AUTHOR NAME: ARJUN MANCHANDA

Blog On Advertisement Sales Prediction Using Machine Learning

Let us understand first why Sales prediction is a need of an Hour

What is Sales Forecasting?

Sales Forecasting is the process of estimating future revenue by predicting the amount of product or services a sales unit (which can be an individual salesperson, a sales team, or a company) will sell in the next week, month, quarter, or year.

At its simplest, a sales forecast is a projected measure of how a market will respond to a company's go-to-market efforts.

Why is Sales Forecasting important?

Forecasts are about the future. It's hard to overstate how important it is for a company to produce an accurate sales forecast. Privately-held companies gain confidence in their business when leaders are able to trust forecasts. For publicly-traded companies, accurate forecasts confer credibility in the market.

Sales forecasting adds value across an organization. Finance, for example, relies on forecasts to develop budgets for capacity plans and hiring. Production uses sales forecasts to plan their cycles. Forecasts help sales ops with territory and quota planning, supply chain with material purchases and production capacity, and sales strategy with channel and partner strategies.

These are only a few examples. Unfortunately, at many companies, these methodologies stay disconnected, which can produce adverse business outcomes. If information from a sales forecast isn't shared, for example, product marketing may create demand plans that don't align with sales quotas or sales attainment levels. This leaves a company with too much inventory, or too little inventory, or inaccurate sales targets—all mistakes that hurt the bottom line. Committing to regular, quality sales forecasting can help avoid such expensive mistakes.

Introduction on Model

Here we are going to do the complete analysis of the Advertising sales prediction model. We will cover all the aspects that we are going to use in ML model and projects and we will also do the complete analysis using the Data visualization to model building and finding the main observations from the analysis which is going to help us a lot for prediction of the best results.

Steps followed during Analysis:

1. Problem Definition: -

- The problem is that whenever any company enters the market, the distribution strategy and channel it uses are keys to success in the market
- The market also knows the knowledge and understanding of the customer
- Effective distribution strategy under supply chain management open doors for attaining competitive advantage and also strong brand equity in the market
- It is a component mix of market which cannot be ignored
- The distribution strategy and channel design have to be right at the first time
- This case study of channels includes the detailed study of the TV, Radio, Newspapers.
- It also predicts the total sale generated from all the sides of the channel.

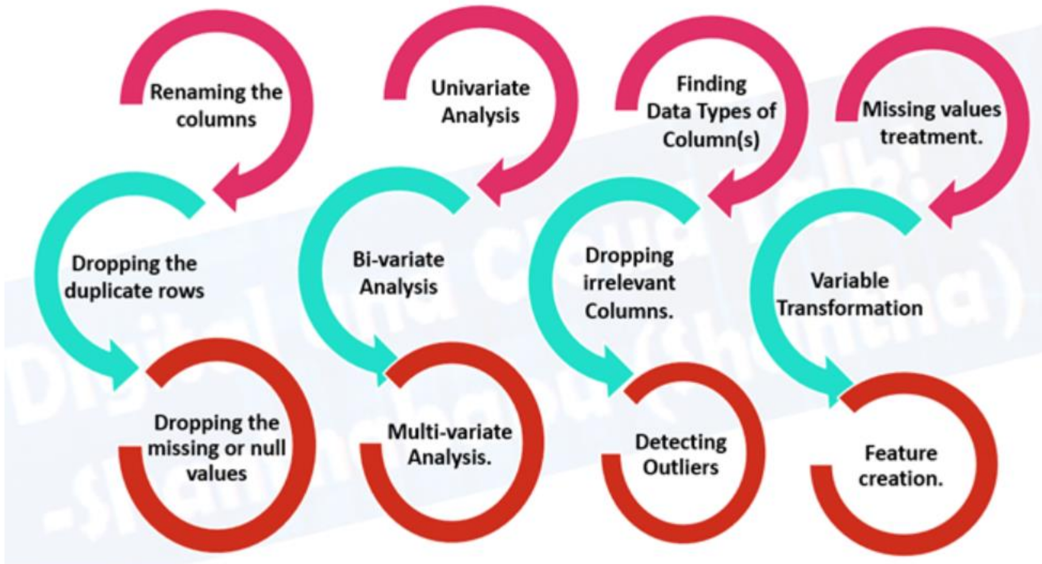
2. Data Analysis: -

Here data set has been provided below and I am sharing the picture of the data set which I am going to use for further analysis

	Unnamed: 0	TV	radio	newspaper	sales
0	1	230.1	37.8	69.2	22.1
1	2	44.5	39.3	45.1	10.4
2	3	17.2	45.9	69.3	9.3
3	4	151.5	41.3	58.5	18.5
4	5	180.8	10.8	58.4	12.9
...	...	...	...	...	...
195	196	38.2	3.7	13.8	7.6
196	197	94.2	4.9	8.1	9.7
197	198	177.0	9.3	6.4	12.8
198	199	283.6	42.0	66.2	25.5
199	200	232.1	8.6	8.7	13.4

200 rows × 5 columns

- We can see that we have 200 rows and 5 columns in the data set.
- We have 5 columns Unnamed :0, TV, radio, newspaper and sales.
- We check the properties like shape, unique, data types etc.
- We have sales column as our target and is continuous in nature, thus it is a Linear Regression problem.
- Now I will do the further analysis according to our problem type, which is linear regression and is a type of Supervised learning.



3. EDA Concluding Remarks: -

The following steps are to be followed to do complete EDA of the data set

- Check missing values
- Statistical Summary
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- To check Skewness

Now we can see whether there are null values present in the data set using is null function

```
Out[8]: Unnamed: 0    0.0
        TV            0.0
        radio         0.0
        newspaper     0.0
        sales         0.0
        dtype: float64
```

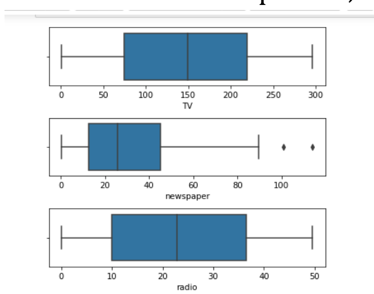
As we can see there is no any null value present in the data set hence the data is clean

- Statistical summary gives information about the mean, median, std, min, max etc.

Main Observations: -

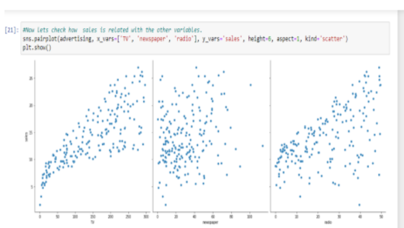
The difference in mean and median is almost similar. There is small difference in 75% percentile and max in columns named Unnamed: 0, TV and radio which shows that no outliers are present in it. we can see the difference in 75% percentile and max in newspaper column which shows that few outliers are present in it.

- For univariate analysis I have plotted boxplots, from these plots we can see the mean, median, max, min and we can also see whether outliers are present or not, as we can see in below picture, we can see that outliers are present in newspaper.

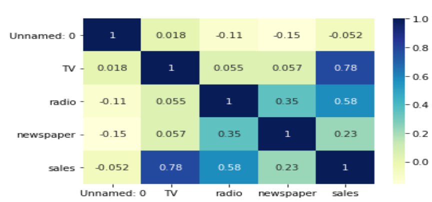


The above box plot shows that there are no considerable outliers present in the data set

- In Bivariate analysis I have used Scatter plot to see the relation of each column with the sales column we can see the scatter plot of TV column with the sales. Here we can see the positive relation between the sales and TV, as the TV advertisement increases sales also increases.

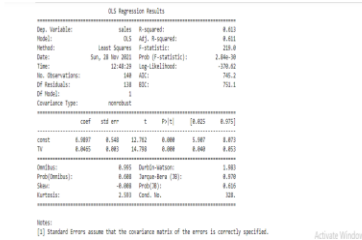


Now let us check the **co-relation** between with the other factors so I have plotted the heatmap



As from the above heatmap and pair plot we can clearly see that the tv is much more related to the sales hence here we will consider tv as our feature variable and let us do the simple **linear regression.**

Now Performing a summary operation lists out of all the different parameters of the regression line fitted



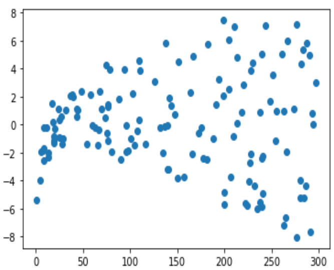
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  sales   R-squared:                       0.613
Model:                            OLS   Adj. R-squared:                  0.611
Method:                 Least Squares   F-statistic:                     219.0
Date:                Sun, 28 Nov 2021   Prob (F-statistic):           2.84e-30
Time:                        12:48:29   Log-Likelihood:                -370.62
No. Observations:                 140   AIC:                             745.2
Df Residuals:                     138   BIC:                             751.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          6.9897      0.548     12.762      0.000       5.907       8.073
TV             0.0465      0.003     14.798      0.000       0.040       0.053
==============================================================================
Omnibus:                        0.995   Durbin-Watson:                   1.983
Prob(Omnibus):                  0.608   Jarque-Bera (JB):                0.970
Skew:                          -0.008   Prob(JB):                        0.616
Kurtosis:                       2.583   Cond. No.                         328.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

The above is the OLS regression results

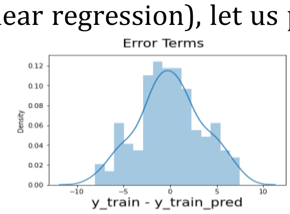
I have also plotted a **Scatterplot**

In [35]:
```
#Patterns in the residuals
plt.scatter(X_train,res)
plt.show()
```

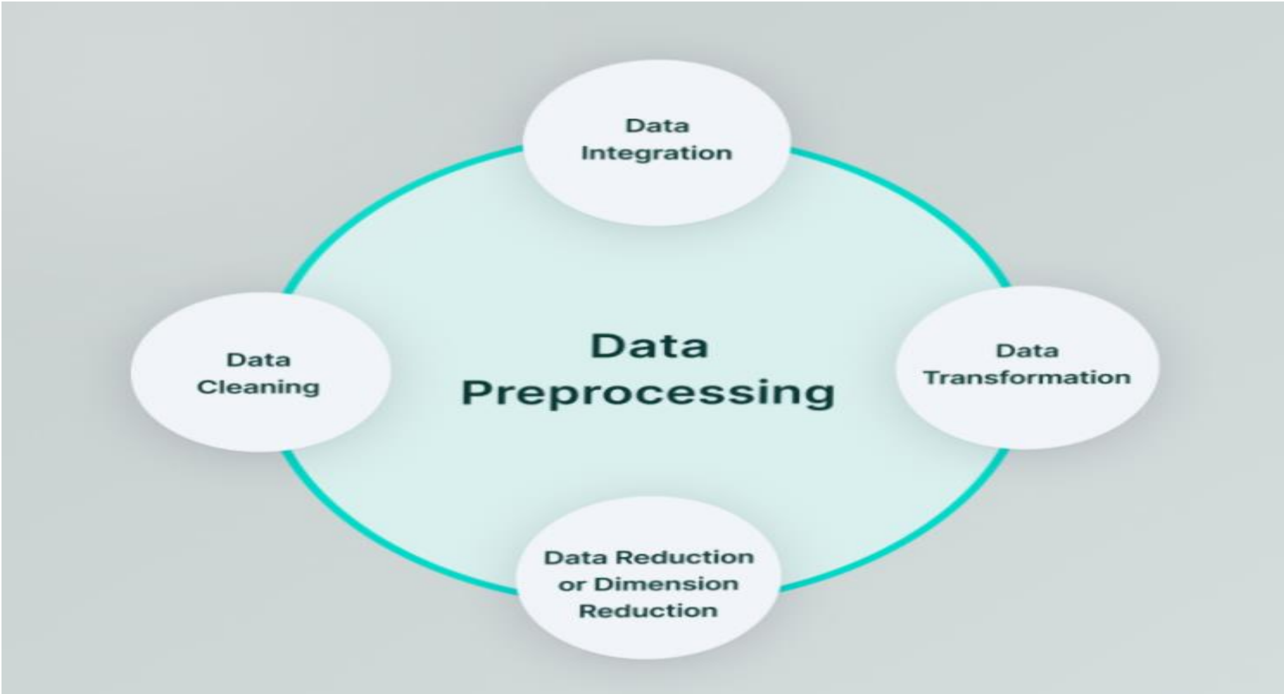


Model Evaluation: -

To validate assumptions of the model, and hence the reliability for inference we need to check if the error terms are also normally distributed (which is in fact, one of the major assumptions of linear regression), let us plot the histogram of the error terms and see what it looks like



From the above graph we can clearly see that the residuals are normally distributed with mean zero. Variance of residuals increasing with X indicates that there is significant variation that this model fit isn't by chance, and has decent predictive power. The normality of residual terms allows some inference on the coefficients.



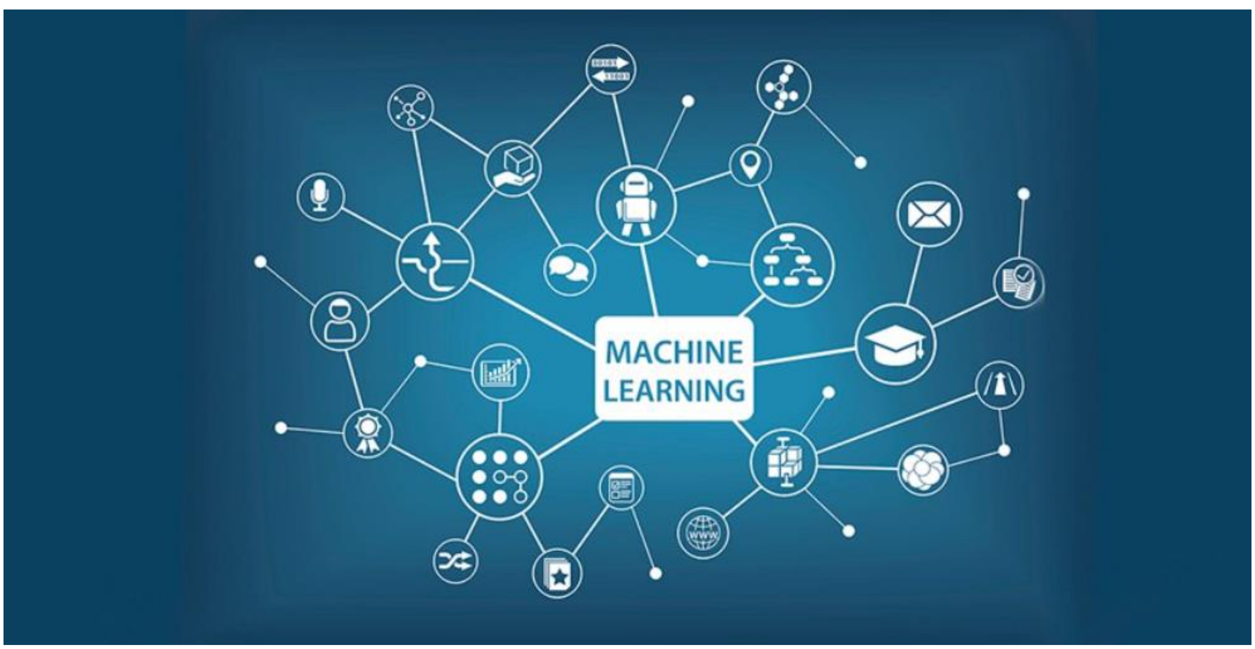
4. Pre-processing pipeline: -

We have many steps included in pre- processing like Data cleaning, Data reduction, Data integration etc. Let us discuss each of them in detail the steps we have done in our project according to our requirement.

- We have drooped the rows which are negatively correlated, here we have dropped the Unnamed: 0 column from the dataset due to its highly negative correlation with the sales column. This column impacts our data negatively, thus we dropped that column.
- Now we are removing the outliers present in our data, we have two methods to remove outliers one is **Zscore** and other is by using IQR method here I have used **Zscore** method to remove all the outliers present in our dataset, I have very less outliers in my dataset thus only two rows are deleted from the dataset as outliers.
- Here we have seen above that no null values are present in our dataset, so here is no need to handle missing values, if there were any missing data then we have to treat it with suitable method, but here no need.
- Next is the we check whether any column is present in string format or not, if any column is present in string format, then we have to change it in integer format by applying Encoding technique, we have two methods in encoding one is One hot encoding and other is Label Encoding, but here in our dataset we have no need to apply encoding technique because all the columns are already present in integer format. So, let's move on to the next step.
- So now we are removing skewness from our dataset, as we have seen that skewness is present in all the columns except the sales column, to remove the skewness we have separated the target variable and the independent variable from the dataset. As we know that

skewness between –0.5 to 0.5 is acceptable but more than it is not, so skewness except this range should be treated by using suitable method. So, here I have used cube Root method to handle negative skewness and square root method to handle positive skewness, till now we have treated the skewness by suitable method and removed skewness from our dataset.

- Next point which comes is the feature engineering but here in this dataset we do not need it, because we don't have special characters, etc. So here we have no need to apply Feature Engineering at all.
- The last thing which I can see is the standardization technique, we use this technique to scale our data. We have two methods to scale our data first one is standard scaler and the second one is min max scaler. We use these techniques only when there is huge difference between the ranges of any 2 columns, that's why we use scaling. Standard scaler is used when data is normally distributed, its changes the mean=0, std=1 and the value ranges between –3 to +3.Min –Max scaler is used when data is not normally distributed this method is also called normalization, it changes the data with mean=0, std=1 but range is 0 to 1.In our dataset we have seen there is no huge difference in ranges of the columns, thus here we have no need to apply standardization technique on our dataset.
- The last which I can see is the PCA technique, this technique is used only when we have large numbers of columns and it difficult to manage them all, but here in our dataset we have only 5 columns thus here we have no need to use this technique.



5.Building Machine Learning Models: -

Now we build a machine learning model, we will use multiple algorithms, as we know we are working on a regression problem so here we will only use regression models like :-

- *Linear Regression*
- *SVR*
- *DecisionTreeRegressor*
- *KNeighborsRegressor*
- *Lasso Regression*
- *Ridge Regression*

To use all of the above, we have to import each model from library scikit learn as follows:

```
from sklearn.metrics import r2_score
from sklearn.linear_model import Lasso,Ridge
```

```
from sklearn.svm import SVR
from sklearn.tree import DecisionTreeRegressor
from sklearn.neighbors import KNeighborsRegressor
```

Firstly, we have to create train test split and thus we use train data for training our model and test data for testing our model performance. Here I have split the data in 70% as train and 30% as test data. I have found the best random state using the regression model which is 52, I used this random state to train all the models. Now we use train data for training our model and test data for testing our model performance. Since we are using regression model, we different evaluation matrix like mean absolute error, mean squared error and Root mean squared error for all the models The error in the models shows the performance of the model if error is least the model is performing well, but if the error is more model is not performing good. For our dataset we have checked the error and founded that the least error is coming from the Decision Tree Regressor. So now let's use evaluation matrix r2_score to see the score for all the models, after checking the r2_score I have found that I am getting maximum r2_score with **DecisionTreeRegressor** is 95%.The picture which comes around us is the cross-validation technique, as we know the score is also due to overfitting, thus we use cross validation method to come over it Cross-Validation is a statistical method of evaluating and comparing learning algorithms by dividing data into two segments: one used to learn or train a model and the other used to validate the model. The basic form of cross-validation is k-fold cross-validation. Here in my model, when I checked the cross validation, I use cv=5 folds, the best cross-validation score is coming out to be for Decision tree regressor. Now I have used hyper parameter tunning to find the best parameter for our model by using GridsearchCV. I have applied gridsearchcv on all the models and found the best parameter for all the models and used these parameters in our model, now I am getting the best accuracy with decision tree regressor with(parameter=mse).

Now we can conclude that decision tree regressor is the best model for our dataset, as we have seen we are getting best r2_score with that model, least error, least difference in r2_score and cross validation score, so it is our best model for this project. So last step is to save our model to use it in future for predictions, we have two techniques for saving the model. First is using joblib and the second is by using pickle. I have saved my model DTR using joblib, so that I can use it in future and predict the sales using this model for future use. The last thing we have to do is concluding remarks:

6) Conclusion Remarks: -

The conclusion in below points:

- The main goal of our project is to solve the problem and predict the sales. For this we have used machine learning skills and solved the problem.
- We have done the complete analysis of the data using EDA, univariate, bivariate, multivariate, checking correlation, checking skewness, checking for outliers, checking for missing values by doing all this analysis.
- We have collected the information about the data, whether it is skewed, having missing values or not etc. Next, I have done the pre-processing of the data and solved all the issues that we found during EDA like Outliers, skewness etc.
- At last model building I used regression algorithm and different evaluation matrix to prepare the models and found decision tree regressor as my best model and at last, we can make predictions for sales channel using our model.