**SIMPLE LINEAR REGRESSION**

**What is regression analysis**

\*\* Regression is a type of regression analysis where the number of independent variables is one and there is a linear relationship between the independent(x) and dependent(y) variable.

\*\* Linear Regression is a supervised machine learning algorithm where the predicted output is continuous and has a constant slope. It's used to predict values within a continuous range, (e.g. sales, price) rather than trying to classify them into categories.

\*\* Here, we fit a curve/ line to the data points ,in such a manner that the differences between the distances of data points from the curve or line is minimized.

**Two types of liner model**

\*\* Linear Regression is generally classified into two types: Simple Linear Regression. Multiple Linear Regression

Regression analysis predict the relationship between two or more features. The benifit of using regresion analysis are as follows:

\*\* It shows the significant relationships between the label and the features. \*\* It indicates the strength of impact of multiple independent variabales on dependent variables

**LINEAR REGRESSION**

\*\* Linear regression is a supervised techmique.

\*\* Linear Regression is a stastical analysis for predicting the values of a quantitative variable.

\*\* In this technique , the dependent variable is continous ,independent variable can be contionous or discrete , and nature of regression line is linear.

\*\* Linear reression establishes a relationship between dependent variable(Y) and one or more independent variables (X)inpendent variables(X) usin a best fit straight line(also known as regression line).

Y=m\*X+b+e

where b is intercept of y

m is the slope of the line

e is the error term

we have several methods and libraries used

basic informations about the dataset are done by the functions and methids like:

df.shape

df.info

df.describe

df.head()

df.tail()

df.columns

type(df)

Dropping the duplicates 🡪 df.drop\_duplicates()

Checking ull values 🡪df.isnull().sum()

Now we need to create a dependnt and independent variables for the data as X and Y

We can also visualize the data by using hist ,scatter etc

Then we need to split the dataset into tarin and test datasets with some percentage of each (X\_train,X\_test,Y\_train,Y\_test) using train\_test\_split model by importing from sklearn

Now we apply linear regression on train dataset and apply the model on test dataset to get the predicated values and compare the actual and predicted outputs

Now visualize the training model and test set results of the data

We also can evaluate confusion metrics, accuracy and AUC roc curve