

گزارش تکلیف دوم

مهسا ناقدی نیا 943611113023

سروناز رحمتی 943611113007

الگوریتم FP-growth را به کمک نرم افزار Rapid Miner پیاده سازی کردیم.

ابتدا با ماژول Read CSV داده ها را از فایل داده هایمان می خوانیم.

سپس ماژول Generate ID را میگذاریم تا برای داده های نتیجه مان آی دی از یک تا تعداد خروجی هایمان تولید کند.

برای استفاده از ماژول Create Association Rule ، داده هایمان باید به صورت باینری و یا true و false به ماژول داده شوند. چون این ماژول چنین ورودی هایی از ما قبول میکند. بدین منظور از ماژول Nominal to Binominal پس از آن استفاده میکنیم تا داده هایمان را از اسمی به باینری تبدیل میکند.

خروجی این روند تا اینجا دیتاستی است که تمام دیتاست به صورت باینری درآمده اند. هر کالا به صورت یک ستون میشود و در صورت خرید آن کالا توسط یک مشتری مقدار آن true و در غیر اینصورت false میشود.

حال این خروجی را به ماژول FP-growth میدهیم تا الگو های پرتکرار را برایمان بدست بیاورد و از آن بتوانیم قوانین انجمنی استخراج کنیم.

و در آخر ماژول Create Association Rules را میگذاریم تا از درخت FP-Growth ساخته شده در مرحله ی قبل قوانین انجمنی استخراج کند.

در ماژول fp-growth، $\text{min support} = 0.005$ در نظر گرفتیم و در ماژول create association rule، $\text{min confidence} = 0.3$ است.

خروجی مان قوانین استخراج شده همراه Support و Confidence و Lift و دیگر اطلاعات مربوط به قوانین خواهد بود.

خروجی برنامه به شکل زیر میشود که با توجه به min_sup و min_conf اولیه ۲۴ قانون انجمنی برایمان استخراج میکند. (البته میتوانیم این تعداد را با تغییر دادن مقدار min_sup و min_conf بیشتر یا کمتر کنیم)

No.	Premises	Conclusion	Support	Confiden...	LaPlace	Gain	p-s	Lift	Convicti...
1	Bought = other vegetables	att2 = whole milk	0.014	0.300	0.969	-0.080	0.011	4.511	1.334
2	att3 = other vegetables, att4 = whole milk	att2 = root vegetables	0.005	0.300	0.988	-0.029	0.005	7.704	1.373
3	att3 = yogurt	att2 = whole milk	0.009	0.315	0.980	-0.050	0.007	4.735	1.363
4	att2 = tropical fruit	Bought = citrus fruit	0.012	0.321	0.976	-0.061	0.010	6.972	1.405
5	att2 = root vegetables	att3 = other vegetables	0.013	0.321	0.975	-0.065	0.011	7.611	1.411
6	att2 = other vegetables	att3 = whole milk	0.018	0.327	0.964	-0.094	0.015	6.361	1.410
7	att4 = other vegetables	att3 = root vegetables	0.009	0.343	0.983	-0.043	0.008	16.196	1.489
8	att3 = whole milk	att2 = other vegetables	0.018	0.356	0.968	-0.085	0.015	6.361	1.465
9	att4 = yogurt	att3 = whole milk	0.007	0.361	0.987	-0.034	0.006	7.024	1.485
10	att3 = tropical fruit	att2 = citrus fruit	0.006	0.392	0.991	-0.023	0.005	16.048	1.604
11	att4 = other vegetables	att5 = whole milk	0.011	0.409	0.985	-0.041	0.010	27.026	1.668
12	att3 = other vegetables	att4 = whole milk	0.017	0.410	0.976	-0.067	0.016	12.790	1.640
13	att3 = other vegetables, att2 = root vegetables	att4 = whole milk	0.005	0.415	0.993	-0.020	0.005	12.946	1.654
14	att3 = root vegetables	att4 = other vegetables	0.009	0.418	0.988	-0.033	0.008	16.196	1.675
15	att3 = curd	att2 = whole milk	0.005	0.510	0.995	-0.015	0.005	7.669	1.905
16	att3 = pip fruit	att2 = tropical fruit	0.008	0.531	0.993	-0.022	0.007	14.700	2.054
17	att4 = whole milk	att3 = other vegetables	0.017	0.540	0.986	-0.047	0.016	12.790	2.081
18	att5 = other vegetables	att6 = whole milk	0.007	0.575	0.995	-0.017	0.007	65.001	2.332
19	att4 = curd	att3 = whole milk	0.005	0.586	0.996	-0.013	0.005	11.394	2.292
20	att4 = butter	att3 = whole milk	0.006	0.667	0.997	-0.012	0.006	12.958	2.846
21	att5 = whole milk	att4 = other vegetables	0.011	0.698	0.995	-0.020	0.010	27.026	3.226
22	att6 = whole milk	att5 = other vegetables	0.007	0.793	0.998	-0.011	0.007	65.001	4.774
23	att2 = root vegetables, att4 = whole milk	att3 = other vegetables	0.005	0.836	0.999	-0.007	0.005	19.814	5.843
24	att2 = sausage	Bought = frankfurter	0.010	1	1	-0.010	0.009	16.957	∞

همچنین با خروجی گرفتن از ماژول FP-Growth به تنهایی الگوهای پرتکرار را بدست میاوریم که خیلی زیاد است در یک عکس جا نمیشود که ضمیمه کنیم!

ماژول Create association rules قوانین جذاب را به ما ارائه میدهد. یعنی برای هر قانونی lift را هم کنترل میکند و چک میکند که الگوها با یکدیگر وابستگی داشته باشند. بنابراین تمام این قوانین استخراج شده توسط ماژول، جذاب هستند و میتوان آنها را به مرکز خرید پیشنهاد داد.

۵ تا از قوانینی که lift آنها از بقیه بزرگتر است و بدین دلیل قوی تر و جذاب تر اند به شکل زیر است:

1. whole milk -> other vegetables
support = 0.007 confidence = 0.79 lift = 65.0
2. tropical fruit -> citrus fruit
support = 0.005 confidence = 0.392 lift = 27.02
3. yogurt -> whole milk
support = 0.007 confidence = 0.361 lift = 7.024
4. sausage -> frankfurter
support = 0.010 confidence = 1 lift = 16.95
5. curd -> whole milk
support = 0.005 confidence = 0.510 lift = 7.66

