Lead Scoring Case Study

BY: Manoj Kumar, S Syed Saddam, Vivek S, Zaki Mehdi

Business Objective

To help X Education select most promising Leads(Hot Leads), the leads that are most likely to convert into paying customers

- Selection of the hot leads
- Focused marketing
- Higher lead conversion rate

Goals of the Case Study

Build a logistic regression model with a score range of 0 to 100 for each lead so that the business can target potential leads. A greater score, on the other hand, would suggest that the lead is hot and likely to convert, while a lower number would indicate that the lead is cold and unlikely to convert.

Approach

- Data understanding and exploring the data
- Data Cleaning and Preparation
- Preparation of Data
- Dummy variable creation
- Scaling
- Model Building
- Logistic Regression Model
- Model Evaluation
- Finding the Optimal Cutoff
- Making Predictions on Test set
- Precision-Recall View
- Precision and recall tradeoff
- Making Predictions on the Test Set

Data Cleaning and Preparation

- Cleaning the data using null values and removing outliers. Removing columns and rows with a significant number of null values, as well as unrelated columns that are no longer required for Exploratory Data Analysis.
- EDA was done on the cleaned data by plotting several types of plots and analysing both continuous and categorical variables, such as doing Univariate analysis against the target variable to discover the patterns.
- According to the findings, persons who spend more time on their jobs have the highest conversion rate.

Model Building

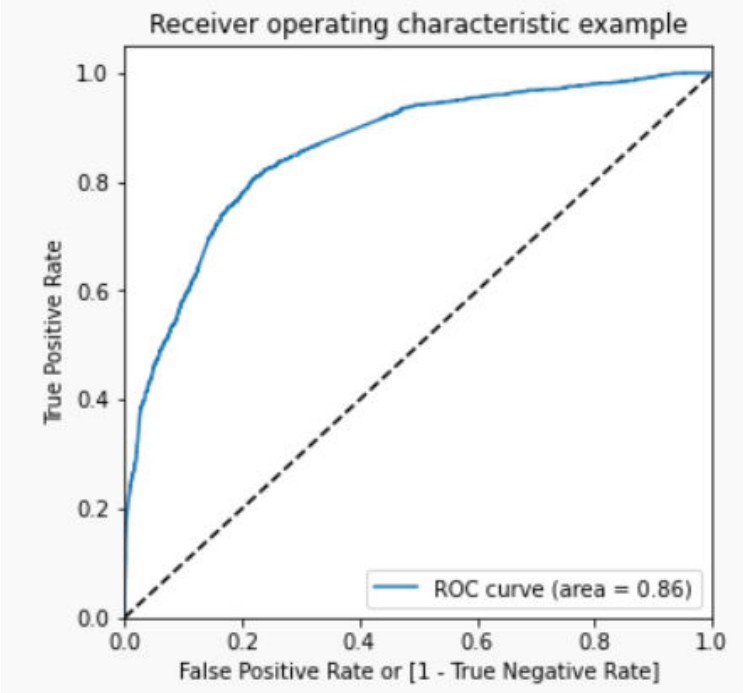
- For categorical columns, dummy variables were established and scaling was performed.
- The RFE technique was used for feature selection, and the processes for fethingthe columns with high value and VIF were followed. The final model is obtained when both the VIF and the p-values are within an acceptable range.

Model Features

Features	Coefficient
TotalVisits	11.3439
Total Time Spent on Website	4.4312
Lead Origin_Lead Add Form	2.9483
Lead Source_Olark Chat	1.4584
Lead Source_Reference	1.2994
Lead Source_Welingak Website	3.4159
Do Not Email_Yes	-1.5053
Last Activity_Had a Phone Conversation	1.0397
Last Activity_SMS Sent	1.1827
What is your current occupation_Housewife	22.6492
What is your current occupation_Student	-1.1544
What is your current occupation_Unemployed	-1.3395
What is your current occupation_Working Professional	1.2743
Last Notable Activity_Had a Phone Conversation	23.1932
Last Notable Activity_Unreachable	2.7868

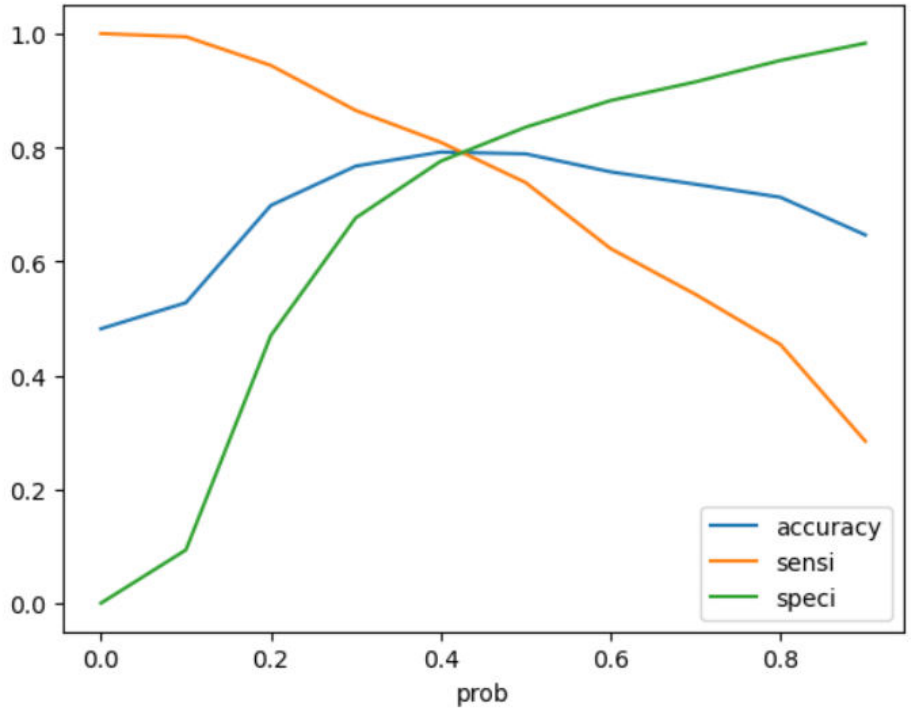
ROC Curve

The area under the curve of the ROC is 0.86 which is quite good. So we seem to have a good model. Let's also check the sensitivity and specificity tradeoff to find the optimal cutoff point.



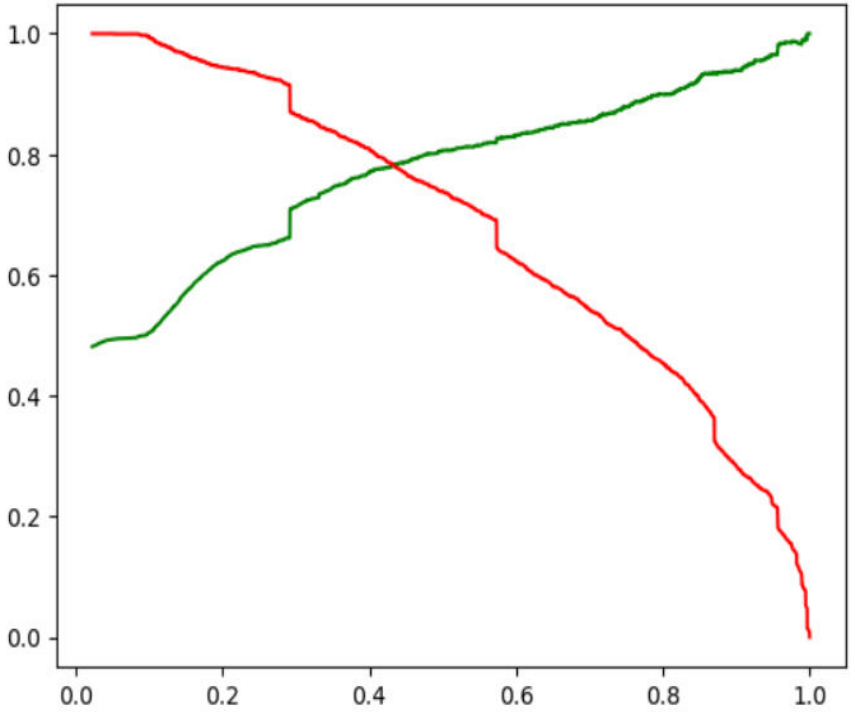
Model Evaluation Matrix

- The best projected likelihood for a customer is roughly 0.42.
- The model is accurate, sensitive, and specific.
- There is a good balance between accuracy, sensitivity, and specificity.



Precision and recall tradeoff

Precision-Recall threshold is as plotted in the diagram



Business Recommendations

Highly likely to be converted leads:-

1. Lead score more than 70.

2. Total time spent on website more than 12 hrs.

Very less likely to be converted leads :-

1. Customers opted for 'Do not email' option. Lead score less than 15.
2. Total time spent on website less than 5 hrs.
3. Lead source Direct Traffic, Referral Sites, Organic Search and Google.
4. Last activity of the customers is any of 'Olark chat conversation', 'page visited on the website', 'Email bounced', 'Form submitted on website', 'Email link clicked'.

Summary

Many leads are created in the first stage (top), but only a small percentage of them become paying visitors from the bottom. To obtain an advanced lead conversion, you must successfully nurture the eventuality leads in the middle stage (i.e. teaching the leads about the product, constantly communicating, etc.).

To begin, separate the trendy prospects from the leads you've generated.The criteria that have the most influence on the likelihood of a lead being converted are TotalVisits, Total Time Spent on Website, and Page Views Per Visit.

Maintain a list of leads so that you can alert them about new courses, services, career prospects, and future further studies. Cover each lead thoroughly to ensure that the information is consistent.

Thank You