PROPOSAL FOR THE USE OF A PREDICTIVE MODEL FOR DETERMINING THE RISK PROFILE OF TAXPAYERS

Beatriz Steinberg

SUMMARY

To control the behavior of individuals and enterprises plays a main role in the implementation of policies that are the duty of AFIP.

This work is just a proposal to improve the control process, focusing on optimizing the detection of risk groups. It proposes to add to the AFIP battery of resources, the use of neural networks for the establishment of risk profiles, guiding the process and making more rigorous the classification of taxpayers based on potential non-compliances.

The Author: *Degree in Computer Science and BA in Psychology from the University of Buenos Aires; Masters in Strategic Business Management, University of Palermo; Master of Software Engineering with a joint degree from the Buenos Aires Technology Institute and the Polytechnic University of Madrid; Control Analyst of the Systems and Design data Department from the Control Branch, AFIP; Associate Professor at the University of Palermo, Faculty of Engineering and a Masters in Software Engineering.*

CONTENT

Introduction

1. The Proposal
2. Development
3. Evaluation
4. Conclusions
5. Bibliography

At present, most countries have been working on the formalization of strategic planning processes of their public administrations in general and particularly of the tax administrations.

A comparative analysis published by the Federal Administration of Public Revenues (AFIP) in January 2007 - Covering Argentina, Chile, Costa Rica, Mexico, Colombia, Nicaragua, Peru, Dominican Republic, Australia, USA, Canada, Venezuela, Ireland, Holland, Brazil, shows that the considered tax administrations mostly develop strategic plans with the tendency to direct the mission of these organizations to provide quality service to taxpayers, ensure the efficient implementation of laws and seek voluntary compliance with tax obligations.

In accordance with this mission surges the trend to set the strategic vision in achieving a quality service to taxpayers and the modernization of organizations relying primarily on the use of new technologies and information systems and a qualified staff. To effectively implement this Vision, the Strategic Objectives focus on the need to optimize the control and fight tax evasion, improving the service provided to the taxpayer, increasing efficiency in managing the organization, the application of new technologies processes and systems, and human resource development.

In the implementation of policies that compete to the AFIP, which rise from the above mentioned strategic objectives, policing the behavior of individuals and enterprises plays an essential role.

The first link in the control process is the research tasks, to analyze cases that are likely to have a significant fiscal interest. The investigations that are of interest are subject to monitoring tasks that may or may not be successful, understood as adjustments made and charged- they point to the taxpayer and the consequent avoidance of disputes, both administrative and judicial-and sanctions that can be applied.

Both research and the different types of existing controls are time-consuming human resources, the choice to initiate and continue some of them implies not to be able to address other options.

In this context it is important to make the process efficient as a whole, and pursue those investigations that will conclude successfully. This problem is not unknown to the AFIP, which invests efforts and resources on the widespread use of computer tools, both in qualifying taxpayers based on their potential risk and in the implementation of management control processes.

1. THE PROPOSAL

This work arises from the detection of a need - which the tax administrations faces when evaluating strategies for the promotion of tax compliance - collects and processes a series of data, and obtains a model for defining and qualifying a particular group of taxpayers - the Large National, according to their risk profile.

The choice of the universe of analysis is based on the smallness of the group - about 0.03% of total taxpayers - and the interest generated in the AFIP for their participation in the collection, more than 48%.

The present proposal is in the context of a working methodology suitable for such developments: CRISP-DM 1.0, hierarchical methodology that provides an overview of the life cycle of a data mining project. [Chapman and others 2000]
Concerns about the feasibility and desirability of the proposal are answered in the following statements:

- The tax administrations have permanent storage for a large volume of data
- There is a high correlation between the reinforcement of the audit tasks and the decrease of tax evasion.
- The perception degree of fiscal behavior of the other taxpayers and the impunity of the big fraudsters are operating as a justification of evasion.
- The characterization of Large National taxpayers as pragmatic compliers - at any time, they decide if they do or do not comply based on a selfish calculation of chance or the result of the equation time spent for complying versus the obtained benefit by complying, for them, a more efficient administration in its oversight task becomes the most effective tool to improve compliance with tax obligations.
- The trend to minimize or avoid the tax burden by taxpayers in the absence of effective control tasks
- The need to identify, as early as possible, related practices, related to, at least, breaches of regulations. [Russo, 2010].
- The AFIP conception of control as a logical and systematic process that requires the development and implementation of processing tools and detailed analysis of individuals, transactions and operations, to identify segments which implement specific actions that lead to increased voluntary compliance and detect and prevent criminal, evasive and elusive maneuvers.

1.1 The innovative nature of the proposal

This paper proposes an innovation process – understood as the application of new ideas to old problems, always looking for significant improvements in efficiency, effectiveness and quality - and the possibility of the use of Information Technology and Communication (ICT) to make possible this process. [Estevez, 2009].

All innovation involves risk, which can be minimized [ANAO, 2009]. In this sense the innovation associated with this work has the following characteristics, all risk-reducing:

a. It is a refinement of an existing process.
b. It does not extend the changes to new areas.
c. It is not a radical change from current practice.
d. It is formulated with clear objectives and precise boundaries.
e. Its implementation does not seem complex once the research stage is over.
f. The application fee does not exist and significant benefits are expected.

1.2 The Information and communication technologies (ICTs)

An analog society is taking place, moving from real objects such as paper to digital ones, where bits are moving through broadband networks and in which information in text, voice and image formats is being unified in the concept of multimedia. According to Juan Hernandez "Talking about Information and Communications Technologies(ICT) for the Tax Administration Service is equivalent to talk about ICT for the facilitation and efficiency in the execution of processes in organizations".

On that basis it is striking that among the technical issues addressed in all CIAT Technical Conference, from 1997 to date, there are paper works related to the ICT

Regarding the linkage between ICT and CIAT tax administrations, today in most of them, the use of information in the fight against tax evasion is no longer discussed. All of them work with different sources of information, whether provided by taxpayers, through legally enforceable claims, or obtained from third parties through agreements, compliance with information conventions, international agreements and integrated information systems- . All are present as essential requirements the quality and security of the information. All prioritize Internet use. Most of them have developed specific IT tools to fight tax evasion.

As in other areas, the introduction of ICT in tax administrations requires placing them to the service of objectives; in this case what is wanted is to identify ways to make more efficient the processes for the management of taxes, to achieve this information technologies must be developed in line with the strategic objectives of the AFIP. This coincidence requires explicit information technology policies, since their absence allows the establishment of implicit policies that, in general, are potentially harmful to the organization, since their foundations are not clear in all cases, they are not documented (or they are but in a very precarious way) and generally responds to the interests of information technology providers, seeking to create a captive market.

Once actions have been taken, it is necessary to measure their impact on the organization, or detect if they have impacted the strategic objectives of the organization or if they were reduced to a mere mechanization, with or without reduction of costs. In the Tax Administrations there are criteria added to the above measurement:

a. The consideration by society in general and taxpayers in particular on the efficiency, transparency and credibility of the administrations
b. The increase of voluntary compliance by taxpayers.
c. The increase in collection.
d. The need to comply with the requirement of publication of accessible services by society.
e The construction of a useful reservoir, from the large volume of data for predictive systems.

Thus an intensive use of ICT should not leave out their use to expand the analytical capacity of tax administrations, transforming data and information processes raw pursuit of knowledge, so as to help improve policy development and decision-making. In this line appears the Data Mining.

1.3 Data Mining

Data Mining can be defined as the exploration and analysis, by automatic or semiautomatic means, of data to discover patterns and rules; the preceding description, using the concept of "discovery" points to patterns and rules that should be hidden until that moment, not known and that it is not necessary to have previous questions or insights to reach them. Similarly Jiawei Han highlights the features that standards and rules should have: be non-trivial, previously unknown, implicit in the data and potentially useful. The emphasis on the notion of discovery

must reconsider the role of verification as part of the taxonomy of data mining.

Tax administrations are among the largest producers, collectors, consumers and disseminators of information in each country. The possession of large amounts of data permanently stored place them in a position to appeal to automatic or semiautomatic procedures on that data to find hidden knowledge to date and interesting hidden patterns, associations, changes, anomalies and significant structures in the data. The great computing power they possess, coupled to said data volume, and enables them to address given data mining processes.

The Data Mining, for what has been said, is then a subsequent process to obtain the data, which looks for generating information similarly to the one,, useful and understandable a human expert would produce, which is, it is a link in a broader knowledge production process and involves the application of algorithms for extracting patterns, using the previously available data, which thus acquire more value.

Because neural networks will be used, some approach to the subject seems appropriate. It is about the adaptation of interconnection of brains neurons with digital computing models. Neural networks are defined by their topology (organization and arrangement of neurons of the network layer), learning mechanism (creation and destruction of connections between neurons as well as changes in their weights trying to minimize the error), the type of association between information input and output (forwards, backwards, recurrent, or any combination of them) and how to represent the data and outputs (values continuous, discrete). They are used for classification problems, estimation, and pattern detection.

1. 4 Segmentation

The formulated problem statement refers to Large National Taxpayers. This implies a previous segmentation, grouping the elements of the universe being studied in homogeneous segments with respect to predefined criteria, which are precisely the determinants of segmentation.

The concept of segmentation in the case of tax administrations generally aims to identify, based on a concept of reliability-defined from the standpoint of taxation, customs and social security, those segments which run, timely, reasonable and affordable different control actions. This is for taxpayers for whom special procedures are defined both in attention and control.

In the CIAT member countries this segmentation process has two levels: while all of them offer the partition of “computer science” from taxpayers to offer a differentiated service according to their size, activity, tax regime, nature of their main income, etc., some take the concept of segmentation to the organization by types of taxpayers, as in Argentina.

The need for segmentation surges in the AFIP on the detection of a small group of taxpayers, with distinct characteristics and a high participation in the collection. The distinct characteristics of this group are:

- Complexity in tax-related operations.
- Propensity to litigate (have advice of professionals from law firms and / or accounting major).
- Rejection of the adjustments identified in inspection stage: involves starting the official process of determination in most inspections.
- Non-acceptance of most of the resolutions and appeals before the National Tax Court Office.

The response of the AFIP is segmentation at various levels.

The structural correlation of the segmentation is the creation of the Central Sub-zone, incorporated into the Organizational Structure of the Internal Revenue Service by Decree No. 1.745/74, which then created the National Taxpayers Directorate

by Resolution 278/87

At the computer level, segmentation acquires influence with the TWO THOUSAND System - Special Differential Control System-which was established in an attempt to minimize tax evasion and tax non-compliance of the most interesting taxpayers and that decentralizes the capture of information in places where it occurs. At the end of 2006 it is absorbed by the 2000 REGIONAL SYSTEM. Finally from July 2008 the system called TAX ACCOUNTS - designed to register and provide information on debts and loans from taxpayers and third parties as well as the means used for their cancellation, is mandatory for the Large National Taxpayers while 2000 REGIONAL remains only for the administration of obligations prior to that time.

At the level of risk analysis, by RG 1974/2005, amended by RG 2166/2006, the computer system "Risk Profile System (SIPER) is approved, in order to categorize taxpayers and / or representatives, previously divided into groups by trading volumes and activity- according to the degree of compliance with their tax formal and / or materials tax obligations, in five (5) categories or segments (A, B, C, D and E), in increasing order indicating the risk of being audited (Category A: low risk of being audited, category E: high risk of being audited). And this system is precisely the starting point for the proposal of this work.

2. DEVELOPMENT

2.1 Basis for Development

It is convenient to explain the critical success factors of the proposed solution both from a business perspective as well as from the process of Data Mining.

Critical success factors of the proposed solution are the maximization of the collection - which should result in a reduction in non-compliance in the National Large Taxpayers segment-, improving the external image of the AFIP - measurable by the number of taxpayers who accept / appreciate the performance and the number of cases in which the assigned risk profile is questioned - the prevention of fraud - which should result in an increasing number of successful audits suggested by the tool and an increase of the recovered amount - and the costs involved in the project - measured in terms of the relationship between resources employed and tax revenue achieved -.

With regard to critical success factors of the process itself, there are the typical measures of efficiency models, the acceptance of the result by experts and display of the results to the community.

The tools to be used are those available from the desktop and, for the discovery of patterns, Weka 3.6.1 (Acronym for Waikato Environment for Knowledge Analysis, produced by the University of Waikato, New Zealand). WEKA is an environment for experimental data analysis that allows applying, analyzing and evaluating relevant techniques of data analysis, mainly those from automatic learning, for any set of user data. It has open source packages -adaptable for any project with potential to be enriched with new algorithms for the users, which include initial technical data preprocessing, as well as classification, clustering, association, and finally displaying of the results.

2.2 Understanding the data

The initial data collection is greatly simplified, due to the high level of computerization of the AFIP, which records in its data base centralized

all the new data of taxpayers and unify them, whenever this is possible, around the Single Tax Identification Code. The initial data to explore are the ones stored quarterly in working files to meet the requirements of the risk profile Measurement System, which bring together, for each taxpayer and in one single registry, all data pertaining to their tax behavior.

The data thus collected are 3547 records with 88 fields per record, including data about the tax behavior of the National Large Taxpayers for the quarter of 2009.

As for the characteristics of variables

- Most of them are of categorical type and indicate the presence or absence of divergence.
- There are discrete numeric variables (number of criminal cases, number of employees) and some continues variables (debt), for the large amount of securities offered, must be treated at the time of using them.
- The existence of out of range values detected in determined variables that must be treated as discrete, creating ranges and grouping in one of them all the uncommon excessively large values.
- There are missing values in all cases for certain attributes, which represent situations that managers have decided not to keep collecting, so they decided to eliminate them
- The dominance of certain values in certain attributes in every category, which leads to the assumption that the variable will have little predictive value
- Redundant attributes are detected, which are eliminated
- The presence of some data that do not apply to the entire universe under study is detected, in which case the calculation of correlations with respect to the class confined to the group to apply throws correlation coefficients not too different from those obtained by working the total universe of Large National, so the issue is not considered a problem.

As to the semantic of the data

- Taxpayers from the selected universe are composed of a 36.48% of individuals and 63.52% of legal persons, of which the majority, 89.3% are corporations.
- In a first approach to data it is possible to detect that the percentage of penalized taxpayers or with detected non-compliance is low and less than the percentage that results from considering the total universe of taxpayers. And the number of trials pending litigation (30%) is high compared to the completed ones in favor of the AFIP, even partially. (5% and 2% respectively).

2.3 Preparation of data

Once the class variable is defined, the preliminary use of the weka tool sheds light on the most significant attributes using a series of selection reviews, thus adding to the use of correlations to delete attributes with very low predictive value.

Missing values for an attribute do not require the construction of special values, on the contrary, the absence of values, when given, far from being a problem concerning an unknown value; it is relevant and marks a real event (e.g. the absence of submitted affidavits when they are not required).

The volume of available records makes working with samples neither necessary nor advisable.

A series of new attributes that group and weight divergences are created, building indices.

Sets of differential data are built in which the prediction attribute is numeric or categorical and in which independent taxes are nominal, using S / N when there are two options and Good / Fair / Poor when working with three options for assigning scores to the divergence.

2.4 Preliminary Modeling.

In a preliminary attempt to create a grouping scheme of taxpayers included in this study using cluster analysis, with the belief that knowledge of available data will increase and that this is a good starting point for all subsequent search for hidden patterns in the data.
This attempt to find a natural grouping between the considered instances according to the similarity that the observed variables show between them is implicit expectation that the grouping sought, when working on attributes that indicate deviations in fiscal performance, resulting in convergence with classification according to the current risk profile and let the "good payers" in a group and the "regular" and "bad" in others.

It is therefore not the result despite the use of the tool in various forms.
It can be assumed that the allocation of categories by the user, by defining fixed and potentially arbitrary cuts in certain scores stop other side of the "frontier" to similar instances, which in a clustering algorithm are left to the same group. In itself the idea of working with score also summarizes those divergences with very low incidence, which a grouping routine does not consider. From the experiences we conclude that cluster analysis cannot generate clusters in solidarity with the categories today defined by the user.

3.5 Neural Networks

The attempt here is to use neural networks for classification of taxpayers in the five categories previously defined by the AFIP. To do this, from an existence of a finite number of classes and assignment to a set of training data, to build a model for each class that can be used for classification of future data

The chosen parameterization, the method of testing during model building and the command used are the followings:

Decay	FALSE	Causes the decrease in the rate of original learning, helping avoid divergences
Autobuild	TRUE	Added and connects the hidden layers of the network.
Hidden Layers	-H a	Defines the number of nodes in the hidden levels of the network, separated by commas. Supports wildcards 'a' = (attribute + classes) / 2, 'i' = attributes, 'or' class = 't' = attribute + classes
LearningRate	-L 0.3	Learning rate or proportion in which the weights are modified
Momentum	0.2-M	Momentum applied to weights during the modification
Nominal Tor Binary Filter	<-B / ~~b~~> <T/F>	Preprocess to instances with a filter. Increases performance if atribut or nominal s and n dat you. It is irrelevant in this case
Normalize Attributes	TRUE	Normalizes the attributes, even nominal, between -1 and 1, to increase performance
Normalize Numeric Class	<-C / ~~b~~> <T/F>	Normalize the class if it is numeric, and only for internal ma, between -1 and 1. It is irrelevant in this case
Reset	TRUE	Allows the process to start automatically re with a lower learning rate if detected divergence.
Seed	-S 0	Used to initialize the random number generator for of the setting of the initial weights of the connectionsbetween nod or s.
Training Time	-N 500	Number of cycles for training.
Validation Set Size	V-0	Percentage of valid set If it is not 0, training continues until the error in the validation set is reduced or training cycles are covered. If 0, the validation set is not used and the training is for the number of indicated cycles
Validation Threshold	-E 20	Used to determine validation. The value indicates how many times within an instance, the error must be reduced for ending the training.

Test Options	Cross Validation 10 Folds	

Java weka. classifiers.functions.MultilayerPerceptron -t-totalweka -L 0.3-M 0.2-N 500-V 0-S 0-E 20-H a-G-R-d modeleclasify.out

The resulting model, with 21 nodes, has a low error level and high level of coverage, as shown in Table 1 which also discriminates both concepts for each class.

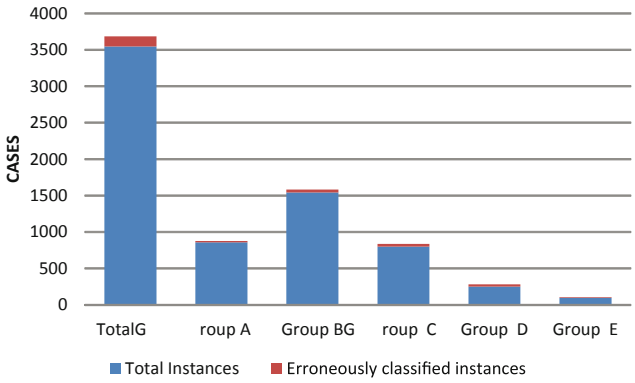
Table 1

Proportion of error in the model

CASES	COVERAGE		ACCURACY	
A: 857	0,989	848/857	0,977	20/868
B: 1543	0,985	1520/1543	0,975	39/1559
C: 799	0,939	750/799	0,955	3 9/789
D: 251	0,849	213/251	0,869	3 2/2 4 5
E: 97	0,804	78/97	0,907	8/86
Total: 3547	0.97		0.961	

Graphic 1

Proportion of error in the model.



3. EVALUATION

The evaluation is an inevitable stage in a Data Mining project and is subject to certain conditions as determined by the type of model (descriptive or predictive), the business to which the model is applied, the initial objectives and the intention of the model recipient. It does not only deals with technical issues, but with the business and can expose issues such as pattern matching not important to the business, poverty in terms of the knowledge generated, "over learning" or need to enrich the basic data in terms of size of records or attributes.

In this case, since it tries to achieve a correct classification of taxpayers based on their compliance, precision is crucial; so for this the metric chosen to test the behavior of the model is determining the percentage of misclassified tuples and it is applied to a new data set, 1362 relative records on the first quarter of 2010, which is performed on the same preparation as the one aforementioned.

It is also important to note that at the time of evaluation it must be established if all misclassification have equal weight or if it is more serious to evaluate as safe in terms of tax compliance a really risky taxpayer than to consider as high risk a taxpayer who is not.

The evaluation result is almost 70% correct (916 cases for 1362)

The first issue to be analyzed in order to understand the errors of the model is to compare the incidence of categories proposed by the model with the real incidence of those categories.

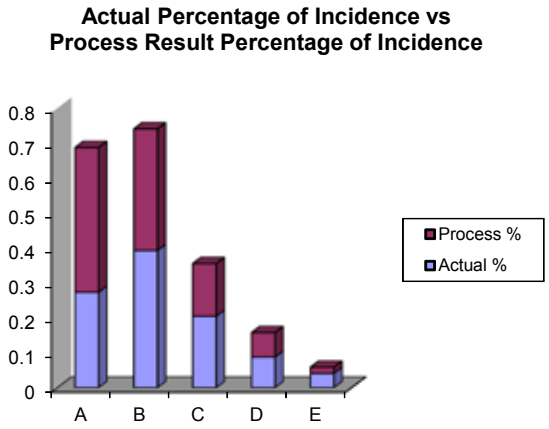
Table No. 2

Incidence of each category after the application of the model

CATEGORY	REAL IMPACT	AWARD
A	0.27312775	0.41409692
B	0.3928047	0.34801762
C	0.20484581	0.15051395
D	0.08810573	0.06975037
E	0.04111601	0.01762115

Graphic No. 2

The actual impact of each category in the evaluation



The model seems to be "generous" in determining the potential level of non-compliant taxpayers, which may be explained by the presence of deviations with little support, that the classification algorithm does not consider. The tool to better place taxpayers in a category better than the one assigned by users, increases the share of category A in the universe, bringing it from 27% to 41% and decreases the participation of category E, moving it from 4% to 2%.

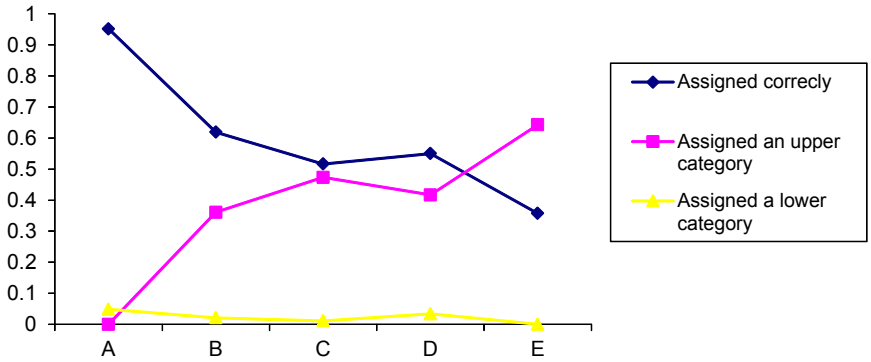
It is necessary then to evaluate for each category the percentages in which the model is correct and those where it is wrong, discriminating whether the error is in the sense of providing a lower risk to the taxpayers (qualifies better) or providing them with increased risk (qualifies worst).

Table No. 3

Details of the categorization of instances in the Assessment

	Hit	BEST RATE	WORST RATE
A	0.9516129	0	0.0483871
B	0.61869159	0.36074766	0.02056075
C	0.51612903	0.47311828	0.01075269
D	0.55	0.41666667	0.03333333
E	0.35714286	0.64285714	0

Graphic No. 3
Successes and failures in the evaluation



Again we can see that the errors have more to do with the location in a category better than assigned by the users, than with a worst location, in this last case the error is less significant since it only occurs in 4% of cases of category A, 2% of cases in category B, 1% of cases in category C and 3% of cases in category D, and in total, the error is less than 3% considered for the universe used in the model evaluation.

It is evident that the model would improve its accuracy if fictitious cases were generated in the presence of deviations of poor support.

4. CONCLUSIONS

The optimization of the control tasks in the tax administrations is essential. While the medium-term objective is to increase voluntary compliance with the obligations, in the short term it is essential to increase the level of compliance; in that context control is a key part, and its optimization is an essential need.

Tax administrations are bound to their achievement, it is possible and comes on the heels of a change of culture that seeks not to increase the number of inspections or enforcement actions, but direct them to qualitatively better results with less consumption of resources, this is based on a series of pillars, between which we can mention the taxpayer segmentation, application of differential measures for these segments, the emergence of specialized areas for their attention, policies for early detection of fraud, the analysis of the political and economic context in which taxpayers are evolving and the impact of economic globalization, the contributions of fiscal sociology, the construction of risk profiles using in a centralized and integrated approach all the information available, and organizational changes.

Risk control is at the center of the new orientation of the control tasks, rather than ex post management it aims to optimize the detection of risk groups.

This raises the need for automated tools that can contribute to the establishment of risk profiles of different groups. For viable innovations using ICT is essential, since, given the volumes of information to process and the geographic areas to cover, it could not be conceived without their use. In this line the growing presence of ICTs in the administrations of the countries included in the CIAT fits, and the reorientation of their use, which evolved from a simple calculation assistant to become a facilitator of cultural change, which places them in the dominant mechanics of communication in the society, facilitates exchanges with other national and international organizations and enable them to facilitate their tasks with taxpayers.

In that context, this paper focuses on the construction of models for the description and classification according to risk of default of Large National Taxpayers, the group of most interest to the AFIP and the search for rules that explain both maintenance and the variation of the pre-sorting.

In particular, it implements neural networks on data in the AFIP about the deviations recorded by the Large National Taxpayers for three periods of 2009.

On the other hand, we proceed to define an experimental environment to validate the results, in order to evaluate the effectiveness and success of the proposed solution. To do this, performance measurement of the degree of precision are used, which is measured as the percentage of misclassified tuples.

Tests using the proposed model to demonstrate that it is possible to apply classification algorithms and have a taxpayer risk prediction model of interest for tax administrations, the degree of confidence found in this work is 70% and is superior to that obtained with other data mining tools, for obtaining rules through trees classifying future taxpayers in 5 categories according to their deviations shows only 61% of matches.

In the future, data collection with more time coverage may provide time series in the search for sequential patterns, the completion of innovative data source by creating test cases covering the entire universe of divergences can increase the precision of predictive models, and work on partnership between divergences can guide research, through the appearing of some

of them, from the presence of others who appear with the first.

In summary,

It is possible to optimize the inspection process using innovative criteria, without major risks, with the help of ICT, providing the tax administrations of useful models for determining risk profiles.

The importance for government to have these types of models is manifold: they make effective and useful use of the large volumes of data, enable universal access to centralized information, guarantee transparency, quality and safety and equal treatment to the same behavior and improve the image that taxpayers have of them.

5. BIBLIOGRAPHY

AFIP, Dirección de Planeamiento y Análisis de Gestión, Dto. Planeamiento: Proceso de Formulación del Plan Estratégico, (Buenos Aires, 2007).

ANAO, Australian National Audit Office: Innovation in the Public Sector - Enabling Better Performance, Driving New Directions Better Practice Guide, (Australia, 2009).

BERRY, M. y LINOFF, G.: Data Mining techniques for Marketing, Sales and Customer Support, (USA, 1977).

BRITOS, Paola Verónica y GARCÍA MARTÍNEZ, Ramón y HOSSIAN, Alejandro y SIERRA, Enrique: Minería de Datos basada en Sistemas Inteligentes,
(Buenos Aires, 2005).

CHAPMAN, P. y CLINTON, J y KERBER, R y KHABASA, T. y REINARTZ, T y SHEARE, C. y WIRH, R: CRISP-DM 1.0 Step-by-step data mining guide in CRISP-DM consortium, (EEUU, 2000).

CONDE, Alberto: Formulación de una política informática para la administración tributaria, en Una Administración Tributaria para el Nuevo Milenio – Escenarios y Estrategias - Conferencias Técnicas del CIAT, (Washington, 2000).

ESPER, Susana C: Factores subjetivos incidentes en la institucionalización de la e-taxation, en Cuadernos del Instituto AFIP, Vol. nº 8, (Buenos Aires, 2009).

ESTÉVEZ, Alejandro: La administración tributaria frente al cambio tecnológico, en Cuadernos del Instituto AFIP, Vol. nº 8, (Buenos Aires, 2009).

HAN, J. y KAMBER, M.: Data mining: Concepts and techniques, (EEUU, 2006).

HERNÁNDEZ BATISTA, Juan: Las tecnologías de la información y de las comunicaciones al servicio de las administraciones tributarias, en Una visión moderna de la administración tributaria - 43ª Asamblea General del CIAT, (Santo Domingo, 2009).

MAIMON, Oded Y ROKACH, Lior: Data Mining and Knowledge Discovery Handbook, (EE.UU, 2005).

MICHALSKI, R.S. y BRATKO, I. y KUBAT, M.: Machine Learning and Data Mining. Methods and Applications, (EE.UU., 1998).

RUSSO, Marcos: Administración de riesgos en Aduana-Una perspectiva Horizontal, en Cuadernos del Instituto AFIP, Vol. nº 10, (Buenos Aires, 2010).

SEGARRA TORMO, Santiago: "El uso de las nuevas tecnologías para facilitar el cumplimiento fiscal ", en La administración tributaria al servicio del ciudadano - Conferencias Técnicas del CIAT, (Sevilla, 2001).

STEINBERG, Beatriz: Aplicación de Minería de Datos al proceso de fiscalización de Grandes Contribuyentes Nacionales en la Administración Federal de Ingresos Públicos, en La Tributación en las Sociedades Digitales Nativas - XXI Encuentro Técnico Internacional de Administradores Fiscales en Argentina, (Mar del Plata, 2011).