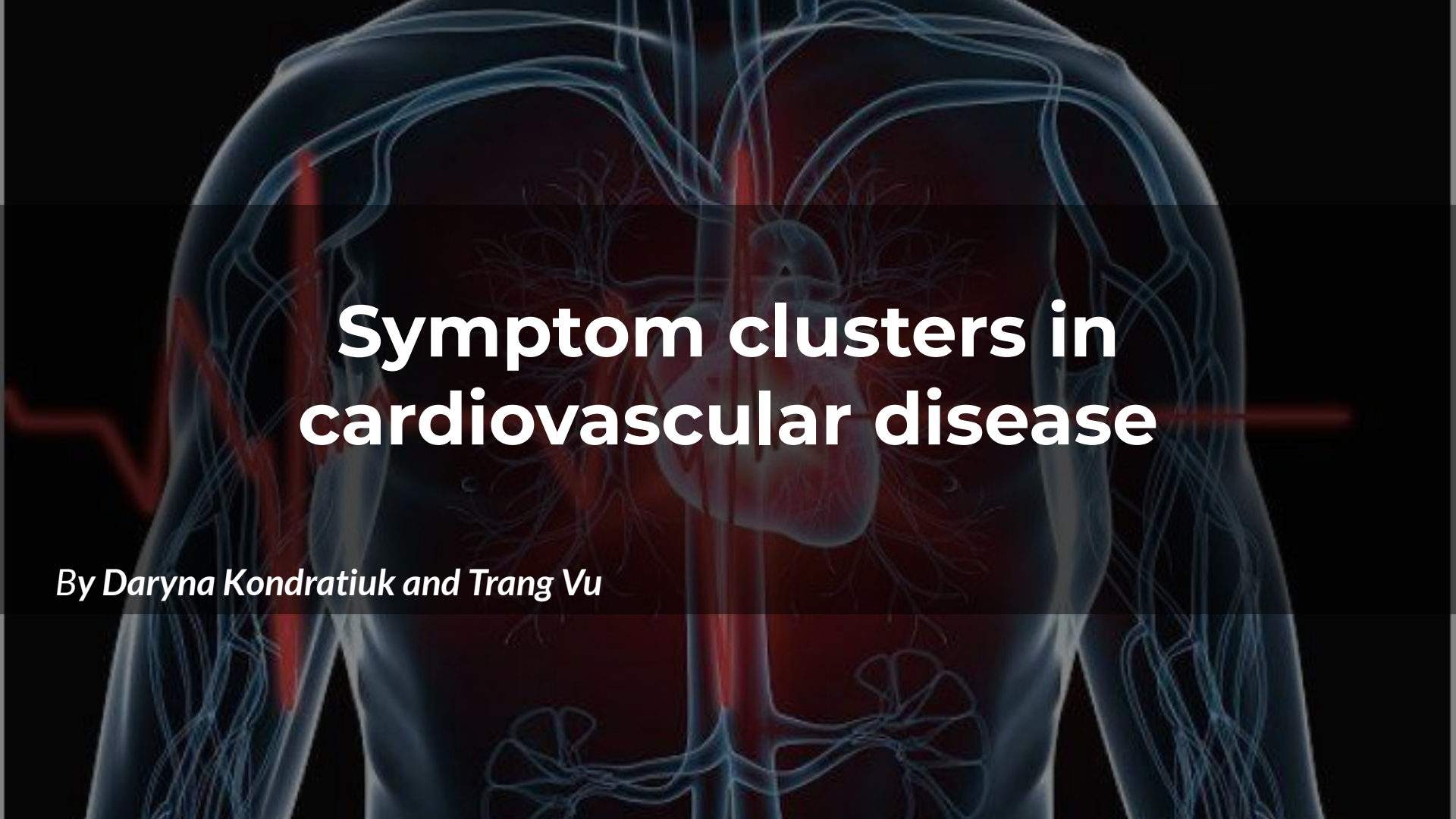


Symptom clusters in cardiovascular disease

By Daryna Kondratiuk and Trang Vu

Objectives:

- Utilizing a dataset to get a comprehensive overview of factors that could trigger a heart disease
- Apply clustering analysis to identify 2 groups of patients who has and who doesn't have heart disease by similar symptoms

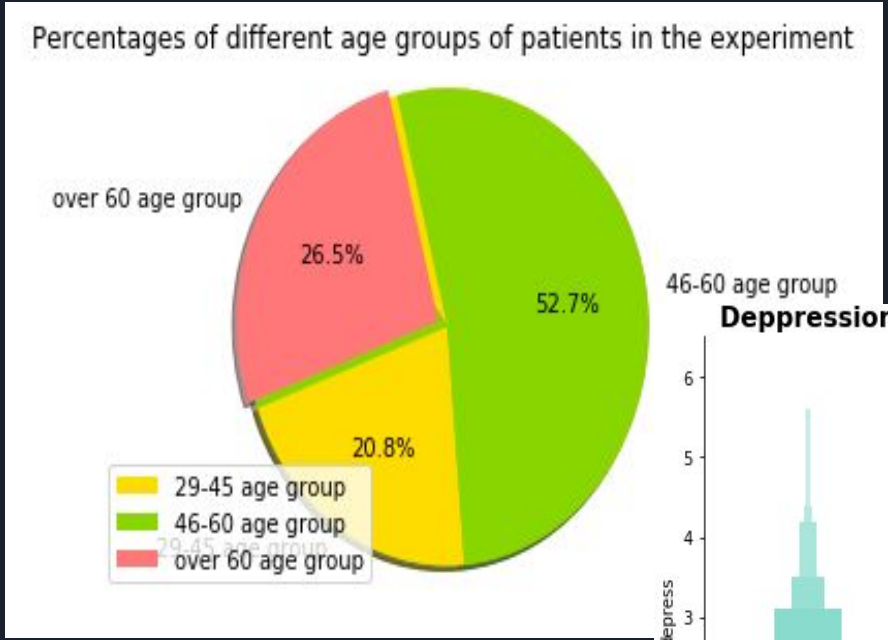
Stakeholder:

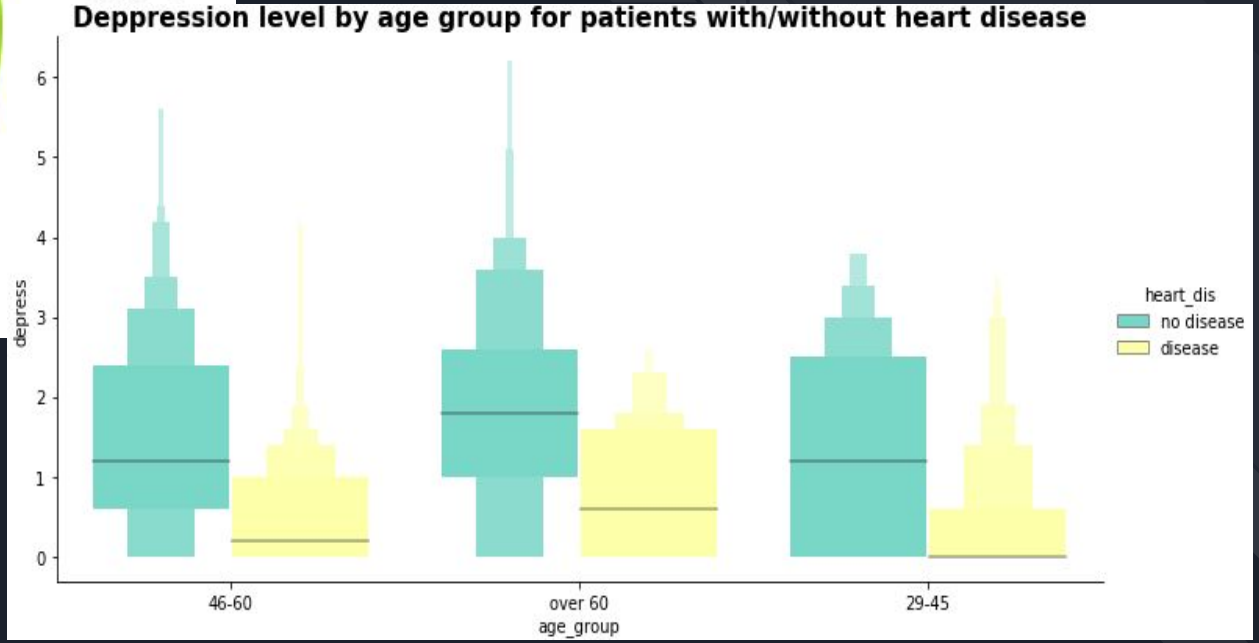
American Heart Association

Heart Disease Dataset

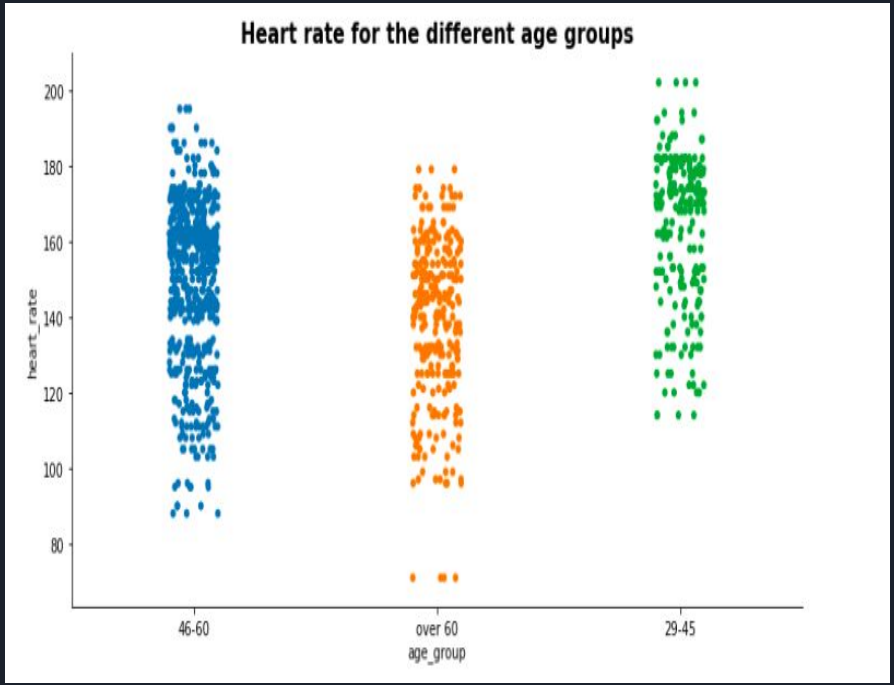
- A dataset includes 14 attributes of patient's health conditions.
- For our analysis we are using age, gender, blood pressure, cholesterol, blood sugar, heart rate, depression level and target variables.
- The "target" field refers to the presence of heart disease for the patient. It is integer valued 0 = no disease and 1 = disease.
- Two classification columns were added: YES/NO Heart disease and age groups.

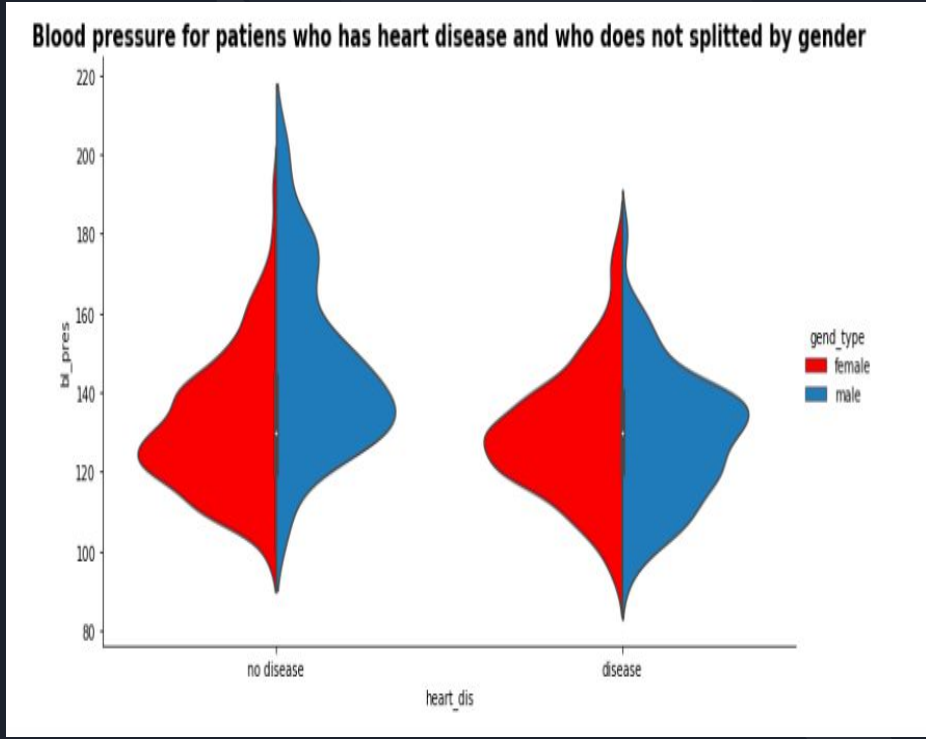
Visualization



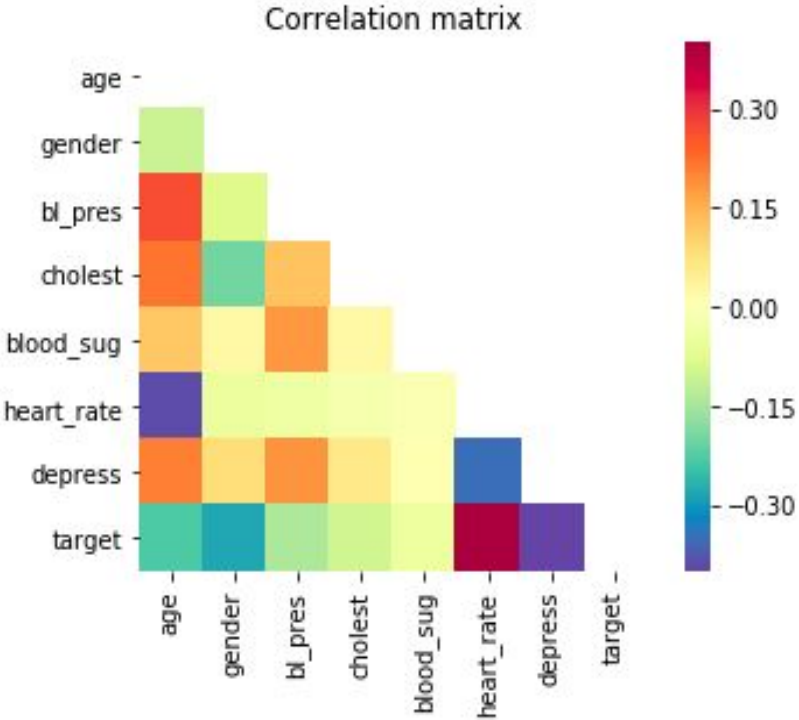


Seaborn Visualization





Correlation Matrix



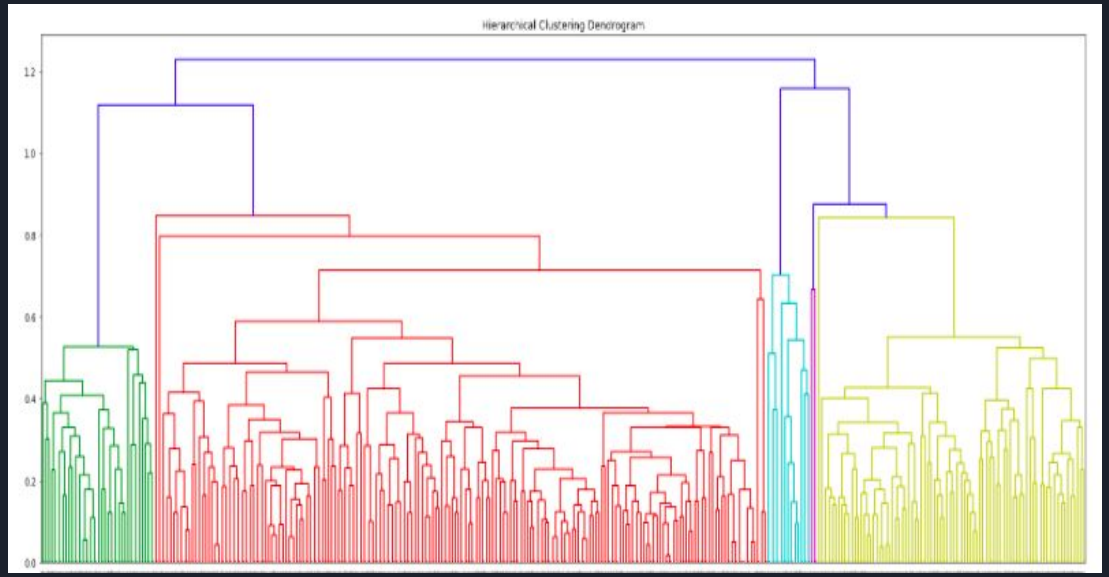
- Heart rate & target
- Depression rate and target
- Heart rate and age
- Blood pressure and age

Dendrogram

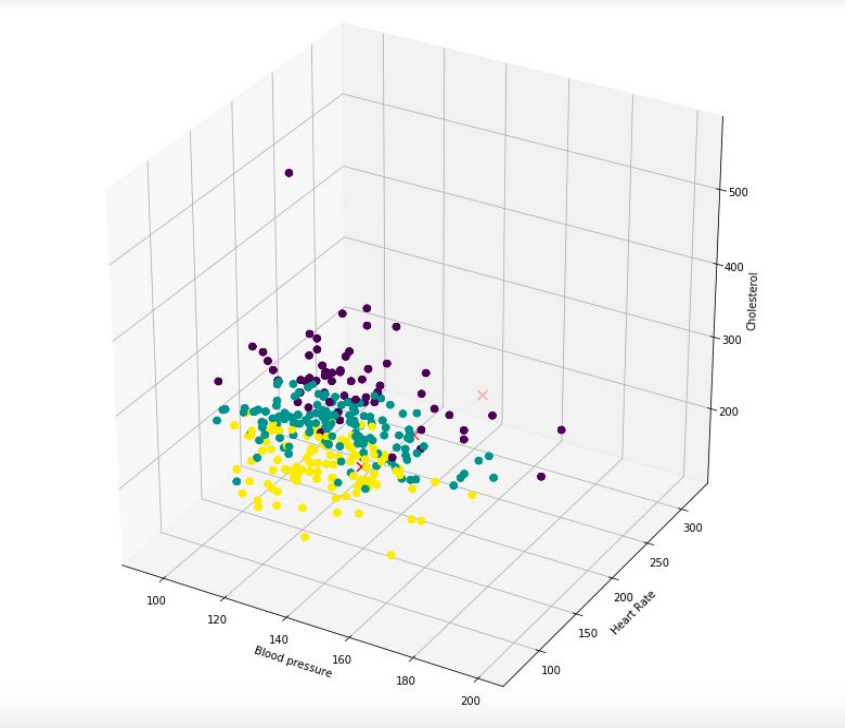
```
In [92]: from sklearn.metrics.pairwise import euclidean_distances
         sub_data2_distance = DataFrame(euclidean_distances(sub_data2))
         sub_data2_distance.astype(int).head()
```

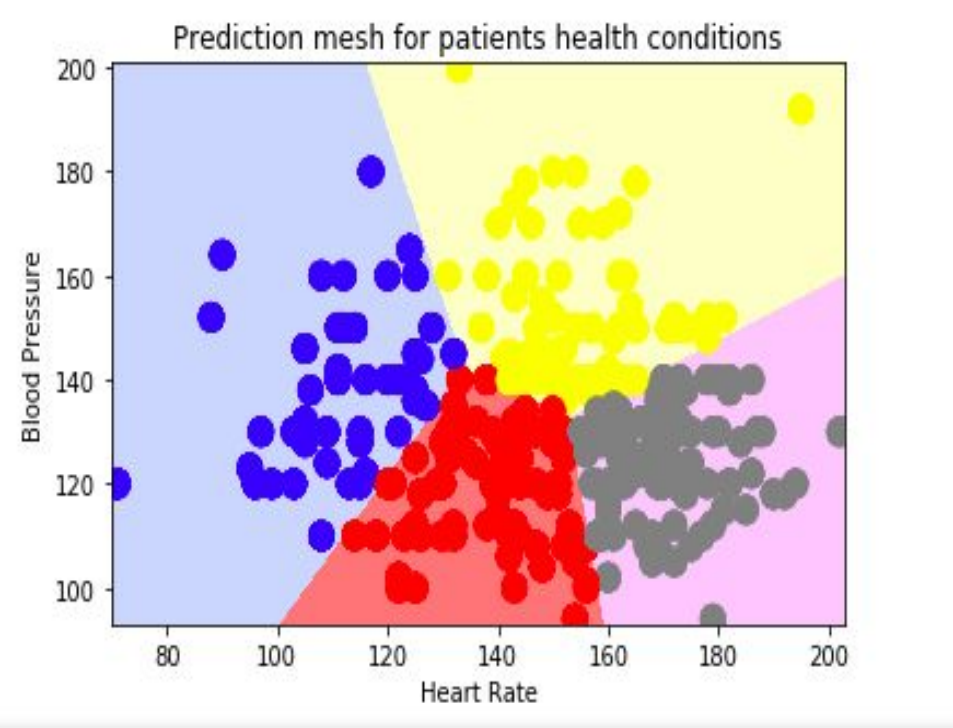
Out[92]:

	0	1	2	3	4	5	6	7	8	9	...	1015	1016	1017	1018	1019	1020	1021	1022	1023	1024
0	0	21	63	27	104	63	110	87	44	90	...	37	72	101	45	29	19	53	81	45	60
1	21	0	45	13	103	68	118	88	51	93	...	30	82	100	44	31	21	59	86	59	48
2	63	45	0	47	121	87	148	118	84	115	...	47	118	115	56	53	62	88	109	95	34
3	27	13	47	0	106	76	121	88	58	98	...	38	80	106	53	42	20	62	93	64	57
4	104	103	121	106	0	61	48	45	63	22	...	82	69	23	137	101	93	51	38	73	108



Kmean


Cholesterol
500
400
300
200
Blood pressure
100
120
140
160
180
200
Heart Rate
300
250
200
150
100


Prediction mesh for patients health conditions
Blood Pressure
200
180
160
140
120
100
Heart Rate
80
100
120
140
160
180
200

KNN Clustering

- Features: Blood pressure and Heart Rate
- Target: integer valued 0 = no disease and 1 = disease
- Split data into train and test data
- Optimal number of neighbors is 1
- Accuracy Score is 0.92. (20% increase)

Prediction

- Predict heart disease based on the health symptom of the patients .
- Blood pressure, heart rate, cholesterol, age, depression.
- 5 functions that take different parameters.
- Heart disease and No heart disease.

```
In [127]: # make prediction for such parameters as Blood Pressure and Heart Rate
          def prediction():
              bl_pres = int(input('Blood Pressure: '))
              heart_rate = int(input('Heart rate : '))
              data_class = knn.predict(np.array([bl_pres, heart_rate]).reshape(1, -1))[0]
              class_name = ["No Heart Disease", "Heart Disease"]
              print('Prediction: Patient X has', data_class, class_name[data_class])
              return prediction

In [128]: prediction()

          Blood Pressure: 155
          Heart rate : 140
          Prediction: Patient X has 1 Heart Disease
```

Conclusion

- The findings demonstrates similarities in symptom during patient's health conditions observation.
- Results may help clinicians to prepare at risk patients for proper treatment seeking and symptom self- management behaviors.