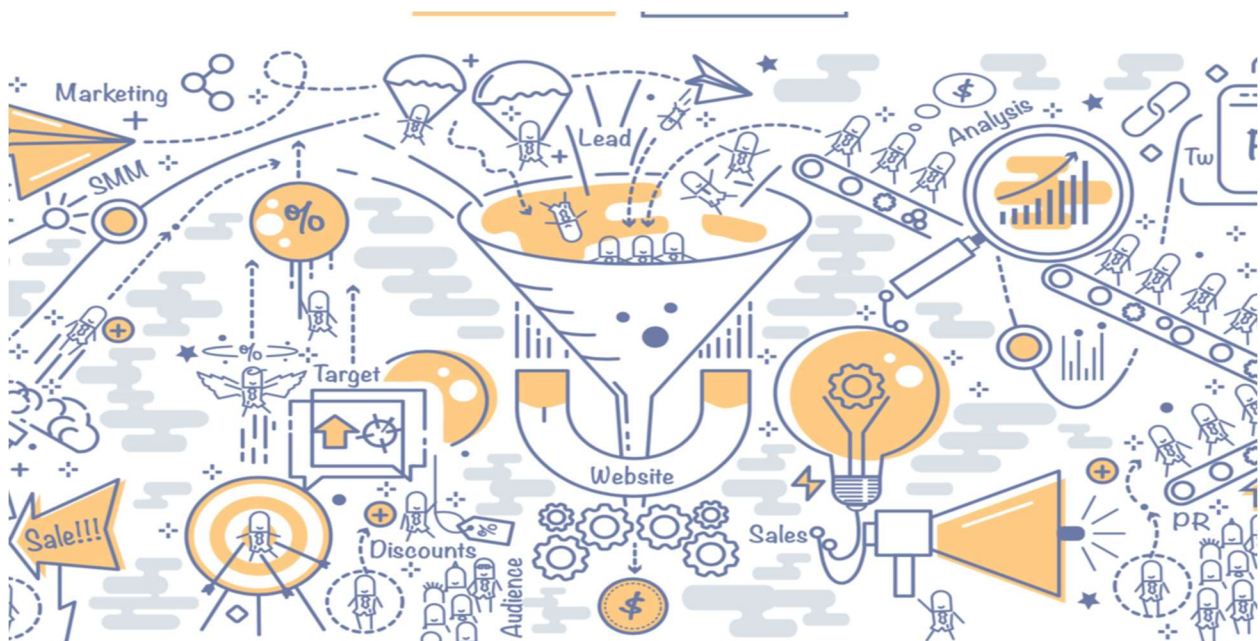


## **Problem Statement**

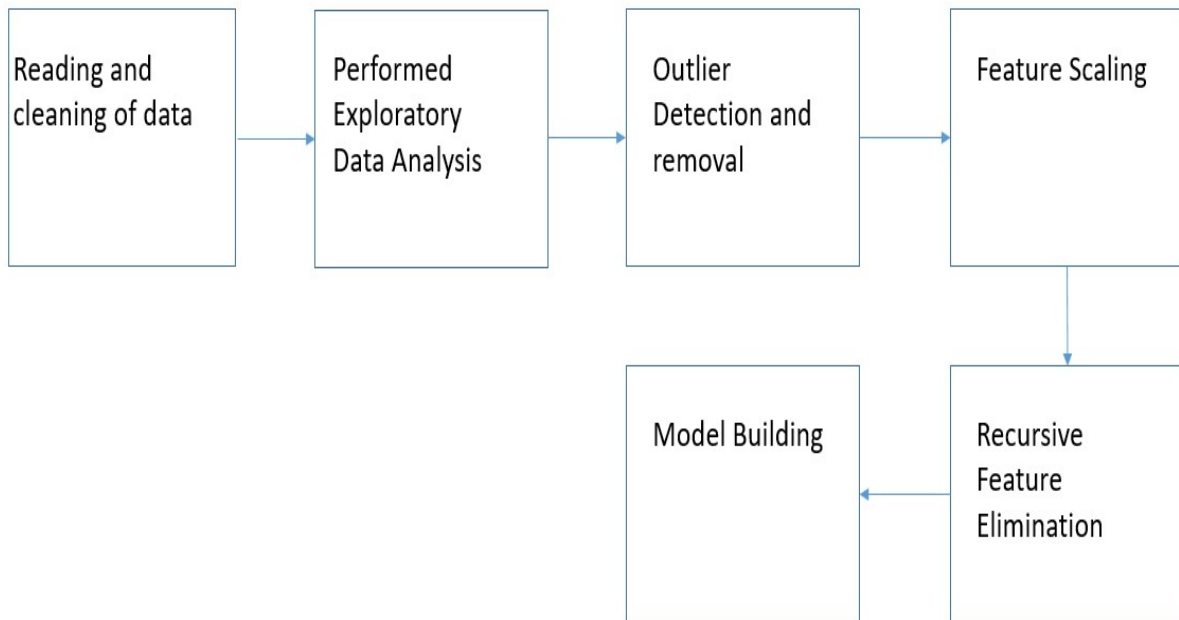
An education company named X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead, although X Education gets a lot of leads, its lead conversion rate is very poor. The task is to build a model wherein you need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

## What is Lead Score?

Lead score is the process of assigning score (rank) to each lead you generate for business. This process helps the sales team to prioritize their work and to focus on their potential customer. With the help of this process a company can work in an optimize way in order to search for their particular customer. With the help of essential features, the machine learning model can easily rack those potential customer. With advancement of Digital Ads, Feedback form, Calls, all these lead to a better machine learning model which will be essential for finding Lead Score.



## Approach



- Reading of the data from csv file.
- As the data consist of very large number of NA values, so in order to fix those issues we have deleted the columns where more than 45% of NA values are there and if it less than 45% we are imputing most frequent values.
- Performed Exploratory Data Analysis on the dataset to get more insight on the data
- For Outlier detection we have used box plot to check all the values are in the range or not.
- Performed Standard Scaling on the updated dataset
- With the help Recursive Feature Elimination, we have identified all the valuable columns are remove that are not useful.
- Iterating over the model creation to check the performance and removing all the unnecessary columns.
- Perform optimal cutoff.
- Created new dataframe with conversion probability.

## Model Evaluation

Using the optimal cut off point if the prediction is done on the model,

We get the confusion Matrix like this

3732 customer were not converted, and model also predicted them as non potential leads

173 customers were wrongly predicted as potential customers while they were not actually

640 customers were converted but the model predicted them as non potential leads

1806 customers who were converted was correctly predicted as potential

Accuracy : 87.2%

Sensitivity : 73.8%

Specificity : 95.6%

False positive rate : 4.2%

Positive Predictive value : 91.2%

Negative Predictive Value : 85.4%

Since the Score is good, we can consider model to be a good performing model and can be used to make predictions on the data.

### **Prediction on Test data**

Accuracy : 86%

Sensitivity : 70.1%

Specificity : 94.8%

Since this score is almost like the sensitivity and specificity score got for trained model the prediction looks correct.

### **Conclusion**

The model has resulted high accuracy results in predicting the leads who can be converted.

So the marketing team can leverage this to make their operations more efficient by reducing the number customer interactions there by improving the conversions as well.