

سوال 3)

روند کلی کد:

ابتدا موارد تکراری در هر ریکورد را حذف میکنیم و سپس هر ایتm را بصورت key value درآورده و سپس میشماریم.

ایتم های پرتکرار را با توجه به ترشولد داده شده بدست میاوریم.

تمام زیرمجموعه های دو عضوی را از ایتm های پرتکرار بدست میاوریم و ساپورت هر کدام را محاسبه میکنیم.

از میان ست های دو عضوی، پرتکرار ها را جدا میکنیم.

تمام زیرمجموعه های سه عضوی را از ست های دو عضوی پرتکرار بدست میاوریم و از میان انها ست های سه عضوی پرتکرار را جدا میکنیم.

ایتم های پرتکرار:

647 تا ایتم پرتکرار داریم که چند مورد اول بدین صورت میباشد. (ایتم ها براساس ساپورت به صورت نزولی مرتب شده اند)

| | | |
|---------------------|---------------------|---------------------|
| ('DAI62779', 6667), | ('FRO32293', 1702), | ('DAI22896', 1219), |
| ('FRO40251', 3881), | ('ELE66810', 1697), | ('GRO85051', 1214), |
| ('ELE17451', 3875), | ('SNA55762', 1646), | ('ELE92920', 1197), |
| ('GRO73461', 3602), | ('DAI22177', 1627), | ('GRO44993', 1193), |
| ('SNA80324', 3044), | ('FRO78087', 1531), | ('ELE14480', 1147), |
| ('ELE32164', 2851), | ('ELE99737', 1516), | ('DAI91290', 1138), |
| ('DAI75645', 2736), | ('GRO94758', 1489), | ('SNA55952', 1094), |
| ('SNA45677', 2455), | ('ELE34057', 1489), | ('SNA72163', 1090), |
| ('FRO31317', 2330), | ('FRO35904', 1436), | ('DAI35347', 1060), |
| ('DAI85309', 2293), | ('FRO53271', 1420), | ('GRO64900', 1046), |
| ('ELE26917', 2292), | ('SNA93860', 1407), | ('SNA62128', 1023), |
| ('FRO80039', 2233), | ('SNA90094', 1390), | ('DAI42493', 1001), |
| ('GRO21487', 2115), | ('GRO38814', 1352), | ('DAI55148', 975), |
| ('SNA99873', 2083), | ('ELE56788', 1345), | ('DAI55911', 974), |
| ('GRO59710', 2004), | ('GRO61133', 1321), | ('GRO24246', 944), |
| ('GRO71621', 1920), | ('DAI88807', 1316), | ('DAI92600', 935), |
| ('FRO85978', 1918), | ('ELE74482', 1316), | ('FRO16142', 931), |
| ('GRO30386', 1840), | ('ELE59935', 1311), | ('FRO66272', 927), |
| ('ELE74009', 1816), | ('SNA96271', 1295), | ('FRO92261', 915), |
| ('GRO56726', 1784), | ('DAI43223', 1290), | ('GRO99222', 906), |
| ('DAI63921', 1773), | ('ELE91337', 1289), | ('SNA59903', 891), |
| ('GRO46854', 1756), | ('GRO15017', 1275), | ('GRO69543', 882), |
| ('ELE66600', 1713), | ('DAI31081', 1261), | ('DAI95741', 878), |
| ('DAI83733', 1712), | ('GRO81087', 1220), | ('SNA88283', 865), |
| | | ('DAI87448', 858), |
| | | ('GRO32086', 856), |
| | | ('GRO88511', 855), |

ایتم ست های دوتایی پرتکرار:

تعداد ایتم ست های دوتایی که از ایتم های پرتکرار بدست آمده برابر 208981 میباشد که در واقع همان candidate pairs ها هستند، که از این تعداد تنها 1334 تا پرتکرار میباشد. تعداد ایتم های منحصر بفرد که جفتهای پرتکرار را ساخته اند برابر 222 میباشد.

ایتم ست های دوتایی پرتکرار در فایل freq_pairs.txt ذخیره شده است. (برای دفعات بعد فایل را میخوانیم چون هر بار ران کردن حدود 2 ساعت طول میکشد به همین خاطر این قسمت از کد یکبار ران شده و سپس کامنت شده است) هر ایتم ست به همراه ساپورت آن نوشته شده است. بدین شکل:

```
[('SNA90258', 'ELE17451'), 113), (('SNA90258', 'GRO99222'), 156), (('SNA90258', 'DAI62779'), 114), (('DAI2289', 'GRO73461'), 123), (('GRO73461', 'FRO31317'), 395), (('GRO73461', 'DAI63921'), 219), (('GRO73461', 'GRO71621'), 305), (('GRO73461', 'FRO47962'), 137), (('GRO73461', 'GRO88511'), 100), (('GRO73461', 'FRO43226'), 110), (('GRO73461', 'FRO40251'), 882), (('GRO73461', 'ELE20398'), 111), (('GRO73461', 'SNA69641', 'DAI62779'), 141), (('GRO39357', 'FRO78087'), 105), (('DAI50921', 'DAI62779'), 106), (('DAI22177', 'DAI22177', 'DAI83733'), 126), (('DAI22177', 'FRO80039'), 152), (('DAI22177', 'GRO46854'), 160), (('ELE59935', 'LE66810', 'ELE66600'), 132), (('ELE66810', 'DAI62779'), 279), (('ELE66810', 'SNA45677'), 119), (('ELE66810', 'E11111', 'DAI75645'), 186), (('ELE11111', 'ELE17451'), 121), (('ELE11111', 'GRO94758'), 101), (('ELE11111', 'GRO72163', 'ELE74009'), 145), (('SNA72163', 'ELE17451'), 272), (('SNA72163', 'GRO30386'), 138), (('SNA72163', 'SNA1317', 'GRO56726'), 281), (('FRO31317', 'ELE56788'), 112), (('FRO31317', 'ELE74482'), 317), (('FRO31317', 'GRO317', 'DAI31081'), 109), (('FRO31317', 'DAI83733'), 159), (('FRO31317', 'FRO80039'), 187), (('FRO31317', 'DAI421', 'GRO21487'), 163), (('DAI63921', 'ELE99737'), 108), (('DAI63921', 'FRO35904'), 121), (('DAI63921', 'ELE342', 'SNA80324'), 174), (('ELE12792', 'DAI75645'), 109), (('ELE12792', 'ELE17451'), 159), (('ELE12792', 'DAI6274', 'DAI88807'), 194), (('ELE32164', 'SNA80324'), 115), (('ELE32164', 'SNA96271'), 101), (('ELE32164', 'SNA380', 'SNA45677'), 224), (('ELE32164', 'FRO43226'), 121), (('ELE32164', 'FRO40251'), 152), (('ELE32164', 'GRO6954', 'ELE17451'), 291), (('DAI88807', 'FRO32293'), 105), (('DAI88807', 'SNA93860'), 109), (('DAI88807', 'GRO59710', 'ELE56788'), 148), (('SNA80324', 'GRO88324'), 106), (('SNA80324', 'ELE74482'), 114), (('SNA80324', 'FRO61354', 'ELE53126'), 101), (('SNA80324', 'ELE37048'), 106), (('SNA80324', 'GRO21487'), 225), (('SNA80324', 'FRO92469', 'DAI55148'), 299), (('DAI75645', 'DAI85309'), 212), (('DAI75645', 'ELE17451'), 547), (('DAI75645', 'ELE26917'), 'ELE34057'), 110), (('DAI75645', 'DAI31081'), 206), (('DAI75645', 'DAI83733'), 155), (('DAI75645', 'FRO80039'), 'GRO46854'), 131), (('DAI67621', 'DAI62779'), 109), (('GRO50832', 'DAI62779'), 150), (('GRO50832', 'FRO40251'), 'RO21487'), 211), (('FRO85978', 'ELE99737'), 110), (('FRO85978', 'FRO35904'), 160), (('FRO85978', 'SNA95666'), 'I62779'), 432), (('ELE74009', 'SNA55762'), 128), (('ELE74009', 'SNA45677'), 233), (('ELE74009', 'FRO40251'), 21487'), 160), (('GRO56726', 'ELE99737'), 119), (('GRO56726', 'SNA90094'), 102), (('GRO56726', 'GRO24246'), 112779'), 237), (('ELE25077', 'GRO89004'), 215), (('ELE25077', 'DAI62779'), 135), (('ELE74482', 'DAI22240'), 103039'), 100), (('GRO71621', 'DAI85309'), 270), (('GRO71621', 'FRO19221'), 109), (('GRO71621', 'ELE17451'), 29077'), 206), (('GRO88511', 'DAI62779'), 175), (('GRO88511', 'FRO40251'), 113), (('GRO88511', 'DAI83948'), 131), 9'), 118), (('GRO81647', 'DAI62779'), 124), (('GRO81647', 'FRO40251'), 113), (('GRO81647', 'GRO21487'), 114), '\n 211) (('DAI85309', 'ELE99737'), 131) (('DAI85309', 'SNA45677'), 283) (('DAI85309', 'FRO40251'), 284) (\n
```

ایتم ست های سه تایی پرتکرار:

تعداد ایتم ست های سه تایی که از ایتم ست های دوتایی پرتکرار بدست آمده برابر **1798940** میباشد که در واقع همان candidate triples ها هستند.

قبل از بدست آوردن سه تایی های پرتکرار، کاندیدها را فیلتر میکنیم، هر کاندید، سه عضو دارد، بنابراین سه زیر مجموعه دو عضوی دارد. هر کاندید در صورتی **میتواند** پرتکرار باشد، که هر سه زیرمجموعه دو عضوی آن پرتکرار باشد (دوتایی های پرتکرار در قسمت قبل بدست آمد). پس از اعمال این شرط، تعداد کاندیدها به **1900** عدد میرسد که به طرز چشمگیری هزینه محاسبات کاهش میابد.

که از این تعداد تنها 79 تا پرتکرار میباشد.

هر ایتم ست به همراه ساپورت آن نوشته شده است. (بصورت نزولی مرتب شده اند) بدین شکل:

```
[('SNA80324', 'DAI75645', 'FRO40251'), 550),
(('SNA18336', 'DAI62779', 'ELE92920'), 432),
(('SNA80324', 'DAI75645', 'DAI62779'), 421),
(('SNA80324', 'ELE17451', 'DAI62779'), 417),
(('ELE17451', 'DAI62779', 'ELE92920'), 345),
(('DAI75645', 'ELE17451', 'DAI62779'), 328),
(('SNA80324', 'DAI75645', 'ELE17451'), 300),
(('GRO73461', 'DAI75645', 'FRO40251'), 293),
(('ELE32164', 'DAI62779', 'DAI43223'), 287),
(('ELE32164', 'ELE17451', 'DAI62779'), 277),
(('GRO73461', 'DAI75645', 'DAI62779'), 261),
(('GRO73461', 'ELE17451', 'DAI62779'), 245),
(('SNA18336', 'ELE17451', 'DAI62779'), 244),
(('SNA18336', 'ELE17451', 'ELE92920'), 228),
(('ELE17451', 'DAI62779', 'DAI43223'), 227),
(('ELE32164', 'ELE17451', 'DAI43223'), 206),
(('GRO59710', 'DAI62779', 'DAI43223'), 205),
(('SNA59903', 'ELE17451', 'DAI62779'), 202),
(('GRO73461', 'FRO40251', 'GRO21487'), 182),
(('GRO73461', 'DAI62779', 'GRO21487'), 173),
(('ELE32164', 'DAI62779', 'ELE92920'), 165),
(('SNA80324', 'DAI75645', 'ELE20847'), 164),
(('ELE17451', 'DAI62779', 'GRO81087'), 160),
(('ELE17451', 'DAI62779', 'SNA45677'), 158),
(('SNA80324', 'ELE20847', 'DAI62779'), 153),
(('GRO73461', 'GRO71621', 'DAI62779'), 153),
(('DAI55148', 'ELE17451', 'DAI62779'), 150),
(('DAI55148', 'ELE17451', 'DAI62779'), 150),
(('ELE17451', 'DAI62779', 'DAI83733'), 147),
(('GRO15017', 'DAI62779', 'ELE92920'), 143),
(('GRO73461', 'FRO19221', 'DAI62779'), 142),
(('SNA18336', 'DAI85309', 'ELE92920'), 139),
(('DAI62779', 'GRO81087', 'ELE92920'), 134),
(('SNA80324', 'ELE26917', 'FRO40251'), 133),
(('SNA80324', 'DAI75645', 'GRO30386'), 131),
(('GRO73461', 'ELE32164', 'DAI62779'), 131),
(('SNA53220', 'FRO19221', 'DAI62779'), 131),
(('SNA80324', 'DAI75645', 'ELE26917'), 130),
(('DAI75645', 'DAI62779', 'ELE92920'), 130),
(('DAI75645', 'ELE26917', 'FRO40251'), 128),
(('DAI75645', 'FRO40251', 'FRO53271'), 123),
(('GRO73461', 'DAI75645', 'ELE17451'), 121),
(('SNA80324', 'DAI75645', 'GRO94758'), 120),
(('SNA59903', 'DAI88807', 'ELE17451'), 120),
(('SNA80324', 'DAI75645', 'FRO92469'), 119),
(('SNA80324', 'FRO40251', 'GRO21487'), 118),
(('SNA38068', 'ELE17451', 'DAI62779'), 118),
(('DAI75645', 'FRO40251', 'GRO38814'), 117),
(('GRO59710', 'DAI62779', 'ELE92920'), 116),
(('ELE78169', 'GRO94758', 'SNA45677'), 115),
(('SNA80324', 'FRO40251', 'FRO53271'), 115),
(('DAI75645', 'ELE20847', 'DAI62779'), 115),
(('SNA96271', 'DAI62779', 'SNA45677'), 115),
(('GRO73461', 'DAI75645', 'GRO21487'), 114),
(('GRO73461', 'SNA96271', 'DAI62779'), 114),
```

```
(( 'SNA80324', 'SNA96271', 'FRO40251'), 112),  
(( 'GRO73461', 'DAI62779', 'SNA45677'), 112),  
(( 'DAI75645', 'FRO40251', 'GRO81087'), 112),  
(( 'SNA80324', 'DAI75645', 'SNA45677'), 111),  
(( 'GRO73461', 'DAI75645', 'FRO47962'), 111),  
(( 'GRO73461', 'ELE32164', 'DAI43223'), 111),  
(( 'GRO73461', 'FRO40251', 'GRO69543'), 111),  
(( 'SNA96271', 'ELE17451', 'DAI62779'), 111),  
(( 'ELE17451', 'GRO15017', 'DAI62779'), 111),  
(( 'SNA80324', 'FRO40251', 'GRO81087'), 109),  
(( 'SNA80324', 'DAI75645', 'GRO81087'), 108),  
(( 'DAI75645', 'FRO40251', 'GRO21487'), 107),  
(( 'SNA72163', 'ELE17451', 'DAI62779'), 107),  
(( 'FRO31317', 'ELE17451', 'DAI62779'), 106),  
(( 'SNA80324', 'DAI75645', 'ELE74009'), 106),  
(( 'SNA80324', 'FRO40251', 'GRO38814'), 106),  
(( 'GRO73461', 'FRO40251', 'GRO38814'), 106),  
(( 'SNA18336', 'GRO15017', 'DAI62779'), 105),  
(( 'GRO73461', 'ELE17451', 'GRO30386'), 103),  
(( 'ELE17451', 'DAI62779', 'SNA90094'), 103),  
(( 'DAI62779', 'DAI83733', 'ELE92920'), 103),  
(( 'SNA80324', 'DAI75645', 'GRO38814'), 102),  
(( 'GRO73461', 'DAI75645', 'GRO46854'), 101),  
(( 'GRO73461', 'DAI75645', 'GRO38814'), 101),  
(( 'DAI75645', 'ELE26917', 'DAI62779'), 101)]
```

محاسبه confidence score :

پس از بدست آوردن ایتm ست های پرتکرار، rule های متناظر به هر ایتm ست را بدست آورده و کانفیدنس آن را حساب میکنیم.

ایتm ست های دوتایی پرتکرار:

برای هر ایتm ست دو rule میتوان نوشت مثلاً برای (a,b) ، $a \rightarrow b$ ، $b \rightarrow a$ را میتوان نوشت. باتوجه به اینکه 1334 ایتm ست دوتایی پرتکرار داریم، تعداد rule ها برابر 2668 خواهد بود که کانفیدنس متناظر با هر کدام محاسبه شده است. در ادامه کانفیدنس های بالاتر از 0.5 ذکر شده است. (مقادیر بصورت نزولی مرتب شده اند)

```
(( 'DAI93865--->', 'FRO40251'), 1.0),
(( 'GRO85051--->', 'FRO40251'), 0.9992),
(( 'GRO38636--->', 'FRO40251'), 0.9907),
(( 'ELE12951--->', 'FRO40251'), 0.9906),
(( 'DAI88079--->', 'FRO40251'), 0.9867),
(( 'FRO92469--->', 'FRO40251'), 0.9835),
(( 'DAI43868--->', 'SNA82528'), 0.973),
(( 'DAI23334--->', 'DAI62779'), 0.9545),
(( 'ELE92920--->', 'DAI62779'), 0.7327),
(( 'DAI53152--->', 'FRO40251'), 0.7179),
(( 'SNA18336--->', 'DAI62779'), 0.7137),
(( 'ELE55848--->', 'GRO32086'), 0.7095),
(( 'GRO89004--->', 'ELE25077'), 0.6981),
(( 'GRO81647--->', 'GRO73461'), 0.6776),
(( 'DAI37288--->', 'ELE32164'), 0.6464),
(( 'SNA18336--->', 'ELE92920'), 0.6417),
(( 'ELE32244--->', 'ELE66600'), 0.6404),
(( 'FRO47962--->', 'DAI75645'), 0.6176),
(( 'FRO73056--->', 'GRO44993'), 0.6016),
(( 'FRO19221--->', 'DAI62779'), 0.5977),
(( 'SNA44451--->', 'DAI18527'), 0.5829),
(( 'FRO17734--->', 'ELE28189'), 0.5816),
(( 'DAI46755--->', 'FRO81176'), 0.5804),
(( 'ELE88583--->', 'SNA24799'), 0.5793),
(( 'SNA95666--->', 'FRO85978'), 0.5759),
(( 'FRO41939--->', 'ELE66810'), 0.5735),
(( 'GRO81045--->', 'GRO46854'), 0.5675),
(( 'DAI16732--->', 'FRO78087'), 0.5668),
(( 'ELE81534--->', 'DAI62779'), 0.5634),
(( 'DAI43223--->', 'ELE32164'), 0.5512),
(( 'ELE59028--->', 'DAI62779'), 0.5498),
(( 'DAI55148--->', 'DAI62779'), 0.5395),
(( 'SNA30859--->', 'GRO24246'), 0.5312),
(( 'ELE20847--->', 'FRO40251'), 0.5306),
(( 'SNA30533--->', 'SNA96271'), 0.509),
(( 'ELE21353--->', 'DAI62779'), 0.5023),
(( 'ELE20847--->', 'SNA80324'), 0.5012),
(( 'GRO30912--->', 'ELE88583'), 0.5),
(( 'DAI23334--->', 'ELE92920'), 0.5),
```

ایتم ست های سه تایی پرتکرار:

برای هر ایتم ست سه تایی، سه rule بصورت 2 به 1 میتوان نوشت. یعنی:

(a,b,c):

(a,b)→c

(a,c)→b

(b,c)→a

با توجه به اینکه 79 تا ایتم ست سه تایی پرتکرار داریم، تعداد rule ها برابر 237 خواهد بود. ($237=3 \times 79$). برای هر کدام از rule های گفته شده مقدار کانفیدنس محاسبه شده است. در ادامه مقادیر بالای 0.5 ذکر شده است. (مقادیر بصورت نزولی مرتب شده اند)

| | |
|--|--|
| (('GRO81087', 'ELE92920--->', 'GRO73461'), 0.9571), | (('DAI75645', 'GRO94758--->', 'DAI22896'), 0.5911), |
| (('SNA18336', 'ELE92920--->', 'GRO73461'), 0.9495), | (('SNA59903', 'DAI62779--->', 'GRO73461'), 0.5906), |
| (('SNA18336', 'GRO15017--->', 'GRO88511'), 0.9292), | (('SNA80324', 'GRO81087--->', 'GRO73461'), 0.5892), |
| (('DAI83733', 'ELE92920--->', 'GRO73461'), 0.9279), | (('DAI75645', 'FRO47962--->', 'GRO73461'), 0.5873), |
| (('SNA18336', 'ELE17451--->', 'GRO71621'), 0.9139), | (('SNA80324', 'FRO53271--->', 'GRO73461'), 0.5867), |
| (('ELE17451', 'ELE92920--->', 'GRO73461'), 0.8984), | (('DAI75645', 'FRO53271--->', 'GRO73461'), 0.5857), |
| (('GRO15017', 'ELE92920--->', 'GRO73461'), 0.8882), | (('SNA80324', 'GRO81087--->', 'DAI22896'), 0.5838), |
| (('SNA18336', 'ELE17451--->', 'FRO47962'), 0.8539), | (('SNA59903', 'ELE17451--->', 'ELE11111'), 0.5755), |
| (('SNA18336', 'DAI62779--->', 'GRO32086'), 0.8538), | (('ELE17451', 'SNA90094--->', 'GRO73461'), 0.5659), |
| (('SNA53220', 'FRO19221--->', 'SNA99873'), 0.8452), | (('ELE78169', 'SNA45677--->', 'SNA90258'), 0.5583), |
| (('GRO59710', 'ELE92920--->', 'DAI22896'), 0.8227), | (('ELE20847', 'DAI62779--->', 'GRO73461'), 0.5564), |
| (('GRO73461', 'FRO47962--->', 'GRO73461'), 0.8102), | (('SNA80324', 'GRO94758--->', 'DAI22896'), 0.5505), |
| (('SNA18336', 'DAI85309--->', 'ELE20847'), 0.8035), | (('SNA80324', 'GRO30386--->', 'DAI22896'), 0.5504), |
| (('DAI75645', 'ELE92920--->', 'GRO73461'), 0.7927), | (('DAI75645', 'ELE17451--->', 'DAI22896'), 0.5484), |
| (('ELE32164', 'ELE92920--->', 'GRO73461'), 0.7366), | (('DAI75645', 'GRO30386--->', 'DAI22896'), 0.5481), |
| (('GRO73461', 'FRO19221--->', 'DAI88079'), 0.703), | (('SNA96271', 'ELE17451--->', 'GRO56726'), 0.5468), |
| (('SNA80324', 'ELE17451--->', 'SNA55952'), 0.6985), | (('DAI75645', 'GRO81087--->', 'GRO73461'), 0.5437), |
| (('ELE17451', 'DAI43223--->', 'GRO73461'), 0.6963), | (('ELE32164', 'ELE17451--->', 'ELE56788'), 0.5421), |
| (('DAI85309', 'ELE92920--->', 'GRO73461'), 0.6915), | (('DAI75645', 'ELE20847--->', 'DAI22896'), 0.5359), |
| (('ELE17451', 'GRO15017--->', 'GRO81647'), 0.6491), | (('DAI75645', 'GRO21487--->', 'GRO73461'), 0.5352), |
| (('SNA38068', 'ELE17451--->', 'DAI55148'), 0.6378), | (('DAI75645', 'GRO46854--->', 'GRO73461'), 0.5316), |
| (('ELE17451', 'DAI43223--->', 'GRO73461'), 0.6319), | (('SNA53220', 'DAI62779--->', 'GRO73461'), 0.5282), |
| (('DAI62779', 'DAI43223--->', 'GRO73461'), 0.6253), | (('SNA80324', 'GRO21487--->', 'GRO73461'), 0.5244), |
| (('ELE17451', 'GRO81087--->', 'GRO73461'), 0.6107), | (('DAI75645', 'GRO81087--->', 'DAI22896'), 0.5243), |
| (('GRO73461', 'DAI43223--->', 'GRO73461'), 0.6033), | (('ELE17451', 'SNA45677--->', 'GRO73461'), 0.5197), |
| (('DAI75645', 'ELE17451--->', 'SNA38068'), 0.5996), | (('SNA80324', 'SNA96271--->', 'FRO53271'), 0.5114), |
| (('ELE17451', 'ELE92920--->', 'GRO73461'), 0.5938), | (('GRO73461', 'SNA96271--->', 'SNA18336'), 0.5067), |
| | (('SNA80324', 'ELE17451--->', 'DAI22896'), 0.5025), |
| | (('DAI75645', 'GRO21487--->', 'GRO73461'), 0.5023), |
| | (('FRO40251', 'GRO81087--->', 'GRO73461'), 0.5022), |
| | (('GRO73461', 'GRO71621--->', 'FRO66272'), 0.5016), |
| | (('SNA18336', 'ELE92920--->', 'GRO73461'), 0.5011), |

بررسی یک مورد:

(('GRO81087' , 'ELE92920-->' , 'GRO73461') , 0.9571)

برای مثال اگر یک مشتری دوکالای (GRO81087) و (ELE92920) را به سبد خود اضافه کرده (یا مثلاً بازدید کرده) احتمالاً کالای (GRO73461) برای بازدید و یا اضافه کردن به سبد خرید مناسب باشد، چرا که مقدار کانفیدنس این rule برابر 0.9571 است که مقدار بالایی محسوب میشود، و هم چنین این rule از کالاهایی ساخته شده است که هر کدام پرتکرار هستند.

به همین صورت میتوان کالاها را به مشتری پیشنهاد داد و احتمالاً یک ریکامنדר سیستم با این منطق ساخت.

* با توجه به ابهامی که در صورت سوال برای پیدا کردن rule ها و محاسبه confidence score وجود داشت، من طبق روشی که در اسلاید درس گفته شده یود عمل کردم. بدین صورت که پس از پیدا کردن frequent itemset ، rule های ممکن از انرا استخراج کرده و سپس مقدار confidence score را برای هر یک محاسبه کردم. * در واقع هر ایتِم ست را دو بخش کرده و rule را برای آن مینویسیم. یک بخش در سمت چپ و بخش دیگر سمت راست قرار میگیرد.

Mining Association Rules

- **Step 1:** Find all frequent itemsets I
 - (we will explain this next)
- **Step 2:** Rule generation
 - For every subset A of I , generate a rule $A \rightarrow I \setminus A$
 - Since I is frequent, A is also frequent
 - **Variant 1:** Single pass to compute the rule confidence
 - $\text{confidence}(A, B \rightarrow C, D) = \text{support}(A, B, C, D) / \text{support}(A, B)$
 - **Variant 2:**
 - **Observation:** If $A, B, C \rightarrow D$ is below confidence, then so is $A, B \rightarrow C, D$
 - Can generate "bigger" rules from smaller ones!
 - **Output the rules above the confidence threshold**