Cognifyz Internship -Level 3 summary PDF

Predictive Modelling & Customer Preference Analysis

Task 1 : modelling

Here we loaded the engineered dataset from level 2.
Splitted the dataset into training and test set (80%-training set, 20%- test set).

Aggregate rating is our target variable

**Selected Metric: R squared **

The primary metric used for model comparison and hyperparameter tuning(refit=)

- R^2 measures how much of the variance in ratings your model explains.
- Higher R^2 = better model.

Pre-Tuning Model Evaluation

Models used Random Forest Regressor and decision tree Regressor.
we use MAE, MSE, and R^2 metrics to evaluate performance of both models.
Decision tree overfitted — achieved perfect performance on the training set (R^2= 1.0, zero error).
Random forest outperformed decsion tree — Higher R^2 - explains 96% of rating variance.

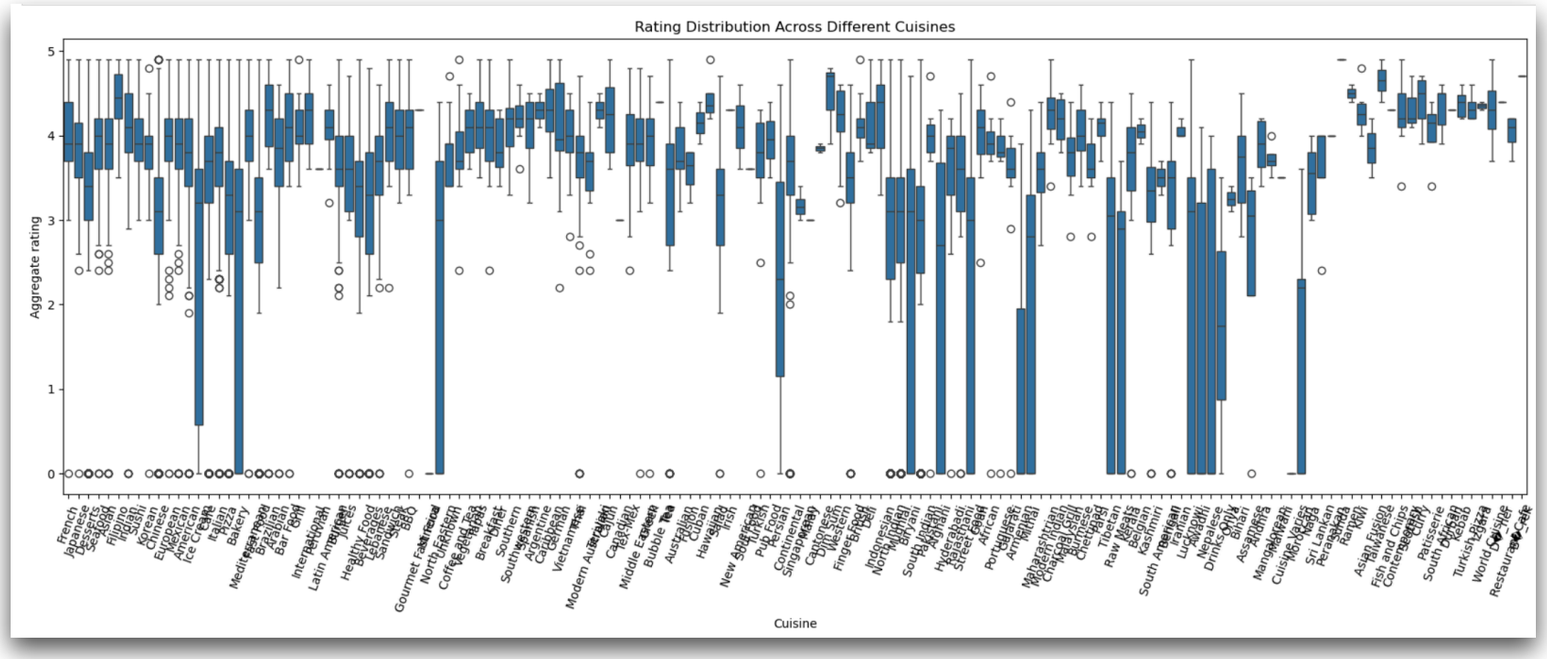
Post-Tuning Model Evaluation : random forest wins

conclusion

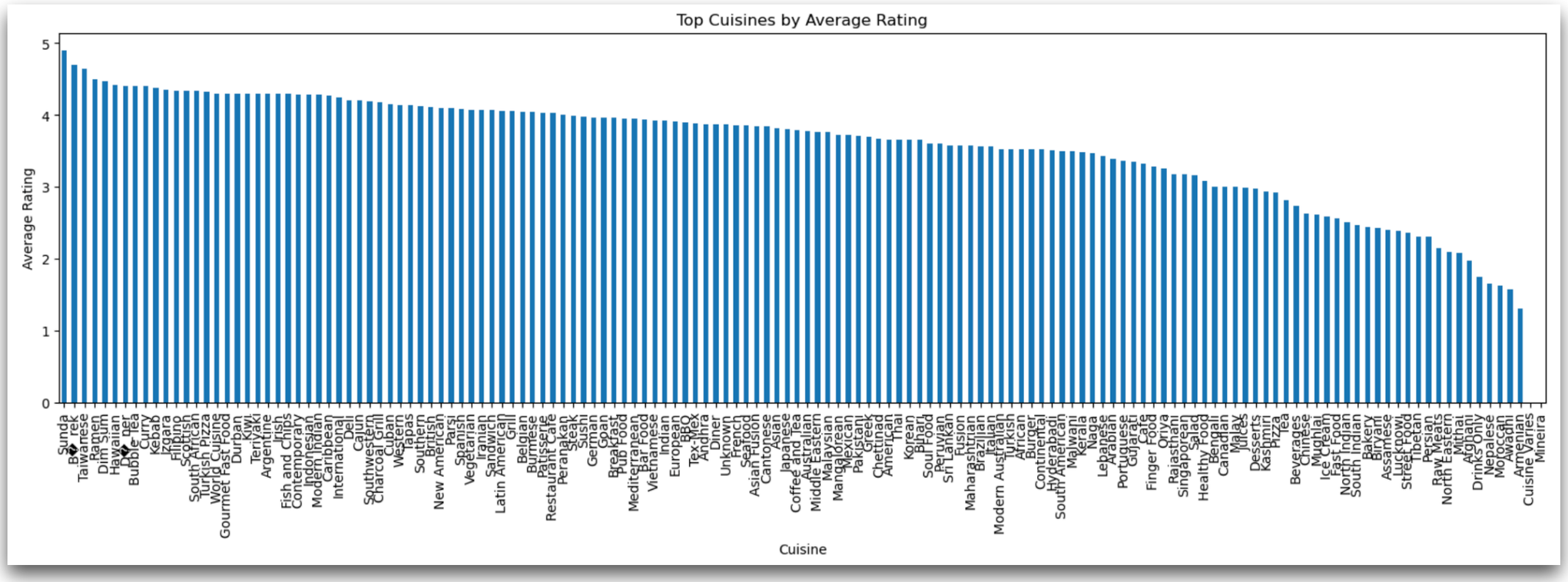
The Random Forest Regressor performs the best among the models tested. It achieves the highest R^2 score (0.9628), meaning it can explain about 96% of the variation in restaurant ratings. It also produces lower prediction errors (MAE & MSE) compared to the Decision Tree. In contrast, the Decision Tree, although improved after tuning, explains around 95% of the variance (R^2 = 0.9534) and has slightly higher errors.Therefore, the Random Forest is the most reliable and accurate model for predicting restaurant ratings in this dataset.

Task 2 : Customer preference Analysis

- Reloaded the original dataset
- Since the analysis is all about individual cuisines not a cuisines combinations. we knew restaurants(rows) have multiple cuisines so we split each cuisine To properly analyse cuisines.
- We Identified the most popular cuisines among customers based on the number of votes.
- Average rating for each cuisine
- Analyse the Relationship btw the type of cuisines and restaurants rating



	Cuisine	Votes
0	North Indian	595981
1	Chinese	364351
2	Italian	329265
3	Continental	288255
4	Fast Food	184058
5	American	183117
6	Cafe	177568
7	Mughlai	151946
8	Desserts	105889
9	Asian	104303



Final Conclusion: Customer Preference Analysis

- *The analysis reveals a clear gap between what customers frequently choose and what they rate highest. North Indian, Chinese, and Italian cuisines dominate in total votes, showing they are the most widely tried and popular choices.*
- *However, the highest average ratings belong to cuisines like Sunda, Taiwanese, Ramen, Bubble Tea, and Dim Sum, which indicates that top-rated cuisines are not necessarily the most popular ones.*

Key Insight

Popularity reflects exposure, not satisfaction.

- Cuisines with the highest votes aren't always the ones customers rate the best.

Quality Consistency (Boxplot Takeaway)

- *Some cuisines have high median with tight rating spread → quality is consistently good (reliable experience).*
- *Others have wide variability → customer experience varies greatly depending on where they eat it.*

1. Consistently high-quality cuisines: E.g., Sunda, Ramen, Dim Sum (high median + small IQR)

2. Inconsistent experience cuisines: E.g., Chinese, Fast Food (large spread despite popularity)