LEAD SCORE CASE STUDY

Submitted By:-
- Apoorva Nagendra
- Raj Pathak
- Charlit Davis

Lead Score Case Study for X Education

Problem Statement :

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. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc.

Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

Business Goal:

X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.
The company needs a model wherein you a lead score is assigned 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%.

Strategy

- Source the data for analysis
- Clean and prepare the data
- Exploratory Data Analysis.
- Feature Scaling
- Splitting the data into Test and Train dataset.
- Building a logistic Regression model and calculate Lead Score. Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and Recall.
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics.

Problem Solving Approach

Data Cleaning & Preparation

- Read the Data from Source
- Convert data into clean format suitable for analysis
- Remove duplicate data &Outlier Treatment
- Exploratory Data Analysis
- Feature Standardization.

Feature Scaling and Splitting Train and Test Sets

- Feature Scaling of Numeric data
- Train & Test Split

Model Building

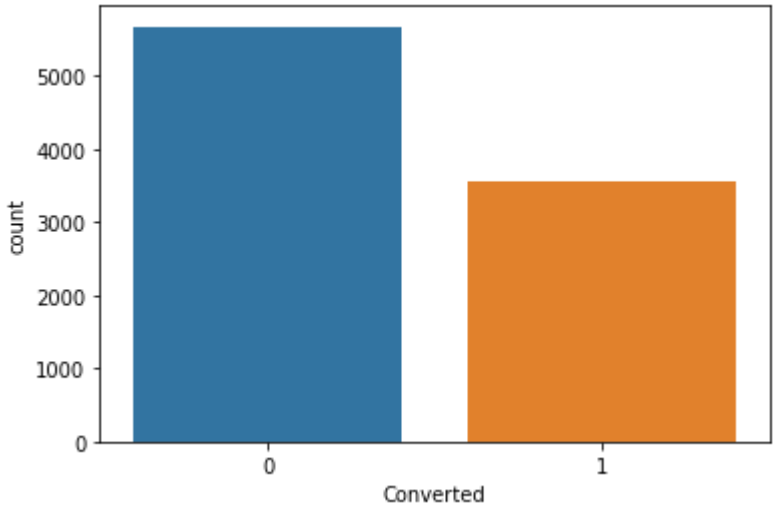
- Feature Selection using RFE
- Determine the optimal model using Logistic Regression
- Calculate various metrics like accuracy, sensitivity, specificity, precision and recall and evaluate the model

Result

- Determine the lead score and check if target final predictions amounts to 80% conversion rate
- Evaluate the final prediction on the test set using cut off threshold from sensitivity and specificity metrics

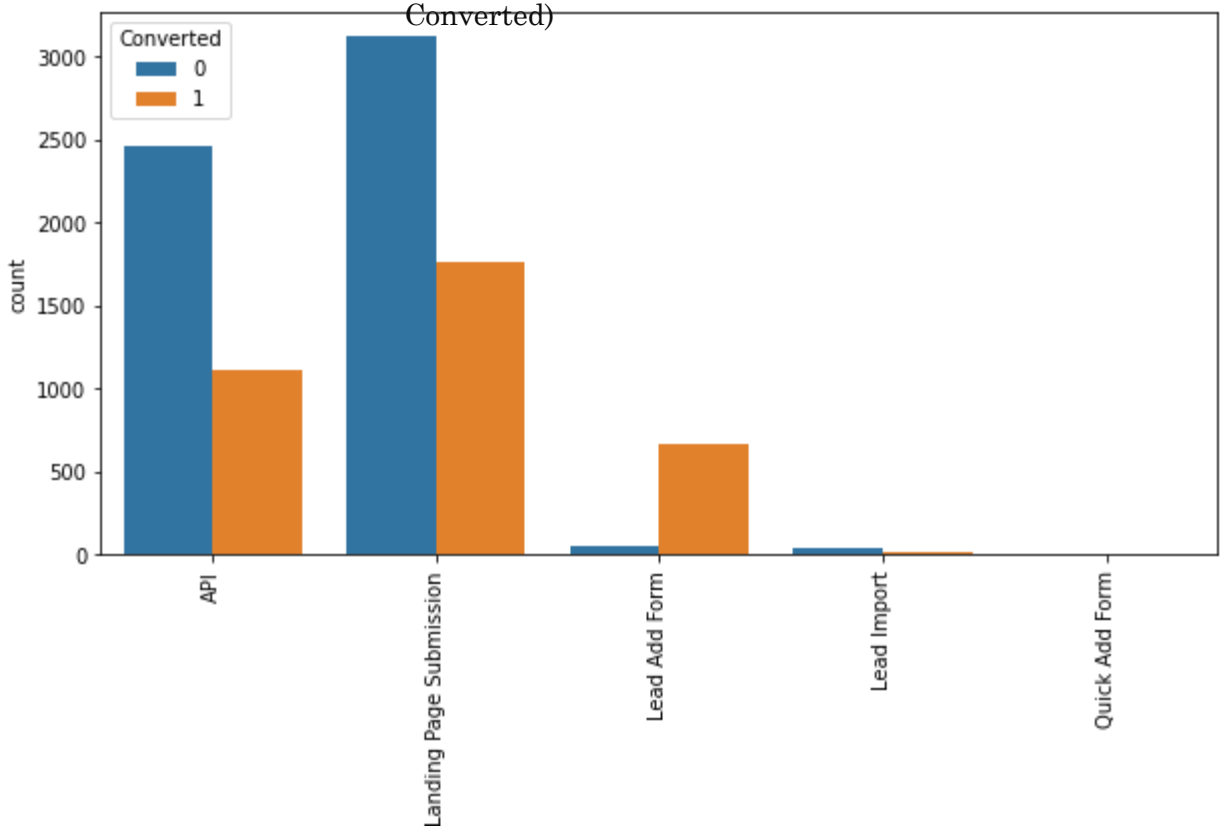
Exploratory Data Analysis

Conversion Ratio



We have around 37.5% Conversion rate in Total

Lead Origin (Converted vs Not Converted)

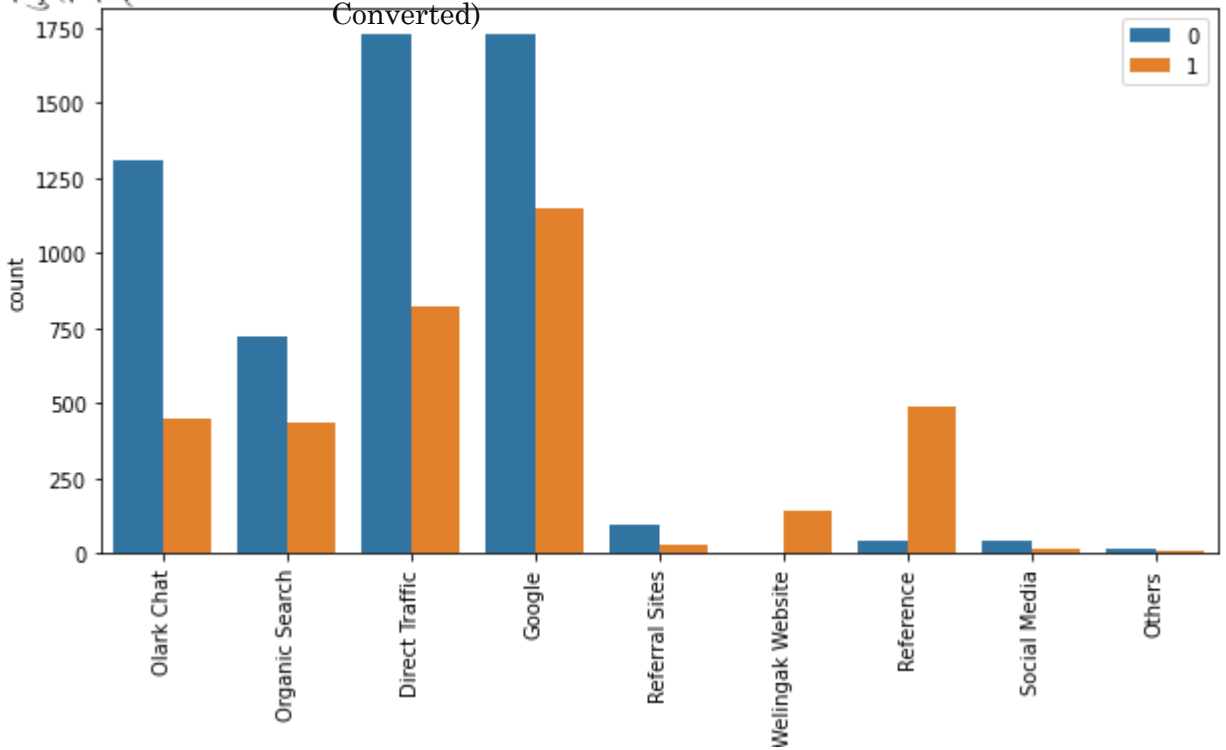


Inferences:-

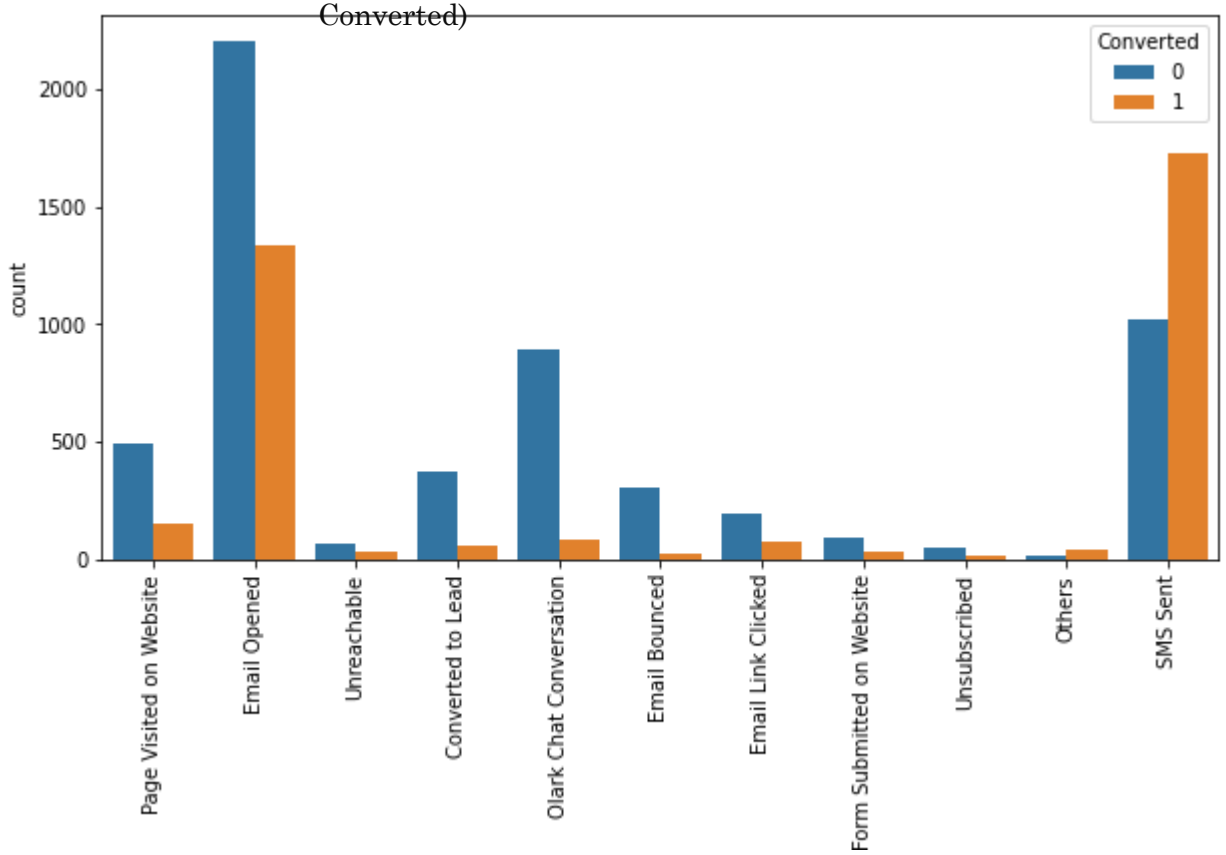
- API and Landing Page Submission bring higher number of leads as well as conversion.
- Lead Add Form has a very high conversion rate but count of leads are not very high.
- Lead Import and Quick Add Form get very few leads.
- In order to improve overall lead conversion rate, we have to improve lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.

Exploratory Data Analysis

Lead Source (Converted vs Not Converted)



Last Activity (Converted vs Not Converted)

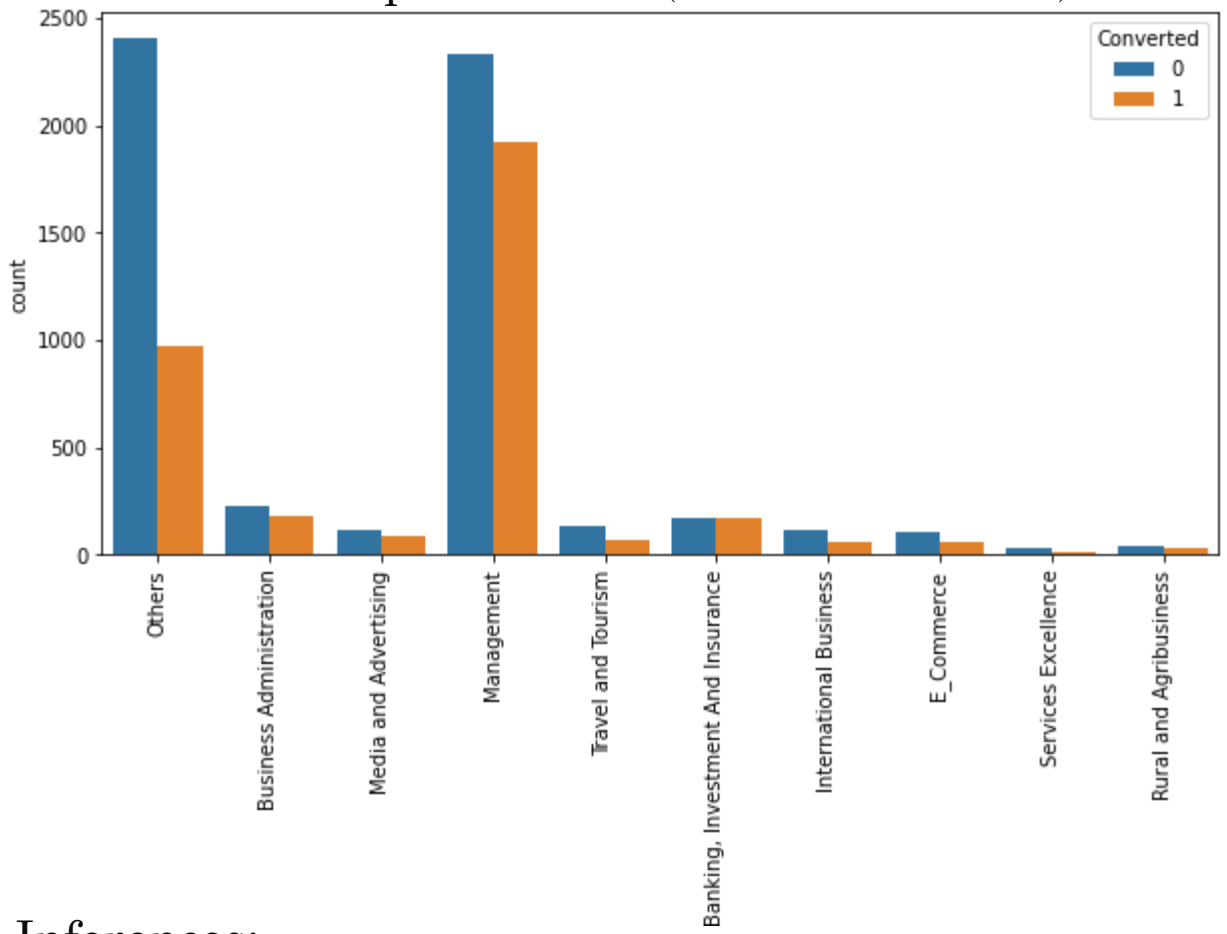


Inferences:-

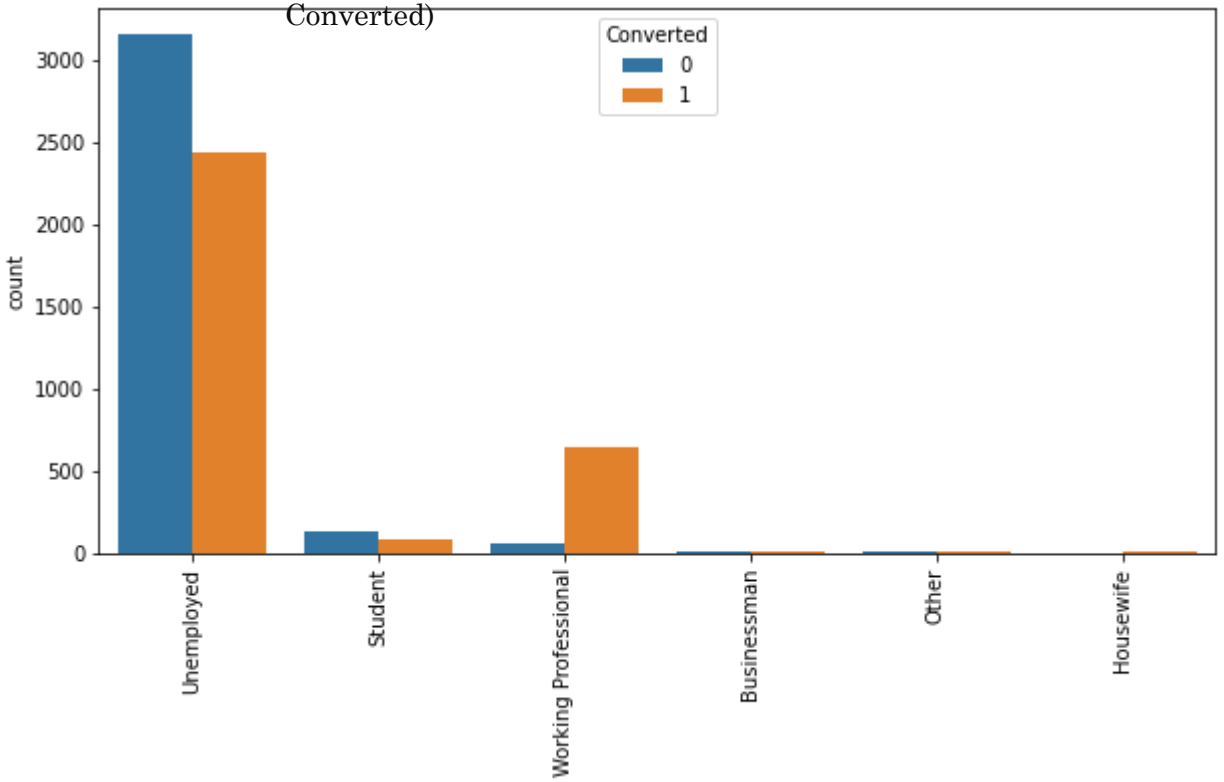
- We Can see leads of Welingak Website & Reference have very higher conversion rate.
- Where as Google, Direct Traffic , Organic Search & Olark Chat has lesser conversion ratio.
- We can clearly see that SMS Sent as last activity has high conversion rate than others

Exploratory Data Analysis

Specialization (Converted vs Not Converted)



Current Occupation (Converted vs Not Converted)



Inferences:-

- Management Specialization Personnel have high lead conversion . For Better conversion , we should also focus on other specialization.
- We can clearly see that Working Professional are more successful leads.

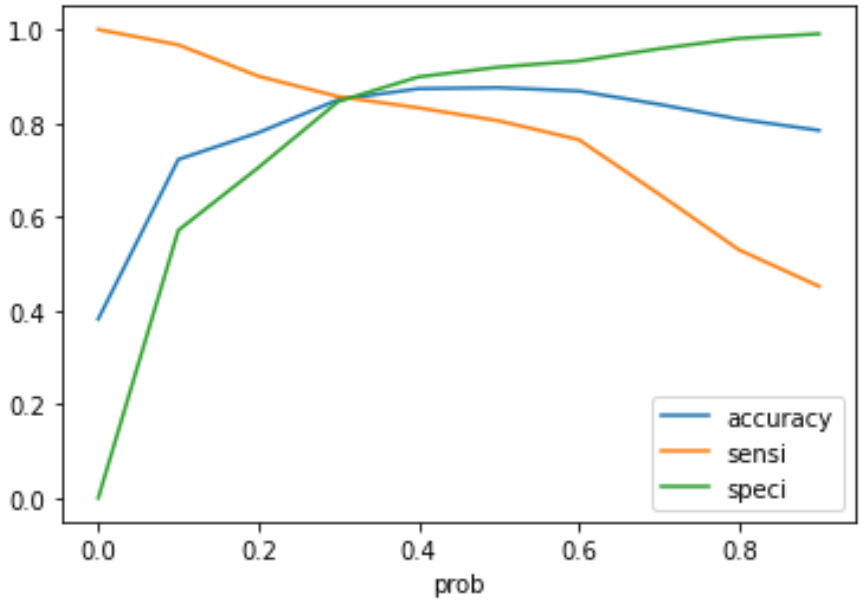
Variables Impacting the Conversion Rate

Top 10 parameters mostly to be considered are as below(In decreasing order of importance)

- Tags_Lost to EINS→ 5.857
- Lead Origin_Lead Add Form → 4.6063
- Last Notable Activity_SMS Sent→ 2.2437
- What is your current occupation_Working Professional →1.9868
- Tags_Will Revert after reading the email →1.7978
- Tags_Busy →1.6882
- Total Time Spent on Website→ 0.9927
- What is your current occupation Unemployed→ -0.6677
- Last Activity_Olark Chat Conversation → -1.09967
- Last Activity_Email Bounced→ -2.3535

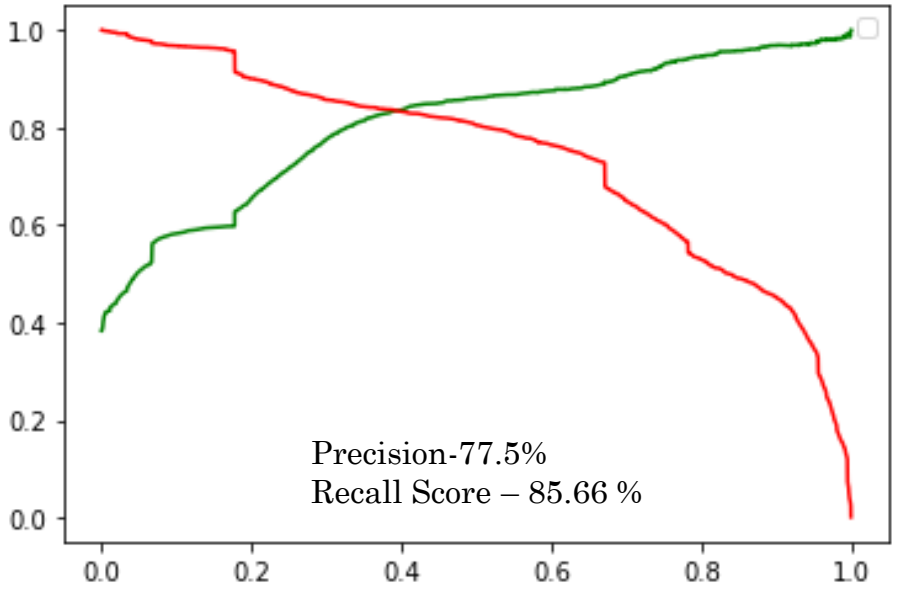
Model Evaluation - on Train Data & Test Data Set

Sensitivity and Specificity on Train Data Set



The graph depicts an optimal cut off of 0.3 based on Accuracy, Sensitivity and Specificity

Precision and Recall on Train Dataset



The graph depicts an optimal cut off of 0.4 based on Precision and Recall

Evaluation	Train Data Set	Test Data Set
	Accuracy - : 85.04% Sensitivity - 85.66% Specificity - 84.66% False Positive Rate – 15.3 % Positive Predictive Value – 77.5 % Negative Predictive Value – 90.5%	Accuracy – 85.7% Sensitivity – 86.16 % Specificity – 85.48 %

Conclusion

- While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction.

- Accuracy, Sensitivity and Specificity values of test set are around 85.7%, 86.16% and 85.48% which are approximately closer to the respective values calculated using trained set.

- The top variables that contribute for lead getting converted in the model are -
  - Tags Lost to EINS
  - Lead Originated from Add Forms
  - Last Notable Activity was SMS Sent
  - Total Time Spent on Website

- As well as X organization should not focus on Unemployed & last activity as Email Bounced.
- In order to improve overall lead conversion rate, we have to improve lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.
- Management Specialization Personnel have high lead conversion . For Better conversion , we should also focus on other specialization.


Thank you!