



LEADS SCORING CASE STUDY

GROUP STUDY ASSIGNMENT

PRESENTED BY -

AANCHAL SHARMA

INDIRA PRIYADARSANI PRADHAN

INDU PRAKASH YADAV

PROBLEM STATEMENT

- X Education specializes in marketing online courses to professionals within various industries.
- X Education generates a significant number of leads; however, their ability to convert these leads into customers is currently suboptimal. To illustrate, if they amass 100 leads in a single day, only around 30 of them ultimately become paying customers.
- In an effort to streamline this process, the company aims to pinpoint the most promising leads, often referred to as 'Hot Leads.'
- If they successfully identify and focus on this particular group of leads, the conversion rate is expected to rise, as the sales team will dedicate more effort to engaging with potential leads, rather than making calls to every lead indiscriminately.

SOLUTION METHODOLOGY

Data cleaning & manipulation

- Check and handle duplicate data
- Check and handle NA values and missing values.
- Drop columns, if it contains a large number of missing values and are not useful for the analysis.
- Imputation of the values, if necessary.
- Check and handle outliers in data.

Exploratory Data Analysis (EDA)

- Univariate data analysis: value count, distribution of variables, etc.
- Bivariate data analysis: correlation coefficients and pattern between the variables etc.
- Feature Scaling & Dummy variables and encoding of the data.
- Classification technique: logistic regression is used for model making and prediction.
- Validation of the model, model presentation, conclusions and recommendations.

BUSINESS OBJECTIVE

➤ X Education is seeking to identify the most favorable leads.

➤ To accomplish this, they aim to construct a model that can recognize these 'hot leads.'

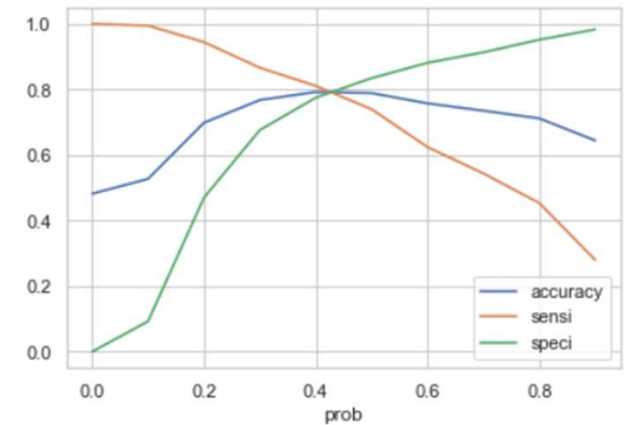
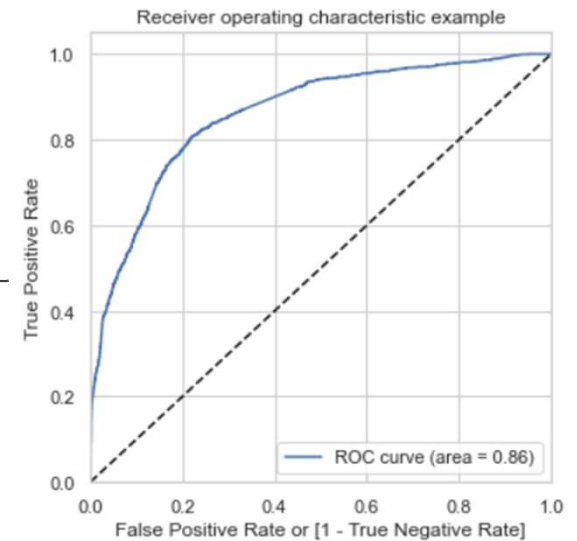
➤ Deployment of the model for the future use.

MODEL BUILDING

- Dividing the Data into Training and Testing Datasets
- The initial fundamental step in regression is to execute a train-test split, with a selected ratio of 70% for training data and 30% for testing data.
- Employing Recursive Feature Elimination (RFE) for Feature Selection
- Utilizing RFE to identify the top 15 variables as the output.
- Constructing the Model by eliminating variables with p-values exceeding 0.05 and variance inflation values (vi) greater than 5
- Making predictions on the test dataset and achieving an overall accuracy of 80%.

ROC Curve

- Finding Optimal Cut off Point
- Optimal cut-off probability is that Probability where we get balanced sensitivity and specificity.
- As you can see that around 0.42, you get the optimal values of the three metrics. So let's choose 0.42 as our cutoff now.



PREDICTION ON TEST SET

- Before making predictions, we standardized the test data and ensured that it had the same columns as our final training dataset.
- Following this, we proceeded to predict the test set, saving the new prediction values in a separate data frame.
- Subsequently, we conducted model evaluation, assessing accuracy, precision, and recall. The accuracy score was approximately 0.79, with precision and recall both around 0.77.
- These results indicate that our test predictions fall within an acceptable range for accuracy, precision, and recall.
- Furthermore, this suggests that our model is stable, demonstrating good accuracy and recall (sensitivity).
- We established a lead score on the test dataset to identify potential leads – the higher the lead score, the greater the likelihood of conversion, while a lower lead score suggests a reduced chance of conversion.

CONCLUSION

It was discovered that the most influential variables in identifying potential buyers, listed in descending order of importance, are as follows:

- 1.The total time spent on the website.
- 2.The total number of visits.
- 3.The lead source, specifically when it is Google, Direct traffic, Organic search, or the Welingak website.
- 4.The last activity, particularly SMS and Olark chat conversations.
- 5.The lead origin, particularly in the format of Lead ads.
- 6.The individual's current occupation as a working professional.

Considering these findings, X Education has a significant opportunity for success, as they are well-positioned to effectively persuade nearly all potential buyers to reconsider and purchase their courses