MCDA5580 - Data and Text Mining

Assignment – 3

SimplyCast.com Association Mining

Submitted By:
Nikhil Bhat (A00434789)
Mehar Singh (A00434701)
Samarth Gupta (A00433096)

Contents

1. Executive Summary

One of the world's best Customer Flow Communication platform – SimplyCast wants to analyze the website clicks data to identify the association between various milestones for users and sessions. As part of this analysis, the 6,65,435 data records of SimplyCast from 20-Jul-2015 to 17-Dec-2015 were analyzed and the findings were reported that may be beneficial to the company in understanding the customers use of various milestones and the association between them.

As a result of the analysis, 76 association rules were identified between various milestones for user level data and the following maximal frequent itemset was identified:

{TestSend}, {SendNow,SimpleProjCreate}, {ReEditProj,SimpleProjCreate}, {ManageTab,SimpleProjCreate}, {ProjPreview,SimpleProjCreate}, {ManageTab,ReEditProj,ReportsTab,SendNow}, {ManageTab,ProjPreview,ReportsTab,SendNow}, {ManageTab,ProjPreview,ReEditProj,ReportsTab}, {ManageTab,ProjPreview,ReEditProj,SendNow}

For the analyzed session level data, 20 association rules were generated and the following maximal frequent itemset was identified:

{TestSend}, {TxtFontSizeColor}, {ManageTab,ReportsTab}, {ProjPreview,ReEditProj}, {ManageTab,ProjPreview,SendNow}, {ManageTab,ReEditProj,SendNow}

2. Objective

The objective of the analysis is to use association mining through Apriori Algorithm to generate maximal frequent itemsets and rules for common associations between milestones for user and session level data.

3. Methodology

Analyze the SimplyCast.com data to understand the association between various milestones by applying association mining through the Apriori Algorithm. The algorithm should be used to generate association rules and maximal frequent itemsets between various milestones for user and session level data.

4. About the data

The SimplyCast.com user click data stored in 'rawdataDec15' table in MYSQL database 'dataset03' is used and it contains the following fields:

S.No.	Column Name	Description
1.	id	Unique identifier for each data row
2.	user_id	Unique identifier for each customer
3.	milestone_name	Name of the milestone on which the user clicked
4.	date	Date when the milestone was recorded
5.	time	Time when the milestone was recorded

Server: http://dev.cs.smu.ca/phpmyadmin/

Database Type: MySQL Database

Database: 'dataset03'

Table Used:

- **rawdataDec15:** Contains the history of user clicks on the website for various milestones from 20-Jul-2015 to 17-Dec-2015

The dataset consists of 6,65,435 records in the dataset and has the data stored for 3,159 users and 24,713 sessions.

5. User Analysis

5.1 Features

For the purpose of association mining of user data, the unique milestones for each user from the 'rawdataDec15' are stored into the table 'UserMilestone'. Since the 'rawdataDec15' contains a lot of repeated milestones for each user, it is crucial to extract only the distinct milestones for each user.

Following is schema of the 'UserMilestone' table for our analysis:

S.No.	Column Name	Description
1.	user_id	Unique identifier for each user
2.	milestone_name	Unique milestone names on which the user clicked

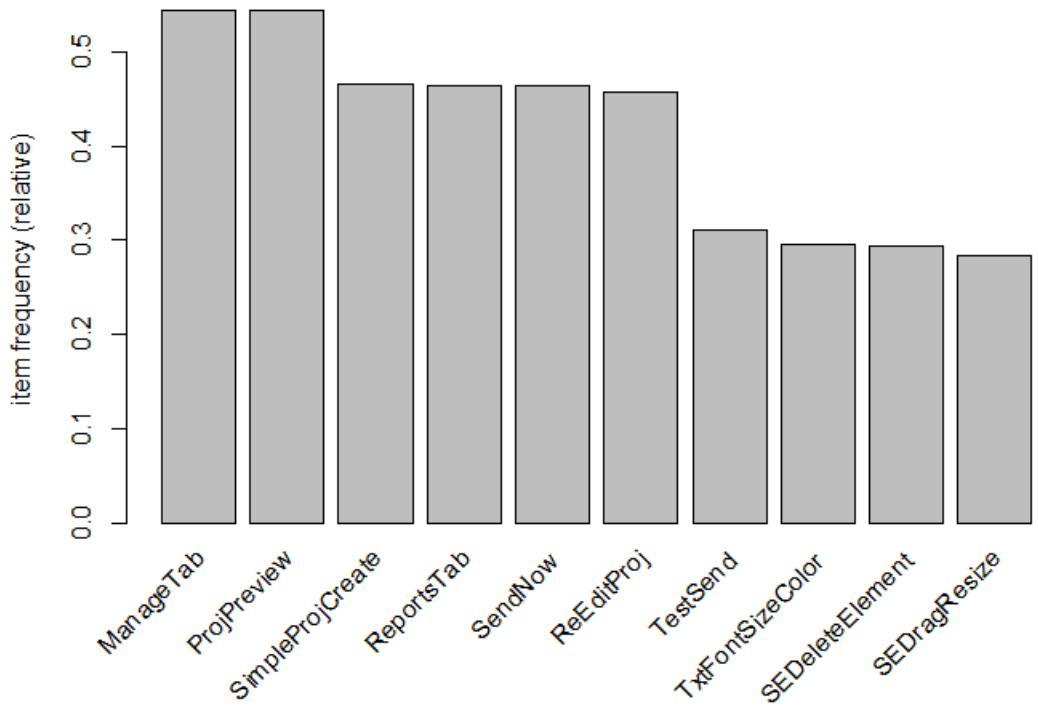
The data is generated by using a distinct combination of user_id and milestone_name fields and contains 39,096 records for 3,159 unique users.

5.2 Data Transformation

For running the Apriori algorithm, the unique milestones for each user should be transposed to concatenate all unique milestones for the user in single row. This step is completed in R and the following summary is generated in order to understand the data:

```
transactions as itemMatrix in sparse format with
 3159 rows (elements/itemsets/transactions) and
 112 columns (items) and a density of 0.1105006

most frequent items:
       ManageTab      ProjPreview SimpleProjCreate       ReportsTab          SendNow          (Other)
            1716             1715             1472             1467             1463            31263

element (itemset/transaction) length distribution:
sizes
  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35
807 241 147 108 101 103 101  71  76  72  70  78  54  56  58  65  55  65  42  50  55  46  48  31  42  40  24  28  23  27  21  32  33  14  22
 36  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60  61  64
 24  22  17  21  12  16  14  12  14  11   8  10  11   5   9  10   5   8   5   5   6   3   1   1   1   1   1

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00    1.00    7.00   12.38   19.00   64.00

includes extended item information - examples:
            labels
1       ABSplitProj
2      ABSplitTools
3 AccountSettingsA
```

The following histogram showcases the relative frequency of the top 10 milestones recorded in the dataset:



5.3 Analysis

After transposing and formatting the data, we run the Apriori Algorithm with the following support and confidence values to generate the rules:

1. Support = 0.2, Confidence = 0.5
   Rules Generated: 661

```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target   ext
        0.5    0.1    1 none FALSE            TRUE       5     0.2      1     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 631

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[112 item(s), 3159 transaction(s)] done [0.00s].
sorting and recoding items ... [21 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 5 6 done [0.00s].
writing ... [661 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

2. Support = 0.3, Confidence = 0.5
   Rules Generated: 76

```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target   ext
        0.5    0.1    1 none FALSE            TRUE       5     0.3      1     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 947

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[112 item(s), 3159 transaction(s)] done [0.01s].
sorting and recoding items ... [7 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 done [0.00s].
writing ... [76 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

3. Support = 0.4, Confidence = 0.5
   Rules Generated: 8

```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target   ext
        0.5    0.1    1 none FALSE            TRUE       5     0.4      1     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 1263

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[112 item(s), 3159 transaction(s)] done [0.00s].
sorting and recoding items ... [6 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing ... [8 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

Based on the above observations, following table is populated and it can be observed that support = 0.3 and confidence = 0.5 gives the optimum number of rules as 76:

S.No.	Support	Confidence	Number of Rules
1.	0.2	0.5	661
2.	0.3	0.5	76
3.	0.4	0.5	8

Note: Confidence value 0.5 is used as it is considered as an optimum value which provides reliable and good number of rules.

5.4 Rules and Itemsets

Sorting the rules generated for support = 0.3 and confidence = 0.5 by lift gives following top 20 rules:

```
     lhs                                    rhs          support   confidence lift     count
[1]  {ManageTab,ReEditProj,ReportsTab}   => {SendNow}    0.3038936 0.9204219  1.987432  960
[2]  {ManageTab,ProjPreview,ReEditProj}  => {SendNow}    0.3206711 0.9200727  1.986678 1013
[3]  {ManageTab,ReEditProj}              => {SendNow}    0.3469452 0.9087894  1.962314 1096
[4]  {ReEditProj,ReportsTab}             => {SendNow}    0.3108579 0.9042357  1.952482  982
[5]  {ManageTab,ProjPreview,ReportsTab}  => {SendNow}    0.3187718 0.8975045  1.937947 1007
[6]  {ManageTab,ReportsTab,SendNow}      => {ReEditProj} 0.3038936 0.8751139  1.918449  960
[7]  {ManageTab,ProjPreview,SendNow}     => {ReEditProj} 0.3206711 0.8747841  1.917726 1013
[8]  {ManageTab,ProjPreview,ReEditProj}  => {ReportsTab} 0.3089585 0.8864668  1.908895  976
[9]  {ManageTab,ProjPreview,ReportsTab}  => {ReEditProj} 0.3089585 0.8698752  1.906964  976
[10] {ReportsTab,SendNow}                => {ReEditProj} 0.3108579 0.8690265  1.905104  982
[11] {ManageTab,ProjPreview}             => {SendNow}    0.3665717 0.8812785  1.902911 1158
[12] {ProjPreview,ReportsTab}            => {SendNow}    0.3266857 0.8805461  1.901329 1032
[13] {ProjPreview,ReportsTab}            => {ReEditProj} 0.3194049 0.8609215  1.887336 1009
[14] {ManageTab,ReEditProj,SendNow}      => {ReportsTab} 0.3038936 0.8759124  1.886167  960
[15] {ManageTab,ProjPreview,SendNow}     => {ReportsTab} 0.3187718 0.8696028  1.872580 1007
[16] {ManageTab,ReEditProj}              => {ReportsTab} 0.3301678 0.8648425  1.862329 1043
[17] {ProjPreview,SendNow}               => {ReEditProj} 0.3374486 0.8494024  1.862083 1066
[18] {ManageTab,ProjPreview}             => {ReportsTab} 0.3551757 0.8538813  1.838726 1122
[19] {ManageTab,ProjPreview}             => {ReEditProj} 0.3485280 0.8378995  1.836867 1101
[20] {ProjPreview,ReEditProj}            => {SendNow}    0.3374486 0.8480509  1.831164 1066
```

The following frequent itemsets are identified:

```
      items                                                 support
[1]   {ManageTab}                                          0.5432099
[2]   {ProjPreview}                                        0.5428933
[3]   {SendNow,SimpleProjCreate}                           0.3222539
[4]   {ReEditProj,SimpleProjCreate}                        0.3114910
[5]   {ManageTab,SimpleProjCreate}                         0.3301678
[6]   {ProjPreview,SimpleProjCreate}                       0.3247863
[7]   {ReportsTab,SendNow}                                 0.3577081
[8]   {ReEditProj,ReportsTab}                              0.3437797
[9]   {ManageTab,ReportsTab}                               0.4283001
[10]  {ProjPreview,ReportsTab}                             0.3710035
[11]  {ReEditProj,SendNow}                                 0.3697373
[12]  {ManageTab,SendNow}                                  0.4169041
[13]  {ProjPreview,SendNow}                                0.3972776
[14]  {ManageTab,ReEditProj}                               0.3817664
[15]  {ProjPreview,ReEditProj}                             0.3979107
[16]  {ManageTab,ProjPreview}                              0.4159544
[17]  {ReEditProj,ReportsTab,SendNow}                      0.3108579
[18]  {ManageTab,ReportsTab,SendNow}                       0.3472618
[19]  {ProjPreview,ReportsTab,SendNow}                     0.3266857
[20]  {ManageTab,ReEditProj,ReportsTab}                    0.3301678
[21]  {ProjPreview,ReEditProj,ReportsTab}                  0.3194049
[22]  {ManageTab,ProjPreview,ReportsTab}                   0.3551757
[23]  {ManageTab,ReEditProj,SendNow}                       0.3469452
[24]  {ProjPreview,ReEditProj,SendNow}                     0.3374486
[25]  {ManageTab,ProjPreview,SendNow}                      0.3665717
[26]  {ManageTab,ProjPreview,ReEditProj}                   0.3485280
[27]  {ManageTab,ReEditProj,ReportsTab,SendNow}            0.3038936
[28]  {ManageTab,ProjPreview,ReportsTab,SendNow}           0.3187718
[29]  {ManageTab,ProjPreview,ReEditProj,ReportsTab}        0.3089585
[30]  {ManageTab,ProjPreview,ReEditProj,SendNow}           0.3206711
```

Based on the user data, following are maximal frequent itemsets:

```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen                    target   ext
         NA    0.1    1 none FALSE            TRUE       5     0.3      1     10 maximally frequent itemsets FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 947

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[112 item(s), 3159 transaction(s)] done [0.00s].
sorting and recoding items ... [7 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 done [0.00s].
filtering maximal item sets ... done [0.00s].
writing ... [9 set(s)] done [0.00s].
creating S4 object  ... done [0.00s].
> inspect(maximal_sets)
    items                                       support   count
[1] {TestSend}                                  0.3111744  983
[2] {SendNow,SimpleProjCreate}                  0.3222539 1018
[3] {ReEditProj,SimpleProjCreate}               0.3114910  984
[4] {ManageTab,SimpleProjCreate}                0.3301678 1043
[5] {ProjPreview,SimpleProjCreate}              0.3247863 1026
[6] {ManageTab,ReEditProj,ReportsTab,SendNow}   0.3038936  960
[7] {ManageTab,ProjPreview,ReportsTab,SendNow}  0.3187718 1007
[8] {ManageTab,ProjPreview,ReEditProj,ReportsTab} 0.3089585  976
[9] {ManageTab,ProjPreview,ReEditProj,SendNow}  0.3206711 1013
```

6. Session Analysis

6.1 Features

For the purpose of association mining of session data, the unique milestones for each user session from the 'rawdataDec15' are stored into the table 'SessionMilestone'. Since the 'rawdataDec15' can contain repeated milestones for sessions, it is crucial to extract only the distinct milestones for each user session.

Following is schema of the 'SessionMilestone' table for our analysis:

S.No.	Column Name	Description
1.	id	Unique identifier for each session. This is an engineered feature and is a combination of the distinct user_id and date to identify unique sessions. Note: Any number of times a user logs in the system during the same date is considered as a unique session.
2.	milestone_name	Unique milestone names on which the user clicked during the session

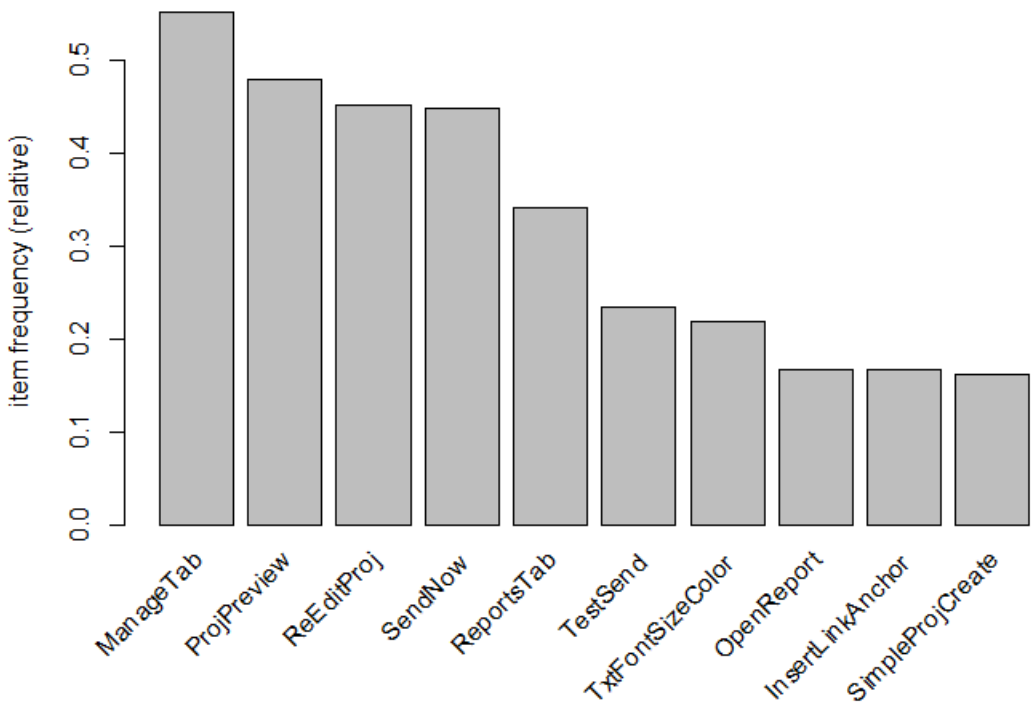
The data is generated by using a distinct combination of user_id, date and milestone_name fields and contains 1,73,082 records for the 24,713 unique sessions for all users.

6.2 Data Transformation

For running the Apriori algorithm, the unique milestones for each session should be transposed to concatenate all unique milestones for the session in single row. This step is completed in R and the following summary is generated in order to understand the data:

```
transactions as itemMatrix in sparse format with
 24713 rows (elements/itemsets/transactions) and
 112 columns (items) and a density of 0.06253288

most frequent items:
  ManageTab ProjPreview  ReEditProj     SendNow  ReportsTab     (Other)
      13636       11845       11173       11100        8421      116907

element (itemset/transaction) length distribution:
sizes
   1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17   18   19   20   21   22   23   24   25   26   27   28
3941 1761 1947 2511 2199 1952 1550 1378 1135  981  869  738  570  491  431  353  338  263  241  215  162  144   96  109   73   67   50   36
  29   30   31   32   33   34   35   36   37   38   39   40   41   42
  17   24   12   17    8    7    8    4    3    4    4    1    2    1

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   3.000   5.000   7.004  10.000  42.000

includes extended item information - examples:
            labels
1      ABSplitProj
2     ABSplitTools
3 AccountSettingsA
```

The following histogram showcases the relative frequency of the top 10 milestones recorded in the dataset:



6.3 Analysis

After transposing and formatting the data, we run the Apriori Algorithm with the following support and confidence values to generate the rules:

1. Support = 0.1, Confidence = 0.5
   Rules Generated: 93

```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target   ext
        0.5    0.1    1 none FALSE            TRUE       5     0.1      1     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 2471

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[112 item(s), 24713 transaction(s)] done [0.02s].
sorting and recoding items ... [23 item(s)] done [0.00s].
creating transaction tree ... done [0.01s].
checking subsets of size 1 2 3 4 done [0.01s].
writing ... [93 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

2. Support = 0.2, Confidence = 0.5
   Rules Generated: 20

```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target   ext
        0.5    0.1    1 none FALSE            TRUE       5     0.2      1     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 4942

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[112 item(s), 24713 transaction(s)] done [0.02s].
sorting and recoding items ... [7 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 done [0.00s].
writing ... [20 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

3. Support = 0.3, Confidence = 0.5
   Rules Generated: 5

```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target   ext
        0.5    0.1    1 none FALSE            TRUE       5     0.3      1     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 7413

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[112 item(s), 24713 transaction(s)] done [0.01s].
sorting and recoding items ... [5 item(s)] done [0.00s].
creating transaction tree ... done [0.01s].
checking subsets of size 1 2 done [0.00s].
writing ... [5 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

4. Support = 0.4, Confidence = 0.5
   Rules Generated: 3

```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target   ext
        0.5    0.1    1 none FALSE            TRUE       5     0.4      1     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 9885

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[112 item(s), 24713 transaction(s)] done [0.02s].
sorting and recoding items ... [4 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing ... [3 rule(s)] done [0.00s].
creating S4 object  ... done [0.01s].
```

Based on the above observations, following table is populated and it can be observed that support = 0.2 and confidence = 0.5 gives the optimum number of rules:

S.No.	Support	Confidence	Number of Rules
1.	0.1	0.5	93
2.	0.2	0.5	20
3.	0.3	0.5	5
4.	0.4	0.5	3

Note: Confidence value 0.5 is used as it is considered as an optimum value which provides reliable and good number of rules.

6.4 Rules and Itemsets

Sorting the rules generated for support = 0.2 and confidence = 0.5 by lift gives following rules:

```
     lhs                          rhs           support   confidence lift     count
[1]  {ManageTab,ReEditProj}    => {SendNow}     0.2413305 0.8135316  1.811244  5964
[2]  {ManageTab,ProjPreview}   => {SendNow}     0.2541982 0.7968037  1.774001  6282
[3]  {ProjPreview,SendNow}     => {ManageTab}   0.2541982 0.9042752  1.638850  6282
[4]  {SendNow}                 => {ManageTab}   0.4047263 0.9010811  1.633061 10002
[5]  {ManageTab}               => {SendNow}     0.4047263 0.7334996  1.633061 10002
[6]  {ReEditProj,SendNow}      => {ManageTab}   0.2413305 0.9000906  1.631266  5964
[7]  {ReEditProj}              => {SendNow}     0.2681180 0.5930368  1.320335  6626
[8]  {SendNow}                 => {ReEditProj}  0.2681180 0.5969369  1.320335  6626
[9]  {ManageTab,SendNow}       => {ReEditProj}  0.2413305 0.5962807  1.318884  5964
[10] {ManageTab,SendNow}       => {ProjPreview} 0.2541982 0.6280744  1.310393  6282
[11] {ProjPreview}             => {SendNow}     0.2811071 0.5864922  1.305764  6947
[12] {SendNow}                 => {ProjPreview} 0.2811071 0.6258559  1.305764  6947
[13] {ProjPreview}             => {ReEditProj}  0.2739854 0.5716336  1.264368  6771
[14] {ReEditProj}              => {ProjPreview} 0.2739854 0.6060145  1.264368  6771
[15] {ReportsTab}              => {ManageTab}   0.2369198 0.6952856  1.260090  5855
[16] {ProjPreview}             => {ManageTab}   0.3190224 0.6655973  1.206285  7884
[17] {ManageTab}               => {ProjPreview} 0.3190224 0.5781754  1.206285  7884
[18] {ReEditProj}              => {ManageTab}   0.2966455 0.6561353  1.189137  7331
[19] {ManageTab}               => {ReEditProj}  0.2966455 0.5376210  1.189137  7331
[20] {}                        => {ManageTab}   0.5517744 0.5517744  1.000000 13636
```

The following frequent itemsets are identified:

```
     items                              support
[1]  {ManageTab}                        0.5517744
[2]  {ManageTab,ReportsTab}             0.2369198
[3]  {ProjPreview,SendNow}              0.2811071
[4]  {ProjPreview,ReEditProj}           0.2739854
[5]  {ManageTab,ProjPreview}            0.3190224
[6]  {ReEditProj,SendNow}               0.2681180
[7]  {ManageTab,SendNow}                0.4047263
[8]  {ManageTab,ReEditProj}             0.2966455
[9]  {ManageTab,ProjPreview,SendNow}    0.2541982
[10] {ManageTab,ReEditProj,SendNow}     0.2413305
```

Based on the session data, following are maximal frequent itemsets:

```
     items                              support   count
[1]  {TestSend}                         0.2353822 5817
[2]  {TxtFontSizeColor}                 0.2192368 5418
[3]  {ManageTab,ReportsTab}             0.2369198 5855
[4]  {ProjPreview,ReEditProj}           0.2739854 6771
[5]  {ManageTab,ProjPreview,SendNow}    0.2541982 6282
[6]  {ManageTab,ReEditProj,SendNow}     0.2413305 5964
```

Appendix A

1. SQL – Users

```
CREATE TABLE UserMilestone AS
SELECT DISTINCT user_id, milestone_name FROM dataset03.rawdataDec15
```

2. SQL – Sessions

```
CREATE TABLE SessionMilestone AS
SELECT DISTINCT concat( date,' ',user_id) as id,milestone_name FROM dataset03.rawdataDec15
```

Appendix B – R Script

```
#install.packages("arules")
#install.packages("plyr",dependencies = TRUE)

library("arules")
library(plyr)
setwd("C:/CDA/Sem 2/Data Mining/Assignment3")

################################################

## User Level Analysis
df_user = read.csv("UserMilestone.csv",header=T)

df_user = ddply(df_user,c("ID"),function(dfl)paste(dfl$Milestone, collapse=","))
df_user$ID = NULL
write.table(df_user,"UserMilestone2.csv",quote=FALSE, row.names = FALSE,col.names = FALSE)
tr = read.transactions("UserMilestone2.csv",format="basket",sep=",")
summary(tr)

itemFrequencyPlot(tr,topN = 10)

#rules = apriori(tr,parameter = list(supp = 0.2, conf=0.5)) #661 rules
#rules = apriori(tr,parameter = list(supp = 0.3, conf=0.5)) #76 rules
#rules = apriori(tr,parameter = list(supp = 0.4, conf=0.5)) #8 rules
#rules = apriori(tr,parameter = list(supp = 0.5, conf=0.5)) #2 rules

#Generating optimal number of rules
rules = apriori(tr,parameter = list(supp = 0.3, conf=0.5))

inspect(sort(rules,by="lift")[1:20])
itemsets = unique(generatingItemsets(rules))

inspect(itemsets)

maximal_sets<- apriori(tr, parameter= list(supp=0.3, conf=0.5, target="maximally frequent itemsets"))
inspect(maximal_sets)

#########################################################

## Session Level Analysis
df_session = read.csv("SessionMilestone.csv",header=T)
df_session = ddply(df_session,c("ID"),function(dfl)paste(dfl$Milestone, collapse=","))
df_session$ID = NULL

write.table(df_session,"SessionMilestone2.csv",quote=FALSE, row.names = FALSE,col.names = FALSE)
tr_session = read.transactions("SessionMilestone2.csv",format="basket",sep=",")
summary(tr_session)
```

```
itemFrequencyPlot(tr_session,topN = 10)

#rules_session = apriori(tr_session,parameter = list(supp = 0.1, conf=0.5)) # 93 rules
#rules_session = apriori(tr_session,parameter = list(supp = 0.2, conf=0.5)) # 20 rules
#rules_session = apriori(tr_session,parameter = list(supp = 0.3, conf=0.5)) # 5 rules
#rules_session = apriori(tr_session,parameter = list(supp = 0.4, conf=0.5)) # 4 rules


rules_session = apriori(tr_session,parameter = list(supp = 0.2, conf=0.5))

inspect(sort(rules_session,by="lift")[1:20])
itemsets_session = unique(generatingItemsets(rules_session))

inspect(itemsets_session)

maximal_sets_session<- apriori(tr_session, parameter= list(supp=0.2, conf=0.5, target="maximally frequent
itemsets"))
inspect(maximal_sets_session)
```