

# LEAD SCORING

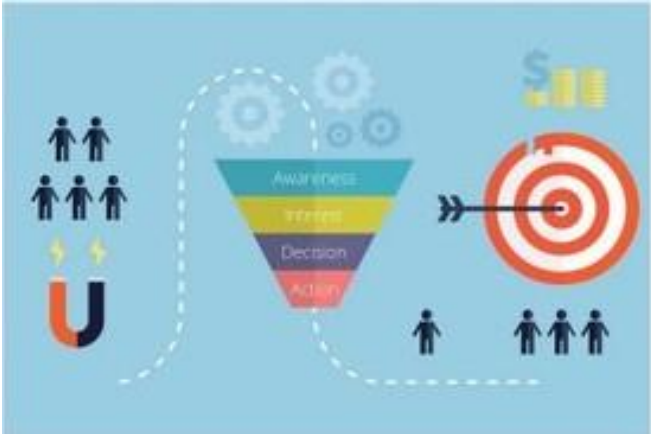
- QUALIFIED
- ACTIVE
- RIGHT FIT



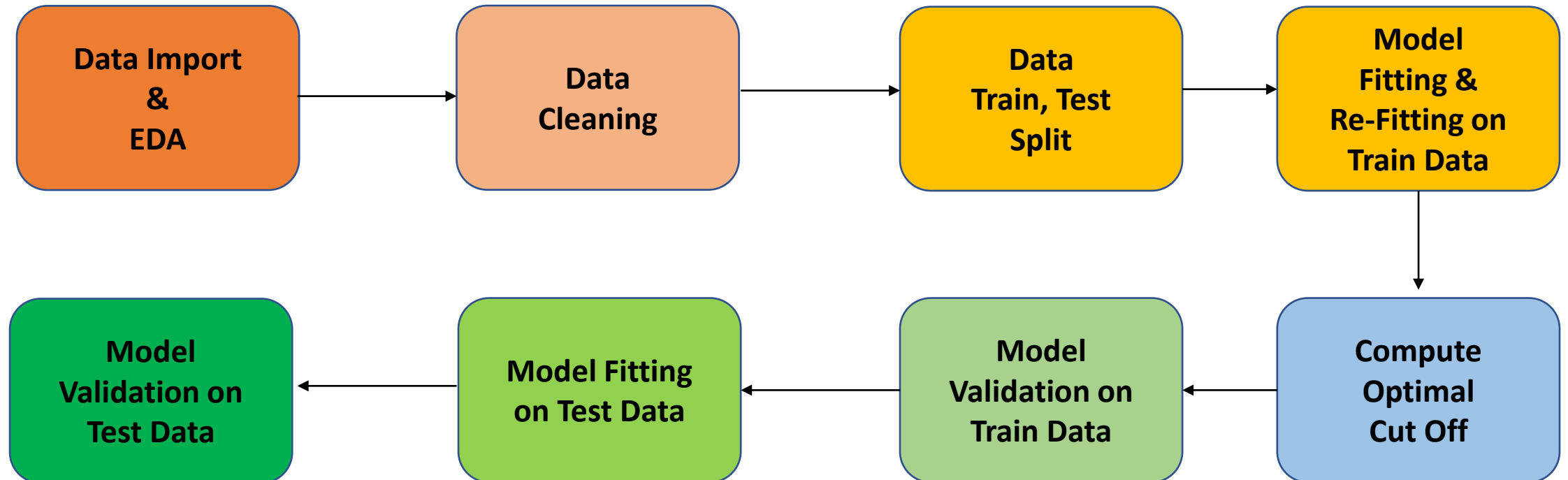
## Lead Scoring Model

Submitted By:  
*Deepak R*

# Process Flow



*Logistic regression model is deployed, as the final output indicates whether leads are Hot/Cold rather than their score.*





# Exploratory Data Analysis

- Size of data: 9240 rows \* 37 columns
- Quantitative parameters: 6
- Parameters with null values > 30%: Tags, Lead Quality, Asymmetrique Activity Index, Asymmetrique Profile Index, Asymmetric Activity Score, Asymmetric Profile. (6 col)
- Visualize parameters( Total Visits, Total Time Spent on Website, Page Views per Visit) for outlier analysis
- Visualize correlation to detect correlation between multiple variables

**E** XPLORATORY

**D** ATA

**A** NALYSIS



# Data Cleaning

- Drop parameters with null values count  $> 30\%$
- Parameters with null values count  $5\% - 30\%$  are replaced with category('others').
- Parameters with null values count  $< 5\%$  are extrapolated on mode().
- Parameters with redundant categories are combined to single category: Lead Source, Country, How did You Hear about X Education, What is your current occupation, What matters most to you in choosing a course, Lead profile, City
- Drop columns with single category
- Create dummy variables for categorical parameters

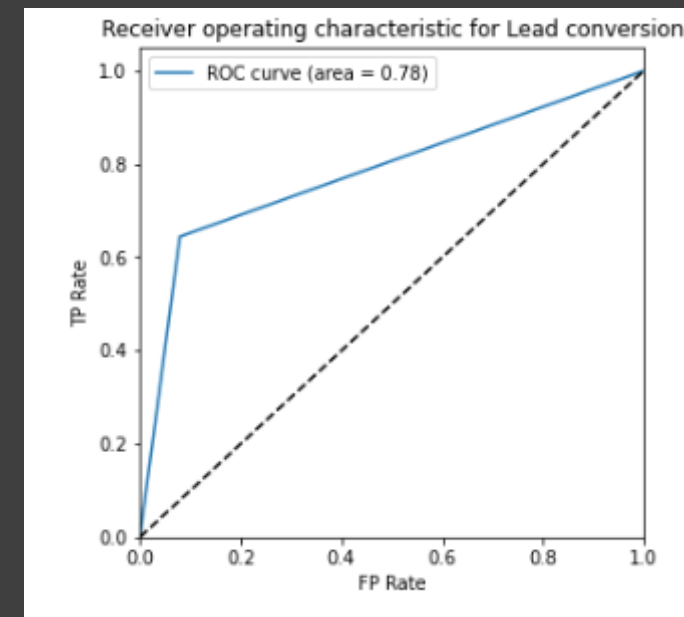


# ROC Curve

- Determines the Area Under the Curve(AUC)
- More the curve towards upper left corner, the model has good performance since most of the area is covered.
- The curve below represents presents the curve which is discovered when the probability of lead conversion is randomly chosen as 60% or 0.6

| Confusion Matrix   |      |      |
|--------------------|------|------|
| Reality/Prediction | 0    | 1    |
| 0                  | 3672 | 316  |
| 1                  | 882  | 1598 |

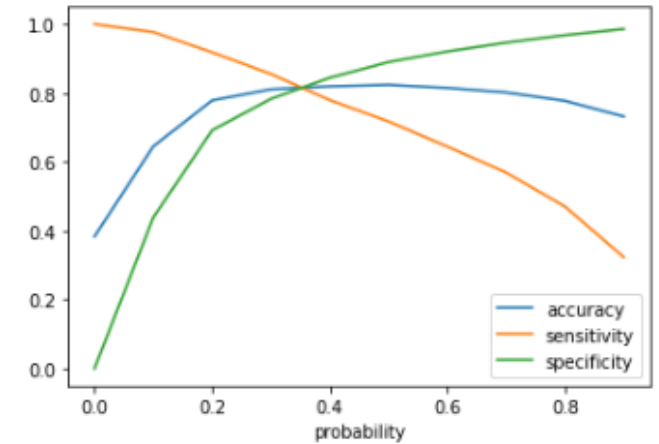
|                  |          |
|------------------|----------|
| Overall Accuracy | 0.814780 |
| Sensitivity      | 0.644355 |
| Specificity      | 0.920762 |



- Since the AUC varies for different probability values and in-order find out optimal point where none of the parameters like accuracy, sensitivity & specificity are low, we find a point of probability where the curves of these parameters meet.



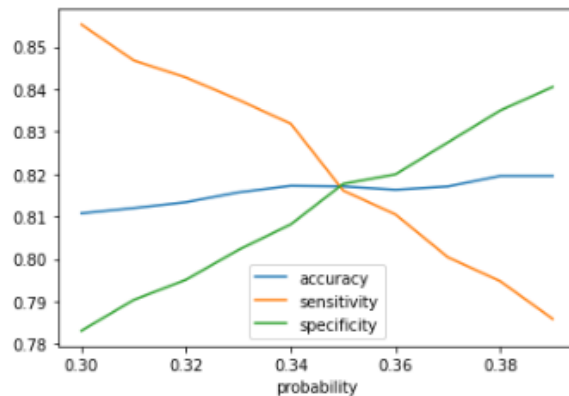
| Probability | Accuracy | Sensitivity | Specificity |
|-------------|----------|-------------|-------------|
| 0.0         | 0.383426 | 1.000000    | 0.000000    |
| 0.1         | 0.645176 | 0.977016    | 0.438816    |
| 0.2         | 0.778602 | 0.918145    | 0.691825    |
| 0.3         | 0.810761 | 0.855242    | 0.783099    |
| 0.4         | 0.819419 | 0.779435    | 0.844283    |
| 0.5         | 0.823748 | 0.717339    | 0.88992     |
| 0.6         | 0.81478  | 0.644355    | 0.920762    |
| 0.7         | 0.801639 | 0.568952    | 0.946339    |
| 0.8         | 0.777211 | 0.470161    | 0.968154    |
| 0.9         | 0.731756 | 0.322984    | 0.985958    |



| Probability | Accuracy | Sensitivity | Specificity |
|-------------|----------|-------------|-------------|
| 0.3         | 0.810761 | 0.855242    | 0.783099    |
| 0.31        | 0.811998 | 0.846774    | 0.790371    |
| 0.32        | 0.813389 | 0.842742    | 0.795135    |
| 0.33        | 0.815708 | 0.8375      | 0.802156    |
| 0.34        | 0.817254 | 0.831855    | 0.808175    |
| 0.35        | 0.8171   | 0.816129    | 0.817703    |
| 0.36        | 0.816327 | 0.810484    | 0.81996     |
| 0.37        | 0.8171   | 0.800403    | 0.827482    |
| 0.38        | 0.819573 | 0.794758    | 0.835005    |
| 0.39        | 0.819573 | 0.785887    | 0.840522    |

# Optimal Cut Off

- Optimal cut off was calculated twice in the interest of specific optimal value
- 1<sup>st</sup> time probability was considered between 0.0 – 0.9.
- After getting to know that there might exist some point between 0.3 and 0.4, again the calculation was done between 0.3 – 0.39



# Model Validation

- With optimal cut off probability of 0.35, after fitting the model on test data, the model has following parameters

Overall Accuracy: 82.864 %

Sensitivity: 84.736 %

Specificity: 81.668 %

| Confusion Matrix   |      |     |
|--------------------|------|-----|
| Reality/Prediction | 0    | 1   |
| 0                  | 1381 | 310 |
| 1                  | 165  | 916 |

- Since all the 3 parameters are almost same and high, model can be considered as good fit to recognizing hot and cold leads

