**Q2) Decision Tree Classification (Titanic)**

Information about the Data

Titanic data set

|  |  |  |
| --- | --- | --- |
| **Variable** | **Definition** | **Key** |
| survival | Survival | 0 = No, 1 = Yes |
| Pclass | Ticket class | 1 = 1st, 2 = 2nd, 3 = 3rd |
| sex | Sex |  |
| Age | Age in years |  |
| sibsp | # of siblings / spouses aboard the Titanic |  |
| parch | # of parents / children aboard the Titanic |  |
| ticket | Ticket number |  |
| fare | Passenger fare |  |
| cabin | Cabin number |  |
| embarked | Port of Embarkation | C = Cherbourg, Q = Queenstown, S = Southampton |
|  |  |  |

Data pre-processing

As we have all the data for the attributes we are going to use to fit the model. No pre-processing was required. Still, there was age and cabin with certain missing values but I didn’t use age as there were many old people who wanted their young generation to survive and hence forfeited their chance for the lifeboat and also the survival depends on fitness to the cold and not the age really.

Model (with its parameters)

Parameters for rpart

To create a model

rpart(formula, data, weights, subset, na.action = na.rpart, method,

model = FALSE, x = FALSE, y = TRUE, parms, control, cost, …)

method used class - as it’s a factor of classification based on the responses

Test data used to predict the survival

predict(object, newdata, type = c("vector", "prob", "class", "matrix"),

na.action = na.pass, ...)

**Q2) Decision Tree Regression (Energy Efficiency)**

Information about the data

X1 Relative Compactness   
X2 Surface Area   
X3 Wall Area   
X4 Roof Area   
X5 Overall Height   
X6 Orientation   
X7 Glazing Area   
X8 Glazing Area Distribution   
y1 Heating Load   
y2 Cooling Load

Data pre processing

All the data are present, missing values N/A

Model – same as Q1 it uses rpart

Method used – anova