The Use of Data Mining Applied In the Accounts Receivable from the Employees of a Government Institution in México

José Alberto Hernández Aguilar[1], Enrique Montoya[1], Alberto Ochoa[2]
José Crispín Zavala[1]and Augusto Renato Pérez Mayo[1]

Abstract

In this research we discussed the application of data mining in financial information, focusing on accounts receivable of the employees of a government office, we applied visual data mining in (993) accounting records using WEKA to identify outliers by geographical region, the frequency and the amount of money missing in a period of time, and analyzed the main factors involved in this process. Our preliminary results show that the insecurity in the region is highly correlated with the amount of money missing in the government institution, overhanging the north and southeast of Mexico as the most affected.

Keywords: Data Mining, financial information, government institution, shortages of money.

1 Introduction

There are various definitions of data mining (Fayyad, 1996) (Corso, García, Ciceri, & Romero, 2014)(Hernández Aguilar, 2013) (Núñez Cárdenas, 2011) (Riquelme, Ruiz, & Gilbert, 2006)(Cevallos Chacon, 2013)(Peña Ayala, 2014) that discuss how data is exploited to obtain knowledge, but a more complete definition according to (Jiawei, Micheline, & Jian, 2012) is "obtaining knowledge from the data, through technical and with the help of applications that allow us to discover patterns that are in the same data, although not all of them can be used for this purpose". At present all the organizations and mainly international consortia generate, process, collect and manage a large amount of information as a result of their operations.

Some of the applications of what is called data mining has been used in the departments of human resources, marketing, production and finance, among others, which indicates that its use is widespread and that there is knowledge of it, where the results are tangible evidence in the measurement of the efficiency of the improvements (Hernandez Aguilar, 2013) (Corso, García, Ciceri, & Romero, 2014).

The analysis of financial patterns in the governmental organizations that manage the money of users and contributors is a very sensitive topic, as it is one of the organizations that are responsible for collecting and managing the toll on the main highways of Mexico, to do this, simply consider the case of OHL Mexico that ended in administrative sanctions for 3.7 million Euros by the imprecision of their accounting records (NTMX, 2016). The factors that influence the determination of patterns of financial behavior could be a large number, therefore, in this work is aimed to determine the patterns of behavior in the toll of some of the states of the Mexican Republic.

The information generated by this behavior is required by different firms of auditors, the Ministry of Finance and Public Credit, the Public Function, IFAI (Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales, or National Institute of Transparency, Access to Information and Protection of Personal Data, in English), as well as of the different internal areas and the Internal Control in the government institution.

---

[1] Universidad Autónomadel Estado de Morelos, Av. Universidad No. 1001, Col. Chamilpa, Cuernavaca, Morelos, 62209, México. Email: jose_hernandez@uaem.mx
[2]Universidad Autónoma de Ciudad Juárez.

The procedure that was used in this work contributes to systematize the development of the required information, which currently is generated manually, which is prone to errors and is obtained outside the deadlines.

Likewise, it allows for the identification of patterns that affect the diversion of resources that are generated and increase the indebtedness of this public institution and with the pass of time, become uncollectible accounts to those responsible for staff operating in the toll. In this research is used visual data mining for the identification of patterns and relationships, and is applied to a considerable amount of transactions. The preliminary results indicate that there is an inverse correlation between the debtors (employees with missing money) with respect to the national criminal rate.

This document is organized as follows, the first section introduces to the issue under discussion. The second section describes the problem and the hypothesis of the research. In the third section is described the data mining process, briefly explains its operation and the methods used in the present analysis, which is based on the case studies where it has been implemented for the identification of fraud and the analysis on accounting records. In the fourth section is described the design of the database used. Fifth section describes the methodology used. Later, we show and discuss the results. Finally, we present the conclusions and future work.

1.1 Problem

There are few studies that analyze the financial information derived from public operations due it is sensitive information. The entity that is analyzed in this research, which from now on will be referred to as the Institution, has as its main source of income the contributions of toll charged in the country's highways, which generates a large amount of operations and therefore constitutes a large volume of information that according to (Jiawei, Micheline, & Jian, 2012) can be analyzed with data mining.

The Institution manages financial information required by different firms of auditors, the Ministry of Finance and Public Credit, the Public Function, IFAI, as well as of the different internal areas like the Internal Organ of control, this information is generated, most of the times, manually and with a strong probability of error, and on many occasions occurs out of time. Examples of this information are: the bank reconciliations of the different banks, reconciliations of accounting-the budget of income and expenses, the integration of the identity cards of investment, the ballot summary of the payroll tax and the income tax, etc. This information is processed in Excel, and is concentrated in reports using a SAP system (SAP, 2016) property of the Institution, and to do this it crosses the information in these reports, and in some cases is matched with bank statements, and produce deliverables for the different requirements of information in a timely manner.

In each monthly closure, the Institution should send information to the Secretariat of Finance and Public Credit, which should be grouped together, organized and summarized. Normally this information is sent out of time, with errors and their processing is done manually which requires many hours man.

Another situation that is presented is the reference to the payment of taxes, which is performed monthly every fifteen day of the month later, this payment is carried out manually and the main problem identified is the large amount of information that is processed, which requires a lot of time, and continually presents delays and errors, resulting in penalty fees or surcharges, since the tax declaration is presented to the SAT each day fifteen of each month, and if the information is not correct, is required to submit a supplementary declaration and this causes a penalty fee.
For the operation of the Institution, it is very important to remark that the toll collection is carried out correctly for the benefit of the same, but also for the user that is using its infrastructure; because during the whole day, all days of the year and throughout the country is using the federal highways and bridges.

From the situations described above, this investigation will focus on the accounts receivable from the employees, given that a vital point for the Agency in study is the missing money in the tolls, due to that it can be very complex the discount track payroll and by the high number of ATMs that become creditors, so that its application and registration is difficult to perform in time and form, causing a great problem to the institution. As the proposed solution is raised by the analysis of the accounts receivable so as to identify patterns derived from the debts, for the different periods that are required (monthly, semi-annual and annual), and to facilitate the decision making through data mining.

These shortages are given in the majority of the places of collection of every country and they are discounted from the ATMS, however, is very difficult for these performing the discount via payroll in a timely form. Another situation that makes difficult the recovery of the missing is when the amount exceeds 30% of the worker's salary, since according to the regulations in vigor, only can deducted that same percentage, another case happens when the worker has many discounts for various reasons and his salary is not enough to cover this debt and then is not deducted.

In an internal audit as in an external, the Institution becomes a creditor, and constitutes an observation on the part of the auditors, where they make it quite clear the recommendation for its implementation and decrease since it constitutes damage to the nation.

Hypothesis. It is possible to apply data mining, to the accounts receivable of employees of a public Institution for the identification of possible mismanagement and fraud.

Research Contribution. The main contribution of this research is the identification of suspicious patterns on the accounting records derived from the shortage of employees of a public Institution.

2 Data mining

Data mining according to (Jiawei, Micheline, & Jian, 2012) is the "reach the knowledge from the data, through technical and with the help of the applications, which allow you to discover patterns that are in the same data, but not all of these, can be used for this purpose". At present all the organizations and mainly international consortia generate, process, collect and manage a large amount of information as a result of its operations, data mining can then be seen as a process which identifies the patterns to get useful and important knowledge, this occurs because they are part of a lot of data which can be included in: databases, a data warehouse, the web and other repositories of information or data that is transmitted in a dynamic way so that these become endless (Jiawei, Micheline, & Jian,2012).
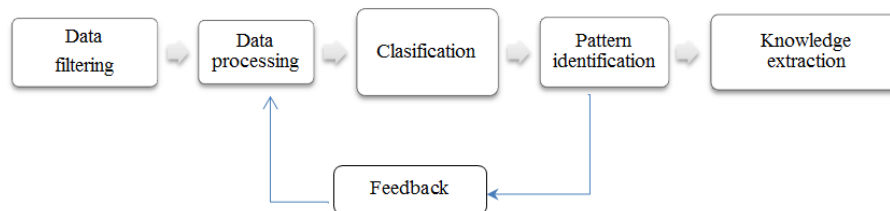

Figure 1.Data Discovery knowledge, Source (Jiawei, Micheline, & Jian, 2012)

The process of knowledge discovery includes the following steps (Jiawei, Micheline, & Jian, 2012):

1. Cleaning of inconsistent data.
2. Integration of data in a single place where multiple data sources can be combined.
3. Selection of data relevant to the task of analysis. These are retrieved from the database.
4. Transformation of data. The data are transformed and consolidated in ways most appropriate for mining through the conduct of operations like summary or aggregation.
5. Data Mining. View as an essential process where intelligent methods are applied to extract patterns of data.
6. Evaluation of the pattern. This step is used to identify the truly interesting patterns that represent the knowledge based on the measures of interest.
7. Knowledge. Where the techniques of representation of visualization and of knowledge are used to present the mined knowledge to users in this same sense, Usama Fayyad defined it as (Fayyad, 1996): "a non-trivial processes of valid identification, novel, potentially useful and understandable patterns that are hidden in the data".

The nature of the data mining is the exploitation and the analysis of the information, also consists in finding and extracting the most important information when having enormous amounts of data. The data warehouses generate the necessary information for the data mining algorithms to function properly.

A particular step of the data mining is to obtain knowledge of the databases by using unique and appropriate algorithms to extract information from the data, using models for this process and with statistics which is the first science that historically allowed to extract information from the data, basically using methodologies from mathematics and has been identified as the key phase of the process, which is formed by one or more of the following functions: classification, regression, grouping, summary, image recovery and removal of rules (Riquelme, Ruiz, & Gilbert, 2006).

When there is an interaction of data mining and a multi-agent system results an innovative tool with a new technology that is the knowledge in the database and with a zero learning curve and a decision-making in different environments (Ruiz, Hernandez, & William, 2009). According to these authors the decision support systems solve problems with different degrees of structure and uncertainty, the decisions for which the information that is obtained from a computational system is only a portion of the total knowledge required for decision making.In this same sense (Molero Castillo &Meda Campaña, 2010) note that: "data mining is the stage that provides usable information in a data group, its main role is the technological support it provides, and that has contributed to give solution to two major challenges: 1) working with sets of data to extract and discover information of interest, and 2) use appropriate techniques to explore, analyze, understand and identify trends and behaviors".

To reduce the uncertainty is needed adequate information, and therefore it will make it possible to make better decisions at all levels of the company. In a credit card fraud is indispensable the immediate response to have a response in line and count with a reduction of response times through the data mining. The infrastructure facilitates interaction with the architecture and the process of the involved parties, learning provides the methods and learning capacities, with the interaction allows the coordination, cooperation and communication configuration also allows the modeling, design and implementation of the user interface (Molero Castillo &MedaCampaña, 2010).For (Núñez Cárdenas, 2011) "data mining is the process of knowledge extraction from the data, which consists of several phases, among which are: the preparation of the data, which consists in the selection, cleaning and processing, evaluation and dissemination. As well as the use of models, that incorporates techniques like decision trees, linear regression, artificial neural networks, Bayesian techniques, support vector machines, and databases".

"From the economic point of view, at present the data mining is a technology for the management and analysis of the information that reduces the cost, and increases the speed for the processing, storage and transmission of data. Which gives as a result find the knowledge contained in large data bases of the organizations" (Altamiranda Echeverri, Peña Retamoza, Ospino De La Rosa, Volpe Barros, Ortega Álvarez, & Cantillo Guerrero, 2013). An interesting fact is that the most used algorithms that were developed three decades ago make possible today software products for data mining that generate results of high reliability that can increase the benefits and reduce costs. Thus "data mining can be understood as a system of removal of relationships, artificial neural networks algorithms, decision trees, induction of rules, fuzzy logic, analysis of the links and the genetic algorithms, it also encompasses, Bayesian and classical statistical methods, models or theories formulated using the data mining application from a narrow view, as examples can be mentioned exploratory data analysis, regression minimum quadratic, logistic regression and the discriminant analysis" (Altamiranda Echeverri, Peña Retamoza, Ospino De La Rosa, Volpe Barros, Ortega Álvarez, & Cantillo Guerrero, 2013).

2.1 Implementation of data mining in Organizations

Data mining has been used in several organizations in the departments of human resources, marketing, production and finance, which indicate that its use is widespread and that there is knowledge on it and tangible evidence in the measurement of the efficiency of itsresults (Hernandez Aguilar, 2013). Nowadays some subjects that cannot be separated in the public and private educational sector are the disciplines of the computer science and mathematics, both key factors in the process of data mining. In this same sense, in the educational sector is used data mining to perform a study and detailed analysis on the behavior with the students in the different systems of learning teaching. The task that performs data mining is (Ballesteros Román, Sánchez-Guzmán, &García Salcedo, 2013): Select and apply the appropriate method, i.e. the realization of a selection of the task for the discovery of knowledge, such methods are the classification, grouping, rules of association, regression, and some algorithms like the decision trees.The bivariate statistics, own of data mining, has the ability to distribute a set of data into groups so that the characteristics of its elements are similar between them, with a strong relationship and relevance that makes it possible to improve the decision-making process (Malbernat, 2013).

According to the article "Analysis and evaluation of the level of risk in the granting of financial credits using data mining techniques" (Tello, Eslava, &Tobías, 2013), and with the intention of applying it to the present investigation, it mentioned that was used WEKA (a data mining visual tool) to analyze multiple records, which are processed with decision trees, and these results were compared with those obtained with other algorithms. These authors indicate that: "when a large amount of information is used it is suitable the use of the decision trees, as they find the relations and underlying models in the data, which allow you to perform the division of the data and predictiontool, generating rules for the information, forming in turn small data sets that are easier to work in its global processing".

One of the main problems in the organizations is the high number of security incidents that occur on computers, so that every organization must place preventive measures and ensure the normal functioning. According to (Corso, García, Ciceri, & Romero, 2014), techniques that offer the statistics do not reflect the true interrelation of variables and therefore, an option considered as interesting and attractive for this is data mining. Within the stages of KDD (Knowledge Discovery in Databases) suggested by (Jiawei, Micheline, & Jian, 2012) and applied in this study of security include: data preparation, selection phase, exploration phase, cleaning phase of data (identification and treatment of deviations also called outliers), phase of transformation of data, hierarchical clustering of the relevant data relating to the computer incident, implementation of non-hierarchical clustering, analysis of the cluster obtained and validation with the user, application of algorithms of association and classification.

According to (Fadairo, Williams, Trotman, &Onyeleku-Eze, summer 2008) around the world, governments spend billions of dollars each year in programs, services and other items. To ensure accountability, governments need to be aware that these payments are valid and meet the requirements of its compliance. In support of that request, the integrity of payment is essential to enable governments to manage effectively. Data mining can perform two basic operations (Fadairo et al., 2008): 1) the prediction of trends and behaviors, and 2) the identification of previously unknown patterns. This technique allows users to analyze the data from many different perspectives, categorize the data and summarize the identified relationships. Predictive Data Mining is more common, and serves for business applications like business analytics,on the other hand, data mining applied for accounting can demonstrate how companies, financial institutions, insurance companies, the tax authorities and other government agencies may be able to use the data of the accounting, prosecutors and tasks related to the audit to identify underlying patterns in the data (Fadairo, et al., 2008).

The models may be used to detect patterns and accounting irregularities, inappropriate practices, questionable transactions, possible fraud and money laundering (. The mining of data allows monitoring transactions, the credit score, the default loan prediction, risk assessment and its reduction to a minimum, as well as the prediction of futures expected in the demand, prices and sales of the products or services, businesses process like the majority of their purchases, sales and payroll transactions in computer equipment. Since this perspective, companies accumulate huge sets of transactions on sales and accounting books, and shopping (Miletti, Aquel, Berri, Diaz, Doffo, Ingrassia, et al. 2010), with many of those transactions that carry a separate set of data to a third party. Data mining then becomes one of the major tools to find a fraud. The identification of cases of fraud in business is not easy. Data mining has its own problems due to the requirement of the data protection legislation. In this case, it is not recommended that the mining of data is done (Fadairo, et al., 2018).

2.2 Decision Making

The decision-making process is a determining factor in all organizations; its success resides in taking an appropriate decision with the information available. The organizations generate large amounts of information, but analysis must be organized and summarized, easy to revise, interpret and be on time (Ochoa Zezzatti, Romo, Bañuelos, Mendhizavili, Iztebegovic, & Hal, 2007)(Calleja Gómez, 2010)(Castro, Sifuentes, González, & Rascón, 2014). On the basis of the traditional reality, decision-making depends on the hierarchical level and its relationship with the goals established in the organizations, in which most of the time there is an ambiguous and uncertain environment where each of its processes inside, the role of information makes the difference between a decision made to permit the resolution of problems and offers an alternative solution, that could rise to a number of consequences not at all favorable (Hernández, Levín and Vázquez, 2011).

An adequate analysis of information reduces the uncertainty in the decision-making and so that occasionally requires the use of the most up-to-date information to feed the decision-making process. In all the organizations the responsible for making the decisions require information that allows them to have feasible answers in relation to its members, which is generated from a diagnosis of the organizational culture. The use of data mining in the activities of the organizations and enterprises is becoming necessary and indispensable. For example, the business of distribution and advertising has employed methods of data mining to reduce costs or increase bids, since it is not the only thing that can be done and has developed applications in the following areas: financial, insurance, scientific, among others (Hernández, Levín and Vázquez, 2011).

3 Design and Implementation of the database

For this experimental work, information was initially located in two Excel files, one file has the balances of the accounts receivable of the employees of the Institution, covering a period of one month (01/01/2015 to 31/01/2015) corresponding to the debts of the employees.

Another file that contains the employee personal data such as: name and surname, date of birth, date of entry, marital status, number of children, schooling. These Excel files come from two reports obtained from the SAP system of the Public Institution.To systematize our analysis, both tables were imported and normalized into a database in Access, using as main key the ID (identifier) of the employee as described in (Connolly & Begg, 2005).
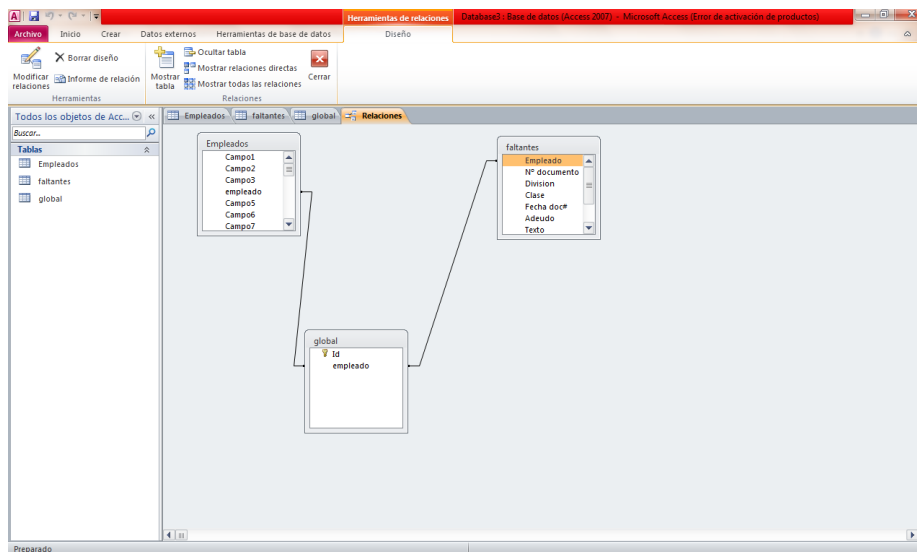

Figure 2.Relational schema of the database (own source)

4 Methodology

For this research work, we used the KDD methodology, shown in figure 1) which is based on the work of (Jiawei, Micheline, & Jian, 2012). The proposed methodology is the following:

1. The first step is obtaining the data which were obtained from the Institution's system which operates in SAP, from two of its reports are generated the tables shown in the database (see figure 2). A complimentary set of data was obtained from CIDAC (2015) corresponding to the Criminal rate by state in Mexico.

2. Filtering data, processing and classification. In this stage we proceed to clean the data, delete those registers that are empty and/or duplicated, leaving only those that are relevant to the data mining process, as there may be different types of debts. It should be noted that at this stage we generated summaries from the missing by division of the Institution, to contrast them with the criminal rate by state where is located the division.

3. When database was filtered and clean, it was exported in CSV format (Comma delimited) and imported in WEKA system, which is a visual data mining software which helps in the processing of information and allows you to identify useful patterns, also allows to identify relationships on the basis of the analysis of main components that serve to rethinking the information to refine the model and find hidden patterns, or generation of reports like employees by segment (age, sex, education, type and size of debt, real income and number of pensions) that do not exist in the Institution.

4. The information is visually analyzed in WEKA to find relationships between the variables that allow identifying patterns in the data set, which subsequently will allow the identification of correlations, and eventually the prediction of suspicious behavior and the generation of knowledge.

5. To evaluate statistically patterns found, an analysis of bivariate correlation in SPSS version 22 is done to verify the validity of the results obtained in the visual way.

5 Results

The preliminary results shown in Figure 3 indicate that there are shortages in the various sections of the Institution, divisions in the north of the country (Tijuana and Reynosa), the southeast (Coatzacoalcos and Veracruz), and those of the central offices. For our analysis, we assume that with data mining can be detected patterns and irregularities to identify possible fraud according to (Jiawei, Micheline, & Jian, 2012) and (Hernández Aguilar, 2011). In this way, the results shown in figure 3 notes that the accounts of Northern Mexico and southeast are the sections most missing have.
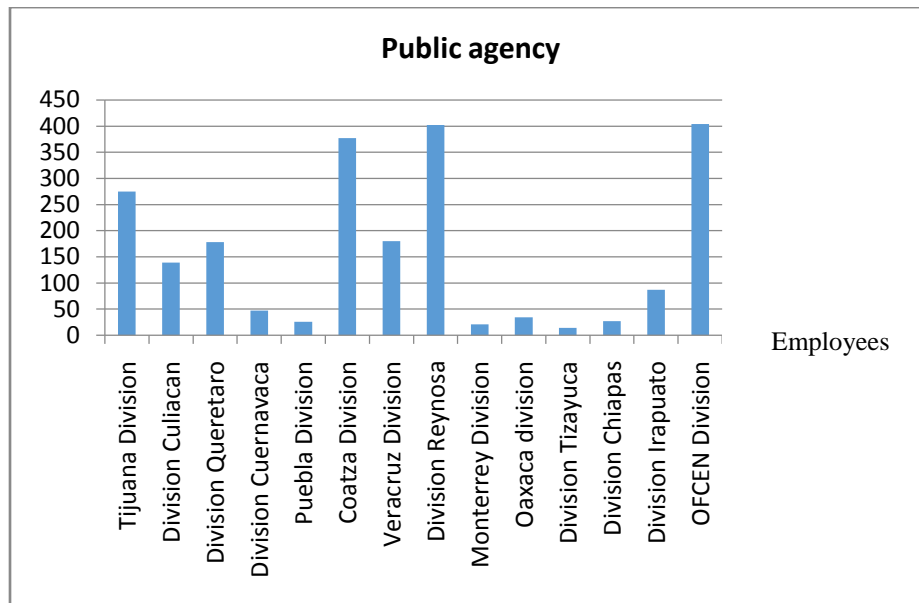

Figure3. Level of missing money by division

By reviewing national indexes, it was identified that there is correlation with the levels of insecurity in the geographical areas of these divisions (see figure 4).

Table 2.Criminal rate by state in México, source (CIDAC, 2015)

Rank	State	Index	Category (insecurity)
1	Tlaxcala	95.9	Moderate
2	Querétaro	95.3	Moderate
3	Aguascalientes	93.9	Moderate
4	Campeche	92.8	Moderate
5	Baja California Sur	92.6	Moderate
6	Hidalgo	92.6	Moderate
7	Yucatan	91.0	Moderate
8	Puebla	90.0	Average
9	Chiapas	89.7	Average
10	Nayarit	88.7	Average
11	San Luis Potosí	87.4	Average
12	Zacatecas	86.8	Average
13	Sonora	85.5	Average
14	Nuevo León	85.4	Average
15	Oaxaca	85.0	Average
16	Veracruz	83.5	Average
17	Distrito Federal	80.5	Serious
18	Tabasco	80.3	Serious
19	Jalisco	79.7	Serious
20	Guanajuato	79.6	Serious
21	Colima	79.4	Serious
22	Coahuila	79.2	Serious
23	Baja California	77.9	Serious
24	Quintana Roo	76.4	Serious
25	Durango	73.3	Severe
26	Michoacán	70.0	Severe
27	Estado de México	69.5	Severe
28	Tamaulipas	67.4	Severe
29	Chihuahua	66.6	Severe
30	Sinaloa	62.8	Severe
31	Morelos	49.4	Severe
32	Guerrero	6.8	Severe

Data of Coatzacoalcos and Veracruz were grouped as well as central offices and Cuernavaca, subsequently we proceeded to analyze the data with SPSS version 22, for this purpose weapplied a Pearson's bivariate correlation analysis as described in (Hernández Aguilar, 2013) of which the following results were obtained:

Table 3.Correlation analysis between the shortage of employees and the criminal rate by source State (own)

		Employees with shortages	Criminal Index
Employees with shortages	The Pearson correlation Coefficient	1	-.557
	Gis (bilateral)		0.60
	N	12	12
Criminal Index	The Pearson correlation Coefficient	-.557	1
	Gis (bilateral)	0.60	
	N	12	12

Table 3 shows that there is an inverse correlation between the numbers of employees with shortages with regard to criminal rate. Although the analysis did not reveal significant differences at 95% confidence level, it should be noted that the divisions of the north of the country (Tijuana and Reynosa) fall into the categories of serious involvement and severe in the criminal rate of the country, the southeast (Coatzacoalcos and Veracruz) are on the verge of average-severe affectation, and those of the central offices and Cuernavaca are contained in the lowest part of the category - severe affectation - according to (CIDAC, 2015).

6 Conclusions and Future Work

The data mining is a tool that allows you to automate the analysis of large databases, which facilitates cleaning, preprocessing and classification of data that can be used to find patterns suspects as are mismanagement and debts of the employees, derived from information from accounting records. Our preliminary results indicate that there is an inverse correlation in the criminal index by state and the level of missing (number of employees with shortages) in the analyzed Institution. Weka is a user-friendly tool that has demonstrated its practical utility for the analysis of large volumes of data from public Institutions. So the hypothesis under study is fulfilled, and it is possible to apply data mining to the receivable accounts of employees of a public organization for the identification of possible mismanagement and fraud.

This work is by no means complete, our future work consists in the validation of other patterns identified in the missing records, the construction of a user-friendly interface to the user based on components of WEKA and Java as well as modeling in decision trees to identify the profile of debtors and anticipate a not desirable behavior.

Acknowledgments

References

Altamiranda Echeverri, L., Peña Retamoza, A. M., Ospino De La Rosa, M., Volpe Barros, I., Ortega Álvarez, D., & Cantillo Guerrero, E. Minería de datos como herramienta para el desarrollo de estrategias de mercadeo B2B ensectoresproductivos, afines a loscolombianos: unarevisión de casos. Sotavento MBA, pp. 126-136. (2013)

Ballesteros Román, A., Sánchez-Guzmán, D., &García Salcedo, R. Minería de datoseducativa: Una herramienta para la investigación de patrones de aprendizajesobreuncontextoeducativo. Latin-American Journal of Physics Education, 7(4). pp. 662-668. (2013).

Calleja Gómez, A. Minería de datos con Weka para la predicción del precio de automóviles de segundamano. Valencia: universidadpolitécnica de valencia. (2010, Diciembre).

Castro, A., Sifuentes, E., González, S., &Rascón, L.H. Uso de Minería de Datosen el manejo de InformaciónGeográfica. InformaciónTecnológica, vol. 25(5).pp. 95-101, (2014).

Cevallos Chacon, A. Guíametodológica para obtenerpatrones de accidentabilidadlaboralusando Data Mining. Tesis de maestríaenDirecciónEstratégicaenTecnologías de la Información. (2013).

CIDAC. Indicedelictivo 8 delitos primero, http://cidac.org/indice-delictivo-8-delitos-primero-actualizacion-2013/?gclid=CPm_krSJt8sCFYGFaQodgjsAxA, last accesed 2015/07/10.

Connolly, T. M., & Begg, C. E. Sistemas de Bases de datos: Un enfoque práctico para diseño, implementación y gestión. Pearson educación.(2005).

Corso, C. L., García, A., Ciceri, L., & Romero, F. (2014). Minería de Datosaplicada a la Detección de factores para la prevención de incidentesinformáticos. XVI Workshop de InvestigadoresenCiencias de la Computación. pp. 168-172 (2014).

Fadairo, S.A., Williams, R.,Trotman, R.,&Onyekelu-Eze, A. Using data mining to Ensure Payment Integrity. Journal of Government Financial Management. Vol. 57(2), 22 (2008).

Fayyad, U. Data Mining and Knowledge discovery: Making sense out data. IEEE Intelligent Systems (5), Vol. 11(5), pp. 20-25. (1996).

Hernández Aguilar, J. A. Generación, tratamiento y análisis de informaciónen las organizaciones. Cuernavaca, Morelos: Juan Pablos Editor, Universidad Autónoma del Estado de Morelos, México, (2013).

Hernández Aguilar, J. A., Levín Kosberg, S., & Vázquez García, A. W. Aplicación de la minería de datos para la toma de decisiones: El Caso de la culturaorganizacionalenunatienda del IMSS. XVI CONGRESO INTERNACIONAL DE CONTADURÍA, ADMINISTRACIÓN E INFORMÁTICA, UNAM (2011).

Jiawei, H., Micheline, K., & Jian, P. Data Mining Concepts and Techniques. USA: Third Edition. (2012).

Miletti, A., Aquel, M., Berri, S., Diaz, A., Doffo, D., Ingrassia, G., et al. Multidimensionalidad de losDatosen la InformaciónContable. DecimoquintasJornadas "Investigacionesen la Facultad de CienciasEconómicas y Estadística. (2010, November).

Molero Castillo, G., &MedaCampaña, M. E. Integración de Minería de Datos y SistemasMultiagente: un campo de investigación y. Ciencias de la Información, vol. 41 (3), pp. 53-56. (2010).

Núñez Cárdenas, F. Introducción a la Minería de Datos. SistemasComputacionales. (2011).

NTMX. Imponenmulta de 71.7 mdpa OHL México; quedaabsuelta de fraude. El Sol de México. (2016, 03 28).

Ochoa Zezzatti, A., Romo, G., Bañuelos, F., Mendhizavili, K., Iztebegovic, H., & Hal, S. Herramientainteligente para la toma de decisionesbasadaenMinería de Datos. RevistaInvestigaciónCientífica. (2007).

Peña Ayala, A. Educational data mining: A survey and a data mining-based analysis of recent works. Expert Systems with Applications. Vol. 41(4), pp. 1432-1462, (2014).

Riquelme, J. C., Ruiz, R., & Gilbert, K. Minería de Datos: Conceptos y Tendencias. Inteligencia Artificial, RevistaIberoamericana de Inteligencia Artificial. Vol. 10 (29), pp. 11-18. (2006).

Ruiz, A., Hernandez, L., & William, G. Aplicación de lossistemas de soporte a la decisión (DSS) en el comercioelectrónico. Ingeniería e Investigación, Vol. 29(2), pp. 94-99. (2009).

SAP. (2016). SAP Latinoamerica. https://go.sap.com/latinamerica/index.html, last accessed 2016/09/03.

Tello, M., Eslava, H., &Tobías, L. Análisis y evaluación del nivel de riesgoen el otorgamiento de créditosfinancierosutilizandotécnicas de minería de datos.RevistaVisiónElectrónica, Vol. 7(1), pp. 13-26, (2013).