TELEMARKETING PREDICTION FOR BANKING / Amogh Singhal

Executive Summary

This is a machine learning model that predicts whether a person will end up buying a term loan or not. The model is able to classify the customers into target segments based on their buying propensity. We can extend this to design a targeting strategy to only go after segments with the most propensity and discard the ones with lower propensity. This will help us two folds :

1. **gain a better success rate for the campaign because we are targeting the segment with the highest potential for conversion.**
2. **we will be able to save our budgeting costs because of the discarded segments.**

P_Rank_Dtree	N	Event Rate
0	0	0
10	4119	52%
9	4119	20%
8	4119	8%
7	4118	6%
6	4119	6%
5	4119	7%
4	4118	5%
3	4119	4%
2	4119	3%
1	4119	3%

Cum ECR	No Model	Cum NECR	KS	Target Zone
0	0	0	0	
46%	10%	5%	0.40317	1
64%	20%	14%	0.49085	1
71%	30%	25%	0.45757	1
76%	40%	35%	0.40027	0
81%	50%	46%	0.34562	0
87%	60%	57%	0.29922	0
91%	70%	67%	0.23925	0
95%	80%	78%	0.16711	0
98%	90%	89%	0.08574	0
100%	100%	100%	0	0

Data Source

[Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

Approach

Kindly follow this notebook to understand the steps done in detail

1. **Load the required libraries and setup the project environment**
2. **Read the dataset and perform basic operations like checking data types and distribution of target variable**
3. **Perform EDA on the features whose discriminative behavior resonates with the target variable. The more the dicrimination, greater is the usefulness of the feature.**
4. **Data cleaning and Data preparation to make it model ready.**
5. **Modelling - we have used three classification models here to test and compare**
    a. **Logistic Regression**
    b. **Decision Trees**
    c. **Random Forest**
6. **Plucking probabilities found from the model and binning them into deciles to create target segments**

Results

We ended up making a product agnostic framework that can be used to design a marketing strategy for any kind of product against any channel. We finalised the decision tree model which was able to give us better bins with higher success rate and better F1-score between the three models. If we target the top group (only 10% of total pop.), my success rate has increased to 52% !

In a no model scenario this was only 11.3% If we target only the top two segments, cumulatively I can gain a success rate of 64% In a no model scenario that would only amount to 20%. These segments can also be studied to design customer personas and build lead scorecards. For more details, please refer to the Gains Curve and presentation.