



# Mini Project

## Data Analysis

Coding Boot Camp – Spring 2024

NUMBER	DEADLINE	TOPICS
2	Sat, 1403/02/05 12:00:00 PM	Python Review

1. پاسخ تمرین را در کارپوشه خود ارسال کرده و پشتیبان فنی(منتور) را منشن نمایید.



2. در صورتی که سوال و یا ابهامی دارید در گروه چت تلگرامی بپرسید.

### جایگاه من کجاست؟

ب.ب.ک (برنامه‌ای بنویسید که) اطلاعات یک مجموعه نقاط را در فضای سه بعدی از فایل پیوست شده خوانده و پس از پاکسازی و حذف داده‌های اشتباه و ناقص، آن‌ها را به K تعداد خوشه، تقسیم بندی نماید.

پ.ن: خوشه بندی یا Clustering یکی از تکنیک‌های علم داده در بحث یادگیری ماشین می‌باشد که به واسطه آن امکان گروه بندی مشاهدات مشابه در یک دسته را فراهم می‌سازد؛ در ادامه مراحل الگوریتم K-Means شرح داده شده:

- تعیین تعداد  $K$  جهت خوشه بندی با ورودی گرفتن از کاربر
  - تعیین روش محاسبه فاصله اقلیدسی یا منهتن با ورودی گرفتن از کاربر
- (برای محاسبه فاصله روش‌های مختلفی وجود دارد که اقلیدسی و منهتن جزو کاربردی‌ترین‌ها هستند)
- خواندن فایل پیوست شده و لود کردن اطلاعات آن در برنامه
  - پاکسازی و پیش پردازش دیتا جهت اجرای الگوریتم
- (بعنوان یک دیتا آنالیزور، داده‌هایی که با آن‌ها سروکار داریم همواره نیاز به پاکسازی و حذف موارد نامربوط دارند)
- ساخت Instance نقاط از روی داده‌های حاصل شده در گام قبل
  - تعیین  $K$  تعداد نقطه مرکزی بصورت تصادفی و رندوم
- (با فرض اینکه مختصات نقاط در هر سه بعد محور بین 10 تا 10- قرار دارند نقاط مرکزی را انتخاب کنید)
- محاسبه فاصله هر نقطه از مراکز با روش انتخاب شده توسط کاربر در گام دوم
  - یافتن نزدیک‌ترین مرکز به ازای هر نقطه و تشکیل خوشه‌ها
  - محاسبه مراکز جدید هر خوشه با میانگین گرفتن از مختصات نقاط آن دسته
  - تکرار مجدد گام‌های 7 الی 9 تا زمانی که مراکز تغییر نکنند

جهت انجام کامل پروژه در زمان ددلاین مشخص شده، پیشنهاد می‌شود بخش‌های زیر را که

سخت‌تر بوده و یا اهمیت کمتری داشته، به ترتیب کنار بگذارید:

1 شئ‌گرایی: تعریف کلاس نقطه و ساخت نمونه از روی آن به ازای هر رکورد از دیتا را صرف نظر کنید.

2 کار با فایل: خواندن اطلاعات از روی فایل را صرف نظر کرده و خودتان مستیقا دیتا را وارد نمایید.

3 پیش پردازش: پاکسازی و حذف موارد نادرست را صرف نظر کنید تا نیازی به بررسی اضافه‌تری نباشد.

4 توابع فاصله: بجای تعریف چندین تابع مختلف برای محاسبه، تنها یک تابع فرمول اقلیدسی بنویسید.

موفق و بی‌باک باشید... :