Association rules 2015

1-30 מילים ראשונות(30 עמודות ראשונות)

##### import numpy as np

import pandas as pd

from apyori import apriori

store\_data = pd.read\_csv('C:\\Datasets\\2015\_reversed.csv')

#st=store\_data.tail(30)

#st=store\_data.head(30)

st=store\_data

print(st)

records = []

for i in range(0, 30):

records.append([str(st.values[i,j]) for j in range(1, 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=0.030, min\_confidence=0.045, min\_lift=1, min\_length=2)#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

RelationRecord(items=frozenset({'a', 'ability'}), support=0.03333333333333333, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'a'}), items\_add=frozenset({'ability'}), confidence=1.0, lift=4.285714285714286), OrderedStatistic(items\_base=frozenset({'ability'}), items\_add=frozenset({'a'}), confidence=0.14285714285714285, lift=4.285714285714286)])

RelationRecord(items=frozenset({'ability', 'abalone'}), support=0.03333333333333333, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'abalone'}), items\_add=frozenset({'ability'}), confidence=1.0, lift=4.285714285714286), OrderedStatistic(items\_base=frozenset({'ability'}), items\_add=frozenset({'abalone'}), confidence=0.14285714285714285, lift=4.285714285714286)])

מילים בין עמודה 40 ל70

##### import numpy as np

import pandas as pd

from apyori import apriori

store\_data = pd.read\_csv('C:\\Datasets\\2015\_reversed.csv')

#st=store\_data.tail(30)

#st=store\_data.head(30)

st=store\_data

print(st)

records = []

for i in range(0, 30):

records.append([str(st.values[i,j]) for j in range(40, 70)])#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=0.045, 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

RelationRecord(items=frozenset({'accepts', 'access'}), support=0.03333333333333333, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'accepts'}), items\_add=frozenset({'access'}), confidence=1.0, lift=7.5), OrderedStatistic(items\_base=frozenset({'access'}), items\_add=frozenset({'accepts'}), confidence=0.25, lift=7.5)])

RelationRecord(items=frozenset({'accepts', 'account'}), support=0.03333333333333333, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'accepts'}), items\_add=frozenset({'account'}), confidence=1.0, lift=3.75), OrderedStatistic(items\_base=frozenset({'account'}), items\_add=frozenset({'accepts'}), confidence=0.125, lift=3.75)])

60-120

תוצאה:

/n

RelationRecord(items=frozenset({'accurate', 'accounting'}), support=0.03333333333333333, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'accounting'}), items\_add=frozenset({'accurate'}), confidence=0.5, lift=2.142857142857143), OrderedStatistic(items\_base=frozenset({'accurate'}), items\_add=frozenset({'accounting'}), confidence=0.14285714285714285, lift=2.142857142857143)])

RelationRecord(items=frozenset({'accounting', 'achieving'}), support=0.03333333333333333, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'accounting'}), items\_add=frozenset({'achieving'}), confidence=0.5, lift=2.5), OrderedStatistic(items\_base=frozenset({'achieving'}), items\_add=frozenset({'accounting'}), confidence=0.16666666666666666, lift=2.5)])

1200-1300

תוצאה:

/n

RelationRecord(items=frozenset({'buhmann', 'brings'}), support=0.03333333333333333, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'brings'}), items\_add=frozenset({'buhmann'}), confidence=0.5, lift=15.0), OrderedStatistic(items\_base=frozenset({'buhmann'}), items\_add=frozenset({'brings'}), confidence=1.0, lift=15.0)])

RelationRecord(items=frozenset({'brings', 'building'}), support=0.03333333333333333, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'brings'}), items\_add=frozenset({'building'}), confidence=0.5, lift=7.5), OrderedStatistic(items\_base=frozenset({'building'}), items\_add=frozenset({'brings'}), confidence=0.5, lift=7.5)])

11000-11050

תוצאה

RelationRecord(items=frozenset({'updates', 'update'}), support=0.2, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'update'}), items\_add=frozenset({'updates'}), confidence=0.5454545454545455, lift=2.337662337662338), OrderedStatistic(items\_base=frozenset({'updates'}), items\_add=frozenset({'update'}), confidence=0.8571428571428572, lift=2.337662337662338)])

RelationRecord(items=frozenset({'updating', 'update'}), support=0.16666666666666666, ordered\_statistics=[OrderedStatistic(items\_base=frozenset({'update'}), items\_add=frozenset({'updating'}), confidence=0.45454545454545453, lift=2.7272727272727275), OrderedStatistic(items\_base=frozenset({'updating'}), items\_add=frozenset({'update'}), confidence=1.0, lift=2.7272727272727275)])