Journal of Industrial and Systems Engineering
Vol. 12, No. 2, pp. 61-94
Spring (April) 2019

Investigation of effective factors in expanding electronic payment in Iran using data mining techniques

Fateme Moslehi[1], Abdorrahman Haeri[1*], Mohammad Reza Gholamian[1]

[1] *School of Industrial Engineering, Iran University of Science & Technology, Tehran, Iran*

moslehi_fateme@ind.iust.ac.ir, ahaeri@iust.ac.ir, Gholamian@iust.ac.ir

Abstract

E-banking has grown dramatically with the development of ICT industry and banks offer their services to customers from different channels. Nowadays, considering the great economic benefits of electronic banking systems, the need to pay attention to the expansion of electronic banking is increasingly felt in terms of reducing costs and increasing the bank's profitability. The purpose of this study is to identify the factors that encourage customers to accept e-banking across the country using the statistics and information retrieved from the Central Bank and the data mining techniques. For this purpose, initially, the K-Means clustering algorithm was appliyed and the provinces of Iran is separated into 3 clusters. In addition, the transactions related to each year were clustered separately, and the formed clusters were compared with each other. In the next step the hidden patterns of the E-payment instrument transactions were detected using the CART algorithm. According to the results obtained from decision tree rules, indices of social-economic and Information and Communication Technology development and business boom were the most effective factors in increasing the usage of electronic payment methods.

Keywords: Machine learning, data mining, electronic payment instruments, classification, clustring.

1- Introduction

Today, the Internet has turned into a fundamental part of economic and financial actions (Mohammad et al, 2014). According to this development, the world payment and banking system changed from coin and papers into electronic procedures, which are faster and reliable than other payment way (Premchand & Choudhry, 2015). Banking is an information-intensive industry in which information technology (IT) is increasingly important (Shih & Fang, 2004). In developing countries the internet revolution has been developing by delays due to the lack of suitable use of infrastructure budgets and their maintenance (Mathew et al., 2014). Today, with advances in ICT industry, face to face communications between banks and customers have been replaced by new e-banking tools and methods. The spread of e-banking services is essential, but this development requires train banking customers (Iyengar& Belvalkar, 2010). The electronic payment system which can be defined as an important platform for world payment system to purchase the goods/services in online way, so development and switch from the conventional cash-based transactions into an online way have great influence on banking services and global business (Roy &

*Corresponding author
ISSN: 1735-8272,

Sinha, 2014). In recent years use of distance services such as Phone banking and Mobile banking that doesn't need any physical presence of customer near bank or payment tools, are more favored by many banks. Various services via several technologies are provided by Mobile and Phone-based payment systems (Mathew et al., 2012). Development of such services will reduce the cost of banks in maintenance due to minimize face to face payment instruments and it has been welcomed by customers due to ease of use,. But since the concept of money in banking services is an important concept in people's lives, expanding the use of these services by customers requires banks to create confidence among costumers (Namazi, 2015).

Unarguably, the financial industry is evolving with an increasing magnitude and fast pace, according to the discernable changes in consumers' choices and expectations, which are stemming from the emerging technology and significant availability of different products and services. As a result, the banking industry has also become highly competitive due to all the threats and disruptions caused by not only direct competitors, but also all the new and innovative entrants (Shirazi and Mohammadi, 2018). With growing use of e-banking services among customers, especially Internet banking services, the number of transactions has increased considerably and banks are also facing with huge amounts of data in their databases. This can be considered as a valuable resource for detection of knowledge and patterns of customers' behaviors and using data mining can be very helpful in this regard (Radfar et al. 2014). The banks are required to observe the behavior of their customers in order to survive and increase their market share in the banking industry. More over in order to increase customers' satisfaction review and concept online customers' behaviors are necessary (Hsieh et al, 2013). The usage of information and communications technology is becoming more and more dependent on socio-economic factors (Cullen, 2003). User features like online skill affect the customers' online behavior (Hargittai & Hinnant, 2008). As Travica (2002) expressed technological, economic, and cultural specificities are important mediating factors in the adoption of e-payment. The Observatory takes place with the aim of creating security for users and obligating banks to provide new services to attract new customers to maintain the market. The banks would monitor and review available tendencies and trends in customers` behavior. As a result of this review, behavior changes and interest in some banking services are predicted and in response to these behavior changes, banks will offer or develop a variety of banking services. Due to the development of ICT infrastructure in Iran and widespread use of them by people, the necessity to extend banking services in the context of ICT is felt more than ever. Data mining is one of the powerful techniques that can be helpful to understand the needs and characteristics of customers. To have the effective presence in the competitive market, the financial industries are turning into dependent on computer and information technologies progressively (Burrell & Folarin, 1997). Several methods have been provided to extract knowledge and convert this knowledge to rules by data science experts, and decision support systems have been developed to infer an outcome according to an input (MacIntyre, 2013).

Many studies had been conducted on e-payment systems to identify affecting factors its use and adoption. The aim of this paper is to survey analyze e-payment transactions trend in Iran's banking system and extracts hidden patterns in available transactional data as a set of rules using data mining techniques. The factors affecting the expansion of e-banking have been investigated in this study.

The present research efforts to answer the following questions:

- What factors affect the customer's desire to use the e-payment instruments?
- Is there a significant relationship between socio-economic indicators of provinces and the volume of e-payment transactions?

Different definitions are provided for e-payment system from scholars. For instance, Kalakota and Whinston (1997) defined a money transfer that takes place online from a payer to payee as an e-payment system. E-payment is also defined as any exchange of funds between the seller and the buyer via an electronic communication channel that customers can manage their bank accounts (2013). Peter and Babatunde (2012) expressed that e-payment is a system that transfers money via the Internet. Another definition suggests that funds transfer for products and services procured through electronic means is e-payment system (2012).

A lot of e-payment researches have been done up to now, which can be summarized into: Historical data of e-banking usages were analyzed by Niyagas et al. (2006) using SOM and K-means algorithms. The customers were divided into 8 clusters according to different features like access time. Logistic regression and decision tree algorithms were selected to construct the predictive model and identify the internet banking customers' behavior in the Mansingh et al. (2010) research. The influence of e-banking in Iran was investigated by Salehi and Alipour (2010). They figured out the Iranian are interested to use e-banking services. Also in their opinion banks should provide high quality e-services to customers for introducing e-banking. In the paper of the Auto (2010), the role of the e-banking in Nigeria's economy was surveyed by using Meyar-Olkin (KMO) approach and Barlett's Test of Sphericity. The critical and influential infrastructure such as power and telecommunication were introduced and noticed that should be supplied to develop e-banking in Nigeria. Özkan et al. (2010) analyzed affecting factors on acceptance of e-banking among customers. They expressed that, from customers' perspective security, advantage and web assurance seals were required factors and perceived risk, trust and usability were relatively adequate factors to acceptance the e-payment system. Hu and Liao (2011) applied fuzzy MCDM as a feed-forward neural network and a genetic-algorithm-based method to analyzed e-service quality in Internet banking sector. These techniques were used to identify the critical criteria for evaluating the quality of services. Benefits, self-efficacy, and ease of use were mentioned as effective measures to expansion of using e-payment from the Malaysian consumers' perspective. The multiple linear regression method was applied to obtain these results (2013). The several algorithms of Data mining to achieve the effective indicators in user trust in e-banking were determined in the study of the Liebana-Calbinallas et al. (2013). In this paper, the best technique of variable selection has been introduced according to the expert's opinion which was the MGA using Mutual Information. Hsieh et al. (2013) aimed to analyze the differences between the urban and rural online activities and e-payment pattern. The results demonstrated that because people in urban were more familiar with internet the usage of e-banking in the urban was higher than rural. Also social indexes were mentioned as the important indexes to development of mobile commerce in India (2013). Another study by Thakur and Mala (2014) surveyed the influence of perceived risk and security concerns on acceptance of mobile payment service. Khobzi et al. (2014) clustered the data of point of sale terminals to analyze profitability of different guild. Gold segments as the most profitable and carrying as the least profitable guilds were pointed. Sales data have used in this research as a mark. An integrated data mining and customer behavior scoring model proposed by Noori (2015) to analyze mobile banking customers in an Iranian bank. Customers were segmented into 6 clusters according to transaction history and RFM background. Özdağoğlu et al. (2016) introduced privacy and contact responsiveness as the most effective indicators to evaluate e-service quality of internet based banking. The importance of demographic features such as age and gender, channel of transmission funds like ATM and Mobile bank and the length of the customer association on e-banking customer retention were noticed. Data mining techniques were used in the study of Keramati et al. (2016). Determinant factors on adoption of mobile-based payment services were recognized in Upadhyay and Jahanyan (2016) study. Some factors such as ease of use, system quality, connectivity and utility were introduced as important factors to increase the usage of mobile payment technologies. Most profitable customers were identified in Farokhi et al. (2016) study. Point of Sales (POS) of data belongs to one of Iranian private banks was split into four clusters by using K-Means and Kohonen algorithms. A study by Wang et al. (2017) the effect of customization and technology adaptability on the customers' tendency to use the e-banking services explored and the results reported that these factors will increase the usage of e-banking services. Expanding the e-banking has set as a target in Pakistani banking field which its trend was analyzed in Hussain et al. (2017) study. The results demonstrated that trust and security have considerable effect on the usage e-banking. Khan et al. (2017) have reviewed the e-payment trend from the perspective of past developments, present impact, and Future Considerations. They represented the number of online mode of payment's customers have increased significantly and also mobile payment methods has had a huge global growth. The different factors that affect the adoption of e-payment tools were surveyed. They proposed that for attracting customers there must be a balance between usability and security. Fadaei Noghani & Moattar (2017) have proposed a hybrid model for credit card fraud detection by adopting decision tree and feature selection methods. In this model an extended wrapper method selected the most effective features and then an ensemble classification was performed. Barker

(2018) investigated fraud prevention and available e-security measures, the legal consequences on co-liability to negate these potential negative consequences to the benefit of both the financial industry and the customer, and proposes a conceptual theoretical framework for e-banking fraud prevention and co-liability through proactive communication.

With regard to literature reviews and compilation of papers, it can be observed that most studies in the field of e-banking in Iran have analyzed the behavioral and financial conditions of customers of one or two specific banks, and the general trend of conducting e-banking transactions and the factors affecting it have not been studied extensively. In this study, the effective demographic factors affecting the development of e-banking have been investigated using data from all provinces of Iran and data mining techniques.

The rest of the paper is organized as follows: Section 2 introduces research methodology. Empirical study and data analysis are described in Section 3 and are discussed in Section 4, in Section 5 conclusions and implications are considered.

2- Research methodology

2-1- CRISP-DM methodology

There are various methods for the implementation of mining projects. One of the powerful and common methods used for data mining projects management is CRISP-DM method. The CRISP-DM methodology provides a structured approach to planning a data mining project. CRISP-DM methodology is recommended when attempting to perform a data mining project, because it has all the available documentation, detailed phases, tasks and activities, and the development of the first phase, that facilitate the problem understanding and its transforming to a data mining problem. It is a better approach if the knowledge of the problem in terms of business is insufficient (Palacios et al., 2017). This method provides a process model for data mining that is an overview of the lifecycle of each mining project. Life cycle of a data mining project consists of six stages: Business understanding, Data understanding, Data preparation, Modeling, Evaluation and Deployment (Chapman et al, 2000):

- First phase – Business understanding: This initial phase focuses on understanding the project objectives and requirements from a business perspective, and then converting this knowledge into a data mining problem definition, and a preliminary plan designed to achieve the objectives. In the other words data miner needs to be familiar with business environments and processes to start the project. In this phase, a data mining project based on the organization's needs and expected demands from the organization is defined. Getting acquainted with different aspects of business takes place in this phase. To do this, previous publications, studies, and documents in the field of e-payment were surveyed and reviewed.
- Second phase - Data understanding: The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses for hidden information. Gathering data and initial analysis is done in this phase. The dataset used in this study consists of 1085 records, statistics related to e-banking services; different payment instruments including Point of sale (POS), Internet banking and Mobile banking are associated with the year of 2013, 2015 and 2016.
- Third phase- Data preparation: The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modeling tool(s)) from the initial raw data. Data preparation tasks are likely to be performed multiple times, and not in any prescribed order. Tasks include table, record, and attribute selection as well as transformation, data integration and cleaning of data for modeling tools. In this stage, related information and statistics like literacy rate and IDI index were added to the original data set.
- Fourth phase – Modeling: This phase is the main purpose of data mining projects. In this phase, various modeling techniques are selected and applied, and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, stepping back to the data preparation phase is often needed. Data mining algorithms are applied to extract knowledge and hidden patterns. In this study, the clustering and classification algorithm was used to model the problem.

- Fifth phase – Evaluation: Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model, and review the steps executed to construct the model, to be certain it properly achieves the business objectives. In this phase, the knowledge obtained in previous phase is analyzed to determine the usefulness and application more over the accuracy and applicability of the model's results are investigated. For example, in the case of predictive models, the model accuracy is determined using test data.
- Sixth phase – Development: Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model, and review the steps executed to construct the model, to be certain it properly achieves the business objectives. In this stage, the knowledge obtained in previous phase is analyzed to determine the usefulness and application. Using results and extracted knowledge for the problem in business takes places in this phase. For example, in the case of predictive models, the model accuracy is determined using test data.

2-1-1- Deployment

Creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that is useful to the customer. The focus of this phase is the integration of knowledge in business processes to solve business major problems. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data or data mining process.

2-2- Clustering

One of the important methods for data analysis is clustering technique. A clustering algorithm forms a set of groups from input data (Xu & Wunsch, 2005). Due to the lack of a labeled clustering method, it is more difficult to implement than supervised data mining methods (Saxena et al., 2017). The K-means clustering algorithm is one of the most widely used clustering techniques. Running K-means algorithm is required to determine the number of clusters and the initial states (He, 2016). At first, the number of clusters (k) is defined. K centroids are chosen randomly as input samples. In the next step, the cluster with the smallest distance to each record is determined and the record is assigned to that cluster. The centroids are updated after adding any record to the clusters. The above steps continue until there is no change in the structure of clusters (Ahmad et al, 2010).

2-3- Decision tree

One of the powerful and common classification algorithms used for data mining projects is the decision tree (Azar & El-Metwally 2013, Breiman et al, 1984). That has been applied for classification and prediction (Murthy, 1998, Pradhan, 2013). A tree-like graph format is used in decision tree. Decision Tree Induction is defined as the learning phase from training data with class labels (Roy, & Urolagin, 2019). This technique divides the feature set into unique scopes, in an ordinal way (Azar & El-Metwally 2013, Breiman et al, 1984, Clark & Pregibon 1992). Decision tree has many algorithms like classification and regression tree (CART) (Breiman et al, 1984, Pradhan, 2013). In this paper, CART algorithm was chosen. Through decision tree classification algorithms, one of the most widely used and popular one is CART algorithm with diverse applications (Zhu et al, 2018).

2-4- Research framework

The used research framework for this study is illustrated diagrammatically in **Error! Reference source not found.** As is shown in **Error! Reference source not found.**, in this study after gathering the e-payment transactions data, additional information related to provinces like population and the economic participation rate was added to the original dataset. Then the average amount feature of transactions is made by dividing the amount of transactions to the number of transactions. After this phase, the provinces were divided into 3 clusters based on transactions information and other features. After that, CART decision tree algorithm was applied to each cluster. Finally, the transactions related to each year were partitioned into 3 clusters separately, and the formed clusters were compared with each other.

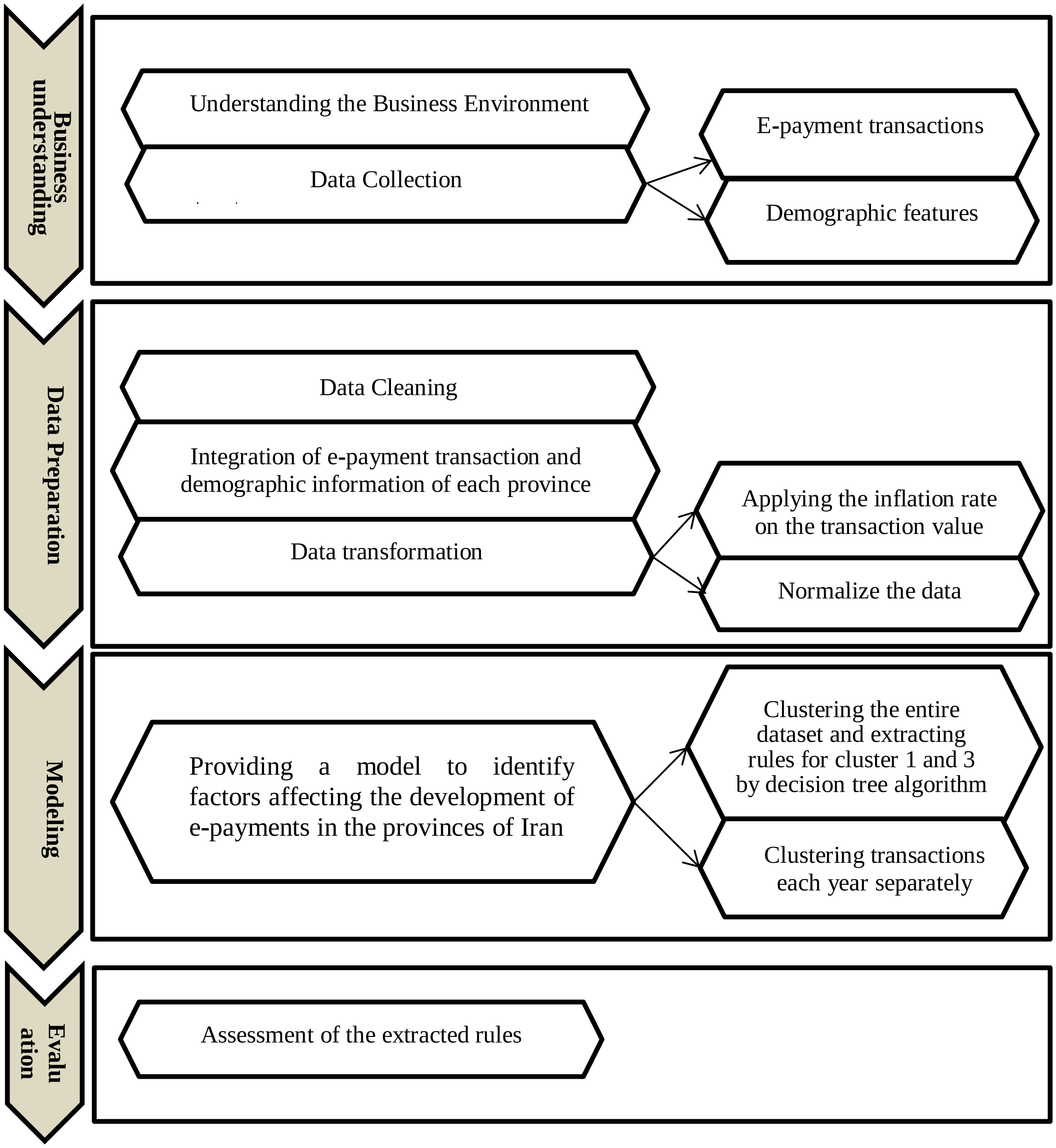


Fig 1. Research flowchart

3- Experiments and results

As CRISP-DM global standard is used to conduct this research, its procedures according to the stages of this standard are described as follows.

3-1- Business understanding

The present research focuses on identification of an organization's purposes and needs. One of the necessities of banking is investigating the present trends in the market. One of the most important areas existing in the banking industry is electronic payment. Following the society's tendency to new tools of electronic payment, banks also found out about the importance of the funds they could absorb through these new trends and the turnover that would be created through not using cash, and tried to increase their activities in this area. Creating new job opportunities, decreasing the costs due to the reduction of in-person visits to different branches, reducing the costs of using banknotes including printing, maintenance, transportation, and their elimination, all are the reasons and needs that have developed this area rapidly. Regarding the importance of payment industry, finding the relations and rules with the help of data mining techniques, which guides bank managers into adopting proper policies seems necessary.

3-2-Data understanding

In this phase published data present on the official website of Central Bank of Iran (www.cbi.ir) was used.The data set including 1085 transactions of point of sale terminals (POS), Internet banking and Mobile banking relating to the years 2013, 2015 and 2016 and all over the country was gathered. The variables of this data set are described in Table 1.

Table 1. Variables definition

Variable Name	Data Type	Variable Name	Data Type
State	Set	Population	Range
Year	Set	Economic participation rate	Range
Month	Set	Unemployment rate	Range
Number of Internet banking transactions	Range	Literacy rate	Range
Amount of Internet banking transactions	Range	Internet penetration rate	Range
Average amount of Internet banking transactions	Range	Number of computer users	Range
Number of Mobile banking transactions	Range	Number of Internet users	Range
Amount of Mobile banking transactions	Range	Number of mobile users	Range
Average amount of Mobile banking transactions	Range	IDI index	Range
Number of POS transactions	Range	Skill IDI index	Range
Amount of POS transactions	Range	Use IDI index	Range
Average amount of POS transactions	Range	Access IDI index	Range
Number of POS device	Range	Development ratio	Range
Deposit	Range	Developmental status	Set
Financial facilities	Range		

3-2-1-Explanation of the variables

- “Average amount of Internet banking transactions”, “Average amount of Mobile banking transactions” and “Average amount of POS transactions” variables have been obtained by dividing amount of transactions by number of transactions.
- *Economic participation rate, Unemployment rate*: it is obtained from the report “an abstract of the results of work force survey design” published on the website of Statistical Center of Iran (www.amar.org.ir) in each season between 2013 and 2016.
- *Population*: the population in 2012 was calculated based on the census results in 2011 and subtracting death rate from and adding birth rate to the population in 2011; the population in 2012 to 2015 was calculated by subtracting death rate from the year before and adding the birth rate of that year to the population. Death and birth statistics of each province were extracted from the website of Civil and Personal Status Registration Authority (www.sabteahval.ir). The population in 2016 has been registered based on the census results in 2016 published on the website of Statistical Center of Iran (www.amar.org.ir).
- *Deposit, Financial facilities*: the total number of loans and deposits which provided by banks and financial institutions have been obtained from the website of Central Bank of Iran (www.cbi.ir) separated based on month and province.
- *Internet penetration rate, Number of computer users, Number of mobile users, Number of Internet users*: statistics of the above mentioned variables related to 2013 and 2015 have been obtained from the results

of "survey design of families and individuals' usage of Information and communication technology" conducted in 2013 and 2015 published on the website of The Ministry of Information and communications technology (www.ict.gov.ir). The statistics related to 2016 have been extracted from www.mis.ito.gov.ir.

- *IDI[2] index, Skill IDI index, Use IDI index, Access IDI index*: it is obtained from the report "Study of the status of Iran's provinces in terms of IDI indicator" published on the website of The Ministry of Information and communications technology (www.ict.gov.ir).
- *Development Ratio*: each province's development rate is obtained from the research "Investigating and comparing selected indices in provinces, emphasizing the census results from 2006 to 2011, and examining the development regarding some combined indices" published on the website of Statistical Research and Training Center (www.srtc.ac.ir).

Table 2. Development status

category	Provinces
Developed provinces	Alborz, Isfahan, Markazi, Qazvin, Qom, Semnan, Tehran and Yazd
Relatively developed provinces	Bushehr, East Azerbaijan, Fars, Gilan, Hamedan, Kermanshah, Khorasan Razavi, Khuzestan & Mazandaran
Undeveloped provinces	Ardabil, Kurdistan, Southern Khorasan, Sistan and Baluchistan, Lorestan, Chaharmahal and Bakhtiari, Zanjan, Golestan, Ilam, Kerman, Kohgiluyeh and Boyer Ahmad, North Khorasan, Hormozgan & West Azerbaijan

- *Development Status*: dividing the provinces into three categories of developed, Relatively developed, and undeveloped has been done considering development indices of the province obtained from the research "Investigating and comparing selected indices in provinces, emphasizing the census results from 2006 to 2011 and examining the development regarding some combined indices" published on the website of Statistical Research and Training Center (www.srtc.ac.ir). This category is shown in Table .

3-3-Data preparation

This phase of CRISP-DM methodology includes data selection, data cleaningand preparing them for data mining process.in this step, null and outlier values were corrected. Since the dataset belongs to a 4-years period, so the money value at the time has changed this affects the results. So the rate of inflation, up to 2016, was applied to amounts. Fields' values were applied in the normalized form to have the same impact in the analysis.

Once data is identified, the data preprocessing step begins. Data preprocessing is a very time consuming process and plays a very important role in a data mining project. Data preprocessing has a significant impact on accuracy and quality of the results of the data mining project. Duos to the real-world datasets include missing, noise and outlier values. The methods used are:

Data cleaning: some features with the noise values are detected in cleaning data and they are eliminated from our analyzis. Then the incompatibilities in the data are corrected.

Data cleaning is the quality control before the data analysis and one of its tasks is to fill in or delete the missing data. In this step, transactions with the incomplete or duplicate information were deleted. Some of these records in the research dataset have several fields with a value of 0. The remaining records were omitted due to duplication; that is, banks did not announce new statistics in some months, and there were duplicate statistics in that month.

Data integration: The dataset, which exists in different places and shapes, is collected in one place as integrated. At this step, data and information for each province, taken from the websites of Center of Statistics and Ministry of Information and Communications, were integrated in the second dataset with the electronic payment transaction information of each province within the period from 2013 to 2016.

[2] ICT Development Index

Data conversion: The normalization is performed on the data. This increases the accuracy and efficiency of the algorithms, especially the algorithms that work with the gap.

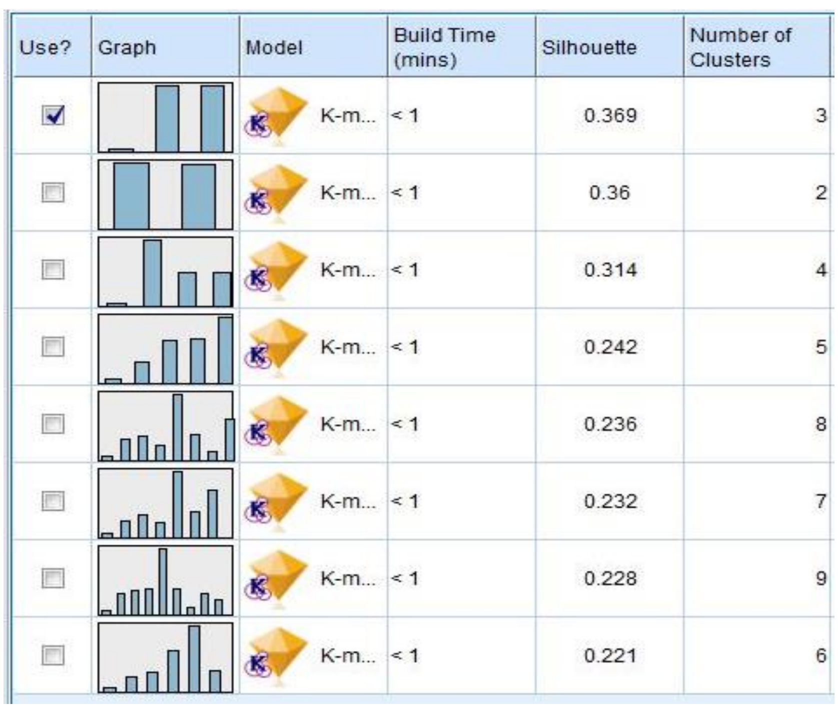
The average variables of the transaction amounts related to the various electronic payment tools were not available in the dataset taken from the Central Bank site. The average transaction value variables of various electronic payment instruments were not available in the dataset taken from the Central Bank site. By dividing the transaction value of each tool into the number of transactions of the same tool, the independent average transaction value variables were added to the dataset. Another activity at this step was applying the inflation rate in the field values. Since the dataset used is related to the multi-year timeline; so the value of money has changed over this period that which affects the analysis provided. For this reason, the inflation rate of each year was applied annually to amount of that year by the year 2016.

3-4-Modelling

In this stage, different modeling methods are selected and applied. Generally, there are some methods for a single type of data mining problem. *Clementine 12* software has been used to apply the methods in the present research.

3-4-1-Clustering

In this stage, Iran's provinces were clustered based on the carried out e-payment transactions and demographic features. To do so, K-Means clustering algorithm was employed. Three clusters were selected for clustering the records. The clustering was examined with different number of clusters 2 to 9 clusters. The silhouette criterion for the different number of clusters was calculated as the number three clusters (K=3) had the highest silhouette measure was selected. The resulting silhouette criterion for the different number of clusters is shown in figure 2.



Use?	Graph	Model	Build Time (mins)	Silhouette	Number of Clusters
✓		K-m...	< 1	0.369	3
		K-m...	< 1	0.36	2
		K-m...	< 1	0.314	4
		K-m...	< 1	0.242	5
		K-m...	< 1	0.236	8
		K-m...	< 1	0.232	7
		K-m...	< 1	0.228	9
		K-m...	< 1	0.221	6

Fig2. The silhouette criterion for the different number of clusters

Table 4 presents the brief results of this clustering. Figure 3 shows how the provinces are located in each cluster.

Fig3. Distribution of the provinces in each cluster

Table 3. Brief results of clustering

Mean / Mod	Cluster 1	Cluster 2	Cluster 3
Number of Internet transactions	180,742.239	33,165,704.057	58,285.445
Amount of Internet transactions	831,216.908	34,764,307.861	390,777.228
Average amount of Internet transactions	4.374	1.273	6.107
Number of Mobile transactions	7,867.5	111,610,843.829	191,469.376
Amount of Mobile transactions	2,991.814	3,898,191.199	1,311.313
Average amount of Mobile transactions	0.706	0.034	0.77
Number of POS transactions	20,781,899.332	162,100,284.743	10,011,878.941
Amount of POS transactions	27,278,447.669	298,065,329.325	11,927,222.242
Average amount of POS transactions	1.47	1.988	1.372
Number of POS device	0.104	0.83	0.044
Deposit	0.023	0.725	0.007
Financial facilities	0.019	0.713	0.007
Population	0.17	0.954	0.08
Economic participation rate	0.64	0.605	0.588
Unemployment rate	0.349	0.284	0.387
Literacy rate	86.321	91.3	81.802
Internet penetration rate	0.637	0.929	0.481
Number of computer users	0.117	0.794	0.046
Number of Internet users	0.145	0.878	0.062
Number of mobile users	0.16	0.95	0.07
IDI index	0.528	0.827	0.346
Skill IDI index	0.516	0.62	0.335
Use IDI index	0.377	0.657	0.255
Access IDI index	0.479	0.737	0.334
Development Ratio	9.562	1	24.429
Developmental Status	Relatively developed: 56.25% Developed: 43.75%	Developed:100%	Undeveloped: 100%
Number of records	560	35	490

According the clustering results formed clusters are described as follow.

Cluster 1: Transactions of more developed provinces

The population, and Mobile, the Internet, and computer users of such provinces are larger in comparison with cluster 3. The value of ICT indices in the provinces of this cluster is higher than that of cluster 3. The lowest mean of Mobile banking transaction number belongs to the records of this cluster. Considering the fact that this cluster's provinces have the larger population, they provide a suitable market for e-banking, and advertisements and marketing among the customers of such cluster will yield favorable results.

Cluster 2: Transactions of Tehran province

The highest mean of the amount and number of the Internet, Mobile, and POS transactions, the highest mean of the average amount of POS transactions, and the lowest mean of the average amount of Mobile and Internet banking transactions belong to this cluster. Customers of such cluster are considered as valuable customers of e-banking due to greater access and making use of information and communications technology as well as their absolute trust in e-payment methods. Discussing the plans for maintaining the customers of Tehran province by bank managers with the aim of increasing these customers' profitability will be useful. One of the influential policies in this regard is designing and presenting various products and customization the services in the field of e-banking. Customization in banking is possible with the use of analytical predictions and combing the analysis of the activities the user's financial status, analysis of big data in usual behavioral patterns, and text analyses. Considering the acceptability of e-banking among the customers of this cluster, offering such services will be welcomed warmly and cause an increase in their profitability and loyalty.

Cluster 3: Transactions of less developed provinces

The value of information technology indices in the provinces of this cluster is lower than in cluster 1. The lowest mean of the number of the Internet and POS transactions, the amount of Mobile banking, Internet banking and POS transactions, and the average amount of POS transactions belong to the records of this cluster. The rate of literacy and skill of IDI index among this cluster's customers is low in comparison with other clusters. It can be inferred that one of the reasons that e-payment methods are less acceptable in these provinces is what was mentioned above. One of the policies that can be effective in line with expanding e-banking in these provinces is planning by bank managers in order to educate the customers of this cluster and make them aware of how to use e-payment methods and their advantages. In addition to this, educating bank stuff so as to guide the customers into using e-payment methods in such provinces will increase the usage of e-banking services. Another point that needs to be considered in designing Internet-based and mobile-based applications is the simplification and easiness of using such applications, which will lead to the customers' warmer welcome.

3-4-1-1-Clustering transactions of each year separately

In the next step records of each year were clustered separately.

Step 1: At first records related to 2013were clustered. Brief results of this clustering are shown in Table .

Table 4.Results from clustering of records related to 2013

Mean / Mod	Cluster 1	Cluster 2	Cluster 3
Number of Internet transactions	52,697.466	25,924,549.36	21 ,770.87
Amount of Internet transactions	293,434.54	12,784,749.94	90,929.633
Average amount of Internet transactions	2.684	0.408	3.119
Number of Mobile transactions	135.932	84,438,350.82	36.571
Amount of Mobile transactions	213.541	1,648,333.711	69.538
Average amount of Mobile transactions	0.973	0.019	1.01
Number of POS transactions	10,292,355.778	101,147,352.1	4,868,490.74
Amount of POS transactions	21,312,165.009	282,503,324.954	9,137,773.218
Average amount of POS transactions	2.057	2.788	1.919
Number of POS device	0.063	0.777	0.023
Deposit	0.014	0.495	0.004
Financial facilities	0.012	0.511	0.004
Population	0.167	0.928	0.076
Economic participation rate	0.603	0.519	0.529

Table 4. Continued

Mean / Mod	Cluster 1	Cluster 2	Cluster 3
Unemployment rate	0.298	0.279	0.345
Literacy rate	85.299	90.46	80.735
Internet penetration rate	0.375	0.775	0.239
Number of computer users	0.077	0.663	0.022
Number of Internet users	0.077	0.725	0.038
Number of mobile users	0.139	0.89	0.05
IDI index	0.295	0.505	0.172
Skill IDI index	0.589	0.695	0.403
Use IDI index	0.131	0.298	0.069
Access IDI index	0.275	0.478	0.183
Development Ratio	9.562	1	24.429
Developmental Status	Relatively developed: 56.25% Developed: 43.75%	Developed: 100%	Undeveloped: 100%
Number of records	176	11	154

Step 2: Then the records related to 2015were clustered. Results of this clustering are shown in table 5

Table 5. Results for clustering of records related to 2015

Mean / Mod	Cluster 1	Cluster 2	Cluster 3
Number of Internet transactions	208,352.5	50,993,243.83	624,48.929
Amount of Internet transactions	991,699.9	33,922,361.53	497,381.008
Average amount of Internet transactions	5.392	0.803	8.554
Number of Mobile transactions	8,770.661	136,797,004.7	17,173.024
Amount of Mobile transactions	3,710.499	4,325,987.717	1,724.768
Average amount of Mobile transactions	0.813	0.032	1.025
Number of POS transactions	21,049,366.339	160,158,586.1	10,169,313.53
Amount of POS transactions	26,889,385.489	280,824,503.7	11,720,795.03
Average amount of POS transactions	1.286	1.752	1.177
Number of POS device	0.11	0.8	0.048
Deposit	0.024	0.744	0.007
Financial facilities	0.019	0.733	0.007
Population	0.171	0.95	0.081
Economic participation rate	0.644	0.559	0.582
Unemployment rate	0.351	0.201	0.39
Literacy rate	85.299	90.46	80.735
Internet penetration rate	0.758	1	0.592
Number of computer users	0.102	0.708	0.037
Number of Internet users	0.15	0.895	0.07
Number of mobile users	0.169	0.956	0.078
IDI index	0.605	1	0.37
Skill IDI index	0.589	0.695	0.403
Use IDI index	0.559	1	0.367
Access IDI index	0.415	0.712	0.257
Development Ratio	9.562	1	24.429

Table 5. Continued

Mean / Mod	Cluster 1	Cluster 2	Cluster 3
Developmental Status	Relatively developed: 56.25% Developed: 43.75%	Developed: 100%	Undeveloped: 100%
Number of records	192	12	168

Step 3: Finally the records related to 2015were clustered. Results of this clustering are shown in table 6

Table 6. Results f clustering of records related to 2016

Mean / Mod	Cluster 1	Cluster 2	Cluster 3
Number of Internet transactions	270,506.39	21,975,889.42	87,593.655
Amount of Internet transactions	1,163,701.09	46,153,482.92	559,033.744
Average amount of Internet transactions	4.90	2.536	6.398
Number of Mobile transactions	14,051.61	111,332,801.6	541,245.798
Amount of Mobile transactions	4,819.88	5,251,358.999	2,036.153
Average amount of Mobile transactions	0.36	0.05	0.294
Number of POS transactions	30,129,847.25	209,171,214.2	14,569,216.87
Amount of POS transactions	33,136,602.29	322,397,514.007	14,690,644.39
Average amount of POS transactions	1.12	1.49	1.059
Number of POS device	0.14	0.874	0.061
Deposit	0.03	0.899	0.009
Financial facilities	0.02	0.863	0.009
Population	0.17	0.95	0.083
Economic participation rate	0.67	0.729	0.647
Unemployment rate	0.39	0.371	0.42
Literacy rate	88.08	90.46	83.846
Internet penetration rate	0.76	1	0.592
Number of computer users	0.17	0.708	0.075
Number of Internet users	0.16	0.895	0.078
Number of mobile users	0.17	0.956	0.08
IDI index	0.67	0.949	0.481
Skill IDI index	0.37	0.476	0.217
Use IDI index	0.42	0.645	0.315
Access IDI index	0.68	0.712	0.55
Development Ratio	9.56	1	24.429
Developmental Status	Relatively developed: 56.25% Developed: 43.75%	Developed: 100%	Undeveloped: 100%
Number of records	192	12	168

According to results of clustering transactions of each year, formed clusters in every 3 steps are similar and described as follow.

Cluster 1: Transactions of more developed provinces. The population, and Mobile, the Internet, and computer users of such provinces are larger in comparison with cluster 3.

Cluster 2: Transactions of Tehran province. The highest mean of the amount and number of all e-payment tools transactions belong to this cluster.

*Cluster 3: Transactions of less developed provinces.*The lowest mean of the amount and number of all e-payment tools transactions belong to this cluster.

3-4-1-2-Comparing the formed Clusters

a. Comparing cluster 1 records in 2013, 2015, and 2016

Comparing cluster 1 reveals that during those four years in more developed provinces, the mean of the amount and numbers of Internet banking, Mobile banking, and POS transactions have been on the rise, while the mean of the average amount of Mobile banking and POS transactions have been on the wane. The volume of Internet transactions of the bank in 2015 has increased in comparison with 2013 and decreased compared to 2016. As can be seen in Table , the highest level of increase during these four years in this cluster belongs to the number of Mobile banking transactions with 10237.23 percent changes and the highest level of decrease belongs to the average amount of Mobile banking transactions with 63.001 percent changes. The number of transactions, the amount of transactions, and the average amount of transactions conducted with different devices in the provinces of cluster 1 in 2013, 2015, and 2016 has been compared respectively in figure 4, figure 5, and figure 6.

Table 7. Changes rate of Cluster 1 records in 2013, 2015, and 2016

Mean of results	Number of Internet transactions	Amount of Internet transactions	Average amount of Internet transactions	Number of Mobile transactions	Amount of Mobile transactions	Average amount of Mobile transactions	Number of POS transactions	Amount of POS transactions	Average amount of POS transactions
Changes rate	413.31% +	296.57% +	82.56% +	10237.23% +	2157.122% +	63.001% −	192.74% +	55.48% +	45.55% −

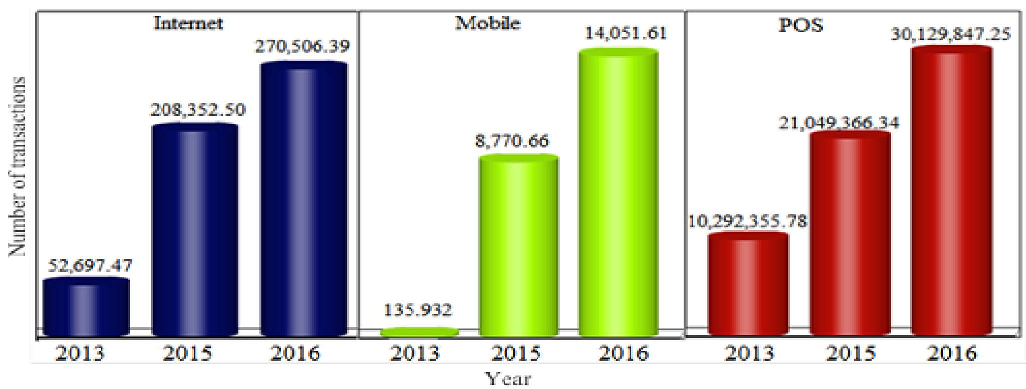


Fig 2. The number of cluster 1 transactions

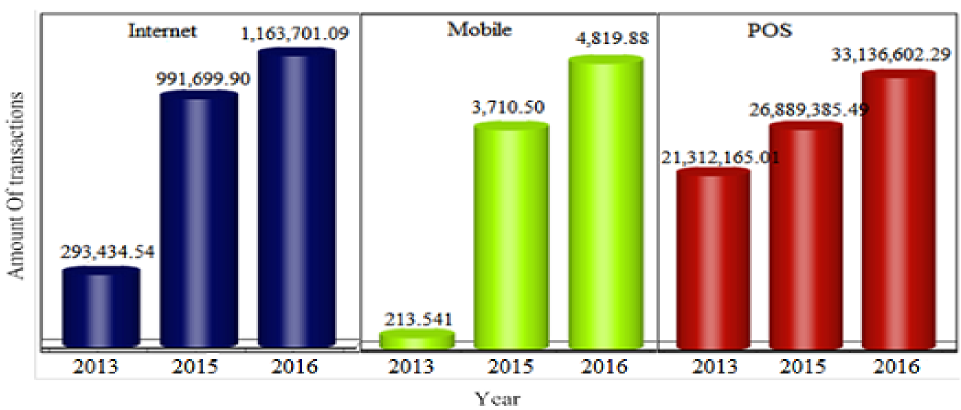


Fig 3. The amount of cluster 1 transactions

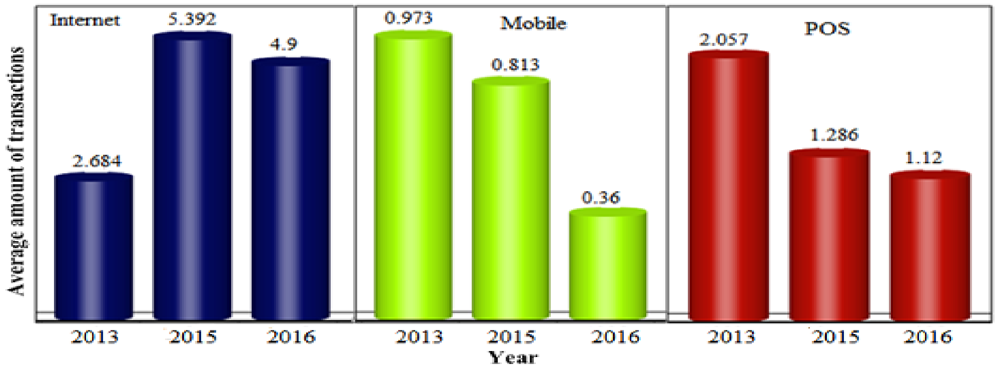


Fig 4. The averge amount of cluster 1 transactions

b. Comparing cluster 2 records in 2013, 2015, and 2016

Comparing cluster 2 during these 4 years shows that the mean of average amount and amount of Internet banking and Mobile banking transactions and the mean of number and amount of POS transactions were on the rise and the mean of average amount of POS transactions was on the wane. The number of mobile banking and Internet banking transactions in 2015 was the largest compared to two other years. As can be seen in Table , the highest rate of increase during these 4 years in this cluster belongs to the mean of amount of Internet banking transactions cost with 521.56 percent changes, while the highest rate of decrease belongs to the mean of average amount of POS transactions with 46.55 percent changes. The number of transactions, the amount of transactions, and the average amount of transactions conducted with different devices in the provinces of cluster 2 in 2013, 2015, and 2016 has been compared respectively in figure 7, figure 8, and figure 9.

Table 8. Changes rate of Cluster 2 records in 2013, 2015, and 2016

Mean of results	Number of Internet transactions	Amount of Internet transactions	Average amount of Internet transactions	Number of Mobile transactions	Amount of Mobile transactions	Average amount of Mobile transactions	Number of POS transactions	Amount of POS transactions	Average amount of POS transactions
changes rate	15.23% −	261.004% +	521.56% +	31.85% +	218.58% +	163.15% +	106.79% +	14.12% +	46.55% −

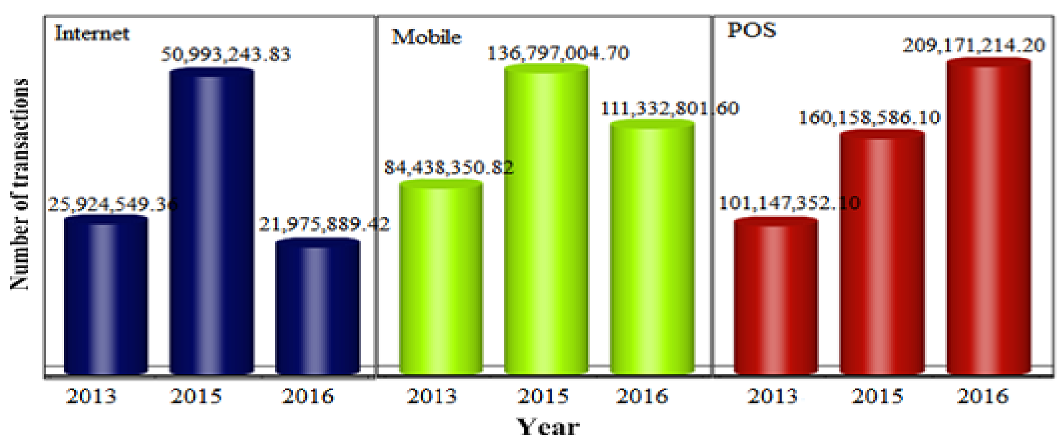


Fig 5. The number of cluster 2 transactions

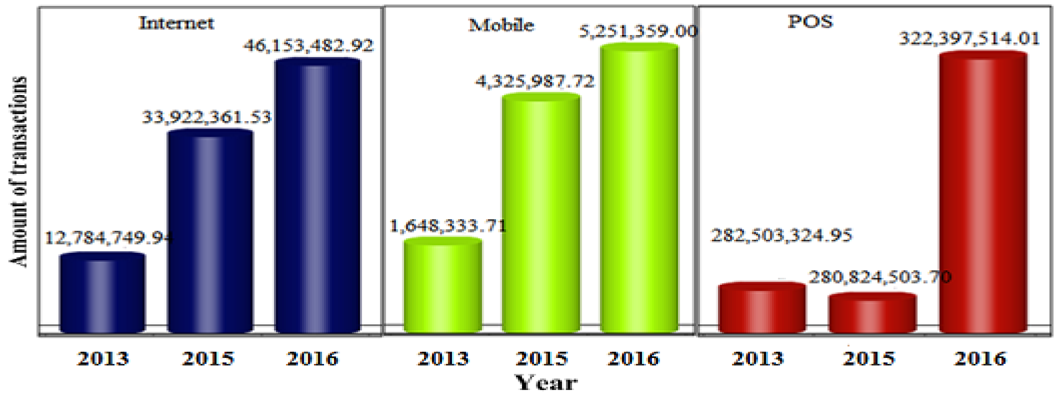


Fig 6. The amount of cluster 2 transactions

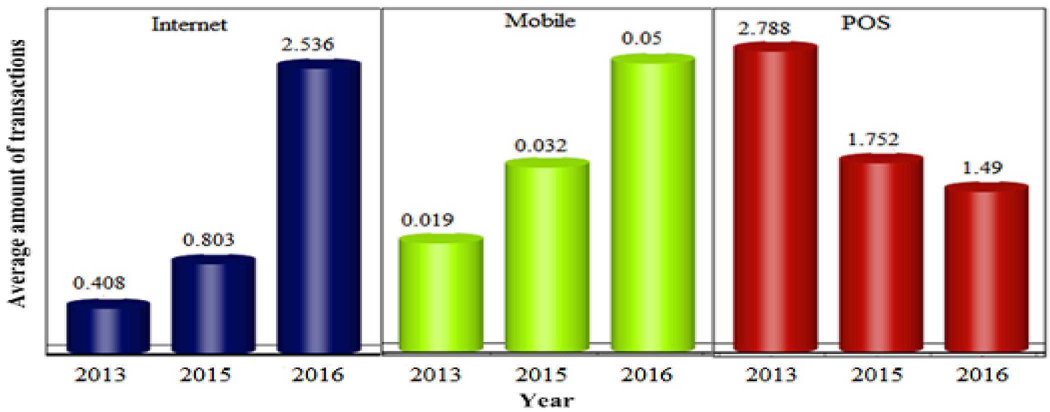


Fig 7. The average amount of cluster 2 transactions

c. Comparing cluster 3 records in 2013, 2015, and 2016

Comparing cluster 3 during these 4 years demonstrates an increase in the mean of the amount and number of Internet banking, Mobile banking, and POS transactions, and a decrease in the average amount of POS and Mobile banking transactions. The average amount of Internet banking transactions in 2015 was the largest compared to two other years. As can be seen in Table 9, the highest rate of increase during these 4 years in this cluster belongs to the the number of Internet banking transactions with 1479886.3 percent changes, while the highest rate of decrease belongs to the average amount of Mobile banking transactions with 70.89 percent changes. The number of transactions, the amount of transactions, and the average amount of transactions conducted with different devices in the provinces of cluster 2 in 2013, 2015, and 2016 has been compared respectively in figure 10, figure 11, and figure 12.

Table 9. Changes rate of cluster 3 records in 2013, 2015, and 2016

Mean of results	Number of Internet transactions	Amount of Internet transactions	Average amount of Internet transactions	Number of Mobile transactions	Amount of Mobile transactions	Average amount of Mobile transactions	Number of POS transactions	Amount of POS transactions	Average amount of POS transactions
changes rate	302.34% +	514.79% +	105.12% +	1479886.3% +	2828.11% +	70.89% −	199.25% +	60.76% +	44.81% −

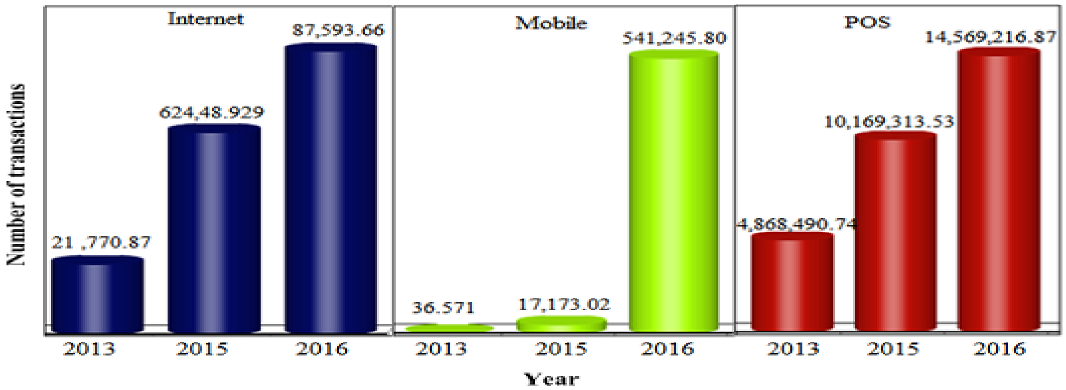


Fig 8. The number of cluster 3 transactions

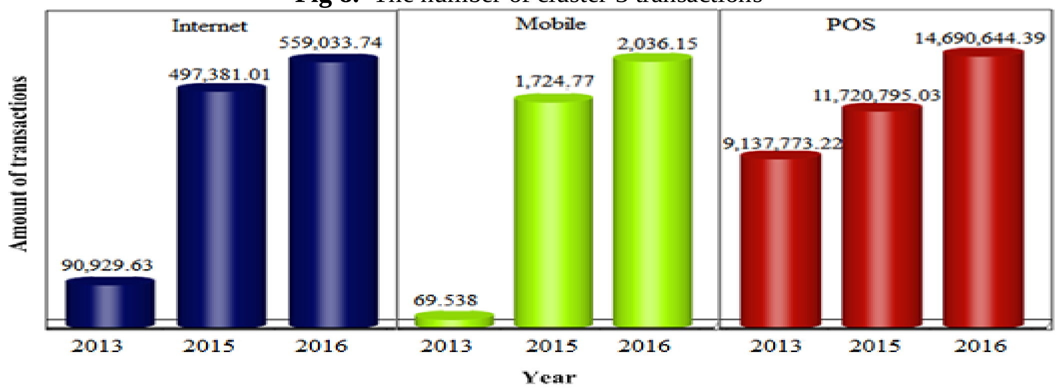


Fig 9. The amount of cluster 3 transactions

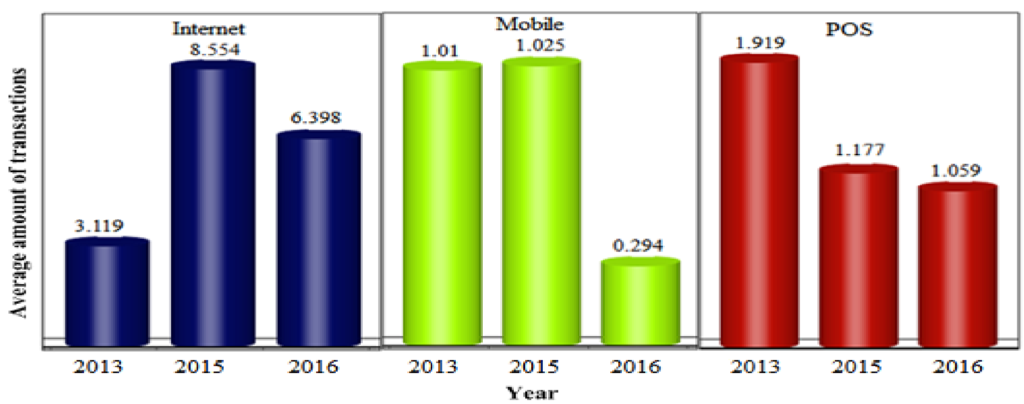


Fig 10. The average amount of cluster 3 transactions

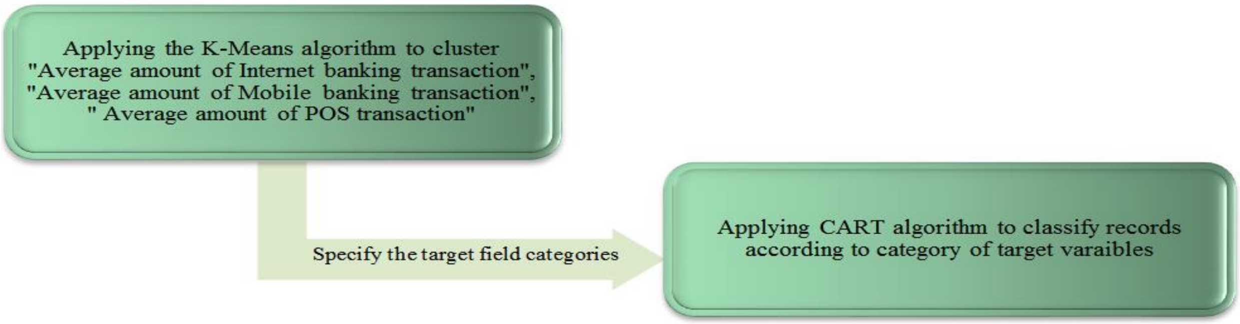


Fig 11. Specify the target field categories

3-4-2-Decision tree

The aim of this research was to identify the effective factors in expanding e-payment in Iran. CART classification algorithm was employed to extract hidden pattern in customers' online behavior. The factors affecting Internet banking, Moblie banking, and POS adoption were surveyed separately. Thus three objective variables were required in classification algorithm implementation. A categorical target variable is required to apply this algorithm. In this study, the target variables were "Average amount of Internet banking transactions", "Average amount of Mobile banking transactions" and "Average amount of POS transactions" which they had continuous numeric values. There is not any threshold to determine low, high or medium amount of transactions in banking literature. There were 2 methods to categorize target variables. Categorize them based on the difference between values or applying clustering technique. So, the clustering algorithm was chosen to categorize target field according to the personal viewpoint of authors. The clustering algorithm was run 3 times. First time for categorizing the variable of "Average amount of Internet banking transactions" and this field was the only input of this implementation, and two more times to category "Average amount of Mobile banking transactions" and "Average amount of POS transactions". These stages are described in Fig 11. This section relates to modeling with the decision tree algorithm, and the rules derived from the decision tree are recorded in these tables from root to leaf. All rules and branches created by decision tree algorithm were examined. Branches in the end nodes that differ in various categories are considered to be the important rules, and the factor that has changed is considered as an effective factor. These important rules are recorded in this section. These rules have been investigated to extract the hidden information from the dataset, and these rules in fact determine the steps of examining different factors to achieve the factor effective in the use of the electronic payment tools. For a better presentation, these rules are drawn out as tree to be more intelligible. As the decision-tree algorithm examines the important features step-by-step to reach a rule, the same process is used in the figures of this section.

Table 10. Results of clustering on the Average Internet banking transactions

Cluster Number	Minimum Value	Maximum value	Record Number	Assigned Label
Cluster 1	0.0005	4.56	578	Low
Cluster 2	12.89	45.76	63	High
Cluster 3	4.64	12.22	444	Medium

Table 11 . Results of clustering on the Average Mobile banking transactions

Cluster Number	Minimum Value	Maximum value	Record Number	Assigned Label
Cluster 1	0	0.65	841	Low
Cluster 2	0.21	14.76	98	High
Cluster 3	0.66	1.92	146	Medium

Table12. Results of clustering on the Average POS transactions

Cluster Number	Minimum Value	Maximum value	Record Number	Assigned Label
Cluster 1	2.31	3.55	84	High
Cluster 2	0.62	1.44	677	Low
Cluster 3	1.45	2.30	324	Medium

Then “Average Amount of Internet banking Transactions”, “Average Amount of Mobile banking Transactions”, and “Average Amount of POS Transactions” fields’ values were divided into three clusters, Medium, Low and High. The labels were assigned according to the average amount of transactions to each record. The results of these clusterings are shown respetively in table 10, table 11, and table 12.

In this stage, CART decision tree algorithms were employed in order to discover the effective factors in the average amount of Internet banking, Mobile banking, and POS transactions. These algorithms were separately run on clusters 1 and 3. Since the records of cluster 2 were specific to the transactions done in Tehran province, no useful rules were extracted from the decision tree. First, the results obtained from running algorithms on cluster 1 have been investigated.

3-4-2-1-Decision tree of Cluster 1 Records

a. The results of CART Decision tree related to Cluster 1 Internet banking transactions

In the first implementation of CART algorithm, the variables of Province, Year, Month, Number of Mobile banking transactions, Amount of Mobile banking transactions, Average amount of Mobile banking transactions, Number of POS transactions, Amount of POS transactions, Average amount of POS transactions, Number of POS device, Deposit, Financial facilities, Population, Economic participation rate, Unemployment rate, Literacy rate, Internet penetration rate, Number of computer users, Number of Internet users, Number of mobile users, IDI index, Access IDI index, skill IDI index, Use IDI index, Development Ratio, Developmental Status were considered as the input variables and the High, Medium and Low labels of the average amount of cluster1 Internet banking transactions was set as the target variable. Table shows the classification accuracy of the algorithm. The extracted rules from CART algorithm is displayed in figure 14 and table 14.

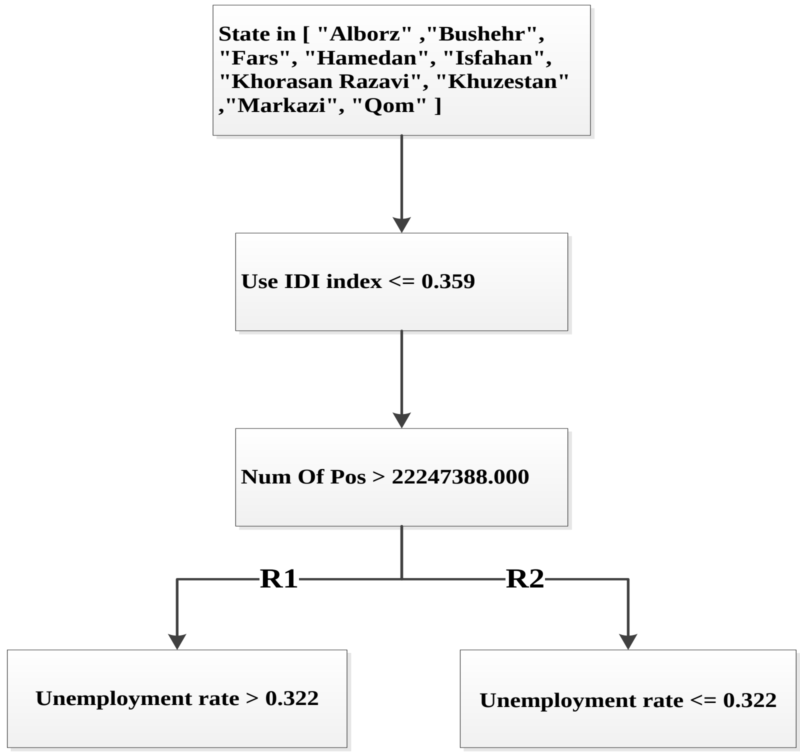


Fig 12. The extracted rules of CART decision tree related to cluster 1 Internet banking transactions

Table13. The accuracy of CART decision tree related to cluster 1Internet banking transactions

Partition	Training		Test	
Correctly classified samples	355	85.13%	91	76.47%
Incorrectly classified samples	62	14.87%	28	23.53%
Total	417		119	

Table 14 . The extracted rules' frequency of R1 and R2

Rule number	**Low Category**		**Medium Category**		**High Category**	
	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency
R_1	86.66%	13	13.33%	2	0%	0
R_2	0%	0	37.5%	3	62.5%	5

Considering R1 and R2 rules in Isfahan and Khorasan Razavi provinces in the months when the rate of unemployment was more than 0.322, the number of Internet banking transactions has dropped.

b. The results of CART Decision tree related to Cluster 1 Mobile banking transactions

In the next implementation of CART algorithm, the variables of Province, Year, Month, Number of Internet banking transactions, Amount of Internet banking transactions, Average amount of Internet banking transactions, Number of POS transactions, Amount of POS transactions, Average amount of POS transactions, Number of POS device, Deposit, Financial facilities, Population, Economic participation rate, Unemployment rate, Literacy rate, Internet penetration rate, Number of computer users, Number of Internet users, Number of mobile users, IDI index, Access IDI index, skill IDI index, Use IDI index, Development Ratio, Developmental Status were considered as the input variables and the High, Medium and Low labels of the average amount of cluster 1 Mobile banking transactions was set as the target variable. Table 15 shows the classification accuracy of the algorithm. The extracted rules from CART algorithm is displayed in figure 15 and table 16.

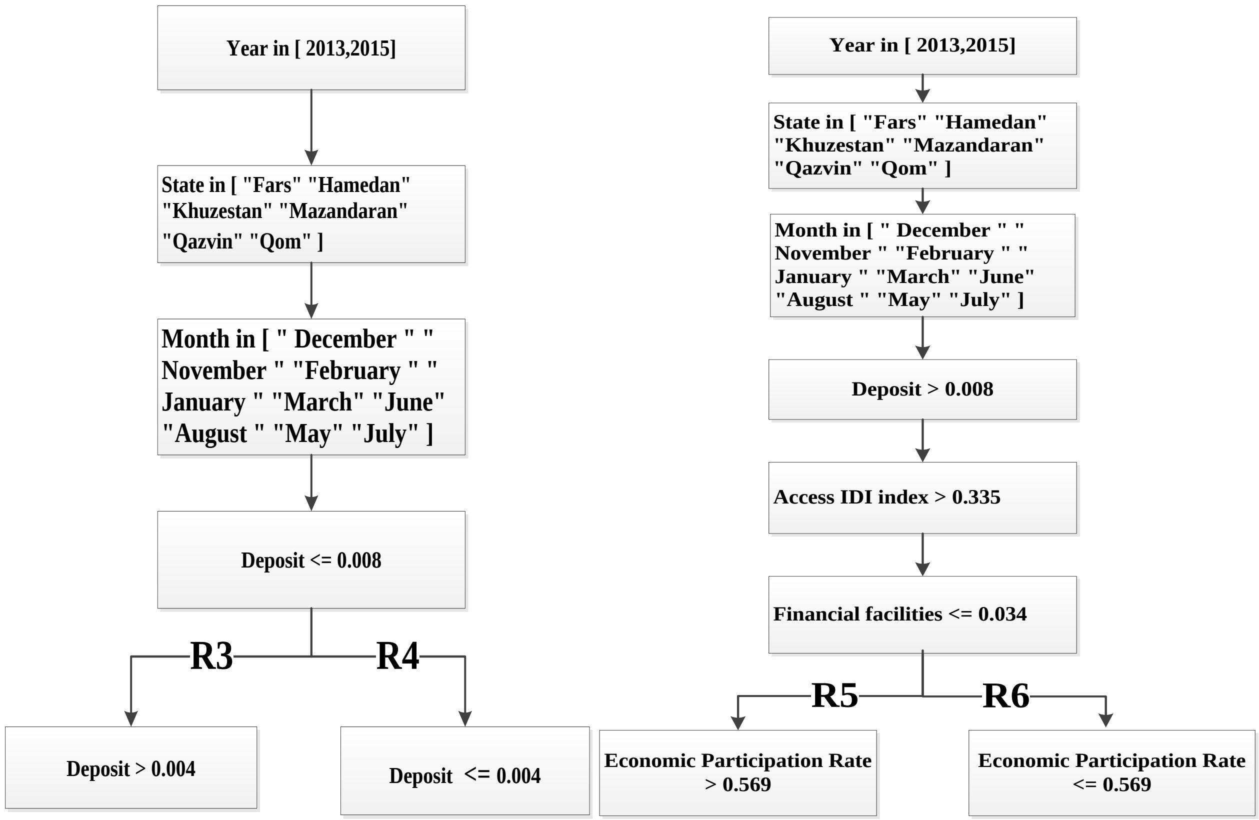


Fig 13. The extracted rules of CART decision tree related to cluster1 Mobile banking transactions

Table15. The accuracy of CART decision tree related to cluster 1mobile banking transactions

Partition	Training		Test	
Correctly classified samples	367	88.01%	90	75.63%
Incorrectly classified samples	50	11.99%	29	24.37%
Total	417		119	

Table 16. The extracted rules' frequency of R3,R4, R5, R6

Rule Number	Low Category		Medium Category		High Category	
	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency
R_3	17.39%	4	73.91%	17	8.69%	2
R_4	100%	7	0%	0	0%	0
R_5	33.33%	7	61.90%	13	4.76%	1

Table 16. Continued

Rule Number	Low Category		Medium Category		High Category	
	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency
R_6	72.72%	8	9.09%	1	18.18%	2

Comparing R3 and R4 rules indicates the positive effect of the quantity of deposits on Mobile banking transactions. The number of transactions in Hamedan, Qom, and Qazvin provinces in the months when the quantity of deposits is more than 0.004 is larger than when the quantity of deposits is less than 0.004. These rules clarify the relationship between bank deposits and welcome to electronic banking. Hence, one of the policies that banks can adopt to expand electronic banking is delivering different products and facilities regarding bank deposits. The banks that perform better in absorbing bank deposits will feel its effect on e-banking expansion.

Considering R5 and R6 rules in Mazandaran, Khuzestan, and Fars where the rate of economic participation is more than 0.569, the number of Mobile banking transactions is more than in Qom and Khuzestan provinces with economic participation rate of less than 0.569. Comparing these rules reveals that bank customers trust mobile as a device for transferring money in the economic activities as well, and mobile is one of the important channels in economic payments. So, delivering various services specific to economic active customers will increase the usage of the mobile device by the customers, which will definitely be welcomed by a wide range of customers because of high accessibility of mobile.

c. The results of CART Decision tree related to Cluster 1 POS transactions

In the next implementation of CART algorithm, the variables of Province, Year, Month, Number of Internet banking transactions, Amount of Internet banking transactions, Average amount of Internet banking transactions, Number of Mobile banking transactions, Amount of Mobile banking transactions, Average amount of Mobile banking transactions, Number of POS device, Deposit, Financial facilities, Population, Economic participation rate, Unemployment rate, Literacy rate, Internet penetration rate, Number of computer users, Number of Internet users, Number of mobile users, IDI index, Access IDI index, skill IDI index, Use IDI index, Development Ratio, Developmental Status were considered as the input variables and the High, Medium and Low labels of the average amount of POS transactions was set as the target variable. Table shows the classification accuracy of the algorithm. The extracted rules from CART algorithm is displayed in figure 16 and table 18.

Comparing R7 and R8 rules confirms the positive effect of economic partnership rate on increasing the number of point of sale transactions. These rules apply to East Azerbaijan, Bushehr, Khuzestan, Kermanshah, and Hamedan provinces in 2013. The transactions have been carried out with high volume in Hamedan and Khuzestan provinces in the months when the economic participation rate is more than 0.439. These rules illustrate the importance of paying attention to economic activities expansion in order to expand e-banking. One of the effects of increasing economic activity is that the number of POS devices will rise. And one of the factors that determine the market size of e-payment services is the number of compliance tools. Increasing in the number of these tools as well as other factors including the number of transactions can be an indicator of access improvement and development of e-payment market. Therefore, the expansion of economic activities has a direct relationship with electronic banking.

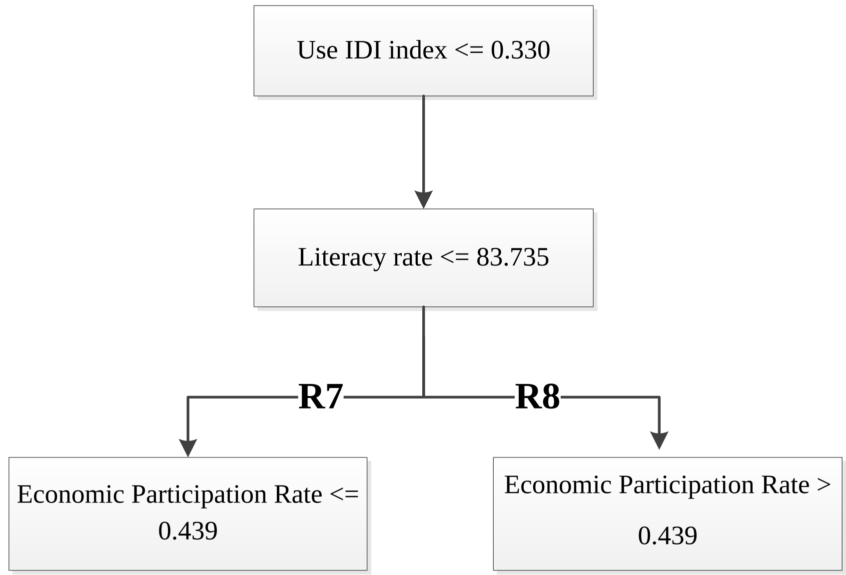


Fig 14. The extracted rules of CART decision tree related to cluster1 POS transactions

Table 17. The accuracy of CART decision tree related to cluster1POS transactions

Partition	Training		Test	
Correctly classified samples	396	94.96%	109	91.6%
Incorrectly classified samples	21	5.04%	10	8.4%
Total	417		119	

Table 18 . The extracted rules' frequency of R7 and R8

Rule Number	Low Category		Medium Category		High Category	
	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency
R_7	0%	0	71.42%	5	28.57%	2
R_8	0%	0	8.1%	3	91.89%	34

3-4-2-2-Decision tree of Cluster 3 Records

a. The results of CART Decision tree related to Cluster 3 Internet banking transactions

In the next implementation of CART algorithm, the variables of Province, Year, Month, Number of Mobile banking transactions, Amount of Mobile banking transactions, Average amount of Mobile banking transactions, Number of POS transactions, Amount of POS transactions, Average amount of POS transactions, Number of POS device, Deposit, Financial facilities, Population, Economic participation rate, Unemployment rate, Literacy rate, Internet penetration rate, Number of computer users, Number of Internet users, Number of mobile users, IDI index, Access IDI index, skill IDI index, Use IDI index, Development Ratio, Developmental Status were considered as the input variables and the High, Medium and Low labels of the average amount of cluster3 Internet banking transactions was set as the target variable. The classification accuracy of the algorithm is indicated in table 19. Figure 17 and table 20 show the extracted rules from CART algorithm.

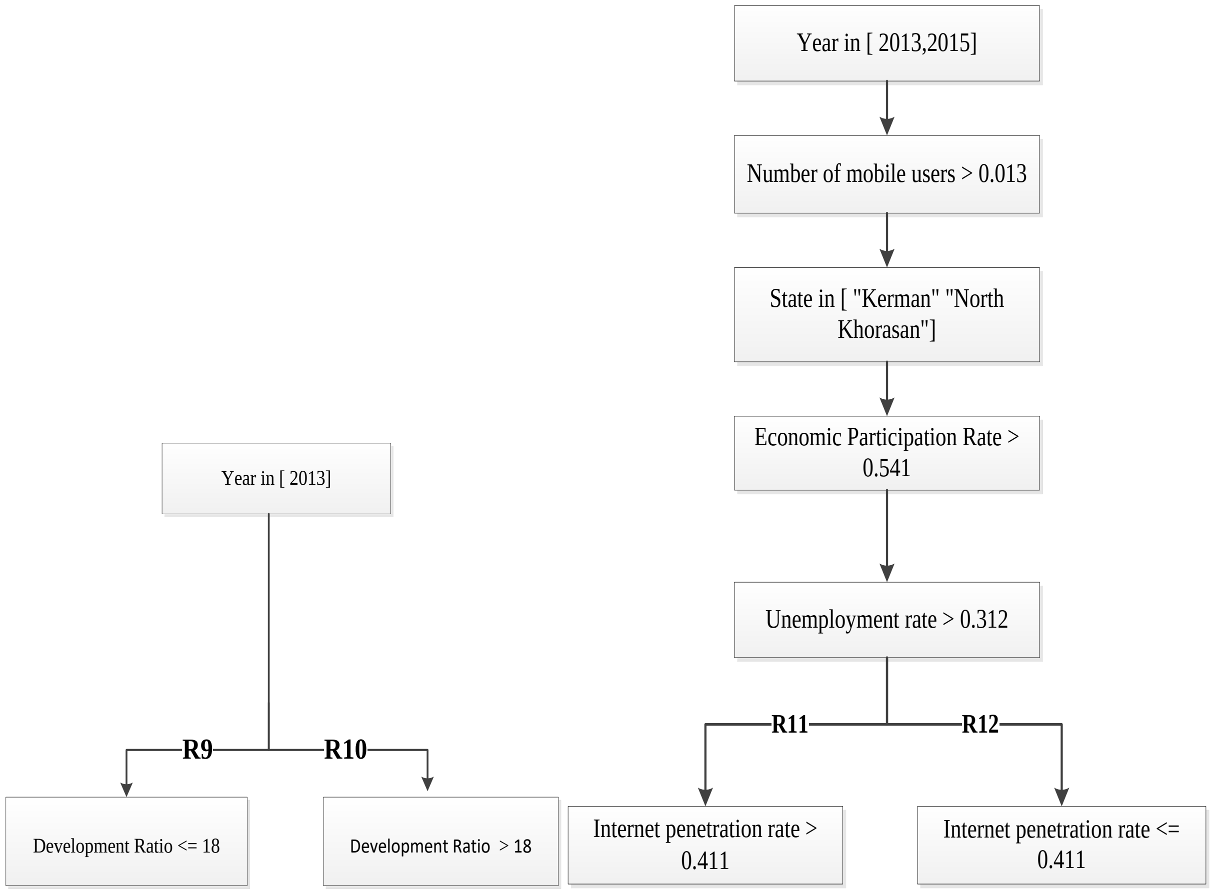


Fig 15. The extracted rules of CART decision tree related to cluster 3 internet banking transactions

Table 19. The accuracy of CART decision tree related to cluster 3internet banking transactions

Partition	Training		Test	
Correctly classified samples	329	87.73%	98	85.22%
Incorrectly classified samples	46	12.27%	17	14.78%
Total	375		115	

Table 20. The extracted rules' frequency of R9, R10, R11 and R12

Rule Number	Low Category		Medium Category		High Category	
	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency
R_9	25%	2	0%	0	75%	6
R_{10}	86.91%	93	12.15%	13	0.93%	1
R_{11}	0%	0	25%	3	75%	9
R_{12}	25%	1	75%	3	0%	0

Table 20. Continued

Rule Number	Low Category		Medium Category		High Category	
	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency

Considering R9 and R10 rules in 2013, Zanjan province with development rate of less than 18 has had more Internet banking transactions than other provinces with development rate of greater than 18. This shows that the customers of less developed provinces require more attention and care from managers and decision makers so that they can benefit from e-banking services. The customers of such provinces have more trust in traditional payment methods and in addition to providing hardware infrastructure, informing them is necessary to increase trust in e-payment methods.

Comparing R11 and R12 rules reveal that the increase in the Internet penetration rate in North Khorasan and Kerman provinces increase the volume of Internet banking transactions. The Internet penetration rate in North Khorasan from June to March, 2015 is more than 0.411, and the number of Internet banking transactions was more in comparison with September to March of 2016 with the Internet penetration rate of greater than 0.411. E-payment is done in the Internet context. So, secure, rapid, and easy access to the Internet is one of the requirements of using e-payment methods. Greater access to the Internet has a direct effect on the expansion of electronic banking. One of the most important required measures to be taken by the authorities in order to expand e-banking is greater access to the Internet and improvement in delivering Internet services.

b. The results of CART decision tree related to cluster 3 mobile banking transactions

In the next implementation of CART algorithm, the variables of Province, Year, Month, Number of Internet banking transactions, Amount of Internet banking transactions, Average amount of Internet banking transactions, Number of POS transactions, Amount of POS transactions, Average amount of POS transactions, Number of POS device, Deposit, Financial facilities, Population, Economic participation rate, Unemployment rate, Literacy rate, Internet penetration rate, Number of computer users, Number of Internet users, Number of mobile users, IDI index, Access IDI index, skill IDI index, Use IDI index, Development Ratio, Developmental Status were considered as the input variables and the High, Medium and Low labels of the average amount of cluster 3 Mobile banking transactions was set as the target variable. The classification accuracy of the algorithm is indicated in table 21. Figure 18 and table 22 show the extracted rules from CART algorithm.

Table 21. The accuracy of CART decision tree related to cluster 3Mobile banking transactions

Partition	Training		Test	
Correctly classified samples	324	86.4%	93	80.87%
Incorrectly classified samples	51	13.6%	22	19.13%
Total	375		115	

Table 22 . The extracted rules’ frequency of R13, R14, R15 and R16

Rule Number	Low Category		Medium Category		High Category	
	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency
R_{13}	70.83%	17	16.66%	4	12.5%	3

Table 22. Continued

Rule Number	Low Category		Medium Category		High Category	
	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency
R_{14}	16.66%	1	16.66%	1	66.66%	4
R_{15}	25%	3	25%	3	50%	6
R_{16}	87.5%	49	7.14%	4	5.35%	3

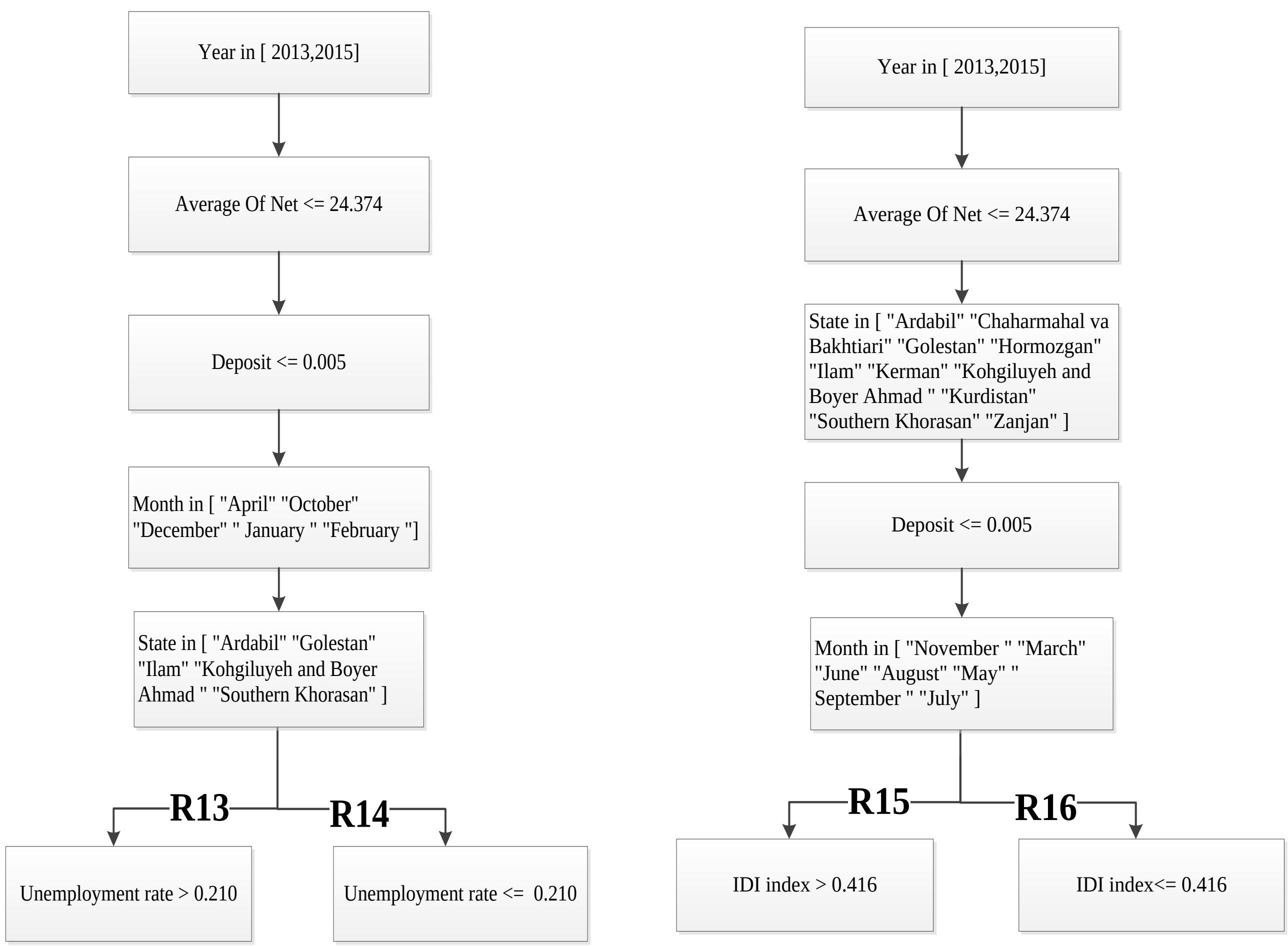


Fig 18. The extracted rules of CART decision tree related to cluster 3 Mobile banking transactions

Considering R13 and R14 rules, the reduction of the unemployment rate has increased mobile banking transactions. Mobile banking transactions in 2013 in North Khorasan with the unemployment rate of lower than 0.21 has been more compared to the provinces with the higher unemployment rate. On condition that business is booming, turnover will certainly be more, which increases the use of e-payment tools.

According to R15 and R16 rules, the increase in IDI index has had a positive effect on the increase in the volume of Mobile banking transactions. In 2015 in Ardebil and Ilam provinces with IDI index of lager than

0.416, the percentage of Mobile banking transactions with high volume has been more in comparison with Chaharmahal and Bakhtiari, Golestan, Hormozgan, Kerman, Kohgiluyeh and Boyer Ahmad, Zanjan, Kurdistan, and North Khorasan with IDI index of smaller than 0.416. The expansion of e-banking has a great dependence on the facilities and infrastructure of information and communication technology. The level of access to information technology equipment and the customers' usage and awareness of it have direct relationship with the customers' welcome to e-banking services. Thus, it is necessary that the authorities place a high priority on the development of hardware facilities of information and communications technology so as to expand e-banking. Moreover, training and informing the customers of e-payment services are among the effective policies that must be taken into account by bank managers.

c. The results of CART Decision tree related to Cluster 3 POS transactions

In the next implementation of CART algorithm, the variables of Province, Year, Month, Number of Internet banking transactions, Amount of Internet banking transactions, Average amount of Internet banking transactions, Number of Mobile banking transactions, Amount of Mobile banking transactions, Average amount of Mobile banking transactions, Number of POS device, Deposit, Financial facilities, Population, Economic participation rate, Unemployment rate, Literacy rate, Internet penetration rate, Number of computer users, Number of Internet users, Number of mobile users, IDI index, Access IDI index, skill IDI index, Use IDI index, Development Ratio, Developmental Status were considered as the input variables and the High, Medium and Low labels of the average amount of cluster 3 POS transactions was set as the target variable. Table 23 indicates the classification accuracy of the algorithm. Figure 19 and table 24 shows the extracted rules from CART algorithm.

Table23. The accuracy of CART decision tree related to cluster 3 POS transactions

Partition	Training		Test	
Correctly classified samples	369	98.4%	111	96.52%
Incorrectly classified samples	6	1.6%	4	3.48%
Total	375		115	

Table24. The extracted rules' frequency of R17, R18, R19 and R20

Row	Low Category		Medium Category		High Category	
	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency	Relative Frequency	Absolute Frequency
R_{17}	0%	0	0%	0	100%	20
R_{18}	0%	0	100%	7	0%	0
R_{19}	0%	0	100%	8	0%	0
R_{20}	100%	20	0%	0	0%	0

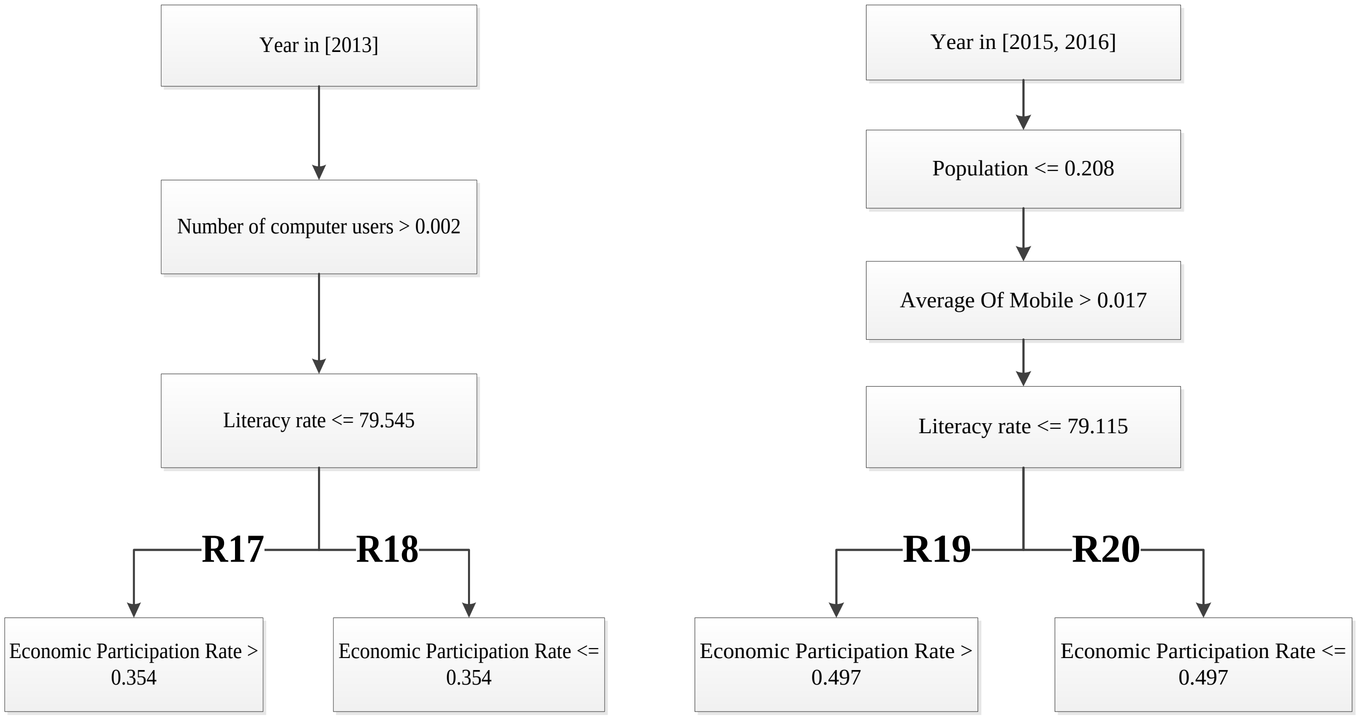


Fig19. The extracted rules of CART decision tree related to cluster 3 POS transactions

A comparison between R17 and R18, and R19 and R20 uncovers the positive effect of economic partnership rate on increasing POS transactions. Economic activity without transferring money is not possible. Obviously, an increase in economic activities will raise money transfer. One of the channels of money transfer is POS terminals. Considering the positive relationship between economic partnership and the volume of point of sale transactions, it can be said that marketing in economy and business field is one of the effective ways to expand e-banking. This field has the high potential for increasing POS terminals and consequently, increasing e-banking transactions. So, it is essential that the banks pay close attention to such field.

According to R17 and R18 rules, Sistan and Baluchistan, Kurdistan, and Eastern Azerbaijan provinces have had more than 0.002 computer users and the literacy rate of less than 79.545 in 2013. Among these three provinces, Sistan and Baluchestan province has had an economic participation rate of less than 0.354 and all of the POS transactions were conducted by high volume. However, the rate of economic partnership in Kurdistan and Eastern Azerbaijan provinces has been more than 0.354 and all of the POS transactions were done with high volume.

Based on R19 and R20 rules, the economic participation rate of Sistan and Baluchistan province in 2015 and 2016 has been less than 0.497, and all of POS transactions were conducted with low volume in this province. And in Kurdistan where economic participation rate has been more than 0.497, all of POS transactions have been conducted with medium volume.

4-Discussions

The applied methods and obtained results are discussed in this section.

4-1-Applied method

The aim of this research was to identify the effective factors in expanding e-payment transactions in Iran. To this aim, Iran's provinces were first divided into three clusters of "more developed provinces", "Tehran province", and "less developed province" based on carried out transactions and demographic features using K-Means algorithm. Also, the transactions of each year were separately clustered, and the formed clusters were compared together then. After the records were clustered in order to discover the effective factors in the volume of conducted transactions in each province, CART decision tree algorithm was run on clusters 1 and 3. The effective factors in the volume of the transactions done through Internet banking, Mobile banking, and POS tools have been investigated.

4-2- The results of cluster analysis

According to the results of clustering, the acceptance of e-payment tools and methods among the people of Tehran is more than other provinces. Therefore, the customers in Tehran are considered as e-banking key customers, suggesting that the banks should pay careful attention to them in order to earn their satisfaction and increase their trust. Hence, delivering different, safe and new services to this group of customers seems necessary. The highest rate of development in using the tools under investigation belongs to Mobile banking method. In fact, the customers’ trust in such a tool has increased by the passage of time, and the customers have displayed more willingness to use this method. Moreover, the highest rate of increase has belonged to the variable of the number of Mobile banking transactions in cluster 3. These results indicate an increase in welcoming Mobile banking method. It can be inferred that focusing on this method of payment will be an appropriate strategic approach in order to expand e-banking. So, advertisements and investment in this field will grow profitability. The leading banks in designing and delivering novel services in the field of mobile-based applications can largely succeed in absorbing new customers to e-banking. Regarding the intense rivalry among the banks in e-banking, achieving success in this field will be considered a competitive advantage for the banks. The increase in the amount of transactions conducted through POS tools has been lower than the increase in the number of transactions through the same tool in all three clusters. As a result, the volume of POS transactions has been on the wane in all provinces. The expansion of using e-payment tools in the provinces of cluster 3, i.e. less developed provinces, in four-year time interval has occurred with the greater percentage than in other clusters. According to the statistics and results obtained from the investigation of the clusters, the number of e-banking customers in the provinces located in cluster 3 is lower than in other provinces, and the need for advertisements and training in the field of e-banking in such provinces is felt more than in other provinces. In fact, these provinces are potential customers of e-banking and have high potential in expanding e- banking. These customers can be absorbed and involved through planning properly and adopting appropriate policies. The change in the behavior and attitude of this cluster’s customers towards e-banking has been considerably positive. Thus, this point can be considered important by the bank managers that the policies of e-banking development have paid off in such provinces, and it is evident that investment in these provinces will produce positive results and increase the profitability and market share of e-banking customers.

4-3-The obtained results of the decision tree rules

According to the rules derived from the decision tree the effective factors in the volume of the transactions done through Internet banking, Mobile banking, and POS tools were investigated and are presented in Fig. According to the figure 20, the development indices of the provinces and the rate of economic activities and business boom in the provinces have had the highest effect in expanding and welcoming electronic banking. the information technology indices of the provinces including the Internet influence coefficient, access to the Internet, and the quality of Internet services have the positive effect on the rate of customers’ welcoming e-payment methods. One of the most important approaches in the expansion of e-banking that must be taken into account by the managers and authorities is proper distribution of the facilities and payment services all around the country and the provision of safe and secure infrastructure to make successful electronic payment transactions. Developing and equipping the facilities and infrastructure of information and communications technology might be costly at first, but it will turn a profit in the long run. The expansion of the banks’ e-services and removal of paper from bank transactions and trades will substantially decrease the costs of production and usage of paper and preserve the environment.

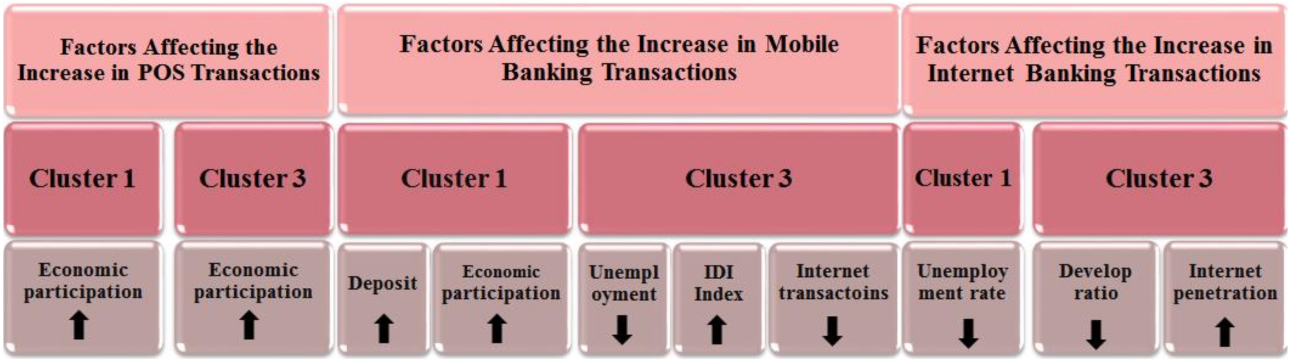


Fig20. Brief results of decision tree rules

Based on the results obtained from the decision tree, another effective factor in increasing warm welcome to e-banking is the increase in economic activities and business boom. Considering the fact that the economic activists have the strong interaction with the banks and the number and amount of their transactions, this group of customers is considered as one of the important and valuable groups of e-banking customers. So, adopting encouraging policies for this group of e-banking customers can have a significant effect on increasing the usage of e-payment services. For example, special facilities could be considered in bank services for those customers with a transaction quantity and/or amount more than a specific number. Providing this group with mobile-based and Windows-based applications for financial and capital management along with analytic tools will have a significant effect on absorbing and maintaining the customers of this field. Another service to be presented to this group of customers is investment Robo-advisors. On the basis of big data analysis, artificial intelligence in banking field can determine investment strategies for the customers, and the customers can start investing by one click using these robots in the user interface.

5-Conclusion and suggestions

Due to competition in the banking industry, banks need to monitor and analyze the behavior of customers. One purpose of the monitoring is to obligate the banks to provide new services to maintain the market. Banks need to analyze their customers' behavior and then extract their behaviors' patterns. The present study aimed to provide a framework for specifying the factors affecting the number and amount of transactions conducted with the E-payment instruments and effective factors in expanding these payment methods. 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. For this purpose, a dataset including 1085 records and the statistics of electronic payment tools such as POS, Internet banking, and Mobile banking in 2013, 2015, and 2016 was received from the Statistical Center of the Central Bank of Iran. Thereafter, the information and statistics related to social, economic, financial, and development indices of information technology of each province were added to the initial data set. Firstly, the K-Means clustering algorithm was applied to separate the transactions conducted in different provinces of the country into three clusters of “more developed provinces”, “Tehran province”, and “less developed province”. Also, the transactions of each year were separately clustered, and the formed clusters were then compared together. For further review, hidden patterns and information in the carried out transactions were discovered with the help of CART decision tree algorithm which was run on clusters 1 and 3. According the obtained results, development indices of the provinces and economic activities and business boom had the highest effect on welcoming and expanding electronic banking. 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.

Conflict of Interest: The authors declare that they have no conflict of interest.

References

Adeoti, O., & Osotimehin, K. (2012). Adoption of point of sale terminals in Nigeria: Assessment of consumers' level of satisfaction. *Research Journal of Finance and Accounting*, *3*(1), 1-6.

Ahmad, N., Alahakoon, D., & Chau, R. (2010). Cluster identification and separation in the growing self-organizing map: application in protein sequence classification. *Neural Computing and Applications*, *19*(4), 531-542.

Auta, E. M. (2010). E-Banking in developing economy: empirical evidence from Nigeria. *Journal of applied quantitative methods*, *5*(2).

Azar, A. T., & El-Metwally, S. M. (2013). Decision tree classifiers for automated medical diagnosis. *Neural Computing and Applications*, *23*(7-8), 2387-2403.

Barker, R. (2018). Knowledge management to prevent fraudulant e-banking transactions. *Communitas*, *23*, 71-86.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Burrell, P. R., & Folarin, B. O. (1997). The impact of neural networks in finance. *Neural Computing & Applications*, *6*(4), 193-200.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide.

Clark, L.A., & Pregibon. D. (1992). Tree-based models. In: Chambers JM, Hastie TJ (eds) Statistical models (chap 9). *S. Chapman & Hall, New York*, pp 377–420.

Cullen, R. (2003). The digital divide: a global and national call to action. *The Electronic Library*, *21*(3), 247-257.

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.

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*, *4*(1), 55-60.

Hargittai, E., & Hinnant, A. (2008). Digital inequality: Differences in young adults' use of the Internet. *Communication research*, *35*(5), 602-621.

He, Z. "Evolutionary K-Means with pair-wise constraints", *Soft Computing*, 20(1), pp. 287-301(2016).

Hsieh, T. C., Yang, K. C., Yang, C., & Yang, C. (2013). Urban and rural differences: Multilevel latent class analysis of online activities and e-payment behavior patterns. *Internet research*, *23*(2), 204-228.

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*, *11*(4), 3764-3770.

Hussain, Z., Das, D., Bhutto, Z. A., Hammad-u-Salam, M., Talpur, F., & Rai, G. (2017). E-Banking Challenges in Pakistan: An Empirical Study. *Journal of Computer and Communications*, *5*(02), 1.

Iyengar, J., & Belvalkar, M. (2010). Case Study of Online Banking in India: User Behaviors and Design Guidelines. In *Human Work Interaction Design: Usability in Social, Cultural and Organizational Contexts* (pp. 180-188). Springer, Berlin, Heidelberg.

Kalakota, R., & Whinston, A. B. (1997). *Electronic commerce: a manager's guide*. Addison-Wesley Professional.

Keramati, A., Ghaneei, H., & Mirmohammadi, S. M. (2016). Developing a prediction model for customer churn from electronic banking services using data mining. *Financial Innovation*, *2*(1), 10.

Khan, B. U. I., Olanrewaju, R. F., Baba, A. M., Langoo, A. A., & Assad, S. (2017). A Compendious Study of Online Payment Systems: Past Developments, Present Impact, and Future Considerations. *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS*, *8*(5), 256-271.

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*, *27*(3), 178-190.

Liébana-Cabanillas, F., Nogueras, R., Herrera, L. J., & Guillén, A. (2013). Analysing user trust in electronic banking using data mining methods. *Expert Systems with Applications*, *40*(14), 5439-5447.

MacIntyre, J. (2013). Applications of neural computing in the twenty-first century and 21 years of Neural Computing & Applications. *Neural Computing and Applications*, *23*(3-4), 657-665.

Mansingh, G., Rao, L., Osei-Bryson, K. M., & Mills, A. (2010). Application of a Data Mining Process Model: A Case Study-Profiling Internet Banking Users in Jamaica. In *AMCIS* (p. 439).

Matthew, N. G., & Laryea, A. E. (2012). A financial performance comparison of foreign vs local banks in Ghana. *International Journal of Business and Social Science*, *3*(21), 82-87.

Ming-Yen Teoh, W., Choy Chong, S., Lin, B., & Wei Chua, J. (2013). Factors affecting consumers' perception of electronic payment: an empirical analysis. *Internet Research*, *23*(4), 465-485.

Mohammad, R. M., Thabtah, F., & McCluskey, L. (2014). Predicting phishing websites based on self-structuring neural network. *Neural Computing and Applications*, *25*(2), 443-458.

Murthy, S. K. (1998). Automatic construction of decision trees from data: A multi-disciplinary survey. *Data mining and knowledge discovery*, *2*(4), 345-389.

Namazi, M. (2015). Predict e-banking customer behavior using data mining, (Case Study Shahr Bank). *2nd International Conference On Modern Research in Management, Economics, and Humanities.* Dubai, United Arab Emirates.

Niyagas, W., Srivihok, A., & Kitisin, S. (2006). Clustering e-banking customer using data mining and marketing segmentation. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, *2*(1), 63-69.

Noori, B. (2015). An Analysis of Mobile Banking User Behavior Using Customer Segmentation. *International Journal of Global Business*, *8*(2), 55.

Özdağoğlu, A., & Güler, M. E. (2016). E-service quality of Internet based banking using combined fuzzy AHP and fuzzy TOPSIS. *Tehnički vjesnik*, *23*(4), 1109-1116.

Özkan, S., Bindusara, G., & Hackney, R. (2010). Facilitating the adoption of e-payment systems: theoretical constructs and empirical analysis. *Journal of enterprise information management*, *23*(3), 305-325.

Ogedebe, P. M., & Jacob, B. P. (2012). E-payment: Prospects and challenges in Nigerian public sector. *International Journal of Modern Engineering Research, 2, 5*.

Palacios, H. J. G., Toledo, R. A. J., Pantoja, G. A. H., & Martínez, Á. A. A comparative between CRISP-DM and SEMMA through the construction of a MODIS repository for studies of land use and cover change. *Advances in Science, Technology and Engineering Systems Journal Vol. 2, No. 3, 598-604 (2017).*

Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers & Geosciences*, *51*, 350-365.

Premchand, A., & Choudhry, A. (2015). Future of Payments–ePayments. *International Journal of Emerging Technology and Advanced Engineering*, *5*, 110-115.

Radfar, R., Nezafati, N., & Yousefi asl, S. (2014). Internet Banking customer classification using data mining algorithms. *Information Technology Management, 6(1), 71-90.*

Thakur, R., & Srivastava, M. (2013). Customer usage intention of mobile commerce in India: an empirical study. *Journal of Indian Business Research*, *5*(1), 52-72.

Roy, S., & Sinha, I. (2014). Determinants of customers' acceptance of electronic payment system in Indian banking sector–A study. *International Journal of Scientific and Engineering Research*, *5*(1), 177-187.

Roy, A. G., & Urolagin, S. (2019). Credit Risk Assessment Using Decision Tree and Support Vector Machine Based Data Analytics. *In Creative Business and Social Innovations for a Sustainable Future* (pp. 79-84). Springer, Cham.

Safarpour, M. (2016). Identification and Ranking the Barriers to Adoption and Development of Electronic Banking in Iran. *Procedia Economics and Finance*, *36*, 374-380.

Salehi, M., & Alipour, M. (2010). E-banking in emerging economy: empirical evidence of Iran. *International Journal of Economics and Finance*, *2*(1), 201.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A. & Lin, C. T. (2017). A review of clustering techniques and developments, *Neurocomputing*, 267, pp. 664-681.

Shih, Y. Y., & Fang, K. (2004). The use of a decomposed theory of planned behavior to study Internet banking in Taiwan. *Internet Research*, 14(3), 213-223.

Shirazi, F., & Mohammadi, M. (2018). A big data analytics model for customer churn prediction in the retiree segment. *International Journal of Information Management*.

Thakur, R., & Srivastava, M. (2014). Adoption readiness, personal innovativeness, perceived risk and usage intention across customer groups for mobile payment services in India. *Internet Research*, *24*(3), 369-392.

Thakur, R., & Srivastava, M. (2014). Adoption readiness, personal innovativeness, perceived risk and usage intention across customer groups for mobile payment services in India. *Internet Research*, *24*(3), 369-392.

Travica, B. (2002). Diffusion of electronic commerce in developing countries: The case of Costa Rica. *Journal of Global Information Technology Management*, *5*(1), 4-24.

Upadhyay, P., & Jahanyan, S. (2016). Analyzing user perspective on the factors affecting use intention of mobile based transfer payment. *Internet Research*, *26*(1), 38-56.

Wang, M., Wang, M., Cho, S., Cho, S., Denton, T., & Denton, T. (2017). The impact of personalization and compatibility with past experience on e-banking usage. *International Journal of Bank Marketing*, *35*(1), 45-55.

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on neural networks*, *16*(3), 645-678.

Zhu, F., Tang, M., Xie, L., & Zhu, H. (2018). A Classification Algorithm of CART Decision Tree based on MapReduce Attribute Weights. *International Journal of Performability Engineering*, *14*(1), 17.