**Homework: This homework will test and develop your knowledge on predictive model construction and using their predictive performance for decision making. Please submit your answers for each section with your explanations in document you compile with the original supporting R script file you used for this assignment.**

Use the BankruptcyData.csv file and the BankruptcyNew.csv data files for this assignment. The attributes to develop a predictive model for predicting bankruptcy of firms. Data dictionary is provided below with the variables’ labels in the data sets and the long descriptive names.



|  |  |  |
| --- | --- | --- |
| **Attribute**  **No** | **Short**  **Name** | **Long**  **Name** |
| **1** | fyear | Fiscal year |
| **2** | at | Assets-total |
| **3** | bkvlps | Book value per share |
| **4** | invt | Inventories- total |
| **5** | Lt | Liabilities-total |
| **6** | rectr | Receivables-total |
| **7** | cogs | Cost of goods sold |
| **8** | dvt | Dividends-Total |
| **9** | ebit | Earnings before interests and taxes |
| **10** | gp | Gross profit |
| **11** | ni | Net income (loss) |
| **12** | oiadp | Operating income after depreciation |
| **13** | revt | Revenue- total |
| **14** | dvpsx\_f | Dividends per share -ex-date fiscal |
| **15** | mkvalt | Market value -Total -Fiscal |
| **16** | prch\_f | Price High -Annual-Fiscal |
| **17** | bankruptcy | Output (Y/N) |

1. Are there any variables with missing values? If there are report which ones and remove records with any missing value. Write your script for checking the missing values and variables with missing values. (1)
2. Build a logistic regression model using all attributes except 1, 12, 13 and the 70% of the data for training and 30% for the testing to predict bankruptcy (attribute 17). Which variables are significant predictors for p<0.05? (1)
3. Present a summary table of the odds ratios. (1)
4. Report the confusion matrix using the logistic regression model and the test data. Calculate model accuracy, precision, sensitivity and F1 score. (1)
5. Plot the ROC curve and report the AUC statistic. (1)
6. Construct a decision tree model (called DT-1) to predict the bankruptcy with 70% of the data for training and 30% for the testing. (1)
7. Report the confusion matrix for DT-1 model using the test data. Calculate model accuracy, precision, sensitivity and F1 score. (1)
8. Plot the ROC curve with the AUC statistic reported for DT-1. (1)
9. Comment on which model has better predictive accuracy using AUC statistic. (0.5)
10. The credit organization received new applications with the attribute values in the new Bankrupcynew.csv data file. Use the logit model you constructed in question 2, to predict if the companies will undergo bankruptcy or not. Report the resulting table for your predictions. (1.5)