**Statistical Machine Learning Approaches to Liver Disease Prediction**

**Team ID: PNT2022TIMD08311**

**Faculty Mentor:** **Team Leader :** Kavuri Swathi

T.Geetha **Team Member :** Jasthi Jyothi

**Team Member :** Karumudi Meghana

**Team Member :** Biridepalli Mounisha

**Splitting the Data-set into Independent and Dependent**

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 (Fitness item, Food product, kitchen goods) is determined by

the Gender (Male, Female), Age, and the Salary of the individual. So, we can say

that Purchased is the dependent variable, the value of which is determined by the

other four variables.

**Skill Tags:**

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. To read the columns, we will use iloc of pandas (used to fix

the indexes for selection) which takes two parameters — [row selection, column

selection].

Let’s split our dataset into independent and dependent variables.

1. The independent variable in the dataset would be considered as 'x'.

2. The dependent variable in the dataset would be considered as 'y'.

Now we will split the data of independent variables.

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

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

2 print(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:

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

2print(Y)

**Output:**

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

In the above code we are creating array or list of the independent variable x with

our selected columns and for dependent variable y we are only taking

the dependent or output or target column.