# Predictive and Logistic Regression, SVM

**Aim:** Designing a predictive regression model that forecasts sales based on the "Advertising.csv" dataset, logistic regression and Support Vector Machines (SVM) to predict defaulters using the "Credit.csv" and "Credit-Modified.csv" datasets.

**Sales Forecasting Using Linear Regression :**
1.**Data Exploration:**
Begin by exploring the "Advertising.csv" dataset for sales forecasting. This involves importing essential libraries such as pandas, numpy, matplotlib.pyplot, seaborn, and scikit-learn.

**2.Data Loading and Selection:** Load the "Advertising" dataset into a pandas DataFrame (df). Choose relevant features ('TV', 'radio', 'newspaper') and define the target variable as 'sales'.

**3.Train-Test Split:** Split the dataset into training and testing sets using the train_test_split function from scikit-learn.

**4.Linear Regression Modeling:** Create a linear regression model, fitting it to the training data, and subsequently making predictions on the test data.

**5.Evaluation Metrics:** Evaluate the performance of the linear regression model by calculating and printing key metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

**6.Scatter Plot:** Visualize the model's predictions by generating a scatter plot that compares actual sales values to the predicted sales.

**Default Prediction Using Logistic Regression and SVM:**

**1.Data Preprocessing:**

For default prediction on the "Credit.csv" and "Credit-Modified.csv" datasets, load both datasets (credit_df and credit_mod_df). Conduct necessary preprocessing steps, including dropping unnecessary columns and creating dummy variables for categorical features.

**2.Model Creation and Fitting:** Create logistic regression and Support Vector Machines (SVM) models for default prediction. Fit these models to the respective datasets.

**3.Model Evaluation :**Make predictions using the trained models and assess their accuracy through metrics such as accuracy scores. Visualize predictions through scatter plots.

**4.Confusion Matrices**: Utilize a function (plot_confusion_matrix) to plot confusion matrices for both logistic regression and SVM models.

**Classification Metrics and ANOVA Test:**

**1.Classification Metrics:** Import precision, recall, and F1 score metrics from scikit-learn. Define a function (print_classification_metrics) to print these classification metrics. Print the metrics for both logistic regression and SVM models.

**2.ANOVA Test:** Conduct an ANOVA test on the features of the "Credit" dataset to analyze the variance between groups. Display the F-statistic, p-value, and a summarized DataFrame of the results.

**3.Logistic Regression Coefficients:** Print the coefficients of the logistic regression model, providing insights into the impact of each feature on the predicted outcome.

**Accuracy of Logistic Regression:**

```
Logistic Regression Accuracy: 0.9416666666666667
```

```
Logistic Regression Metrics:
Precision: 0.7727
Recall: 0.8947
F1 Score: 0.8293
```

**Accuracy of SVM:**

```
SVM Accuracy: 0.8833333333333333
```

```
SVM Metrics:
Precision: 0.6923
Recall: 0.4737
F1 Score: 0.5625
```

**Confusion Matrix for SVM:**

```
Confusion Matrix for SVM:
[[97  4]
 [10  9]]
```

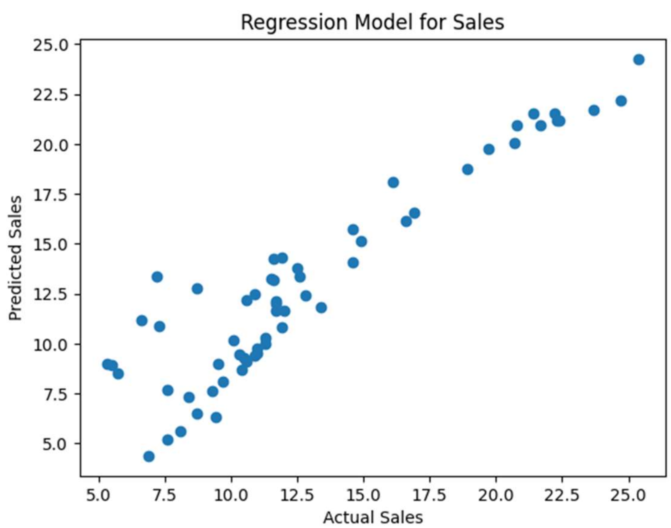
**Confusion Matrix for Loggistic Regreesion:**

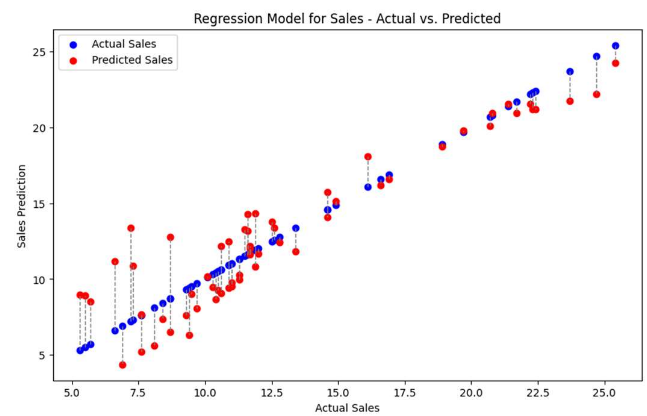
```
Confusion Matrix for Logistic Regression:
[[96  5]
 [ 2 17]]
```

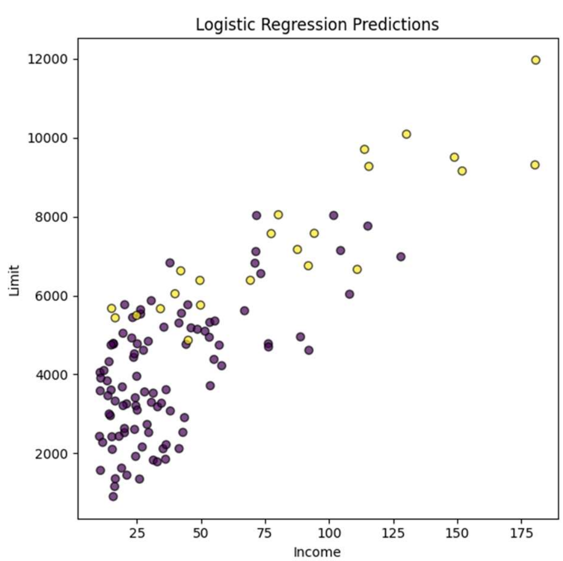
Anova Test Results:

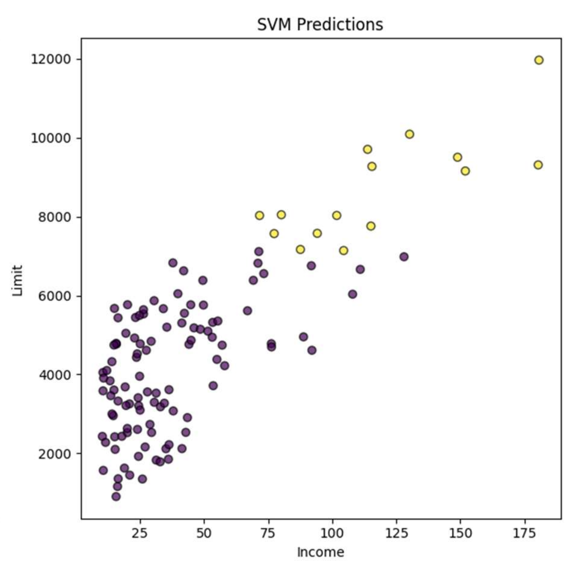
| | Feature | F-Statistic | P-Value |
|---|---|---|---|
| 0 | Income | 78.609099 | 2.571368e-17 |
| 1 | Limit | 252.248530 | 2.396566e-44 |
| 2 | Rating | 259.693257 | 2.465673e-45 |
| 3 | Cards | 4.025334 | 4.549784e-02 |
| 4 | Age | 0.579367 | 4.470113e-01 |
| 5 | Education | 0.036607 | 8.483654e-01 |
| 6 | Balance | 494.736145 | 8.163230e-72 |
| 7 | Gender_Male | 0.004563 | 9.461796e-01 |
| 8 | Student_Yes | 33.242414 | 1.633990e-08 |
| 9 | Married_Yes | 0.313386 | 5.759248e-01 |
| 10 | Ethnicity_Asian | 0.036340 | 8.489131e-01 |
| 11 | Ethnicity_Caucasian | 0.045334 | 8.314999e-01 |

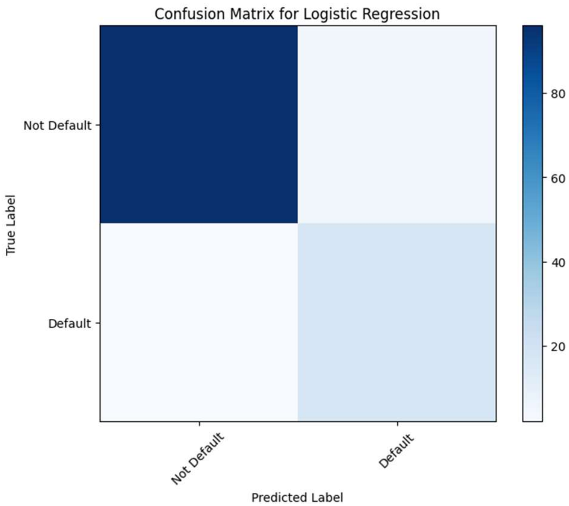
**Plots:**

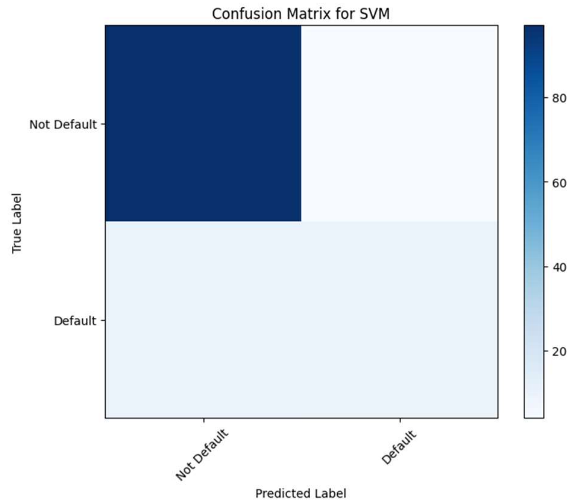



Confusion Matrix for Logistic Regression
Not Default
Default
True Label
Not Default
Default
Predicted Label
80
60
40
20


Confusion Matrix for SVM
Not Default
Default
True Label
Not Default
Default
Predicted Label
80
60
40
20