

An abstract black and white wavy pattern, resembling a topographical map or a liquid marbled effect, occupies the left half of the slide. It features thick black lines and thinner grey lines flowing and swirling across a white background.

SC1015 - SC3

INTRO TO DATA SCIENCE AND AI

{Andrew *Leung* | Daniel *Tay* | Jonathan *Ng*}

TABLE OF CONTENTS

01

MOTIVATION

Why Food Delivery?

02

EDA, CLEANING

Insights on consumer demographics, prepping data

03

DATA PREP, RESAMPLING

Classification Modelling

04

CONCLUSION

Results, Evaluation and Conclusion

05

CONCLUSION

Results, Evaluation and Conclusion

01

MOTIVATION

Why Food Delivery?



COVID-19 AND DELIVERY SERVICES

- Impact of COVID-19 caused many stranded in their homes
- Many unable to leave their houses to buy groceries or dine in restaurants
- Hence lead to increase in users using delivery services
- Eg. Food Panda, GrabFood and Uber Eats
- How can we help restaurants attract consumers via food delivery services?



OUR DATASET

Dataset: “Online Food Delivery Preferences – Bangalore region”




SURVEY

Results from a survey
investigating user's
experiences of food delivery
apps and food



INVESTIGATION

By analysing data from the
survey results, we aim to
identify factors that increase
customer purchases



**What are the optimal factors
to attract consumers via
food delivery service?**

PROBLEM DEFINITION

INDEPENDENT VARIABLE

Various categories of user opinions on how important certain factors are when making delivery purchases

DEPENDENT VARIABLE

Whether or not the user will purchase the food from the delivery service and restaurant again

Outcome: Predict which factors are most important to users

An abstract graphic on the left side of the slide, featuring a series of black and white wavy, organic shapes that flow from the left edge towards the center. The shapes are layered, creating a sense of depth and movement. The entire slide is framed by a thin black border with decorative diamond-shaped corner ornaments.

02

EDA & DATA CLEANING

Insights on consumer demographics,
and prepping data

BASIC INFORMATION

AGE

24

STATUS

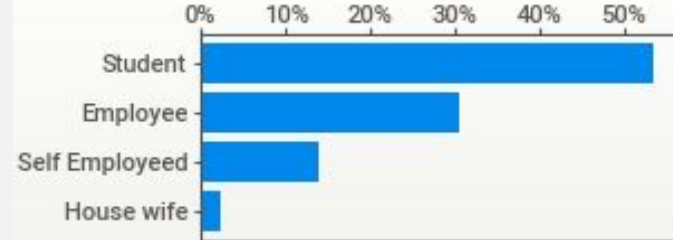
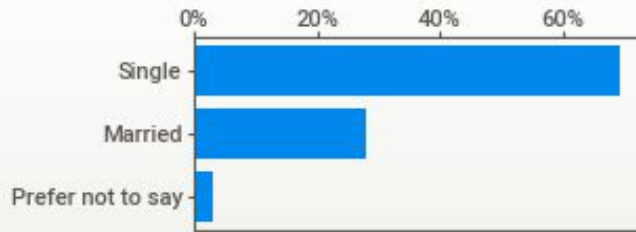
Single

WORK

Students

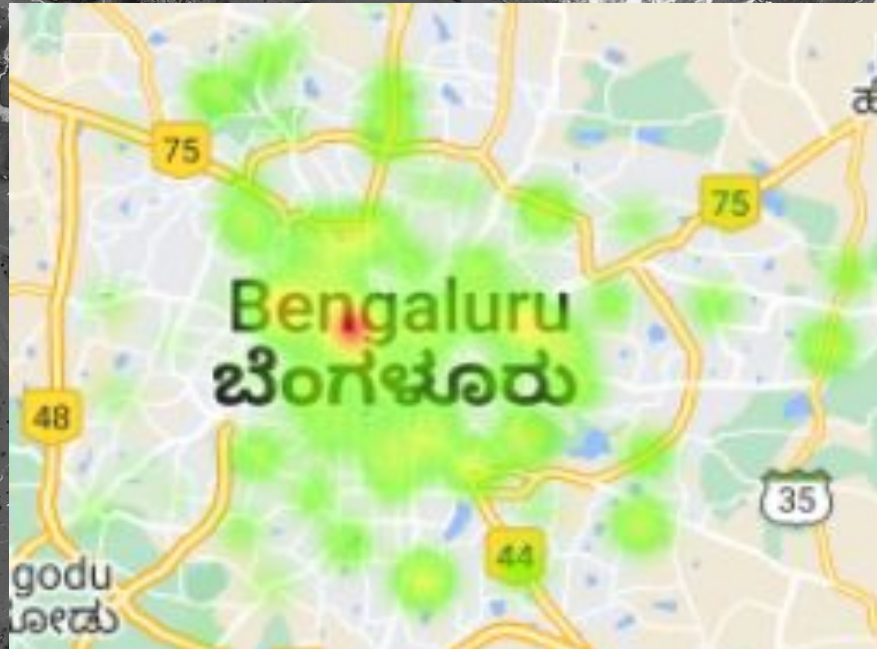
INCOME

No
Income



CONSUMER DEMOGRAPHICS

RESIDENCE

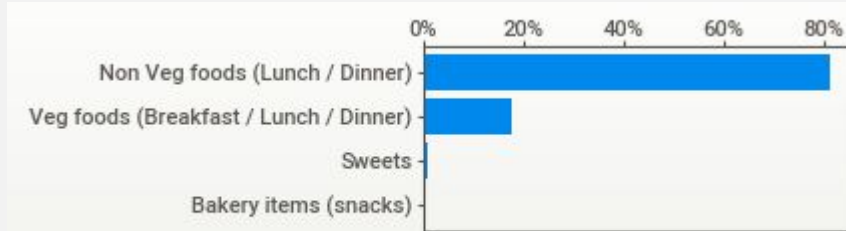


DELIVERY PREFERENCES



PREFERENCE

Non-Veg Foods



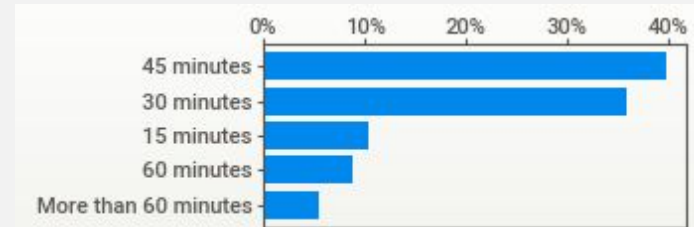
TIME OF DAY

Lunch and Dinner



DELIVERY TIME

30-40 minutes



DATA CLEANING

01 DATA CHECK

We checked for data fields that
may be entered as NaN or empty

```
foodDelivery.isnull().sum()
```

02 ENCODING

Encoding categorical data
into numerical data

```
Cleanup_nums = {"Gender":  
{"Male": 0, "Female": 1} }
```



03

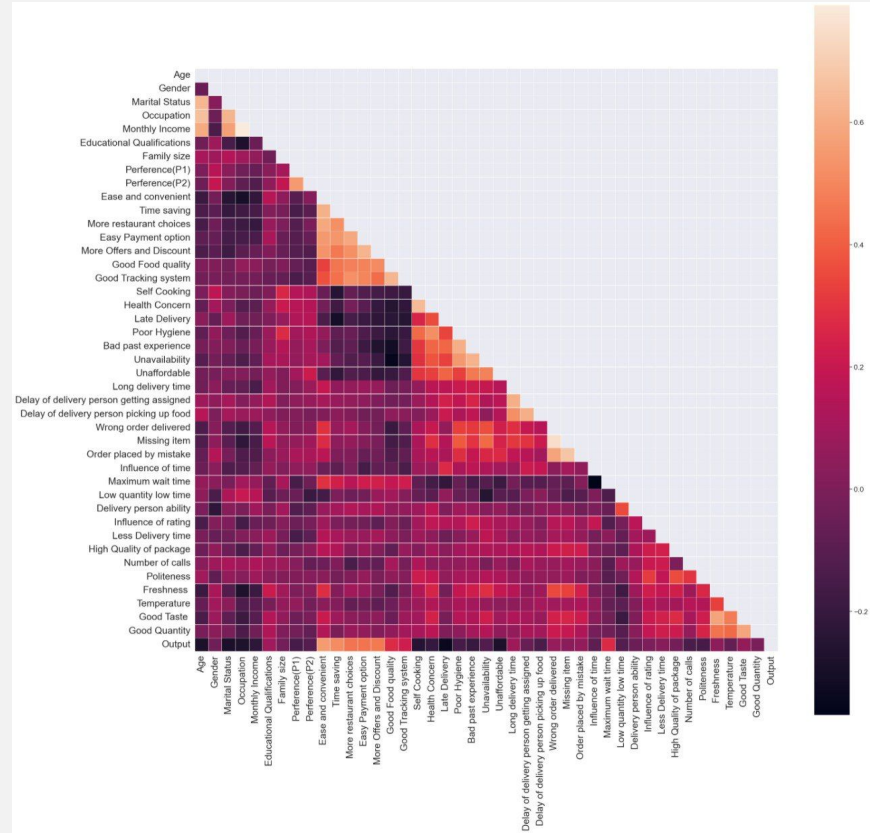
DATA PREPARATION AND RESAMPLING

Classification Modelling



CORRELATION MATRIX

- Evaluate linear relationship between variables
- Spearman's coefficient
- Bottom row - relationship with output

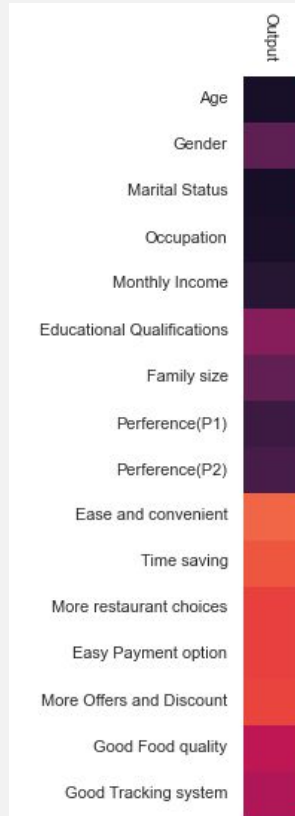


A CLOSER LOOK

Taking a closer look at the correlation matrix of the dataset, we can see that factors such as:

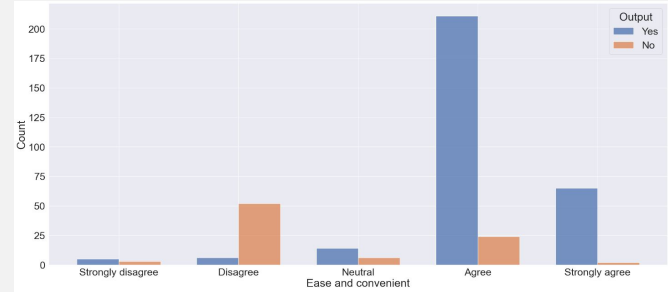
- Ease and Convenience
- Timesaving
- More Restaurant Choices
- Easy Payment Option
- More Offers and Discounts
- Good Food Quality
- Good Tracking System

Have a highest correlation with users purchasing again, suggesting that these factors are of significance to restaurant owners and the delivery service companies

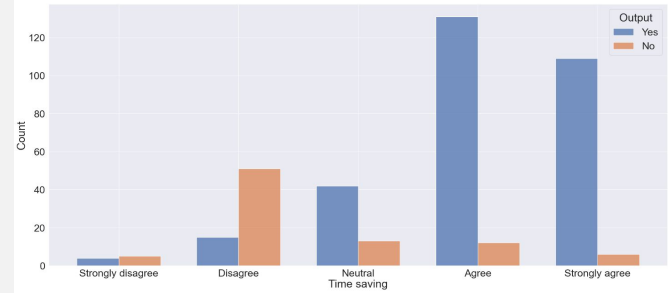


BIVARIATE STATISTICS

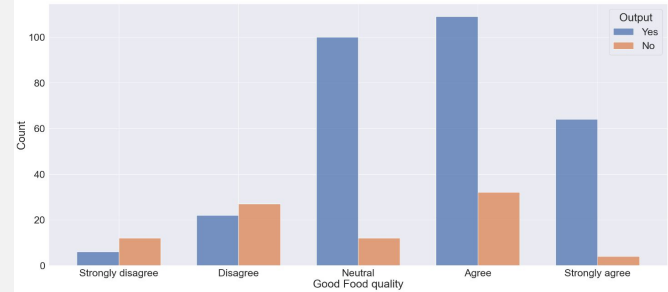
01 CONVENIENCE



02 TIME-SAVING



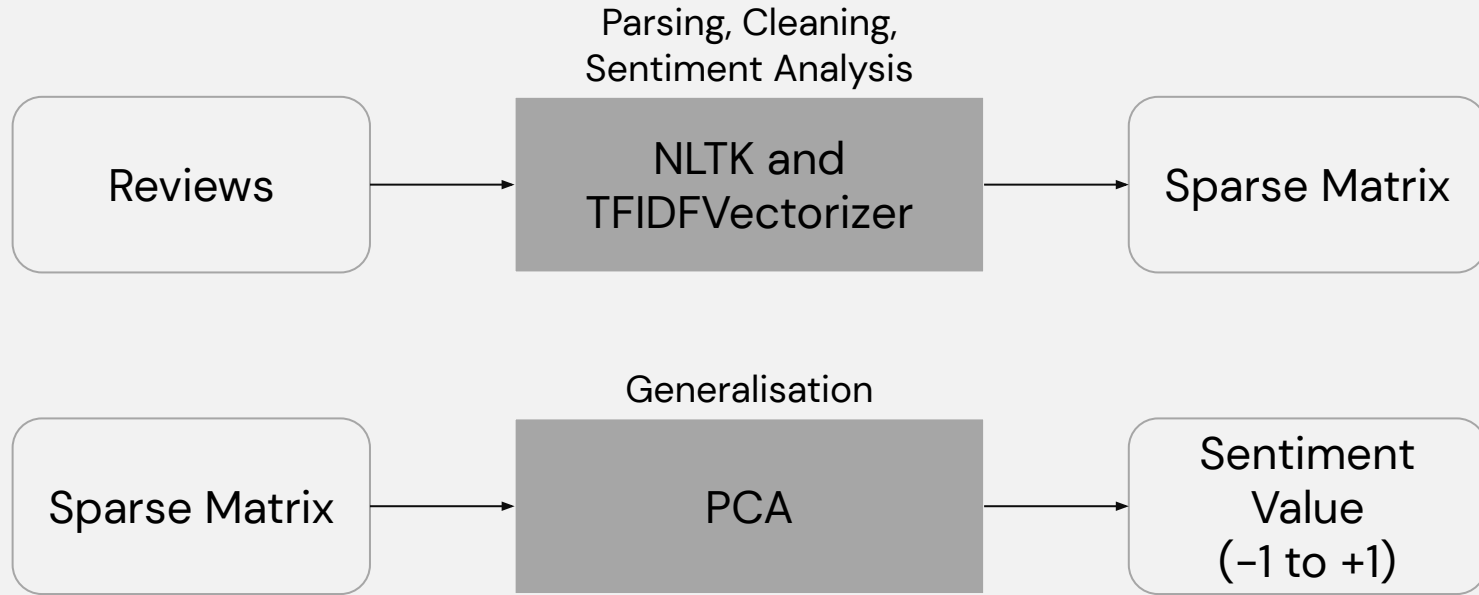
03 FOOD QUALITY



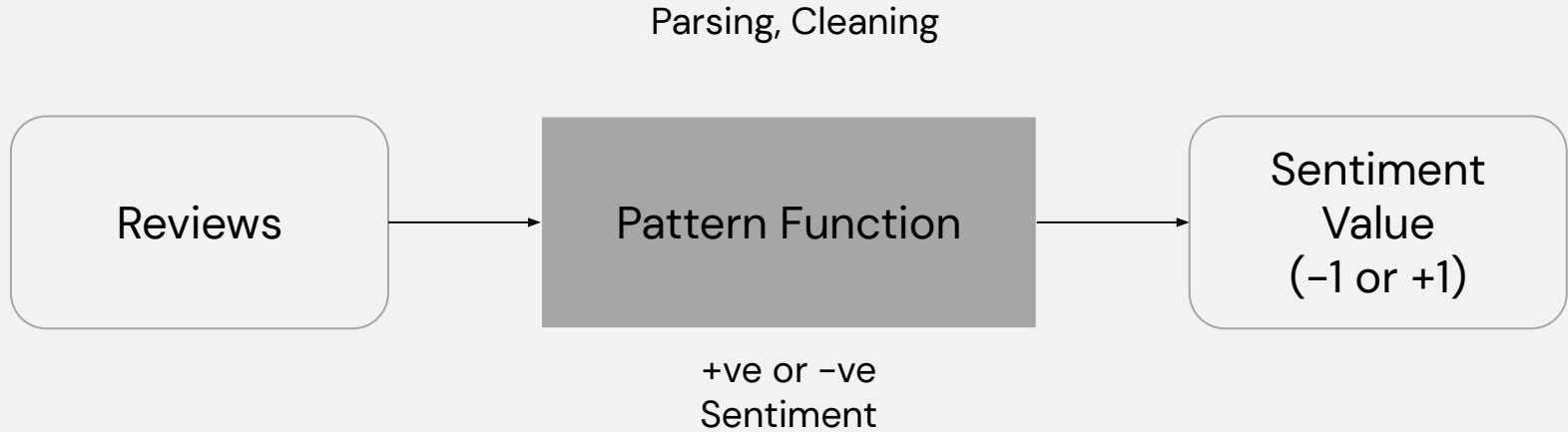
Reviews were included in the surveys, where users filled in some comments on their experiences with the delivery service and the food.



METHOD 1



METHOD 2





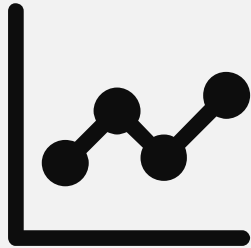
04

CLASSIFICATION MODELLING

Training and Evaluating Models

CLASSIFICATION MODELLING

01



MODELLING

1. Logistic Regression
2. KNN Classifier
3. Random Forest
4. XGBoost

02



TUNING

1. Cross Validation
2. GridSearch

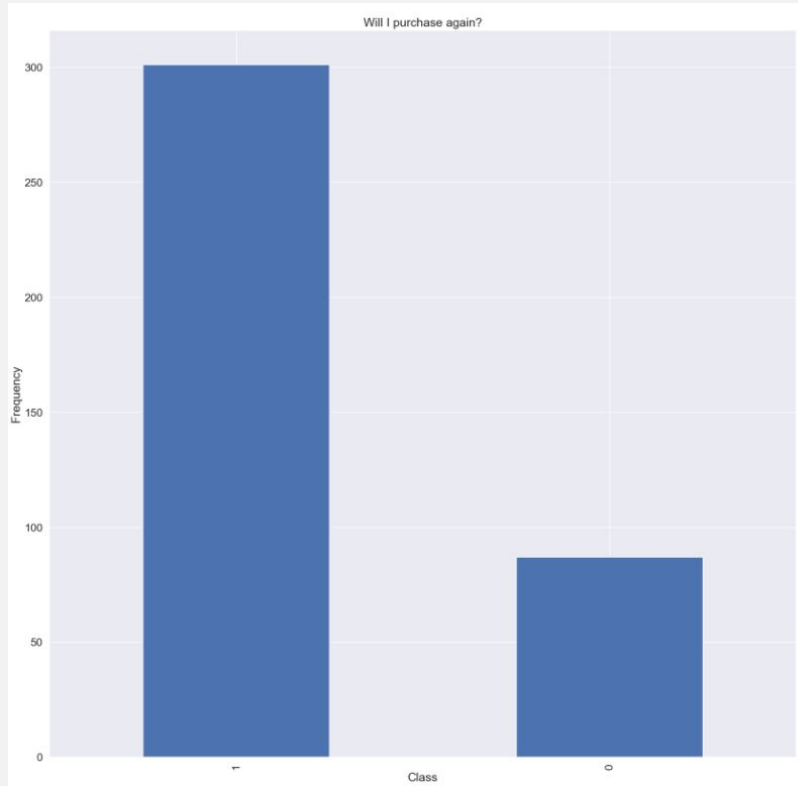
03



EVALUATION

1. ROC_AUC
2. F1 Score

IMBALANCED DATASET



Imbalance in small dataset may lead to problems in training models

SMOTE (Synthetic Minority Oversampling Technique)

Overcome the imbalanced distribution between the categories 'yes' and 'no'

Synthesize elements for the minority class

LOGISTIC REGRESSION MODELLING

01

SPLIT

02

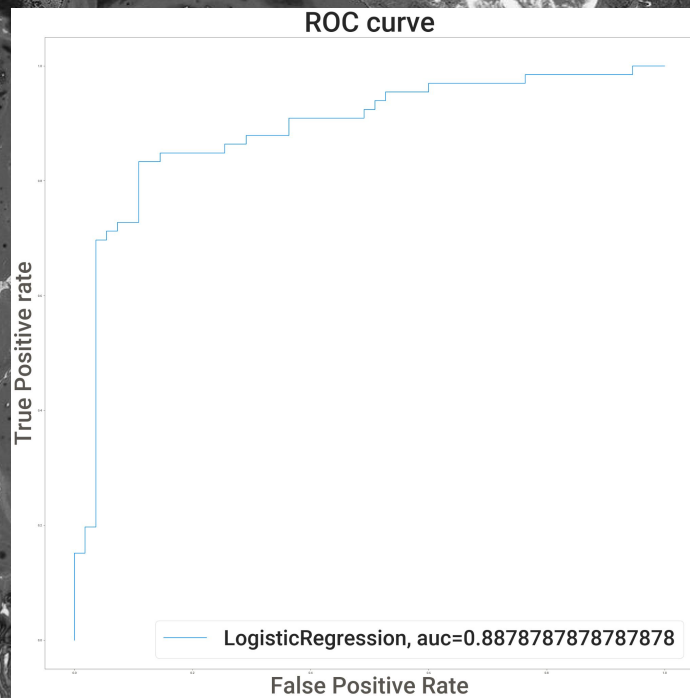
TRAIN

03

EVALUATE

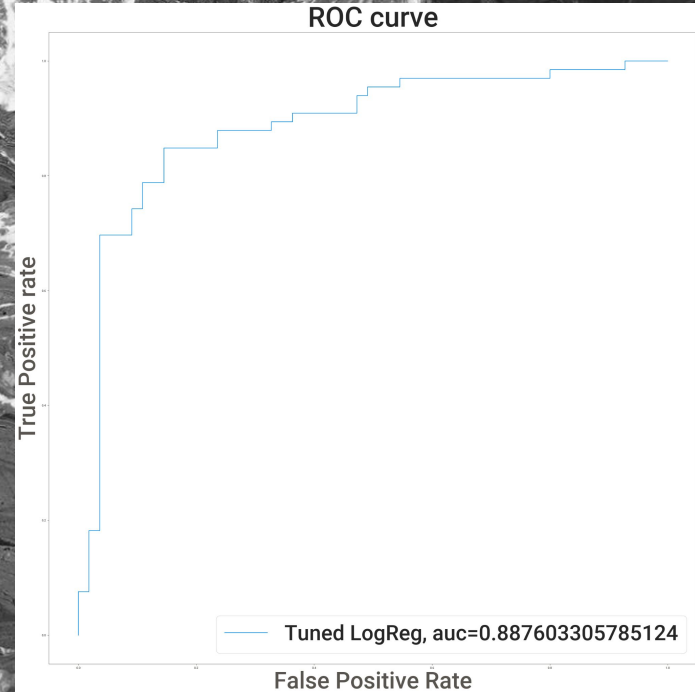
Goodness of Fit of Model	Test Dataset
Explained Variance (R^2)	: 0.8347107438016529
Mean Squared Error (MSE)	: 0.1652892561983471
Root Mean Squared Error (RMSE)	: 0.4065578140908709
False Positive Rate	: 0.10909090909090909
True Positive Rate	: 0.7878787878787878
Accuracy: 0.8347107438016529	
Precision: 0.896551724137931	
Recall: 0.7878787878787878	

ROC_AUC SCORE (Before Tuning)



ROC_AUC SCORE (After Tuning)

Goodness of Fit of Model	Test Dataset
Explained Variance (R^2)	: 0.8347107438016529
Mean Squared Error (MSE)	: 0.1652892561983471
Root Mean Squared Error (RMSE)	: 0.4065578140908709
False Positive Rate	: 0.10909090909090909
True Positive Rate	: 0.7878787878787878



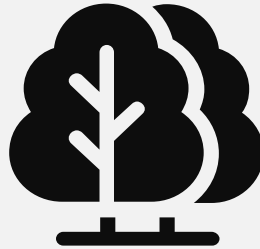
OTHER MODELS

02



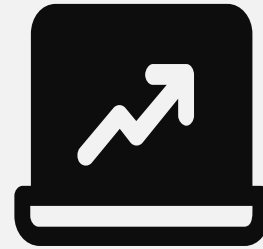
KNN CLASSIFIER

03



RANDOM FOREST

04



XGBOOST

An abstract graphic on the left side of the slide, featuring a series of black and white wavy, organic shapes that flow from the left edge towards the center. The shapes are layered, creating a sense of depth and movement. The entire slide is framed by a thin black border with decorative diamond-shaped corner ornaments.

05

CONCLUSION

Results, Evaluation and Conclusion

RESULTS

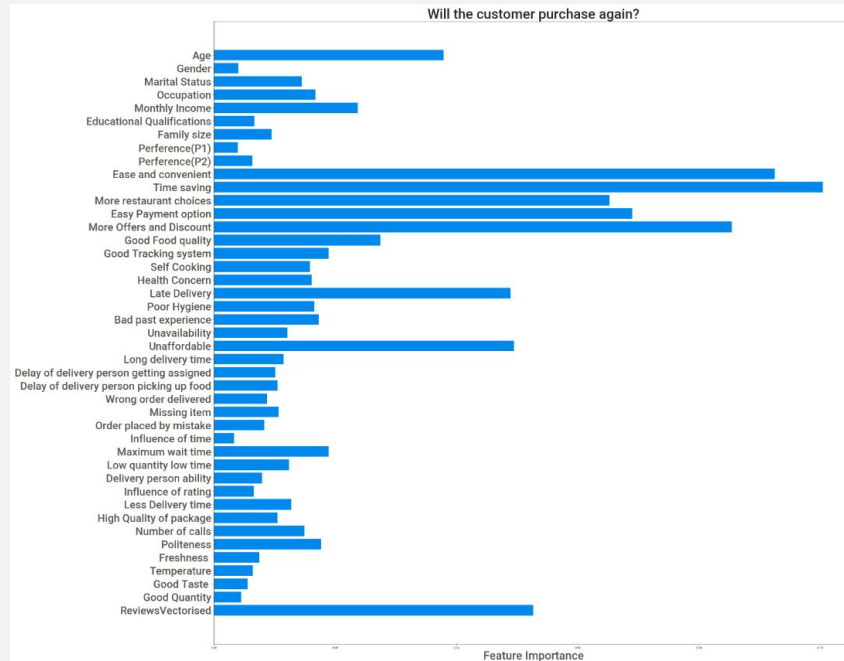
		Logistic Regression	KNN Classification	Random Forest	XGBoost
Before Tuning	TFIDF	0.888	0.927	0.950	0.947
	Pattern	0.922	0.955	0.967	0.968
After Tuning	TFIDF	0.888	0.938	0.966	0.955
	Pattern	0.923	0.956	0.986	0.972

EVALUATION

- All 4 models have and AUC Score of over 85%
- Random Forest was the most accurate, followed by XGBoost
- SMOTE may have influenced the KNN classification model
- TFIDF produced good results, but Pattern was better

OUTCOME

- Random Forest model can be used to find the most important variables:



1. Time Saving
2. Ease and Convenience
3. More Offers and Discounts

CONCLUSION

For a restaurant in Bangalore to **attract the majority of its customers, which are students around the age of 22–25, via delivery services**, it should have a delivery service that is **easy and convenient**, its delivery time needs to be **fast**, and the restaurant **gives offers and discounts regularly**.

THANKS!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution

