LEADS SCORE CASE STUDY

SUBMITTED BY : ***SRINIVASAN. G, NAVEED.J, DENNY.J AND KUMAR.A***

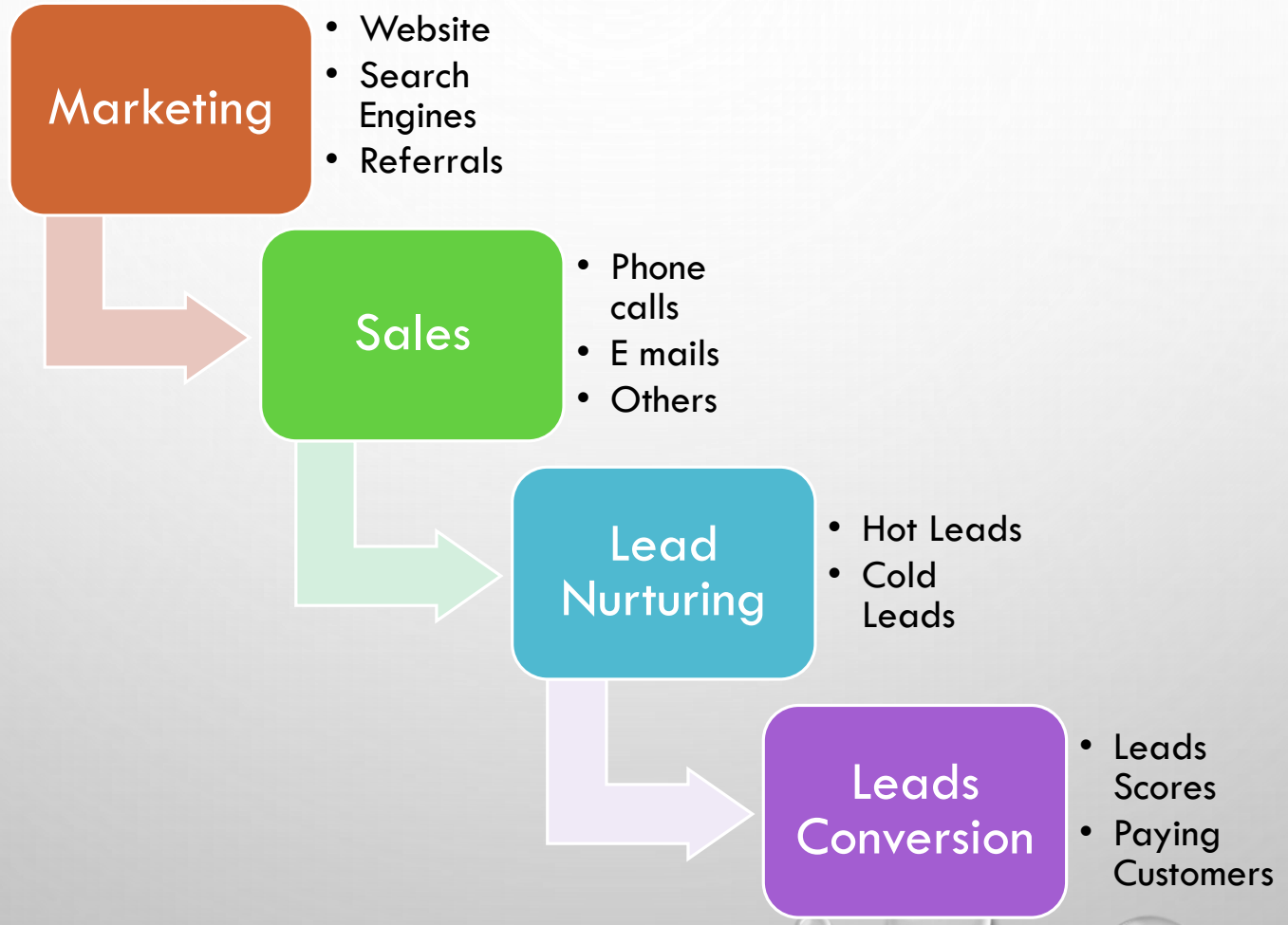
DATE OF SUBMISSION : ***03-MARCH-2019***

BATCH : ***SEPTEMBER - 2018***

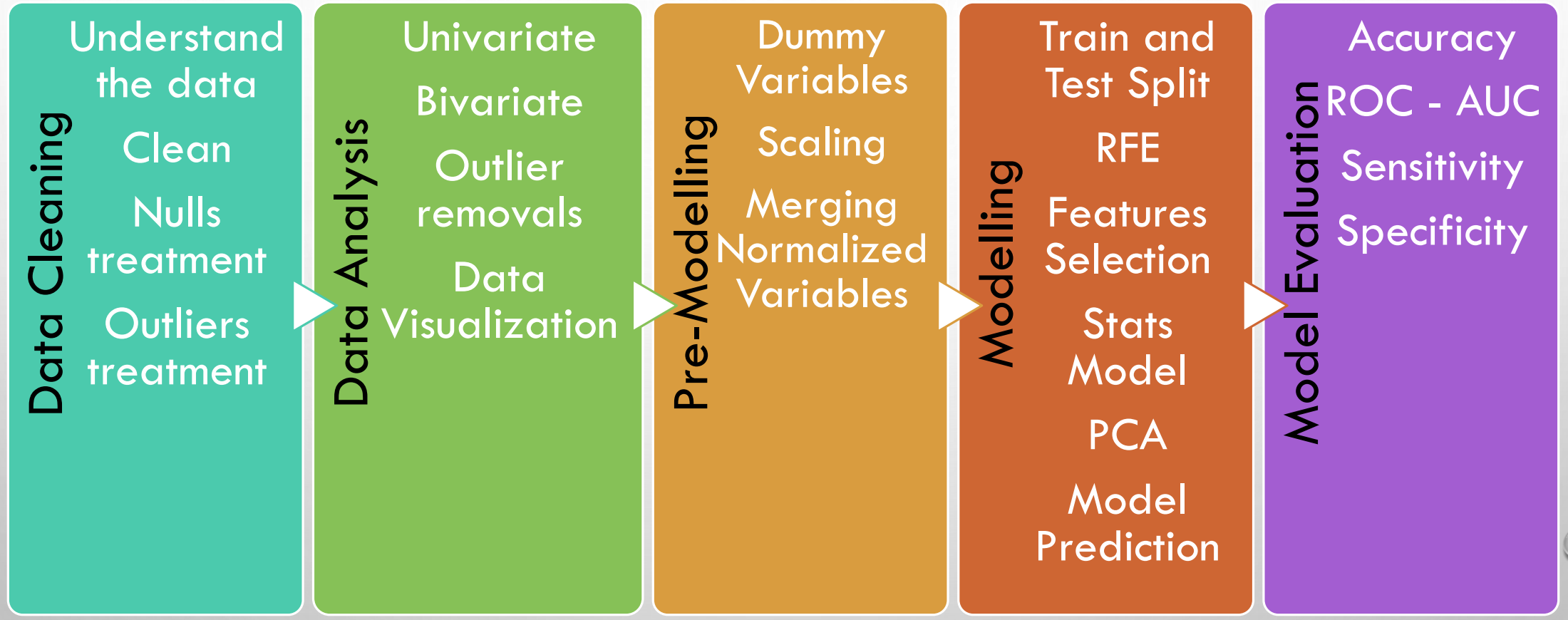
PROBLEM STATEMENT

- X EDUCATION SELLS ONLINE COURSES ONLINE TO INDUSTRY PROFESSIONALS
- MANY LEADS ARE GENERATED VIA SEVERAL WEBSITES AND SEARCH ENGINES
- X EDUCATION MANAGES TO CONVERT ONLY 30% OF THIS LEADS I.E. CONVINCES THE CUSTOMERS TO JOIN THEIR ONLINE COURSE
- WE ARE GIVEN A TARGET TO CONVERT 80% OF THIS LEADS BY THE CEO OF X EDUCATION
- WE NEED TO IDENTIFY THE TOP 3 FEATURES THAT NEEDS TO BE FOCUSED ON CONVERTING THE MOST LEADS
- WE NEED TO PROVIDE A STRATEGY TO THEIR SALES TEAM OF AN AGGRESSIVE LEAD CONVERSION TARGET

X EDUCATION BUSINESS PROCESS



TECHNICAL PROCESS



DATA CLEANSING

Null

- Categorical Columns with more than 30% Null were dropped
- Categorical Columns with less Null value % were replaced with ‘others’
- Rows with less Null % were dropped

Impute

- Categorical Columns with Null were replaced with ‘unknown’
- Categorical Columns with ‘Select’ were replaced with ‘unknown’
- Categorical Columns with ‘other’ were replaced with ‘others'
- Categorical columns such as Asymmetric Index with blanks were replaced with 0
- Numerical columns with blanks were replaced with mean

Cleansed Data

- 3 columns Dropped
- 1.5% Rows with Null dropped

DATA ANALYSIS

Univariate

- Boxplots to check Outliers
- Distribution plots
- Removal of Outliers identified in TotalVisits and PagesPerView columns

Bivariate

- CountPlots to understand some key independent variables Vs Target ‘Converted’ variable
- Correlation matrix to check highly correlated variables

LOGISTICAL REGRESSION MODEL

Dummy variables

- Replace Yes/No value rows with 1/0
- Create Dummy Categorical Values

Scaling

- Standardize Numerical Values except 1/0 value columns
- Merge normalized data with Original data

Build Model

- Train and Test Data Split
- RFE
- Stats Model
- Re-build model with RFE support variables
- Make Predictions

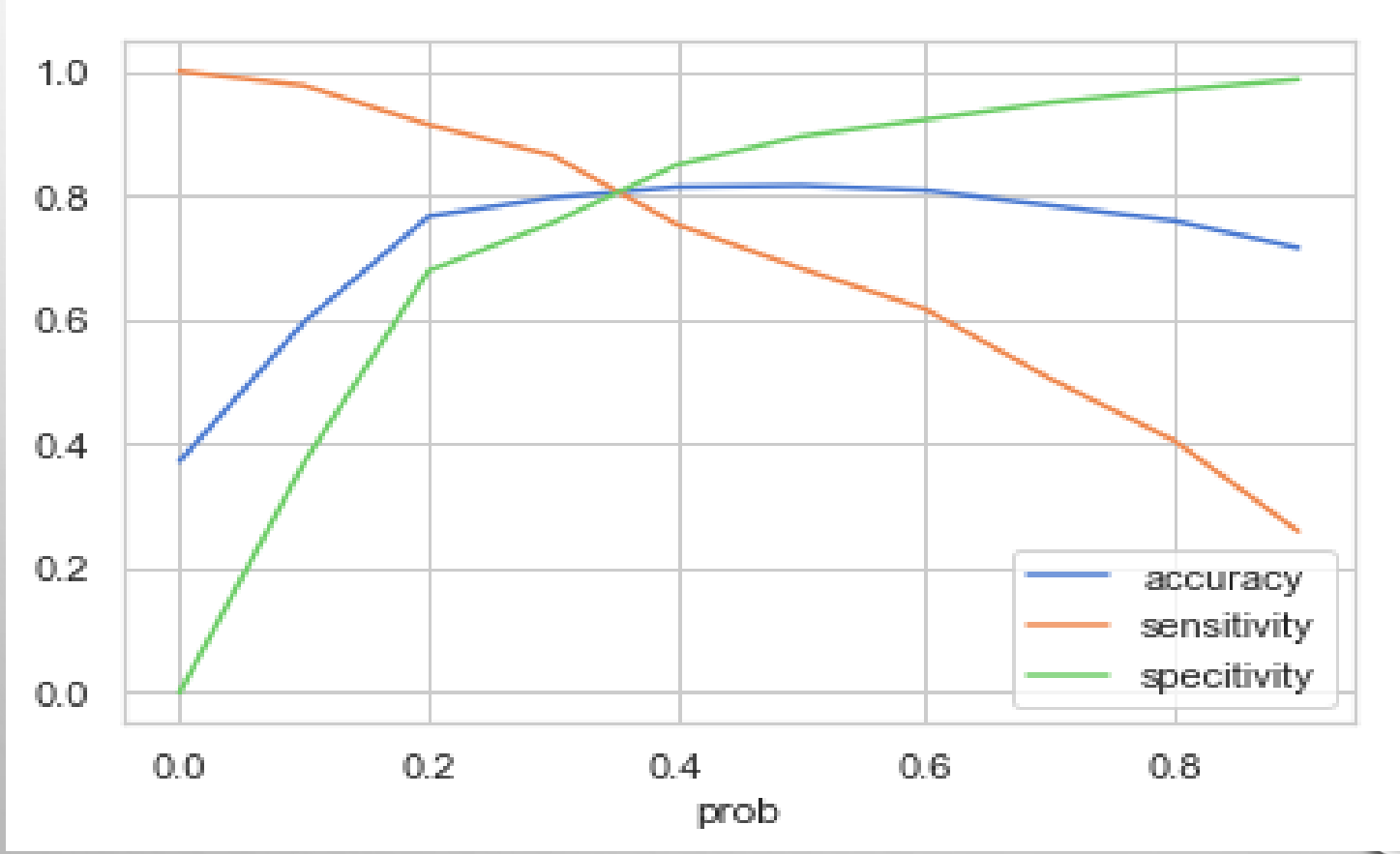
Model Evaluation(RFE)

- Confusion Matrix
- Accuracy Score
- Sensitivity
- Specifity
- ROC-AUC

Model (PCA)

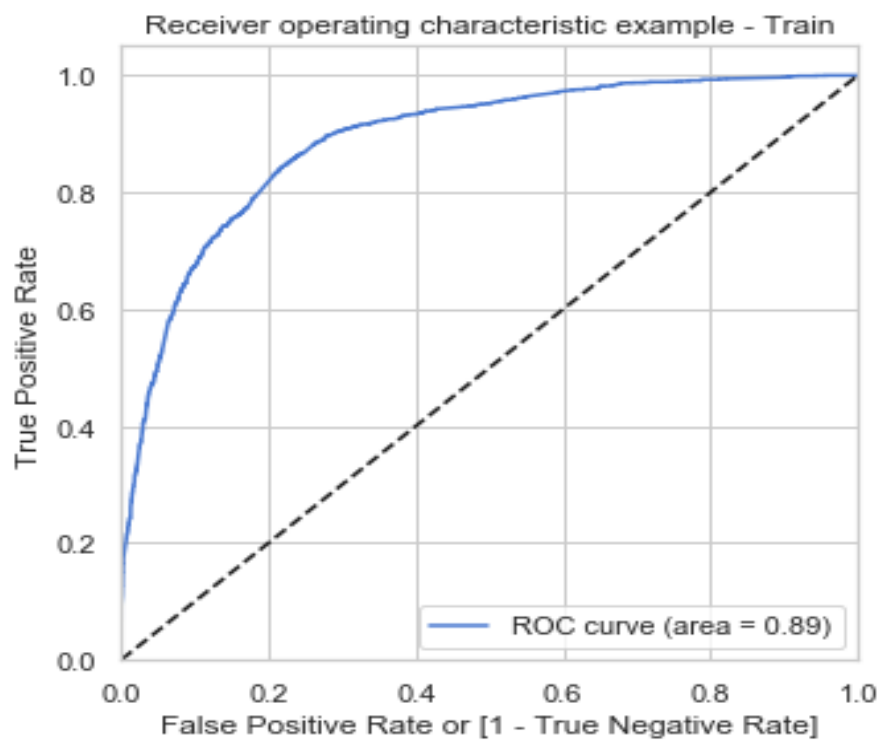
- Train and Test data split
- PCA Analysis
- Build Logistical model
- Evaluate
- Compare RFE and PCA model Accuracy score

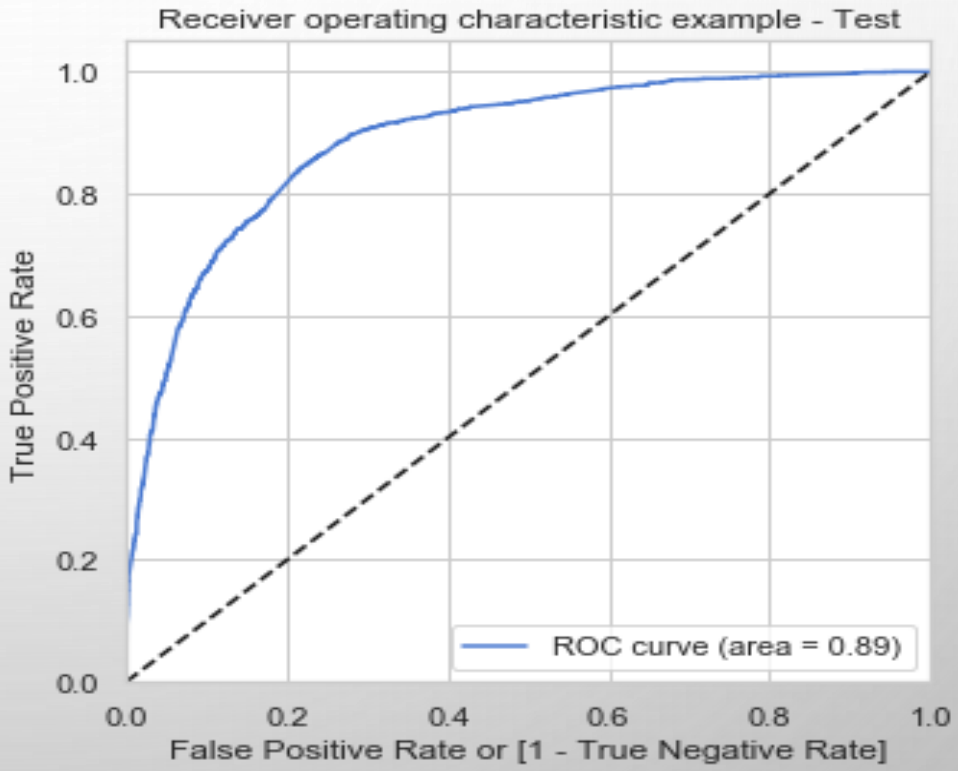
ACCURACY SENSITIVITY AND SPECIFICITY FOR VARIOUS PROBABILITIES



Optimal Cut-off = 0.35

ROC - AUC – TRAIN VS TEST





Both
0.89

SUMMARY & CONCLUSION

- **BOTH RFE (RECURSIVE FEATURE ELIMINATION) AND PCA (PRINCIPAL COMPONENTS MODEL) WERE BUILD TO FIND OUT THE ACCURACY OF THE ACCURACY OF THE MODEL.**
- **OPTIMUM CUT-OFF WAS DEFINED AS `0.35' AND THE PREDICTED TARGET CONVERSION HAS BEEN RESET FOR BOTH RFE AND PCA MODEL.**
- **THE STATISTICS OBTAINED FROM BOTH MODELS (TEST STATISTIC) ARE AS BELOW;**
  - **THE ACCURACY SCORE COMES OUT IDENTICAL IN THE BOTH CASES;**
    - ***ACCURACY SCORE USING RFE - 0.81***
    - ***ACCURACY SCORE USING PCA - 0.81***
- **THE PRECISION SCORE ARE ;**
  - ***RFE - 0.71***
  - ***PCA - 0.73***
- **THE RECALL SCORES ARE ;**
  - ***RFE - 0.84***
  - ***PCA - 0.82***
- **ALSO OTHER METRICS INCLUDES ;**
  - ***SENSITIVITY - 0.84***
  - ***SPECIFICITY - 0.80***
  - ***FALSE POSTIVE RATE - 0.20***
  - ***POSITIVE PREDICTIVE VALUE - 0.71***
  - ***NEGATIVE PREDICTIVE VALUE - 0.88***