# Task 1

If we see into our heart dataset we see there have a target column. Which contain two value that is 1/0. This is a categorical problem. We know that this type of problem can solve using Unsupervised ml algorithm. So I choice SVM to solve this problem because of Support Vector Machine is a linear model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is simple. The algorithm creates a line or a hyperplane which separates the data into classes.

Why I choose SVM:

* The SVM provides a very useful technique within it known as kernel and by the application of associated kernel function we can solve any complex problem
* SVM generally do not suffer condition of overfitting and performs well when there is a clear indication of separation between classes
* SVM Algorithm is that it is able to handle High dimensional data
* Support Vector Machine is useful in finding the separating Hyperplane, finding a hyperplane can be useful to classify the data correctly between different groups

Note: I attached my code into the zip file.

# Task 2

A dataset in machine learning is, quite simply, a collection of data pieces that can be treated by a computer as a single unit for analytic and prediction purposes. This means that the data collected should be made uniform and understandable for a machine that doesn't see data the same way as humans do.

Columns: column arranges data vertically from top to bottom. In my dataset I have 14 columns.

Row:  row arranges data horizontally from left to right. In my dataset I have 1025 rows

Data Size: File size is a measure of how much data a computer file contains or, alternately, how much storage it consumes. My dataset size is 38114 bytes.

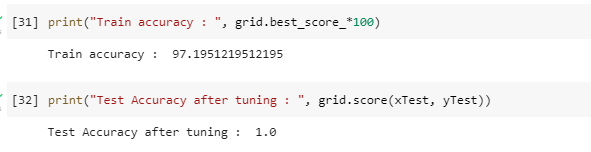
# Task 3

I choice SVM as my model for solving above dataset problem.

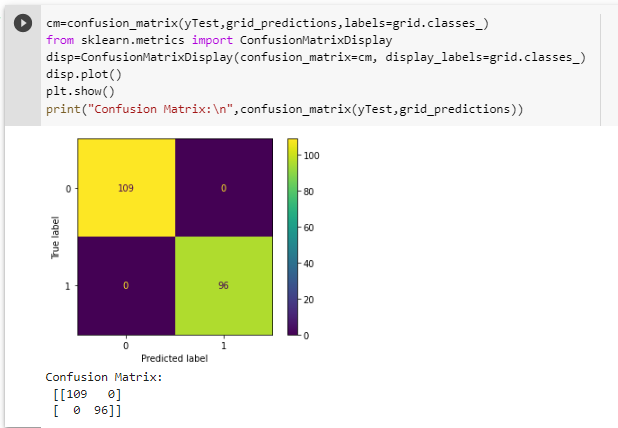
I get accuracy of the model that is:

In Training: 97%

In Testing: 100%



Confusion Matrix: A confusion matrix is a table that is used to define the performance of a classification algorithm.



Confusion matrices represent counts from predicted and actual values.

TP = True Positive

TN = True Negative

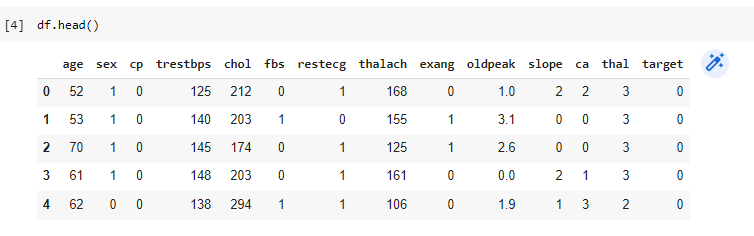
FP = False Positive

FN = False Negative

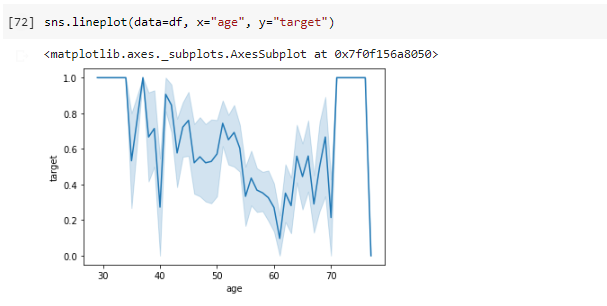
In my dataset I found TP=109, TN=96, FP=0, and FN=0.

# Task 4

Visualize the dataset: In the heart dataset there have 13 feature columns and one target columns



Different types of graph I use to show my dataset.

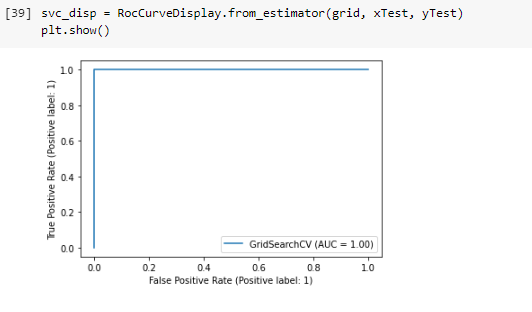




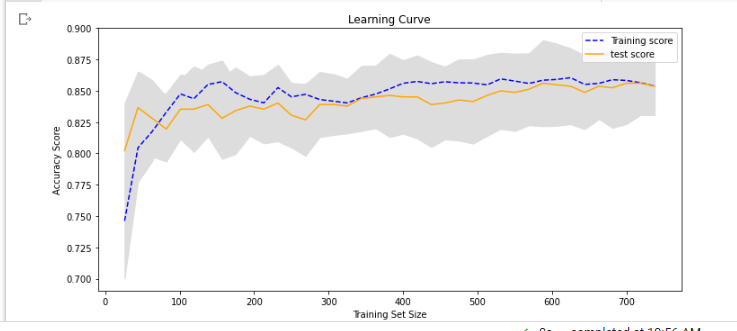
# Task 5

ROC curve:

A ROC curve is constructed by plotting the true positive rate (TPR) against the false positive rate (FPR). The true positive rate is the proportion of observations that were correctly predicted to be positive out of all positive observations (TP/(TP + FN)). Similarly, the false positive rate is the proportion of observations that are incorrectly predicted to be positive out of all negative observations (FP/(TN + FP)).



# Task 6

Learning curve

Train and Test Error Graph:

