**Employee salary Prediction**

Problem Statement:

Predicting total yearly compensation from years of experience ,

base salary ,years at company.

INPUT: base salary,years at company,years of experience

OUTPUT:Total yearly compensation.

Base salary:Base salary is nothing but actual salary.

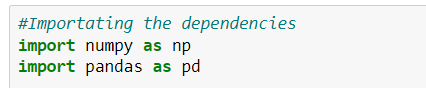
Which the organization give to an employer.

Total Yearly Compensation means Basesalary+Incentive compensation.Total compensation includes the base salary,but it also includes the value of any benefits received in addition to their salary.Some of the benefits that are most commonly provided within a total compensation package include:Bonuses,commissions,paidtimeoff(vacation days,sick days and holidays),Dental insurance,Life insurance,Disability insurances etc….

Importing libraries:

Numpy:Numpy is a python library used for working with arrays.

Pandas:pandas is a python library to deal with sequential and tabular data



Converting .csv file into dataframe

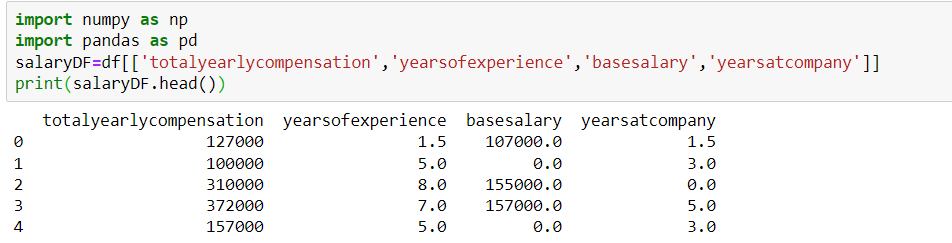


Selecting four columns from the dataframe:

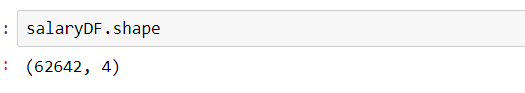
* **Totalyearlycompensation**
* **Year of experience**
* **Years at company**
* **Base salary**

**Preprocessing the data:**

**Head():Head function is used to access the first n rows**

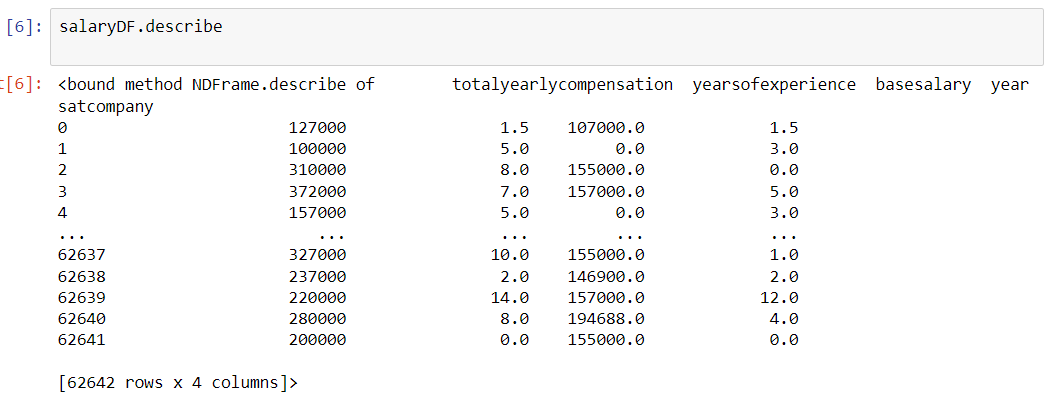
****

**Shape():shape function is used to obtain the shape of dataframe for example:Dataframe has a shape(80,10) this implies that the dataFrame is madeup of 80 rows and 10 columns**

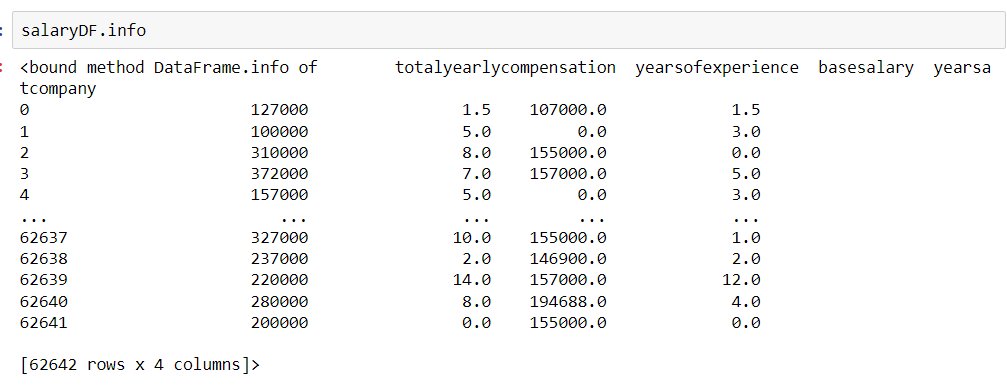
****

Describe():Describe() method returns description of the data in the

dataframe

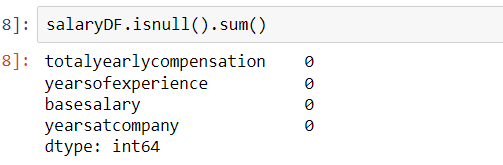


Info():The info method prints information about the Dataframe



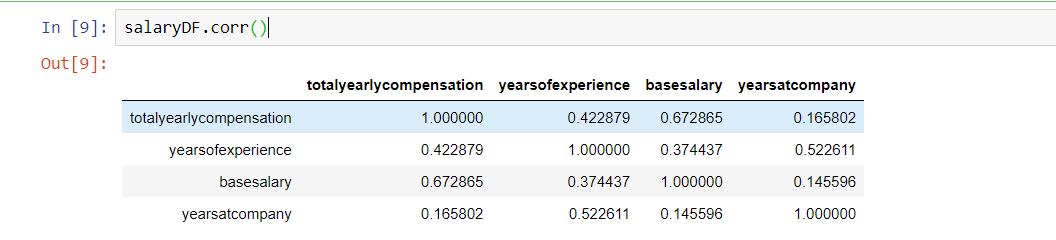
Isnull().sum():The Isnull() function returns the number of missing

values in dataframe



Corr():Corr() is used to find the pairwise correlation of all columns in

the pandas Dataframe in python



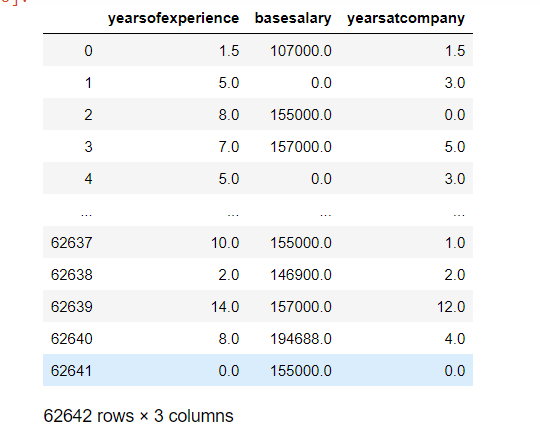
Splitting the features and target:

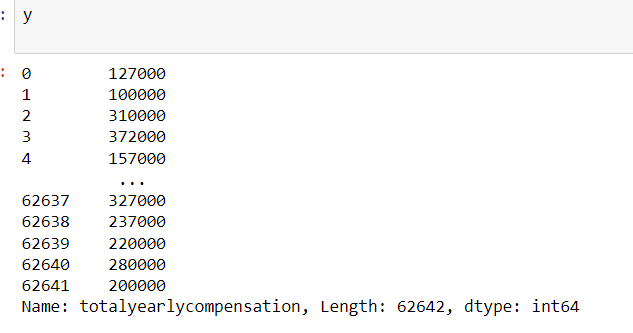
From above columns take “Y”(totalyearlycompensation) as target

And remaining columns taken as Features “x”(yearsofexperience,basesalary,yearsatcompany)

**Applying LinearRegression between X and Y variable**:



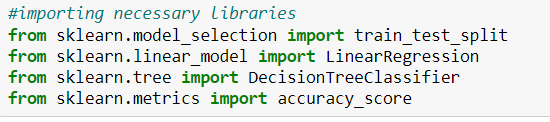




LinearRegression():It defines the relation between

Dependent variables (X) and independent variable(Y)

**Libraries used**:



**1.from sklearn.model\_selection import train\_test\_split**:

train\_test\_split is afunction is Sklearn model selection for splitting data arrays into

two subsets:for training data and for testing data.

2. from sklearn.linear\_model import LinearRegression:

Linear Regression is a machine learning algorithm based on supervised learning.It performs a regression task.It mostly used for finding out the relationship between

variables and forecasting

3. from sklearn.tree import DecisionTreeClassifier:

DecisionTreeClassifier is a class of performing multi-class Classification

On a dataset.Incase that there are multiple classes with the same and

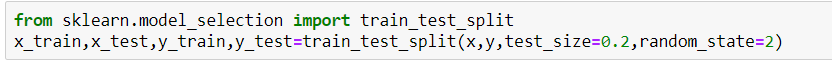
Highest probability,the classifier will predict the class with the lowest

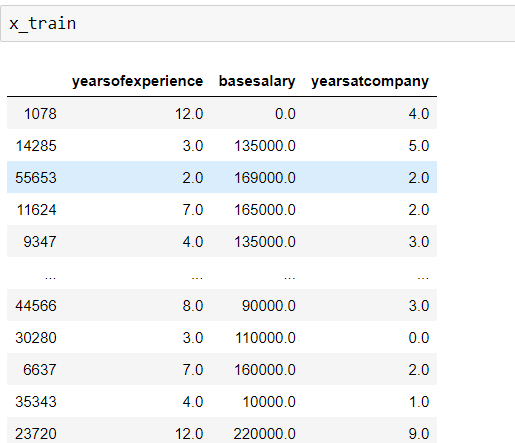
index amongst those classes

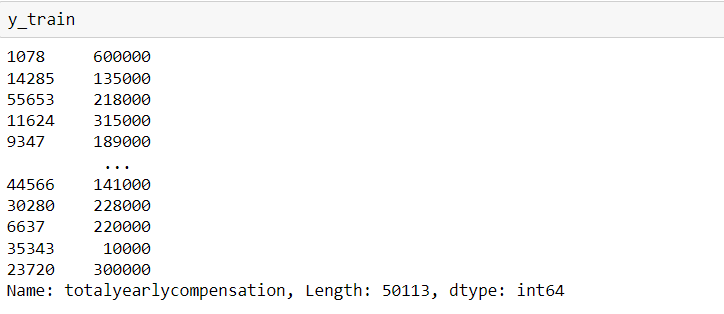
4.from sklearn.metrics import accuracy\_score:

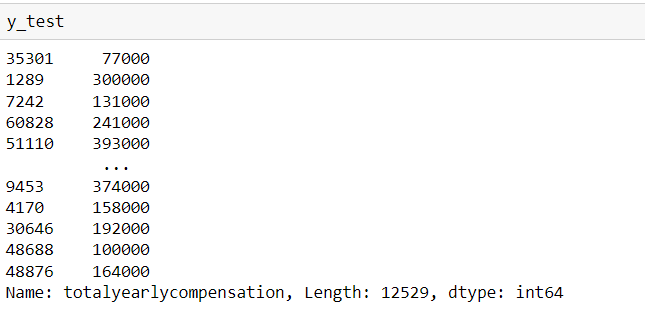
Calculates the accuracy score for a set of predicted labels against the true labels

**Splitting the data into training data and testing data**

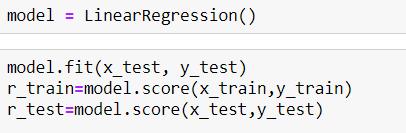




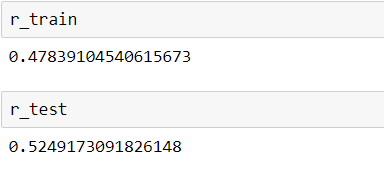


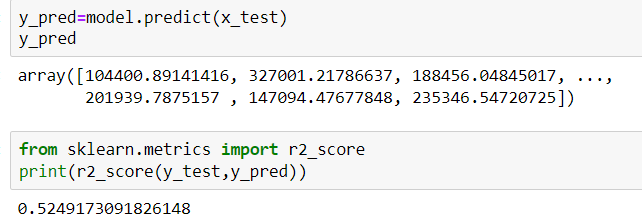


**Model training:**

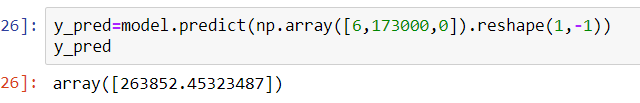


**Model evaluation :(using linearRegression)**

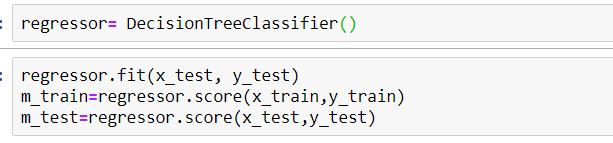


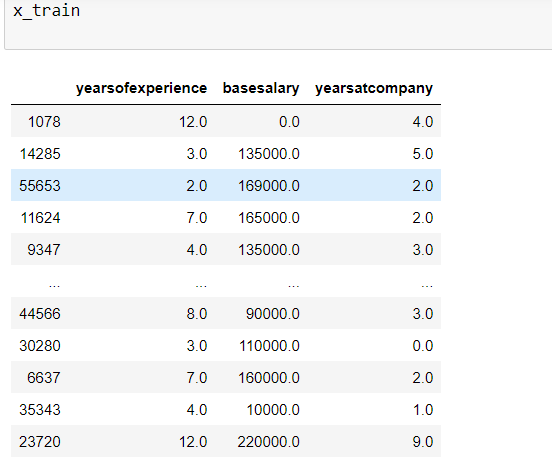


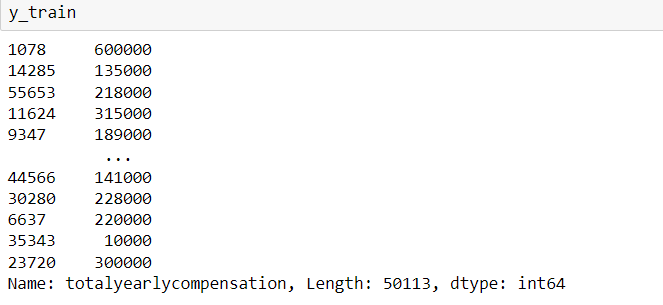
**Building a Predictive system:**

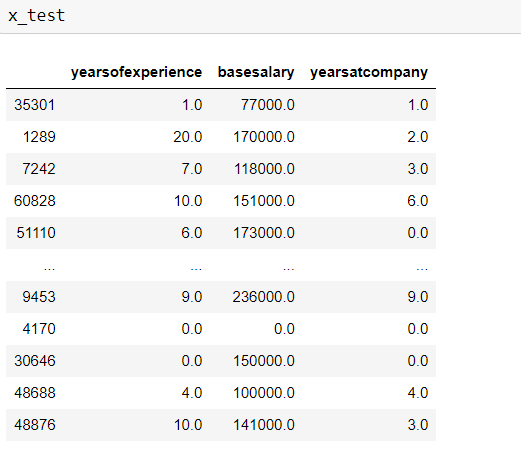


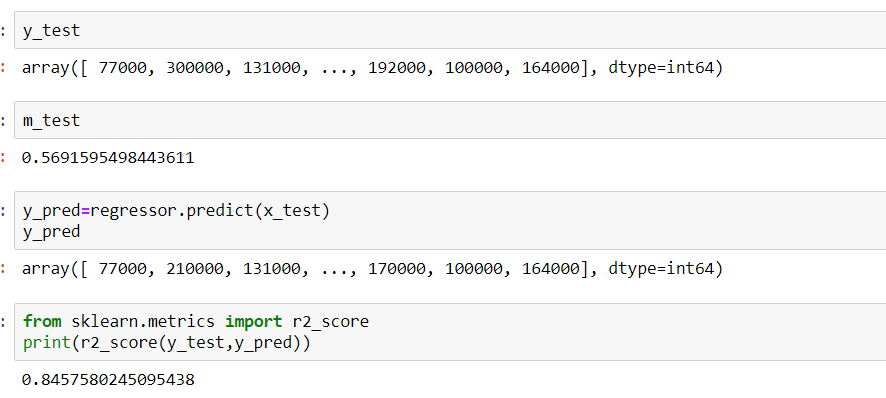
**Model Evaluation:(using DecisionTreeClassifier)**



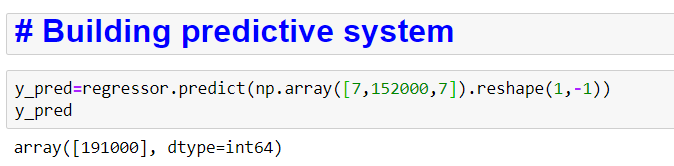








**Building a Predictive system:**



Advantages of total yearly compensation:

According to this model the organization is easily predict the total yearly compensation of a particular employee.

Total Yearly Compensation stimulates employees to work harder,thus increasing productivity and enhancing job performance.

Github link:

<https://github.com/monikaganga/EmployeesalaryPrediction.git>

Name:K.monika ganga

Class:cse-3c

ID NO:S180148