

"بسمه تعالی"

## گزارش پروژه پیاده سازی الگوریتم CMAR

درس : داده کاوی

استاد : جناب آقای دکتر احمدی

تهیه کننده : فرخنده زینالی آق قلعه

شماره دانشجویی : ۹۶۱۱۲۷۴

در این پروژه قصد داریم با استفاده از الگوریتم CMAR دسته بندی تراکنش های یک Supermarket را انجام دهیم. فایل داده این پروژه به فرمت arff است اطلاعات آن شامل خصوصیات Attribute و داده ها Data می باشد. این پروژه با استفاده از نرم افزار Microsoft Visual Studio 2010 نوشته شده است که در زیر روند انجام عملیات توضیح داده شده است.

## خواندن فایل داده های سوپرمارکت

همانطور که گفته شد فایل داده ها به فرمت arff می باشد که توسط برنامه خواند شده و خصوصیات یا آیتم ها به اضافه آخرین خصوصیت که همان کلاس های دسته بندی می باشند و همینطور داده های مربوط به تراکنش ها بدست می آید.

پس از انتخاب فایل و کلیک کردن بر روی دکمه بارگزاری تراکنش ها عملیات خواندن اطلاعات انجام می شود. همانطور که جدول مربوطه ملاحظه می کنید تمامی آیتم ها مقدار کلاس false دارند غیر از آخرین آیتم که همان دسته بندی ها می باشد (مقادیر low و high).

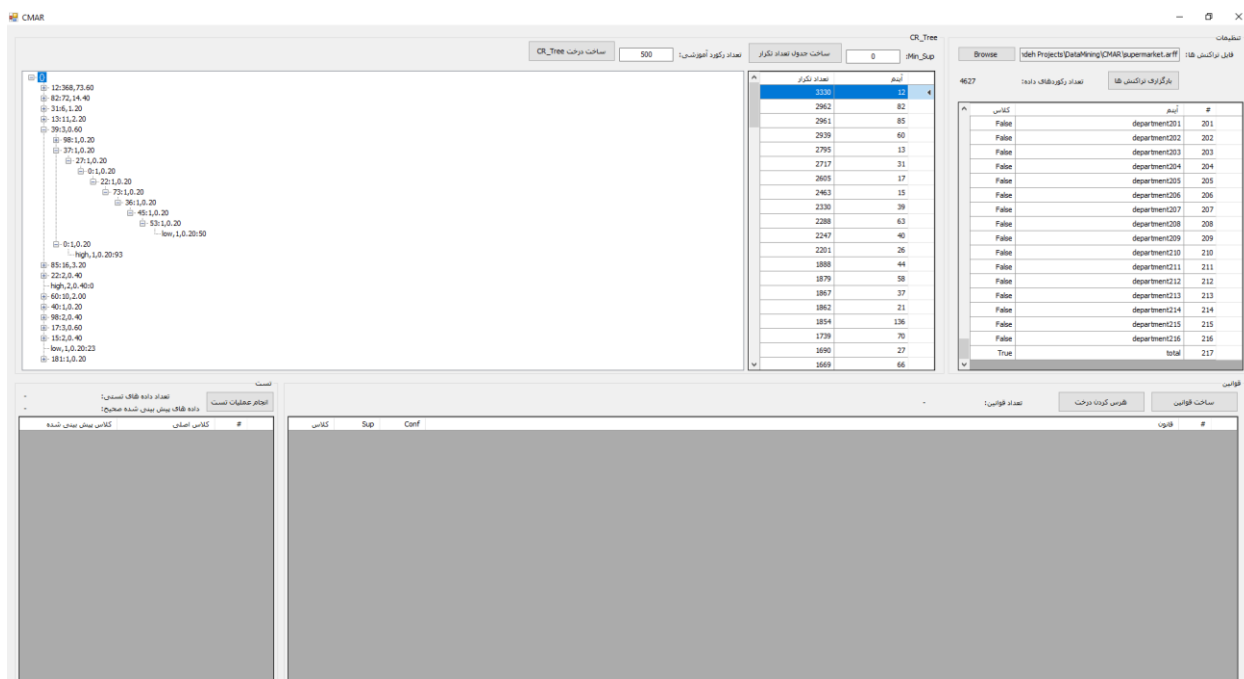
The screenshot shows the CMAR software interface. The main window is titled "CMAR" and contains several panels. The top right panel, labeled "تنظیمات" (Settings), shows the file path "C:\deh Projects\DataMining\CMAR\supermarket.arff" and the number of records "4627". A button labeled "بارگزاری تراکنش ها" (Load Transactions) is highlighted. The bottom right panel, labeled "قوانین" (Rules), shows a table with columns "#", "کلاس" (Class), and "آیتم" (Item). The table lists 217 items, with the last item being "total" with a class of "True". The bottom left panel, labeled "تست" (Test), shows a table with columns "#", "کلاس اصلی" (Main Class), and "کلاس پیش بینی شده" (Predicted Class). The bottom middle panel, labeled "قوانین" (Rules), shows a table with columns "#", "قانون" (Rule), "Sup", and "Conf".

#	کلاس	آیتم
201	False	department201
202	False	department202
203	False	department203
204	False	department204
205	False	department205
206	False	department206
207	False	department207
208	False	department208
209	False	department209
210	False	department210
211	False	department211
212	False	department212
213	False	department213
214	False	department214
215	False	department215
216	False	department216
217	True	total

## ساخت درخت CR\_Tree

پس از بارگزاری اطلاعات جهت رسیدن به قوانین دسته بندی باید یک درخت از داده ها را توسعه داد که در اینجا از درخت FP\_Growth استفاده شده است که مقادیر Sup و Conf نیز به آن اضافه شده اند که به ترتیب تعداد تکرار آیتم مربوطه و درصد تکرار آن در کل داده ها می باشند.

برای ساخت درخت FP\_Tree مقدار Min\_Sup که مربوط به الگوریتم FP\_Growth است را وارد کرده سپس با کلیک بر روی ساخت جدول تعداد تکرار، جدول مربوطه ساخته می شود و پس از آن با وارد کردن تعداد رکوردهای تست که برای ساخت درخت CR\_Tree و آموزش مورد استفاده قرار می گیرد در خت مربوطه توسعه داده می شود. در زیر درخت CR\_Tree که با Min\_Sup برابر ۰ و تعداد ۵۰۰ رکورد آزمایشی را ملاحظه می کنید.



## ساخت قوانین انجمنی

پس از ساخت درخت CR\_Tree نوبت به پیمایش و استخراج قوانین از این درخت می رسد که با کلیک بر روی دکمه ساخت قوانین و اتمام این عملیات قوانین استخراج شده به صورت جدول مربوطه که در زیر مشاهده می کنید نمایش داده می شود.

CMAR

CR\_Tree

ساخت درخت CR\_Tree: 500

تعداد رکورد آموزش: 0

ساخت جدول تعداد بزرگتر

تعداد رکورد های داده: 4627

فایل ترانژکشن ها: C:\Users\user\Documents\Idea Projects\CMAR\supermarket.arff

توضیحات

فایل ترانژکشن ها

تعداد رکورد های داده: 4627

کلاس

#	کلاس
201	department201
202	department202
203	department203
204	department204
205	department205
206	department206
207	department207
208	department208
209	department209
210	department210
211	department211
212	department212
213	department213
214	department214
215	department215
216	department216
217	total

کلاس

تعداد داده های آموزش: 469

تعداد رکورد درخت: 469

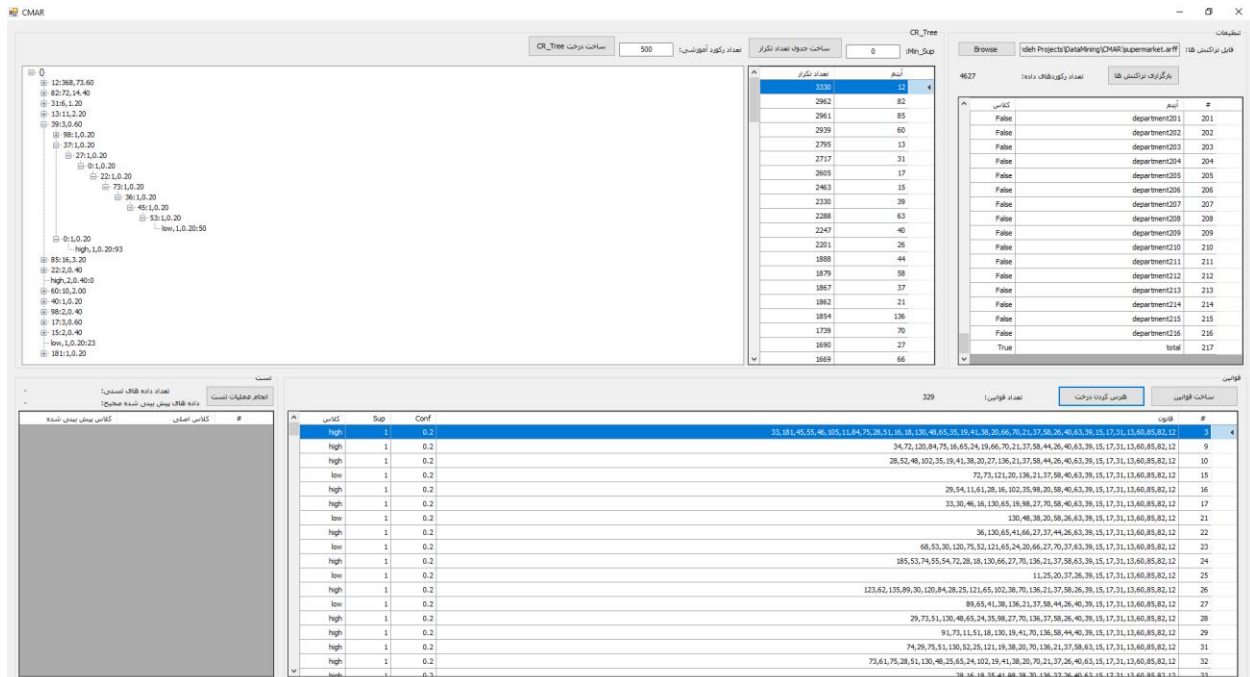
ساخت قوانین

قوانین

#	قانون
1	high, 1, 0.2, 1.94, 129.25, 72, 105.9, 1, 120.6, 184.28, 16, 130.0, 52, 40, 121.65, 102.35, 41, 98.38, 20.66, 27, 70, 136.21, 37.58, 26, 40, 63.39, 15, 17, 31, 13.60, 85.82, 12
2	high, 1, 0.2, 35, 18, 1, 45, 35, 46, 105, 11, 84, 75, 20, 51, 16, 136, 136, 40, 65, 35, 19, 41, 38, 20, 66, 70, 21, 37, 58, 26, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12
3	high, 1, 0.2, 75, 120, 61, 22, 94, 26, 16, 130, 0, 121, 65, 24, 102, 19, 41, 38, 20, 66, 70, 21, 37, 58, 26, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12
4	high, 1, 0.2, 129, 45, 74, 75, 120, 61, 16, 136, 136, 120, 85, 24, 35, 19, 41, 38, 66, 70, 37, 58, 26, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12
5	high, 1, 0.2, 34, 22, 75, 51, 130, 52, 40, 25, 102, 19, 41, 38, 20, 66, 70, 136, 21, 58, 26, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12
6	high, 1, 0.2, 120, 51, 16, 130, 0, 52, 40, 25, 121, 102, 35, 41, 98, 38, 66, 136, 21, 58, 26, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12
7	high, 1, 0.2, 9, 32, 74, 54, 46, 72, 105, 36, 91, 22, 75, 51, 18, 40, 25, 121, 102, 35, 41, 98, 38, 66, 70, 21, 37, 58, 44, 26, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12
8	high, 1, 0.2, 34, 72, 120, 84, 75, 16, 65, 24, 19, 66, 70, 21, 37, 58, 44, 26, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12
9	high, 1, 0.2, 28, 52, 40, 102, 35, 19, 41, 38, 20, 27, 136, 21, 37, 58, 44, 26, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12
10	high, 1, 0.2, 30, 54, 46, 105, 36, 91, 61, 29, 51, 52, 40, 65, 24, 102, 35, 19, 41, 98, 38, 20, 136, 21, 37, 58, 44, 26, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12
11	high, 1, 0.2, 90, 104, 89, 54, 120, 11, 28, 51, 0, 52, 40, 25, 121, 35, 19, 41, 38, 20, 66, 70, 136, 21, 37, 58, 44, 26, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12
12	high, 1, 0.2, 71, 62, 104, 93, 76, 91, 75, 61, 84, 28, 16, 136, 40, 25, 65, 24, 19, 25, 27, 70, 136, 37, 58, 44, 26, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12
13	high, 1, 0.2, 209, 30, 46, 105, 36, 70, 11, 84, 51, 18, 52, 102, 35, 19, 38, 20, 27, 21, 37, 44, 26, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12
14	high, 1, 0.2, 72, 73, 121, 20, 136, 21, 37, 58, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12
15	high, 1, 0.2, 29, 54, 11, 61, 28, 56, 102, 35, 98, 20, 58, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12
16	high, 1, 0.2, 33, 30, 46, 16, 130, 65, 19, 98, 27, 70, 58, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12
17	high, 1, 0.2, 30, 46, 16, 130, 65, 19, 98, 27, 70, 58, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12
18	high, 1, 0.2, 30, 46, 16, 130, 65, 19, 98, 27, 70, 58, 40, 63, 39, 15, 17, 31, 13.60, 85.82, 12

## هرس کردن قوانین انجمنی

پس از استخراج قوانین از روی درخت CR\_Tree جهت بالا رفتن کارایی الگوریتم با استفاده از سه روش توضیح داده شده در فایل ضمیمه پروژه به هرس کردن (Pruning) قوانین اقدام می کنیم که همانطور که ملاحظه می کنید با کلیک بر روی دکمه هرس کردن درخت این عملیات انجام شده و تعداد قوانین استخراج شده با تنظیمات فوق الذکر از ۴۹۹ به ۳۲۹ کاهش می یابد.



## دسته بندی داده های تستی

پس از استخراج قوانین و هرس آنها باقیمانده داده ها را به الگوریتم داده تا دسته آنها مشخص گردد. پس از کلیک بر روی انجام عملیات تست و اتمام آن تعداد داده های تستی که برای آنها قانون منطبق یافته شده و تعداد و درصد پیش بینی صحیح از میان آن داده های تستی به همراه جدول مربوطه قابل رویت خواهد بود که در شکل زیر ملاحظه می فرمایید.

