Dear Participants,

Please find below graded individual assignment.

You are requested to create an India credit risk(default) model, using the data provided in the spreadsheet raw-data.xlsx, and validate it on validation_data.xlsx. Please use the logistic regression framework to develop the credit default model.

Hints :

Data description - Please direct them to the video - Default Risk Prediction. After removing variables for multicollinearity, we should try to take at least one variable for creating the model from each of the 4 factors namely -

1) Profitability
2) Leverage
3) Liquidity
4) Company's size

In Dr. Sarkar's video of Default Risk Estimation, he has clearly bifurcated all the variables in different buckets.

Creation of new variables - This is an important step in the project as the company which is the biggest in size, will also have bigger asset size, cash flows etc. (Hint: We need to think in terms of ratios - Equity to asset ratio, debt to equity ratio etc)

Dependent variable - We need to create a default variable which should take the value of 1 when net worth is negative & 0 when net worth is positive.

Validation Dataset - We need to build the model on raw dataset and check the model performance measures on validation dataset.

Please find attached the files to be referred.