# Customer Default Identification Report

## Data Summary

Approximately 30,000 Credit One customers’ balance history was evaluated using **Regression** and **Classifier Machine Learning Models** to determine if Credit One can minimize the number of defaulted loans.

## Data Dictionary

Credit Given – Individual consumer credit and his/her family credit

Gender – Boolean entries as male (1) or female (2)

Education – Graduate school (1), University (2), High School (3), Others (4)

Marital Status – Married (1), Single (2), Divorced (3), Others (4)

Age – Reported in years

Payment History – reported 6-month payment status (from April 2005 to September 2005)

Bill Statement Amount – reported 6-month bill statement (from April 2005 to September 2005)

Previous Payment Amount – reported 6 months payments (from April 2005 to September 2005)

Client’s behavior – Boolean entries provided for not-defaulting (0), and defaulting (1)

## Analysis

Three Regression Models were used to evaluate the Credit Given as the dependent variable: Random Forest Regression, Linear Regression, and Support Vector Regression. All three models had a **low R2 score** (coefficient of determination) indicating that the models were poor at correlating features to the dependent variable. All three models also had a **high RMSE** (Square Root of Mean Squared Error) indicating that the models had a high error.

Chart, scatter chart

Description automatically generated

|  |  |  |
| --- | --- | --- |
| Model Name | R2 Score | RMSE |
| Random Forest Regression | 0.472 | 93531.652 |
| Linear Regression | 0.353 | 103482.441 |
| Support Vector Regression | -0.037 | 131044.619 |

Figure 1. Regression ML Models

Due to the Regression Models’ poor fit to the data, a Classification Model was used to determine if Default as a dependent variable could be predicted with high accuracy. The Decision Tree Classifier was effective at reaching an 82% accuracy score, where **Payment History Status appeared to be the most important feature needed to predict Default customers**. When narrowing to only Payment History the model was able to predict Default customers with 78% accuracy. **The least important feature was the amount of Credit Given**.

A final Classification Model was evaluated to determine if discretized Credit Given could be determined based on the demographic data alone. The model was 70% accurate at determining Credit Given to customers with Age being the most important feature. The Decision Tree showed that customers < 25 years old will likely have a Credit Balance of either below $40k or between the $40k – 80k range. Customers > 25 years old will likely have a Credit Balance between the $80k – 500k range. The most important demographic features for determining the Credit balance was age > education > gender > marital status.

# Conclusions

## How do you ensure that customers can/will pay their loans?

From the machine learning models, the most significant feature that can accurately predict a customer’s likelihood of paying their loans is their payment history. In this data set, the most important of months were the two most recent months (August and September). One option that Credit One can provide to customers is to respond more quickly to customers’ late payments. **Credit One can contact customers to help them with a payment plan or resources so that they do not carry over a late payment into their next statement**.

## Can we approve customers with high certainty?

The machine learning model showed that there was a strong relationship between the age of a customer and the amount of Credit Given to a customer. Younger customers were likely to hold smaller loan amounts between $0-80k. While it may appear that the smaller loan amounts account for the greatest number of defaults, the machine learning model showed that age was not a contributing factor. None of the features that related to customer’s demographic background was significant in determining whether or not a customer was going to default on their loan. **We cannot predict a customer’s behavior but checking a customers’ credit history may provide excellent insight in determining if their past payment history has shown late payments**.