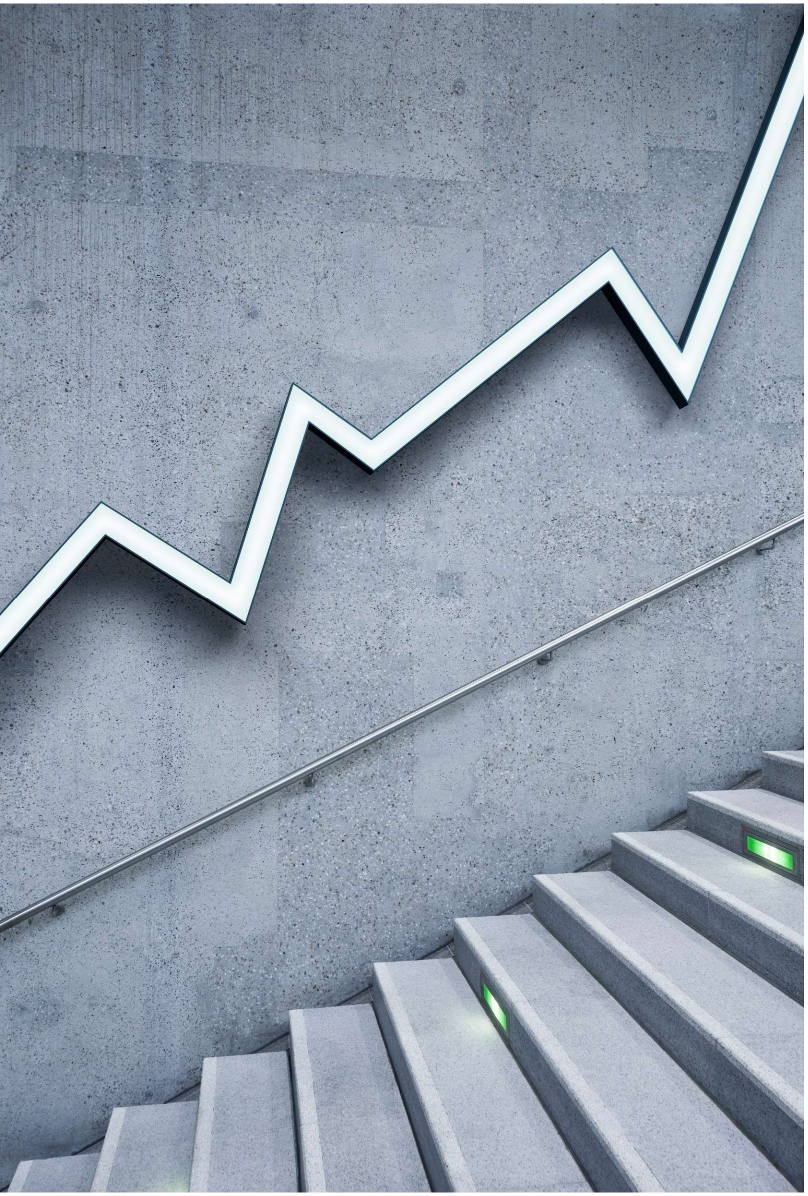


LEAD SCORING CASE STUDY

- G P SHIVA PRASAD
- CHENNIBABU DOGIPARTHI

Problem Statement & Goals

Problem Statement :

- X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals.
- 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%.

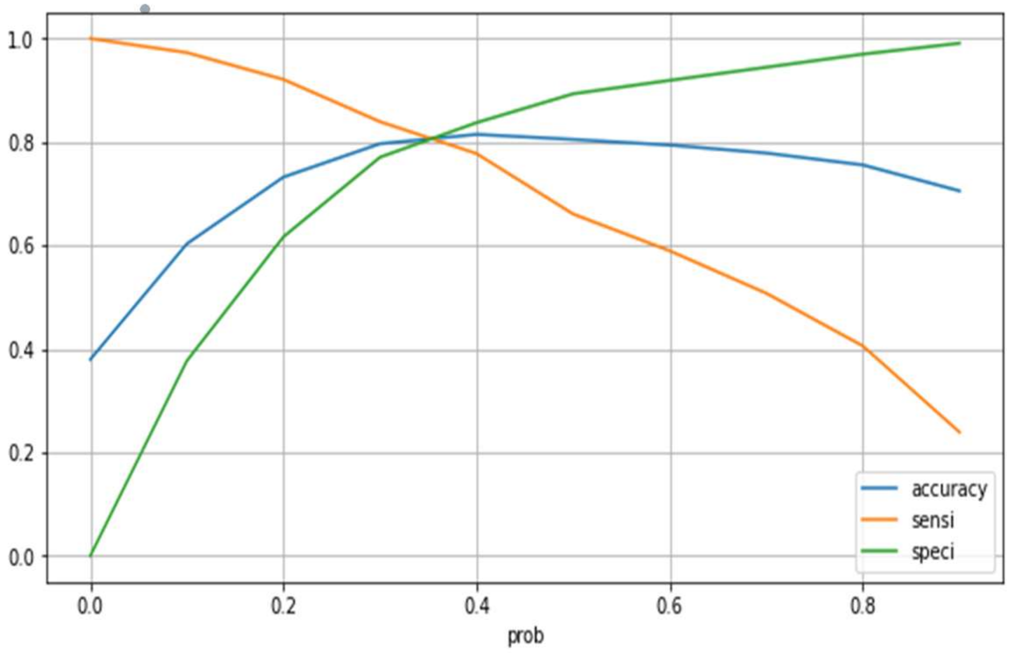
Business Goal:

- The company needs a model wherein a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

APPROACH

- Reading & Cleaning the data from null values, duplicate values and “Select” value which is a default value for some column and thus converted to null values and remained with around 98% rows left after removing the null values
- Handled the Outliers by capping the outliers (Highest outliers to 95 percentile and lowest outliers to 5 percentile).
- Dropped some columns which don’t be much useful for analysis after doing EDA on columns
- Converted Binary Variables to (yes/no) to 1/0 and created dummy variables for categorical columns.
- Performed Train-Test split with train size at 80% and performed Feature scaling using standardization
- Built a model
- Evaluate the model based on metrics such as Accuracy, Sensitivity, Specificity, Precision and Recall
- Predicting the test dataset with the model
- Create a new column “Lead Score” which gives score to the leads. Highest score means hot lead and lowest means lead most probably wont join the course

Determining the Threshold for probability



- Plotting the Accuracy, Sensitivity and Specificity of the model for various probabilities.
- From the chart, we can get the threshold value (optimum cut- off value) for the predictions
- The optimal cut-off (Threshold value) for the predictions is “0.34”

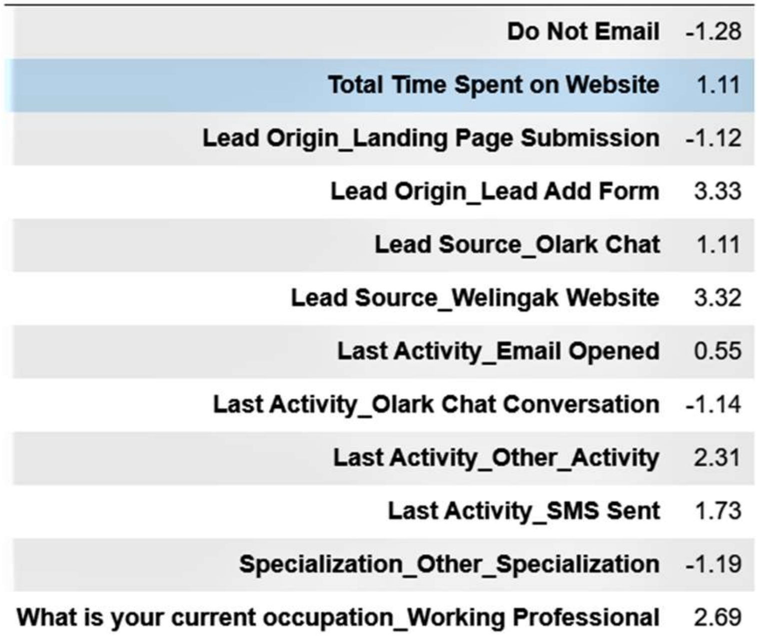
Model Performance
(Threshold Value = 0.34)

- On the Train Set
- Accuracy = 0.81
- Sensitivity = 0.82
- Specificity = 0.801
- Precision = 0.79
- Recall = 0.66
- F1 – Score = 0.72

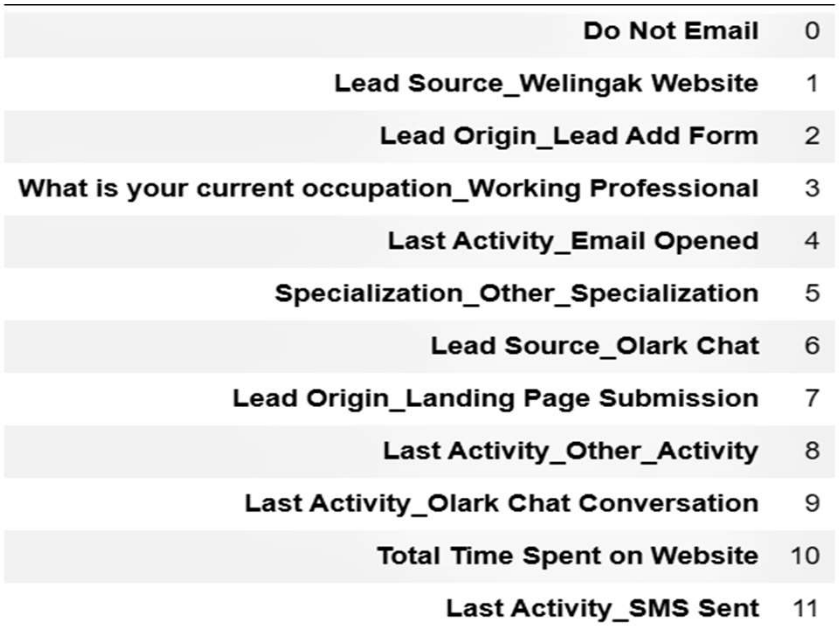
- On the Test Set
- Accuracy = 0.8
- Sensitivity = 0.815
- Specificity = 0.801
- Precision = 0.715
- Recall = 0.77
- F1 – Score = 0.74

Determining Feature Importance

(Feature with their Probability Coefficients(Left)/With Ranking representing the feature important(Right))

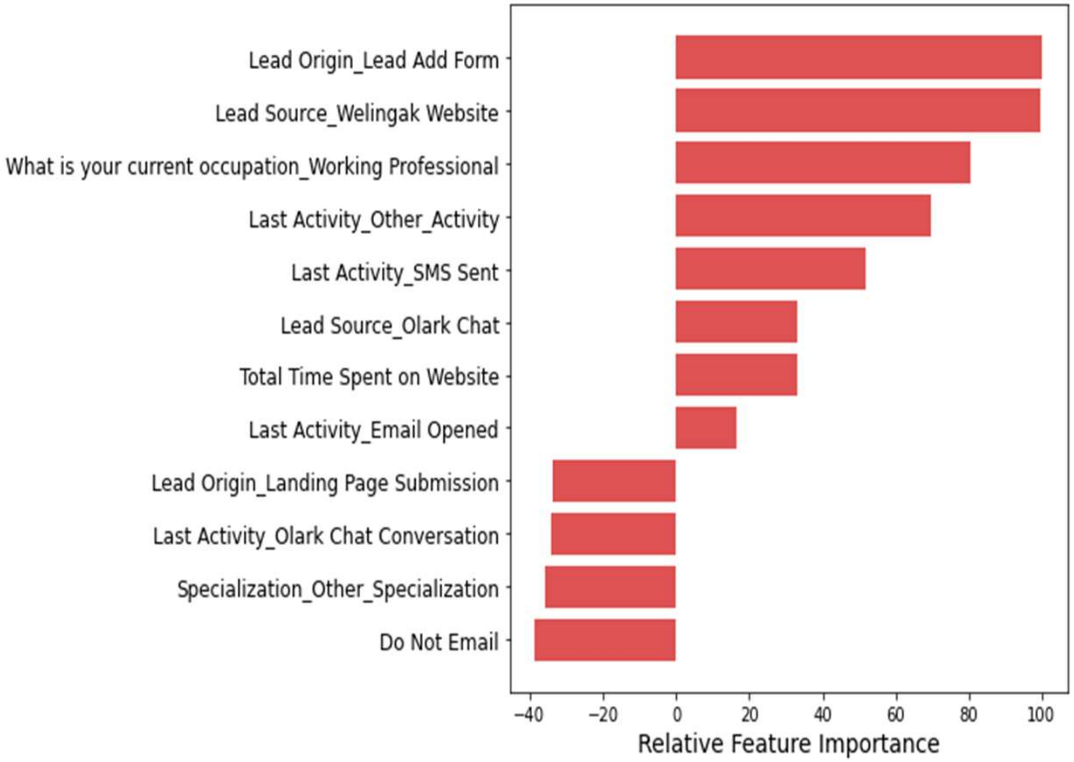


Do Not Email	-1.28
Total Time Spent on Website	1.11
Lead Origin_Landing Page Submission	-1.12
Lead Origin_Lead Add Form	3.33
Lead Source_Olark Chat	1.11
Lead Source_Welingak Website	3.32
Last Activity_Email Opened	0.55
Last Activity_Olark Chat Conversation	-1.14
Last Activity_Other_Activity	2.31
Last Activity_SMS Sent	1.73
Specialization_Other_Specialization	-1.19
What is your current occupation_Working Professional	2.69



Do Not Email	0
Lead Source_Welingak Website	1
Lead Origin_Lead Add Form	2
What is your current occupation_Working Professional	3
Last Activity_Email Opened	4
Specialization_Other_Specialization	5
Lead Source_Olark Chat	6
Lead Origin_Landing Page Submission	7
Last Activity_Other_Activity	8
Last Activity_Olark Chat Conversation	9
Total Time Spent on Website	10
Last Activity_SMS Sent	11

Recommendation



- Website Should be made more intuitive and engaging with more information added because it's a strong predictor with a strong positive coefficient

- Lead Add Form in Lead Origin should be given high importance

- Welingak Website in Lead Source is a reliable one in attracting new leads.

Thank You