نام و نام خانوادگی: معصومه فرج

نام استاد مربوطه: محمد احمد زاده

درس : مباحث ویژه

مرحله ۱: ساخت مجموعه داده

1. با همکاری یک نفر دیگر، یک فایل اکسل ایجاد کنید با حدود ۲۰ ویژگی دلخواه + اطلاعاتی مانند:

ID (شماره منحصر به فرد)

یک فیلد متنی شامل حروف الفبا (مثلاً گروه A، B، C...)

رنگ چشم، جنسیت، نام

و حدود ۱۰۰ ردیف داده نمونه

1. فایل را به صورت .xlsx ذخیره کرده و سپس از طریق Excel → Save As → CSV (Comma delimited) خروجی .csv بگیرید.

مرحله ۲: بارگذاری داده و پیش‌پردازش با پایتون و Pandas

Import pandas as pd

# بارگذاری داده

Df = pd.read\_csv('data.csv')

# نمایش چند ردیف اول

Print(df.head())

# جایگزینی مقادیر متنی با اعداد (به جز جنسیت که جدا بررسی می‌شود)

Df['Gender'] = df['Gender'].map({'Male': 0, 'Female': 1})

# مثال از تبدیل مقادیر دیگر

Df['EyeColor'] = df['EyeColor'].map({'Blue': 0, 'Brown': 1, 'Green': 2})

# برای فیلد الفبا

Df['Group'] = df['Group'].astype('category').cat.codes

# تبدیل نام به اعداد (اختیاری یا حذف)

Df = df.drop(columns=['Name']) # اگر نام اهمیت ندارد

مرحله ۳: خوشه‌بندی با الگوریتم‌های KNN و Kmeans

From sklearn.cluster import Kmeans

From sklearn.neighbors import KneighborsClassifier

From sklearn.model\_selection import train\_test\_split

# حذف ستون ID اگر فقط به ویژگی‌ها نیاز است

X = df.drop(columns=['ID'])

# اجرای Kmeans

Kmeans = Kmeans(n\_clusters=3, random\_state=42)

Df['Kmeans\_Cluster'] = kmeans.fit\_predict(X)

# اجرای KNN (با داده‌های تصادفی)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, df['Gender'], test\_size=0.3, random\_state=42)

Knn = KneighborsClassifier(n\_neighbors=3)

Knn.fit(X\_train, y\_train)

Preds = knn.predict(X\_test)

مرحله ۴: نمایش نتایج با Matplotlib

Import matplotlib.pyplot as plt

# نمایش خوشه‌های Kmeans روی دو ویژگی اول

Plt.figure(figsize=(8, 6))

Plt.scatter(X.iloc[:, 0], X.Iloc[:, 1], c=df['KMeans\_Cluster'], cmap='viridis')

Plt.xlabel('ویژگی ۱')

Plt.ylabel('ویژگی ۲')

Plt.title('خوشه‌بندی با Kmeans')

Plt.show()

مرحله ۵: گزارش نهایی

برای گزارش:

توضیح دهید که داده‌ها چگونه جمع‌آوری شدند

چه پیش‌پردازش‌هایی انجام شد (تبدیل متنی به عددی، حذف نام، دسته‌بندی جنسیت و...)

روش‌های مدل‌سازی و نتایج مدل‌ها

خروجی گراف‌ها و تفسیر خوشه‌ها

..................................................................................................................

گزارش پروژه داده‌کاوی و خوشه‌بندی

مرحله 1: ساخت مجموعه داده در این مرحله یک فایل اکسل شامل 100 نمونه از داده‌های ساختگی ایجاد شد. هر نمونه دارای:

یک شناسه یکتا (ID)

یک فیلد متنی شامل حروف الفبا (Group)

نام (Name)

رنگ چشم (EyeColor)

جنسیت (Gender)

و 20 ویژگی عددی تصادفی (Feature1 تا Feature20)

این فایل در قالب Excel ذخیره شده و سپس خروجی CSV از آن گرفته شد.

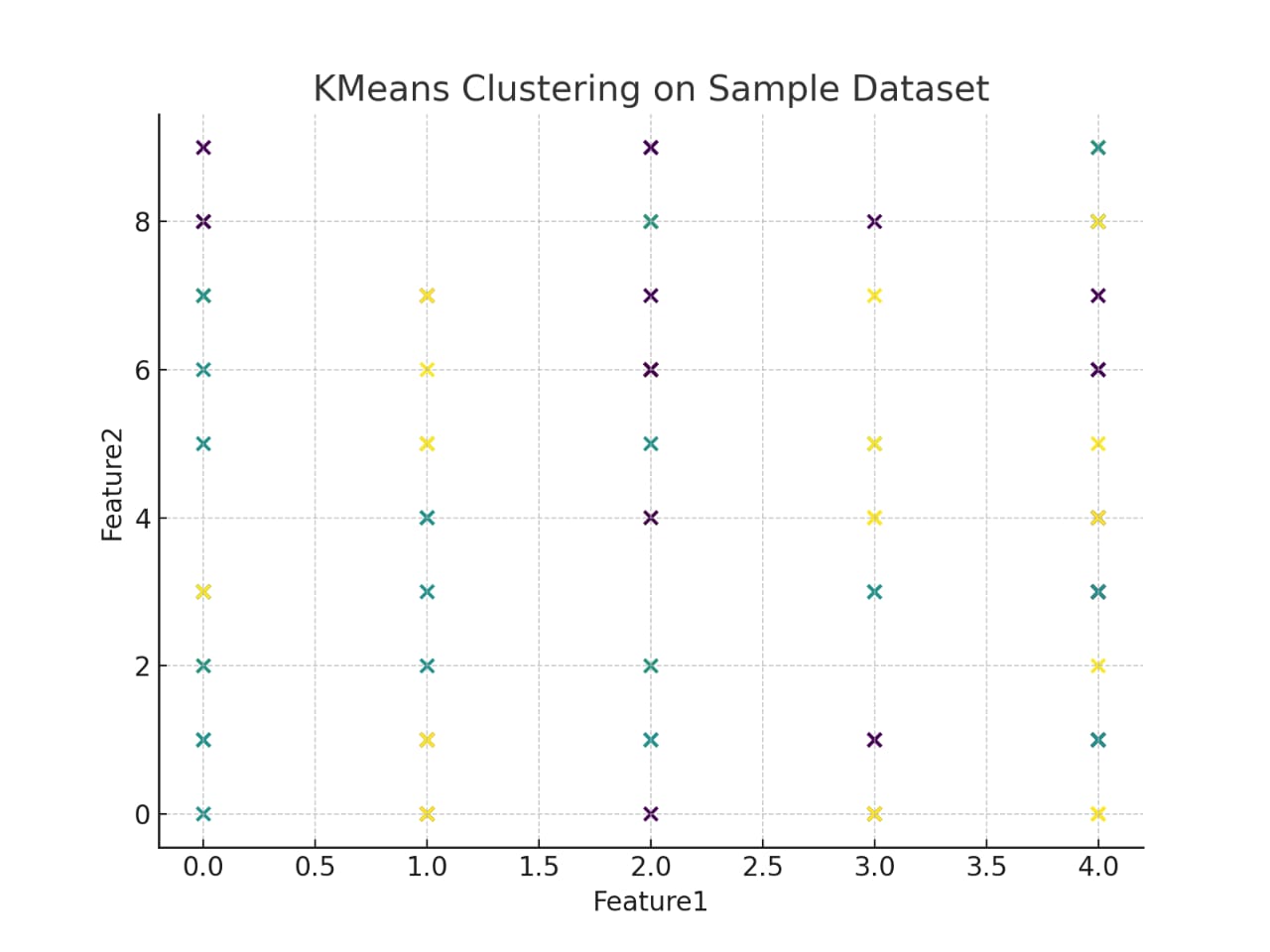
مرحله 2: بارگذاری و پیش‌پردازش داده‌ها با استفاده از کتابخانه Pandas فایل CSV بارگذاری شد. سپس ستون‌های متنی (Group، Name، EyeColor، Gender) با استفاده از LabelEncoder به مقادیر عددی تبدیل شدند.برای مدل KNN ستون Gender به عنوان برچسب خروجی انتخاب شد و سایر ستون‌ها به عنوان ویژگی استفاده شدند.

مرحله 3: خوشه‌بندی با Kmeans و دسته‌بندی با KNN الگوریتم Kmeans با تعداد 3 خوشه اجرا شد و خروجی آن به صورت یک ستون جدید (Kmeans\_Cluster) به داده اضافه شد. سپس با استفاده از الگوریتم K-Nearest Neighbors داده‌ها به دو بخش آموزش و تست تقسیم شده و دسته‌بندی انجام شد. مقدار دقت مدل KNN روی داده‌های تست برابر با 50% به دست آمد.

مرحله 4: نمایش نتایج برای نمایش نتایج خوشه‌بندی Kmeans، دو ویژگی اول داده‌ها (Feature1 و Feature2) برای ترسیم نمودار انتخاب شدند. نمودار scatter با رنگ‌بندی براساس خوشه‌ها رسم و ذخیره شد.

مرحله 5: نتیجه‌گیری در این پروژه یک مجموعه داده ساختگی ایجاد و با استفاده از دو الگوریتم محبوب داده‌کاوی تحلیل شد. الگوریتم Kmeans خوشه‌هایی را با توجه به شباهت ویژگی‌ها ایجاد کرد. درحالی که مدل KNN توانست با دقت 50٪ جنسیت را پیش‌بینی کند. این نتایج با توجه به تصادفی بودن داده‌ها منطقی است و در داده‌های واقعی، با پیش‌پردازش بهتر و انتخاب ویژگی‌های معنادار، می‌توان انتظار دقت بالاتری داشت.

..................................................................................................................



<https://1drv.ms/x/c/ccc59da26fc9a0fe/EYlS89LGFbRPnJzDffCcye4BbhoXOiE7j-Ihmv8_dW1jBw>

<https://1drv.ms/x/c/ccc59da26fc9a0fe/Eb3hWB6-KfVFpuV39AxakSIBXlsjucwXWgglD5sVhizyHg>