

به نام خدا



دانشگاه صنعتی اصفهان

دانشکده مهندسی برق و کامپیوتر

عنوان: تکلیف پنجم داده کاوی

افروز شیخ الاسلامی

9729393

بهار 1401

مدرس: دکتر ناصر قدیری

فهرست مطالب

3..... سوال اول

3..... قسمت h

5..... سوال دوم

5..... قسمت g

6..... سوال سوم

6..... قسمت C

## سوال اول

### قسمت h

تفسیر یکی از قوانین به دست آمده:

در این نمودار، rule های به دست آمده قابل مشاهده است.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(BISCUITS - CORE & NON CORE)	(CLEANERS - UTENSIL)	0.283564	0.046631	0.011194	0.039477	0.846593	-0.002028	0.992553
1	(CLEANERS - UTENSIL)	(BISCUITS - CORE & NON CORE)	0.046631	0.283564	0.011194	0.240064	0.846593	-0.002028	0.942757
2	(SALTY SNACKS (2/97))	(BISCUITS - CORE & NON CORE)	0.106086	0.283564	0.025651	0.241789	0.852677	-0.004432	0.944903
3	(BISCUITS - CORE & NON CORE)	(SALTY SNACKS (2/97))	0.283564	0.106086	0.025651	0.090458	0.852677	-0.004432	0.982817
4	(BISCUITS - CORE & NON CORE)	(TOILET SOAPS (04/00))	0.283564	0.059901	0.013715	0.048366	0.807437	-0.003271	0.987879

برای تحلیل، rule با بیشترین support و confidence انتخاب کرده ام:

Salty Snacks → Biscuits, Core & Non-Core

یعنی افرادی که چیپس یا اسنک نمکی خریده اند، بیسکویت هم خریداری کرده اند. معیار support این قانون برابر 0.28 و confidence آن برابر 0.241 شده است.

قانون دیگری که جالب به نظر می رسد، قانون زیر است:

Biscuits, Core & Non-Core → Cleaners, Utensil

یعنی افرادی که بیسکویت خریداری کرده اند، به احتمال زیاد شوینده و وسایل آشپزخانه هم خریداری می کنند.

منظور از leverage و conviction:

$$\text{leverage}(A \rightarrow C) = \text{support}(A \rightarrow C) - \text{support}(A) \times \text{support}(C), \quad \text{range: } [-1, 1]$$

عبارت  $\text{support}(A \rightarrow C)$  در واقع مقدار پشتیبانی مشاهده شده را نشان می دهد. در حالی که عبارت  $\text{sup}(A) * \text{sup}(C)$  میزان پشتیبانی قابل انتظار را نشان می دهد. به طور کلی، leverage

، نشان دهنده تفاوت بین این دو مقدار است و اگر صفر باشد، یعنی مقدار پشتیبانی مورد انتظار با مقدار پشتیبانی مشاهده شده برابر است و A و C مستقل از هم هستند.

$$\text{conviction}(A \rightarrow C) = \frac{1 - \text{support}(C)}{1 - \text{confidence}(A \rightarrow C)}, \quad \text{range: } [0, \infty]$$

این معیار، احتمال ظاهر شدن A، بدون ظاهر شدن C در صورتی که مستقل باشند، با فراوانی واقعی آمدن A بدون وجود C را مقایسه می کند. مقدار بالا conviction، به معنی وابستگی زیاد consequent (نتیجه) به antecedent (مقدم) است. زیرا در حالتی که confidence بیشترین مقدار خود را داشته باشد، مخرج فرمول بالا صفر می شود و مقدار conviction برابر بینهایت می شود. بنابراین به طور کلی مقدار conviction=1 به این معنی است که آیتم ها به هم مرتبط نیستند، در حالی که مقدار conviction > 1 نشان دهنده رابطه قوی بین آیتم ها است و هرچه این مقدار بیشتر باشد، رابطه آن ها قوی تر است.

## سوال دوم

### قسمت g

قوانین به دست آمده از الگوریتم apriori :

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(Age_binned_(20.999, 25.0])	(Outcome_0)	0.347656	0.651042	0.289062	0.831461	1.277124	0.062724	2.070486
1	(Age_binned_(25.0, 36.0])	(Outcome_0)	0.321615	0.651042	0.196615	0.611336	0.939012	-0.012770	0.897841
2	(Age_binned_(36.0, 81.0])	(Outcome_0)	0.330729	0.651042	0.165365	0.500000	0.768000	-0.049954	0.697917
3	(Age_binned_(36.0, 81.0])	(Outcome_1)	0.330729	0.348958	0.165365	0.500000	1.432836	0.049954	1.302083
4	(BMI_binned_(-0.001, 28.9])	(Outcome_0)	0.338542	0.651042	0.287760	0.850000	1.305600	0.067356	2.326389
5	(BMI_binned_(28.9, 34.733])	(Outcome_0)	0.328125	0.651042	0.188802	0.575397	0.883810	-0.024821	0.821846
6	(BMI_binned_(34.733, 67.1])	(Outcome_0)	0.333333	0.651042	0.174479	0.523438	0.804000	-0.042535	0.732240
7	(BMI_binned_(-0.001, 28.9], Age_binned_(20.999, 25.0])	(Outcome_0)	0.154948	0.651042	0.149740	0.966387	1.484370	0.048862	10.381510
8	(BMI_binned_(-0.001, 28.9], Outcome_0)	(Age_binned_(20.999, 25.0])	0.287760	0.347656	0.149740	0.520362	1.496772	0.049698	1.360075
9	(Age_binned_(20.999, 25.0], Outcome_0)	(BMI_binned_(-0.001, 28.9])	0.289062	0.338542	0.149740	0.518018	1.530146	0.051880	1.372371
10	(BMI_binned_(28.9, 34.733], Age_binned_(20.999, 25.0])	(Outcome_0)	0.098958	0.651042	0.079427	0.802632	1.232842	0.015001	1.768056
11	(BMI_binned_(34.733, 67.1], Age_binned_(20.999, 25.0])	(Outcome_0)	0.093750	0.651042	0.059896	0.638889	0.981333	-0.001139	0.966346
12	(Outcome_1, Age_binned_(20.999, 25.0])	(BMI_binned_(34.733, 67.1])	0.058594	0.333333	0.033854	0.577778	1.733333	0.014323	1.578947
13	(BMI_binned_(-0.001, 28.9], Age_binned_(25.0, 36.0])	(Outcome_0)	0.095052	0.651042	0.075521	0.794521	1.220384	0.013638	1.698264
14	(BMI_binned_(28.9, 34.733], Age_binned_(25.0, 36.0])	(Outcome_1)	0.108073	0.348958	0.055990	0.518072	1.484625	0.018277	1.350911
15	(BMI_binned_(34.733, 67.1], Age_binned_(25.0, 36.0])	(Outcome_0)	0.118490	0.651042	0.069010	0.582418	0.894593	-0.008131	0.835663
16	(Age_binned_(36.0, 81.0], BMI_binned_(-0.001, 28.9])	(Outcome_0)	0.088542	0.651042	0.062500	0.705882	1.084235	0.004856	1.186458
17	(BMI_binned_(-0.001, 28.9], Outcome_1)	(Age_binned_(36.0, 81.0])	0.050781	0.330729	0.026042	0.512821	1.550575	0.009247	1.373766
18	(Age_binned_(36.0, 81.0], BMI_binned_(28.9, 34.733])	(Outcome_1)	0.121094	0.348958	0.063802	0.526882	1.509870	0.021545	1.376065
19	(Age_binned_(36.0, 81.0], BMI_binned_(34.733, 67.1])	(Outcome_1)	0.121094	0.348958	0.075521	0.623656	1.787193	0.033264	1.729911

قوانین به دست آمده از الگوریتم fpgrowth :

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(Age_binned_(36.0, 81.0])	(Outcome_1)	0.330729	0.348958	0.165365	0.500000	1.432836	0.049954	1.302083
1	(Age_binned_(36.0, 81.0])	(Outcome_0)	0.330729	0.651042	0.165365	0.500000	0.768000	-0.049954	0.697917
2	(BMI_binned_(-0.001, 28.9], Outcome_1)	(Age_binned_(36.0, 81.0])	0.050781	0.330729	0.026042	0.512821	1.550575	0.009247	1.373766
3	(Age_binned_(36.0, 81.0], BMI_binned_(-0.001, 28.9])	(Outcome_0)	0.088542	0.651042	0.062500	0.705882	1.084235	0.004856	1.186458
4	(Age_binned_(36.0, 81.0], BMI_binned_(34.733, 67.1])	(Outcome_1)	0.121094	0.348958	0.075521	0.623656	1.787193	0.033264	1.729911
5	(BMI_binned_(28.9, 34.733])	(Outcome_0)	0.328125	0.651042	0.188802	0.575397	0.883810	-0.024821	0.821846
6	(Age_binned_(36.0, 81.0], BMI_binned_(28.9, 34.733])	(Outcome_1)	0.121094	0.348958	0.063802	0.526882	1.509870	0.021545	1.376065
7	(BMI_binned_(28.9, 34.733], Age_binned_(20.999, 25.0])	(Outcome_0)	0.098958	0.651042	0.079427	0.802632	1.232842	0.015001	1.768056
8	(BMI_binned_(-0.001, 28.9])	(Outcome_0)	0.338542	0.651042	0.287760	0.850000	1.305600	0.067356	2.326389
9	(BMI_binned_(-0.001, 28.9], Age_binned_(20.999, 25.0])	(Outcome_0)	0.154948	0.651042	0.149740	0.966387	1.484370	0.048862	10.381510
10	(BMI_binned_(-0.001, 28.9], Outcome_0)	(Age_binned_(20.999, 25.0])	0.287760	0.347656	0.149740	0.520362	1.496772	0.049698	1.360075
11	(Age_binned_(20.999, 25.0], Outcome_0)	(BMI_binned_(-0.001, 28.9])	0.289062	0.338542	0.149740	0.518018	1.530146	0.051880	1.372371
12	(Age_binned_(25.0, 36.0])	(Outcome_0)	0.321615	0.651042	0.196615	0.611336	0.939012	-0.012770	0.897841
13	(BMI_binned_(-0.001, 28.9], Age_binned_(25.0, 36.0])	(Outcome_0)	0.095052	0.651042	0.075521	0.794521	1.220384	0.013638	1.698264
14	(BMI_binned_(34.733, 67.1], Age_binned_(25.0, 36.0])	(Outcome_0)	0.118490	0.651042	0.069010	0.582418	0.894593	-0.008131	0.835663
15	(BMI_binned_(28.9, 34.733], Age_binned_(25.0, 36.0])	(Outcome_1)	0.108073	0.348958	0.055990	0.518072	1.484625	0.018277	1.350911
16	(Age_binned_(20.999, 25.0])	(Outcome_0)	0.347656	0.651042	0.289062	0.831461	1.277124	0.062724	2.070486
17	(BMI_binned_(34.733, 67.1])	(Outcome_0)	0.333333	0.651042	0.174479	0.523438	0.804000	-0.042535	0.732240
18	(Outcome_1, Age_binned_(20.999, 25.0])	(BMI_binned_(34.733, 67.1])	0.058594	0.333333	0.033854	0.577778	1.733333	0.014323	1.578947
19	(BMI_binned_(34.733, 67.1], Age_binned_(20.999, 25.0])	(Outcome_0)	0.093750	0.651042	0.059896	0.638889	0.981333	-0.001139	0.966346

قواعد به دست آمده از دو الگوریتم، کاملاً مشابه هم هستند.

سوال سوم

قسمت C

اعداد زیر:

44, 13, 11