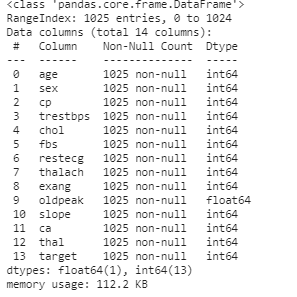
|  |  |
| --- | --- |
| **السكشن** | **الاسم** |
| 3 cs | محمد عادل حلمي الكردي |
| 4cs | مصطفي منير شهاب |
| 3cs | محمد سامي الحبشي |
| 3cs | ماركو مجدي كيرلس |

**Project Name: heart disease**

## Introduction

World Health Organization has estimated 12 million deaths occur worldwide, every year due to Heart diseases. Half the deaths in the United States and other developed countries are due to cardio vascular diseases. The early prognosis of cardiovascular diseases can aid in making decisions on lifestyle changes in high risk patients and in turn reduce the complications.

we use data set called (heart\_disease ) .



We represented age,sex,cp(chest pain type),trestbps(resting blood pressure),chol(serum cholestroral),pbs(fasting blood sugar),vestecg(resting electrocardiographic),thalach(maximum heart rate achieved),exang(exercise include agina),oldpeak(st depression include by exercise).

In this project we will use 3 classification models:

1. Decision Tree Classifier
2. k-nearest neighbors Classifier

3-Random Forest Classifier

**Decision Tree Classifier**

**What Is a Decision Tree?**

A decision tree is a managerial tool that presents all the decision alternatives and outcomes in a flowchart type of diagram, like a tree with branches and leaves. Each branch of the tree represents a decision option, its cost and the probability that it is likely to occur. The leaves at the end of the branches show the possible payoffs or outcomes. A decision tree illustrates graphically all the possible alternatives, probabilities and outcomes and identifies the benefits of using decision analysis.

## Advantages of Decision Tree Classifier:

## Easy to Use

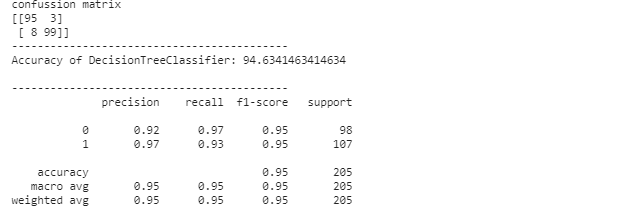
## Decision trees are easy to use and explain with simple math, no complex formulas. They present visually all of the decision alternatives for quick comparisons in a format that is easy to understand with only brief explanations.

## Specific

## Decision trees assign specific values to each problem, decision path and outcome. Using monetary values makes costs and benefits explicit. This approach identifies the relevant decision paths, reduces uncertainty, clears up ambiguity and clarifies the financial consequences of various courses of action.

## Disadvantages of Decision Tree Classifier:

* For a Decision tree sometimes calculation can go far more complex compared to other algorithms.
* Decision tree often involves higher time to train the model.
* Decision tree training is relatively expensive as the complexity and time has taken are more.
* **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**



**KNN Classifier**

**What Is a KNN Classifier?**

* KNN is a very simple algorithm used to solve classification problems. KNN stands for K-Nearest Neighbors. K is the number of neighbors in KNN. Lets find out some advantages and disadvantages of KNN algorithm

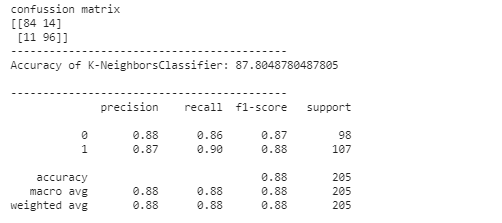
**Advantages of KNN Classifier:**

* Since the KNN algorithm requires no training before making predictions, new data can be added seamlessly which will not impact the accuracy of the algorithm.
* KNN is very **easy to implement**. There are only two parameters required to implement KNN i.e. the value of K and the distance function (e.g. Euclidean or Manhattan etc.)

**Disadvantages of KNN Classifier:**

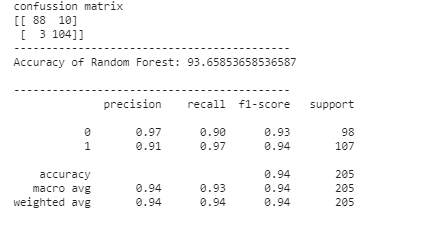
* **Does not work well with large dataset:**In large datasets, the cost of calculating the distance between the new point and each existing points is huge which degrades the performance of the algorithm
* **Sensitive to noisy data, missing values and outliers**: KNN is sensitive to noise in the dataset. We need to manually impute missing values and remove outliers.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**



Random Forest

Random Forest is a powerful algorithm in Machine Learning. It is based on the Ensemble Learning technique (bagging). Following are the advantages and disadvantages of Random Forest algorithm.  
  
**Advantages of Random Forest**  
  
**1.** Random Forest is based on the **bagging** algorithm and uses **Ensemble Learning** technique. It creates as many trees on the subset of the data and combines the output of all the trees. In this way it **reduces overfitting** problem in decision trees and also**reduces the variance**and therefore **improves the accuracy**.  
  
**2.** Random Forest can be used to **solve both classification as well as regression problems**.  
  
**3.** Random Forest works well with both **categorical and continuous variables**.  
  
**4.** Random Forest can automatically **handle missing values**.  
  
**5. No feature scaling required:** No feature scaling (standardization and normalization) required in case of Random Forest as it uses rule based approach instead of distance calculation.  
  
**6. Handles non-linear parameters efficiently:** Non linear parameters don't affect the performance of a Random Forest unlike curve based algorithms. So, if there is high non-linearity between the independent variables, Random Forest may outperform as compared to other curve based algorithms.  
  
**7.** Random Forest can automatically **handle missing values**.  
 **8.** Random Forest is usually **robust to outliers** and can handle them automatically.  
 **9.** Random Forest algorithm is very **stable**. Even if a new data point is introduced in the dataset, the overall algorithm is not affected much since the new data may impact one tree, but it is very hard for it to impact all the trees.  
  
**10.** Random Forest is comparatively **less impacted by noise**.  
  
**Disadvantages of Random Forest**  
  
**1. Complexity:** Random Forest creates a lot of trees (unlike only one tree in case of decision tree) and combines their outputs. By default, it creates 100 trees in Python sklearn library. To do so, this algorithm requires much more computational power and resources. On the other hand decision tree is simple and does not require so much computational resources.  
  
**2. Longer Training Period:** Random Forest require much more time to train as compared to decision trees as it generates a lot of trees (instead of one tree in case of decision tree) and makes decision on the majority of votes.



**The Best Model In 3 Model is (Decision tree classifier )because it has the best Accuracy**

