Cardio Vascular disease prediction

```
data = pd.read_csv('cardio_train.csv',sep=";")
```

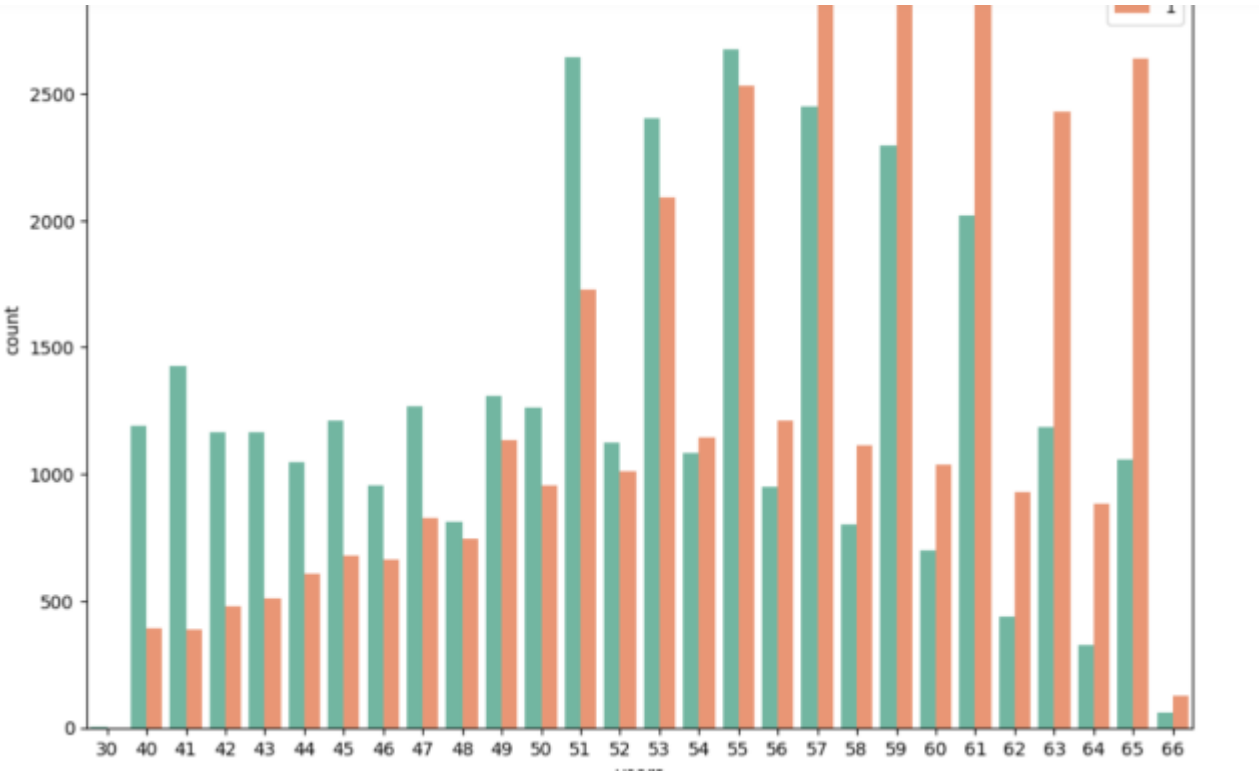
```
data.head()
```

	id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
0	0	18393	2	168	62.0	110	80	1	1	0	0	1	0
1	1	20228	1	156	85.0	140	90	3	1	0	0	1	1
2	2	18857	1	165	64.0	130	70	3	1	0	0	0	1
3	3	17623	2	169	82.0	150	100	1	1	0	0	1	1
4	4	17474	1	156	56.0	100	60	1	1	0	0	0	0

```
data.drop("id",axis=1,inplace=True)
```

```
data['years'] = (data['age'] / 360).round().astype('int')
```

```
data["bmi"] = data["weight"] / (data["height"]/100)**2
```



```
data.head()
```

	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio	years	bmi
0	18393	2	168	62.0	110	80	1	1	0	0	1	0	51	21.967120
1	20228	1	156	85.0	140	90	3	1	0	0	1	1	56	34.927679
2	18857	1	165	64.0	130	70	3	1	0	0	0	1	52	23.507805
3	17623	2	169	82.0	150	100	1	1	0	0	1	1	49	28.710479
4	17474	1	156	56.0	100	60	1	1	0	0	0	0	49	23.011177

```
data.columns
```

```
Index(['age', 'gender', 'height', 'weight', 'ap_hi', 'ap_lo', 'cholesterol',
       'gluc', 'smoke', 'alco', 'active', 'cardio', 'years', 'bmi'],
      dtype='object')
```

In [45]:
```
models = {
    "LR": LogisticRegression(),
    "KNN": KNeighborsClassifier(),
    "DT": DecisionTreeClassifier(),
    "RF": RandomForestClassifier(),
    "XGB": XGBClassifier(),
    "Naive Bayes": GaussianNB(),
    "SVC": SVC()

}
```

Testing Accuracy:

```
------------------------------
Training Model KNN
--------------
Training Accuracy: 0.7786428571428572
Testing Accuracy: 0.68
Testing Confusion Matrix:
[[4948 2121]
 [2359 4572]]
Testing Recall: 0.659645072861059
Testing Precesion: 0.6831017480950247
Testing F-1: 0.6711685261303583
Testing F-Beta: 0.6782778981099606
------------------------------
```

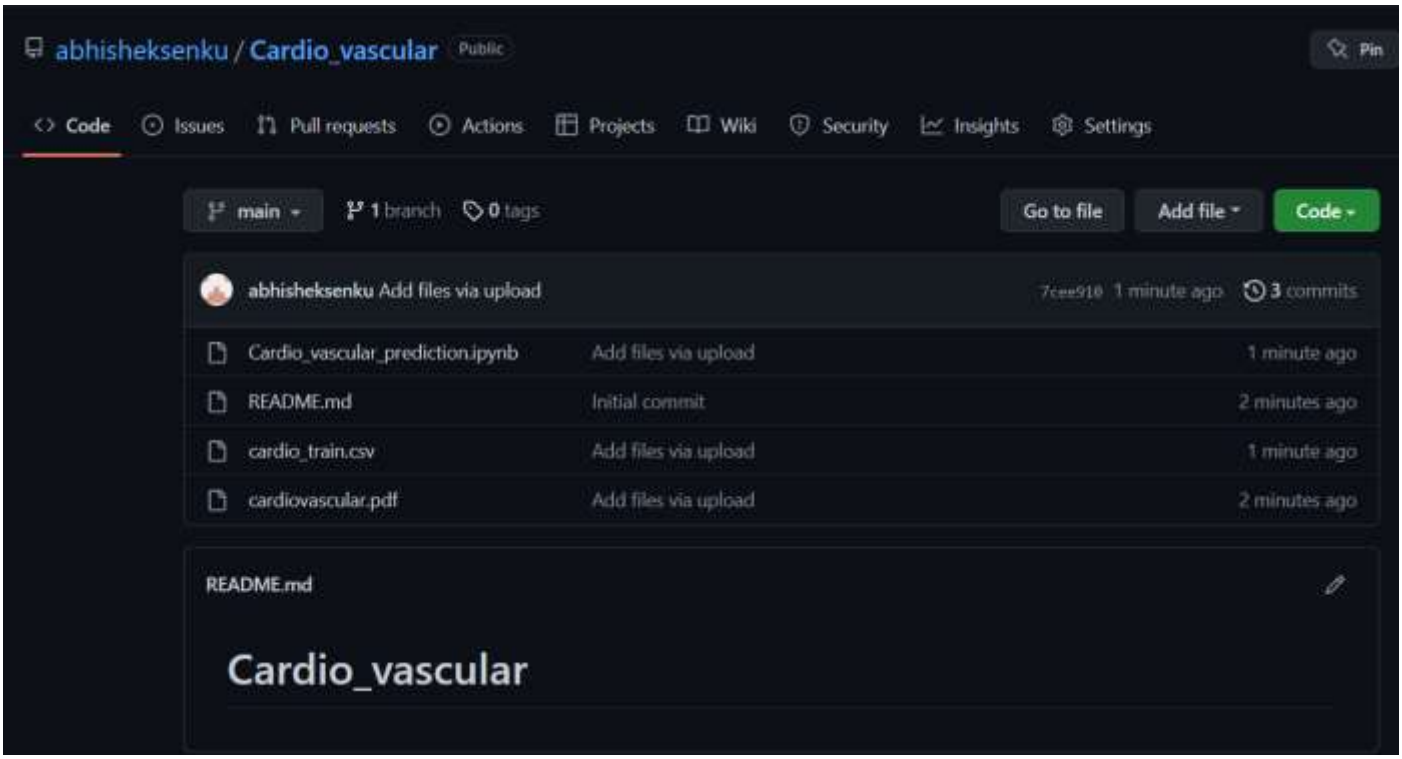
```
Training Model DT
--------------
Training Accuracy: 0.9998035714285715
Testing Accuracy: 0.6329285714285714
Testing Confusion Matrix:
[[4460 2609]
 [2530 4401]]
Testing Recall: 0.634973308324917
Testing Precesion: 0.6278174037089872
Testing F-1: 0.6313750806972241
Testing F-Beta: 0.6292356523977011
------------------------------
```

```
Training Model RF
--------------
Training Accuracy: 0.9997857142857143
Testing Accuracy: 0.7111428571428572
Testing Confusion Matrix:
[[5148 1921]
 [2123 4808]]
Testing Recall: 0.6936949935074304
Testing Precesion: 0.7145192450587011
Testing F-1: 0.7039531478770131
Testing F-Beta: 0.7102549708984548
------------------------------
```

```
Training Model XGB
--------------
Training Accuracy: 0.7689642857142858
Testing Accuracy: 0.7310714285714286
Testing Confusion Matrix:
[[5468 1601]
 [2164 4767]]
Testing Recall: 0.6877795411917472
Testing Precesion: 0.7485866834170855
Testing F-1: 0.7168960072185879
Testing F-Beta: 0.7355800388852883
```

```
Training Model Naive Bayes
--------------
Training Accuracy: 0.5953571428571428
Testing Accuracy: 0.5960714285714286
Testing Confusion Matrix:
[[6239  830]
 [4825 2106]]
Testing Recall: 0.30385225797143267
Testing Precesion: 0.7173024523160763
Testing F-1: 0.4268774703557312
Testing F-Beta: 0.563855421686747
```

```
Training Model SVC
--------------
Training Accuracy: 0.6054642857142857
Testing Accuracy: 0.5981428571428572
Testing Confusion Matrix:
[[4600 2469]
 [3157 3774]]
Testing Recall: 0.5445101716923965
Testing Precesion: 0.6045170591061989
Testing F-1: 0.572946713223015
Testing F-Beta: 0.5914804250383976
------------------------------
```



Data source link : https://github.com/abhisheksenku/Cardio_vascular