Project Name: Indian Liver Patient

السكشن	الاسم
2 CS	شروق فرحات عبدالغفار حسين

we use data set called (indian_liver_patient_weka_dataset) and it has **583** records .

Based on certain data, we determine whether a person is Indian Liver Patient or not .

The Content of our Data IS:

Instances: 583
Attributes: 11
Tasks: Classification
Downloads: 2149
Year Published: 2012
Missing Values: No

Attribute Details:

Name	Type	Description
age	integer	Age of the patient in years
gender	string	Patient Gender: Male or Female
TB	float	Total Bilirubin
DB	float	Direct Bilirubin
alkphos	float	Alkaline Phosphotase
sgpt	float	Alamine Aminotransferase
sgot	float	Aspartate Aminotransferase
TP	float	Total Proteins
ALB	float	Albumin
A_G	float	Ratio of Albumin and Globulin
class	float	Predictor Class: 1 if patient has Liver Disease and 2 if they do not

We represented male as 1 and female as 0.

and we represented Patient as 1 and non-patient as 0.

and there are no missing values because we don't make preprocessing on Data.

In this project we will use 3 classification models:

1- Decision Tree Classifier … (30% Test & 70% training).
2- k-nearest neighbors Classifier … K=3 … (20% Test & 80% training).
3- Neural Network … (30% Test & 70% training).

Loaded data:-

```
In [1]: runfile('E:/Final Project/MY.py', wdir='E:/Final Project')
   age  gender    TB   DB  alkphos  sgpt  sgot   TP  ALB   A_G  label
0   65       0   0.7  0.1      187    16    18  6.8  3.3  0.90      1
1   65       0   0.7  0.1      187    16    18  6.8  3.3  0.90      1
2   62       1  10.9  5.5      699    64   100  7.5  3.2  0.74      1
3   62       1   7.3  4.1      490    60    68  7.0  3.3  0.89      1
4   58       1   1.0  0.4      182    14    20  6.8  3.4  1.00      1
-----------Model1 Acurracy------------
```

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.
- _____

```
-----------Model1 Acurracy-------------
Accuracy: 0.6477272727272727
Confusion matrix:
[[24 32]
 [30 90]]
              precision    recall  f1-score   support

           0       0.44      0.43      0.44        56
           1       0.74      0.75      0.74       120

    accuracy                           0.65       176
   macro avg       0.59      0.59      0.59       176
weighted avg       0.64      0.65      0.65       176
```
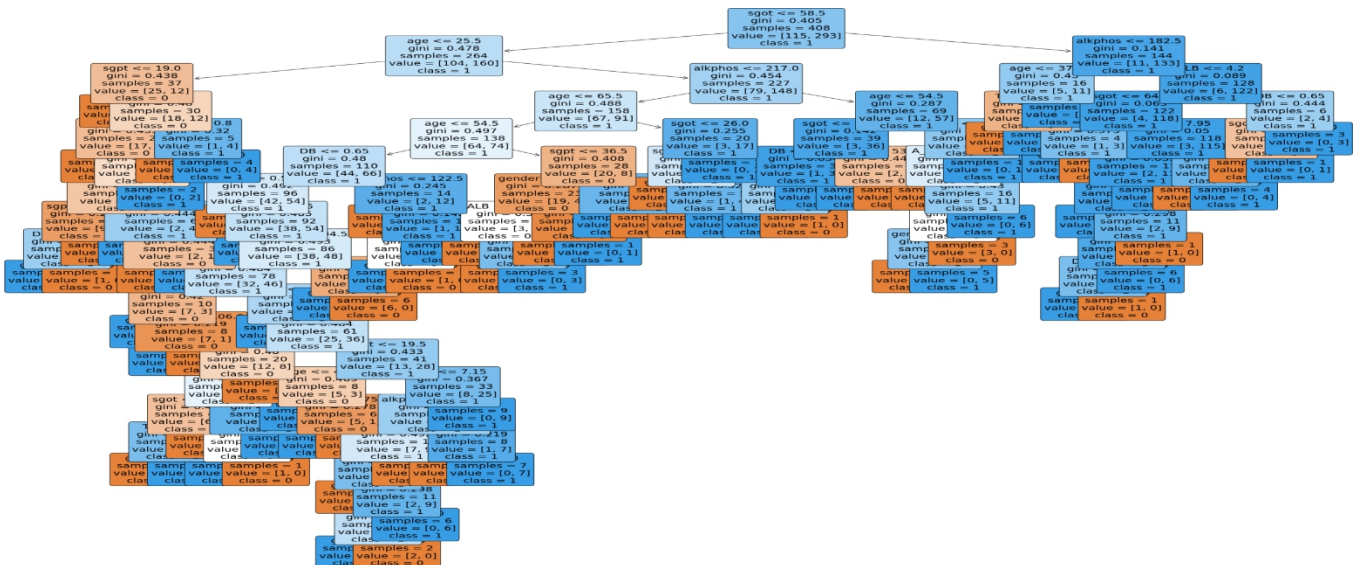
We make evaluate the model and out confusion matrix and Accuracy of it and the result is :

Accuracy = 0.6534090909090

TP = 32, FP = 33, FN = 28, TN = 83.

Then calculate precision, recall, f1-score and support.

Then we make visualization to our decision tree: -

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.

```
-----------Model2 Acurracy------------
Accuracy: 0.6923076923076923
Confusion matrix:
[[ 9 19]
 [17 72]]
              precision    recall  f1-score   support

           0       0.35      0.32      0.33        28
           1       0.79      0.81      0.80        89

    accuracy                           0.69       117
   macro avg       0.57      0.57      0.57       117
weighted avg       0.68      0.69      0.69       117
```

We make evaluate the model and out confusion matrix and Accuracy of it and the result is :

Accuracy = 0.7094017094017094

TP = 15, FP = 19, FN = 15, TN = 68.

Then calculate precision, recall, f1-score and support

Neural Network Classifier

- neural networks are relatively crude electronic networks of neurons based on the neural structure of the brain. They process records one at a time, and learn by comparing their classification of the record (i.e., largely arbitrary) with the known actual classification of the record. The errors from the initial classification of the first record is fed back into the network, and used to modify the networks algorithm for further iterations.

Advantages of Neural Networks:

- **Storing information on the entire network:** Information such as in traditional programming is stored on the entire network, not on a database. The disappearance of a few pieces of information in one place does not prevent the network from functioning**.**
- **Ability to make machine learning:** Artificial neural networks learn events and make decisions by commenting on similar events.

Disadvantages of Neural Networks:

- **Hardware dependence:** Artificial neural networks require processors with parallel processing power, in accordance with their structure. For this reason, the realization of the equipment is dependent.
- Difficulty of showing the problem to the network.

_____

```
-----------Model3 Acurracy-------------
Accuracy: 0.7102272727272727
Confusion matrix:
[[  0  51]
 [  0 125]]
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        51
           1       0.71      1.00      0.83       125

    accuracy                           0.71       176
   macro avg       0.36      0.50      0.42       176
weighted avg       0.50      0.71      0.59       176
```

We make evaluate the model and out confusion matrix and Accuracy of it and the result is :

Accuracy = 0.65909090909091

TP = 0, FP = 0, FN = 60, TN = 116.

Then calculate precision, recall, f1-score and support.

The Best Model In 3 Model is (The Neural Network)because it has the best Accuracy

```
 The Neural Network is best
C:\Users\Shorouk\anaconda3\lib\site-packages\sklearn\metrics\_classification.py:
1272: UndefinedMetricWarning: Precision and F-score are ill-defined and being set
to 0.0 in labels with no predicted samples. Use `zero_division` parameter to
control this behavior.
  _warn_prf(average, modifier, msg_start, len(result))

In [2]: runfile('E:/Final Project/MY.py', wdir='E:/Final Project')
```