Supervised Learning Challenge Analysis

Based on the two models, I feel that the logistic regression model is a better fit because it will perform better for this given dataset. Both models were good for classification purposes, however I believe with over 80 variables to take into consideration, the logistic model is the better fit. The random forest is recommended for simpler classification problems.

Evaluating both models, without scaling the data, the random forest model performed better on smaller test data. Once the data was scaled, the logistic regression became the better model. It had an accuracy score of .66. The random forest model had an accuracy score from .64 to .5. From what I gathered; this is expected as the random forest model doesn’t rely on scaled data while the logistic model does.

This would not be a good model to use for credit worthiness rating purposes. It contains too many inaccuracies which would result in money lost on a regular basis if loan decisions were made using these models.