



# Lead Scoring Case Study

Submitted by

Gitartha Pal

Harish Gupta

Bhavika Gupta

# Contents

- Problem statement
- Problem approach
- EDA
- Correlations
- Model Evaluation
- Observations
- Conclusion

# Problem Statement

- ❖ An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. They have process of form filling on their website after which the company that individual as a lead.
- ❖ 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%. Now, this means if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as Hot Leads in order to let the conversation rate go up

# Business Objective

- ❖ X Education wants to build a model to give every lead a lead score between 0 -100 . So that they can identify the Hot leads and increase their conversion rate as well.
- ❖ The CEO want to achieve a lead conversion rate of 80%
- ❖ They want the model to be able to handle future constraints as well like Peak time actions required, how to utilize full man power and after achieving target what should be the approaches

# Problem Approach

- Importing the data and inspecting the data frame
- Data preparation
- EDA
- Dummy variable creation
- Test-Train split
- Feature scaling
- Correlations
- Model Building (RFE R squared VIF and p-values)
- Model Evaluation
- Making predictions on test set

## EDA – Data Cleaning

There are a few columns in which there is a level called 'Select' which is having maximum count

```
In [25]: leads['How did you hear about X Education'].value_counts()
```

```
Out[25]: Select                5043
Online Search                808
Word Of Mouth                348
Student of SomeSchool        310
Other                        186
Multiple Sources             152
Advertisements               70
Social Media                 67
Email                       26
SMS                          23
Name: How did you hear about X Education, dtype: int64
```

```
In [24]: leads['Lead Profile'].astype('category').value_counts()
```

```
Out[24]: Select                4146
Potential Lead                1613
Other Leads                   487
Student of SomeSchool         241
Lateral Student               24
Dual Specialization Student   20
Name: Lead Profile, dtype: int64
```

```
In [26]: leads['Specialization'].value_counts()
```

```
Out[26]: Select                1942
Finance Management            976
Human Resource Management     848
Marketing Management          838
Operations Management         503
Business Administration       403
IT Projects Management        366
Supply Chain Management       349
Banking, Investment And Insurance 338
Travel and Tourism            203
Media and Advertising         203
International Business        178
Healthcare Management         159
Hospitality Management        114
E-COMMERCE                    112
Retail Management             100
Rural and Agribusiness         73
E-Business                    57
Services Excellence           40
Name: Specialization, dtype: int64
```

\*Leads from HR, Finance & Marketing management specializations are high probability to convert

# Correlation

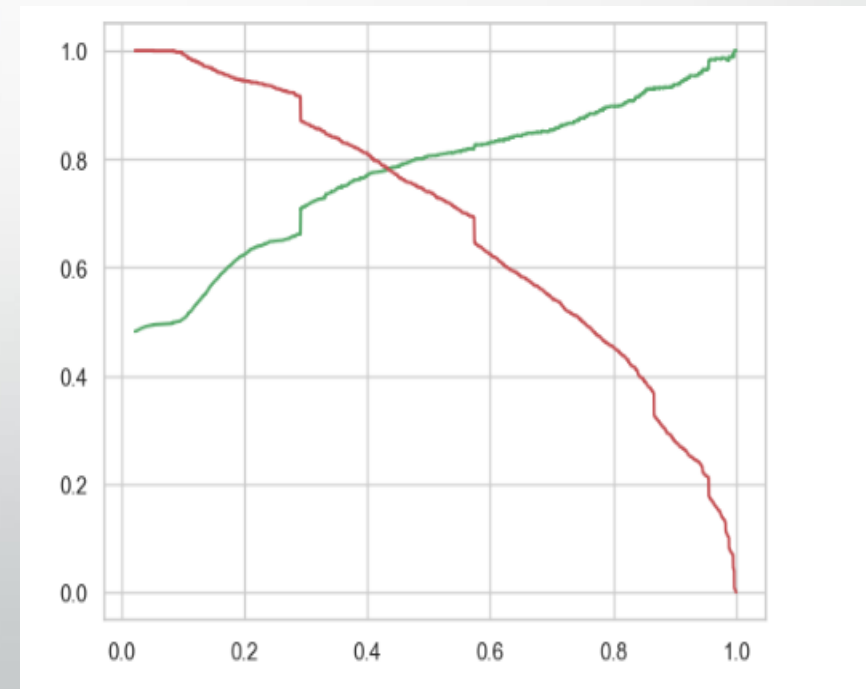
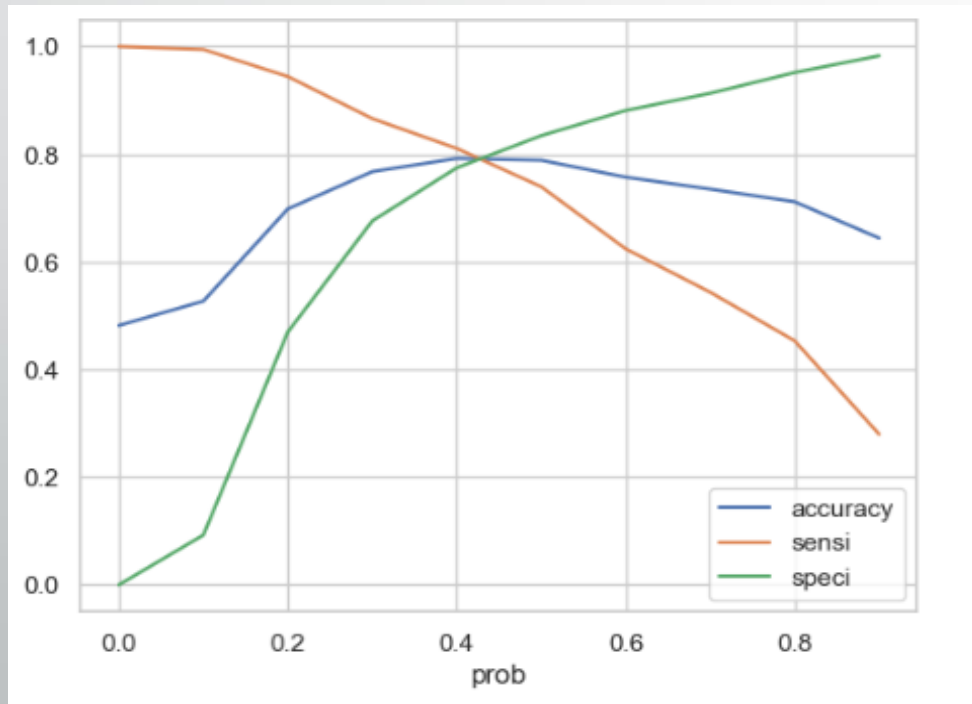
There is no correlation between the variables



# Model Evaluation

## ROC curve

0.42 is the tradeoff between Precision and Recall - Thus we can safely choose to consider any Prospect Lead with Conversion Probability higher than 42% to be a HOT LEAD





# Observations

## **Train Data:**

Accuracy : 80%

Sensitivity : 77%

Specificity : 80%

## **Test Data:**

Accuracy : 80%

Sensitivity : 77%

Specificity : 80%

## Final Features list

- Lead Source\_Olark Chart
- Specialization\_Others
- Lead Origin\_Lead Add Form
- Lead Source\_Welingak Website
- Total Time Spent on Website
- Lead Origin\_Landing Page Submission
- What is your current occupation\_Working Professionals
- Do Not Email

# Conclusion

- We see that the conversion rate is 30-35% (close to average) for API and Landing page submission. But very low for Lead Add form and Lead import. Therefore we can intervene that we need to focus more on the leads originated from API and Landing page submission
- We see max number of leads are generated by google/ direct traffic. Max conversion ratio is by reference and welingak website
- Leads who spent more time on website, more likely to convert
- Most common last activity is email opened. Highest rate is SMS sent. Max conversion with working professional