Results of Course Project : Space Debris Detection

**Data Cleaning and Pre-Processing(Class balancing):**

* Removing irrelevant columns
* Removing NA values
* Removing blank values.
* Converting categorical columns to factor.
* Class Balancing using ROSE.

Output :

----------Before Pre-Processing----------

Null Values Per Columns :

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 39 0 14372 0 0 0 0

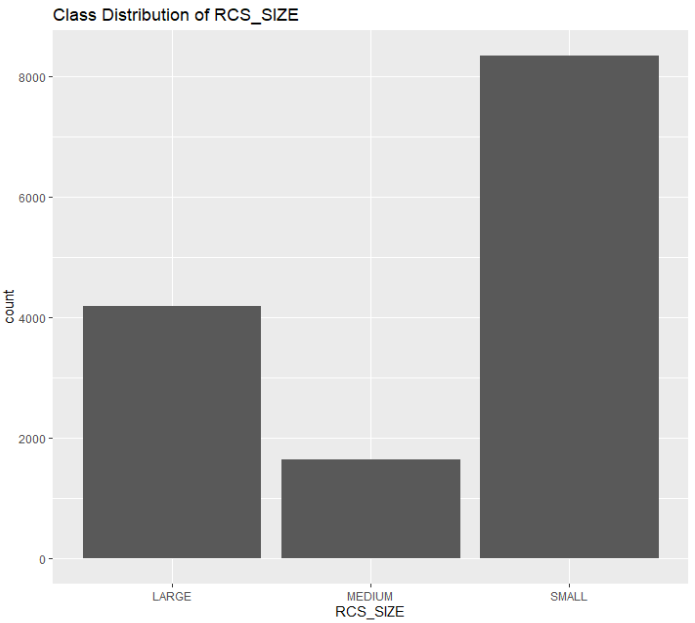
Empty Values Per Columns :

0 0 0 0 0 39 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 198 39 NA 39 NA 0 0 0 0

Number of rows in Dataset : 14372

Number of columns in Dataset : 40

Data Distribution (Blank, Large, Medium, Small) : 198 4189 1639 8346



----------After Pre-Processing----------

Null Values Per Columns :

0 0 0 0 0 0 0 0 0 0 0 0 0 0

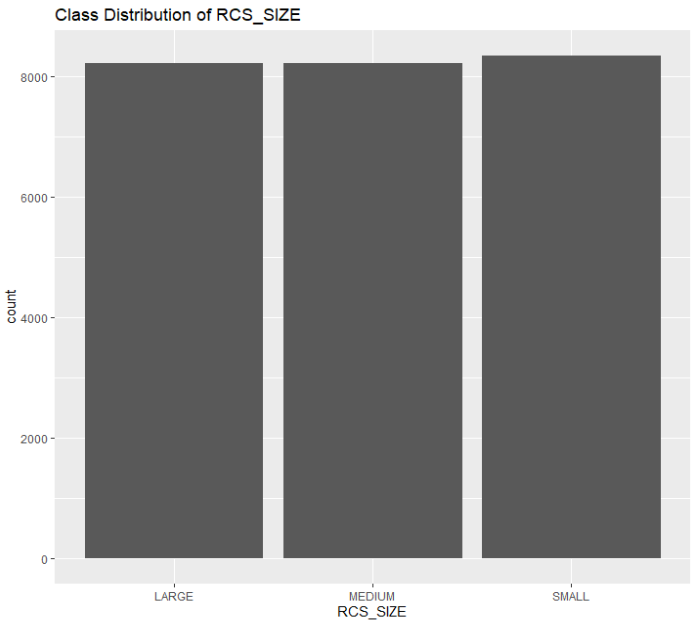
Empty Values Per Columns :

0 0 0 0 0 0 0 0 0 0 0 0 0 0

Number of rows in Dataset : 24794

Number of columns in Dataset : 14

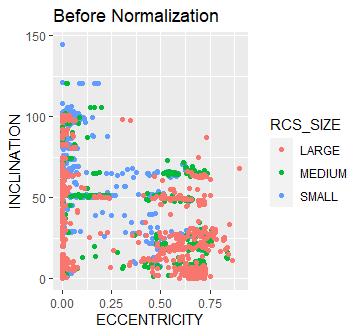
Data Distribution (Large, Medium, Small) : 8224 8224 8346

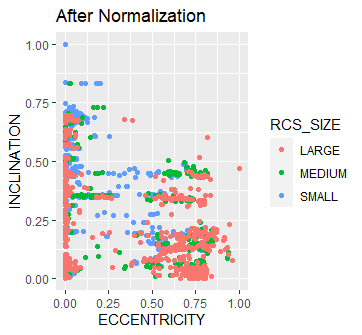


**Pre-Processing(Normalization and Outlier Removal):**

* Normalization (min-max method)
* Removing Outliers (Using Inter quartile Range)

Output:





----------Before Removing Outliers----------

Number of rows in Dataset : 24794

Number of columns in Dataset : 14

----------After Removing Outliers----------

Number of rows in Dataset : 10696

Number of columns in Dataset : 14

**Model Training:**

Gradient Boosting Machine (GBM)

[1] "Accuracy: 0.890603085553997"

[1] "Precision Large: 0.960552268244576"

[1] "Precision Medium: 0.807387862796834"

[1] "Precision Small: 0.922196796338673"

[1] "Recall Large: 0.855887521968366"

[1] "Recall Medium: 0.9"

[1] "Recall Small: 0.90561797752809"

[1] "F1 Score Large: 0.905204460966543"

[1] "F1 Score Medium: 0.851182197496523"

[1] "F1 Score Small: 0.913832199546485"

[1] "Specificity Large: 0.987261146496815"

[1] "Specificity Medium: 0.899931459904044"

[1] "Specificity Small: 0.945556445156125"

Random Forest (RF)

[1] "Accuracy: 0.977559607293128"

[1] "Precision Large: 0.991087344028521"

[1] "Precision Medium: 0.94413407821229"

[1] "Precision Small: 0.996519721577726"

[1] "Recall Large: 0.977152899824253"

[1] "Recall Medium: 0.994117647058824"

[1] "Recall Small: 0.965168539325843"

[1] "F1 Score Large: 0.984070796460177"

[1] "F1 Score Medium: 0.968481375358166"

[1] "F1 Score Small: 0.980593607305936"

[1] "Specificity Large: 0.996815286624204"

[1] "Specificity Medium: 0.972583961617546"

[1] "Specificity Small: 0.99759807846277"

Multinomial Logistic Regression (MLR)

[1] "Accuracy: 0.809724170172978"

[1] "Precision Large: 0.817184643510055"

[1] "Precision Medium: 0.705722070844687"

[1] "Precision Small: 0.893939393939394"

[1] "Recall Large: 0.785588752196837"

[1] "Recall Medium: 0.761764705882353"

[1] "Recall Small: 0.861797752808989"

[1] "F1 Score Large: 0.801075268817204"

[1] "F1 Score Medium: 0.732673267326733"

[1] "F1 Score Small: 0.877574370709382"

[1] "Specificity Large: 0.936305732484076"

[1] "Specificity Medium: 0.85195339273475"

[1] "Specificity Small: 0.927141713370697"

Naive Bayes (NB)

[1] "Accuracy: 0.711547452080411"

[1] "Precision Large: 0.640416047548291"

[1] "Precision Medium: 0.604060913705584"

[1] "Precision Small: 0.838857142857143"

[1] "Recall Large: 0.757469244288225"

[1] "Recall Medium: 0.525"

[1] "Recall Small: 0.824719101123595"

[1] "F1 Score Large: 0.694041867954911"

[1] "F1 Score Medium: 0.561762391817467"

[1] "F1 Score Small: 0.831728045325779"

[1] "Specificity Large: 0.845859872611465"

[1] "Specificity Medium: 0.839616175462646"

[1] "Specificity Small: 0.8871096877502"

Support Vector Machine (SVM)

[1] "Accuracy: 0.814866760168303"

[1] "Precision Large: 0.857142857142857"

[1] "Precision Medium: 0.700261780104712"

[1] "Precision Small: 0.891203703703704"

[1] "Recall Large: 0.769771528998243"

[1] "Recall Medium: 0.786764705882353"

[1] "Recall Small: 0.865168539325843"

[1] "F1 Score Large: 0.811111111111111"

[1] "F1 Score Medium: 0.740997229916897"

[1] "F1 Score Small: 0.877993158494869"

[1] "Specificity Large: 0.953503184713376"

[1] "Specificity Medium: 0.843043180260452"

[1] "Specificity Small: 0.924739791833467"

Decision Tree (DT)

[1] "Accuracy: 0.770920991117345"

[1] "Precision Large: 0.846827133479212"

[1] "Precision Medium: 0.609409190371991"

[1] "Precision Small: 0.91796875"

[1] "Recall Large: 0.680140597539543"

[1] "Recall Medium: 0.819117647058824"

[1] "Recall Small: 0.792134831460674"

[1] "F1 Score Large: 0.754385964912281"

[1] "F1 Score Medium: 0.698870765370138"

[1] "F1 Score Small: 0.850422195416164"

[1] "Specificity Large: 0.955414012738854"

[1] "Specificity Medium: 0.7553118574366"

[1] "Specificity Small: 0.949559647718175"

**Majority Ensembling (Voting Based):**

[1] “Accuracy: 0.8503974”

[1] "Precision Large: 0.887429643527205"

[1] "Precision Medium: 0.762845849802372"

[1] "Precision Small: 0.92969696969697"

[1] "Recall Large: 0.83128295254833"

[1] "Recall Medium: 0.851470588235294"

[1] "Recall Small: 0.861797752808989"

[1] "F1 Score Large: 0.858439201451906"

[1] "F1 Score Medium: 0.804725503822099"

[1] "F1 Score Small: 0.894460641399417"

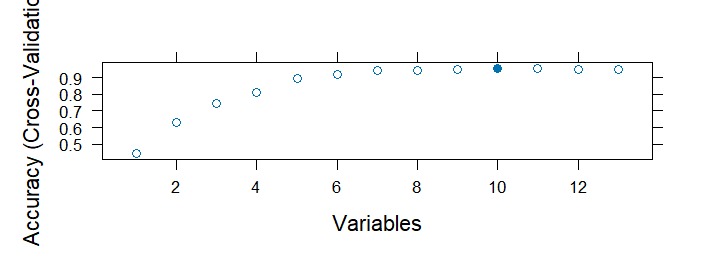
[1] "Specificity Large: 0.961783439490446"

[1] "Specificity Medium: 0.861549006168609"

[1] "Specificity Small: 0.953562850280224"

**Feature Selection:**

Recursive Features Elimination (RFE):



Resampling performance over subset size:

Variables Accuracy Kappa AccuracySD KappaSD

1 0.4422 0.08971 0.014637 0.024974

2 0.6290 0.42505 0.063307 0.108686

3 0.7437 0.60432 0.074823 0.122065

4 0.8149 0.71681 0.014029 0.021664

5 0.8995 0.84661 0.010691 0.016357

6 0.9203 0.87855 0.007684 0.011760

7 0.9441 0.91478 0.006900 0.010541

8 0.9462 0.91807 0.006031 0.009205

9 0.9535 0.92920 0.006342 0.009663

10 0.9562 0.93335 0.005744 0.008751

11 0.9551 0.93164 0.006987 0.010649

12 0.9540 0.92994 0.007081 0.010786

13 0.9529 0.92823 0.006891 0.010511

Mutual Information:

[1] "OBJECT\_TYPE" "OBJECT\_AGE" "INCLINATION" "CENT\_FOCUS\_DIST"

[5] "ECCENTRICITY" "PERIOD" "SEMIMAJOR\_AXIS" "MEAN\_MOTION"

[9] "APOAPSIS" "PERIAPSIS" "MEAN\_ANOMALY" "ARG\_OF\_PERICENTER"

[13] "RA\_OF\_ASC\_NODE"

**Model Training After Feature Selection (with 10 columns(X variables)):**

Gradient Boosting Machine (GBM)

[1] "Accuracy: 0.886395511921459"

[1] "Precision Large: 0.968253968253968"

[1] "Precision Medium: 0.802110817941952"

[1] "Precision Small: 0.912200684150513"

[1] "Recall Large: 0.857644991212654"

[1] "Recall Medium: 0.894117647058824"

[1] "Recall Small: 0.898876404494382"

[1] "F1 Score Large: 0.909599254426841"

[1] "F1 Score Medium: 0.845618915159944"

[1] "F1 Score Small: 0.905489530277306"

[1] "Specificity Large: 0.989808917197452"

[1] "Specificity Medium: 0.897189856065798"

[1] "Specificity Small: 0.938350680544436"

Random Forest (RF)

[1] "Accuracy: 0.976157082748948"

[1] "Precision Large: 0.991071428571429"

[1] "Precision Medium: 0.939058171745152"

[1] "Precision Small: 0.997666277712952"

[1] "Recall Large: 0.975395430579965"

[1] "Recall Medium: 0.997058823529412"

[1] "Recall Small: 0.960674157303371"

[1] "F1 Score Large: 0.983170947741364"

[1] "F1 Score Medium: 0.967189728958631"

[1] "F1 Score Small: 0.978820835718374"

[1] "Specificity Large: 0.996815286624204"

[1] "Specificity Medium: 0.969842357779301"

[1] "Specificity Small: 0.99839871897518"

Multinomial Logistic Regression (MLR)

[1] "Accuracy: 0.80551659654044"

[1] "Precision Large: 0.813528336380256"

[1] "Precision Medium: 0.701369863013699"

[1] "Precision Small: 0.888631090487239"

[1] "Recall Large: 0.78207381370826"

[1] "Recall Medium: 0.752941176470588"

[1] "Recall Small: 0.860674157303371"

[1] "F1 Score Large: 0.797491039426523"

[1] "F1 Score Medium: 0.726241134751773"

[1] "F1 Score Small: 0.874429223744292"

[1] "Specificity Large: 0.935031847133758"

[1] "Specificity Medium: 0.850582590815627"

[1] "Specificity Small: 0.923138510808647"

Naive Bayes (NB)

[1] "Accuracy: 0.683496961196821"

[1] "Precision Large: 0.558638083228247"

[1] "Precision Medium: 0.605150214592275"

[1] "Precision Small: 0.8375"

[1] "Recall Large: 0.778558875219684"

[1] "Recall Medium: 0.414705882352941"

[1] "Recall Small: 0.828089887640449"

[1] "F1 Score Large: 0.650513950073421"

[1] "F1 Score Medium: 0.492146596858639"

[1] "F1 Score Small: 0.832768361581921"

[1] "Specificity Large: 0.777070063694268"

[1] "Specificity Medium: 0.873886223440713"

[1] "Specificity Small: 0.88550840672538"

Support Vector Machine (SVM)

[1] "Accuracy: 0.808789153810192"

[1] "Precision Large: 0.823747680890538"

[1] "Precision Medium: 0.700542005420054"

[1] "Precision Small: 0.892111368909513"

[1] "Recall Large: 0.780316344463972"

[1] "Recall Medium: 0.760294117647059"

[1] "Recall Small: 0.864044943820225"

[1] "F1 Score Large: 0.8014440433213"

[1] "F1 Score Medium: 0.729196050775741"

[1] "F1 Score Small: 0.877853881278539"

[1] "Specificity Large: 0.939490445859873"

[1] "Specificity Medium: 0.848526387936943"

[1] "Specificity Small: 0.925540432345877"

Decision Tree (DT)

[1] "Accuracy: 0.770920991117345"

[1] "Precision Large: 0.846827133479212"

[1] "Precision Medium: 0.609409190371991"

[1] "Precision Small: 0.91796875"

[1] "Recall Large: 0.680140597539543"

[1] "Recall Medium: 0.819117647058824"

[1] "Recall Small: 0.792134831460674"

[1] "F1 Score Large: 0.754385964912281"

[1] "F1 Score Medium: 0.698870765370138"

[1] "F1 Score Small: 0.850422195416164"

[1] "Specificity Large: 0.955414012738854"

[1] "Specificity Medium: 0.7553118574366"

[1] "Specificity Small: 0.949559647718175"

**Majority Based Ensembling (Voting Based):**

[1] “Accuracy: 0.8401122"

[1] "Precision Large: 0.840070298769772"

[1] "Precision Medium: 0.741633199464525"

[1] "Precision Small: 0.926150121065375"

[1] "Recall Large: 0.840070298769772"

[1] "Recall Medium: 0.814705882352941"

[1] "Recall Small: 0.859550561797753"

[1] "F1 Score Large: 0.840070298769772"

[1] "F1 Score Medium: 0.77645409950946"

[1] "F1 Score Small: 0.891608391608392"

[1] "Specificity Large: 0.94203821656051"

[1] "Specificity Medium: 0.869773817683345"

[1] "Specificity Small: 0.951160928742994"