THE ROLE OF INFORMATION TECHNOLOGY AND DATA PROCESSING IN THE MARKETING SECTOR OF COMPANIES PROVIDER PAYMENT SERVICES USING THE DATA MINING METHOD OF DECISION TREES

[1]ROUHOLLAH MALEKIAN, [2]MEHDI ESMAEILI

[1,2]Engineering Department of Information Technology, Islamic Azad University, Electronics Unit, Tehran, Iran
Email: [1]infopecw@gmail.com, [2]msxp12@yahoo.com

I. INTRODUCTION

Economy in today's world has different business sections compared to the close past so that in the area of free trade or trades limited to the borders of the countries, a new type of trade has emerged as e-commerce that has a specific place in the economy covering a significant amount of goods and services trading in traditional and electronic markets, in this space the payment service provider companies are formed, which are the maker of electronic payment industry and its developer.

On the other hand, in these electronic transactions, electronic funds are transferred from one side to another, this is the same point that banks and financial organizations as the main part of the value chain of systems exploiting the benefits of electronic commerce services have stepped toward developing these systems, so the rapid growth and development of bank cards as part of the premises and the infrastructure of creating and developing the electronic markets has been predicted for payment in transactions and companies providing the payment services are working with banks and deliver services to holders of bank cards in the framework of financial transactions as payer of electronic funds and adopters of bank cards as vendors providing goods and services.

Due to the large investments done by the companies and its trading partners and also a competitive market that providers of payment services are facing, lack of using the services of these companies, for any reason, will cause great losses for firms providing payment services and other value chain. By studying the behavior of existing customers and determining their value, we can establish an infrastructure of marketing strategies, this process will be possible by applying data mining. Today the power of data mining has been completely revealed for firms[2]and lots of academic research have been carried out from the past up to now to integrate the applicable software, because customers' behavior become more and more complex. Organizations and companies should focus on new marketing strategies based on modern marketing techniques and customer preferences. The management of relationship with customers is one of the systems that may has an important ability and competence on organizational decision-making processes. The management of relationship with customers helps companies collect information to influence the current and future needs and demands of customers. The most valuable asset of companies is information that relates to the variety of customer purchasing samples.

Data mining is one of the latest advances in databases management knowledge for marketing, and it can also help companies to move toward customer-oriented marketing. In this regard, the segmentation of the market are classified in line with the establishment of marketing strategies and customers [5]. The marketing unit seeks to maximize the customers' revenue and customer's profiting, more intersecting selling and these goals are possible using data mining. Data mining plays an important role in increasing corporates' profits by improving the performance of marketing. Decision making and data mining support systems are used to organize and improve marketing [6].

Marketing has a direct, strategic, and inseparable relationship with customers' relationship management, and some researchers have been performed that aimed at studying the e-commerce to integrate customers' relationship management with a structural model of data mining to improve business forecasting and create a strategy for establishing the strategic customers' relationship management that puts the customer at the center of their attention [8]. In this research, in order to exploit the vast majority of financial resources in the field of e-commerce, the concepts of developing intelligent marketing and service strategies are emphasized using the data mining tool in the field of clustering.

II. MATERIAL AND METHODS

2.1.Data Mining

Data mining is an unknown method of extracting the information, it is the discovery of useful, profitable and potential knowledge, and the purification of data from raw data stored in data warehouse using pattern recognition techniques. The two main goals of data mining are to predict and describe. Data mining with

the aim of forecasting is a systemic model that can predict with specific data. Its overall goal is to provide a model for classification, forecasting and evaluation. Descriptive data mining refers to new information based on a set of data, and analytics systems utilize the knowledge and communication that they create. In the research [6], it is said that the methods that enable companies to impact on their customers have been changed. The customer does not guarantee the long-term business continuity and success.

Companies have realized that they need to understand their customers better and respond to their needs and desires quickly. Additionally, the time response frame has been reduced. Companies cannot wait for a long time to observe the signs of customer's dissatisfaction, but before any dissatisfaction, special action must be taken. In order to succeed, companies must always be active and predict what customers want and prefer. Data mining is not an abstract and low-value extraction of a novel, but rather it is an understandable and documented knowledge of a set of data. Data mining is a process that uses data analysis and various models to discover patterns and relationships in order to construct and create an accurate prediction. Data mining will help you select and focus on the right prediction and offer the value of needed additions to existing customers and it identifies good customers who may leave the organization. Finally, the result is that the revenue is improved because it enhances the ability of companies to create any personalized communication. In this research, the formation of data mining is presented in Fig. 1.

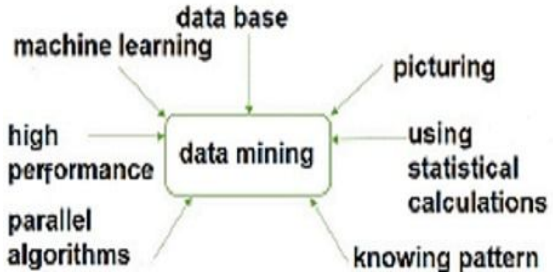


Figure 1. The Sciences that Participate in Data Mining [6]

Data mining is a process of discovering interesting business information and extracting knowledge from a large amount of data. The process of data mining is being considered in countless domains, where data mining can prove promising tools to retrieve useful information and provide researchers with a new perspective [9]. Data mining will discover new patterns and customer behavior rules, as well as discovering the consumer buying behavior to help companies focusing on a global and regional phenomenon, and it provides coordination and jointly decision making on organizational levels of a business for customer sectors and market. And it gives the companies the opportunity to make the best decisions and get the top spot in the competition market [10], [11]. In the research [12], data mining has been defined as the process of extracting, transforming, and loading data into a warehouse, storing and managing a multi-dimensional database system, providing data access conditions for business analysts and IT professionals. Data mining is a multi-stage process including diagnosis of abnormality, learning the community rules, classification, regression, and summarization and clustering. In data mining, the data are searched using two methods of learning with or without a supervisor. Figure 2.

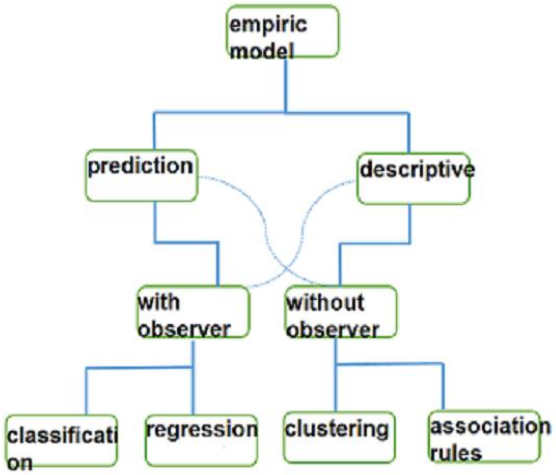


Figure 2. Learning Methods with Observation and without Observation in Data Mining

The Tasks of Prediction

The goal of predicting the value of a feature and its specific features based on the number and value of other features. The feature, which is used to build a prediction, is called the independent variable. Other values that are predicted are called the target or dependent variable.

Supervised Learning

In this type, data mining is based on training, data in use, at the same time, put the questions and answers for learning to the learner model that is being produced in its place. These methods are fast and accurate. Neural networks models, multilayer perceptron, decision trees are some of monitoring algorithms.

Classification Algorithms

The overall shape of classification algorithms of data mining resembles the others and as an example the algorithm of ID3is considered to predict the unknown amount of features. Note the following steps:

- Start
- Selecting the categories to be classified
- Implementation of Basic Classification Algorithm
- Receiving the results
- Conclusion
- The end

The Algorithms of Prediction

The algorithm ID3is useful to predict the amount of unknown samples and includes the following steps:

- Start
- Getting the name of the product to be foreseen
- Applying the algorithm ID3
- Receiving the result of both yes and no
- The end

Predicting with a decision tree is a tree, whose branches are split from the nodes, and it is a choice between the elections and each leaf represents a decision. A decision tree is commonly used to obtain information in order to make a decision. Decision tree begins with a root node and from this node, users divide recursively according to the learning algorithm of the decision tree. The final result of a decision tree is that each branch represents a possible scenario of its decision and outcome. Decision tree classifies by scrolling from the root node to the leaf node. We start from the root of the decision tree of the experiment and determine the character by this node and move toward the down branches of Pine according to the character value delivered in data set, this process is repeated in each sub-sector.
ID3 is a simple learning algorithm of decision tree.

III. RESEARCH METHOD

3.1. Collecting Data

Data collection is considered as one of the vital stages of data mining, whose type and quality has a significant impact on data mining and exploitation of the potential of various algorithms and data mining methods. In this section, data collection is carried out in two areas:
- Using the database that includes financial transaction information of card readers.
- Using the experts of the periodic and emergency inspection unit of the active companies providing payment services. In this research, 12 people are used to collect data in field sectors.

Collecting Data From the Database

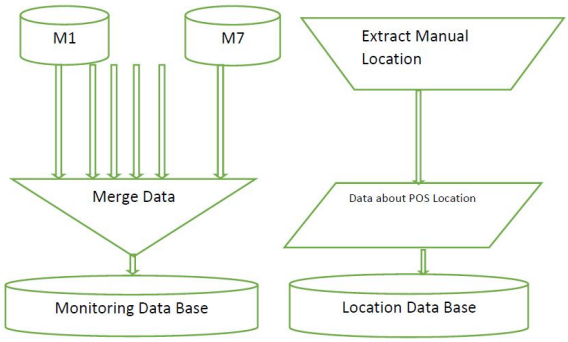
It refers to data that is entered in the database system in a one-month period, that is, the transactions resulting from the operation of the cardholders of the store are as follows, and the aggregation of micro-transactions is included in the number and sum of their value. Figure 3-1

The registration of the data of regions and the routes of the recipients, which confirm the declared addresses by referring the experts to the shops, and the region and the route for it is created and entered in the system.

IV. DATA TABLES

Monthly Monitoring Table

It is a table that contains information of the recipients and it consists of 42 fields.



(Figure 3-1) The nature and sources of online and offline data used in the research Field Data Collection of Areas

Regions' Table

In this table which is formed from 6 fields, the information of the recipients' mailing addresses is referred to the two sections of the region and the route.

Data Planning

In order to prepare data for the continuation of data mining operations, it is necessary to aggregate the tables and the required characteristics for data mining methods be gathered, so at this stage, the financial transaction tables, the areas under investigation and their aggregation process is begun.

Accumulation of 7-month-old Store Terminal Tables

The store terminal tables that are used as "Tables of Monitoring" in this study, include all store card readers tools, which are from the year 2005 until the end of the monitoring file of that month in 2015 and 2016, and includes the general information related to card readers and aggregates of the micro-transactions of a terminal including the sum of the resulting amounts. The monitoring tables have 42 properties, the names of their properties and their type have already been mentioned, but in order to prepare the tables for aggregation and use in the processes, first of all, among 42 of the characteristics, some of them that are not used in data mining operations should be eliminated, and this process is selected manually by an expert. The reason for deleting additional fields is their inactivity and their existence in the last monitoring table in August 2016. 7 monitoring tables are connected to each other based on the terminal number. How to do the process has been described in the shape (3-2).

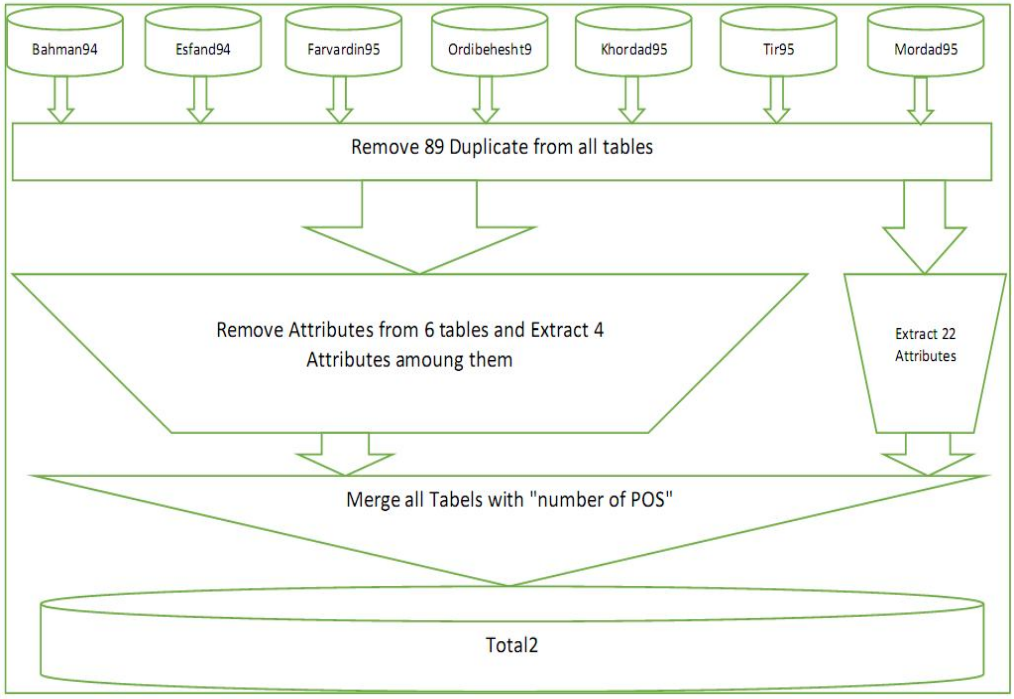


Figure 3-2(Aggregation of monthly monitoring of total transactions)

Table of Total 2 that is produced from the combination of seven monitoring tables will be used in the next sections to add other tables of areas, latitudes and longitudes, periodic and emergency surveys.

Classification of the Receptive Based on Transactional Levels

Classification Through Existing Concepts

In this section, due to the understanding of the business concepts of payment service companies, a classification model was carried out based on the number of transactions by the researcher, which resulted in the division of owners of the card reader at one month in the area of transaction volume at the levels of Great, Very good, Good, Middle, Weak, Wonder full, Not Active as in Table (3-3):

(Table 3-3) Classification of Transactions per Month

Up to transaction	From transaction	class	row
0	0	Not Active	1
8	1	Weak	2
249	9	Middle	3
500	250	Good	4
1000	501	Very good	5
1500	1001	Great	6
Bigger	1501	Wonder full	7

This action is performed via Rapid Miner and by the Discretize, Loop Attributes, Retrieve operators as follows, the Discretize operator is used within the Loop Attributes operator. In the sequel to this process, we combine the fields containing monthly transactions together and we create a field called Number of the numerical transaction, which are divided in the levels of Great, Very good, Good, Middle, Weak, Wonder full, Not Active, as in Table (3-4):

(Table 3-4) Classification of samples based on the number of transactions in 7 months

Up to transaction	From transaction	class	row
0	0	Not Active	1
56	1	Weak	2
2000	57	Middle	۳
3500	2001	Good	4
7000	3501	Very good	5
10500	7001	Great	6
Bigger	10501	Wonder full	7

Generating the Base Table of Total 2

In this section, the base table of Total 2 contains 17205 samples and 84 properties that the properties of Number of PM, PM + EM, Number of service_EM have been mentioned by creating a fully-qualified copy of the fields and with changing the type of integer to polynomial using the Generate Copy operator and Numerical to Polynomial operator ended. Performing an operation to create a conformance version and modifying the type of properties in order to make use of a number of tree decision methods is made.

Transaction Asnaf Table

The Asnaf Transaction table contains the aggregation of the transaction number that has taken place over the seven months of the research period in each class. The value attributed to the Row Labels expresses the

concept of how many stores exist in a particular class, and the values attributed to the property of the Sum of number of transaction Numerical determine determines the number of sum of the transactions performed in all its stores of classes.

Senf Mone Table

The Senf money table includes the aggregation of financial amounts resulting from the number of transactions that occurred during the seven months of the research period in each class. A value attributed to the Row Labels attributes the concept of how many stores exist in a particular class, and the values attributed to the Sum of Sum of Money attributes the total amount of transactions transacted in all stores of that class. The two values of Sum of Sum of Money and Sum of number of transaction numerical are added to the TOTAL 2 table. It can be seen that the number of transactions in the guilds is of a considerable importance due to the high number of stores in that class, while there are also a number of trades that are not so much but they have high transaction potential, which is also worth considering in the financial amounts resulted from transactions. For this reason, due to the normalization of the value of the guilds, we divide the characteristics into the number of stores in that class so that the real value of each class is determined using the averaging criterion, the result of this operation is the creation of two new qualities under the title of average Transaction and average money.

V. DATA ANALYSIS

5.1.Classification with Data Mining Techniques

The Beginning of the data Mining Process

After performing the preprocessing steps and determining the dependencies of properties on each other, we begin data mining operations in following domains and formed with simple expressions.

- Coordination of marketing channels
- Marketing Strategies

Coordination in Marketing Channels

Owners of marketing channels in the business of attracting transactions of card readers are looking for two objectives that seem to be separated from one another, so that companies delivering the payment services seek to increase the number of transactions and share the broad market and on the other hand, banks are seeking the proceeds of transactions, and the analysis of how to achieve these goals and lack of their conflict with each other is a useful step to implement the decision of the owners of marketing channels.

Decision Trees

Using a decision tree in the processes of the companies providing payment services, can be an effective step in predicting the events and how to provide services to the customer, so that a predictive model for analyzing is able to reduce doubt on strategies. In the following, a number of algorithms of decision trees will be analyzed for data mining in the field of obtaining information from aggregated knowledge in our previous discussions.

explanations	execution time	Accuracy	Tag attribute	Characteristics used in the model	Number samples	Model	row
Pay attention to the abundance of property values Sticker in the drawing The decision tree, some of the values of the attribute are not plotted in the decision tree	5 Second	87.9% 3	the number of transaction	Sum of money, Number of transaction	172 0 5	Decision Tree	1
Not paying attention to the abundance of all the values of the label's attribute and the tree's drawing with all its values.	2 Second	% 87.7 4	the number of transaction	Sum of money, Number of transaction	172 0 5	Random Forest	3

Marketing Strategies

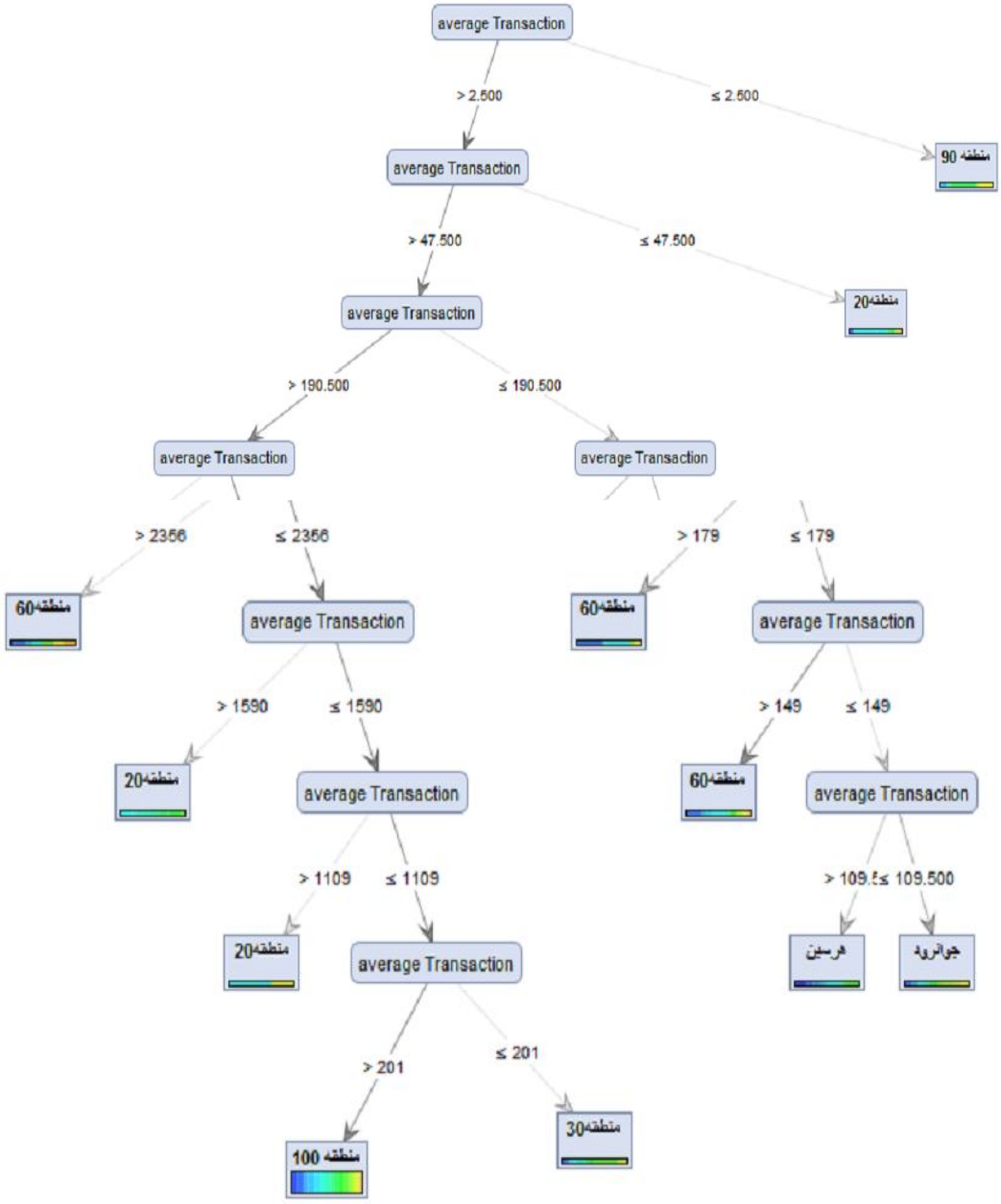
Marketing Strategies According to the most Valuable Business and Areas

Marketing strategies can be reviewed and changed based on different characteristics. To this end, we analyze some of the features in different models:

Analyzing the effect of features like average money, average transaction, trade and bank on the economic areas takes place as feature of "name of region".

Random Forest Model

In this section, especial properties of Average money, Average transaction, trade and regions are applied in the model of Random Forest and the feature of areas is analyzed as sticker, the important point is that in the analyzed output, the especial properties of Average money and the class is not used in this type of tree (Fig. 4: 1).

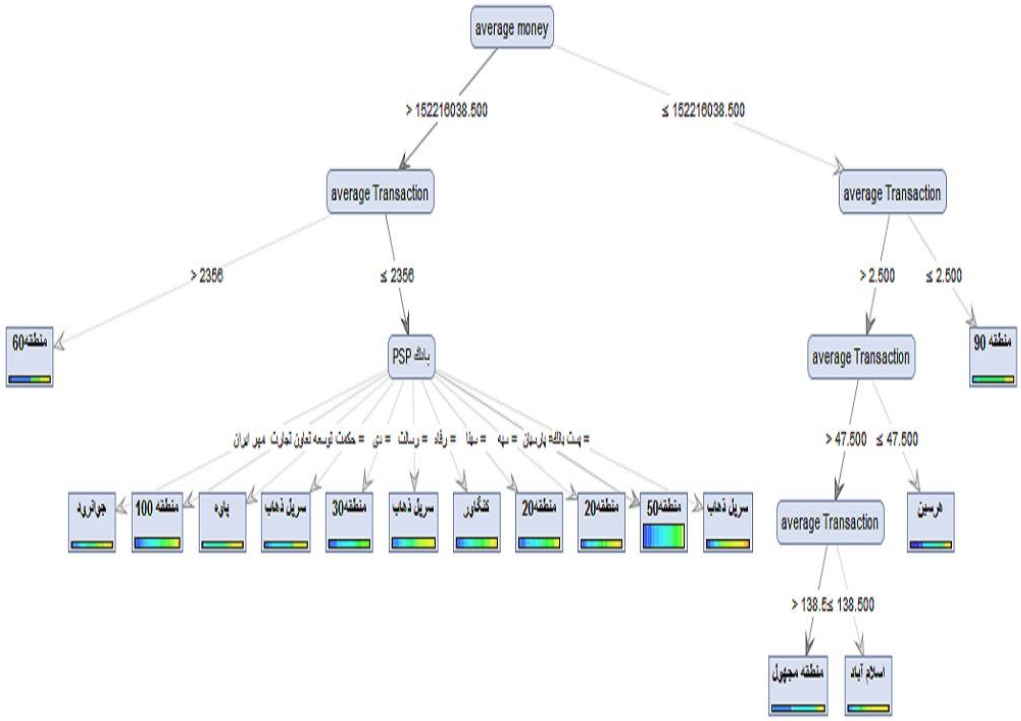


(Figure4-1)Reviewing the average money, average transaction, trade on areas in the model Random Forest

Marketing Strategies According to the Most Valuable Guilds, Areas and Banks

In this section the especial properties of Average money, Average transaction, Guilds, areas and Banks are used in the Models of Decision tree and Random Forest. The output of analysis expresses the full-transaction Guild sin related areas that are appeared in the outputs.

In some of outputs of Random Forest Model it is seen that with adding features of banks ,the behavior of Random Forest at ignoring the feature of guilds and Average money in previous analysis will be solved. The output of this model in addition to emphasizing the conscious marketing strategy, indirectly refers to the lack of strategy conflict in the marketing channels in categories of (Bank branches deployed at areas).



(Figure4-2)reviewing the average money, average transaction, Bank, Class on regions in Random Forest Model

With Changing the Sticker Feature in Decision Tree Model, the Strategy Marketing of active banksis analyzed at card reader area, the Output of This Model introduces Active Banks in the area off ull-transaction stores more successful than other banks and it even expresses that the project guilds of plaquere placement , Etka and Kosar stores, Prison that are absorbed by Parsian Banks, the Financial and credit institute of Kosar and Pecco don't have successful marketing strategy until the time of analysis. The output of all Modes of Decision Tree, Random Forest analyze the Issue of Marketing Strategy that delivers it to the banks and companies delivering the payment services at the area of card reader based on the geographical regions and beneficial guilds. At analysis that is considered with Decision tree and CHAID models in line with the impact of the feature of region's name (Parent) and class on Number of Transaction provided that those samples are used that have the transaction levels of Wonder full, very good, great, we intend to analyze the impact of other features on the above mentioned -levels.

At Above Results it is Specified that Models of -Decision tree and CHAID do not consider the impact of Independent Features of Guild and Areas and they analyze the model Decision Stump ,the quality of "Guild" and the model ID3 enters the Features of-"Guild" and "Region" at the analysis. The analysis of above mentioned models is in line with the improvement of Marketing Strategy and it expresses this issue that which guilds in which areas have the most trans action. Also which areas and guilds do shops with the greatest transaction exist? In the following, we added the quality of the bank to the total independent qualities to observe its impact as ID3. It is seen that in this part in addition to the independent variables of area, the bank variable process is also analyzed in the transaction process. We used other models of CHAID, Decision Tree, Random Forest and Decision Stump, whose output will be mentioned in the conclusion table.

(Table4-3)The Output of decision trees at marketing Strategy based on the most valuable Guilds, Banks and areas

Row	Model	Number of Samples	Features used in the Model	The Feature of Sticker	Target	Description
1	Decision tree	17205	Average money, Average transaction , class and areas, bank	Area	Marketing Strategies According to the most valuable Guilds,	The important point is that despite the Random Forest at analyzed Output the properties especially the Average money and Class are used at this Kind of Tree. This Output analyzes

Row	Model	Number of Samples	Features used in the Model	The Feature of Sticker	Target	Description
					areas And Banks	the Capillary Marketing Strategy based on the Important Principle of Guilds' Value and areas Analysis You can **A.**
2	Random Forest	17205	Average money, Average transaction , class and areas, bank	Area	Marketing Strategies According to the most valuable Guilds, areas And Banks	The important point is that at analyzed Output the properties especially the Average money and Class are not used at this Kind of Tree.
3	Random Forest	17205	Average money, Average transaction , class and areas, bank	Area	Marketing Strategies According to the most valuable Guilds, areas And Banks	At Some of Outputs of Random Forest Model it is observed that by adding the quality of banks, the behavior of Random Forest model in ignoring the quality of business and Average money at past Analysis is solved. the output of these models in addition to emphasizing the cautious Marketing Strategy, indirectly refers to the lack of Strategy conflict at the channels of Marketing At categories of (bank branches Deployed At Areas).
4	Decision tree	17205	Average money, Average transaction , class and areas, bank	Area	Marketing Strategies According to the most valuable Guilds, areas And Banks	The analyzed output expresses the Full transaction Guilds at related areas that appears at outputs.
5	Decision tree	17205	Average money, Average transaction , class and areas, bank	Banks	Marketing Strategies According to the most valuable Guilds, areas And Banks	The output determines these active Banks at the Area of Full transaction stores as more Succeeded than other Banks and it even Expresses that specified Guilds that are absorbed by Special Banks, do not benefit from a Successful Marketing strategy until the time of analysis.

DISCUSSION AND CONCLUSION

In this study, by aggregating the field data of postal addresses and transactions of card reader devices that have been Distributed at different geographic areas, the conditions of creating the smart marketing strategyis analyzed using previous data of classes and geographic regions, on this analysis that is done using decision trees' tools in data mining, models of ID3 and Decision Tree and CHAID and Random Forest are used and data mining is carried out based on Classroom and areas fields, Number of Transaction and the obtained results include appropriate accuracy of models CHAID and Decision Tree and it expresses that for Marketing we should work in what areas and routes and on what groups of trades. For future research we will use the latitude and longitude in the area of marketing.

REFERENCES

[1] Khobzi, H. Akhondzadeh-Noughabi, E. and Minaei-Bidgoli, B. (2014). A New Application of RFM Clustering for Guild Segmentation to Mine the Pattern of Using Banks'-Payment Services. Journal of Global Marketing, 27:178–190

[2] Ahlemeyer-Stubbe, A. and Coleman, SH. (2014) A Practical Guide to Data Mining for Business and Industry, This edition first published 2014 John Wiley & Sons, Ltd.

[3] Gupta, G. and Aggarwal, H. (December-2013). Overview of CRM using Data Mining. International Journal for Multi-Disciplinary Engineering and Business Management, Volume-1, Issue-1

[4] Sunil Kumar, S. (July 2015). Increasing the Efficiency of CRM Process Using Data Mining Practices.

[5] International Journal of Advance Research in Computer Science and Management Studies, Volume 3, Issue 7.106-11

[6] Jiang, H. Yu, Q. Liu, C. Zhu, Q. and Guo, L. (2013 IEEE). The Analysis of CRM Customer Information Based on Data Mining. 2013 Ninth International Conference on Natural Computation (ICNC). 978-983

[7] Senkamalavalli, R. and Bhuvaneshwari, T. (2014 IEEE). Data Mining Techniques for CRM. ICICES2014 - S.A.Engineering College, Chennai, Tamil Nadu, India, ISBN No.978

[8] Ms. Simi Sojan, Ms.Sify K Raphy, Ms. Pinky Thomas. (2014). TECHNIQUES USED IN DECISION

[9] SUPPORT SYSTEM FOR CRM- A REVIEW. International Journal of Information Technology

[10] Infrastructure. Volume 3, No.1

[11] Sadat, L. (April 2013). Data Mining in E-Commerce: A CRM Platform. International Journal of Computer Applications (0975 – 8887) Volume 68– No.24

[12] Lu, N. Lin, H. and Lu, J. Member. (2011), A Customer Churn Prediction Model in Telecom Industry Using Boosting. IEEE and Guangquan Zhang.

[13] Xiuhong,W. and Fudong,W. Review on the application of data mining technology in apparel logistics, Proceedings of the 7th International Conference on Innovation & Management: 2104–.9012

[14] Kimball, R., Reeves, L., Ross, M. and Thornthwaite, W. (1998). The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses, John Wiley & Sons.

[15] Kaur Mann, A. and Kaur, N. (April 2013). Survey Paper on Clustering Techniques. International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 4

[16] Kaur Mann, A. and Kaur, N. (April 2013). Survey Paper on Clustering Techniques. International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 4

[17] Yadav, S.; Desai, A. and Yadav, V. (July-2013). Knowledge Management in CRM using Data mining Technique. International Journal of Scientific & Engineering Research, Volume 4, Issue 7

★★★