

# Portuguese Bank Acquisition Analytics

**Submitted by :**  
Akshay Nasarpuri

## Problem Statement :

A Portuguese bank had conducted a telemarketing campaign somewhere in late 2010 to sell term deposit. The bank collected data of around 41,000 prospects. Aim is to analyze this historic data so that the bank can launch another telemarketing campaign to sell the same product with better cost management.

As an analyst, below are the questions to be answered:

1. Once the likelihood of response is predicted, how many prospects should be targeted?
2. By how much can the marketing cost be reduced using the model and how many prospects can be acquired?

In short, the problem statement is:

- To reduce the customer acquisition cost by targeting the ones who are likely to buy.
- To improve the response rate.

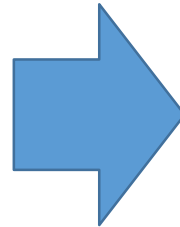
## Business Goal:

Business objective is achieving 80% of total responders at the minimum possible cost. The total number of responders is the total number of prospects who responded, from the available data of about 41,000 prospects.

- ☐ Source the data for analysis
- ☐ Clean and prepare the data
- ☐ Exploratory Data Analysis.
- ☐ Feature Scaling
- ☐ Splitting the data into Test and Train dataset.
- ☐ Building a logistic Regression model with Principal Component Analysis.
- ☐ Evaluating the model by using different metrics - Specificity and Sensitivity.
- ☐ Applying the best model in Test data based on the Sensitivity and Specificity Metrics.
- ☐ Analysis Gain and Lift Chart.
- ☐ Performing the Cost Analysis.

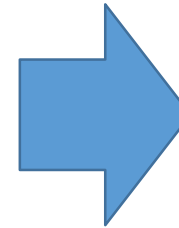
## Data Sourcing , Cleaning and Preparation

- Read the Data from Source
- Convert data into clean format suitable for analysis
- Remove duplicate data
- Outlier Treatment
- Exploratory Data Analysis



## Feature Scaling and Splitting Train and Test Sets

- Standardization of Numeric data
- Splitting data into train and test set.



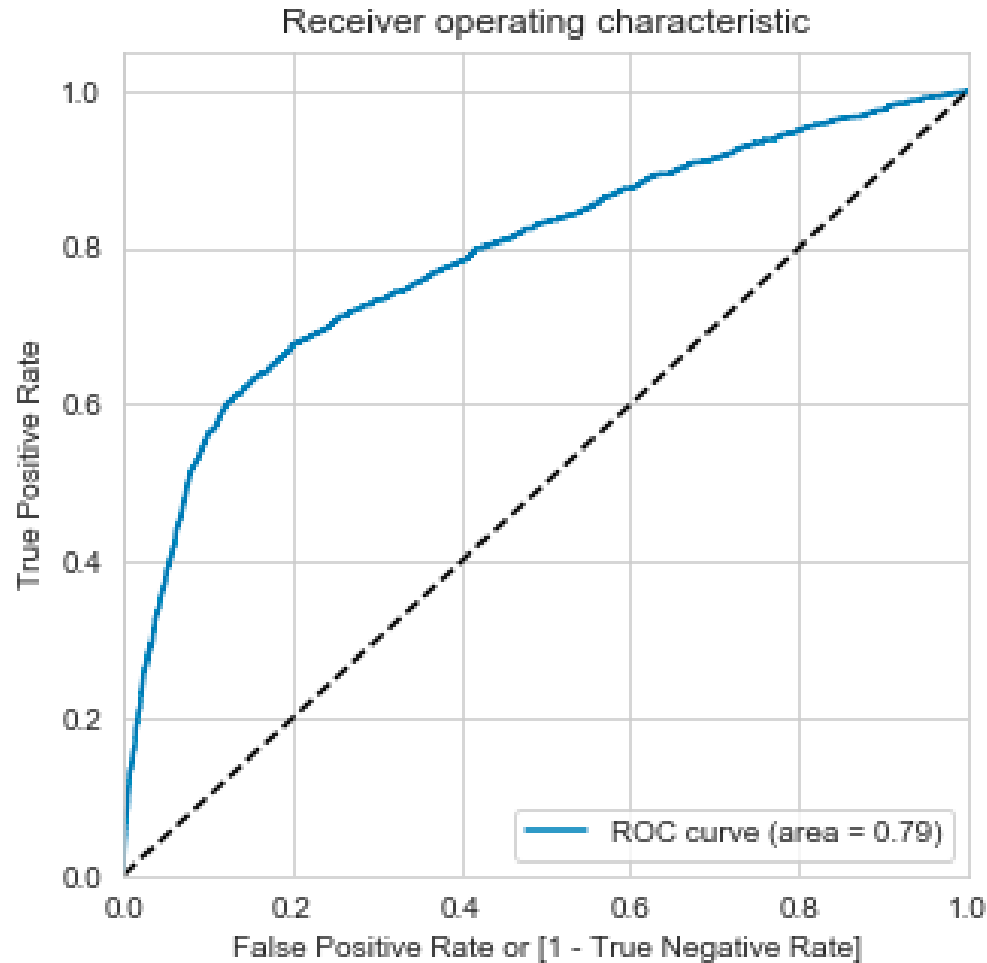
## Model Building

- Feature Extraction using PCA.
- Determine the optimal model using Logistic Regression
- Calculate various metrics like accuracy, sensitivity, specificity to evaluate the model.



## Result

- Determine the response and check if target final predictions amounts to 80% conversion rate.
- Evaluate the final prediction on the test set using cut off threshold from sensitivity and specificity metrics

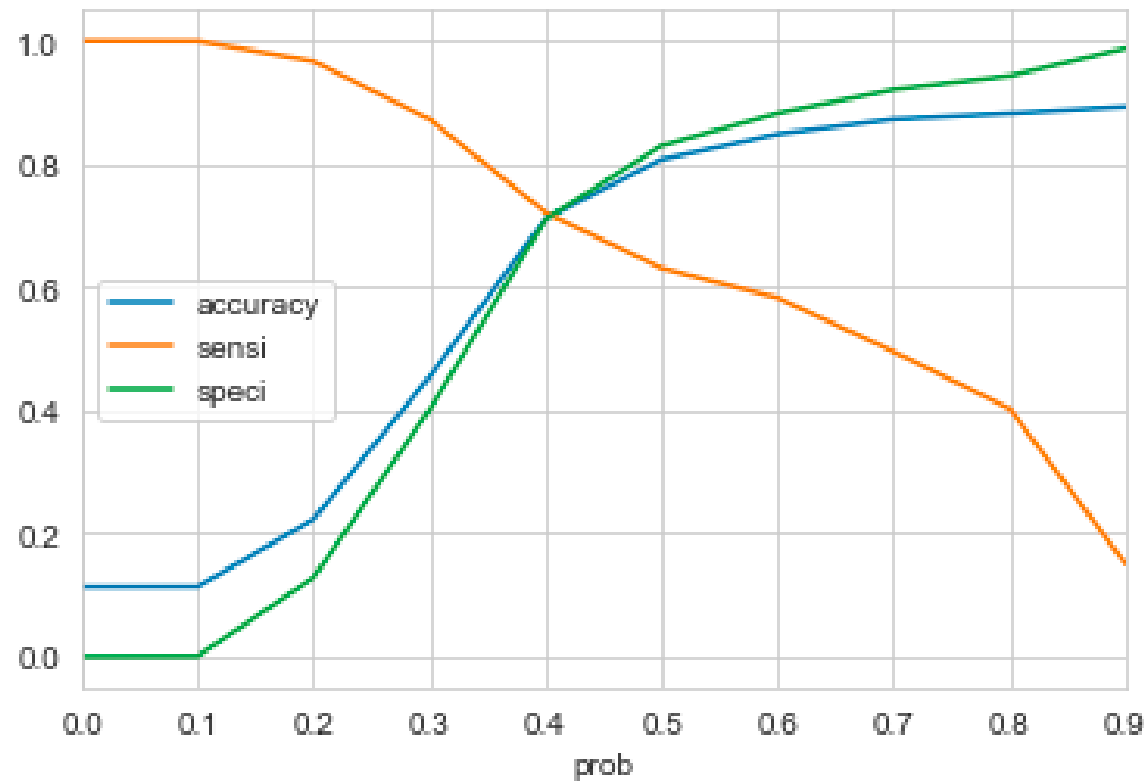


As we can see from the graph on your left, we have plotted a ROC curve. We can see from the graph that the area under the curve is 0.79

So we can say that The accuracy of the model seems to be good.

## Model Evaluation - Sensitivity and Accuracy on Train and Test Data Set

The graph depicts an optimal cut off of 0.40 based on Accuracy, Sensitivity and Specificity



### Train Data Set : Confusion Matrix

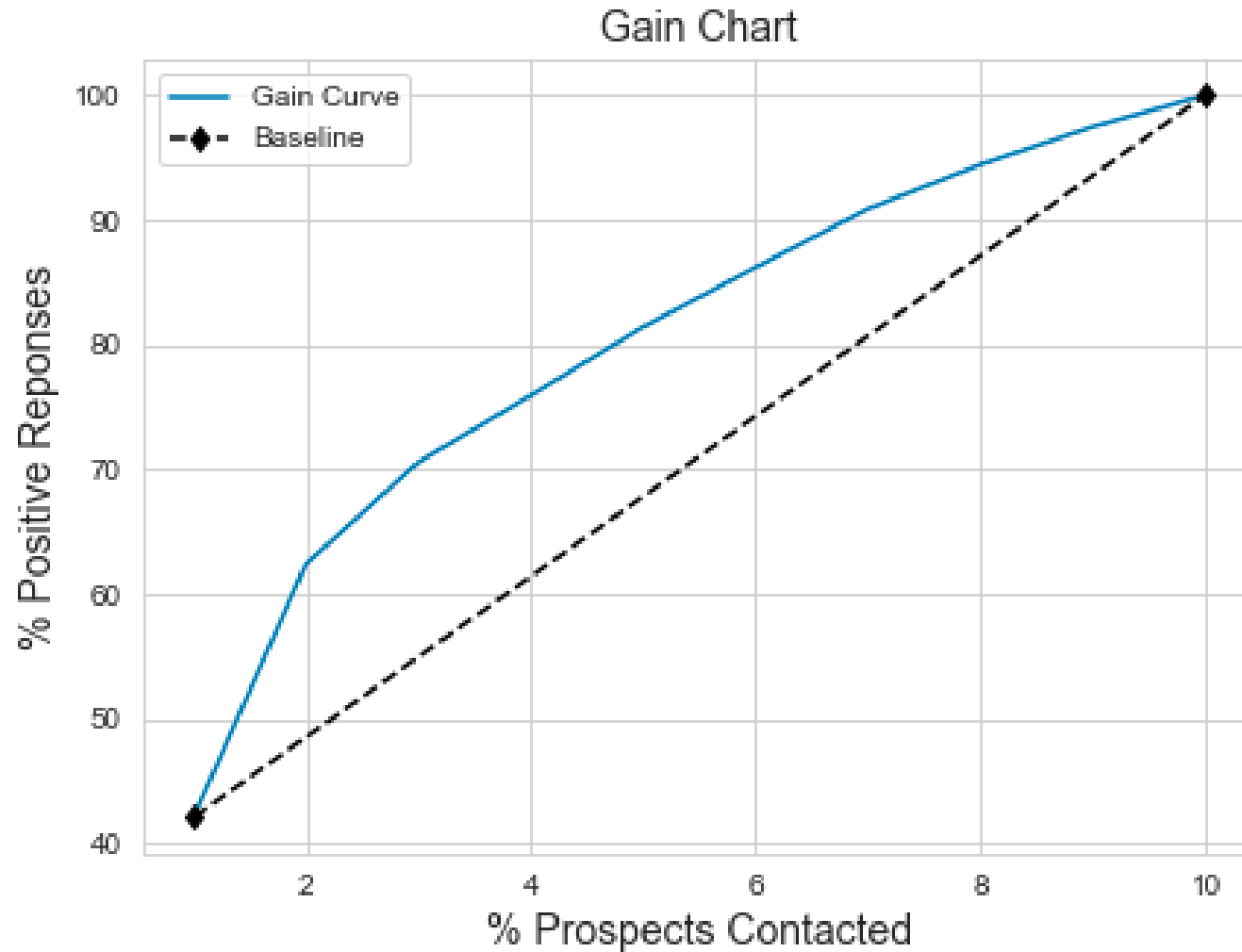
18163	7403
908	2349

- Accuracy - 71%
- Sensitivity - 72 %
- False Positive Rate - 29 %
- Positive Predictive Value - 76 %

### Test Data Set : Confusion Matrix

7811	3160
378	1004

- Accuracy - 71 %
- Sensitivity - 73 %



It plots the responders captured against the number of prospects targeted.

**Y-axis:** Shows the percentage of positive responses. This is a percentage of the total possible positive responses

- 1,382 as the total number of responses
- 11.19 % as the percentage.

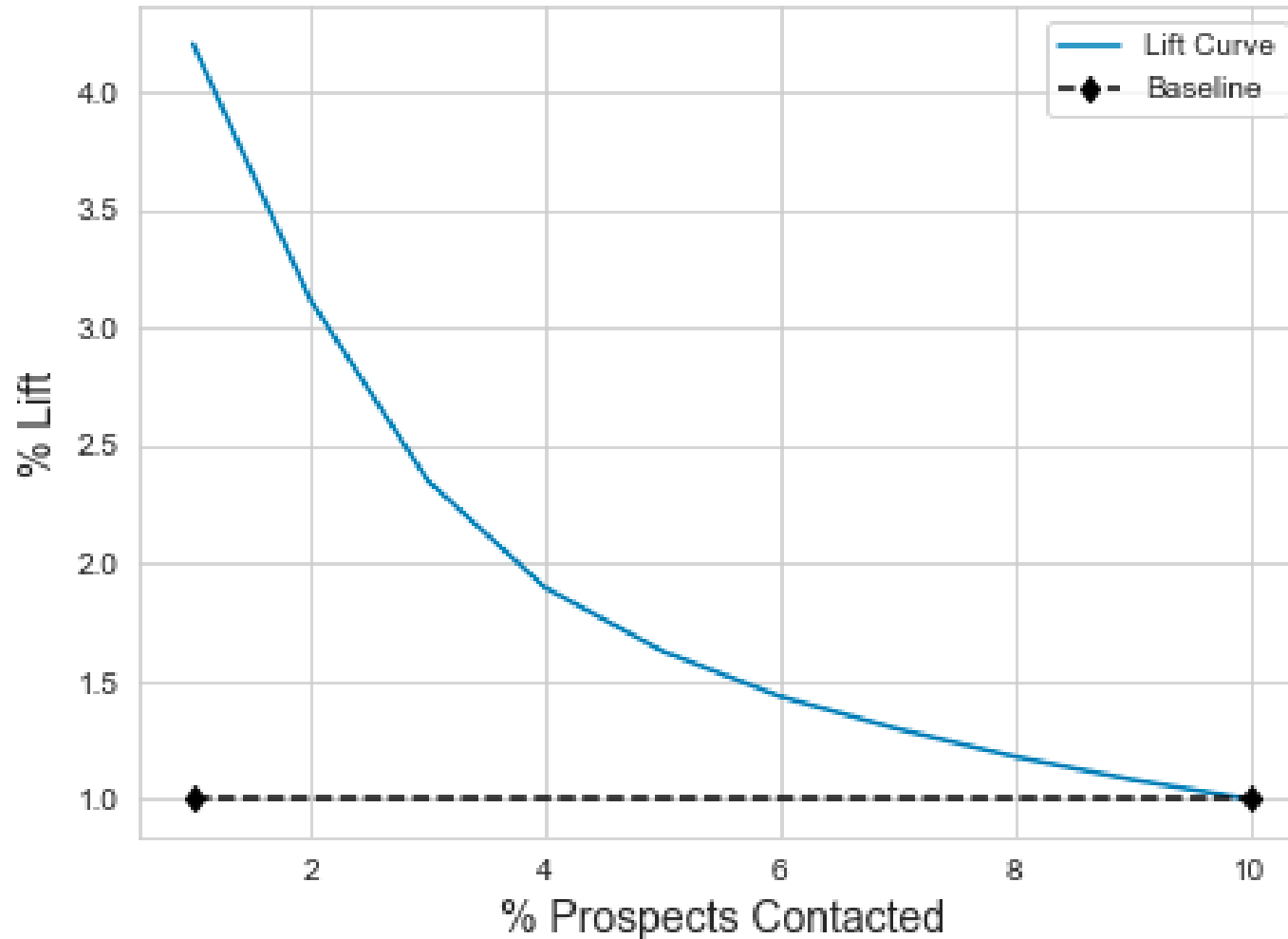
**X-axis :** Shows the percentage of customers contacted, which is a fraction of the 12,353 total customers.

**Baseline (overall response rate):** If we contact X% of customers then we will receive X% of the total positive responses (When model is not used)

**Gain Curve:** Using the predictions of the response model, 80% of the total responders can be captured by targeting the top 50% of the prospects

- Top 50% prospects: 6176
- 80% of response: 1125

Lift Chart



It compares the response rate with and without using the model; Shows the actual lift.

**Y-axis:** Shows the ratio between the result predicted by our model and the result using no model.

**X-axis:** Shows the percentage of customers contacted, which is a fraction of the 12,353 total customers.

**Baseline (overall response rate):** If we contact X% of customers then we will receive X% of the total positive responses (When model is not used)

**Lift Curve:** Using the predictions of the response model, by targeting the **top 50%** of the prospects, we can get 1.63 times the original response rate.

- Original response rate = 11.19
- Lift of 1.63 =>  $11.19 \times 1.63 = \mathbf{18.22\%}$



## Conclusion

- We have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction.
- Accuracy and Sensitivity values of test set are around 71% and 73% which are approximately closer to the respective values calculated using trained set.
- The average call duration for targeting the top 50% of the total prospects is 255 minutes.
- Total cost for targeting the top 50% prospects is Rs. 58,460.
- Cost Benefit:
  - Cost when no Model: Rs. 3,58,319
  - Cost with Model: Rs. 58,460
  - Cost of approx. 6 times less with the model.
- 80% of the total responses can be captured by targeting the top 50% of the prospects.
- This can be achieved by utilizing around 16% of the total initial cost.
  - Saving of around 84%.
- Hence, overall this model seems to be good.