**Introduction:**

For banks, risk management and default detection has always been a crucial part in issuing credit cards. Though the law has stringent measures against credit card defaulting, it is still prevalent in most parts. Predicting the nature of a customer of whether he might be a defaulter or not is a complex function. Statistic models can play the role of assistive banker in ensuring default less credits and safe banking. In order to reduce or even prevent loss of this kind, banks need to determine appropriate given credit for each special client based on their information.

Predictive modelling is the process of applying statistical technique to derive a model between response variables and predictors and approach to the best. In this project, the data taken from the file train and we are provided 20,000 credit card customers consists of a single binary response, indicating default in the October 2005 payments and 23 predictor variables. We are going to demonstrate the process of deriving the best fit predictive model, which represents the relationship between default of credit card payment and some significant features. The task is to produce a function takes a n × 23 matrix of predictor variables as its input and produces a n-dimensional vector of predictions with 0 for no default and 1 for default as its output.

At the very first parts of this paper, we will present missing data handling and denoting, defective data cleaning, making sure variables are appropriately categorised into quantitative (continuous or discrete) or categorical. Histogram and bar-chart will be plotted to illustrate their distribution respectively. As the prior process of selecting model, bivariate analysis will be included continuous and categorical variables and conducted by producing box-plot and scatter-plot to examine the effect of predictors on defaulting, and the relationship will be stated clearly. The cross validation method also used to select the final model. The following section is the most significant where we derive the final model by comparing the best model that derivate by algorithms involve backwards, forwards, stepwise and automatic. And in our last sections, statistic summary and figures will be illustrated to support the assumptions checking of our final model. As verification of our model selection, a prediction will be made on the default of a customer by using the final model and some new data.