Journal of AI and Data Mining
Published online:

Analyzing and Investigating the Use of Electronic Payment Tools in Iran using Data Mining Techniques

F. Moslehi, A. Haeri*, and A. Moini

Dept. of Industrial Engineering, Iran University of Science & Technology, Tehran, Iran.

Received 28 January 2017; Revised 22 April 2017; Accepted 19 September 2017
**Corresponding author: ahaeri@iust.ac.ir (A. Haeri).*

Abstract
In today's world, most financial transactions are carried out using electronic instruments and in the context of information technology and internet. Disregarding the application of new technologies in this field and sufficing to traditional ways will result in financial loss and customer dis-satisfaction. The aim of the present work is to survey and analyze the use of electronic payment instruments in banks across the country using the statistics and information retrieved from the Central Bank and the data mining techniques. For this purpose, initially, according to the volume of the transactions carried out and using the K-Means algorithm, a label was dedicated to any record. Then the hidden patterns of the E-payment instrument transactions were detected using the CART algorithm. The results obtained enable bank administrators to balance their future policies in the field of E-payment and in the bank and customers' interest direction based on the detected patterns, and provide higher quality services to their customers.

Keywords*: Banking, Data Mining, Electronic Payment Instruments, Classification, CRISP-DM.*

1. Introduction
In the recent years, information and communication technology (ICT) has made a significant progress.
Over time, the role of information technology (IT) in all industries has become more prominent, and the banking industry has not been an exception to this. Banking has been a strong industry for innovation concerning the information systems and technologies [1]. One of the impacts of IT in this area has been the development of electronic banking. From the 90s to the present time, electronic banking (E-banking) has become a distribution channel with the greatest potential for financial institutions [2]. Development of E-commerce is another advantage of IT as E-banking is an essential tool for the implementation of E-commerce. ICT enables the banks to reach the market by reduction in costs, and provides excellent customer services by overcoming the physical infra-structures and physical distance [3].ICT has changed the way of conducting business transactions, and has met the growing demands of customers for most organizations. The promise of ICT in the banking sector is its potentials to increase customer base, reduce transaction costs, improve quality and timeliness of response, enhance opportunities for advertising and branding, facilitate self-service and service customization, and improve communication and relationship with customer [4].
Data mining (DM) is a relatively new and promising technology, which can be defined as the process of discovering meaningful new correlations, patterns, and trends by digging into (mining) large amounts of data stored in a warehouse using statistics, machine learning, artificial intelligence (AI), and data visualization techniques. DM is concerned with finding the hidden relationships present in business data in

order to allow businesses to make predictions for future use. It is the process of data-driven extraction of not so obvious but useful information from large databases. Data mining has emerged as a key business intelligence technology [5].

According to the competitive environment in the banking industry, in order to maintain the existing customers and attract the new ones, each bank should adopt appropriate decisions and policies. In this regard, banks with understanding of customer preferences and existing trends in this industry provide services tailored to the customers' requests. E-banking is one of the areas in which the customers are very interested. This is why banks require accurate analysis and exact decisions for expanding E-banking to survive in this competitive environment. One of the technologies that can be useful in this regard is DM. Due to the high volume of data available in the banks, this research work attempted to use DM as one of the information technology tools to analyze the E-payment transaction trends in Iran banking system, and shows hidden patterns in the available transactional data as a set of rules. Also the results of this work will increase the ability of a bank manager to provide better E-banking services and E-banking future policies adjusted in the interests of both the customers and the bank based on the electronic payment analysis and discovered patterns.

The present research work aimed to answer the following questions:

- What factors affect the number and amount of transactions conducted with the E-payment instruments?
- Is there a significant relationship between the period and the volume of the transaction?
- Is there a significant relationship between the bank and the province in which transactions are carried out with the E-payment instruments?

The rest of the paper is organized as what follows. Section 2 reviews the E-banking and data mining literature. Section 3 presents the research methodology. Empirical study and data analysis are presented in Section 4. Finally, in Section 5, conclusions and implications are considered.

2. Literature Review

2.1. Electronic Banking (E-Banking)

With the advent of the internet, the original notion of electronic banking was formed in 1991, meaning that customers not being in a branch are able to do their banking affairs by going to the cyberspace [6]. E-banking has many advantages such as no time and spatial limitation, easy access to the information, reduction in costs and saving customers` time, which have caused the rapid growth of E-banking services [7]. Using E-banking services is one of the solutions for banks to acquire the competitive advantage, and has created a close competition in this field. In these conditions, the customers' expectation level to receive these services has increased greatly [8].

2.2. Data mining (DM)

DM refers to the derivation and extraction of knowledge and useful information from a set of data [9].

In other words, DM is presented as a process of extracting knowledge from a set of data using the intelligent techniques. In fact, data mining is considered as a combination of several techniques: data management, statistics, machine learning, and visualization. The information and knowledge obtained from the data can be used for different applications such as management, trade and business, production control, market analysis, and design engineering. Generally, the DM techniques are used for two purposes: description and prediction. Prediction includes classification, regression, etc. The descriptive techniques include clustering, anomaly detection, and so on [10].

2.3. Payment systems in Iran

E-payment in Iran is not long ago. The introduction of new electronic payment instruments dates back to 1370, when Sepah Bank ATM service was inaugurated, which offered the first prototype of the cards with withdrawing money from the ATM terminals to the bank network customers. The bank card network in the Islamic Republic of Iran has been integrated by inter-exchange network information (SHETAB) since 1381, and now all of the issued bank cards are accepted at all terminals installed across the country. SHETAB manages E-payment in Iran.

2.4. Related works

2.4.1. DM applications in banking industry

Due to the flexibility and power of machine learning techniques to predict, these methods have been used in various fields.

For instance, Gharipour et al. have used these techniques to compare the approximate economic behavior ability of artificial neural networks (ANNs) and support vector machines using a set of data in some Middle East countries. For their experiments, they used a World Bank data set including 200 financial indicator data. The results obtained showed that neural network outperforms support vector machines both in terms of generalization from training dataset and accuracy of approximation [11]. In the field of banking, different methods have been used for different purposes; among these studies, we can refer to the research work performed by Nie et al. They built a customer churn prediction model using the logistic regression and DT-based techniques within the context of the banking industry [12]. Hu and Liao have evaluated the electronic service quality for internet banking with the fuzzy MCDM as a feed-forward neural network and a genetic algorithm-based method to automatically determine degrees of importance of respective criteria. They identified critical criteria for evaluating the service quality using these techniques [13]. Adeoti has surveyed the dimensions of ATM frauds and has proposed solutions that will decrease the ATM frauds in the Nigerian banking system. He used the chi-square statistical technique to analyze the data and test the hypothesis raised. The conclusion is that both the bank customers and the bankers have a joint role to prevent the delinquent of ATM frauds in the banks [14]. Khashei et al. have investigated credit risk evaluation using the DM classification techniques. The authors used the hybrid model FMLP (fuzzy MLP) for credit scoring assessment based on ANN and fuzzy logic. The article utilized fuzzy numbers to update the weight and bias of MLPs so as to achieve better results when both uncertain and complex data exist. This FMLP model outperforms other models such as SVM, KNN, QDA, and LDA [15]. Liébana-Cabanillas et al. have reviewed different DM techniques to achieve effective measures to attract the users trust in E-banking, and have concluded that the best method to perform variable selection of the best method to perform variable selection according to the expert's opinion is MGA using Mutual Information [16]. Khobzi et al. have proposed a new application for guild segmentation using (RFM)-based clustering and point of sale data. The most profitable guild is the gold segment, and carrying is the least profitable one; this research work used the point of sales data [17]. Farquad et al. have used SVM to predict customer churn from bank credit cards. They introduced a hybrid approach to extract rules from SVM for customers` relationship management purposes. The approach composed of three phases, where: 1) SVM-recursive feature elimination was applied to reduce the feature set; 2) the dataset obtained was used to build the SVM model; and 3) using NB, tree rules were generated [18]. Calis et al. analyzed the credit risk of the customers who had loans, and customers' repayments were estimated. For this purpose, the K-Means, CART, and C5.0 techniques were used. The most important variable in the analysis process using the CART algorithms was the amount of monthly salary, and the C5.0 algorithm was the education level [19]. Farokhi et al. have investigated the application of DM approach in the implementation of the customer relationship management system. The information gathered from Point of Sales (POS) in one of the Iranian private banks was used in this research work. They used the two methods K-Means and Kohonen to detect the most profitable customers, and clustered them into four segments [20]. Fadaei Noghani & Moattar have suggested a novel DM model for credit card fraud detection. Their proposed model was to consider the feature selection and the decision cost for accuracy enhancement of credit card fraud detection.

In this work, first, the best and most effective features were selected by an extended wrapper method, and then an ensemble classification was performed. After selecting the feature, the C4.5 decision tree was applied as a wrapper approach [21].

2.4.2. DM applications in trend analysis

In this work, the E-payment trend in Iran was investigated using DM techniques. In addition to the banking industry, the DM methods are used for trend analysis in other industries. In what follows, some examples of trend analysis studies in other industries are expressed.

Kudyba & Lawrence have analyzed the role of DM methods to increase product sales in an E-commerce platform. They believed that the DM techniques could improve online market effectiveness by analyzing the historical data for the previous purchasing patterns of customers [22]. Kaur & Kang have identified the changing trends of market data during the time using the association rule mining. They used "Extended bakery" datasets.

In this paper, a new algorithm was proposed to investigate the customer behavior, and assists in sales growth [23]. Karamizadeh & Zolfagharifar have used the DM techniques to analyze the third party insurance of Iran insurance company auto. They concluded that application of clustering algorithms could provide a model to identify the affecting factors and to determine their effects on the profit and loss of the auto third party insurance [24]. Many packages of products with different specifications and prices are offered in telecommunication (TC) companies. Insani & Soemitro have investigated the trend of providing services and products in the TC industry. They determined the factors affecting profitability using Market Basket Analysis (MBA) and detecting the customer behavior patterns [25]. Plessas-Leonidis et al. have presented a case study of the implementation of DM methods in revealing sales trends. They used sales transactions of the publishing industry.

In this work, the factors affecting sales such as the popularity of the books' authors and ineffective sales factors such as physical characteristics of a book were determined [26]. The reviewed works in this part are summarized and compared in table 1.

Reviewing the recent and related works indicated that there were no further studies about e-payment transaction trend analysis by DM techniques in Iran's banking industry. In this study with help of DM algorithms Iran's e-banking industry status is investigated.

3. Research methodology

3.1. CRISP-DM methodology

There are various methods for the implementation of mining projects. One of the powerful and common methods used for DM project management is the CRISP-DM method. The CRISP-DM methodology provides a structured approach to planning a DM project. This method provides a process model for DM that is an overview of the lifecycle of each mining project. The life cycle of a DM project consists of six stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. With the help of these steps, the data mining project is formed. Each phase has separate goals:

- Business understanding: This phase focuses on understanding the requirements of a business perspective, and then converting this knowledge into a DM problem definition. The literature and previous studies in this area and reports of Central Bank of Iran were reviewed in order to become familiar with the banking business.
- Data understanding: The data understanding phase starts with an initial data collection and proceeds with activities in order to become familiar with the data. In this research work, E-banking transactions were collected at first. This dataset includes 49425 records related to the ATM and Pin Pad transactions of bank, province, and month separately during the years 1386 to 1394.
- Data preparation: Cleaning and converting the data format take place in this phase. Preparation and refinement of the dataset are conducted. For instance, the duplicated records were merged and records with null values were omitted.
- Modeling: In this phase, various techniques are used to analyze the data and extract knowledge out of them. In this step, modeling using clustering and decision tree algorithms and related software was done. Data analysis by decision tree algorithms requires a categorical target variable. Our target variables were continuous numeric, and therefore, clustering technique was applied to assign a label to each record. At the end, the hidden patterns and rules in the dataset were detected based on the defined classes by decision tree algorithms.
- Evaluation: In this phase, the knowledge obtained in the previous phase is analyzed to determine the usefulness and application.
- Deployment: In this phase, the extracted knowledge is used to solve the business major problems.

The steps of a DM project are shown in figure 1.

Table 1 . Summary of reviewed works.

Title	Writers	Research scope	DM method
A Comparative Approximate Economic Behavior Analysis of Support Vector Machines and Neural Networks Models.	Gharipour, A., Sameti, M., & Yousefian, A. (2010).	Economic	ANN, SVM
Credit cards churn forecasting by logistic regression and decision tree.	Nie, G., Rowe, W., Zhang, L., Tian, Y., & Shi, Y. (2011).	Banking	Logistic Regression, Decision Tree
Finding critical criteria of evaluating electronic service quality of internet banking using fuzzy multiple-criteria decision-making. Applied Soft Computing,	Hu, Y. C., & Liao, P. C. (2011).	E-banking	Neural Network, Genetic-Algorithm
Automated teller machine (ATM) frauds in Nigeria: the way out.	Adeoti, J. O. (2011).	E-banking	Chi-square
A bi-level neural-based fuzzy classification approach for credit scoring problems.	Khashei, M., Rezvan, M. T., Hamadani, A. Z., & Bijari, M. (2013).	Banking	Fuzzy MLP
Analyzing user trust in electronic banking using data mining methods.	Liébana-Cabanillas, F., Nogueras, R., Herrera, L. J., & Guillén, A. (2013).	E-banking	Multi-objective Selection Genetic Algorithm,
A new application of RFM clustering for guild segmentation to mine the pattern of using banks' E-payment services.	Khobzi, H., Akhondzadeh-Noughabi, E., & Minaei-Bidgoli, B. (2014).	E-banking	RFM
Churn prediction using comprehensible support vector machine: an analytical CRM application.	Farquad, M. A. H., Ravi, V., & Raju, S. B. (2014).	Banking	SVM
Data mining application in banking sector with clustering and classification methods.	Caliş, A., Boyaci, A., & Baynal, K. (2015).	Banking	K-Means, CART, C5.0
A new application of clustering for segmentation of banks' E-payment services based on profitability.	Farokhi, S., Teimourpour, B., Shekarriz, F., & Masoudi, M. (2016).	E-banking	K-Means, Kohonen
Ensemble classification and extended feature selection for credit card fraud detection.	Fadaei Noghani, F., & Moattar, M. (2017).	Banking	C4.5 decision tree
Enhancing information management through DM analytics to increase product sales in an E-commerce platform.	Kudyba, S. and Lawrence, K. (2008).	E-commerce platform	Bayesian Network
Market Basket Analysis: identify the Changing Trends of Market Data using Association Rule Mining.	Kaur, M., & Kang, S. (2016).	Bakery	Association Rule mining
Market Basket Analysis: identify the Changing Trends of Market Data using Association Rule Mining	Kaur, M., & Kang, S. (2016).	Insurance	FP Growth, K-Means, Kohonen, Two Steps
DM for marketing in TC industry.	Insani, R., & Soemitro, H. L. (2016, May).	Telecommunication	K-Means, Kohonen, Association Rule mining
Revealing sales trends through DM.	Plessas-Leonidis, S., Leopoulos, V., & Kirytopoulos, K. (2010, February).	Publishing	k-Means, Naïve Bayes, Neural Networks, k-Nearest Neighbors

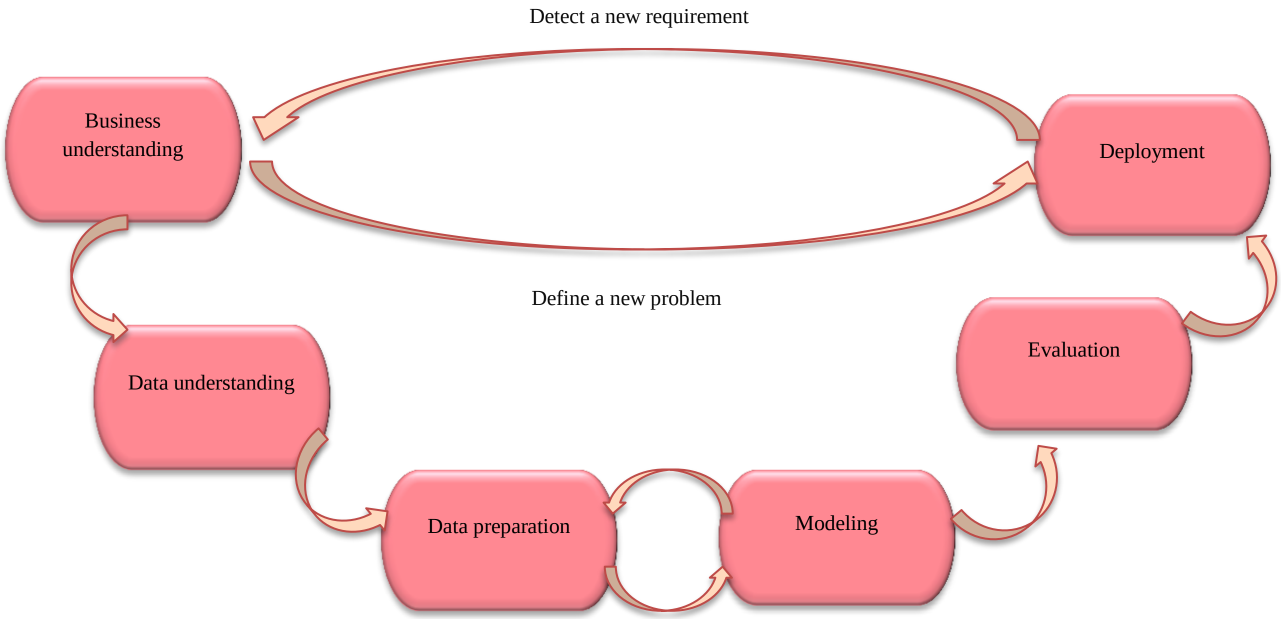

Figure 1. Steps of a DM project.

3.3. DM techniques

Clustering and classification techniques are used in this research work.

3.3.1. Clustering

Cluster analysis group data objects are based upon the information found in the data that describes the objects and their relationships. The goal is that the objects in a group be similar (or related) to one another and different from (or unrelated to) the objects in other groups. Number, shape, and properties of clusters are unknown at the beginning of the process of clustering, and since there is no prior knowledge of clusters, the clustering technique is called unsupervised technique. In this research work, the clustering algorithm was used for assigning a label to each record and determining the target field categories. To do so, the k-means algorithm was selected. The k-means algorithm is a simple iterative clustering algorithm that partitions a given dataset into a user-specified number of clusters, *k*. The algorithm is simple to implement and run, relatively fast, easy to adapt, and common in practice. It is historically one of the most important algorithms in DM.

3.3.2. Classification

Classification is the task of assigning objects to their respective categories. Classification is a supervised method in DM. It means that the purpose of classification is clear, and there are variables whose values are predictable from the values of other variables and data. The goal of supervised predictive models is to find a model or mapping that will correctly associate the inputs with the targets. In this research work, the classification algorithm was applied to analyze and investigate the factors affecting the number and amount of transactions conducted using the E-payment instruments. For this purpose, the CART decision tree was selected.

The CART decision tree is a binary recursive partitioning procedure capable of processing continuous and nominal attributes as targets and predictors. The data is handled in its raw form; no binning is required or recommended. Beginning with the root node, the data is split into two children, and each child is, in turn, split into grandchildren. Trees are grown to a maximal size without the use of a stopping rule, essentially the tree-growing process stops when no further splits are possible due to the lack of data.

The proposed process for specifying the factors affecting transactions conducted with the E-payment instruments is shown in figure 2.

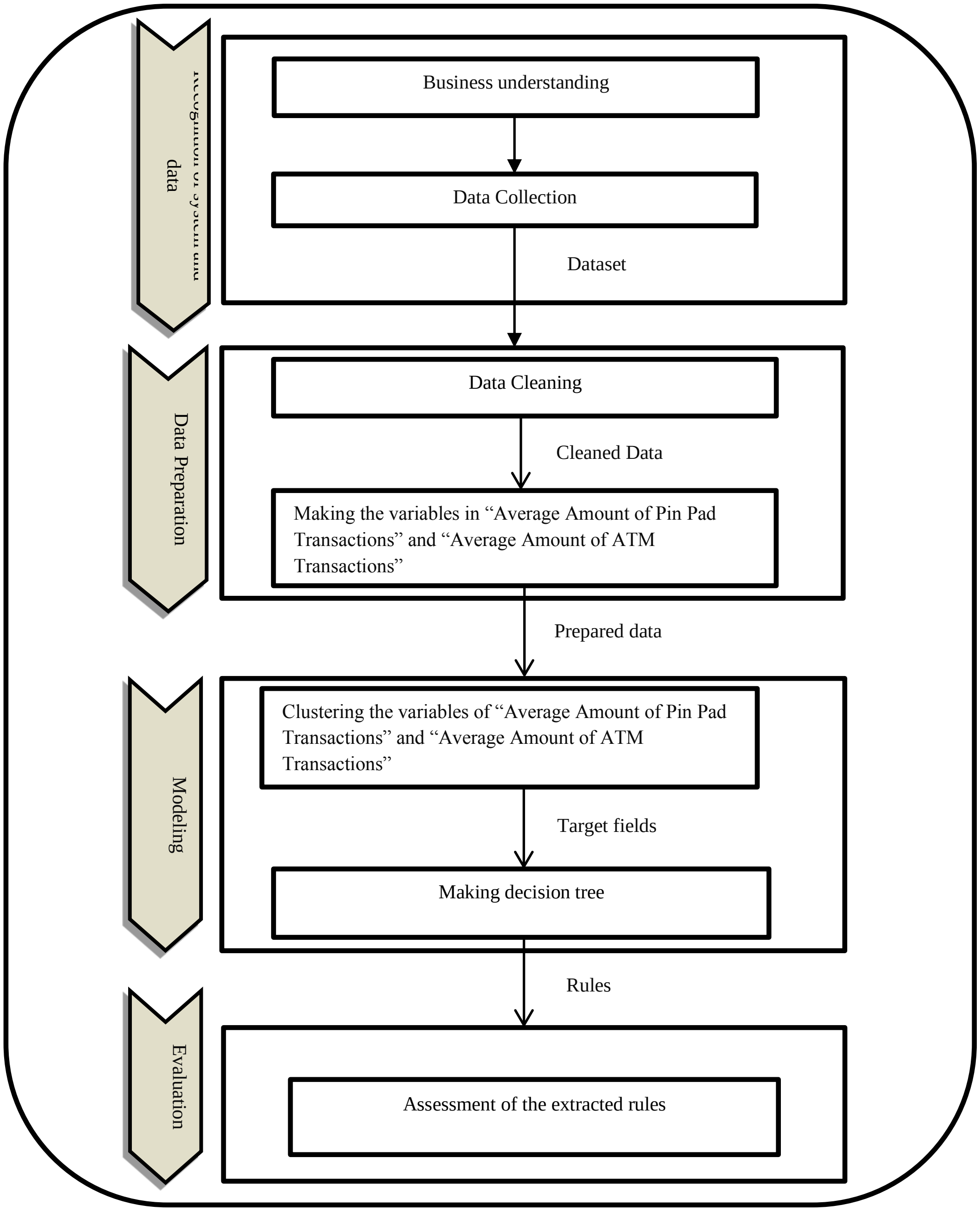

Figure 2. Proposed process for this work.

4. Research framework

In this research work, DM techniques were selected to analyze the transactions since the order and priority of variables could not be determined in the descriptive statistics methods but there is a possibility of analytic hierarchy in DM techniques using the decision tree. The formation of tree branches and priority of variables are determined using the information gain in the decision tree, and this advantage will cause a higher accuracy and a better detection of results. Another reason for using the DM techniques instead of descriptive statistics methods is the ease of achieving the desired results. In order to achieve the rules obtained by statistics methods, different filters should be applied on different variables in various time periods for each bank and province, and the results obtained need to be compared. This process is too time-consuming and might not get all the results. The desired results were achieved easily and quickly using the DM techniques.

A research method followed based on the CRISP-DM methodology to achieve the research goals. There are different ways to implement the data analysis projects, and one of these powerful methods is CRISP-DM.

As CRISP-DM world standard is used for the process implementation, the executive research structure is explained based on the standard levels, as follows:

4.1. Business understanding

This section focuses on understanding the project goals and essentials from the viewpoint of business that the client's financial transactions management is one of the bank business aims. In the today's world, most financial transactions are carried out through electronic instruments, and in the context of IT and internet. Disregarding the application of new technologies in this field and sufficing to traditional methods will result in financial loss and customers` discontent. Regarding this issue, finding relations through DM, which can help the Iranian banks to promote better E-banking services, seems to be necessary. In this work, the purpose of DM was to classify and extract knowledge and rules in each category based on DM algorithms.

4.2. Data understanding

In this phase, the published data present on the official website of the Central Bank of Iran (Www.cbi.ir) was used. Payment instruments in Islamic Republic of Iran have no long history but in the recent years, especially with the establishment of the SHETAB center as the country`s bank card national switching, the development of such tools has been accelerated. The Central Bank has published performance statistics of electronic payment tools including ATM and Pin Pad, separated based on the bank and province since 1386 until now for each month. In this work, the corresponding data was taken to 1394. In this work, 12 variables were used (Table 2).

Table 2. Variable definition.

Variable Name	Data Type	Variable Modality
Province	Set	qualitative
Bank	Set	qualitative
Year	Set	qualitative
Month	Set	qualitative
Number of ATM	Range	quantitative
Number of ATM Transactions	Range	quantitative
Amount of ATM Transactions	Range	quantitative
Average Amount of ATM Transactions	Range	quantitative
Number of Pin Pad	Range	quantitative
Number of Pin Pad Transactions	Range	quantitative
Amount of Pin Pad Transactions	Range	quantitative
Average Amount of Pin Pad Transactions	Range	quantitative

4.3. Data preparation

This phase of the CRISP-DM methodology includes data selection and data cleaning, and readies them for DM. Data cleaning actually is a qualitative control phase before data analyzing, and one of its tasks is to fill or remove the missing data. At this point, the transactions with incomplete or duplicated data were excluded. The dataset record number decreased from 65696 to 49425. Some of these records had a few fields with a value of 0. Other records were removed as being replicated, meaning that banks in some months had not announced new statistics, and duplicate statistics were inserted for that month. Another activity performed at this stage was applying the inflation rate to the field values. Since the data collection relates to a 9-years period, the money value at the time has changed, which affects the presented analysis. That is why the inflation rate, up to 1394, was applied to each year amounts.

4.4. Modeling

In this phase, the types of modeling techniques are selected and used. In general, there are several solutions for one type of DM problem. In this work, **Clementine 12** was used for applying methods.

The data used includes 49425 records related to the performance of ATM and Pin Pad electronic payment of bank, province, and month separately during the years 1386 to 1394. The main purpose of this work was to identify the factors affecting the E-payment tools in Iran according to the average amount of transactions. To do this, classification algorithm was applied to extract rules. Two objective variables were required in the classification algorithm implementation. Using this algorithm requires a categorical target variable. In this work, the target variables are “Average Amount of ATM Transactions” and “Average Amount of Pin Pad Transactions”. These variables had continuous numeric values. There were 2 ways to categorize target variables. Categorize them according to the distance of values and using clustering algorithm. There is no threshold to specify low, high or medium amount of transactions in the banking literature. Therefore, the clustering method was selected to categorize the target variables. In this work, the clustering algorithm was used to determine the target field categories, not for grouping dataset records. The clustering algorithm was applied to not categorize the target field based on the personal viewpoint of authors. The clustering algorithm was run for 2 times. First, time for categorizing the variable of “Average Amount of ATM Transactions”, and this field was the only input of this implementation. At the second time, “Average Amount of Pin Pad Transactions” field was categorized, and this field was the only input of this implementation. These steps are shown in Figure 3.

First, the K-Means algorithm was implemented with three clusters on the variable “Average Amount of ATM Transactions”. The average ATM transactions field values were divided into three clusters, Low, Medium, and High. The label was assigned according to the average ATM transaction amount to each record. Table 3 indicates the results of this clustering. Graphical representation of clustering result is shown in figure 4.

Table 3. Results of clustering on Average ATM transactions.

Cluster Number	Minimum Value	Maximum value	Record Number	Assigned Label
Cluster 1	0.29	2500165.22	42498	Low
Cluster 2	10831689.52	36011478.22	87	High
Cluster 3	2500805.92	10805639.21	6871	Medium

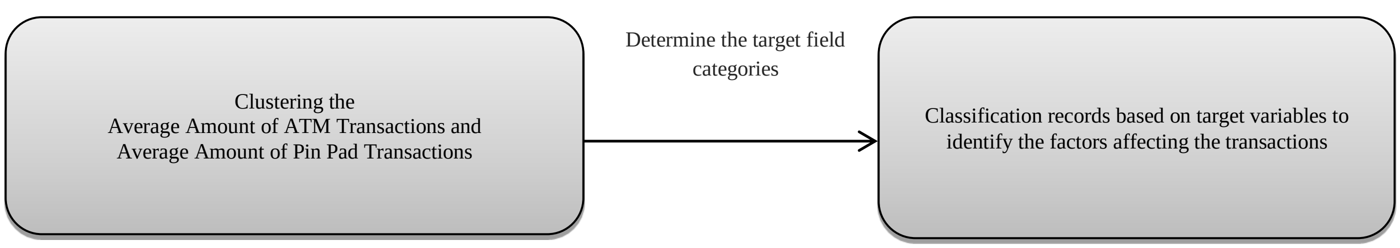

Figure 3. Determination of target field categories.

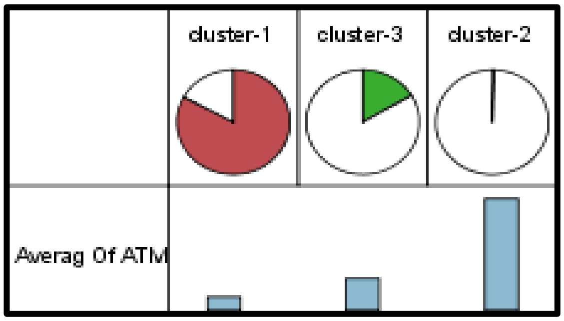

Figure 4. Clusters of variable of "Average Amount of ATM Transactions".

In the next step, "Average Amount of Pin Pad Transactions" field values were divided into three clusters, Medium, Low and High. The label was assigned according to the average of Pin Pad transaction amount to each record. Table 4 indicates the results of this clustering. The graphical representation of the clustering result is shown in figure 5. At the next stage, the CART decision tree algorithm was used to explore the factors affecting the average amount of ATM and the average amount of Pin Pad transactions. This algorithm was run for 4 times.

Table 4. Results of clustering on Average Pin Pad transactions.

Cluster Number	Minimum Value	Maximum value	Record Number	Assigned Label
Cluster 1	0	43290407.94	41118	Low
Cluster 2	520108047.03	1025370182.77	12	High
Cluster 3	43498385.89	68223635.48	8326	Medium

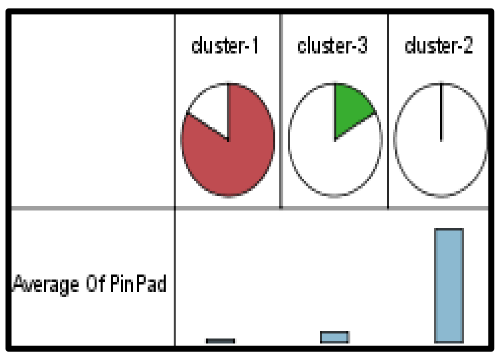

Figure 5. Clusters of variable "Average Amount of PinPad Transactions".

4.4.1. Results of decision tree of ATM transactions regardless of month and year

In the first implementation of the CART algorithm, the variables province, bank, ATM number, and Pin Pad number were considered as the input variables, and the High, Medium, and Low labels of the average amount of ATM transactions were set as the target variable. Table 5 shows the classification accuracy of the algorithm, and the extracted rules are displayed in table 6.

Table 5. Accuracy of decision tree of ATM transactions regardless of month and year.

Partition	Training		Test	
Correctly classified samples	29825	86.36%	12798	85.96%
Incorrectly classified samples	4712	13.64%	2090	14.04%
Total	34537		14888	

Table 6. Extracted rules of decision tree of ATM transactions regardless of month and year.

Row	Rule	Low Category		Medium Category		High Category	
		Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency
R_1	**Bank:** (Ayandeh, Middle East, Pasargad)	31.16%	834	67.9%	1817	0.93%	25
R_2	**Num. of ATM** ≤ 72 **Province**: (East Azerbaijan, Khuzestan, Bushehr, Isfahan, Tehran, Free zones) **Bank:** (Tejarat, Refah, Sarmaye, Saderat, Melat, Meli, Industry and Mine)	73.63%	852	23.76%	275	2.59%	30
R_3	**Bank:** (Qavamin, Keshavarzi, Maskan, Tose'e Ta'avon, Export Development, Parsian, Post Bank, Tourism, Sepah, Sina, Saman, Hekmat Iranian, Ansar, Iran Zamin, Eghtesad Novin)	95.66%	16491	4.25%	733	0.08%	14
R_4	**Bank:** (Tejarat, Dey, Refah, Sarmayeh, Saderat, Melat, Meli, Industry and Mine, Karafarin, Qarzol-Hasaneh Mehr Iran) **Num. of ATM** > 72 **Province:** (Gilan, Golestan, Markazi, Mazandaran, Lorestan, Kerman, East Azerbaijan, Khorasan Razavi, North Khorasan, Chaharmahal va Bakhtiari, Zanjan, Semnan, Qazvin, Khuzestan)	72.81%	1350	26.8%	497	0.37%	7

4.4.2. Results of decision tree of ATM transactions with respect to month and year

In the second implementation of the CART algorithm, the variables province, bank, month, year, ATM number, and Pin Pad number were considered as the input variables, and the High, Medium, and Low labels of the average amount of ATM transactions were set as the target variable. Table 7 shows the classification accuracy of the algorithm, and the extracted rules are indicated in table 8.

Table 7. Accuracy of decision tree of ATM transactions with respect to month and year.

Partition	Training		Test	
Correctly classified samples	31027	89.84%	13389	89.93%
Incorrectly classified samples	3510	10.16%	1499	10.07%
Total	34537		14888	

Table 8. Extracted rules of decision tree of ATM transactions with respect to month and year.

Row	Rule	Low Category		Medium Category		High Category	
		Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency
R_5	**Bank:** (Ayandeh, Middle East, Pasargad) **Year:** (1386, 1387, 1388) **Num. of ATM** ≤ 72	97.05%	462	2.94%	14	0%	0
R_6	**Bank:** (Ayandeh, Middle East, Pasargad) **Year:** (1389, 1390, 1391, 1392, 1393, 1394)	16.90%	372	81,95%	1803	1.13%	25
R_7	**Year:** (1391, 1392, 1393, 1394) **Province:** (Yazd, Qom, Hormozgan, Kermanshah, Free zones, Kurdistan, Fars, East Azerbaijan, West Azerbaijan, Khorasan Razavi, Khuzestan, Tehran, Bushehr, Isfahan) **Bank:** (Saderat, Meli, Melat)	24.94%	310	67.98%	845	7.08%	88
R_8	**Year:** (1391, 1392, 1393, 1394) **Bank:** (Tejarat, Dey, Refah, Sarmayeh, Industry and Mine, Karafarin, Qarzol-Hasaneh Mehr Iran) **Province:** (Free zones, Tehran, Isfahan)	25.49%	116	74.28%	338	0.22%	1
R_9	**Year:** (1391, 1392, 1393, 1394) **Bank:** (Tejarat, Dey, Refah, Sarmayeh, Industry and Mine, Karafarin, Qarzol-Hasaneh Mehr Iran) **Province:** (Yazd, Qom, Hormozgan, Kermanshah, Kurdistan, Fars, East Azerbaijan, West Azerbaijan, Khorasan Razavi, Khuzestan, Bushehr) **Month:** (Ordibehesht, Khordad, Farvardin)	84.39%	357	15.63%	66	0%	0
R_{10}	**Year:** (1391, 1392, 1393, 1394) **Bank:** (Tejarat, Dey, Refah, Sarmayeh, Industry and Mine, Karafarin, Qarzol-Hasaneh Mehr Iran) **Month:** (Tir, Mordad, Shahrivar, Mehr, Aban, Azar, Dey, Bahman, Esfand) **Province:** (Yazd, Qom, Fars, East Azerbaijan, west Azerbaijan, Khuzestan, Bushehr) **Num. of Pin Pad** ≤ 4	62.74%	261	37.26%	155	0%	0
R_{11}	**Year:** (1391, 1392, 1393, 1394) **Bank:** (Tejarat, Dey, Refah, Sarmayeh, Industry and Mine, Karafarin, Qarzol-Hasaneh Mehr Iran) **Month:** (Tir, Mordad, Shahrivar, Mehr, Aban, Azar, Dey, Bahman, Esfand) **Province:** (Hormozgan, Kermanshah, Kurdistan, Khorasan Razavi)	72.96%	386	27.03%	143	0%	0
R_{12}	**Bank:** (Tejarat, Dey, Refah, Sarmayeh, Saderat, Melat, Meli, Industry and Mine, Karafarin, Qarzol-Hasaneh Mehr Iran) **Year:** (1386, 1387, 1388, 1389, 1390)	96.02%	6526	3.91%	266	0.05%	4

Ayandeh Bank, Middle East Bank, and Pasargad Bank, which have progressed in average amount of ATM transactions, can be observed by comparing the rules R_5 and R_6 resulting from the decision tree. Decision tree branches of rules 5 and 6 are indicated in figure 6. 2.94% of the records of these three banks in the years 1386 to 1388 lied placed in the Medium ATM transactions category, while 81.95% of them was placed in Medium category between 1389 and 1390, and such increase suggests the progress of these banks. The performance of banks was evaluated separately. Ayandeh Bank statistics have been recorded since 1391. During the years 1391 to 1394, 71.2% of its total records were placed in the Medium category. The performance of the bank during a period of 4 years is indicated in table 9 and figure 7. With regard to the following table in 1393, the bank has the best performance in ATM transactions.

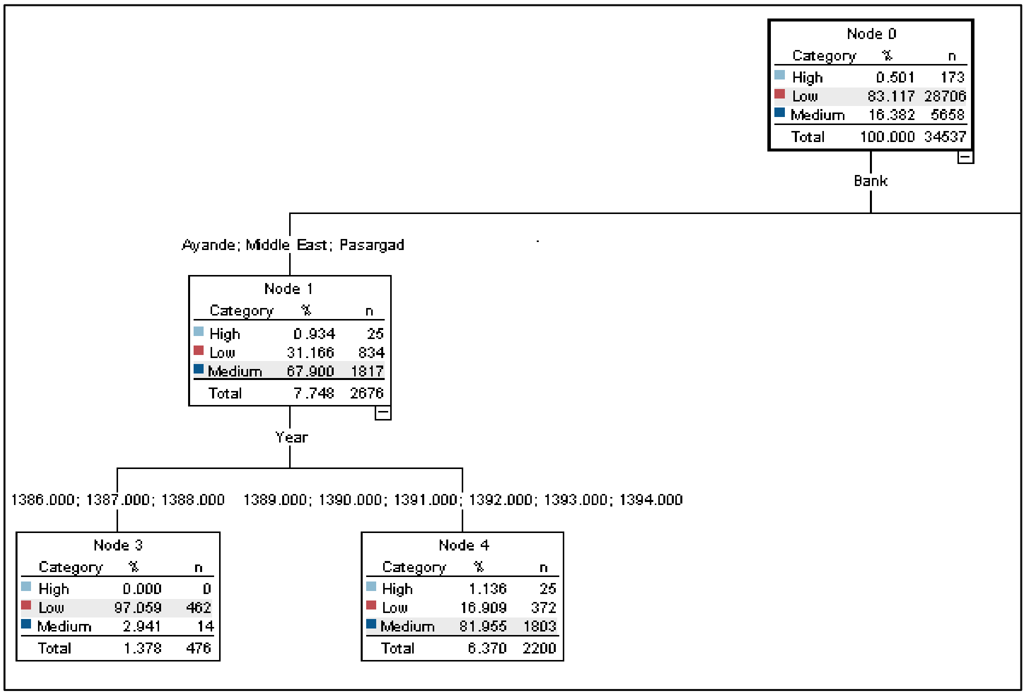

Figure 6. Branch of rules 5 and 6.

Table 9. Ayandeh Bank performance.

Year	Records of Medium class
1391	65.4%
1392	65.8%
1393	78.5%
1394	68.03%

The Middle East Bank statistics in Tehran, East Azerbaijan, Isfahan, Fars, and Khorasan Razavi provinces has been registered since 1393 up to the present time.

In the last 2 years, 44.7% of the total ATM transaction records in the bank were in the Medium

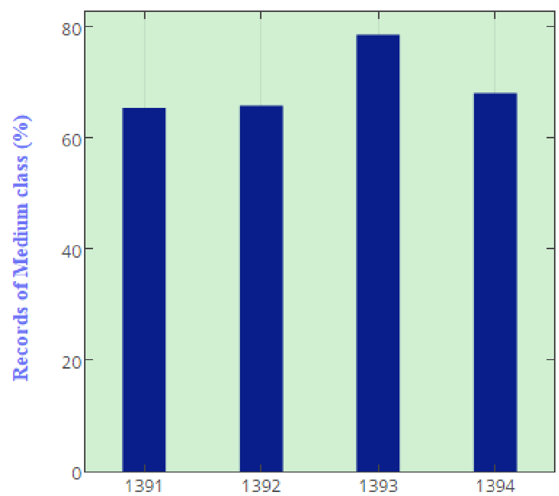

Figure 7. Ayandeh Bank performance.

class. Most of the records are related to the Tehran province.

Pasargad Bank, during the period of this research work, has had a growing trend. 3.33% of the bank records in the years 1386 to 1388 were placed in the Medium ATM transactions class, while that statistic was 89.1% for the years 1389 to 1390. The performance of this bank is shown in table 10 and figure 8. According to the table below, the percent of the high volume ATM transaction of the bank has an increasing trend every year. Comparing the R_8 and R_9 rules indicates that there is a significant difference between the average amount of ATM transaction of Tejarat Bank, Dey Bank, Refah Bank, Sarmayeh Bank, Industry and Mine Bank, Karafarin Bank, and Qarzol-Hasaneh Mehr Iran Bank of Tehran province in the years 1391 to 1394 and between the first 3 months of the year and other months of the year. According to the R_8 rule, 74.28% of these bank records are placed in the Medium category; and also according to the rule R_9, 15.63% of these bank records is placed in the Medium category in three months of the year. The 2 rules include Tehran and some other provinces. The Tehran province records were examined separately to determine whether these rules were applicable for the Tehran province alone or not. According to the survey, 62.5% of the Tejarat Bank, Dey Bank, Refah Bank, Sarmayeh Bank, Bank of Industry and Mine, Karafarin Bank, and Mehr Iran Bank records in the first 3 months, and 76% of the whole year records in the Tehran province in the years 1391 to 1394 are placed in the Medium class. It is observed that the difference between rules R_8 and R_9 is not true for the Tehran province.

Table 10. Pasargad Bank performance.

Year	Records of Medium class
1389	60%
1390	76.6%
1391	89%
1392	94.3 %
1393	97.3%
1394	97.3%

Comparing the decision tree R_9 and R_{10} rules suggests that there is a significant difference between the average amount of ATM transactions of Tejarat Bank, Dey Bank, Refah Bank, Sarmayeh Bank, Bank of Industry and Mine, Karafarin Bank and Qarzol-Hasaneh Mehr Iran Bank in Yazd, Qom, Fars, Khuzestan, Bushehr, West Azerbaijan, and East Azerbaijan provinces within the first 3 months of the year comparing the other months in the period of 1391 to 1394. The records of these provinces and banks in the first 3 months of the year were compared with the rest of the year separately. According to table 11 as well as the decision tree $\mathbf{R_9}$ And $\mathbf{R_{10}}$ rules, the amount of ATM transactions of Tejarat Bank, Dey Bank, Refah Bank, Sarmayeh Bank, Bank of Industry and Mine, Karafarin Bank, and Qarzol-Hasaneh Mehr Iran Bank in the Yazd, Qom, Fars, Khuzestan, Bushehr, West Azerbaijan and East Azerbaijan provinces in years 1391 to 1394 have increased significantly in summer, autumn, and winter than spring. These results are shown in figure 9.

4.4.3. Results of decision tree of Pin Pad transactions regardless of month and year

In the third implementation of CART algorithm, the variables province, bank, ATM number, and Pin Pad number were considered as the input variables, and the High, Medium, and Low labels of average amount ATM transactions were set as the target variable. Table 12 shows the classification accuracy of the algorithm, and the extracted rules are indicated in table 13.

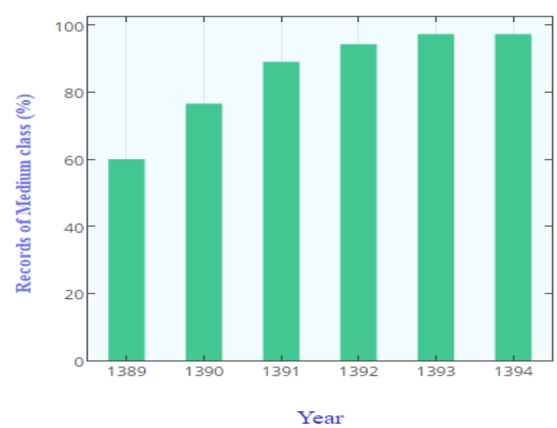

Figure 8. Pasargad Bank performance.

Table 11. Comparison between spring and rest of year's transactions.

Province	Records of Medium class (%)	
	3 First months of year	Other months of year
East Azerbaijan	14.28%	51.61%
West Azerbaijan	26.7%	50.53%
Qom	12. 5%	47.31%
Fars	10.71%	43.54%
Khuzestan	37.5%	52.15%
Yazd	28.57%	47.31%

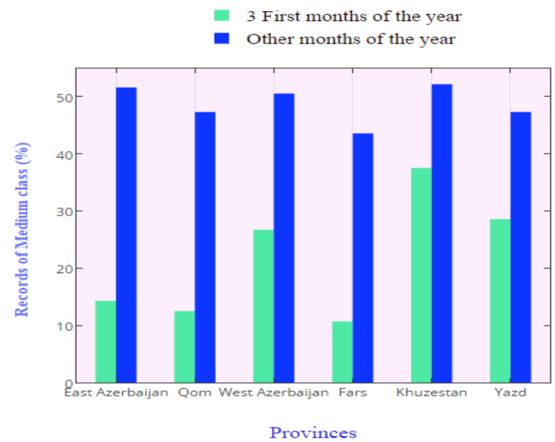

Figure 9. Comparison between spring and rest of year's transactions

According to the R_{17} and R_{18} rules, increasing the number of ATMs has increased the average amount of Pin Pad transactions. Decision tree branches of rules 17 and 18 are indicated in figure 10. In order to obtain more accurate result records related to the Tat Bank, Tejarat Bank, Melat Bank, Sarmayeh Bank, Tourism Bank and Parsian Bank in Gilan, Isfahan, Khorasan Razavi, Khuzestan, Fars, Kerman, Lorestan, Mazandaran, and East Azerbaijan provinces were investigated. From the listed banks, only Tejarat Bank and Melat Bank in these provinces had more than 125 ATMs. Tejarat Bank and Melat Bank records in Isfahan, Khorasan Razavi, Khuzestan and Fars provinces after 1391 had more than 125 ATMs, and were assigned to the Medium category. Records relating to the provinces of Lorestan and Gilan did not have more than 125 ATMs. These 2 bank records in the Kerman province after 1393 had more than 125 ATMs and assigned to the Medium category. Melat Bank records from 1391 and Tejarat Bank from 1392 in the Mazandaran province had more than 125 ATMs, and was assigned to the Medium category. Melat Bank records from 1393 and Tejarat Bank from 1392 in the East Azerbaijan province had more than 125 ATMs, and was assigned to the Medium category.

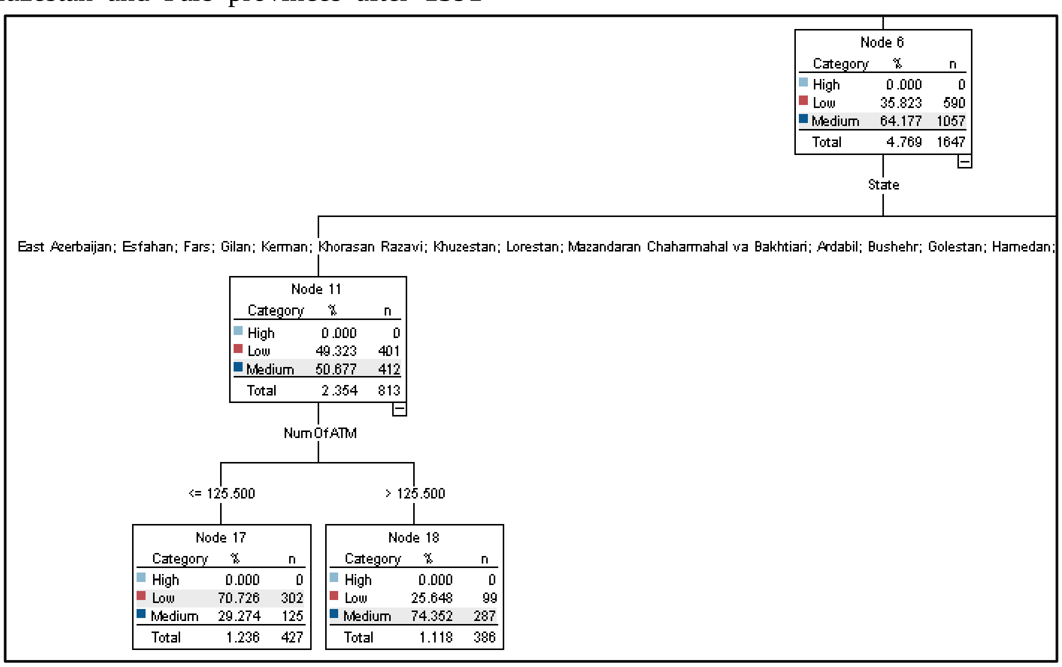

Figure 10. Branches of Rules 16 and 17.

Table 12. Accuracy of decision tree of Pin Pad transactions regardless of month and year.

Partition	Training		Test	
Correctly classified samples	29794	86.27%	12702	85.32%
Incorrectly classified samples	4743	13.73%	1499	14.68%
Total	34537		14888	

Table 13. Extracted rules of decision tree of Pin Pad transactions regardless of month and year.

Row	Rule	Low Category		Medium Category		High Category	
		Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency
R_{13}	**Bank:** (Keshavarzi, Tose'e Ta'avon, Post Bank, Hekmat Iranian, Ansar, Iran Zamin, Refah, Karafarin)	97.05%	9793	2.94%	297	0%	0
R_{14}	**Bank:** (Ayandeh, Tejarat, Sarmayeh, Qavamin, Melat, Parsian, Tourism) **Num. of ATM** > 68 **Province:** (Yazd, Hormozgan, Free zones, Kermanshah, Kurdistan, Qom, Tehran, Bushehr, west Azerbaijan, Ardabil, Golestan, Hamedan, Markazi, North Khorasan, Zanjan, Semnan, Qazvin, Sistan and Baluchestan, Chaharmahal va Bakhtiari **)**	22.66%	189	77.33%	645	0%	0
R_{15}	**Num. of ATM** ≤ 68 **Bank:** (Qavamin)	18.27%	70	81.72%	373	0%	0
R_{16}	**Bank:** (Ayandeh, Tejarat, Sarmayeh, Melat, Parsian, Tourism) **Province:** (Gilan, Kerman, Lorestan, Fars, East Azerbaijan, Khorasan Razavi, Khuzestan, Isfahan, Mazandaran) **Num. of ATM** > 125	25.64%	99	74.35%	287	0%	0
R_{17}	**Bank:** (Ayandeh, Tejarat, Sarmayeh, Melat, Parsian, Tourism) **Province:** (Gilan, Kerman, Lorestan, Fars, East Azerbaijan, Khorasan Razavi, Khuzestan, Isfahan, Mazandaran) **Num. of ATM** ≤ 125	70.37%	302	29.27%	125	0%	0

4.1.4. Results of decision tree of Pin Pad transactions with respect to month and year

In the fourth implementation of the CART algorithm, the variables province, bank, month, year, ATM number and Pin Pad number were considered as the input variables, and the High, Medium, and Low labels of average amount ATM transactions were set as the target variable. Table 14 shows the classification accuracy of the algorithm, and the extracted rules are indicated in table15.

Table 14. Accuracy of decision tree of Pin Pad transactions with respect to month and year.

Partition	Training		Test	
Correctly classified samples	30487	88.27%	12798	87.27%
Incorrectly classified samples	4050	11.73%	1895	12.73%
Total	34537		14888	

Table 15. Extracted rules of decision tree of Pin Pad transactions with respect to month and year.

Row	Rule	Low Category		Medium Category		High Category	
		Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency
R_{18}	**Year:** (1386, 1387, 1388, 1389, 1390) **Bank:** (Melat, Tejarat, Sarmayeh)	93.04%	2319	6.53%	162	0%	0
R_{19}	**Year:** (1391, 1392, 1393, 1394) **Bank:** (Melat, Qavamin)	11.08%	152	88.91%	1219	0%	0
R_{20}	**Year:** (1391, 1392, 1393, 1394) **Bank:** (Parsian)	90.24%	768	9.27%	83	0%	0
R_{21}	**Year:** (1391, 1392, 1393, 1394) **Province :(**Chahar Mahal and Bakhtiari, Hormozgan, Markazi, Hamedan, Kermanshah, Qom, Kurdistan, Fars, Bushehr, Ardabil, Isfahan Tehran, free zones, West Azerbaijan and East Azerbaijan) **Bank:** (Sarmayeh ,Tejarat, tourism)	25.59%	270	74.40%	785	0%	0
R_{22}	**Year:** (1391, 1392, 1393, 1394) **Province:** (Southern Khorasan, Yazd, Kerman, North Khorasan, Khorasan Razavi, Khuzestan, Zanjan, Semnan, Qazvin, Sistan and Baluchestan, Lorestan, Gilan, Golestan)	69.78 %	984	30.21%	426	0%	0

According to the R_{18} Rule, 6.53% of Melat Bank, Tejarat Bank, and Sarmayeh Bank records in the years 1386 to 1390 are in the Medium category; and regarding the R_{19} rule, 88.91% of Melat Bank and Qavamin Bank in the years 1391 to 1394 are in the Medium category. Melat Bank is repeated in these 2 rules. Decision tree branches of rules 18 and 21 are indicated in figure 11. Regarding the significant difference in the results of these 2 rules, Melat Bank performance in these two periods was examined separately. As shown in table 16 and figure 12, Melat Bank in the years 1391 to 1394 in the Pin Pad payments during the period 1386 to 1390 had a significant progress. This is due to a great increase in the amount of Pin Pad transactions conducted by the bank from 1391 than before.

Comparing the $\mathbf{R_{18}}$ and $\mathbf{R_{21}}$ decision tree rules suggests the growing trend of Sarmayeh Bank and Tejarat Bank Pin Pad payments in the Chahar Mahal and Bakhtiari, Hormozgan, Markazi, Hamedan, Kermanshah, Qom, Kurdistan, Fars, Bushehr, Ardabil, Isfahan Tehran, Free Zones, West Azerbaijan, and East Azerbaijan provinces in the period 1391 to 1394 compared to 1386 to 1390. In order to obtain more accurate results, the Sarmayeh Bank and Tejarat Bank payment trends in the mentioned provinces in these two periods were investigated separately.

Sarmayeh Bank did not have any records with a Medium tag from 1386 to 1390. However, 77.23% of the bank records were in the Medium category from 1391 to 1394. The high volume of the bank Pin Pad transactions, due to the low number of transactions, was conducted by the bank.

Tejarat Bank records indicated that 19.5% of the bank records in the years 1386 to 1390, mostly belonging to the provinces of Tehran and Free Zones were in the Medium category. However, 74.95% of the bank records were in the Medium category from 1391 to 1394. Considering table 16, Tejarat Bank in the field of Pin Pad payment in 1393 had the best performance.

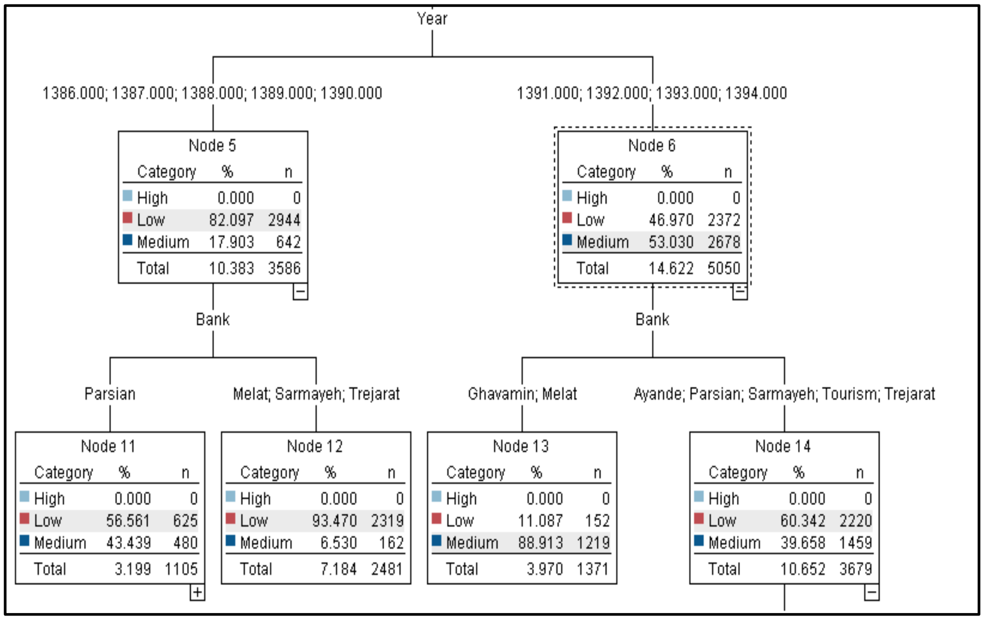

Figure 11. Branches of rule 18 & rule 19.

Table 16. Melat Bank performance.

Year	Records of Medium class (%)
1386 to 1390	2.5%
1391	83%
1392	93.2%
1393	97.7%
1394	95.1%

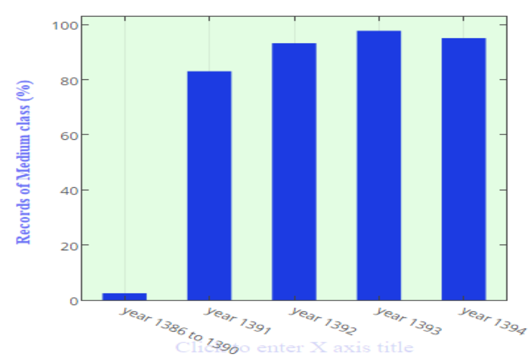

Figure 12. Melat Bank performance.

5. Discussion

DM is an important tool for effective utilization of data, and is one of the most important technologies for exploitation of Big Data. Banks require understanding and compliance with the recent changes in the behavior of customers to provide better customer service and survive in the dynamic markets. The focus of this work was on identifying the factors affecting the electronic payment transactions in Iran using the DM techniques. Therefore, a dataset including 49425 ATM and Pin Pad transactions, separated based on the bank and province relating to the years 1386 to 1394 and from all over the country, was received from the Statistical Center of the Central Bank of Iran. In order to extract the rules related to the ATM and Pin Pad transactions, the CART decision tree algorithm was used. In implementation of the classification algorithm, two target variables are required. Due to the continuous numeric values, the variables "Average Amount of ATM Transactions" and "Average Amount of Pin Pad Transactions" were awarded 2 labels to any records by the K-Means algorithm. At first, the K-Means algorithm was run with 3 clusters on the variables of "Average Amount of ATM Transactions". The Average Amount of ATM Transaction field values were divided into the three clusters Low, Medium, and High. The labels were allocated to each record according to the average ATM transaction amount. In the next implementation of the K-Means algorithm, Average Amount of Pin Pad Transaction field values were divided into the three clusters Low, Medium, and High. In the next step, the CART algorithm was run to dedicate the hidden patterns of E-payment instruments. In this research work, the DM techniques were selected to analyze the transactions because the order and priority of variables could not be determined in descriptive statistics methods; however, there is the possibility of analytic hierarchy in data mining techniques using the decision tree. The formation of tree branches and priority of variables are determined using the Gini index in the decision tree, and this advantage will cause a higher accuracy and a better detection of results. Another reason for using DM techniques instead of the descriptive statistics methods is the ease of achieving the desired results. To achieve the rules obtained by the statistics methods, different filters on different variables in various time periods for each bank and province should be applied, and the results obtained should be compared. This task is too time-consuming and might not get all the results. The desired results were achieved easily and quickly using the DM techniques. For example, one of the results of this algorithm determined that the ATM transaction volume of Tejarat Bank, Dey Bank, Refah Bank, Sarmayeh Bank, Industry and Mine Bank, Karafarin Bank, and Qarzol-Hasaneh Mehr Iran Bank in the Yazd, Qom, Fars, Khuzestan, Bushehr, West Azerbaijan, and East Azerbaijan provinces in the years 1391 to 1394 in summer, autumn, and winter than in spring has increased significantly. Achieving this rule by the statistical methods requires several filtering on different variables, which is very time-consuming.

Also increasing the number of Melat Bank and Tejarat bank ATM terminals has increased the average amount of Pin Pad transactions of these banks in the Isfahan, Khorasan Razavi, Khuzestan, Fars, Kerman, and East Azerbaijan provinces. Reviewing the decision tree rules specifies the growing trend of Ayandeh Bank, Middle East Bank, and Pasargad Bank in the volume of ATM transactions and Sarmayeh Bank and Tejarat Bank in the volume of Pin Pad transactions. The order and priority of variable determination was so important in extracting this rule, and this order was determined by the Gini index in decision tree, and this rule was obtained.

6. Conclusion and suggestions

In this research work, a process was proposed for specifying the factors affecting the number and amount of transactions conducted with the E-payment instruments. The purpose of this research work was not to propose a new algorithm but to focus on the execution and the understanding of the model. A suitable design of the systematical way to build a model could be helpful to execute the rules. To do this, the data mining techniques were chosen. According to the research purpose and subject, data mining techniques, in comparison with the descriptive statistics methods, were more efficient, and the desired results were achieved more easily and accurately.

Therefore, a dataset including 49425 ATM and Pin Pad transactions, separated based on the bank and province relating to the years 1386 to 1394 and all over the country, was received from the Statistical Center of the Central Bank of Iran. The K-Means algorithm was used for assigning labels to each record and determining the target field categories. In order to extract the rules relating to the ATM and Pin Pad transactions, the CART decision tree algorithm was used.

The results of this work can be used to survey different bank customer behaviors by bank performance classification and developed marketing strategies to attract customers. Also the results of this research work will increase the ability of a bank manager to provide better E-banking services and E-banking future policy adjustments in the interests of customers and bank based on the electronic payment analysis and discovered patterns.

This work is limited to the ATM and Pin Pad transaction information. Future analytical works could be done using other electronic payment tools such as telephone banking, internet banking, mobile banking, and POS for a comprehensive review of the bank performance in electronic payment.

Moreover, considering the demographic characteristics of each province like population, educational level, mean age, and average salary will help to better analyze the cause of desire or lack of desire to use the tools and methods of electronic payment in different provinces.

References

[1] Shu, W., & Strassmann, P. A. (2005). Does information technology provide bankswith profit? Information & Management, vol. 42, no. 5, pp. 781–787.

[2] Karjaluoto, H., Mattila, M., & Pento, T. (2002). Factors underlying attitude formation toward online banking in Finland. International Journal of Bank Marketing, vol. 20, no. 2, pp. 261–272.

[3] Chilukuri, S. S., & Madhav, V. V. (2015). ICT Evolution in Banking–A Strategy of Financial Inclusion. Advances in Economics and Business Management (AEBM), vol. 2, no. 12, pp. 1142-1147.

[4] Garau, C. (2002) Online banking in transition economies: The implementation and development of online banking systems in Romania. International Journal of Bank Marketing, vol. 20, no. 6, pp. 285-296.

[5] Sumathi, S., & Sivanandam, S. N. (2006). Introduction to data mining principles. Introduction to Data Mining and its Applications, pp. 1-20.

[6] Hashemian, M., Isaee, M, T., Mikaeli, F., & Tabatabee, M. (2012). Factors affecting the electronic banking instruments acceptance by the customer (Saman Bank survey). Information Technology Management, vol. 4, no. 11, pp. 155-174.

[7] Divandari, A., Abedi, E., & Naserzadeh, M, R. (2013). Presenting a conceptual model to explain the key factors affecting the quality of the internet banking service provider systems (Melat Bank survey). Information Technology Management, vol. 5, no. 1, pp. 19-36.

[8] Rasuli, H., & Manian, A. (2012). Fuzzy inference system design for choosing electronic banking services systems (Sepah Bank survey). Information Technology Management, vol. 4, no. 12, pp. 41-64.

[9] Edward, F.R, & Mishkin F.S.(1995).The decline of traditional banking: implication for financial stability and regulatory policy.Federal Reserve Bank of New York policy Review, pp. 27-45.

[10] Han, J., & Kamber, M. (2006). Data Mining:Concepts and Techniques. United States of America: Morgan Kaufmann publishers.

[11] Gharipour, A., Sameti, M., & Yousefian, A. (2010). A Comparative Approximate Economic Behavior Analysis of Support Vector Machines and Neural Networks Models. Iranian Economic Review, vol. 15, no. 26, pp. 17-40.

[12] Nie, G., Rowe, W., Zhang, L., Tian, Y., & Shi, Y. (2011). Credit card churn forecasting by logistic regression and decision tree. Expert Systems with Applications, vol. 38, no. 12, pp. 15273-15285.

[13] Hu, Y. C., & Liao, P. C. (2011). Finding critical criteria of evaluating electronic service quality of Internet banking using fuzzy multiple-criteria decision making. Applied Soft Computing, vol. 11, no. 4, pp. 3764-3770.

[14] Adeoti, J. O. (2011). Automated teller machine (ATM) frauds in Nigeria: The way out. Journal of Social Sciences, vol. 27, no. 1, pp. 53-58.

[15] Khashei, M., Rezvan, M. T., Hamadani, A. Z., & Bijari, M. (2013). A bi-level neural-based fuzzy classification approach for credit scoring problems. Complexity, vol. 18, no. 6, pp. 46–57.

[16] Liébana-Cabanillas, F., Nogueras, R., Herrera, L. J., & Guillén, A. (2013). Analyzing user trust in electronic banking using data mining methods. Expert Systems with Applications, vol. 40, no. 14, pp. 5439-5447.

[17] Khobzi, H., Akhondzadeh-Noughabi, E., & Minaei-Bidgoli, B. (2014). A new application of rfm clustering for guild segmentation to mine the pattern of using banks'e-payment services. Journal of Global Marketing, vol. 27, no. 3, pp. 178-190.

[18] Farquad, M. A. H., Ravi, V., & Raju, S. B. (2014). Churn prediction using comprehensible support vector machine: An analytical CRM application. Applied Soft Computing, vol. 19, pp. 31-40. Elsevier B.V. Retrieved from http://dx.doi.org/10.1016/j.asoc.2014.01.031.

[19] Caliş, A., Boyaci, A., & Baynal, K. (2015). Data mining application in banking sector with clustering and classification methods. In Industrial Engineering and Operations Management (IEOM), 2015 International Conference on (pp. 1-8). IEEE.

[20] Farokhi, S., Teimourpour, B., Shekarriz, F., & Masoudi, M. (2016). A new application of clustering for segmentation of banks'e-payment services based on profitability. Uncertain Supply Chain Management, vol. 4, no. 1, pp. 55-60.

[21] Fadaei Noghani, F., & Moattar, M. (2017). Ensemble classification and extended feature selection for credit card fraud detection. Journal of AI and Data Mining, vol. 5, no. 2, pp. 235-243.

[22] Kudyba, S. & Lawrence, K. (2008). Enhancing information management through data mining analytics to increase product sales in an e-commerce platform. Int. J. Electronic Marketing and Retailing, vol. 2, no. 2, pp .97–104.

[23] Kaur, M., & Kang, S. (2016). Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining. Procedia Computer Science, vol. 85, no. Cms, pp. 78-85. http://dx.doi.org/10.1016/j.procs.2016.05.180

[24] Karamizadeh, F., & Zolfagharifar, S. A. (2016). Using the clustering algorithms and rule-based of data mining to identify affecting factors in the profit and loss of third party insurance, insurance company auto. Indian Journal of science and Technology, vol. 9, no. 7, pp. 1-9.

[25] Insani, R., & Soemitro, H. L. (2016, May). Data mining for marketing in telecommunication industry. In Region 10 Symposium (TENSYMP), 2016 IEEE (pp. 179-183). IEEE.

[26] Plessas-Leonidis, S., Leopoulos, V., & Kirytopoulos, K. (2010, February). Revealing sales trends through data mining. In Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on vol. 1, pp. 682-687. IEEE.