LEAD SCORING CASE STUDY

Submitted by:
Asim Khan
Upasana Mishra
Namrata Dash

PROBLEM STATEMENT

An education company named 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%.

OBJECTIVE

The company requires us to build a model where we need to assign a lead score 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%.

STRATEGIES

- Import the data for analysis
- Clean and prepare the data
- Exploratory Data Analysis.
- Feature Scaling
- Splitting the data into Test and Train dataset.
- Building a logistic Regression model and calculate Lead Score.
- Evaluating the model by using different metrics -Specificity and Sensitivity or Precision and Recall.
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics.

VARIABLES

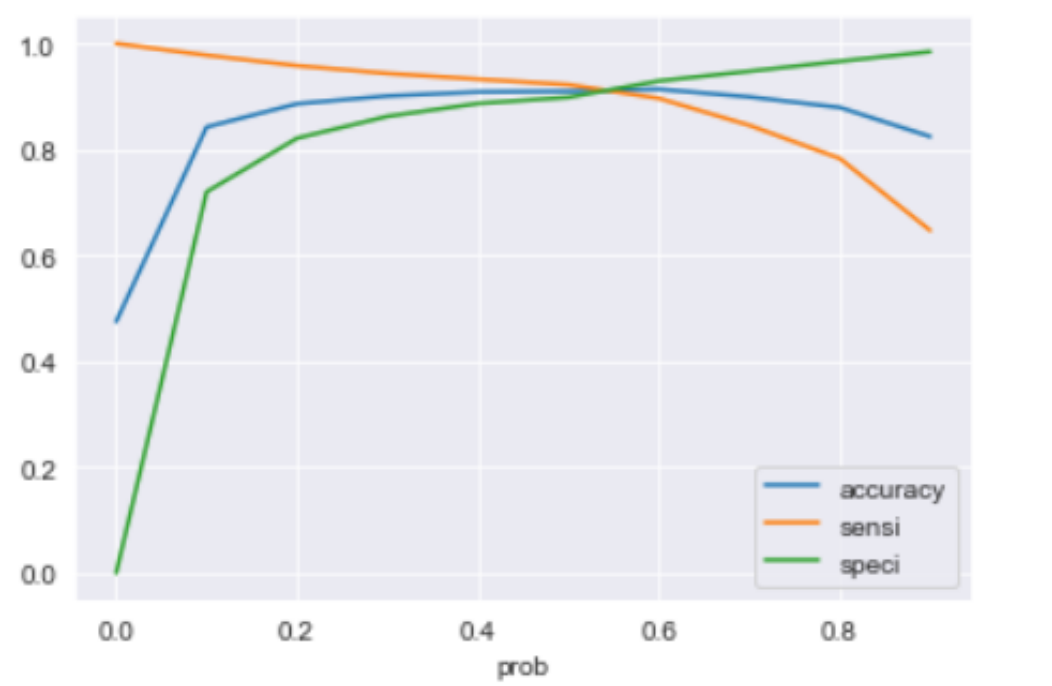
5

IMPACTING CONVERSION RATE

TotalVisits
Total Time Spent on Website
Lead Origin_Lead Add Form
Do Not Email_Yes
Last Activity_SMS Sent
Last Notable Activity_Modified
Last Notable Activity_Olark Chat Conversation
Tags_Busy
Tags_Lost to EINS
Tags_Ringing
Tags_Will revert after reading the email
Tags_switched off

MODEL EVALUATION

Sensitivity and Specificity on Training Data Set



The above graph shows an optimal cut off threshold of 0.55 based on Accuracy, Sensitivity and Specificity.

Confusion Matrix

2110	239
165	1950

- Accuracy – 91%
- Sensitivity – 92%
- Specificity – 90%
- False Positive Rate – 10%
- Positive Predicitive Value – 89%
- Negative Predicitve Value – 92%

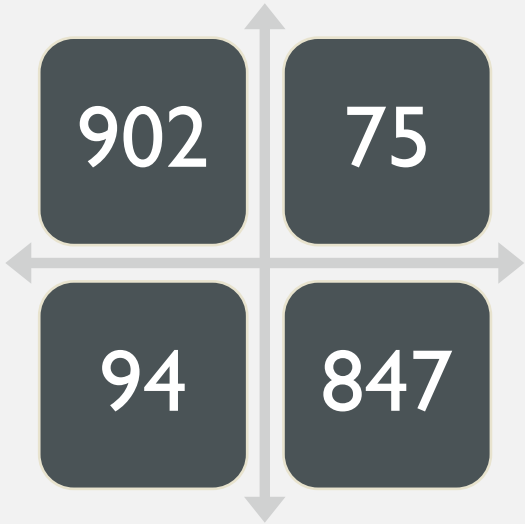
MODEL EVALUATION

7

Prediction on Test Data Set

Confusion Matrix

- Accuracy - 91%
- Sensitivity – 90%
- Specificity – 92%
- False Positive Prediction – 7%

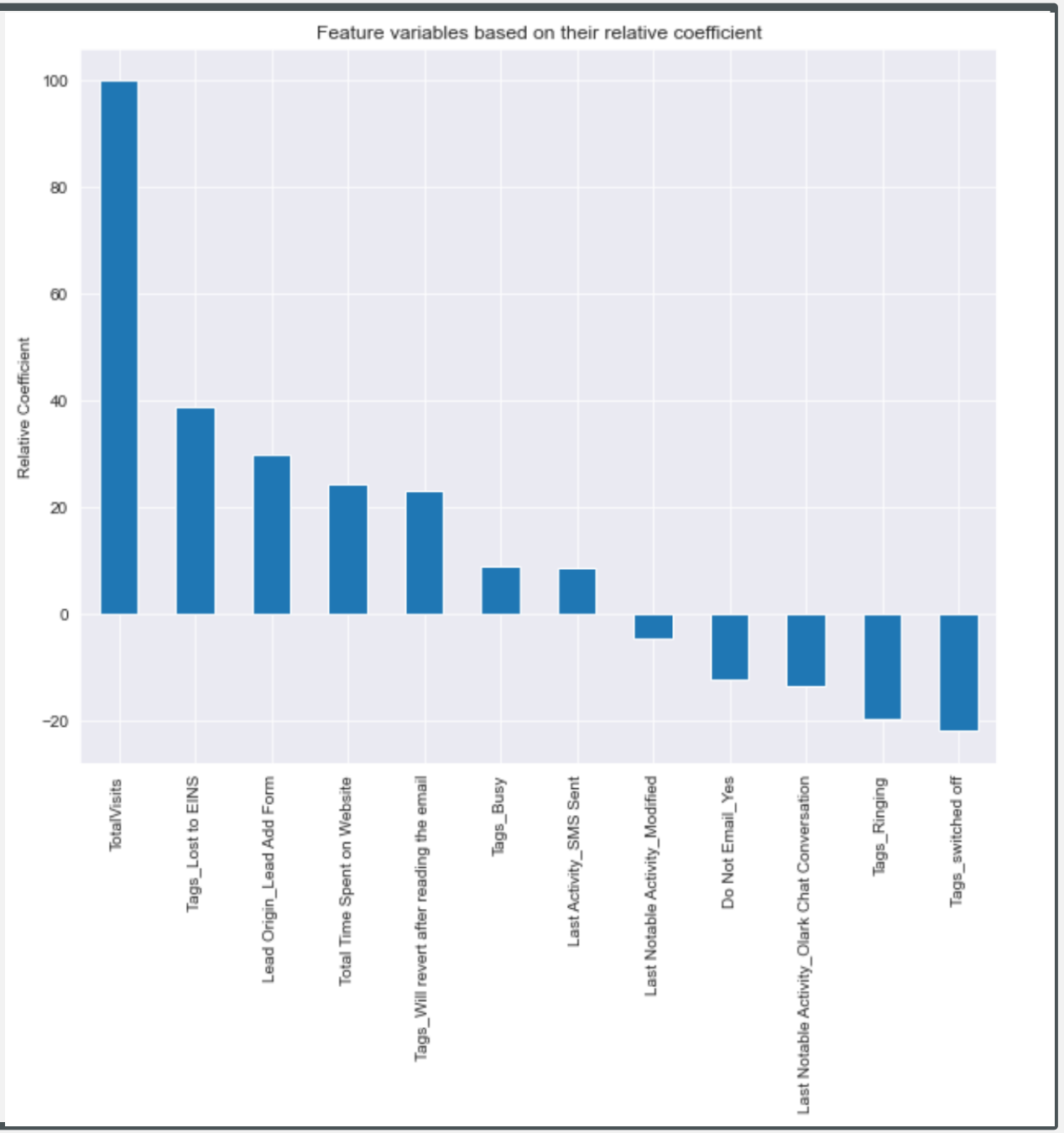


FEATURE VARIABLES

Feature Variables based on their coefficient values

8 Index values of the top 3 variables:

	index	0
0	TotalVisits	100.00
8	Tags_Lost to EINS	38.75
2	Lead Origin_Lead Add Form	29.90



TOP 3 FEATURES

TOP 3 FEATURES WHICH CONTRIBUTE MOST TOWARDS THE PROBABILITY OF A LEAD GETTING CONVERTED:

Top 3 variables that contributing to convert a lead are:

- TotalVisits
- Tags_Lost to EINS
- Lead Origin_Lead Add Form

Top 3 variables that need improvement to convert a lead are:

- Tags_switched off
- Tags_Ringing
- Last Notable Activity_Olark Chat Conversation

CONCLUSION

10

- Here we have compared both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction.
- The lead score calculated indicates the conversion rate on the final predicted model is around 89% (in train set) and 91% in test set.
- The top 3 variables that impact to lead getting converted in the model are:
  - TotalVisits
  - Tags_Lost_to EINS
  - Leas Origin_Lead Add Form
- Hence, overall this model seems to be good.

THANK YOU