

"بسمه تعالی"

پیاده سازی دسته بندی سلسله مراتبی به روش AGNES

درس : داده کاوی

استاد : جناب آقای دکتر احمدی

تهیه کننده : فرخنده زینالی آق قلعه

شماره دانشجویی : ۹۶۱۱۲۷۴

خوشه بندی (clustering) سلسله مراتبی بر دو نوع می باشد یکی تجمیعی و دیگری تقسیمی که در اینجا یکی از روش های تجمیعی به نام AGNES را مورد استفاده قرار می دهیم.

در روش های تجمیعی ابتدا هر نمونه در یک خوشه قرار گرفته سپس با ادغام خوشه ها بر اساس معیار مشخصی خوشه ها کمتر شده تا در نهایت همه نمونه ها در یک خوشه قرار گیرند.

در پیاده سازی حاضر دیتاستی که باید مورد بررسی قرار گیرد ابتدا می بایست جهت استفاده در برنامه تنظیم و تصحیح شود. مراحل تنظیم دیتاست به شرح زیر انجام می شود:

- از آنجایی که برنامه برای داده های دو بعدی نوشته شده است لذا از میان خصوصیات نمونه ها در فایل دیتاست خصوصیت های ForkVA و ForkW را که اعدادی اعشاری از نوع real هستند را نگه می داریم.

- انواع نونه ها که شامل موارد

'Bank' , 'AutomobileIndustry' , 'BpoIndustry' , 'CementIndustry' , 'Farmers1' , 'Farmers2' , 'HealthCareResources' , 'TextileIndustry' , 'PoultryIndustry' , 'Residential(individual)' , 'Residential(Apartments)' , 'FoodIndustry' , 'ChemicalIndustry' , 'Handlooms' , 'FertilizerIndustry' , 'Hostel' , 'Hospital' , 'Supermarket' , 'Theatre' , 'University'

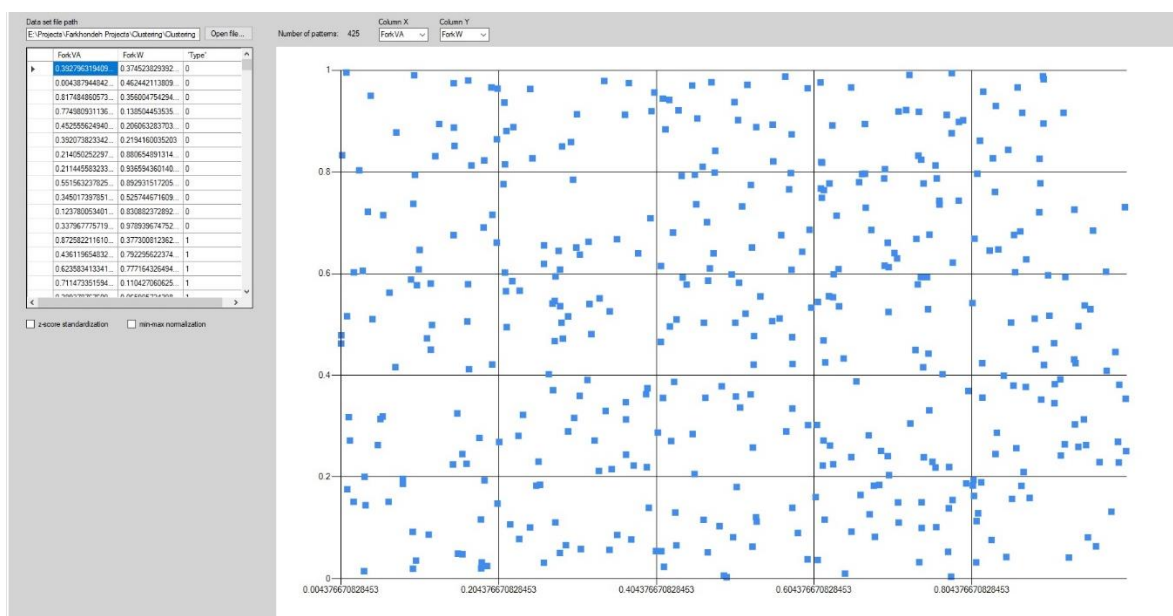
می باشد را به ترتیب از شماره 0 تا 19 جایگزین می کنیم.

- از آنجایی که تعداد نمونه ها بسیار زیاد است لذا به صورت رندوم تعدادی از آنها را انتخاب کرده و بقیه را حذف می کنیم.

پس از تنظیم دیتاست می توانیم مراحل خوشه بندی را آغاز کنیم.

۱- معرفی فایل دیتاست به برنامه

فایل دیتا ست را به برنامه معرفی می کنیم. سپس برنامه با خواندن دیتا ست جدول نمونه ها را پر کرده و همچنین روی دستگاه مختصات نیز آنها را نمایش می دهد.



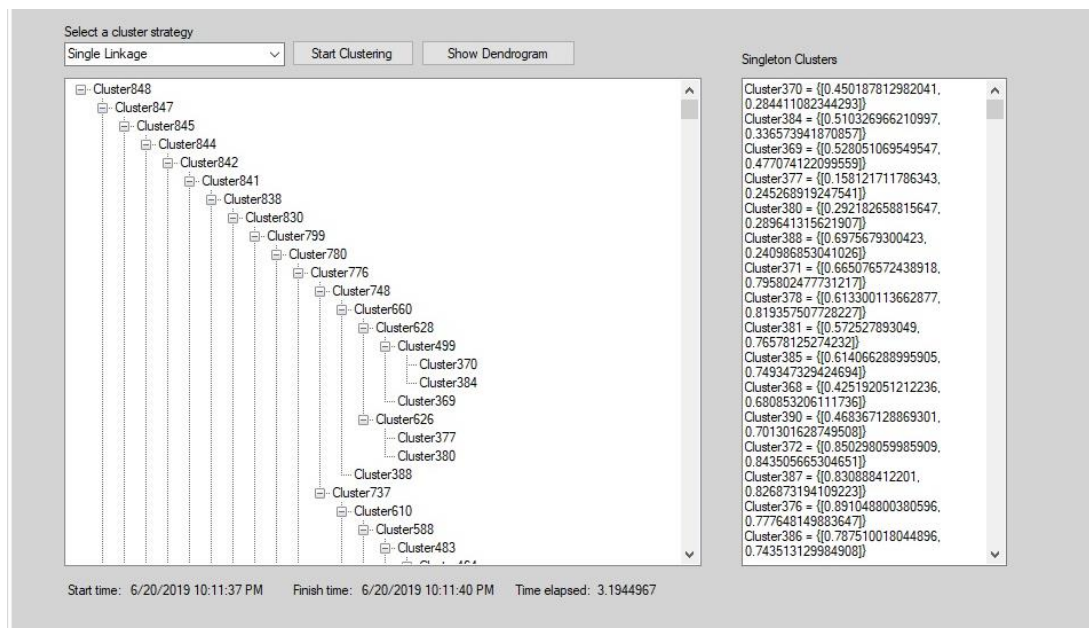
۲- عملیات خوشه بندی

در این مرحله عملیات خوشه بندی انجام می شود همانطور می دانید خوشه بندی بر اساس معیار های خاصی می تواند انجام شود که انواع آن به شرح زیر است:

- Single Linkage
- Complete Linkage
- Average Linkage WPGMA
- Average Linkage UPGMA

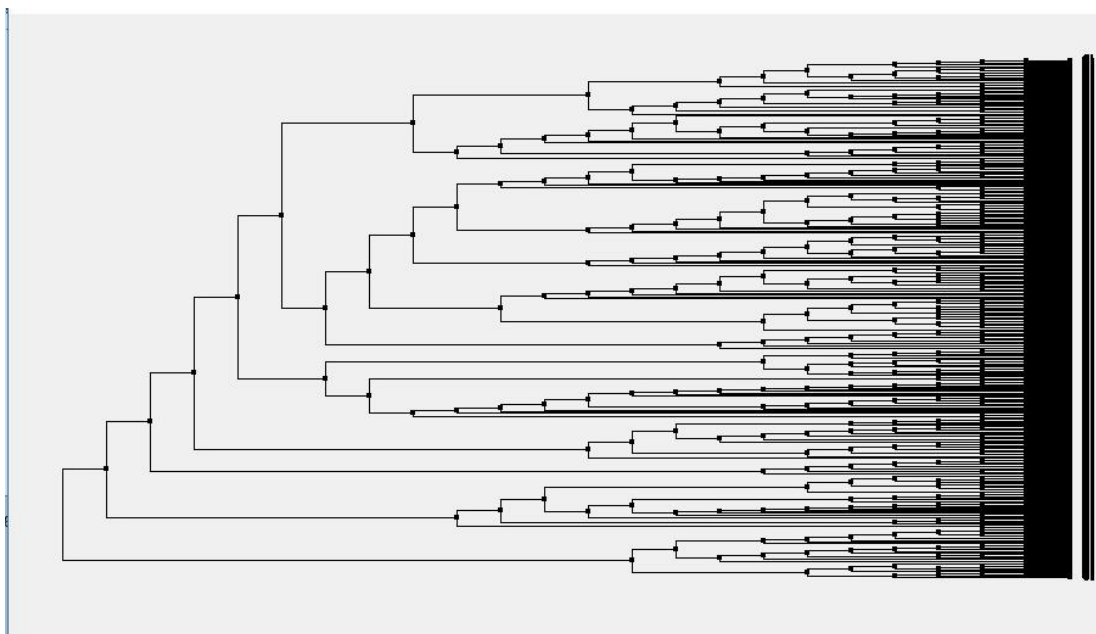
گزینه اول یعنی single linkage را انتخاب کرده عملیات خوشه بندی را انجام می دهیم.

پس از پایان خوشه بندی درختی از خوشه ها که سلسله مراتب خوشه بندی را نمایش می دهد قابل ملاحظه خواهد بود.



۳- مشاهده دندروگرام

همانطور که می دانیم جهت نمایش خوشه بندی های سلسله مراتبی می توان از دندروگرام استفاده نمود که نمایش مناسبی برای خوشه بندی محسوب می شود پس از انجام عملیات خوشه بندی دندروگرام خوشه ها نیز قابل مشاهده خواهد بود.



۴- مشاهده خوشه بندی نمونه ها

پس از انجام عملیات خوشه بندی نمونه ها در هر یک خوشه قرار گرفته می شوند که می توان آنها را در دستگاه مختصات مشاهده و اعتبارسنجی نمود. بر اساس اینکه چند خوشه مورد نظرمان باشد می توان نمونه ها را تنظیم کرد. این یکی از قابلیت های منحصر بفرد خوشه بندی سلسله مراتبی می باشد. در شکل های زیر نمونه ها را در ۲۰ خوشه و ۵ خوشه ملاحظه می کنید.

