**Feature Selection Report**

1. Introduction

This report outlines the feature selection process undertaken for the restaurant recommendation system project. Feature selection is crucial for building effective machine learning models by identifying the most relevant and informative features that contribute significantly to the prediction of user preferences.

2. Feature Selection Techniques:

Filter Methods:

Correlation Analysis: Calculated correlation coefficients (e.g., Pearson, Spearman) between features and the target variable (user ratings). Selected features with high correlation.

Chi-Square Test: Applied the chi-square test for feature selection in cases where both the feature and the target variable are categorical.

Information Gain: Measured the information gain provided by each feature in reducing the uncertainty about the target variable.

Wrapper Methods:

Recursive Feature Elimination (RFE): Trained a model (e.g., Support Vector Machine) and iteratively removed the least important features based on their coefficients or feature importance scores.

Embedded Methods:

L1 Regularization (Lasso): Utilized L1 regularization in models like Lasso regression, which automatically performs feature selection by setting the coefficients of less important features to zero.

Decision Tree-based Methods: Employed decision tree algorithms (e.g., Random Forest) to assess feature importance based on their contribution to the decision-making process.

3. Feature Selection Results:

Top-performing features:

[List the top-performing features identified through the feature selection process, e.g., user age, user location, cuisine type, price range, past dining history, review sentiment, restaurant ratings.]

Removed features:

[List the features that were removed due to low relevance or high redundancy.]

4. Justification:

The selected features were chosen based on their statistical significance, predictive power, and interpretability.

Features with low correlation, low information gain, or high redundancy were removed to improve model performance and reduce computational complexity.

5. Conclusion:

The feature selection process resulted in a reduced set of highly informative features that are expected to improve the accuracy and efficiency of the recommendation system. The selected features will be used to train and evaluate the machine learning models in the subsequent stages of the project.