**Week-3 Assignment 23/50**

**1)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Confusion matrix** | | |  |
|  | **Prediction Positive** | **Prediction Negative** | **Measures** |
| **Actual POSITIVE** | **TP – True Positive (Hit)** | FN – False Negative (type II error) |  |
| **Actual NEGATIVE** | FP – False Positive (type I error) | **TN – True Negative (Correct Rejection)** |  |

The below table shows the format of confusion matrix.

In this question, We have 3 cutoffs 0.25, 0.5, 0.75 , Here we have to develop confusion matrix for each cutoff.

**For cutoff (0.25):**

|  |  |  |  |
| --- | --- | --- | --- |
| **Confusion matrix** | | |  |
|  | **Prediction Positive** | **Prediction Negative** | **Measures** |
| **Actual POSITIVE** | **3** | 0 |  |
| **Actual NEGATIVE** | 8 | **9** |  |

Error Rate = = = 0.4

Accuracy = = = 12/20 =0.6

= = 1

= = 0.53

**For Cutoff (0.5):**

|  |  |  |  |
| --- | --- | --- | --- |
| **Confusion matrix** | | |  |
|  | **Prediction Positive** | **Prediction Negative** | **Measures** |
| **Actual POSITIVE** | **3** | 0 |  |
| **Actual NEGATIVE** | 2 | **15** |  |

Error Rate = = = 0.1

Accuracy = = = 18/20 =0.9

= = 1

= =15/17= 0.88

**For Cutoff (0.75):**

|  |  |  |  |
| --- | --- | --- | --- |
| **Confusion matrix** | | |  |
|  | **Prediction Positive** | **Prediction Negative** | **Measures** |
| **Actual POSITIVE** | **2** | 1 |  |
| **Actual NEGATIVE** | 0 | **17** |  |

Error Rate = = = 0.05

Accuracy = = = 19/20 =0.95

= = 0.66

= = 1

**2)**

**a)** From the given information, We are applying the classification model for the

Sample(n) = 800, Here we have 2 categories a)Frauds and b) non-Frauds

Number of correctly classified Frauds =310

Number of correctly classified non-Frauds = 270

Number of missed frauds = 90 (means model predicted actual frauds as non- frauds)

Number of incorrectly classified as frauds = 130 (model predicted non-frauds as frauds)

|  |  |  |  |
| --- | --- | --- | --- |
| **Confusion matrix** | | |  |
|  | **Prediction Frauds** | **Prediction Non-frauds** | **Measures** |
| **Actual Frauds** | **310** | 90 |  |
| **Actual Non-Frauds** | 130 | **270** |  |

**b) Adjusted Misclassification Rate:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Prediction Frauds** | **Prediction Non-frauds** |  |
| **Actual Frauds** | **310** | 90(FP) | 400 |
| **Actual Non-Frauds** | 130(FN) | **270** | **400** |
|  | 440 | **360** | **800** |

**Misclassification rate: = FP + FN / Total**

**= 90+130 / 800**

**= 0.275**

Missing the adjusted CM for the original dataset -7

You have to put your interpretation/explanation/output data in the text only code is not enough

In #4 the following codes give error message. Can you say why? -10

CoursetopicsRules1<- apriori(Coursetopics.trans, parameter = list(support=0.01, conf=0.1))

# Sort the rules by lift in decreasing order

CoursetopicsRules\_byLift<- sort(CoursetopicsRules, by ="lift", decreasing = TRUE)

In #5 the following codes give an error message can you say why? Also, do you know how to write the lm() function parameters? Also this error causes error in the rest of the codes -10

# F. Get the training set and remove the attributes CHAS, INDUS, and AGE

train.df1 <- train.df[, !names(BostonHousing.df) %in% c("CHAS", "INDUS", "AGE")]

Regeression\_model<-lm(BostonHousing.df1$MEDV ~ ., data=train.df1)

summary(Regeression\_model)