

Applying decision tree models to SMEs: A statistics-based model for customer relationship management

Ayad Hendalianpour^{a*}, Jafar Razmi^a and Arefe Rameshi Sarvestani^b

^a*School of Industrial Engineering, College of Eng, Tehran University, Tehran, Iran*

^b*Department of Mathematics, Shahrood University of Technology, Shahrood, Iran*

CHRONICLE

Article history:

Received February 5, 2016
Received in revised format April 15, 2016
Accepted May 7, 2016
Available online
May 9, 2016

Keywords:

Customer Relationship Management (CRM)
SMEs
Decision tree
C&RT
C4.5
ID3

ABSTRACT

Customer Relationship Management (CRM) has been an important part of enterprise decision-making and management. In this regard, Decision Tree (DT) models are the most common tools for investigating CRM and providing an appropriate support for the implementation of CRM systems. Yet, this method does not yield any estimate of the degree of separation of different subgroups involved in analysis. In this research, we compute three decision-making models in SMEs, analyzing different decision tree methods (C&RT, C4.5 and ID3). The methods are then used to compute ME and VoE for the models and they were then used to calculate the Mean Errors (ME) and Variance of Errors (VoE) estimates to investigate the predictive power of these methods. These decision tree methods were used to analyze small- and medium-sized enterprises (SME's) datasets. The paper proposes a powerful technical support for better directing market trends and mining in CRM. According to the findings, C&RT shows a better degree of separation. As a result, we recommend using decision tree methods together with ME and VoE to determine CRM factors.

© 2016 Growing Science Ltd. All rights reserved.

1. Introduction

In recent decades, Customer Relationship Management (CRM) has specifically reflected the crucial role of the customer as a factor that helps a company's profitability and operation meet. The idea behind CRM is to learn the real requirements of the customer and to leverage this knowledge to increase the firm's profitability in the long term (Stringfellow et al., 2004). In the rapidly changing business culture, the economics of customer relationships is changing, and firms are facing the need to apply new solutions to address these changes (Ritter & Geersbro, 2011). In this context, the advent of information technology (IT) has transformed the way marketing is stated and how firms manage information about their customers (Stringfellow et al., 2004).

* Corresponding author.

E-mail address: hendalianpour@ut.ac.ir (A. Hendalianpour)

Customer Relationship Management (CRM) is an enterprise management strategy which concentrates on customers (Stringfellow et al., 2004). CRM applies modern IT to increase the capability of enterprises to maintain and recognize their customers through Business Process Reengineering (BPR), and to maximize the profitability. With the development of economy and competition among different organizations, there is an increase awareness of the fact that to acquire customers, only having good products and wide distribution networks are sufficient. In fact, enterprises could absorb faithful customers and be dominant in fierce market competitions only by considering customers' requirements, increasing the response speed, and providing customers with constant one-to-one services.

CRM can help enterprises better manage different activities concerning customers and render the business routines programmed and automated. In any small firm, to improve customer relationship process, a working group forms a committee to apply data mining algorithms to improve customer relationships, to increase process efficiency and to take necessary steps for organizational purposes.

Considering the increase in the complexity and accumulation of customer information, most organizations may succeed only by performing critical activities such as analyzing complex customer data, identifying customers' values, detecting the trend of customers' behaviors, appreciating the real value of customers, and analyzing customers from a lifecycle marketing perspective. However, all of these factors depend on data mining (DM), and the more the data are accumulated in a database, the more useful DM techniques are. On the other hand, DM is a process of gathering necessary information and knowledge hidden in a large reservoir of data which are incomplete, noisy, fuzzy, and random. DM is recognized by concepts, rules and patterns, etc. Compared with common database inquiry, the excellence of DM depends on how it could discover connections between businesses as well as the latent trends and patterns. For example, the problem of consumer attrition is sometimes out of control because of not having any early warnings, but DM can set up a model disclosing the rate of lost clients in advance. By using this model, enterprises could predict which clients they may lose in near future. Armed by this information, sellers, relying on dynamic marketing activities, could keep the clients who have the idea of leaving; even if clients are lost, the model assists sellers overcome their passivity that led to consumer attrition.

In general, through applying DM technology to analyze the data and determine the related knowledge and regulations, the whole CRM system may become a closed loop and work better. Customer data and information technology tools shape the foundation to build successful strategy (Ngai et al., 2009). With this trend, technologies such as data mining and data warehousing have changed CRM as a new area in which most organizations may gain competitive advantage. Data mining offers a sophisticated set of tools to excavate customer data in an analytical CRM framework. The data mining methods are indeed considered a leading CRM tool deeply influenced by IT.

There is no doubt, the upcoming advents in IT call for business organizations to shape the way CRM functions in their business culture. In the literature, many researchers have studied the potential advantages of data mining systems in managing customers (Ngai et al., 2009), but among them few have applied the principles of such strategic tools for small organizations. The interesting point here is that, although Small and Medium sized Enterprises (SMEs) have a strong focus on customers, suppliers, employees and other stakeholders, they are considered business enterprises which have not well-implemented the principals of e-business in their infrastructural marketing strategies. Data mining is one of those technology-driven tools of e-business that has gained less attention in smaller firms, compared to larger ones.

Considering prior research, one can argue that most of the studies did not propose a well-formulated technique to be used in SEMs. The present study applies three models (decision trees) and analyzes them to find the best model. Furthermore, the study tries to select the best method for SEMs in Fars province, Iran, by relying on a case study. The major contributions of the study is answering two critical

questions: (a) how can use data mining algorithm for CRM in SMEs? and (b) which of the widely used decision tree methods is better to be used in SMEs?

The organization of this paper is as follows: section 2 is literature review. Section 3 describes the CRM in SMEs. Section 4 briefly introduces the decision trees methods (C&RT; C4.5; ID3). In Section 5, the statistics methods for validation is presented and the best method from C&RT; C4.5; ID3 is chosen. In section 6, we present how to use decision trees methods CRM with a case study in Small and Medium sized Enterprises in Fars Province. At the end, a concluding remark is given in section 7.

2. Literature review

Decision trees are produced by algorithms that identify various ways of splitting a data set into branch-like segments (Moon et al., 2012). These segments construct an inverted decision tree that originates with a root node at the top of the tree. The object of analysis is reflected in this root node as a simple, one-dimensional display in the decision tree interface (Nefeslioglu et al., 2010). The name of the field of data that is the object of analysis is usually displayed, along with the spread or distribution of the values that are contained in that field. A sample decision tree is illustrated in Fig 1, which shows that the decision tree can reflect both a continuous and categorical object of analysis. The display of this node reflects all the data set records, fields, and field values that are found in the object of analysis (Wu et al., 2009; Ture et al., 2009). The discovery of the decision rule to form the branches or segments underneath the root node is based on a method that extracts the relationship between the object of analysis (that serves as the target field in the data) and one or more fields that serve as input fields to create the branches or segments. The values in the input field are used to estimate the likely value in the target field. The target field is also called an outcome, response, or dependent field or variable (Nefeslioglu et al., 2010; Sindhu et al., 2010).

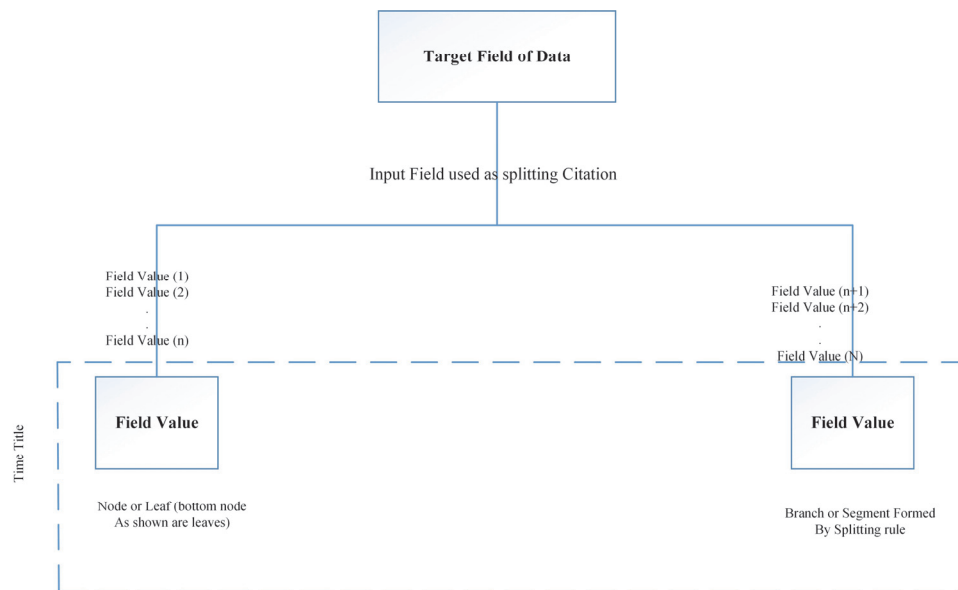


Fig. 1. Illustration of the Decision Tree

Many decision-tree algorithms have been developed. The most famous algorithms are C&RT, C4.5, ID3 that are simple decision tree learning algorithms developed by Ross Quinlan (T'sou et al., 2000, True et al., 2009), whose choice of split attribute is based on information entropy. The basic idea of ID3 algorithm is to build the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node. C4.5 is an extension of ID3 developed by Prather et al. in 1997 (Parther et al., 1997, Ture et al., 2009; Yıldız, 2011). It improves computing efficiency, deals with continuous values, handles attributes with missing values, avoids over-fitting, and performs other

functions. To deal with continuous data, CART (classification and regression tree) algorithms have been proposed. CART is a data-exploration and prediction algorithm similar to C4.5, which is a tree construction algorithm developed by Martinez-Munoz and Suárez (2004). Breiman et al. (1984) summarized the classification and regression trees. Instead of information entropy, it introduces measures of node impurity. It relies on a variety of different problems, such as the detection of chlorine from the data contained in a mass spectrum, as proposed by Berson and Smith in 1997. CHAID (Chi-square automatic interaction detector) is similar to CART, but their difference lies in that it chooses a split node. It depends on a Chi-square test used in contingency tables to determine which categorical predictor is farthest from independence with the prediction values that proposed with Bittencourt and Clarke (2003). It also has an extended version, Exhausted-CHAID.

3. CRM in SMEs

Numerous researches have been carried out in the context of data mining in larger firms (Treacy & Wiersema, 1997), but the literature is underdeveloped in case of SMEs. The reason for this underdeveloped research trend is well justified: in many cases, smaller firms cannot afford the initial capital for the installation of a systematic data mining framework. In fact, data mining is used when there is a huge expanse of data with important implied information. Managing the information can help companies be successful in market competition.

Nowadays, in the competitive e-business environment, business organizations are switching from product-oriented business strategies to customer-oriented ones. Globalization, increasing competition, and advances in information and communication technology, have forced firms to concentrate on managing customer relationships to efficiently maximize revenues (Azad & Ahmadi, 2015). As a strategy to optimize the lifetime-value of the customers, CRM can help firms succeed in the world of e-business (Rygielski et al., 2002; Hsu, 2009), define CRM as a cross-functional procedure to attain a continuous exchange of ideas with customers, across all their contacts and access points, with a personalized treatment of the most valuable customers to increase customer retention and the effectiveness of marketing networks.

Both large and multinational enterprises and SMEs are rigorously seeking to deploy CRM to gain competitive advantage to build their long-term profitability. Particularly, customer retention is important to SMEs because of their limited resources (Özgener & İraz, 2006; Reddick, 2011). Skaates and Seppänen (2002) point out the major contribution of CRM to the competence development of small firms. Furthermore, SMEs are embracing CRM as a major element of business strategy, while technological applications allow for a precise segmentation, provide profiling and targeting of customers, and bring about a customer-centric culture due to competitive pressures (Day et al., 2002, Sulaiman, 2011; Skaates & Seppänen, 2002; Gurău et al., 2003).

4. Methods

4.1. Decision Tree

During the late 1970s and early 1980s, Quinlan, a researcher in machine learning, developed a decision tree algorithm known as ID3 (Iterative Dichotomiser). This work was expanded on earlier findings about concept learning systems described by others and Quinlan later presented C4.5 (a successor of ID3), which became a benchmark to which newer supervised learning algorithms are often compared. In 1984, a group of statisticians (Chang et al., 2012) published the book *Classification and Regression Trees* (CART), which described the generation of binary decision trees. ID3 and CART were invented independently of one another at around the same time, although they followed a similar approach to learning decision trees from training tuples. These two cornerstone algorithms spawned a flurry of work on decision tree induction (Gurău et al., 2003).

ID3, C4.5, and CART adopt a greedy (i.e., non-backtracking) approach in which decision trees are constructed in a top-down recursive divide-and-conquer manner. Most algorithms for decision tree induction also follow such a top-down approach, which starts with a training set of tuples and their associated class labels. The training set is recursively partitioned into smaller subsets as the tree is being built. In the following sections, we briefly introduce the C&RT, C4.5, ID3 methods.

3.1.1. Classification and regression tree

Classification and regression tree (C&RT) is a recursive partitioning technique implemented for both regression and classification. C&RT partitions the data into two subsets so that the records within each subset are more homogeneous than those in the previous subset. It is a recursive process through which each of those two subsets is then split again, and the process is repeated until the homogeneity criterion is reached or until some other stopping criterion is met (as in case of all of the tree-growing methods). The objective is to generate subsets of the data which are as homogeneous as possible with respect to the target variable (Breiman et al., 1984). In this study, measure of Gini impurity was used for categorical target variables.

Gini Impurity Measure:

The Gini index at node t , $g(t)$, is defined as:

$$g(t) = \sum_{j \neq i} p(j|t)p(i|t), \quad (1)$$

where i and j are categories of the target variable. The equation for the Gini index can also be written as:

$$g(t) = 1 - \sum_j p^2(j|t). \quad (2)$$

Thus, when the cases in a node are evenly dispersed across the categories, the Gini index takes its maximum value of $1 - (1/k)$, where k is the number of categories for the target variable. When all cases in the node are found to be in the same category, the Gini index equals 0. If costs of misclassification are specified, the Gini index is computed via:

$$g(t) = \sum_{j \neq i} c(i|j)p(j|t)p(i|t), \quad (3)$$

where $C(i-j)$ is the probability of misclassifying a category j case as category i . The Gini criterion function for split s at node t is defined as:

$$\phi(s, t) = g(t) - p_L g(t_L) - p_R g(t_R), \quad (4)$$

where p_L , the proportion of cases in t , is sent to the left child node, and p_R is the proportion sent to the right child node. Split s is chosen to maximize the value of $\phi(s, t)$. This value is reported along with the progression in the tree (Breiman et al., 1984).

4.1.2. Commercial version 4.5

C4.5 is a kind of supervised learning classification algorithm which is used to build decision trees from the data (Treacy & Wiersema, 1997). Most empirical learning systems are given a set of pre-classified cases, each explained by a vector of attribute values and constructed from a mapping of attribute values to classes. C4.5 is such a system that learns decision tree classifiers. It applies a divide-and-conquer approach to growing decision trees (Azad & Ahmadi, 2015). The main difference between C4.5 and other similar decision tree building algorithms is in the test selection and evaluation process.

Let attributes be denoted $A = \{a_1, a_2, \dots, a_m\}$, cases be denoted $D = \{d_1, d_2, \dots, d_n\}$, and classes be denoted $C = \{c_1, c_2, \dots, c_k\}$. For a set of cases D , a test T_i is a split of D derived from attribute at . It splits D into mutually exclusive subsets D_1, D_2, \dots, D_p . These subsets of cases are single-class collections of cases. If a test T is chosen, the decision tree for D consists of a node identifying the test T , and one branch for each possible subset D_i . For each subset D_i , a new test is then chosen for further splits. If D_i satisfies a stopping criterion, the tree for D_i is a leaf associated with the most frequent class in D_i . One reason for stopping is that cases in D_i belong to one class.

C4.5 decision tree algorithm uses a modified splitting criterion called *gain ratio*. It uses $\arg \max (gain(D, T))$ or $\arg \max (gain \text{ ratio } (D, T))$ to choose tests for split:

$$Info(D) = -\sum_{i=1}^k p(c_i, D) \log_2(p(c_i, D)), \quad (5)$$

$$Split(D, T) = -\sum_{i=1}^p \frac{|D_i|}{|D|} \log_2 \left(\frac{|D_i|}{|D|} \right), \quad (6)$$

$$Gain(D, T) = Info(D) - \sum_{i=1}^p \frac{|D_i|}{|D|} Info(D_i), \quad (7)$$

$$GainRatio(D, T) = Gain(D, T) / Split(D, T), \quad (8)$$

where, $p(c_i, D)$ denotes the proportion of cases in D that belong to the i -th class. C4.5 selects the test that maximizes gain ratio value Quinlan (1996). Once the initial decision tree is constructed, a pruning procedure is initiated to decrease the overall tree size and decrease the estimated error rate of the tree (Bradley, 1998).

4.1.3. ID3 algorithm

In decision-tree analysis based approaches, the most influential analysis is represented by the ID3 algorithm which was proposed by Quinlan (1996). It uses the decreasing speed of entropy as the criterion of selecting the attribute to be studied. This method uses well-known sample categories to obtain the ordered testing attributes, until all of the known samples have been classified. In the process of forming a decision tree, a method of informatics is used which can provide the largest information increase at any time, i.e. the largest entropy decrease (Zhuge, 2011).

For N samples, they belong to classes $C_i (i = 1, 2, \dots, C)$. There are N_i samples in class C_i ; every sample has K attributes, and every attribute has J_k values. The process of forming decision involves:

(1) Computing the initial entropies.

$$entropy(I) = \sum_{i=1}^c -(N_i/N) \log_2(N_i/N) = \sum_{i=1}^c -p_i \log_2 p_i \quad (9)$$

For the given training sample aggregate, all the sample classes are known. So the sample aggregate constitutes the initial entropies of the system.

(2) Selecting an attribute as the root node of the decision tree.

- For attribute $A_k (i = 1, 2, \dots, K)$ according to the J_k values (a_{kj}) of A_k , the initial samples are divided into the first grade sample aggregates. Although (a_{kj}) branch has (a_{kj}) samples, in general, the n_{kj} samples do not belong to the single class.

- For n_{kj} samples of the branch, assuming that the number of samples belonging to class C_i is $n_{kj}(i)$, we can use following formula to get the entropy of this branch.

$$\text{entropy}(I, A_k, j) = \sum_{i=1}^c \left[-n_{kj}(i)/n_{kj} \right] \log_2 \left[n_{kj}(i)/n_{kj} \right] \quad (10)$$

$$\text{entropy}(I, A_k) = \sum_{j=1}^{J_k} \sum_{I=1}^C (n_{kj}/N) \times \left[-n_{kj}(i)/n_{kj} \right] \log_2 \left[n_{kj}(i)/n_{kj} \right] \quad (11)$$

- Calculating the entropy decrease due to attribute A_k , i.e.

$$\Delta \text{entropy}(k) = \text{entropy}(I) - \text{entropy}(I, A_k) \quad (12)$$

- Selecting attribute A_{k_0} which causes the maximum entropy decrease via:

$$\Delta \text{entropy}(A_{k_0}) > \Delta \text{entropy}(A_k) \quad (K = 1, 2, \dots, K; k \neq k_0) \quad (13)$$

- So, the attribute A_{k_0} is the root node of decision tree.

(3) Attribute A_{k_0} has J_{k_0} values and divides the sample into J_{k_0} subsets. For every subset, we use the above method in turn and select an attribute A_k as the next lower grade node of the decision tree, in which the maximum entropy decrease is gained.

(4) Following step 3, we continuously construct the next lower grade nodes until all of subsets have only one class. Here, the entropies are zero and the decision tree has been accomplished.

5. Model Validation

Validating the model is a critical step in the process. It allows us to determine if we have successfully performed all the prior steps. If a model is not validated well, it can be due to data problems, poorly fitting variables, or problematic techniques. There are several methods for validating models. In this study, regarding recent studies of statistical methods for validation, we used two statistical methods: the first one is Mean Errors (ME) and the second one is Variance of Errors (VoE) (Montgomery & Runger, 2010). To make a statistical comparison of the two methods, mean errors (ME) and variance of errors (VoE) are adopted as defined by Eq. (14) and Eq. (15) respectively.

$$ME = \frac{1}{n} \sum_{h=1}^n \frac{|y_h^A - y_h^M|}{y_h^A} \cdot 100, \quad (14)$$

$$VoE = \frac{1}{n-1} \left(\sum_{h=1}^n \frac{|y_h^A - y_h^M|}{y_h^A} \cdot 100 - ME \right)^2, \quad (15)$$

where n is the total number of datasets y_h^A and y_h^M are the Actual values (from survey data) and the Model outputs (the proposed approach or statistical regression) for h th dataset respectively. It is noted that the smaller the ME and VoE values, the more accurate the model is. According to validation methods, a method is optimal if both mean errors and variance of errors have the minimum value. On the other hand, if in special cases, both mean errors and variance of errors are not in minimum value for the same method, we consider variance of errors because computation of the mean errors is in the variance of errors. For more details regarding to ME and VoE a reader may see Montgomery and Runger's contribution (Montgomery & Runger, 2010)

6. Case Study

Considering the significance of decision tree and its application in unfolding relationships for managing communication with costumers, in the following sections three widely used decision tree algorithms, which were explained in the precious sections, are investigated in a case study. This study uses the algorithms to discover the internal relationships of data and to select the best method commonly used in SEMs.

The above-mentioned decision tree methods are used in this section for measuring the relationship of management regarding companies as well as customer opinions. Fig. 2 illustrates the implementation of three tree methods in a SMEs through a step-by-step flowchart.

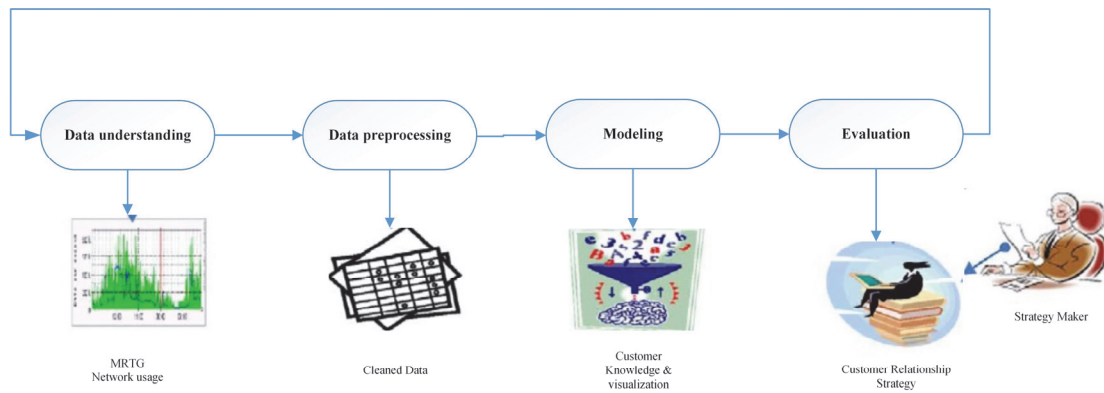


Fig. 2. Framework of tree methods

Fig. 2 shows the implementation process of the research through the data gathered from SEMs in Fars province, Iran. At the first stage of data understanding, considering experts' opinions, the main attributes of the research were extracted and, based on this primary structure, the data were collected. At the data preprocessing stage, statistical procedures were conducted on the collected data. Also, considering the possibility that some records might not involve any value in the effective attributes, decision-making, in such cases, is normally conducted by eliminating or substituting them; in this study we used the elimination technique. At the third stage called *modeling*, the collected data were computed using three decision tree models. Finally, given the outputs obtained from the three decision models, the defined criteria for selecting the best method in the SEMs were considered, and appropriate strategies were suggested. The effective specifications of the relationship between customers are shown in Table 1, in which Q11 is class attribute. In this data set, we try to predict who can find by this attribute the relationship of product quality and satisfaction, and then suggest a solution is to improve the problems with respect to output models.

Table 1

Set of all attributes

Attributes	Description	Type
Age	Between 15olds to above	Quantitative
Gender	Males or females	Binary
Education	Basic school, High school, University degree	Linguistic
Q1	Price compared to quality	Quantitative
Q2	Price compared to other companies	Quantitative
Q3	Price compared to expectations	Linguistic
Q4	Image of the store you deal with	Quantitative
Q5	Employees treated you politely and with respect when you complained	Linguistic
Q6	Services	Quantitative
Q7	Reliability of the products	Quantitative
Q8	Product variety	Binary
Q9	Product designing	Linguistic
Q10	Quality improvement	Linguistic
Q11	Overall satisfaction	Quantitative

Table 2
Statistical analysis of data set

Attribute	Statistical Methods		
	Mean	Standard deviation	Variance
Age	4.00	0.671	0.143
Gender	5.00	0.921	0.196
Education	3.82	0.853	0.182
Q1	3.36	0.727	0.155
Q2	4.09	0.811	0.173
Q3	3.23	1.020	0.218
Q4	4.09	0.684	0.146
Q5	3.59	0.734	0.157
Q6	3.18	0.853	0.182
Q7	3.86	0.889	0.190
Q8	3.82	0.795	0.169
Q9	3.36	1.002	0.214
Q10	3.41	1.008	0.215
Q11	4.55	0.671	0.143

This case study included 2000 individuals, as customers of SMEs in Fars Province, Iran (specific on food industrials). For this purpose, a questionnaire was designed which includes questions about product quality, price, services, and so on (a list of the questions are shown in Table 1). Next, through the above-mentioned decision tree methods, the customer relationship management of customers was evaluated. To do this, the decision tree (C&RT, C4.5 and ID3) methods were implemented in the C++ encoding language and after solving the proposed decision tree (C&RT, C4.5 and ID3) methods, the accuracy of the evaluation methods was calculated. Table 3 illustrates the values of statistics related to each of the attributes.

Table 3
Comparison of the error rates of models

Index	Models		
	C 4.5	CHAID	C&RT
Training-MAE	1425.057	1211.031	1415.059
Testing-MAE	1349.232	1190.404	1343.345
Training-Linear Correlation	0.246	0.515	0.335
Testing-Linear Correlation	0.165	0.462	0.187

In Table 3 we calculate the mean error for each of the four criteria for evaluating the algorithms employed in this study we have discussed. These criteria were used in both training and testing. These criteria indicate that the error rate for each of the algorithms used in the training set and the test is to what extent. Parameters for each algorithm, comparing the values specified in Table 6 is based on the method of least error. Regarding to output the C&RT is the maximum linear correlation.

According to validation method, from among C&RT, C4.5 and ID3, C&RT was found to be the best, because both ME and VoE showed the minimum value for this method. In this research, to justify the application of the appropriate method, we used a bar chart and scatter graphs. The result of each of these experiments is depicted in Fig. 3.

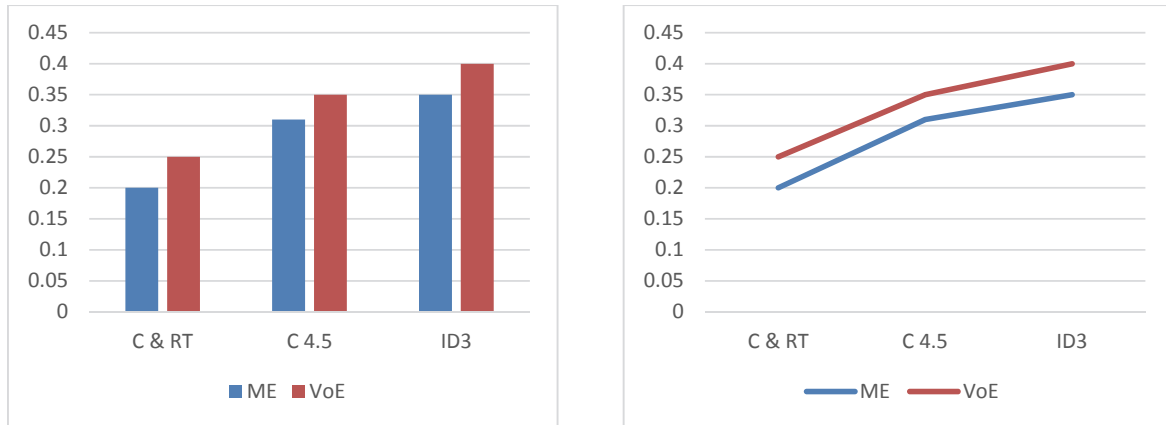


Fig. 3. Validation of decision tree methods with ME and VoE comparison

Table 4

Compared existing of decision tree methods

Index	Models		
	C&RT	C4.5	ID3
ME	0.2	0.31	0.35
VoE	0.25	0.35	0.4

Table 4 compares the outputs of the methods under study. Considering Table 4, one can observe that C&RT, compared to the two other methods, showed the minimum value of error in both of the statistical tests. Therefore, SMEs for customer communication management is more appropriate than the two other methods.

Finally, from among the approaches to decision tree algorithms, certainly C4.5, compared to the other two algorithms, had a better performance in development model of CRM in SMEs. Therefore, C4.5 algorithm is useful for predicting the behavior of customers. Otherwise, respectively CRAT and ID3 algorithms can more appropriately predict behavior of customers in the long-term.

7. Conclusion

CRM involves important measures and necessary steps for enterprises to keep their competitiveness in market economy. The development and application of CRM, as a result, are taken seriously by enterprises. Presently, there are a number of well-established products of CRM. DM technology is the rapidly developing key technology which, through development and perfection, can help accomplish the aims of CRM. DM can completely activate CRM and lay a good foundation For CRM. Further development in DM technology will produce more extensive applications for the future and market values of CRM. Recently, decision tree models for this purpose have incorporated novel ideas and interpretations into the solution of such problems. In this study, we found that C&RT performed better than C4.5 and ID3 techniques. As a result, we recommend the application of decision tree methods altogether, especially C&RT with statistical comparisons (Mean Errors and Variance of errors) to select the best known approach for Small and Medium sized Enterprises.

Then, the proposed models were meticulously compared by using a real data set in order to provide helpful information on the general tendency of data, assess the effect of specific variables on survival in data sets, and help Small and Medium sized Enterprises to select the best method for solving CRM

problems. There are limited data on the sufficiency of classification efforts by applying just one approach. Based on our observations, we suggest that data should be better explored and processed through high performance modeling procedures. Future studies can be conducted on applying hybrid robust models for this purpose.

References

- Azad, N., & Ahmadi, F. (2015). The customer relationship management process: its measurement and impact on performance. *Uncertain Supply Chain Management*, 3(1), 43-50.
- Bittencourt, H. R., & Clarke, R. T. (2003, July). Use of classification and regression trees (CART) to classify remotely-sensed digital images. In INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (Vol. 6, pp. VI-3751).
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees Belmont, CA: Wadsworth International Group.
- Bradley, P. S., Fayyad, U. M., & Reina, C. (1998, August). Scaling Clustering Algorithms to Large Databases. In *KDD* (pp. 9-15).
- Chang, S. H., Wang, K. Y., Chih, W. H., & Tsai, W. H. (2012). Building customer commitment in business-to-business markets. *Industrial Marketing Management*, 41(6), 940-950.
- Day, G. S., & Van den Bulte, C. (2002). Superiority in customer relationship management: Consequences for competitive advantage and performance. Marketing Science Institute.
- Gurău, C., Ranchhod, A., & Hackney, R. (2003). Customer-centric strategic planning: Integrating CRM in online business systems. *Information Technology and Management*, 4(2-3), 199-214.
- Hsu, C. H. (2009). Data mining to improve industrial standards and enhance production and marketing: An empirical study in apparel industry. *Expert Systems with Applications*, 36(3), 4185-4191.
- Moon, S. S., Kang, S. Y., Jitpitaklert, W., & Kim, S. B. (2012). Decision tree models for characterizing smoking patterns of older adults. *Expert Systems with Applications*, 39(1), 445-451.
- Nefeslioglu, H. A., Sezer, E., Gokceoglu, C., Bozkir, A. S., & Duman, T. Y. (2010). Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul, Turkey. *Mathematical Problems in Engineering*, 2010.
- Martinez-Munoz, G., & Suárez, A. (2004). Aggregation ordering in bagging. In *Proc. of the IASTED International Conference on Artificial Intelligence and Applications* (pp. 258-263).
- Montgomery, D. C., & Runger, G. C. (2010). *Applied statistics and probability for engineers*. John Wiley & Sons.
- Ngai, E. W. (2005). Customer relationship management research (1992-2002) an academic literature review and classification. *Marketing intelligence & planning*, 23(6), 582-605.
- Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, 36(2), 2592-2602.
- Özgener, Ş., & İraz, R. (2006). Customer relationship management in small-medium enterprises: The case of Turkish tourism industry. *Tourism Management*, 27(6), 1356-1363.
- Prather, J. C., Lobach, D. F., Goodwin, L. K., Hales, J. W., Hage, M. L., & Hammond, W. E. (1997). Medical data mining: knowledge discovery in a clinical data warehouse. In *Proceedings of the AMIA annual fall symposium*(p. 101). American Medical Informatics Association.
- Quinlan, J. R. (1996, August). Bagging, boosting, and C4. 5. In AAAI/IAAI, Vol. 1 (pp. 725-730).
- Reddick, C. G. (2011). Customer Relationship Management (CRM) technology and organizational change: Evidence for the bureaucratic and e-Government paradigms. *Government Information Quarterly*, 28(3), 346-353.
- Ritter, T., & Geersbro, J. (2011). Organizational relationship termination competence: A conceptualization and an empirical test. *Industrial Marketing Management*, 40(6), 988-993.
- Rygielski, C., Wang, J. C., & Yen, D. C. (2002). Data mining techniques for customer relationship management. *Technology in Society*, 24(4), 483-502.

- Sulaiman, S., Ariffin, M. K. A., Esmailian, G. R., Faghihi, K., & Baharudin, B. T. H. T. (2011). Customer knowledge management application in Malaysian mobile service providers. *Procedia Engineering*, 15, 3891-3895.
- Sindhu, S. S. S., Geetha, S., & Kannan, A. (2012). Decision tree based light weight intrusion detection using a wrapper approach. *Expert Systems with Applications*, 39(1), 129-141.
- Skaates, M. A., & Seppänen, V. (2002). Managing relationship-driven competence dynamics in professional service organizations. *European Management Journal*, 20(4), 430-437.
- Stringfellow, A., Nie, W., & Bowen, D. E. (2004). CRM: Profiting from understanding customer needs. *Business Horizons*, 47(5), 45-52.
- T'sou, B. K., Lai, T. B., Chan, S. W., Gao, W., & Zhan, X. (2000, October). Enhancement of a Chinese discourse marker tagger with C4. 5. In *Proceedings of the second workshop on Chinese language processing: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 12* (pp. 38-45). Association for Computational Linguistics.
- Treacy, M., & Wiersema, F. D. (1997). *The discipline of market leaders: Choose your customers, narrow your focus, dominate your market*. Basic Books.
- Ture, M., Tokatli, F., & Kurt, I. (2009). Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4. 5 and ID3) in determining recurrence-free survival of breast cancer patients. *Expert Systems with Applications*, 36(2), 2017-2026.
- Wu, D. (2009). Supplier selection: A hybrid model using DEA, decision tree and neural network. *Expert Systems with Applications*, 36(5), 9105-9112.
- Yıldız, O. T. (2011). Model selection in omnivariate decision trees using Structural Risk Minimization. *Information Sciences*, 181(23), 5214-5226.
- Zhuge, H. (2011). Semantic linking through spaces for cyber-physical-socio intelligence: A methodology. *Artificial Intelligence*, 175(5), 988-1019.



© 2016 by the authors; licensee Growing Science, Canada. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).