

Big HW. LazyFca

Denis Sharakshinov

1. Description of dataset

I have chosen Heart Attack Analysis Prediction Dataset About this dataset

1. Age: displays the age of the individual.
2. Sex: displays the gender of the individual using the following format :
1 = male
0 = female
3. Chest-pain type ("cp"): displays the type of chest-pain experienced by the individual using the following format :
1 = typical angina
2 = atypical angina
3 = non — anginal pain
4 = asymptotic
4. Resting Blood Pressure("trestbps"): displays the resting blood pressure value of an individual in mmHg (unit)
5. Serum Cholestrol("chol"): displays the serum cholesterol in mg/dl (unit)
6. Fasting Blood Sugar("fbs"): compares the fasting blood sugar value of an individual with 120mg/dl. If fasting blood sugar > 120mg/dl then :
1 (true)
else : 0 (false)
7. Resting ECG("restecg") : displays resting electrocardiographic results
0 = normal
1 = having ST-T wave abnormality
2 = left ventricular hyperthrophy
8. Max heart rate achieved : displays the max heart rate achieved by an individual.
9. Exercise induced angina :
1 = yes
0 = no
10. ST depression induced by exercise relative to rest: displays the value which is an integer or float.
11. Peak exercise ST segment :
1 = upsloping
2 = flat
3 = downsloping

12. Number of major vessels (0–3) colored by flourosopy : displays the value as integer or float.
13. Thal : displays the thalassemia :
 3 = normal
 6 = fixed defect
 7 = reversible defect
14. Diagnosis of heart disease : Displays whether the individual is suffering from heart disease or not :
 0 = absence
 1 = present

2. Data Pre-Processing

The dataset is shown below:

	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

First of all, we should binarize it (details in code):

	age29_40: 0	age29_40: 1	age41_60: 0	age41_60: 1	age61_77: 0	age61_77: 1	male: 0	male: 1	female: 0	female: 1	...	caa4: 0	caa4: 1	thal0: 0	thal0: 1	thal1: 0	tha
0	True	False	False	True	True	False	True	False	False	True	...	True	False	True	False	True	F
1	True	False	False	True	True	False	False	True	True	False	...	True	False	True	False	True	F
2	True	False	False	True	True	False	False	True	True	False	...	True	False	True	False	True	F
3	True	False	True	False	False	True	False	True	True	False	...	True	False	True	False	True	F
4	True	False	False	True	True	False	False	True	True	False	...	True	False	True	False	True	F

3. Comparison with classical classification algorithms

Here, I decided to compare the results with the other classification algorithms. There were used next classifiers:

1. Random Forest Classifier

2. K - nearest neighbors algorithm (KNN)

3. Decision Tree

Classifier	Accuracy	
Lazy_fca	83.09	
f1_score	85.76	
KNN	87.86	
RandomForest	87.09	
Decision Tree	83.87	