# ML-Final Report



**Student:**

**MOHAMMAD ALSHAHROUR**

**Date:**

2021/6/8

* **Time**

It took 1 week to cleanup to pre-preparing the dataset before we start the algorithms and training

Then it took 2 week to implement the main algorithms (KNN , DT , SVM , RandomForest,logistiv) then we added some extra algorithms (K-MEANS,DBSCAN,AGGLOMERATIVE) then we implemented the Feature Selection and Ensemble Learning (K-FOLD AND GRID SEARCH, PCA,AUC).

* **Dataset:**

We used dataset about Survived on Titanic Ship of 419 people are in ship in 1985, from [kaggle.com](file:///D:\COLLAGUE\2020-2\ML\LAB\المشروع%20كامل\kaggle.com)

-After we discover the dataset we found that the data have a lot of

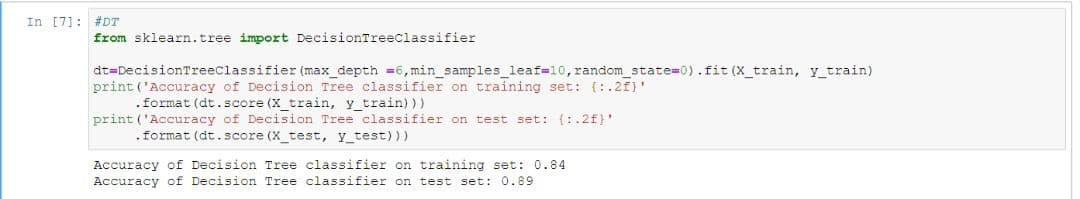
null value in column cabin.

* **Algorithms**
  + **Decision Tree**

Decision tree is the best algorithm scored in our project we put those parameters :

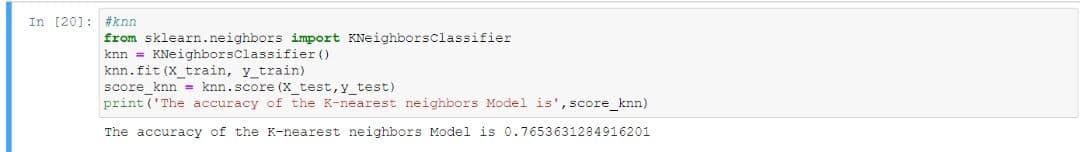
1. max\_depth =6 because we choose 6 columns on dataset
2. min\_samples\_leaf=10

as a result It scored 89%



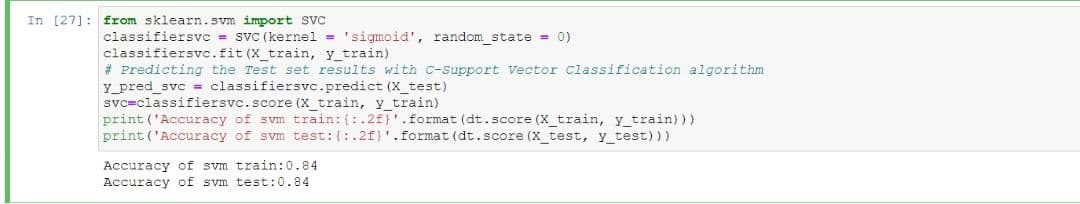
* + **KNN Regression**

Logically KNN is a bad idea to apply to our project because not all the features should be weighted the same as others so distance between nodes are useless, but we implemented it any way, we tried the N neighbor is the maximum the prediction is score was 76.5%, and its unexpected result, it is better than we thought.



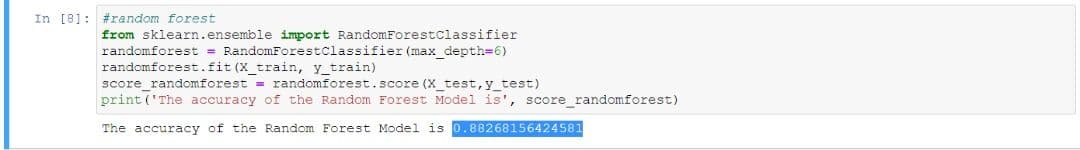
* + **Support-Vector Machine (SVM)**

We used SVM in our project to train and predict then it reached 84% result as the highest attempt



* + **Random Forest Algorithms:**

We applied random forest on the data with max depth = 6 and it did well, 88.2% was the result of this algorithm

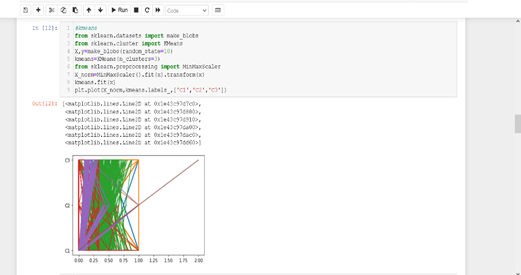


.

Also, we tried PCA on Pipeline with other algorithms for feature selection and it didn’t improve the result with any of the algorithms because it score:67%.



**K-means**

we applied k-means algorithm that use clustering and we use 3 clustering with statment (n\_clusters=3)

**gird search and k-fold:**

**we use that two in our project**

* **Environment**

We used anaconda for implementing the pre-processing and data science section and implementing non-time-expensive algorithms then when we wanted to train the time-expensive algorithms we used Google Colab platform, it really helped us and saved a much time than local training.

**Submission Students:**

**MOHAMMAD ALSHAHROUR**