# PROJECT REPORT

# APRIORI ALGORITHM



Course: Data Mining

Prepared By: Vinay Basavaraja

GitHub link: <a href="https://github.com/vinayb004/Apriori\_Algorithm">https://github.com/vinayb004/Apriori\_Algorithm</a>

# Introduction

Apriori Algorithm is the classical algorithm which is used in data mining. It is used for mining frequent item sets and relevant association rules for those item sets. It is devised to operate on a database containing more transactions, for instance, items brought by customers in a store.

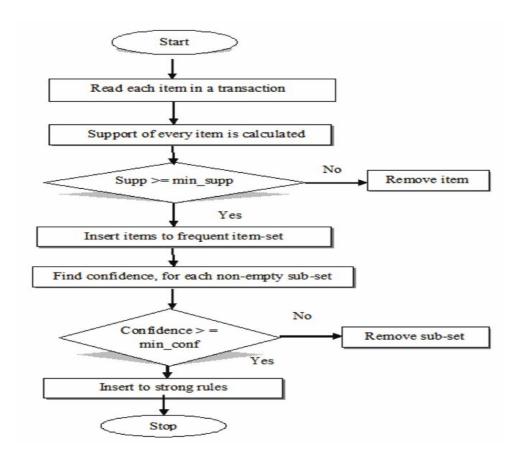
It is important for effective Market Basket Analysis and it helps the users in purchasing their products with more ease which provides more sales in the markets. It has been even used in the healthcare industry for the finding of adverse drug reactions. It produces association rules that shows what all combinations of medications and patient characteristics lead to ADRs.

# **Implementation**

An implementation of Apriori Algorithm to generate Association rules with user input values for support and confidence. The support and confidence values are given in decimal values. It gives associations rules for different datasets: Amazon, BestBuy, KMART, Nike, Grocery.

High values of support and confidence gives an out of index value. Select a numerical value (input) for obtaining the respective data set's association rules. The csv files are uploaded too. The screenshots are attached in the file that is submitted. Check the file for report and for basic software and hardware used to implement the project, the Project is not software or hardware specific.

# **Implementation Architecture**



# **Software Configuration**

- 1. Environment: Anaconda-Navigator
- 2. IDE: Jupiter Notebook 6.1.4
- 3. Python version: 3.8
- 4.Libraries csv (for reader)

# **Source of Data:**

The data used are the ones provided in the dataset examples in the midterm folder. I have also created a custom dataset for grocery store. All the data from these files are written into csv files. The datasets used are:

- 1. Amazon
- 2. BestBuy
- 3. KMART
- 4. Nike
- 5. Custom/Grocery

# **Hardware Configuration**

1. Laptop: Apple MacBook Pro

2. Processor: M1 / Apple Silicon chip

#### Code

Coding is complete and these are the few functions used in the code

1. append\_data: The purpose is to load the data set and return the dataset that contains a list of transactions and further, each transaction contains several items in it.

2. gen\_Candidate1: Generate and return a set(dataset) which contains all frequent candidate 1-itemsets.

3. apriori : To check whether a frequent candidate k-itemset satisfy Apriori property and return true/false accordingly.

4. gen\_Cank : Generate and return candidate k, a set which contains all frequent candidate k-itemsets.

5. gen\_Lisk : Generate and return a list/set which contains all frequent k-itemsets.

6. gen\_L : Generate all frequent itemsets by the value of minimum support's input.

7. association\_rules: Generate and return association rules.

8. Main: Execution of functions/code and take user input for minimum support and minimum confidence.

9. Encoding used: windows-1252

#### **Difficulties faced:**

To implement the understood logic of the functions as this was my first python project.

#### **Screenshots:**

Screenshots are attached for different dataset and different user input values for support and confidence.

#### **CSV** files.:

CSV files are attached in the zip file.

**GitHub Link:** <a href="https://github.com/vinayb004/Apriori\_Algorithm">https://github.com/vinayb004/Apriori\_Algorithm</a>

## **Source Code:**

from csv import reader

```
def append_data():
```

#The purpose is to load the data set and return the dataset that contains a list of transactions and further, each transaction contains several items in it.

print("Hello, welcome to association rule mining\nPlease choose the dataset for which you would like to get the Association rules")

print("Please type \n1 for Amazon\n2 for BestBuy\n3 for KMART\n4 for Nike\n5 for Custom")

```
while True:
  datachoice = input()
  if (datachoice == '1'):
    data ='/Users/VINAY/midterm_datamining_vinay/amazon.csv'
    print('You selected Amazon Dataset')
    break
  elif(datachoice == '2'):
    data = '/Users/VINAY/midterm_datamining_vinay/bestbuy.csv'
    print('You selected BestBuy Dataset')
    break
  elif(datachoice == '3'):
    data = '/Users/VINAY/midterm_datamining_vinay/kmart.csv'
    print('You selected KMART Dataset')
    break
  elif(datachoice == '4'):
    data = '/Users/VINAY/midterm_datamining_vinay/nike.csv'
    print('You selected Nike Dataset')
    break
  elif(datachoice == '5'):
    data = '/Users/VINAY/midterm_datamining_vinay/grocery.csv'
```

```
print('You selected Custom Dataset')
       break
     else:
       print("Please select a valid choice from the above list")
  with open(data, 'r',encoding="windows-1252") as read_obj:
     # pass the file object to reader() to get the reader object
     csv_reader = reader(read_obj)
     # Pass reader object to list() to get a list of lists
     dataset = list(csv_reader)
     for items in dataset:
       for j in range(0, len(items)):
          for items1 in items:
            if items1 == "":
               items.remove("")
  return dataset
def gen_Candidate1(dataset):
  # Generate and return a set(dataset) which contains all frequent candidate 1-itemsets.
  Can1 = set()
  for t in dataset:
     for item in t:
       item_set = frozenset([item])
       Can1.add(item_set)
  return Can1
```

```
def apriori(Cank_item, Lisk1):

#To check whether a frequent candidate k-itemset satisfy Apriori property and return true/false accordingly.
```

```
for item in Cank_item:
     sub_Cank = Cank_item - frozenset([item])
     if sub_Cank not in Lisk1:
       return False
  return True
def gen_Cank(Lisk1, k):
  #Generate and return candidate k, a set which contains all frequent candidate k-itemsets.
  Cank = set()
  length\_Lisk1 = len(Lisk1)
  listof_Lisk1 = list(Lisk1)
  for i in range(length_Lisk1):
     for j in range(1, length_Lisk1):
       list1 = list(listof_Lisk1[i])
       list2 = list(listof_Lisk1[j])
       list1.sort()
       list2.sort()
       if list1[0:k-2] == list2[0:k-2]:
          Cank_item = listof_Lisk1[i] | listof_Lisk1[j]
          if apriori(Cank_item, Lisk1):
```

return Cank

Cank.add(Cank\_item)

```
def gen_Lisk(dataset, Cank, minsup, sup_data):
  #Generate and return a list/set which contains all frequent k-itemsets.
  Lisk = set()
  item_count = {}
  for t in dataset:
     for item in Cank:
       if item.issubset(t):
          if item not in item_count:
            item_count[item] = 1
          else:
            item_count[item] += 1
  t_num = float(len(dataset))
  for item in item_count:
    if (item_count[item] / t_num) >= minsup:
       Lisk.add(item)
       sup_data[item] = item_count[item] / t_num
  return Lisk
def gen_L(dataset, k, minsup):
  #Generate all frequent itemsets by the value of minimum support's input.
  sup\_data = \{\}
  Can1 = gen_Candidate1(dataset)
  List1 = gen_Lisk(dataset, Can1, minsup, sup_data)
  Lisksub1 = List1.copy()
  List = []
  List.append(Lisksub1)
  for i in range(2, k+1):
```

```
Ci = gen_Cank(Lisksub1, i)
     Li = gen_Lisk(dataset, Ci, minsup, sup_data)
     Lisksub1 = Li.copy()
     List.append(Lisksub1)
  return List, sup_data
def association_rules(List, sup_data, minconf):
  #Generate and return association rules.
  asso_rules = []
  sublist = []
  for i in range(0, len(List)):
     for freq_set in List[i]:
       for sub_set in sublist:
          if sub_set.issubset(freq_set):
            conf = sup_data[freq_set] / sup_data[freq_set - sub_set]
            rule = (freq_set - sub_set, sub_set, conf)
            if conf >= minconf and rule not in asso_rules:
               # print freq_set-sub_set, " => ", sub_set, "conf: ", conf
               asso_rules.append(rule)
       sublist.append(freq_set)
  return asso_rules
if __name__ == "__main__":
  #print("Please enter the min_support")
  #inp = input("Enter the minimum support: ")
```

```
dataset = append_data()
  List, sup_data = gen_L(dataset, k=3, minsup = float(input("Enter the minimum support: ")))
  rules_list = association_rules(List, sup_data, minconf= float(input("Enter the minimum
Confidence: ")))
  for Lisk in List:
     print("-*"*40)
     print("frequent " + str(len(list(Lisk)[0])) + "-itemsets\t\tsupport")
     print("-*"*40)
    for freq_set in Lisk:
       print(freq_set, sup_data[freq_set])
  print("\n")
  print("#-"*40)
  print("Association Rules with Confidence")
  print("#-"*40)
  for item in rules list:
     print(item[0], "=>", item[1], "confidence: ", item[2])
```

# **Screenshots:**

# 1.Amazon

```
Frequent 1-itemets support

frozenets(('Java For Dumies') 0.6

frozenets(('Java For Dumies') 0.7

frozenets(('Java For Du
```

# 2.BestBuy

```
C:\Users\VINAY\midterm_datamining_vinay>python midterm_apriori.py
lello, welcome to association rule mining
llease choose the dataset for which you would like to get the Association rules
lease type
I for Amazon
I for SetBuy
I for KMART
I for Nike
I for Custom
     ou selected BestBuy Dataset
nter the minimum support: 0.15
nter the minimum Confidence: 0.2
        oxenset('External Hand-Drive')) 0.45

oxenset('Anti-Virus')) 0.7

oxenset('Microsoft Office')) 0.55

oxenset('Jab roo,') 0.55

oxenset('Lab roo,') 0.65

oxenset('Lab roo,') 0.65

oxenset('Lab roo,') 0.65

oxenset('Lab roo,') 0.65

oxenset('Islah Drive') 0.65

oxenset('Desk roo,') 0.3

oxenset('Desk roo,') 0.3

oxenset('Desk roo,') 0.5
          requent 2-itemsets support
                          nset({'Flash Drive', 'Speakers'}) 0.3
nset({'Speakers', 'Printer'}) 0.25
nset({'Speakers', 'Printer'}) 0.25
nset({'External Hard-Drive', 'Anti-Virus'}) 0.45
nset({'Lab Top', 'Speakers'}) 0.25
nset({'Desk Top', 'Microsoft Office'}) 0.25
nset({'Desk Top', 'Anti-Virus'}) 0.2
nset({'Desk Top', 'External Hard-Drive'}) 0.15
nset({'External Hard-Drive', 'Digital Camera'}) 0.15
nset({'Lab Top Case', 'Flash Drive'}) 0.45
nset({'Microsoft Office', 'Flash Drive'}) 0.55
nset({'Microsoft Office', 'Printer'}) 0.45
                 censet(('Microsoft Office', 'Printer')) 8.45

Command Prompt

censet(('Anti-Virus', 'Flash Drive')) 8.5

censet(('Desk Top', 'Speakers')) 8.2

censet(('Desk Top', 'Speakers')) 8.35

censet(('Anti-Virus', 'Printer')) 8.35

censet(('Anti-Virus', 'Printer')) 8.35

censet(('Anti-Virus', 'Printer')) 8.35

censet(('Anti-Virus', 'Printer')) 8.25

censet('Iab Top, 'Digital Camera')) 8.25

censet('Iab Top Case', 'Flash Drive', 'Speakers')) 8.25

censet('Iab Top Case', 'Speakers', 'Printer')) 8.35

censet('Osak Top', 'Microsoft Office', 'Speakers')) 8.15

censet('Osak Top', 'Microsoft Office', 'Speakers')) 8.35

censet('Iab Top Case', 'Speakers', 'Printer')) 8.35

censet('Lab Top Case', 'Speakers', 'Printer')) 8.15

censet('Catt-Virus', 'Speakers', 'Printer')) 8.15

censet('External Hard-Drive', 'Flash Drive', 'Speakers')) 8.3

censet('External Hard-Drive', 'Printer')) 8.15

censet('External Hard-Drive', 'Printer')) 8.15

censet('External Hard-Drive', 'Iab Top', 'Printer')) 8.15

censet('External Hard-Drive', 'Speakers', 'Digital Camera')) 8.3

censet('External Hard-Drive', 'Speakers', 'Lab Top')) 8.45

censet('Flash Drive', 'Speakers', 'Digital Camera')) 8.25

censet('External Hard-Drive', 'Lab Top Case') 8.3

censet('External Hard-Drive', 'Lab Top Case') 8.4

censet('External Hard-Drive', 'Lab Top Case') 8.2

censet('External Hard-Drive', 'Lab Top Case') 8.2

censet('External Hard-Drive', 'Lab Top Case') 8.2

censet('Microsoft Office', 'External Hard-Drive', 'Speakers') 8.3

censet('Microsoft Office', 'External Hard-Drive', 'Speakers') 8.2

censet('Microsoft Office', 'External Hard-Drive', 'Anti-Virus') 8.25

censet('Microsoft Office', 'External Hard-Drive',
Command Prompt
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         O
```

## 3.KMART

```
Select Command Prompt
                                                                                                                                                                                                                                                                                                                                                                                                                                                          X
 Hello, welcome to association rule mining
Please choose the dataset for which you would like to get the Association rules
 Please type
        for Amazon
      for BestBuy
for KMART
4 for Nike
5 for Custom
П
    frozenset(('Socka')) => frozenset(('Modern Pants')) confidence: 0.8

Select Command Drompt
frozenset(('Solming Shirt', 'Dry Fit V-Nick')) => frozenset(('Rash Guard')) confidence: 0.8

frozenset(('Solming Shirt', 'Rash Guard')) => frozenset(('No Fit V-Nick')) confidence: 0.9

frozenset(('Solming Shirt', 'Rash Guard')) => frozenset(('No Fit V-Nick')) confidence: 0.9

frozenset(('Solming Shirt')) >> frozenset(('Solming Shirt', 'Dry Fit V-Nick')) confidence: 0.85

frozenset(('Solming Shirt')) >> frozenset(('Solming Shirt', 'Rash Guard')) confidence: 0.81

frozenset(('Treb Pants', 'Dry Fit V-Nick')) >> frozenset(('Fash Guard')) confidence: 0.8

frozenset(('Treb Pants', 'Dry Fit V-Nick')) >> frozenset(('Treb Pants')) confidence: 0.8

frozenset(('Treb Pants', 'Nash Guard')) >> frozenset(('Treb Pants')) confidence: 0.8

frozenset(('Treb Pants')) >> frozenset(('Treb Pants', 'Dry Fit V-Nick')) >> frozenset(('Treb Pants', 'Or Fit V-Nick')) >> frozenset(('Treb Pants', 'Dry Fit V-Nick')) >> frozenset(('Solming Shoe', 'Nodern Pants')) >> frozenset(('Solming Shoe')) >> frozenset(('Solming Shoe')) >> frozenset(('Nodern Pants')) >> frozenset(('Nodern Pants')
```

# 4.Nike

```
equent 1-itemsets support

ozenset(['Swimming Shirt']) 0.55

ozenset(['Nedem Pants']) 0.45

ozenset(['Medem Pants']) 0.5

ozenset(['Swimming Shoe']) 0.5

ozenset(['Summing Shoe']) 0.7

ozenset(['Sweatchirts']) 0.5

ozenset(['Nedem']) 0.65

ozenset(['Reddies']) 0.4

ozenset(['Reddies']) 0.4

ozenset(['Reddies']) 0.5
                                      enset(('Dry Fit V-Nick')) 0.5

unent 2-itemsets

support

enset(('Hoodies', 'Dry Fit V-Nick')) 0.35

enset(('Swimming Shor', 'Rash Guard')) 0.35

enset(('Swimming Shor', 'Rash Guard')) 0.5

enset(('Swimming Shor', 'Nowenthint's')) 0.5

enset(('Swimming Shor', 'Swimmint's')) 0.5

enset(('Rash Guard', 'Dry Fit V-Nick')) 0.5

enset(('Rash Guard', 'Dry Fit V-Nick')) 0.4

enset(('Rash Guard', 'Dry Fit V-Nick')) 0.4

enset(('Hoodies', 'Rash Guard')) 0.4

enset(('Hoodies', 'Rash Guard')) 0.5

enset('Socks', 'Wodern Pants') 0.4

enset('Gocks', 'Wodern Pants') 0.5

enset('Running Shor', 'Socks')) 0.55

enset('Running Shor', 'Nodern Pants')) 0.45

enset('Cryming Shor', 'Nodern Pants')) 0.45

enset('Cryming Shor', 'Nodern Pants')) 0.45

enset('Cryming Shor', 'Nodern Pants')) 0.35

enset('Cryming Shor', 'Nodern Pants')) 0.35
Select Command Prompt
                   ovenset(('Socks', 'Sweatshirts', 'Modern Pants')) 0.4
ovenset(('Hoodies', 'Tech Pants', 'Dry Fit V-Nick')) 0.35
ovenset(('Gwimming Shirt', 'Tech Pants', 'Dry Fit V-Nick')) 0.35
ovenset(('Guimming Shirt', 'Rash Guard', 'Swimming Shirt')) 0.35
ovenset(('Swimming Shirt', 'Rash Guard', 'Dry Fit V-Nick')) 0.45
ovenset(('Rich Pants', 'Rash Guard', 'Dry Fit V-Nick')) 0.4
ovenset(('Running Shoe', 'Socks', 'Modern Pants')) 0.4
ovenset(('Running Shoe', 'Socks', 'Sweatshirts')) 0.5
ovenset(('Nunning Shoe', 'Socks', 'Sweatshirts')) 0.5
ovenset(('Hoodies', 'Rash Guard', 'Dry Fit V-Nick')) 0.35
        | Second Common | Second Commo
     frozenset(('Nodern Pants')) => frozenset(('Modern Pants')) confidence: 0.8

Select Command Prompt
frozenset(('Wodern Pants')) => frozenset(('Socks')) confidence: 0.8

Select Command Prompt
frozenset(('Swimming Shirt', 'Dry Fit V-Nick')) => frozenset(('Dry Fit V-Nick')) confidence: 0.9
frozenset(('Rash Guard')) => frozenset(('Swimming Shirt', 'Dry Fit V-Nick')) confidence: 0.9
frozenset(('Rash Guard')) => frozenset(('Swimming Shirt', 'Rash Guard')) confidence: 0.8
frozenset(('Ony Fit V-Nick')) => frozenset(('Swimming Shirt', 'Rash Guard')) confidence: 0.8
frozenset(('Tech Pants', 'Dry Fit V-Nick')) => frozenset(('Tech Pants')) confidence: 0.8
frozenset(('Tech Pants', 'Rash Guard')) => frozenset(('Tech Pants')) confidence: 0.8
frozenset(('Tech Pants', 'Rash Guard')) => frozenset(('Tech Pants')) confidence: 0.8
frozenset(('Tech Pants')) => frozenset(('Tech Pants')) confidence: 0.8
frozenset(('Tech Pants')) => frozenset(('Tech Pants', 'Ury Fit V-Nick')) confidence: 0.8
frozenset(('Rash Guard')) => frozenset(('Tech Pants', 'Ury Fit V-Nick')) confidence: 0.8
frozenset(('Rash Guard')) => frozenset(('Tech Pants', 'Ury Fit V-Nick')) confidence: 0.8
frozenset(('Running Shoe', 'Socks')) => frozenset(('Tech Pants', 'Ury Fit V-Nick')) confidence: 0.8
frozenset(('Running Shoe', 'Socks')) => frozenset(('Running Shoe')) confidence: 0.8
frozenset(('Running Shoe', 'Socks')) => frozenset(('Running Shoe')) confidence: 0.8
frozenset(('Running Shoe')) => frozenset(('Running Shoe')) confidence: 0.8
frozenset(('Running 
Select Command Prompt
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        o ×
```

## 5.Custom

```
es_Command Prompt

::\Users\VINWY\midterm_datamining_vinay>python midterm_apriori.py

tello, welcome to association rule mining

lease choose the dataset for which you would like to get the Association rules

lease type

for Amazon

for BestBuy

for NMART

for Nike

for Custom
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          n.
    rozenset(| WAGGI | ) 0.25
rozenset(| MAGGI | ) 0.25
rozenset(| MAGGI | ) 0.35
rozenset(| ORNIFLAKES' | ) 0.3
rozenset(| ORNIFLAKES' | ) 0.3
rozenset(| GOCT | ) 0.15
rozenset(| OFFEE' | ) 0.4
rozenset(| MAGGI | ) 0.3
    requent 2-itemsets support

rozenset((TEA', 'BREAD')) 0.2

rozenset((TEA', 'CORN-LAKES')) 0.1

rozenset((TEA', 'CORN-LAKES')) 0.1

rozenset((SEREA), 'BOLNITA')) 0.1

rozenset((BREAD), 'BOLNITA')) 0.15

rozenset((MILK', 'BISCUIT')) 0.1

rozenset((MILK', 'BISCUIT')) 0.1

rozenset((GORN-LAKES', 'COK')) 0.1

rozenset(('CORN-LAKES', 'NILK')) 0.1

rozenset(('SREAD, 'NILK')) 0.2

rozenset(('SUKER', 'BOLRNITA')) 0.1
Command Prompt
  Frozenset(('BREAD', 'BISCUIT')) 0.2
frozenset(('GNEAD'), 'BISCUIT')) 0.1
frozenset(('COK', 'BISCUIT')) 0.1
frozenset(('JAM', 'MAGGI')) 0.1
frozenset(('JAM', 'MAGGI')) 0.1
frozenset(('BREAD', 'JAM', 'MAGGI')) 0.1
frozenset(('REA', 'BISCUIT', 'MAGGI')) 0.1
frozenset(('TEA', 'BISCUIT', 'MAGGI')) 0.1
frozenset(('TEA', 'BISCUIT', 'MAGGI')) 0.1
frozenset(('GNEAD', MILK', BISCUIT')) 0.1
frozenset(('GNEAD', 'MILK', BISCUIT')) 0.1
frozenset(('GOFFEE', 'GOKNELAKES', 'GOK')) 0.1
frozenset(('GOFFEE', 'BREAD', 'GOFFEE')) 0.1
frozenset(('GOFFEE', 'GOK', 'BISCUIT')) 0.1
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       ri i
```

```
Command Prompt
   et({'COFFEE'}) => fro
   Ø
Command Prompt
                          0.28571428571428575
 \Users\VINAY\midterm_datamining_vinay>
```

n