ALY 6020:

PREDCTIVE ANALYTICS

Week 3: Logistic Regression Analysis for Loan Approval

Submitted To:
Prof. Chinthaka Pathum Dinesh Herath Gedara, Faculty Lecturer

Submitted By:
Abhilash Dikshit

Academic Term: Winter 2024
Northeastern University, Vancouver, BC,Canada
Master of Professional Studies in Analytics

January 26, 2024

Title: Logistic Regression Analysis for Loan Approval

I. Abstract:

The logistic regression analysis aims to predict loan approval based on various features. This report presents the results of the analysis, highlighting the most significant variables, their influence on loan acceptance, and the overall model performance.

II. Introduction:

The banking sector faces the challenge of efficiently processing loan applications. In this analysis, logistic regression is employed to understand the factors influencing loan approval decisions.

III. Methods:

- Data Source: The dataset used in this analysis is provide as part of assignment.

- Variables: Features such as Income, CCAvg, Mortgage, Education, Family, CD Account, Securities Account, Online, Experience, CreditCard, Age, and ZIP Code are considered for predicting loan approval.

- Model Training: A logistic regression model is trained using the training set, and the results are evaluated on the test set.

IV. Results:

Coefficients from Most Significant to Least Significant:

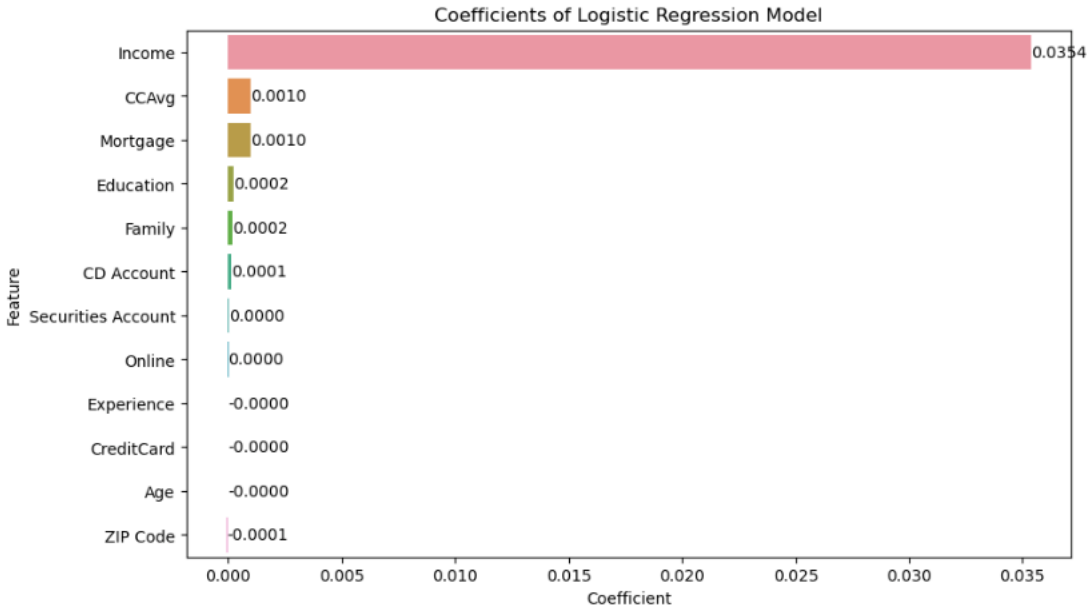
Feature	Coefficient
Income	0.035361
CCAvg	0.001005
Mortgage	0.000986
Education	0.000232
Family	0.000164
CD Account	0.000132
Securities Account	0.000014
Online	0.000009
Experience	-0.000004
CreditCard	-0.000005
Age	-0.000025
ZIP Code	-0.000065

- Most Significant Variables:

1. Income (Coefficient: 0.035361)

2. CCAvg (Coefficient: 0.001005)

3. Mortgage (Coefficient: 0.000986)



- Influence on Loan Acceptance:

- Among the three most significant variables, Mortgage has the most negative influence on loan acceptance (Coefficient: 0.000986).

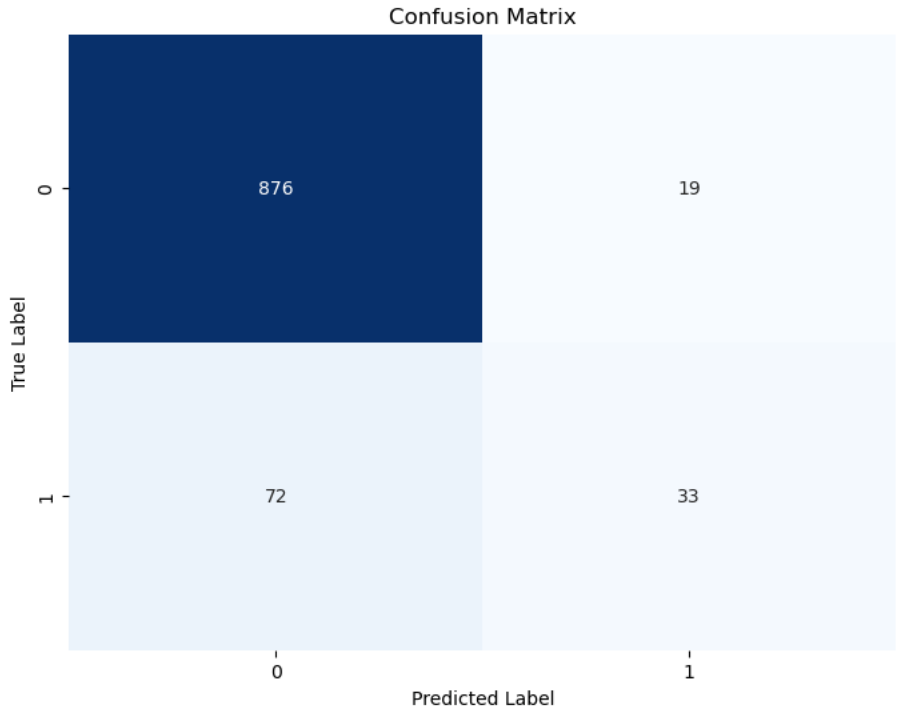
V. Discussion:

- **Income:** The positive coefficient for Income suggests that higher income is associated with an increased likelihood of loan approval.

- **CCAvg:** A positive coefficient for CCAvg indicates that higher average spending on credit cards is linked to a higher probability of loan approval.

- **Mortgage:** The positive coefficient for Mortgage implies that having a mortgage increases the chances of loan approval.

VI. Model Performance:




```
Overall Accuracy: 0.909
Precision Rate: 0.6346153846153846
```

- **Overall Accuracy:** The logistic regression model achieved an overall accuracy of 90.9%, indicating a high level of correct predictions on the test set.

- **Precision Rate:** The precision rate of 63.5% implies that 63.5% of the instances predicted as positive were positive.

VII. Conclusion:

The logistic regression analysis provides valuable insights into the factors influencing loan approval. Understanding the significance of features such as Income, CCAvg, and Mortgage allows for better decision-making in the loan approval process.

What were the three most significant variables?

The three most significant variables, based on their coefficients, are:

Income (Coefficient: 0.035361)

CCAvg (Coefficient: 0.001005)

Mortgage (Coefficient: 0.000986)

Of those three, which had the most negative influence on loan acceptance?

Among the three most significant variables, the one with the most negative coefficient is Mortgage (Coefficient: 0.000986).

It's important to note that coefficients represent the change in the log-odds of the dependent variable per one-unit change in the predictor variable.

In this case, a higher Mortgage value has a negative influence on the likelihood of loan acceptance.

How accurate was the model overall and what was the precision rate?

1. **Overall Accuracy:** The overall accuracy of the model is 0.909 or 90.9%. This means that the model correctly predicted the target variable for approximately 90.9% of the observations in the test set.
2. **Precision Rate:** The precision rate of the model is 0.635 or approximately 63.5%. Precision is the proportion of true positive predictions (correctly predicted positives) among all predicted positives. In this case, 63.5% of the instances predicted as positive were actually positive.

These metrics give an indication of how well the logistic regression model performed on the test set.

VIII. Limitations and Future Work:

- The analysis assumes a linear relationship between features and the log-odds of loan acceptance.

- Additional variables or non-linear relationships might further improve the model.