PROJECT PROPOSAL

TITLE: PREDICT RESTAURANT RATINGS

OBJECTIVE:

The objective of this project is to develop a machine learning model that accurately predicts the aggregate rating of a restaurant based on various features. By leveraging data science techniques, we aim to provide valuable insights to restaurant owners and stakeholders to help them improve their ratings and enhance customer satisfaction.

STEPS:

1. Data Preprocessing:

- ✓ Handle missing values: Identify and handle missing values in the dataset using appropriate techniques such as imputation or removal.
- ✓ Encode categorical variables: Convert categorical variables into numerical representations using techniques like label encoding or one-hot encoding.
- ✓ Split the data: Divide the dataset into training and testing sets to facilitate model training and evaluation.

2. Model Selection and Training:

- ✓ Choose regression algorithms: Evaluate and select suitable regression algorithms such as linear regression, decision tree regression, or ensemble methods based on the nature of the problem and dataset characteristics.
- ✓ Train the model: Train the selected regression model on the training data to learn the underlying patterns and relationships between features and the target variable.

3. Model Evaluation:

- ✓ Performance metrics: Assess the performance of the trained model using standard regression evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), and R-squared (R2) score.
- ✓ Cross-validation: Employ cross-validation techniques to ensure robustness and generalization of the model.

4. Interpretation and Analysis:

- ✓ Interpret model results: Analyze the coefficients or feature importance scores generated by the model to understand the relative impact of different features on restaurant ratings.
- ✓ Identify influential features: Identify the most influential features affecting restaurant ratings and provide actionable insights to stakeholders for improving their ratings.

IMPLEMENTATION DETAILS:

- ✓ Utilized Python programming language and popular libraries such as pandas, NumPy, scikit-learn, and matplotlib for data preprocessing, model development, and evaluation.

- ✓ Conducted exploratory data analysis (EDA) to gain insights into the distribution and relationships between variables.
- ✓ Implemented both linear regression and decision tree regression models to compare their performance in predicting restaurant ratings.
- ✓ Evaluated model performance using a holdout test set and validated results through cross-validation techniques.

CONCLUSION:

By following the outlined steps, we aim to develop a robust and accurate machine learning model for predicting restaurant ratings. This project has the potential to provide valuable insights to restaurant owners and stakeholders, enabling them to make data-driven decisions to enhance customer satisfaction and business performance.