Sales Channel Prediction

-Sumit Dhandhania

➢ Prelude:

In any firm sales form, the bread and butter to an organization. Sales is one of the factors and quite crucial in determining the survival and prosperity of an organization. So, continuing our discussion, this project is regarding the marketing mix an organization uses to boost sales. For this task, one analyst has to determine the perfect ratio of media channels engaged by the firm to maximise the sales by allocating optimum budget to each medium of communication.

➢ About the Independent Variables:

We are given three continuous variables which are as follows:

1. TV: This variable talks about expenditure in Television advertisements
2. Radio: This variable talks about expenditure in Radio advertisements
3. Newspaper: This variable talks about expenditure in Newspaper advertisements

➢ About the Dependent Variable:

Sales: This continuous variable talks about the sales generated by employing the above independent variables.

➢ Purview:

Planning the budget allocation for marketing mix has many decision-making importance. For instance, an ideal budget allocated to marketing mix of media channels for the purpose of advertisement not only boosts sales but also helps in saving excess money which otherwise would have been wasted on advertising by not generating more revenue. Moreover, the saved penny could be used either for other function of the business or could have been used in future.

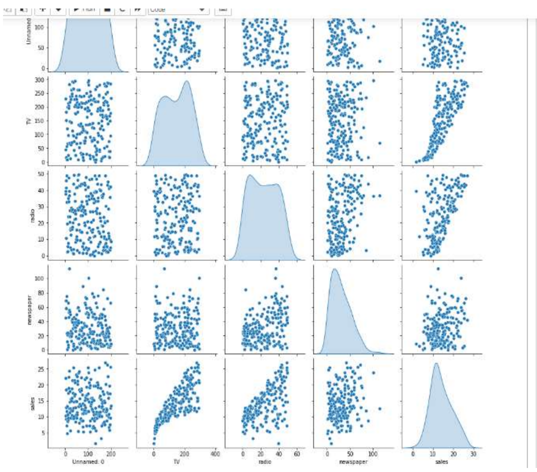
➢ Problem Definition:

We have to build a best-fit prediction model for allocation of budget in media channels to gain optimum sales. This model will help us to understand which marketing medium is influencing the sales of the firm and in what manner.

➢ Data Analysis:

1. We need to import the necessary libraries to upload the data in python environment, analysing, data visualization and even for further processing.
2. After importing the libraries, we will have to upload the data.
3. Primarily we will check the fundamentals of the data uploaded. This includes, schema, null values, character, top and bottom few rows of the data.

4. There is no missing value in the data
5. For this data we see that there are only 5 columns and 200 rows
6. All 5 columns are having numeric data-type.
7. Moving on, we will now learn about the descriptive statistics.
8. The data follows a normal distribution
9. The first column- “unnamed:0” is having serial numbers and not important to our model building.
10. Average expenditure on TV ads is 147.04 and average sales generated is 12089.37
11. Average expenditure on Radio ads is 23.26 and average sales generated is 3021.92
12. Average expenditure on Newspaper ads is 30.55 and average sales generated is 3874.63
13. Average sales generated because of media channels is 14.02
14. Newspaper has outliers
15. There is no issue of multi-collinearity in the data
16. Cumulative effect of spending on media results in increasing sales

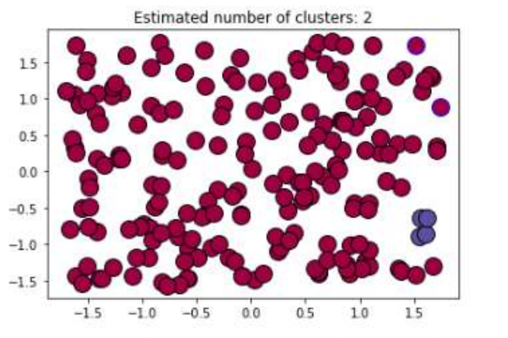


➢ Exploratory Data Analysis (EDA) Concluding Remarks:

1. Since, all the variables are continuous in nature, we use correlation to test multicollinearity and significance.
2. Here, we find column “Unnamed:0” to be insignificant. Therefore, we drop the column from our dataset.
3. The data is positively skewed.
4. TV and radio show a positive correlation. With increase in expense on these variables sales also increases.
5. As we find an outlier or rare occurrence value in the column newspaper, we will have to treat the value to align with the data.
6. This treatment will help the machine to learn the pattern and predict values for the coefficient

➢ Pre-Processing Pipeline:

1. Let's start by dropping the "Unnamed:0" column
2. Let's proceed with treating the outlier in the column newspaper. We can do that by either toning the value to either 1st quartile or the 3rd quartile of the column variable or applying the z-score. I will choose the former process.
3. Since we do not have missing value issue, we don't do anything in that regard. In case, we had that problem we could drop the values, fill the values using "fillna" or apply a strategy depending upon the situation.
4. To check the noise or outliers in the data we can even resort to DB Scan for better visualization.



5. We have to treat the variables for skewness as the data is not following a normal distribution. I choose either "box-cox" or "Yeo-Johnson" methods depending upon the situation. For this situation, I am going ahead with "Yeo-Johnson" method. Box-Cox is only for the positive values and in some manner inferior to "Yeo-Johnson". For this, we have to drop the target variable "Sales".

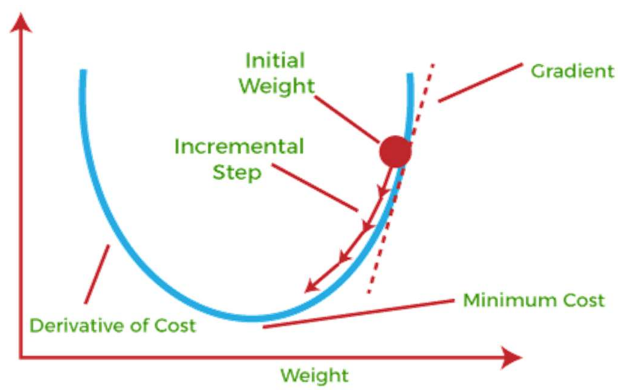
```
# Removing skewness using Yeo-Johnson
# Removing data skewness
pt = PowerTransformer(method='yeo-johnson',standardize='True')

data_noskew= pt.fit_transform(data_features2.values)
data_noskewtab= pd.DataFrame(data_noskew)
data_noskewtab.columns = ['TV', 'radio', 'newspaper']

print(data_noskewtab)
```

6. Now, we split the data into train and test taking test size to be 30% of the data.
7. After splitting we scale the data. For scaling the data, I check different strategies to adopt the scaling of the data before applying it. As each situation is different and different methods fit better in certain situations. In this case, I have selected robust scaling technique as it had scored maximum.
8. We can also do Principal Component Analysis to check the interrelations among the independent variables. Since our dataset is small, this step is not so important.

9. We can also use polynomials to increase the accuracy of the model. But let's follow principle of Occam's Razor and keep our model simplified.

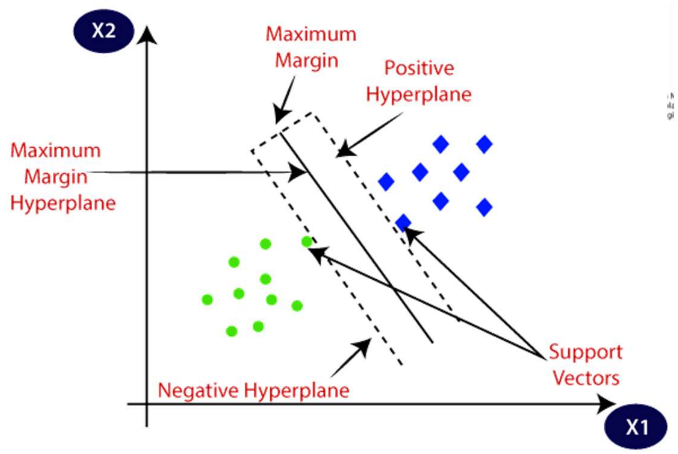
➢ Building Machine Learning Models

1. Firstly, I will be trying my hands on Linear Regression Model. The Linear Regression Model did a decent job but our accuracy can be better. The Mean Absolute Error (Mean Absolute Error) is 14.4 and Root Mean Square Error (Root Mean Square Error) is 14.76. However, R-square and Adj R-square are 0.038 and 0.024 respectively. This means that independent variables are not being explained by the dependent variable well.
2. Moving to our second model which is Lasso or L1. This model also belongs to the family of Linear Regression Model. Lasso performed better than Linear as the errors reduced. Mean Absolute Error= 1.9 and Root Mean Square Error=2.61. We now do cross validation to check the performance of our model. After cross validation we see some room of improvement for our model. Therefore, I now do hyperparameter tuning to improve the model. The new Mean Absolute Error= 1.35 and Root Mean Square Error= 1.87.



3. Further, I try my luck with Ridge Regression (L2) Model. It is another member from the family of Linear Regression Prediction Model. It performed better than earlier model. The Mean Absolute Error = 1.36 and Root Mean Square Error= 1.89. Similarly, even for Ridge Regression we do cross validation and hyperparameter tuning to improve our model. The new Mean Absolute Error = 1.87 and Root Mean Square Error= 2.21
4. The next model belonging the family of Linear Regression is Elastic Net (L1.5). This model is combination of the two models above. Mean Absolute Error=2.51 and Root Mean Square Error=3.36. Next, we do cross validation and hyperparameter tuning. The new Mean Absolute Error= 1.33 and Root Mean Square Error= 1.84. Our model has improved its accuracy.

5. Next model in the line for me is Support Vector Regression Model. This is a supervised machine learning model which develops support vector closer to the hyperplane. These vectors influence the position and orientation of hyperplane.



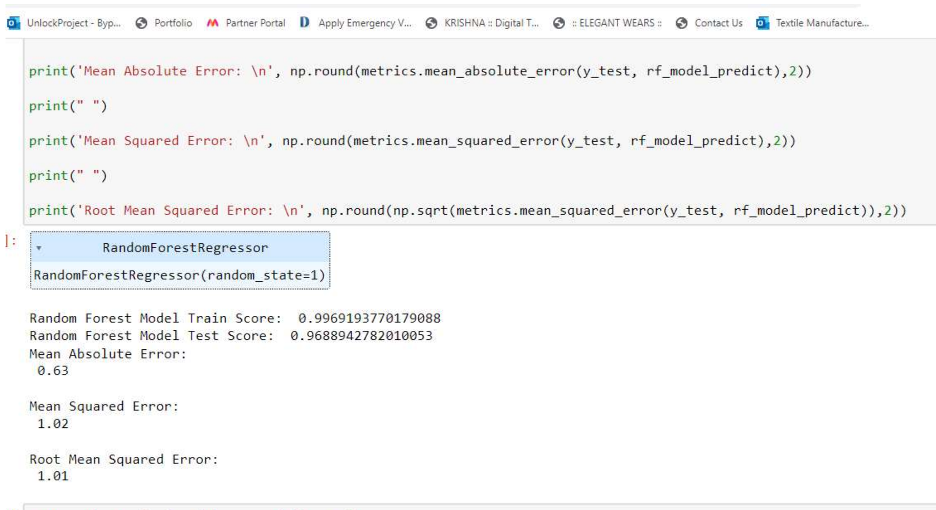
The Mean Absolute Error= 0.92 and Root Mean Square Error= 1.62. These metrics show that our model is performing better than the previous ones.

6. Decision Tree Regressor, happens to be my next model. This is a bagging technique algorithm which is used to reduce the variance in the model prediction. It is an ensemble machine learning technique. Mean Absolute Error= 1.02 and Root Mean Square Error= 1.41

```
print('Mean Absolute Error: \n', np.round(metrics.mean_absolute_error(y_test, y_predict),2))

print(" ")

print('Mean Squared Error: \n', np.round(metrics.mean_squared_error(y_test, y_predict),2))

print(" ")

print('Root Mean Squared Error: \n', np.round(np.sqrt(metrics.mean_squared_error(y_test, y_predict)),2))
```

```
DecisionTreeRegressor
DecisionTreeRegressor(random_state=1)

Decision Tree Model Train Score:  1.0
Decision Tree Model Test Score:  0.9398799882225034
Mean Absolute Error:
 1.02

Mean Squared Error:
 1.98

Root Mean Squared Error:
 1.41
```

7. Further in the position is Random Forest Regressor. This is a boosting algorithm in machine learning ensemble technique. This model will increase the complexity but will also provide more accurate result. Mean Absolute Error= 0.63 and Root Mean Square Error= 1.01.



```
print('Mean Absolute Error: \n', np.round(metrics.mean_absolute_error(y_test, rf_model_predict),2))
print(" ")
print('Mean Squared Error: \n', np.round(metrics.mean_squared_error(y_test, rf_model_predict),2))
print(" ")
print('Root Mean Squared Error: \n', np.round(np.sqrt(metrics.mean_squared_error(y_test, rf_model_predict)),2))
```

```
RandomForestRegressor
RandomForestRegressor(random_state=1)

Random Forest Model Train Score:  0.9969193770179088
Random Forest Model Test Score:  0.9688942782010053
Mean Absolute Error:
 0.63

Mean Squared Error:
 1.02

Root Mean Squared Error:
 1.01
```

➢ Assumptions:

The following assumptions were taken into account:

- The relationship between independent variables and dependent variable is linear.
- The variance residual is same as the value of independent variable.
- Observations are independent of each other.
- Number of values is at least thrice more than the number of columns
- The data is normally distributed.
- For any, fixed value of a particular independent variable, the value of dependent variable is normally distributed
- The average of the residual should be zero

➢ Conclusion:

Amalgamation of technology with business has proven to be a boon for the modern-day business. Data science plays a very vital role in this finding. Data being facts help us to understand the problem and opportunities of the business. Taking this case study for example, we learn about how different variables impact sales. Now, let’s take a broader point of view. Say, an organization is planning to launch an advertisement campaign in western rural region of India. They have a very limited budget. Using the information on media channel, it’s coverage, cost, frequency and section of coverage can help the organization to spend at the right channel. This in return will help the organization to penetrate the market. So, by using this prediction model we can achieve optimum utilization of resources. In this case, budget allocated for advertising.