LEAD SCORING CASE STUDY

SUBMITTED BY:

PRIYA

DEEPIKA GUPTA

SIMRANJEET KAUR

Lead Score Case Study for X Education

PROBLEM STATEMENT

- X Education is an organization which provides online courses for industry professional. The company marks its courses on several popular websites like Google.
- X Education wants to select most promising leads that can be converted to paying customers.
- Although the company generates a lot of leads only a few converted into paying customers, wherein the company wants a higher lead conversion. Leads come through numerous modes like email, advertisements on websites, google searches etc.
- The company has had 30% conversion rate through the whole process of turning leads into customers by approaching those leads which are to be found having interest in taking the course. The implementation process of lead generating attributes are not efficient in helping conversions.

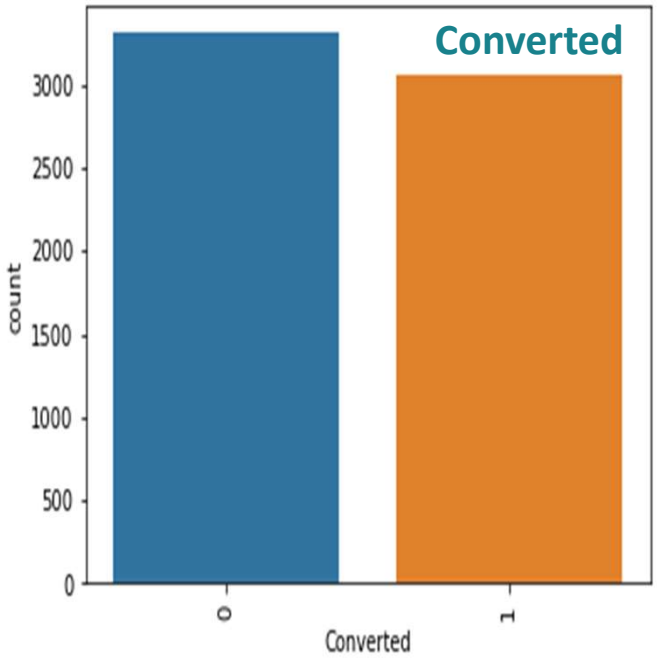
BUSINESS GOAL

- The company requires a model to be built for selecting most promising leads.
- Lead score to be given to each leads such that it indicates how promising the lead could be. The higher the lead score the more promising the lead to get converted the lower it is the lesser the chances of conversion.
- The model to be built in lead conversion rate around 80% or more.

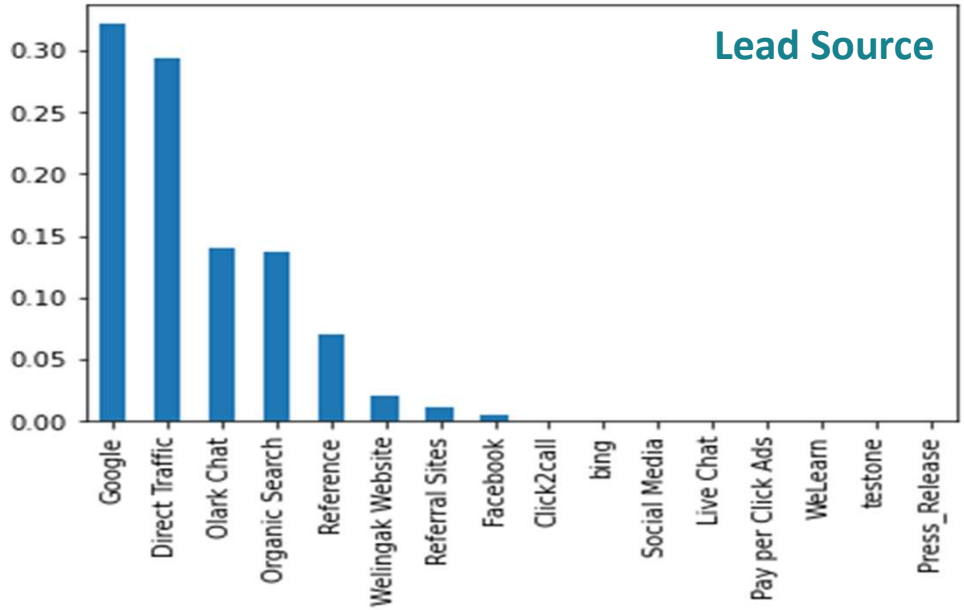
SOLUTION METHODOLOGY

- Reading, understanding and visualizing the data
- Preparing the data for modelling(train-test split, rescaling etc.)
- Training the model
- Model Building & Feature selection using RFE
- Prediction and evaluation on the test set
- Metrics & Plotting ROC Curve
- Finding Cutoff Point & Making Prediction
- Measure the accuracy of the model and other metrics for evaluation

Exploratory Data Analysis

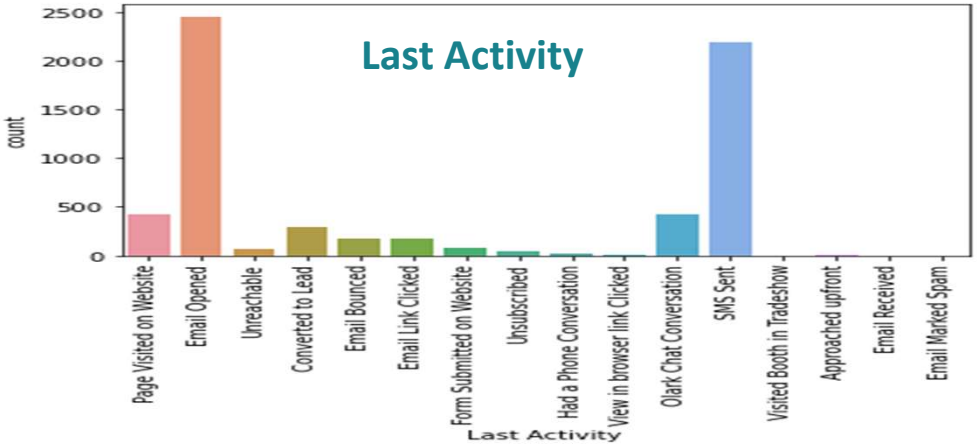


We have around 39% Conversion rate in Total

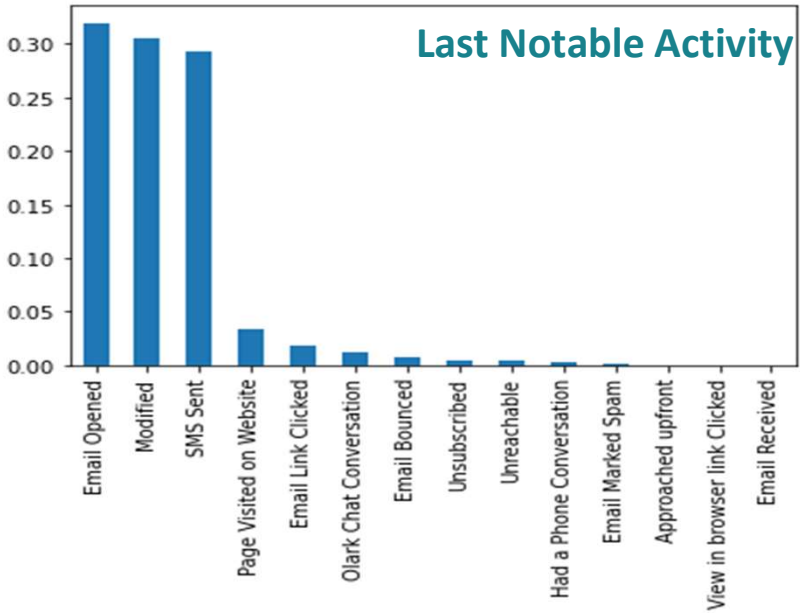


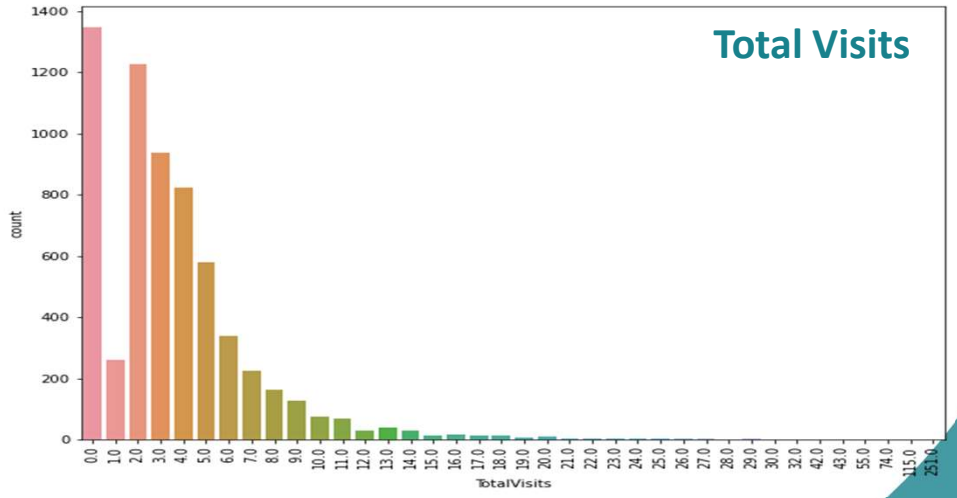
Major conversion in the lead source is from Google

Major conversion has happened from Emails Opened and SMS Sent in Last Activity



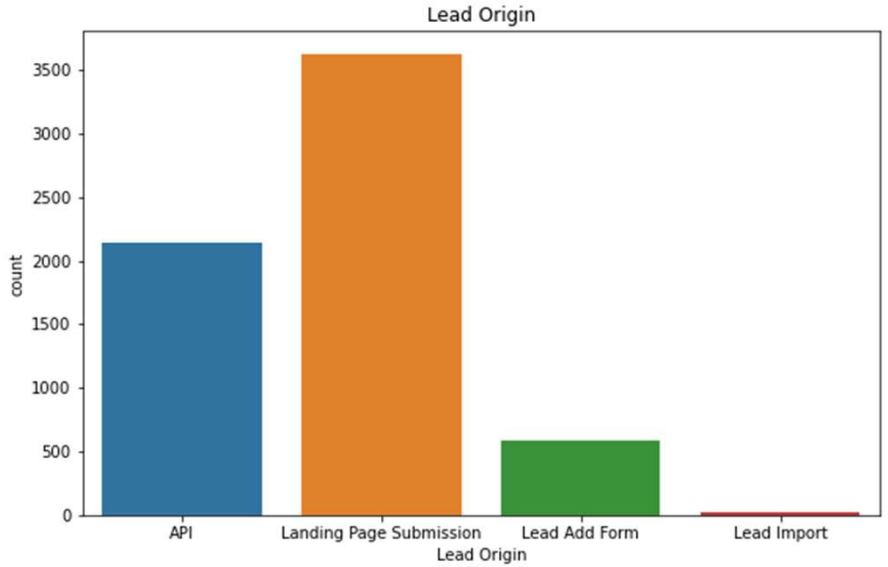
Most leads are converted with message. Emails also induce leads.

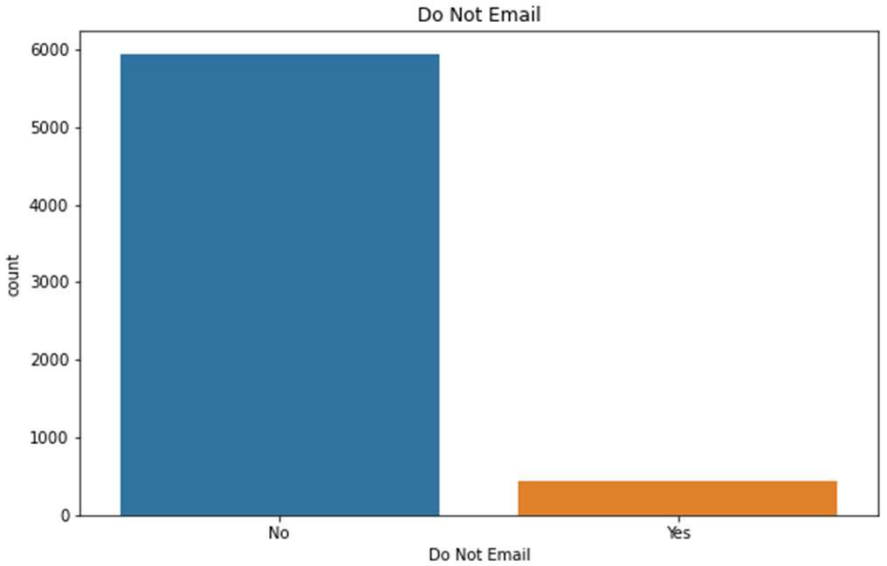




The conversion rates were high for Total Visits

Categorical Variable Relation

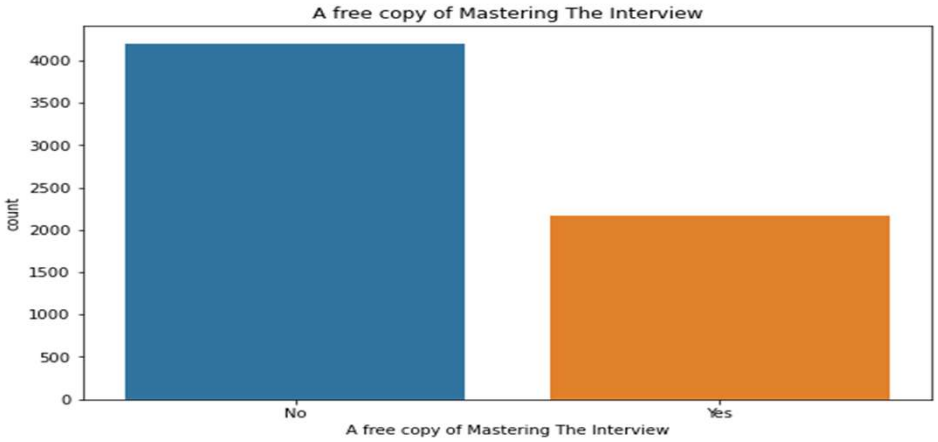




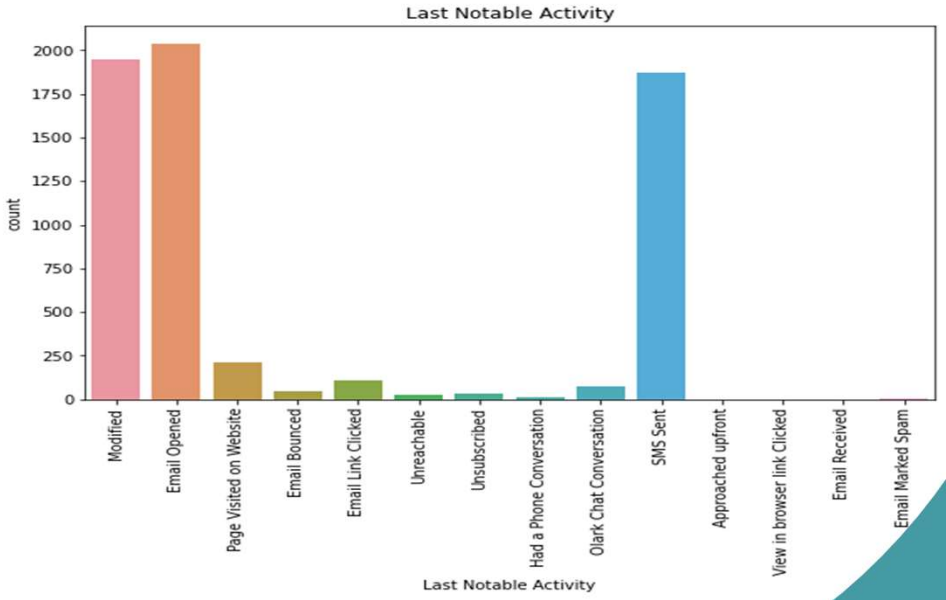
Landing Page Submission has had high lead conversions

Most leads prefer not to informed through Email

Leads prefer less copies of interviews



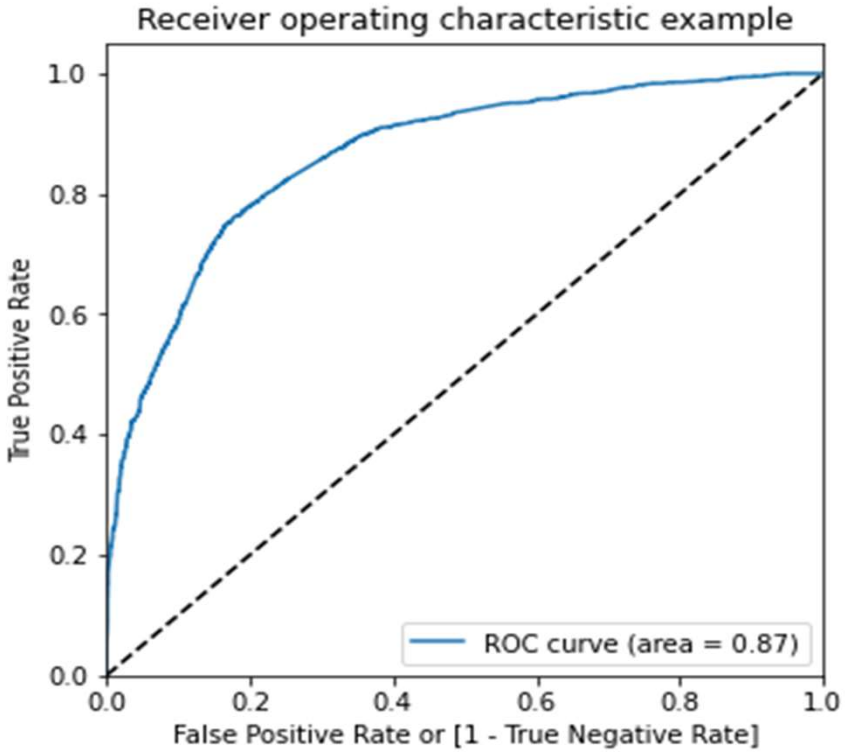
SMS has shown to be a promising method for getting higher confirmed leads, emails also has high conversion



Model Building

- Splitting the Data into Training and Testing Sets
- The first basic step for regression is performing a train-test split
- Use RFE for feature selection
- Assessing the model with StatsModels
- Check VIF value for all the existing columns
- Building model by removing the variable whose P-value is greater than 0.05 and VIF value is greater than 5
- Evaluate accuracy and other metric
- Predictions on test data set
- Precision and recall analysis on test predictions

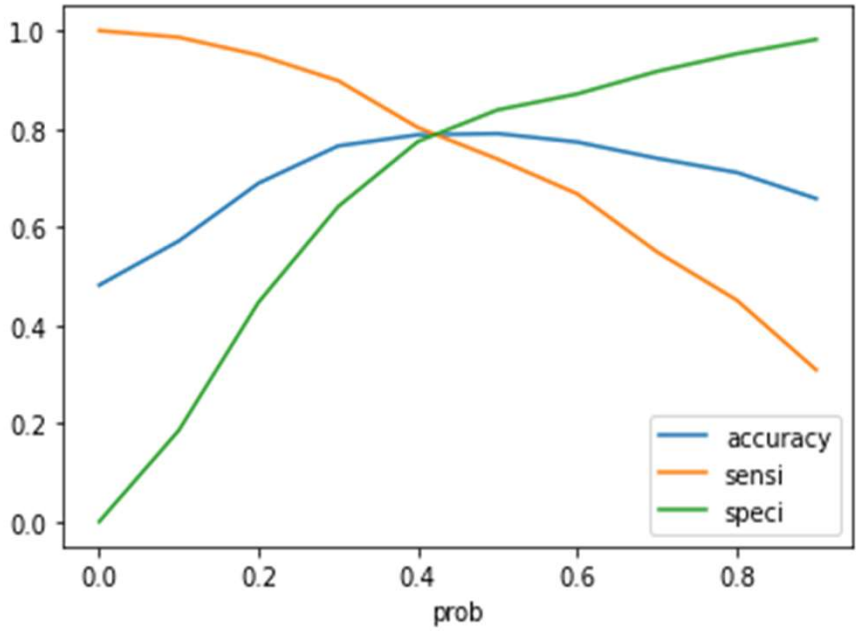
ROC Curve



The area under the curve of the ROC is 0.87 which seems good. So we seem to have a good model. Let's also check for sensitivity and specificity.

Model Evaluation -Sensitivity and Specificity on Train Data Set

The graph depicts an optimal cut off of 0.41 based on Accuracy, Sensitivity and Specificity



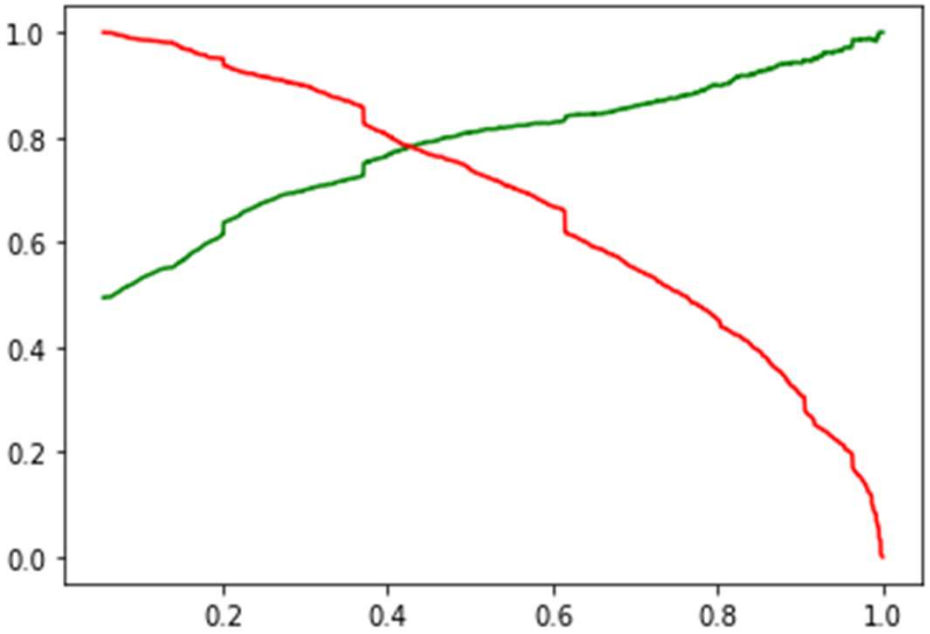
Confusion Matrix

1811	501
441	1708

- **Accuracy - 78%**
- **Sensitivity - 79%**
- **Specificity - 78%**
- **False Positive Rate - 21%**
- **Positive Predictive Value - 77%**
- **Negative Predictive Value - 80%**

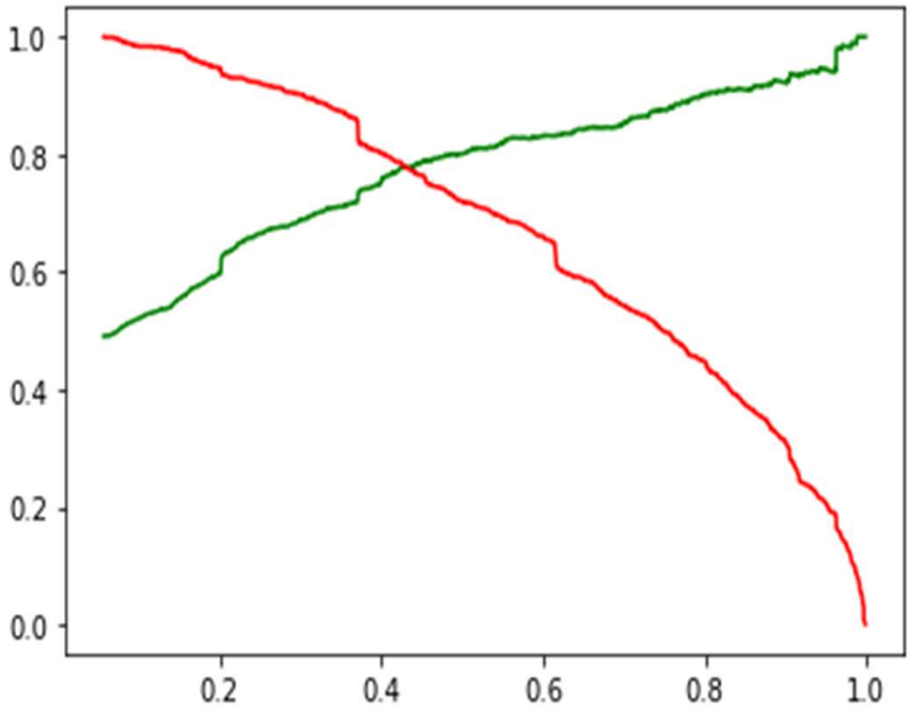
Model Evaluation-Precision and Recall on Train Dataset

The graph depicts an optimal cut off of 0.42 based on Precision and Recall



- **Accuracy - 79%**
- **Precision - 80%**
- **Recall - 73%**

Model Evaluation (Test)



Confusion Matrix

782	214
195	721

- **Accuracy - 78%**
- **Precision - 77%**
- **Recall - 78%**

OBSERVATIONS:

After running the model on the Test Data , we obtain:

- **Accuracy : 78.60%**
- **Sensitivity : 78.71 %**
- **Specificity : 78.51 %**

RECOMMENDATION:

To solve the problem of lead scoring in a case study, my recommendations are as follows:

- Clearly outline the objectives and challenges of lead scoring.
- Establish a lead scoring framework with defined thresholds.
- The company should make calls to leads who spent more time on website as these are more likely to get converted
- The company should not make calls to the leads whose specialization was others as they are not likely to get converted
- The company should make calls to the leads whose last activity was SMS sent as they are more likely to get converted
- The company should not make calls to the leads who chose the option of Do not Email as “yes” as they are not likely to get converted

CONCLUSION

- People spending higher than average time are promising leads, so targeting them and approaching them can be helpful conversions.
- Landing page submissions can help find out more leads.
- SMS messages can have a high impact on lead conversion.
- Marketing management, human resources management has a high conversion rates. People from these specialization can be promising leads.

Logistic Regression Model:

- The model shows high close to 78% accuracy.
- The model shows 79% sensitivity and 78% specificity.
- The model finds correct promising leads and leads that have less chances of getting converted
- Overall this model proves to be accurate.

Thank you