

# Case Study - Leads Scoring

Deeshma M

# Steps followed

## 1) Data Cleaning:

- drop columns with more than 40% null values
- Deleted *Country, How did you hear about X Education, Lead Profile, and City* column
- Deleted columns such as *Do Not Email, Do Not Call, Search, Magazine, Newspaper Article, X Education Forums, Newspaper, Digital Advertisement, Through Recommendations, Receive More Updates About Our Courses, Update me on Supply Chain Content, Get updates on DM Content, I agree to pay the amount through cheque, A free copy of Mastering The Interview, What matters most to you in choosing a course, Tags* as they do not have interpretable info

## 2. Dummy variable creation

*'Lead Origin','Lead Source', 'Last Activity', 'What is your current occupation',  
'Last Notable Activity', Specialization*

3. Train - Test Split -The split was done at 70% and 30% for train and test data respectively.

## 4. Scaling

## 5. Model Building using RFE

6. Removed columns such as *Lead Origin\_Lead Add Form* , *Last Notable Activity\_Had a Phone Conversation*, *'What is your current occupation\_Housewife*, *What is your current occupation\_Working Professional*

## 7. Model Evaluation

- *Sensitivity of train dataset – 78.8%*
- *Specificity of train dataset – 79.2%*
- *Sensitivity of test data set – 78.05%*
- *Specificity of test dataset – 79.8%*

## 8. Final Model

**9.1347\* TotalVisits + 4.4306\* Total Time Spent on Website + 1.4429\* Lead Source\_Olark Chat + 4.1882\* Lead Source\_Reference + 6.2565\* Lead Source\_Welingak Website + -1.6866\* Last Activity\_Email Bounced + 2.8035\* Last Activity\_Had a Phone Conversation + 1.1130\* Last Activity\_SMS Sent + -2.3539\* What is your current occupation\_Student + -2.5480\* What is your current occupation\_Unemployed + 2.8083\* Last Notable Activity\_Unreachable**

# Inference

It was found that the variables that mattered the most are:

1. TotalVisits
2. Lead Source\_Welingak Website
3. Total Time Spent on Website
4. Lead Source\_Reference
5. Last Activity\_Had a Phone Conversation
6. Last Notable Activity\_Unreachable
7. Lead Source\_Olark Chat
8. Last Activity\_SMS Sent
9. Last Activity\_Email Bounced
10. What is your current occupation\_Student
11. What is your current occupation\_Unemployed