**Summary for Lead Scoring Case Study:**

For an education company, we need to identify the most promising leads, i.e., the leads that are most likely to convert into paying customers. We build a Logistic regression model and using the optimal cut-off, we will predict the value for converted variable. The sales team has to focus on leads with Converted value 1 in order to improve the conversion rate.

1. Import all necessary libraries and load the data set.

2. Inspect the data frame by getting information about columns in the data set.

3. Prepare the data set by handling missing values and outliers.

4. Map binary variables to 0 and 1. Create dummy variables using one-hot encoding for categorical variables.

5. Split the data set into train and test set.

6. Apply feature scaling using **MinMax** scaling.

7. Check correlation matrix and remove highly correlated variables.

8. Using RFE feature selection, build the first training model and iterate model building till all variables with high p-value(>0.05) and high VIF(>5) are removed.

9. The primary variables that are significant in predicting the Converted flag are

* + **Total Time Spent on Website**
  + **Lead Origin\_Lead Add Form**
  + **What is your current occupation\_Working Professional**
  + **Lead Source\_Welingak Website**
  + **Last Notable Activity\_SMS Sent**
  + **Total Visits**

10. The sales team has to focus on leads having **high values** for above features.

11. Find the optimal cut-off for the train set.

12. For this data set, the accuracy, Specificity and Sensitivity curves meets at **0.36**. So, the optimal cut-off for the model is 0.36.

13. Calculate the lead score using the cut-off value and make predictions for Converted variable.

14. Calculate metrics like accuracy score, sensitivity, specificity, precision, recall, positive predictive rate, false positive rate for the training set using the optimal cut-off.

15. Apply the model on test set and make predictions for the Converted variable.

16. Calculate the lead score and metrics like accuracy score, sensitivity, specificity, precision, recall, positive predictive rate, false positive rate for the test set using the optimal cut-off.

17. If all the metrics on both train and test set are good, the model built is considered as a good model. All the metrics calculated using the model are good, hence the **model can be concluded as a good model**.