**Overview**

The purpose of the analysis was to assess the risk level prospective loans and endeavor to predict which prospective loans represent good risks or bad risks for the financial institution in question.

The data utilized covered information pertinent to the loan such as loan size and interest rate along with data about the prospective borrower such as income, number of accounts, derogatory marks, and total debt. The goal was to predict an eventual loan status with a 0 indicating a loan in good standing and a 1 indicating a loan determined to be high risk.

A logistical regression model, A Decision Tree model, and a K Neighbors model were each implemented utilizing a train-test split to fine tune accuracy before being run on the data itself. After the models were run, a confusion matrix was created so that the evaluation could be assessed for overall accuracy.

**Results:**

* **Logistical Regression Model:**
* + precision recall f1-score support
* 0 0.99 1.00 0.99 18760
* 1 0.85 0.84 0.85 624
* **Decision Tree Model**
* precision recall f1-score support
* 0 0.99 1.00 1.00 18702
* 1 0.93 0.84 0.88 682
* **K Neighbors Model**
* precision recall f1-score support
* 0 0.99 1.00 1.00 18702
* 1 0.93 0.84 0.88 682

**Summary:**

Both the K neighbors and Decision Tree models slightly out performed the Logistical Regression model in terms of precision correctly predicting high risk loans at 93 percent accuracy versus 85 percent accuracy for the Logistical Regression model. All three models did well with 99 percent of predicted good loans being accurate.

In terms of recall, the models produced uniform results. Once again, the good loans were identified and predicted at very high rate, perfect in fact. However, there was a drop on all three models predicting high risk loans with all three coming in at 84 percent accuracy. This might suggest that the data is insufficient to predict at a higher rate of accuracy. Perhaps another variable or data point would prove indicative in determining greater risk in providing a loan as the models continuously returned the same value.

Given that the identification of high risk loans is likely of paramount importance to a financial institution, the accuracy in identifying which loans could be high risk should be considered the key stat here. Of course, a major variable is missing, what is the current rate of high risk loans being given under the current protocols? Obviously, if 16 percent of loans being high risk is an improvement on that number, a statistical model could prove highly useful. Generally speaking, 84 percent can plausibly be categorized as quite accurate, so even the low end of the model is still very accurate.

In any event, my recommendation would be the gathering of more demographic data related to prospective loans in the hopes that more data will lead to a statistical model that is even more accurate in identifying potential high risk loans.