Market Analysis for Bank Data

Akash Soni (0873602), Chirag Patel (0883437), Hardil Mehta (0898141),
Vatsal Patel (0883974)

Lakehead University, Thunder Bay, Ontario

Abstract

The main problem definition of our project is to predict the success of calls made by the bank to sell their long-term deposits. In addition, how strongly are the customers engaged with their bank can be found out. Dataset taken is of the time span of 2 years starting from 2008. A large dataset of 22 features is being analyzed. We compared two models: Logistic Regression and Decision tree. The findings are the accuracy of both the models: Logistic Regression (80.01%) and Decision Tree (76.03%). The insight that we've gained is a Logistic Regression model works more accurately in our database for predicting whether a customer would open a bank account or not in Financial Institution.

Keywords: Bank deposits, Telemarketing, Savings, Classification, Variable selection, ROC, AUC, Visualization, Data Analytics, Linear Regression, Logistic Regression, Decision Tree

---

1. Introduction

Nowadays, companies prefer to contact their clients and customer directly to meet a specific goal[1]. For doing so, any company needs to analyze their best bunch of customers. For this a strong predictive model is needed to accomplish the task. The goal is to design a data-driven model that learns an unknown underlying function that will be able to map several input variables, which will categorize an item, with one labeled output target (In our case: opening of an account by customer: "failure" or "success").

In the past 15 years, Bayesian statistical models have evolved a lot. In 2005, Rossi, Allenby, and McCulloch[11] proved that Markov Chain Monte Carlo methods works quet efficiently than Bayesian Models. Afterwards, in 2010 Kaplan and Haenlein[12], research found an important relation of customer-firm relationship. Using which decision tress came into the picture and was proved more accurate model to use. The motivation behind choosing decision tree is the results shown in these researches. Further, we found Classification, Confusion matrix, ROC, and AUC. The achievements of the work are as follows:

- Developed a better model for prediction using Logistic Regression with accuracy of 80.01%.

- Successful in applying the Sensitivity - Specificity analysis, ROC Curve and at last computed AUC which gave us a final accuracy.

2. Related Work

In Bank direct marketing, there are two types of methods for enterprises to sell their products or services: mass campaigns and directed marketing which targets a specific set of contacts (Ling and Li, 1998)[14]. Directed marketing is focused on targets that assumes that these types of campaigns are more attractive because of to its efficiency (Ou et al. 2003)[15]. However, directed marketing has some disadvantages, for example due to the risk of privacy, it might trigger a negative attitude towards banks (Page and Luding 2003)[16].

In Business Intelligence and Data Mining, according to Turban et al. (2010)[17], Business Intelligence is a term that includes architectures, different databases, tools, applications and several methodologies with the goal of using data for decisions of business managers. Data Mining is a Business Intelligence technology that uses model that is data-driven to find out meaningful knowledge from complex and large data (Witten and Frank, 2005)[18].

The process known as CRoss-Industry Standard Process for Data Mining (CRISP-DM) is a very famous methodology for increasing the success of projects of Data Mining (Chapman et al., 2000)[19]. CRISP-DM is a cyclic process, where many iterations are used to allow final result more accurate towards the business goals. When we identify the goal we want to achieve , analysis of data is needed for analyzing (Data Understanding) and processed (Data Preparation). Witten and Frank (2005) [18] says that the data knows what needs to be learned, independent records which are related to it's occurrence and attributes that shows an aspect of a given instance. The Modeling phase helps to build the model which will show the learned knowledge (e.g. the given model can be used to do prediction of the target variable that represents the goal which is defined). In addition, the model is being analyzed in the Evaluation phase with respect to its performance and utility[8]. For example, tasks of classification tasks focus to predict a discrete target variable, common metrics which is basically the confusion matrix (Kohavi and Provost 1998) [20] and the ROC (Receiver Operating Characteristic) curve (Fawcett 2005) [21]. In case, the obtained model isn't good enough to be used in support business, then there's a need to define a new iteration for the CRISP-DM[10]. If not, the model will be implemented in real time environment which is also known as Deployment phase.

Given the interest in marketing campaigns, there are various works which use Data Mining on marketing campaigns to improve campaigns (Ling and Li, 1998)[14] (Hu, 2005)[22] (Li et al., 2010)[3]. More commonly, classification Data Mining approach works, where the objective is to build a predictive model that learns to label the data points into one of many predefined classes, binary or multiple. Various Data mining algorithms are available for classifying marketing contacts; each one has its purposes and capabilities. Some of the widely used mining techniques are Decision Trees (DT) (Aptea and Weiss, 1997)[4], Support Vector Machines (SVM)(Cortes and Vapnik, 1995)[5], and Naive Bayes (NB) (Zhang, 2004)[6].

Classification metrics such as ROC or accuracy rate can be used to measure the performance of thes classifier. However, in the context of marketing campaigns, most commonly used metric to evaluate prediction models is the Lift (Coppock 2002)[7]. Specifically, the cumulative Lift curve is used, that is a percentage graph which divides the population into deciles, in which predicted probability of response is used to place the population members. The deciles are sorted in descending order meaning that the highest responders are sorted on the first decile. Marketing managers use Lift as a useful marketing tool to make essential marketing decisions like how many contacts to make (from the original set) and also can be used to check if there is an alternate better model for some goal of target responses

How our work is Novel/Unique?

We've surveyed the previous research papers and chosen the best optimal methods from it. We try to obtain maximum accuracy in the models that we've applied. We've tried to apply compare different models, found the accuracy, ROC, AUC and finally used the model which has higher accuracy. Comparison of different model and getting optimal solution makes our work unique which helps us to define a relation between an Financial Institution and the customer.

3. Methodology

3.1. Data Source

The data set is collected from www.kaggle.com where the data had been provided by the UCI Machine Learning Repository[2]. This repository is a collection of databases available for use, for the empirical analysis of data set using machine learning algorithms. The data set selected is of Portuguese Banking Institution related to direct marketing campaigns (Phone Call). The data set is a Multi-variate Demographic dataset in the Financial domain and has 45211 instances with 17 attributes. It is designed to integrate multiple factors like education, marital status, job, etc. and provide a comprehensive picture of the available prospects for acquisition by the bank.

Statistical Information of the data-set:

- Publisher: UCI Machine Learning Repository
- Number of rows: 41.1k
- Number of Columns: 20

Attribute Information of the data-set:

Table 1: Description of the columns

Column Name	**Description**
age	Age of the user
job	Description of the type of job. For example : White-collar, or technician
marital	Status of marraige. For example: Marrried or Unmarried
education	level of education i.e. high school, or graduate
default	Do the user have credit or not?
housing	Do the user have housing loan or not?
loan	Do the user have personal loan or not?
contact	Details for how to communicate
month	The month when the user was lastly contacted
day_of_week	Day of the last contact done
duration	Duration of the last contact done
campaign	Total number of people contacted during the campaign organized
pdays	Number of days passed after the user was lastly contacted
previous	number of contacts which were done before campaign
poutcome	Previous campaigns outcome
emp.var.rate	Employement Variation Rate
cons.price.idx	Consumer Price Index
cons.conf.idx	Consumer Confidence Index
euribor3m	rate of 3 months
nr.employed	Total number of employees
Output y	Whether a user opens an account or not

3.2. Data Preprocessing

- Variable Selection

Variable selection is the base for construction of a model. Variable that contribute to build a model ought to be chosen and if the data is immense then a subset of those variables is needed to be selected. Not only should the subset of these variables to be chosen, yet the copies of those indicators ought to be removed. Excessive co-linearity between explanatory variable and target variable, can often prevent identification of an optimal set of explanatory variables for a statistical model.

For Variable Selection, We use step-wise Variance Inflation factors (VIF). VIF is an easy to calculate and straight forward technique. VIF and co-linearity are directly proportional on each other. Higher the VIF, Higher is the co-linearity. In R, VIF functionality is provided by "car" package. The function uses step-wise selection of variable and hence start with a model with no predictors. Then it will start Removing individual variables with high VIF values. Though this is insufficient in the initial comparison using the full set of explanatory variables.The VIF values will change after each variable is removed. It will remove all the variables step-by-step until all the values are below the threshold.

```
> vif(model_2)
                     jobadmin.                    jobretired                    jobstudent
                      1.413710                      1.170345                      1.130520
                 jobtechnician               maritaldivorced    educationPrimary_Education
                      1.280400                      1.018266                      1.411819
   educationTertiary_Education               contactcellular                      monthapr
                      1.321083                      2.347471                      3.804666
                      monthjul                      monthjun                      monthmar
                      3.854532                      8.726474                      1.397140
                      monthmay                      monthnov                      monthoct
                      3.602450                      2.330151                      1.456207
               day_of_weekfri                day_of_weekmon                      campaign
                      1.065908                      1.068281                      1.046169
pdaysContacted_in_first_10days   pdaysContacted_after_10days               poutcomefailure
                      1.220221                      1.034852                      1.130321
                  emp.var.rate                cons.price.idx                 cons.conf.idx
                    139.738077                     52.278115                      2.608831
                   nr.employed  `previousMore than_3_times`
                     71.853803                      1.078169
>
```

Figure 2: VIF value of all the indipendent variables

As we can see that some of the variables such as "nr.employed" has VIF value 71.85 which is more than 10 so should not consider this variable in model. VIF should be less than 5.

- Preprocessing Data

Preprocessing data is an important part of any data analysis process. Data preprocessing is a data mining technique that helps to understand the raw data. It helps correct various inaccuracies, inconsistency and missing trend(s) in the data. This processing of data includes techniques like Data Cleansing, Data selection, Data Integration, Data Transformation, Data Segmentation and Categorization.

In preprocessing, we have added a prospect number to each of the candidate that needs to be evaluated for targeting. We change the textual response of our target variable to a numerical one. We Categorize our data into various categories. Like Age is binned using equal interval binning with bins being (16,20),(20,30),(30,40),(40,50), (50,60),(60,70),(70,80).

Similarly, we categorize Job, Martial status, education, day's after contact(pday) and poutcome.

- Missing Data and Inconsistency

On summing all the Na's in the data-set, we found a zero output indicating there are no Missing Data points in our dataset. There are various Inconsistency and Outliers in our dataset, which we explain in Outlier Detection Section.

- Outlier Detection in Data

Box plot allows to study the distributional characteristics in a vector of values as well as the level of the scores. For Outlier Detection and finding inconsistency, Box plot is an useful tool. We have selected box-plot for the task. Using plotting tools, quantile and box-plot function of r, we found out that most of 99percentile of our data is below age 71. There are few outliers found using box-plot. hence we capped the age-limit to 71. Similarly, this technique is used on duration which is capped to 1271 seconds.

- Standardizing in Data

To allow comparison between different attributes and better prediction Standardizing of data needs to be done. The data-set consists of numerical value and most of the data in the data-set is between an finite interval of range. After Categorization and Outlier Detection the data-set is standardized.

- Information Value or WOE as feature

To Understand, how the categories of Dependent Variables are separated by Independent Variables. In general, it is a variable reduction technique which can help identify optimal binning structure for a data set. The Main risk of using Information Value is that it assumes same linear Relationship for all the variables. On Visualization of data, we found that our data set holds a non-linear relationship and hence we have opted not to use Information Value or WoE as feature.

- Segmentation of the data

Data segmentation is needed for getting a better identification, categorization, and labelling in the data-set. As we've taken a huge data set, we need to analyze each and every columns. Segmentation is done using inbuilt function "dummy.data.frame" in package "dummies". After Segmentation, each of the categories in a vector transforms into their individual vector of values.

3.3. Visualization of data

- Box plot

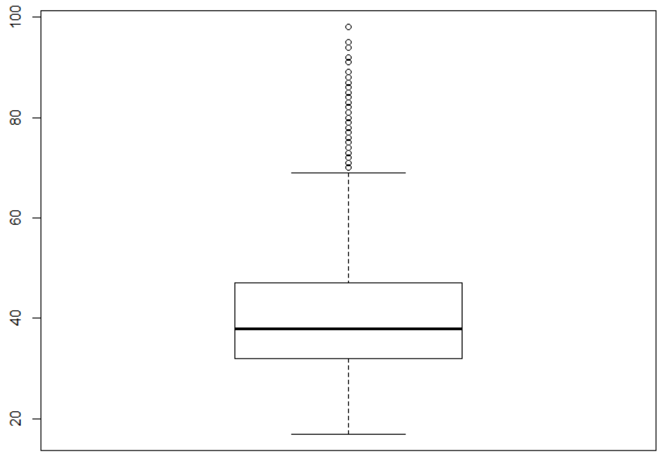

Figure 1: Outlier detection using boxplot on "Age"

The above boxplot shows outlier in age attribute. As one can see, age above 71 is an outlier and hence we put a cap on this value. Similarly we use capping on duration, and campaign days.Smaller the box in box-plot, more is the data aggregated together. as we can see most of our data in campaign attribute is located below 14. So, we cap the attribute on this value.

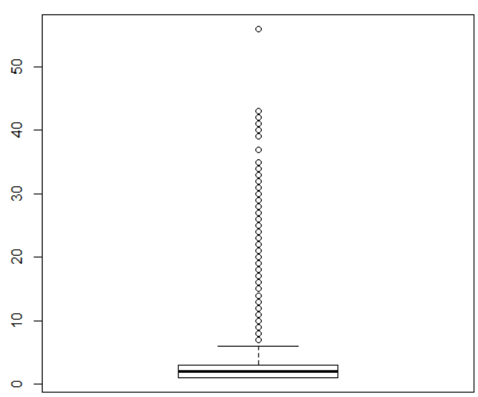

Figure 2: Outlier detection using boxplot on attribute "campaign"

- Histogram

Below, We plot Response rate for different attributes. Response here means that the person gave a positive response on being contacted. I.E. the customer opened an account in the financial institution. Using this plots, we understand which type of attributes contributes more towards a positive response for our Target variable.

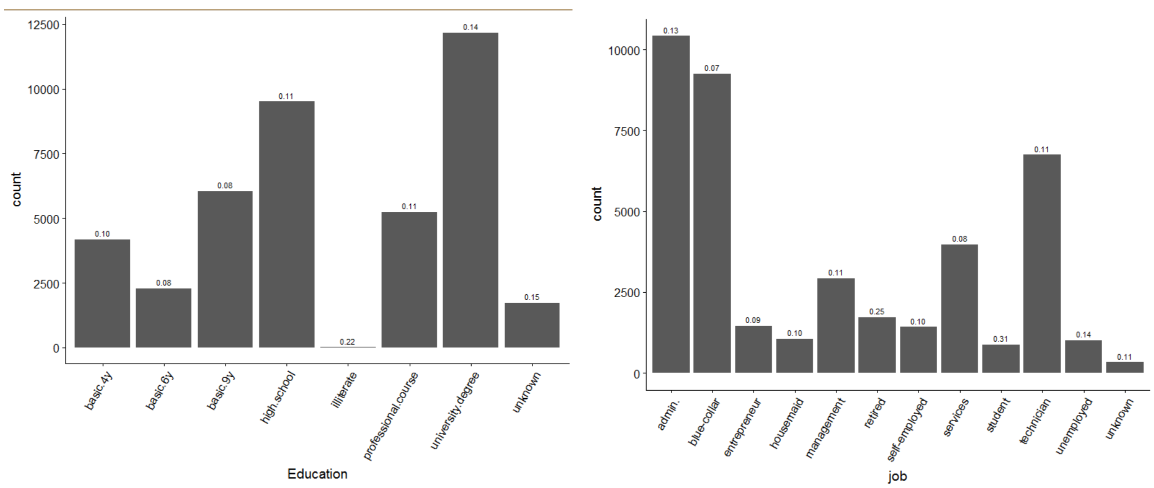

Figure 3: Response Count Vs Education (Left Image) and Response count Vs Job (Right Image)

People having University degree, showed the highest response rate, while people with no education showed the least response rate.Similarly, we understand all the attributes in the table by plotting, processing and standardizing them.

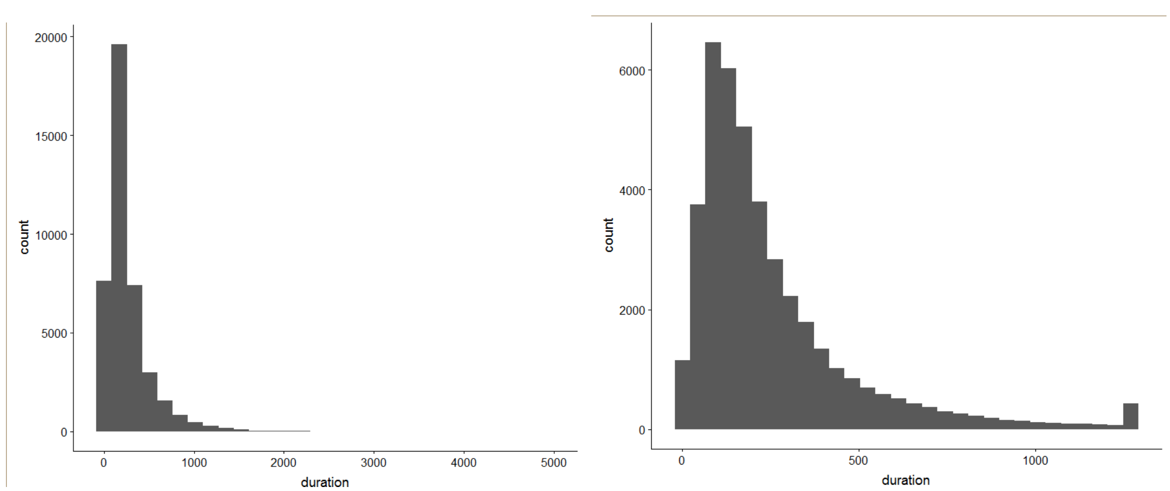

Figure 4: Duration Before Capping values (Left Image) and Duration after Capping values (Right Image)

- Pie-chart

To better understand percentage or proportional data pie chart is really useful visualization method. Percentage represented every category is provided next to the corresponding slice of pie. We have illustrated pie chart of "job" and "education" attribute.

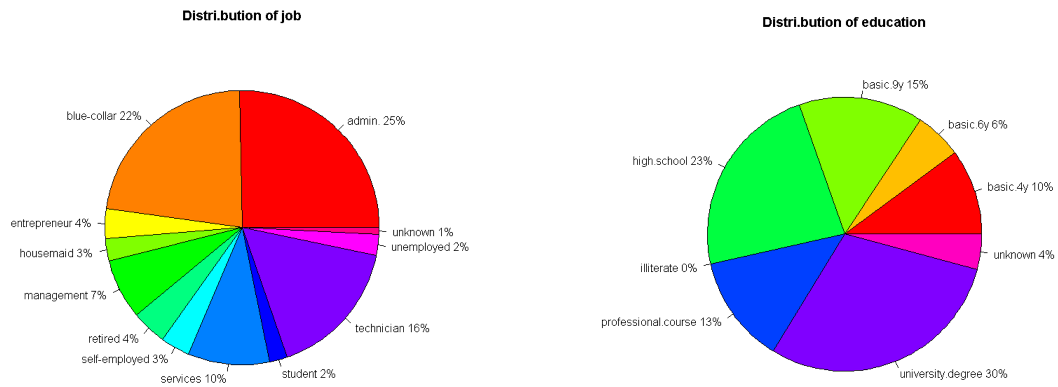

Figure 3: Pie chart representing distribution of job and education.

we can observer from the graph that almost half of the total observations belongs to class blue-collar(22%) and admin job(25%). And from second chart we came to know that highest number customer has has university degree(30%).

3.4. Predictive Analysis

Predictive analytical methods are used to extract crucial information from data. The outcome of the models can be used to predict the behavior of new observations. This process includes many areas of research such as data modeling, machine learning, AI, deep learning algorithms and data mining. Predictive modeling is divided into seven step.

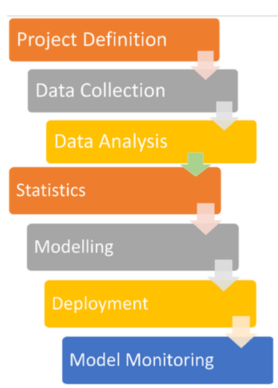

Figure 8: Flowchart of Predictive analytical model building.

Before any project begins one need to understand the requirements, aim, business objectives, data sets that are going to be used in the project. We are building a model to predict that whether customer will open a account with a Portuguese bank or not based on their characteristic data. Data is crucial for any analytical process. This data may belong to multiple sources. A good dataset can lead to successful implementation of the model. Before the process of modeling begins data analysis is needs to be done. This process consists of observing, cleaning and modeling data with the goal of extracting useful information, arriving at a conclusion. As part of preprocessing of data we have used some of the technique as outliers detection, data segmentation, data categorization. Statistical Analysis empowers to validate the assumptions, hypothesis and test them utilizing standard statistical models. In this paper we have discussed several predictive models such as logistic regression and decision tree to predict the target variable. Once the model is built we can deploy its results into everyday decision making process. We have used confusion matrix to review the model performance. Apart from that we have implemented ROC curve and AUC.

Logistic Regression

Logistic regression is a technique for analysis which falls under predictive analysis and can be utilized direct analysis of dependent variable being binary. It can be used to define the relation between one dependent variable with one or more independent variables and the output can either be coded as True or False, 1 or 0.

In our data-set, we start building the model, Regressing response on all the variables. Then, starting from this maximal model, we perform backward model selection and checking variance inflation factor at each iteration, optimize the model.For finding the predictive capability, we analyze deviance residuals and the error and Z values of the final model.

Table 2: Deviance Residuals

	Deviance Residuals
Minimum	-2.0760
1st Quartile	-0.3866
Median	-0.3305
3rd Quartile	-0.2553
Maximum	2.9032

Table 3: Calculated z values for the independent varaibles

	Estimate	**Std. Error**	**z value**	Prob(> $\|z\|$)
(Intercept)	-99.336138	5.230996	-18.990	<2e-16
contactcellular	0.715658	0.003342	25.10	<2e-16
monthmar	1.158768	0.116340	9.960	< 2e-16
monthmay	-0.601780	0.055565	-10.830	<2e-16
monthnov	-0.441666	0.074972	-5.891	3.84e-09
dayofweekmon	-0.296324	0.053783	-5.510	3.60e-08
campaign	-0.048562	0.011825	-4.107	4.01e-05
pdaysContactedinfirst10days	1.362811	0.084219	16.182	<2e-16
pdaysContactedafter10days	1.203113	0.176803	6.805	1.01e-11
poutcomefailure	-0.573881	0.065389	-8.776	<2e-16
emp.var.rate	-0.742657	0.020199	-36.767	< 2e-16
cons.price.idx	1.049754	0.056247	18.663	< 2e-16
cons.conf.idx	0.033678	0.004169	8.079	6.54e-16

In Logistic regression model, distribution is done according to the standard binomial family of distributions,whose dispersion parameter is 1. Null deviance: 20299 with 28831 degrees of freedom. Residual deviance: 16016 on 28819 degrees of freedom. AIC:16042. Number of Fisher Scoring iterations conducted for achieving convergence: 6.

Decision Tree

A decision tree is another predictive model that can be imagined in a tree frame-like structure. The decision tree is derived from different decision connected on a variable to open them to right and left splits on the tree. The decision tree should give a general decision to a set of variables not very specific to a unique variable, in this case, the decision tree made has over-fitted the data.

In this decision tree we are predicting response (yes or No) and we are using rpart library for that in which we are taking all the attribute against the target variable classification. Final output is classification in this two categories with some impurities in it. Rpart considered only few attributes to construct the decision tree. Tree starts with nr.employed attribute and gives condition as values having more than 5088 and less than that. it goes on using different variables to derive the leaf nodes. Other two attributes are duration and pdays. Other thing we can observe from decision tree is how the data values is distributed. For example 88% values of nr.employed is greater than 5088. And 81% data has nr.employed higher than 5088 and duration less than 607.

Below is the decision tree created using **rpart** library in R :

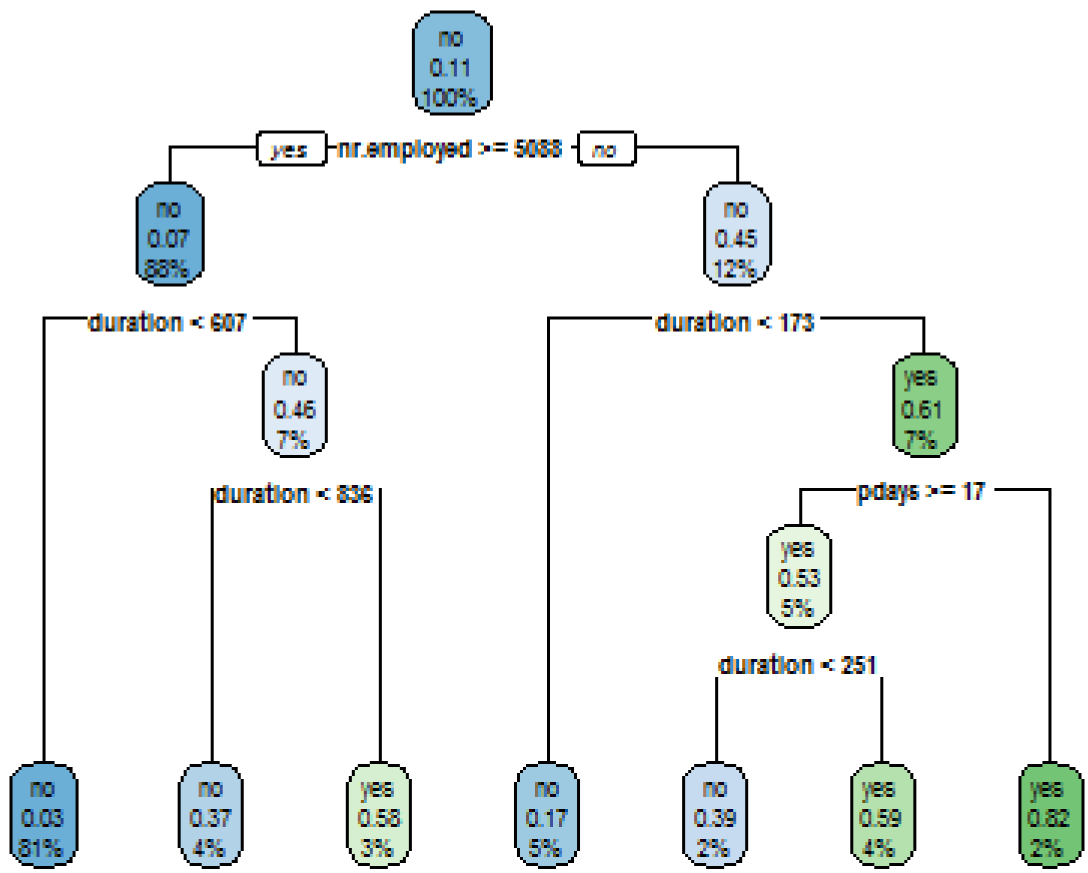

Figure 5: Decision Tree

In 2005, Rossi,Allenby, and McCulloch[11] proved that Markov Chain Monte Carlo methods worksquet efficiently than Bayesian Models. Afterwards, in 2010 Kaplan and Haenlein[12],research found an important relation of customer-firm relationship. Using which decision tress came into the picture and was proved more accurate model to use. These are the paper references and motivation behind the choice of techniques in our project.

4. Results

The percentage(%) of predictions correctly done using the model when compared with the real/actual classification is defined as Accuracy Measure. Hence, it's the way to check model's ability to classify a label correctly in unseen test case. Accuracy is calculated in two ways: i. Categorical data(classification): The rate of time in which a case will be correctly labeled with the correct category. ii. Continuous: The average distance between the correct value and predicted label. Our Problem is a classification problem, where we want to analyze, whether the person would open account in the financial institution or not.

Analyzing the dataset, it is found that students and retired personnel's are least interested in opening a long term deposit in a bank.Also, Campaigns done during the months of April, June, July, October are least effective, and if a Targeted customer has been contacted more than 3 times before then the target would most likely wont be opening a long term deposit in the bank.

4.1. Logistic Regression

Using different values for cutoff can result in different predictions. We analyze the output for different cutoff. We initially start with point 0.50 cutoff. Using this cutoff, we predict 11914 "No" output and 442 "Yes" output. The accuracy of the model using 0.5 cutoff is 90%. The confusion matrix for this cutoff is as shown below:

```
Confusion Matrix and Statistics

          Reference
Prediction    no   yes
       no  10821  1093
       yes   143   299

               Accuracy : 0.9
                 95% CI : (0.8945, 0.9052)
    No Information Rate : 0.8873
    P-Value [Acc > NIR] : 3.456e-06

                  Kappa : 0.2874
 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.21480
            Specificity : 0.98696
         Pos Pred Value : 0.67647
         Neg Pred Value : 0.90826
             Prevalence : 0.11266
         Detection Rate : 0.02420
   Detection Prevalence : 0.03577
      Balanced Accuracy : 0.60088

       'Positive' Class : yes
```

Figure 6: Confusion matrix

We find sensitivity, specificity and accuracy for different cutoff values and is as shown below.

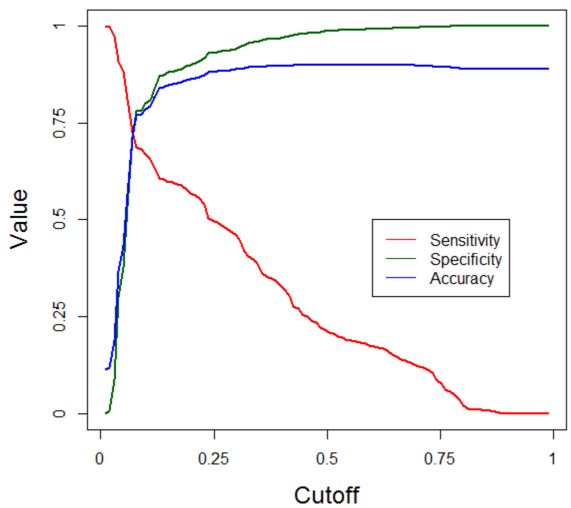

Figure 7: Cutoff Vs Values (Accuracy, Sensitivity, Specificity)

At cutoff value 0.5, we have a high accuracy and specificity but our model is not good at predicting True positive outcomes as the sensitivity suffers.So, We select a Cutoff value, where there isn't any bias toward sensitivity and specificity. From, graph we can see that this occurs for cutoff value 0.1125. For this cutoff value, the confusion matrix is as shown below:

```
Confusion Matrix and Statistics

          Reference
Prediction   no  yes
       no  8984  490
       yes 1980  902

               Accuracy : 0.8001
                 95% CI : (0.7929, 0.8071)
    No Information Rate : 0.8873
    P-Value [Acc > NIR] : 1

                  Kappa : 0.3186
 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.6480
            Specificity : 0.8194
         Pos Pred Value : 0.3130
         Neg Pred Value : 0.9483
             Prevalence : 0.1127
         Detection Rate : 0.0730
   Detection Prevalence : 0.2332
      Balanced Accuracy : 0.7337

       'Positive' Class : yes
```

Figure 8: Confusion Matrix with 0.1125 cutoff value.

As we can see, accuracy decreases due to the trade-off in bias(sensitivity vs specificity). But our model is more robust in producing unbiased and optimal results.Below graph represents the ROC curve for our model.

prospect_id	duration	response	prob_of_response	pred_response	cost_of_call
39337	136	yes	0.8840965	yes	5.288
39345	187	yes	0.8840965	yes	6.971
39349	272	yes	0.8840965	yes	9.776
39160	153	yes	0.8790268	yes	5.849
39197	498	yes	0.8790268	yes	17.234
39315	910	yes	0.8790268	yes	30.830
39354	128	no	0.8790268	yes	5.024
39165	221	yes	0.8737671	yes	8.093
39225	212	yes	0.8737671	yes	7.796
39257	422	yes	0.8667022	yes	14.726
39292	263	yes	0.8667022	yes	9.479
39176	456	yes	0.8609913	yes	15.848
39379	277	yes	0.8609913	yes	9.941
39269	192	yes	0.8501126	yes	7.136
39199	264	yes	0.8438191	yes	9.512
39324	173	yes	0.8438191	yes	6.509

Figure 8: Confusion Matrix with 0.1125 cutoff value.

The given tables illustrates actual response and predicted response with probability of it. For example first observation has actual response as Yes and it was correctly predicted as Yes with 88% probability.

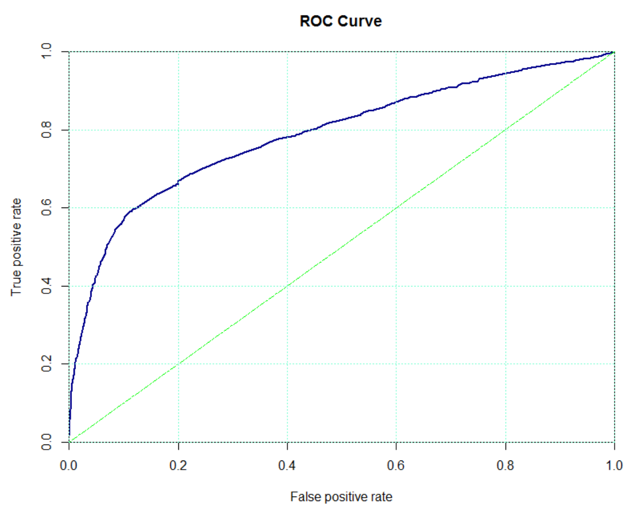

Figure 9: ROC Curve

Area Under Curve(AUC) calculated on above ROC is 0.75014, which represents that it is a fairly good model.

4.2. Decision Trees

Performance Evaluation: Accuracy Measure and Confusion Matrix
On testing our Decision Tree, we got the following results: Confusion matrix represent the performance of predictive model. In this case we can see that 8911 data values were "No" and they were correctly predicted as "No" and 412 data values were "No" but they were wrongly predicted as "Yes". Same way it works for "Yes" class. On testing the model on Test data set, we find our model to be 76.03% accurate. The confusion matrix is as shown below:

Table 4: Confusion Matrix

	No	**Yes**
No	8911	412
Yes	2549	485

We calculate specificity and sensitivity from confusion matrix and find the value of sensitivity to be 0.1598 and the value of specificity to be 0.955. Apart from that we have plotted graph of ROC which basically represent True Positive Rate (Sensitivity) VS True Negative Rate (1 - Specificity).

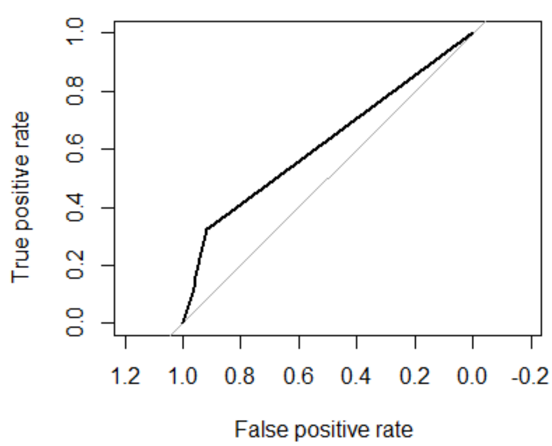

Figure 10: ROC Curve for decision Tree

AUC (Area Under the Curve): 62.04%
AUC is used to determine that which model gives the best output. If AUC is more than 50%, it means that model is better than any random model. 90% AUC is considered as a good model and 98-99% is too good to be true and model would be having problem of over fitting in this case.

5. Conclusion

Logistic Regression was able to correctly predict the output with 80% accuracy. While decision tree was able to correctly predict the output with 76% accuracy.Hence, after successful implementation and optimization of logistic regression and Decision Tree, we can now compare their results. Area Under Curve (AUC) for logistic regression is greater than AUC for Decision Tree, which states that Logistic Regression outperforms Decision tree for prediction of Target variable on bank data. Still, We need to optimize the sensitivity vs specificity using various techniques for decision tree and then compare the output. Using a small subset of this data-set, testing can be done on more computationally expensive machine learning algorithms like SVM or neural networks and compare the models. So future work would be analyzing SVM and Neural Network model to compare their accuracy and more optimal model for the dataset.

6. References

[1] Mahboubeh Khajvand, Kiyana Zolfaghar, Sarah Ashoori, Somayeh Alizadeh. Estimating customer lifetime value based on RFM analysis of customer purchase behavior: case study. Procedia Computer Science 3 (2011) 5763

[2] Razieh qiasi, Malihe baqeri-Dehnavi, Behrooz Minaei-Bidgoli, Golriz Amooee. Developing a model for measuring customers loyalty and value with RFM technique and clustering algorithms. TJMCS Vol. 4 No. 2 (2012) 172 181

[3] Li, W., Wu, X., Sun, Y. and Zhang, Q., 2010. "Credit Card Customer Segmentation and Target Marketing Based on Data Mining", In Proceedings of International Conference on Computational Intelligence and Security, 73-76.

[4] Aptea, C. and Weiss, S. 1997. "Data mining with decision trees and decision rules", Future Generation Computer Systems 13, No.2-3, 197210.

[5] Cortez, P. and Embrechts, M. 2011. "Opening Black Box Data Mining Models Using Sensitivity Analysis. In Proceedings of IEEE Symposium on Computational Intelligence and Data Mining (Paris, France), 341-348.

[6] Zhang, H. 2004. "The Optimality of Nave Bayes", In Proceedings of the 17th FLAIRS conference, AAAI Press.

[7] Coppock, D. 2002. Why Lift? Data Modeling and Mining, Information Management Online (June).

[8] 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

[9] S. Moro, R. Laureano and 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, 2011. EUROSIS.

[10] Laura Orfanedes, Fiveworx, Linda Dethman, Jamie Lalos, Fiveworx, Charting the Future: How to Use Customer Engagement Strategies to Ensure Energy Savings and Persistence, 2016 ACEEE Summer Study on Energy Efficiency in Buildings.

[11] Rossi, Peter E., Greg M. Allenby, and Rob McCulloch (2005), Bayesian Statistics and Marketing. Hoboken, NJ: John Wiley.

[12] Kaplan, Andreas M. and Michael Haenlein (2010), "Users of the World, Unite: The Challenges and Opportunities of Social Media," Business Horizons, 53 (1), 59-68.

[13] Bank Marketing Data Set
URL: "https://www.kaggle.com/janiobachmann/bank-marketing-dataset"

[14] Ling, X. and Li, C., 1998. "Data Mining for Direct Marketing: Problems and Solutions". In Proceedings of the 4th KDD conference, AAAI Press, 7379.

[15] Ou, C., Liu, C., Huang, J. and Zhong, N. 2003. "On Data Mining for Direct Marketing". In Proceedings of the 9th RSFDGrC conference, 2639, 491498.

[16] Page, C. and Luding, Y., 2003. "Bank managers direct marketing dilemmas customers attitudes and purchase intention". International Journal of Bank Marketing 21, No.33, 147163.

[17] Turban, E., Sharda, R. and Delen, D. 2010. Decision Support and Business Intelligence Systems 9th edition, Prentice Hall Press, USA.

[18] Witten, I. and Frank, E. 2005. Data Mining Pratical Machine Learning Tools and Techniques 2nd edition, Elsevier, USA.

[19] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. 2000. CRISP-DM 1.0 - Step-by-step data mining guide, CRISP-DM Consortium.

[20] Kohavi, R. and Provost, F. 1998. "Glossary of Terms", Machine Learning 30, No.23, 271274.

[21] Fawcett, T. 2005. An introduction to ROC analysis, Pattern Recognition Letters 27, No.8, 861874.

[22] Hu, X. 2005, "A data mining approach for retailing bank customer attrition analysis", Applied Intelligence 22(1):47-60.