Machine Learning Internship Report

Task 1: Predict Restaurant Ratings

Objective

Build a machine learning model to predict the aggregate rating of a restaurant based on features like cuisine type, pricing, services, and more.

Data Overview

- **Total Records: 9,551**
- **Total Features: 21**
- **Missing Values: 9 (in `'Cuisines'` column)**

Data Preprocessing

- **Filled missing 'Cuisines' with `'Unknown'`**
- **Created new feature: Cuisine_Count**
- **One-hot encoded popular cuisines: *Italian, Indian, Chinese, etc.***
- **Converted table booking and online delivery into numeric (binary) format**

Exploratory Data Analysis

Visualizations:

- **Histogram of rating distribution**

- Average rating by price range
- Average rating by cuisine
- Correlation matrix

Model Performance

Model	RMSE	MAE	R^2
Linear Regression	1.2905	1.0733	0.2684
Random Forest	1.2438	0.9835	0.3203
Gradient Boosting	1.2080	0.9694	0.3589

✅ Best Model: Gradient Boosting Regressor

Feature Importance

Feature	Importance
Price range	0.5393
Average cost for two	0.2378
Has_Online_delivery	0.1352
Cuisine_Count	0.0427
Has_Indian	0.0206
Table booking	0.0045
Has_Chinese	0.0185
Has_Japanese	0.0092

Insights

- **Price and cost are the strongest predictors**
- **Online delivery positively impacts ratings**
- **Cuisine diversity influences rating**
- **Table booking has minimal effect**
- **Indian and Chinese cuisines contribute more than Japanese**

Limitations & Improvements

- **$R^2 \approx 0.36$ → indicates limited predictive power**
- **Suggested enhancements:**
  - **Add location data (e.g., proximity to landmarks)**
  - **Incorporate review text sentiment**
  - **Include operating hours**
  - **Explore hybrid and deep learning models**

Task 2: Recommendation System

Goal

Build a content-based recommender system to suggest restaurants based on user preferences.

Example 1: High-End Italian Restaurant Lover

Restaurant	City	Cost	Rating
Ristorante Prego	Noida	₹1200	3.1
San Gimignano - The Imperial	New Delhi	₹5000	3.7

The Krib	**New Delhi**	**₹2500**	**3.2**
Prego - The Westin	**Gurgaon**	**₹3000**	**3.8**
Bella Cucina	**Gurgaon**	**₹4000**	**4.1**

Diversity Metrics:

- **Cities: 3**
- **Cuisine Entropy: 0.000**
- **Price Range Entropy: 0.7219**

Example 2: Budget Asian Food Lover

Restaurant	City	Cost	Rating
Red	**Gurgaon**	**₹1200**	**3.0**
Soi Thai	**New Delhi**	**₹1500**	**3.6**
Noshi	**New Delhi**	**₹1000**	**4.2**
China Bowl	**New Delhi**	**₹800**	**3.9**
Shagun	**New Delhi**	**₹800**	**3.9**

Diversity Metrics:

- **Cities: 2**
- **Cuisine Entropy: 1.5774**
- **Price Range Entropy: 0.9710**

Example 3: Food Quality Enthusiast

Restaurant	City	Cost	Rating
Keventers	**New Delhi**	**₹400**	**3.9**
RollsKing	**Noida**	**₹300**	**3.8**

Tonico Cafe	Kochi	₹400	4.0
Atul Chaat Corner	New Delhi	₹100	3.8
Novelty Dairy	New Delhi	₹200	4.1

Diversity Metrics:

- Cities: 3
- Cuisine Entropy: 0.0000
- Price Range Entropy: 0.0000

Implementation Highlights

- Preprocessing: filled missing values, binary cuisine indicators
- Feature engineering: extracted key cuisines
- Similarity metric: Cosine similarity
- Evaluation: Entropy and uniqueness of results

Future Improvements

- Add location-based filtering
- Implement hybrid recommender systems
- Consider restaurant operating hours
- Integrate user feedback loops

Task 3: Rating Classification

Goal

Classify restaurant ratings into categories:

- **Low, Medium, High, Excellent**

Model Results

Model	Accuracy	Macro F1-score
Logistic Regression	0.5568	0.42
Random Forest	0.5835	0.48
SVM	0.5667	0.43
Gradient Boosting	0.6112	0.47

✅ **Best Classifier: Gradient Boosting Classifier**

Task 4: Geographic Analysis

Dataset Stats

- **Unique cities: 141**
- **Unique localities: 1,208**
- **Unique countries: 15**
- **Final Dataset Size: (9,551, 24)**

Geographic Range

- **Latitude: -41.330 to 55.976**
- **Longitude: -157.948 to 174.832**
- **Map File:** `global_restaurant_distribution.html`

Top 10 Cities by Restaurant Count

1. **New Delhi — 5473**
2. **Gurgaon — 1118**
3. **Noida — 1080**
4. **Faridabad — 251**
5. **Ghaziabad — 25**
   ...

Top Cities by Average Rating

1. **Inner City — 4.90**
2. **Quezon City — 4.80**
3. **Makati City — 4.65**
4. **Pasig City — 4.63**
5. **Mandaluyong — 4.63**

Top Cities by Cuisine Type (Example: New Delhi)

- **North Indian: 2425**
- **Chinese: 1638**
- **South Indian: 411**
- **Fast Food: 1304**
- **Italian: 376**

Conclusion

This project successfully addressed key machine learning challenges in the restaurant domain:

- **Pricing and delivery options are critical for predicting ratings**
- **Content-based recommendation aligns well with user profiles**
- **Gradient Boosting models achieved best regression & classification performance**
- **Geographic visualizations reveal cuisine preferences and restaurant clustering**
