**Advertising Sales Channel Prediction**



**Session-2021-2022**

SUBMITTED TO SUBMITTED BY

Data Trained Education YATENDRA JHA

Batch -1835

**Table content**

* Introduction
* Problem Definition
* EDA
* Data analysis
* Building Machine Learning Models
* Concluding Remarks

Introduction:

In this project, I have built and evaluated multiple regression models using Python. I have used scikit-learn to calculate the regression, while using pandas for data manipulation and seaborn for data Visualization. The dataset for this project consists of the very popular Advertising Dataset to predict sales revenue based on advertising spending through media such as TV, radio and newspaper.

I got the data set for UCI Machine learning Repository:

<https://archive.ics.uci.edu/ml/datasets/TV+News+Channel+Commercial+Detection+Dataset>

Problem Statement:

When a company enters a market, the distribution strategy and channel it uses are keys to its success in the market, as well as market know-how and customer knowledge and understanding. Because an effective distribution strategy under efficient supply-chain management opens doors for attaining competitive advantage and strong brand equity in the market, it is a component of the marketing mix that cannot be ignored.

The distribution strategy and the channel design have to be right the first time. The case study of Sales channel includes the detailed study of TV, radio and newspaper channel. The predict the total sales generated from all the sales channel

Attribute Information:

1-TV (Sales Generated by advertising on TV)

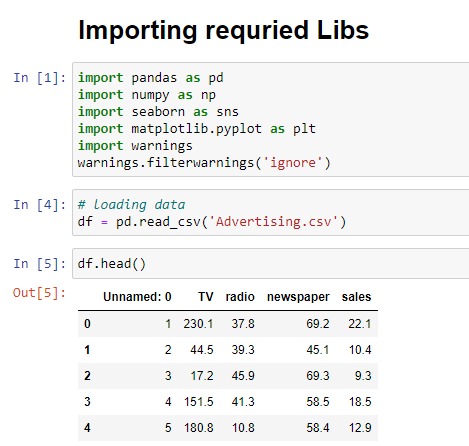
2-Radio (Sales Generated by advertising on Radio)

3-Newspaper (Sales Generated by advertising on Newspaper)

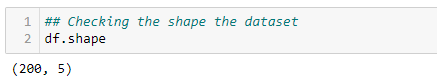
4-Sales (Total Sales)

Importing our Dataset:

First, we will import required library to import our dataset.

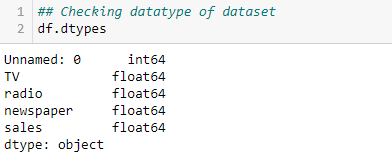


Now we will check shape and Data type of our attributes,



1. We have 200 rows and 5 attributes including our Target Attribute.
2. Sales in our target Attribute in our Dataset.

Checking data Type,



This indicate that in our dataset we have all the attribute of float and int data type.

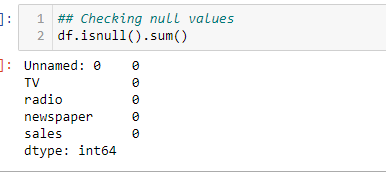
When our target attribute has float data type, it mean that that it is a Regression problem statement.

Let’s Start Our EDA:

Generally, in EDA we try to know more about our dataset. During the dataset we check do data wrangling, checking data type, checking null values, data cleaning and dropping unnecessary columns which have no significance to our Machine Learning Model.

**Checking null values**:

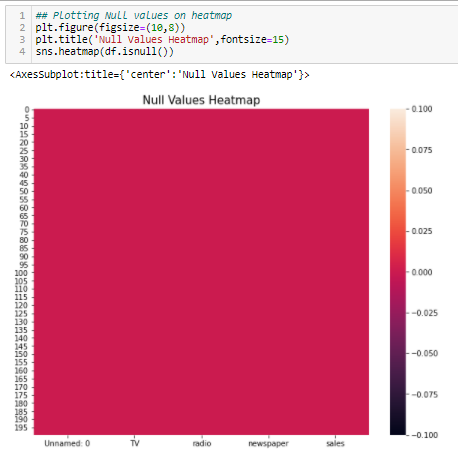
We can check null values for this code



We could see in our dataset there is no missing values.

We need to be sure, In some times it could be possible that in our dataset some erroneous data like (?,Nan,@,--) could be present. So we always check and remove before moving forward. In our dataset we have no such values so we can move further.

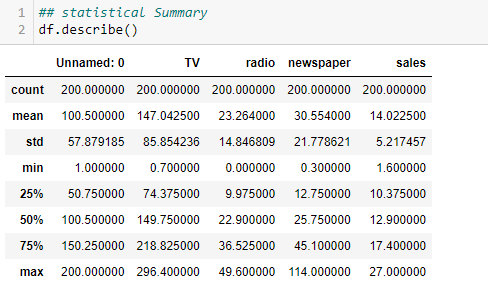
**Visualization of Null values in Heatmap:**



We could see color uniformity in our heatmap which means no missing values. If there were some missing values present in our dataset we could see some horizontal lines (based on number of missing values) in this heatmap but fortunately there is no missing values.

**Statistical Summary:**

Statistical Summary is very import information in order to know about distribution/skewness/outliers present in our dataset. Statistical summary generates information only for integer/float datatype.



When we get above table, we need to observe things very carefully. Let’s explore more about this dataset.

1-Unnamed:0--> We can see that count value and max values have similar number it means it is just a counting number/serial number which generally has no significance in prediction so we can drop it in our further process.

2-TV- We can compare the mean value with median values (50%). We know when mean=median=mode, our data distribution Normally Distribution. Here approximately both values are near to each other means our data is normally distributed and there are no outliers.

3-Radio- Minimum values of Radio Advertisement is zero. But Median and mean values are nearby to each other hence data is normally distributed and no outliers are present in our dataset.

4-Newspaer- In newspaper column we could see that mean>median value which indicates that our data distribution is skewed and when we compare the 3rd quantile(75%) with maximum value we could see there is a big difference which indicate that outliers are present in our dataset.

5- Sales- This attribute is approximately normally distributed.

**Data Visualization** **& Analysis**:

Here we will do univariate and bi-variate analysis with the help of data visualization methods.

Univariate Analysis-**Univariate** is a term commonly used in **statistics** to describe a type of data which consists of observations on only a single characteristic or attribute .

Bi-variate Analysis- **Bivariate analysis** is one of the simplest forms of quantitative (statistical) **analysis**. It involves the **analysis** of two variables (often denoted as X, Y), for the purpose of determining the empirical relationship between them

**Univariate Analysis:**

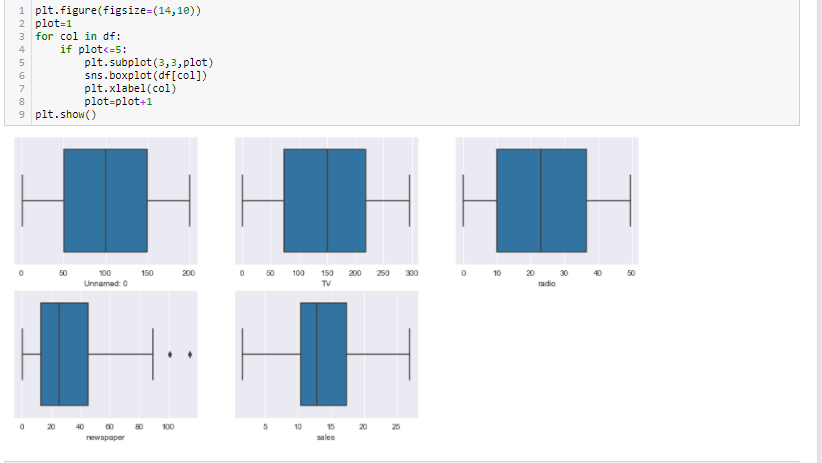
In this we will data visualization we will see the distribution of our numeric attribute.



1. It shows that our dataset is approximately bell shaped in distribution means normally distributed.2- Newspaper attribute is a little right skewed.3- our target attribute is Normally distributed.

**Outliers Detection using BOX-PLOT:**

This is also a very important visualization method to detect the outliers present in our dataset. There are various other methods also present like scatterplot, z-score,IQR,PCA to detect outliers. Here we will use Box Plot.



We could see, there is some outliers present in newspaper attribute. Apart from news paper there is no outliers present on our dataset.

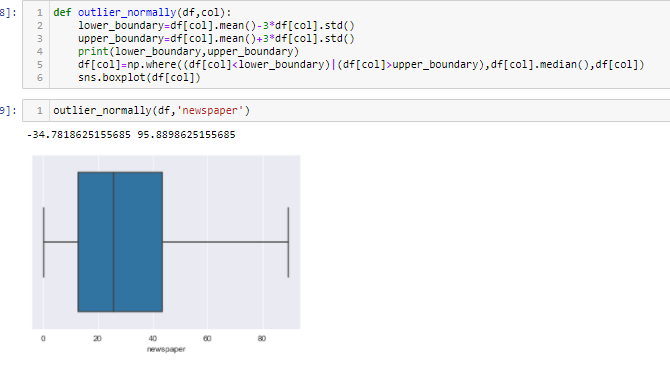
**Outlier Removal:**

After detecting outliers, it is very important to treat those outliers by taking proper action. Here I will share two method.

IQR Method—When our data distribution is skewed (right skewed/left skewed) we will use IQR method to remove the outliers.

Z-score Method—We apply Z-score method when our data distribution is normally distributed. When z score is more than 3, we consider those points as an outlier.

We can either drop the outliers’ rows or change it with upper limit or we can change outliers with median values, all these things depend upon domain knowledge and problem statement.

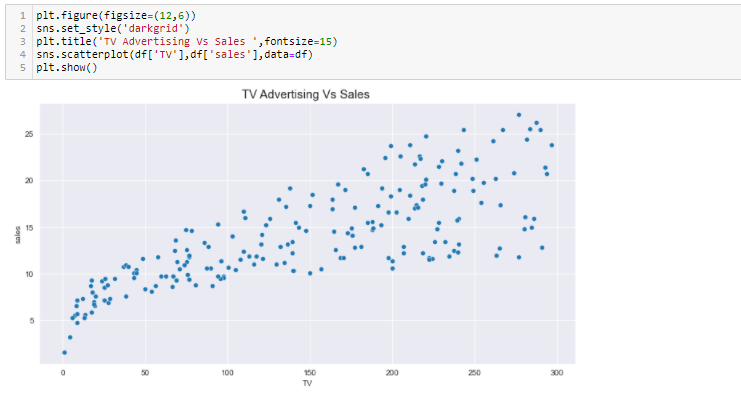


In above firstly we define a function and in that we use z-score method . In our dataset instead of removing outliers we change the outliers with the median values. We should always avoid the practice of dropping / removing the data because data is very expensive. With the help of that we could see that our outliers successfully removed.

**Bi-Variate Analysis**-

In this we will plot graphs between two attribute and try to find some meaning full insights out from it.

Tv Vs Sale (Scatter Plot)



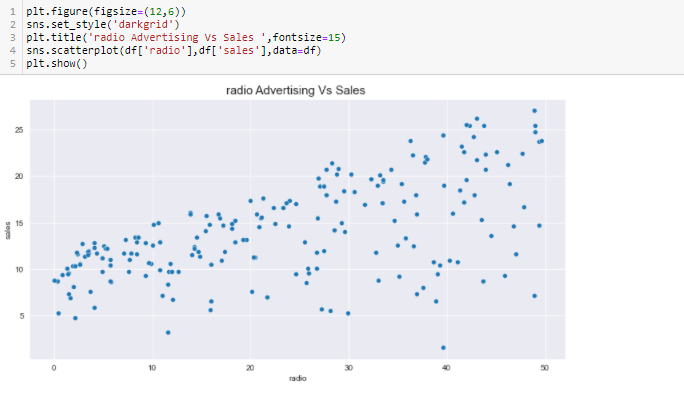
1- There is a linear positive correlation between TV advertising and Sales.

2- As Tv advertising increases sales also increases.

Radio Vs Sales (Scatter Plot)

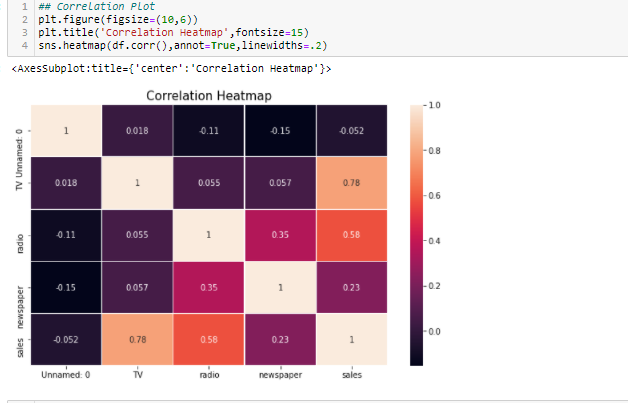
1- This plot shows a positive linear relation between radio advertising and sales.

2- As radio advertising increases, sales also get increases



**Correlation Heat-Map:**

Correlation plot: **Correlation plots** can be used to quickly find insights. It is used to investigate the dependence between multiple variables at the same time and to highlight the most **correlated** variables in a data table. In this visual, **correlation** coefficients are colored according to the value.

****

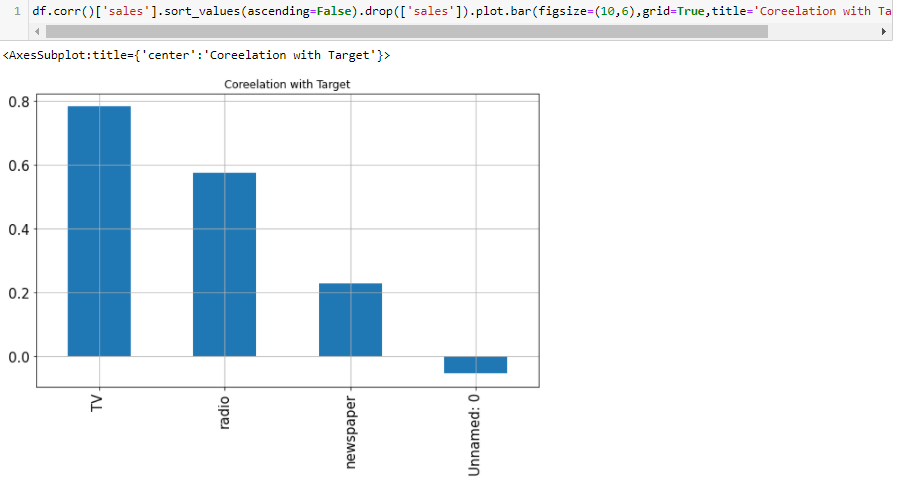
1-Unnamed:0 have least and negative correlation with our target attribute.

2- TV and radio has a strong and positive correlation with target attribute.

3-Newspaper and radio are correlated to each other.

**Correlation with target Attribute:**

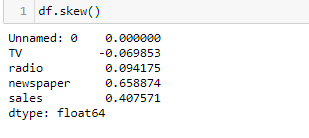
Here we will check our correlation with target attribute and plot it on a bar plot.

****

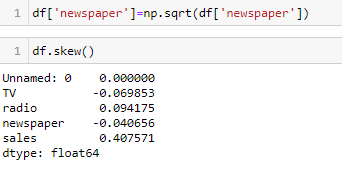
We could see that TV is highly correlated among all. Unnamed has negative and nearly zero correlation with our target attribute. Correlation help us to determine which attribute has highest/least impact on our target attribute.

**Skewness:**

**Skewness** refers to a distortion or asymmetry that deviates from the symmetrical bell curve, or normal distribution, in a set of data. If the curve is shifted to the left or to the right, it is said to be skewed.

****

We should keep in our mind that skewness should always between [-0.55,+.055]. This is a practical value. If skewness shows a too high or too low value, then we apply some transformation methods like square root, cube root, log transformation ,reciprocal transformation etc. Here in our dataset newspaper skewness is more that 0.55 so will apply some transformation method.



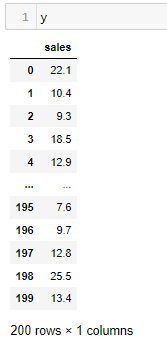
We use square root transformation method and skewness has been removed. We need to do some try and error method in order to get the right transformation method.

**Model Building:**

Before moving to build our final model, we need to split our data in dependent and independent variable. In this we will split our data in X and Y where X will be independent variable Y will be dependent variable.







**Feature Scaling:**

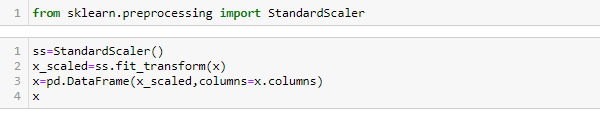
**Feature Scaling** is a technique to standardize the independent **features** present in the data in a fixed range. ... If **feature scaling** is not done, then a **machine learning** algorithm tends to weigh greater values, higher and consider smaller values as the lower values, regardless of the unit of the values.

Standard Scaling:

Standardization is another **scaling** technique where the values are centered around the mean with a unit **standard** deviation. This means that the mean of the attribute becomes zero and the resultant distribution has a unit **standard** deviation.

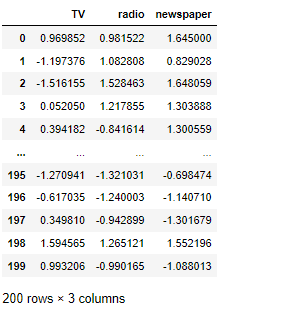
Why scaling?

It is a step of data Pre-Processing which is applied to independent variables to normalize the data within a particular range. It also helps in speeding up the calculations in an algorithm.

****

Scaled Data:

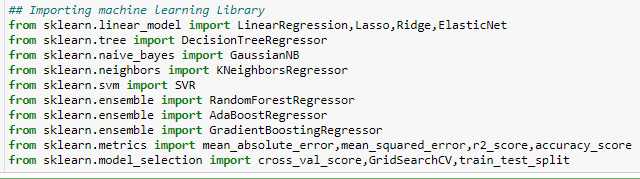
Now our dependent values are in same scale and it ranges between -1 to 1 variance.



**Model Building:**

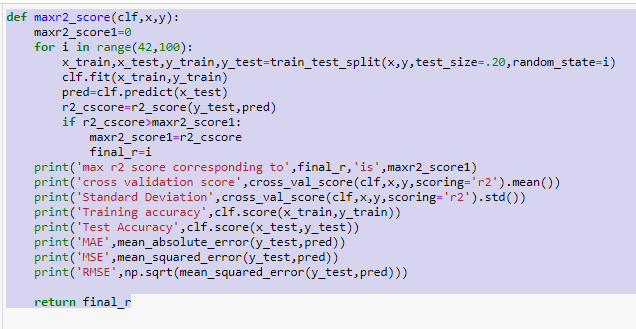
Finally, I standardize my data, Now is going to model creation.

Now I will import some important some machine learning library.



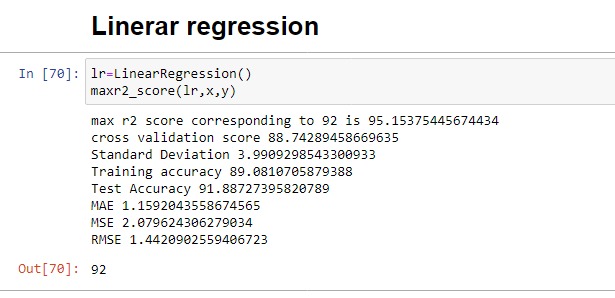
Mean\_absolute\_error, mean\_squred\_error,r2\_score are our performance matrix.

Here am making a function for finding the best random state for better r\_2score in model.



Now we will apply some machine learning.

1-Linerar Model



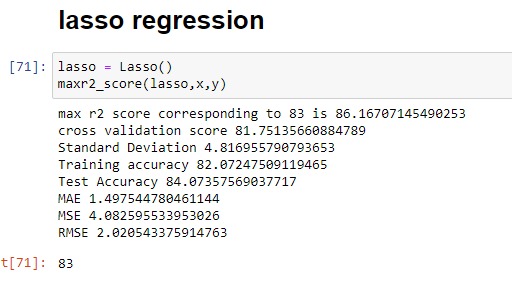
We could see that at random state 92 we get highest r\_2 score 95.15.

Our training and test score of linear regression is .89 & .91.

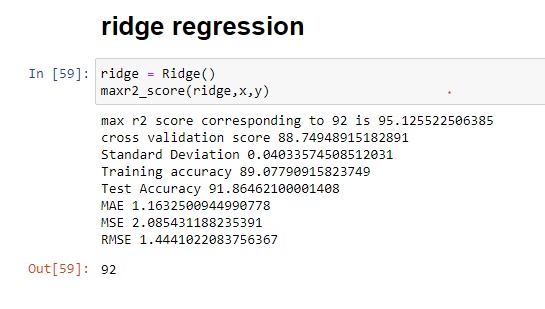
But we can’t choose merely based on max r\_2 score. There could be chances that our model would be under fit or over fit.

So we will check some other models also.

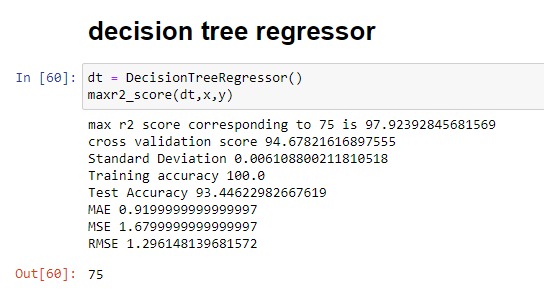
Lasso Regression:



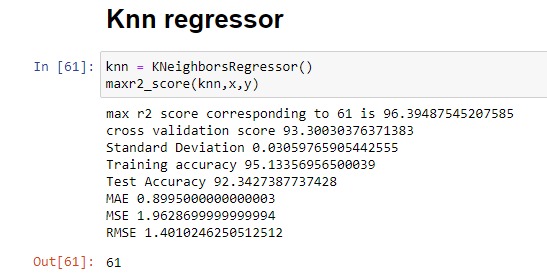
Ridge Regression:



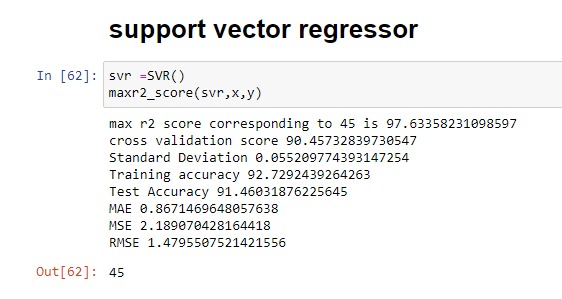
Decision Tree:



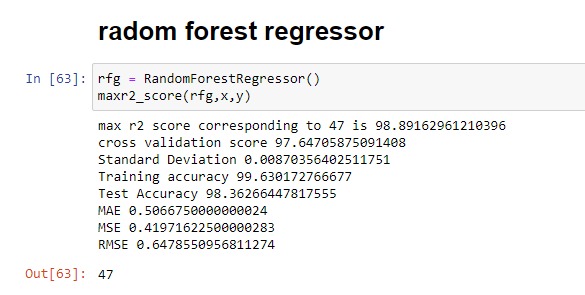
KNN:



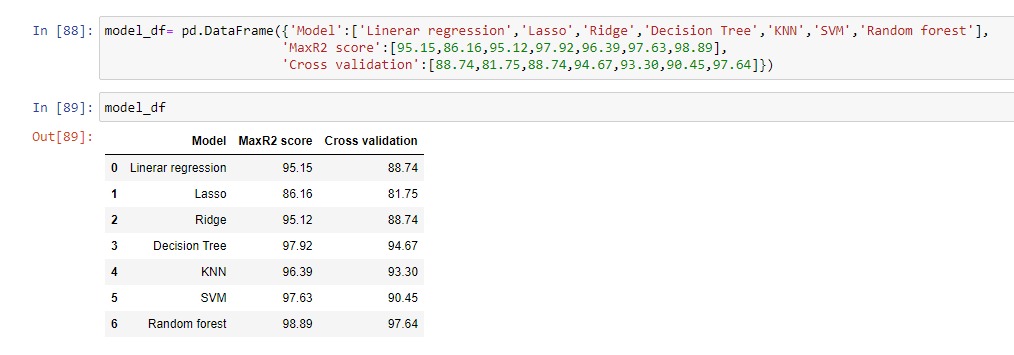
Support Vector Machine:



Random Forest:



Creating Data Frame of all of our machine learning model:



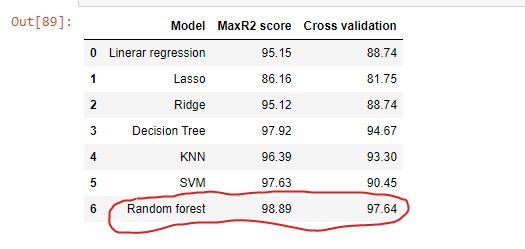
**How to choose our Best Model??**

We can’t choose any model based on height R\_2 score. Why? Because it could be possible that model is overfit which mean giving a very good performance with respect to our Training data set. So how to counter that thing how choose the best model out of all?

Cross validation:

**Cross**-**validation** is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold **cross**-**validation**.

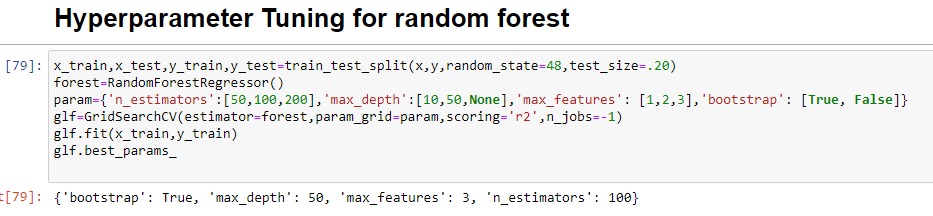
Over Fit Checking:

If the model is the least difference with mean cross validation score and r\_2 score at random state then we consider that model as a best model.

We could see that our random forest model givers us a most generalized model. The difference between R\_2 score and cross validation score is minimum is this case.

Hypermeter Tuning:

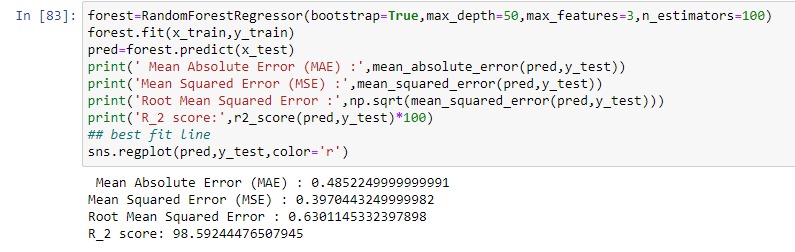
**Hyperparameters** are **crucial** as they control the overall behavior of a machine learning model. The ultimate goal is to find an optimal combination of **hyperparameters** that minimizes a predefined loss function to give better results.



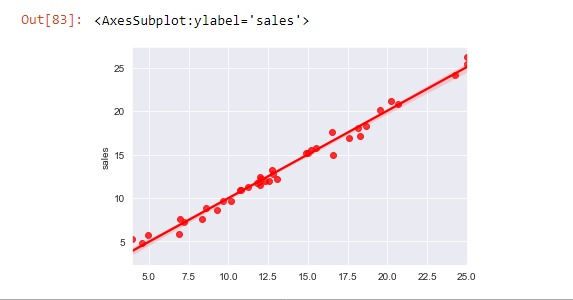
Applying best parameter:

In order to improve our overall result we do some hyperparameter tuning with the help of Grid Search CV. In this we put some values and automatically based on our

Scoring parameter the Grid Search CV gives us the best parameter.



We could see that by applying Hyperparameter tuning our R\_2 score has improved.



Saving Our Model:



**Conclusion:**

so, we got the final conclusion that with the Random Forest Model our model is performing very well we are going to finalise the model.

Thanks

