Lead Scoring Case Study using Logistic regression

Submitted by –
Awnish Praveen Toppo

CONTENT

- **Problem statement**
- **Business Objective**
- **Problem Approach**
- **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.
- After obtaining these leads, sales team members begin calling, emailing, and so on. Some of the leads convert during this process, but the majority do not.
- At X education, the lead conversion rate is typically 30%. This implies that just around 30 of the leads, say, that they generate each day will actually be converted. The organization wants to find the most potential leads, or "Hot Leads," in order to streamline this procedure.
- 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

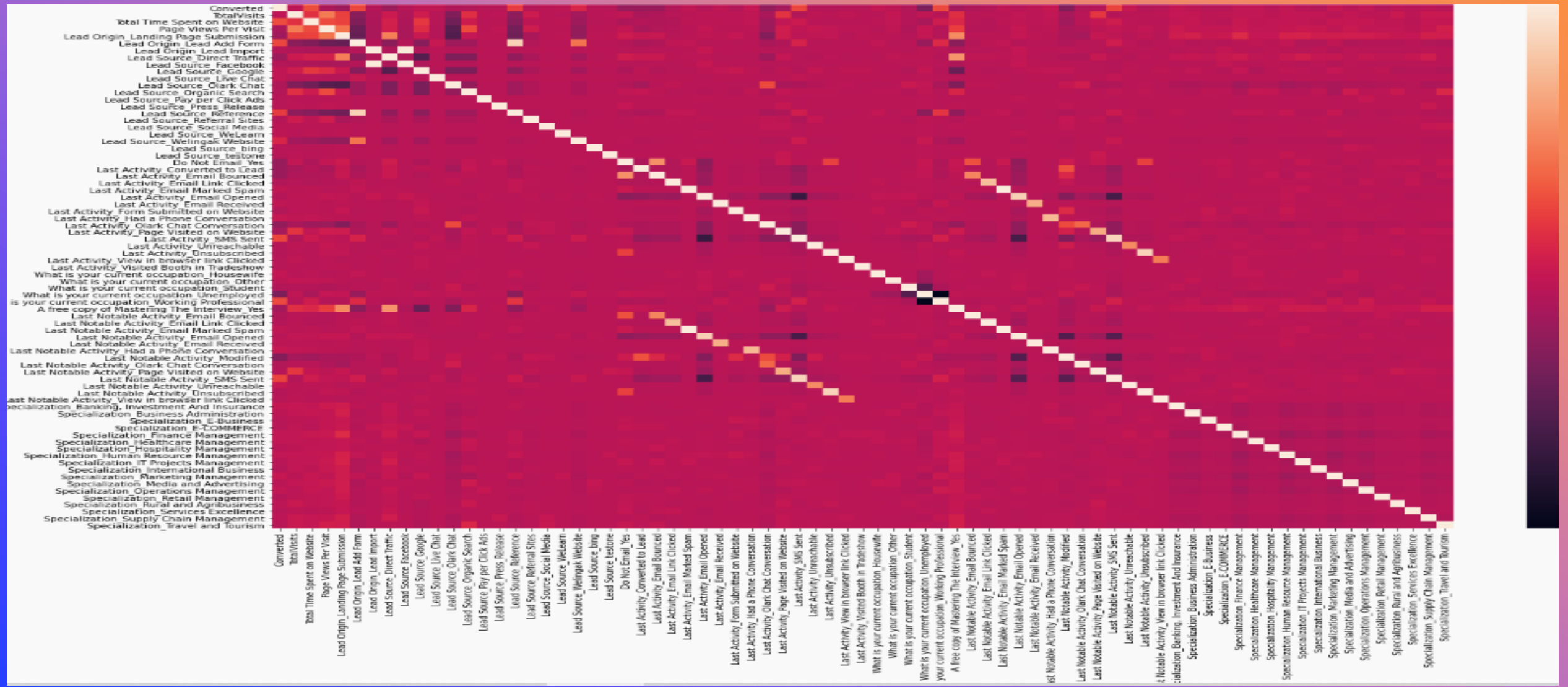
BUSINESS OBJECTIVE

- Lead X wants us 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%.
- Future limitations such as the need for peak time activities, how to use all available manpower, and what to do when a target is reached should also be handled by the model.

PROBLEM APPROACH

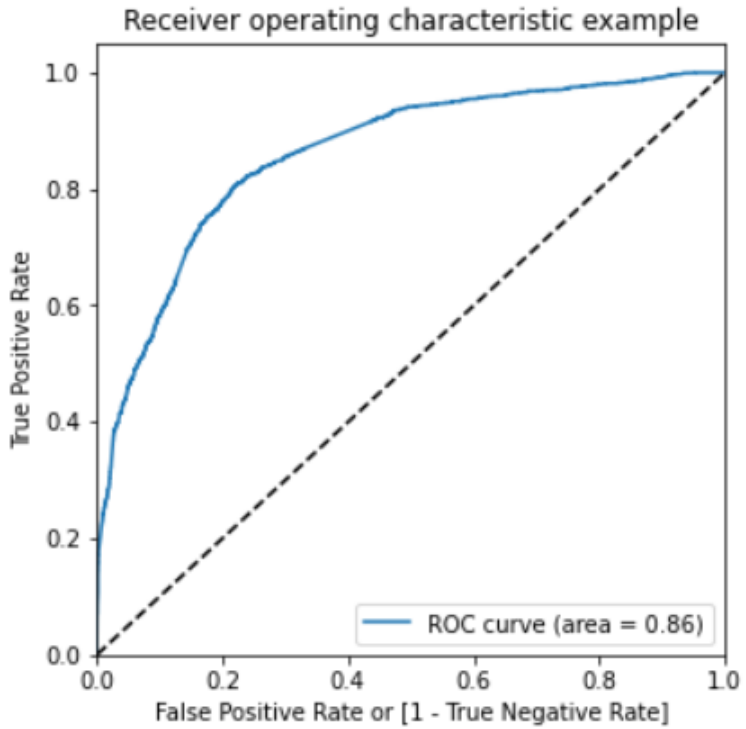
- Importing the data and Understanding the data frames
- Data Cleaning and Preparation
- Dummy variable creation
- Test-Train split
- Scaling
- Correlations
- Model Building (RFE Rsquared VIF and p- values)
- Model Evaluation
- Making predictions on test set

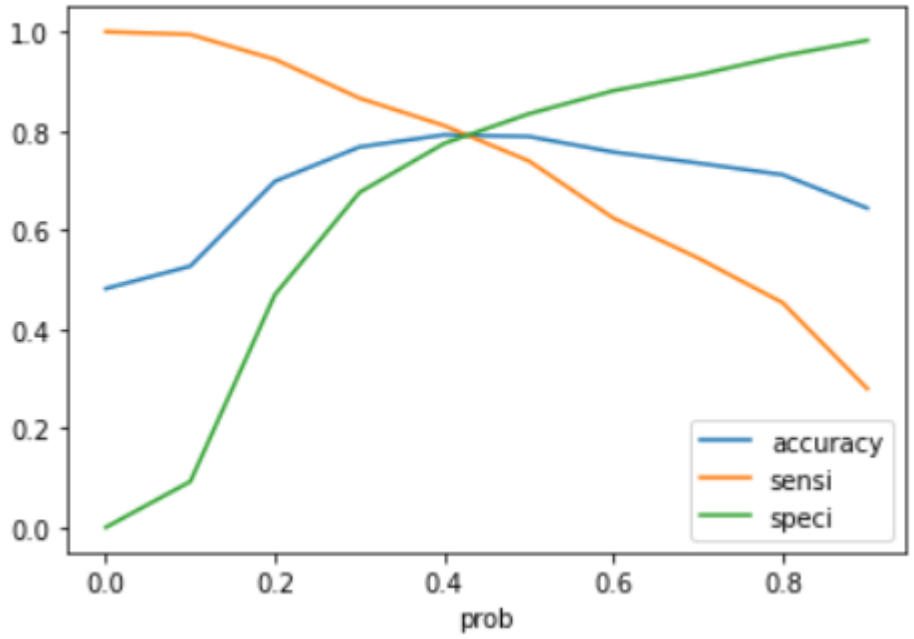
CORRELATION



MODEL EVALUATION

0.42 is the trade-off 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%

CONCLUSION

- The conversion rate for landing page and API submissions is 30–35%, which is almost typical. yet extremely low for Lead import and Lead Add form. We may thus step in and say that we should pay more attention to the leads that come from submitting landing pages and APIs.
- We may observe that Google and direct traffic create the greatest amount of leads. The Welingak website and reference provide the maximum conversion ratio.
- Leads who spent more time on website, more likely to convert.
- Most common last activity is email opened. highest rate = SMS Sent. Max are unemployed. Max conversion with working professional.

THANK YOU