1. For the unscaled data I think the logistic regression model will perform better. I imagine the factors to determine a person’s loan risk would be linear as there is a threshold that must be crossed to go from low risk to high risk. Logistic regression tends to work better than random forest on linear data.
2. On the unscaled data the random forest classifier performed better. This could be due to the data being more complex and less linear than I originally thought. Being a low or high risk doesn’t appear to be a black and white area.
3. I think after the data is scaled both models will perform better but I think the logistic regression model will outperform the random forest classifier. This is due to similar reasons as the first prediction, but the scaled data will hopefully appear more linear for the logistic regression model.
4. Using the scaled data the logistic regression model not only improved but also performed better than the random forest classifier. The random forest did not improve like I also thought, it scored about the same as the unscaled. The scaled data appears to have been easier for the logistic regression model to read.