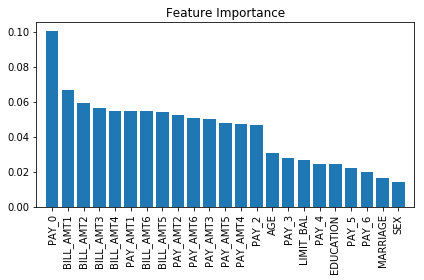
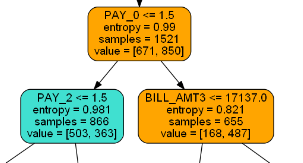
Phillip Roman

Task 3: Final Report Concerning Credit One Analysis

 Credit One customers are currently defaulting at a rate of 22%. Lowering the default rate will save the company money. After reviewing the available customer attribute and credit information, we have determined the best course of action to increase the likelihood that customers are going to pay their loans. There were a few demographic attributes that were deemed to have a low impact on likelihood to default. The Feature Importance graph below ranks the importance of each feature when attempting to predict default. All customer demographic features where removed due to a low importance. Also, past payments from month 3 and older where removed. Bill amount and payment amount where the most important categories to predicting default. The remaining features can be used in a model to help predict if the customer will default on their upcoming payment. Several models that can be implemented are based on “trees” where a loan’s features determine the prediction the loan is categorized with. In the example below, if the PAY\_0 is less than or equal to 1.5 than PAY\_2 is evaluated.

This approach is carried on until a determination of default or paid is achieved. A total of 5 models were addressed and three used a form of a tree to predict. The model that had the greatest accuracy was called Random Forest Classification. This model correctly predicted the correct outcome 80% of the time. Since we know that 11 loans will default for every 50 loans, it is important to know how accurate the model is at predicting each individual possibility. The model could predict a defaulted loan 70% of the time and a paid loan 83% of the time. A random binary guess results in a 50% accuracy. Such an implementation does have some risk but the benefits are greater. In the dataset, there were 1,991 loans that defaulted. The model was able to predict 681 loans. The model could have potentially reduced up to 34% of the defaulted loans. However, the model misclassification 297 paid loans as defaulted loans. This misclassification would have reduced the number of paid loans from 7,009 to 6,712 (a 4% reduction). The costs of having 681 defaulted loans is greater than the profit from having kept the 297 loans paid loans. Therefore, it is recommended to implement a similar Random Forest model in your credit process.