Lead Scoring Case Study Presentation

Submitted by:
Anshika Sharma
Dakshita Sharma
Yash Sharma

PROBLEM STATEMENT:

- X Education sells online courses to industry professionals.
- X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- In order to increase the lead conversion rate, the company first should identify the most potential leads, also known as ‘Hot Leads’.
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

GOAL

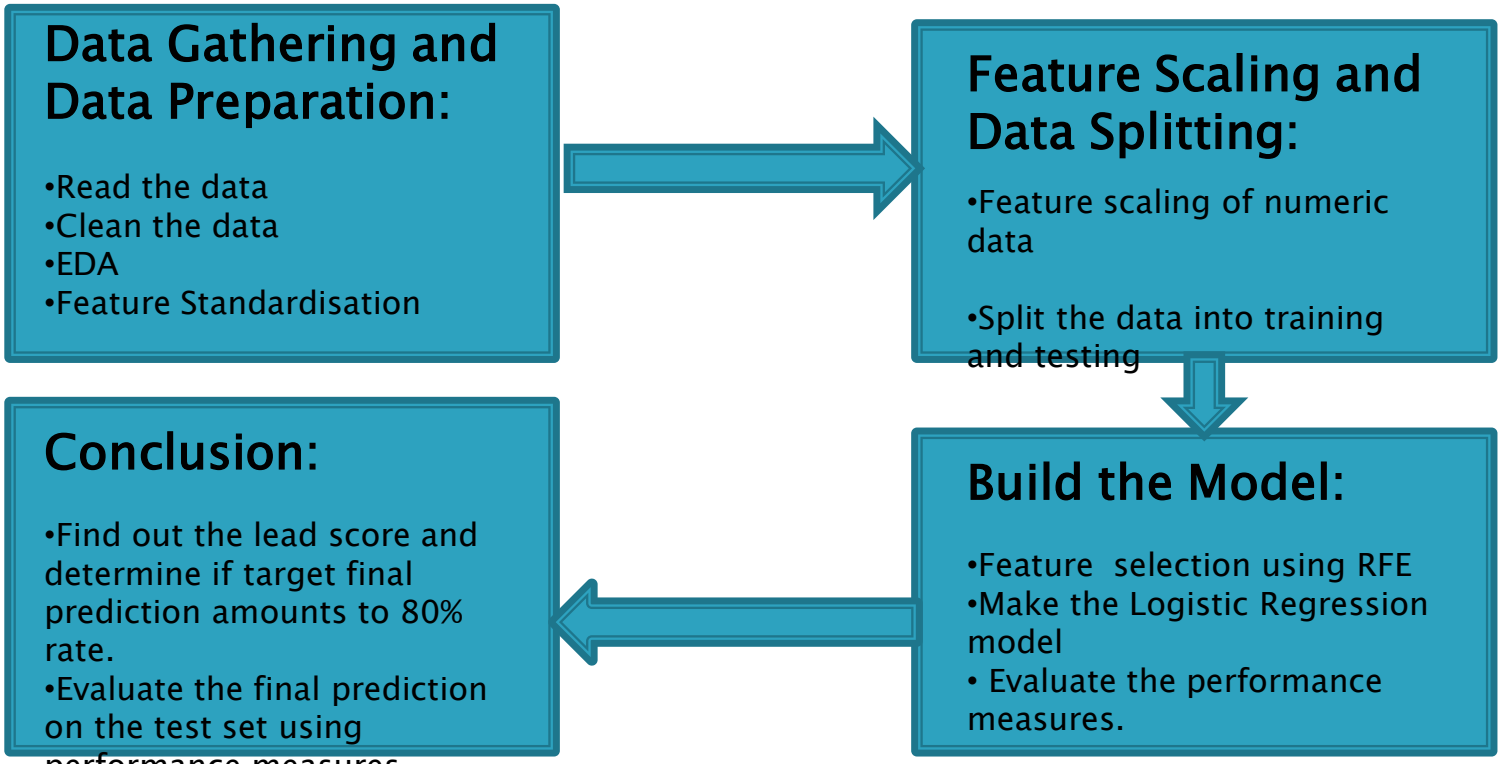
- What are the most promising leads? asks X education.
- They aim to create a model that identifies the hot leads for that purpose.
- The model's deployment for upcoming use.

TECHNIQUES:

- Data gathering
- Data preparation
- Exploratory data analysis
- Feature scaling
- Data splitting into training and testing
- Build a Logistic Regression model
- Evaluate the performance measure such as Accuracy, Sensitivity and Specificity.
- Applying the best model in test data with regards to Sensitivity and Specificity.

METHODOLOGY:

Follow the steps:

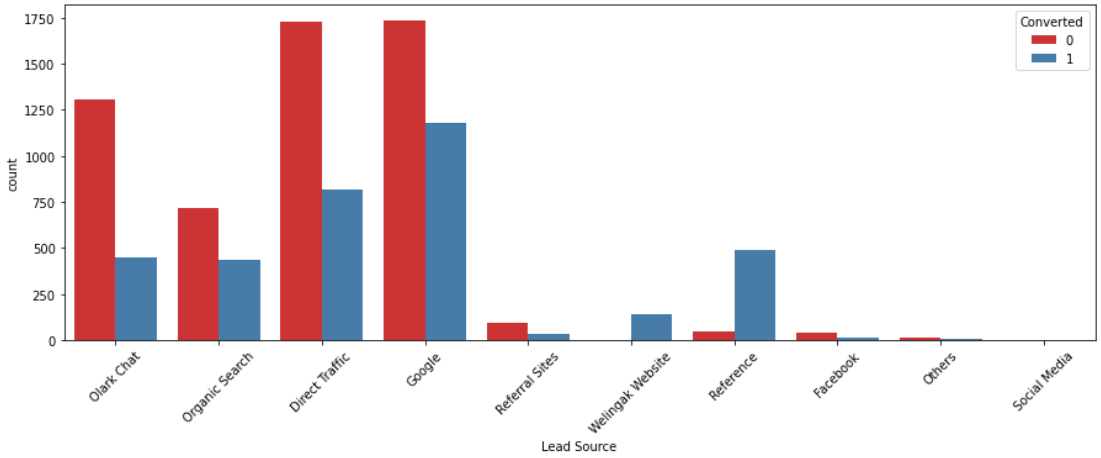


EDA:

visualizing count of Lead Source Variable based on Converted value

Inference

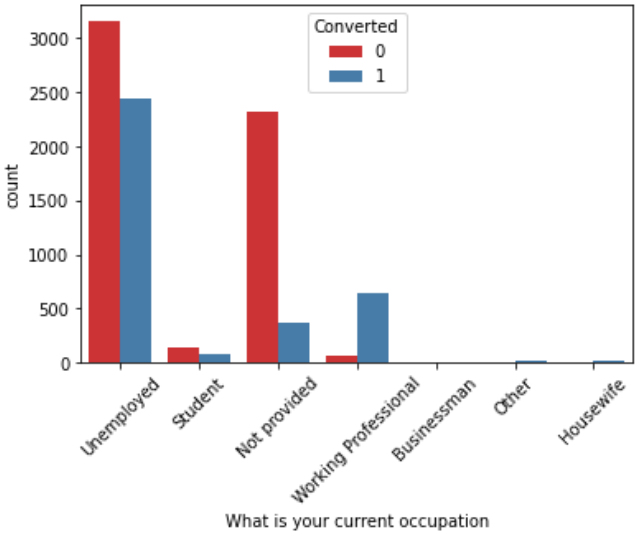
- Google and Direct Traffic are generating Maximum leads
- There is high Conversion rate of Reference leads and Welinkgak Website



EDA:

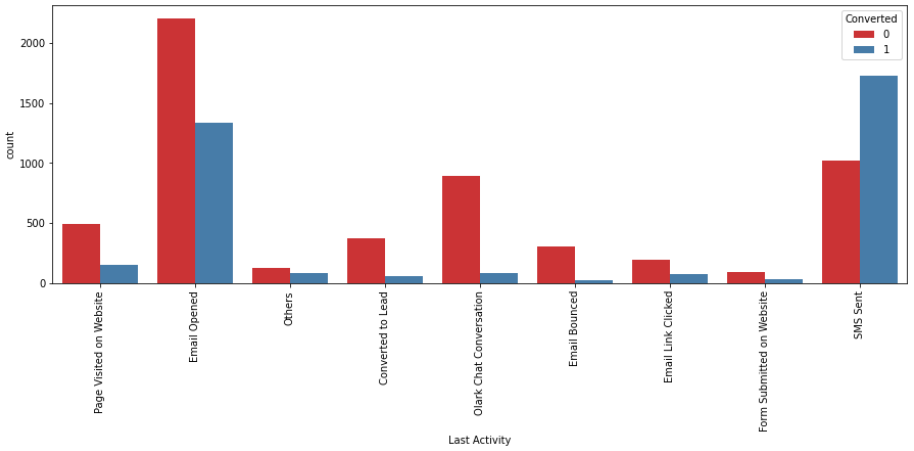
Inference

•most of the leads are generated by the students and more then 60% are converted.
•convertion rate for the working professional is very high so we need to focus on them



Inference

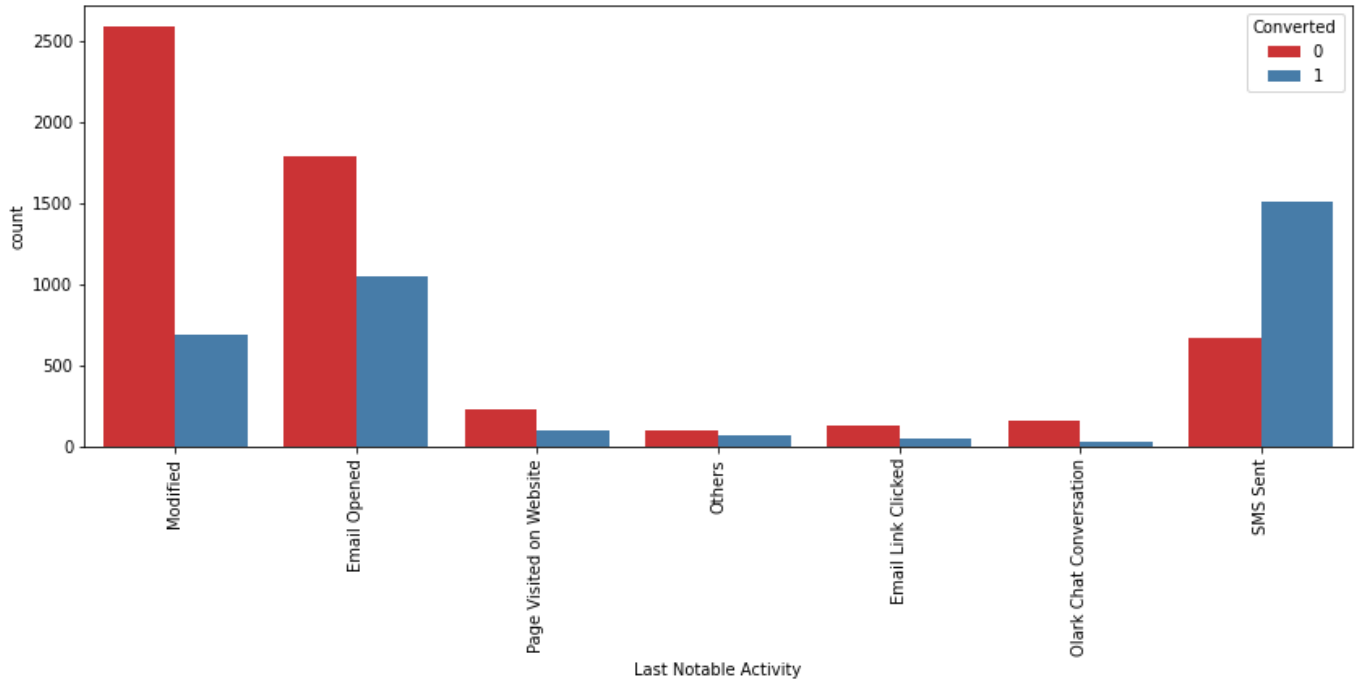
•Maximum leads are from email opened.
•the highest coversion rate is of SMS send.



EDA:

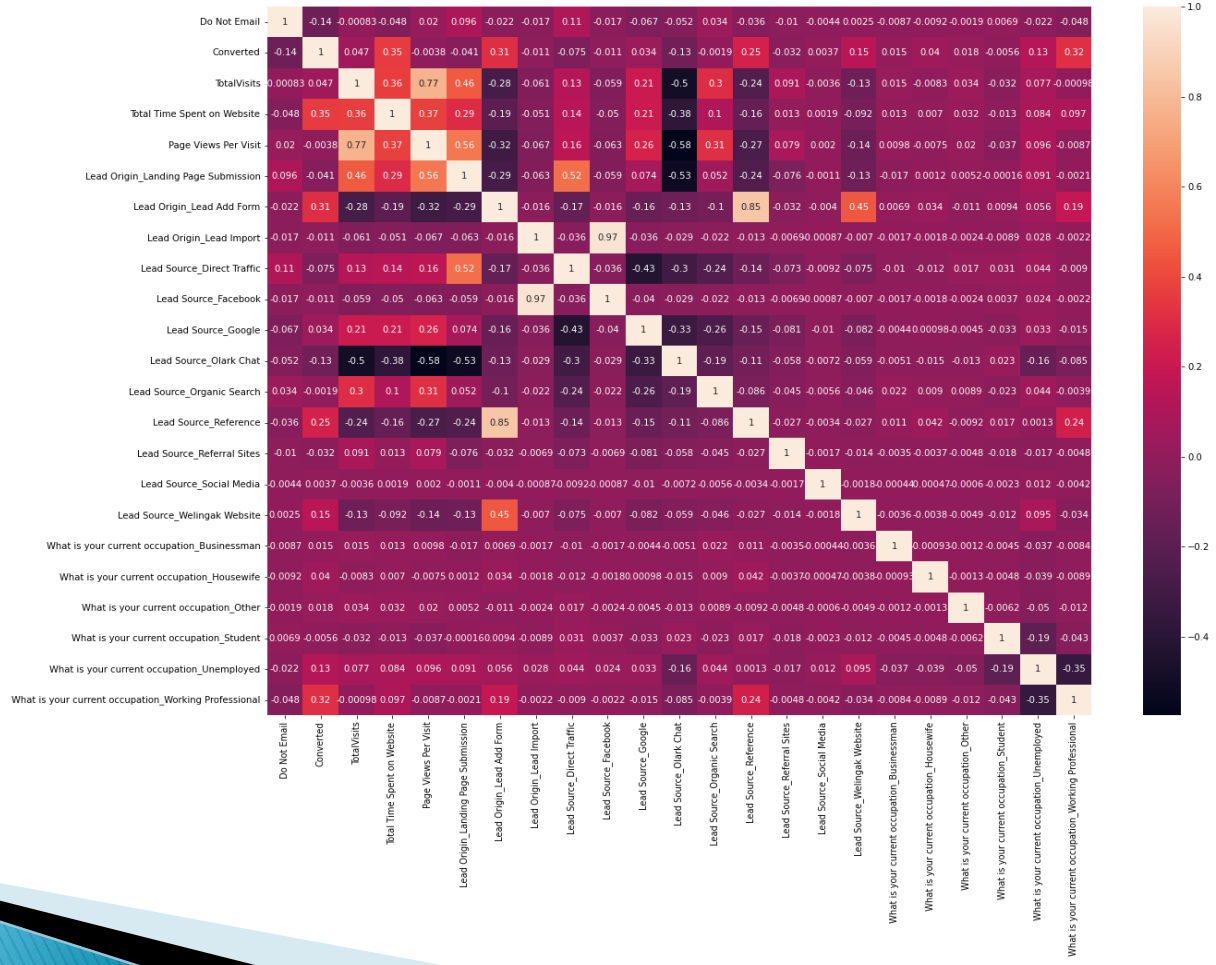
Inference

- Maximum leads are generated having last activity as Email opened
- highest conversion rate is from SMS Sent.



EDA:

Correlation matrix



Result:

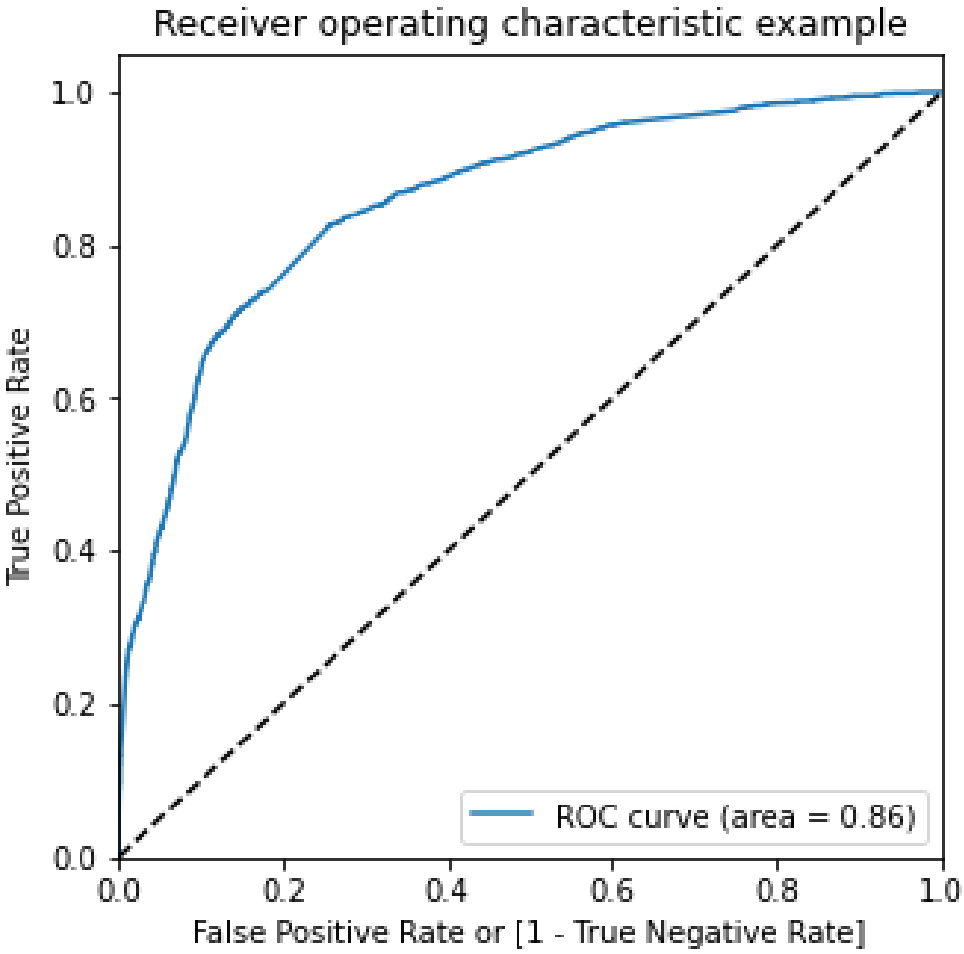
Calculating VIF

Each value of VIF is good for further proceeding for making further predictions using this model.

	Features	vif
8	What is your current occupation_Unemployed	1.61
3	Lead Source_Direct Traffic	1.38
5	Lead Source_Olark Chat	1.33
1	Total Time Spent on Website	1.29
2	Lead Origin_Lead Add Form	1.28
9	What is your current occupation_Working Profes...	1.16
7	What is your current occupation_Student	1.03
0	Do Not Email	1.02
4	Lead Source_Facebook	1.02
6	Lead Source_Referral Sites	1.01

Cont…

The ROC Curve should be a value close to 1. We are getting a good value of 0.85 indicating a good predictive model.



Performance Measure:

From this curve, 0.3 is the optimum point to take it as a cutoff probability.

Inference:

So the model seems to work correctly. And the ROC curve value is 0.86. We have the following values for the Train **Data**:

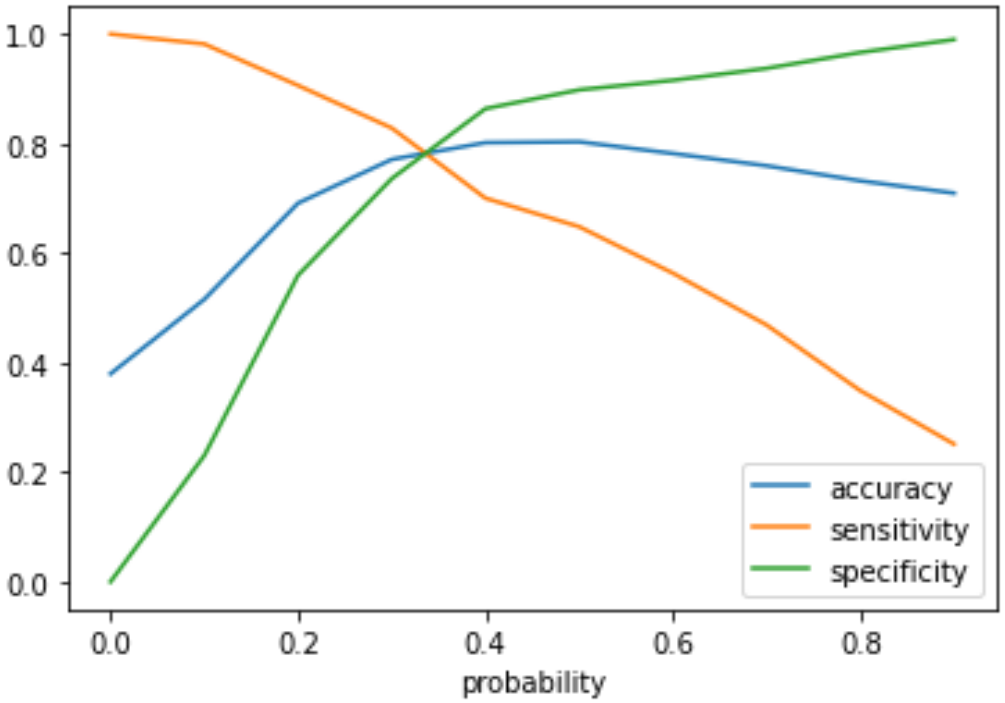
Accuracy : 77.10%

Sensitivity :82.84%

Specificity : 73.58%

Confusion Matrix:

2909	1044
415	2004



Conclusion:

Important features responsible for good conversion rate or the ones' which contributes more towards the probability of a lead getting converted are :

- Lead Origin_Lead Add Form
- What is your current occupation_Working Professional
- Total Time Spent on Website

Recomendations:

- The lead from "Lead Origin_Lead Add Form" is more likely to convert. Hence company should make a call to this lead.
- The lead from "What is your current occupation_Working Professional" is more likely to convert. Hence company should make a call to this lead.
- The lead from "What is your current occupation_Unemployed " is more likely to convert. Hence company should make a call to this lead.
- The lead from "What is your current occupation_Student " is more likely to convert. Hence company should make a call to this lead.

Cont…

- The lead from "Lead Source_Facebook " is not more likely to convert. Hence company should not make a call to this lead.
- The lead from "Lead Source_Olark Chat" is not more likely to convert. Hence company should not make a call to this lead.
- The lead from "Lead Source_Direct Traffic " is not more likely to convert. Hence company should not make a call to this lead.
- The lead from "Do Not Email" is not more likely to convert. Hence company should not make a call to this lead.
- The lead from "Lead Source_Referral Sites" is not more likely to convert. Hence company should not make a call to this lead.
- The lead from "const" is not more likely to convert. Hence company should not make a call to this lead.