Credit Default Classification Memo

I have chosen to do a classification analysis for my final project, and, for that, I have chosen the default credit data set to classify whether which individuals based on a certain characteristic are likely to default. I am highly interested in micro-lending in developing countries, and I believe this analysis could serve as a building block to what is to come. Africa is experiencing the fastest growth and as its economy is growing and adapting to the latest financial technologies, there are financial opportunities that emerge. Particularly, I am interested in the financial lending opportunity for small business individuals who seek credit to expand their existing business. A major task we face is how to classify people with no little to no digital data which can be to quantify their credit worthiness. There are many companies with bigger resources already attempting this, but as some with limited capital and skill, I plan to learn through doing a shadow study on credit default and using the result to predict credit default in my area of interest. That is, I hope the result from this analysis can give me the predictive insight to be scaled and transform into the African market. This will require great precision in my predictive power. In other words, I want to low test error.

For my analysis, I would like to predict credit default based on certain attributes of the factors included in the dataset. This data contains a mixture of qualitative and quantitative variables, which in my opinion are useful for predicting credit default. Even though this data is from 2005, it bares similarity to our current period where their easy access to credit due to the extended quantitative easing that was implemented by the Fed to support the injured economy. Two key groups in my data are University student and grad student. These are the people facing some large credit balance but with amounting pressure to repay them in the future considering all future expected expenses. If time permits, I pan to compare 2005 payment amount to recent payment amount to see if other conditions held constant would result in the same classification. To achieve that, I will certainly need to do some merging or cleaning. To be specific, I will need to do some replacement. That is, instead of the 2005 payments and bill amounts, I will include the most recent monthly bill and payment made by students with varying completed education to their loan amount. However, credit limit balance will be kept unchanged.

While this project does not reflect my market of interest, I believe a thorough classification analysis will not only deepen my understanding of credit default analysis, but also give me the tools to implement a similar algorithm for my market of interest. I hope that I can achieve a high predictive accuracy, which if applied elsewhere (different market), would still give a reasonable predictive power.

Citation:

Yeh, I. C., & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients

https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients