ISSN : 2230-9519 (Online) | ISSN : 2231-2463 (Print)

Predictive Modeling to Improve Success Rate of Bank Direct Marketing Campaign

Vaidehi R
Velammal Engineering College, Chennai, India

Abstract
The increasingly vast number of marketing campaigns over time has reduced its effect on the general public. Economical pressures and competition have led marketing managers to invest on directed campaigns with a strict and rigorous selection of contacts: lesser contacts should be done, with better success rate. Although telemarketing is a direct mode of communication with the prospective customer, this may make customers grumpy.

In this paper, Predictive modeling is used in determining the main characteristics that affect success and selection of potential buying customers. GLM (Generalized Linear model), Rule based and R Part Tree algorithms were used to build the model using the most popular tool R and the appropriate model is selected based on ROC, ACC (Accuracy) value and False Negative value (FN).

Keywords
Data Mining, Prediction, and Variable Importance, ROC

I. Introduction
The most successful direct marketing is to focus on the quality of prospect data, attempting to predict the expected customers that have a higher probability to use the service by using data mining technique. To understand customer behavior, many banks have adopted the predictive technique based on the data mining to predict the customer data for classifying the customers before offering special services. The prediction or classification is the most important task in the data mining that is usually applied to classify the group of data. In classification, the outcome is a categorical variable and several combinations of input variable are used to build a model and the model that gives a better prediction with good accuracy is chosen to target the prospective customers,

II. Proposed Approach
In this study, the proposed approach used to enhance the predictive rate of the bank telemarketing dataset is using three models namely Generalized Linear Model, Decision Tree and Rule Based model and appropriate model is selected based on Receiver Operating Characteristics (ROC) curve.

A. Generalized Linear Model
In a Generalized Linear Model (GLM), each outcome of the dependent variables, Y, is assumed to be generated from a particular distribution in the exponential family, a large range of probability distributions that includes the normal, binomial, Poisson and gamma distributions, among others. The mean, μ, of the distribution depends on the independent variables, X, through:

$$\mathrm{E}(\mathbf{Y}) = \boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\beta})$$

where E(Y) is the expected value of Y; Xβ is the linear predictor, a linear combination of unknown parameters β; g is the link function.

B. Decision Tree
A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm.

C. Rule-based Modeling
Rule-based modeling is a modeling approach that uses a set of rules that indirectly specifies a mathematical model. The rule-set can either be translated into a model such as Markov chains or differential equations, or be treated using tools that directly work on the rule-set in place of a translated model, as the latter is typically much bigger

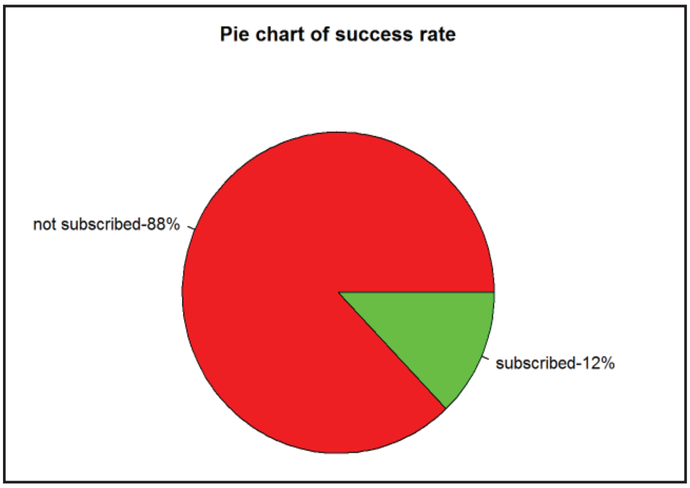
III. Data Preprocessing
This study considers real data provided by The UCI Machine Learning Repository. This dataset was collected from a Portuguese retail bank, from May 2008 to June 2013. The dataset is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

There are 45211 records, which of which is composed of 16 features and 1 binary response. The following table gives a brief introduction of all features.

IV. Atrributes
Table 1:

Num.	Attribute Name	Description	Type
1	Age	It is age of client.	Numeric
2	Job	It is type of client's job.	Categorical
3	Marital	It is client's marital status.	Categorical
4	Education	What is the highest education of client?	Categorical
5	Default	Does client has credit?	Categorical
6	Housing	Does client has housing loan?	Categorical
7	Loan	Does client has personal loan?	Categorical
8	Contact	What is a contact communication type of client?	Categorical
9	Month	What is the last month of the year contracting to the client?	Categorical
10	Day of Week	What is the last day of the week contracting to the client?	Categorical
11	Duration	How long does it contact to the client?	Numeric
12	Campaign	Number of contacts performed during this campaign and for this client	Numeric
13	Pdays	Number of days that passed by after the client was last contacted from a previous campaign	Numeric
14	Previous	Number of contacts performed before this campaign and for this client	Numeric
15	Poutcome	Outcome of the previous marketing campaign	Categorical
16	Label	Does the client has subscribed a term deposit?	Categorical

V. Basic Plots



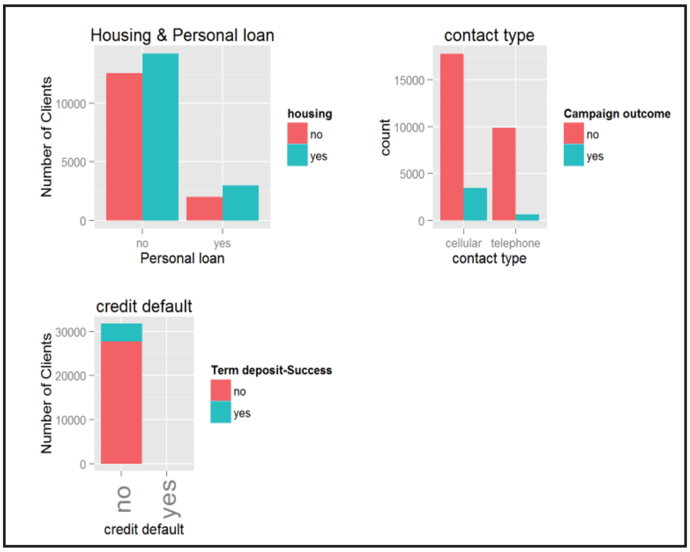


Fig. 1:

The above plot (Fig. 1) shows that credit default is not a predictor of campaign success. (A client with default credit can not be the target customer)

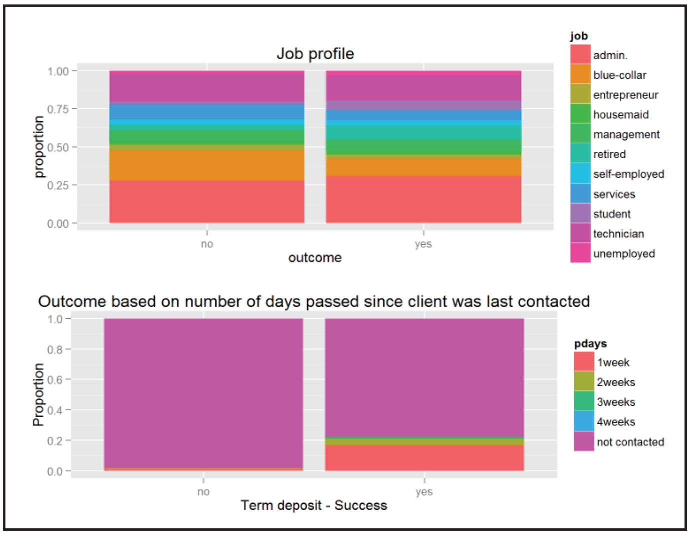


Fig. 2:

The above plot (Fig. 2) shows that marketing success may vary based on customer's job profile. Admin, technician and retired people are more likely to be the target customer. Since, often more than one contact to the same client was required, in order to access if the product would be or not subscribed, the success is more likely when contacted frequently.

Fig. 3:

The above plot (Fig. 3) shows clearly that when a client talks for more duration, it is more likely that he may subscribe a term deposit. The term deposit subscription is more among the customers with age between 25 and 40 and those above 60. The same is true from the job distribution plot (previous slide) which shows that the success rate is more among retired people and technician.

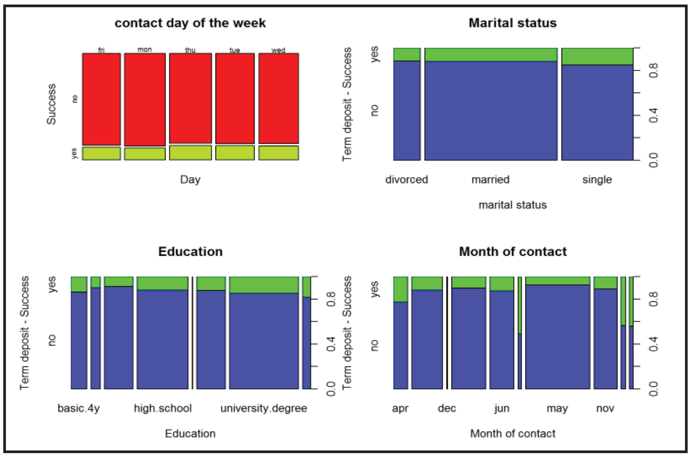
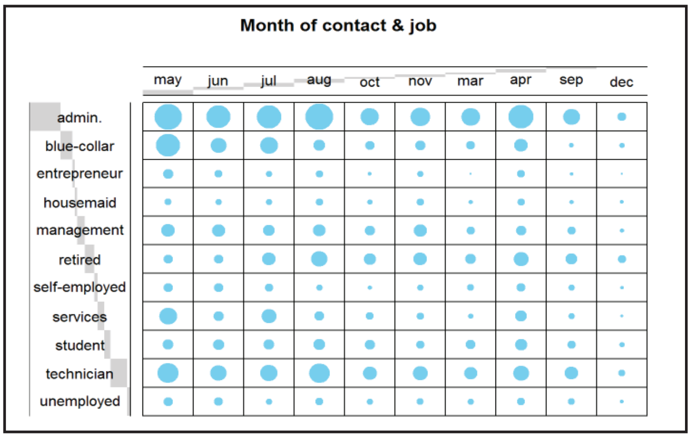


Fig. 4:

The above plot (Fig. 4) shows that the contact day of week and marital status do not affect the campaign success.
The success rate varies based on the month of contact.



 ISSN : 2230-9519 (Online) | ISSN : 2231-2463 (Print)

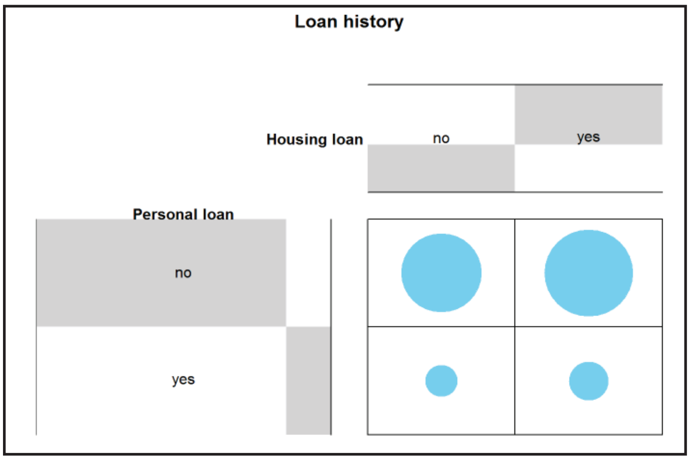


Fig. 5:

From the above balloon plot (Fig. 5), success distribution is same among people having different types of loan and hence loan may not influence the outcome.

VI. Model Selection

Table 2:

Algorithm	Model	Accuracy	FN	ROC
Rule based	Variable importance	0.910400	1041	0.7453
Rule based	AIC	0.908804	1031	0.7586
Tree	Variable importance	0.908340	1012	0.8133
Tree	Full model	0.908300	1012	0.8133
Tree	AIC	0.908300	1012	0.8133
GLM	Variable importance	0.906500	1466	0.9101
GLM	AIC	0.906000	1494	0.9049
GLM	Campaign related	0.905500	1492	0.9001
Rule based	Full model	0.904400	1095	0.7690
GLM	Plots	0.906700	1467	0.9092
Zero R	Zero R	0.888100	1843	0.5000

GLM algorithm with variables selected from variable importance has more accuracy and better ROC.

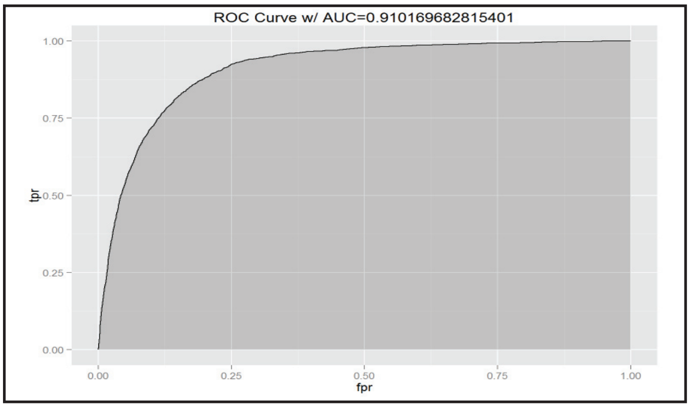


Fig. 6:

The following are the variables selected for the model

- Duration
- P days - number of days passed since last contact made with client
- Age
- Contact - cellular/ Telephone
- Month
- Job

The influence of the above variables on bank direct marketing success is also seen in the plots

VII. Conclusion

The following variables seem to be the most relevant inputs in predicting the Success rate of bank direct marketing campaign

- Duration - call duration
- Pdays - Number of days since last contact
- Month - month of contact
- Age - customer age
- Contact - cellular/ Telephone
- Job

A client is more likely to subscribe term deposit if customer talks for more duration. Campaign is more likely to be successful during March, September, December (end of every trimester). A Customer with default credit would not go for term deposit subscription. Campaign reach is good among blue collar, admin, retired and housemaids.

Reference

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

[2] S. Moro, R. Laureano, P. Cortez,"Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology", In P. Novais et al. (Eds.), Proceedings of the European Simulation and Modelling Conference - ESM'2011, pp. 117-121, Guimaraes, Portugal, October, 20 `11. EUROSIS. [bank.zip]