# credit-risk-classification

**credit-risk-classification**

Machine learning techniques are used to analyze a dataset of historical lending activity from a peer-to-peer lending services company to build a model that can identify the creditworthiness of borrowers.

**Overview of the Analysis**

Factors considered in the analysis included data on:

* the size of the loan
* its interest rate
* the borrower's income
* the debt to income ratio
* the number of accounts the borrower held
* derogatory marks against the borrower
* the total debt

The dataset (77,536 data points) was split into training and testing sets. The training set was used to build an initial logistic regression model (Logistic Regression Model ) using the LogisticRegression module from [scikit-learn](https://scikit-learn.org/stable/index.html). Logistic Regression Model was then applied to the testing dataset. The purpose of the model was to determine whether a loan to the borrower in the testing set would be low- or high-risk and results are summarized below.

This model was drawing from a dataset that had 75,036 low-risk loan data points and 2,500 high-risk data points.

**Logistic Regression Model:**

* Precision: 92% (an average--in predicting low-risk loans, the model was 100% precise, though the model was only 85% precise in predicting high-risk loans)
* Accuracy: 92%
* Recall: 95% (an average--the model had 99% recall in predicting low-risk loans, but 91% recall in predicting high-risk loans)

**Summary:**

Logistic Regression Model is less likely to predict false negative results. Logistic Regression Model predicted slightly more false positives (low-risk when the actual was high-risk).

If the goal of the model is to determine the likelihood of high-risk loans, no model scores above 90% precision. Logistic Regression Model had fewer false predictions of the testing data overall and would be the best model to use based on the high accuracy and recall of this model.