



# Lead Scoring Case Study

---

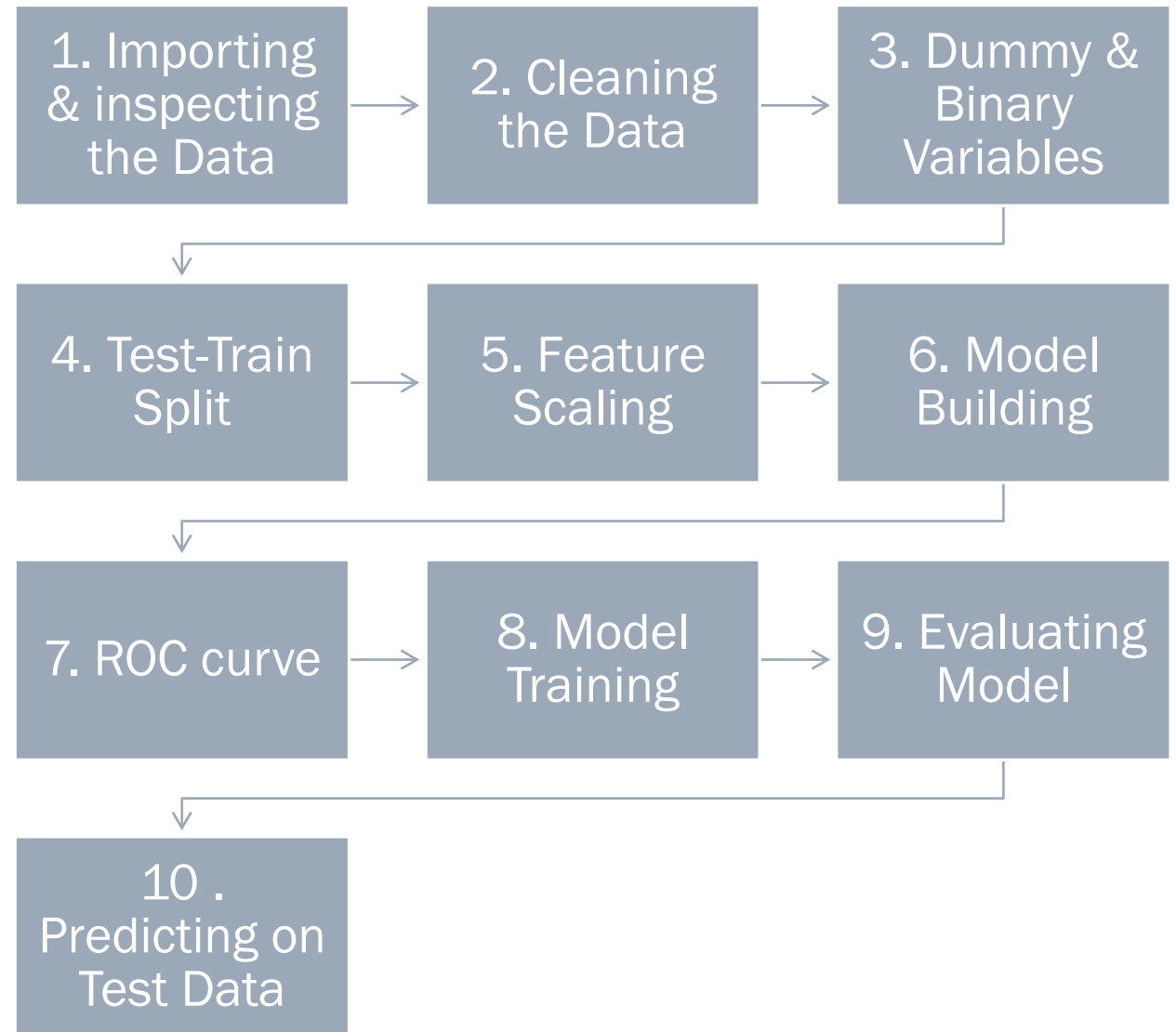
NEERAJ - SWATI (DS C24 BATCH)

*An education company named X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google. When people fill up the form on website redirected from search engines providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.*

*We need to build a model wherein we need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance. The CEO has given a ballpark of the target lead conversion rate to be around 80%.*

## PROBLEM STATEMENT

# Steps involved in Analysis



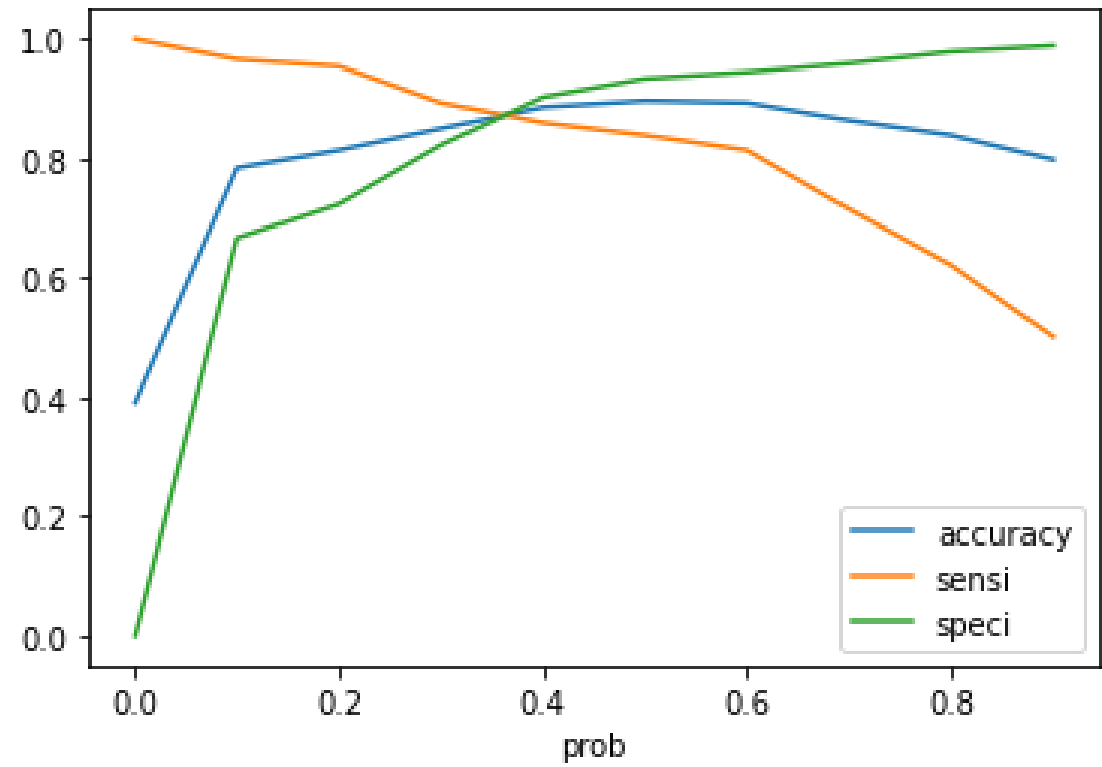
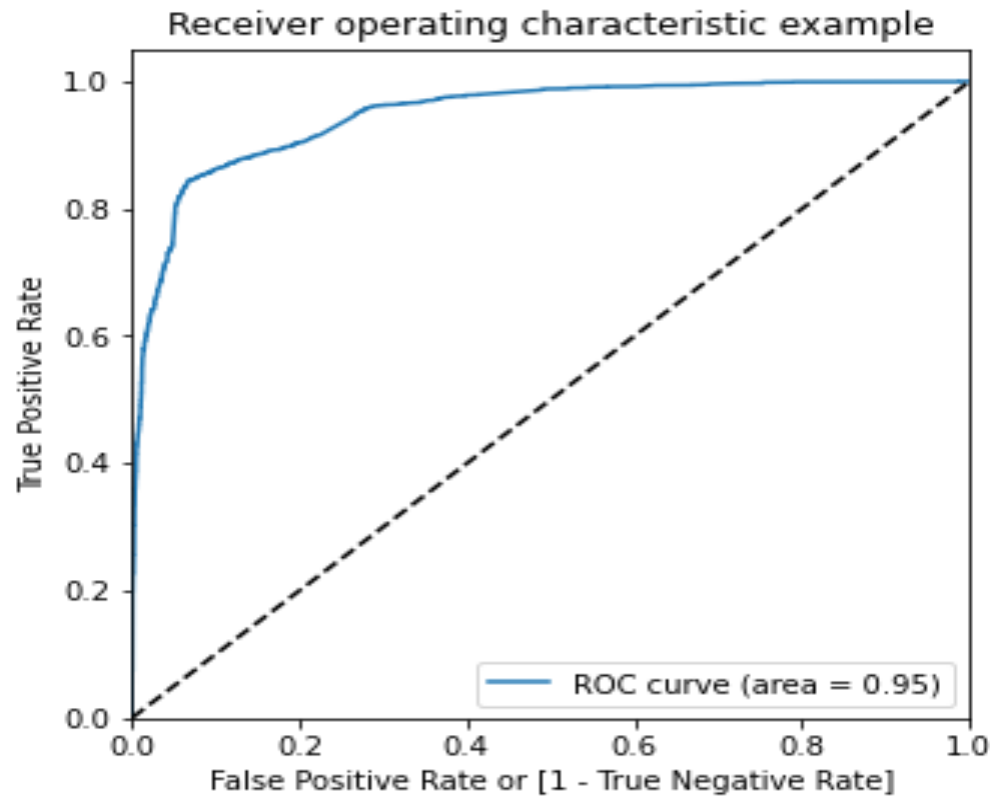
# Steps continued

---

1. Importing the necessary libraries and data required for analysis.
2. Removing the null values and treating the outliers in the data to clean it.
3. Creating dummy variables and converting 'Yes/No' to '1/0' respectively for categorical variables.
4. Split the data into Train-Test in 70%-30%.
5. Scaling the numerical variables to same scale for better prediction.
6. Building the model using features selected by RFE and eliminating some based-on p-values and VIFs.
7. Selecting the optimal cutting point using plot of accuracy, sensitivity and specificity for various probabilities.
8. Training the model on train data set and predicting the values for train data.
9. Evaluating the model based on specificity, sensitivity, precision, recall and accuracy.
10. Finally, predicting values for test data set.

# ROC Curve

# Accuracy, Sensitivity and Specificity plot



# Metrics

---

Accuracy:  $\frac{TN+TP}{TN+TP+FN+FP} = 87.33\%$

Specificity:  $\frac{TN}{TN+FP} = 86.96\%$

Sensitivity:  $\frac{TP}{TP+FN} = 87.86\%$

Precision:  $\frac{TP}{TP+FP} = 81.24\%$

Recall:  $\frac{TP}{TP+FN} = 87.86\%$

# Business Terms

---

ROC curve and Metric values shows that the model build is pretty good in predicting the leads which can be converted.

X Education needs to target clients based on lead score of each client to be converted. Based on the business needs they can choose which score to target (i.e., 50 and above or 75 and above).

# Hot Leads (More than 75% lead score)

---

	Converted	Conversion Probability	Prospect ID	final_predicted	Lead Score
5688	1	0.997940	f1d1bef7-e296-47bd-90bc-05beff0b595c	1	99
5657	1	0.992159	f4a5dfba-b297-464e-8c34-0fefe0af1b27	1	99
1957	1	0.997380	42892ba2-48ab-4736-98c8-c670a29c1018	1	99
5293	1	0.999985	b1f90f08-8cc5-4e9c-a629-02078ece8999	1	99
2727	1	0.990975	649caa1e-3042-45e3-9eda-abf5354bb2a3	1	99
...	...	...	...	...	...
5631	1	0.762196	82bbedbe-5253-49ad-908f-0f05d02ba96a	1	76
1875	0	0.764709	d338bbe4-ef5d-4a6e-9b02-9cec2bfb36cd	1	76
6534	1	0.768134	2f823185-6723-41a4-81d9-a804a4c49bd0	1	76
200	1	0.766761	a996259c-fab6-4fb6-bf41-7c68fbb22114	1	76
6920	0	0.769703	135b97f0-c9a9-4aef-863e-849eb4c8dff4	1	76

752 rows × 5 columns