



Data Science Project

Zomato Restaurant Rating Predictor

Problem Statement :-

“Analyze the Zomato dataset and build a regression model to predict restaurant ratings based on various features such as location, online order availability, booking table option, cost, type of service, etc.?”

Main Objective :-

The main agenda of this project is:

- ❖ Perform extensive Exploratory Data Analysis(EDA) on the Zomato Dataset.
- ❖ Build an appropriate Machine Learning Model that will help various Zomato Restaurants to predict their respective Ratings based on certain features.
- ❖ Deploy the Machine learning model via Flask that can be used to make live predictions of restaurants ratings.

Project Description :-

The project titled "Zomato Restaurant Rating Predictor: Analyzing and Predicting Restaurant Ratings Using Machine Learning" focuses on analyzing the Zomato dataset and developing a regression model to predict restaurant ratings. The code provided performs various data preprocessing and visualization tasks to understand the dataset and prepare it for model training.

The project starts by importing the necessary libraries, such as pandas, numpy, matplotlib, seaborn, and plotly, which are used for data manipulation, visualization, and building machine learning models. It also imports specific classes from scikit-learn for linear regression and ensemble regression models.

The code then reads the Zomato dataset from a CSV file and displays its shape, data types, and missing values. Unnecessary columns like "url" and "phone" are dropped, and duplicate rows are removed. The dataset is further cleaned by dropping rows with missing values.

Data exploration and visualization techniques are employed to gain insights from the dataset. Visualizations include bar plots, histograms, and pie charts to visualize various attributes such as famous restaurant chains, table booking availability, online order delivery, cost distribution, and rating distribution.

The dataset is preprocessed by transforming categorical variables into numeric using label encoding. The final dataset is split into input features (X) and the target variable (y) for model training.

Two regression models, namely Linear Regression and Random Forest Regression, are trained on the dataset using the scikit-learn library. The trained models are evaluated using the R-squared score, which measures the goodness of fit of the models. The Random Forest Regression model outperforms the Linear Regression model in predicting restaurant ratings.

Finally, the best-performing Extra Trees Regression model is saved using pickle for future use.

Summary :-

In summary, the project involves data analysis, preprocessing, and visualization of the Zomato dataset. It utilizes machine learning techniques to train regression models for predicting restaurant ratings based on various features. The project aims to provide insights into factors influencing restaurant ratings and to develop a predictive model that can assist in decision-making for restaurant owners and customers.



Improvement Needed :-

1. The project would benefit from a more comprehensive evaluation of different machine learning algorithms to identify the best-performing model for restaurant rating prediction.
2. Adding additional features or incorporating external data sources could enhance the predictive accuracy of the model and provide more meaningful insights for restaurant owners and customers.



Conclusion :-

The project "Zomato Restaurant Rating Predictor: Analyzing and Predicting Restaurant Ratings Using Machine Learning" successfully achieved its objectives of analyzing the Zomato dataset and building a regression model to predict restaurant ratings. Here are the key conclusions drawn from the project:

1. **Dataset Analysis** : The Zomato dataset contains valuable information about restaurants, including attributes like location, cuisines, cost, online ordering, and more. Exploratory data analysis helped understand the distribution and relationships between different variables.
2. **Data Preprocessing** : The dataset underwent several preprocessing steps, including handling missing values, removing duplicates, and converting categorical variables into numeric form using label encoding. This ensured that the data was in a suitable format for model training.
3. **Data Visualization** : Visualizations provided insights into various aspects of the dataset, such as the most famous restaurant chains, table booking availability, online order delivery, cost distribution, and rating distribution. These visualizations helped identify patterns and trends within the data.
4. **Model Training and Evaluation** : Two regression models, Linear Regression and Random Forest Regression, were trained on the dataset. The models were evaluated using the R-squared score, which measures the proportion of variance in the target variable that can be explained by the input features. The Random Forest Regression model outperformed the Linear Regression model, indicating its better predictive capability for restaurant ratings.
5. **Model Deployment** : The best-performing Extra Trees Regression model was saved using pickle, making it available for future use. This allows the model to be deployed in a production environment for real-time predictions.

Overall, the project demonstrated the potential of machine learning techniques in predicting restaurant ratings based on various features. The developed model can assist restaurant owners in understanding the factors that influence ratings and make informed decisions to improve their offerings. Additionally, customers can utilize the model to make informed choices when selecting restaurants based on predicted ratings.