

A Decision Tree based Algorithm for Forecasting the Proposed Service Acceptance by Store Customers

Ali Mirzapour¹

¹(Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran)

Corresponding Author: Ali Mirzapour

Abstract: Organizations are under intense pressure today to respond quickly to changing circumstances and innovation since business environment is increasingly becoming complex and changing. Furthermore, data analysis and data mining will create a competitive advantage. Hence, customer relationship management has been very much considered in recent years. Accordingly, in order to understand the effective criteria for accepting the proposed service to customers through the database, the present study aimed to classify the information about the store customers and extract the appropriate pattern using the decision tree algorithm so that it will be possible to properly predict the accepting the proposed service and ultimately, meet the advertising and marketing requirements for the society. According to the results obtained from implementing the proposed system on the collected store data, evaluated during 2016-2018 through the Clementine 12.0 software, the system accuracy level is at a desirable level.

Date of Submission: 22-02-2019

Date of acceptance: 08-03-2019

I. Introduction

Customer relationship management seeks to attract, recognize, deliver services and anticipate customer needs. Continue to survive in such an environment require organizations to be agile and make very quickly their strategic, tactical and operational decisions. Such decisions require a large amount of data, information and knowledge that must be processed in the form of required decisions and be quickly processed, on a timely and continuous basis. As a business process, customer relationship management collects and integrates customer information, manages the customer relationships, and informs about the purchases of products or services by the organization. This process helps organizations track customer interactions with the organization and allow employees to track service records and review and evaluate past customer sales and significant orders or unsolved issues. Customer Relationship Management gathers all the customer information that small business owners register in their office, including the names of customers, items they bought, related issues, and so on in a database. Customer relationship management helps coordination of sales, marketing, and customer service teams units to continuously work to meet the customer's need. The recent research and papers on this topic, has mostly ignored the impact of factors such as sales tax and profit. It has been tried in the present study to examine the impact of the above factors, as well as other influential factors such as income category, gender, marital status, number of children, education and residence to conduct the necessary forecast. Forecasting the service acceptance will be done through the decision tree algorithm as stated above.

II. Literature Review

Data-mining:

As an attempt to form a human-computer collaboration, data mining provides the best results by balancing the knowledge of specialists in describing issues and goals with computer search capabilities [1]. In [2], an intelligent method is proposed from a human-computer collaboration for an industrial application. In this method, fuzzy systems apply the knowledge of specialists to control two valves in water distribution systems so that the customer satisfaction is met. It is predicted according to the online magazine that data mining was expected to revolutionize the next decade. MIT has also selected data mining as one of the top ten emerged technologies that will revolutionize the world [3]. Data-mining algorithms have been used in many industrial applications such as [4] where the authors use a support vector machine based method as a new technology for intelligent fault location in a power grid. In [5] the authors effectively used data mining for energy management of a residential community to reduce the power fluctuations in presence of uncertainties.

Definition of Data-mining:

Data mining is considered as a part of the knowledge discovery process in databases which refers to the process aimed to find valid, useful, novel, and intelligible patterns in data [6].

Data mining is defined as the analysis of a large amount of data by automatic and semi-automatic tools to extract significant patterns and trends [7].

III. Problem Statement

The present study aimed to classify store customers according to factors such as location, education, marriage status, number of children, income category, gender, purchasing power, and corporate income tax and sold items profits to categories, and determine the role of the most important factors on the purchase rate and acceptance of the proposed service through advertising and marketing.

The utilized dataset:

The data used in this project is provided by a store database. This data is related to 30 customer that is in the form of an Excel file with 1124 records. The proposed model to predict sales growth in the store, based on store database data, is described below:

The various stages of the data mining process, including data collection, preparation and pre-processing of data, are performed on store data in the proposed model, and the decision tree algorithm is used in this regard.

A series of pre-processing operations are performed on the data for a better performance of the data mining algorithm. Also, the numeric attributes have turned into the equivalent group attributes after aggregating the data into a file. For example, all customers are divided into two single and married groups. The customer education is also divided into four groups of under diploma, diploma, associate degree or a bachelor's degree and senior management degree. Table 1 shows the characteristics used and the desired coding.

Table no 1: The status of the fields and its encoding

Field (attribute)	Type	Coding	scale
Gender	Male and female	Male = 0, Female = 1	Qualitative - nominal
Customer's education	Master Degree, Bachelor and associate degree, Diploma, lower diploma	Master and above = 1, Associate degree and Bachelor = 2, Diploma = 3 lower diploma = 4	Qualitative - nominal
Marriage status	Single and Married	Single = 0 Married = 1	Qualitative - nominal
No. of Children	numeric	As numbers	numeric
Dep_ID	Food, clothing, home appliances, cultural products, sporting goods	Food = 1, Clothing = 2, Home appliances = 3 Cultural products = 4, Sports products = 5	Qualitative - nominal
sale	numeric	As numbers	numeric
location	North of Tehran, West Tehran, East of Tehran and south of Tehran	North of Tehran = 1 West of Tehran = 2, East of Tehran = 3, South of Tehran = 4	Qualitative - nominal
Customer's income level	Above 1,700 million Between 1,200 and 1,700 million, Between 500 and 1,200 million, less than 500 thousand Tomans	;Above 1,700 million = 1 Between 1,200 and 1,700 ,million = 2 Between 500 and 1,200 ,million = 3 Less than 500 thousand Tomans = 4	Qualitative - nominal
Sales profits	numeric	As numbers	numeric
Sales tax	numeric	As numbers	numeric
Service acceptance	Accept/reject	Accepted = 1 Reject = 0	Qualitative - nominal

The customer revenues range is presented as options in their membership form to choose among one of these options since customers usually do not disclose their exact amount. It should be considered that our goal in this analysis, is merely to know the revenue category of the customer, in order to make our advertising and marketing according to their income level. So we considered them as qualitatively-nominal scales.

IV. Introduction to modeling

As a set of rules, formulas, or equations, a model can be used to predict a result based on a set of variables or input fields. For example, a model may be used in order to identify the customers who are likely to welcome cultural products by gathering data on age, income, and education.

The most important and time-consuming step in data mining projects is the data preparation phase. Since the data are as the input in these projects, the more accurate they are, the more efficient the output. The data preparation operation was performed by the Excel program and the Clementine 12.0 software provided by the SPSS Company.

Decision tree algorithm:

A tree node (C & R tree) classification and regression is a tree-based classification and forecasting method that uses recursive divisions to separate records into sections with values similar to fields in the output. This case implements the C & R tree for a marketing study whose target field is the tendency to buy a service. . One of the fundamental feature of the tree is the elimination of the attributes based on the importance [8]. Pattern recognition and optimization algorithms can be addressed over various computational, and combinatorial models such as Data Envelopment Analysis (DEA) [9-10]and fuzzy expert systems [11]. The forecasting fields include education, income category, gender, marriage status, number of children, place of residence and customer desirable department, purchase and tax rates on goods, and the companies' profits. Tree based classification is also widely used for decision making in engineering. In reference [12], this approach is successfully applied in a prioritization algorithm to define the novel concept of energy price tag, which the results are pivotal for research in field of energy storage systems. In regards to data mining classification algorithms such as decision tree and association rules algorithm such as Apriori other respected works concentrated on various aspect Data mining is a foremost expertise with exceptional prospective to aid thecompanies.

The number of samples tested is 1,124, during the tree division, of which 328 customers, which is equivalent to 29.181% do not accept the proposed service, and 796 samples equivalent to 70.819% accept them. This sample can be analyzed in more detail by dividing the tree, and factors affecting customer service acceptance can be evaluated. Fig.1 displays output of decision tree.

Analyses review:

The decision tree is formed in 6 divisions, that each of them were briefly analyzed and evaluated. In the first division, the marital status impact has been evaluated on service acceptance. The node 1 indicates that 29.7% of the total customer service is accepted by single customers. In the next division in node 6, the impact of customer's residence place is evaluated on the client's acceptance that according to which, 61.9% of the married people living in the north of the city welcomed the service. In the third division, the impact of the number of children has been evaluated, and one-child families had the lowest rate with 11.5%. The customer's education was evaluated in the next division, that the clients with diploma education and higher accepted the service that constitute the 26.7% of the sample. Finally, customer purchase rate were analyzed, that the customers with less than \$ 26300 purchase more welcomed the service, and 71.4 of them accepted the service.

The profit table has been examined in the next step for this tree. The percentage of the profit index tells you which proportion the target group in each node differs others.

The indifferent responses of 0 and excited responses equal to 1 can be selected as the target group to display the profit. We set the Target Category to 1 for our analysis, that is, our target analysis is on admission. The profit table and output of decision tree can be displayed in Fig. 1 and Fig.2 respectively.

Nodes	Node: n	Node (%)	Gain: n	Gain (%)	Response (%)	Index (%)
1	334.00	29.72	258.00	32.41	77.25	109.08
17	301.00	26.78	229.00	28.77	76.08	107.43
25	28.00	2.49	20.00	2.51	71.43	100.86
11	130.00	11.57	86.00	10.80	66.15	93.41
6	321.00	28.56	199.00	25.00	61.99	87.54
26	10.00	0.89	4.00	0.50	40.00	56.48

Fig.1. Profit table

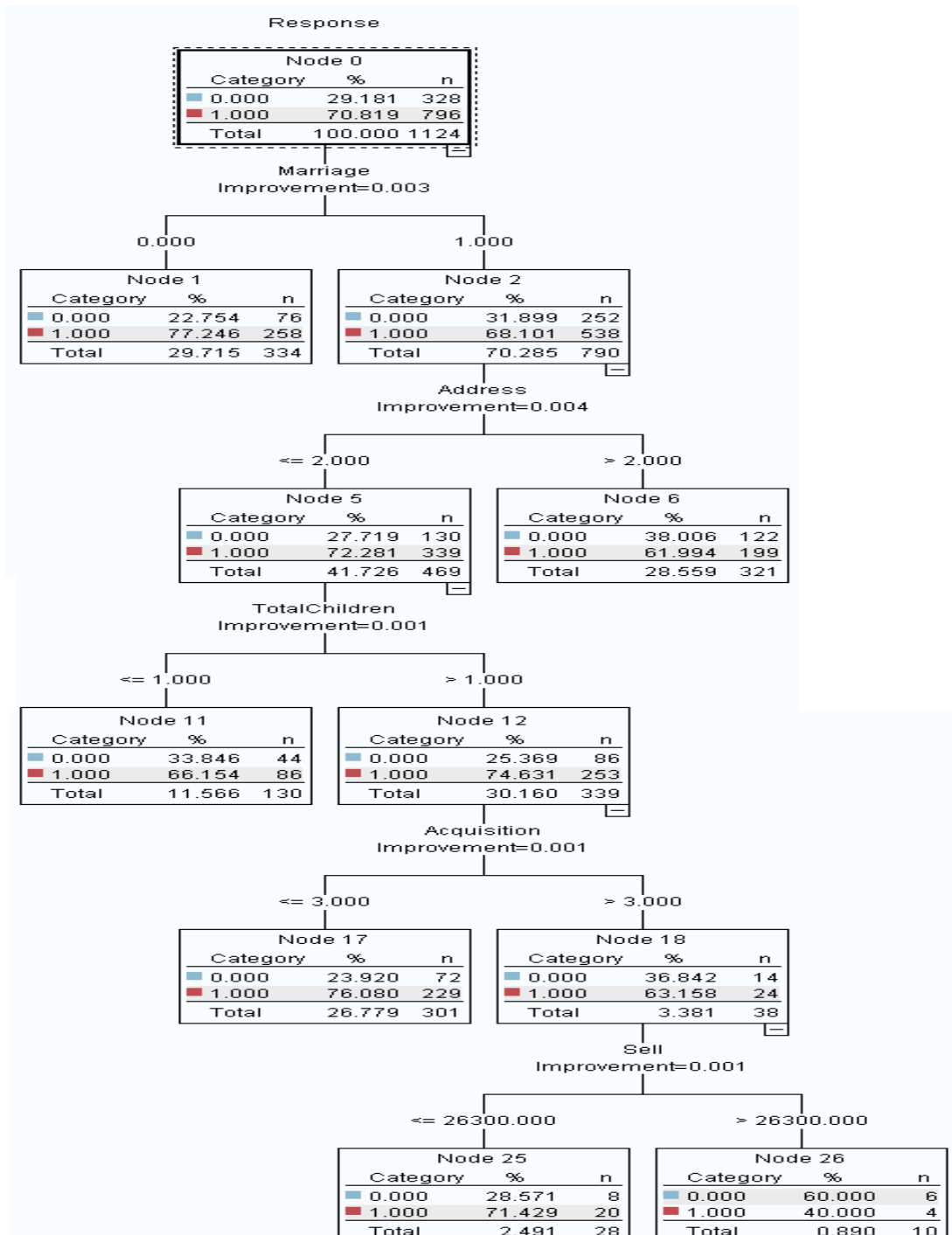


Fig.2. Output of decision tree

The nodes 1, 17, 25, 11, 6 and 26 indicate the highest index in acceptance. Nodes with values greater than 100 show a greater chance of acceptance. The index values in this sample indicate that Node 1, 17, and 25 have a greater chance of accepting than other nodes. Node 1 has a maximum percentage rate of around 109.08. That is, the customers of node 1 with a probability of approximately 10.9, node 17 with probability 10 or 7, and node 25 with a probability of 10 times, accept the services. It can be concluded, looking at the status of the nodes, that the single customers are far better because they accept 10 or 9 times the service. On the other hand, married clients with more than one child who have a postgraduate education, are likely to accept services 10 or 7 times in all regions except the northern regions of the city. Growth charts are shown in Fig.3.

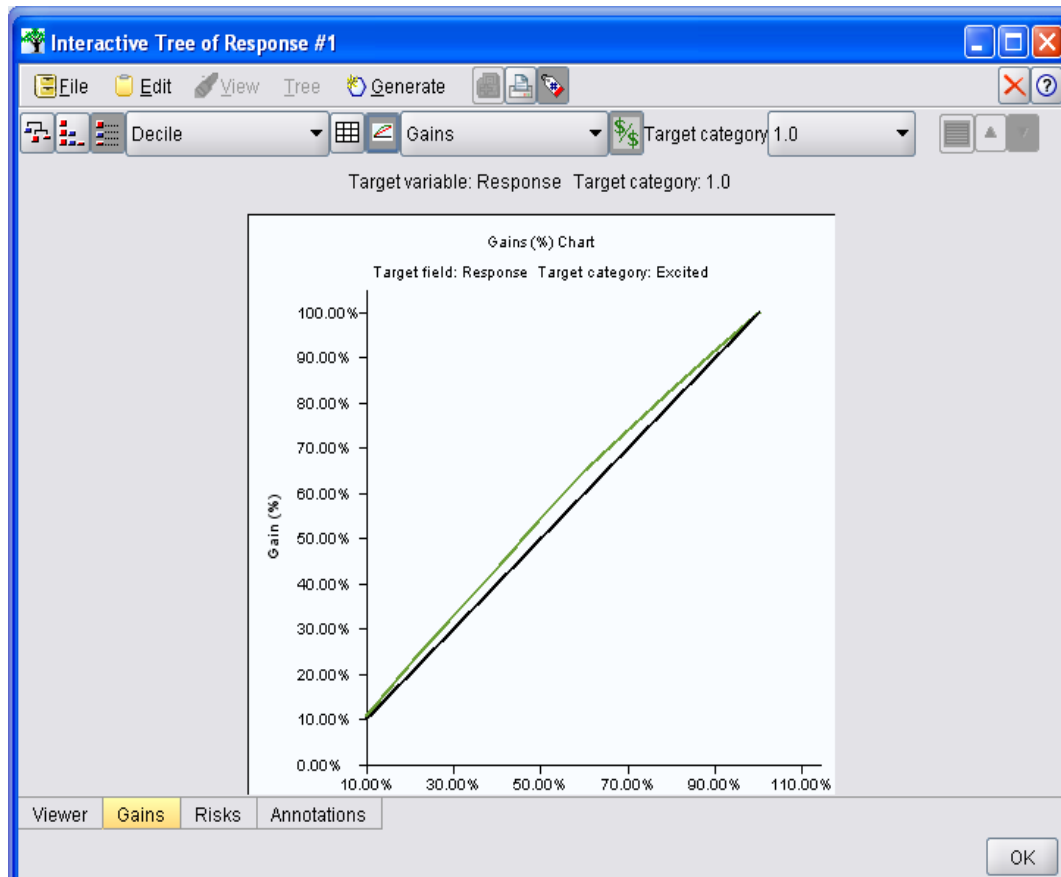


Fig.3. Growth charts

V. Conclusion

The case studied in this paper has showed how to use the C & R node tree and decision tree in a marketing study. Accordingly, it is possible to determine the customers who want to buy the suggested service. Using the predictors such as gender, education, income category, number of children, location, and intended departments, it is possible to predict and categorize responses to obtain the highest rates of response. Tree models are easy to understand. The system creates the tree and each branch is recursively divided to one or more criteria. Then, the best predictor is automatically selected based on the modeling method used in each division. As shown in this case, the tree can be used to generate a model or select a node and running it for the future data set. The necessary decisions on the customer's purposeful choices and the necessary advertising for them can be adopted utilizing this model

References

- [1]. M. Kantardzic, *Data Mining: Concepts, Models, Methods, Algorithms*, IEEE press, 2003.
- [2]. O. A. Gashteroodkhani, M. Majidi, M. Etezadi-Amoli, "A Fuzzy-based Control Scheme for Recapturing Waste Energy in Water Pressure Reducing Valves" *IEEE Power and Energy Society General Meeting (PESGM)*, pp. 1-5, Portland, OR, Aug 2018.
- [3]. O.R. Zaiane, *Web Usage Mining for a Better Web-Based Learning Environment*, 2002.
- [4]. O. A. Gashteroodkhani, M. Majidi, M. Etezadi-Amoli, A. F. Nematollahi, B. Vahidi, "A hybrid SVM-TT transform-based method for fault location in hybrid transmission lines with underground cables" *Electric Power Systems Research*, vol. 170, pp. 205-214, 2019.
- [5]. S. Aznavi, P. Fajri and M. Rasheduzzaman, "Hierarchical Energy Management Strategy for a Community of Multi Smart Homes," *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, Washington, DC, 2018, pp. 176-181.
- [6]. U.M Fayyad, G. Piatetsky-shapiro, P. Smyth, and R. Uthurusamy, *Advances in Discovery and Data mining*, Menlo Park, CA: AAAI/MIT press, 1998.
- [7]. C. Romeo, S Ventura (2007). *Educational data mining: A survey from 1995 to 2005*. *Expert Systems with Applications*, Vol. 33,1,PP. 135-146.
- [8]. M. Mirmozaffari, "Eco-Efficiency Evaluation in Two-Stage Network Structure: Case Study: Cement Companies". *Iranian Journal of Optimization (IJO)*. Dec. 16, 2018.
- [9]. M. Mirmozaffari and A. Alinezhad, "Ranking of heart hospitals using cross-efficiency and two-stage DEA," *2017 7th International Conference on Computer and Knowledge Engineering (ICCCKE)*, Mashhad, 2017, pp. 217-222.

- [10]. M. Mirmozaffari, M. Zandieh, and S.M Hejazi. An Output Oriented Window Analysis Using Two-stage DEA in Heart Hospitals. International Conference on Innovation in Science, Engineering Computer and Technology (ISECT). Dubai (UAE) Oct. 17-19, 2017.
- [11]. M. Mirmozaffari, "Developing an Expert System for Diagnosing Liver Diseases", *EJERS*, vol. 4, no. 3, pp. 1-5, Mar. 2019.
- [12]. S. Aznavi, P. Fajri and A. Asrari, "Smart Home Energy Management Considering Real-Time Energy Pricing of Plug-in Electric Vehicles," in 2018 IEEE Energy Conversion Congress and Exposition (ECCE), Portland, OR, USA, 2018, pp. 67-72.