**Splitting The Dataset Into Dependent And Independent Variable**

**Project Name: statistical machine learning Team ID : PNT2022TMID04163**

**approaches to liver disease prediction**

**Splitting The Dataset Into Dependent And Independent Variable**

* In machine learning, the concept of dependent variable (y) and independent variables(x) is important to understand. Here, Dependent variable is nothing but output in dataset and independent variable is all inputs in the dataset.
* With this in mind, we need to split our dataset into the matrix of independent variables and the vector or dependent variable. Mathematically, Vector is defined as a matrix that has just one column.

## Dependent and Independent Variables

**See the below equation:**

1y = 10a + 2b - 4.3c

It demonstrates that the value of y is dependent on the value of a, b, and c. So, y is referred to as dependent feature or variable and a, b, and c are independent features or variables. Any predictive mathematical model tends to divide the observations (data) into dependent/ independent features in order to determine the causal effect. It should be noted that relationship between dependent and independent variables need not be linear, it can be polynomial. It is common practise while doing experiments to change one independent variable while keeping others constant to see the change caused on the dependent variable.

### Splitting the Data-set into Independent and Dependent Features

In machine learning, the concept of dependent and independent variables is important to understand. In the above dataset, if you look closely, the first four columns (Item\_Category, Gender, Age, Salary) determine the outcome of the fifth, or last, column (Purchased). Intuitively, it means that the decision to buy a product of a given category

### Splitting the Dataset into the Independent Feature Matrix:

1X = df.iloc[:, :-1].values

2print(X)

Output:

1[['Fitness' 'Male' 20 30000]

2['Fitness' 'Female' 50 70000]

3['Food' 'Male' 35 50000]

4['Kitchen' 'Male' 22 40000]

5['Kitchen' 'Female' 30 35000]]

### Extracting the Dataset to Get the Dependent Vector

1Y = df.iloc[:, -1].values

2print(Y)

Output:

1['Yes', 'No', 'Yes', 'No', 'Yes']