# AI Chat Review Report

## Introduction

This report provides an overview of the assistance provided for completing the Data Mining assignment using the SEMMA methodology. The steps include data sampling, exploration, modification, modeling, and assessment, with recommendations and observations on model performance.

## SEMMA Steps Review

### 1. Sample

The data was successfully loaded and split into training and testing sets. Techniques such as SMOTE were utilized to address class imbalances, ensuring a balanced dataset for model training.

### 2. Explore

Data exploration was conducted to understand the distribution of features, check for missing values, and analyze correlations. Visualizations like histograms and heatmaps were created to identify patterns and relationships within the data.

### 3. Modify

Data preprocessing was performed, including handling missing values, encoding categorical variables, and scaling numerical features. Feature engineering was also applied to create new features and drop irrelevant ones.

### 4. Model

Multiple models were trained, including Logistic Regression, Random Forest, and XGBoost. Hyperparameter tuning was performed using GridSearchCV to optimize model performance. The tuned Random Forest model showed improved metrics.

### 5. Assess

Model performance was evaluated using metrics such as accuracy, precision, recall, and F1 score. The tuned Random Forest model demonstrated an F1 score of 61.86%, indicating a balanced performance between precision and recall.

## Conclusion and Recommendations

The SEMMA methodology was successfully implemented, and model performance improved after hyperparameter tuning. To further enhance the model, consider additional feature engineering, experimenting with different algorithms, and using ensemble methods. Cross-validation can be performed to ensure model generalizability.