Predicting customer response in direct marketing using data mining

Aarushi[1*], Rainu Nandal[2], Himanshi Verma[3]
1 Research Scholar, University Institute of Engineering & Technology, Maharshi Dayanand University, Rohtak, India
2 Assistant Professor, University Institute of Engineering & Technology, Maharshi Dayanand University, Rohtak, India
3 Research Scholar, University Institute of Engineering & Technology, Maharshi Dayanand University, Rohtak, India

Abstract

Data mining is a technique to extract required information from pre existing database. In the present focused world, it has turned out to be essential for companies to apply a methodology before as well as after introducing their product in the market. This research explains the application of data mining techniques such as naive bayes algorithm and decision trees in the field of direct marketing. Direct marketing is a form of promoting your product/service in which companies communicate directly to consumers with the help of different types of media without any middleman involved. The main issue in direct marketing is that there are approximately 5% people who revert back. So, the main focus of this research is to determine those 5% people with the assistance of available historical data. By doing so, companies can save their time as well as money on spending those people who are not supposed to revert back. The next main focus is to differentiate between the results of various techniques of classification in data mining applied on the historical data available.

Keywords: *Data mining; Decision Tree; Direct marketing; Naïve Bayes; WEKA.*

1. Introduction

1.1. Data Mining

Data Mining is a technique to find new useful information or to extract new patterns or to generate some rules from the pre existing database. It is a process which involves following steps [11]:
1. Data Cleansing
2. Selection and Transformation
3. Data Mining
4. Post processing and Visualization

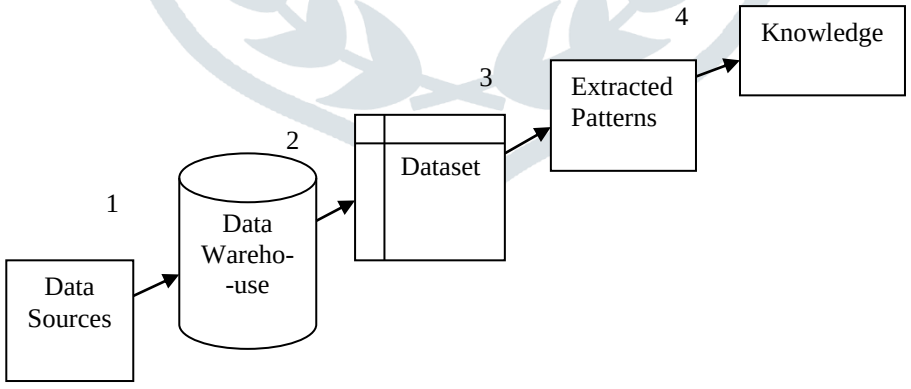


Fig. 1: Process of data mining [34]

1.2. Direct Marketing

Direct marketing is the act of promoting a service or product by presenting the details directly to the target audience without involvement of any middleman. The main focus of direct marketing is to present data to a person that has been predicted to be a purchaser. Common forms of direct marketing include:
1. Phone calls
2. Text messages
3. Emails

4. Brochures
5. Catalogs
6. Online display ads on social media
7. Newsletters
8. Post cards
9. Promotional coupons

While some marketing methods mean to expand some mindfulness or to instruct the audience about a company's items or administrations, the prime objective of direct marketing is to convince the recipient to atleast reply back. Contrasting to mass publicizing, which does not involve a target audience, direct marketing is exhibited upon individuals who are suspected to possess some interest or need in your organization's articles, on the basis of the data assembled.

It is successful because of the following reasons:

- You can make the message individual, influencing the beneficiary that he/she feels that the ad is implied only for them.
- It is more cost-effective to advertise to customers who have been identified as willing to buy.
- Higher rate of profitability is there.
- It is measurable. It keeps observing the successfulness of each campaign thus providing knowledge about how to do better next time.
- It reduces the distance between the company and the customer.

The main challenge to direct marketing is that there exists inverse correlation between the money to be spent for buying the product and the chances that a person will buy it because if a product/service is costly, less people will buy it.

1.3. Classification

It is a technique of data mining to classify various items in different known classes on the basis of similarities and dissimilarities. It is similar to arranging books in the library in different sections according to similarities (subject). Existence of various classes is known to the classifier. It involves a target, various predictors and cases.

In the training process of a model, the relationships among predictors' and the targets' values is found with the help of classification [13]. Testing of model is done by comparing the predicted values with known targets. The recorded information for an arrangement venture is commonly partitioned into two informational collections: one is used to build the model and also known as training of data; the other is used for testing the trained model.

There are many applications of classification using data mining such as in credit analysis, consumer segmentation, marketing, and biomedical, business modelling and drug response modelling, etc. In this research work distinction among various techniques of classification in data mining applied on that historical data available to predict who will response and who will not, will be done.

1.3.1. Decision Tree

Decision tree model demonstrate as a structure similar to tree. It splits a dataset into little subsets similar to how a tree splits into various branches. At last a tree with decision node as well as leaf nodes is generated. The highest node in the tree is called as root node. A root node has at least two branches. Other nodes have as many branches or no branches. These trees can deal with both categorical as well as numerical information.

How to build a decision tree? [31]

Step 1: At first ensure that data (attributes to be classified) must be discrete in value and not continuous. It is called as training data. Put all of the training data in a single node of tree.

Step 2: If all of the instances belong to same class or say that if all instances have similarities and no dissimilarity to differentiate between them then halt.

Step 3: Divide the next node into sub nodes. Among the given attributes, choose an attribute that is optimal for splitting data into subsets and in this way decision node is created.

Step 4: Node is divided on the basis of values of attribute chosen.

Step 5: If any of the following conditions meet, halt, else go to step 3:

a) When this splitting results into data that belong to a same class and there is no requirement of further split.

b) When no other attributes remained for splitting.

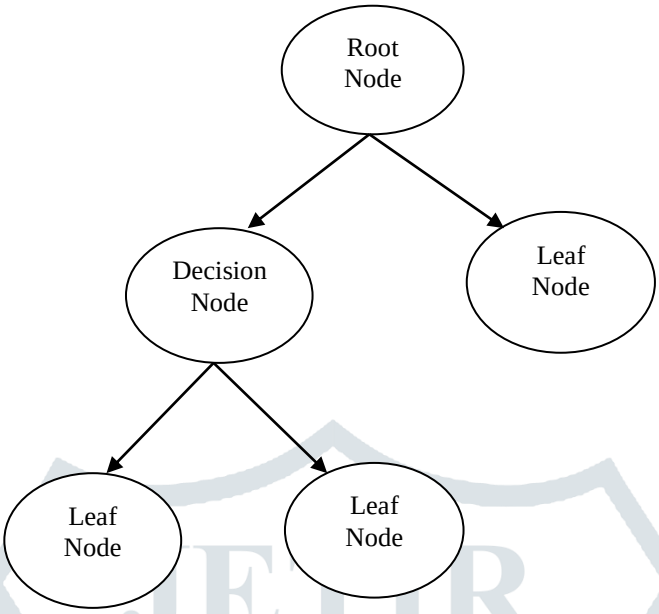


Fig. 2: Sample Decision Tree

ID3 Algorithm

Ross Quinlan has given the ID3 (Iterative Dichotomiser 3) algorithm in 1986 [23]. This classification depends on the entropy and information gain as calculated with the help of qualities. The point is to choose the best traits as a node with a specific end goal to develop the most basic choice state or choose that one which is near simple state.

Steps of ID3 algorithm are as follows [16]:

1. Entropy E(S) is calculated:

$$E(S) = - \sum_{u=1}^{c} P_u * \log_2 P_u \quad (1)$$

- S: set of all objects.
- u: represents class from 1 to n.
- c: no. of classes in S.
- P_u = (no. of objects in class u) / (no. of objects in S).

2. Sub-entropy $E^+(X)$ generated by node with attribute X is calculated:

$$E^+(X) = \sum_{v=1}^{w} (n_v / n) * E(S_v) \quad (2)$$

- S_v: subset of S with the same object attribute v.
- $E(S_v)$: entropy of S_v.
- n: total no. of the objects in S.
- n_v: no. of objects in subset S_v.
- w: total no. of subsets in S.

3. Information gain Gain (X) is calculated:

$$Gain(X) = E(S) - E^+(X) \quad (3)$$

Attribute having highest information gain is chosen for splitting at each level after the calculation of Gain (X) for each attribute.

C4.5 Algorithm

In C4.5, the decision tree model is constructed in the similar way as in ID3. Following are the improvements in C4.5 algo [30]:

- It takes care of both continuous and discrete properties - keeping in mind the end goal to deal with continuous data, C4.5 makes an edge and after that parts the divide into those whose quality is over the limit and those that are not exactly or equivalent to it.
- It deals with information with missing traits - C4.5 permits those qualities to be set as ? for missing. Missing characteristic qualities are basically not utilized as a part of information gain calculations.
- It takes care of properties with varying costs.
- It prunes trees after creation – Once the tree has been made, C4.5 visits the tree once again and it excludes the branches that don't help in getting better results by supplanting them with leaf nodes.

J48 Algorithm

C4.5 algorithm's open source Java implementation in WEKA is known as J48 algorithm used for classification [30]. J48 is one of the algorithms that are commonly used in the research.

1.3.2. Naive Bayesian classifier

Bayes' theorem is the main basis of the Naive Bayesian classifier This classifier is easy to build. It includes no complicated iterative parameter estimation and that's why it is very useful even for very large datasets.

Bayes theorem calculates the posterior probability, P(s|z), with the help of P(s), P(z), and P(z|s). An assumption called as conditional independence is taken into consideration by naive bayes classifier. Assumption is that there is no effect of other predictors on the predictor(z) on a given class(s).
Bayes theorem [15]:

$$P(s|z) = (P(z|s) * P(s)) / P(z) \quad (4)$$

where

$$P(s|z) = P(z_1|s) * P(z_2|s) *.....* P(z_n|s) * P(z|s) * P(s) \quad (5)$$

- P(s|z): posterior probability of target class when predictor (attribute) is given.
- P(s): prior probability of class.
- P(z|s): posterior probability of predictor when class is given.
- P(z): prior probability of predictor.

2. Literature Review

Various researchers told and showed that data mining can be a boon if used beforehand and after launching a new product in the market in business areas, E.W.T. Ngai, Li Xiu, D.C.K. Chau (2009) [2] and Chris Rygielski, Jyun-Cheng Wang, David C. Yen (2002) [26] and Charles X. Ling, Chenghui Li (1998) [27] and Sankar Rajagopal (2011) [3]. According to Megha Gupta, Naveen Aggarwal (2010) [20], future patterns and practices are anticipated with the help of data mining tools which enables organization to make better decisions for the organization. Hasan Ziafat, Majid Shakeri (2014) [6] told that with the help of data mining techniques patterns can be extracted that even the most skilled expert person can't. Esra Kahya Ozyirmidokuza, Kumru Uyarb Mustafa Hakan Ozyirmidokuzc (2015) [1] showed how to extract patterns from firms marketing channel. Rabiya Shamsheer, Parul Agarwal (2017) [7] surveyed different data mining functions helping in customer relationship management. Charles Dennis, David Marsland et al. (2001) [4] showed that available data about customer can be used to identify what are the needs of customers and which customers (segment) a company wants to target. A.Bharathi, E.Deepankumar (2014) [19] told that one of the data mining method known as classification is used to predict data instances' membership to a group.
Shu-Ching Wang et al. (2014) [16] told that the corporate basic analyses are for the most part gathered from the history information. A data framework is utilized to acquire the variables and to direct surveying with those components. In this investigation amid the choice procedure of the promoting, the items data and past advertising background or information are utilized to first analyse and afterwards to discover valuable, unobvious, or ineffectual strategies for direct marketing. Some researchers told that direct marketing can be done by sending emails, calling, etc. to customers directly. They had implemented decision tree for direct marketing. Algorithms such as ID3 and C4.5 were used. They also told the problems in classification for direct marketing, M Suman, T Anuradha, K Manasa Veena (2011) [5].

Chun Fu Lin, Yu Hsin Hung, Ray I Chang (2013) [17] build up an expert system in which relevant information is recognized with the help of decision tree. Some researchers told that out of total people mailed (if email is chosen for direct marketing), only 5%

people respond back. So the main problem is to determine those people with the help of historical data, Ke Wang et al. (2005) [8]. Some of them showed that the naive bayes classifier accurately chose those people who will respond to the mail sent by organization for advertisement of their product, Chris Fleizach, Satoru Fukushima (1998) [18].

3. Technique

3.1. WEKA

Tool used for data mining is WEKA which refers to Waikato Environment for Knowledge Analysis. It is written in Java language. This tool consists of inbuilt algorithms such as decision tree algorithm, clustering algorithm, association rule algorithm, etc. required for analysing the data. It has an interactive interface which makes it easy to use and view results clearly.

Table 1: Dataset's meta information

Attribute Name	Description	Type
Age	Client's age	Numeric
Job	Client's occupation	Categorical: • admin • management • entrepreneur • services • technician • blue-collar • self-employed • retired • housemaid • student • unemployed • unknown
Marital	Client's marital status	Categorical: • divorced • single • married • unknown
Education	Client's academic qualification	Categorical: • professional.course • basic.9y • basic.6y • basic.4y • high.school • university.degree • illiterate • unknown
Housing	Client has taken loan for house?	Categorical: yes or no
Loan	Client has taken loan other than for housing?	Categorical: yes or no
Contact	Communication 's type	Categorical: cellular or telephone
Month	Month when client was last contacted	Categorical: Jan-Dec
Day_of_week	Day on which client was last contacted	Categorical: Mon-Fri

Duration	Duration of contact	Numeric
Cci	Consumer confidence index	Numeric
Campaign	No. of times client is contacted during this campaign	Numeric
Cpi	Consumer price index	Numeric
Pdays	No. of days passed since client was last contacted	Numeric
Previous	No. of times client is contacted before this campaign.	Numeric
Poutcome	Output of last campaign	Categorical: • success • nonexistent • failure
Evr	Employment variation rate	Numeric
Y	Output	Categorical: yes or no

The dataset taken consisted of 41188 instances and 18 attributes which consist of information about clients as shown above in Table 1 along with their description [32]. Dataset [35] is taken from UCI Machine Learning Repository [33].

Points of interest of Weka [36]:

- It is freely accessible.
- It is portable, as written above that it is written in Java language and thus runs on any processing system.
- Information pre processing and displaying strategies are completely gathered at one platform.
- Convenience because of its graphical UIs.

Fig. 3: WEKA Startup Screen [10]

3. Results

When the classification model is trained and tested by J48 tree algorithm, following results are obtained:

Confusion Matrix resulting from J48 algorithm applied on the dataset:

a	b	← classified as
10505	380 |	a = no
729	742 |	b = yes

In a confusion matrix, instances at top left and bottom right are correctly classified instances and instances on top right and bottom left are incorrectly classified instances. As shown, 10505 are correctly classified instances as no, 742 are correctly classified instances as yes, 380 are incorrectly classified instances as yes when actually they are no and 729 are incorrectly classified as no when actually they are yes.

When the classification model is trained and tested by naive bayes algorithm, following results are obtained:

Confusion Matrix resulting from naive bayes algorithm applied on the dataset:

a	b	classified as
10216	669 |	a = no
746	725 |	b = yes

10216 are correctly classified instances as no, 725 are correctly classified instances as yes, 669 are incorrectly classified instances as yes when actually they are no and 746 are incorrectly classified as no when actually they are yes.

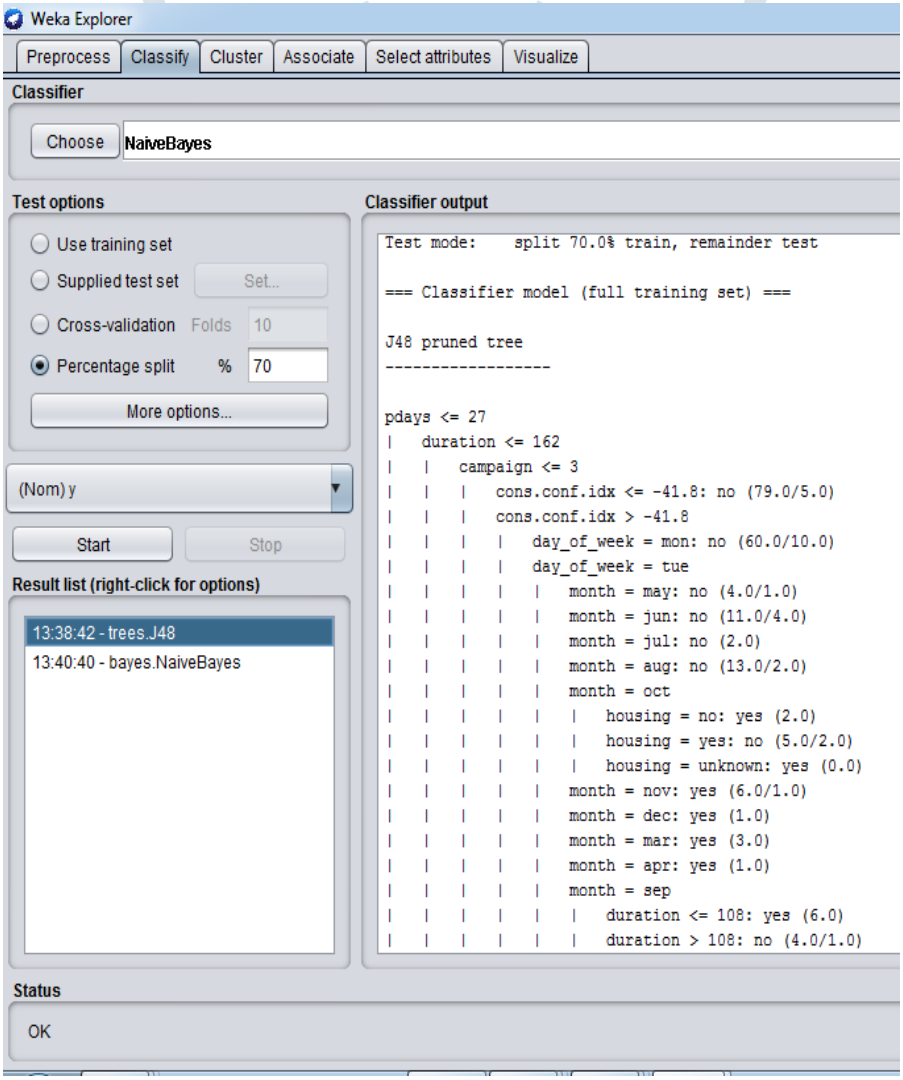


Fig. 4: Snapshot of J48 tree

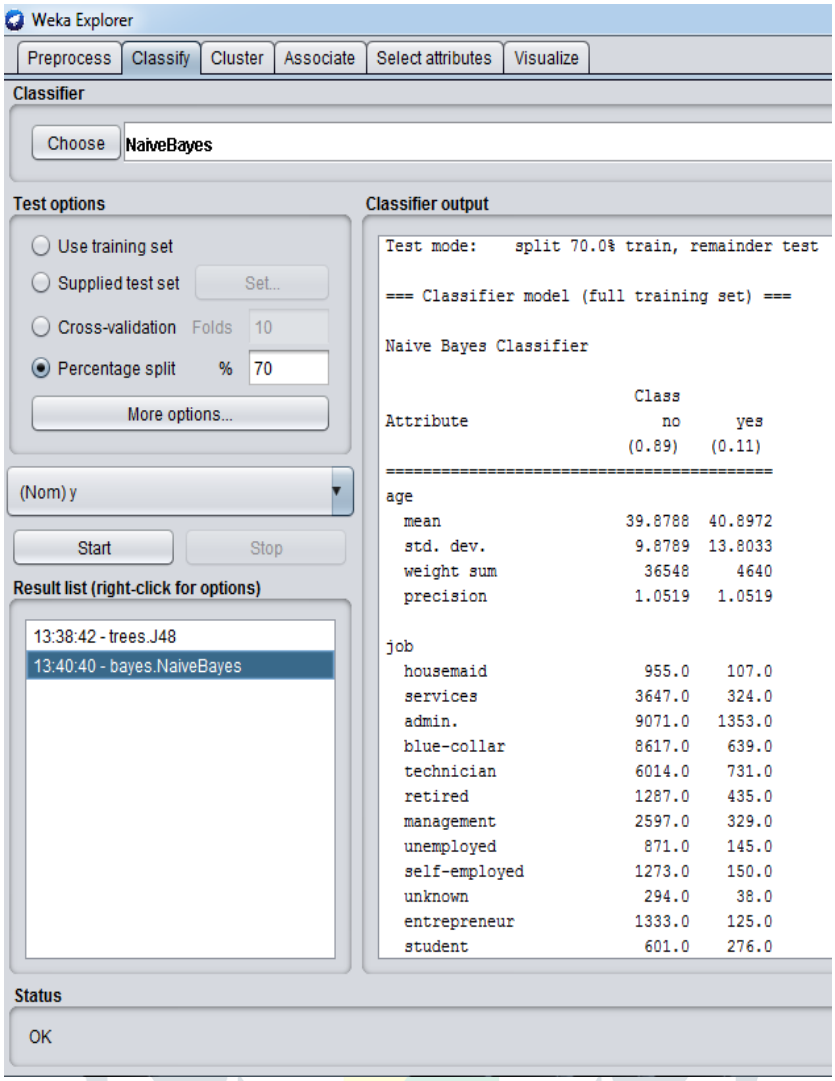
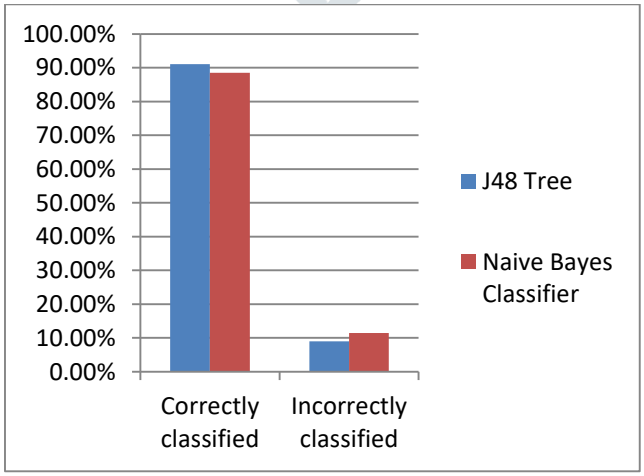


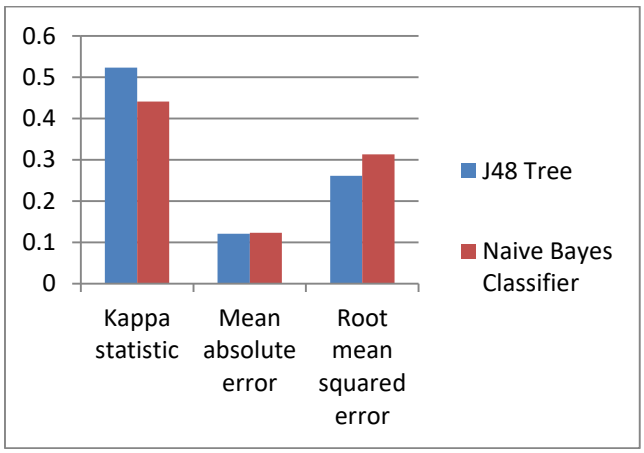
Fig. 5: Snapshot of naïve bayes classifier

As shown in Graph 1, 91.0246% instances are correctly classified by J48 tree whereas naive bayes classifier classified 88.5481% instances correctly. 8.9754% instances are incorrectly classified by J48 tree whereas naive bayes classifier classified 11.4519% instances as incorrect.

Graph 1: Proportion of correctly & incorrectly classified instances resulting from application of both algorithms

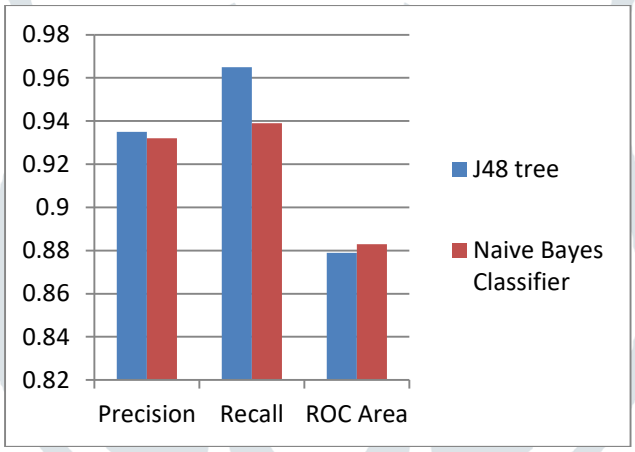


Graph 2: Kappa statistic, mean absolute error, root mean squared error resulting from application of both algorithms

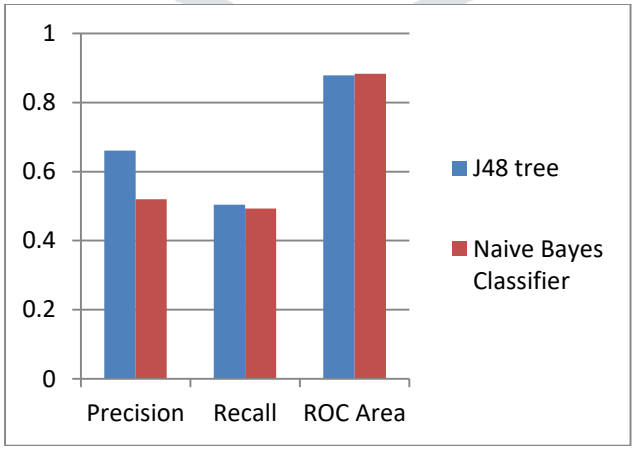


As shown in Graph 2, Kappa Statistic resulted from J48 tree algorithm is 0.5232 whereas from naive bayes classifier it is 0.4414. Mean absolute error resulted from J48 tree algorithm is 0.1206 and from naive bayes classifier, it is 0.1228. Root mean squared value resulted from J48 algorithm is 0.2615 and from naive bayes algorithm, it is 0.3132.

Graph 3: Precision, Recall, ROC Area for class no



Graph 4: Precision, Recall, ROC Area for class yes



As shown in Graph 3 (on next page), for class no, Precision value resulted from J48 tree algorithm is 0.935 whereas from naive bayes classifier it is 0.932. Recall value resulted from J48 tree algorithm is 0.965 and from naive bayes classifier, it is 0.939. ROC area value resulted from J48 algorithm is 0.879 and from naive bayes algorithm, it is 0.883.
As shown in Graph 4 (on next page), for class yes, Precision value resulted from J48 tree algorithm is 0.661 whereas from naive bayes classifier it is 0.520. Recall value resulted from J48 tree algorithm is 0.504 and from naive bayes classifier, it is 0.493. ROC area value resulted from J48 algorithm is 0.879 and from naive bayes algorithm, it is 0.883.

From previous 4 graphs, it is observed that J48 decision tree algorithm is performing better than naive bayes algorithm according to the available data.

4. Conclusion

Direct marketing is becoming important day by day due to its increasing applicability in various industries specially banking, fundraising, retailing & insurance. A main and serious challenge to it is about inverse correlation between the probability to buy and the money to be spent on buying any product/ service.

In this paper, two classification algorithms are explained and applied on data set required for direct marketing which shows how data mining helps in overcoming those challenge faced by direct marketing and the classification model is trained and further tested. This research will help various business sources and companies to find out those people who are expected to respond back while company makes contact to them with any of the means of direct marketing. It will also help them in choosing better data mining technique. This will lead to less wastage of money as well as time that company waste on people who are expected not to respond back.

References

[1] Esra Kahya Ozyirmidokuza, Kumru Uyarb, Mustafa Hakan Ozyirmidokuzc (2015). "A Data Mining Based Approach to a Firm's Marketing Channel" , *Proceedings of the 22nd International Economic Conference – IECS 2015 "Economic Prospects in the Context of Growing Global and Regional Interdependencies"*, pp. 77 – 84, Available online at www.sciencedirect.com.

[2] E.W.T. Ngai a, Li Xiu b, D.C.K. Chau (2009) "Application of data mining techniques in customer relationship management", *Expert Systems with Applications*, 36, pp. 2592–2602.

[3] Sankar Rajagopal (2011) "Customer data clustering using data mining technique", *International Journal of Database Management Systems (IJDMS)*, Vol.3, No.4.

[4] Charles Dennis, David Marsland and Tony Cockett (2001), "Data Mining for Shopping Centres–Customer Knowledge-Management Framework", *Journal of Knowledge Management*, 5 (4), 368-374, 1367-3270.

[5] M Suman, T Anuradha, K Manasa Veena (2011), "Direct Marketing with the Application of Data Mining", *Journal of Information Engineering and Applications*, ISSN 2224-5758 (print), ISSN 2224-896X (online), Vol 1, No.6, Available online at www.iiste.org

[6] Hasan Ziafat, Majid Shakeri (2014), "Using Data Mining Techniques in Customer Segmentation", *International. Journal of Engineering Research and Applications*, www.ijera.com ISSN: 2248-9622, Vol. 4, Issue 9(Version 3), pp. 70-79, Available Online at www.ijera.com

[7] Rabiya Shamsheer, Parul Agarwal (2017), "Exploring the Techniques of Classification Rule Mining", *International Journal of Advanced Research in Computer Science*, Vol. 8, No. 5, pp. 1092-1094, Available online at www.ijarcs.info.

[8] Ke Wang et al. (2005), "Mining Customer Value: From Association Rules to Direct Marketing", *Data Mining and Knowledge Discovery*, Vol. 11, pp. 57–79.

[9] Su-Yeon Kim, Tae-Soo Jung, Eui-Ho Suh (2006), "Customer segmentation and strategy development based on customer lifetime value: A case study", *Expert Systems with Applications*, Vol. 31, pp. 101–107, Available online at www.elsevier.com/locate/eswa

[10] (2010)[Online]Available:https://www.ibm.com/developerworks/library/os-weka1/

[11] (2015) [Online]Available: http://www.wideskills.com/data-mining-tutorial/data-mining-processes

[12] (2017)[Online]Available: http://www.zentut.com/data-mining/data-mining-techniques

[13] (2008)[Online]Available:https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/classify.htm#DMCON004

[14] [Online]Available:http://www.saedsayad.com/decision_tree.htm

[15] [Online]Available:http://www.saedsayad.com/naive_bayesian.htm

[16] Shu-Ching Wang et al. (2014), "Systematic Approach for Digital Marketing Strategy through Data Mining Technology", *Journal of Computers*, Vol. 25, No. 3, pp. 32-51.

[17] Chun Fu Lin, Yu Hsin Hung, and Ray I Chang (2013), "Mining Customer Behavior Knowledge to Develop Analytical Expert System for Beverage Marketing", *International Journal of Computer Trends and Technology (IJCTT)*, Vol. 4, Issue 4, pp. 579-584, Available online at http://www.ijcttjournal.org

[18] Chris Fleizach, Satoru Fukushima, "A naive Bayes classifier on 1998 KDD Cup", Available online at http://www.sysnet.ucsd.edu

[19] A.Bharathi, E.Deepankumar (2014), "Survey on Classification Techniques in Data Mining", *International Journal on Recent and Innovation Trends in Computing and Communication* , Vol. 2, Issue 7, pp. 1983-1986, Available online at http://www.ijritcc.org
[20] Megha Gupta, Naveen Aggarwal (2010), "Classification Techniques Analysis", *Proceedings of the NCCI 2010 -National Conference on Computational Instrumentation CSIO , Chandigarh, INDIA*, pp. 128-131.
[21] Ritu Sharma, Mr. Shiv Kumar, Mr. Rohit Maheshwari (2015), "Comparative Analysis of Classification Techniques in Data Mining Using Different Datasets", *International Journal of Computer Science and Mobile Computing (IJCSMC)*, Vol. 4, Issue. 12, pg.125-134, Available online at www.ijcsmc.com
[22] Radhakrishnan B, Shineraj G, Anver Muhammed K.M (2013), "Application of Data Mining In Marketing", (*IJCSN) International Journal of Computer Science and Network*, Vol. 2, Issue 5, pp. 41-46, Available online at www.ijcsn.org
[23] Bharat S. Makhija, Dr. Anjali B. Raut (2016), "Classification Algorithms of Data Mining", *International Journal of Science, Engineering and Technology Research (IJSETR)*, Volume 5, Issue 5, pp. 1639-1642.
[24] Mansi Gera, Shivani Goel (2015), "Data Mining - Techniques, Methods and Algorithms: A Review on Tools and their Validity", *International Journal of Computer Applications*, Vol. 113, No. 18, pp. 22-29.
[25] (2018)[Online]Available:https://www.tutorialspoint.com/data_mining/dm_rbc.htm
[26] Chris Rygielski, Jyun-Cheng Wang, David C. Yen (2002), Technology in Society , 24, pp. 483-502, Available online at www.elsevier.com/locate/techsoc
[27] Charles X. Ling and Chenghui Li, "Data Mining for Direct Marketing: Problems and Solutions", *KDD-98*, pp. 73-79.
[28] Xindong Wu et al. (2008), "Top 10 algorithms in data mining, Knowledge and Information Systems", *Knowledge and Information Systems*, Vol. 14, Issue 1, pp 1-37.
[29] Dr. Neeraj Bhargava et al. (2013), "Decision Tree Analysis on J48 Algorithm for Data Mining", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 3, Issue 6, pp. 1114-1119, Available online at www.ijarcsse.com
[30] [Online]Available:https://en.wikipedia.org/wiki/C4.5_algorithm
[31] G.K.Gupta, *Introduction to Data Mining with Case Studies*, Asoke K. Ghosh, (2012), pp.115-116.
[32] (2005)[Online]Available:https://www2.1010data.com/documentationcenter/prod/Tutorials/MachineLearningExamples/BankMarketingDataSet.html
[33] [Online]Available:http://archive.ics.uci.edu/ml/datasets/Bank+Marketing
[34] Aarushi, Rainu Nandal (2018), "Application of Data Mining in Customer Relationship Management: A Review", *International Journal of Emerging Technologies and Innovative Research*, Vol. 5, Issue 1, pp. 743-745.
[35] S. Moro, P. Cortez and P. Rita (2014), "A Data-Driven Approach to Predict the Success of Bank Telemarketing", *Decision Support Systems, Elsevier*, 62, 22-31.
[36] [Online]Available:https://en.wikipedia.org/wiki/Weka_(machine_learning)