MULTI LINEAR REGRESSION

SUMMARY OF DATASET

Date RentedBikeCount Hour Temperature Humidity WindSpeed

01-01-18: 24 Min. : 0.0 Min. : 0.00 Min. :-17.80 Min. : 0.00 Min. :0.000

01-02-18: 24 1st Qu.: 191.0 1st Qu.: 5.75 1st Qu.: 3.50 1st Qu.:42.00 1st Qu.:0.900

01-03-18: 24 Median : 504.5 Median :11.50 Median : 13.70 Median :57.00 Median :1.500

01-04-18: 24 Mean : 704.6 Mean :11.50 Mean : 12.88 Mean :58.23 Mean :1.725

01-05-18: 24 3rd Qu.:1065.2 3rd Qu.:17.25 3rd Qu.: 22.50 3rd Qu.:74.00 3rd Qu.:2.300

01-06-18: 24 Max. :3556.0 Max. :23.00 Max. : 39.40 Max. :98.00 Max. :7.400

(Other) :8616

Visibility DewPointTemperature SolarRadiation Rainfall Snowfall

Min. : 27 Min. :-30.600 Min. :0.0000 Min. : 0.0000 Min. :0.00000

1st Qu.: 940 1st Qu.: -4.700 1st Qu.:0.0000 1st Qu.: 0.0000 1st Qu.:0.00000

Median :1698 Median : 5.100 Median :0.0100 Median : 0.0000 Median :0.00000

Mean :1437 Mean : 4.074 Mean :0.5691 Mean : 0.1487 Mean :0.07507

3rd Qu.:2000 3rd Qu.: 14.800 3rd Qu.:0.9300 3rd Qu.: 0.0000 3rd Qu.:0.00000

Max. :2000 Max. : 27.200 Max. :3.5200 Max. :35.0000 Max. :8.80000

Seasons Holiday FunctioningDay

Autumn:2184 No :8328 No : 295

Spring:2208 Yes: 432 Yes:8465

Summer:2208

Winter:2160

STRUCTURE OF THE DATASET

> str(SeoulBikeData)

'data.frame': 8760 obs. of 14 variables:

$ Date : Factor w/ 365 levels "01-01-18","01-02-18",..: 12 12 12 12 12 12 12 12 12 12 ...

$ RentedBikeCount : int 254 204 173 107 78 100 181 460 930 490 ...

$ Hour : int 0 1 2 3 4 5 6 7 8 9 ...

$ Temperature : num -5.2 -5.5 -6 -6.2 -6 -6.4 -6.6 -7.4 -7.6 -6.5 ...

$ Humidity : int 37 38 39 40 36 37 35 38 37 27 ...

$ WindSpeed : num 2.2 0.8 1 0.9 2.3 1.5 1.3 0.9 1.1 0.5 ...

$ Visibility : int 2000 2000 2000 2000 2000 2000 2000 2000 2000 1928 ...

$ DewPointTemperature: num -17.6 -17.6 -17.7 -17.6 -18.6 -18.7 -19.5 -19.3 -19.8 -22.4 ...

$ SolarRadiation : num 0 0 0 0 0 0 0 0 0.01 0.23 ...

$ Rainfall : num 0 0 0 0 0 0 0 0 0 0 ...

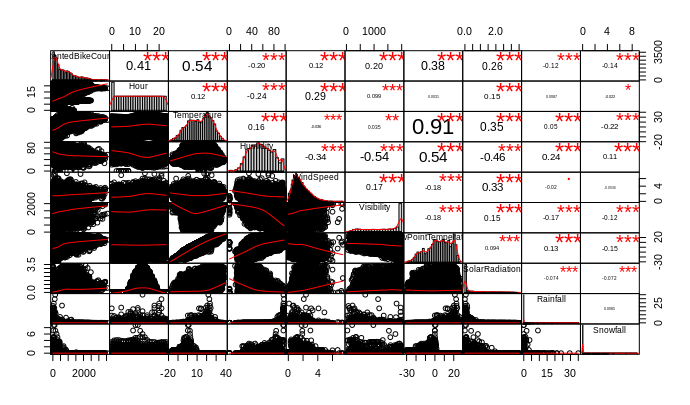
$ Snowfall : num 0 0 0 0 0 0 0 0 0 0 ...

$ Seasons : Factor w/ 4 levels "Autumn","Spring",..: 4 4 4 4 4 4 4 4 4 4 ...

$ Holiday : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...

$ FunctioningDay : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...

CHART CORRELATION WITH 10 ATTRIBUTES



SUMMARY MULTI LINEAR REGRESSION 10 ATTRIBUTES

Summary(model\_mlr10)

Call:

lm(formula = RentedBikeCount ~ Hour + Temperature + Humidity +

WindSpeed + Visibility + DewPointTemperature + SolarRadiation +

Rainfall + Snowfall, data = train10)

Residuals:

Min 1Q Median 3Q Max

-1414.1 -283.1 -38.9 225.3 2278.9

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 521.37237 117.63037 4.432 9.49e-06 \*\*\*

Hour 27.88231 0.94401 29.536 < 2e-16 \*\*\*

Temperature 26.19671 4.60862 5.684 1.37e-08 \*\*\*

Humidity -8.72315 1.30911 -6.663 2.91e-11 \*\*\*

WindSpeed 9.00089 6.56316 1.371 0.17029

Visibility 0.03328 0.01228 2.709 0.00676 \*\*

DewPointTemperature 6.39858 4.85428 1.318 0.18751

SolarRadiation -81.55466 9.80307 -8.319 < 2e-16 \*\*\*

Rainfall -58.76048 5.56980 -10.550 < 2e-16 \*\*\*

Snowfall 28.42363 13.87876 2.048 0.04060 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 469.6 on 6122 degrees of freedom

Multiple R-squared: 0.4777, Adjusted R-squared: 0.4769

F-statistic: 622.1 on 9 and 6122 DF, p-value: < 2.2e-16

ANOVA LINEAR MODEL WITH 10 ATTRIBUTES

anova(model\_mlr10)

Analysis of Variance Table

Response: RentedBikeCount

Df Sum Sq Mean Sq F value Pr(>F)

Hour 1 446339306 446339306 2024.0497 < 2.2e-16 \*\*\*

Temperature 1 635280730 635280730 2880.8571 < 2.2e-16 \*\*\*

Humidity 1 101247323 101247323 459.1341 < 2.2e-16 \*\*\*

WindSpeed 1 580836 580836 2.6340 0.1047

Visibility 1 6826510 6826510 30.9567 2.750e-08 \*\*\*

DewPointTemperature 1 4098470 4098470 18.5857 1.650e-05 \*\*\*

SolarRadiation 1 14501426 14501426 65.7608 6.099e-16 \*\*\*

Rainfall 1 24816426 24816426 112.5370 < 2.2e-16 \*\*\*

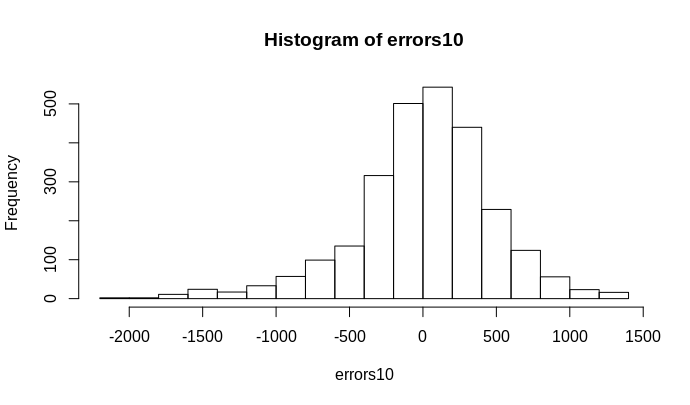
Snowfall 1 924915 924915 4.1943 0.0406 \*

Residuals 6122 1350010928 220518

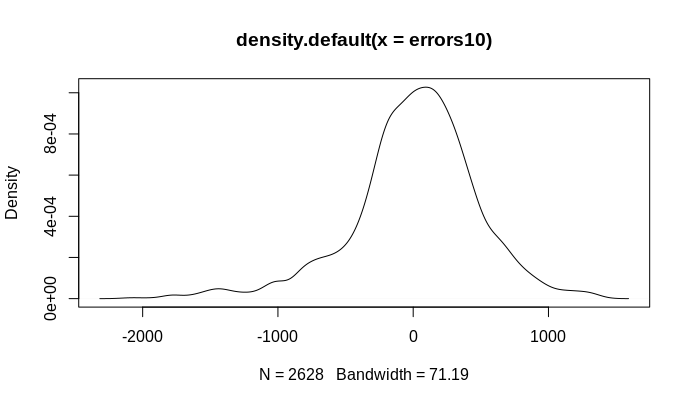
---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

HISTOGRAM OF ERRORS 10 ATTRIBUTES



DENSITY OF ERRORS 10 ATTRIBUTES



ROOT MEAN SQUARED ERRORS 10 ATTRIBUTES

rmse10 = 469.942

FORWARD SELECTION 10 ATTRIBUTES

|  |
| --- |
| summary(stepf10)  Call:  lm(formula = RentedBikeCount ~ Temperature + Hour + Humidity +  Rainfall + SolarRadiation + Visibility + Snowfall, data = bike10at)  Residuals:  Min 1Q Median 3Q Max  -1397.9 -284.7 -39.4 222.5 2304.4  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 435.94835 35.99450 12.112 <2e-16 \*\*\*  Temperature 31.57131 0.52947 59.628 <2e-16 \*\*\*  Hour 27.50284 0.76429 35.985 <2e-16 \*\*\*  Humidity -7.47595 0.39161 -19.091 <2e-16 \*\*\*  Rainfall -59.24499 4.59693 -12.888 <2e-16 \*\*\*  SolarRadiation -79.21727 7.74988 -10.222 <2e-16 \*\*\*  Visibility 0.02319 0.01027 2.258 0.024 \*  Snowfall 20.09109 11.99636 1.675 0.094 .  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 469.6 on 8752 degrees of freedom  Multiple R-squared: 0.4704, Adjusted R-squared: 0.47  F-statistic: 1110 on 7 and 8752 DF, p-value: < 2.2e-16  SUMMARY BACKWARD ELIMINATION 10 ATTRIBUTES  summary(stepb10)  Call:  lm(formula = RentedBikeCount ~ Hour + Temperature + Humidity +  Visibility + SolarRadiation + Rainfall + Snowfall, data = bike10at)  Residuals:  Min 1Q Median 3Q Max  -1397.9 -284.7 -39.4 222.5 2304.4  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 435.94835 35.99450 12.112 <2e-16 \*\*\*  Hour 27.50284 0.76429 35.985 <2e-16 \*\*\*  Temperature 31.57131 0.52947 59.628 <2e-16 \*\*\*  Humidity -7.47595 0.39161 -19.091 <2e-16 \*\*\*  Visibility 0.02319 0.01027 2.258 0.024 \*  SolarRadiation -79.21727 7.74988 -10.222 <2e-16 \*\*\*  Rainfall -59.24499 4.59693 -12.888 <2e-16 \*\*\*  Snowfall 20.09109 11.99636 1.675 0.094 .  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 469.6 on 8752 degrees of freedom  Multiple R-squared: 0.4704, Adjusted R-squared: 0.47  F-statistic: 1110 on 7 and 8752 DF, p-value: < 2.2e-16  CHECKING THE BEST COMBINATION BETWEEN 10 ATTRIBUTES  as.data.frame(sub.sum10$outmat)  Hour Temperature Humidity WindSpeed Visibility DewPointTemperature SolarRadiation Rainfall  1 ( 1 ) \*  2 ( 1 ) \* \*  3 ( 1 ) \* \* \*  4 ( 1 ) \* \* \* \*  5 ( 1 ) \* \* \* \* \*  6 ( 1 ) \* \* \* \* \* \*  7 ( 1 ) \* \* \* \* \* \*  8 ( 1 ) \* \* \* \* \* \* \*  Snowfall  1 ( 1 )  2 ( 1 )  3 ( 1 )  4 ( 1 )  5 ( 1 )  6 ( 1 )  7 ( 1 ) \*  8 ( 1 ) \*  CHART CORRELATION 13 ATTRIBUTES    SUMMARY MULTI LINEAR REGRESSION 13 ATTRIBUTES  summary(model\_mlr13)  Call:  lm(formula = RentedBikeCount ~ Hour + Temperature + Humidity +  WindSpeed + Visibility + DewPointTemperature + SolarRadiation +  Rainfall + Snowfall + Holiday + FunctioningDay + Spring +  Summer + Autumn + Winter, data = train13)  Residuals:  Min 1Q Median 3Q Max  -1381.98 -279.66 -37.17 219.83 2279.02  Coefficients: (2 not defined because of singularities)  Estimate Std. Error t value Pr(>|t|)  (Intercept) 5.163e+02 1.155e+02 4.470 7.95e-06 \*\*\*  Hour 2.806e+01 9.346e-01 30.026 < 2e-16 \*\*\*  Temperature 1.826e+01 4.568e+00 3.998 6.47e-05 \*\*\*  Humidity -9.959e+00 1.287e+00 -7.740 1.15e-14 \*\*\*  WindSpeed 9.891e+00 6.512e+00 1.519 0.12884  Visibility 9.655e-03 1.265e-02 0.763 0.44522  DewPointTemperature 6.562e+00 4.774e+00 1.375 0.16932  SolarRadiation -8.032e+01 9.684e+00 -8.295 < 2e-16 \*\*\*  Rainfall -5.773e+01 5.305e+00 -10.882 < 2e-16 \*\*\*  Snowfall 4.008e+01 1.416e+01 2.829 0.00468 \*\*  Holiday -1.223e+02 2.772e+01 -4.412 1.04e-05 \*\*\*  FunctioningDay NA NA NA NA  Spring 2.518e+02 2.368e+01 10.633 < 2e-16 \*\*\*  Summer 2.884e+02 3.578e+01 8.060 9.13e-16 \*\*\*  Autumn 2.907e+02 2.491e+01 11.669 < 2e-16 \*\*\*  Winter NA NA NA NA  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 463.5 on 6118 degrees of freedom  Multiple R-squared: 0.486, Adjusted R-squared: 0.4849  F-statistic: 445 on 13 and 6118 DF, p-value: < 2.2e-16  ANOVA 13 ATTRIBUTES  anova(model\_mlr13)  Analysis of Variance Table  Response: RentedBikeCount  Df Sum Sq Mean Sq F value Pr(>F)  Hour 1 423881125 423881125 1973.2977 < 2.2e-16 \*\*\*  Temperature 1 623854326 623854326 2904.2348 < 2.2e-16 \*\*\*  Humidity 1 103495933 103495933 481.8056 < 2.2e-16 \*\*\*  WindSpeed 1 1071957 1071957 4.9903 0.0255259 \*  Visibility 1 4731881 4731881 22.0284 2.745e-06 \*\*\*  DewPointTemperature 1 3210897 3210897 14.9477 0.0001117 \*\*\*  SolarRadiation 1 15234076 15234076 70.9193 < 2.2e-16 \*\*\*  Rainfall 1 25607688 25607688 119.2117 < 2.2e-16 \*\*\*  Snowfall 1 495316 495316 2.3058 0.1289395  Holiday 1 6391654 6391654 29.7551 5.094e-08 \*\*\*  Spring 1 4241288 4241288 19.7445 9.008e-06 \*\*\*  Summer 1 1122412 1122412 5.2252 0.0222962 \*  Autumn 1 29250451 29250451 136.1699 < 2.2e-16 \*\*\*  Residuals 6118 1314198417 214809  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Warning message:  In predict.lm(model\_mlr13, interval = "prediction", newdata = test13) :  prediction from a rank-deficient fit may be misleading  HISTOGRAM OF ERRORS WITH 13 ATTRIBUTES |
|  |
| DENSITY OF ERRORS 13 ATTRIBUTES    ROOT MEAN SQUARED ERRORS 13 ATTRIBUTES  rmse13 = 459.02  FORWARD SELECTION 13 ATTRIBUTES  summary(stepf13)  Call:  lm(formula = RentedBikeCount ~ Temperature + Hour + Humidity +  Winter + Rainfall + SolarRadiation + Holiday + Spring + Snowfall +  WindSpeed + DewPointTemperature, data = bike13at)  Residuals:  Min 1Q Median 3Q Max  -1414.68 -276.65 -41.97 220.90 2269.76  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 845.754 93.496 9.046 < 2e-16 \*\*\*  Temperature 17.304 3.897 4.441 9.08e-06 \*\*\*  Hour 28.189 0.776 36.324 < 2e-16 \*\*\*  Humidity -10.234 1.061 -9.644 < 2e-16 \*\*\*  Winter -303.263 20.248 -14.977 < 2e-16 \*\*\*  Rainfall -58.698 4.555 -12.886 < 2e-16 \*\*\*  SolarRadiation -77.525 7.920 -9.789 < 2e-16 \*\*\*  Holiday -130.143 22.959 -5.668 1.49e-08 \*\*\*  Spring -49.241 13.365 -3.684 0.000231 \*\*\*  Snowfall 38.757 11.958 3.241 0.001195 \*\*  WindSpeed 13.218 5.402 2.447 0.014423 \*  DewPointTemperature 6.964 4.070 1.711 0.087119 .  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 462.1 on 8748 degrees of freedom  Multiple R-squared: 0.4875, Adjusted R-squared: 0.4868  F-statistic: 756.4 on 11 and 8748 DF, p-value: < 2.2e-16  BACKWARD ELIMINATION 13 ATTRIBUTES  summary(stepb13)  Call:  lm(formula = RentedBikeCount ~ Hour + Temperature + Humidity +  WindSpeed + DewPointTemperature + SolarRadiation + Rainfall +  Snowfall + Holiday + Spring + Summer + Autumn, data = bike13at)  Residuals:  Min 1Q Median 3Q Max  -1426.57 -276.29 -42.49 221.53 2277.29  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 552.4411 93.5909 5.903 3.71e-09 \*\*\*  Hour 28.0793 0.7823 35.893 < 2e-16 \*\*\*  Temperature 17.5497 3.9032 4.496 7.00e-06 \*\*\*  Humidity -10.3056 1.0631 -9.694 < 2e-16 \*\*\*  WindSpeed 13.5875 5.4120 2.511 0.01207 \*  DewPointTemperature 7.3514 4.0854 1.799 0.07198 .  SolarRadiation -78.2127 7.9440 -9.845 < 2e-16 \*\*\*  Rainfall -58.5804 4.5562 -12.857 < 2e-16 \*\*\*  Snowfall 38.7816 11.9580 3.243 0.00119 \*\*  Holiday -132.4700 23.0554 -5.746 9.46e-09 \*\*\*  Spring 244.5871 19.8449 12.325 < 2e-16 \*\*\*  Summer 279.4335 29.5966 9.441 < 2e-16 \*\*\*  Autumn 299.1785 20.5831 14.535 < 2e-16 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 462 on 8747 degrees of freedom  Multiple R-squared: 0.4875, Adjusted R-squared: 0.4868  F-statistic: 693.5 on 12 and 8747 DF, p-value: < 2.2e-16  CHECKING THE BEST COMBINATION BETWEEN 13 ATTRIBUTES Warning message:  In leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax, force.in = force.in, :  2 linear dependencies found  as.data.frame(sub.sum13$outmat)  Hour Temperature Humidity WindSpeed Visibility DewPointTemperature SolarRadiation Rainfall  1 ( 1 ) \*  2 ( 1 ) \* \*  3 ( 1 ) \* \* \*  4 ( 1 ) \* \* \*  5 ( 1 ) \* \* \* \*  6 ( 1 ) \* \* \* \* \*  7 ( 1 ) \* \* \* \* \*  8 ( 1 ) \* \* \* \* \*  9 ( 1 ) \* \* \* \* \*  Snowfall Holiday FunctioningDay Spring Summer Autumn Winter  1 ( 1 )  2 ( 1 )  3 ( 1 )  4 ( 1 ) \*  5 ( 1 ) \*  6 ( 1 ) \*  7 ( 1 ) \* \*  8 ( 1 ) \* \* \*  9 ( 1 ) \* \* \* \*  MACHINE LEARNING  CONFUSION MATRIX 11 ATTRIBUTES   |  | | --- | | cm11  Predicted  Actual 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15  1 689 102 3 17 2 0 0 0 0 0 0 0 0 0 0  2 222 244 12 10 0 0 0 0 0 0 0 0 0 0 0  3 53 128 67 39 12 1 0 0 0 0 0 0 0 0 0  4 12 41 70 61 51 18 0 0 0 0 0 0 0 0 0  5 4 12 30 70 110 31 5 3 0 0 0 0 0 0 0  6 0 4 12 13 48 39 17 10 0 0 0 0 0 0 0  7 0 1 4 3 27 37 21 18 0 0 0 0 0 0 0  8 0 0 1 2 11 17 12 19 13 0 0 0 0 0 0  9 0 0 0 3 4 7 8 10 27 9 0 0 0 0 0  10 0 0 0 0 0 7 2 3 26 23 0 0 0 0 0  11 0 0 0 0 0 1 2 0 10 13 0 0 0 0 0  12 0 0 0 0 0 0 0 0 1 9 0 0 0 0 0  13 0 0 0 0 0 0 0 0 1 6 0 0 0 0 0  14 0 0 0 0 0 0 0 0 0 6 0 0 0 0 0  15 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 | |  | | ACCURACY OF THE MODEL FOR 11 ATTRIBUTES = 49.467%  CONFUSION MATRIX 14 ATTRIBUTES   |  | | --- | | cm14  Predicted  Actual 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15  1 685 109 1 1 1 4 0 0 0 0 0 0 0 0 0  2 202 269 32 6 2 0 0 0 0 0 0 0 0 0 0  3 52 82 96 31 18 6 0 0 0 0 0 0 0 0 0  4 14 35 74 58 82 18 1 0 0 0 0 0 0 0 0  5 18 6 23 43 105 23 4 0 0 0 0 0 0 0 0  6 8 4 10 28 48 35 8 0 4 0 0 0 0 0 0  7 1 2 12 10 43 41 9 10 7 1 0 0 0 0 0  8 1 1 4 3 20 14 9 14 16 2 0 0 0 0 0  9 1 1 0 2 18 12 7 12 19 6 0 0 0 0 0  10 0 0 0 1 6 4 2 4 12 17 0 0 0 0 0  11 0 0 0 0 2 1 1 4 2 11 0 0 0 0 0  12 0 0 0 0 0 0 0 2 2 2 0 0 0 0 0  13 0 0 0 0 0 0 0 0 0 4 0 0 0 0 0  14 0 0 0 0 0 0 0 0 3 3 0 0 0 0 0  15 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 | |  | | ACCURACY OF THE MODEL FOR 14 ATTRIBUTES = 49.467%   |  | | --- | |  | |  |  | | --- | |  | |  |  | | --- | |  | |