**EEE 591**

**Project 1**

**Trisha Ashok**

**1226289911**

**AIM:**

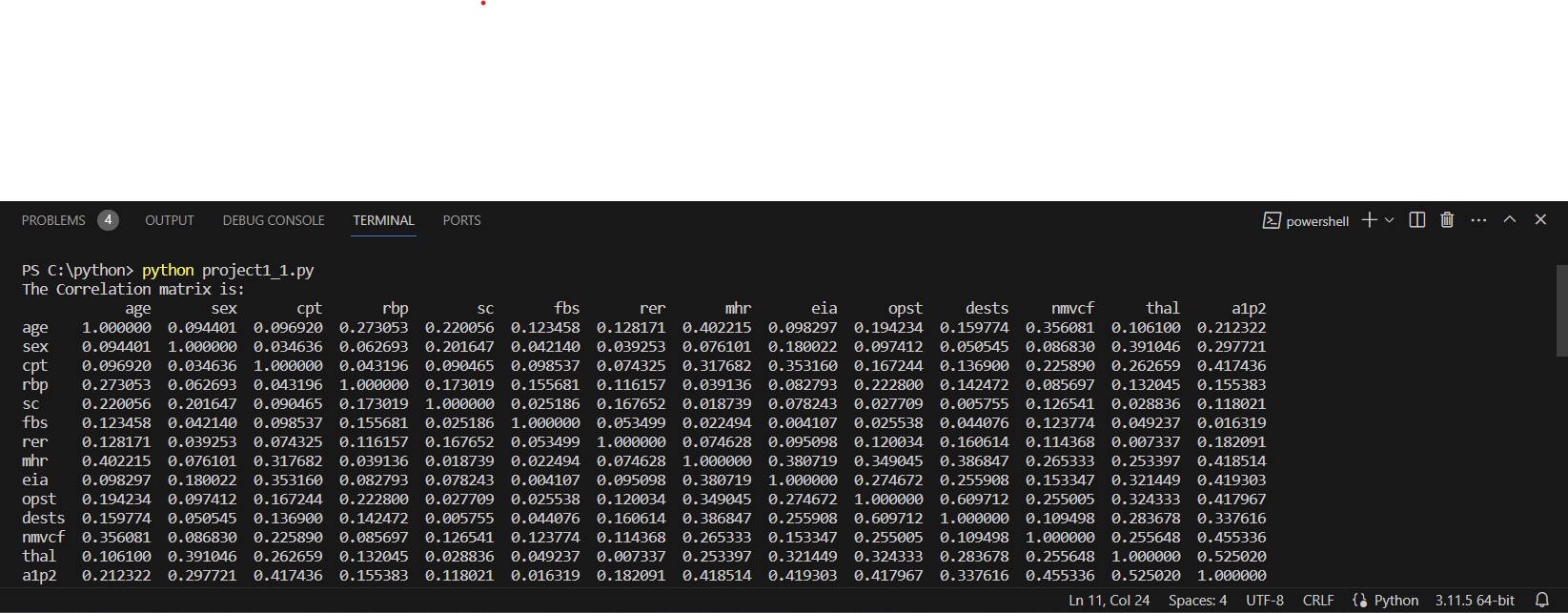
To obtain the correlation and covariance for each variable in the dataset pertaining to the heart disease prediction for the doctors from AMAPE. Also, to determine the highest achievable accuracy for predicting heart disease using machine learning algorithms.

The Machine Learning Algorithms to be used:

* Perceptron
* Logistic Regression
* Support Vector Machine
* Decision Tree
* Rain forest
* K Nearest Neighbor

**Problem1:**

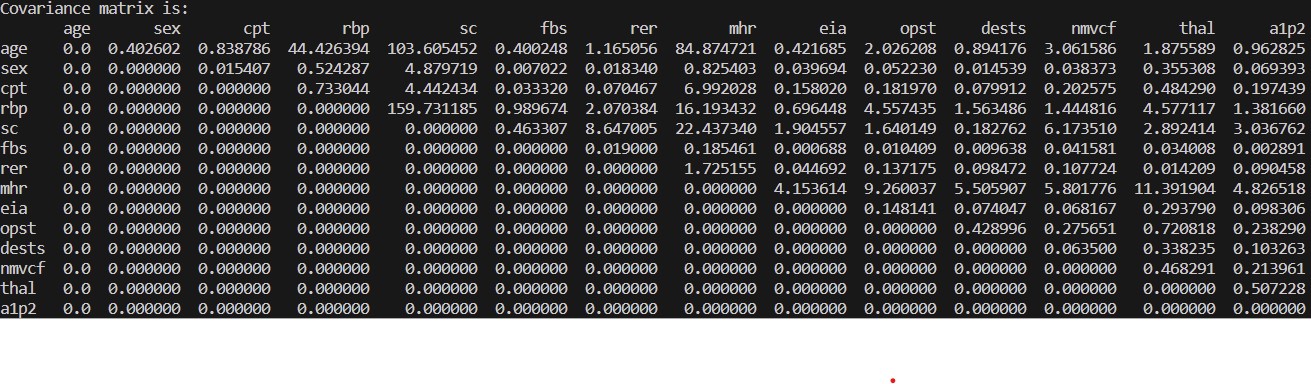
Correlation matrix:



**Result:**

It can be observed from the dataset that the **a1p2**(absence and presence of a heart disease) is highly correlated with **thal.** Also**,** there is a slight dependency of **mhr,eia, opst,cpt** with **a1p2** and all of them almost has same correlation. In contrast, any changes **i**n **fbs** will not affect **a1p2** since they are independent to each other.

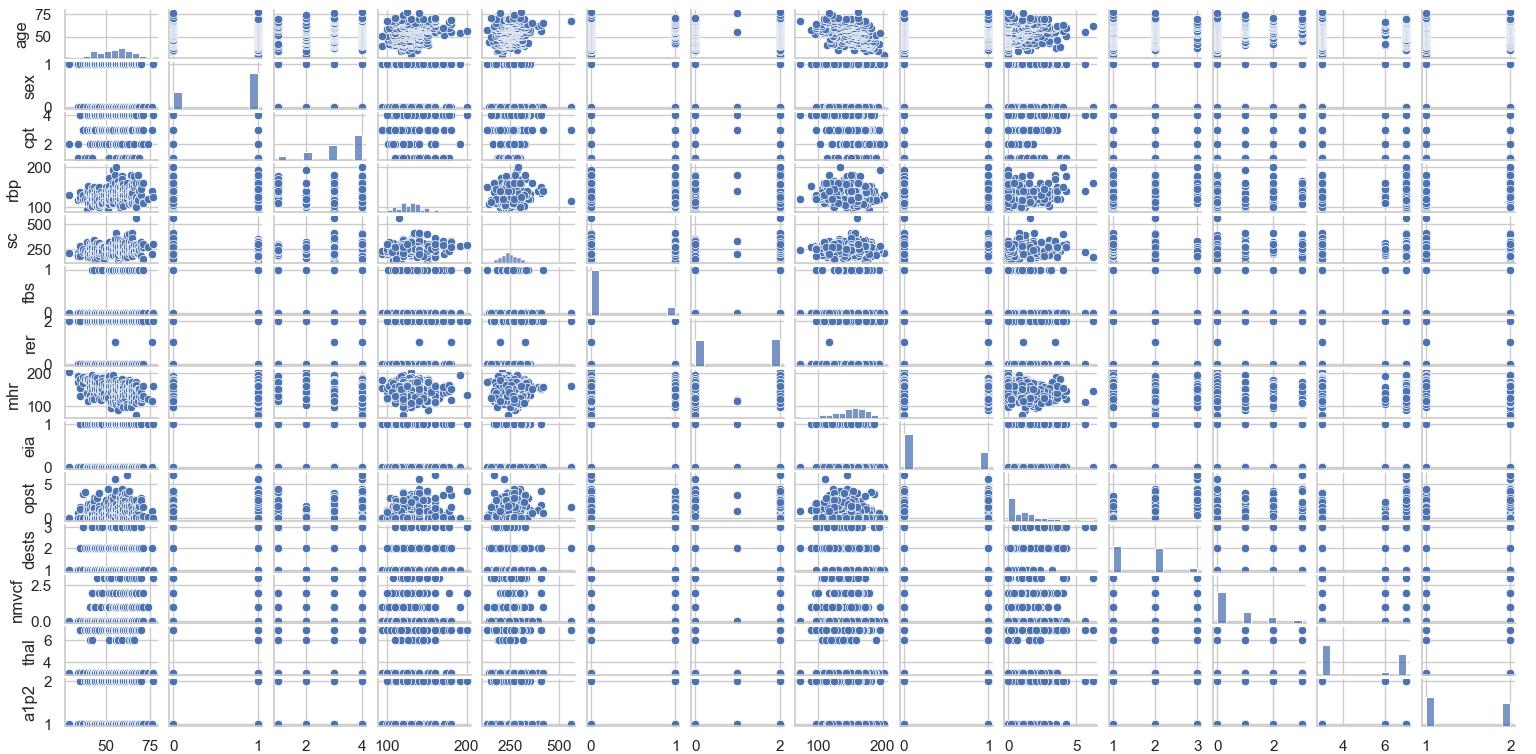
Covariance matrix:

****

**Result:**

It can be observed from the above matrix that the prediction of heart disease is has highest covariance with **sc.** Also**, thal** has slightly measurable covariance for prediction. In contrast, **sex** doesn’t matter for this prediction, so it has zero covariance.

Pair plot:





**Result:**

All the correlations can be observed here with the pair plot. This shows all the best correlated predictions for heart disease. More information can be obtained from **a1p1** with those that are underlined because the data is not overlapped with each other. At some point, it is very difficult to obtain the information as the data seems to be overlapped too much.

**Problem2:**

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **Accuracy** | **Combined Accuracy** |
| Perceptron | 75% | 80% |
| Logistic Regression | 77% | 86% |
| SVM | 75% | 86% |
| Decision Tree | 77% | 92% |
| Random Forest | 75% | 93% |
| K Nearest Neighbor | 73% | 92% |

**Result:**

Based on the observations from the table above, it appears that RANDOM FOREST is the most accurate Machine Learning Algorithm for heart disease. So, it will predict the heart disease for a person accurately using the app. Other models are comparatively producing less accurate results for the doctors but still they can be used.

Remarks:

Hence the correlation and covariance between each and every variable from the data set has been predicted along with the accuracy of the disease using the algorithms.