranean_Grill-Singapore.html#REVIEWS	68	120	Pistachio_Middle_Eastern_Mediterranean_Grill	Pistachio Middle Eastern Mediterranean Grill

Data cleaning

Raw Text

I’ve come to entre nous periodically over the past 10 years it seems and the service and food have always been consistently great. Geraldine and her husband are clearly invested in maintaining the highest standards and I’m so happy the restaurant is still going strong...even through COVID

Data cleaning

Raw Text

I've come to entre nous periodically over the past 10 years it seems and the service and food have always been consistently great. Geraldine and her husband are clearly invested in maintaining the highest standards and I'm so happy the restaurant is still going strong...even through COVID

Cleaned Text

come entre nous periodically past years seems service food always consistently great geraldine husband clearly invested maintaining highest standards happy restaurant still going strong even covid

Data cleaning

Raw Text

I've come to entre nous periodically over the past 10 years it seems and the service and food have always been consistently great. Geraldine and her husband are clearly invested in maintaining the highest standards and I'm so happy the restaurant is still going strong...even through **COVID**

Cleaned Text

come entre nous periodically past years seems service food always consistently great geraldine husband clearly invested maintaining highest standards happy restaurant still going strong even **covid**

Data cleaning

Cleaned Text

come entre nous
periodically past years
seems service food
always consistently great
geraldine husband clearly
invested maintaining
highest standards happy
restaurant still going
strong even covid

Cleaned Lemmatized Text

come entre nous
periodically past year
seem service food
always consistently
great geraldine
husband clearly invest
maintain high standard
happy restaurant still go
strong even covid

Data cleaning

Cleaned Text

come entre nous periodically past **years** **seems** service food always consistently great geraldine husband clearly **invested** **maintaining** **highest** **standards** happy restaurant still **going** strong even covid

Cleaned Lemmatized Text

come entre nous periodically past **year** **seem** service food always consistently great geraldine husband clearly **invest** **maintain** **high** **standard** happy restaurant still **go** strong even covid

Data cleaning

cleaned and lemmatized

Nan

Workflow

Web scraping

Restaurant reviews
from TripAdvisor

Data cleaning

- restaurant name
- tokenized reviews
- lemmatized words

Exploratory Analysis

- top words (overall)
- ‘authentic’ counts
- sentiment analysis

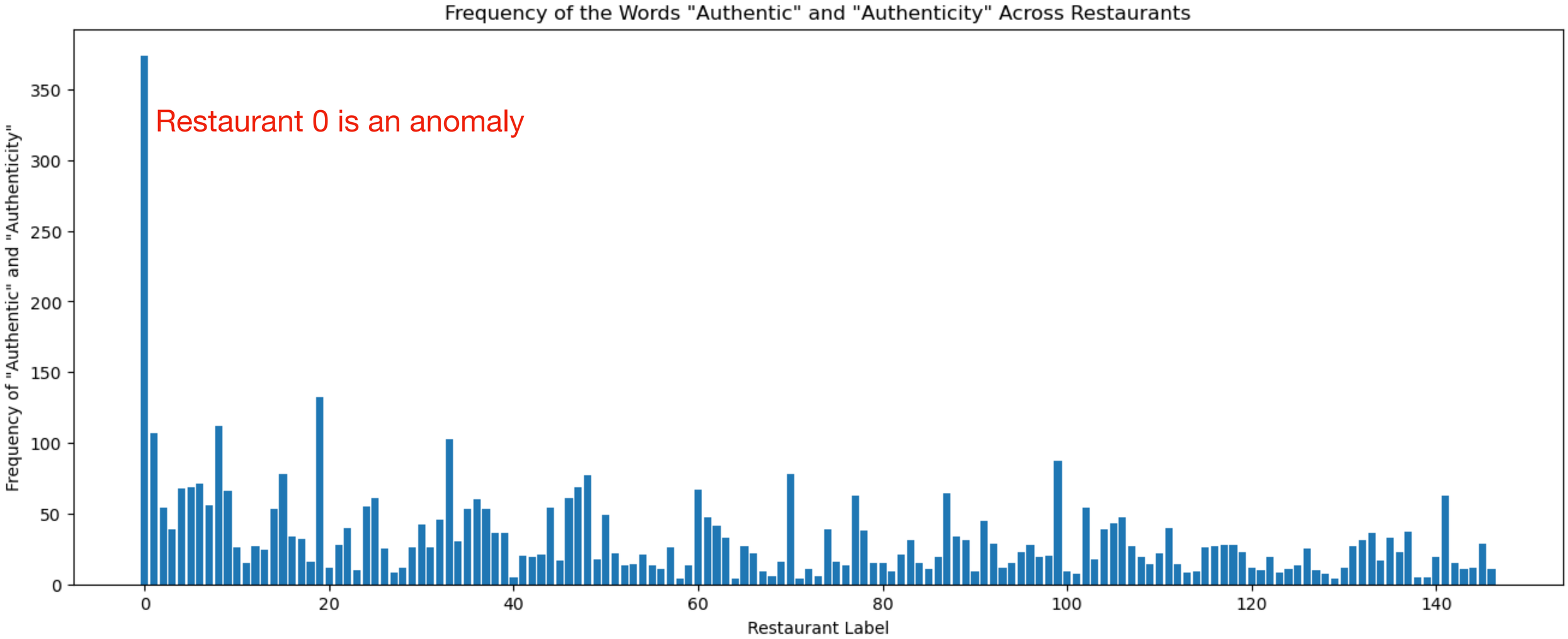
Pre-processing
Modelling

1. Topic Modelling
2. Classification

App

Streamlit

Few reviews comment on the authentic taste of the dish

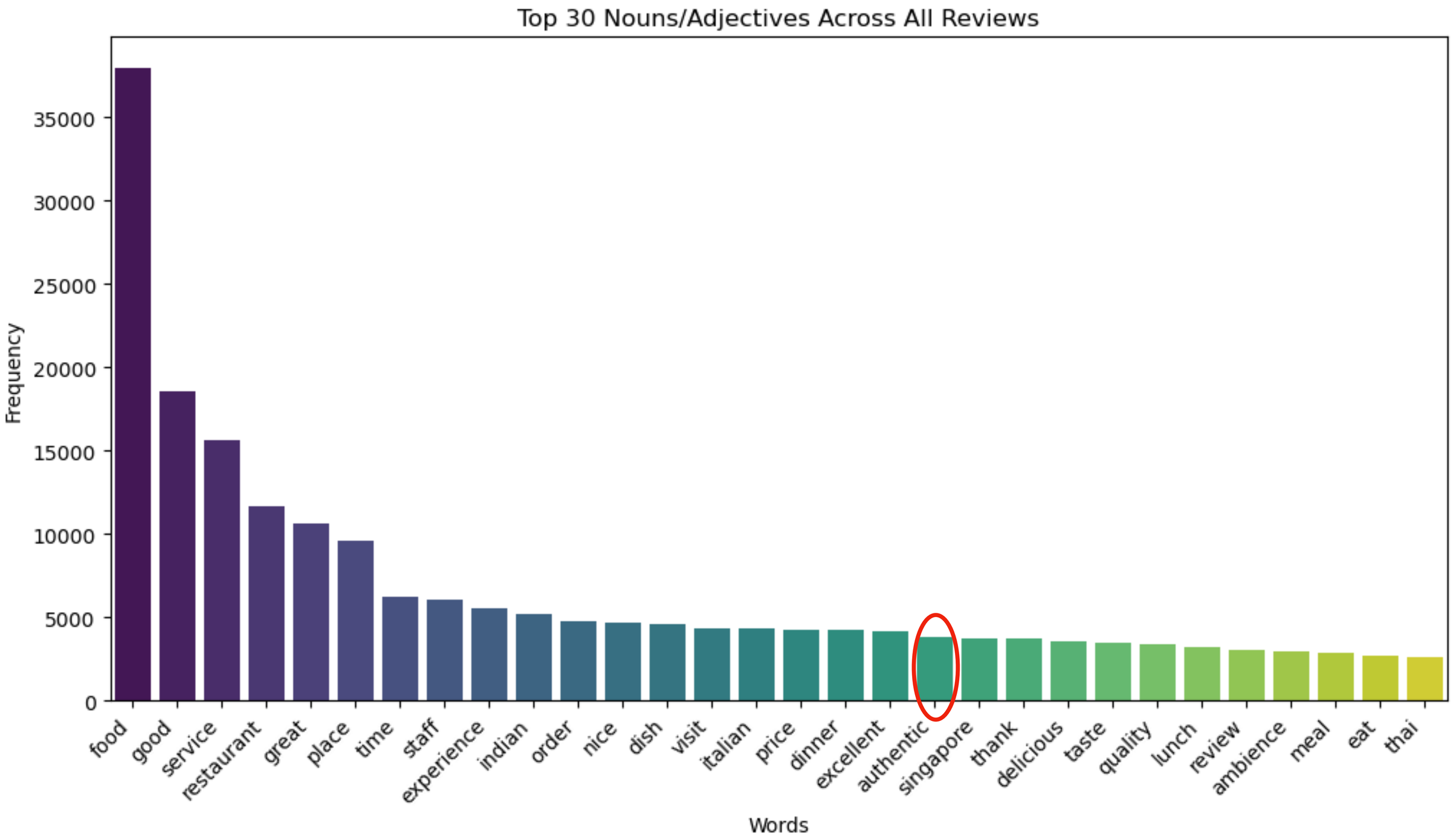


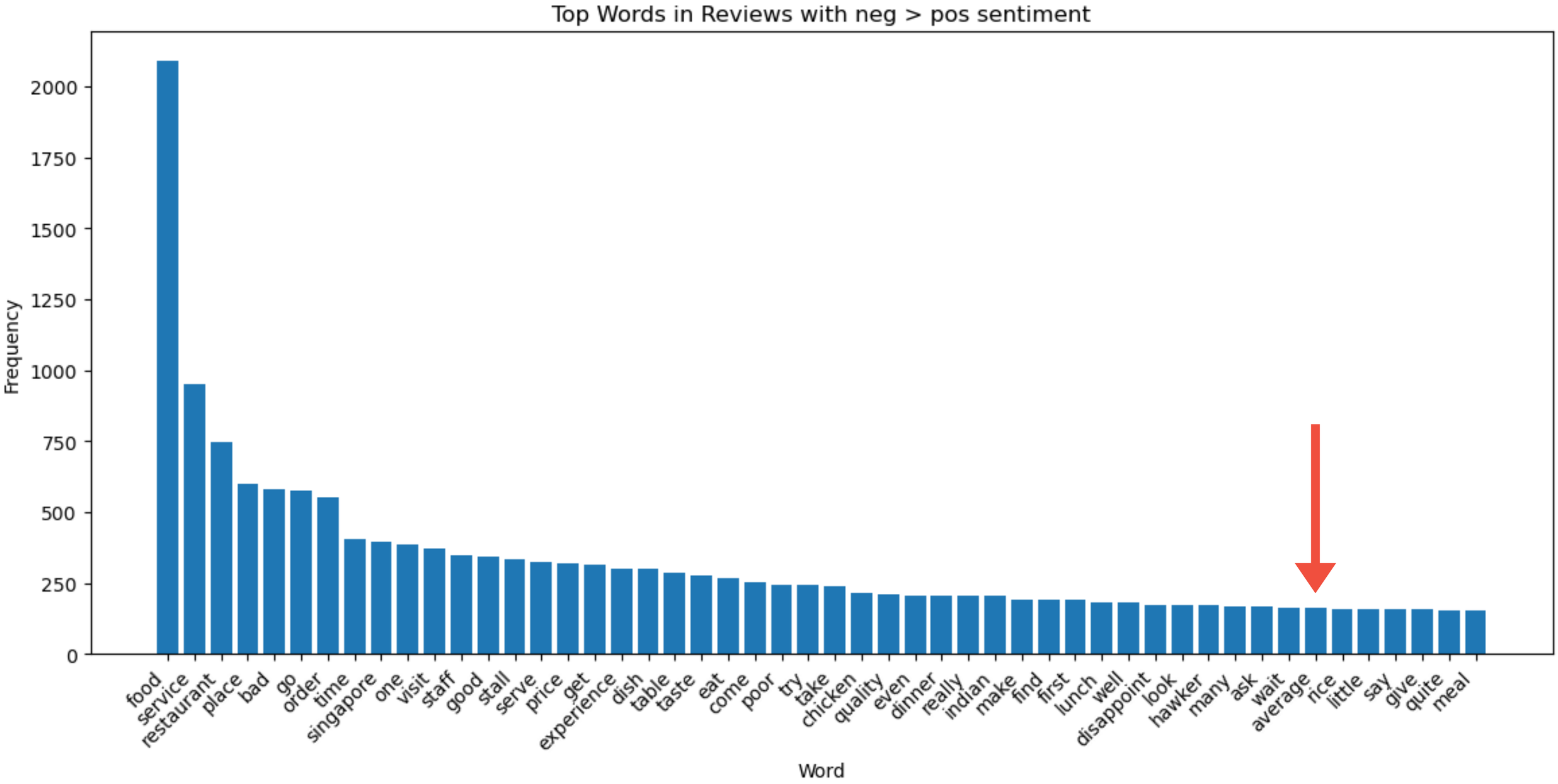
Top Words for Restaurant 0

crepe
thank
much
nous
creperie
entre
nous
delicious
dear
authentic
gluten free
sweet crepe
galette
recommend
entre
take time
great
lunch
friendly
french crepe
recommend
welcome
soon back
always
taking time
cr pe
find
happy know
excellent
owner
definitely
savoury sweet
really
authentic french
little
dinner
salt caramel
thank
crepe dear
perfect
much recommend
price
option
menu
disappoint
like
geraldine
cider
brittany crepe
buckwheat crepe
meal
last week
husband
hear enjoy
raffle hotel
wonderful review
review recommend
would
enjoy crepe
glad know
staff
much taking
always pleasure
amazing comment
choice
write lovely
still
great day
sunday
delivery
drink
ever
amazing review
authentic crepe
crepe delicious
see soon
know love
guest
eat
must
breton
escargot
people
much dining
atmosphere
simple
favorite
twice
wait
right
kind word
quality
brittany galette
french cr
homemade
even
real
set
warm
cozy
amazing
forward welcome
everything
wife
crispy
yummy
birthday
apple cider
really happy
small
day hope
get
home
savoury
ingredient
want
last weekend
evening
fill
savory
enjoy
happy hear
say
think
couple
sg
since
bring
gen
cr perie
write amazing
real pleasure
post wonderful
delighted hear
creperie last
come back
attentive
bit
quite
crepe ever
tasty
taste
france
hope
welcome
lovely
visit
write wonderful
enjoy brittany
much recently
family
savoury crepe
love
time
write
choose entre
brittany
ice cream
fantastic
lovely review
know appreciate
year
highly recommend
creperie take
crepe artisanal
favourite
could
super
first
mean lot
crepe great
especially
dessert
never
creperie happy
look
offer
come
know
time recommend
make
much take
experience
friend
quality
wonderful comment
we
recommend
know enjoy
look forward
time
lady
we happy
sweet
delighted know
delicious
sweet crepe
dear

Top Words Across All Reviews

Top Words Across All Reviews


Top 30 Nouns/Adjectives Across All Reviews
Frequency
35000
30000
25000
20000
15000
10000
5000
0
food
good
service
restaurant
great
place
time
staff
experience
indian
order
nice
dish
visit
italian
price
dinner
excellent
authentic
singapore
thank
delicious
taste
quality
lunch
review
ambience
meal
eat
thai
Words


Top Words in Reviews with neg > pos sentiment
Frequency
0
250
500
750
1000
1250
1500
1750
2000
Word
food
service
restaurant
place
bad
go
order
time
singapore
one
visit
staff
good
stall
serve
price
get
experience
dish
table
taste
eat
come
poor
try
take
chicken
quality
even
dinner
really
indian
make
find
first
lunch
well
disappoint
look
hawker
many
ask
wait
average
rice
little
say
give
quite
meal

Authenticity in other words/phrases…

Authenticity in other words/phrases…

“Feel like in itsly.”

“classic”

“traditional”

“best regional restaurant”

“average food”

“Taste has nothing to do with the originals.”

“Just a tourist trap.”

“unpretentious italian dishes”

“average ____ restaurant”

“A place where you can find the classic hawker food found in Malaysia”

Workflow

Web scraping

Restaurant reviews
from TripAdvisor

Data cleaning

- restaurant name
- tokenized reviews
- lemmatized words

Exploratory Analysis

- top words (overall)
- ‘authentic’ counts
- sentiment analysis

Pre-processing
Modelling

1. Topic Modelling
2. Classification

App

Streamlit

1 Topic modelling - pre-trained BERT

Bidirectional
Encoder
Representations from
Transformers

1 Topic modelling - pre-trained BERT

Bidirectional Encoder Representations from Transformers

BERT is a popular large language model created and published in 2018.

Its original model was trained on over 800 million words from BooksCorpus and over 2.5 billion words from Wikipedia. It was originally trained on two tasks: language modeling and next sentence prediction.

Source: Python Topic Modelling with BERT model (2022)
https://deepgram.com/learn/python-topic-modeling-with-a-bert-model

1 Topic modelling - pre-trained BERT

Bidirectional Encoder Representations from Transformers

Topics = ‘food’, ‘service’, ‘ambiance’, ‘authentic’

chao serve one good authentic teochew cuisine singapore small restaurant table well make reservation avoid disappointment going wait seat outside restaurant face noisy beach roadchao serve authentic teochew cuisine

food_score = 0.379

service_score = 0.210

ambiance_score = 0.241

authentic_score = 0.490

overall_score = 0.330

Overall score

Topics = ‘food’, ‘service’, ‘ambiance’, ‘authentic’

Overall score

Topics = ‘food’, ‘service’, ‘ambiance’, ‘authentic’

Individual reviews

whether the review is a detractor or promoter

>= 0.3 **promoter**

< 0.3 detractor

Each restaurant

NPS score

Overall score

Topics = ‘food’, ‘service’, ‘ambiance’, ‘authentic’

Individual reviews

whether the review is a detractor or promoter

>= 0.3 **promoter**

< 0.3 detractor

Each restaurant

NPS score

Average overall score grouped reviews by restaurant

Compare themselves against competitors

Overall score

Topics = ‘food’, ‘service’, ‘ambiance’, ‘authentic’

Individual reviews

whether the review is a detractor or promoter

>= 0.3 **promoter**

< 0.3 detractor

Each restaurant

NPS score

Average overall score grouped reviews by restaurant

Compare themselves against competitors

Through

% of **promoter**-% of detractor

:) or :(

Overall score

Topics = ‘food’, ‘service’, ‘ambiance’, ‘authentic’

Individual reviews

whether the review is a detractor or promoter

>= 0.3 **promoter**

< 0.3 detractor

2 Classification modelling

Each restaurant

NPS score

Average overall score grouped reviews by restaurant

Compare themselves against competitors

Through

% of **promoter**-% of detractor

:) or :(

Imbalanced promoter (1) and detractor (0)

Before resampling

1	22967
0	9789

After SMOTE

1	22967
0	22967

After TomekLinks

1	22967
0	9789

Imbalanced promoter (1) and detractor (0)

Before resampling

1	22967
0	9789

After SMOTE

1	22967
0	22967

After TomekLinks

1	22967
0	9789

Modelling and tuning - XGBoost

vectorizer	classifier	f1_train	f1_test	f1_diff	elapsed_time
TF-IDF	Logistic Regression	0.99898	0.89608	0.10290	148.922
TF-IDF	**XGBoost**	**0.94997**	**0.89454**	**0.05543**	**168.464**
TF-IDF	XGBoost	0.94837	0.89400	0.05437	942.953
TF-IDF	Logistic Regression	0.98602	0.89191	0.09411	45.061
Count Vectorizer	Multinomial Naive Bayes	0.91391	0.87308	0.04083	6.069
TF-IDF	Multinomial Naive Bayes	0.99806	0.84533	0.15273	18.784
TF-IDF	Random Forest	0.72504	0.81488	-0.08884	3936.678
TF-IDF	Random Forest (baseline)	0.73618	0.78897	-0.02579	3197.552

TF-IDF adds weight to words.

taking into account unique words that could improve reviews classification

Modelling and tuning - XGBoost

vectorizer	classifier	f1_train	f1_test	f1_diff	elapsed_time
TF-IDF	Logistic Regression	0.99898	0.89608	0.10290	148.922
TF-IDF	**XGBoost**	**0.94997**	**0.89454**	**0.05543**	**168.464**
TF-IDF	XGBoost	0.94837	0.89400	0.05437	942.953
TF-IDF	Logistic Regression	0.98602	0.89191	0.09411	45.061
Count Vectorizer	Multinomial Naive Bayes	0.91391	0.87308	0.04083	6.069
TF-IDF	Multinomial Naive Bayes	0.99806	0.84533	0.15273	18.784
TF-IDF	Random Forest	0.72504	0.81488	-0.08884	3936.678
TF-IDF	Random Forest (baseline)	0.73618	0.78897	-0.02579	3197.552

Informs restaurants <u>how well</u> the model could identify if a review is a detractor or promoter of the restaurant.

Workflow

Web scraping

Restaurant reviews
from TripAdvisor

Data cleaning

- restaurant name
- tokenized reviews
- lemmatized words

Exploratory Analysis

- top words (overall)
- ‘authentic’ counts
- sentiment analysis

Pre-processing
Modelling

1. Topic Modelling
2. Classification

App

Streamlit

App - Restaurant and food recommender chatbot

Retrieval Augmented Generation (RAG)

App - Restaurant and food recommender chatbot

Retrieval Augmented Generation (RAG)

Data in csv format

	restaurant	review_text	review_rating	overall_score
0	Entre Nous creperie	come entre nous periodically past y	5	0.42841488
1	Entre Nous creperie	absolutely delicious menu lovely off	5	0.35523906
2	Entre Nous creperie	thank much choose entre nous crep	5	0.3769728
3	Entre Nous creperie	lovely little french restaurant really	4.5	0.37279508
4	Entre Nous creperie	dear rebecca thank much dining ent	4	0.37502688

App - Restaurant and food recommender chatbot

Retrieval Augmented **Generation** (RAG)

Data in csv format

	restaurant	review_text	review_rating	overall_score
0	Entre Nous creperie	come entre nous periodically past y	5	0.42841488
1	Entre Nous creperie	absolutely delicious menu lovely off	5	0.35523906
2	Entre Nous creperie	thank much choose entre nous crep	5	0.3769728
3	Entre Nous creperie	lovely little french restaurant really	4.5	0.37279508
4	Entre Nous creperie	dear rebecca thank much dining ent	4	0.37502688

Train questions

Would you recommend Paulaner Brauhaus Singapore to someone looking for authentic food?

What is the overall rating of Beng Hiang Restaurant?

How would you rate the authenticity of the food at Khansama Tandoori Restaurant?

App - Restaurant and food recommender chatbot

Retrieval Augmented Generation (RAG)

Train questions

Would you recommend Paulaner Brauhaus Singapore to someone looking for authentic food?

What is the overall rating of Beng Hiang Restaurant?

How would you rate the authenticity of the food at Khansama Tandoori Restaurant?

Evaluation questions

Is Entre Nous creperie known for serving authentic food?

Did the restaurant offer authentic Greek cuisine?

Were the food portions at BAKALAKI Greek Taverna worth the price?

App - Restaurant and food recommender chatbot

Retrieval Augmented Generation (RAG)

Train questions

Would you recommend Paulaner Brauhaus Singapore to someone looking for authentic food?

What is the overall rating of Beng Hiang Restaurant?

How would you rate the authenticity of the food at Khansama Tandoori Restaurant?

Evaluation questions

Is Entre Nous creperie known for serving authentic food?

Did the restaurant offer authentic Greek cuisine?

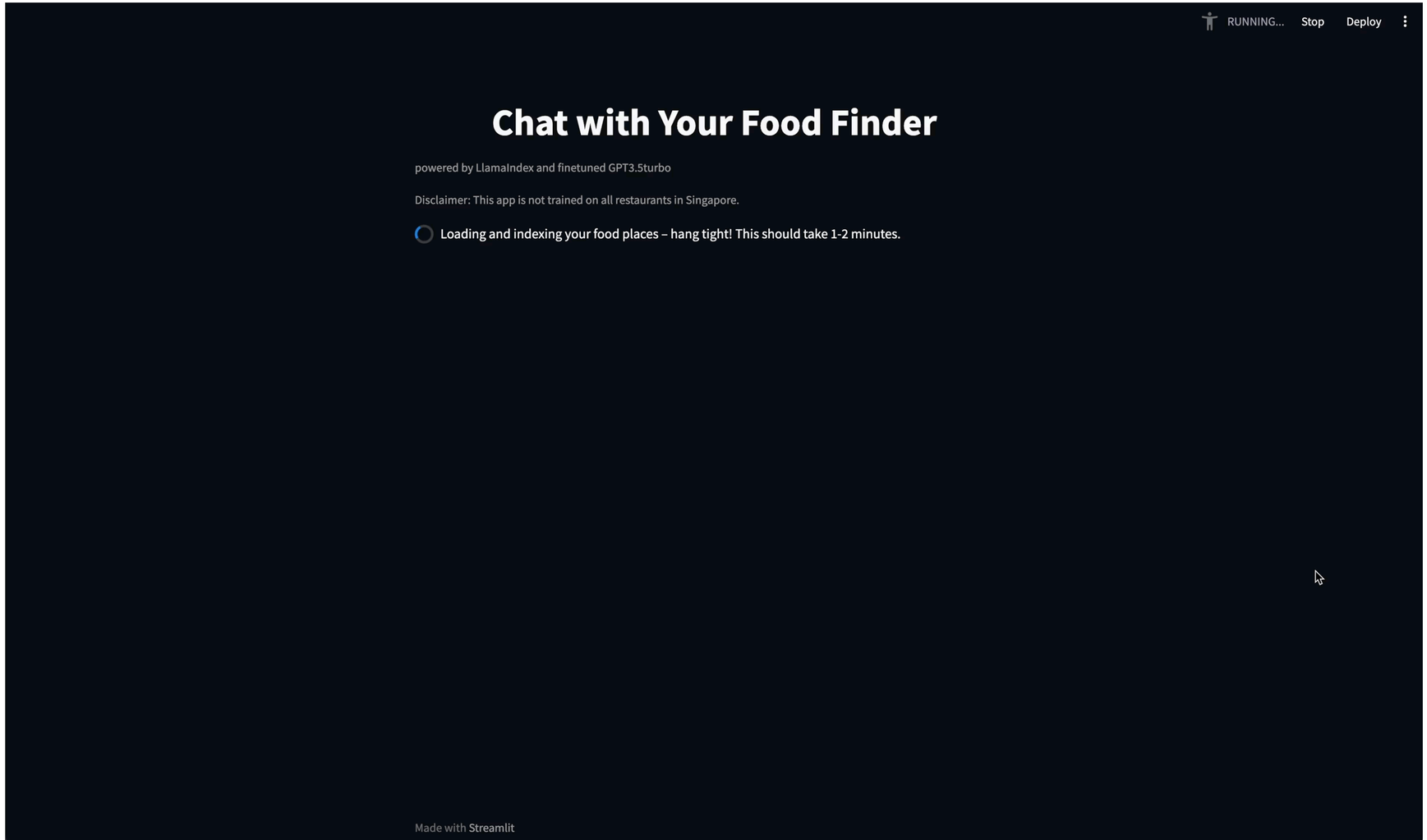
Were the food portions at BAKALAKI Greek Taverna worth the price?

{'faithfulness': 0.9125, 'answer_relevancy': 0.9643}

measures factual accuracy of the generated answers

measures how relevant the answers are to the question

App demo


RUNNING...
Stop
Deploy
Chat with Your Food Finder
powered by LlamaIndex and finetuned GPT3.5turbo
Disclaimer: This app is not trained on all restaurants in Singapore.
Loading and indexing your food places – hang tight! This should take 1-2 minutes.
Made with Streamlit

RUNNING... Stop Deploy

Chat with Your Food Finder

powered by LlamaIndex and finetuned GPT3.5turbo

Disclaimer: This app is not trained on all restaurants in Singapore.

Loading and indexing your food places – hang tight! This should take 1-2 minutes.

Made with Streamlit

Limitations

App

- limited to the reviews and the details of restaurants that are trained on
- needs specificity, ie which restaurant are we referring to?

Data

- reviews may be biased and subjective
- require periodic checks for updates on NPS-like score

Further improvements

- Experiment with prompt engineering
- Real-Time NPS Score Monitoring for Businesses/Restaurants
- Cuisine Classification

Recommendations

Reviewers/Customers

- describe the food with extra details

eg. authentic nasi lemak with a rich coconut milk flavoured rice, with sambal that is sweet but not that spicy and crispy, tender chicken.

Restaurants/Businesses

- Cultural events and collaborations

eg. Singapore Food Festival, Ramadhan Bazaar, Food Expo

- Actively seeking customer feedback on food authenticity

Thank You

Sources - images

- Instagram accounts: @twl_sg, @themalayancouncil.sg, @goingom
- Komala Vilas https://komalavilas.com.sg/
- Rumah Makan Minang https://minang.sg/
- The Dim Sum Place https://www.thedimsumplace.sg/