Lead score case study

Submitted by :

Roopesh Radhakrishnan

Ambika kumari

Problem Statement :-

X Education sells online courses to industry professionals. any given day, many professionals who are interested in the courses land on their website and browse for courses.

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.

some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

Business Solution:-

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you 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% .

Approaches of Case study :-

- ❖ Reading data and understanding data
- ❖ Data cleaning
- ❖ Exploratory data analysis
- ❖ Data preparation / Dummy Creation
- ❖ Train-Test Split
- ❖ Feature Scaling
- ❖ Model Building
- ❖ Recursive feature elimination
- ❖ Plotting ROC Curve
- ❖ Finding optimal threshold
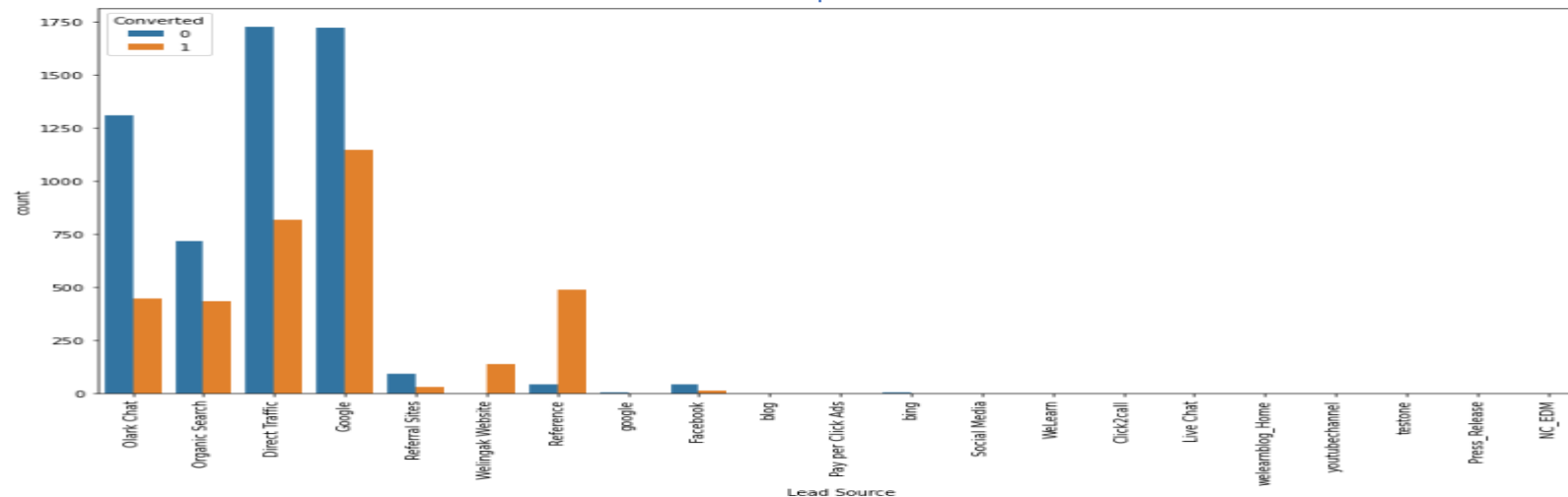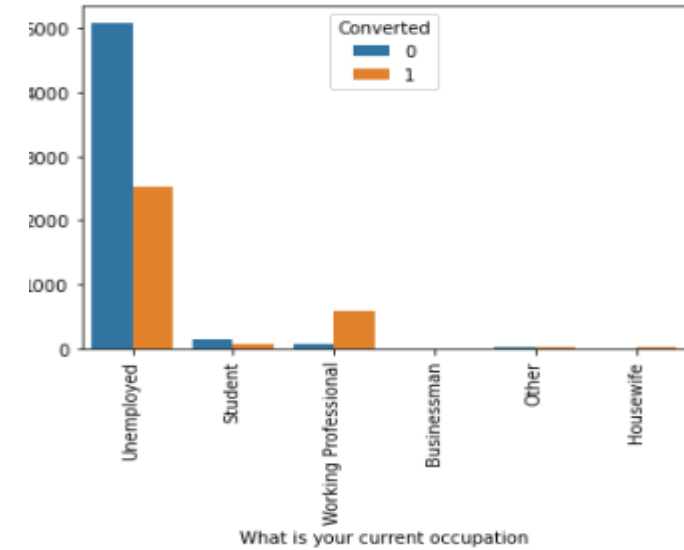- ❖ Making predictions on test set

EDA
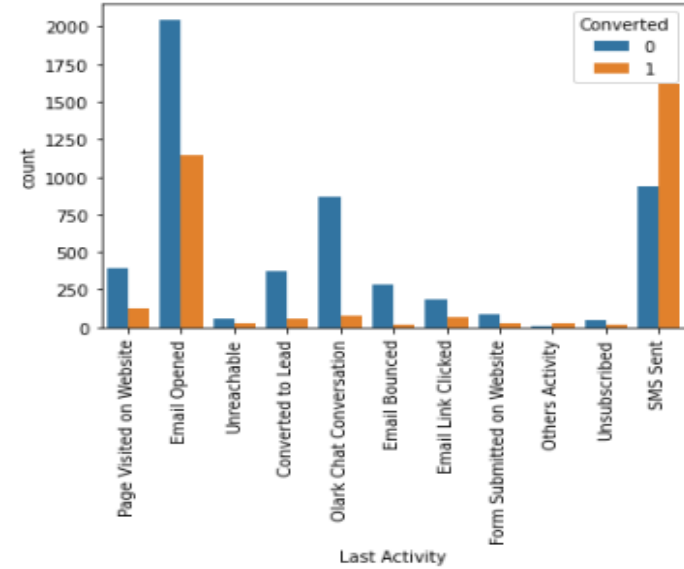
Maximum conversion in Last activity Email Opened and SMS sent.

Major conversion happened from do Not Email and Do not call made.
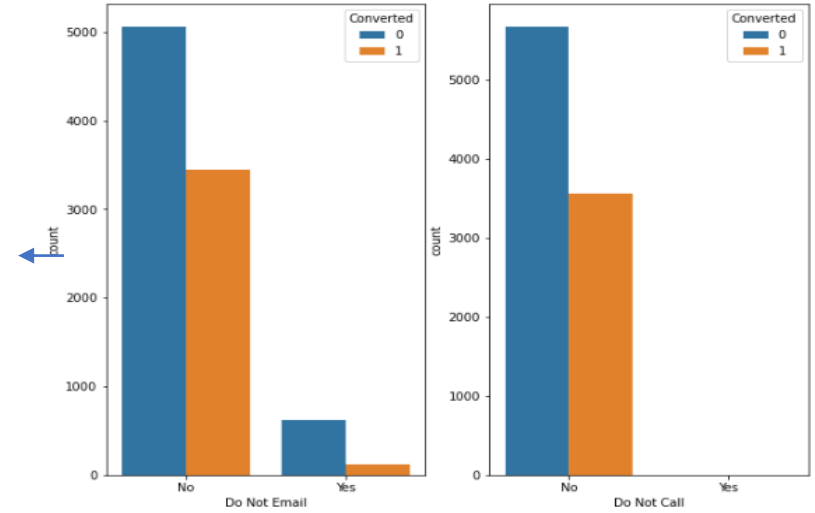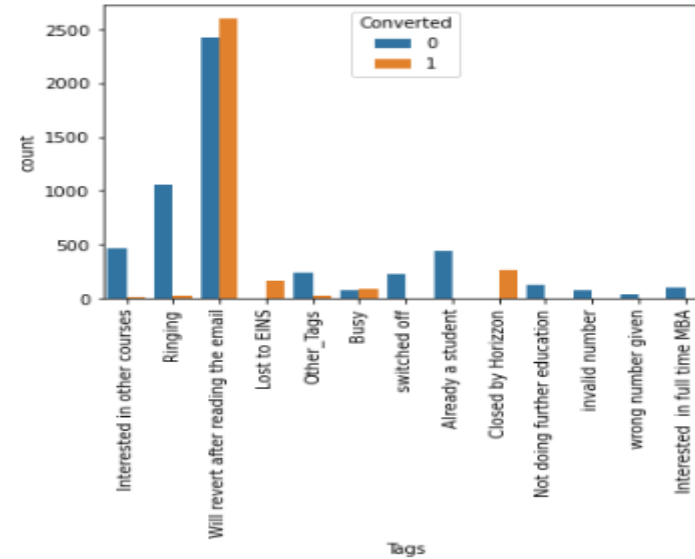
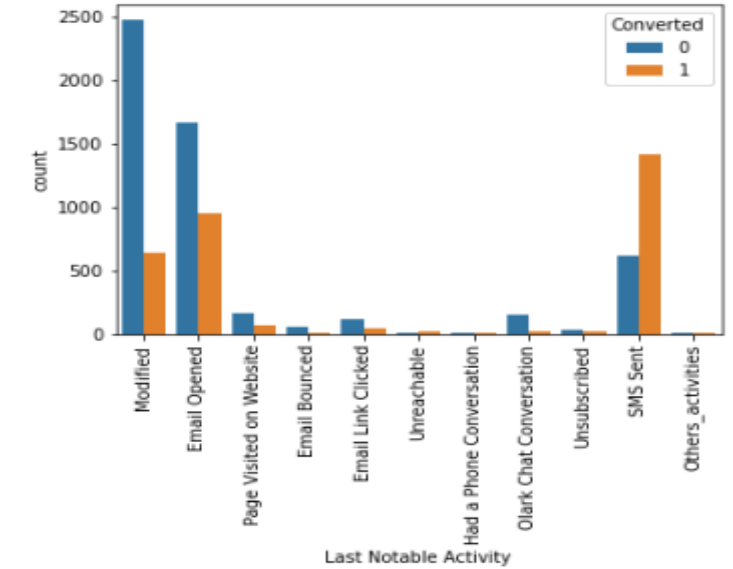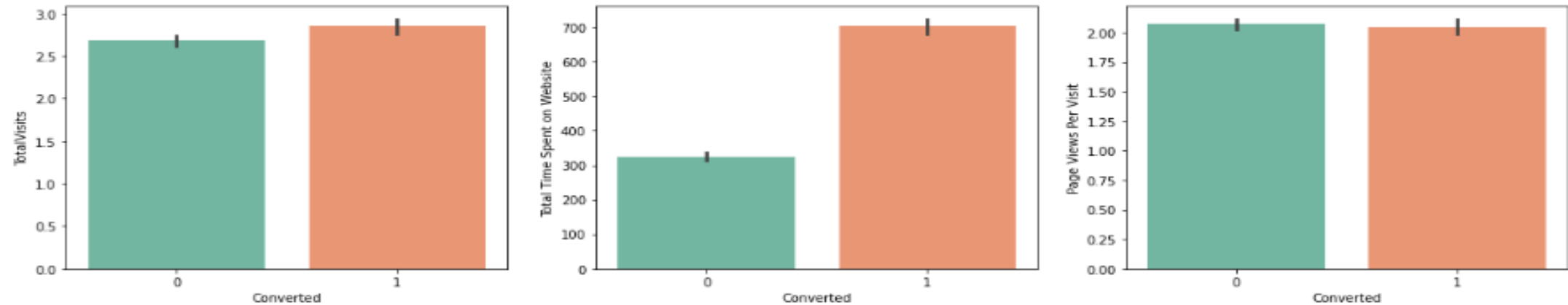Major conversion in lead source from google .

In the what is your current occupation major conversion is Unemployed .

In the tags columns major conversion is will revert after reading the email .

In the Last notable activity columns maximum conversion is in modified .

The conversion rated were high for Total Visits, Total Time Spent on Website and Page Views Per Visit
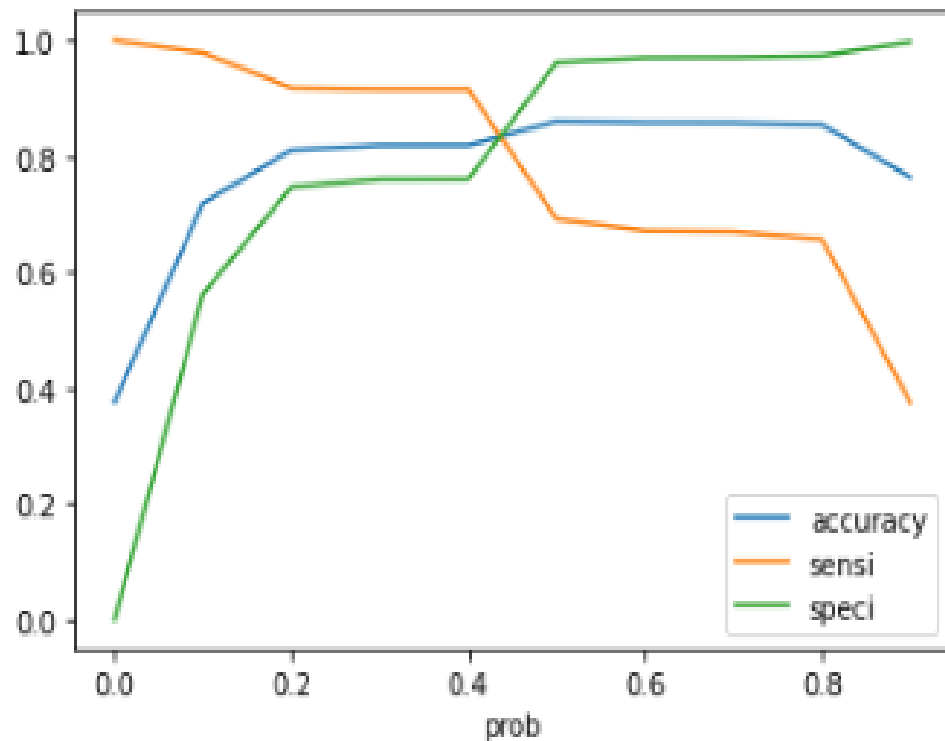
In the data set conversion rated is very high for Total visits , Total time spent on website and pages views per visits .

ROC Curve

Receiver operating characteristic example
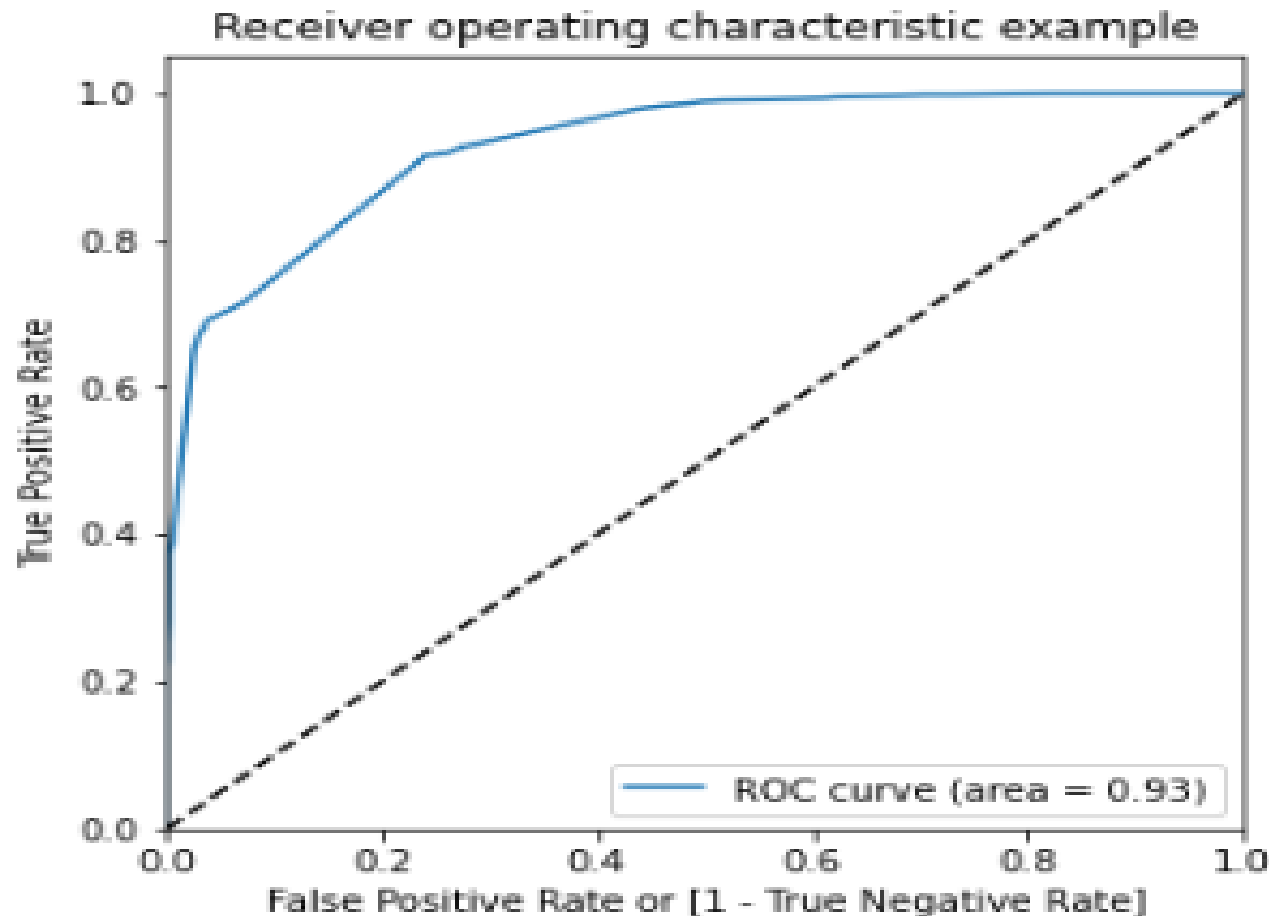
- The Roc curve Shows that 93 % of area is under the curve.

- The classification of probability of lead conversion (1/0) is very high by the model .
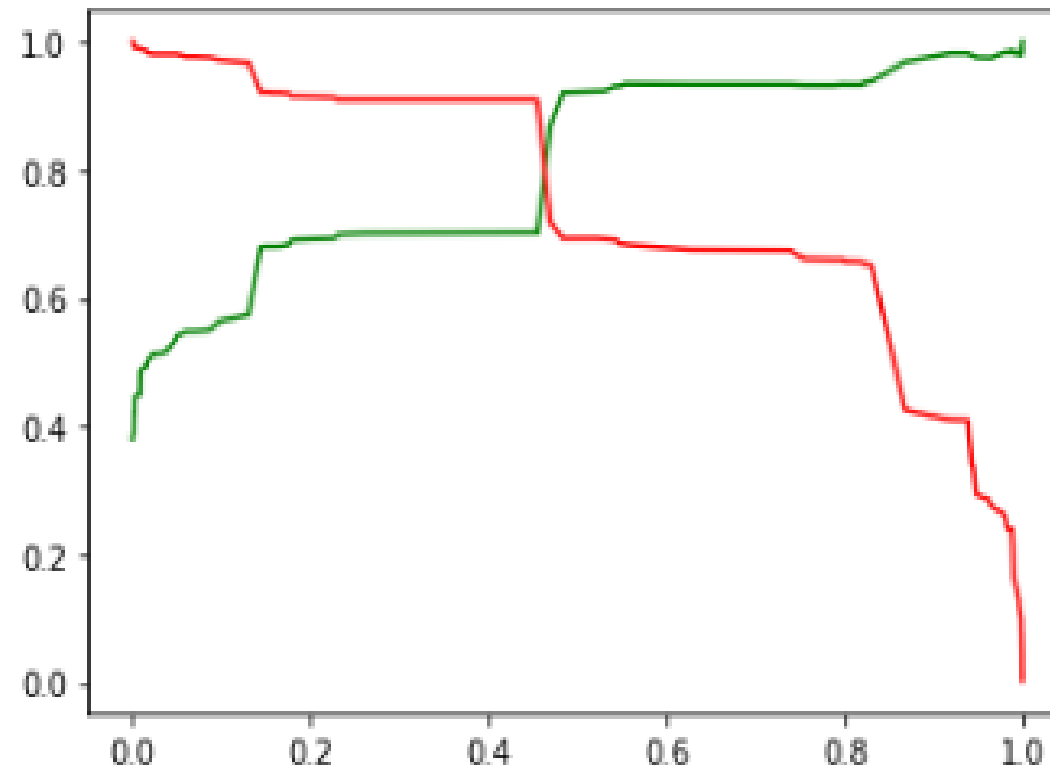
Finding optimal threshold

- The line of Accuracy, sensitivity, specificity intersect on the point is 0.42 probability and the point is called optimum cut off threshold probability.
- With Optimum point the model
  - Accuracy = 86%
  - Specificity = 76%
  - Sensitivity = 91%
  - Precision = 69%
  - Recall = 91%

Confusion Matrix on test data set

Confusion Matrix

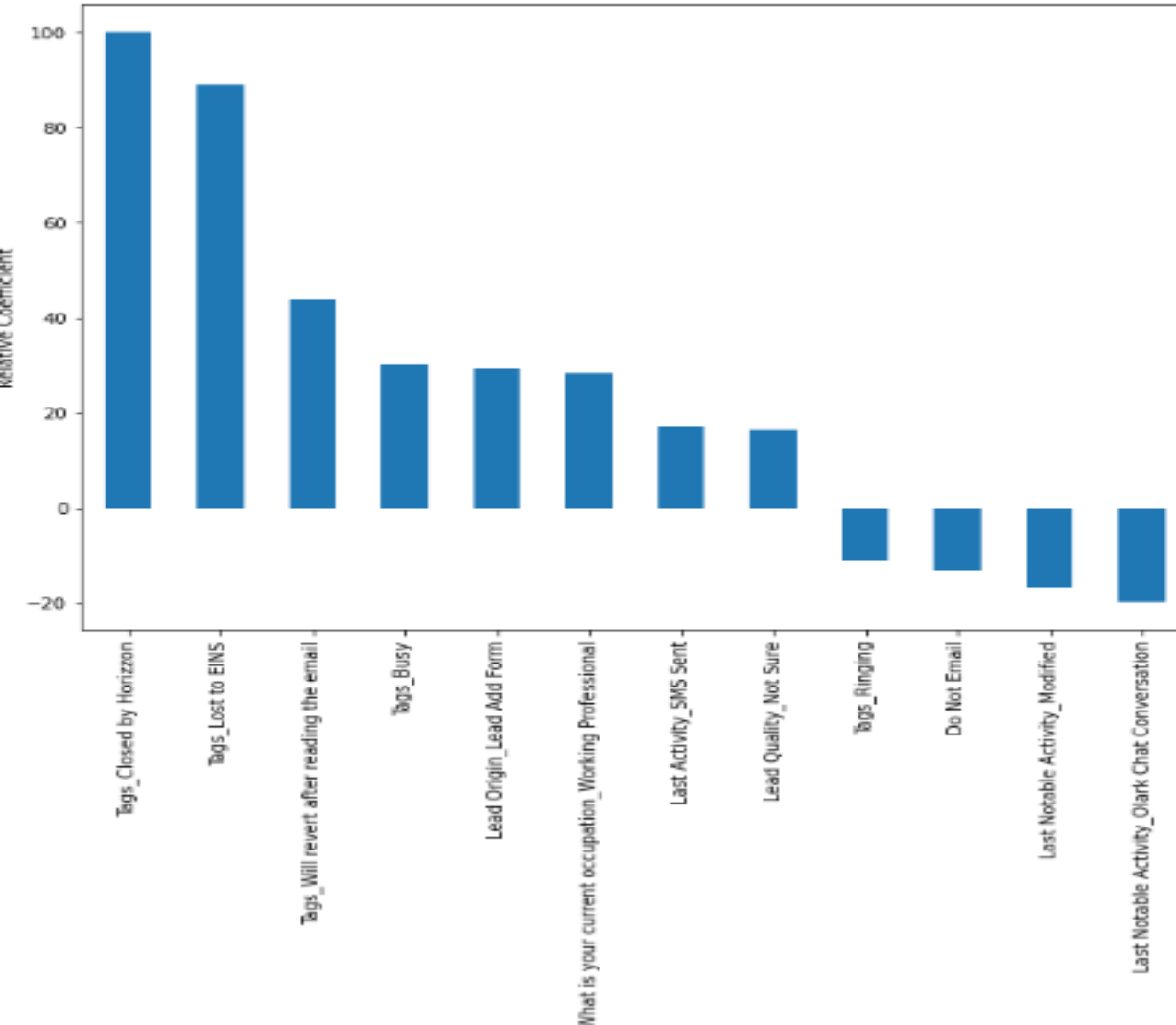| 1208	372
86	877

Accuracy = 81%

Specificity = 76%

Sensitivity = 91%

Precision = 70%

Recall = 91%

Important features

Feature variables based on their relative coefficient

we have checked sensitivity, specificity as well as precision and Recall and optimal point also considered. .cut off based on sensitivity , specificity for calculating final prediction .

Top 3 variables are converted as a lead:
. Tags_Closed by Horizzon
. Tags_Lost to EINS
. Tags_Will revert after reading the email

Conclusion

❖ We have checked Specificity , sensitivity , as well as precision , recall metrics we have considered optimum cutoff based on sensitivity , specificity for calculating the final calculating .

❖ we have checked sensitivity, specificity as well as precision and Recall and optimal point also considered. .cut off based on sensitivity , specificity for calculating final prediction .

❖ Top 3 variables are converted as a lead:
   Tags_Closed by Horizzon
   Tags_Lost to EINS
   Tags_Will revert after reading the email

❖ Top 3 variables have need impovement these are:-
   Do not Email
   Last Notable Activity_modified
   Last Notable Activity_Olark_chat_conversation

Thank you