8th International Scientific Conference
Technics and Informatics in Education
Faculty of Technical Sciences, Čačak, Serbia, 18-20th September 2020

Session 2: *IT Education and Practice*

Professional paper
UDC: 004.891:336.2

Expert Systems as a Means in Detecting Tax Evasion

Dalibor Bogdanović [1*], Elmir Babović [2**]
[1] University of East Sarajevo, Faculty of Electrical Engineering, East Sarajevo, BiH
[2] Džemal Bijedić University, Faculty of Information Technology, Mostar, BiH
[*]daliborbogdanovic@hotmail.com [**]elmir.babovic@edu.fit.ba

Abstract: *The paper presents an overview of currently existing methods for the detection of tax evasion in databases using expert systems and compares the method of improving the detection of tax evasion with the proposed expert system. Detection of tax evasion and its efficiency in recognizing them are essential for timely and efficient detection of them and improving the knowledge base of the expert system. The paper presents an abstract solution using an expert system on the domain of tax evasion as well as a performance model. Expert system builder is a GUI interface for the work of employees in the specified expert system. This way of realizing the detection of tax evasion enables facilitation in the work of the competent tax services. The results of such a proposed system of expert systems are presented in the paper and give an enviable level of detection of tax evasion patterns.*

Keywords: *expert system; expert system builder.*

1. INTRODUCTION

Tax evasion has been present since the very beginning of the economic development of human civilization, where only the instruments, methods and techniques have changed, in accordance with the technological possibilities of society. Also, the appearance of tax havens in the world allows individuals and legal entities to place their funds earned in their home countries to deposit in destinations known as tax havens and thus commit tax fraud or evasion. In this way, the perpetrators of these acts keep their capital and do not pay taxes in the countries from which they come.

Today, artificial intelligence and expert systems [1] are indispensable in the work process and everyday life, from simple technical software or machine aids to complex systems. With the development of science and technology, their importance grows and has a great influence in business, education, economics, informatics, etc.…

In the continuation of the paper, some ways of tax evasion-fraud will be presented. Simple or more sophisticated-sophisticated such as tax evasion by trading on virtual stock exchanges where cryptocurrency is traded.

These trends, the models of tax evasion available to "criminals", are constantly changing and improving in order to hide the traces of illegally acquired funds. Over time, tax evasion also improves in line with the development of technology and globalization. Due to these trends, all world institutions and financial regulators must be ready and willing to reveal patterns in order to reduce this phenomenon to a tolerable level and preserve their own economy and development.

2. APPLIED EXPERT SYSTEM IN DETECTION OF TAX EVASION

The use of expert systems and artificial intelligence in the process of discovering patterns of tax fraud has been documented in a number of papers published in journals and professional-scientific conferences around the world. In order to form the basis for comparative analysis of the possibility of their application in specific conditions of analysis of the fiscal system in BiH, as well as evaluation of the obtained results, it is necessary to analyze the selected set of available references taken from tax systems in BiH and in the world. In the process of analysis of a representative sample of references, emphasis was placed on documented examples of expert systems and mechanisms in the process of analyzing data on performed financial transactions within a set of organizational systems.

2.1. The use of artificial intelligence to detect fraud in Brazilian customs

This approach to identifying possible fraud is based on the interaction between customs service and Decision Support System DSS developed under the name Carancho [2], and it finds suspicious operations by detecting outliers (out-of-interval values). Such a system assumes that most international trade is legal, ie they are in

accordance with the law and the products are classified accordingly. Therefore, an Information System for Products and Foreign Exporters (PFEIS) [2] has been developed that uses orthographic verification features that suggest possible duplicates and assist in their classification.

Such fraud detection systems are important to reduce the manual work of customs officers in product inspection and to achieve the maximum amount of fraud detection. This system must be presented with several problems in its work, such as a large number of attributes with which the system operates, then the imbalance of the database and inaccurate spelling of the product in relation to which the data in the database are compared.

2.2. Income tax audits using the Bayesian network method in Brazil

As is generally known in all countries of the world there is a sales tax and income tax as tax deductions to individuals and legal entities in their businesses. This method primarily refers to the audit of personal income tax in Brazil. The basis of her work is on a probabilistic analysis of income for individuals in real time. In this scientific study, a conceptual predictive model was modeled, which gives quite good results in predicting whether a particular taxpayer's income is harmonized with the regulations of the regulator (tax administration of Brazil). Data analysis of all taxpayers is based on machine learning and CRISP-DM (Cross industry standard processing for data mining) standard. By training the Naive Bayesian network as well as by training the pronounced decision tree of the Bayesian network, a model of detecting the detection of compliance or non-compliance of taxpayers' tax revenues in Brazil is achieved. [3]

2.3. Detection of value added tax evasion of business entities in Kazakhstan

The basis of this paper is on data discovery of knowledge, ie on data mining for the target group of business entities in the state of Kazakhstan. This is a machine-oriented approach to learning the data that businesses have in their work. The analysis is taken for the basic characteristics (attributes) that are significant and a comparative analysis is made in relation to the method conducted by the tax administration of the state of Kazakhstan. [4]

Data exploration or data mining is based on Kohonen's self-organizing maps and clustering of samples. Such maps are suitable for easy review of the multidimensional attributes of the samples being analyzed. Data collection is in accordance with the rules of the Tax Administration of Kazakhstan with a special target group of taxpayers whose turnover does not amount to less than one million tenge of Kazakhstan. Visualization of this data is by means of self-organizing maps. Only the calculation and detection of possible deviations or tax evasion is based on the static approach of the Gaussian distribution. It was taken into account that the proposed approach was tested on a specially selected set of business entities.

3. ABSTRACT SOLUTION MODEL IN DETECTION OF TAX EVASION PATTERNS

Depending on the given case on the domain in the tax system of one institution or state, the applied expert system can give a pretty good assessment of the case of tax fraud.

Figure 1 presents an abstract model of solutions in the detection of potential tax evasion or fraud in the tax systems of Bosnia and Herzegovina.

As previously stated, the basic facts and facts available to the knowledge base of the expert system are stored in the repository of knowledge by extracting cases from the Criminal Code which prescribes penal provisions for certain cases known over the domain of evasion. Figure 1 shows an abstract solution in the form of an expert system or knowledge base in which there are combinations of knowledge provided by an expert in the field of tax fraud and the Criminal Code. By combining the two knowledge in building an expert system, the knowledge engineer gives the final look of the knowledge base. Also the knowledge engineer designs and designs what the reasoning machine will be like and what its appearance or design will be like.

Figure 1 shows the reasoning mechanism that gives the result of the search requested by the user depending on the case that interests him over the domain of tax evasion or fraud.

Figure 1 shows the cyclical or interactive exchange of data and knowledge exchanged by users of the expert system and the system itself intended to solve the specific problem of tax fraud. Through the application programming interface (API), the user queries using tools such as Java system shell, Prolog, Lisp or Expert system shell for individual cases that interest him over the "evasion" domain. he expert system or the reasoning mechanism mechanically searches the knowledge base using all the resources of one expert system specifically designed for special cases. The inference mechanism works in a way that it cyclically and cross-searches the knowledge base and draws conclusions that it gives to the user in real time.

Figure 1 shows the basic four segments consisting of the User Interface (user interaction tools and natural language), then the inference mechanism (Heuristics, reasoning schemes, inference), then the rule library (knowledge-managed rules) and the database. knowledge (facts and data that represent expert knowledge). This pattern of search and knowledge extraction mechanism is very efficient

and gives good results in cases of tax fraud in support of the work of users (tax officials or others).

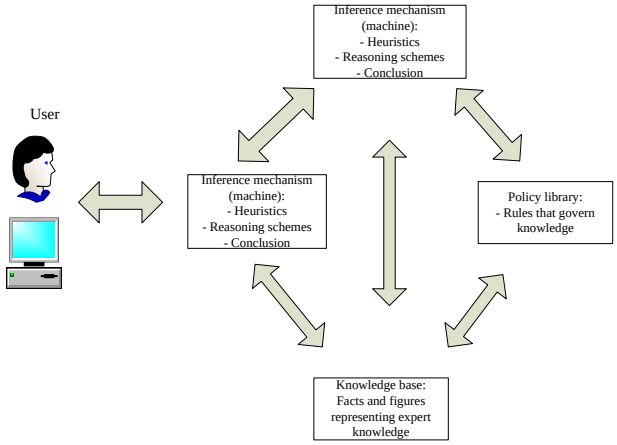

Figure 1. *abstract model of solutions in the detection of potential tax evasion or fraud in the tax systems of Bosnia and Herzegovina*

4. IMPLEMENTATION MODEL FOR DETECTION OF TAX EVASION BY EXPERT DATA SEARCH

Figure 2 presents an implementation model of a solution for the prevention, detection and prevention of tax evasion. This model consists of three blocks, which are (reporting-data collection block), then (transformation or data translation block) in which the data are prepared for the block of expertise or expert data collection. shows the expertise of the data collected by the expert module for knowledge extraction. Each taxpayer from his workstation via a link on the Internet reports or reports to the regulator on the transaction of the party or makes a report to the competent tax administration.

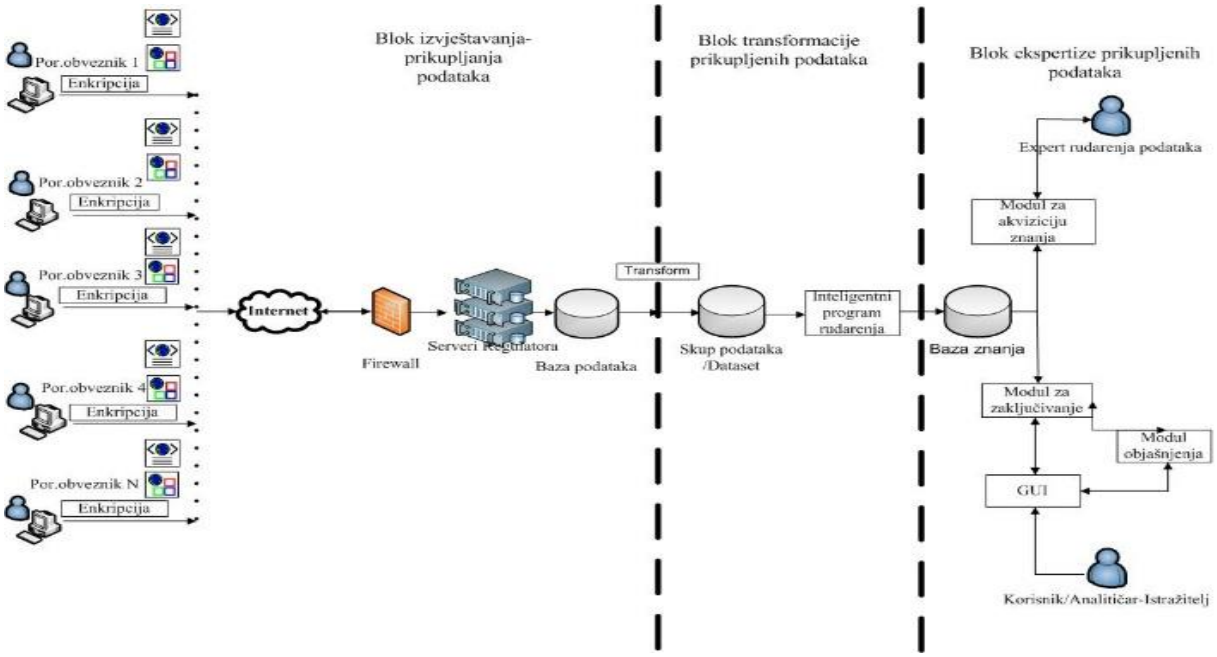

Figure 2. *Acquisition of discovered knowledge*

Taxpayers can report completed transactions of parties in single or batch mode, ie they can file one or more tax returns at once. Figure 4.8 shows the method of filing created by an authorized person for tax returns that are sent via a web service to the server of the tax regulator that is on the Internet. These file-transactions are encrypted and protected from attacks on the Internet. Reported transactions are transmitted via the Internet, protected by encryption, to the controller's server. The section shows the firewall that separates the Internet from the local intranet within the tax institution. All transactions are received on the servers of the regulator and they are stored in the database warehouse for storage.

By transforming structured data through in which a set of data suitable for expertise and analysis is created. The next step in is the application of an expert detection program over the data set and the extraction of the pattern from the data set. Permanent expert search of data forms a knowledge base. A collection of all the patterns found creates a knowledge base.

The next step in of the implementation solution is the acquisition of the discovered knowledge (acquisition module in Figure 2) which serves for the analysis and expertise of the knowledge collected from the database by the data analysis expert. Based on this module, the data analysis expert simplifies the knowledge found and helps all interested users of the system. This part of is designated for experts in the field of data analysis. Furthermore, the module in is a conclusion module that serves primarily users, and these are officials who investigate the occurrence of "tax evasion". This module interacts with system users through two modules, namely the explanation module and the GUI (graphic user interface). Through the GUI, the system user uses the knowledge found to analyze and investigate tax evasion.

5. EXPERT SYSTEM BUILDER

A practical example is given in Figure 3, which is the reasoning mechanism on the basis of which the conclusion is made and assistance in the analysis of transactions from the tax aspect of the tax officer. This reasoning mechanism is based on the decision tree as shown in Figure 3 The starting node is a transfer of funds, which is logical, because it is most likely according to the theory of information gain (Shannon's theory) [5] if it starts from this starting point.

Figure 3 is the starting point for whether the origin of funds is known and this is the basic premise of any analytical approach to tax evasion research. Without this evidence, there is no need to analyze every transaction within the fiscal system because the security factor (factor certainy is approximately 1) if the funds do not originate to be tax fraud or evasion.

This reasoning mechanism provides exact data to each tax officer and is a clear approach that can be used to analyze each set of data collected by each tax administration and which they have at their disposal in their databases. Furthermore, an important decision tree of the decision tree is the amount of the transaction that works in accordance with the analysis with (Police Agencies at all levels of government in BiH). If the amount of the transaction is less than 30,000.00 KM (cash transaction, non-cash, cash related) it has a high probability that it is related to tax fraud or tax evasion. (Safety factor is greater than 50%). A further important node of the analysis of this financial transaction is the review (view) of whether the legal or natural person regularly settles tax liabilities and especially for the performed transaction that is being analyzed and which is important for the tax investigation. This approach again gives a safety factor of approximately 1 according to Shannon's theory.

A further important node is the examination of whether direct and indirect tax debts have been settled by a natural or legal person, and especially for the analyzed transaction which is important for the investigation. This approach again gives a safety factor of approximately 1 according to Shannon's theory.

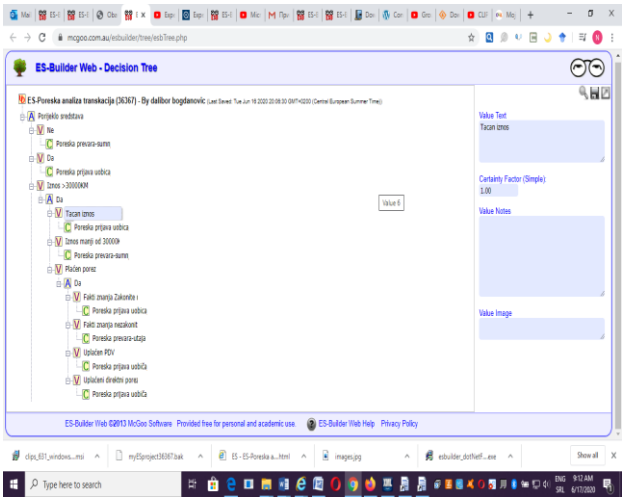

Figure 3. *Expert system builder for this model*

6. EXPERIMENTAL RESULTS

This chapter presents the experimental result of the success and efficiency of the expert system in working on a data set of an institution that fights against tax fraud or tax evasion. This approach can be applied to the global payment base in Bosnia and Herzegovina with appropriate empirical assumptions. As previously stated in the paper, it is shown that it is based on rule based rules (Legislation regarding taxpayers' tax returns) . In order to determine the results of the information search over the data set, the standard classification procedure for calculating the reported tax data was applied.

According to the required classification into "tax evasion" and "ordinary tax returns", the classification of the database can be simply described by a binary classification. The ID3 binary tree algorithm was used for this particular case. This algorithm is simple and practical to apply over almost all data sets. By applying this classification algorithm to the data set available to the financial institution over the data set, a clear model of tax evasion pattern recognition can be obtained. The first step of the ID3 algorithm of classification "origin of funds". The second step is "the type of transaction reported", the third step is the amount of the transaction, the fourth step is "whether the tax has been paid" and the fifth step is the final and these are the facts available to the expert. In binary data set classification, precision and recall are the two main parameters.

A simple metric was taken for this practical case. Two basic parameters precision and recall are

defined and they are represented by equations (1) and (2). [6]

$$precision = \frac{t_p}{t_p + f_p} \quad (1)$$

$$recall = \frac{t_p}{t_p + f_n} \quad (2)$$

The database of one tax institution contains a total of 3.597.933 reported tax returns. Of that amount, 1,891 were reported as "tax evasions", and 3.596.042 were reported as "ordinary tax returns".

For the purposes of this research, one set of data was taken and extracted from the database of the tax institution. This dataset contains transactions characterized as "tax evasion" and "ordinary tax transactions". This situation is not realistic because it is based on the legal recognition of "rule based" and it hides a large number of "tax evasions". This is the main goal of the study, recognizing or discovering hidden "tax evasion".

The current situation does not reflect the true situation in the set of tax transactions. It is clear that the accuracy of the data set of the tax institution is approximately 100% but the response over the data set of tax transactions is very small and is approximately 5%. It is clearly concluded that the accuracy of the proposed expert system is only slightly less than the accuracy of the current tax system, but the response is ten times better, ie for the order of magnitude of tax transactions.

Table 1 shows the parameters of precision, recall for 2019.

Table 1. *Results*

Transaction collection period (years)	Precision (%)	Recall (%)
2019	95.8	51.1

7. CONCLUSION

The paper specifically develops a model that, based on legally defined models (database of facts) of tax evasion, finds new models of tax evasion for a given set of data or data sets available to tax administrations or other institutions in the country. The analysis of the found new forms determines the relevance of the found forms for the next use (reusability) and they are automatically added to the knowledge base available to the system. This model is dynamic and can be used for a large number of data that are in the database and that arrive online by taxpayers who report on tax returns. All found patterns are compared with known models of tax fraud to gain additional relevance. This expert system provides a fast and efficient way to detect new models of tax fraud or evasion of tax systems in Bosnia and Herzegovina.

This system is the future in the research of new forms of tax fraud and relies on artificial intelligence, which in the foreseeable future will be much more widespread in this type of problem as well as in other fields of science and economy. The application of artificial intelligence in the detection of tax fraud patterns also reduces the error of users who deal with the occurrence of tax evasion, which is the goal of any decision support system.

REFERENCES

[1] Topolovec,V., Bajgorić,N., (1989). Inteligentni informacijski sistemi za podršku u odlučivanju. Journal of Information and Organizational Sciences. Vol. 7, br.13, str 220.

[2] Digiampietri, A., L., Norton T., R., Meira, L. A. A., Filho, J., J., Ferreira, C., D., Kondo, A., A. (2008). Uses of Artificial Intelligence in the Brazilian Customs Fraud Detection System. The Proceedings of the 9th Annual International Digital Government Research Conference.

[3] Leon S. D. S., Rigitano H., Carvalho, R. N., Souza JCF. (2016). Bayesian Networks on Income Tax Audit Selection - A Case Study of Brazilian Tax Administration.

[4] Assylbekov, Z., Bekishev, R., (2016). Detecting Value-Added Tax Evasion by Business Entities of Kazakhstan. Nazarbayev University Kazakstan.

[5] Sergio, V. (1998, Oktobar). Fifty Years of Shannon Theory. IEEE transactions on informations theory, Vol. 44, No. 6.

[6] Nitesh, V., (2009). DATA MINING FOR IMBALANCED DATASETS: AN OVERVIEW Chawla Department of Computer Science and Engineering University of Notre Dame IN 46530, USA.