

پروژه ارشد هوش مصنوعی

عنوان: خوشه‌بندی داده‌ها روی نقشه

دانشجو: صدرا صمدی طاهرگورابی

شماره دانشجویی: ۹۹۰۱۹۷۴۵۱

استاد: مجید ایرانپور مبارکه

دانشگاه: پیام نور واحد نجف‌آباد

پاییز ۱۴۰۱

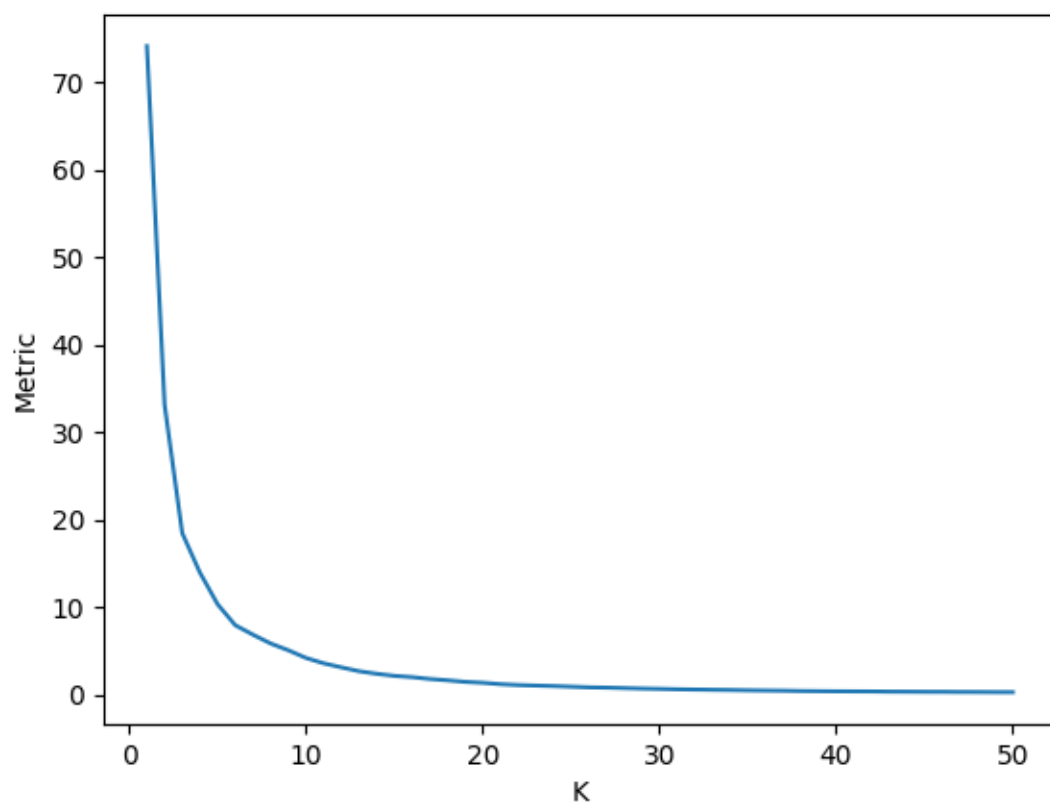
توضیحات

در فاز اول این پروژه ما مشخصات مکانی مجموعه‌ای از نقطه‌ها در اختیار داریم که می‌خواهیم ابتدا آن‌ها را در پنجره‌های یک کیلومتر در یک کیلومتر جداسازی کرده و سپس روی این پنجره‌ها الگوریتم خوشه‌بندی را پیاده‌سازی کنیم.

این پروژه شامل دو بخش Backend و Frontend است.

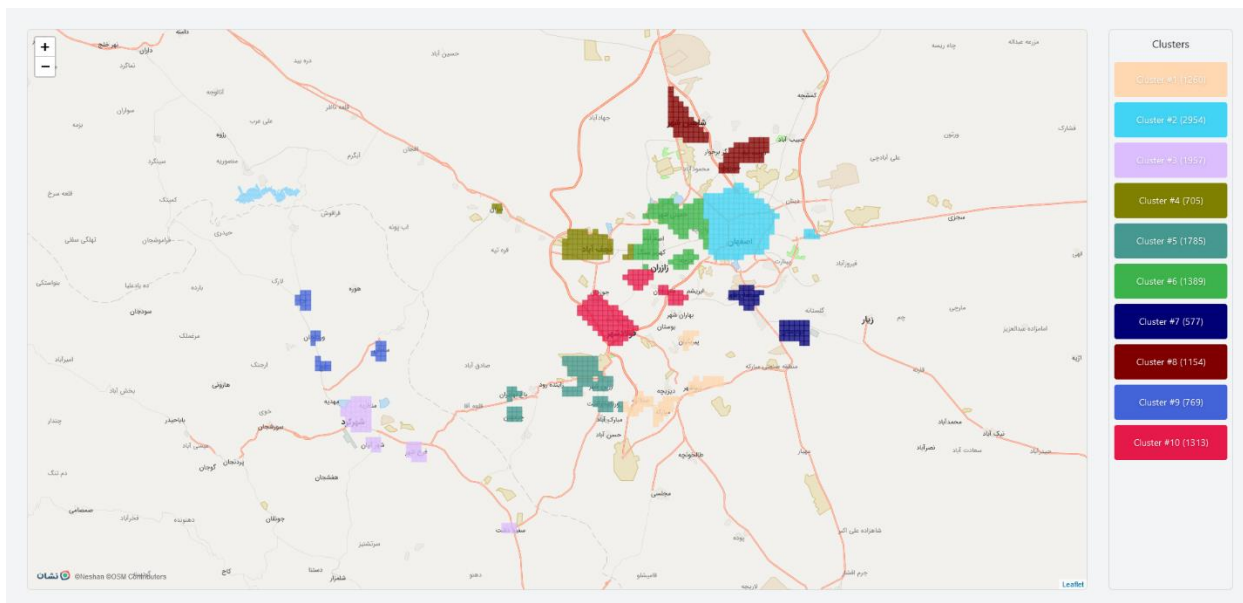
در بخش Backend که توسط زبان برنامه نویسی پایتون نوشته شده است مراحل اصلی یعنی پردازش داده‌ها و اجرای الگوریتم خوشه‌بندی انجام می‌شود که به شرح زیر است:

۱. ابتدا داده‌های موجود در فایل ورودی را به صورت یک جدول (DataFrame) بارگذاری می‌کنیم.
۲. داده‌های ناقص را حذف کرده و مشخصات مکانی دقیق هرکدام را به صورت Longitude و Latitude در ستون‌های مختلف قرار می‌دهیم.
۳. ناحیه‌ای که نقطه‌ها در آن قرار دارند را به پنجره‌های یک در یک کیلومتر تقسیم کرده و مشخصات مکانی و تعداد نقاط موجود در هرکدام را مشخص می‌کنیم.
۴. از آنجا که نقطه‌ها به صورت Longitude و Latitude هستند و این دو پارامتر مقیاس‌های متفاوتی دارند بهتر است که آن‌ها را Normalize کرده و در بازه‌ی صفر تا یک قرار دهیم.
۵. الگوریتم خوشه‌بندی K-Means را با تعیین کردن حداکثر تعداد خوشه (K) روی پارامترهای مکانی نرمال شده پنجره‌ها اجرا می‌کنیم تا مشخص شود هر خانه به کدام خوشه تعلق دارد.
۶. برای یافتن تعداد مناسب خوشه‌ها از روش Elbow بهره می‌بریم، یعنی به ازای تعداد خوشه‌های مختلف الگوریتم را اجرا کرده، مجموع مربعات خطا (SSE) را محاسبه و در نمودار نمایش می‌دهیم. نقطه‌ای که در آن نمودار به شکل آرنج یک دست درآید معمولاً بهترین انتخاب است:



۷. نتیجه نهایی را در یک فایل خروجی ذخیره می‌کنیم.

در بخش Frontend که به صورت یک وب اپ و به کمک کتابخانه ReactJS طراحی شده است، اطلاعات موجود در فایل خروجی تولید شده در قسمت قبل را با یک رابط کاربری ساده نمایش می‌دهیم.



همانطور که در نمونه تصویر بالا مشخص است، پنجره های موجود روی نقشه نمایش داده شده و خوشه‌های مربوط به هر کدام با رنگ‌های مختلف متمایز شده‌اند. با قرار دادن نشانگر موس روی هر کدام، تعداد داده های موجود در آن نمایش داده می‌شود. همچنین تعداد کل داده‌ها در هر خوشه نیز در سمت راست قرار داده شده‌اند.

راه اندازی بخش Backend

برای این بخش نیاز است که نسخه ۳.۸ یا بالاتر نرم افزار Python و pip روی سیستم شما نصب باشد.

<https://www.python.org/>

کدهای اصلی این قسمت در مسیر `/src/app/main.py` و فایل های ورودی و خروجی در مسیر `/data` قرار داده شده اند.

جهت اجرا مراحل زیر را به ترتیب انجام دهید:

۱. یک خط فرمان (command line, cmd, terminal, powershell, ...) را دقیقاً در مسیر پوشه `/backend` اجرا کنید.
۲. دستورات زیر را جهت ساخت یک محیط مجازی و نصب پکیج های استفاده شده در پروژه اجرا کنید:
3. `python -m venv .venv`
4. `.\.venv\Scripts\activate`
5. `pip install poetry`
6. `poetry install`
۷. منتظر بمانید تا پکیج ها دریافت و نصب شوند (اولین اجرا ممکن است کمی زمان بر باشد).
۸. جهت نمایش نمودار مجموع مربعات خطا دستور زیر را وارد کنید (این قسمت اختیاری است):
9. `poetry run metrics`
۱۰. جهت اجرای الگوریتم خوشه بندی و ذخیره اطلاعات در یک فایل خروجی دستور زیر را وارد کنید:
11. `poetry run export`
۱۲. به طور پیش فرض الگوریتم برای ۱۰ خوشه اجرا می شود. برای تغییر این مقدار می توانید در انتهای دستور قبل تعداد مورد نظر را هم وارد کنید، مثال:
13. `poetry run export 15`
۱۴. اگر همه مراحل بدون مشکل اجرا شوند، یک فایل خروجی در مسیر `/data/output.xlsx` ساخته می شود.

راهنمایی بخش Frontend

برای این بخش نیاز است که یکی از نسخه‌های (ترجیحا به‌روز) نرم‌افزار Node.JS و npm روی سیستم شما نصب باشد.

<https://nodejs.org/en/>

کدهای اصلی این قسمت در مسیر `/src/app.tsx` قرار داده شده‌اند.

جهت اجرا مراحل زیر را به ترتیب انجام دهید:

۱. یک خط فرمان (... , powershell, terminal, cmd, command line) را دقیقاً در مسیر پوشه `/frontend` اجرا کنید.

۲. دستور زیر را جهت نصب پکیج‌های استفاده شده در پروژه اجرا کنید:

3. `npm install`

۴. منتظر بمانید تا پکیج‌ها دریافت و نصب شوند (اولین اجرا ممکن است کمی زمان بر باشد).

۵. فایل خروجی که در بخش Backend ساخته شد را در مسیر `/frontend/public/output.xlsx` کپی کنید.

۶. دستور زیر را اجرا کنید:

7. `npm run start`

۸. پس از چند ثانیه وارد آدرس زیر در مرورگر خود شوید:

9. <http://localhost:3000/>

۱۰. اگر همه مراحل بدون مشکل اجرا شوند، رابط کاربری برنامه در صفحه مرورگر نمایش داده می‌شود.

۱۱. برای بستن برنامه در خط فرمان دکمه‌های `Ctrl` و `C` کیبورد را دوبار به صورت همزمان فشار دهید.