Adv. Math. Fin. App., 2022, **7(**4), P. 1075-1098

Advances in Mathematical Finance & Applications

www.amfa.iau-arak.ac.ir
Print ISSN: 2538-5569
Online ISSN: 2645-4610
Doi: 10.22034/AMFA.2021.1942827.1644

Applied-Research Paper

Developing a Model for Managing the Risk Assessment of Import Declarations in Customs based on Data Analysis Techniques

Hassan Ali Khojasteh Aliabadi[a], Saeed Daei-Karimzadeh[b*], Majid Iranpour Mobarakeh[c], Farsad Zamani Boroujeni[d]

[a] *Department of Public-Financial Management, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran.*
[b] *Department of Economics, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran.*
[c] *Department of Computer Engineering and IT, Payam Noor University, Tehran, Iran.*
[d] *Faculty of Engineering, Department of Computer, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran.*

ARTICLE INFO
Article history:
Received 2021-10-18
Accepted 2021-12-29

Keywords:
risk
risk management
data analysis
customs
import declaration

ABSTRACT

In customs management, the main problem is balancing the needs of trade facilitation as a process of simplifying and accelerating foreign business on the one hand and countering illegal trade, reducing government revenue, capital sleep and the level of controls and interventions on the other. Also, due to the financial crisis in recent years, risk management has been reconsidered, although this attention is related to various financial branches. Since risk analysis and identification is the main component of risk management, developing a suitable model for data analysis is of particular importance. The purpose of this study was to use data data analysis techniques to develop an intelligent model to timely predict the risk of import declarations in customs and thus prevent irreparable losses. In this study, data analysis techniques have been used according to the statistical population which is data-driven. Statistical data were extracted from www.eplonline.ir with 575006 import declarations of all Iranian customs during 2019-2020. having pre-processed and prepared the data using PCA, LDA and Fast ICA methods, attribute reduction and effective attribute extraction were performed using 14 data analysis algorithms. Using Python software, algorithms were trained and modeled with 80% of the final data. Then, 14 obtained models were tested and validated with 20% of the data. Finally, the results of these models were compared with each other and the model obtained from the random forest algorithm was selected as a comprehensive model for predicting and determining the level of risk of import declarations at customs.

1 Introduction

Customs offices around the world are responsible for implementing governments' extensive policies on revenue collection, compliance with trade laws and regulations, community protection, cultural heritage, intellectual property rights, statistics collection, and environmental protection [50]. On the

* Corresponding author. Tel.: +98-913-329-8618
E-mail address: *saeedkarimzade@yahoo.com*

other hand, in recent years, customs have faced emerging and complex threats [47]. The most prominent threats are non-compliance with customs laws and illegal transactions which affect government revenues on the one hand and threaten the security of society on the other hand[38]. Because of declining customs revenues due to economic integration, tariff reform processes, economic cooperation agreements and large-scale fraud, customs need to use effective customs controls to detect financial fraud in order to increase revenue [55]. These controls cause significant delays in trade [4]. Customs delays also have a negative impact on the foreign sales of companies and importers. These effects are more severe for time-sensitive goods. On the other hand, customs delays increase the cost of capital sleep and reduce government revenues [43]. Therefore, customs control methods should be based on the use of modern effective methods of combating fraud, protecting the health of consumers and the safety and promotion of legitimate trade [4]. Also, due to the financial crisis in recent years, risk management has been reconsidered, although this attention is related to various financial branches. The crisis, even for non-financial organizations, identified special importance: identifying, measuring and overcoming risk in order to maintain their existence and continue their work [24]. WTO studies show that developing countries that have benefited from risk management have had higher economic growth and development than other developing countries [52].

The experience of the World Customs Organization also shows that a risk-based approach to customs controls, while creating a balance in facilitating trade on the one hand and security on the other, leads to the efficiency of human resources [38]. Therefore, the necessity of using risk management tools, especially in organizations such as customs which is responsible for collecting import and export duties and taxes, as well as import, export and transit of goods is strongly felt due to the sensitivity of the profession[6]. Therefore, risk management needs to adopt an appropriate strategy for each category of customers in accordance with the risk level and predict customer performance in the future [30]. Given that databases hold all customer transaction data, it is possible to review them to provide a suitable model to predict customer performance in the future [34]. The importance of data analysis has increased in the last decade due to the increasing development and use of information technology and increasing access to large databases [17]. Traditional statistical methods have lost their effectiveness today for two reasons: Firstly, increasing the number of observations and secondly, increasing the number of variables related to an observation [33]. One of the best ways to extract customer behavior patterns is to use data analysis algorithms. Data analysis algorithms use statistical methods and artificial intelligence to extract patterns from very large sets [34]. Due to the large number of customs exchanges and the multiplicity of risks, risk analysis has not been performed sufficiently to help identify crimes at customs. New challenges must evolve to solve this problem [55]. Therefore, risk analysis seems to be a priority for the modernization of customs systems in developing countries [38].

A review of the background of domestic and foreign studies shows that many studies in the field of risk assessment and classification in various fields of banking, insurance, stock exchange, medicine, etc. have been done using data mining techniques. But there is little scientific research in the field of customs risk management. Due to the lack of comprehensive scientific and research studies in the field of customs risk assessment which is the most important element of customs risk management, effective measures were taken in this study. Considering the launch of Iran's customs cross-border trade system from the end of 1392 and the creation of an appropriate database and the obligation to refrain from judging decisions regarding the risk of import declarations, the researcher decided to address this issue. Also in the target community, in the customs offices of the World Customs Organization, data mining techniques have rarely been used in customs risk assessment management. Therefore, research is innovative in this regard. The process of data mining and risk assessment in this study is innovative according to the proposed model and research method used and has a favorable range in the use of 14 various algorithms as well as providing various models. Various methods were used to understand the data and select effective features and prepare the data. Fourteen different data mining techniques were modeled and validated by three methods: PCA, LDA and Fast ICA. 42 models were made. The results of the models were compared with the desired techniques and the superior model was selected. Using the selected model, the risk of imported declarations in Iranian customs can be assessed online and using the obtained codes.

2 Research Literature

2.1 Risk in Customs

According to Longman culture, risk is defined as " the possibility that something bad, unpleasant, or dangerous may happen" . Risk is the uncertainty about future events, and in fact, risk is a type of uncertainty about the future that can be calculated [24]. Increasing international trade and the speed of exchanges and the high volume of financial transactions in the absence of modern systems for detecting fraud and lack of control and inspection resources, increasing violations of customs laws and fraud, financial and non-financial threats have caused:

-Financial threats typically relate to evasion of import duties and taxes on imported goods and are generally related to the entry of goods into the customs territory, such as intentional misrepresentation (smuggling), non-declaration of imported goods by the declarant, incorrect description or declaration, incorrect report of the origin of the goods and under-declaration of value [53]. According to the Green Line Institute in 2010, non-financial threats include terrorism, counterfeiting, money laundering, migration, human diseases, dangerous goods, strategic goods, environmental threats, endangered species, and environmental threats. Risks are usually divided into two categories: systematic risk (extra-organizational factors) and non-systematic risk (intra- organizational factors). operational risk is the most important type of non-systematic risk being specific to the organization [39].

2.1.1 Operational Risk

Operational risk arises from the failure of internal systems or the error of the people who run the organization. The British Banking Association defines operational risk as: Operational risk is the risk of direct or indirect losses resulting from the inadequacy or inefficiency of internal processes, individuals and systems, or external events. In another definition of operational risk, Professor Simmons also considers it to be the result of a defect in the core operations, construction, or processing capacity of an organization. Examples of operational risks include computer system malfunctions, management errors, misuse and fraud of documents, theft and embezzlement, settlement problems, intentional and unintentional human error, and fraud (A situation where traders give wrong information) as well as technical error [24].

2.2 Risk Management

Risk management is a preventive activity protecting the organization from taking advantage of opportunities and preventing the occurrence of threats and spending too much time and resources by relying on standard methods [6]. The American Project Management Knowledge Standard defines risk management as the process of maximizing the results of positive events and minimizing the likelihood of the occurrence or impact of adverse consequences on project objectives.

2.2.1 Risk Management in Customs

Increasing changes in the strategic outlook of the customs operating environment along with the increasing growth of trade and volume of travel have affected the way customs is managed and approached. These conditions along with increasing uncertainty have led many of the customs around the world to pursue the plan and implementation of structured risk management. Today, customs risk management has become one of the key tools in meeting the requirements of the operating environment [8]. Customs risk management is a topic that is studied in the field of laws and a way to effectively avoid customs risk. Its main purpose is to combine the limited resources of customs and the best effect of risk control with the optimal combination of skills such as risk management planning, risk analysis and identification, risk quantification and ranking, risk aversion, risk monitoring and control [16].

Many global agreements are aimed at implementing customs risk management to expedite trade; agreements such as the Agreement on the Transfer of Low-Risk Goods and Supply Chain Facilitation within the framework of the standards of safe and easy trade of the World Customs Organization and the revised Kyoto Convention. Risk management is also one of the ten principles of the action plan prepared by the World Customs Organization under the title of Customs Strategy in the 21st Century which outlines the customs strategy for the future. Items such as the World Customs Organization Risk Management Guidelines, Global Information Strategy, Risk Standards Assessment, the World Customs Organization High Risk Index Guidelines and e-learning materials, as well as Article 7 of the 1994 GATT Agreement are the legal basis for customs risk management. The relationship between risk management components is shown in Fig. 1.

2.2.2 Risk Assessment

ISO 1998 defines risk assessment as a logical way to quantify the quantitative and qualitative risks and to assess the possible consequences of potential accidents on people, materials, equipment and the environment. In fact, it determines the effectiveness of existing control methods and provides valuable data for decisions to reduce risks, hazards, improve control systems and plan to respond to them. It is also one of the pillars of risk management and its purpose is to measure risks based on various indicators such as impact and probability of occurrence and the more accurate the results of this step can be said that the risk management process is performed with a higher degree of confidence. As a result, the decision maker can plan for the amount of resources available to deal with each risk [29]. According to Richard. E. Olson, in high-risk systems, there is a strong tendency for complete reliability of statistical probabilities because using a number seems to be an easy way to measure safety and the probability of a defect or accident, but it is necessary to try to Such measurements, limitations and basic principles of such methods along with previous experience are well known and quantitative acceptance parameters are fully defined and predictable, provable and most importantly useful, meaning that they can be easily converted into design criteria. On the other hand, it is possible for a system security analyst and its management to become so interested in statistical methods that they forget other very convenient and important ways of expressing the problem [24].

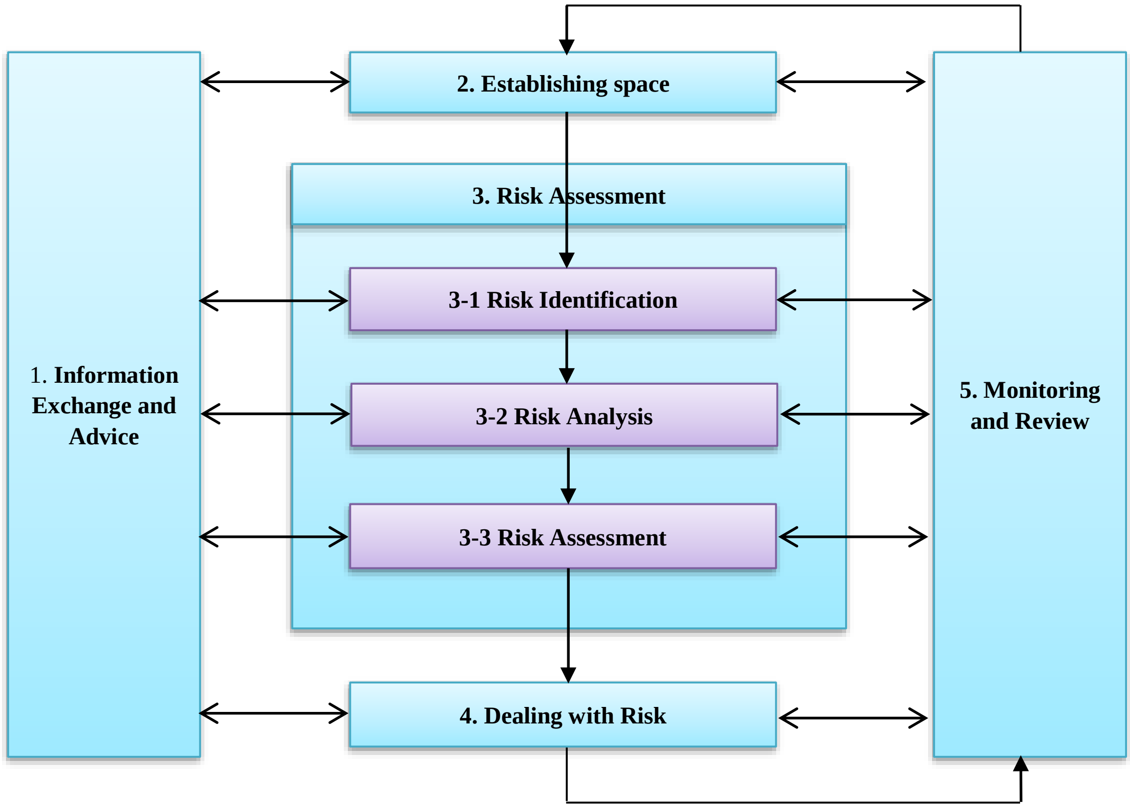

Fig. 1: Risk Management Process in the International Standard of Risk Management System ISO 31000

2.2.3 Developing the Risk Policy

Defining and explaining the specific risk of organization is the first step in developing a risk policy for an organization. The second step is to determine the level of risk aversion based on the willingness of management and others to take the risk or how much the organization wants to be exposed to risk. The third step is to identify the goals of the risk management policy. At this stage, the organization's senior manager must decide which areas are so important to apply risk management and how much resources should be consumed. The last step is to identify the areas in which risk management is required, as well as to identify the individuals or groups responsible for risk management [52]. The policy of implementing risk management in customs processes allows the application of different kinds selective controls including complete control and evaluation of the declaration to unannounced inspection, referred to as red, yellow and green channels.

-Red channel: Documentary examination and physical inspection of good. Declarations in this direction have a very high level of risk. In addition to the documentary inspection, there is also a physical inspection, meaning that all control measures have been taken in this direction, and in addition to the fact that the mentioned document are thoroughly examined, the goods are observed physically and the accuracy of the information contained in the declaration is carefully checked.

-Yellow channel: Documentary examination. The yellow path conveys the notion that existing declarations have a medium level of risk. It is inspected by documents, but there is no need for physical inspection and direct viewing of the goods, and as it is clear, this route is in terms of priority after the red route. Sampling is also done in this direction if necessary.

-Green Chanel: Immediate discharge without examination. The declarations that are placed in this direction do not seem to pose any risk to customs and have a very low level of risk. Therefore, the declarations of this route are only briefly reviewed. In this route, we are faced with a delay option that indicates a time delay in terms of hours, for the automatic evaluation of these declarations, which after the expiration of this time in case failure to review by the expert, the declaration is automatically evaluated by the system [11,56].

2.2.4 Conditions for Implementing the Risk Management Strategy

Acceptability: The determined strategies must be meaningful and acceptable to the people who implement them.
Compatibility: Strategies must be sensible in the context of the stated values and goals of management and have a logical flow from period to period.
Quality: The effect of strategies on improving management decisions should be seen [53].

3 Data Analysis

Data analysis is defined as the process of discovering meaningful connections, new patterns and trends by sifting large amounts of data stored in data warehouses using pattern recognition technologies along with statistical and mathematical methods [37]. Data analysis helps organizations explore and predict future patterns and behaviors by exploring the data in a system. By analyzing past events, data analysis provides an automatic and predictive analysis and answer questions that have not been possible or required to be answered in the past [44]. In the risk management cycle, data analysis can reduce the pressure from low-risk customers to high-risk customers by segmenting and analyzing customers. Data analysis may answer the following questions:

- Can certain patterns of unknown data be formed?
- Can big data analysis identify patterns?
- Can classification algorithms be used to improve and create a risk management system to control infringing importers, organized crime, smuggling forecasting, revenue collection and electronic management and influence the customs administration?

By analyzing the data stored in customs databases, it is possible to better identify customers and increase the efficiency of customs risk management by focusing and optimally allocating resources in high-risk areas. In order to use data analysis, Fayad et al. suggested the following steps in Fig. 2.
1. Extracting data from a large set of databases,
2. Selecting work-related subset,
3. Deciding on appropriate sampling methods, deleting lost data and sorting data,
4. Using a suitable method for data preprocessing,
5. Creating a model of preprocessed data [18].

In 1996, also, the standard CRISP-DM cross-industry process for data mining was developed by analysts by DaimlerChrysler, SPSS and NCR. CRSIP-DM stands for Cross-industry standard process for data mining and means "cross-industry standard data mining process". The Crisp process is a standard open source process model that describes the general approaches of data mining professionals. This method is one of the most widely used analytical models. According to Crisp data analysis method, a life cycle project consists of six steps. These successive steps begin with understanding the basic needs of the business and end with the presentation of useful solutions. These steps are applied continuously and repetitively throughout the data analysis process in Fig. 3.

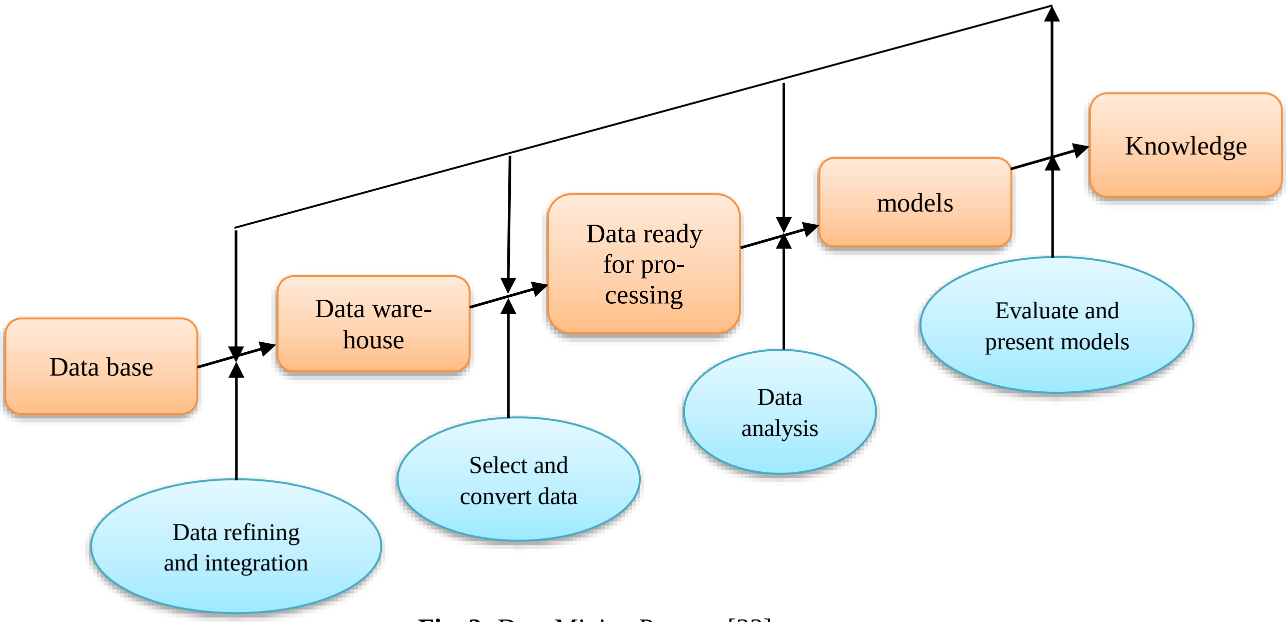

Fig. 2: Data Mining Process [33]

As mentioned, the CRSIP-DM method is a data-driven process model that provides a systematic, useful, and practical solution for data dimensions as well as data clustering. Since clustering is one of the most important and fundamental issues in data mining, it is good to mention clustering briefly. A cluster is a collection of similar data, and clustering is the process by which you can break a set of objects into separate groups.

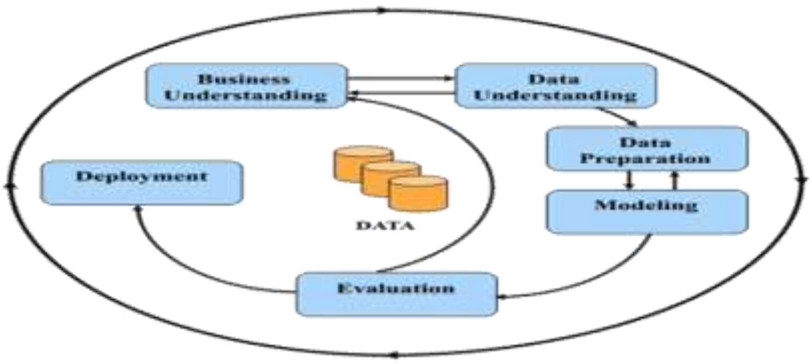

Fig. 3: Crisp Methodology [15]

4 Research Method

According to the above methods, the method proposed in this study consists of several steps which its general outline is shown in Fig. 4.

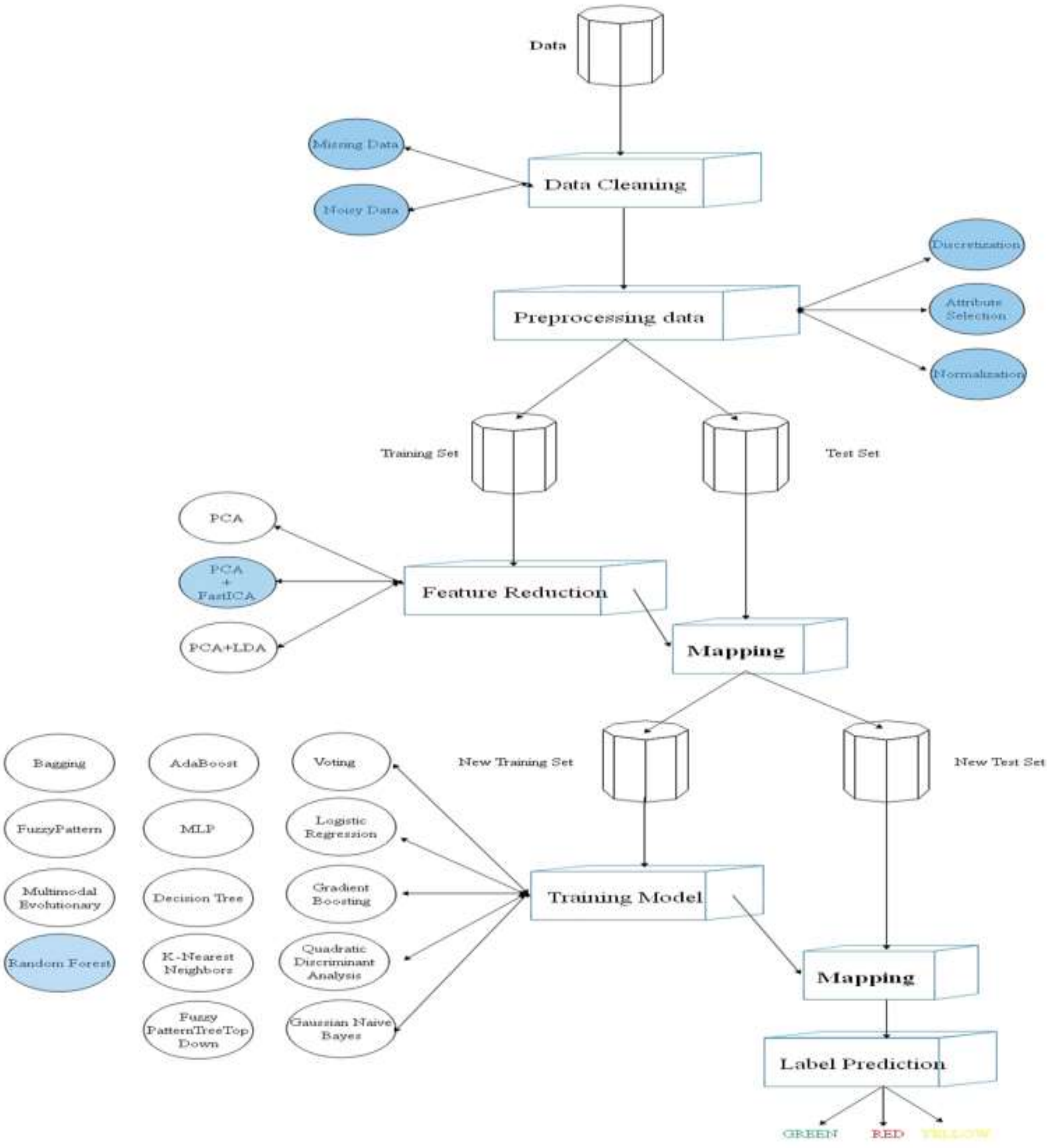

Fig. 4: Research Implementation

4.1 Data Collection and Comprehension Step

One of the main parts of any decision making is data collection and data comprehension. If this is done in a regular and correct way, the work of analyzing and concluding the data will be done with good speed and accuracy.

4.1.1 Data Collection Step

First and foremost, we need to focus on the data. In this step we are looking for data related to our work (data for analysis). Data are collected based on the information contained in the data mining query. Because a data mining query is usually only relevant to a portion of the database, selecting relevant data will not only make the search more efficient, but the results will be more productive and meaningful than exploring the entire database [33]. The present study includes 575006 data related to the information of import declarations during 2020-2019 extracted from www. eplonline.ir.

4.1.2 Data Comprehension Step

Choosing right attributes is one of the important factors in creating a model is. The first and most important step in categorizing customers based on risk is to identify risk factors. It may be difficult to identify specific set of related traits. Thus, the researcher may select some specific attributes that feels are important and omit some other specific attributes that may play a role in the description. In order to identify the effective characteristics and better understand the data and form a risk profile in this research, identification is performed in two phases.

4.1.2.1 Studying and Reviewing the Existing Research

In this section, based on the existing studies, the mentioned items are divided into revenue groups (collecting customs duties), money laundering, documentary issues, customs procedures, transportation, demographics and the history of the owner of the goods and the declarant. Tariff changes, devaluation in order to avoid paying taxes and duties, change of certificate of origin, import of unsanitary and non-standard goods are of cases that can be cited as potential risks or obstacles to customs and prevent the implementation of regulations [56]. Laporte (2011) includes risk profiles for six different criteria based on the level of violation, such as importer, transport agents, HS nomenclature (tariff book), certificate of origin, health and safety, importer / exporter, and customs procedure. Commercial fraud includes declaring the value of goods, the tariff on goods, the quantity of goods, the country of origin or destination, tax violations, the correct deposit of guarantees [51].

4.1.2.2 Survey of Experts

After extracting the characteristics from the existing studies, a questionnaire was designed and completed by experts who had sufficient experience and knowledge about the damages and customs violations to determine the characteristics. The statistical population was 2496 people selected from the total customs of Iran. The 333 people were selected as sample for distribution of the questionnaire by Cochran's formula and the questionnaire was distributed by random sampling. Banner (2006) used the Cattell test method to select the final factors and analyze the opinions of experts. Cattell test is a criterion that can be used to extract the appropriate factors. Cattell test actually depicts the eigenvalues and values of each factor on the chart from ascending to descending. This determines the breaking point in the selection of factors. The most important factors including value, tariff, country of origin, number of goods and customs procedure were identified.

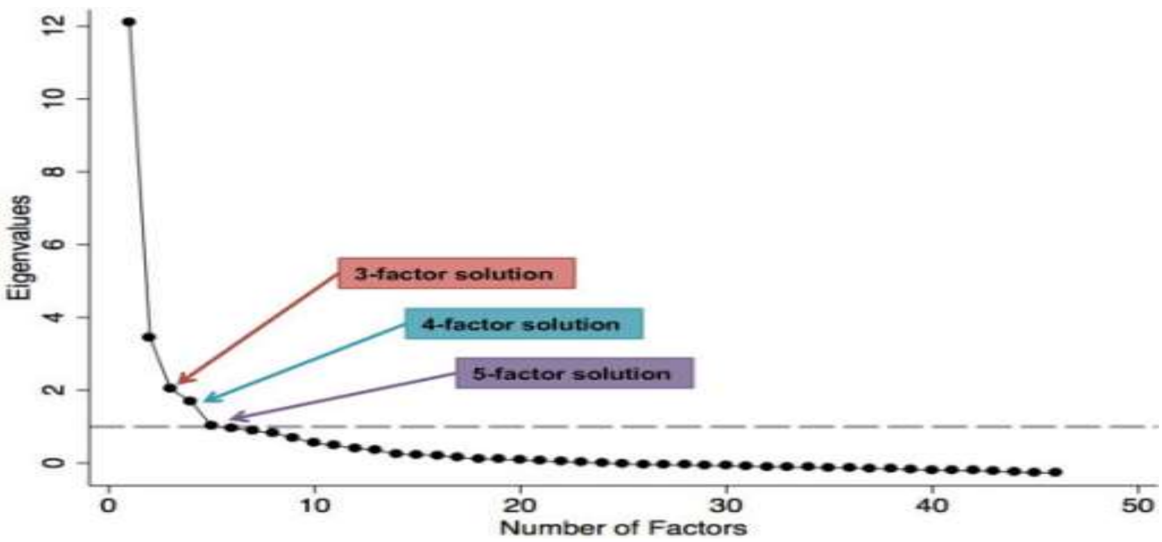

Fig. 5: Cattel Test Curve

Due to the nature of data analysis and the fact that other variables of the declaration of import of goods may include a specific pattern, finally, so the risk profile in this study includes the titles of variables used in import declarations, the same information required to complete the international declaration called SAD1 (administrative unit document) in www.eplonline.ir (cross-border trade unit) [20]. For this reason, all available information is used, that is the contents of approved import declarations and inspection results during a reference period [26]. The risk profile is based on the Table 1.

Table 1: Variables in the Customs Declaration before Data Preparation

Type of transport	Input customs	Product Type	The declarant	product owner
Value components	Customs procedure	Country of origin	Total invoice amount	Country of the transaction
Terms of delivery	The value of the product pen	Eight-digit tariff code	Source of input rights	Customs value of goods item
Input rights	payment method	Number of units of goods	Gross weight	net weight

4.2 Data Preparation Step

In the book Data Preparation for Data analysis, Dorian Pyle estimates that data preparation time is 60% of the total data analysis process time [37]. Data preparation is conduscted in three steps.

4.2.1 Data Labeling

At this step, import declarations were labeled in three levels: green (low risk), yellow (medium risk) and red (high risk).

4.2.1.1 Red Label

In order to identify import labels with a red label with a high level of risk, the customs need to identify the declarations that have led to violations in a previous period (for example, during the last twelve months) [38]. In this regard, 63321 declarations were identified. Declarations consisting of violations like miscalculation in order to evade customs duties, false statement, incorrect entry of tariff code, declaration of prohibited goods, under-declaration, over-declaration, change of certificate of origin, abuse of exemptions, abuse of laws entrepreneurship (such as CKD & SKD goods), smuggling of goods, security issues, health issues, money laundering that have been discovered and violated by evaluation, virtual expertise, inspection, inspection, security and auditing were verified after clearance.

4.2.1.2 Yellow Label

In order to identify import declarations with a yellow label with a medium level of customs risk, identifying declarations that have been examined in the previous period on suspicion of violation is necessary, but their violation has not been proven. In this regard, 189041 declarations were identified.

4.2.1.3 Green Label

In order to identify low-risk green label import declarations, other declarations that have not been placed in either high-risk or medium-risk levels in a previous period are reviewed. Thus, 322644 declarations were identified.

4.2.2 Data Pre-Processing

All data are not of the required quality and cannot be modified and are not eligible to enter the final model. Factors such as data defects, the presence of irrelevant data, data management whose main fields are missing or incorrectly recorded that may include a specific pattern, data duplication, outdated data (some statistical methods are sensitive to the presence of such values and may result in instability) and elimination of correlated variables emphasizing on a particular predictor component at the expense of decreasing the effect of other variables. Therefore, the raw data is pre-processed before modeling to have acceptable quality.

In the first step, the data set contains 575006 samples of import declaration data including 20 attributes. All data are not of the required quality and cannot be modified and are not eligible to enter the final model. After applying the data screening phase and deleting noisy data, the final data set was reduced to 574843.

Table 2: Data Set after Data Preparation

<table>
<tr><th>Method</th><th>Number of features</th><th colspan="3">number of samples</th></tr>
<tr><td>Data</td><td>20</td><td colspan="3">575006</td></tr>
<tr><td>Data after Data Cleaning</td><td>16</td><td colspan="3">574843</td></tr>
<tr><td>Data after Data Transformation</td><td>16</td><td colspan="3">574843</td></tr>
<tr><td rowspan="3">Number of samples in each class</td><td rowspan="3">16</td><td colspan="3">574843</td></tr>
<tr><td>Red channel</td><td>Yellow channel</td><td>Green Channel</td></tr>
<tr><td>63292</td><td>188988</td><td>322563</td></tr>
</table>

In the pre-processing phase, the number of attributes was reduced from 20 to 16 by removing similar attributes. In this step, 4 similar attributes have been removed. It should be noted that the attribute, like the tariff code is extracted based on the tariff book (HS eight-digit code). After preparing the data, the number of samples is 322563 samples for the green label, 63292 samples for the red label and 188988 samples for the yellow label (Table 2). Many of the data in the Import Declaration section contain non-quantitative or non-discrete data. In this regard, this type of attribute was identified and discretization was performed on the data. Also, the types of normalization methods were evaluated and the extracted attributes were normalized in the range of zero to one, green label with number zero label, red label with number one label and yellow label with number two label, respectively.

4.2.3 Reducing Attributes and Extracting Effective Attributes

According to the application of risk management, also, three methods of PCA principal component analysis, LDA linear differential analysis and fast independent fast ICA component analysis are used to reduce the attribute to extract effective attributes and evaluate these types of attributes.

4.2.3.1 Principal Components Analysis (PCA)

PCA is a multivariate statistical analysis that selects a smaller number of factors as the principal components from among the primary factors so that several insignificant data are removed. In the first extracted basic component, the maximum amount of data scatter is in the entire data set, i.e. the first component is correlated with at least several variables. The second extracted component has two important features: this component considers the largest data set not calculated by the first component, and it is not correlated with the first component. In other words, regardless of the previous component, by passing from the initial component to the final components, each component describes less variance. It means that the first principal component always describes the maximum amount of variance and the last components describe the least variance; thus, such information will not be lost by deleting the last components [33].

4.2.3.2 Liner Differential Analysis (LDA)

LDA is a statistical method to reduce the dimensions of an issue and identify categories by maximizing the ratio of scatters between groups to scatters within groups. The linear diagnostic analysis approach is similar to and borrowed from the method used by Ronald Fisher to determine the degree of differentiation in groups. It became the basis for variance analysis and thus is sometimes called "linear differential analysis". The linear diagnostic analysis is very close to variance and regression analyses; all three statistical methods model the dependent variable as a linear combination of other variables. However, variance and regression analyses take the dependent variable as an interval one, while linear differential analysis is used for nominal or ordinal dependent variables. Therefore, linear differential analysis is more similar to logistic regression. The linear diagnostic analysis is also similar to principal components analysis and factor analysis which are used to linearly combine variables in a way that best describes the data. A major application of both of these methods is to reduce the number of data dimensions. However, these methods differ significantly: in linear differential analysis, class differences are modeled, while in principal component analysis, class differences are ignored [37].

4.2.3.3 Fast Independent Component Analysis (Fast ICA)

This is an efficient algorithm to detect underlying factors or components of multivariate data dependent on signal separation. Fast ICA is an optimized method of independent component analysis. This method shows a faster convergence of results than the independent component analysis method. Fast ICA is based on a fixed point algorithm that has a known performance speed. Compared to conventional fixed point algorithms, this algorithm has been modified to provide higher performance. It is also similar to some neural algorithms, and is computationally simple, and requires less computational memory [33].

In the proposed method, PCA method is used to evaluate the attribute. The number of attributes has not been reduced for this purpose and the same 16 attributes have been included. For LDA and Fast ICA methods, only 2 and 3 attributes were used, respectively (Table. 3). In these steps, a number of properties have been selected according to special vector diagrams. Equation (1) is also used to select the number of attributes.

$$\frac{\sum_{i=1}^{k}\lambda_i}{\sum_{j=1}^{f}\lambda_j} \geq 0.98 \tag{1}$$

In this relation, k is the number of selected properties, f is the total number of properties, and λ is the values of the special vector.

Table 3: LDA and Fast ICA Methods

Method	Number of features	number of samples	
Data after Feature Reduction	PCA+LDA, 2 Feature		
	PCA + Fast ICA, 3 Feature		
Sample segmentation	16	test %20	training %80

At this step, the test and training data sets were mapped to the same situation according to the weights extracted from the training set which resulted in a noticeable superiority and the choice of PCA + Fast ICA method over the other two methods. After applying the attribute reduction methods, new training and experimental datasets were created.

4.3 Data Modeling

Machine learning includes a range of analytical tools that can be classified as "supervised" and "unsupervised" learning tools [14]. At this step, supervised learning tools are used. The modeling process is also a two-step process which in the first step the model is created by a set of training data that identifies the class label of all its examples. This step is known as the learning or training step. In the second step, by experimental data set in which the class label is usually unknown, the obtained model is validated or tested. In fact, the evaluation of the model is calculated according to the class of how many experimental data samples are correctly estimated [37].

4.3.1 Modeling Using Educational Data

This step is known as the learning or teaching step. Due to the high volume of data in this research, the software used is Python software. In this section, 80% of the data were allocated by random sampling for training and the training was repited 10 times. According to machine learning methods, 14 algorithms including Decision Trees, Gaussian Naïve Bayes, Fuzzy Pattern, Random Forests, Logistic Regression, k-Nearest Neighbors, Ada Boost, Voting, Gradient Boosting, Bagging, Multimodal Evolutionary, MLP, Fuzzy Pattern Tree Top Down, Quadratic Discriminant Analysis for risk management in the training were used and separate model was made for each algorithm. It should be noted that different combinations of PCA, LDA and Fast ICA methods have been used at this step. As mentioned before, it is noteworthy that the import declarations were divided into three categories according to the level of risk: green with label number zero, red with label number one and yellow with label number two. Each category represents the level of risk. In this regard, the training models are for the three mentioned outputs.

4.3.2 Validating (Testing) Models

The obtained models are validated using an experimental data set in which the class label is usually unknown. Therefore, the validity and predictive power of the models are tested by experimental data without class labels (green, yellow, and red) that they have not encountered before. In this section, 20% of the data were assigned for the allocation test due to random sampling and the test were repeated 10 times. For this reason, experimental data are considered as data for training models.

4.4 Evaluating and Selecting Model

Many questions may arise after creating a model. Questions such as:

- How accurate is the model for predicting importers' behavior?
- What is the accuracy of a model and how is it estimated?
- How can a reliable estimate of the model be obtained?
- With more than one model, how do you choose the best model among others?

4.4.1 Metrics to Evaluate the Efficiency of Import Declaration Models

- Accuracy: The accuracy of a model on experimental data set is defined as the percentage of data in that set that is correctly labeled by the model.
- Precision: This measurement is considered as a correctness measurement. Percentage of data that is labeled positive (correct) and their class is really positive (correct) but does not focus on data that is true and incorrectly labeled.
- Recall: It is a measure of integrity. Percentage of correct data that is correctly categorized but does not focus on data that is incorrect and labeled as correct.

Speaking generally, precision and recall measures are inversely related to each other, and increasing one may decrease the other. So the solution to using them is to use the F1-score.

- F1-score of combination of recall and precision : The measure F is the harmonic mean of the two measures of precision and recall.

4.4.2 Comparison of Models based on ROC Curves

They are a useful visual tool for comparing two models. The ROC curve drawn for a model shows the relationship between the correct positive ratio and the false positive ratio. The farther the ROC curve of a model is from the diagonal line, the higher the accuracy of the model and vice versa [33].

Table 4: Analysis of the Findings of Applying Different Models with 16 Features

Classification method	Accuracy	macro a vg	Label	Precision	Recall	f1-score
Voting(Random Forest, Bagging, Fuzzy Pattern)	66	54.5	0	61	60	61
			1	68	88	77
			2	50	8	14
Bagging	68	60	0	73	57	64
			1	69	90	78
			2	44	17	24
Fuzzy Pattern	51	41.5	0	36	59	45
			1	65	61	63
			2	23	10	14
Multimodal Evolutionary	38	39	0	35	42	38

Table 4: Analysis of the Findings of Applying Different Models with 16 Features

Classification method	Accuracy	macro a vg	Label	Precision	Recall	f1-score
			1	60	34	44
			2	21	44	29
Random Forest	69	58.5	0	78	58	67
			1	69	92	79
			2	48	13	21
Fuzzy Pattern Tree Top Down	60	43.5	0	56	21	30
			1	60	95	74
			2	52	1	1
K-Nearest Neighbors	68	61	0	64	72	68
			1	76	77	77
			2	42	33	37
Decision Tree	53	47	0	48	49	48
			1	67	62	64
			2	26	31	28
MLP	66	52	0	71	48	57
			1	66	93	77
			2	41	5	10
Ada Boost	62	50	0	59	42	49
			1	63	91	75
			2	52	2	4
Gaussian Naive Bayes	57	44	0	42	50	46
			1	63	78	70
			2	44	2	4
Quadratic Discriminant Analysis	59	46.5	0	46	52	49
			1	64	82	72
			2	45	2	4
Gradient Boosting	61	46	0	60	29	39
			1	61	94	74
			2	61	1	3
Logistic Regression	64	51	0	70	39	50
			1	63	95	76
			2	59	2	4

5 Findings

5.1 Findings from Fourteen Models using Principal Component Analysis with 16 Features

Findings obtained of the test of 14 research models in Table 4 using PCA principal component analysis method with 16 attributes showed that the risk prediction accuracy of the random forest model is 69% higher than other models in this method. As can be seen in ROC curve in Fig. 6, the comparison of the findings of 14 research models using the PCA principal component analysis method with 16 characteristics, larger area under the ROC curve of the random forest model than other models indicates shows the higher accuracy of the risk level prediction of this model than other models.

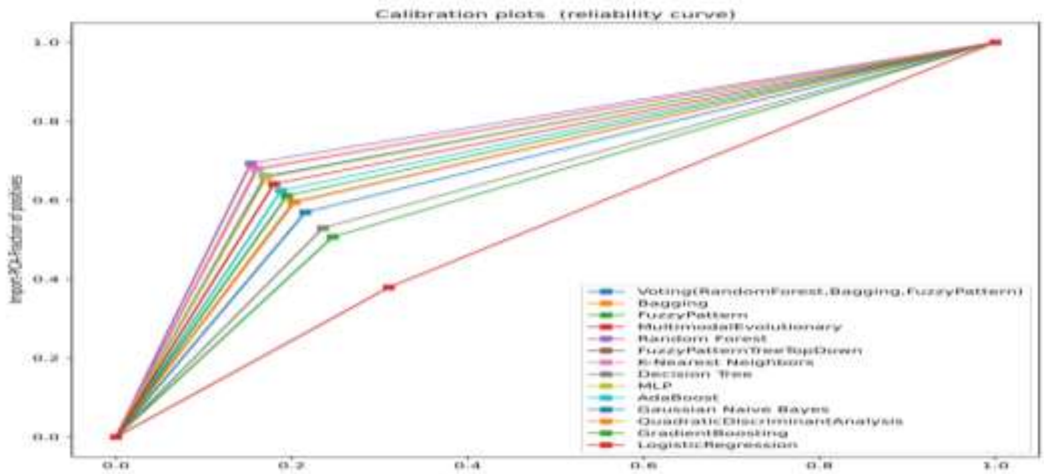

Fig . 6: ROC Curve

5.2 Findings from Fourteen Models using Fast Independent Component Analysis Method with 3 Features

Summary of the findings of 14 research models in Table 5 using the fast independent component analysis method with 3 features showed that the accuracy of risk prediction of the logistic regression model is 61% higher than other models in this this method.

Table 5: Analysis of the Findings of Applying Different Models with 3 Features

Classification method	Accuracy	macro a vg	Label	precision	recall	f1-score
Voting(Random Forest, Bagging, Fuzzy Pattern)	57	43	0	45	39	42
			1	63	80	70
			2	26	9	13
Bagging	55	43	0	45	34	38
			1	62	78	69
			2	25	13	17
Fuzzy Pattern	39	38.5	0	33	60	42
			1	61	35	45
			2	22	27	24
Multimodal Evolutionary	44	31	0	20	2	3
			1	54	66	59
			2	21	30	25
Random Forest	57	42	0	47	29	36
			1	61	85	71
			2	26	9	13
Fuzzy Pattern Tree Top Down	57	33	0	40	9	14
			1	58	96	73
			2	6	0	0
K-Nearest Neighbors	50	41	0	36	44	39
			1	63	63	63
			2	24	17	20
Decision Tree	46	39	0	33	35	34
			1	61	59	60
			2	21	22	22
MLP	59	37	0	49	23	32

Table 5: Analysis of the Findings of Applying Different Models with 3 Features

Classification method	Accuracy	macro a vg	Label	precision	recall	f1-score
			1	60	93	73
			2	0	0	0
Ada Boost	59	37.5	0	54	23	32
			1	60	94	73
			2	0	0	0
Gaussian Naive Bayes	57	33	0	41	12	18
			1	59	95	72
			2	0	0	0
Quadratic Discriminant Analysis	57	34	0	39	17	23
			1	59	93	72
			2	0	0	0
Gradient Boosting	58	25	0	0	0	0
			1	58	100	73
			2	0	0	0
Logistic Regression	61	43	0	62	23	34
			1	61	96	74
			2	32	0	0

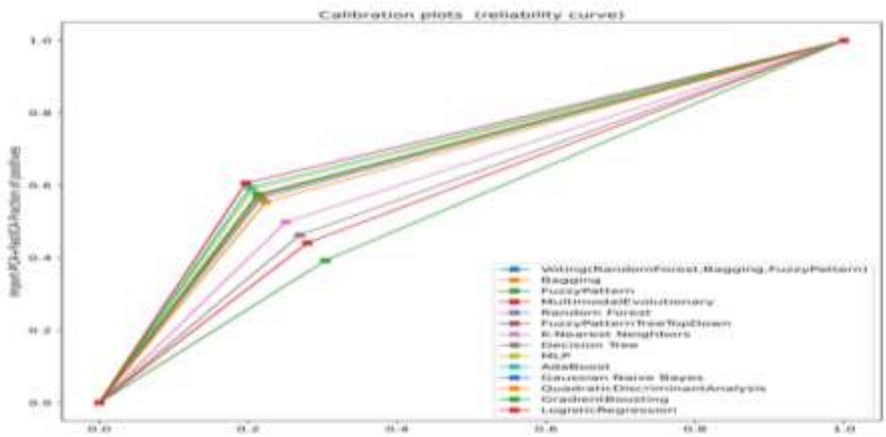

Fig. 7: ROC Curve

As can be seen in ROC curve in Fig. 7 , the comparison of the findings of 14 research models using the fast independent component analysis method with 3 features, larger area under the ROC curve of the logistic regression model than other models indicates shows the higher accuracy of the risk level prediction of this model than other models.

5.3 Findings from Fourteen Models using Linear Differential Analysis Method with 2 Features

Summary of the results of testing 14 research models in Table 6 using the method of linear differential analysis with 2 features showed that the accuracy of risk prediction of the logistic regression model with 62% and an macro average of 48.5 is higher than other models in this method.

Table 6: Analysis of the Findings of Applying Different Models with 2 Features

Classification method	Accuracy	macro a vg	Label	precision	recall	f1-score
Voting(Random Forest, Bagging, Fuzzy Pattern)	53	41	0	43	31	36
			1	61	76	68
			2	22	14	17
Bagging	42	40.5	0	42	31	36
			1	61	74	67
			2	22	15	18
Fuzzy Pattern	36	40.5	0	40	51	45
			1	61	24	35
			2	22	53	31
Multimodal Evolutionary	42	43	0	37	60	46
			1	67	36	47
			2	23	39	29
Random Forest	54	41	0	44	30	36
			1	61	77	68
			2	22	13	16
Fuzzy Pattern Tree Top Down	58	38	0	62	0	0
			1	58	1	73
			2	50	0	0
K-Nearest Neighbors	49	39.5	0	35	41	38
			1	62	64	63
			2	22	14	17
Decision Tree	47	39	0	36	35	35
			1	61	60	60
			2	21	23	22
MLP	62	47.5	0	71	27	39
			1	61	96	75
			2	55	2	3
Ada Boost	62	47.5	0	68	25	37
			1	61	96	75
			2	62	1	3
Gaussian Naive Bayes	61	45	0	58	32	41
			1	62	92	74
			2	40	2	4
Quadratic Discriminant Analysis	61	45	0	57	32	41
			1	62	92	74
			2	40	2	4
Gradient Boosting	61	47	0	59	30	40
			1	61	94	74
			2	60	1	3
Logistic Regression	62	48.5	0	71	26	38
			1	61	97	75
			2	66	1	3

As can be seen in ROC curve in Fig. 8, the comparison of the findings of 14 research models using the linear differential analysis method with 2 features, larger area under the ROC curve of the logistic regression model than other models indicate shows the higher accuracy of the risk level prediction of this model than other models.

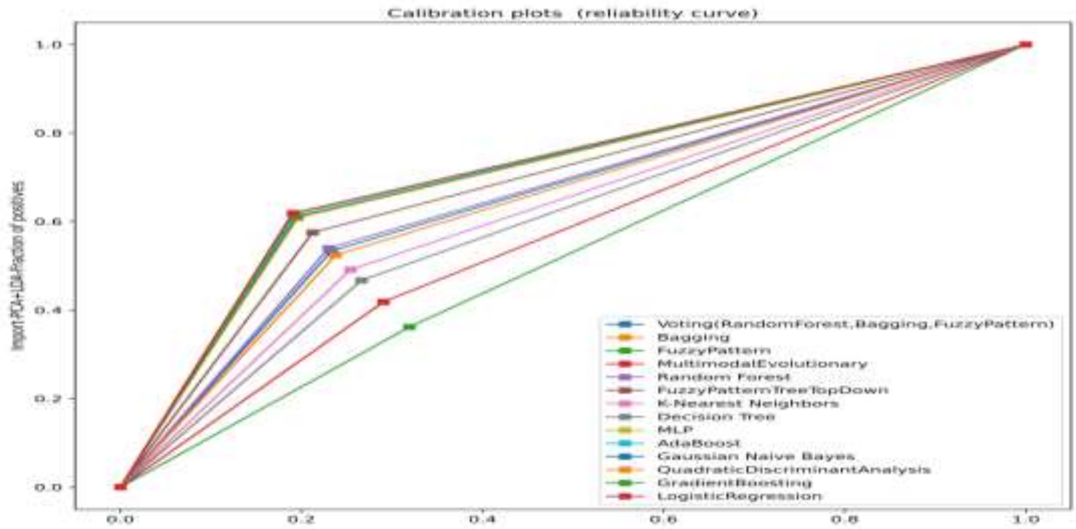

Fig. 8: ROC Curve

Finally, comparison of the precision of prediction of the superior models in the three tables (4, 5 and 6) showed a higher risk prediction accuracy in the random forest model. By this model and in the case of having special input attributes, the value of the specific attribute of the target can be estimated. In other words, the declarations can be atrributed to one of the three defined risk levels. Finally, since the accuracy of the random forest model is higher than other models, so this model can be used in risk assessment management in the customs organization in the proper identification of the risk category of import declarations as a suitable organizational knowledge.

6 Conclusion

Failure to use risk management as well as early information system as a subset of daily routines and the tendency to use past data without study are blind spots in understanding and receiving risk as well as preventing crisis. Major negligence in estimating risks is the most important reason for financial crises. Risk rating is one of the main bases of risk management and provides the opportunity to provide appropriate and timely response to risks. In this paper, we presented fourteen supervised learning method and applied them to customs data. Tests show that data analysis techniques allow for effective targeting declarations. Since the it is important for decision makers to know which model is right and appropriate for them and to prepare accordingly, the results of this study showed that the random forest model was more precised in identifying and determining the level and rate of risk of imported declarations than other models used. The random forest model provides reliable results and the ranking of risks by this method is valid by considering several features at the same time, the flexibility of the method and the analytical nature of the results. The rules resulting from this model are considered a hidden pattern in the Iranian customs database and can be used to predict import declarations with a level of risk in red, yellow and green channels and apply policies to manage the risk of import declarations. The most important purpose of the selected model is to provide the wasted financial revenues of the government from the place of salaries and input duties and with different sub-objectives as follows:

- Prevent tax evasion and entry fees and access to lost government revenues,
- Preventing the sleep of capital in customs,
- Facilitate and accelerate trade and contribute to economic growth,
- Prevent smuggling, money laundering and organized crime,
- Facilitate and accelerate the supply chain of goods,
- Increase controls if necessary,
- Targeted regulation of goods inspection programs,
- Formation of risk profiles and analysis of risk information,
- Identify potential security threats.

The application of this algorithm can be extended to solve other problems. The goal of financial management organizations or companies should be to quickly identify threats to their reputation and maintain their reputation through long-term precautionary measures. Despite all these bitter experiences, risk management is still often considered as one of the side tasks, while the main focus of the organization's senior managers should be and all decisions should be made in coordination and in parallel with risk management. Although, there is a very good discourse in favor of the adoption of risk management practices by customs at the highest levels of decision-making by government, there are few studies on the principles of implementation and application of customs risk management [48]. However, from a research point of view, there seem to be two main reasons: First, few theories have been explained and formulated to embed risk management in the body of the customs organization and to institutionalize it. Second, the data available in the field of customs risk management do not show much desire to share information with researchers because traders may misuse the published data and statistics in line with their illegal greed and calculated avoidance of customs inspections. Therefore, despite the importance of risk management for customs, related techniques have not been fully studied empirically and theoretically [52].

But Sience the concept of risk-based management is applicable in almost every business and government field, many experiences can be shared with the subject of customs [12]. In an article entitled "Amending the association rules of customs crime information based on repeated motives", using the association rules, introduced a way to discover the knowledge used in customs crime data. They came with a set of rules of knowledge to predict and target specific hazard conditions [55]. According to [37], the decision tree is the most appropriate data mining technique. Using two techniques of self-organizing network and k-means algorithms, conducted a study entitled "Comparison of two data mining methods in risk-based clustering of vehicle insurance customers (Case study: Mellat Insurance Company)". The results showed that the outputs of the two systems are very close to each other. The customers were clustered based on risk and the strengths and weaknesses of each technique were reported [30]. In another study, in an article entitled "Investigating the usefulness of data mining techniques in detecting fraud in financial statements" examined a sample of 202 companies, including 101 fraudulent companies and 101 non-fraudulent ones. The results of applying data mining techniques such as multilayer feedforward neural networks, support vector machines, genetic programming, group data modeling, logistic regression, and probabilistic neural network indicated the superiority of probabilistic neural networks without feature extraction in the discovery of the fraudulent financial statements. In the case of including feature extraction, genetic programming and possible neural networks performed with almost equal accuracy [49]. In a paper entitled "Risk management systems: the use of data mining in the customs of developing countries" examined several simple statistical regression methods, logit, and probit in Senegal. They indicated that using statistical risk indexing methods, 96.6% of declaration violations were detected by inspecting only 20.6% of them. Therefore, it can be claimed that the number of intrusive inspections can be reduced by 80% [38].

References

[1] Aguirre, J. G., Villar, A. D., *Administrative Costs and Tariff Rates in the Presence of Customs Evasion: Evidence from Ecuador*, Journal of Economies, 2021, **21**(9), P. 1-16. Doi:10.3390/economies9010021

[2] Asadollahi, *S.Y.*, Taherabadi, *A, A.*, and Khairollahi, F., *Modeling Liquidity Risk Management in Banking Using System Dynamics Approach*, Advances in Mathematical Finance and Applications, 2020, **6**(3), P. 493-514. Doi: 10.22034/AMFA.2020.1899914.1424

[3] Annamalah, S., Raman, M., Marthandan, G., and Logeswaran, A. K., *Implementation of Enterprise Risk Management (ERM) Framework in Enhancing Business Performances in Oil and Gas Sector,* Journal of Economies, 2018, **4**(6), P. 1-12. Doi:10.3390/economies6010004

[4] Afanasieva, V., Ivanov, L., and Yanushkevych, D., *Modern Approaches to Risk management and Their Use in Customs,* Traektoriâ Nauki, 2017, **3**(4), P. 1-14. Doi: 10.22178/pos.21-6

[5] Alipour Shirsavar, H., Shirinpour, M., *The effect of electronic customs administration on facilitating the export activities of export companies based in Gilan, Iran,* Intellectual Economics, 2016, **10**, P. 114–121. Doi: 10.1016/j.intele.2017.03.004

[6] Ali Asghari, S., Ahmadi Abkenar, F., and Shah Bahrami, A., *Identification of systemic and business risks and risk management in customs.* National Conference on Organizational Risk Management, Tehran, Center for Productivity and Human Resources Studies, 2015, (in Persian).

[7] Alipour, M., Badiee, H., and Heydari, M., *Requirements and strategies for exercising control after customs clearance of goods*, Journal of Management Accounting and Auditing Knowledge, 2012, **4**, P. 56-72, (in Persian).

[8] Arabi, M., *Strategic planning of the customs of the Islamic Republic of Iran. Tehran*, Nil Publications, 2004, 136 pages (in Persian).

[9] Basir, A., Satyadini, A. E., and Barata, A., *Modern Customs Risk Management ramework: Improvement towards Institutional Reform*, International Journal of Innovative Science and Research Technology, 2019, **4**(7).

[10] Braganza, A., Brooks, L., Nepelski, D., Ali, M., and Moro, R., *Resource management in big data initiatives: Processes and dynamic capabilities*, Journal of Business Research, 2017, **70**, P.328– 337. Doi: 10.1016/j.jbusres.2016.08.006

[11] Biljan, J., Trajkov, A., *Customs Risk Management – time of Declaration processing by contorol caneels*, Economy and Market Communication Review, 2016, **6**(1), P. 93-108. Doi: 10.7251/EMC16093B

[12] Biljan, J., Trajkov, A., *Risk management and Customs performance improvements: The case of the Republic of Macedonia*, Procedia - Social and Behavioral Sciences, 2012, **44**, P. 301– 313. Doi:10.1016/j.sbspro.2012.05.033

[13] Coombs, C., Hislop, D., Taneva, S K., and Barnard, S., *The strategic impacts of Intelligent Automation for knowledge and service work: An interdisciplinary review*, Journal of Strategic Information Systems, 2020, **29**. Doi: 10.1016/j.jsis.2020.101600

[14] Chermiti, B., *Establishing risk and targeting profiles using data mining: Decision trees*, World Customs Journal, 2019, **13**(2), P.46-57.

[15] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R., *CRISP-DM Step-Data Mining Guide*, SPSS Inc.CRISPMWP, 2000.

[16] Duan, J-H., & Pan, Y-L., *A Study of Customs Tax Risk Management. Advances in Economics, Business and Management Research (AEBMR)*, International Conference on Management Science and Management Innovation, 2018, 54.

[17] De Wispelaere, F., Pacolet, J., *Data Mining for More Efficient Enforcement*, Learning Resource Paper, 2017, P.1-20.

[18] Fayyad, G., Piatestsky-Shapiro, p., and Symth, P., *From data mining to Knowledge discovery in databases*, Al Magazin, 1996, **17**, P. 37-54.

[19] Garcia, G., Caballero, M., *A Multi-Objective Bayesian Approach with Dynamic Optimization (MOBADO). A Hybrid of Decision Theory and Machine Learning Applied to Customs Fraud Control in Spain*, Mathematics, 2021, **9**. Doi: 10.3390/math9131529

[20] Goli, R., *Explanation and analysis of the Customs Affairs Law and its executive regulations*, Tehran University route, 2018, 383 pages, (in Persian).

[21] Grover, V., Chiang, RH.L., Liang, T-P., and Zhang, D., *Creating Strategic Business Value from Big Data Analytics: A Research Framework*, Journal of Management Information Systems, 2018, **35**(2), P. 388- 423. Doi: 10.1080/07421222.2018.1451951

[22] Gunther, W. A., Rezazade Mehrizi, M. H., Huysman, M., and Feldberg, F., *Debating big data: A literature review on realizing value from big data*, Journal of Strategic Information Systems, 2017, **26**, P. 191– 209. Doi: 10.1016/j.jsis.2017.07.003

[23] Ghazanfari, M., Alizadeh, S., and Timourpour, B., *Data mining and knowledge discovery*, Tehran University of Science and Technology, 2016, (in Persian).

[24] Gleissner,W., Berger, T., *Risk Management in Simple Language*. Translated by Shahram Amin Torabi, Tehran: World Economy Publication, 2015, 136 pages, (in Persian).

[25] Geourjon, A-M., Laporte, B., Coundoul, O., and Gadiaga, M., *Inspecting less to inspect better: The use of data mining for risk management by customs administrations*, Working Paper Development Police, 2012, **46**, P. 1- 19.

[26] Geourjon, A-M., Laporte, B., and Graziosi, R., *How to modernise risk analysis and the selectivity of customs controls in developing countries?* WCONEWS, 2010, **62**, P. 29- 31.

[27] Herusantoso, Kh., Dwi Saputra, A., *Factors affecting the customs clearance time at prime customs office type an of tanjung priok*, Customs Research and Application Journal, 2020, **2** (2).

[28]Hoffman, A., Grater, S., C Venter, W., Maree, J., and Liebenberg, D., *Designing a new methodology for customs risk models*, World Customs Journal, 2019, **1**(13), P. 31-55.

[29] Hayati, M., Atai, M., Khalokakaei, R., and Sayadi, A., *Risk assessment and ranking in the supply chain using taxonomic analysis method (Case study: Isfahan Steel Complex)*, Journal of Operations Research in its Applications, 2013, **1**(40), P. 85-103, (in Persian).

[30] Hanafizadeh, P., *Comparison of two data mining methods in segmenting car body insurance customers based on risk (Case study: Mellat Insurance Company)*, Industrial Management Studies, 2013, **30**, P. 77- 97, (in Persian).

[31] Haji Heidari, N., Samarand, K., and Farahi, A., *Risk Classification of Car Body Insurance Insurers Using Data Mining Algorithms (Case Study: An Insurance Company)*, Insurance Research Journal (formerly Insurance Industry), 2010, **4**, P. 107- 129, (in Persian).

[32] Jellis, V., David, M., and Bruno, P., *Customs fraud detection : assessing the value of behavioural and high-cardinality data under the imbalanced learning issue*, Pattern analysis and applications, 2020, **23**, P. 1457-1477. Doi: 10.1007/S10044-019-00852-W

[33] Jiawei, H., Camber, M., Pi, J., *data mining: Concepts and techniques*. Translated by Mehdi Ismaili, Tehran, Niaz Danesh Publications, 2017, 600 pages, (in Persian).

[34] Jamaat, A., Asgari, F., *Credit risk management in the banking system with a data mining approach*, Quarterly Journal of Quantitative Studies in Management, 2010, **11**, P.115- 126, (in Persian).

[35] Kavoya, J., *Machine learning for intelligence driven Customs management*, African Tax and Customs Review, 2020, **1**(3), P. 50–58.

[36] Kitchens, B., dobolyi, D., LI, J., abbasi, A., *Advanced customer analytics: strategic value through integration of relationship-oriented big data*, Journal of Management Information Systems, 2018, **35**(2), P.540–574. Doi: 10.1080/07421222.2018.1451957

[37] Larus T. Daniel., *Data Acquisition an Introduction to Data Mining*. Translated by Ali Zeinal Hamedani, Farhad Ebrahimian, Hadith Yaghoubzadeh, Isfahan: Isfahan University of Technology Publishing, 2013, 295 pages, (in Persian).

[38] Laporte, B., *Risk management systems: using data mining in developing countries customs administrations*, World Customs Journal, 2011, **5**(1), P. 17- 29.

[39] Moslemi, A., Pourzamani, Z., and Jahanshad, A., *Ranking of Banks' Risk Reporting Using Data Envelopment Analysis*, Advances in Mathematical Finance and Applications, 2021, **6**(4), P. 695-715. Doi: 10.22034/AMFA.2021.1899631.1436

[40] Mohammadi, M., Yazdani, Sh., and Khanmohammadi, M., *Presenting a Model for Financial Reporting Fraud Detection using Genetic Algorithm*, Advances in Mathematical Finance and Applications, 2021, **6**(2), P. 377-392. Doi: 10.22034/amfa.2019.1872783.1252

[41] Marzuki, M. M., Abdul majid, W. Z. N., Azis, N.K., Rosman, R., and Abdulatiff, N. K. H., *Fraud Risk Management Model: A Content Analysis Approach*, Journal of Asian Finance, Economics and Business, 2020, **10**(7), P. 717– 728. Doi: 10.13106/jafeb.2020.vol7.no10.717

[42] Mikalef, P., Boura, M., Lekakos, G., and Krogstiea, J., *Big data analytics and firm performance: Findings from a mixed-method approach, Journal* of Business Research, 2019, **98**, P.261– 276. Doi: 10.1016/j.jbusres.2019.01.044

[43] Martincus, C. V., Carballo, J., and Graziano, A., *Customs*, Journal of International Economics, 2015, **96**, P. 119– 137. Doi: 10.1016/j.jinteco.2015.01.011

[44] Meshkani, A., Nazemi, A., *An introduction to data mining*. Mashhad: Ferdowsi University Publishing Institute, 2009, 450 pages, (in Persian).

[45] Nemirova, G. I., Savelyeva, T. I., *Significance assessment of the risk management system to improve the quality customs service provision*, Economic consultant, 2020, **31**(3), P. 42- 52.

[46] Nelson, C., *Machine learning for detection of trade in strategic goods: an approach to support future customs enforcement and outreach*, World Customs Journal, 2020, **2**(14), P. 124-133.

[47] Okazaki, Y., *Implications of Big Data for Customs -How It Can Support Risk Management* Customs, Journal WCO Research Paper, 2017, **39**, P. 1- 24.

[48] Rukanova, B., Tan, Y-H., Slegt, M., Molenhuis, M., Rijnsoever, B. V., Migeotte, J., Labare, M. L.M., Plecko, K., Caglayan, B., Shorten, G., Meij, O. V. D., and Post,S., *Identifying the value of data analytics in the context of government supervision: Insights from the customs domain*, Government Information Quarterly, 2021, **38**, 101496. Doi: 10.1016/j.giq.2020.101496

[49] Ravisankar P, Ravi V, Rao GR, Bose I., *Detection of financial statement fraud and feature selection using data mining techniques*, Decision Support Systems, 2011, **50**, P. 491- 500.

[50] Shishechiha, M., *Risk management in customs affairs*, Tehran, Basic Science Development Publications, 2015, 144 pages, (in Persian).

[51] Salarzaei, A., Jamshidi, S., *New Risk and Customs Management in Khorramshahr*, Master Thesis, Sistan and Baluchestan University, 2012, (in Persian).

[52] Yousefi, M., *Comparative study of customs risk management*, Tehran, Dara Publications, 2016, 179 pages, (in Persian).

[53] Yousefi, M., *Modern customs programs in the 21st century*, Tehran, *Dara Publications*, 2016, 235 pages, (in Persian).

[54] Zhou, X., *Data mining in customs risk detection with cost-sensitive classification*, World Customs Journal, 2019, **2**(13), P. 115-128.

[55] Zehero, B-B., Soro, E., Gondo, Y., Brou, P., and Asseu, O., *Elicitation of Association Rules from Information on Customs Offences on the Basis of Frequent Motives*, Engineering, 2018, **10**, P. 588- 605. Doi: 10.4236/eng.2018.109043

[56] Zarepour, Z., Chamani, G., *Risk management and its executive requirements in customs*, paper Second National Conference on Modern Management sciences, Gorgan, 2013, (in Persian).