**📊 Level 3 – Task 1: Predictive Modeling**

**📌 Objective**

The goal of this task is to build a **regression model** to predict the **aggregate rating** of a restaurant based on available features.

**🎯 Key Goals:**

1️⃣ **Build a regression model** to predict restaurant ratings.  
2️⃣ **Split the dataset into training and testing sets** to evaluate model performance.  
3️⃣ **Experiment with different algorithms** (Linear Regression, Decision Trees, Random Forest) and compare their performance.

This analysis helps in understanding:  
✔ **Which factors influence restaurant ratings the most**.  
✔ **Which machine learning model gives the best predictions**.  
✔ **How well we can predict customer preferences based on restaurant features**.

**1️⃣ Step 1: Prepare the Dataset for Predictive Modeling**

**🔹 Process:**

✔ Selected relevant features:

* **Price Range** (pricing level of the restaurant).
* **Votes** (total customer votes).
* **Has Table Booking** (binary: 1 = Yes, 0 = No).
* **Has Online Delivery** (binary: 1 = Yes, 0 = No).
* **Restaurant Name Length** (length of the restaurant name).  
  ✔ Dropped missing values to ensure a clean dataset.

**📊 Summary of the Prepared Dataset:**

* **Total Entries:** 9,551
* **Columns Used:** 5 predictor variables + 1 target variable (Aggregate Rating).

**2️⃣ Step 2: Split the Dataset into Training and Testing Sets**

✔ Split the dataset into:

* **Training Set (80%) → 7,640 samples**.
* **Testing Set (20%) → 1,911 samples**.  
  ✔ Ensured that the model was trained on one part of the data and tested on unseen data.

**3️⃣ Step 3: Train the Regression Models**

**🔹 Linear Regression Model**

* **Intercept:** 1.28
* **Coefficients:**
  + **Price Range (0.6577)** → Higher price → **Higher ratings**.
  + **Votes (0.00067)** → More votes → **Slightly better ratings**.
  + **Has Table Booking (-0.2376)** → Table booking **slightly lowers ratings**.
  + **Has Online Delivery (0.6415)** → Online delivery **increases ratings**.
  + **Restaurant Name Length (-0.0035)** → **No significant impact** on ratings.

**📊 Model Performance (Linear Regression)**

* **MAE:** 1.07 (average prediction error).
* **MSE:** 1.67 (error magnitude).
* **R²:** 0.26 (**only 26% variance explained** → Poor performance).

**4️⃣ Step 4: Train a Decision Tree Model**

✔ **Decision Tree Model** was trained to see if it performed better.

**📊 Model Performance (Decision Tree)**

* **MAE:** 0.32 ✅ (better than Linear Regression).
* **MSE:** 0.24 ✅ (lower error).
* **R²:** 0.89 ✅ (explains 89% of the variation in ratings).

🔹 **Key Insight:** Decision Tree performed **significantly better** than Linear Regression.

**5️⃣ Step 5: Train a Random Forest Model**

✔ **Random Forest Model** was tested to compare its accuracy with Decision Tree.

**📊 Model Performance (Random Forest)**

* **MAE:** 0.24 ✅ (**Best**).
* **MSE:** 0.14 ✅ (**Lowest error**).
* **R²:** 0.94 ✅ (**Explains 94% of rating variations**).

🔹 **Key Insight:** **Random Forest is the best model** for predicting restaurant ratings.

**6️⃣ Final Model Comparison**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | **Model** | **MAE (Lower is Better)** |  | **MSE (Lower is Better)** | **R² (Higher is Better)** | | --- | --- | --- | --- | --- | | **Linear Regression** | 1.07 ❌ |  | 1.67 ❌ | 0.26 ❌ | | **Decision Tree** | 0.32 ✅ |  | 0.24 ✅ | 0.89 ✅ | | **Random Forest** | ✅ **0.24** (Best) |  | ✅ **0.14** (Best) | ✅ **0.94** (Best) | |

🔹 **Final Takeaways:**  
✅ **Random Forest is the best model** with the lowest error and highest accuracy.  
✅ The **most influential factors** for predicting ratings are **Price Range, Votes, and Online Delivery**.  
✅ **Table booking negatively impacted ratings**, which may indicate customer dissatisfaction with booking experiences.

**🎯 Final Summary**

✔ Successfully **built and tested predictive models** to estimate restaurant ratings.  
✔ Discovered that **Random Forest outperforms Linear Regression and Decision Trees**.  
✔ Identified **key factors influencing restaurant ratings**.