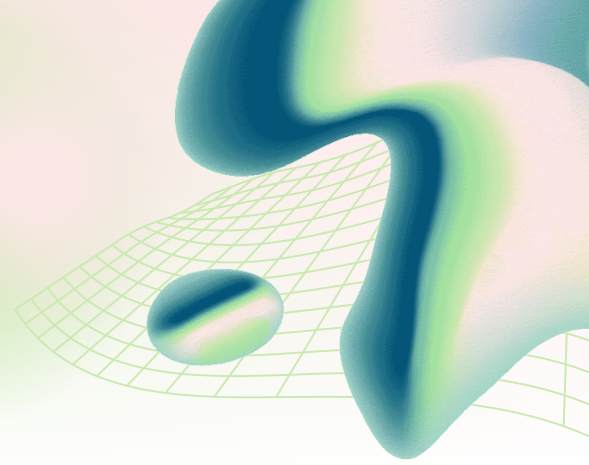


MACHINE LEARNING INTERNSHIP REPORT

COGNIFYZ TECHNOLOGIES

Mihir Milind Ughade



Summary

This report contains the complete implementation and summary of three machine learning tasks completed as part of the internship with Cognifyz Technologies. The tasks demonstrate practical applications of machine learning, including regression, recommendation systems, and geospatial data analysis on a real-world restaurant dataset.

Task 1: Predict Restaurant Ratings

Objective:

To build a machine learning regression model that predicts the aggregate rating of a restaurant based on available features such as cuisine, price range, location, and customer votes.

Approach:

- Performed data cleaning and handled missing values
- Encoded categorical variables such as city, cuisine, and currency
- Trained and evaluated two models: Linear Regression and Decision Tree Regressor
- Evaluated performance using Mean Squared Error (MSE) and R^2 Score
- Interpreted results and identified key features affecting ratings

Tools & Libraries:

Python, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn

Task 2: Restaurant Recommendation System

Objective:

To develop a content-based recommendation system that suggests similar restaurants based on user input such as restaurant name, preferred cuisine, and price range.

Approach:

- Created combined profiles using cuisines and price range for each restaurant
- Used TF-IDF Vectorization and Cosine Similarity to calculate similarity between restaurants
- Developed an interactive front-end using Gradio for real-time recommendations

- Displayed top 5 similar restaurants based on input criteria

Tools & Libraries:

Python, Pandas, Scikit-learn, TF-IDF Vectorizer, Cosine Similarity, Gradio

Task 3: Location-Based Restaurant Analysis

Objective:

To perform a geographic and city-level analysis of restaurants using features like latitude, longitude, city, average ratings, and cuisine diversity.

Approach:

- Plotted restaurant locations on an interactive Folium map using coordinates
- Grouped data by city to analyze:
 - Number of restaurants
 - Average aggregate rating
 - Cuisine diversity
- Visualized insights using bar charts and heatmaps

Insights:

- Identified cities with the highest restaurant concentration and average ratings
- Found location-based trends in cuisine preferences and restaurant pricing

Tools & Libraries:

Python, Pandas, Matplotlib, Seaborn, Folium

Conclusion

This internship enabled the practical application of machine learning techniques in real-world scenarios. Each task contributed to enhancing technical skills in data preprocessing, model building, evaluation, and data visualization.

By working on this project, I gained hands-on experience in developing production-ready solutions including an interactive recommendation system and geospatial insights using real datasets.