

```
1 import pandas as pd # type: ignore
2 from sklearn.preprocessing import StandardScaler # type: ignore
3 from sklearn.decomposition import PCA # type: ignore
4 from sklearn.cluster import KMeans # type: ignore
5 import matplotlib.pyplot as plt # type: ignore
```

فراخوانی تمامی کتابخانه های مورد نیاز از جمله Pandas و K-Means

```
df = pd.read_csv("dataset-uci.csv")
```

تبدیل فایل .xlsx به .csv.

```
df_clean = df.dropna()
```

حذف داده های ناقص

```
features = df_clean.select_dtypes(include=["float64", "int64"])
```

انتخاب فقط ستون های عددی

```
16 scaler = StandardScaler()
17 X_scaled = scaler.fit_transform(features)
```

نرمال سازی داده ها

```
19 pca = PCA(n_components=2)
20 X_pca = pca.fit_transform(X_scaled)
```

کاهش بعد با PCA برای رسم نمودار

خوشه بندی با K-Means

```
25 plt.figure(figsize=(8, 6))
26 plt.scatter(X_pca[:, 0], X_pca[:, 1], c=df_clean["Cluster"], cmap="viridis", s=50)
27 plt.title("K-Means Clustering (PCA Projection)")
28 plt.xlabel("PCA 1")
29 plt.ylabel("PCA 2")
30 plt.grid(True)
31 plt.show()
```

رسم نمودار نتایج

```
33 df_clean.to_excel("clustering_result.xlsx", index=False)
```

ذخیره فایل خروجی در Excel

گزارش نهایی خوشه‌بندی داده‌ها

در این پروژه، ابتدا یک دیتاست شامل ۲۰ ویژگی و ۲۰۰ نمونه از افراد مختلف مورد پردازش قرار گرفت. مراحل به شرح زیر انجام شد:

1. پیش‌پردازش:

ابتدا داده‌های ناقص (Missing Values) حذف شدند و فقط ستون‌های عددی برای تحلیل انتخاب شدند.

2. نرمال‌سازی: با استفاده از StandardScaler، مقیاس تمامی ویژگی‌ها استاندارد شد.

3. کاهش ابعاد: با استفاده از روش PCA ابعاد داده‌ها به دو بعد کاهش داده شد تا امکان رسم نمودار فراهم شود.

4. خوشه‌بندی:

- با الگوریتم K-Means خوشه‌بندی داده‌ها انجام شد (تعداد خوشه‌ها = ۳).

- سپس با الگوریتم **K-Medoids** نیز همین کار انجام شد تا نتایج مقایسه شوند.

تحلیل خوشه‌ها:

با بررسی نمودارهای PCA، خوشه‌بندی‌ها به طور منطقی داده‌ها را تفکیک کرده‌اند. خوشه‌ها تقریباً از هم جدا و با تراکم یکنواخت هستند.

با استفاده از میانگین مقادیر ویژگی‌ها برای هر خوشه نشان داد که خوشه‌ی شماره در الگوریتم K-Means از نظر برخی ویژگی‌ها مانند X1، X7 و X14 دارای مقادیر بالاتر و متمایزتری بود. این موضوع نشان‌دهنده شاخص‌بودن این خوشه نسبت به سایر خوشه‌هاست.

این موضوع نشان می‌دهد که افراد موجود در این خوشه دارای ویژگی‌های
شاخص‌تری در برخی معیارها هستند.

نتیجه‌گیری:

در این پروژه مشاهده شد که خوشه‌بندی به خوبی توانست الگوهای موجود در
داده‌ها را آشکار کند. الگوریتم K-Medoids به دلیل مقاومت بالاتر در برابر
داده‌های پرت، در برخی موارد نتایج پایدارتر و واضح‌تری ارائه داد. در نهایت فایل
اکسل نهایی شامل شماره خوشه به ازای هر فرد ارائه شده است.

