1. Use the file “german\_credit\_data.csv”
2. Partition in 3 parts, Training, Validation and Test in proportion of about 50%, 30% and 20%
3. Build “Logistic Regression” using Training data. Apply the model on Validation data. Select best subset of variables using Validation data. Best subset of variable should minimize cost using following data
   1. cost of predicting “not Creditworthy” as “Credit worthy” is Rs. 15,000 per case
   2. cost of predicting “Creditworthy” as “not Credit worthy” is Rs. 10,000 per case
4. Apply final model on test data to find out cost per case

Additional Exercise:

Whenever person applies for bank loan, a process is followed involving various persons to determine whether to give loan or not. It costs Rs. 5,000 for processing each case. The error involved in this process (manual decision) is negligible and is to be ignored for this case.

A data mining technique is planned to replace this process, for suggesting giving / denying bank loan. Technique predicts whether person is loan worthy or not.

If person is “loan worthy” but is predicted as “not loan worthy” and denied the loan, expected loss is Rs. 10,000. On other hand if a person is “not loan worthy” but predicted as “loan worthy” and given loan, one looses Rs. 25,000.

For building the model, 5,000 samples are collected from past data. The samples are partitioned in “Training” (60%) and “Test” (40%). Model is built using “Neural Network”. The confusion matrix on “Test” data is as follows:

|  |  |  |
| --- | --- | --- |
|  | **Predicted Class** | |
| **Actual Class** | Loan Worthy | Not Loan Worthy |
| Loan Worthy | 800 | 200 |
| Not Loan Worthy | 100 | 900 |

|  |  |
| --- | --- |
| **ITEM** | **Amount (INR)** |
| One time cost of building data mining model : Hardware, Consultant, Tools, etc | 5,000,000 |
| Recurring monthly cost of running data mining model: Hardware, Tools, etc | 1,000,000 |

Typically, every month, 1,000 persons apply for the loan. Using the data given,

1. Calculate ROI
2. Calculate Pay Back Period