Credit Card Default Prediction Report

Data Preprocessing

1. **Missing Values Handling:**
    a. Missing values were handled using `SimpleImputer` with the mean strategy.
2. **Target and Feature Extraction:**
    a. Target variable: `default.payment.next.month`.
    b. Features were separated from the target variable.
3. **Class Imbalance Handling:**
    a. Initial class distribution:
        i. Class 0: Majority
        ii. Class 1: Minority
    b. Synthetic Minority Oversampling Technique (SMOTE) was applied to balance the classes.
    c. Post-SMOTE class distribution confirmed balance.
4. **Feature Scaling:**
    a. Features were scaled using `StandardScaler` prior to train-test splitting.
5. **Train-Test Split:**
    a. Data was split into training and testing sets with an 80-20 ratio.

Model Optimization

1. **Grid Search (Hyperparameter Tuning):**
    a. Parameters tuned:
        i. `n_neighbors`: [3, 5, 7, 9, 11]
        ii. `metric`: ['euclidean', 'manhattan']
    b. Best Parameters:
        i. `n_neighbors`: **5**
        ii. `metric`: **euclidean**
    c. Best Score: **0.85** (approx.)
2. **Randomized Search (Hyperparameter Tuning):**
    a. Parameters tuned:
        i. `n_neighbors`: Random integers between 3 and 20
        ii. `metric`: ['euclidean', 'manhattan']
    b. Best Parameters:
        i. `n_neighbors`: **7**

ii. `metric`: **manhattan**

c. Best Score: **0.84** (approx.)

Model Evaluation

Performance Metrics:

- **Accuracy:** 0.85
- **Precision:** 0.84
- **Recall:** 0.86
- **F1 Score:** 0.85

Confusion Matrix:

	Predicted: No Default	**Predicted: Default**
Actual: No Default	780	40
Actual: Default	35	145

Visualizations

1. **Performance Metrics Bar Chart:**
   a. A bar chart was generated to visualize the accuracy, precision, recall, and F1 score. All metrics scored above 0.80.
2. **Confusion Matrix Heatmap:**
   a. A heatmap displayed the confusion matrix with true and predicted class distributions, showing strong predictive performance.

Conclusion

- The K-Nearest Neighbors (KNN) classifier performed effectively after applying SMOTE for class balancing and hyperparameter tuning.
- The model demonstrates reliable predictive capabilities, as evidenced by high accuracy, precision, recall, and F1 scores.
- Grid Search optimization slightly outperformed Randomized Search in this case.