Lead Score Case Study

-By Keshav Mishra

Problem Statement

▶ X Education sells online courses to industry professionals.

▶ X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.

▶ To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.

▶ If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone

Business Objective:

▶ X education wants to know most promising leads

▶ For that they want to build a Model which identifies the hot leads.

▶ Deployment of the model for the future use.

Modelling Approach

- ▶ Importing the dataset.

- ▶ Understanding the data.

- ▶ Preparing the data for modelling.

- ▶ Selecting the logistic model for the target variable.

- ▶ Splitting the data to train and test dataset.

- ▶ Training the model on train dataset.

- ▶ Plotting the ROC curve to check the fit.

- ▶ Finding optimal cut-off point for converting conversion probability to converted.

- ▶ Testing the fit with test dataset.

Data Conversion

- Scaling Numerical Columns Using Robust Scaler.
- Categorical Columns are converting to Binary and Dummy Variables Using One-Hot Encoding.
- Total Rows for Analysis – 9204
- Total Columns for Analysis - 65

Model Building

▶ Splitting the Data into Training and Testing Sets.

▶ The first basic step for regression is performing a train-test split, I have chosen 70:30 ratio.

▶ Use RFE for Feature Selection, Running RFE with 15 variables as output.

▶ Building Model by removing the variable whose p-value is greater than 0.05 and vif value is greater than 5.

Final Model Output

	coef	std err	z	P>|z|	[0.025	0.975]
const	-4.3242	0.203	-21.311	0.000	-4.722	-3.927
Total Time Spent on Website	1.3534	0.086	15.713	0.000	1.185	1.522
Do Not Email_Y	-1.5346	0.261	-5.888	0.000	-2.045	-1.024
Lead Add Form	1.7526	0.388	4.516	0.000	0.992	2.513
Welingak Website	4.3063	1.087	3.960	0.000	2.175	6.438
SMS Sent	2.5925	0.126	20.623	0.000	2.346	2.839
Busy	2.6552	0.293	9.070	0.000	2.081	3.229
Closed by Horizzon	8.7408	1.030	8.485	0.000	6.722	10.760
Lost to EINS	7.3504	0.622	11.818	0.000	6.131	8.569
Not tagged	2.2578	0.204	11.066	0.000	1.858	2.658
Ringing	-1.2543	0.297	-4.230	0.000	-1.835	-0.673
Will revert after reading the email	6.5835	0.255	25.839	0.000	6.084	7.083
switched off	-1.7437	0.630	-2.767	0.006	-2.979	-0.509

All p-values are below the 0.05 threshold, indicating statistical significance

VIF(s) of the Final Model

	Features	VIF
2	Lead Add Form	1.77
10	Will revert after reading the email	1.69
4	SMS Sent	1.56
0	Total Time Spent on Website	1.42
3	Welingak Website	1.34
6	Closed by Horizzon	1.29
8	Not tagged	1.15
9	Ringing	1.10
1	Do Not Email_Y	1.06
5	Busy	1.04
7	Lost to EINS	1.04
11	switched off	1.04

All VIF values are below 5, indicating no significant multicollinearity
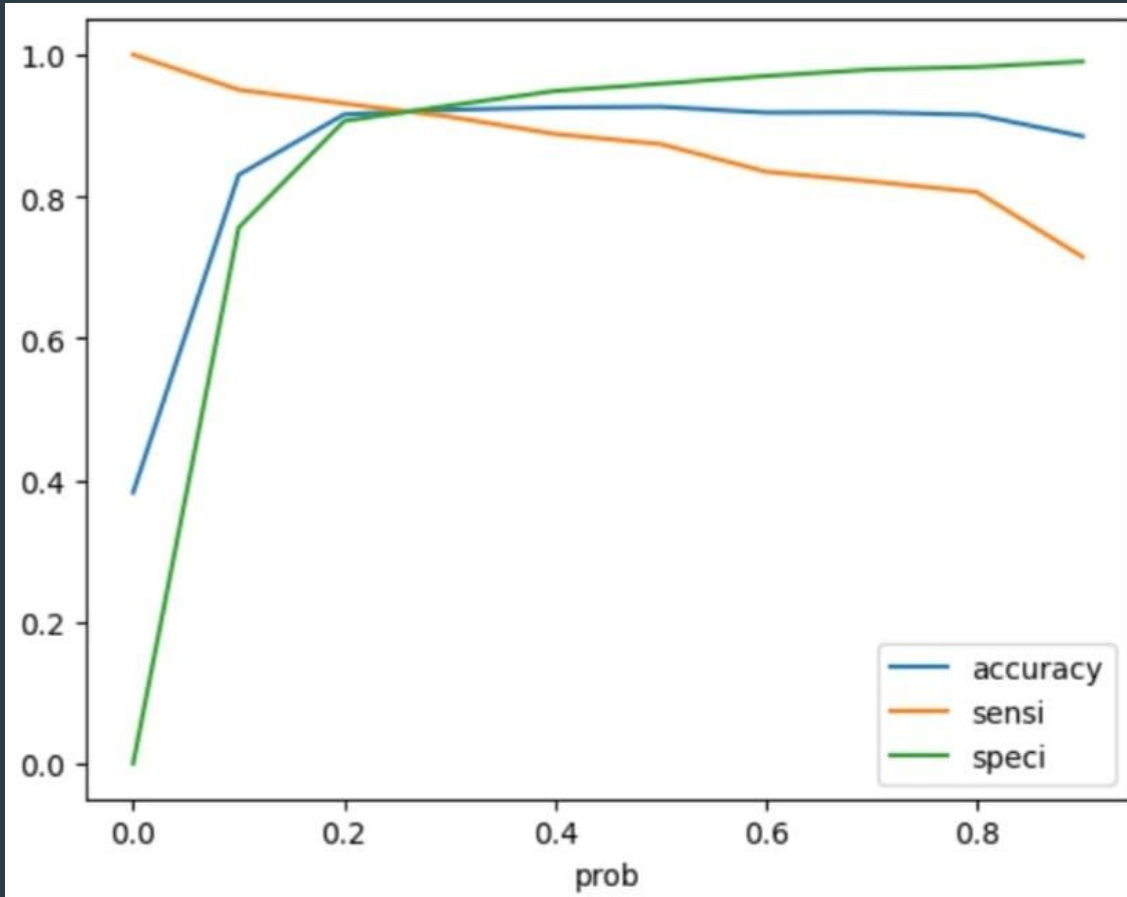
ROC CURVE
Trade-off between Specificity and Sensitivity
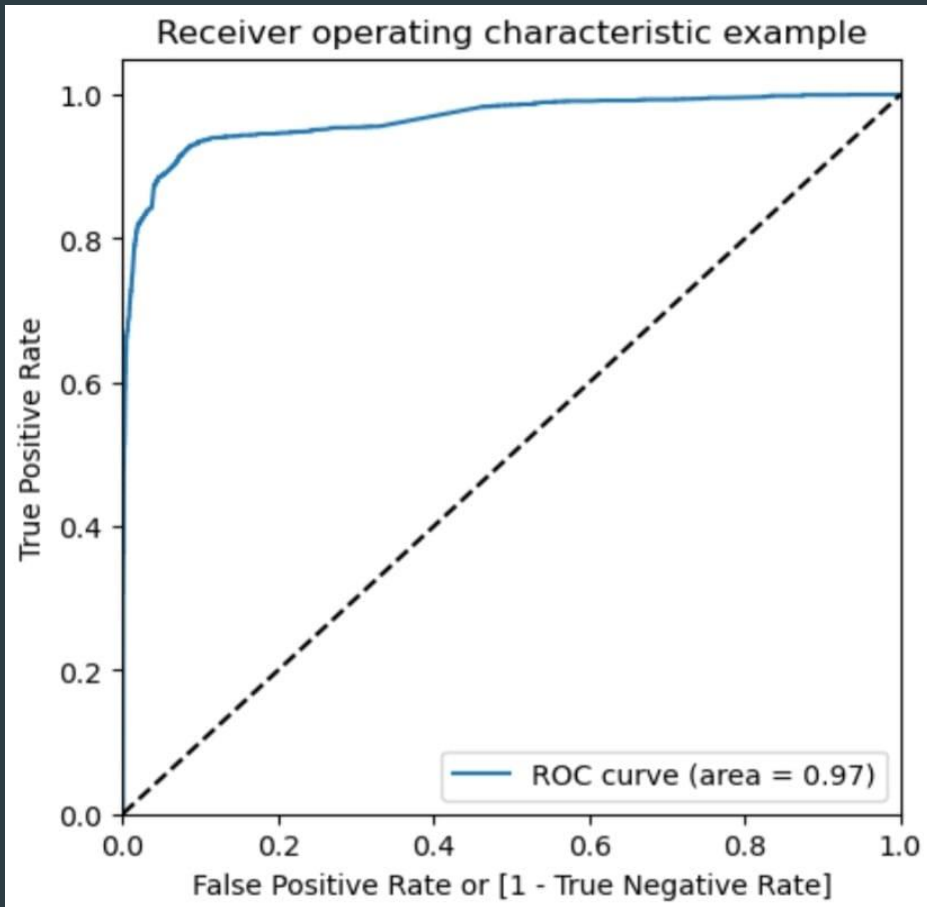
The ROC Curve rises sharply and stays close to the True Positive Rate (TPR) axis, indicating good model performance

Selecting the Optimal Cut-Off point for Probability Conversion

The optimal cutoff point is identified at a probability threshold of 0.3

Final Analysis at Probability 0.3

- ▶ Model Prediction On Train Dataset-

    Accuracy Score of Train Model is 92%

    Sensitivity of Train Model is 91%

    Specificity of Train Model is 92%

- ▶ Model Prediction On Test Dataset-

    Accuracy Score of Test Model is 92%

    Sensitivity of Test Model is 92%

    Specificity of Test Model is 93%

Conclusion and Recommendation

▶ Based on the model output, the top three variables impacting the conversion rate are: 'Closed by Horisson', 'Lost to EINS', and 'Will revert after reading the email'.

▶ Based on the original variable names, the most important variables influencing the conversion rate are: 'Tags', 'Lead Source', and 'Last Notable Activity.

▶ It is recommended that the company prioritize customers who are tagged as 'Closed by Horisson', 'Lost to EINS', or 'Will revert after reading the email', particularly those whose last notable activity is 'SMS Sent' and whose lead source is the 'Welingak Website'.