University of Colombo School of Computing

Data Analytics
IS4003
Assignment 2

Registration Number: 2015/IS/038

Index Number: 15020381

Table of Contents

I have performed association rule mining using Apriori algorithm for IBM Employee Attrition Dataset available in Kaggle data repository.

1. Objectives

There is an immense impact on employee attrition on the functioning of an organization. Employee attrition has an immense impact on the functioning of the organization. Both voluntary as well as the involuntary employee attrition, is a major loss to the organization, affecting the company's productivity, causing delays in project deadlines, increase in poor services, eventually leading to disappointed clients. Looking for right replacements, organizing interviews and hiring & training involve investment of considerable amount of time and money. Attrition also may lead to a" Snowball Effect", influencing the other employees to leave as well. The losses and issues generated due to employee attrition, makes it very important to identify attributes that leads to attrition of employees.

The objective of association rule mining for IBM Employee Attrition Dataset is to identify meaningful and interesting associations between different attributes such as age, gender, education level, job satisfaction etc. and employee attrition, so that it will eventually help the organizations to understand their employee's behaviors and hence take actions to retain them. It would assist in uncovering the factors that lead to employee attrition and reaping benefits in the employee attrition control domain.

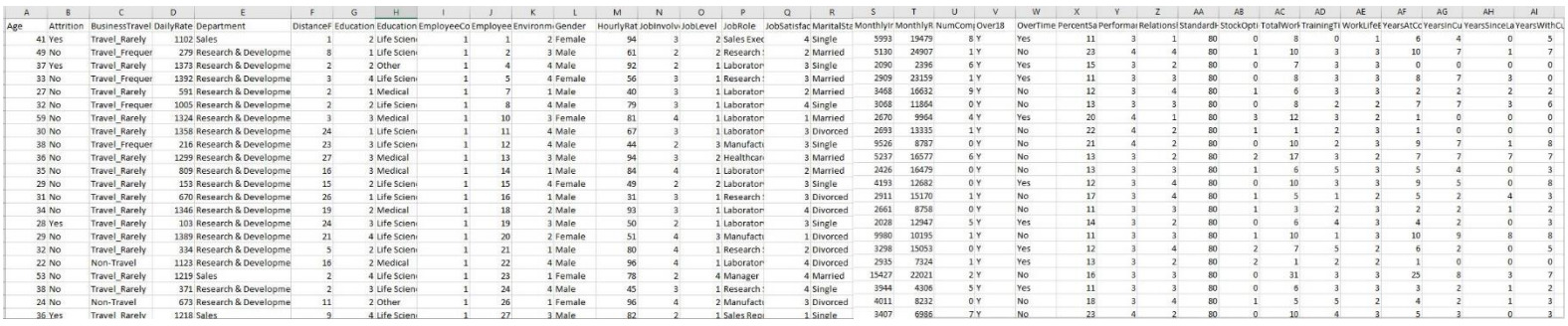
2. Dataset Description

The dataset I used for this assignment was IBM Employee Attrition Dataset which was downloaded from the Kaggle data repository. It is a fictional dataset created by the IBM data scientists. The dataset contains 35 attributes and 1470 instances. Following table provides details about the attributes given in the dataset.

Attribute Name	**Description**
Age	Numerical Value
Attrition	No, Yes
BusinessTravel	No Travel, Travel Frequently, Travel Rarely
DailyRate	Numerical Value - Salary Level
Department	Human Resources, Research & Development, Sales
DistanceFromHome	Numerical Value - The Distance from Work to Home
Education	1-Below College, 2-College, 3-Bachelor, 4-Master, 5-Doctor
EducationField	Human Resources, Life Sciences, Marketing, Medical Sciences, Others, Technical
EmployeeCount	Numerical Value

EmployeeNumber	Numerical Value
EnvironmentSatisfaction	1-Low, 2-Medium, 3-High, 4-Very High
Gender	Male, Female
HourlyRate	Numerical Value
JobInvolvement	1-Low, 2-Medium, 3-High, 4-Very High
JobLevel	Numerical Value
JobRole	Healthcare Representative, Human Resources, Laboratory Technician, Manager, Managing Director, Research Director, Research Scientist, Sales Executive, Sales Representative
JobSatisfaction	1-Low, 2-Medium, 3-High, 4-Very High
MaritalStatus	Divorced, Married, Single
MonthlyIncome	Numerical Value
MonthlyRate	Numerical Value
NumCompaniesWorked	Numerical Value
Over18	Y
OverTime	Yes, No
PercentSalaryHike	Numerical Value
PerformanceRating	1-Low, 2-Good, 3-Excellent, 4-Outstanding
RelationshipSatisfaction	1-Low, 2-Medium, 3-High, 4-Very High
StandardHours	Numerical Value
StockOptionLevel	Numerical Value
TotalWokingYears	Numerical Value
TrainingTimesLastYear	Numerical Value
WorkLifeBalance	1-Bad, 2-Good, 3-Better, 4-Best
YearsAtCompany	Numerical Value
YearsInCurrentRole	Numerical Value
YearsSinceLastPromotion	Numerical Value
YearsWithCurrManager	Numerical Value

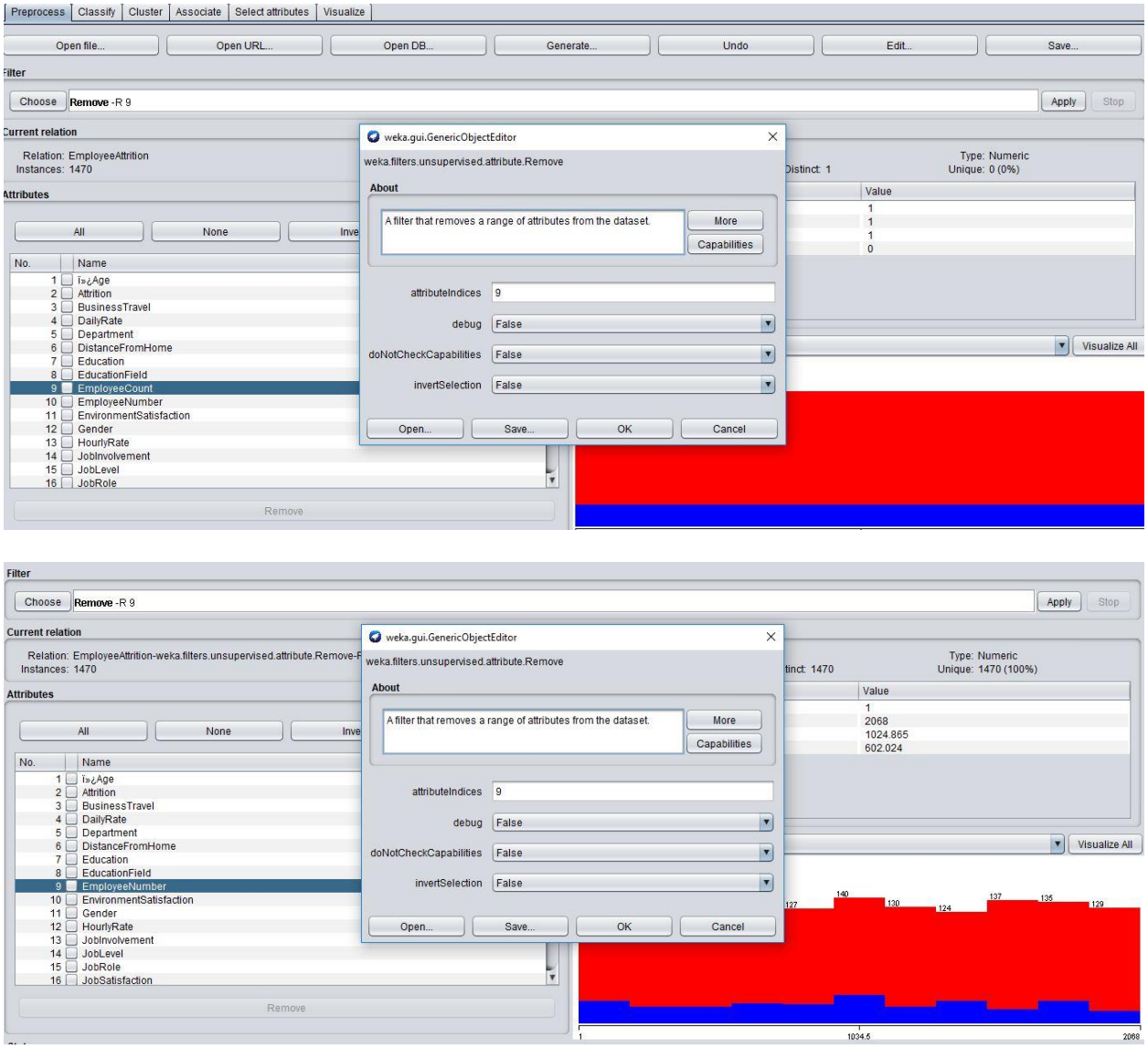
Following diagram shows a sample of the employee attrition dataset.



Age	Attrition	BusinessTravel	DailyRate	Department	DistanceF	Education	Education	EmployeeCo	Employee	Environm	Gender	HourlyRat	JobInvolv	JobLevel	JobRole	JobSatisfac	MaritalSte	MonthlyIr	MonthlyR	NumComp	Over18	OverTime	PercentSa	Performar	RelationsI	Standard	StockOpti	TotalWor	TrainingTi	WorkLifeE	YearsAtCc	YearsInCu	YearsSince	YearsWithC
41	Yes	Travel_Rarely	1102	Sales	1	2	Life Scien	1	1	2	Female	94	3	2	Sales Exec	4	Single	5993	19479	8	Y	Yes	11	3	1	80	0	8	0	1	6	4	0	5
49	No	Travel_Frequer	279	Research & Developme	8	1	Life Scien	1	2	3	Male	61	2	2	Research	2	Married	5130	24907	1	Y	No	23	4	4	80	1	10	3	3	10	7	1	7
37	Yes	Travel_Rarely	1373	Research & Developme	2	2	Other	1	4	4	Male	92	2	1	Laboratori	3	Single	2090	2396	6	Y	Yes	15	3	2	80	0	7	3	3	0	0	0	0
33	No	Travel_Frequer	1392	Research & Developme	3	4	Life Scien	1	5	4	Female	56	3	1	Research	3	Married	2909	23159	1	Y	Yes	11	3	3	80	0	8	3	3	8	7	3	0
27	No	Travel_Rarely	591	Research & Developme	2	1	Medical	1	7	1	Male	40	3	1	Laboratori	2	Married	3468	16632	9	Y	No	12	3	4	80	1	6	3	3	2	2	2	2
32	No	Travel_Frequer	1005	Research & Developme	2	2	Life Scien	1	8	4	Male	79	3	1	Laboratori	4	Single	3068	11864	0	Y	No	13	3	3	80	0	8	2	2	7	7	3	6
59	No	Travel_Rarely	1324	Research & Developme	3	3	Medical	1	10	3	Female	81	4	1	Laboratori	1	Married	2670	9964	4	Y	Yes	20	4	1	80	3	12	3	2	1	0	0	0
30	No	Travel_Rarely	1358	Research & Developme	24	1	Life Scien	1	11	4	Male	67	3	1	Laboratori	3	Divorced	2693	13335	1	Y	No	22	4	2	80	1	1	2	3	1	0	0	0
38	No	Travel_Frequer	216	Research & Developme	23	3	Life Scien	1	12	4	Male	44	2	3	Manufacti	3	Single	9526	8787	0	Y	No	21	4	2	80	0	10	2	3	9	7	1	8
36	No	Travel_Rarely	1299	Research & Developme	27	3	Medical	1	13	3	Male	94	3	2	Healthcar	3	Married	5237	16577	6	Y	No	13	3	2	80	2	17	3	2	7	7	7	7
35	No	Travel_Rarely	809	Research & Developme	16	3	Medical	1	14	1	Male	84	4	1	Laboratori	2	Married	2426	16479	0	Y	No	13	3	3	80	1	6	5	3	5	4	0	3
29	No	Travel_Rarely	153	Research & Developme	15	2	Life Scien	1	15	4	Female	49	2	2	Laboratori	3	Single	4193	12682	0	Y	Yes	12	3	4	80	0	10	3	3	9	5	0	8
31	No	Travel_Rarely	670	Research & Developme	26	1	Life Scien	1	16	1	Male	31	3	1	Research	3	Divorced	2911	15170	1	Y	No	17	3	4	80	1	5	1	2	5	2	4	3
34	No	Travel_Rarely	1346	Research & Developme	19	2	Medical	1	18	2	Male	93	3	1	Laboratori	4	Divorced	2661	8758	0	Y	No	11	3	3	80	1	3	2	3	2	2	1	2
28	Yes	Travel_Rarely	103	Research & Developme	24	3	Life Scien	1	19	3	Male	50	2	1	Laboratori	3	Single	2028	12947	5	Y	Yes	14	3	2	80	0	6	4	3	4	2	0	3
29	No	Travel_Rarely	1389	Research & Developme	21	4	Life Scien	1	20	2	Female	51	4	3	Manufacti	1	Divorced	9980	10195	1	Y	No	11	3	3	80	1	10	1	3	10	9	8	8
32	No	Travel_Rarely	334	Research & Developme	5	2	Life Scien	1	21	1	Male	80	4	1	Research	2	Divorced	3298	15053	0	Y	Yes	12	3	4	80	2	7	5	2	6	2	0	5
22	No	Non-Travel	1123	Research & Developme	16	2	Medical	1	22	4	Male	96	4	1	Laboratori	4	Divorced	2935	7324	1	Y	Yes	13	3	2	80	2	1	2	2	1	0	0	0
53	No	Travel_Rarely	1219	Sales	2	4	Life Scien	1	23	1	Female	78	2	4	Manager	4	Married	15427	22021	2	Y	No	16	3	3	80	0	31	3	3	25	8	3	7
38	No	Travel_Rarely	371	Research & Developme	2	3	Life Scien	1	24	4	Male	45	3	1	Research	4	Single	3944	4306	5	Y	Yes	11	3	3	80	0	6	3	3	2	1	2	2
24	No	Non-Travel	673	Research & Developme	11	2	Other	1	26	1	Female	96	4	2	Manufacti	3	Divorced	4011	8232	0	Y	No	18	3	4	80	1	5	5	2	4	2	1	3
36	Yes	Travel_Rarely	1218	Sales	9	4	Life Scien	1	27	3	Male	82	2	1	Sales Repi	1	Single	3407	6986	7	Y	No	23	4	2	80	0	10	4	3	5	3	0	3

2.1. Data Preprocessing

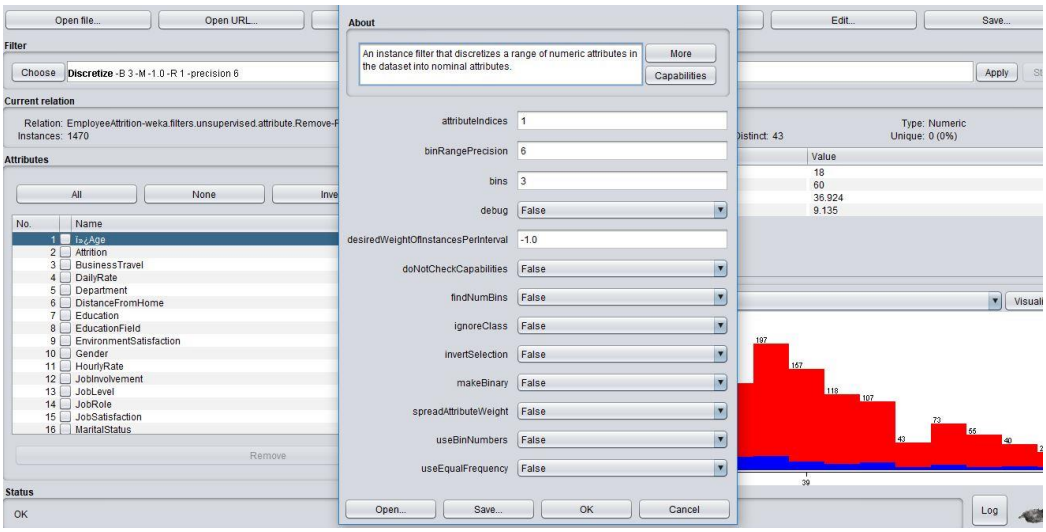
The dataset did not contain any missing values. Therefore, no preprocessing steps were carried out to remove the missing values. Among the attributes in the dataset EmployeeCount, EmployeeNumber and Over18, StandardHours attributes were identified not giving any useful information relevant to employee attrition. All the instances had EmployeeCount as 1, Over18 as Y and StandardHours as 80. EmployeeNumber is a unique identifier for each employee. Therefore, they were removed using Remove attribute filter in WEKA. The following diagram shows the removal of EmployeeCount and EmployeeNumber attributes.



The diagram given below shows the remaining set of attributes after removing the above two attributes.

```
@attribute Age numeric
@attribute Attrition {Yes,No}
@attribute BusinessTravel {Travel_Rarely,Travel_Frequently,Non-Travel}
@attribute DailyRate numeric
@attribute Department {Sales,'Research & Development','Human Resources'}
@attribute DistanceFromHome numeric
@attribute Education numeric
@attribute EducationField {'Life Sciences',Other,Medical,Marketing,'Technical Degree','Human Resources'}
@attribute EnvironmentSatisfaction numeric
@attribute Gender {Female,Male}
@attribute HourlyRate numeric
@attribute JobInvolvement numeric
@attribute JobLevel numeric
@attribute JobRole {'Sales Executive','Research Scientist','Laboratory Technician','Manufacturing Director','Healthcare Representative',Manager,'Sales
Representative','Research Director','Human Resources'}
@attribute JobSatisfaction numeric
@attribute MaritalStatus {Single,Married,Divorced}
@attribute MonthlyIncome numeric
@attribute MonthlyRate numeric
@attribute NumCompaniesWorked numeric
@attribute OverTime {Yes,No}
@attribute PercentSalaryHike numeric
@attribute PerformanceRating numeric
@attribute RelationshipSatisfaction numeric
@attribute StandardHours numeric
@attribute TotalWorkingYears numeric
@attribute TrainingTimesLastYear numeric
@attribute WorkLifeBalance numeric
@attribute YearsAtCompany numeric
@attribute YearsInCurrentRole numeric
@attribute YearsSinceLastPromotion numeric
@attribute YearsWithCurrManager numeric
```

After removing the above mentioned attributes, the dataset contained 23 numeric attributes including Age, DailyRate, DistanceFromHome, Education, EnvironmentSatisfaction, HourlyRate, JobInvolvement,JobLevel,JobSatisfaction,MonthlyIncome,MonthlyRate,NumCompaniesWorked ,PercentSalaryHike,PerformanceRating,RelationshipSatisfaction,StockOptionLevel,TotalWorking Years,TrainingTimesLastYear,WorkLifeBalance,YearsAtCompany,YearsInCurrentRole,YearsSince LastPromotion, and YearsWithCurrManager. Since association rule mining can only be performed on categorical data, discretization of these numeric attributes was required. In order to perform discretization on Age, DailyRate, DistanceFromHome, HourlyRate, MonthlyIncome, MonthlyRate,NumCompaniesWorked,PercentSalaryHike,TotalWorkingYears,TrainingTimesLastY ear,YearsAtCompany,YearsInCurrentRole,YearsSinceLastPromotion, and YearsWithCurrManager attributes WEKA was used. Each of the attributes were discretized into 3 bins. Following diagrams show discretization applied for some of the attributes listed above.



Selected attribute

Name: ï»¿Age
Missing: 0 (0%)
Distinct: 3
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	'(-inf-32]'	516	516.0
2	'(32-46]'	714	714.0
3	'(46-inf)'	240	240.0

Name: DailyRate
Missing: 0 (0%)
Distinct: 3
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	'(-inf-567.666667]'	484	484.0
2	'(567.666667-1033.333333]'	490	490.0
3	'(1033.333333-inf)'	496	496.0

Name: DistanceFromHome
Missing: 0 (0%)
Distinct: 3
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	'(-inf-10.333333]'	1026	1026.0
2	'(10.333333-19.666667]'	215	215.0
3	'(19.666667-inf)'	229	229.0

Education,EnvironmentSatisafaction,JobInvolvement,JobLevel,JobSatisfaction,PerformaceRating, RelationshipSatisfaction, StockOptionLevel, WorkLifeBalance attributes were discretized manually using a text editor where the “Numeric” keyword was placed with the relevant set of discrete values.

```
@attribute Age {'\'(-inf-32]\'','\'(32-46]\'','\'(46-inf)\''}
@attribute Attrition {Yes,No}
@attribute BusinessTravel {Travel_Rarely,Travel_Frequently,Non-Travel}
@attribute DailyRate {'\'(-inf-567.666667]\'','\'(567.666667-1033.333333]\'','\'(1033.333333-inf)\''}
@attribute Department {Sales,'Research & Development','Human Resources'}
@attribute DistanceFromHome {'\'(-inf-10.333333]\'','\'(10.333333-19.666667]\'','\'(19.666667-inf)\''}
@attribute Education {1,2,3,4,5}
@attribute EducationField {'Life Sciences',Other,Medical,Marketing,'Technical Degree','Human Resources'}
@attribute EnvironmentSatisfaction {1,2,3,4}
@attribute Gender {Female,Male}
@attribute HourlyRate {'\'(-inf-53.333333]\'','\'(53.333333-76.666667]\'','\'(76.666667-inf)\''}
@attribute JobInvolvement {1,2,3,4}
@attribute JobLevel {1,2,3,4,5}
@attribute JobRole {'Sales Executive','Research Scientist','Laboratory Technician','Manufacturing Director','Healthcare Representative',Manager,'Sales
Representative','Research Director','Human Resources'}
@attribute JobSatisfaction {1,2,3,4}
@attribute MaritalStatus {Single,Married,Divorced}
@attribute MonthlyIncome {'\'(-inf-7339]\'','\'(7339-13669]\'','\'(13669-inf)\''}
@attribute MonthlyRate {'\'(-inf-10395.666667]\'','\'(10395.666667-18697.333333]\'','\'(18697.333333-inf)\''}
@attribute NumCompaniesWorked {'\'(-inf-3]\'','\'(3-6]\'','\'(6-inf)\''}
@attribute OverTime {Yes,No}
@attribute PercentSalaryHike {'\'(-inf-15.666667]\'','\'(15.666667-20.333333]\'','\'(20.333333-inf)\''}
@attribute PerformanceRating {3,4}
@attribute RelationshipSatisfaction {1,2,3,4}
@attribute StockOptionLevel {0,1,2,3}
@attribute TotalWorkingYears {'\'(-inf-13.333333]\'','\'(13.333333-26.666667]\'','\'(26.666667-inf)\''}
@attribute TrainingTimesLastYear numeric
@attribute WorkLifeBalance {1,2,3,4}
@attribute YearsAtCompany {'\'(-inf-13.333333]\'','\'(13.333333-26.666667]\'','\'(26.666667-inf)\''}
@attribute YearsInCurrentRole {'\'(-inf-6]\'','\'(6-12]\'','\'(12-inf)\''}
@attribute YearsSinceLastPromotion {'\'(-inf-5]\'','\'(5-10]\'','\'(10-inf)\''}
@attribute YearsWithCurrManager {'\'(-inf-5.666667]\'','\'(5.666667-11.333333]\'','\'(11.333333-inf)\''}
```

The diagram given above shows the saved .arff file after the discretization process.

WEKA has assigned its own labels to each of the value ranges for the discretized attributes. Therefore, to replace these labels with more succinct and readable ones the global search/replace functions in the text editor was used. The manual re-labeling process was carried out for Age, DailyRate, DistanceFromHome, HourlyRate, MonthlyIncome, MonthlyRate, NumCompaniesWorked,PercentSalaryHike,TotalWorkingYears,TrainingTimesLastYear,YearsAtCompany, YearsInCurrentRole, YearsSinceLastPromotion, and YearsWithCurrMgr attributes. The following diagram shows the .arff file after re-labeling.

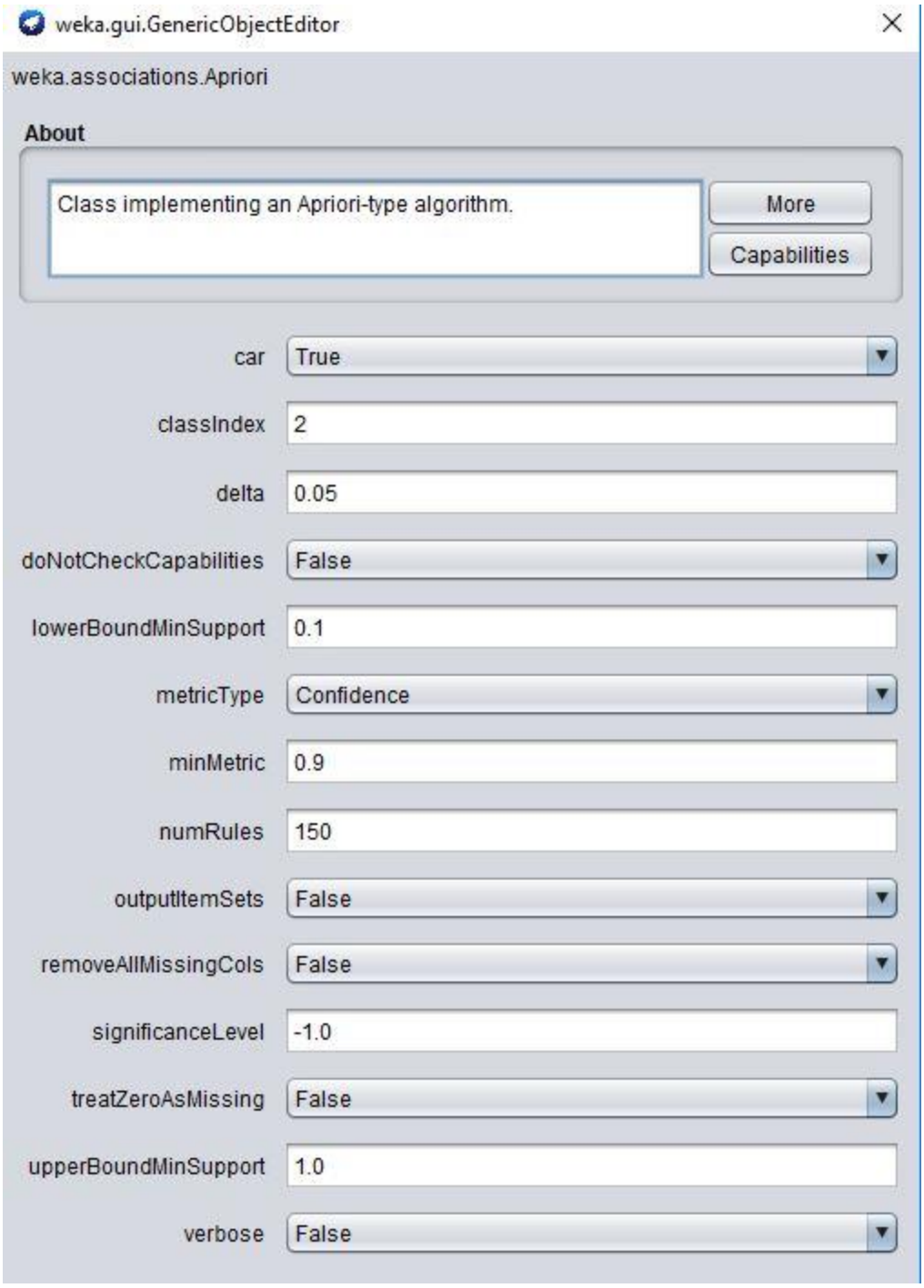
```
@attribute Age {18_32,33_46,47_60}
@attribute Attrition {Yes,No}
@attribute BusinessTravel {Travel_Rarely,Travel_Frequently,Non_Travel}
@attribute DailyRate {102_568,569_1033,1034_max}
@attribute Department {Sales,'Research & Development','Human Resources'}
@attribute DistanceFromHome {1_10,11_20,21_max}
@attribute Education {1,2,3,4,5}
@attribute EducationField {'Life Sciences',Other,Medical,Marketing,'Technical Degree','Human Resources'}
@attribute EnvironmentSatisfaction {1,2,3,4}
@attribute Gender {Female,Male}
@attribute HourlyRate {30_53,54_77,78_100}
@attribute JobInvolvement {1,2,3,4}
@attribute JobLevel {1,2,3,4,5}
@attribute JobRole {'Sales Executive','Research Scientist','Laboratory Technician','Manufacturing Director','Healthcare Representative',Manager,'Sales
Representative','Research Director','Human Resources'}
@attribute JobSatisfaction {1,2,3,4}
@attribute MaritalStatus {Single,Married,Divorced}
@attribute MonthlyIncome {min_7339,7340_13669,13670_max}
@attribute MonthlyRate {min_10396,10397_18697,18698_max}
@attribute NumCompaniesWorked {0_3,4_6,7_9}
@attribute OverTime {Yes,No}
@attribute PercentSalaryHike {11_16,17_20,21_25}
@attribute PerformanceRating {3,4}
@attribute RelationshipSatisfaction {1,2,3,4}
@attribute StockOptionLevel {0,1,2,3}
@attribute TotalWorkingYears {0_13,14_27,28_40}
@attribute TrainingTimesLastYear {0_2,3_4,5_6}
@attribute WorkLifeBalance {1,2,3,4}
@attribute YearsAtCompany {0_13,14_27,28_40}
@attribute YearsInCurrentRole {0_6,7_12,13_18}
@attribute YearsSinceLastPromotion {0_5,6_10,11_15}
@attribute YearsWithCurrManager {0_6,6_11,12_17}
```

The preprocessing of the dataset was completed through the above steps.

3. Rule Mining Process

In order to mine association rules on the IBM employee attrition dataset using WEKA, Apriori algorithm was used. Apriori is one of the simple and widely used algorithms for association rule mining. Features of Apriori algorithm that were considered when choosing it for association rule mining on employee attrition dataset are usage of large item set property, easily parallelized, simple and easy to implement, efficient for finding all frequent item sets and very effective for extracting repeated groups for Boolean association rules.

The Apriori algorithm was excercised to get association rules that have minSupport=0.2 and minConfidence=0.9. The parameters set for the Arpriori algorithm is given below.



Since the intention was to find associations between attrition attribute and other attributes, the "car" parameter was changed to True and index of the attrition attribute was given as the classIndex. Weka took about 1 second to generate 150 association rules based on the parameters provided on the employee attrition dataset.

When loweBoundMinSupport was set to 0.1 and numRules were set to 100, the rule generated with maximum confidence had only 0.94. But when the numRules were 150, maximum confidence of the rule generated was 0.96.

4. Resulting Rules

The 150 association rules generated by Weka are given below. They depict interesting associations between attrition and the rest of the 30 attributes in the dataset.

```
Apriori
=======

Minimum support: 0.2 (294 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 52

Size of set of large itemsets L(2): 385

Size of set of large itemsets L(3): 933

Size of set of large itemsets L(4): 1308

Size of set of large itemsets L(5): 1055

Size of set of large itemsets L(6): 447

Size of set of large itemsets L(7): 89

Size of set of large itemsets L(8): 7
```

```
Best rules found:

 1. i»¿Age=33_46 NumCompaniesWorked=0_3 OverTime=No 328 ==> Attrition=No 314    conf:(0.96)
 2. i»¿Age=33_46 Department=Research & Development OverTime=No 336 ==> Attrition=No 317    conf:(0.94)
 3. i»¿Age=33_46 OverTime=No WorkLifeBalance=3 321 ==> Attrition=No 302    conf:(0.94)
 4. i»¿Age=33_46 OverTime=No TotalWorkingYears=0_13 336 ==> Attrition=No 316    conf:(0.94)
 5. i»¿Age=33_46 OverTime=No TotalWorkingYears=0_13 YearsAtCompany=0_13 336 ==> Attrition=No 316    conf:(0.94)
 6. i»¿Age=33_46 OverTime=No PerformanceRating=3 YearsSinceLastPromotion=0_5 360 ==> Attrition=No 338    conf:(0.94)
 7. i»¿Age=33_46 MonthlyIncome=min_7339 OverTime=No 353 ==> Attrition=No 331    conf:(0.94)
 8. OverTime=No StockOptionLevel=1 YearsAtCompany=0_13 368 ==> Attrition=No 345    conf:(0.94)
 9. OverTime=No StockOptionLevel=1 429 ==> Attrition=No 402    conf:(0.94)
10. i»¿Age=33_46 DistanceFromHome=1_10 OverTime=No 363 ==> Attrition=No 340    conf:(0.94)
11. i»¿Age=33_46 OverTime=No PerformanceRating=3 YearsAtCompany=0_13 362 ==> Attrition=No 339    conf:(0.94)
12. i»¿Age=33_46 OverTime=No PerformanceRating=3 437 ==> Attrition=No 409    conf:(0.94)
13. Department=Research & Development OverTime=No WorkLifeBalance=3 YearsAtCompany=0_13 358 ==> Attrition=No 335    conf:(0.94)
14. i»¿Age=33_46 OverTime=No YearsSinceLastPromotion=0_5 418 ==> Attrition=No 391    conf:(0.94)
15. JobLevel=2 MonthlyIncome=min_7339 OverTime=No 356 ==> Attrition=No 333    conf:(0.94)
16. OverTime=No StockOptionLevel=1 YearsSinceLastPromotion=0_5 356 ==> Attrition=No 333    conf:(0.94)
17. DistanceFromHome=1_10 JobLevel=2 MonthlyIncome=min_7339 325 ==> Attrition=No 304    conf:(0.94)
18. i»¿Age=33_46 MonthlyIncome=min_7339 OverTime=No YearsAtCompany=0_13 325 ==> Attrition=No 304    conf:(0.94)
19. i»¿Age=33_46 OverTime=No 510 ==> Attrition=No 477    conf:(0.94)
20. DistanceFromHome=1_10 MaritalStatus=Married OverTime=No 340 ==> Attrition=No 318    conf:(0.94)
21. OverTime=No YearsInCurrentRole=7_12 324 ==> Attrition=No 303    conf:(0.94)
22. OverTime=No StockOptionLevel=1 YearsAtCompany=0_13 YearsSinceLastPromotion=0_5 324 ==> Attrition=No 303    conf:(0.94)
23. i»¿Age=33_46 OverTime=No PercentSalaryHike=11_16 321 ==> Attrition=No 300    conf:(0.93)
24. i»¿Age=33_46 OverTime=No PercentSalaryHike=11_16 PerformanceRating=3 321 ==> Attrition=No 300    conf:(0.93)
25. JobLevel=2 OverTime=No YearsSinceLastPromotion=0_5 336 ==> Attrition=No 314    conf:(0.93)
```

26. Department=Research & Development OverTime=No WorkLifeBalance=3 411 ==> Attrition=No 384   conf:(0.93)
27. Department=Research & Development DistanceFromHome=1_10 NumCompaniesWorked=0_3 OverTime=No 334 ==> Attrition=No 312   conf:(0.93)
28. ï»¿Age=33_46 OverTime=No PerformanceRating=3 YearsAtCompany=0_13 YearsSinceLastPromotion=0_5 318 ==> Attrition=No 297   conf:(0.93)
29. ï»¿Age=33_46 OverTime=No YearsAtCompany=0_13 422 ==> Attrition=No 394   conf:(0.93)
30. JobLevel=2 OverTime=No 388 ==> Attrition=No 362   conf:(0.93)
31. ï»¿Age=33_46 DistanceFromHome=1_10 NumCompaniesWorked=0_3 323 ==> Attrition=No 301   conf:(0.93)
32. DistanceFromHome=1_10 NumCompaniesWorked=0_3 OverTime=No WorkLifeBalance=3 323 ==> Attrition=No 301   conf:(0.93)
33. StockOptionLevel=1 WorkLifeBalance=3 352 ==> Attrition=No 328   conf:(0.93)
34. Department=Research & Development OverTime=No WorkLifeBalance=3 YearsSinceLastPromotion=0_5 352 ==> Attrition=No 328   conf:(0.93)
35. BusinessTravel=Travel_Rarely Department=Research & Development DistanceFromHome=1_10 OverTime=No 349 ==> Attrition=No 325   conf:(0.93)
36. JobLevel=2 OverTime=No YearsAtCompany=0_13 YearsSinceLastPromotion=0_5 319 ==> Attrition=No 297   conf:(0.93)
37. JobLevel=2 OverTime=No PerformanceRating=3 332 ==> Attrition=No 309   conf:(0.93)
38. JobLevel=2 MonthlyIncome=min_7339 OverTime=No YearsAtCompany=0_13 332 ==> Attrition=No 309   conf:(0.93)
39. JobSatisfaction=4 OverTime=No 317 ==> Attrition=No 295   conf:(0.93)
40. Department=Research & Development OverTime=No PerformanceRating=3 WorkLifeBalance=3 330 ==> Attrition=No 307   conf:(0.93)
41. ï»¿Age=33_46 OverTime=No YearsAtCompany=0_13 YearsSinceLastPromotion=0_5 373 ==> Attrition=No 347   conf:(0.93)
42. OverTime=No PerformanceRating=3 StockOptionLevel=1 358 ==> Attrition=No 333   conf:(0.93)
43. DistanceFromHome=1_10 JobLevel=2 357 ==> Attrition=No 332   conf:(0.93)
44. Department=Research & Development OverTime=No WorkLifeBalance=3 YearsAtCompany=0_13 YearsSinceLastPromotion=0_5 328 ==> Attrition=No 305   conf:(0.93)
45. JobLevel=2 MonthlyIncome=min_7339 NumCompaniesWorked=0_3 326 ==> Attrition=No 303   conf:(0.93)
46. ï»¿Age=33_46 BusinessTravel=Travel_Rarely OverTime=No 353 ==> Attrition=No 328   conf:(0.93)
47. BusinessTravel=Travel_Rarely NumCompaniesWorked=0_3 OverTime=No WorkLifeBalance=3 322 ==> Attrition=No 299   conf:(0.93)
48. JobLevel=2 OverTime=No YearsAtCompany=0_13 362 ==> Attrition=No 336   conf:(0.93)
49. BusinessTravel=Travel_Rarely Department=Research & Development NumCompaniesWorked=0_3 OverTime=No 334 ==> Attrition=No 310   conf:(0.93)
50. Department=Research & Development EducationField=Life Sciences OverTime=No 319 ==> Attrition=No 296   conf:(0.93)
51. BusinessTravel=Travel_Rarely OverTime=No WorkLifeBalance=3 YearsAtCompany=0_13 401 ==> Attrition=No 372   conf:(0.93)
52. ï»¿Age=33_46 Department=Research & Development DistanceFromHome=1_10 318 ==> Attrition=No 295   conf:(0.93)

53. BusinessTravel=Travel_Rarely DistanceFromHome=1_10 OverTime=No WorkLifeBalance=3 330 ==> Attrition=No 306   conf:(0.93)
54. BusinessTravel=Travel_Rarely OverTime=No WorkLifeBalance=3 YearsSinceLastPromotion=0_5 396 ==> Attrition=No 367   conf:(0.93)
55. MaritalStatus=Married MonthlyIncome=min_7339 OverTime=No 341 ==> Attrition=No 316   conf:(0.93)
56. BusinessTravel=Travel_Rarely OverTime=No WorkLifeBalance=3 YearsAtCompany=0_13 YearsSinceLastPromotion=0_5 366 ==> Attrition=No 339   conf:(0.93)
57. BusinessTravel=Travel_Rarely OverTime=No WorkLifeBalance=3 460 ==> Attrition=No 426   conf:(0.93)
58. MaritalStatus=Married OverTime=No YearsSinceLastPromotion=0_5 403 ==> Attrition=No 373   conf:(0.93)
59. OverTime=No YearsWithCurrManager=6_11 347 ==> Attrition=No 321   conf:(0.93)
60. JobLevel=2 NumCompaniesWorked=0_3 359 ==> Attrition=No 332   conf:(0.92)
61. MaritalStatus=Married OverTime=No YearsAtCompany=0_13 YearsSinceLastPromotion=0_5 371 ==> Attrition=No 343   conf:(0.92)
62. BusinessTravel=Travel_Rarely DistanceFromHome=1_10 NumCompaniesWorked=0_3 OverTime=No 357 ==> Attrition=No 330   conf:(0.92)
63. DistanceFromHome=1_10 JobLevel=2 YearsAtCompany=0_13 330 ==> Attrition=No 305   conf:(0.92)
64. JobInvolvement=3 OverTime=No WorkLifeBalance=3 369 ==> Attrition=No 341   conf:(0.92)
65. ï»¿Age=33_46 Department=Research & Development PerformanceRating=3 YearsSinceLastPromotion=0_5 329 ==> Attrition=No 304   conf:(0.92)
66. BusinessTravel=Travel_Rarely MonthlyIncome=min_7339 OverTime=No WorkLifeBalance=3 329 ==> Attrition=No 304   conf:(0.92)
67. MaritalStatus=Married MonthlyIncome=min_7339 OverTime=No YearsAtCompany=0_13 326 ==> Attrition=No 301   conf:(0.92)
68. BusinessTravel=Travel_Rarely OverTime=No PerformanceRating=3 WorkLifeBalance=3 YearsAtCompany=0_13 339 ==> Attrition=No 313   conf:(0.92)
69. BusinessTravel=Travel_Rarely Department=Research & Development OverTime=No PerformanceRating=3 YearsSinceLastPromotion=0_5 363 ==> Attrition=No 335   conf:(0.92)
70. MaritalStatus=Married NumCompaniesWorked=0_3 OverTime=No 337 ==> Attrition=No 311   conf:(0.92)
71. MaritalStatus=Married OverTime=No PerformanceRating=3 YearsSinceLastPromotion=0_5 337 ==> Attrition=No 311   conf:(0.92)
72. BusinessTravel=Travel_Rarely OverTime=No PerformanceRating=3 WorkLifeBalance=3 YearsSinceLastPromotion=0_5 337 ==> Attrition=No 311   conf:(0.92)
73. Department=Research & Development DistanceFromHome=1_10 OverTime=No 490 ==> Attrition=No 452   conf:(0.92)
74. MonthlyIncome=min_7339 NumCompaniesWorked=0_3 OverTime=No WorkLifeBalance=3 333 ==> Attrition=No 307   conf:(0.92)
75. ï»¿Age=33_46 NumCompaniesWorked=0_3 PerformanceRating=3 397 ==> Attrition=No 366   conf:(0.92)
76. DistanceFromHome=1_10 MonthlyIncome=min_7339 OverTime=No WorkLifeBalance=3 320 ==> Attrition=No 295   conf:(0.92)
77. MaritalStatus=Married OverTime=No YearsAtCompany=0_13 422 ==> Attrition=No 389   conf:(0.92)
78. BusinessTravel=Travel_Rarely Department=Research & Development OverTime=No PerformanceRating=3 422 ==> Attrition=No 389   conf:(0.92)
79. Department=Research & Development DistanceFromHome=1_10 OverTime=No PerformanceRating=3 409 ==> Attrition=No 377   conf:(0.92)

80. BusinessTravel=Travel_Rarely Department=Research & Development OverTime=No 495 ==> Attrition=No 456   conf:(0.92)
81. JobInvolvement=3 OverTime=No WorkLifeBalance=3 YearsAtCompany=0_13 330 ==> Attrition=No 304   conf:(0.92)
82. BusinessTravel=Travel_Rarely Department=Research & Development OverTime=No PerformanceRating=3 YearsAtCompany=0_13 366 ==> Attrition=No 337   conf:(0.92)
83. BusinessTravel=Travel_Rarely Department=Research & Development OverTime=No YearsSinceLastPromotion=0_5 429 ==> Attrition=No 395   conf:(0.92)
84. JobLevel=2 NumCompaniesWorked=0_3 YearsAtCompany=0_13 328 ==> Attrition=No 302   conf:(0.92)
85. DistanceFromHome=1_10 StockOptionLevel=1 416 ==> Attrition=No 383   conf:(0.92)
86. BusinessTravel=Travel_Rarely OverTime=No PerformanceRating=3 WorkLifeBalance=3 390 ==> Attrition=No 359   conf:(0.92)
87. Department=Research & Development NumCompaniesWorked=0_3 OverTime=No 475 ==> Attrition=No 437   conf:(0.92)
88. DistanceFromHome=1_10 OverTime=No WorkLifeBalance=3 450 ==> Attrition=No 414   conf:(0.92)
89. Department=Research & Development DistanceFromHome=1_10 OverTime=No PerformanceRating=3 YearsSinceLastPromotion=0_5 350 ==> Attrition=No 322   conf:(0.92)
90. BusinessTravel=Travel_Rarely StockOptionLevel=1 YearsAtCompany=0_13 362 ==> Attrition=No 333   conf:(0.92)
91. BusinessTravel=Travel_Rarely Department=Research & Development OverTime=No YearsAtCompany=0_13 435 ==> Attrition=No 400   conf:(0.92)
92. Department=Research & Development DistanceFromHome=1_10 OverTime=No YearsSinceLastPromotion=0_5 422 ==> Attrition=No 388   conf:(0.92)
93. ï»¿Age=33_46 Department=Research & Development PerformanceRating=3 397 ==> Attrition=No 365   conf:(0.92)
94. BusinessTravel=Travel_Rarely Department=Research & Development OverTime=No PerformanceRating=3 YearsAtCompany=0_13 YearsSinceLastPromotion=0_5 334 ==> Attrition=N
95. DistanceFromHome=1_10 NumCompaniesWorked=0_3 OverTime=No 507 ==> Attrition=No 466   conf:(0.92)
96. BusinessTravel=Travel_Rarely StockOptionLevel=1 420 ==> Attrition=No 386   conf:(0.92)
97. NumCompaniesWorked=0_3 OverTime=No WorkLifeBalance=3 456 ==> Attrition=No 419   conf:(0.92)
98. DistanceFromHome=1_10 OverTime=No WorkLifeBalance=3 YearsSinceLastPromotion=0_5 382 ==> Attrition=No 351   conf:(0.92)
99. DistanceFromHome=1_10 NumCompaniesWorked=0_3 OverTime=No YearsSinceLastPromotion=0_5 431 ==> Attrition=No 396   conf:(0.92)
100. Department=Research & Development NumCompaniesWorked=0_3 OverTime=No PerformanceRating=3 YearsSinceLastPromotion=0_5 332 ==> Attrition=No 305   conf:(0.92)
101. Department=Research & Development NumCompaniesWorked=0_3 OverTime=No PerformanceRating=3 392 ==> Attrition=No 360   conf:(0.92)
102. DistanceFromHome=1_10 OverTime=No WorkLifeBalance=3 YearsAtCompany=0_13 392 ==> Attrition=No 360   conf:(0.92)
103. Department=Research & Development NumCompaniesWorked=0_3 OverTime=No YearsSinceLastPromotion=0_5 404 ==> Attrition=No 371   conf:(0.92)
104. ï»¿Age=33_46 DistanceFromHome=1_10 PerformanceRating=3 YearsSinceLastPromotion=0_5 355 ==> Attrition=No 326   conf:(0.92)
105. DistanceFromHome=1_10 OverTime=No WorkLifeBalance=3 YearsAtCompany=0_13 YearsSinceLastPromotion=0_5 354 ==> Attrition=No 325   conf:(0.92)
106. Department=Research & Development NumCompaniesWorked=0_3 OverTime=No YearsAtCompany=0_13 415 ==> Attrition=No 381   conf:(0.92)

```
107. ï»¿Age=33_46 MaritalStatus=Married 329 ==> Attrition=No 302    conf:(0.92)
108. Department=Research & Development NumCompaniesWorked=0_3 OverTime=No PerformanceRating=3 YearsAtCompany=0_13 341 ==> Attrition=No 313    conf:(0.92)
109. DistanceFromHome=1_10 StockOptionLevel=1 YearsAtCompany=0_13 353 ==> Attrition=No 324    conf:(0.92)
110. Department=Research & Development DistanceFromHome=1_10 OverTime=No YearsAtCompany=0_13 438 ==> Attrition=No 402    conf:(0.92)
111. DistanceFromHome=1_10 Gender=Male NumCompaniesWorked=0_3 OverTime=No 328 ==> Attrition=No 301    conf:(0.92)
112. MaritalStatus=Married OverTime=No TotalWorkingYears=0_13 340 ==> Attrition=No 312    conf:(0.92)
113. MaritalStatus=Married OverTime=No TotalWorkingYears=0_13 YearsAtCompany=0_13 340 ==> Attrition=No 312    conf:(0.92)
114. MaritalStatus=Married OverTime=No PerformanceRating=3 YearsAtCompany=0_13 352 ==> Attrition=No 323    conf:(0.92)
115. NumCompaniesWorked=0_3 OverTime=No WorkLifeBalance=3 YearsSinceLastPromotion=0_5 387 ==> Attrition=No 355    conf:(0.92)
116. BusinessTravel=Travel_Rarely Department=Research & Development OverTime=No YearsAtCompany=0_13 YearsSinceLastPromotion=0_5 399 ==> Attrition=No 366    conf:(0.92)
117. BusinessTravel=Travel_Rarely OverTime=No TotalWorkingYears=0_13 WorkLifeBalance=3 326 ==> Attrition=No 299    conf:(0.92)
118. BusinessTravel=Travel_Rarely OverTime=No TotalWorkingYears=0_13 WorkLifeBalance=3 YearsAtCompany=0_13 326 ==> Attrition=No 299    conf:(0.92)
119. Department=Research & Development DistanceFromHome=1_10 OverTime=No PerformanceRating=3 YearsAtCompany=0_13 YearsSinceLastPromotion=0_5 326 ==> Attrition=No 299
120. Department=Research & Development DistanceFromHome=1_10 OverTime=No PerformanceRating=3 YearsAtCompany=0_13 362 ==> Attrition=No 332    conf:(0.92)
121. ï»¿Age=33_46 DistanceFromHome=1_10 PerformanceRating=3 422 ==> Attrition=No 387    conf:(0.92)
122. DistanceFromHome=1_10 JobInvolvement=3 OverTime=No PerformanceRating=3 YearsSinceLastPromotion=0_5 325 ==> Attrition=No 298    conf:(0.92)
123. Department=Research & Development DistanceFromHome=1_10 OverTime=No YearsAtCompany=0_13 YearsSinceLastPromotion=0_5 397 ==> Attrition=No 364    conf:(0.92)
124. DistanceFromHome=1_10 OverTime=No PerformanceRating=3 WorkLifeBalance=3 YearsSinceLastPromotion=0_5 324 ==> Attrition=No 297    conf:(0.92)
125. Department=Research & Development MonthlyIncome=min_7339 NumCompaniesWorked=0_3 OverTime=No 359 ==> Attrition=No 329    conf:(0.92)
126. NumCompaniesWorked=0_3 OverTime=No WorkLifeBalance=3 YearsAtCompany=0_13 394 ==> Attrition=No 361    conf:(0.92)
127. DistanceFromHome=1_10 OverTime=No PerformanceRating=3 WorkLifeBalance=3 382 ==> Attrition=No 350    conf:(0.92)
128. MaritalStatus=Married OverTime=No 487 ==> Attrition=No 446    conf:(0.92)
129. DistanceFromHome=1_10 JobInvolvement=3 OverTime=No YearsSinceLastPromotion=0_5 379 ==> Attrition=No 347    conf:(0.92)
130. DistanceFromHome=1_10 JobInvolvement=3 OverTime=No 438 ==> Attrition=No 401    conf:(0.92)
131. OverTime=No WorkLifeBalance=3 639 ==> Attrition=No 585    conf:(0.92)
132. BusinessTravel=Travel_Rarely NumCompaniesWorked=0_3 OverTime=No YearsSinceLastPromotion=0_5 437 ==> Attrition=No 400    conf:(0.92)
133. DailyRate=1034_max OverTime=No 354 ==> Attrition=No 324    conf:(0.92)

134. DistanceFromHome=1_10 OverTime=No PerformanceRating=3 WorkLifeBalance=3 YearsAtCompany=0_13 330 ==> Attrition=No 302    conf:(0.92)
135. DistanceFromHome=1_10 NumCompaniesWorked=0_3 OverTime=No PerformanceRating=3 YearsSinceLastPromotion=0_5 365 ==> Attrition=No 334    conf:(0.92)
136. BusinessTravel=Travel_Rarely StockOptionLevel=1 YearsSinceLastPromotion=0_5 353 ==> Attrition=No 323    conf:(0.92)
137. Department=Research & Development NumCompaniesWorked=0_3 OverTime=No TotalWorkingYears=0_13 353 ==> Attrition=No 323    conf:(0.92)
138. Department=Research & Development NumCompaniesWorked=0_3 OverTime=No TotalWorkingYears=0_13 YearsAtCompany=0_13 353 ==> Attrition=No 323    conf:(0.92)
139. JobInvolvement=3 NumCompaniesWorked=0_3 OverTime=No 423 ==> Attrition=No 387    conf:(0.91)
140. DistanceFromHome=1_10 JobInvolvement=3 OverTime=No PerformanceRating=3 376 ==> Attrition=No 344    conf:(0.91)
141. JobLevel=2 WorkLifeBalance=3 329 ==> Attrition=No 301    conf:(0.91)
142. MonthlyIncome=min_7339 OverTime=No WorkLifeBalance=3 458 ==> Attrition=No 419    conf:(0.91)
143. Department=Research & Development NumCompaniesWorked=0_3 OverTime=No YearsAtCompany=0_13 YearsSinceLastPromotion=0_5 375 ==> Attrition=No 343    conf:(0.91)
144. Department=Research & Development OverTime=No 690 ==> Attrition=No 631    conf:(0.91)
145. Department=Research & Development OverTime=No PerformanceRating=3 573 ==> Attrition=No 524    conf:(0.91)
146. Department=Research & Development OverTime=No PerformanceRating=3 YearsSinceLastPromotion=0_5 491 ==> Attrition=No 449    conf:(0.91)
147. Department=Research & Development MonthlyIncome=min_7339 NumCompaniesWorked=0_3 OverTime=No YearsAtCompany=0_13 339 ==> Attrition=No 310    conf:(0.91)
148. OverTime=No WorkLifeBalance=3 YearsAtCompany=0_13 561 ==> Attrition=No 513    conf:(0.91)
149. OverTime=No WorkLifeBalance=3 YearsSinceLastPromotion=0_5 548 ==> Attrition=No 501    conf:(0.91)
150. DistanceFromHome=1_10 NumCompaniesWorked=0_3 OverTime=No PerformanceRating=3 431 ==> Attrition=No 394    conf:(0.91)
```

Some of the interesting association rules discovered using Apriori are discussed below.

One of the association rules has the maximum confidence, 0.96. It shows the following association.

- Employees between age 33-46, number of companies worked between 0-3 and don't do overtime are not likely to show attrition

21 rules are having 0.94 confidence. The associations revealed by them include,

- Employees between 33-46, in Research & Development department and don't do overtime are not likely to show attrition
- Employees between 33-46, don't do overtime and better work life balance are not likely to show attrition
- Employees between 33-46, don't do overtime and total working years between 0-13 are not likely to show attrition
- Employees between 33-46, don't do overtime, total working years between 0-13 and have spent between 0-13 years in the company are not likely to show attrition
- Employees between 33-46, don't do overtime, excellent performance rating and years since last promotion has been between 0-5 years are not likely to show attrition
- Employees between 33-46, monthly income between the minimum value and 7339, and don't do overtime are not likely to show attrition

- Employees who don't do overtime, stock option level=1 and years at the company is between 0-13 are not likely to show attrition
- Employees who don't do overtime and stock option level=1 are not likely to show attrition
- Employees between 33-46, distance from home between 1-10 and don't do overtime are not likely to show attrition
- Employees between 33-46, don't do overtime, excellent performance rating and years at the company is between 0-13 years are not likely to show attrition
- Employees between 33-46, don't do overtime and excellent performance rating are not likely to show attrition
- Employees in Research & Development department, don't do overtime and have better work life balance are not likely to show attrition
- Employees between 33-46, don't do overtime and years since last promotion has been between 0-5 years are not likely to show attrition
- Employees with Job level=2, monthly income between minimum and 7339 and don't do overtime are not likely to show attrition
- Employees who don't do overtime, with stock option level=1 and years since last promotion has been 0-5 years are not likely to show attrition
- Employees with distance from home between 1-10, with job level=2, monthly income between minimum and 7339 are not likely to show attrition
- Employees between age 33-46, monthly income between minimum and 7339, don't do overtime and has been in the company 0-13 years are not likely to show attrition
- Employees between age 33-46 and don't do overtime are not likely to show attrition
- Employees with distance from home between 1-10, married and don't do overtime are not likely to show attrition
- Employees that don't do overtime, has been in the current role for 7-12 years are not likely to show attrition
- Employees that don't do overtime, with stock option level=1, years at the company between 0-13 and years since last promotion has been 0-5 years are not likely to show attrition

37 rules are having 0.93 confidence, 79 rules are having 0.92 confidence and 12 rules are having 0.91 confidence. All these rules can be identified as interesting rules since the support and confidence for each of these rules are greater than the minimum support and the minimum confidence.

5. Recommendations

The generated association rules show non attrition of the employees. Therefore, attributes and combination of attributes that cause non attrition can be identified using the set of association rules generated. Through identification of these attributes it would help organizations to improve the quality of those attributes to retain their employees. Management of organizations can plan their future activities according to the identified associations so that it would improve the employee satisfaction ultimately paving way for benefit of the organization itself.