"بسمه تعالی"

گزارش پروژه پیاده سازی الگوریتم CMAR

درس : داده کاوی

استاد : جناب آقای دکتر احمدی

تهیه کننده : فرخنده زینالی آق قلعه

شماره دانشجویی : 9611274

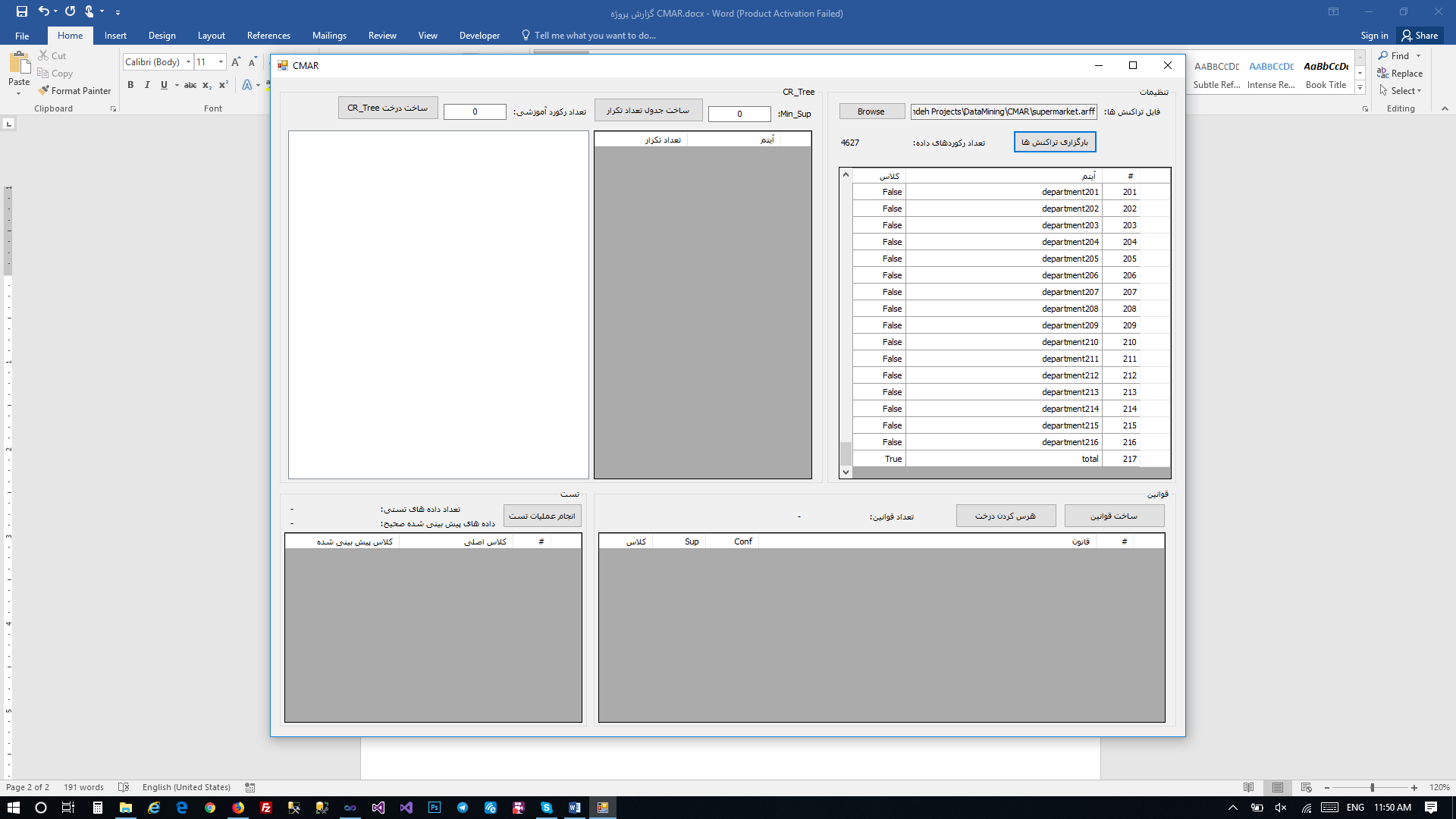
بهار 98

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.

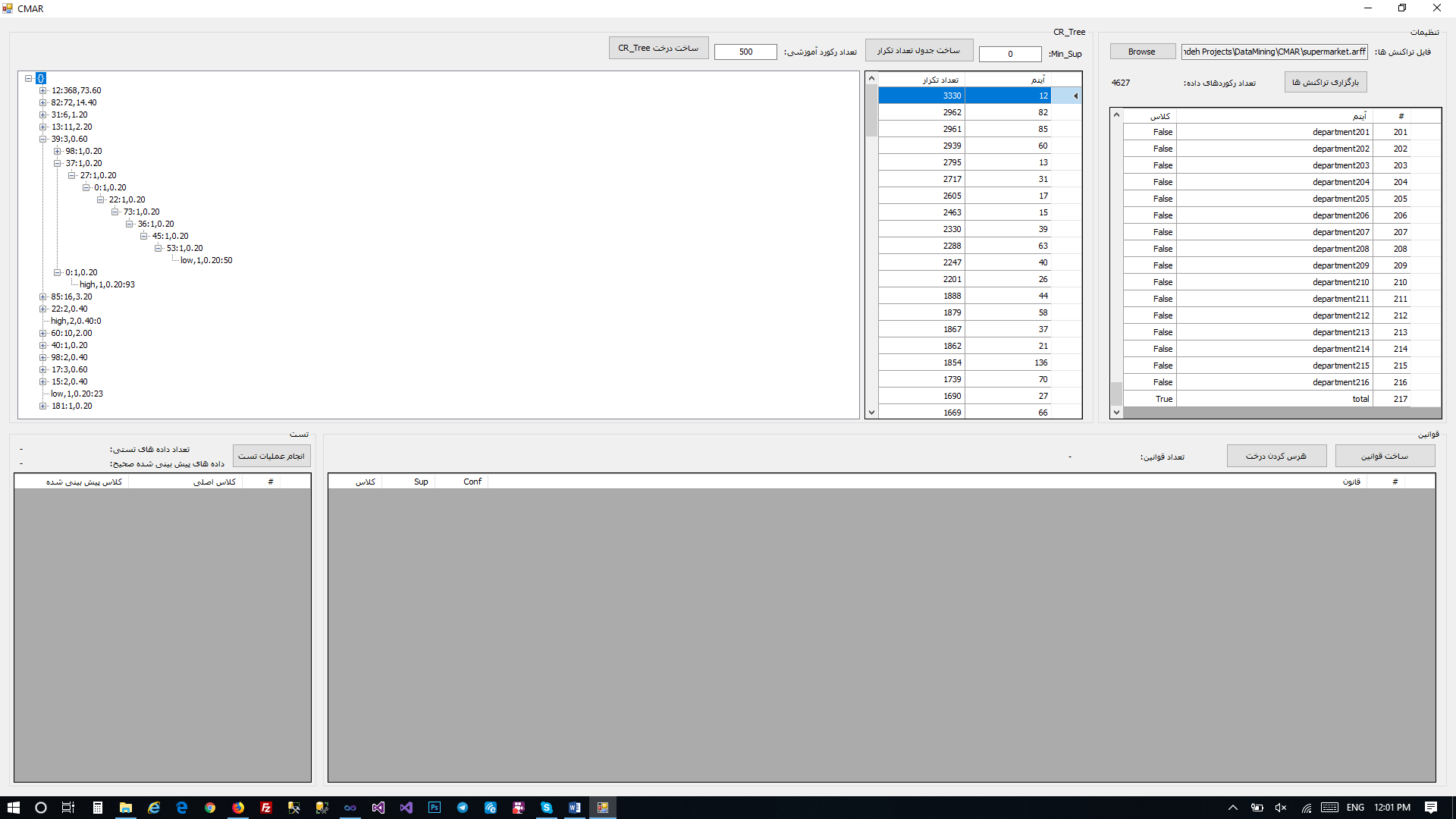
پس از انتخاب فایل و کلیک کردن بر روی دکمه بارگزاری تراکنش ها عملیات خواندن اطلاعات انجام می شود. همانطور که جدول مربوطه ملاحظه می کنید تمامی آیتم ها مقدار کلاس false دارند غیر از آخرین آیتم که همان دسته بندی ها می باشد (مقادیر low و high).



**ساخت درخت CR\_Tree**

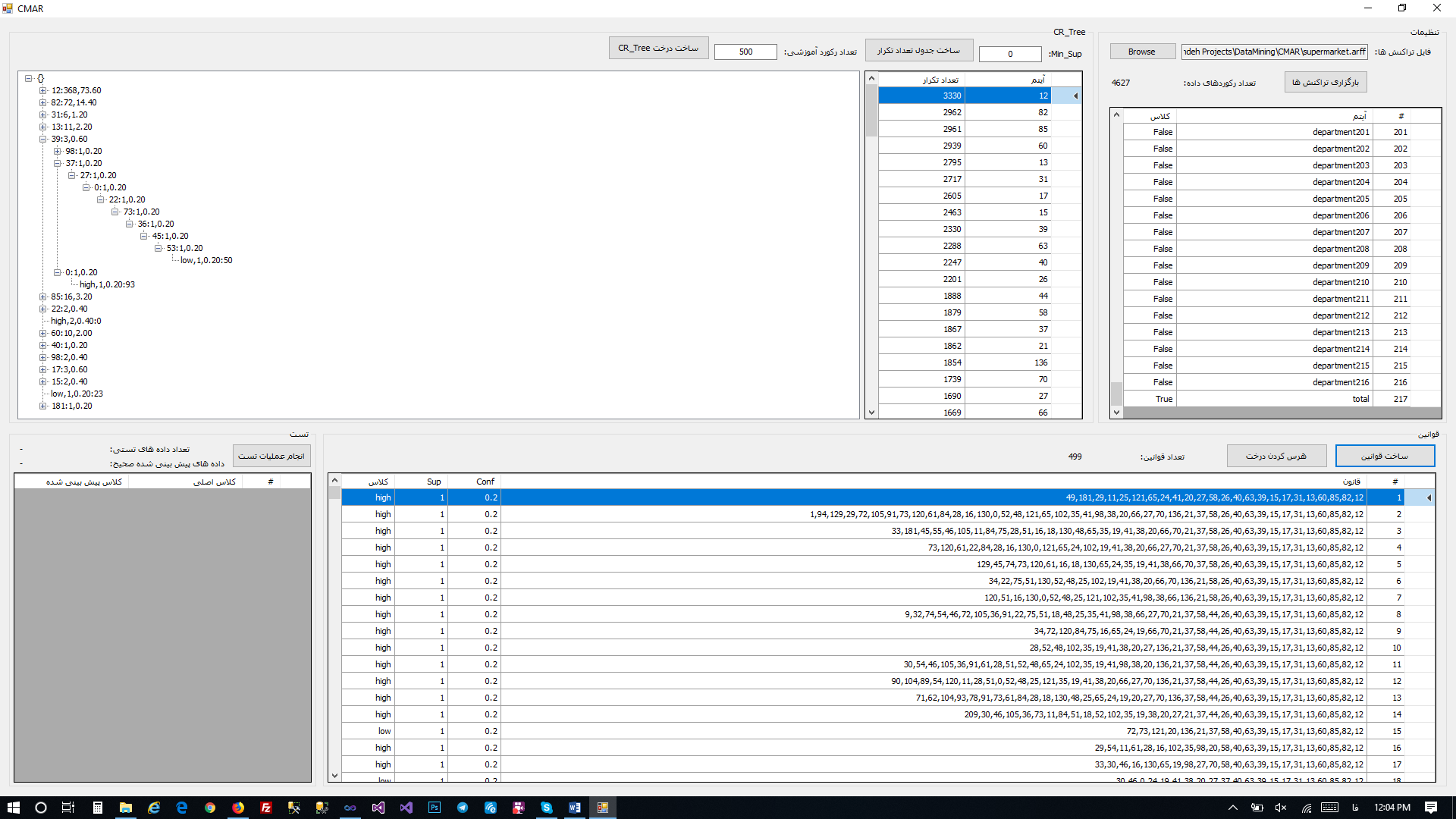
پس از بارگزاری اطلاعات جهت رسیدن به قوانین دسته بندی باید یک درخت از داده ها را توسعه داد که در اینجا از درخت FP\_Growth استفاده شده است که مقادیر Sup و Conf نیز به آن اضافه شده اند که به ترتیب تعداد تکرار آیتم مربوطه و درصد تکرار آن در کل داده ها می باشند.

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 0 و تعداد 500 رکورد آزمایشی را ملاحظه می کنید.



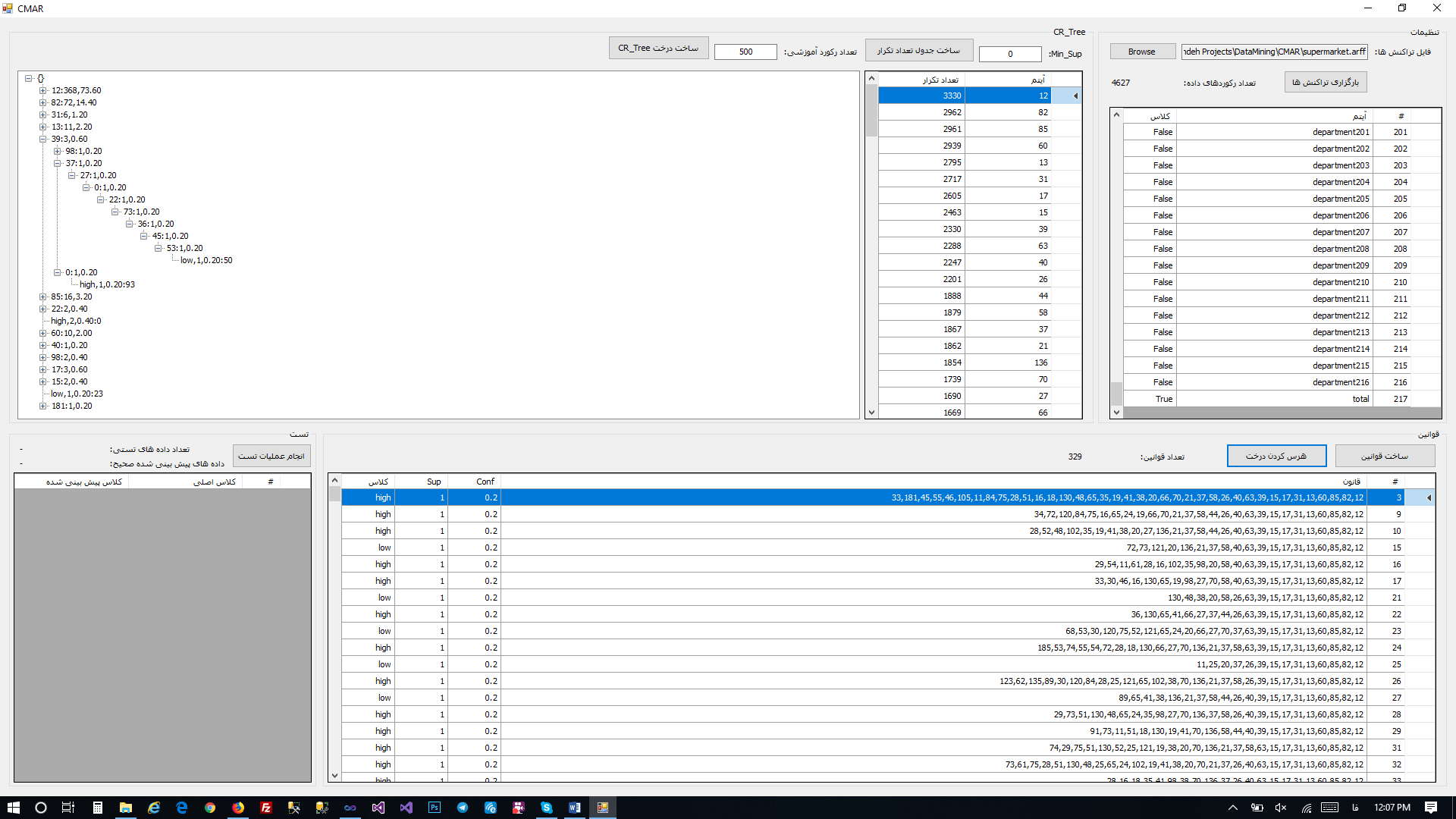
**ساخت قوانین انجمنی**

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 مشاه ده می کنید نمایش داده می شود.



**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 499 به 329 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 تعداده 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.

