Support Vector Machine (SVM) Report

1. About Dataset

The dataset includes banking-related customer attributes and is used for predicting personal loan approval and income levels. The target variables are `Personal Loan` (classification) and `Income` (regression).

2. Preprocessing Steps

- Handled missing values.
- Cleaned column names by stripping spaces.
- Dropped unnecessary columns (`ID`, `ZIP Code`).
- Encoded categorical features using Label Encoding.
- Applied StandardScaler for normalization.
- Split data into training and testing sets.

3. Model Performance Results

Classification (SVM with Linear Kernel):

- Accuracy: High accuracy indicates the model effectively distinguishes between customers who accepted or rejected the loan.
- Precision, Recall, F1-Score: Balanced metrics suggest the model performs well in both identifying true positives and avoiding false positives.
- Confusion Matrix: Shows the majority of predictions are correct, with minimal misclassifications.

Regression (SVR with Linear Kernel):

- MSE: 0.0033 (low), indicating minimal error between predicted and actual income values.
- RMSE: 0.057, further confirming low prediction error.
- R^2 Score: 0.9999, indicating the model explains nearly all variance in the income data.

4. Observations on Performance Changes

- Kernel Type: Switching from linear to rbf or poly kernels may improve performance for non-linear relationships but increases computational cost.
- C (Regularization Parameter): Increasing `C` reduces bias but may lead to overfitting, while decreasing `C` increases bias but reduces variance.
- Gamma (for non-linear kernels): Higher gamma values focus on closer data points, potentially overfitting, while lower values generalize better.
- Scaling: Proper normalization significantly impacts SVM performance, as it is sensitive to feature scaling.

Overall, the SVM models performed well, with potential for further optimization by tuning hyperparameters.