



Lead Scoring Case Study

Contribution:

Sourabh Hulawale

Baisali Mukherjee



Agenda

- The purpose is to optimize the lead scoring mechanism based on their fit, demographics, behaviours, buying tendency etc., by implementing explicit & implicit modelling with lead point system,

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.
- 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.
- there are a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers from the bottom. In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion.
- 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 a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.

Goals

- There are quite a few goals for this case study.
- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot that is is most likely to convert whereas the lower score would mean that the lead is cold and will mostly not get converted. there are some more problems presented by the company which your model should be able to adjust to if the companies requirement changes in the future so you will need to handle these as well.



Approach

- Source the data for analysis
- Reading and understanding the data
- Data cleaning
- EDA
- Feature scaling
- Splitting the data into test and train data set
- Prepare the data for modeling
- Model building
- Model evaluation: specificity and sensitivity or precision recall
- Making prediction on the test set

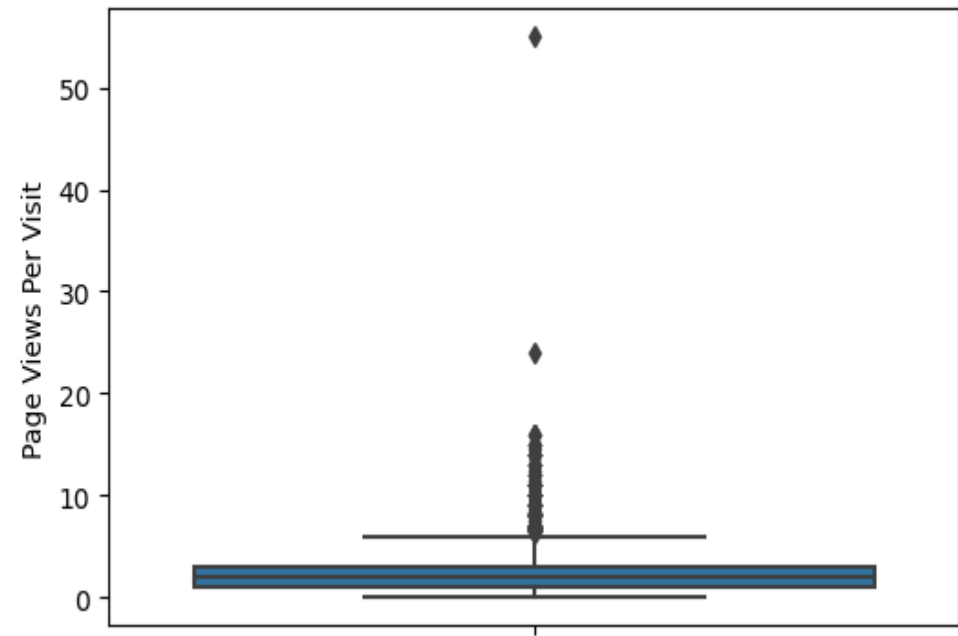
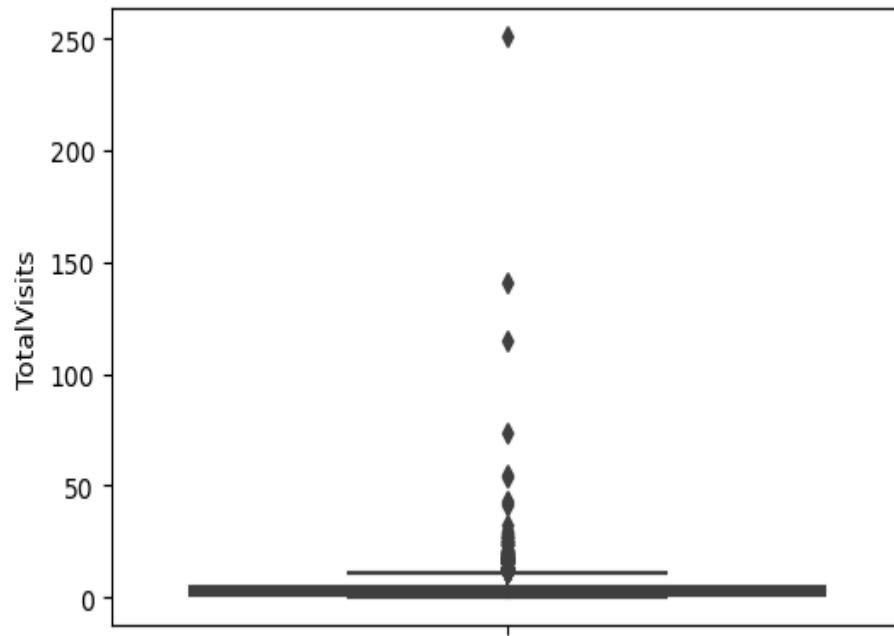


Data Sourcing Cleaning & Preparation

- Read the data from csv file
- Outlier treatment
- Data cleaning - handling null values and removing higher null values data
- Imputing null values
- Exploratory data analysis
- Feature standardization

Outlier handling

- Total visits, page views per visit have outliers: After checking percentiles, removed top & bottom 1% of the column outlier values from the data



Data Preparation

- Converted binary variables into 0 & 1
- Created dummy variables for categorical variables

```
In [101]: leads.head()
```

```
Out[101]:
```

	Converted	TotalVisits	Total Time Spent on Website	Page Views Per Visit	Lead Origin_Landing Page Submission	Lead Origin_Lead Add Form	Lead Origin_Lead Import	What is your current occupation_Housewife	What is your current occupation_Other	What is your current occupation_Student	...	Last No Activity_ Conver:
0	0	0.0	0	0.0	0	0	0	0	0	0	...	
1	0	5.0	674	2.5	0	0	0	0	0	0	...	
2	1	2.0	1532	2.0	1	0	0	0	0	1	...	
3	0	1.0	305	1.0	1	0	0	0	0	0	...	
4	1	2.0	1428	1.0	1	0	0	0	0	0	...	

5 rows × 57 columns



Feature Scaling & Splitting Train & Test Datasets

- Feature scaling of numeric data
- Splitting data into train & test sets



Model Building

- Feature selection using RFE
- Determined optimal model using logistic regression
- Calculated accuracy, sensitivity, specificity, precision & recall to evaluate model

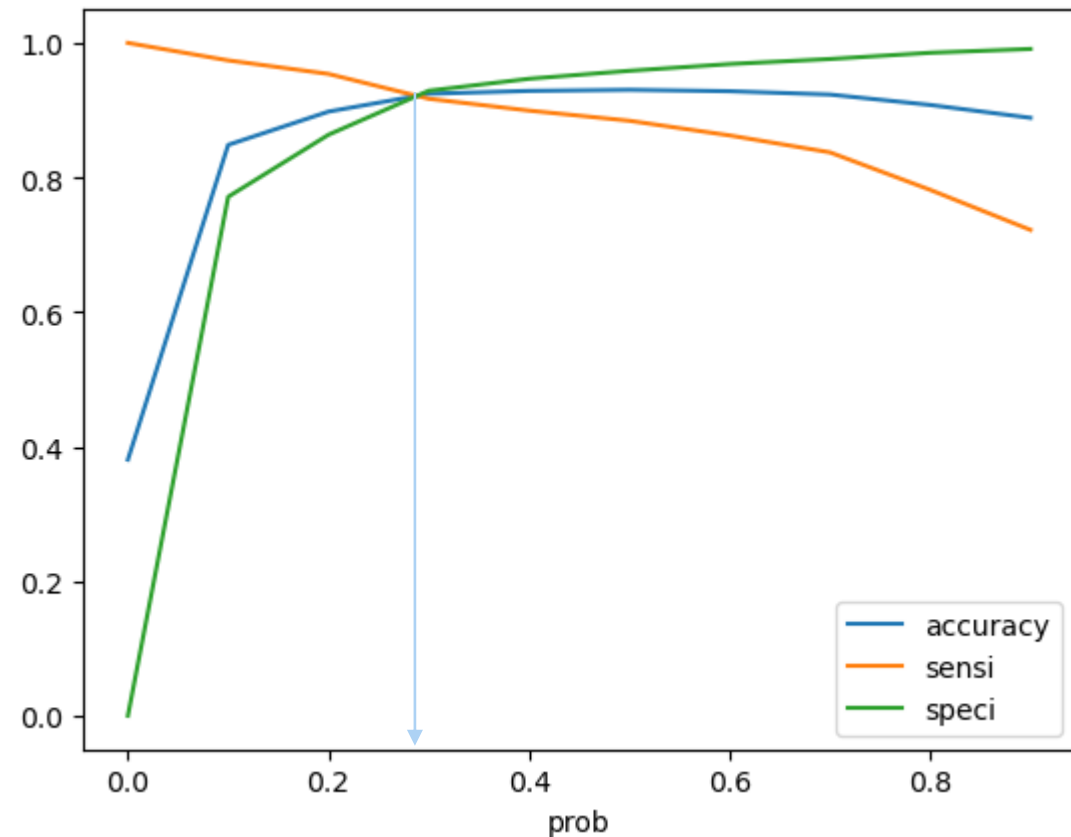


Variables impacting the conversion rate

- Total Visits
- Total Time Spent on Website
- Lead Origin_Lead Add Form
- Last Activity_SMS Sent
- What is your current occupation_Working Professional
- Tags_Closed by Horizzon
- Tags_Lost to EINS
- Tags_Will revert after reading the email
- Lead Source_Welingak Website

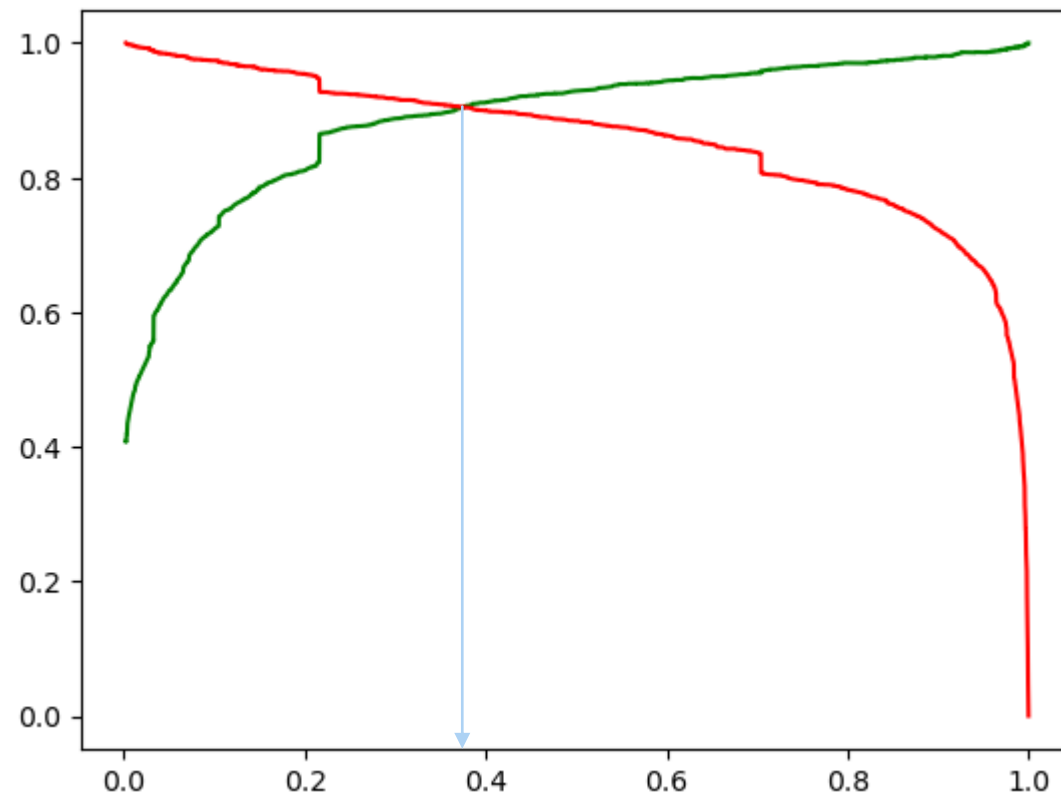
Model Evaluation: Sensitivity & Specificity on Train Dataset

- Graph depicts an optimal cut off of 0.3 based on Accuracy, Sensitivity & Specificity.
- - Accuracy: 92.16%
- - Sensitivity: 91.99%
- - Specificity: 92.27%



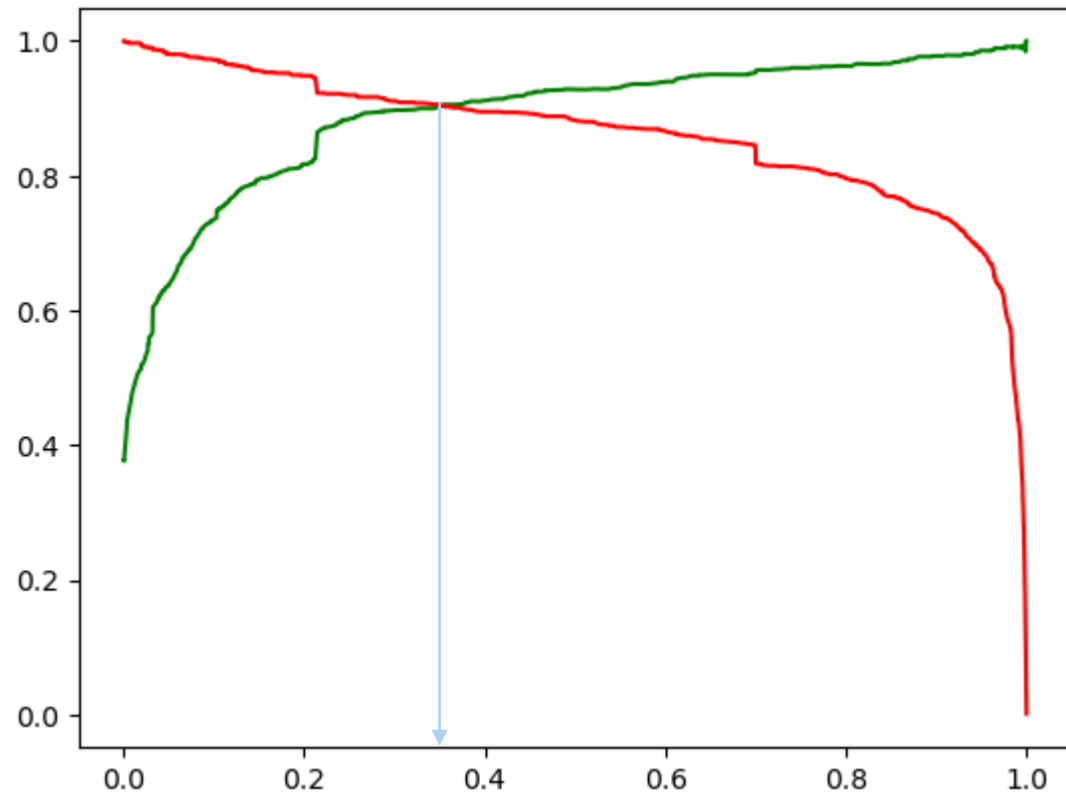
Model Evaluation: Precision & Recall on Train Dataset

- The graph depicts optimal cut off of 0.4 based on precision & recall
- Precision: 88.79%
- Recall: 91.69%



Model Evaluation: Sensitivity & Specificity, Precision & Recall on Test Dataset

- Accuracy: 92.59%
- Sensitivity: 91.38%
- Specificity: 93.31%
- Precision: 89.75%
- Recall: 91.08%



Result

- Accuracy sensitivity and specificity values of training and test set are close to training set
- Accuracy sensitivity and specificity values of training set are 92.16%, 91.99%, 92.27% respectively
- Accuracy, sensitivity and specificity values of test set are 92.59%, 91.38%, 93.31% respectively
- We have done the prediction on the test set using cut off threshold from sensitivity and specificity metrics

Conclusion

- While we have checked both sensitivity specificity as well as precision and recall metrics
- We have considered the optimal cutoff based on sensitivity and specificity for calculating the final prediction
- Accuracy sensitivity and specificity values of test data set are 92.59%, 91.38%, 93.31% respectively which are approximately closer to values of train data set
- Hence overall model seems to be good.

Summary

- There are a lot of leads generated in the initial stage but only a few of them come out as being customers from the bottom. In the middle stage, you need to nurture the potential leads well (educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion. First sort out the best prospects from the leads you have generated. 'Total visits', 'total time spent on website', 'page views per visit' which contribute most towards the probability of a lead getting converted. Then you must keep a list of leads handy so that you can inform them about new courses, services, job offers and future higher studies. Monitor each lead carefully so that you can tailor the information you send to them. Carefully provide job offerings information or courses that suits best according to the interest of the leads. A proper plan to chart the needs of each lead will go a long way to capture the lead as prospects. Focus on converted leads. Hold question answer sessions with leads to extract the right information you need about them. Make further enquiries and appointments with the leads to determine their intention and mentality to join online courses.