Aassociation rules popular words

##### import numpy as np

import pandas as pd

from apyori import apriori

store\_data = pd.read\_csv('C:\\Datasets\\reversed.csv')

#st=store\_data.tail(30)

#st=store\_data.head(30)

st=store\_data

print(st)

records = []

for i in range(1, 99):

records.append([str(st.values[i,j]) for j in range(15, 30)])#רשימה חדשה עם כל הפריטים שאני יכול למצוא. תהפוך לסטרינג של כל הערכים במיקומים אי ג'י(מטריצה) והעמודות בין 1 ל20

print(records)

association\_rules = apriori(records, min\_support=0.045, min\_confidence=0.030, min\_lift=1, min\_length=3)#תמצא קשר בעזרת אפריורי על הרקורדז שבניתי שהמינימים סופורט הוא 0.00045 אחוז כלומר מחפש סלים שמקיימים שלפחות 0.00045% מהדאטה סט מכיל אותם, סלים הכל מגודל מסויים זה האורך

association\_results = list(association\_rules)

#print(len(association\_rules))

print('/n')

print(association\_results[0])

print(association\_results[1])

15-30 עמודות (מילים)

תוצאה:

RelationRecord(items=frozenset({'also'}), support=0.4897959183673469, ordered\_statistics=[OrderedStatistic(items\_base=frozenset(), items\_add=frozenset({'also'}), confidence=0.4897959183673469, lift=1.0)])

RelationRecord(items=frozenset({'based'}), support=0.5102040816326531, ordered\_statistics=[OrderedStatistic(items\_base=frozenset(), items\_add=frozenset({'based'}), confidence=0.5102040816326531, lift=1.0)])

##### import numpy as np

import pandas as pd

from apyori import apriori

store\_data = pd.read\_csv('C:\\Datasets\\reversed.csv')

#st=store\_data.tail(30)

#st=store\_data.head(30)

st=store\_data

print(st)

records = []

for i in range(1, 99):

records.append([str(st.values[i,j]) for j in range(50,60)])#רשימה חדשה עם כל הפריטים שאני יכול למצוא. תהפוך לסטרינג של כל הערכים במיקומים אי ג'י(מטריצה) והעמודות בין 1 ל20

print(records)

association\_rules = apriori(records, min\_support=0.045, min\_confidence=0.030, min\_lift=1, min\_length=3)#תמצא קשר בעזרת אפריורי על הרקורדז שבניתי שהמינימים סופורט הוא 0.00045 אחוז כלומר מחפש סלים שמקיימים שלפחות 0.00045% מהדאטה סט מכיל אותם, סלים הכל מגודל מסויים זה האורך

association\_results = list(association\_rules)

#print(len(association\_rules))

print('/n')

print(association\_results[0])

print(association\_results[1])

תוצאה (50-60 עמודות)

RelationRecord(items=frozenset({'analysis'}), support=0.6224489795918368, ordered\_statistics=[OrderedStatistic(items\_base=frozenset(), items\_add=frozenset({'analysis'}), confidence=0.6224489795918368, lift=1.0)])

RelationRecord(items=frozenset({'example'}), support=0.7142857142857143,

ordered\_statistics=[OrderedStatistic(items\_base=frozenset(), items\_add=frozenset({'example'}), confidence=0.7142857142857143, lift=1.0)])

##### import numpy as np

import pandas as pd

from apyori import apriori

store\_data = pd.read\_csv('C:\\Datasets\\reversed.csv')

#st=store\_data.tail(30)

#st=store\_data.head(30)

st=store\_data

print(st)

records = []

for i in range(1, 99):

records.append([str(st.values[i,j]) for j in range(70,85)])#רשימה חדשה עם כל הפריטים שאני יכול למצוא. תהפוך לסטרינג של כל הערכים במיקומים אי ג'י(מטריצה) והעמודות בין 1 ל20

print(records)

association\_rules = apriori(records, min\_support=0.030, min\_confidence=0.030, min\_lift=2, min\_length=3)#תמצא קשר בעזרת אפריורי על הרקורדז שבניתי שהמינימים סופורט הוא 0.00045 אחוז כלומר מחפש סלים שמקיימים שלפחות 0.00045% מהדאטה סט מכיל אותם, סלים הכל מגודל מסויים זה האורך

association\_results = list(association\_rules)

#print(len(association\_rules))

print('/n')

print(association\_results[0])

print(association\_results[1])

תוצאה 70-85 עמודות

RelationRecord(items=frozenset({'variables', 'step', 'size', 'classification', 'paper'}), support=0.20408163265306123, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'size', 'step', 'variables', 'classification'}), items\_add=frozenset({'paper'}), confidence=0.8695652173913043, lift=2.028985507246377)])

RelationRecord(items=frozenset({'variables', 'step', 'size', 'note', 'paper'}), support=0.1836734693877551, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'size', 'note', 'step', 'variables'}), items\_add=frozenset({'paper'}), confidence=0.8571428571428572, lift=2.0000000000000004)])

##### import numpy as np

import pandas as pd

from apyori import apriori

store\_data = pd.read\_csv('C:\\Datasets\\reversed.csv')

#st=store\_data.tail(30)

#st=store\_data.head(30)

st=store\_data

print(st)

records = []

for i in range(1, 99):

records.append([str(st.values[i,j]) for j in range(60,75)])#רשימה חדשה עם כל הפריטים שאני יכול למצוא. תהפוך לסטרינג של כל הערכים במיקומים אי ג'י(מטריצה) והעמודות בין 1 ל20

print(records)

association\_rules = apriori(records, min\_support=0.030, min\_confidence=0.030, min\_lift=1, min\_length=3)#תמצא קשר בעזרת אפריורי על הרקורדז שבניתי שהמינימים סופורט הוא 0.00045 אחוז כלומר מחפש סלים שמקיימים שלפחות 0.00045% מהדאטה סט מכיל אותם, סלים הכל מגודל מסויים זה האורך

association\_results = list(association\_rules)

#print(len(association\_rules))

print('/n')

print(association\_results[0])

print(association\_results[1])

תוצאות: 60-75

RelationRecord(items=frozenset({'bound'}), support=0.6530612244897959, ordered\_statistics=[OrderedStatistic(items\_base=frozenset(), items\_add=frozenset({'bound'}), confidence=0.6530612244897959, lift=1.0)])

RelationRecord(items=frozenset({'class'}), support=0.5612244897959183, ordered\_statistics=[OrderedStatistic(items\_base=frozenset(), items\_add=frozenset({'class'}), confidence=0.5612244897959183, lift=1.0)])

##### import numpy as np

import pandas as pd

from apyori import apriori

store\_data = pd.read\_csv('C:\\Datasets\\reversed.csv')

#st=store\_data.tail(30)

#st=store\_data.head(30)

st=store\_data

print(st)

records = []

for i in range(1, 99):

records.append([str(st.values[i,j]) for j in range(75,90)])#רשימה חדשה עם כל הפריטים שאני יכול למצוא. תהפוך לסטרינג של כל הערכים במיקומים אי ג'י(מטריצה) והעמודות בין 1 ל20

print(records)

association\_rules = apriori(records, min\_support=0.030, min\_confidence=0.030, min\_lift=2, min\_length=3)#תמצא קשר בעזרת אפריורי על הרקורדז שבניתי שהמינימים סופורט הוא 0.00045 אחוז כלומר מחפש סלים שמקיימים שלפחות 0.00045% מהדאטה סט מכיל אותם, סלים הכל מגודל מסויים זה האורך

association\_results = list(association\_rules)

#print(len(association\_rules))

print('/n')

print(association\_results[0])

print(association\_results[1])

תוצאה טובה(ארוכה) 75-90

RelationRecord(items=frozenset({'parameter', 'size', 'note', 'paper'}), support=0.20408163265306123, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'parameter', 'size', 'note'}), items\_add=frozenset({'paper'}), confidence=0.8695652173913043, lift=2.028985507246377)])

RelationRecord(items=frozenset({'parameter', 'rate', 'size', 'paper'}), support=0.20408163265306123, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'parameter', 'rate', 'size'}), items\_add=frozenset({'paper'}), confidence=0.8695652173913043, lift=2.028985507246377)])