

## توضیحات فاز دوم

سوگند رضوی-امین عرفانیان

لینک پروژه:

<https://colab.research.google.com/drive/1uqg7nlHNLX5Xs3gRtaV1MH7JZKpS8bay?usp=sharing>

### بخش اول:

برای این بخش ابتدا داده های مورد نیاز را categorical میکنیم.

- برای BMI به ۴ دسته `underweight, Normal, Overweight, Obese` تقسیم میکنیم.

- برای Age ۲۰ تا ۲۰ تا جدا میکنیم.

- برای cholesterol به ۳ دسته `normal, high, very high` تقسیم میکنیم.

- برای Exercise hours per week به ۳ دسته `low, medium, high` تقسیم میکنیم.

- برای Triglycerides به ۳ دسته `high, borderline high, high` تقسیم میکنیم.

- برای stress level به ۳ دسته `low, medium, high` تقسیم میکنیم.

- برای heart rate به ۴ دسته `low, normal, elevated, high` تقسیم میکنیم.

سپس با استفاده از `pd.getdummies` شروع به encode میکنیم که برای پیدا کردن الگوهای مکرر لازم هست.

**نتیجه الگو مکرر بین جنسیت و خطر حمله قلبی:**

در مردها خطر حمله قلبی بیشتر است.

**نتیجه الگو مکرر بین سن و خطر حمله قلبی:**

در افراد بین رنج سنی ۲۰ تا ۴۰ خطر حمله قلبی بیشتر است.

**نتیجه الگو مکرر بین کلسترول و خطر حمله قلبی:**

در افراد که کلسترول `very high` دارند خطر حمله قلبی بیشتر است.

نتیجه الگو مکرر بین ساعات ورزش و خطر حمله قلبی:

در افرادی که ساعات ورزش آن ها در هفته کم هست خطر حمله قلبی بیشتر است.

نتیجه الگو مکرر بین رژیم و خطر حمله قلبی:

در افرادی که رژیم آن ها **unhealthy** هست خطر حمله قلبی بیشتر است.

نتیجه الگو مکرر بین فشار خون و خطر حمله قلبی:

در افرادی که فشار خون بالایی دارند خطر حمله قلبی بیشتر است.

نتیجه الگو مکرر بین شاخص توده بدنی و خطر حمله قلبی:

افرادی که شاخص توده بدنی آنها در رنج **obese** قرار دارد خطر حمله قلبی بالاتری دارند.

نتیجه الگو مکرر بین استرس و خطر حمله قلبی:

افراد با استرس بیشتر خطر حمله قلبی بالاتری دارند.

### بخش دوم:

برای اعمال کلاسترینگ و PCA باید ابتدا انکد انجام بدهیم.

شکل کلاسترینگ در لینک پروژه قرار گرفته است.

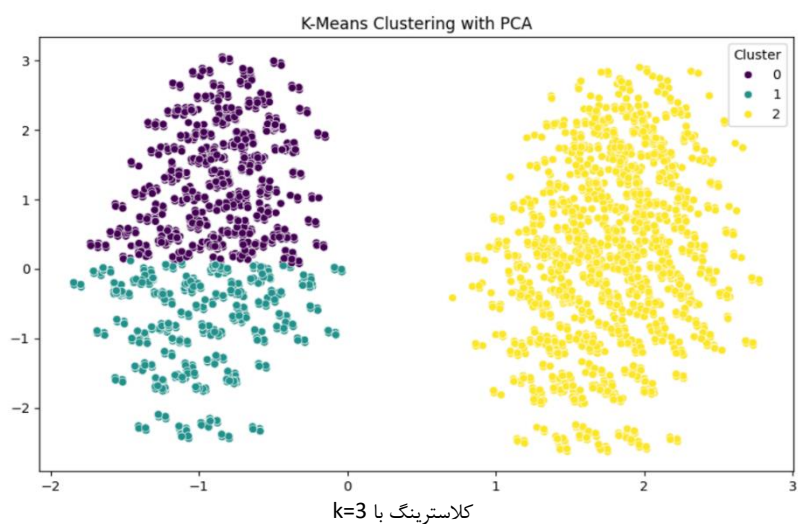
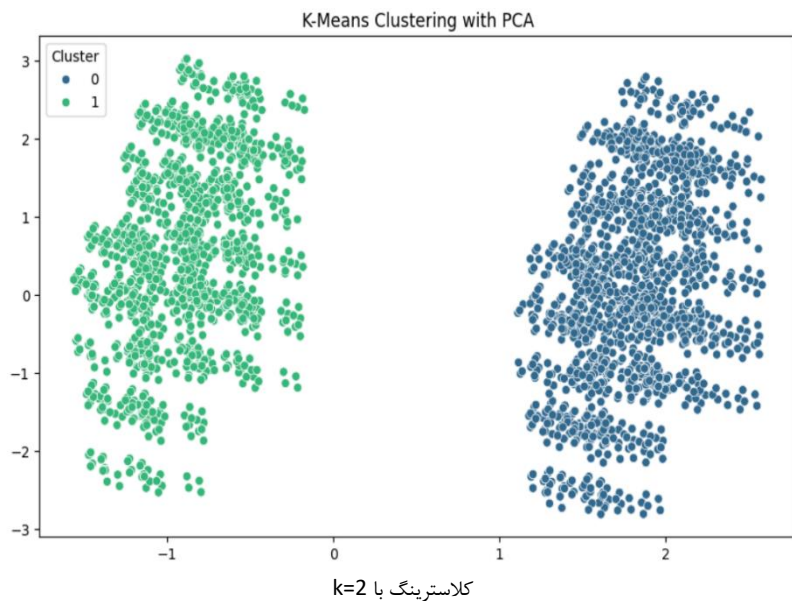
برای **k-means** متوجه میشویم که  $k=4$  بهتر از بقیه عمل میکند که با استفاده از **silhouette score** متوجه این منظور میشویم.

و همچنین از الگوریتم **meanshift** هم استفاده کردیم که اشکال آن در لینک پروژه موجود است.

### بخش سوم:

در لینک پروژه توضیحات به صورت کامل قرار گرفته است.

نتایج:



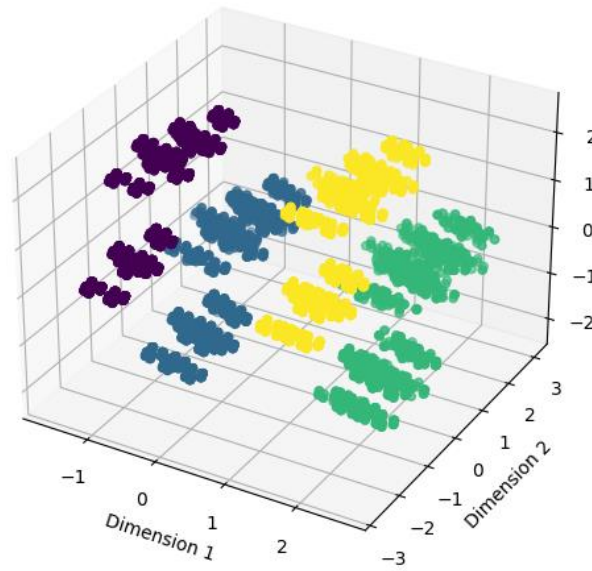


کلاسترینگ با  $k=4$

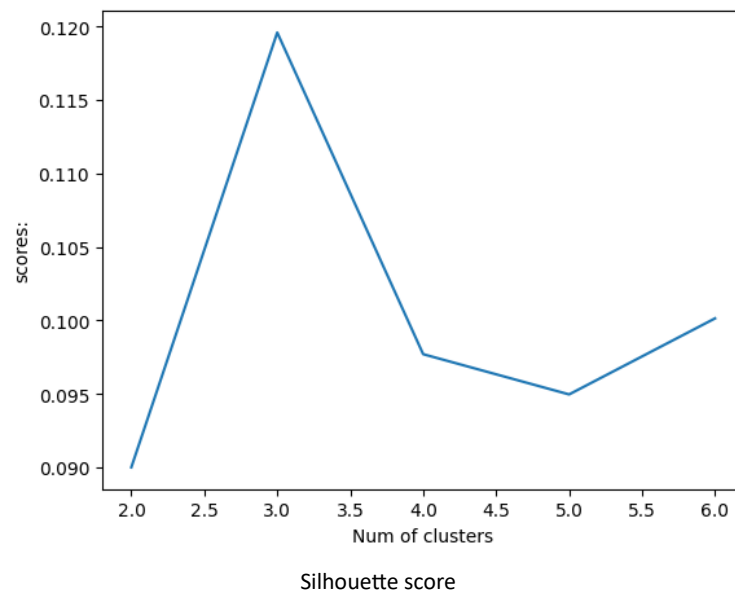


کلاسترینگ با  $k=5$

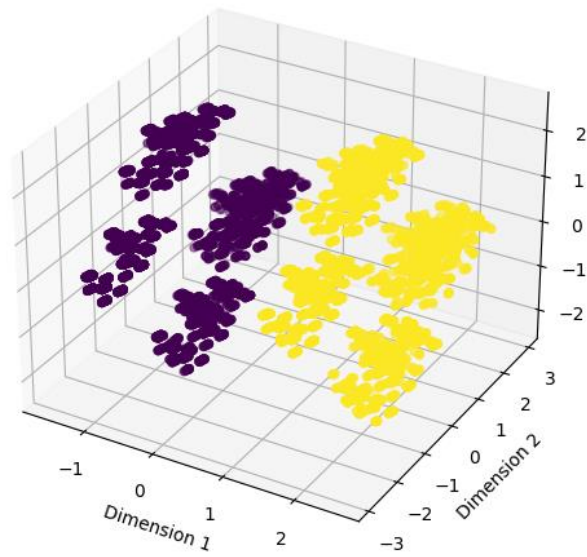
K-Means Clustering with PCA (3D)



کلاسترینگ با  $k=4$  (سه بعدی)



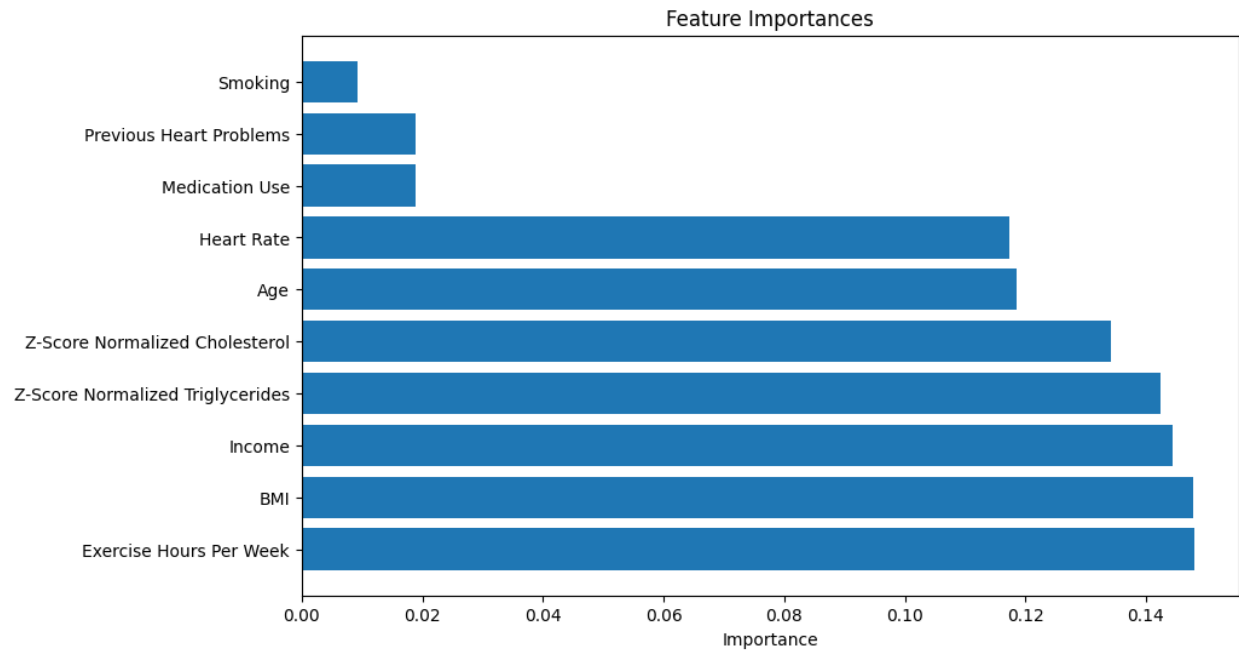
Mean Shift Clustering with PCA (3D)



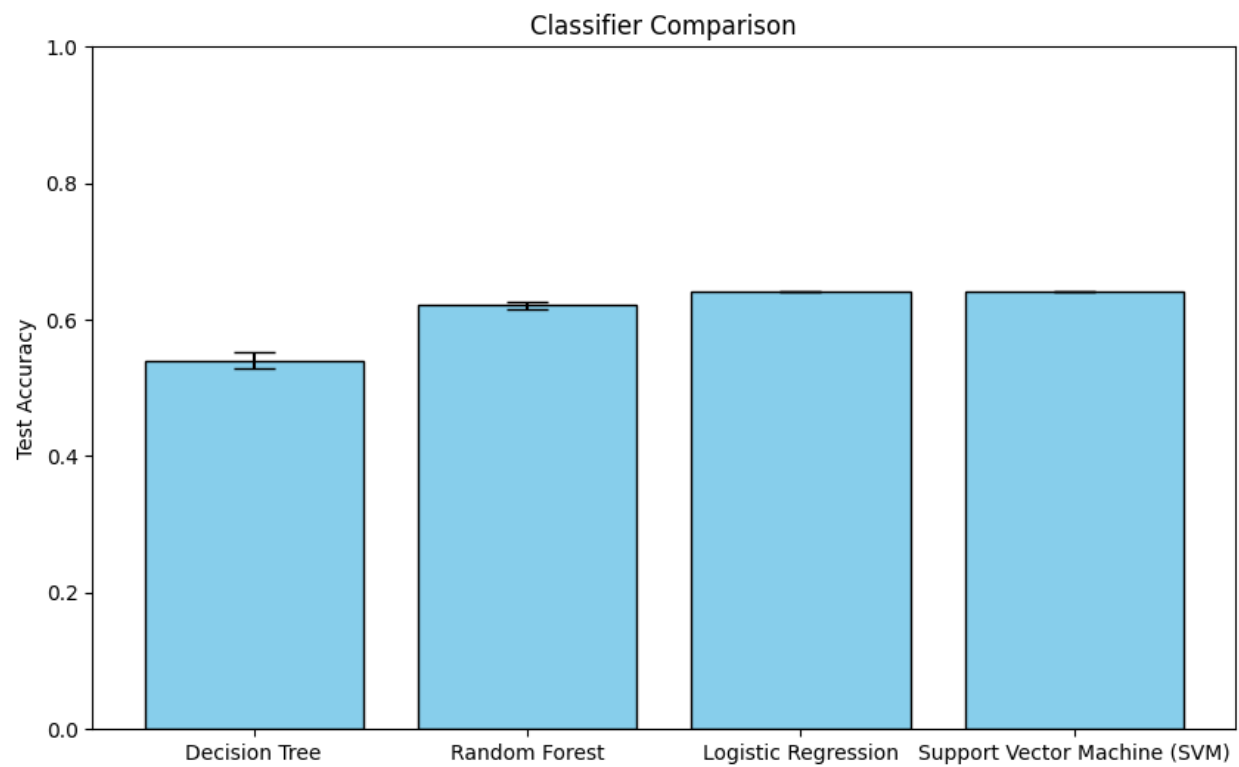
Mean shift algorithm (3D)

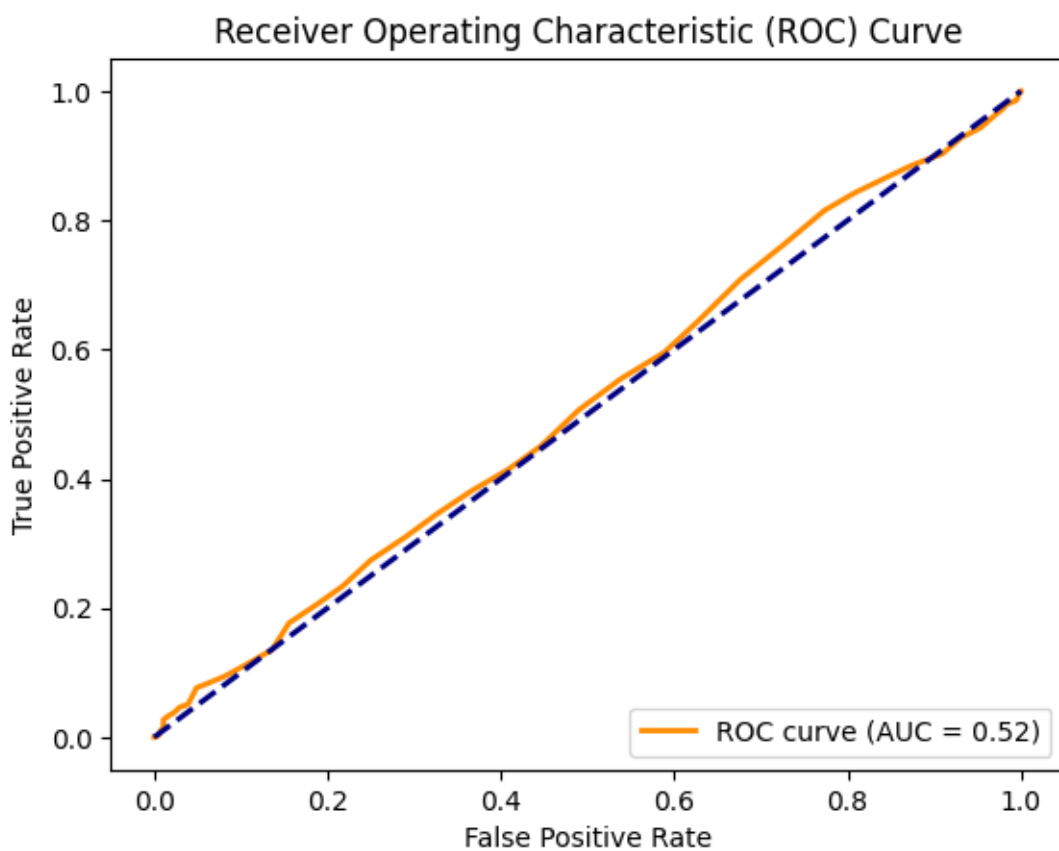
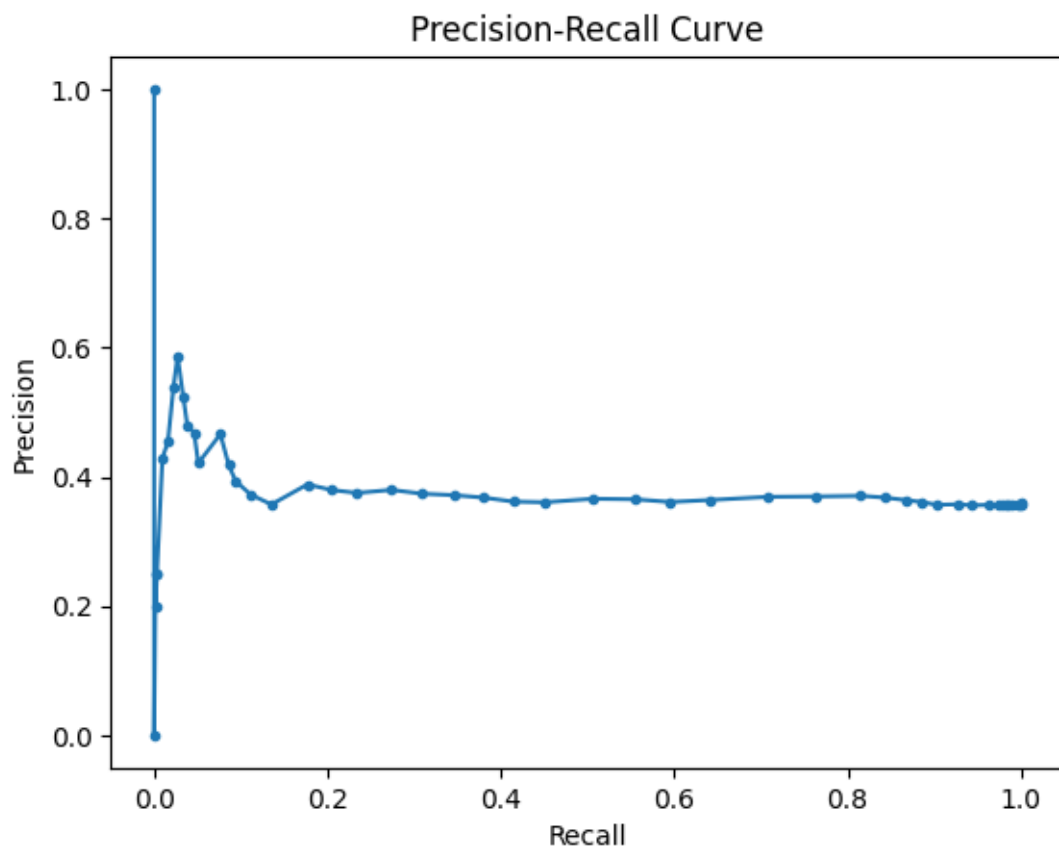


Mean shift algorithm (2D)



Identify and compare attributes







Confusion Matrix

