-:Assignment Data Mining:-

CART AND CHAID ANALYSIS

**Decision tree on the basis of CART algorithm**

**Below are the results**

n= 2940

node), split, n, loss, yval, (yprob)

\* denotes terminal node

1) root 2940 474 No (0.83877551 0.16122449)

2) OverTime=No 2108 220 No (0.89563567 0.10436433)

4) TotalWorkingYears\_bkt=2,3,4,5,6,7 1982 174 No (0.91220989 0.08779011) \*

5) TotalWorkingYears\_bkt=1 126 46 No (0.63492063 0.36507937)

10) BusinessTravel=Non-Travel,Travel\_Rarely 110 32 No (0.70909091 0.29090909)

20) MaritalStatus=Divorced,Married 54 6 No (0.88888889 0.11111111) \*

21) MaritalStatus=Single 56 26 No (0.53571429 0.46428571)

42) JobRole=Research Scientist 24 4 No (0.83333333 0.16666667) \*

43) JobRole=Human Resources,Laboratory Technician,Sales Representative 32 10 Yes (0.31250000 0.68750000) \*

11) BusinessTravel=Travel\_Frequently 16 2 Yes (0.12500000 0.87500000) \*

3) OverTime=Yes 832 254 No (0.69471154 0.30528846)

6) MonthlyIncome\_bkt=2,3,4,5,6 692 158 No (0.77167630 0.22832370)

12) JobRole=Healthcare Representative,Human Resources,Manager,Manufacturing Director,Research Director,Research Scientist 398 50 No (0.87437186 0.12562814)

24) JobLevel>=1.5 296 24 No (0.91891892 0.08108108) \*

25) JobLevel< 1.5 102 26 No (0.74509804 0.25490196)

50) DailyRate\_bkt=1,2,3,4,6 88 14 No (0.84090909 0.15909091) \*

51) DailyRate\_bkt=5 14 2 Yes (0.14285714 0.85714286) \*

13) JobRole=Laboratory Technician,Sales Executive,Sales Representative 294 108 No (0.63265306 0.36734694)

26) StockOptionLevel>=0.5 170 34 No (0.80000000 0.20000000) \*

27) StockOptionLevel< 0.5 124 50 Yes (0.40322581 0.59677419)

54) MonthlyIncome\_bkt=3 70 28 No (0.60000000 0.40000000)

108) BusinessTravel=Non-Travel,Travel\_Rarely 58 18 No (0.68965517 0.31034483)

216) NumCompaniesWorked\_bkt=1,2 44 8 No (0.81818182 0.18181818) \*

217) NumCompaniesWorked\_bkt=3,5 14 4 Yes (0.28571429 0.71428571) \*

109) BusinessTravel=Travel\_Frequently 12 2 Yes (0.16666667 0.83333333) \*

55) MonthlyIncome\_bkt=2,4,5 54 8 Yes (0.14814815 0.85185185) \*

7) MonthlyIncome\_bkt=1 140 44 Yes (0.31428571 0.68571429)

14) DailyRate\_bkt=5,6 62 28 No (0.54838710 0.45161290)

28) MonthlyRate\_bkt=3,4,5 42 12 No (0.71428571 0.28571429) \*

29) MonthlyRate\_bkt=1,2 20 4 Yes (0.20000000 0.80000000) \*

15) DailyRate\_bkt=1,2,3,4 78 10 Yes (0.12820513 0.87179487) \*

As is clear from the results, OVERTIME is the most significant variable

**Confusion matrix PREDICTED CLASS**

|  |  |  |
| --- | --- | --- |
| **ACTUAL CLASS** | **YES** | **NO** |
| **YES** | 208(TRUE “YES”) | 266(FALSE ”NO”) |
| **NO** | 32(FALSE “YES”) | 2434(TRUE “NO”) |

Precision: 208/240

Recall: 208/474

**CHAID TREE**

Summary:

Model formula:

Attrition ~ Age\_bkt + BusinessTravel + DailyRate\_bkt + Department +

DistanceFromHome\_bkt + Education + EducationField + EnvironmentSatisfaction +

Gender + HourlyRate\_bkt + JobInvolvement + JobLevel + JobRole +

JobSatisfaction + MaritalStatus + MonthlyIncome\_bkt + MonthlyRate\_bkt +

NumCompaniesWorked\_bkt + OverTime + PercentSalaryHike\_bkt +

PerformanceRating + RelationshipSatisfaction + StockOptionLevel +

TotalWorkingYears\_bkt + TrainingTimesLastYear + WorkLifeBalance +

YearsAtCompany\_bkt + YearsInCurrentRole\_bkt + YearsSinceLastPromotion\_bkt +

YearsWithCurrManager\_bkt

Fitted party:

[1] root

| [2] OverTime in No

| | [3] TotalWorkingYears\_bkt in 1

| | | [4] BusinessTravel in Non-Travel: No (n = 22, err = 9.1%)

| | | [5] BusinessTravel in Travel\_Frequently: Yes (n = 16, err = 12.5%)

| | | [6] BusinessTravel in Travel\_Rarely

| | | | [7] RelationshipSatisfaction in 1, 3

| | | | | [8] WorkLifeBalance in 1, 3

| | | | | | [9] DailyRate\_bkt in 1, 5: Yes (n = 8, err = 25.0%)

| | | | | | [10] DailyRate\_bkt in 2, 3, 4: No (n = 18, err = 0.0%)

| | | | | | [11] DailyRate\_bkt in 6: No (n = 4, err = 50.0%)

| | | | | [12] WorkLifeBalance in 2, 4: Yes (n = 16, err = 0.0%)

| | | | [13] RelationshipSatisfaction in 2, 4

| | | | | [14] MaritalStatus in Divorced, Married: No (n = 24, err = 0.0%)

| | | | | [15] MaritalStatus in Single: No (n = 18, err = 33.3%)

| | [16] TotalWorkingYears\_bkt in 2, 4, 7

| | | [17] WorkLifeBalance in 1

| | | | [18] EnvironmentSatisfaction in 1, 4

| | | | | [19] YearsWithCurrManager\_bkt in 1, 4, 5, 6, 7: Yes (n = 18, err = 22.2%)

| | | | | [20] YearsWithCurrManager\_bkt in 2, 3: No (n = 16, err = 12.5%)

| | | | [21] EnvironmentSatisfaction in 2, 3: No (n = 24, err = 0.0%)

| | | [22] WorkLifeBalance in 2, 3, 4

| | | | [23] BusinessTravel in Non-Travel: No (n = 74, err = 0.0%)

| | | | [24] BusinessTravel in Travel\_Frequently

| | | | | [25] JobSatisfaction in 1, 2, 4

| | | | | | [26] TrainingTimesLastYear in 0, 1, 2, 5, 6: No (n = 48, err = 0.0%)

| | | | | | [27] TrainingTimesLastYear in 3, 4

| | | | | | | [28] YearsAtCompany\_bkt in 1, 5, 6, 7: Yes (n = 4, err = 0.0%)

| | | | | | | [29] YearsAtCompany\_bkt in 2, 3, 4

| | | | | | | | [30] DailyRate\_bkt in 1: Yes (n = 6, err = 33.3%)

| | | | | | | | [31] DailyRate\_bkt in 2, 3, 4, 5, 6: No (n = 26, err = 0.0%)

| | | | | [32] JobSatisfaction in 3

| | | | | | [33] Age\_bkt in 1, 2, 4

| | | | | | | [34] DailyRate\_bkt in 1, 2, 3, 4, 6: Yes (n = 16, err = 12.5%)

| | | | | | | [35] DailyRate\_bkt in 5: No (n = 4, err = 0.0%)

| | | | | | [36] Age\_bkt in 3: No (n = 8, err = 0.0%)

| | | | [37] BusinessTravel in Travel\_Rarely

| | | | | [38] Department in Human Resources: No (n = 10, err = 40.0%)

| | | | | [39] Department in Research & Development

| | | | | | [40] YearsAtCompany\_bkt in 1, 2, 4, 5, 6

| | | | | | | [41] NumCompaniesWorked\_bkt in 1, 2, 3, 5

| | | | | | | | [42] DistanceFromHome\_bkt in 1, 2, 3, 4, 6: No (n = 194, err = 0.0%)

| | | | | | | | [43] DistanceFromHome\_bkt in 5: No (n = 12, err = 16.7%)

| | | | | | | [44] NumCompaniesWorked\_bkt in 4: No (n = 14, err = 14.3%)

| | | | | | [45] YearsAtCompany\_bkt in 3

| | | | | | | [46] PerformanceRating in 3: No (n = 44, err = 4.5%)

| | | | | | | [47] PerformanceRating in 4: No (n = 8, err = 50.0%)

| | | | | | [48] YearsAtCompany\_bkt in 7: Yes (n = 6, err = 33.3%)

| | | | | [49] Department in Sales

| | | | | | [50] JobInvolvement in 1: Yes (n = 8, err = 25.0%)

| | | | | | [51] JobInvolvement in 2, 3, 4

| | | | | | | [52] NumCompaniesWorked\_bkt in 1, 2, 4

| | | | | | | | [53] JobRole in Healthcare Representative, Human Resources, Laboratory Technician, Manager, Manufacturing Director, Research Director, Research Scientist, Sales Executive: No (n = 78, err = 0.0%)

| | | | | | | | [54] JobRole in Sales Representative: No (n = 34, err = 11.8%)

| | | | | | | [55] NumCompaniesWorked\_bkt in 3, 5

| | | | | | | | [56] JobLevel in 1, 2, 4, 5

| | | | | | | | | [57] Education in 1, 2, 5: Yes (n = 6, err = 33.3%)

| | | | | | | | | [58] Education in 3, 4: No (n = 18, err = 0.0%)

| | | | | | | | [59] JobLevel in 3: Yes (n = 4, err = 0.0%)

| | [60] TotalWorkingYears\_bkt in 3, 5, 6

| | | [61] JobSatisfaction in 1

| | | | [62] StockOptionLevel in 0

| | | | | [63] YearsWithCurrManager\_bkt in 1, 5, 6, 7

| | | | | | [64] Education in 1, 3: No (n = 22, err = 0.0%)

| | | | | | [65] Education in 2, 4, 5

| | | | | | | [66] Department in Human Resources, Research & Development: Yes (n = 14, err = 42.9%)

| | | | | | | [67] Department in Sales: No (n = 16, err = 12.5%)

| | | | | [68] YearsWithCurrManager\_bkt in 2, 3

| | | | | | [69] RelationshipSatisfaction in 1, 2, 3

| | | | | | | [70] TrainingTimesLastYear in 0, 1, 2, 4, 5, 6: Yes (n = 20, err = 0.0%)

| | | | | | | [71] TrainingTimesLastYear in 3: No (n = 8, err = 25.0%)

| | | | | | [72] RelationshipSatisfaction in 4: No (n = 12, err = 16.7%)

| | | | | [73] YearsWithCurrManager\_bkt in 4: No (n = 24, err = 0.0%)

| | | | [74] StockOptionLevel in 1, 2, 3

| | | | | [75] JobRole in Healthcare Representative, Laboratory Technician, Manager, Manufacturing Director, Research Director, Research Scientist, Sales Executive, Sales Representative

| | | | | | [76] Age\_bkt in 1: No (n = 6, err = 33.3%)

| | | | | | [77] Age\_bkt in 2, 3, 4

| | | | | | | [78] TrainingTimesLastYear in 0, 1, 2, 3, 5: No (n = 136, err = 0.0%)

| | | | | | | [79] TrainingTimesLastYear in 4, 6: No (n = 14, err = 14.3%)

| | | | | [80] JobRole in Human Resources: No (n = 4, err = 50.0%)

| | | [81] JobSatisfaction in 2, 3, 4

| | | | [82] EnvironmentSatisfaction in 1

| | | | | [83] JobInvolvement in 1: Yes (n = 14, err = 42.9%)

| | | | | [84] JobInvolvement in 2, 4

| | | | | | [85] WorkLifeBalance in 1: No (n = 4, err = 50.0%)

| | | | | | [86] WorkLifeBalance in 2, 3, 4: No (n = 72, err = 0.0%)

| | | | | [87] JobInvolvement in 3

| | | | | | [88] YearsWithCurrManager\_bkt in 1, 3, 7

| | | | | | | [89] DistanceFromHome\_bkt in 1, 2, 5, 6

| | | | | | | | [90] BusinessTravel in Non-Travel, Travel\_Rarely: No (n = 24, err = 0.0%)

| | | | | | | | [91] BusinessTravel in Travel\_Frequently: Yes (n = 6, err = 33.3%)

| | | | | | | [92] DistanceFromHome\_bkt in 3, 4: Yes (n = 8, err = 0.0%)

| | | | | | [93] YearsWithCurrManager\_bkt in 2, 4, 5, 6

| | | | | | | [94] Department in Human Resources, Research & Development: No (n = 72, err = 0.0%)

| | | | | | | [95] Department in Sales

| | | | | | | | [96] MaritalStatus in Divorced, Single: No (n = 8, err = 50.0%)

| | | | | | | | [97] MaritalStatus in Married: No (n = 12, err = 0.0%)

| | | | [98] EnvironmentSatisfaction in 2, 4

| | | | | [99] EducationField in Human Resources, Life Sciences, Marketing, Medical

| | | | | | [100] WorkLifeBalance in 1: No (n = 12, err = 16.7%)

| | | | | | [101] WorkLifeBalance in 2, 3, 4

| | | | | | | [102] StockOptionLevel in 0, 1, 2

| | | | | | | | [103] RelationshipSatisfaction in 1: No (n = 66, err = 3.0%)

| | | | | | | | [104] RelationshipSatisfaction in 2, 3, 4: No (n = 326, err = 0.0%)

| | | | | | | [105] StockOptionLevel in 3

| | | | | | | | [106] MaritalStatus in Divorced, Single: No (n = 18, err = 0.0%)

| | | | | | | | [107] MaritalStatus in Married: No (n = 4, err = 50.0%)

| | | | | [108] EducationField in Other, Technical Degree

| | | | | | [109] HourlyRate\_bkt in 1, 3, 4, 5: No (n = 62, err = 0.0%)

| | | | | | [110] HourlyRate\_bkt in 2

| | | | | | | [111] YearsAtCompany\_bkt in 1, 2, 4, 5, 6, 7: Yes (n = 10, err = 20.0%)

| | | | | | | [112] YearsAtCompany\_bkt in 3: No (n = 10, err = 0.0%)

| | | | [113] EnvironmentSatisfaction in 3

| | | | | [114] Education in 1, 2, 3, 4

| | | | | | [115] TrainingTimesLastYear in 0: No (n = 4, err = 50.0%)

| | | | | | [116] TrainingTimesLastYear in 1, 2, 3

| | | | | | | [117] StockOptionLevel in 0, 2: No (n = 96, err = 12.5%)

| | | | | | | [118] StockOptionLevel in 1, 3

| | | | | | | | [119] EducationField in Human Resources, Life Sciences, Marketing, Medical, Other: No (n = 106, err = 0.0%)

| | | | | | | | [120] EducationField in Technical Degree: No (n = 8, err = 25.0%)

| | | | | | [121] TrainingTimesLastYear in 4, 5, 6: No (n = 72, err = 0.0%)

| | | | | [122] Education in 5: No (n = 12, err = 33.3%)

| [123] OverTime in Yes

| | [124] MonthlyIncome\_bkt in 1

| | | [125] DailyRate\_bkt in 1, 2: Yes (n = 36, err = 0.0%)

| | | [126] DailyRate\_bkt in 3, 4

| | | | [127] TotalWorkingYears\_bkt in 1, 3, 4, 6, 7: Yes (n = 24, err = 0.0%)

| | | | [128] TotalWorkingYears\_bkt in 2, 5: No (n = 18, err = 44.4%)

| | | [129] DailyRate\_bkt in 5, 6

| | | | [130] JobSatisfaction in 1: Yes (n = 14, err = 14.3%)

| | | | [131] JobSatisfaction in 2, 3

| | | | | [132] StockOptionLevel in 0, 2: Yes (n = 18, err = 33.3%)

| | | | | [133] StockOptionLevel in 1: No (n = 16, err = 12.5%)

| | | | | [134] StockOptionLevel in 3: Yes (n = 2, err = 0.0%)

| | | | [135] JobSatisfaction in 4: No (n = 12, err = 0.0%)

| | [136] MonthlyIncome\_bkt in 2

| | | [137] MaritalStatus in Divorced, Married

| | | | [138] BusinessTravel in Non-Travel, Travel\_Rarely

| | | | | [139] JobInvolvement in 1: Yes (n = 4, err = 0.0%)

| | | | | [140] JobInvolvement in 2

| | | | | | [141] TotalWorkingYears\_bkt in 1, 3, 4, 5, 6, 7: No (n = 12, err = 0.0%)

| | | | | | [142] TotalWorkingYears\_bkt in 2: Yes (n = 10, err = 20.0%)

| | | | | [143] JobInvolvement in 3, 4

| | | | | | [144] YearsInCurrentRole\_bkt in 1, 3, 4, 5

| | | | | | | [145] EducationField in Human Resources, Life Sciences, Marketing, Medical, Other

| | | | | | | | [146] Gender in Female: No (n = 4, err = 50.0%)

| | | | | | | | [147] Gender in Male: No (n = 16, err = 0.0%)

| | | | | | | [148] EducationField in Technical Degree: Yes (n = 4, err = 0.0%)

| | | | | | [149] YearsInCurrentRole\_bkt in 2: No (n = 36, err = 0.0%)

| | | | [150] BusinessTravel in Travel\_Frequently

| | | | | [151] TotalWorkingYears\_bkt in 1, 3, 4, 5, 6, 7: Yes (n = 18, err = 22.2%)

| | | | | [152] TotalWorkingYears\_bkt in 2: No (n = 8, err = 0.0%)

| | | [153] MaritalStatus in Single

| | | | [154] JobSatisfaction in 1, 2, 3

| | | | | [155] EnvironmentSatisfaction in 1, 2, 4: Yes (n = 22, err = 0.0%)

| | | | | [156] EnvironmentSatisfaction in 3: No (n = 8, err = 50.0%)

| | | | [157] JobSatisfaction in 4

| | | | | [158] YearsWithCurrManager\_bkt in 1, 3, 4, 6, 7: Yes (n = 6, err = 33.3%)

| | | | | [159] YearsWithCurrManager\_bkt in 2: No (n = 12, err = 0.0%)

| | | | | [160] YearsWithCurrManager\_bkt in 5: Yes (n = 2, err = 0.0%)

| | [161] MonthlyIncome\_bkt in 3, 4, 5, 6

| | | [162] Department in Human Resources, Research & Development

| | | | [163] NumCompaniesWorked\_bkt in 1

| | | | | [164] Age\_bkt in 1: Yes (n = 2, err = 0.0%)

| | | | | [165] Age\_bkt in 2, 3, 4

| | | | | | [166] YearsSinceLastPromotion\_bkt in 1, 2, 3, 6

| | | | | | | [167] PerformanceRating in 3: No (n = 82, err = 0.0%)

| | | | | | | [168] PerformanceRating in 4: No (n = 12, err = 16.7%)

| | | | | | [169] YearsSinceLastPromotion\_bkt in 4, 5

| | | | | | | [170] Education in 1, 5: Yes (n = 2, err = 0.0%)

| | | | | | | [171] Education in 2, 3: No (n = 14, err = 0.0%)

| | | | | | | [172] Education in 4: No (n = 8, err = 50.0%)

| | | | [173] NumCompaniesWorked\_bkt in 2, 3

| | | | | [174] TrainingTimesLastYear in 0, 2, 3, 4, 5, 6

| | | | | | [175] JobLevel in 1, 2, 3, 4: No (n = 124, err = 0.0%)

| | | | | | [176] JobLevel in 5: No (n = 12, err = 16.7%)

| | | | | [177] TrainingTimesLastYear in 1: No (n = 8, err = 25.0%)

| | | | [178] NumCompaniesWorked\_bkt in 4, 5

| | | | | [179] EnvironmentSatisfaction in 1, 2: Yes (n = 18, err = 33.3%)

| | | | | [180] EnvironmentSatisfaction in 3, 4

| | | | | | [181] TotalWorkingYears\_bkt in 1, 2, 3, 4, 5, 6: No (n = 34, err = 0.0%)

| | | | | | [182] TotalWorkingYears\_bkt in 7: Yes (n = 4, err = 0.0%)

| | | [183] Department in Sales

| | | | [184] MaritalStatus in Divorced, Married

| | | | | [185] YearsSinceLastPromotion\_bkt in 1, 3, 4, 5, 6

| | | | | | [186] JobSatisfaction in 1, 2, 4: No (n = 86, err = 2.3%)

| | | | | | [187] JobSatisfaction in 3

| | | | | | | [188] RelationshipSatisfaction in 1, 2: Yes (n = 14, err = 42.9%)

| | | | | | | [189] RelationshipSatisfaction in 3, 4: No (n = 22, err = 0.0%)

| | | | | [190] YearsSinceLastPromotion\_bkt in 2: Yes (n = 18, err = 33.3%)

| | | | [191] MaritalStatus in Single

| | | | | [192] YearsSinceLastPromotion\_bkt in 1, 6

| | | | | | [193] WorkLifeBalance in 1, 3

| | | | | | | [194] JobInvolvement in 1, 2, 4: Yes (n = 10, err = 20.0%)

| | | | | | | [195] JobInvolvement in 3

| | | | | | | | [196] MonthlyRate\_bkt in 1, 2, 3, 5: No (n = 20, err = 0.0%)

| | | | | | | | [197] MonthlyRate\_bkt in 4: Yes (n = 8, err = 25.0%)

| | | | | | [198] WorkLifeBalance in 2: Yes (n = 8, err = 0.0%)

| | | | | | [199] WorkLifeBalance in 4: No (n = 4, err = 0.0%)

| | | | | [200] YearsSinceLastPromotion\_bkt in 2, 3, 4, 5: Yes (n = 20, err = 0.0%)

Number of inner nodes: 90

Number of terminal nodes: 110

Result Analysis

Overtime is deciding factor on topmost node in CHAID tree as well.