Restaurant Sales & Status Prediction

【Machine Learning】【Prediction】【PROM02】

July 2021

Chapter 1 - Introduction

- Restaurant owner want to open a new restaurant but don't know anything about the area, district, or other variables
- They will have harder time planning and budgeting than people who has a restaurant based on data and evidence.

Goal: -

I will use restaurant point of sale(POS) data, public data from Hong Kong government, and property agency to determine the best approach for opening a restaurant in Hong Kong.

Benefit : -

- Accelerate the decision-making processes
- We use machine learning to uncover hidden patterns and deliver location insights and recommendations by integrating diverse data sources.

Problem Statement

01

Problem: To get a new restaurant site up and running, a significant amount of time and investment is required.
Outcome: When the wrong location is chosen for a restaurant brand, the site closes within months and incurs operating losses.

02

Problem: Ambros provide over 5000 restaurant data but limited on operational data which may or may not be able to resolve the problem.
Solution: According to number of reference journals, we should add more data such as shop size, location specific data, rental price etc…

Goal: This research is to predict a selection of locations that can survive and the revenue of the restaurants in the given dataset using machine learning algorithms from previously established restaurant data.

Chapter 2 - Literature Review

Point out related researches and then examine the theories that underpin the Problem Statement

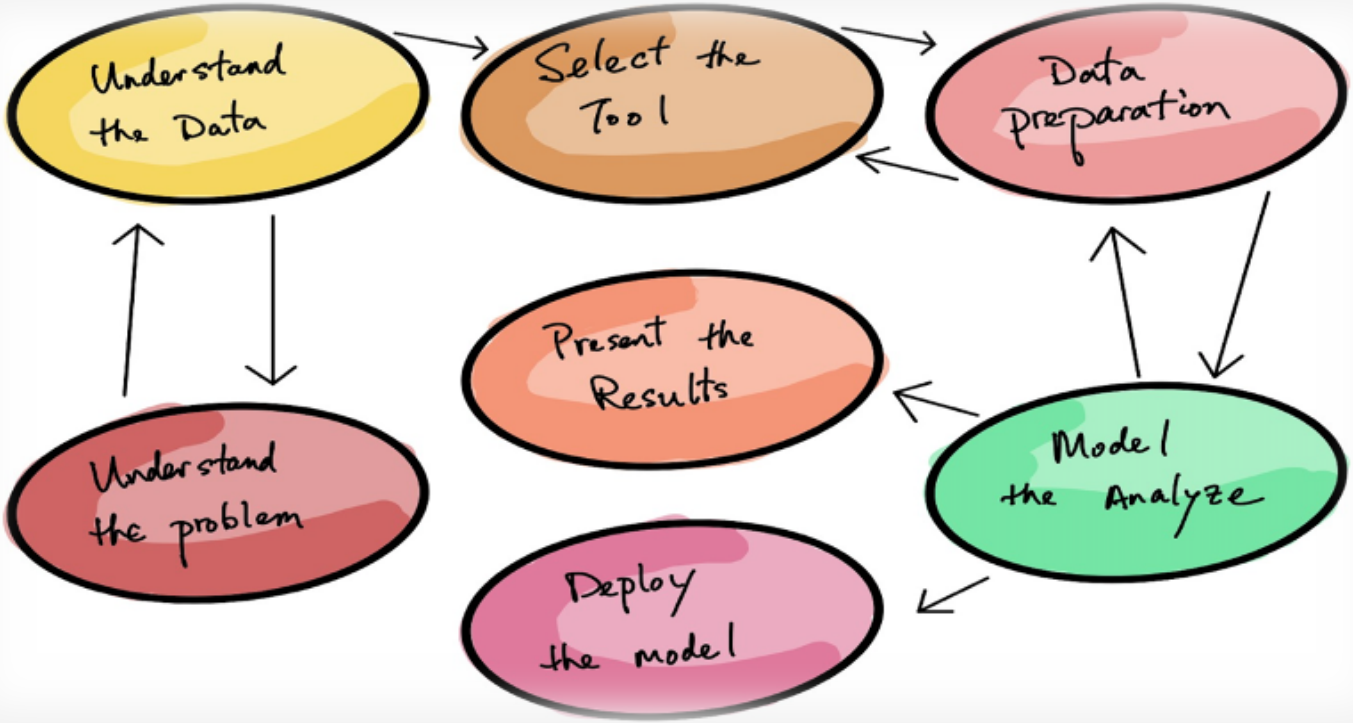
Select feature set, target and method algorithms – SVM, KNN, Radom forest regression, ANN, etc…

Why Use Machine Learning?

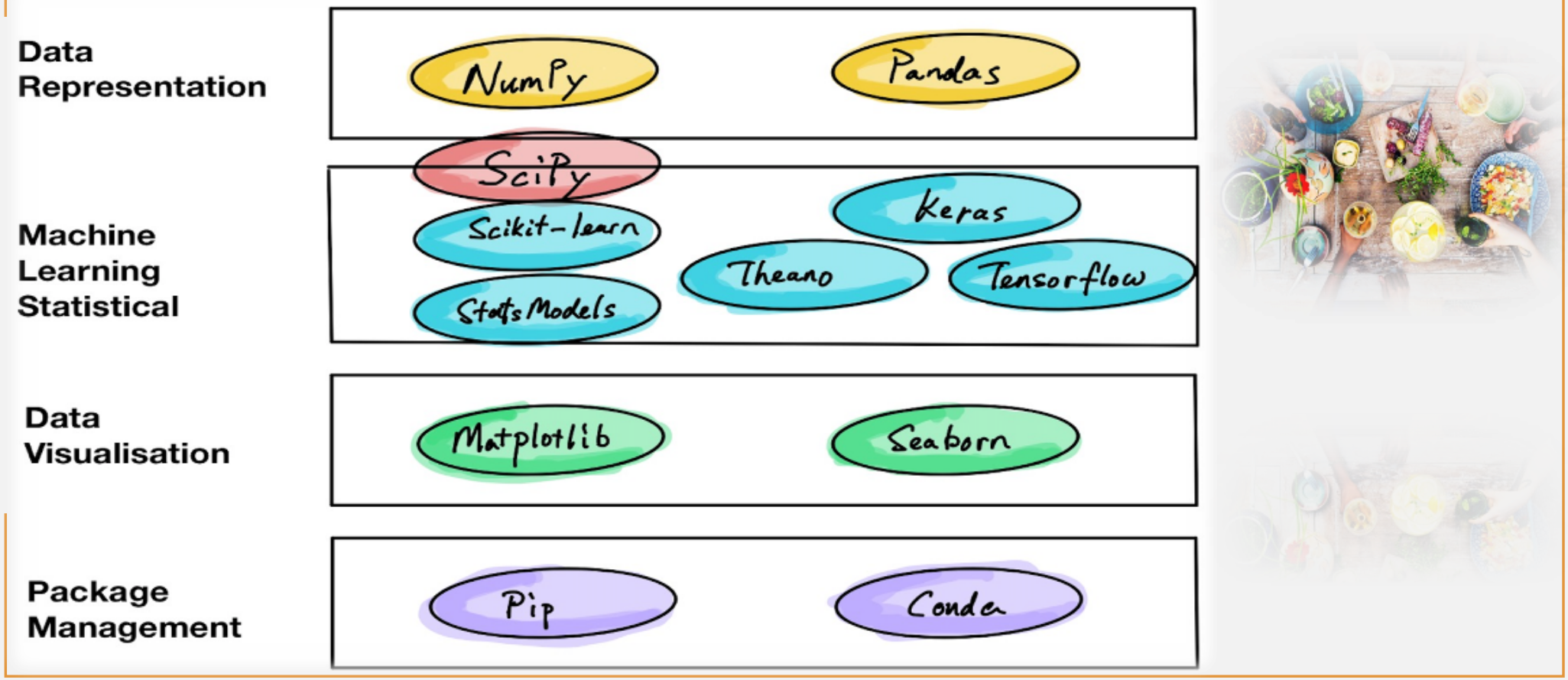
- ML algorithm can simplify code and perform better.
- ML can find complex problems
- ML can easy adapt new data
- Easy to get insights of large amount of data

Chapter 3 - Research Methodology

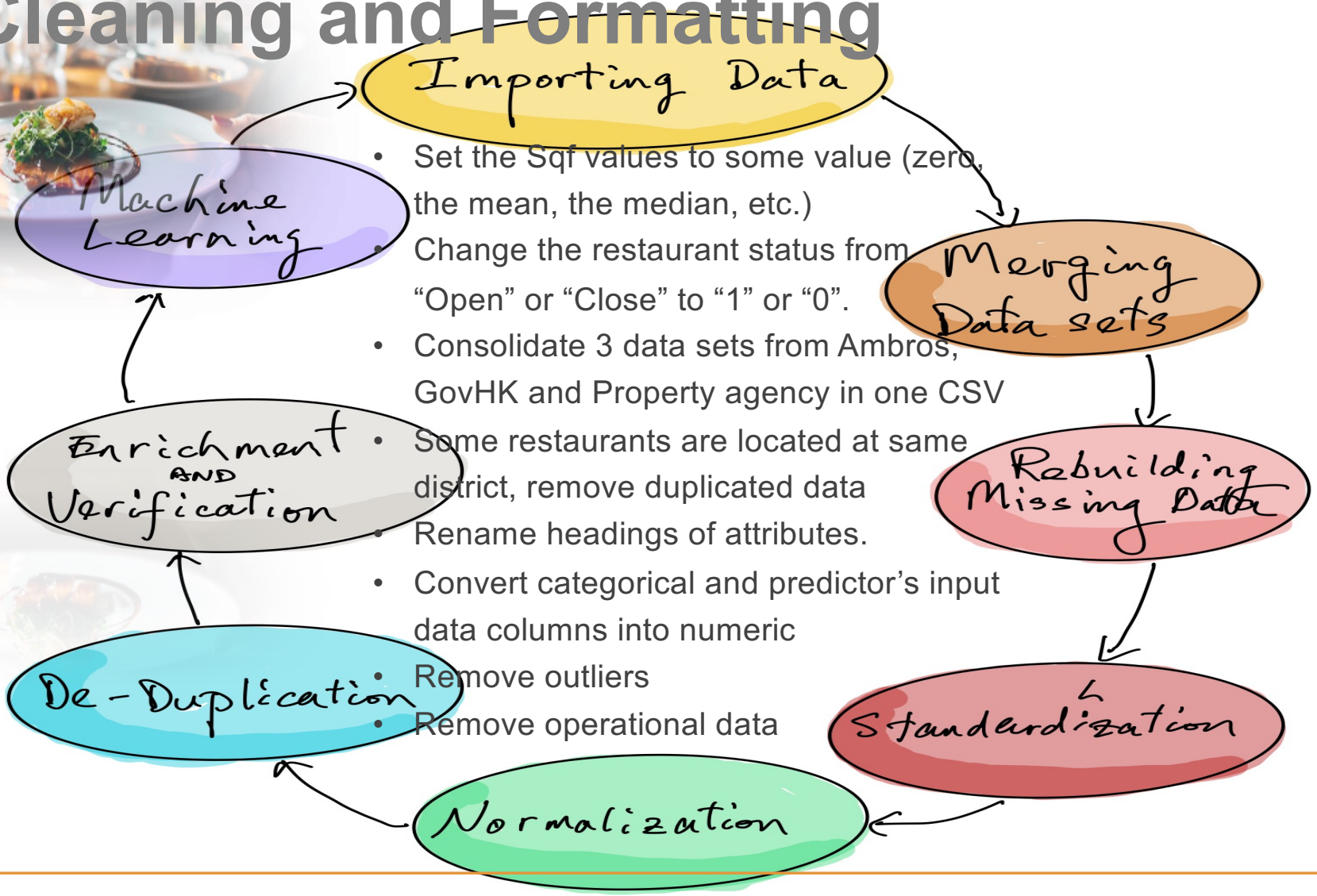
Define data analysis process:
From Raw data to structured data, model building, evaluation and parameters tunings, visualization and model deployment.



This research of tools in Python



Data Cleaning and Formatting



- Set the Sqf values to some value (zero, the mean, the median, etc.)
- Change the restaurant status from "Open" or "Close" to "1" or "0".
- Consolidate 3 data sets from Ambros, GovHK and Property agency in one CSV
- Some restaurants are located at same district, remove duplicated data
- Rename headings of attributes.
- Convert categorical and predictor's input data columns into numeric
- Remove outliers
- Remove operational data

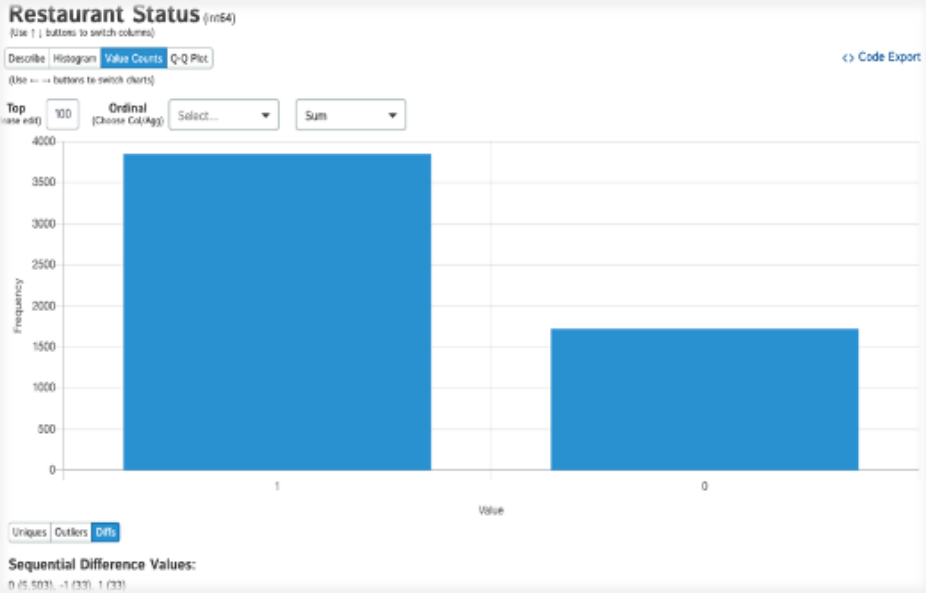
Dataset consists of following fields/feature description:

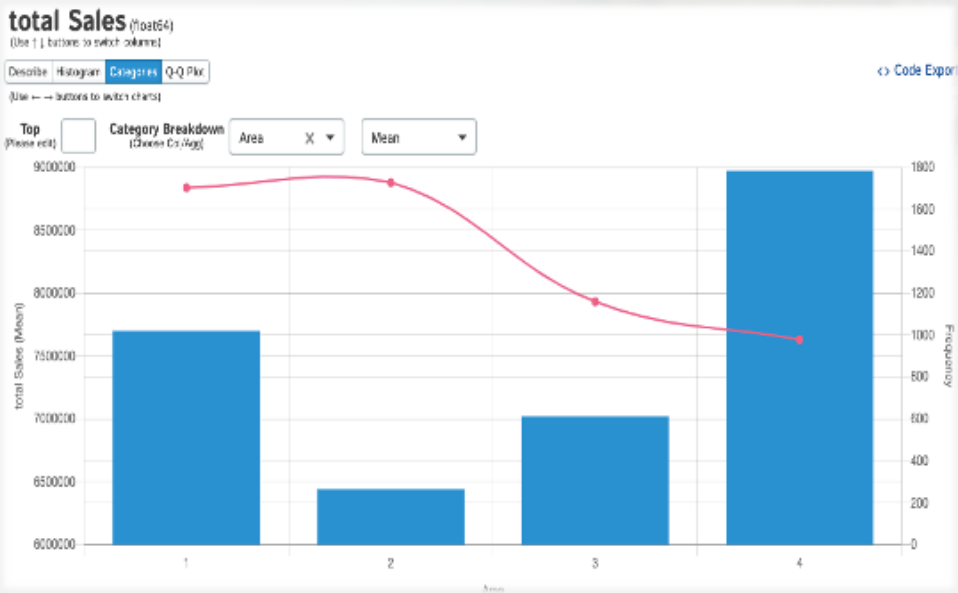
- Restaurant id: the ID of restaurant
- Restaurant type: type of restaurant ie. Chinese, western, fast food, etc..
- Area: restaurant location
- District: HK official district name
- Start Day, month, year
- District about number of people able to work 15 to 65+: Age range is still working in that district
- Area number of worker 15 to 65+: Age range is still working in that location.
- Total Sales: revenue of the restaurant
- Payment type: ie. cash,visa/master, others
- Number of transaction: transaction of the restaurant
- Restaurant Sqf: size of restaurant
- Monthly rent: rental fees
- Location type: eg. shop in the street, mall, school, etc…
- Restaurant Status: Closed or Open

Dataset

Insight of Dataset

- There are 5570 rows and 41 features in training data
- Restaurant Status 1 = Open, 0 = Close
- Area 1 = HK island, 2 = Kowloon, 3 = New Territory West, 4 = New Territory East, It shows the new territory east total sales revenue are the highest vs others.



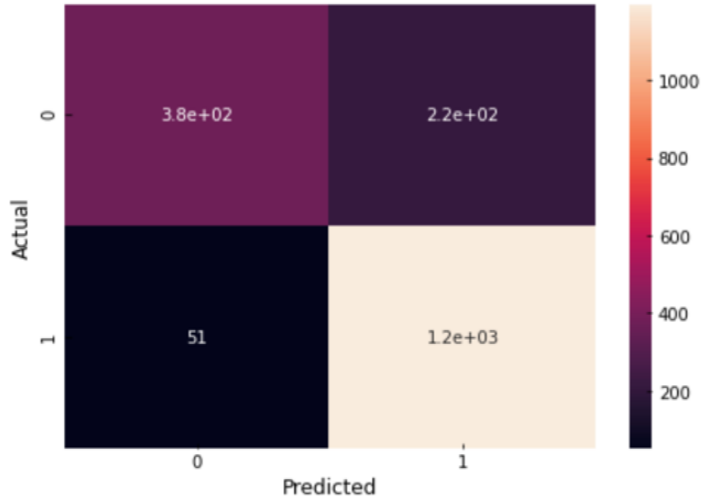


Chapter 4 - Results and Evaluations

One of ML - Random Forest Classifier testing here. The accuracy result at 85% and showing Random Forest classifier is a good algorithm for this purpose.

A Random Forest Classifier model to predict whether restaurant open or not

Model	Train Acc	Test Acc
Random Forest Classifier	89 %	85 %



Confusion Matrix: Random Forest Classifier

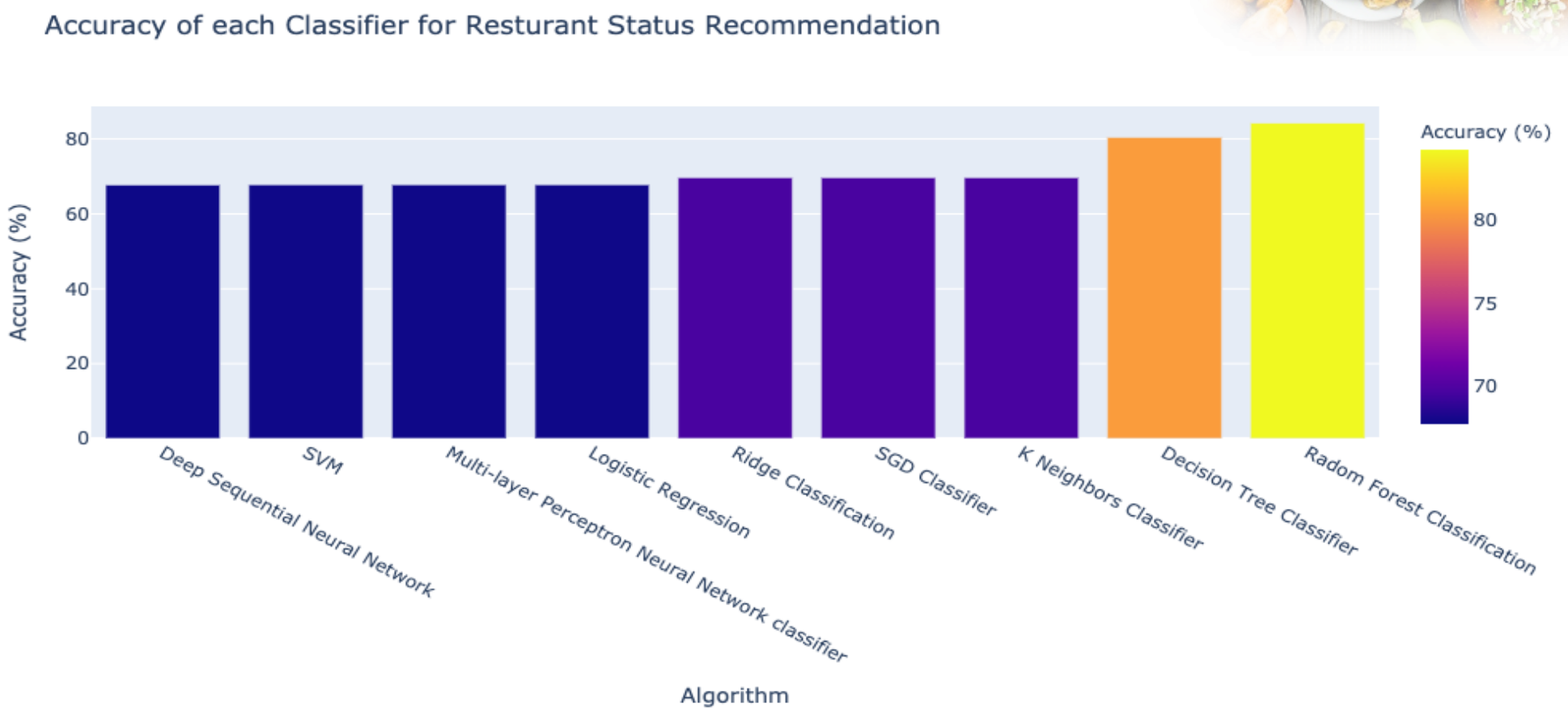
```
[[ 377  215]
 [  51 1196]]
```

Classification Report:Random Forest Classifier

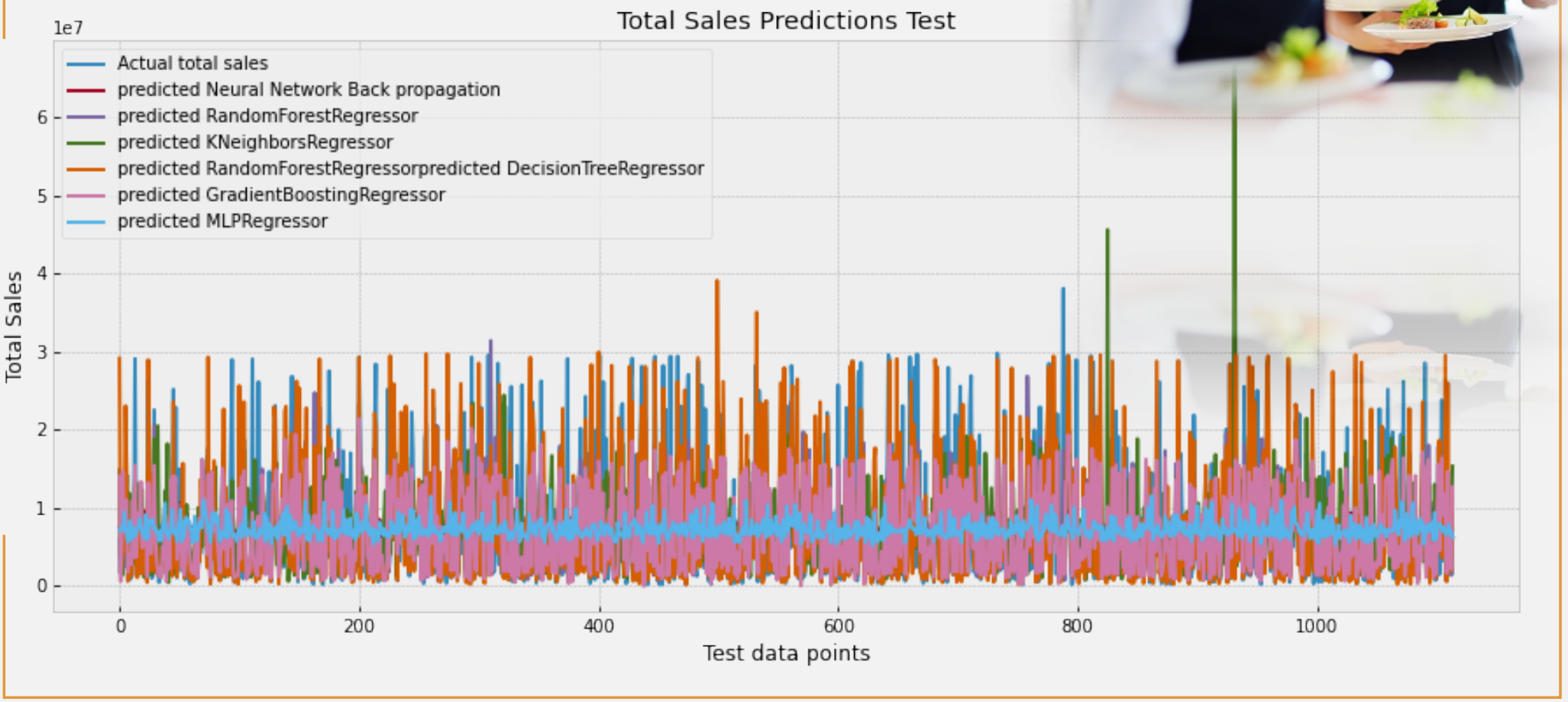
	precision	recall	f1-score	support
0	0.88	0.64	0.74	592
1	0.85	0.96	0.90	1247
accuracy			0.86	1839
macro avg	0.86	0.80	0.82	1839
weighted avg	0.86	0.86	0.85	1839

('Random Forest Classifier Accuracy:', 85.53561718325177)

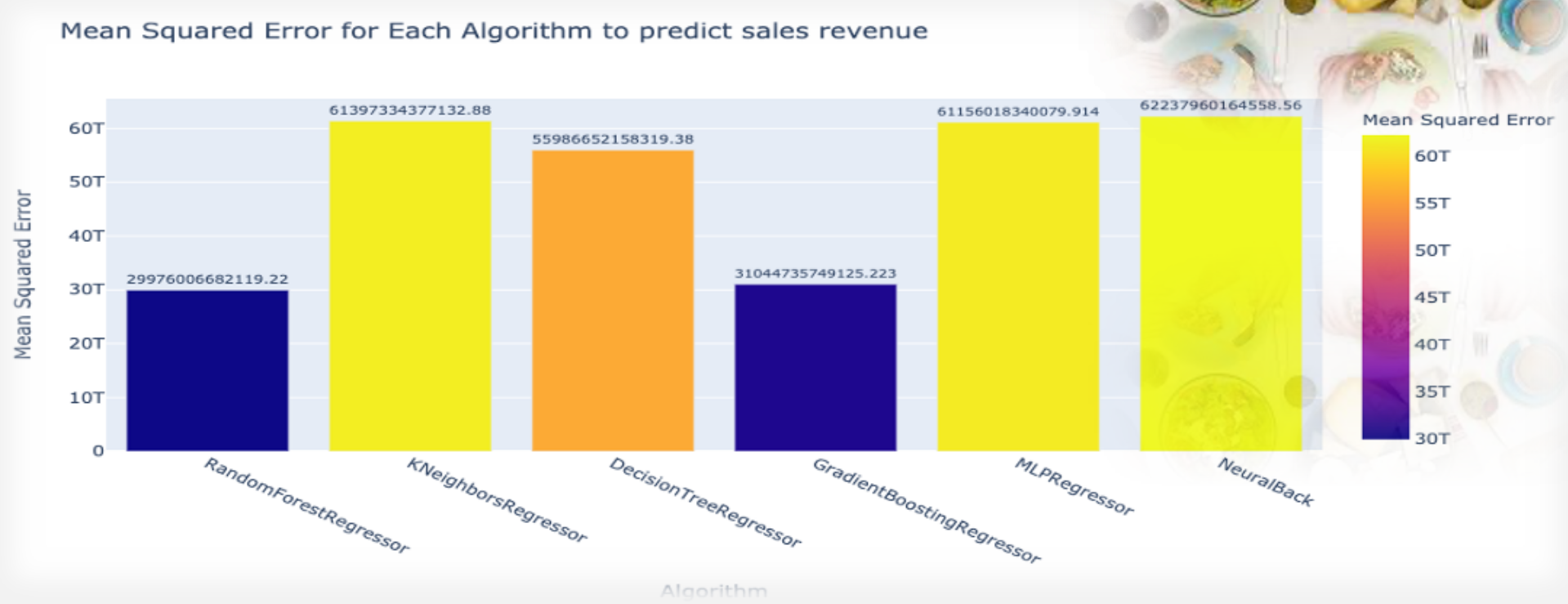
Classifier Forecast Accuracy



Total Sales Predictions Test



Mean Squared Error for Each Algorithm to predict total sales



Chapter 5 - Conclusions

- The research shows that the Random Forest Classification is the best method for restaurant's status prediction
- K Neighbour Regression is the best method for restaurant's sales prediction.

The research found that the machine learning selection must be randomly selected for better accuracy.

End

【Machine Learning】【Prediction】【Algorithms】

May 2021