Summary : Lead Scoring Case Study

Problem Statement :

An education company named X Education get leads via various sources and methods. Their current conversion rate is approx. 30%, i.e. 30% leads get converted to real subscription for their program. Need to create a ML model which predicts and assign a Lead Score so that Sales team can target only "Hot Leads", i.e. which have extremely high chances of getting converted. Target is to achieve 80% positive conversion ratio by optimising resources.

Analysis Approach:

- Understand data and explore various features of it.

- Data cleaning
  - Handle missing values,
  - Impute for unwanted values
  - Handling outliers. Etc.

- Prepare the data for modelling
  - Handling imbalance in data
  - Converting categorical columns to Binary
  - Creating Dummy variables for multi-layered categorical columns
  - Split data to train-test
  - Scaling the data

- Building the model
  - Train the model
  - Test the model

Model Outcome :

We assigned a LEAD SCORE to each lead. A lead score tell the probability of lead getting converted in an actual subscription. Higher the lead score, more are chances of lead getting converted.

After analysis we found optimal probability cut-off as 0.35.

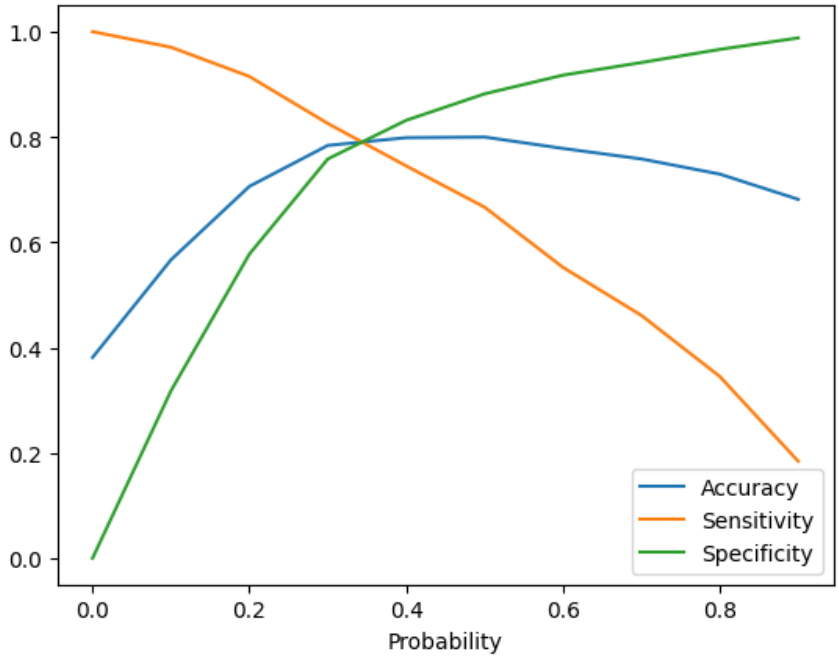
This means if lead score is more than 35, lead is likely to be converted.

Average lead score of converted leads => 69

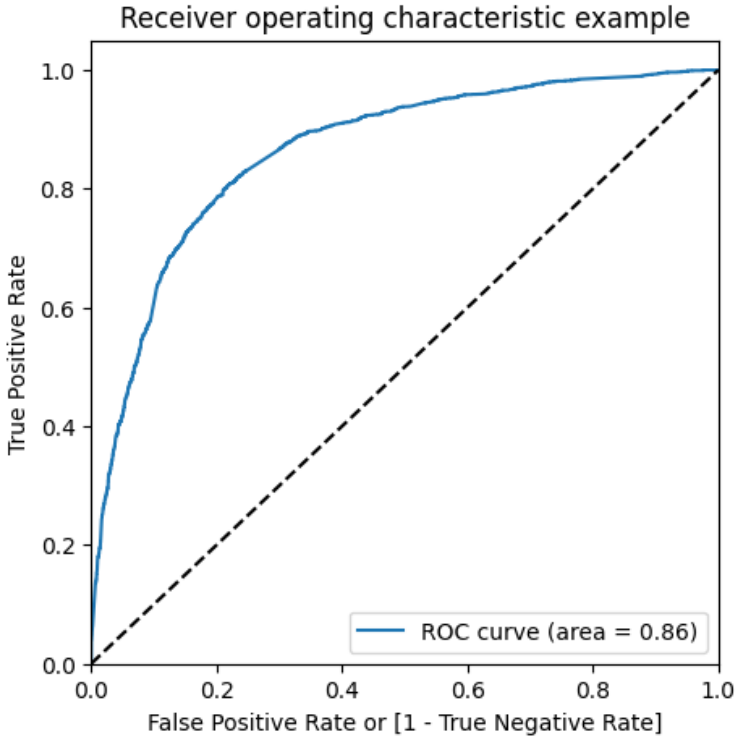
Average lead score of not-converted leads => 15

Important Measures of the Model:

Measure	Train Set	Test Set
Accuracy	0.79	0.79
Sensitivity	0.79	0.79
Specificity	0.79	0.80



Gini for the Model:



Conclusion :

The model demonstrates strong performance with high accuracy, sensitivity, and specificity. In summary, it effectively applies the knowledge gained from the training set to perform well in the test set.

Recommendations to Company :

Leads which to be considered as “Hot Lead” and to put effort by sales team to put resources on them :

- Leads coming from Welingak website.
- Leads referenced by someone.
- Leads who have spent considerable time (>12 Hrs) on X-Education Website.
- Leads who have spent SMS to the company.

Leads which are less likely to be converted are :

- Time spent on website less than 5 Hours
- Leads who have opted for “Do not email” or “Do not call” option.
- Last activity of the customers is any of ‘Olark chat conversation’, ‘page visited on the website’, ‘Email bounced’

In conclusion, the assignment offered practical exposure to various aspects of data processing, including data cleansing, exploratory analysis, preparation, modeling, and assessment. It underscored the significance of selecting suitable evaluation criteria aligned with the business challenge and navigating the trade-offs among different metrics. Lastly, the analysis shed light on factors influencing lead conversion rates and proposed strategies to enhance these rates.