

به نام خدا

پروژه درس داده کاوی : خوشه بندی

استاد درس: دکتر رضانی

مسئول پروژه: خزایی

تاریخ تحویل پروژه: در کانال اعلام خواهد شد

محل تحویل: سامانه lms

#### نکات:

- لطفا تمام مراحل پروژه را با دقت و به صورت انفرادی انجام دهید.
- مشاهده هر گونه شباهت در تحلیل و کد، منجر به کسر نمره خواهد گردید.
- گزارش را به صورت کامل همراه با تحلیل خود از نتایج تهیه نمایید.

در این پروژه به بررسی الگوریتم‌های خوشه‌بندی DBScan، KMeans و Agglomerative (Hierarchical) می‌پردازیم. برای این منظور مجموعه داده مربوط به مشتریان ۱۰ فروشگاه که از محصولات خریداری شده خود در فیسبوک پست منتشر کرده‌اند، در اختیار شما قرار داده شده است. این مجموعه داده حاوی ۷۰۵۱ سطر و تعداد ۱۲ ستون می‌باشد. در این پروژه باید مشتریان را خوشه‌بندی کنید. لطفا مراحل زیر را همراه با تحلیل در گزارش ارائه دهید.

۱. در گام اول می‌بایست مجموعه داده مورد نظر را پیش‌پردازش نمایید. برای این کار توضیحات خود را برای هر تصمیم خود در پیش‌پردازش توضیح دهید.

۲. در گام دوم و قبل از خوشه‌بندی داده‌ها، لازم است تا درک خوبی از داده‌ها داشته باشید. برای این کار باید نموداری از توزیع داده‌ها نمایش دهید و توضیحاتی درباره ستون‌ها بنویسید. سپس مشخص کنید کدام یک از ستون‌ها از نظر شما اهمیت بالاتری برای خوشه‌بندی کاربران دارند.

۳. در گام سوم لازم است تا داده‌ها خوشه‌بندی شوند:

۳-۱. با استفاده از تابع `interia` بهترین تعداد خوشه را تشخیص دهید (با رسم نمودار `elbow`).

۳-۲. با استفاده از الگوریتم `KMeans` داده‌ها را با بهترین تعداد خوشه‌ها، خوشه‌بندی نمایید. خوشه‌ها را نمایش دهید.

۳-۳. الگوریتم‌های `DBScan` و `Agglomerative (Hierarchical)` را برای خوشه‌بندی به کار ببندید، سپس خوشه‌ها را نمایش دهید.

۳-۴. خوشه‌های به دست آمده در سه الگوریتم را با هم مقایسه کنید و تحلیل خود را از این خوشه‌ها و کیفیت آن‌ها بیان کنید.

۳-۵. نقاط مرکزی خوشه‌ها را در شکل نمایش دهید.