

Enhancing Direct Marketing using Data Mining: A Case of YAA Asantewaa Rural Bank Ltd. in Ghana

Anokye Acheampong Amponsah
(Corresponding Author)
Department of Computer Science
Kwame Nkrumah University Science and
Technology

Kwaku Agyepong Pabbi
Department of Computer Science
Kwame Nkrumah University Science and
Technology

ABSTRACT

With fields like the banking industry, data are being generated massively on a regular basis. So Managers and Administrators are finding ways to turn these data into very valuable information. Marketing risk is one of the most serious problems of every bank. Fortunately, all marketing promotions are highly dependent on the data about customers stored in electronic format. Data mining, machine learning and artificial intelligence promise the use of models to go through and analyse huge data that are very difficult for human brains. This study is conducted to demonstrate with practical methods, experiments and dataset that data mining can be used to assist in direct marketing. The classifiers used are J48 decision tree and Naïve Bayes. The University of Waikato open source data mining software (Weka) is used to perform all the experiments. Confusion matrix is used to calculate the accuracy, sensitivity and specificity which are used to evaluate the performance of the classifiers. Receiver Operating Characteristic curves are also used to pictorially display the performance of the model. The results of the experiments show DT performs better than NB. It also denotes with high precision that the models can be used in detecting prospects for marketing campaigns. The decision tree and the Naive Bayes classifiers produced an accuracy of 92.5% of 91.6% respectively.

General Terms

Confusion matrix, Receiver Operating Characteristics Curve, Entropy, Information Gain, CRISP-DM,

Keywords

Direct Marketing, Data Mining, Decision Tree, J48, banking sector

1. INTRODUCTION

All over the world companies and organizations generate huge amounts of data on daily basis which are accumulated automatically or manually by clearing a customer with barcode reader, filling a form on a webpage, clicking on a button to save the entered data, clicking on a link to other websites or accessing corporate databases using an android app on a smartphone. Ghana is considered one of the developing countries in West Africa, with numerous organizations and institutions that generate a lot of data on a regular basis. These institutions/organizations include banking, financial, educational, supermarkets, shopping malls, insurance companies, non-governmental organizations and security agencies. These institutions are operating in a different realm with the development of ICT [1]. Yaa Asantewaa Rural Bank Ltd is located at Ejisu in the Ashanti region of Ghana. The Bank has been in operation since 2012 and has been serving its customers with a number of loan

products. There is marketing team (staffs) which develops a weekly “root plan” that guides them during their direct marketing process as to who to contact, what product to sell and whom to sell to, the route to take and so on. This process is aimed at selling the Bank’s products to the community people whiles expanding the customer base by acquiring new customers. Although the Bank cannot state the exact amount of money spent on this campaign as there is no specific documentation covering the activities, it is considered to be expensive since prospective customers who reject the offered products are contacted several times more until there is no reason to contact them anymore. One problem of this marketing method is that, it is not efficient and cost effective as about 50% or less of the prospects approached, only respond by purchasing a product. There are other rural and private banks in the district which also dispatch their mobile bankers to sell products. With this, it is quite obvious that marketing decisions on how to pinpoint customers who have high tendency of purchasing a loan so as to limit the number of contact per person are very crucial for the bank in the district. In order to stay on top of the competition, this study is undertaken to use a data mining model to detect prospects who are more likely to respond and by doing that all scares marketing resources would be directed to them.

2. DATA MINING AND DATA MINING TECHNIQUES

Data Mining (DM) can be defined as the method of evaluating big data from different sections and generating useful information. The basic purpose of DM is helping companies to make good use of their data by discovering buried links among a group of customers and also identifying fresh and innovative connection between data [2].

Data mining techniques can be grouped into six (6) main classes according to [3] and these are;

- Classification
- Estimation
- Prediction
- Affinity grouping or association rules
- Clustering
- Description and visualization

3. THE CROSS - INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM)

This work is guided by CRISP-DM which is a framework for guiding data mining projects. When followed carefully, there

is a higher chance of successfully completing any data mining project [4]. It is made up of six main phases.

3.1 Business Understanding

After conducting series of interviews with the bank authorities, it was realized that the bank spends more money when selling its products to customers and trying to win more customers. 50 per cent or less of the prospects approached only accept the offered products. Therefore the business objective is to identify the number of prospects who responded so as to use their characteristics to target future prospects during marketing campaigns. The effect of this is to limit the scope of marketing and enhance administrative focus.

3.2 Data understanding

The dataset used for this study is a survey data. The community folks were contacted by the bank's mobile bankers to sell a loan product to them. Some of them responded during the campaigns, others responded later and most of them did not respond at all. All the activities were recorded. The dataset contains 1000 record, 12 attributes but only 10 were used by the model because of privacy issues. Table 1 is the structure and description of the dataset. No prospect was contacted more than once.

Table 1: Description of the dataset

No.	Attribute	Type	Values
1	Name	Cat	NA
2	Gender	Cat	F-female; M-male
3	Age	Cat	20-30; 31-40; 41-50; 51-60; 61-70; 71-100
4	Marital status	Cat	S-single; M-married
5	Occupation	Cat	Trader; Carpenter; Driver; Tailor; Farmer; Pastor; Nurse; Mechanic; Teacher; Seamstress
6	Children	Numeric	1; 2; 3; 4; 5; 6; 7; 8
7	Income	Cat	Low; Medium; High
8	Education	Cat	Primary; Secondary; Tertiary
9	House	Cat	N-no; Y-yes
10	Contact Period	Cat	First; Second; third; Fourth
11	Residence	Cat	NA
12	Class	Cat	N-no; Y-yes

3.3 Data Preparation

Taking into consideration the privacy of data mining and confidentiality issues associated with customer data, some explicit attributes that could have been used to uniquely identify participants were removed (Names and addresses). There are no missing values and misspelt entries (eg. "taecher" instead of teacher) are taken care of. Also, the Age attribute are changed into ranges. The Income attribute is also encoded to protect the actual earnings of participants. Low is any amount less or equal to 500, medium is any amount between 500 and 1000 and lastly, high is any amount above 1000. The currency is Ghana Cedis.

3.4 System Modeling

At this point, a number of models are used so as to detect the model that does the best classification. As a result, Naïve Bayes, Support Vector Machine (SVM) and J48 decision tree classifiers are chosen. J48 decision tree yields the highest results and as a result, is chosen for the classification. The SVM classifier did not run at all. All the experiments were done with Weka data mining software.

3.5 Deployment

The deployment of the model is the exclusive decision to be made by the administrators of the bank (Main Manager, Branch Manager, Loan Officer and Information Technology Head). However, there is a high anticipation that this study will enhance the weekly "root" plan development and go a long way to enhance their process of direct marketing.

4. DECISION TREE MODELING AND PERFORMANCE EVALUATION

4.1 Decision Tree Modeling

Decision tree works by transforming the information space into a structure that can be used to classify and predict new instances. Despite the development of many decision tree classifiers so that the small group of attributes needed to build the classifier are obtained, it is difficult to find the best tree structure for any given data set [5] and [6]. However, all DTs that employ C4.5 algorithm are based on ID3 that was developed by Quinlan John Ross. ID3 uses a core that operates by using Entropy and Information Gain [7] but according to [8], these two terms describe decision tree algorithms.

4.2 Entropy Reduction or Information Gain

Information gain defines a brilliant notion for ensuring purity in leaves. If a leaf is totally pure, then all the classes in the leaf can be easily described and understood. If a leaf is considered as highly impure, it means all classes in the leaf are complex and difficult to understand [8]. Entropy may be discussed as a number of Y/N (yes or no) questions that should be answered to define the form of a system. If there exist eight (8) states of a system, it will take $\log_2(8)$ or three bits to number all of them or to uniquely distinguish one [9]. Extra information which will decrease the entropy by reducing the number of questions that needs to be answered in order to define the form of the system is termed as information gain [9]. In this view, information gain can be said to mean the same as entropy reduction [8]. Initial ID algorithms used seat-of-the-pants method to perform inductions but Peter Gacs recommended the use of information base technique which is influenced by two assumptions [10]. These are:

1. A correct DT constructed for D will categorize the same objects as part of D. A random object that is chosen by probability will belong to class M with $m/(m+t)$ and also to class T with $t/(m+t)$.
2. If a DT is adopted to group a set of items, the outcome will be a class. Therefore a DT can be considered as a message source of 'M' or 'T', with the information required to produce the message stated by;

$$I(m, t) = - \frac{m}{m+t} \log_2 \frac{m}{m+t} - \frac{t}{m+t} \log_2 \frac{t}{m+t} \quad (1)$$

If attribute P contains items (P₁, P₂, ..., P_n) and it is implemented as the root node of the DT, it will divide D into (D₁, D₂, ..., D_n) where D_i holds all the objects in D which have P_i of P. Let D_i contain m_i items of class M and t_i items of class T the required information for the sub tree for D_i is I(m_i, t_i). Then the required information for the tree which has P as root node will be calculated as weighted average

$$E(P) = \sum_{i=1}^n \frac{mi + ti}{m + t} I(mi, ti) \quad (2)$$

where the value for the *i*th section is the part of items in D that exists for D_i. Therefore by branching on attribute P the information gained is calculated as

$$Gain(P) = I(m, t) - E(P) \quad (3)$$

4.3 Performance Evaluation

4.3.1 Confusion matrix

Most data miners use this method to evaluate classification models. It comprises of evidence of actual and predicted classifications made by a classification model [11]. The two classifiers that ran are the J48 DT and NB. Looking at the values in table 2, the J48 DT correctly classifies 600 negative items as negative and 325 positive items as positive. Also 47 positive items are wrongly classified as negative. The same applies to 28 negative items that are wrongly classified as positive. The NB also classifies 536 positive items as positive and 313 negative items as negative. Also 59 negative items are wrongly classified as positive likewise 92 positive items as negative. These values are used to calculate the accuracy, specificity and sensitivity of the two classifiers in table 3.

Table 2: Confusion matrix for the J48 DT and NB

		PREDICTED			
		J48		NB	
		Negative	Positive	Negative	Positive
ACTUAL	Negative	600	28	313	59
	Positive	47	325	92	536

4.3.2 Accuracy, Specificity and Sensitivity

The accuracy of a classification algorithm is termed as the proportion of the number of instances that are classified correctly which is equal to TP plus TN divided by the total amount of instances in the information space N [12].

$$Accuracy = \frac{TP + TN}{N} \quad (5)$$

Sensitivity points to the number of positive instances that are correctly classified and it is equal to TP divided by TP plus TN [13].

$$Sensitivity = \frac{TP}{TP + FN} \quad (6)$$

Specificity refers to the proportion of negative instances that are correctly classified and it is equal to the proportion of TN divided by TN plus FP [12].

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

From table 3, it is apparent that J48 DT performs better than the NB in terms of the evaluation parameters.

Table 3: Accuracy, sensitivity and specificity of J48 DT and NB

	Accuracy	Sensitivity	Specificity
J48	0.925	0.8737	0.9554
NB	0.849	0.6313	0.3907

4.3.3 Receiver Operating Characteristic Curves and detailed Accuracy by class of J48 and NB

Recall is the same as sensitivity. It refers to part of real positive instances that are predicted correctly as positives. The values obtained in table 4 are calculated by Weka. It is very important if Precision - Recall curve model performance evaluation are to be calculated. However they are used in the generation of ROC curves.

Table 4: Precision, recall and F-Measure of J48 DT and NB

	Precision	Recall	F-Measure	Class
J48 DT	0.927	0.955	0.941	No
	0.921	0.874	0.897	Yes
Weighted Avg.	0.925	0.925	0.925	
NB	0.901	0.854	0.877	No
	0.773	0.841	0.806	Yes
Weighted Avg.	0.853	0.849	0.850	

The DT produces ROC value of 0.925 in figure 1a and NB produces 0.9162 in figure 1b. These mean that there are 92.5 per cent and 91.6 per cent chances the two classifiers will make correct classification in future. The ROC value for a perfect classifier is 1.

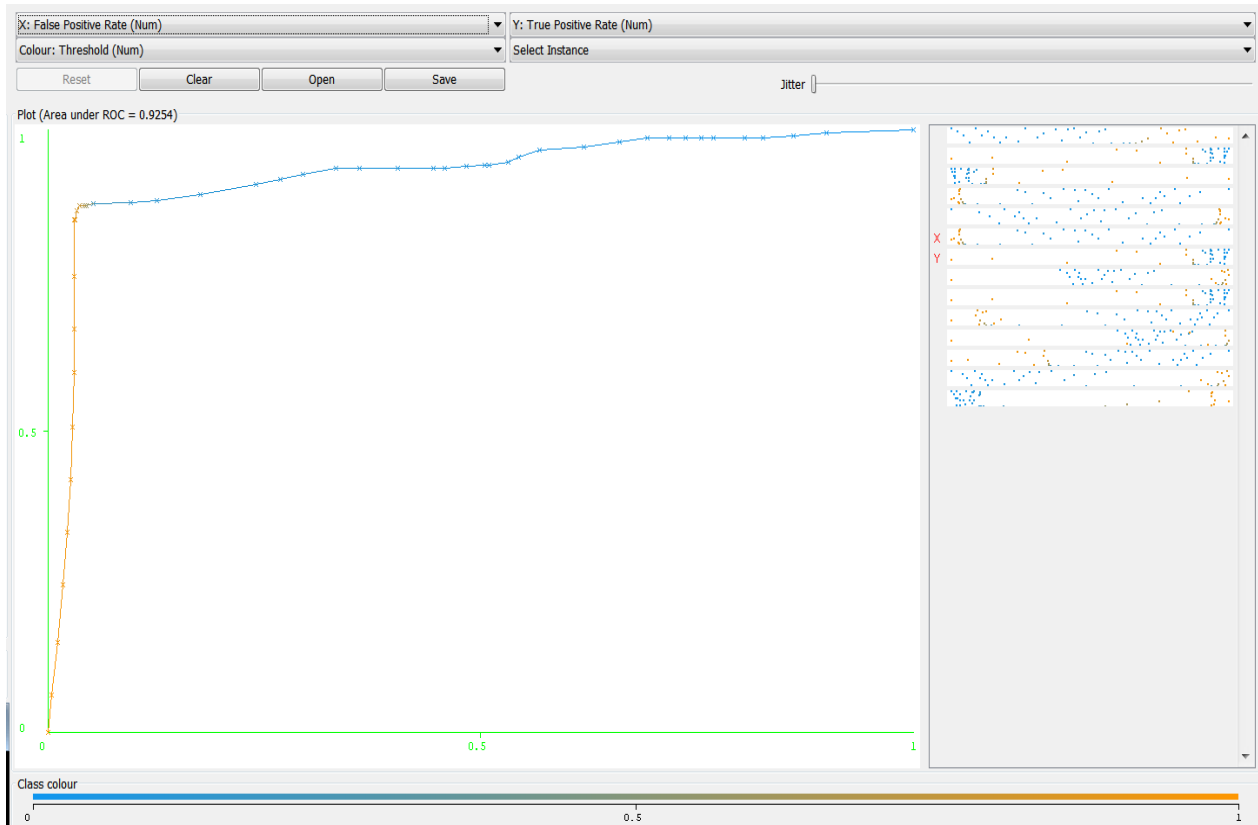


Figure 1a: ROC curve for the J48 Decision tree Classifier

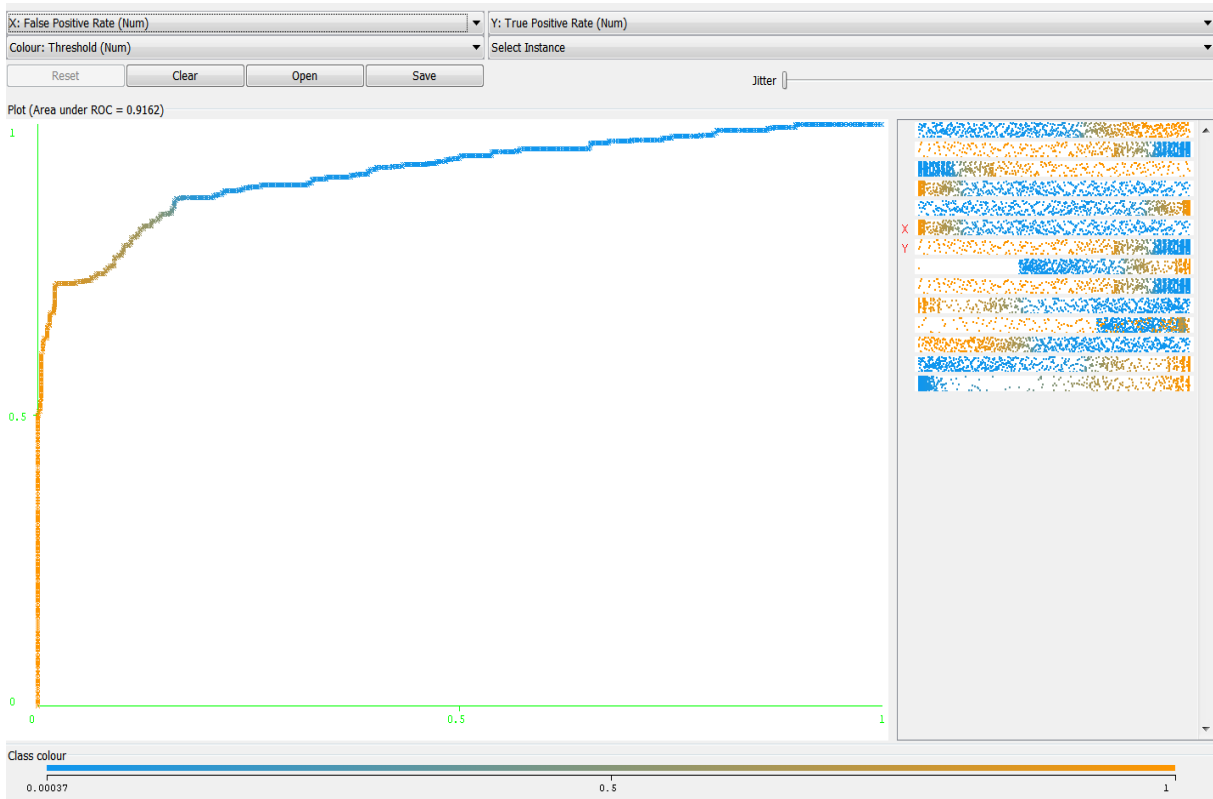


Figure 2b: ROC curve for the Naïve Bayes Classifier

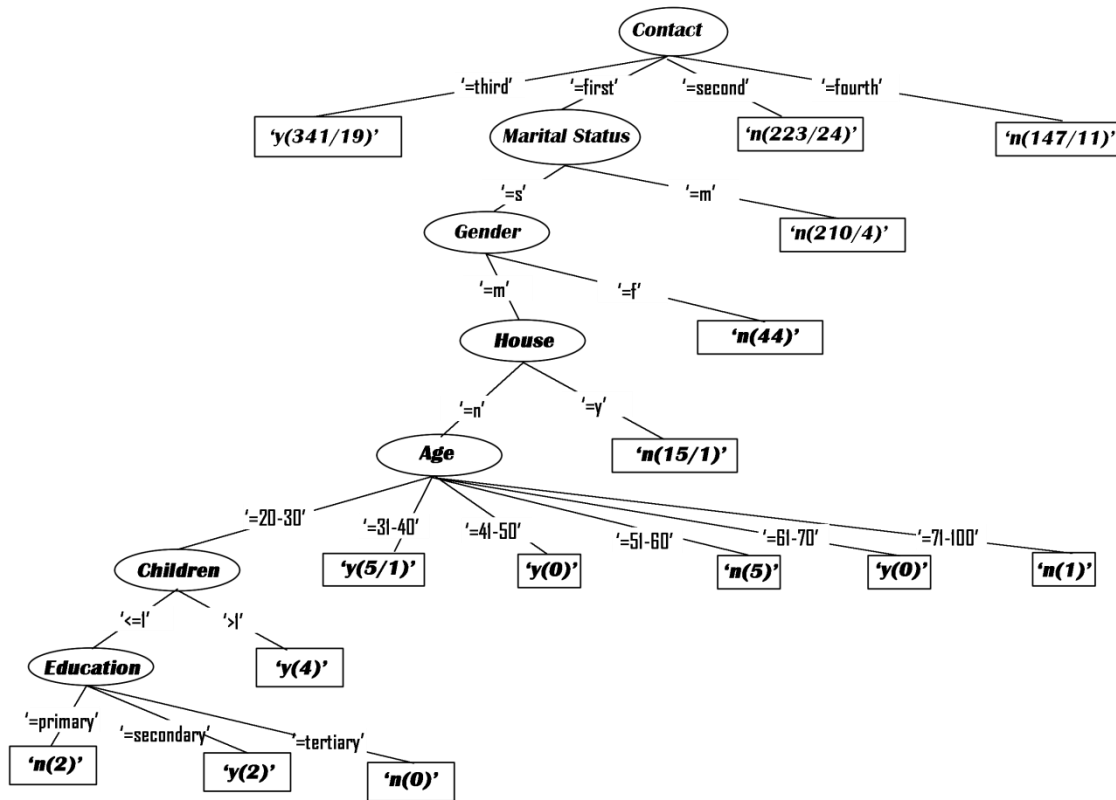


Figure 3: The Decision tree generated to make classification to enhance direct marketing

5. EXPERIMENTAL RESULTS AND ANALYSIS

The number of leaves of the DT classifier is 15, the tree size is 22 and it took 0.02 seconds to build the classifier. The number of correctly classified records is 925 instances representing 92.5 per cent. The number of incorrectly classified instances is 75 which represent 7.5 per cent. The Kappa statistics is also 0.8378. the Kappa value is considered as “almost perfect agreement” by [14]. The NB correctly classified 849 and incorrectly classified 151 instances.

5.1 Statistics for J48 Decision tree

Number of Leaves: 15

Size of the tree: 22

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 925 92.5 %

Incorrectly Classified Instances 75 7.5 %

Kappa statistic 0.8378

Mean absolute error 0.1237

Root mean squared error 0.2581

Relative absolute error 26.4637 %

Root relative squared error 53.3959 %

Total Number of Instances 1000

The generated tree used to make classification for the identification of prospects is shown in figure 2.

5.2 Statistics of Naïve Bayes

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 849 84.9 %

Incorrectly Classified Instances 151 15.1 %

Kappa statistic 0.6826

Mean absolute error 0.1582

Root mean squared error 0.3318

Relative absolute error 33.8591 %

Root relative squared error 68.6379 %

Total Number of Instances 1000

5.3 Classification rules

These classification rules are generated after the J48 decision classifier is used. They can be transformed into simple if statements.

contact = third: y (341.0/19.0)

contact = first

| marital status = s

| | gender = m

| | | house = n

| | | | age = 20-30

| | | | | children <= 1

| | | | | | education = primary: n (2.0)

| | | | | | education = secondary: y (2.0)

| | | | | | education = tertiary: n (0.0)

| | | | | children > 1: y (4.0)

| | | | | age = 31-40: y (5.0/1.0)

| | | | | age = 41-50: y (0.0)

| | | | | age = 51-60: n (5.0)

| | | | | age = 61-70: y (0.0)

| | | | | age = 71-100: n (1.0)

| | | | | house = y: n (15.0/1.0)

| | | | | gender = f: n (44.0)

| | | | | marital status = m: n (210.0/4.0)

contact = second: n (223.0/24.0)

contact = fourth: n (148.0/11.0)

5.4 Attributes and Percentage of contribution

With respect to the J48 classifier, the three attributes that contribute massively in the classification are **Contact**, **Education** and **Age** with the results graphed below in Figure 3. As a result of this, the results analyses and discussion would be made around the three most contributing attributes.

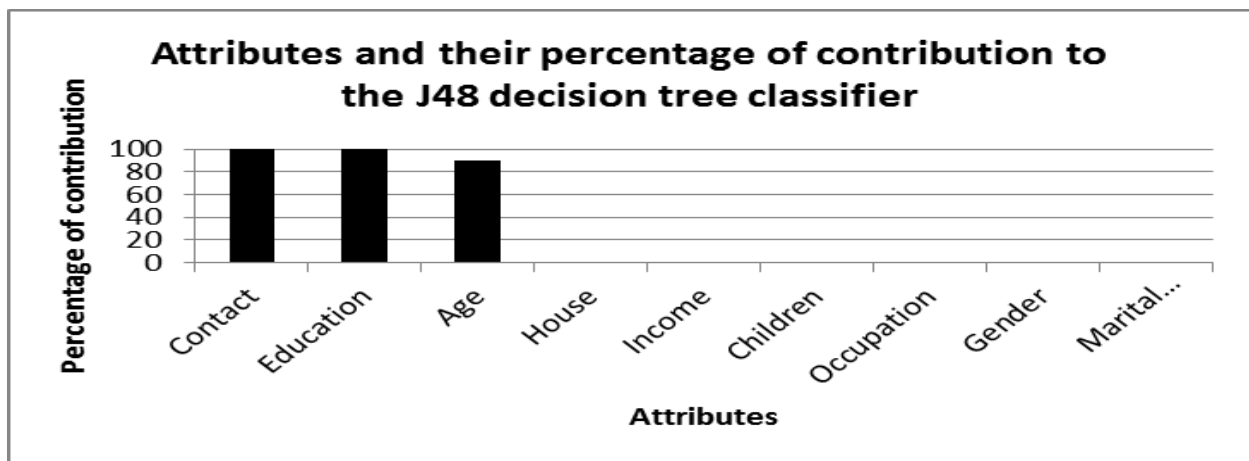


Figure 4: Attributes and their percentage of contribution to the J48 decision tree classifier

It is revealed by the model that majority of the prospects who were contacted in the third week responded to purchase the loan product. Out of the 341 contacts made, 322 (94.43%)

responded and 19 (5.57%) did not. Comparing the findings, it could be analyzed that prospects who are contacted in third weeks are more likely to respond to the offer. The next highest contacts are made in the first week which are 288 contacts and for this, 15 (5.21%) responded and 273 (94.79%) did not. The second week contact records 24 (10.76%) positive responses and 199 (89.24%) negative responses. Likewise the fourth week contact also records 11 (7.43%) positive responses and 137 (92.57%) negative responses. Therefore it is better for the bank to market its loan products in third weeks.

5.4.1 Contact Attribute

The prospects were approached on four different weeks in one month – first, second, third, and fourth week. It is discovered that most of the prospects were contacted in the third week with the total value of 341 contacts.

5.4.2 Education Attribute

All the prospects contacted are grouped into three different categories for the education attribute – Primary, Secondary

and Tertiary. The results indicated that most of the people (448) contacted had secondary school as their highest educational level at the time the campaign was being done.

The rest are having 284 and 268 people having primary and tertiary as their highest educational level respectively.

It is also discovered that majority of the people who responded to the loan product offer had tertiary education. Out of the 268 prospective customers contacted, 186 (69.40%) responded positively while 82 (30.60%) responded negatively. On the other hand, the group that had primary as their highest level of education recorded 88 (30.90%) positive responses and 196 (69.10%) negative responses. For the 448 prospects with secondary as their highest education, 98 (21.87%) responded positively and 350 (78.13%) responded negatively. Although the number of contacts made for secondary education prospects is very high, when compared with tertiary education in terms of positive responses, tertiary is significantly higher. This means that people with tertiary education have a higher possibility of responding to a loan offer and that the bank can focus on this group of people for loan marketing.

5.4.3 Age Attribute

Under the age attribute, there are six categories. These are 20-30, 31-40, 41-50, 51-60, 61-70, and 71-100. It is realized that the age group that responded massively to the offer is 31-40 with 150 (50%) positive and 150 (50%) negative responses. The second group that responded quite well with 150 contacts is 51-60, which recorded 72 (48%) positive and 78 (52%) negative responses. Again the third group that also responded

quite significantly is 41-50 with 91 (24.60%) positive and 279 (75.40%) negative responses. Therefore it can be analyzed that most of the prospective customers who responded to the offer are in the age group of 31-40 followed by 51-60 and that the bank could target this group for loan marketing and make it very attractive for the local people.

5.4.4 Class Attribute

The main objective for implementing the J48 decision tree algorithm was to determine the number of prospective customers who are likely to respond when approached with an offer to buy a loan product from the bank and then their characteristics would be used in direct marketing. The model discovers that, out of the 1000 records used by the classifier 372 prospects representing 37.2 % responded positively while 628 prospects representing 62.8 % responded negatively.

6. RECOMMENDATION

In light of the general nature of data mining and the findings of this study, the following recommendations are made to help bank administrators (Managers, Loan officers) to fully leverage the benefits of data mining during marketing campaigns.

- The marketing team should market loans in third weeks rather than any other weeks.
- Prospects with tertiary as their educational level should be targeted more because they are very promising for loan marketing.
- The age group 31-40 is more likely to respond when approached to buy a loan product.

7. CONCLUSION

Using the human brain to analyze the huge data stored in databases and data warehouses is nearly impossible. With this, fields like data mining, machine learning and artificial intelligence allow the development or generation of well defined, robust and simple to use models to sort through and analyze huge data for knowledge discovery. This study has demonstrated with practical methods and experiments that data mining can be used to assist in marketing. Typically it was realized that out of the 341 contacts made in the third week, 322 responded and 19 did not respond as against 288 contacts made in the first week where 15 responded and 273 did not respond..

8. FUTURE WORK

The scope of this study is limited to only one branch of the case bank. In future, data from other branches of the case and other banks in the region would be included. With this, it is expected that very interesting discoveries could be made to enhance the marketing processes of the banks in the region. Also data mining will be leveraged to automate credit scoring and reduce other operational risks.

9. ACKNOWLEDGEMENT

Our sincere gratitude goes to the bank authorities (Main manager, branch manager, loan officer and head of I.T. unit for their cooperation and support.

10. REFERENCES

- [1] Du, X. (2006). Data Mining Analysis and Modeling for Marketing Based on Attributes of Customer Relationship
- [2] Chopra, B., Bhambri, V., Krishan, B. (2011), Implementation of Data Mining Techniques for Strategic CRM Issues, *Int. J. Comp. Tech. Appl.*, Vol 2 (4), 879-883 Available online at www.ijcta.com (Accessed Wednesday, August 26, 2015, 10:29:24 PM)
- [3] Radhakrishnan B., Shineraj G., Anver Muhammed K.M. (2013), Application of Data Mining In Marketing, *International Journal of Computer Science and Network*, Volume 2, Issue 5, October 2013, 41 – 46, available online at www.ijcsn.org
- [4] 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.
- [5] Patil, N., Lathi, R., & Chitre, V. (2012). Comparison of C5.0 & CART classification algorithms using pruning technique. *International Journal of Engineering Research and Technology* (Vol. 1, No. 4 (June-2012)). ESRSA Publications.
- [6] Sun, H., Huai, J., Liu, Y., & Buyya, R. (2008). RCT: A distributed tree for supporting efficient range and multi-attribute queries in grid computing. *Future Generation Computer Systems*, 24(7), 631-643.
- [7] http://www.saedsayad.com/decision_tree.htm (accessed Friday, October 09, 2015 09:41:36 AM)
- [8] Linoff, G. S., & Berry, M. J. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.
- [9] Fuchs, C. A., and Peres, A. (1996). Quantum-state disturbance versus information gain: Uncertainty relations for quantum information. *Physical Review A*, 53(4), 2038.
- [10] Quinlan, J.R. (1986). *Induction of Decision Trees*, *Machine Learning*, vol. 1, pp. 81-106, 1986.
- [11] Kohavi, R. & Provost, F. (1998). Glossary of Terms. *Machine Learning*, 30(2-3), 271-274.
- [12] Nisbet, R., Miner, G., & Elder IV, J. (2009). *Handbook of statistical analysis and data mining applications*. Academic Press.
- [13] Elsalamony, H. A. (2014). Bank Direct Marketing Analysis of Data Mining Techniques. *Network*, 5, 0. *International Journal of Computer Applications (0975 – 8887)* Volume 85 – No 7, January 2014
- [14] Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam Med*, 37(5), 360-363.