



دانشگاه تهران  
پردیس فارابی  
دانشکده مهندسی  
گروه مهندسی کامپیوتر

مبانی داده کاوی

گزارش تمرین دوم

محمد مهدی حیدری

220797042

بسمه تعالی

مقاله مورد نیاز حل این تمرین به طور کامل خوانده شد.

با توجه به مقداری دشوار بودن فهم الگوریتم، بیشتر تمرکز در بخش پیاده‌سازی الگوریتم، بر روی قسمت شبهه کدهای الگوریتم قرار گرفت.

مقاله از هفت الگوریتم استفاده کرده است که شبهه کد آن‌ها را در خود مقاله آورده است.

و همچنین برای بهبود کد و گرفتن خروجی مورد انتظار از کد قرار گرفته بر روی گیت هاب مرتبط با این مقاله نیز استفاده شد و از آن کمک گرفته شد.

ابتدا کد فایل `executiveCode` نوشته شد که در آن ابتدا، دیتاست‌های مورد نظر خوانده و سپس به فایل دیگری به نام `sostream` ارسال می‌شوند تا عملیات خوشه بندی انجام شود.

سپس نمودارهای دو دیتاست موردنظر نمایش داده می‌شود و مراکز خوشه‌ها در آن‌ها با رنگ دیگری مشخص می‌شوند.

و همچنین موارد خواسته شده در صورت تمرین نیز بدست آورده می‌شود.

در فایل `sostream` مقادیر موردنیاز کد ایجاد می‌شوند و با استفاده از یک حلقه بر روی تمام ردیف‌های دیتاست‌ها، تمام ردیف‌ها به تابع `sostream` ارسال و پردازش می‌شوند.

تابع `sostream` با استفاده از شبهه کد مقاله و کمک از گیت هاب معرفی شده پیاده سازی گردیده است که در آن ابتدا یک نقطه را به عنوان `win` در نظر می‌گیرد و سپس به بررسی آن نقطه و ایجاد خوشه و استفاده از تابع‌های `merge` و ... می‌پردازد.

در مرحله ما آرایه‌ای به نام `centroidsArr` در اختیار داریم که خوشه‌های بدست آمده تا آن لحظه را در آن قرار می‌دهیم و همیشه آخرین خانه آن که جدیدترین خانه آن است جدیدترین بروز رسانی از خوشه بندی های ما است.

در هربار اجرا، آخرین خانه این آرایه را در متغیری موقت کپی میکنیم و به مقایسه نقطه جدید داده شده به تابع اصلی، به پردازش آن میپردازیم که اگر کلاستری جدید است ایجاد و یا بررسی شود در کلاستری دیگر قرار میگیرد یا... .

در تمامی توابع سعی شده از شبه کد مقاله استفاده شود و الگوریتم معرفی شده در مقاله پیاده سازی شود.

در پایان چهار نمودار قرار داده شده است که نمودارهای مراکز کلاسترها در حضور کل دیتا در یک نمودار است.

و همچنین دو نمودار که تغییرات تعداد خوشه بندی ها را در طول پردازش کل دیتا نمایش میدهد.

زمان انجام هر کدام از دیتاستها و همچنین تعداد کلاسترهای ادغام شده و کلاسترهای حذف شده نیز در اسکرینی قرار داده شده است.