CS 634 Data Mining Midterm Project

APRIORI ALGORITHM

Submitted by: Sowmya Kothapalli

UCID: sk272

Email: sk272@njit.edu

Date: 03/28/2021

Programming Language: Python

How to run the file:

python Apriori.py datasetname {minimum support (in percentage)} {minimum confidence (in percentage)}

E.g.: python3 Apriori.py bestbuy.csv 50 70

Input: For dataset Best buy

Output: For dataset Best buy

```
Est. C:\Windows\System32\cmd.exe
Total number of transactions: 20
Frequent sets for k= 1
                                | Frequency |
         Printer
      Flash Drive
   Microsoft Office
         Speakers
        Anti-Virus
      Lab Top
Lab Top Case
                                           12
14
 requent set for k= 2
                                                                  | Frequency |
  ('Printer', 'Flash Drive')
('Flash Drive', 'Microsoft Office')
('Flash Drive', 'Anti-Virus')
('Anti-Virus', 'Lab Top')
('Anti-Virus', 'Lab Top Case')
('Lab Top', 'Lab Top Case')
                                                                            10
   ====== List of all frequent item sets and support levels ========
                  Frequent Item sets
                                                                  | Support in (%) |
  ('Printer', 'Flash Drive')
('Flash Drive', 'Microsoft Office')
('Flash Drive', 'Anti-Virus')
('Anti-Virus', 'Lab Top')
('Anti-Virus', 'Lab Top Case')
('Lab Top', 'Lab Top Case')
                                                                                60
```

Selected sets	Predecessor	Result	Support in (%)	Confidence in (%)
Selected Sets	Fredecessor	Result	3upport in (%)	confidence in (%)
('Printer', 'Flash Drive')	('Printer',)	('Flash Drive',)	50	100
('Flash Drive', 'Printer')	('Flash Drive',)	('Printer',)	50	77
('Flash Drive', 'Microsoft Office')	('Flash Drive',)	('Microsoft Office',)	55	85
('Microsoft Office', 'Flash Drive')	('Microsoft Office',)	('Flash Drive',)	55	100
('Flash Drive', 'Anti-Virus')	('Flash Drive',)	('Anti-Virus',)	50	77
('Anti-Virus', 'Flash Drive')	('Anti-Virus',)	('Flash Drive',)	50	71
('Anti-Virus', 'Lab Top')	('Anti-Virus',)	('Lab Top',)	50	71
('Lab Top', 'Anti-Virus')	('Lab Top',)	('Anti-Virus',)	50	83
('Anti-Virus', 'Lab Top Case')	('Anti-Virus',)	('Lab Top Case',)	60	86
('Lab Top Case', 'Anti-Virus')	('Lab Top Case',)	('Anti-Virus',)	60	86
('Lab Top', 'Lab Top Case')	('Lab Top',)	('Lab Top Case',)	50	83
('Lab Top Case', 'Lab Top')	('Lab Top Case',)	('Lab Top',)	50	71

::\Users\sowmy\OneDrive\Desktop\NJIT\DataMining\Project\AprioriAlgorithm>

Input: For dataset Nike

Output: For dataset Nike

```
ः C:\Windows\System32\cmd.exe
Total number of transactions: 20
requent sets for k= 1
             Frequency
  Item sets
 Running Shoe
                     14
    Socks
 Sweatshirts
  Rash Guard
                     12
 requent set for k= 2
                           | Frequency |
         Item sets
 ('Socks', 'Sweatshirts') |
  ====== List of all frequent item sets and support levels ========
                           | Support in (%) |
    Frequent Item sets
 ('Socks', 'Sweatshirts') |
                                    60
  ====== Association and Confidence levels =========
                                                                     Support in (%) | Confidence in (%) |
      Selected sets
                               Predecessor
                                                      Result
 ('Socks', 'Sweatshirts') |
('Sweatshirts', 'Socks') |
                                 ('Socks',)
                                                 ('Sweatshirts',)
                             ('Sweatshirts',)
 :\Users\sowmy\OneDrive\Desktop\NJIT\DataMining\Project\AprioriAlgorithm>_
```

Input: For dataset Kmart

Output: For dataset Kmart

```
C:\Windows\System32\cmd.exe
Frequent sets for k= 1
                           Frequency
       Item sets
 Decorative Pillows
                                   10
         Shams
     Kids Bedding
      Bed Skirts
          Sheets
                                   10
Frequent set for k=2
               Item sets
                                             Frequency
 ('Shams', 'Kids Bedding')
('Shams', 'Bed Skirts')
('Kids Bedding', 'Bed Skirts')
('Kids Bedding', 'Sheets')
('Bed Skirts', 'Sheets')
                                                    10
                                                    10
requent set for k=3
                                                         | Frequency |
                      Item sets
 ('Kids Bedding', 'Bed Skirts', 'Sheets') |
```

Input: For dataset Generic

```
C:\Windows\System32\cmd.exe

C:\Users\sowmy\OneDrive\Desktop\NJIT\DataMining\Project\AprioriAlgorithm>python3 Apriori.py generic.csv 50 75
file name generic.csv
Support 50
Confidence 75
======= Input Transactions ========

A, B, C
A, B, C
A, B, C
A, B, C
A, B, C, D
A, B, C, D, E
A, B, D, E
A, D, E
A, D, E
A, E
A, E
A, C, E
A
```

Output: For dataset Generic

```
C:\Windows\System32\cmd.exe
requent sets for k= 1
 Item sets | Frequency |
     Α
requent set for k=2
 Item sets | Frequency
====== List of all frequent item sets and support levels ========
 Frequent Item sets | Support in (%) |
     ('A', 'C')
('A', 'E')
                             64
   ----- Association and Confidence levels ------
 Selected sets | Predecessor | Result | Support in (%) | Confidence in (%) |
                     ('C',)
('E',)
   ('C', 'A')
('E', 'A')
                                ('A',)
('A',)
                                                64
                                                                    100
                                                                    100
:\Users\sowmy\OneDrive\Desktop\NJIT\DataMining\Project\AprioriAlgorithm>_
```

Input: For dataset Amazon

Output: For dataset Amazon

```
Beginner's Guide, Java: The Complete Reference, Java For Dummies, Android Programming: The Big Nerd Ranch Beginner's Guide, Java: The Complete Reference, Java For Dummies, Android Programming: The Big Nerd Ranch ead First Java 2nd Edition, Beginning Programming with Java, Java 8 Pocket Guide ndroid Programming: The Big Nerd Ranch, Head First Java 2nd Edition
Beginner's Guide, Java: The Complete Reference, Java For Dummies
Total number of transactions: 20
requent sets for k= 1
                                                       | Frequency |
               A Beginner's Guide
       Java: The Complete Reference
Java For Dummies
Android Programming: The Big Nerd Ranch
requent set for k= 2
                                                                        | Frequency |
                           Item sets
 ====== List of all frequent item sets and support levels =====
                    Frequent Item sets | Support in (%) |
('Java: The Complete Reference', 'Java For Dummies') | 50
 ====== Association and Confidence levels =======
                                                                                                                                                         | Support in (%) | Confidence in (%) |
                       Selected sets
                                                                                       Predecessor
 ('Java: The Complete Reference', 'Java For Dummies') | ('Java: The Complete Reference',) | ('Java For Dummies',) |
:\Users\sowmy\OneDrive\Desktop\NJIT\DataMining\Project\AprioriAlgorithm>_
```

Source Code:

Apriori.py

```
import sys
from itertools import combinations
from typing import Dict
from prettytable import PrettyTable

filename = sys.argv[1]
fileobject = open(filename, "r")
lines = fileobject.readlines()
total_no_trans = 0
support_of_all_item_set = {}
min_supp = int(sys.argv[2])
min_conf = int(sys.argv[3])

all_trans_support = {} # type: Dict[str, int]
all_selected_trans_support = {} # type: Dict[str, int]
all_selected_trans = []

print("file_name ", filename)
```

```
def get frequent set(selected set, rejected set, all trans, n):
   global all selected trans
           selected set.append(key)
   return selected set, rejected set
```

```
all trans support.update(c1)
print("----
print()
print("total number of transactions: ", total no trans)
        rejected set.append(key)
print(t)
```

CS 634 Data Mining Midterm Project