Leads Scoring Case Study Presentation

Problem Statement

- Help X Education identify the most promising leads, the leads that are most likely to convert into paying customers
- Lead Conversion rate is 30%. X Education wants to increase this to 80%
- Build a model to identify the attributes that impact lead conversion
- Model should assign lead score to each lead 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

Analysis Approach

- Logistic Regression model is used to calculate the lead score
- Steps include
  - Data Understanding - Read and analyse the data
  - Data Cleaning and Preparation
    - Data type alignment
    - Null value (including Select) treatment
    - Incorrect entries treatment
  - EDA
  - Model Building
    - Dummy Variables
    - Scaling using StandardScaler
    - **Model building using Stats Model & RFE**
  - Model Evaluation on Train and Test data set
    - Key Metrics calculation (Accuracy, Sensitivity, Specificity
    - ROC Curve
    - Optimal Cutoff point

Model Result Business

- Model developed can help business to accurately predict the lead conversion with 92.5% accuracy
- Top attributes impacting positively impacting lead conversion are Tags (Closed by Horizzon, Lost to EINS, Will revert after reading the email), Lead Source (Welingak Website), Specialization, Lead Origin (Lead Add Form) and Total Time Spent on Website
- Top attributes negatively impacting lead conversion are Tags (switched off, Ringing, Already a student, Interested in other courses)
- Sales team should be provided leads which have positively impacting attributes. Leads with negatively impact attributes should be filtered. This will improve conversion rate and productivity of the sales team
- Few findings from EDA
  - Providing free copy of Mastering The Interview has no impact on lead conversion
  - Newspaper advertisement channel is not generating leads and can be stopped
  - Recommendations are one of the best source of leads with high conversion

Model Evaluation

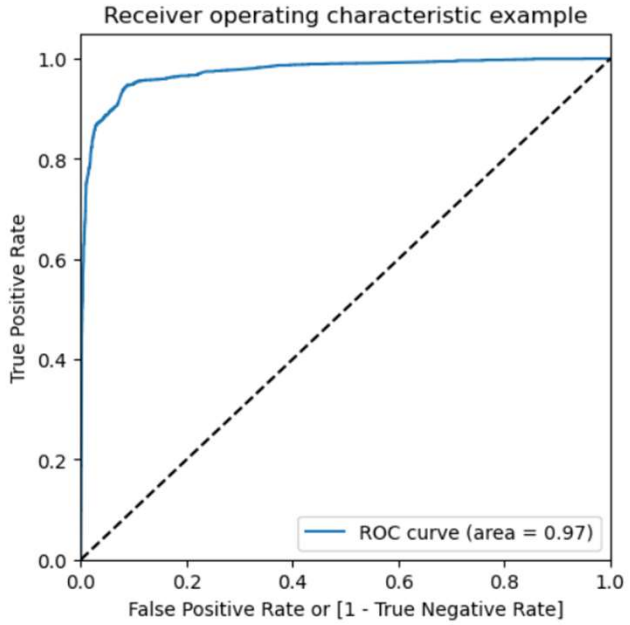
- Logistic Regression model can be deployed to help the business accurately predict outcomes (e.g., lead conversion) with 92.5% accuracy
- The high precision and recall for both classes ensure that the model minimizes both false positives (incorrectly predicting a lead would convert) and false negatives (missing out on actual lead conversions)
- A high ROC AUC score of 0.97 indicates the model can strongly differentiate between both classes, further enhancing its performance

Model Result Final Attributes

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.6303	0.090	-18.163	0.000	-1.806	-1.454
Total Time Spent on Website	0.9644	0.055	17.499	0.000	0.856	1.072
Lead Origin_Lead Add Form	1.2712	0.392	3.242	0.001	0.503	2.040
Specialization_Available	1.8821	0.141	13.355	0.000	1.606	2.158
Lead Source_Welingak Website	2.5625	0.846	3.028	0.002	0.904	4.221
Last Activity_SMS Sent	1.8767	0.117	15.976	0.000	1.646	2.107
Last Notable Activity_Modified	-1.6149	0.129	-12.489	0.000	-1.868	-1.361
Last Notable Activity_Olark Chat Conversation	-1.3958	0.442	-3.161	0.002	-2.261	-0.530
Tags_Already a student	-3.6965	0.612	-6.038	0.000	-4.896	-2.497
Tags_Closed by Horizzon	6.6630	0.741	8.987	0.000	5.210	8.116
Tags_Interested in other courses	-2.6609	0.376	-7.072	0.000	-3.398	-1.923
Tags_Lost to EINS	6.3305	0.733	8.633	0.000	4.893	7.768
Tags_Other_Tags	-2.2703	0.244	-9.320	0.000	-2.748	-1.793
Tags_Ringing	-3.8900	0.252	-15.448	0.000	-4.384	-3.396
Tags_Will revert after reading the email	4.6040	0.191	24.104	0.000	4.230	4.978
Tags_switched off	-4.1740	0.638	-6.544	0.000	-5.424	-2.924

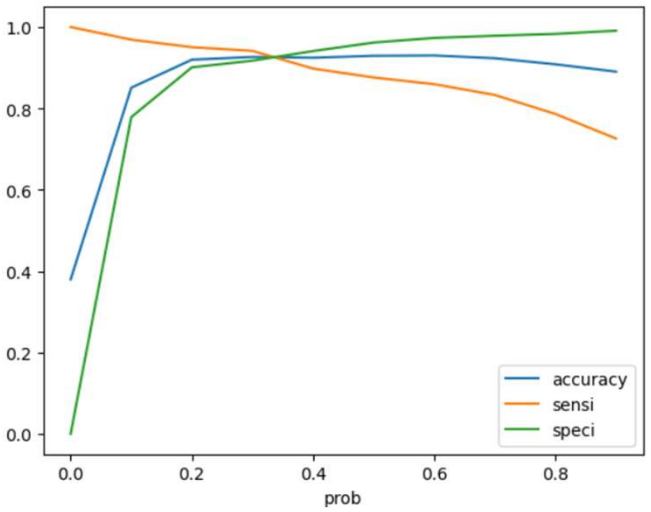
Model Evaluation Results

Metrics	Train Data	Test Data
Accuracy	92.63%	92.64%
Sensitivity	94.13%	93.28%
Specificity	91.73%	92.24%
Precision	87.44%	88.12%
Recall	94.13%	93.28%

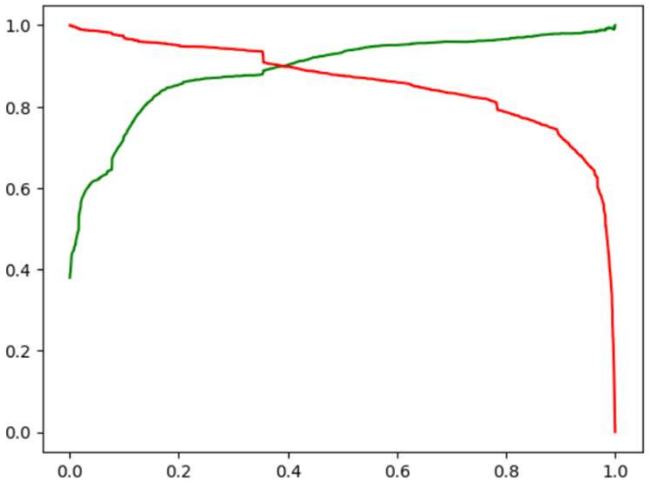


ROC value of 0.97

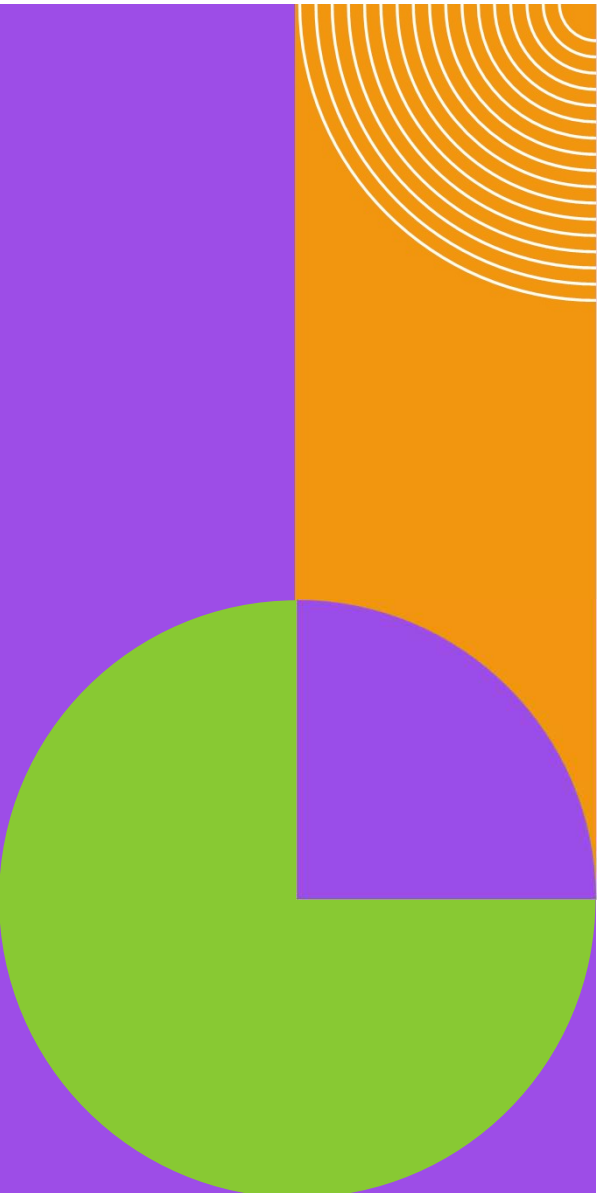
Model Evaluation



Optimal Cutoff Point 0.3



Precision Recall Curve



Lead Scoring and Business Insights

Lead Score: Assigned based on predicted probability of conversion, prioritizing high-potential leads.

Marketing Optimization:

- ✓ Target leads from high-converting lead origins and sources.
- ✓ Focus on users with high engagement on the website.
- ✓ Tailor campaigns based on user attributes and preferences.

Recommendations and Conclusion

Recommendations:

- Implement lead scoring in marketing campaigns.
- Personalize customer interactions based on lead score.
- Continuously monitor and refine model performance.

Conclusion:

- Lead scoring model effectively identifies high-potential leads.
- Enables data-driven decision-making for marketing optimization.
- Drives conversion rates and enhances business outcomes for X Education.