به نام خدا

تمرین سری سوم

ابتدا الگوریتم Apriori را روی basket اجرا کردیم . روش انجام کار به این صورت بود که ابتدا باید از قسمت processesو سپش open file داده های basket را وارد میکردیم و سپس از قسمت choose و filter و سپس supervised گزینه ی discretize را انتخاب میکردیم تا به بازه گسسته تبدیل کنیم و در نتیجه با رفتن به نوار associate و انتخاب الگوریتم apriori خروجی زیر را مشاهده میکنیم:

*داده های basket همه nominal هستند.

#basket

```
=== Run information ===


Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:     MarketBasket
Instances:    1000
Attributes:   11
              fruitveg
              freshmeat
              dairy
              cannedveg
              cannedmeat
              frozenmeal
              beer
              wine
              softdrink
              fish
              confectionery
=== Associator model (full training set) ===
Apriori
=======

Minimum support: 0.1 (100 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 22

Size of set of large itemsets L(2): 170

Size of set of large itemsets L(3): 601

Size of set of large itemsets L(4): 920

Size of set of large itemsets L(5): 967

Size of set of large itemsets L(6): 480

Size of set of large itemsets L(7): 330

Size of set of large itemsets L(8): 165

Size of set of large itemsets L(9): 15

Best rules found:
 1. cannedveg=T beer=T fish=F confectionery=T 118 ==> wine=T 109    <conf:(0.92)> lift:(1.3) lev:(0.02) [24] conv:(3.39)

 2. fruitveg=T freshmeat=T cannedveg=F softdrink=T 147 ==> dairy=T 135    <conf:(0.92)> lift:(1.12) lev:(0.01) [14] conv:(2)

 3. freshmeat=T wine=F confectionery=T 117 ==> dairy=T 107    <conf:(0.91)> lift:(1.11) lev:(0.01) [10] conv:(1.88)

 4. fruitveg=T freshmeat=T cannedveg=F softdrink=T confectionery=T 113 ==> dairy=T 103    <conf:(0.91)> lift:(1.11) lev:(0.01) [10] conv:(1.82)

 5. fruitveg=T freshmeat=T cannedveg=F cannedmeat=T softdrink=T 112 ==> dairy=T 102    <conf:(0.91)> lift:(1.11) lev:(0.01) [9] conv:(1.8)

 6. fruitveg=T cannedveg=F softdrink=T confectionery=T 128 ==> dairy=T 116    <conf:(0.91)> lift:(1.1) lev:(0.01) [10] conv:(1.74)

 7. fruitveg=T freshmeat=T cannedveg=F softdrink=T fish=T 117 ==> dairy=T 106    <conf:(0.91)> lift:(1.1) lev:(0.01) [9] conv:(1.73)

 8. freshmeat=T cannedveg=F frozenmeal=F softdrink=T 114 ==> dairy=T 103    <conf:(0.9)> lift:(1.1) lev:(0.01) [9] conv:(1.68)

 9. freshmeat=T cannedveg=T fish=F confectionery=T 124 ==> wine=T 112    <conf:(0.9)> lift:(1.27) lev:(0.02) [23] conv:(2.74)

10. cannedveg=T fish=F confectionery=T 144 ==> wine=T 130    <conf:(0.9)> lift:(1.27) lev:(0.03) [27] conv:(2.76)
```

اگر در همین الگوریتم min support به صورت Minimum metric <confidence>: 0.5تغییر بدهیم و instance 650 داشته باشیم خروجی زیر مشاهده میشود:

```
Apriori

=======

Minimum support: 0.65 (650 instances)

Minimum metric <confidence>: 0.5

Number of cycles performed: 7


Generated sets of large itemsets:


Size of set of large itemsets L(1): 11
Size of set of large itemsets L(2): 6

Best rules found:

softdrink=T 816 ==> freshmeat=T 675    <conf:(0.83)> lift:(1.01) lev:(0.01) [8] conv:(1.05) .۱
freshmeat=T 817 ==> softdrink=T 675    <conf:(0.83)> lift:(1.01) lev:(0.01) [8] conv:(1.05) .۲
softdrink=T 816 ==> dairy=T 674    <conf:(0.83)> lift:(1) lev:(0) [2] conv:(1.01) .۳
freshmeat=T 817 ==> dairy=T 673    <conf:(0.82)> lift:(1) lev:(0) [0] conv:(1) .۴
cannedmeat=T 796 ==> freshmeat=T 654    <conf:(0.82)> lift:(1.01) lev:(0) [3] conv:(1.02) .۵
cannedmeat=T 796 ==> softdrink=T 654    <conf:(0.82)> lift:(1.01) lev:(0) [4] conv:(1.02) .۶
dairy=T 823 ==> softdrink=T 674    <conf:(0.82)> lift:(1) lev:(0) [2] conv:(1.01) .۷
dairy=T 823 ==> freshmeat=T 673    <conf:(0.82)> lift:(1) lev:(0) [0] conv:(1) .۸
cannedmeat=T 796 ==> dairy=T 650    <conf:(0.82)> lift:(0.99) lev:(-0.01) [-5] conv:(0.96) .۹
softdrink=T 816 ==> cannedmeat=T 654    <conf:(0.8)> lift:(1.01) lev:(0) [4] conv:(1.02) .۱۰
```

اگر در همین الگوریتم min support را از ۰,۱ به 0.8 تغییر بدهیم و instance 600 و دلتا را به ۰,۰۸ تغییر بدهیم خروجی زیر مشاهده میشود:

```
priori
=======
Minimum support: 0.6 (600 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 5

Generated sets of large itemsets:

Size of set of large itemsets L(1): 11

Size of set of large itemsets L(2): 7
```

```
Best rules found:

۱. confectionery=T 724 ==> dairy=T 603    <conf:(0.83)> lift:(1.01) lev:(0.01) [7] conv:(1.05)
۲. softdrink=T 816 ==> freshmeat=T 675    <conf:(0.83)> lift:(1.01) lev:(0.01) [8] conv:(1.05)
۳. freshmeat=T 817 ==> softdrink=T 675    <conf:(0.83)> lift:(1.01) lev:(0.01) [8] conv:(1.05)
۴. softdrink=T 816 ==> dairy=T 674    <conf:(0.83)> lift:(1) lev:(0) [2] conv:(1.01)
۵. freshmeat=T 817 ==> dairy=T 673    <conf:(0.82)> lift:(1) lev:(0) [0] conv:(1)
۶. cannedmeat=T 796 ==> freshmeat=T 654    <conf:(0.82)> lift:(1.01) lev:(0) [3] conv:(1.02)
۷. cannedmeat=T 796 ==> softdrink=T 654    <conf:(0.82)> lift:(1.01) lev:(0) [4] conv:(1.02)
۸. dairy=T 823 ==> softdrink=T 674    <conf:(0.82)> lift:(1) lev:(0) [2] conv:(1.01)
۹. dairy=T 823 ==> freshmeat=T 673    <conf:(0.82)> lift:(1) lev:(0) [0] conv:(1)
۱۰. cannedmeat=T 796 ==> dairy=T 650    <conf:(0.82)> lift:(0.99) lev:(-0.01) [-5] conv:(0.96)
```

سپس همین دیتاست basket را با FPGrowth اجرا میکنیم به این صورت که ابتدا از گزینه open file فایل را لود میکنیم سپس باید از مسیر مقبل فیلتر nominal to binary را انتخاب کنیم:

Choose>filters>supervised >attribute>nominal to binary

بعد از انجام این کار apply را میزنیم و سپس فیلتر discretize را از همان مسیر انتخاب میکنیم و خروجی را مشاهده میکنیم که به صورت زیر است:

#basket

```
=== Run information ===

Scheme:       weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Relation:     MarketBasket-weka.filters.supervised.attribute.NominalToBinary-weka.filters.supervised.attribute.Discretize-Rfirst-last-precision6
Instances:    1000
Attributes:   11
              fruitveg=F
              freshmeat=F
              dairy=F
              cannedveg=F
```

```
        cannedmeat=F
        frozenmeal=F
        beer=F
        wine=F
        softdrink=F
        fish=F
        confectionery
=== Associator model (full training set) ===


FPGrowth found 18660 rules (displaying top 10)


 1. [softdrink=F='All']: 1000 ==> [fruitveg=F='All']: 1000   <conf:(1)> lift:(1) lev:(0) conv:(0)
 2. [fruitveg=F='All']: 1000 ==> [softdrink=F='All']: 1000   <conf:(1)> lift:(1) lev:(0) conv:(0)
 3. [softdrink=F='All']: 1000 ==> [frozenmeal=F='All']: 1000   <conf:(1)> lift:(1) lev:(0) conv:(0)
 4. [frozenmeal=F='All']: 1000 ==> [softdrink=F='All']: 1000   <conf:(1)> lift:(1) lev:(0) conv:(0)
 5. [softdrink=F='All']: 1000 ==> [freshmeat=F='All']: 1000   <conf:(1)> lift:(1) lev:(0) conv:(0)
 6. [freshmeat=F='All']: 1000 ==> [softdrink=F='All']: 1000   <conf:(1)> lift:(1) lev:(0) conv:(0)
 7. [softdrink=F='All']: 1000 ==> [fish=F='All']: 1000   <conf:(1)> lift:(1) lev:(0) conv:(0)
 8. [fish=F='All']: 1000 ==> [softdrink=F='All']: 1000   <conf:(1)> lift:(1) lev:(0) conv:(0)
 9. [softdrink=F='All']: 1000 ==> [dairy=F='All']: 1000   <conf:(1)> lift:(1) lev:(0) conv:(0)
10. [dairy=F='All']: 1000 ==> [softdrink=F='All']: 1000   <conf:(1)> lift:(1) lev:(0) conv:(0)
```

محاسبه انحراف معیارو میانگین و مینیم و ماکزیمم که هر داده ای عددی باشد قابل مشاهده است و در دیتاست churn دیده میشود که چند تا را در زیر می اورم:

Account_Length :

Minimum	1
Maximum	243

Mean 101.065

StdDev 39.822

Area_Code’

Minimum 408

Maximum 510

Mean 437.182

StdDev 42.371

الان به سراغ دیتاست churn می‌رویم و طبق همان basket ان را اجرا می‌کنیم و خروجی زیر را مشاهده می‌کنیم:

#churn

```
=== Run information ===

Scheme:       weka.associations.Apriori -R -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Relation:     churn-weka.filters.supervised.attribute.Discretize-Rfirst-last-precision6-
weka.filters.unsupervised.attribute.Remove-weka.filters.unsupervised.attribute.Remove-
weka.filters.unsupervised.attribute.Remove-weka.filters.unsupervised.attribute.Remove-R4

Instances:    3333

Attributes:   20

              State

              Account_Length

              Area_Code’

              Int’l_Plan’

              VMail_Plan’
```

```
        VMail_Message’
        Day_Mins’
        Day_Calls’
        Day_Charge’
        Eve_Mins’
        Eve_Calls’
        Eve_Charge’
        Night_Mins’
        Night_Calls’
        Night_Charge’
        Intl_Mins’
        Intl_Calls’
        Intl_Charge’
        CustServ_Calls’
        Churn?’
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.95 (3166 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 1

Generated sets of large itemsets:

Size of set of large itemsets L(1): 7
Size of set of large itemsets L(2): 21

Size of set of large itemsets L(3): 35

Size of set of large itemsets L(4): 35

Size of set of large itemsets L(5): 21

Size of set of large itemsets L(6): 7

Size of set of large itemsets L(7): 1

Best rules found:

 1. Area_Code’='All' 3333 ==> Account_Length='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 2. Account_Length='All' 3333 ==> Area_Code’='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 3. Day_Calls’='All' 3333 ==> Account_Length='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 4. Account_Length='All' 3333 ==> Day_Calls’='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 5. Eve_Calls’='All' 3333 ==> Account_Length='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 6. Account_Length='All' 3333 ==> Eve_Calls’='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 7. Night_Mins’='All' 3333 ==> Account_Length='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 8. Account_Length='All' 3333 ==> Night_Mins’='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 9. Night_Calls’='All' 3333 ==> Account_Length='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. Account_Length='All' 3333 ==> Night_Calls’='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
```

که با کاهش ویژگی ها از ۲۰ تا به ۱۰تا خروجی زیر مشاهده شد:

```
=== Run information ===
Scheme:       weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Relation:     churn-weka.filters.unsupervised.attribute.Remove-R4-13-
weka.filters.unsupervised.attribute.Remove-R1-weka.filters.supervised.attribute.Discretize-Rfirst-last-
precision6

Instances:    3333

Attributes:   10

              Account_Length

              Area_Code’

              Night_Mins’

              Night_Calls’

              Night_Charge’

              Intl_Mins’

              Intl_Calls’

              Intl_Charge’

              CustServ_Calls’

              Churn?’

=== Associator model (full training set) ===



Apriori

=======


Minimum support: 0.95 (3166 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 1


Generated sets of large itemsets:


Size of set of large itemsets L(1): 5
Size of set of large itemsets L(2): 10


Size of set of large itemsets L(3): 10


Size of set of large itemsets L(4): 5


Size of set of large itemsets L(5): 1


Best rules found:


 1. Area_Code’='All' 3333 ==> Account_Length='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 2. Account_Length='All' 3333 ==> Area_Code’='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 3. Night_Mins’='All' 3333 ==> Account_Length='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 4. Account_Length='All' 3333 ==> Night_Mins’='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 5. Night_Calls’='All' 3333 ==> Account_Length='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 6. Account_Length='All' 3333 ==> Night_Calls’='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 7. Night_Charge’='All' 3333 ==> Account_Length='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 8. Account_Length='All' 3333 ==> Night_Charge’='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
 9. Night_Mins’='All' 3333 ==> Area_Code’='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. Area_Code’='All' 3333 ==> Night_Mins’='All' 3333    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
```

اجرای churrn با الکوریتم FPGrowth :

```
Run information ===


Scheme:       weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
```

```
Relation:     churn-weka.filters.unsupervised.attribute.Remove-R1,4,15-17-
weka.filters.unsupervised.attribute.Remove-R12-weka.filters.supervised.attribute.NominalToBinary-
weka.filters.supervised.attribute.Discretize-Rfirst-last-precision6

Instances:    240

Attributes:   15

AccountLength

AreaCode

IntlPlan=no

VMailPlan=no

VMailMessage

DayMins

DayCalls

DayCharge

EveMins

EveCalls

EveCharge

IntlCalls

IntlCharge

CustServCalls

Churn

Associator model (full training set) === ===


FPGrowth found 173052 rules (displaying top 10)


[VMailPlan=no='All']: 240 ==> [VMailMessage='All']: 240   <conf:(1)> lift:(1) lev:(0) conv:(0)  ۱.

[VMailMessage='All']: 240 ==> [VMailPlan=no='All']: 240   <conf:(1)> lift:(1) lev:(0) conv:(0)  ۲.

[VMailPlan=no='All']: 240 ==> [IntlPlan=no='All']: 240   <conf:(1)> lift:(1) lev:(0) conv:(0)  ۳.

[IntlPlan=no='All']: 240 ==> [VMailPlan=no='All']: 240   <conf:(1)> lift:(1) lev:(0) conv:(0)  ۴.

[VMailPlan=no='All']: 240 ==> [IntlCharge='All']: 240   <conf:(1)> lift:(1) lev:(0) conv:(0)  ۵.
```

۶. [IntlCharge='All']: 240 ==> [VMailPlan=no='All']: 240   <conf:(1)> lift:(1) lev:(0) conv:(0)

۷. [VMailPlan=no='All']: 240 ==> [IntlCalls='All']: 240   <conf:(1)> lift:(1) lev:(0) conv:(0)

۸. [IntlCalls='All']: 240 ==> [VMailPlan=no='All']: 240   <conf:(1)> lift:(1) lev:(0) conv:(0)

۹. [VMailPlan=no='All']: 240 ==> [EveMins='All']: 240   <conf:(1)> lift:(1) lev:(0) conv:(0)

۱۰. [EveMins='All']: 240 ==> [VMailPlan=no='All']: 240   <conf:(1)> lift:(1) lev:(0) conv:(0)

محاسبه‌ی میانه :

ابتدا فایل arff را تبدیل به csv نمودم و سپس در این فایل میانه‌های داده‌های عددی churn را محاسبه کردم که در basket قابل محاسبه نبود و این اعداد برای churn را محاسبه کردم:

Median c	101
median D	415
median H	0
median I	179.4
median J	101
median K	30.5
median L	201.4
median m	100
median N	17.12
Median O	201.2
median P	100
median Q	9.05
median R	10.3
median S	4
median T	2.78
median U	1