Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Artificial Intelligence Impact on Auditing Activity: Solving Bottlenecks in Fiscal Transfers from Union to Subnational Entities

Sérgio Tadeu Neiva Carvalho

ABSTRACT

Objectives: This research aims to understand how the use of Artificial Intelligence (AI) is sufficiently accurate to dispense with the conventional analysis of accountability of agreements. The use of technological solutions such as Machine Learning (ML) and AI has been expanding in the private sector with positive impacts in preventing fraud and increasing efficiency. However, studies on its use in Public Administration are still incipient. Therefore, this work also aims to understand the impacts of AI on the Government. The methodology made use of the implementation of a computational environment to test ML architectures, parameterizing them according to the volume and the identity of the granting agency. The results indicate that the identity of the granting agency influences ML performance and as the size of the training sequence grows, performance increases. Understanding AI performance will validate this innovative approach, streamlining the public sector workforce.

Key-Words: *Control, Federalism, Artificial Intelligence, Public Administration, Discretionary Transfers*

Classification JEL: H83, H77, M42, P35

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

SUMARY

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

1.INTRODUCTION

This work aims to analyze how the integration of technology with internal audit can be useful in the control of the Union's Voluntary Transfers, intermediated by the "Plataforma + Brasil" [1]. From this perspective, we analyze the case of the "Malha Fina de Convênios" system, an innovative technological approach involving the Federal Comptroller General - CGU - and the Ministry of Economy - ME. In this context, the relevance of two central points emerges: the role of Control and the application of Technology for its performance.

The use of Technology reveals its importance because it considers the variable nature of records and the incorporation of non-traditional sources of data in the audit domains (Vasarhelyi, Kogan and Tuttle 2015). The new opportunities for audit analysis provided by the use of Technology meet the need for changes in control. However, there are challenges in the way of its use, in addition to its potential benefits.

The overcoming of these obstacles route s in Public Administration has nuances intrinsic. Wirtz, Weyerer and Geyer (2019) propose 4 different categories of challenges to be overcome, specifically for the public sector: (i) implementation of technology, (ii) legislation and regulation on the applicability of Technology, (iii) ethics and (iv) behavioral paradigm. In turn, the paradigm behavioral entails a more complex debate over other category s , given that addresses issues tortuous a bureaucracy (Barzelay 2019 ; Bresser - Pereira 2008 ; Mintzberg 1979 ; Olivieri 2008), as a replacement and the transformation of the workforce, the acceptance and the trust of the adoption of technologies at the expense of conventional working methods, and, finally, the interaction human- machine. Issa, Sun and Vasarhelyi (2016) propose a reflection on the effects that automation generates in replacing auditors as a workforce.

However, the adoption of Technology in the audit activity should be encouraged. Once the auditor learns to properly deal with data mining techniques, artificial intelligence, Benford's

---

1 The Federal Government's Covenant System - SICONV - was renamed Plataforma + Brasil (Decree No. 10,035, 2019)

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Law, among others, there is an enlightening break with current methods, leading him to question how audits were carried out without the use of these technological modernities (Huang and Vasarhelyi 2019; C. Zhang 2019) . Notably, the Public Administration must constantly seek the best was in control of the performance for the political public are evaluated and monitored with lower cost and faster. The use of technological innovations in the Control function allows progress in this area (Maciejewski 2017).

While, on the one hand, the technology applied to the control is relevant for the implementation of public policies, on the other hand, the implementation of them in Brazil is a huge challenge. We are a country with continental dimensions, profound cultural diversities, configured in a sumptuous number of independent subnational entities, comprising 27 federation units and 5570 municipalities (IBGE, 2020). In turn, the Union's Discretionary Transfers are a useful option for the implementation of public policies in the country.

The Voluntary transfers are funds transferred by the Union without constitutional or legal determination, other federal agencies or private non-profit organizations to carry out works or services of common interest. The Brazilian federal state uses federalism through cooperation to achieve the fundamental principles and objectives defined in the Brazilian Federal Constitution of 1988 - CF88 -, articles 1 and 3, in particular national development and the reduction of social and regional inequalities.

However, the institutional arrangement stipulated by CF88 assigns competing competencies between entities, not always harmonious, to the extent that the structure federation is very uneven among local, state and federal levels (Fajardo, 2016). Indeed, the implementation of public policies by the Federal Government through partnerships and agreements with the subnational entities is imperative, but not always the resources transfers are redistributive, as demonstrated by research undertaken by Amorim Neto and Simonassi (2013), Brollo and Nannicini (2012), Ferreira and Bugarin (2008), Meireles (2019) and Soares and Melo (2016).

In turn, the assessments of the public policy of the Union Voluntary Transfers indicate an imbalance striking between the amount of covenant entered into and the ability of the granting

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

agencies are analyzing the benefits of accounts of covenants entities (Brazil, 2018a). The expression of this reality is the continuous accumulation of accountability pending analysis. The consequence practice of this phenomenon is the existence of a stock incremental of agreements liabilities, which provides ç will account does not have any prospect of being analyzed timely, removing the constitutional principle of the duty of accountability. Furthermore, the requirements imposed by control agencies, notably the CGU, the grantors bodies and conventions en tes as the re qu i expense of evidence ments contribute to the slowness of the accountability process. Therefore, CGU is also part of the problem, albeit indirectly.

Faced with this paradoxical scenario, in which the performance of the control agency implies the deficiency of the public policy that it controls, CGU encouraged the proposition of a solution through the use of technology in the control actions. This solution, called "Malha Fina de Convênios", facilitates the analysis of accountability of covenants, observing the expressive volume of instruments entered into and the limited capacity of analysis of the staff of the granting agencies. The “Malha Fina de Convênios” system is a predictive model of Artificial Intelligence that allows indicating, with a certain degree of certainty, the result of the analysis of the rendering of accounts of the covenants at the moment when their accounts are presented by the covenants (subnational entities that receive resources) to the grantors (federal agencies that transfer resources).

In this context, the potential of the use of the “Malha Fina de Convênios”, their impacts already demonstrated and their limitations are still not clear and definitive. Thus, the objective of the work is to analyze some of these aspects, specifically with regard to the precision and accuracy of the “Malha Fina de Convênios” in different circumstances. The influence of the size and identity of the resources transferring agencies will be analyzed, in different scenarios of training sequence, in the accuracy and precision of the algorithms (Alpaydin 2014; Breiman 2001; Domingos 2012). The implementation of these different scenarios will be conducted through the use of supervised machine learning Random Forest, executed in Phyton language.

To understand the impact of the “Malha Fina de Convênios” in the real world of Voluntary Transfers from the Union, this study focuses on the theme of the relationship between audit and

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

technology (Issa, Sun, and Vasarhelyi 2016), a classic question for the Control function in science Contemporary administration, since the strong expansion of the delegation of public policies to sub-national entities since the Federal Constitution of 1988 increased the relevance of the connection between the increase of resources transferred by the Union and the need to monitor them efficiently (Arretche 2010; Abrucio e Franzese 2007; Afonso, Araújo, and Fajardo 2016). This issue will be addressed both from a theoretical and empirical point of view, in this case, analyzing the applicability of the “Malha Fina de Convênios” in the Federal Government.

This study will analyze this theme in 5 sections, in addition to this Introduction. Section 2, subsequent to this introductory section, consists of a literature review, starting with a brief overview of the Union's Voluntary Transfers, the concepts of Control and their importance in Public Administration, in addition to highlighting the role of technology in redesign of the audit activity, addressing concepts of machine learning.

Then, section 3 addresses the problem of accountability for federal transfers transferred through Plataforma + Brasil. Also in this section, the system “Malha Fina de Convênios” is exposed as a proposed solution for the liability of accounts to be analyzed, as well as the characterization of its limitations and potential. In turn, the research problem is outlined and the hypotheses to be tested are listed. Subsequently, in section 4, the adopted methodological procedures are presented, describing the strategy for simulating machine learning tests and detailing the database of Plataforma + Brasil, including its time cut. In continuous action, section 5 presents the results obtained in the simulations and the empirical data. The analyzes are presented according to two architectures configured to simulate the performance of the machine learning algorithms: identity of the resource transferor and volumetry of the training sequence. Finally, in section 6, the final considerations are presented, explaining the main conclusions, as well as the limitations of the present study and some work proposals for future research.

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

2.THEORETICAL FRAMEWORK

This section addresses the theoretical framework used to support the study, that is, it seeks to review the literature already developed, highlighting the potential contributions of this work. Having made the considerations on federative relations and their influence on Voluntary Transfers, we move on to the discussion of Control within the rational-legal framework of Public Administration and then to the contextualization of the role of Technology applied to Control in the paradigm of the "Digital Era Governament ". Thus, the objective of this section is to refer the applicability of a new technology for the use of Control in order to solve a real problem of public policy, namely Voluntary Transfers.

2.1.Federative Relations and Voluntary Transfers

The Union's Voluntary Transfers are a public policy in which the Union transfers federal public resources at its discretion to States, Municipalities, the Federal District or private non-profit entities. The purpose for the transfer of federal funds to loved federated or to social organizations is to execute these policies in the areas of education, health, sanitation, construction and rehabilitation of roads, water supply, housing and urban and rural energy. After all, the citizen lives in the municipality (Amorim Neto and Simonassi 2013) and the most assertive implementation of public policy could only occur in this subnational entity (Soares and Melo 2016).

Voluntary transfers play an important role in Brazilian federative relations, mainly between the Union and municipalities. According to Fajardo (2016), the Federal Constitution of 1988 (CF88) was a milestone in the institutional arrangement of the Brazilian federation, since it gave municipalities the category of federative entity, making them of singular importance, in view of their greater representativeness in quantity in relation to the States. In turn, Arretche (2010) and other academics, such as Afonso, Araújo and Fajardo (2016) and Loureiro and Abrucio (2004), highlight the peculiarity of the advent of the municipal entity in the model of Brazilian federalism, each in its perspective arising from your searches. In summary, most of these perspectives are negative because the empirical evidence from these studies shows that the small collection of these subnational entities and their large share in the composition of

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

public spending threaten the sustainability of the municipal entity. Likewise, these negative perspectives are corroborated by the statement by Amorim Neto and Simonassi (2013) in which the capture of political support prevails over criteria of equity in transfers from the Union to subnational entities. Nevertheless, the municipality plays a central role in the Union's voluntary transfers. In a way, the role of the municipalities in the public policy of the Union's voluntary transfers can be seen in Figure 1, as it is the recipient with greater volume and materiality agreements, surpassing even private non-profit entities.

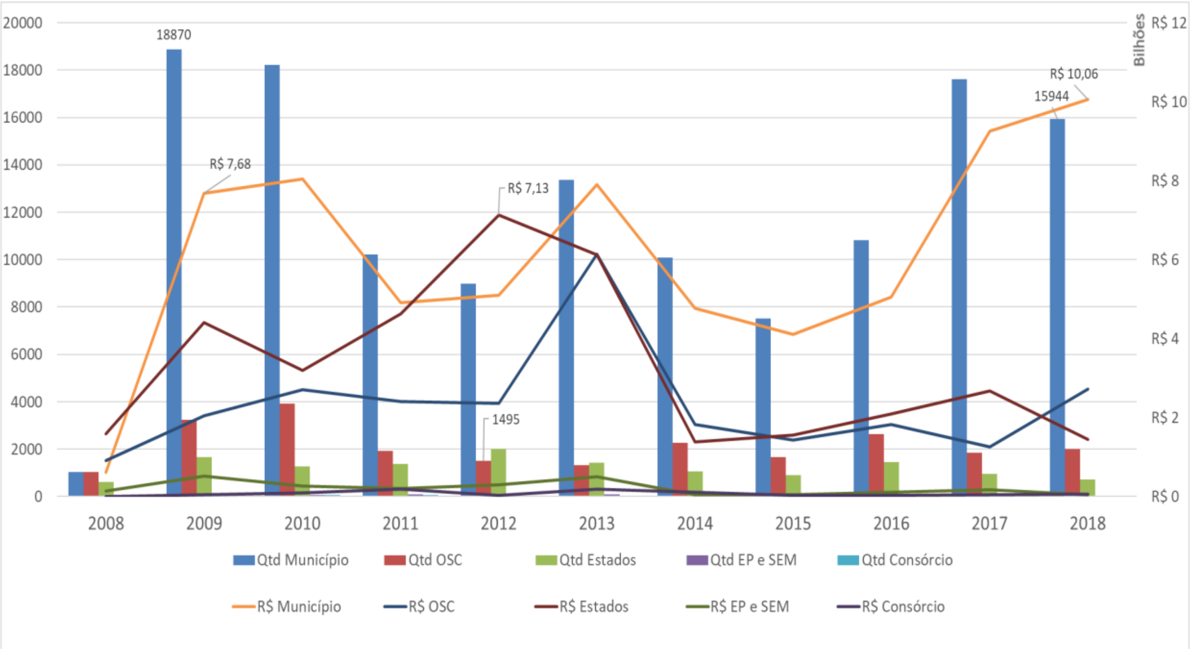

Figure 1 – Distribution of the quantity and materiality of instruments operated on the Plataforma + Brasil . Source: Plataforma + Brasil. Prepared by the author[2].

Aside from the blatant supremacy of the municipalities in relation to the states in terms of the number of instruments, Figure 1 illustrates a less expressive panorama in relation to the materiality of transfers made to the Municipalities in relation to the States, even though these are superior to the first recipient when compared to the second. Figure 1 represents the quantities of instruments by the bars and the values of resources transferred by the lines. In effect, the blue bars, which symbolize the number of instruments in which the Municipalities configure themselves as beneficiaries, are much greater in magnitude than any other category

[2] EP- Empresa Pública, SEM – Sociedade de Economia Mista, OSC – Organização da Sociedade Civil.

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

of bar, that is, the States and non-profit entities. However, the orange line, which symbolizes the total value of resources transferred to the Municipalities, is not so far from the lines that symbolize the values transferred to the States and non-profit entities, highlighting that in 2012 the volume of contributions made to the States surpassed that done to the municipalities. However, Table 1 allows us to infer that the level of resources transferred to the Municipalities is almost double that transferred to the States.

Legal Nature	Quantidade	Valor
County/Municipality	136429	R$ 69.319.120.959,84
Civil Society Organization	24118	R$ 26.614.140.309,16
States	13.624	R$ 36.598.277.602,35
State Owned Enterprises	852	R$ 2.331.638.055,77
Public Consortium	433	R$ 815.383.238,07
Total	**175456**	**R$ 135.678.560.165,19**

Table 1 - Quantitative and Total Values of instruments operated on Plataforma + Brasil by legal nature of the recipient. Prepared by the author. Source: Plataforma + Brasil.

To a certain extent, Figure 1 and Table 1 strengthen the role that the Municipality exercises in the public policy of the Union's voluntary transfers and consequently in the federative arrangement (Arretche 2010). Furthermore, the presented scenario shows that the instruments entered into with the municipalities are quantitatively bulky and have low values, as shown in Figure 2. This contributes to the dispersion of resources, which can generate adverse effects, given that there is not always coordination for the execution of programs by the federal government and subnational governments adopt exaggerated autonomous positions (Abrucio 2005; Fajardo 2016). Naturally, public policies are shaded, the vast majority of Brazilian municipalities do not have an adequate bureaucratic structure for the management of a large number of agreements, there is an overload in the operational capacity to manage them in the organs of the Union and, mainly, in the municipality, and lastly and most impactfully, the fractioning of the transferred resources materializing in low value agreements (Figure 2) makes the effectiveness of government programs low, opposing the hypothesis that the resources are applied in a centralized way, in large scale. Likewise, the costs involved in managing the life

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

cycle of high value covenants are no different from the cost of low value covenants. These challenges sometimes affect the proper application of public resources from voluntary transfers.

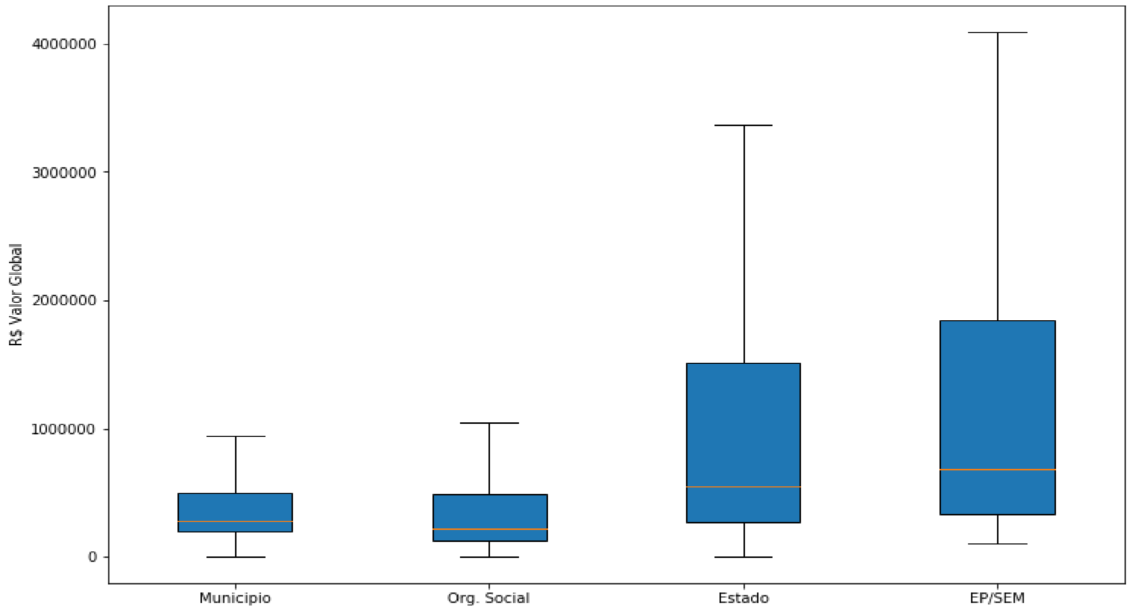

Figure 2 - Boxplot of the Global Values of the agreements segregated by Entitiy. Prepared by the Author. Source: Plataforma + Brasil.

The 1988 Constituent adopted a federative formula that envisaged vertical intergovernmental cooperation, with voluntary transfers being a tool for achieving this objective. The creation of municipalities after the promulgation of CF88 is in line with this precept and radiates effects from a political perspective, which refers to the existence of an articulation conceived over time, prior to the promulgation event of the Constitutional Charter. Fajardo (2016), on the other hand, argues, anchored in the research promoted by Abrucio (2005) and Arretche (2010), that CF88 was a process that started well before its promulgation, promoting consequences under administrative and fiscal cousins, in addition to the political point of view, whose greatest impact was the creation of the municipal entity.

However, the plurality of subjects with competing competences among the federated entities and the wide diversity of related legislation, ranging from health, through education, security, culture, to social assistance, makes it very difficult to have an accurate classification of the types of transfers from the Union, implying divergences in doctrine and academia (Brasil, 2008; Dallaverde, 2014). Although the existence of these storms, decision-making bodies for the

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

policies of federative relationship and management of public expenditure, located in a central position in the Federal Government, made efforts to clarify the panorama of the fiscal transfers of the Union. A group formed by the extinct Ministry of Planning constituted (SOF and SLTI), Ministry of Finance (STN) and CGU, led by STN, whose result of the work was the categorization of fiscal transfers from the Union, as provided in Technical Note No. 14/2015 / COINT / SURIN / STN / MF- DF.

According to the aforementioned note, transfers from the Union admit the (i) mandatory and (ii) discretionary classifications, as shown in Figure 3. Mandatory transfers comprise those that have an express provision in the Federal Constitution or Law and are transferred directly, one that since there is no need to sign a formal legal instrument between the granting party and the beneficiary. Continuous act, transfers emanating directly from the Constitution order are subclassified as constitutional, while those in which the obligation is established by law are subclassified as legal. Examples of mandatory constitutional transfers are those made to the Municipality Participation Fund - FPM - or State Participation Fund - FPE. On the other hand, mandatory legal transfers can be found in public basic education policies such as in the PNAE, PNAT, PDDE, as well as in health policies in the Unified Health System - SUS - and in the National Health Fund - FNS.

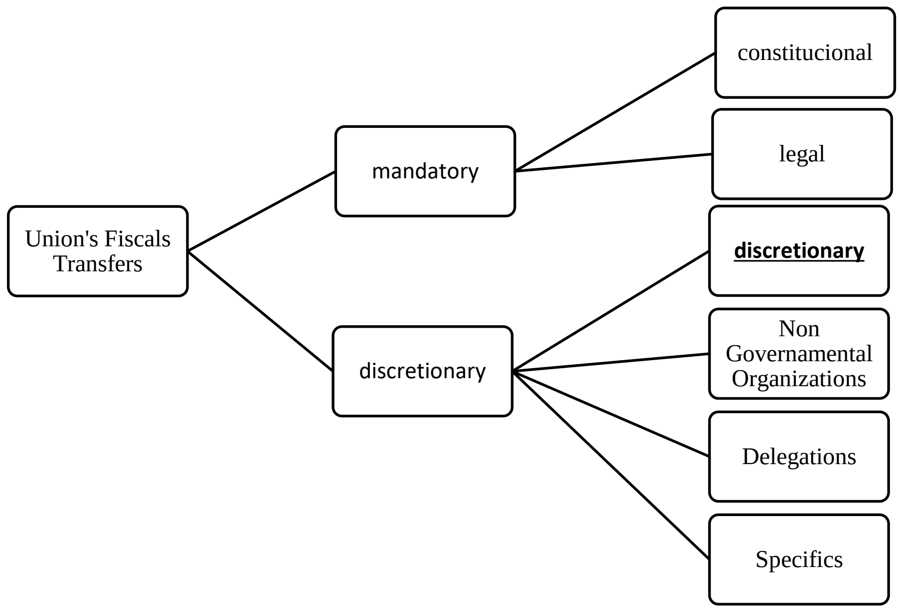

Figure 3 – Classification of the STN for the types of federal tax transfers. Prepared by the Author.

In turn, discretionary transfers comprise funds transferred under the aegis of a legal instrument between two parties and are subclassified into 4 categories. The first

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

subclassification, and most important for the purpose of this study, is voluntary discretionary transfers, consisting of those in which the delivery of resources to subnational entities seeks cooperation, assistance or financial assistance. Discretionary voluntary transfers are not intended for SUS, nor are they included in the hypotheses of mandatory transfers. Commonly referred to as voluntary transfers in the daily life of the federal bureaucracy, voluntary discretionary transfers are operated on the Plataforma + Brasil and basically consist of two types of legal instruments: the agreement and the transfer agreement. The difference between these instruments lies in the fact that the first is celebrated directly between the parties, on the one hand a federal agency and on the other a subnational agency, and the second is celebrated through a federal financial agent on paper as a representative of the Union. Currently, this role is exercised by CAIXA. Generally, the objects of agreements have scope for the acquisition of goods and contracting of services while the objects of onlending contracts consist of civil construction works.

The importance of voluntary transfers (voluntary discretionary transfers) in the federal institutional arrangement of the Brazilian state is not reflected in its ordinary legislation. In fact, the volume of voluntary transfers is significantly lower than the volume of the mandatory transfers category. In 2018, voluntary transfers totaled more than R$ 14 billion[3],while mandatory transfers reached a level above R$ 350 billion [4]. However, this is not a reason to label those as insignificant or even meaningless in relation to these. On the contrary, the resources are highly expressive for the entities that depend on federal resources for the realization of new investments (Fajardo 2016), considering that the revenues of a large part of the Municipalities, whether from their own collection or transferred by force of constitutional determination, they end up being almost fully employed in current expenses for the maintenance of the administrative machinery (Abrucio 2005).

---

3 Data obtained from Plataforma + Brasil. This is the sum of the global values of all instruments registered in the system. The global values are the portion transferred by the Union and consideration provided by the convention. The total amount of agreements signed in 2018 reached R $ 14,335,621,107.03..

4 Data obtained from the Integrated Financial Administration System - SIAFI. Mandatory transfers include the State and Federal District Participation Fund - FPE; Municipality Participation Fund - FPM; IPI - Export; CIDE-Fuels; Fund for Maintenance and Development of Basic Education and Valorization of Education Professionals - Fundeb; Royalties; Tax on Rural Territorial Property - ITR; IOF-Ouro; and Lei Kandir (LC 87/96). The total amount transferred to states and municipalities in 2018 reached R $ 351,447,109,424.57.

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

It turns out that mandatory transfers have robust regulatory support, while voluntary discretionary transfers stand out due to the absence of detailed ordinary legislation.

The normative framework for the Union's voluntary transfers emanates from the express legal provision in article 116 of Federal Law no. 8666/93. There are studies indicating that voluntary transfers are supported by the constitutional provision (art. 241), including Dallaverde (2016), although this line of thought is silent in the Brazilian Administrative Law Doctrine, notably in Pietro (2015):

> "(…)
> *Art. 241. A União, os Estados, o Distrito Federal e os Municípios disciplinarão por meio de lei os consórcios públicos e os* ***convênios de cooperação entre os entes federados****, autorizando a gestão associada de serviços públicos, bem como a transferência total ou parcial de encargos, serviços, pessoal e bens essenciais à continuidade dos serviços transferidos.*
> (…)" **my emphasis**.

In a way, the helplessness of legal diplomas regulating the public policy of voluntary transfers is a reflection of the recent contemporary scope of the municipality for the federated entity, an innovation provided by CF88. Voluntary transfers, as it is a discretionary policy with applicability on a wide spectrum of issues, does not encourage the actors in the process to discipline it through legislation approved by the National Congress, relegating them to non-legal rules (Pietro 2015). In this regard, the legal gap was filled with a myriad of ordinances and decrees issued over time, as shown in Figure 4. It can be said that voluntary transfers take on radically different contours after the creation of SICONV in 2007 (Meireles 2019), to the extent that the entire management process of the agreements with the federal government started to be made in this system. Previously, the agreements entered into with the Union were registered with SIAFI, but this system lacked functionalities to allow assertive control over transfers, impairing the transparency of information and the disclosure of data, and, on the other hand, favoring the scenario of payment of benefits. outstanding accounts and the misuse of public resources. Now, SIAFI was developed to serve as a financial execution system for the Federal Government (Abrucio and Loureiro 2018) and not a system for managing covenants.

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

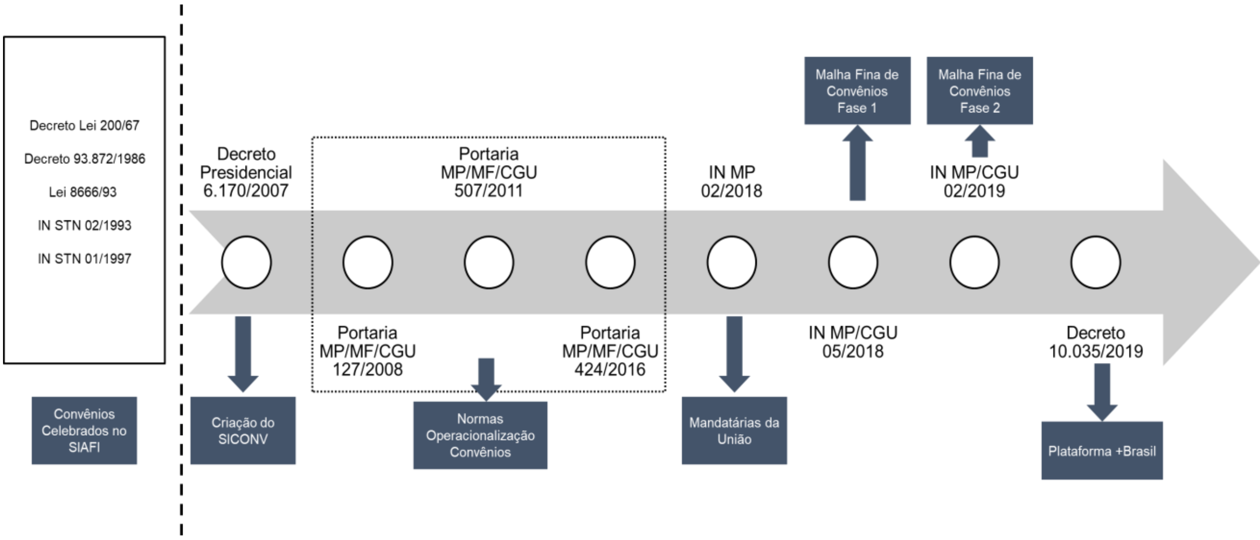

Figure 4 -

Figure 4 - Evolution of the normative framework for voluntary transfers. Prepared by the Author.

Normative Instruction No. 01/97 of the STN stood out as an important regulatory framework for voluntary transfers, having lasted until the publication of Decree no. 6,170 of 2007, which revolutionized the process with the creation of SICONV. In 2008, Interministerial Ordinance 127 was issued with the purpose of conveying the operational rules for the execution of Decree no. 6,170 (2007) and was succeeded by Interministerial Ordinances 507 and 424 of 2011 and 2016, respectively. Then, the chronology shown in Figure 4 shows another major regulatory framework, Normative Instruction n. 2 of 2018, which tried to establish guidelines and rules for the execution of onlending contracts (type of instrument of voluntary discretionary transfers) through the Mandataries of the Union. Also in 2018, the first phase of the “Malha Fina de Agreements ”, with the edition of Normative Instruction CGU / MP n. 5 of 2018, enabling an automated alternative for accountability analysis. Accordingly, in 2019, the second phase of the “Malha Fina de Convênios” began, through the publication of Normative Instruction CGU / MP n. 2 of 2019. While the first phase restricted the scope of the automated analysis only to the stock of liability for rendering of accounts pending analysis, the second phase allowed the use of automated analysis in rendering of accounts anachronically, as long as the stock was sanitized.

The considerations discussed above illustrate that voluntary transfers from the union do not have an isolated public policy, much less an end in itself, since its primary purpose is to

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

enable, as a cooperation, the financing of public services in subnational entities. . At this point, a question may arise whether the financial cooperation received by a poorer entity, to the detriment of its small contribution compared to a richer entity, contradicts the balance of the federative pact. In response, the modulation that the designs of federative relations apply to voluntary transfers is enshrined, as the federative balance is solidified with the mechanism of transferring resources (Afonso, Araújo, and Fajardo 2016; Amorim Neto and Simonassi 2013; Fajardo 2016). The objective is to include the most needy and needy entities in the development process, and not to exclude the most favored entities, who are often responsible for boosting the Federation's economy.

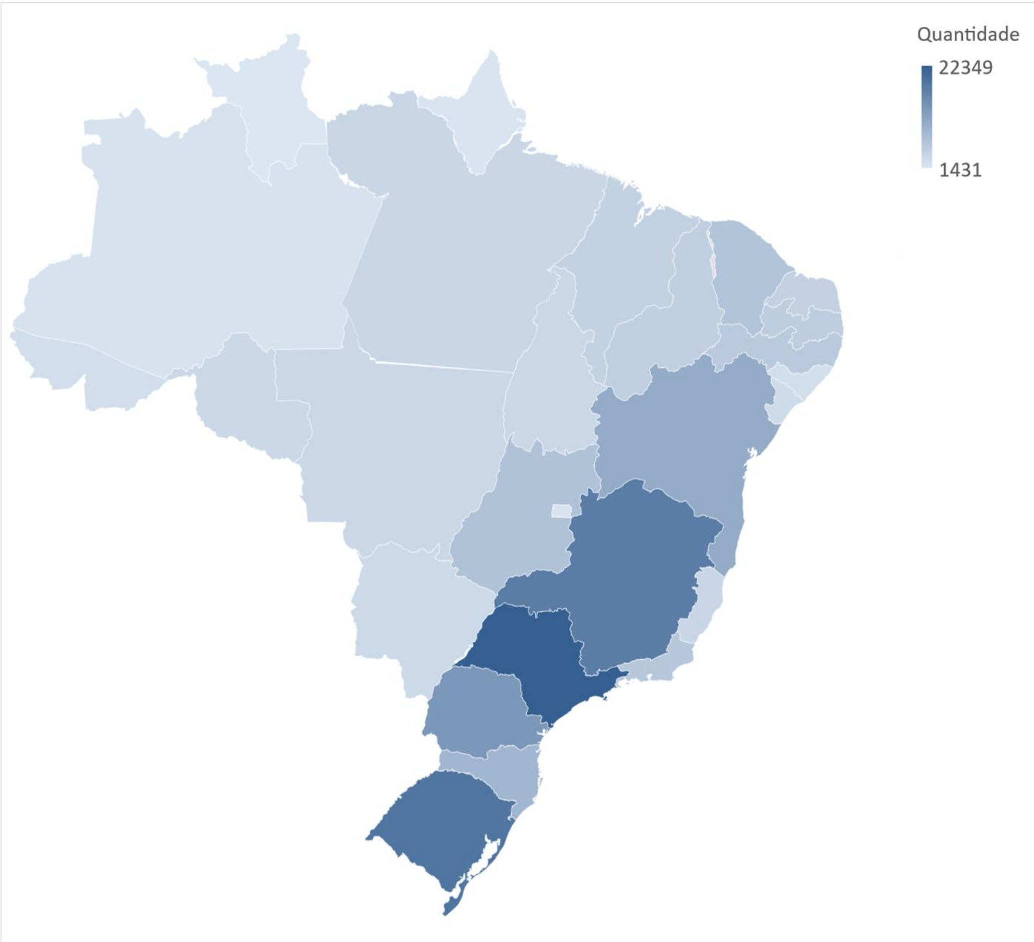

Figura 5 –

Figure 5 -Concentration of Instruments celebrated by Federation Unit. Prepared by the Author. Source: Plataforma + Brasil.

Ultimately, it is essential that the equity criteria overlap with the political objectives in the transfers resulting from voluntary transfers, as this resource would contribute to the reduction

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

of interregional inequalities in the Brazilian federation. However, the empirical evidence from research by Amorim Neto and Simonassi (2013), Brollo and Nannicini (2012), Limongi and Figueiredo (2005), Meireles (2019) and Soares and Melo (2016) indicate the opposite, portraying that subnational entities with political alignment with the Union representative or with greater representation in the National Congress receive more resources, regardless of other criteria. Figure 5, above, endorses this scenario, given that the poorest regions of the country are discredited in relation to the richest in terms of the number of instruments celebrated.

2.2.O papel do uso da tecnologia aplicada ao Controle

In view of the above, based on the referenced bibliographic review, it is assumed that new information technologies are an important instrument for the exercise of government internal audit activity.

Dunleavy et al (2006) argue that New Public Management tends to die and make room for a new era of digital governance, called "Digital-Era Governance". This new paradigm takes advantage of technological advances and aims to make them central to public service offerings. In other words, technology is no longer an instrument that enables efficiency gains to become a solution that helps to shape the public service itself.

In the first analysis, New Public Management is no longer new (Barzelay 2019), recognizably becoming a middle-aged model, which, applied in contemporary bureaucratic reality, can generate adversities in the expected results. It is urgent to reverse this deficit, emphasizing that Information Technology and Information Systems are key mechanisms for the continuation of the Weberian rationalization of public administration processes in the 2020s. The future has arrived. Naturally, it highlights the great importance that technology assumes in the new way of operating management systems and methods of interacting with citizens and public service consumers (Barzelay and Gallego 2006).

The roots of the New Public Management prestige of are rooted in neglect of this model to recognize that the hindrance to the simplification of public administration embraced by overcoming the situation of "creating trouble selling facilities" (Bresser-Pereira 1996). The problem manifests itself when civil servants acting as auditors establish boutique bureaucracies in which they create complex levels of formal requirements . Perverting rationality to develop strong legalistic management in government agencies is seen as problematic, because it

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

generates management attitudes obsessed with intermediary organizational objectives, rather than providing services or effectiveness (Hood, James and Scott 2000 ; Secchi 2009) .

Nevertheless, the "Digital-Era Governance" model influences the audit and control activity. This is because auditing and control (Dunleavy et al. 2006) are part of social science Administration (Fayol 1949; Mintzberg 1979; Taylor 1911). In turn, this influence implies wide-ranging effects since it is not restricted to imminently technological applications (Newell and Simon 1976). Contrary to what is suggested, the influence of "Digital-Era Governance" in the audit activity is not restricted to technology instruments, as this influence causes the association of a wide architecture of concepts involving cognition, behavior, organizational, political and cultural change linked to information technology. The advent of "Digital-Era Governance" is currently the most widespread and peculiar influence on how governance arrangements are changing to exercise government control.

It is clear that the current paradigm makes auditors shy away from adopting obsolete techniques (Vasarhelyi and Halper 1991) such as physical paper exams, examining the authenticity of documents through visual comparison or correlating information through face-to-face observation. In the same way that typists were extinguished from Public Administration because typewriters are no longer used, auditors must be careful not to have the same purpose. Outdated auditing techniques should be left out, while the provision of public services is predominantly digital, requiring the public servant invested in the audit activity to operate electronic information systems. In this respect, there is a need to face the new relationships of the technological disruption ecosystem (Margetts and Dunleavy 2002) created by the "Digital-Era Governance".

According to Parker, Jacobs and Schmitz (2019) the more innovative the technology, the greater the prospect that auditors propose new paradigms for the operation of the business and for the manager's decision making. In turn, the greater the complexity of the technology, the greater the possibility of innovation, but the difficulty of its implementation is increased, precisely by the complexity of its handling by the auditors. This implementation of new technologies requires a change in approach and behavior. Every action causes a reaction. Consequently, there is resistance to the adoption of new technologies by the auditors (Kim, Mannino, and Nieschwietz 2009).

Naturally, the Brazilian bureaucratic apparatus exercises and will continue to exercise resistance to change the logic of acceptance of innovations (Heber, 2014) in which the rule is the obtuse interpretation of the principle of legality (Campana 2017). After all, it is taken

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

literally "Dura Lex, Sed Lex," the law is hard, but it is the law. In the public context, it is only allowed to do what is within the limits of the legislation, while any act that goes beyond the normative rules can be considered illegal (Bresser-Pereira 2008). This implies two nefarious aspects: it inhibits the good auditor from proposing non-traditional and innovative alternatives, and makes the bad auditor supported by plausible justifications for his inaction.

However, according to the authors Kim, Mannino and Nieschwietz (2009) there is progress in the acceptance of technologies in the activity of internal audit, making an innovative approach anchor management.

Additionally, the culture of fear, disseminated in public management (Alves and Calmon 2008), was a preponderant factor, and continues to be, for the fullness of the development of innovative practice, the system "Malha Fina de Convênios". Within the scope of Public Administration, there is currently the formation of a dispute in which the litigating poles are, on the one hand, the public managers and on the other the control bodies. In turn, the configuration of this scenario is very harmful, inasmuch as its consequences are harmful to the machinery of the public machine (Power, 2003b, 2003a).

Precipitatively, control has the primary objective of ensuring good governance (Brasil 2017a; COSO 2013). However, public managers often point out that the control bodies transpose the duties they are intended for, causing inappropriate management interference. In turn, these interferences guide management, given that they cause some effects on managers: fear of budget execution, failure to implement innovation, aversion to taking risks, and finally restlessness in taking on positions to execute public policies.

Given the above, the control paradox is configured: its performance improves public governance as it causes adverse effects on management. So, what would be the relationship between control actions and fear-oriented management?

The triumph over this obstacle that was put in front of the implementation of the " *Malha Fina de Convênios* ", that is, the fear of managers to define an appetite for risk to be able to use the mesh of covenants and then approve accounts automatically, demanded the use of persistence as a key factor (Power 2009) . In fact, the proposal for a new procedural archetype for instant accountability analysis creates discomfort for managers when they compare it with the traditional accountability process.

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

The traditional accountability process involves handling processes, which often consist of physical paper, drafting opinions, diligence with the agreement, verification of receipts and invoices, processing in the hierarchical chain, and is subject to all weather conditions. and a common administrative process of a rational bureaucracy (Ramos 1983; Tenorio 1981). The alternative of an analysis that generates an instant result, exempting the necessary steps from a traditional process, is seen as a revolution by the decision maker. However, it generates fear.

Thus, the persistence and resilience adopted by the "Malha Fina de Convênios" team were essential to prevent the initiative from fading . Therefore, the main factors that contributed to the implementation of the " *Malha Fina De Convênios* " were persistence, resilience and teamwork with the Department of Voluntary Transfers of the Union of the Ministry of Economy.

2.3.The Sophistication of Auditing Methods : Machine Learning Performance Measurement Techniques

The design and nature of audit techniques are changing as a result of the emergence of "Digital-Era Governance", supported by frameworks such as Big Data, Internet of Things (IoT) and Artificial Intelligence. In this new paradigm, machine learning (Machine Learning) emerges with a technique covered with protagonism in the audit activity (Brown-Liburd and Vasarhelyi 2015), as it is very versatile to correlate cause and effect, sophisticating the prediction methodology results. Stanisic, Radojevic and Stanic (2019) claim that a line of previous research was based on the classic use of statistical models, mainly logistic regression and probit, to design a prediction methodology. In turn, these same authors claim that more recent research demonstrates that the use of Machine Learning techniques, to the detriment of statistical models, has a better performance in prediction.

There are several different machine learning algorithms, the main distinction between which is in the supervised and unsupervised learning categories. Among the supervised algorithms, Random Forest stands out, since it presents the best performance in comparison with other supervised algorithms, among them Naive Bayesian, Bayesian Belief Network, Artificial Neural Network, Support Vector Machine, Decision Tree and Random Tree. In due course, this

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

empirical finding finds support in Breiman (2001), Cecchini et al. (2010), Tiwari and Hooda (2018) and Grover, Bauhoff and Friedman (2019), as it states that Random Forest builds a multiple decision tree mechanism and selects the best scenario for combining variables to provide a prediction and therefore has a better performance among the other categories of machine learning algorithm. Random Forest uses the combination of a myriad of decision trees to compute a score, which means the degree of classification prediction. This score is calculated based on the average forecast for each individual tree. Consequently, Random Forest provides a more assertive risk score to measure the probability of approval or disapproval of accounts, for example.

In turn, the accuracy of classification algorithms is linked to their performance (Fawcett 2006). In the literature, there are several approaches to benchmarking machine learning algorithms, with Stanisic, Radojevic and Stanic (2019), Zhang (2019a) and Bao et al. (2020) underline that there is no consecrated one. In this scenario, a common approach to assess performance is to proceed with cross-validation tests (Kohavi 1995). However, the data used for training, which are disapproved and approved agreements over the years, are of an intertemporal nature, which makes the performance measurement through cross-validation inadequate (Bao et al. 2020). In due course, the prediction of the result of the accountability analysis can be categorized into two mutually exclusive classes: approved and rejected. Consequently, it is feasible to measure performance by means of metrics that assess eventual categorization errors in classes (i) approved accountability or (ii) rejected accountability (Bao et al. 2020; Stanisic, Radojevic and Stanic 2019 ; Zhangb 2019a).

In this specific configuration in which machine learning algorithms classify instances, in this case agreements, into two exclusive categories (approved or rejected), Alpaydin (2020) and Fawcett (2006) propose an analytical approach to classification problems when there are only two categories. In this scenario, there are four possible results. On the one hand, if the prediction of the algorithm is correct for an approved agreement, we have a true positive; if the prediction is incorrect for an approved agreement, there is a false negative. On the other hand, if the prediction of the algorithm correctly indicates a failed agreement, we have a true negative; if the prediction incorrectly points to a failed agreement, there is a false positive. Thus, the set of results of the predictions of the classifier algorithm allows the elaboration of a contingency

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

table with two rows and two columns, commonly called the confusion matrix, which represents the dispositions of the result set. The confusion matrix allows to identify which types of incorrect classification occur, that is, which categories (approved or disapproved agreements) are frequently confused.

	Approved Result	*Result Rejected*
Approval Hypothesis	**True Positive (TP)**	**False Positive (FP)**
Rejection Hypothesis	**False Negative (FN)**	**True Negative (TN)**

Figure 6 - Confusion matrix. Adapted from Fawcett (2006) and Alpayadin (2020).

Figure 6 shows a confusion matrix. The numbers on the main diagonal (TP and TN) represent the correct predictions of the algorithm, and the numbers on the secondary diagonal (FP and FN) represent the errors, the confusion between the two categories. In turn, the confusion matrix forms the basis of 3 metrics for measuring the performance of classification algorithms generated by machine learning: accuracy, specificity and sensitivity. These metrics are widely used in literature, as seen by Alpaydin (2020), Fawcett (2006), Hooda, Bawa, and Rana (2020), Grover, Bauhoff and Friedman (2019), Tiwari and Hooda (2018) and Zhang (2019a), despite Hooda, Bawa, and Rana (2020) and Tiwari and Hooda (2018) adopting a different way of computing these metrics. The methodology for calculating the accuracy, specificity and sensitivity adopted mainly in the literature is:

$$\textbf{Accuracy}= \frac{TP + TN}{TP + TN + FN + FP}$$

$$\textbf{Sensitivity}(\mathit{Recall}) = \frac{TP}{TP + FN}$$

$$\textbf{Specificity} = \frac{TN}{TN + FP}$$

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

However, the metrics for performance measurement usually used (accuracy, specificity and sensitivity) are not appropriate when there is an unbalanced balance of the training sequence, in which the occurrence of one category is much greater than the other (Grover, Bauhoff and Friedman 2019 ; Bao et al. 2020; Alpaydin 2020).

According to Bao et al. (2020), one could use a metric, called balanced precision, as an alternative to properly measure the performance of a predictive model. This balanced metric is defined as the average between the simple average between the sensitivity and the specificity of prediction of the instances (provision of covenant accounts). In turn, the balanced precision metric is:

$$\text{Balanced Precision} = \frac{\text{Sensitivity} + \text{Specificity}}{2}$$

On the other hand, Larcker and Zakolyukina (2012) consider that there is a limitation in the use of the balanced precision metric as a performance evaluation parameter. Said metric is designed based on the assertiveness of class prediction, in the case in question, approval or rejection. Objectively, this means that the prediction depends on the labeling of the data in the sequence of training, that is, the conclusive result of the analysis that the granting agencies carried out in the presented accounts. It turns out that the learning algorithms are quite sensitive to the imbalance of the training sequence. Once the granting agencies disapprove a very small number of agreements in proportion to the total, this will influence the distribution of the prediction, consequently in sensitivity and specificity, overestimating them (Fawcett 2006).

Thus, one way to overcome this limitation is the use of the ROC (Receiver Operating Characteristics) curve and measurement of the AUC (Area Under the Curve) metric (Bao et al., 2020; Fawcett, 2006; Stanisic, Radojevic and Stanic 2019). The ROC curve is a two-dimensional representation of the performance of a classifier, in this research represented by the algorithm of the “Malha Fina de Convênios”. In turn, the ROC curve combines the rate of occurrence of true positives, the sensitivity, and the rate of occurrence of false positives, the subtraction value of 1 and specificity, in a graph with two axes. Therefore, it is possible to infer the performance of a predictive model through a single scalar when calculating the AUC metric.

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Since the AUC is a part of the unit's square area, its value will always be between 0 and 1.0. To the extent that a random selection produces the diagonal line between (0, 0) and (1, 1), which has an area of 0.5, no plausible prediction algorithm should have an AUC less than 0.5. The value up to 0.5 means a model without any discriminating power and the value 1 means a perfect model. The AUC metric is equivalent to the probability that a disapproved agreement chosen at random will be classified with a higher score (true negative) than an approved agreement (true positive), also chosen at random (Bao et al. 2020; Breiman 2001; Fawcett 2006; Stanisic, Radojevic and Stanic 2019).

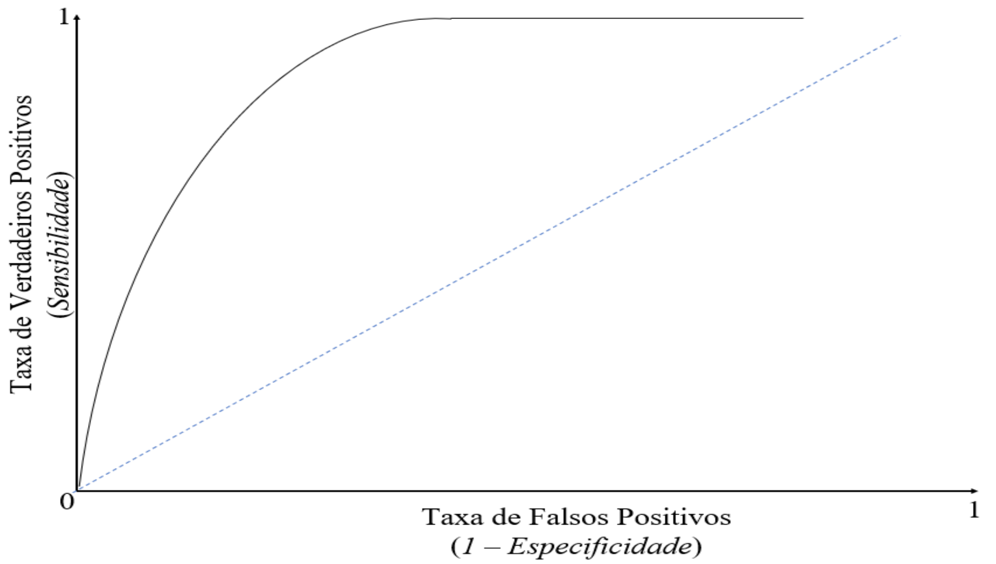

Figure 6 – ROC curve. Adapted from Alpaydin (2020) and Fawcett (2006). Prepared by the Author.

ROC curves are a very useful tool for visualizing the performance of classification algorithms (Cecchini et al. 2010). They are able to provide insight into performance measurement, in addition to scalar measures such as accuracy, sensitivity, sensitivity or error rates. However, Fawcett (2006) stresses that the ROC curve must be used sparingly, because it has limitations as with any evaluation metric. Thus, it is suggested that its adoption be compared with the other scalar metrics.

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

3. CONTEXTUALIZATION OF THE PROBLEM

The objective of this section is to deal with the operational bottleneck that was configured in the procedural flow of the Voluntary Transfers of the Union. the continuous expansion of the number of accounts rendered pending conclusive analysis. This section is divided into 4 more parts, in addition to this introduction. In the first, we will describe the incentives that led to the conception of “Malha Fina de Convênios”. In the second part, we deal with what this solution consists. Then, in the third part, the problems and uncertainties in its use are presented. Finally, in the fourth section, we highlight the importance of these problems, to the point of motivating this work.

3.1. Voluntary Transfers and Plataforma + Brasil (formerly SICONV)

The Union's Voluntary Transfers process handled more than R $ 135 billion between 2008 and 2019 through another 170 thousand instruments among the Federation's entities, according to data from Plataforma + Brasil. In November 2019, precisely, it had 175,456 instruments registered (Brazil, 2019b).

However, the granting of voluntary transfers by government agencies and entities constitutes a major challenge for public administrators, both with regard to the desired smoothness, as well as the agile and effective operationalization of the thousands of instruments destined to the implementation of public policies in the Brazilian States and Municipalities. This challenge occurs due to the number of agents and the multiplicity of objects present in the concession agreement.

Among the public policies that have received funds through federal transfers in the last 10 years are: planning, management and urban development (R $ 11.3 billion); support for tourism (R $ 9.3 billion); specialized health care (R $ 7.5 billion); family farming (R $ 5.5 billion); water resources (R $ 4.3 billion); public security (R $ 4.3 billion); indigenous policies (R $ 4.2 billion); sports for major events (R $ R $ 3.7 billion); sports and leisure (R $ 2.8 billion); and agriculture and livestock (R $ 2.7 billion).

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

In spite of the deficiencies notably acknowledged in Plataforma + Brasil (formerly SICONV), this system has been improved over the last decade and today it allows meeting fundamental requirements of good public governance, which requires transparency as a fundamental pillar. In turn, the process of voluntary transfers is still excessively slow, inefficient and effective (Brasil 2018a). According to data from an CGU audit report, the average life cycle time of a federal money transfer process at SICONV reaches more than 5 years. In addition, CGU detected a major imbalance between the operating capacity of the granting agencies and the volume of work expended to analyze the rendering of accounts of the transfers made. The number of transfers made would require an analysis effort much greater than the available analysis capacity of the transferring agencies.

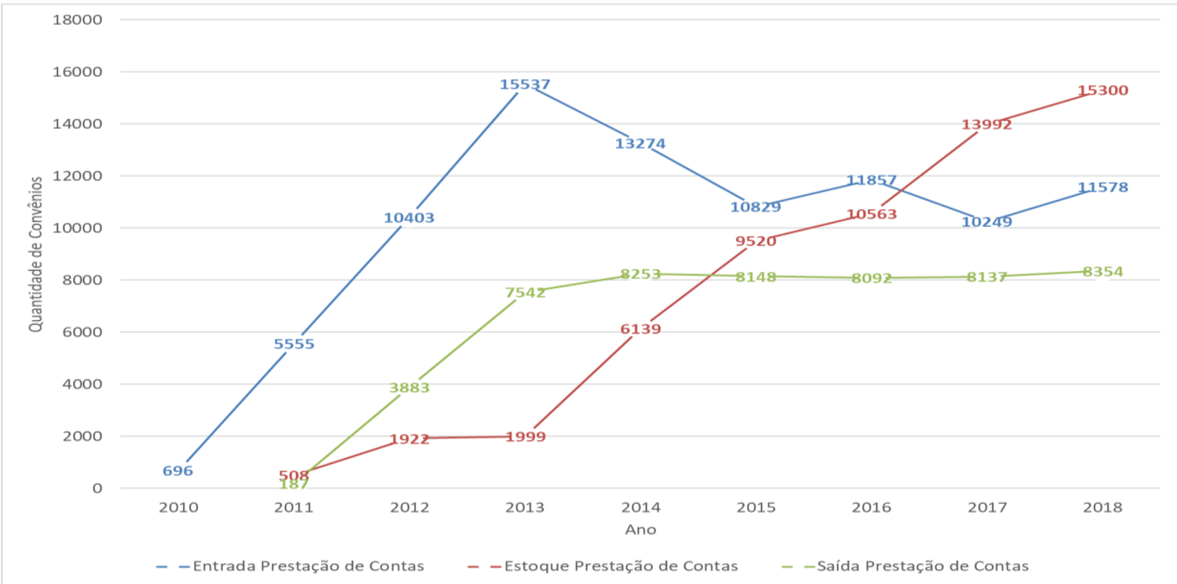

Figure 7 - Convenant stock pending analysis. Prepared by the Author. Adapted from Brazil (2018a).

Source: Plataforma + Brasil.[5]

In view of this operational bottleneck, the problem of stock of accountability pending analysis by Organs transferring agencies emerges. The criticality of the problem is characterized by the continued growth of this stock. From the perspective of government auditing, there are two

[5] Figure 8 was elaborated from the base's history (photograph) on the last day of each year. Only then would it be possible to verify the real situation year-on-year of the situation of the agreements, insofar as the Platform + Brazil database has inconsistencies. Over 10 years, many interventions and updates to the Plataforma + Brasil system have followed. Thus, it would not be reliable to illustrate the evolution of the accountability liability pending analysis by means of a simple arithmetic account, subtracting the amount of accountability submitted for analysis in a given year by the amount that was analyzed in the subsequent year.

lines of action (Brasil 2017; IIA 2012; Olivieri 2011) for the situation illustrated in Figure 8.

The first line is the independent and objective evaluation of the voluntary transfer operation process, aiming at improving risk management, internal controls and governance (Brasil 2017a). The immediate result of acting in this approach would be some recommendations whose content would demand that the resources transferring agencies make efforts to carry out the accountability analysis, or if they do not have the analysis capacity, refrain from promoting new transfers. These are relevant and useful, but obvious, recommendations that are just the same (Power 2003b).

On the other hand, the second line of action consists of consultancy, in which the audit provides advisory and advisory services in strategic matters (Brazil 2017a) of the process of voluntary transfers from the union. The result of the consultancy is usually a product or a solution built together with the manager. In the case at hand, the consultancy produced a solution for the stock of accountability of onlending pending analysis.

After the completion of the performance of the internal audit following the precepts of the way of acting in consultancy, a system was delivered that allows to solve the flagrant imbalance between the operational capacity of the granting agencies and the volume of work required to analyze the rendering of accounts. This imbalance generated a liability of more than 15 thousand instruments, which correspond to approximately R $ 17 billion pending analysis. This system was called Malha Fina de Convênios and consists of the use of artificial intelligence.

Based on the characteristics of each agreement or transfer agreement, the tool recognizes standards and allows to predict, with a high degree of precision, the result of the analysis of accounts, if a manual evaluation was carried out by employees of the granting federal agencies (Huang and Vasarhelyi 2019 ; Issa, Sun, and Vasarhelyi 2016) . In practice, the application of the "Malha Fina de Convênios" verifies the instruments operationalized in Plataforma + Brasil, uses machine learning with the supervised algorithm *Random Forest* and provides a *score* to measure the probability of approval or disapproval of the accounts. The methodology also combines the issuing of alerts generated in the audit trails applied by CGU, in search of predefined patterns of indications of improprieties or irregularities.

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

3.2. The Solution "*Malha Fina de Convênios*"

The "Malha Fina de Convênios" system is a predictive model created by CGU that allows indicating, with a certain degree of certainty, the result of the analysis of the rendering of accounts of the agreements when they are presented by the agreement (subnational entities that receive funds) to the grantors (Union agencies that transfer resources). In other words, Malha Fina allows to infer whether the accounts of the agreements will be approved or rejected.

In summary, this system was the result of the development of a solution that uses a machine learning algorithm based on the characteristics of the agreements whose accounts have already been analyzed. Between September 2008 and December 2017, more than 61,000 agreements (Brasil 2019b) had their accounts analyzed by the grantors, providing a satisfactory amount of data for the learning of the algorithm to provide results with precision. 104 variables from each agreement were used continuously in the supervised learning algorithm Random Forest, because their performance is better compared to others (Breiman 2001; Domingos 2012).

Likewise, the "Malha Fina de Convênios" will allow greater agility in triggering the trigger of administrative measures to be taken aiming at the avoidance of damage to the Treasury. Between 2018 and 2019, 2,869 Special Account Taking (TCE) processes were initiated related to agreements and onlending contracts with indications of irregularities, which seek to recover losses estimated at more than R$ 6 billion[6].

This system was officially made available to the Federal Public Administration on 11/7/2018 with the publication in the Official Gazette of the Union of Interministerial Normative Instruction nº 5, of 6 November 2018 (Brazil 2018b), edited jointly by CGU, the extinct Ministry of Finance and the extinct Ministry of Planning, Development and Management. The aforementioned Normative Instruction stipulated the date of 08/31/2018 as a time limit to define the inventory liability. Thus, all account payments sent to Organs granting agencies until 08/31/2018 are considered stock and their analysis can be supported by the "Malha Fina de Convênios" (Table 2).

[6] Data extracted from the e-TCE system, maintained by the Comptroller General of the Union and the Federal Court of Accounts.

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Then, the Normative Instruction ME / CGU nº 1, of February 14, 2019 (Brazil, 2019a), expanded the scope of applicability of the "Malha Fina de Convênios" for rendering of accounts submitted for analysis as of 09/01 / 2018, although the liability stock related to the base date of 08/31/2018 is cleared.

Thus, the two aforementioned Normative Instructions outlined the scope of coverage of the "Malha Fina de Convênios" and established the debut milestone. The operation of the system at Plataforma + Brasil started with the publication of IN MP / MF / CGU nº 05, of 11/06/2018. This Normative Instruction contributed significantly to the reduction of the average term of the accountability phase, which in 2018 exceeded 2.5 years. With the innovation, there was an immediate benefit resulting from the lower costs of administrative processes, such as the number of servers related to the analysis of the existing liabilities (Issa, Sun, and Vasarhelyi 2016; Wirtz, Weyerer, and Geyer 2019).

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Granting Superior Agency	Total Covenants		Total Value of Agreements	
Ministry of Health	2243	14.66%	R $ 2,064,572,925.18	12.36%
Ministry of Tourism	2115	13.82%	R $ 1,070,053,176.45	6.41%
Ministry of Sport	1416	9.25%	R $ 1,301,841,452.22	7.80%
Ministry of Justice	1205	7.88%	R $ 1,302,123,217.50	7.80%
Ministry of National Integration	1109	7.25%	R $ 1,023,379,009.39	6.13%
Ministry of Education	1059	6.92%	R $ 1,444,256,052.34	8.65%
Ministry of Agriculture	894	5.84%	R $ 516,133,850.15	3.09%
Presidency of the Republic	830	5.42%	R $ 495,708,688.49	2.97%
Ministry of Culture	823	5.38%	R $ 587,646,518.43	3.52%
Ministry of Cities	760	4.97%	R $ 371,683,423.51	2.23%
Ministry of Social Development	653	4.27%	R $ 2,879,625,133.53	17.24%
Ministry of Human Rights	545	3.56%	R $ 370,170,155.79	2.22%
Secretariat for Family Agriculture and Agrarian Development	469	3.07%	R $ 648,547,655.40	3.88%
Ministry of Agrarian Development	305	1.99%	R $ 473,913,885.79	2.84%
Ministry of Labor and Employment	258	1.69%	R $ 518,660,571.00	3.11%
Ministry of Defense	230	1.50%	R $ 637,759,438.63	3.82%
Ministry of Science, Technology, Innovations and Communications	192	1.25%	R $ 649,926,740.64	3.89%
Ministry of Labor and Social Security	70	0.46%	R $ 138,538,169.93	0.83%
Ministry of the Environment	53	0.35%	R $ 39,800,858.50	0.24%
Ministry of Development, Industry and Foreign Trade	52	0.34%	R $ 146,587,135.28	0.88%
Ministry of Mines and Energy	16	0.10%	R $ 15,770,059.31	0.09%
Ministry of s Transport	1	0.01%	R $ 1,430,704.88	0.01%
Electoral justice	1	0.01%	R $ 677,247.82	0.00%
Ministry of Planning, Development and Management	1	0.01%	R $ 777,825.00	0.00%
Total Geral	**15300**	**100,00%**	**R$ 16.699.583.895,16**	**100,00%**

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Table 2 - Stock of agreements on 08/31/2018. Prepared by the Author. Source: Plataforma + Brasil. .

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

The need to present a solution to the adversity imposed by the increasing number of agreements without conclusive accountability was encouraged by the recurring plea from the managers involved in the operation of the agreements. Even before the institution of the former SICONV in 2007, the problem of liability for accountability under an agreement has been an inconvenient hitch for Organs granting bodies and the control bodies that routinely need to relax the requirements imposed for deadlines for the completion of accountability. In this way, the order for research and criteria that could support the analysis of rendering of accounts arises more quickly (Brazil 2017a, 2017b) in view of the constant increase in the number of instruments in stock.

The main objective of the “Malha Fina de Convênios” is to solve the critical problem of lack of operational capacity that involves the process of voluntary transfers from the Union. . By the end of August 2018, the number of agreements inserted in a liability stock awaiting accountability had already totaled 15.3 thousand instruments, in the total amount of R $ 16.7 billion.

The target audience of this initiative is all the granting agencies of discretionary transfer funds operated in Plataforma + Brasil through agreements and onlending contracts. Specifically, the sectors that have benefited the most or may benefit from the Malha Fina system of agreements are those responsible for the financial execution and accountability analysis within the agencies. Generally, the administrative structure in force in the bureaucratic apparatus of the federal government calls these sectors as the Directorate for Internal Management, the Directorate for Finance and Accounting, the Secretariat for Planning, Budget and Administration, the Directorate for Administration, among others. Thus, the target audience that determines the “Malha Fina de Convênios” is 140 agencies of the Federal Public Administration Direct, Autarchic and Foundational that sign agreements on the aegis of the Platform + Brazil rules.

The essence of the "Malha Fina de Convênios" lies in the application of an algorithm that assigns an individual score to each agreement, ranging from 0 to 1. The closer the score is to 0, the greater the chance that the agreement will have its accounts disapproved. Alternatively, the closer to 1, the greater the chances that the agreement will have its accounts rejected. Consequently, the rejection of the accounts of an agreement entitles the grantor to take the

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

appropriate measures to recover the damage to the Treasury, such as, for example, the establishment of a Special Accountability (ECA).

In addition, it is worth noting that the “Malha Fina de Convênios” system meets the premises established in the “Digital-Era Governance” model and endorses the rupture in relation to “New Public Management” (Dunleavy et al. 2006). There is an efficiency gain when implementing the “Malha Fina de Convênios”, while the routine work of public servants who analyze agreements is now rationalized weberian, in its strict sense, since from the results from the data processing performed through the system, it will be determined how many and which covenants should be evaluated more closely by people.

Therefore, “Malha Fina de Convênios” is not just a technological innovation that makes it possible, for example, to digitize the rendering accounts, and that makes work more efficient by streamlining accountability through treatment and analysis data (formerly New Public Management view); rather, a new role given to technology that is part of public policy itself, in this case of the Voluntary Transfers of the Union (Margetts and Dunleavy 2002; Barzelay 2018; Bresser-Pereira 1998).

3.3. Problem Description

The “Malha Fina de Convênios” would help to assist in the analysis of accountability, since this system has all the qualities of what Huang and Vasarhelyi (2019), Sun (2019) and Zhang (2019b) call Robotic Process Automation (RPA). An RPA serves to automate well-defined and repetitive tasks, in this case the analysis of accountability of agreements by Organs granting agencies. It is enough that the agency stipulates a minimum score before which all agreements classified below it are approved. However, are there any hypotheses and conditions in which the adoption of the “Malha Fina de Convênios” in the accountability analysis is more assertive?

A strong objection against the logic of the “Malha Fina de Convênios” systematics emerges: to what extent the precision of the artificial intelligence algorithm (Sun 2019; Zhang 2019) would prevent its adoption for automatic account analysis, without the need for conventional evaluation by a public servant of the granting agency?

TESOURONACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

First, in order to answer this question, it is essential to measure the accuracy of the algorithm according to the variation in the range of scores (scores), which can range from 0 to 1. Invariably, the adoption of Malha Fina for analysis of accountability will focus on the definition of a score threshold in which approved agreements would be admitted. All agreements with a score above this threshold would be considered reprehensible, requiring a conventional analysis. Thus, the distribution of grades by their accumulated interval becomes very relevant, since the error rate of the algorithm must be considered in the entire interval after a certain threshold.

As an example, looking at the data shown in Table 3 and the illustration in Figure 9, if a particular agency stipulates a score of 0.8 as its threshold, it means that 79.4% of its agreements may be subject to tacit approval, and among these, 4.62% would be inadvertently approved, as the classification algorithm is not perfect. Notably, the granting agency's rating threshold reflects its risk appetite. According to the curves in Figure 9, the greater the risk appetite, the higher the threshold score and the greater the chances of tacitly approving reprehensible agreements.

Accountability Situation	**Accumulated *Score* Interval**						
	[0,0 a 0,4)	**[0,0 a 0,5)**	**[0,0 a 0,6)**	**[0,0 a 0,7)**	**[0,0 a 0,8)**	**[0,0 a 0,9)**	**[0,0 a 1,0]**
Approved	201	436	639	842	1.012	1.192	1.305
Approved with caveats	9	21	44	84	162	264	336
Rejected		1	6	12	22	90	307
Total	**210**	**458**	**689**	**938**	**1.196**	**1.546**	**1.948**

Table 3 - Distribution of the accumulated range of grades awarded by the "Malha Fina de Convênios" of the covenants analyzed between 01/01/2018 and 04/18/2018.

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

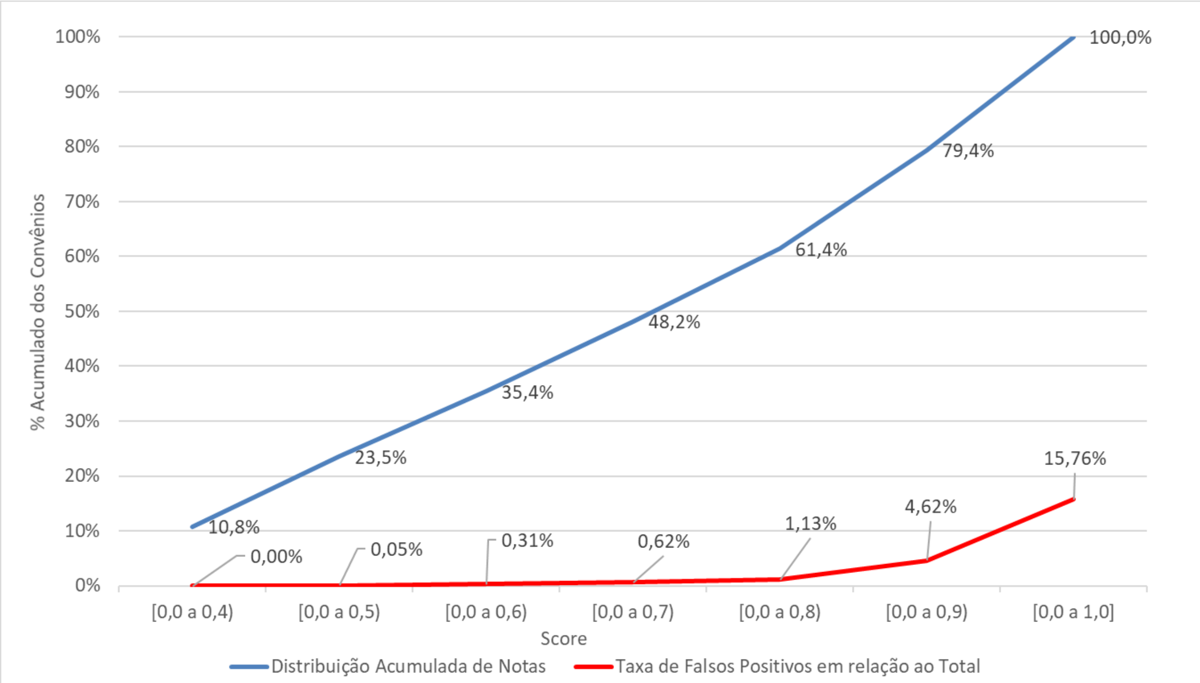

Figure 8 – Accuracy of the " Malha Fina de Convênios". Covenants analyzed between 01.01.2018 and 04.18.2018. Prepared by the Author.

However, the nature of action of public policies of agencies that transfer funds is plural and very different from each other. The curves shown in Figure 9, contemplate the entire spectrum of instruments celebrated in Plataforma + Brasil by federal agencies, in an aggregate manner. In other words, the error rates shown do not take into account the intrinsic qualities of the transferring body, of the public policy that is competent, let alone the context in which it is inserted. Consequently, these variables give rise to the possibility that the error rates are different according to the identity of the granting agency, which motivates the proposition of 2 hypotheses:

Hypothesis 1: would the accuracy of the algorithm be better if machine learning was applied to each grantor separately?

Hypothesis 2: would the performance of the algorithm behave differently if there was segregation by the granting agency?

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

In turn, the data obtained on the disapproval rates of health insurance accounts presented by the beneficiaries, was restricted to a period of less than 4 months, with repercussions between 01/01/2018 and 04/18/2018. The algorithm has learned to classify the agreements based on all those with analyzed accounts (approved, approved with reservation or rejected) from the beginning of data registration in the platform of the Platform + Brazil between September 2008 until 12/31/2017.

After realizing that algorithms based on machine learning by Random Forest make mistakes, in line with what is argued by Alpaydin (2020), Issa, Sun and Vasarhelyi (2016) and Sun (2019), a central question is established about volumetry of the training sequence used for the algorithm to learn. This is the optimum point of the size of the training sequence that would provide security in the adoption of the system, minimizing the risk of approval of reprehensible agreement accounts (false positives) and maximizing automatic analysis, without the need for conventional evaluation by a public servant granting agency. As seen in Figure 9, in cases of evaluation performed automatically by the system, there is an inherent rate of probability of “diagnostic errors”, that is, disapproved agreements, but classified with a score below the threshold stipulated for tacit approval.

In this sense, there is the possibility of carrying out simulations in which it would be possible to measure and measure the reliability of learning performance according to the aggregation of data over time. For example, the algorithm could be run in annual batches considering the state of the base on 12/31. Subsequently, the result of the analysis of the rendering of accounts in the subsequent year and the comparison with the marks attributed by the system would be verified. Therefore, the prospect of clarifying these uncertainties justifies the proposition of two more hypotheses:

Hypothesis 3: Does the machine learning algorithm's classification accuracy increase as there is more data in the learning universe?

Hypothesis 4: is it possible to predict the existence of a saturation point in which the amount of data is no longer relevant for the machine learning algorithm to classify covenants?

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Associated Construct	Hypotheses
Identity of the resource transferor	**H1**: would the accuracy of the algorithm be better if machine learning were applied to each grantor separately? **H2**: would the performance of the algorithm behave differently if there was segregation by the granting agency?
Data training size	**H3:** Does the machine learning algorithm's classification accuracy increase as there is more data in the learning universe? **H4:** is it possible to predict the existence of a saturation point in which the amount of data is no longer relevant for the machine learning algorithm to classify covenants?

Table 4 - Relationship between Constructs and Research Hypotheses.

3.4. Importance of the Problem

The continuous increase in the number of account payments pending conclusive analysis has very serious negative impacts. The existence of a large number of agreements awaiting analysis of their accountability generates consequences such as: the creation of a stock; ineptitude in the evaluation and monitoring of public resources transferred to subnational entities; inability of the Union to assess the effectiveness of public policy on voluntary transfers; and non-compliance with the legal duty to render accounts, which reflects a total absence of accountability (Schedler 1999). In effect, the “Malha Fina De Convênios” directly attacks the cause of the problems (Figure 10), which will provide a reversal of the deleterious effects that

TESOURONACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

the stock of accountability generates in the public policy of voluntary transfers from the union.

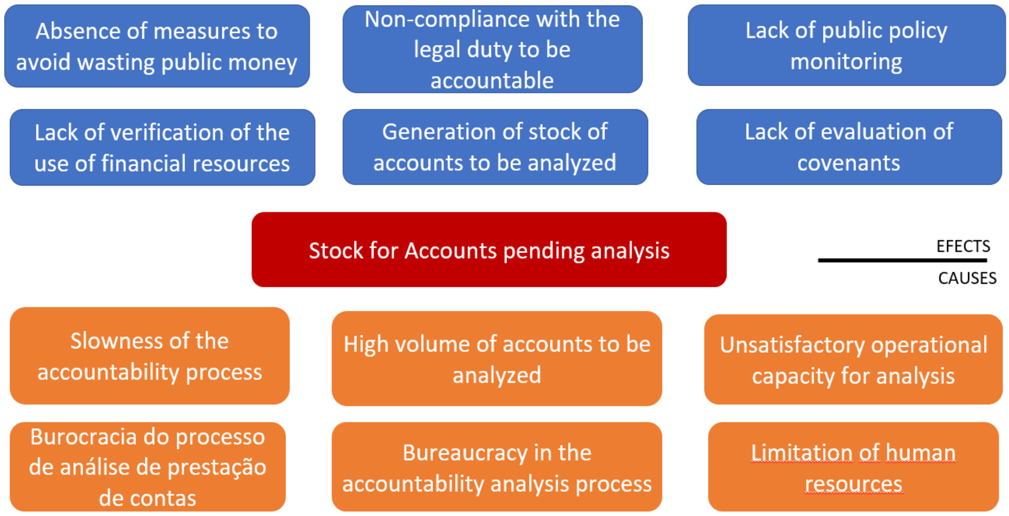

Figure 9 – Problem tree of the accountability phase.

Assuming that the “Malha Fina de Convênios” system can be seen as a project, it is to be expected that it offered a result aiming to solve the liability problem by employing a temporary effort. Thus, the means and resources employed in this temporary effort were related to the problems illustrated in Figure 10. The methodology of the objectives tree is properly applied in the development project of Malha Fina. Jackson (1997) states that once the objectives are defined, the analysis of the project's action strategy can become more logical, showing the relationship between the listed problems and the proposed objectives. Consequently, the changes brought about with the purpose of solving the problem on screen aimed at achieving substantial objectives such as: the elimination of stock; aptitude in the evaluation and monitoring of public resources transferred to subnational entities; the Union's ability to assess the effectiveness of public policy on voluntary transfers in a timely manner; and, last but not least, the observance of accountability through the fulfillment of the legal duty to be accountable (Figure 11).

Therefore, the validation of the automated accountability method is essential for the continuity of the innovative approach proposed by the “Malha Fina de Convênios”. When clarifying the conditions in which the adoption of Malha Fina in the accountability analysis is more assertive, notably through the constraints Identity of the resource transferor (H1 and H2) and Size of data training (H3 and H4), possible objections against its implantation will be discarded.

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

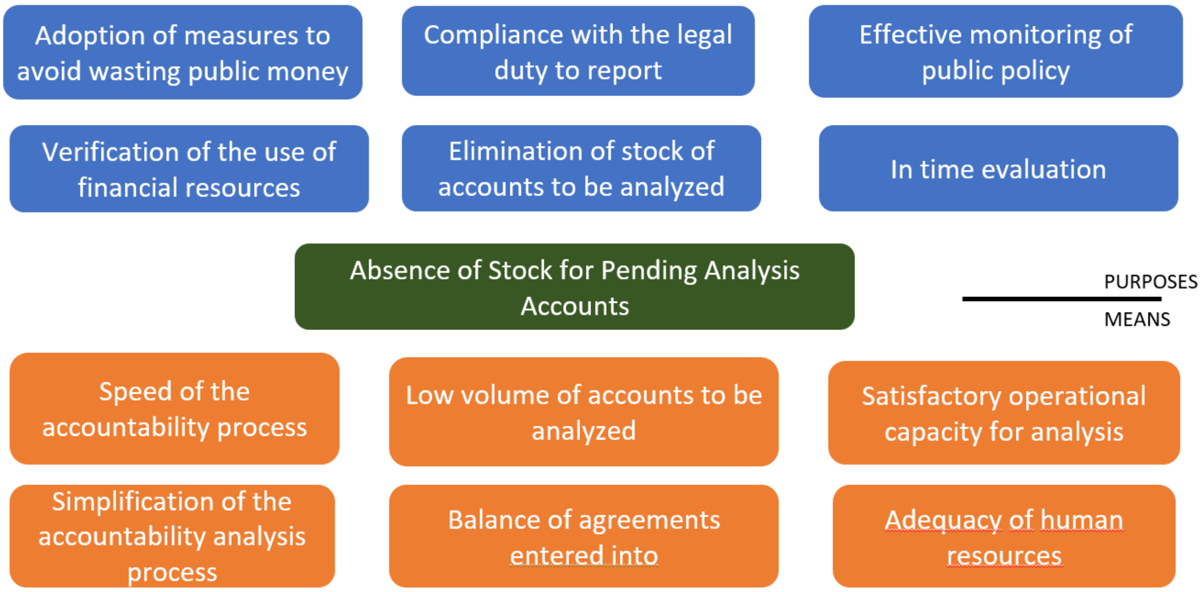

Figure 10 - Objectives tree to solve the accountability bottleneck.

Furthermore, the benefits arising from the use of Malha Fina are very relevant to minimize the risk of audit activity, a fact that further strengthens the importance of testing the proposed constructs. Indeed, the audit activity, like any other operation of an organization, is susceptible to the risk of failure (Brasil, 2017a). In addition, auditors need to make decisions aimed at safety in issuing opinions on what is being assessed (Huang and Vasarhelyi 2019; Stanisic, Radojevic, and Stanic 2019). However, these decisions are contrasted between the scarcity of inputs and the pressure to ensure and certify operations whose auditable universe would consume much greater scope, time and resources than those available (GAO 2011; Zhang 2019b). This dichotomy creates the risk of auditing.

The audit risk is the auditor's inability to detect an error in a sample (Al-qudah, Baniahmad, and Al-fawaerah 2013). Failure to detect a sample error can be caused by fatigue or inattention by the auditor, but it can also be caused by the application of an inappropriate audit procedure (IIA, 2012). In the light of this scenario, the mechanisms of robotic automation and artificial intelligence that enjoy the recognition of patterns for defining recurring events are configured as a solution to solve the problem between the scarce audit resources versus the security in issuing opinions about the operation audited (Power 2009). In this perspective, the "Malha Fina de Convênios" consists of an audit mechanism of this category.

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

4. METHODOLOGY

In this chapter, the methodological strategy, inspired by Bao et al. (2020), Cecchini et al. (2010), Grover, Bauhoff and Friedman (2019), Hooda, Bawa and Rana (2020), Stanisic, Radojevic and Stanic (2019) and Zhang (2019a). This strategy, in turn, mainly used the empirical implementation of a computational environment in phyton language. Therefore, the performance of the algorithm of the “Malha Fina de Convênios” (Random Forest) will be evaluated against several scenarios designed exclusively to test the constraints Identity of the resource transferor (H1 and H2) and Size of data training H3 and H4).

Naturally, the achievement of the objectives of this research was made possible through the following activities:

- ✓ Mapping the life cycle of an agreement operated by Plataforma + Brasil and listing all the situations that await analysis by the granting federal agency in the rendering of accounts presented by the sub-national entity (BRASIL, 2019b), as well as a description of the universe of agreements in the Plataforma + Brasil, including the time cut.
- ✓ Setting up a test environment, whose computational performance was satisfactory, with databases from different dates so that the “Malha Fina de Convênios” algorithm trains with different data series, as emphasized by Stanisic, Radojevic and Stanic (2019) and Zhang (2019a).
- ✓ Identification of the necessary requirements for the construction of an analysis model that demonstrates the relationship of the different scenarios to be tested in the hypotheses, ensuring that the available computational resources are not an obstacle to performance (Breiman 2001; Domingos 2012). Therefore, there is a need to develop a viable architecture for the execution of the test scenarios, given that these are very numerous, in the order of hundreds;
- ✓ Analysis and interpretation of the results of the adopted model, providing elements necessary for the conclusions set out in the present research, in the light of the script suggested in the research conducted by Sun (2019).

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

According to Creswell (2007), there are three research techniques: Quantitative, Qualitative, or Mixed Methods. The most appropriate research technique to study the proposed hypotheses is Quantitative, as this technique uses data collection and statistical instruments. Necessarily, the use of comparative data, derived from the different supervised machine learning algorithms Random Forest developed in the computational environment in Phyton language with the purpose of testing the hypotheses, will be a key point for the unfolding of the research.

The greatest risk inherent in the process of the “Malha Fina De Convênios” system is the inadvertent classification of agreements whose accounts were rejected with a score close to zero (false positive). Thus, the performance of the AI algorithms designed to test the hypotheses will be measured by the rate of occurrence of false positives over the score domain between 0 and 1. In turn, the accuracy of the algorithm is checked by checking the distribution of covenants failed over the interval [0, 1].

While the Random Forest supervised learning algorithms are of high performance (Bao et al. 2020; Hooda, Bawa and Rana 2020; Breiman 2001; Grover, Bauhoff, Friedman 2019), this has little value in the proposed constructs. The main objective in testing the hypotheses proposed in this research is to accurately detect as many disapproved agreements as possible, without mistakenly classifying them as approved, that is, the rate of false positives in relation to the total. In this respect, it is important to highlight a peculiarity in relation to the specificity and sensitivity metrics, as these metrics are not related to the total population, but to the approved (sensitivity) and disapproved (specificity) covenants categories.

Furthermore, the metrics used to measure performance generally used (accuracy, sensitivity and specificity) are not appropriate for tackling this research problem, given the imbalance in the balance of the training sequence (Alpaydin 2020; Bao et al. 2020). Failed covenants represent less than 3% of the total. Naturally, the strategy of assessing specificity through the correct rate of classification of failed agreements would lead to very high accuracy, which would be naive to assess performance.

The imbalance of the training sequences, as portrayed in the literature (Bao et al. 2020; Breiman 2001; Fawcett 2006; Domingos 2012; Stanisic, Radojevic, Stanic 2019), can be seen in Table 5, below. While the extinct Secretariat for Agricultural Development and Cooperativism did not reject any account rendering of the 1,815 agreements in which that Secretariat was a

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

resource transferor, the Ministry of Tourism failed 779 rendering of accounts among the 6,796 agreements signed by this Ministry as a transferor of resources.

Agency Description	Accountability			
	Rejected	Approved	Approved with Caveats	Total
Secretariat for Agricultural Development and Cooperatives	0	1795	20	1815
Ministry of Cities	2	9012	233	9247
Ministry of Justice and Public Security	7	688	151	846
National Health Foundation	16	677	297	990
Sport Ministery	29	3553	273	3855
Ministry of Agriculture, Livestock and Supply	46	12453	249	12748
Ministry of Health	84	2295	3148	5527
defense Ministry	134	1104	206	1444
Ministry of Citizenship	137	2274	377	2788
Ministry of Tourism	779	4972	1045	6796
Total 10 Agencies	**1234**	**38823**	**5999**	**46056**
Total All Agencies	**1533**	**43575**	**6737**	**51845**

Table 6 – Distribution of the accountability situation of the covenants of the 10 most representative bodies of the population. Source: Plataforma + Brasil. Base date: November 2019.

Likewise, Figure 12 illustrates the difference in accuracy in the analysis of accountability among the 10 elected bodies (Table 6) for testing the hypotheses of the construct “Identity of the resource transferor”. The Ministries of Tourism, Citizenship and Defense clearly reject more accountability than the defunct Secretariat for Agricultural Development and Cooperativism and the Ministries of Cities and Justice and Public Security. The first bodies are presumed to be more austere and measured in accountability analyzes, while the second are more lenient. However, it should be stressed that such a statement is merely suggestive, as it would be necessary to assess the merit of each accountability presented to those bodies with a low failure rate to support any allegation about the zeal and diligence about its analysis, after all the transfers of resources to subnational entities are surrounded by idiosyncrasies (Amorim Neto and Simonassi 2013; Abrucio and Franzese 2007; Fajardo 2016). The fact that the agreements

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

signed in these bodies are conducted with the proper and regular application of public resources cannot be ruled out, which is why the accounts of their agreements are never rejected.

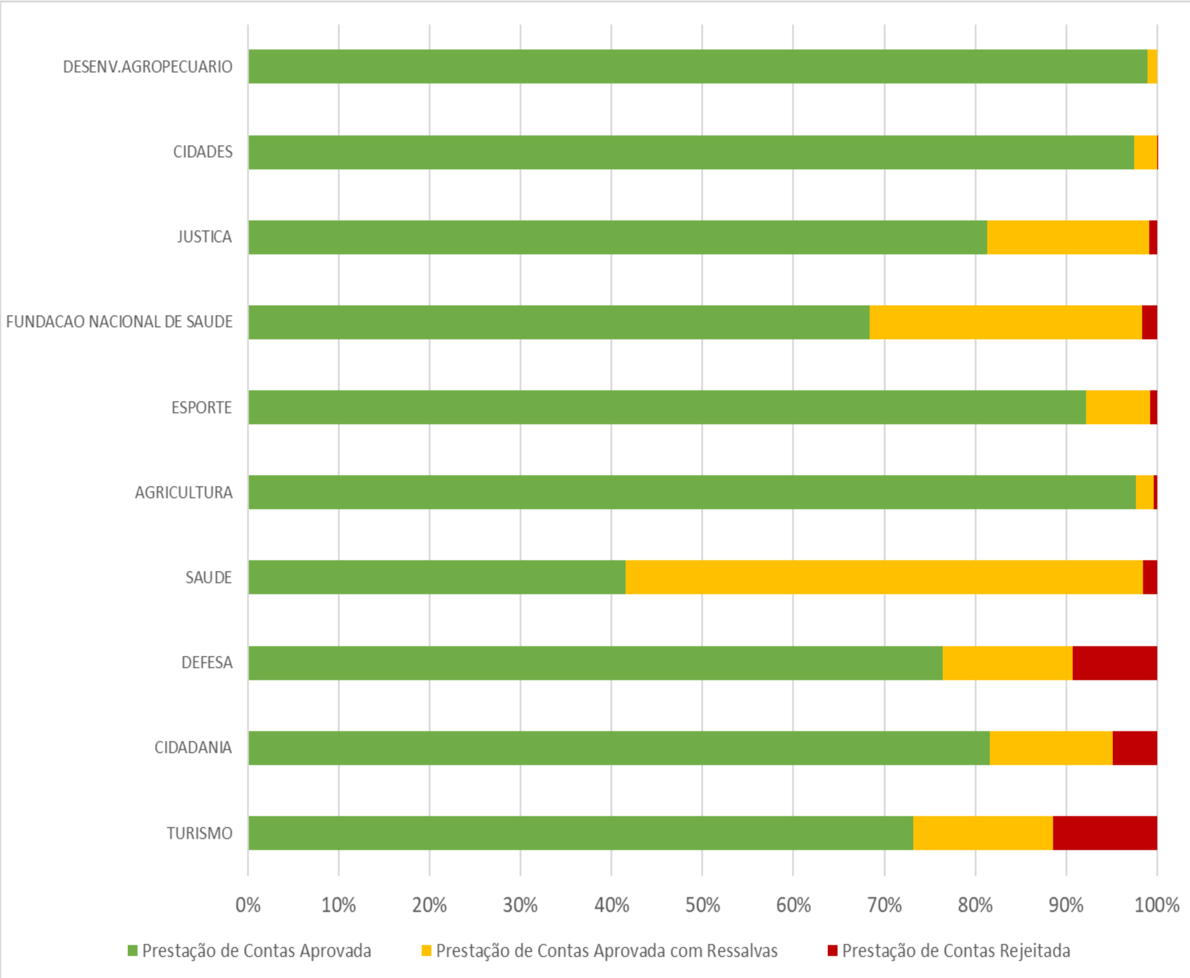

Figure 11 - Percentage distribution of resultad the analysis of the agreements of the checks and balances of the 10 most representative bodies of the population. Source: Plataforma + Brasil. Base date: November 2019.

Therefore, this research methodology will use the metric AUC, obtained by calculating the area of the ROC curve, in addition to proposing a new scalar metric derived from accuracy (Alpaydin 2020), which we call inaccuracy (ε). The inaccuracy value (ε) of an algorithm trained by a sample of covenants in a given interval is:

$$\varepsilon_{ij} = \frac{FP_{ij}}{TP_{ij} + TN_{ij} + FN_{ij} + FP_{ij}}$$

In which:

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

FP is False Positive;

TP is True Positive;

TN is True Negative;

FN is False Negative;

i: score ranges [0,0 to 0,1); [0,1 to 0,2); [0,2 to 0,3); [0,3 to 0,4); [0,4 to 0,5); [0,5 to 0,6); [0,6 to 0,7); [0,7 to 0,8); [0,8 to 0,9); and [0,9 to 1,0];

j: algorithm trained by a certain sample of agreements, according to the volume or identity of the grantor;

Therefore, the inaccuracy (ε) of an algorithm for an accumulated range in the score range is::

$$\varepsilon_j = \sum_{0}^{i} \frac{FP_{ij}}{VP_{ij} + VN_{ij} + FN_{ij} + FP_{ij}}$$

Additionally, the AUC metric for a given algorithm trained by a sample of covenants is calculated by the integral of the ROC curve, calculating the function of the rate of true positives in relation to false positives:

$$\mathrm{AUC}_j = \int_{0}^{1} \frac{VP_j}{VP_j + U - (VP_j + VN_j + \boldsymbol{FP_j})} d\boldsymbol{FP_j}$$

In which:

FP is False Positive;

TP is True Positive;

TN is True Negative;

U is the Universe of Convenants, U – (TP +TN +FP) é igual a FN;

j: algorithm trained by a certain sample of agreements, according to the volume or identity of the grantor;

Encouraged by Cecchini et al. (2010), Fawcett (2006) and Grover, Bauhoff and Friedman (2019), the proposition of the metric inaccuracy (ε) aims to measure the real performance of the algorithm on the test conditions imposed by the constraints Identity of the resource transferor (H1 and H2) and Data training size (H3 and H4) focusing on the rate of false positives in relation to the total number of covenants. This is due to the fact that the specificity metric

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

reflects the proportion of the approved covenants category, while the sensitivity metric depicts the proportion of the failing covenants category, that is, they are metrics that address performance separately in subsets of the population universe. In turn, the accuracy metric measures the rate of true positives and negatives jointly under the two categories of approved and disapproved agreements, making it inappropriate, since it is intended to analyze performance from the perspective of false positives in relation to the total.

In view of the above, the performance of the algorithms generated by parameterizing the identity of Organs granting bodies and the volume of the training sequence will be evaluated by the metrics AUC and inaccuracy (ε). The AUC metric will reflect the intrinsic quality of the generated algorithm in relation to its training sequence (Larcker and Zakolyukina 2012) and the inaccuracy metric (ε) will reflect the accuracy of the algorithm when it classifies agreements different from those used in its training (Cecchini et al. 2010). In all the scenarios of this research, the test data were those agreements with life cycle closed at Plataforma + Brasil.

On the other hand, it is worth noting that the decision making in this research by the AUC and inaccuracy (ε) metrics to evaluate the performance of Machine Learning algorithms was based on a recurring problem in the Control and Audit research field. Within the social science Administration (Fayol 1949; Taylor 1911), research focusing on the accuracy of AI algorithms undertake efforts to clarify the effectiveness of replacing human intervention, notably an auditor, by a computational mechanism to issue an opinion on an operation business organization. These operations include reports on financial statements, regularity of processes or proof of expenses. In this regard, Bao et al. (2020), Cecchini et al. (2010), Grover, Bauhoff and Friedman (2019), Fawcett (2006), Hooda, Bawa and Rana (2020), Larcker, Zakolyukina (2012), Stanisic, Radojevic and Stanic (2019) and Zhang (2019a) use in their researches the AUC metric and a derivation of the metric inaccuracy (ε) to the detriment of other metrics used in research in the exact sciences such as those portrayed by Domingos (2012).

4.1. Relevance for the AI algorithm in distinguishing the identity of resource transferors

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Evidently, each granting body has its own accountability analysis process, with its inherent characteristics and properties. Therefore, by universalizing the machine learning process without distinction among all resource transferors, more stringent bodies in the accountability analysis may suffer an unwanted mitigation in the risk classification of their covenants by adding less stringent organs in the analysis.

The construct "Identity of the resource transferor" in Table 4 was designed with the aim of resolving the doubt raised about the accuracy (Bao et al. 2020; Zhang 2019) in different circumstances regarding the grantor, since the performance metrics of the algorithm were established uniformly for all the resources transferring agencies.

In these terms, the AI algorithm was trained separately for those granting agencies with the greatest representativeness in the universe of the Plataforma + Brasil agreements whose life cycle is closed. This implies that the rendering of accounts of these agreements was analyzed by the granting agency, receiving a classification as: a) approved rendering of accounts; b) accountability approved with reservations; or c) rejected accountability.

The study population consists of a total of 51,845 agreements that total more than R $ 24 billion (R $ 24,481,496,942.56). To the extent that the agreements signed by Organs granting agencies in the sample in Table 6 mirror approximately 90% of the quantitative and 80% of materiality, the statistical representativeness of this sample in relation to the population is satisfied.

Thus, the algorithm was trained separately with 10 different data sets, one for each grantor. Consequently, the same procedure used to generate the "Malha Fina de Convênios" algorithm will be resumed, however, with a substantial difference. This new procedure aims to train the machine learning algorithm, exclusively, according to the identity of the resource transferor, that is, 10 different algorithms will be generated, each one reflecting the way that the resources transferring agencies analyze the accountability presented. Thus, there will be a beacon that will allow the comparison of the metrics AUC and inaccuracy (ε) (Bao et al. 2020; Zhang 2019a;) between (i) the application of the algorithm generated by the entire universe of agreements in each of the forwarding agencies of resources and (ii) the application of the algorithm generated by agreements of a given transfer agency in its own agreements.

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Granting Agency	Amount		Value	
	#	% of the Universe	R $ Million	% of the Universe
Ministry of Agriculture	12828	24.74%	R $ 4,103.58	16.76%
Ministry of Cities	9306	17.95%	R $ 3,632.21	14.84%
Ministry of Tourism	6838	13.19%	R $ 2,997.27	12.24%
Ministry of Health	5642	10.88%	R $ 2,616.30	10.69%
Ministry of Sport	3883	7.49%	R $ 1,307.82	5.34%
Ministry of Citizenship	2820	5.44%	R $ 2,252.18	9.20%
Secretariat for Agricultural Development and Cooperatives	1816	3.50%	R $ 466.27	1.90%
Ministry of Defense	1447	2.79%	R $ 849.52	3.47%
Foundation National of Health	991	1.91%	R $ 516.64	2.11%
Ministry of Justice and Public Security	861	1.66%	R $ 967.54	3.95%
Total Sample	**46432**	**89,56%**	**R$ 19.709,34**	**80,51%**
Universe	**51845**	**100%**	**R$ 24.481,50**	**100%**

Table 5 - Sample used for the tests of the “Identity of the resource transferor” construct.

Thus, the algorithm was trained separately with 10 different data sets, one for each grantor. Consequently, the same procedure used to generate the “Malha Fina de Convênios” algorithm will be resumed, however, with a substantial difference. This new procedure aims to train the machine learning algorithm, exclusively, according to the identity of the resource transferor, that is, 10 different algorithms will be generated, each one reflecting the way that the resources transferring agencies analyze the accountability presented. Thus, there will be a beacon that will allow the comparison of the metrics AUC and inaccuracy (ε) (Bao et al. 2020; Zhang 2019a;) between (i) the application of the algorithm generated by the entire universe of agreements in each of the forwarding agencies of resources and (ii) the application of the algorithm generated by agreements of a given transfer agency in its own agreements.

This strategy will allow the machine learning behavior to be assessed separately by the grantor. It is intended to detect whether training sequences differentiated by the identity of the granting agency, would produce algorithms with different performances to the point of preventing the extrapolation of a single algorithm generated from the learning of the universe.

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Figure 13, below, illustrates the strategy for testing the hypothesis of the “Identity of the resource transferor” construct. Ten different algorithms will be generated according to the resource transfer agency to be compared with the algorithm generated by a training sequence that contemplates the universe of agreements registered in Plataforma + Brasil.

The expected results were to obtain data that enable the accuracy of AI algorithms to be measured using two metrics: inaccuracy (ε) and AUC, as disciplined in Bao et.al (2020), Breiman (2001), Fawcett (2006) and Zhang (2019a). These metrics were used to measure accuracy in two configurations involving the identity of Organs granting agencies. The first configuration, scenario 1, is that resulting from the application of the algorithm generated by the entire universe of agreements in each of the resources transferring agencies. The second, scenario 2, is one in which the algorithm generated by covenants from a given transfer agency was applied to classify their own covenants (Table 6).

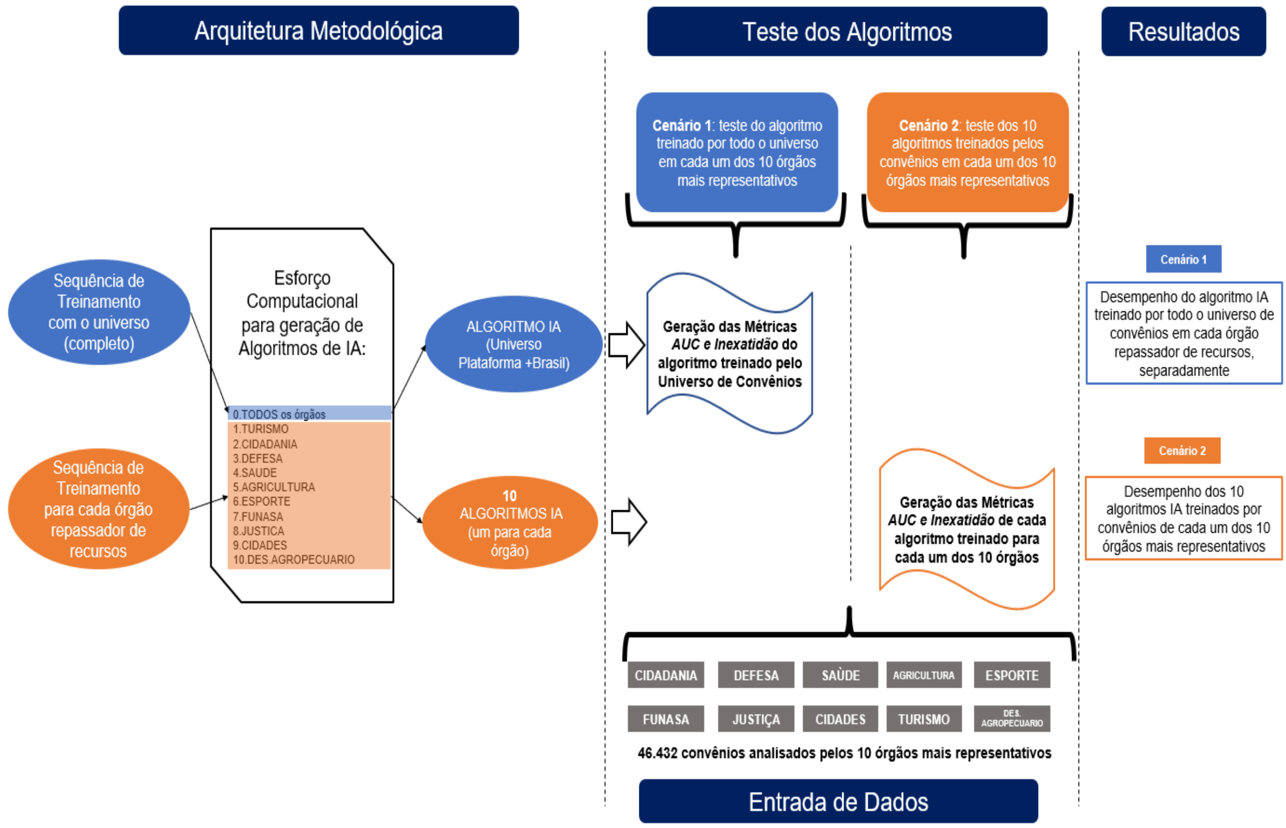

Figure 12 - Methodology to verify the impact of the resource transferor identity on the precision (AUC metrics and inaccuracy) of the artificial intelligence algorithm. Prepared by the author.

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

4.2. Size relevance of the training sequences for the AI algorithm

The construct "Size of data training" (Table 4) was established in order to ascertain the impact that the volumetry of the training series has on the accuracy of the machine learning algorithm. At the heart of the proposed intervention is the verification of the accuracy behavior of different algorithms, which were generated by incremental training sequences, segmented by year of signing the agreements. The behavior of the results of the algorithms will be analyzed as more agreements serve as input for machine learning (Breiman 2001; Zhang 2019).

In supervised learning, algorithms need to be trained first on the training data set and then tested on the test data set. The general rule in the division between training set and test is to choose about 75% of the sample as training data and the rest as test data (Alpaydin, 2020; Hooda, Bawa and Rana 2020; Zhang 2019a;). However, this is not an established rule, as the guarantee of machine learning performance depends on the availability of data labeling (Sun 2019). Another important issue, as emphasized by Cecchini et al. (2010), Grover, Bauhoff, Friedman (2019) and Zhang (2019a) is how to guarantee the performance of machine learning, given the availability and balance of data that are labeled, in the case of the approved or rejected accounts. In this context, the potential of the use of the Malha Fina de Convênios, their impacts already demonstrated and their limitations are still not clear and definitive. Therefore, the motivation for proposing the construct "Size of data training" aims at understanding this scenario.

In January 2020, Plataforma + Brasil had 175,456 registered agreements. Of these, 53,002 completed their life cycle, that is, the fund transfer agency assessed the accountability presented by the agreement definitively, approving it, approving it with reservations, or rejecting it, as shown in Table 7. The essence evaluation of this construct is in the verification of the precision behavior of the "Malha Fina de Convênios" system algorithm over time.

In this way, the algorithm was trained with different data segments according to the year the agreement was signed, thus admitting 9 different data series between the years 2008 to 2016. These series will be analyzed specifically regarding the accuracy behavior of the algorithm.

Additionally, it is important to define that the training sequence of the algorithm must

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

correspond only to agreements whose life cycle has ended, that is, it has passed through the phases of celebration, execution and accountability. Consequently, this requires a reliable and stable source of the Plataforma + Brasil database, which reflects the continuity of the situations of the agreements. As the process of voluntary transfers from the union is inherently volatile (Brollo and Nannicini 2012; Meireles, 2019), changes in the situations of the agreements may occur, especially in those recently concluded. Consequently, the year 2017 was adopted as the cutoff point for the population of health plans to be used as a training sequence.

Year	**Accumulated Qty**	**Accountability**					
		Approved	**Approved with caveats**	**Rejected**	**Total**	**Accumulated**	**% Accumulated**
2008	2746	797	165	182	1144	1144	2,16%
2009	26643	10515	1711	631	12857	14001	26,42%
2010	50186	9918	1495	462	11875	25876	48,82%
2011	63833	5855	1038	84	6977	32853	61,98%
2012	76465	4727	635	40	5402	38255	72,18%
2013	92712	6044	562	67	6673	44928	84,77%
2014	106225	3491	719	30	4240	49168	92,77%
2015	116386	1541	313	19	1873	51041	96,30%
2016	131413	1535	133	9	1677	52718	99,46%
2017	151931	200	22	1	223	52941	99,88%
2018	170683	55	6		61	53002	100,00%
2019	175418	0	0	0	0	53002	100,00%
Total	**175418**	**44678**	**6799**	**1525**	**53002**	**53002**	**100,00%**

Tabela 6 - Amostra utilizada para realizar os testes do construto “*Tamanho de treinamento de dados*”. Data base janeiro de 2020. Fonte: Plataforma +Brasil.

However, by observing Table 7, it is clear that after 2017, only 284 agreements had their life cycles closed. In turn, this quantitative low has no significance for the training sequence. This fact led to the delimitation of the scope in 9 series between the years 2008 and 2016, excluding

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

the 2017 series.

Figure 14 illustrates the process performed to determine different artificial intelligence algorithms with the gradual increase in training sequences. The 9 series used as a training sequence produced 9 distinct algorithms, capable of classifying any Platform + Brazil agreement regarding the probability of the accounts being approved or rejected. In turn, the 9 algorithms generated were tested by analyzing the metrics AUC and inaccuracy (ε) (Bao et al. 2020; Fawcett 2006; Zhang 2019a).

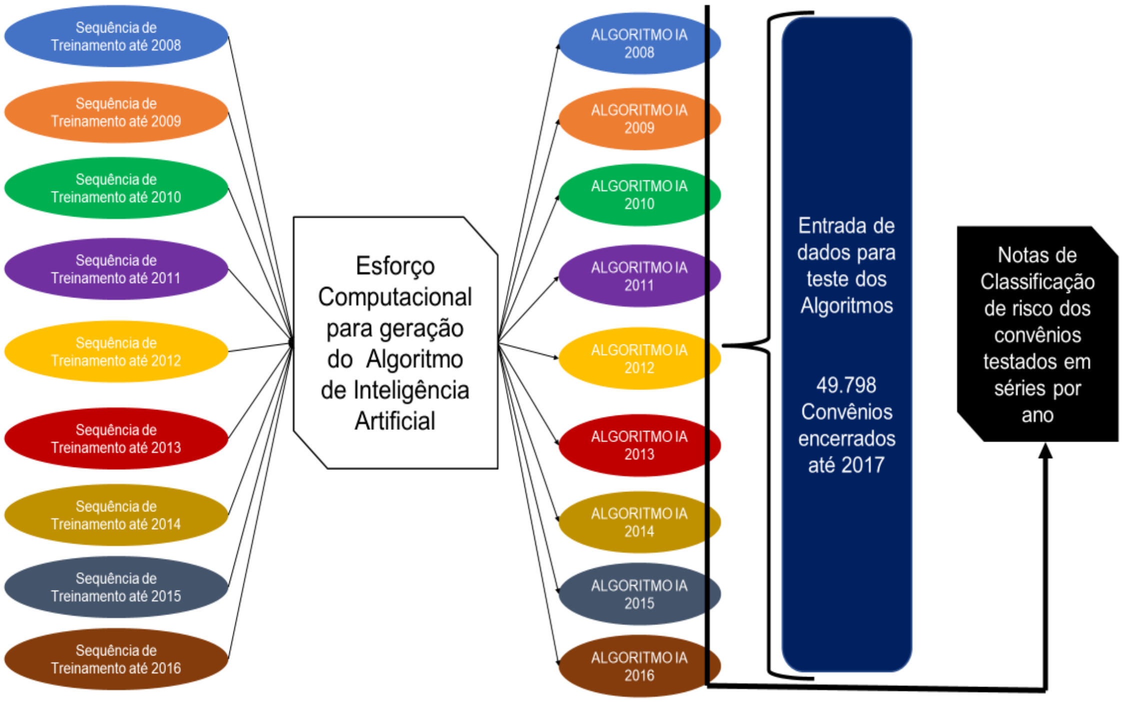

Figure 13 – Methodology to verify the impact of volumetry in the training sequence on the precision (AUC metrics and inaccuracy) of the AI algorithm. Prepared by the author.

Finally, the testing of the algorithms could only be performed with agreements whose life cycle has ended, similarly to the population used for the training sequence. Now, the most assertive way to test the accuracy of machine learning algorithms is to compare them with the actual result of the agreement. Likewise, the algorithm learns from closed agreements, that is, from the variables that led them to have their accounts rejected or closed. Thus, the test of the 9 algorithms consisted of classifying the completeness of the agreements in Table 7.

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

5. ANALYSIS OF RESULTS

The results obtained and their analysis are presented below. The results consist of the metrics AUC and inaccuracy (ε) generated in the different machine learning scenarios, which were conceived through the parameterization of the identity of the resource transferor and the volume of the training sequence.

5.1.1. AUC metrics and inaccuracy (ε) according to the identity of the granting agency

The evaluation of the performance of the accuracy of the trained algorithms according to the identity of the granting agency from the perspective of the AUC metric allows us to conclude that some perform better than the algorithm trained with the complete population of the agreements. However, the evaluation under the perspective of the metric inaccuracy (ε) points out that no algorithm performs better than that trained by the universe. As the performance surpassing from the point of view of the AUC metric occurs in only 4 algorithms, in a residual way, and under the burden of resulting in 6 algorithms with inferior performance, 2 of which are very unsatisfactory, the H1 of the Identity construct of the resource transferor is not confirmed under certain circumstances, which do not entail the adoption of the strategy of training algorithms separately according to the resource transferor body. In turn, H2 is confirmed because performance is different depending on the identity of the granting agency.

5.1.1.1.Scenario 1: the AI algorithm trained with the universe and applied to each grantor separately

The performance of the algorithm trained from the universe is satisfactory from the point of view of the AUC metric, since it assumes a value of 0.96, as shown in Figure 15. This inference is corroborated by Fawcett (2006) and Zhang (2019a) , because they claim that the closer the AUC approaches 1, the better the performance of a classifier algorithm. However, the analysis of the metric inaccuracy (ε) indicates that the algorithm has a different performance over the interval [0 to 1], as there is an increase in the probability of false positives in the agencies whose history of disapproval of covenant accounts is greater than the average, as shown in Table 8. On the other hand, the opposite occurs in the bodies with a history of disapproval of accounts

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

below the average[7], as there is a decrease in false positives. However, the metric inaccuracy (ε) only presents a discrepancy from the interval (0.8 to 1).

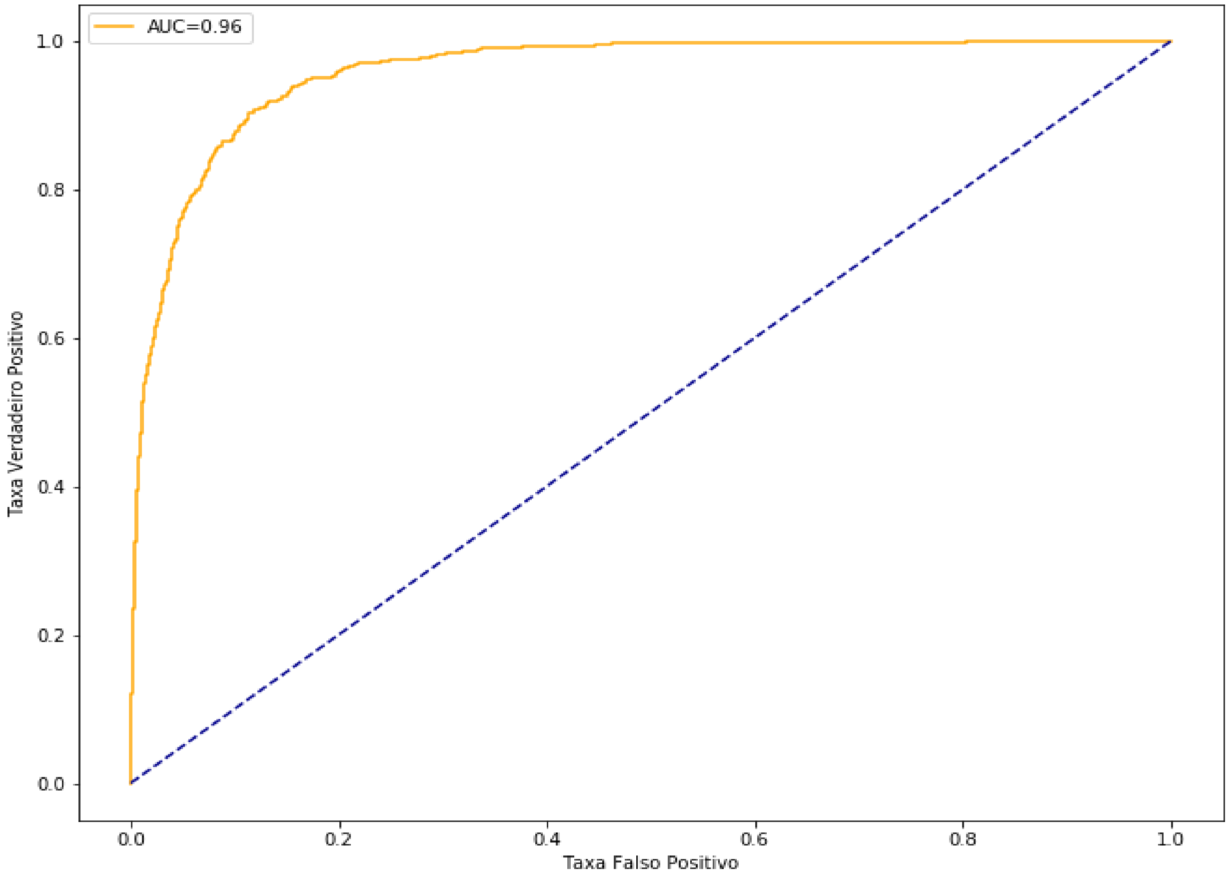

Figure 14 - ROC curve and AUC metric of the AI algorithm trained with the universe..

The finding that the metric inaccuracy (ε) shows deficient values only from the interval (0.8 to 1) is inferred through the analysis of the curves in Figure 16, generated from Table 8. It is observed that the inaccuracy (ε) presents very good values between the range [0 to 0.8). However, the discrepancy of inaccuracy (ε) observed from the 0.8 score is due to overestimation and underestimation of false positives in different identities of Organs granting agencies. The "DEFESA", "TURISMO" and "CIDADANIA" curves, associated with the agreements signed by the Ministries of Defense, Tourism and Citizenship, respectively, are significantly higher than the "All Organs" curve from the interval (0.8 to 1) It is inferred, therefore, that the

[7] The average rejection of accountability in the sample of the 10 bodies in Table 9 (Ministry of Tourism, Ministry of Citizenship, Ministry of Defense, Ministry of Health, Ministry of Agriculture, Ministry of Sport, Funasa, Ministry of Justice, Ministry of Cities , Agricultural Development) is 137,111.

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

discrepant behavior of the curves of the imprecision metric, when the algorithm is applied in these 3 bodies, results from the greater rejection of covenant accounts promoted by these bodies in relation to the average, making the training sequence be less unbalanced, this conclusion is in line with what Alpaydin (2020) and the research findings of Bao et al. (2020) and Cecchini et al. (2010).

In summary, the analysis of the empirical results of the metric inaccuracy (ε) allows us to conclude that the algorithm trained from the entire population of Plataforma + Brasil, when applied separately in each granting agency, presents a lower performance for the Ministries of Defense, Tourism and Citizenship a from the score 0.8, with an increase in performance degradation from the score 0.9. Additionally, the results indicate that the performance of the algorithm in the Ministry of Citizenship agreements presents a moderate degradation start from the 0.8 score. This observation implies that the probability of false positives in this Ministry has a broader spectrum than the Ministries of Defense and Tourism.

In addition, there is a discrepancy from the 0.9 score, implying two perspectives. The first is that the inaccuracy (ε) of 7 organs is underestimated by around -2%, meaning a moderate deviation that does not generate adverse consequences, as it will result in an imposition of conservatism on these bodies in the analysis of their accounts. covenants. Now, the underestimation of risk consists of the indication by the AI algorithm of more false positives, that is, more agreements would be disapproved than they really are. However, overestimating errors has adverse consequences, given that there is a classification of disapproved agreements as possibly approved by the algorithm. This underestimation occurs for the Ministries of Defense, Tourism and Citizenship, reaching divergence rates of approximately 6%, 8% and 2%, respectively.

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Granting Agency	Accounts Rejected	Accumulated *Score* Interval						
		[0.0 to 0.4)	[0.0 to 0.5)	[0.0 to 0.6)	[0.0 to 0.7)	[0.0 to 0.8)	[0.0 to 0.9)	[0.0 to 1.0]
All Agencies	1490	0.00000%	0.00195%	0.02140%	0.03890%	0.10310%	0.277428%	2.89838%
Ministry of Tourism	779	0.00000%	0.01471%	0.04414%	0.07357%	0.17657%	0.44144%	11.46263%
Ministry of Citizenship	137	0.00000%	0.00000%	0.03587%	0.03587%	0.177934%	0.86083%	4.91392%
Ministry of Defense	134	0.00000%	0.00000%	0.00000%	0.00000%	0.00000%	0.27701%	9.27978%
Ministry of Health	84	0.00000%	0.00000%	0.01809%	0.03619%	0.05428%	0.21712%	1.51981%
Ministry of Agriculture	46	0.00000%	0.00000%	0.00000%	0.01569%	0.04707%	0.07844%	0.36084%
Ministry of Sport	29	0.00000%	0.00000%	0.00000%	0.02594%	0.05188%	0.15564%	0.75227%
Funasa	16	0.00000%	0.00000%	0.20202%	0.20202%	0.20202%	0.30303%	1.61616%
Ministry of Justice	7	0.00000%	0.00000%	0.11820%	0.11820%	0.11820%	0.23641%	0.82742%
Ministry of Cities	two	0.00000%	0.00000%	0.01081%	0.01081%	0.01081%	0.01081%	0.02163%
Agricultural Development	0	0.00000%	0.00000%	0.00000%	0.00000%	0.00000%	0.00000%	0.00000%

Table 7 – Inaccuracy (ε) metric of the AI Algorithm trained from the universe (False Positive Probability).

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

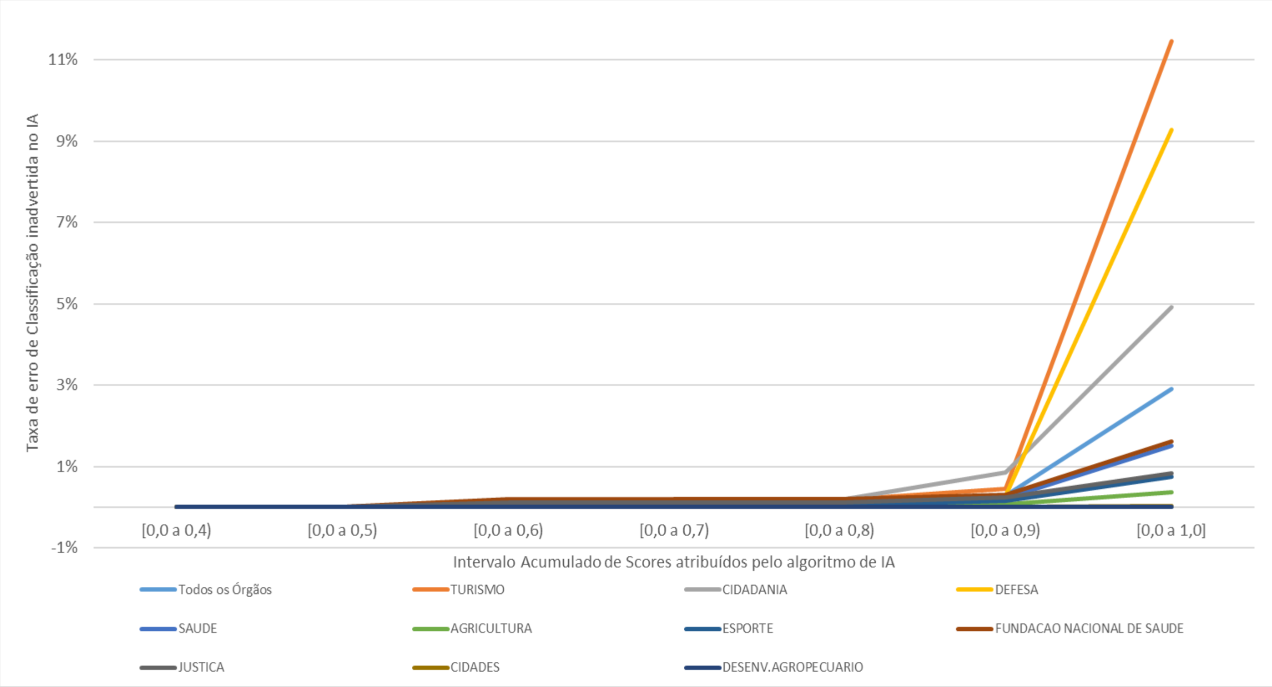

Figure 15 – Inaccuracy (ε) metric of the AI Algorithm trained from the universe (False Positive Probability).

5.1.1.2.Scenario 2: AI algorithm trained with covenants according to the identity of the granting agency

The performances of the algorithms trained based on the agreements of each granting body present different behaviors from the perspective of the AUC metric, some showing very satisfactory performance, while others quite unsatisfactory. Considering the algorithm trained from the universe as benchmarking, whose AUC is 0.96, the algorithms trained by agreements of the Ministries of Tourism, Agriculture, Sport and Cities, whose AUC metrics assumed 0.97, 0.97, 0.99 and 0.98, respectively (Figure 17), obtained very satisfactory performance. The superior performance of the algorithm trained by the Ministry of Tourism agreements is due to the better quality in balancing the training sequence among all Ministries, corroborating the findings of the research conducted by Bao et. al (2020) and Zhang (2019a). In turn, the superior performance of the algorithms trained by the agreements of the Ministries of Agriculture and Cities are justified by the volume of the training sequence, as they are the two granting bodies with the largest amount of accountability analyzed, 24.74% and 17.95% respectively, in line with what Cecchini et al. (2010) and Larcker and Zakolyukina (2012). On the other hand, the

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

almost ideal performance of the Ministry of Sport's algorithm, with AUC assuming the value of 0.99, has no basis in volumetry or in the balance of the training sequence, since this Ministry is only the fifth in volumetry, with 7 , 49% of the total, and the amount of accountability rejected, 29, is well below the average of 137. This uniqueness is explained by Alpaydin (2020), Breiman (2001) and Domingos (2012) when some characteristic of the sequence of training allows unambiguous identification of the instances to be classified, causing false positives and false negatives to be reduced to almost zero.

Furthermore, Figure 17 illustrates the performance of the algorithms trained by the agreements of the Ministries of Citizenship, Defense and Health, whose AUC metrics assumed 0.91, 0.91 and 0.92, respectively. As emphasized by Fawcett (2006), his performance is satisfactory due to the AUC being greater than 0.9, but the empirical data obtained in this research demonstrate that his performance did not exceed that obtained in the algorithm trained by the universe (AUC = 0.96) . In contrast, the algorithms of Funasa and the Ministry of Justice achieved very unsatisfactory performance, since their AUC were 0.63 and 0.77, respectively. These results support the conclusions of Bao et al. (2020) and Zhang (2019a), as the volume of the agreements of these two bodies are the two lowest among all bodies (1.91%, Funasa, 1.66%, Ministry of Justice) and the balance is very unbalanced (16 failures for Funasa and 7 for the Ministry of Justice). In turn, there was no possibility of producing an algorithm for the Secretariat for Agrarian Development, as this agency never failed to render accounts. Consequently, there are no metrics for measuring performance for this body.

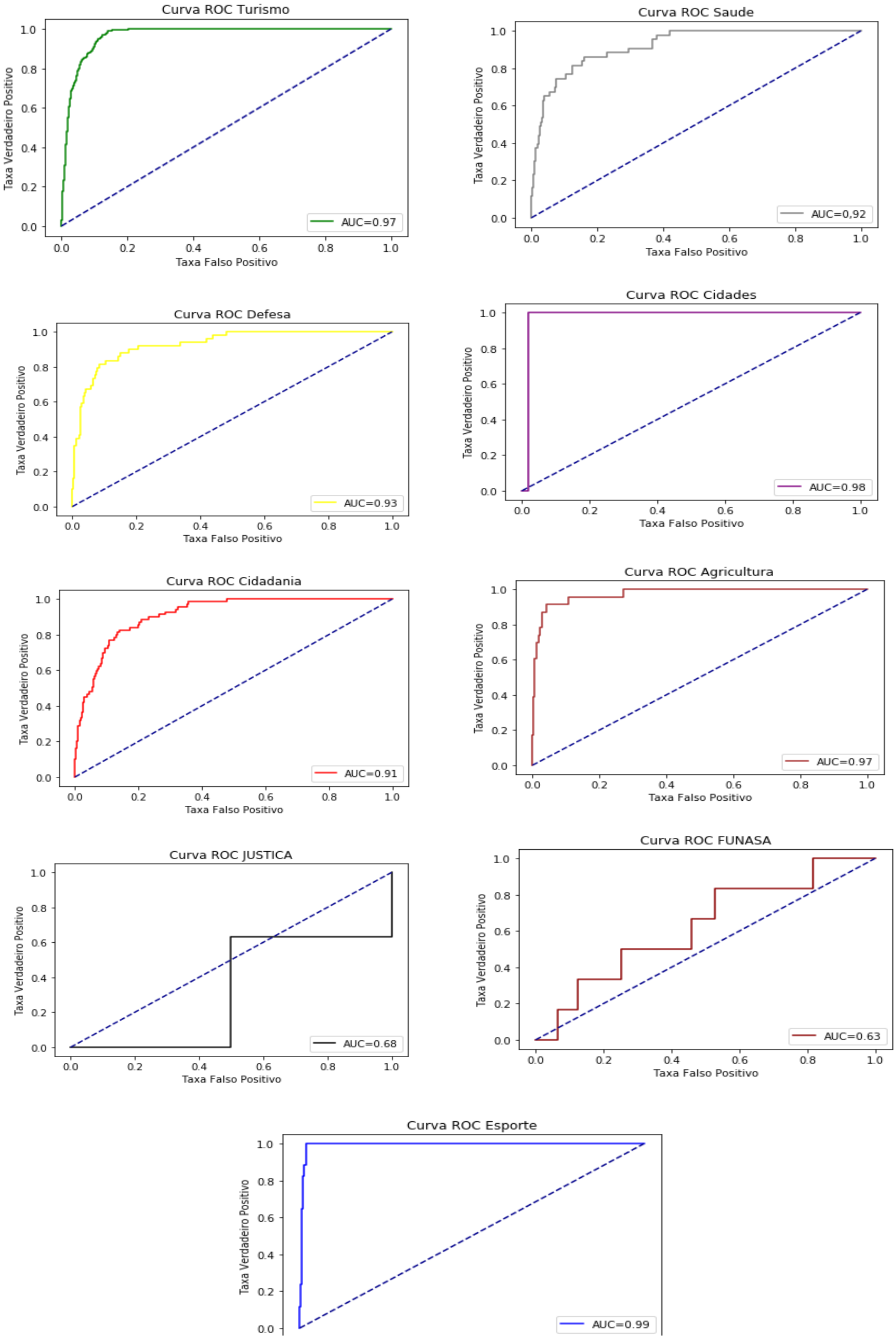

Figure 16 - AUC metric. ROC curve of the trained algorithms according to the identity of the granting agency.

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Table 9 allows to infer that, under the perspective of the metric inaccuracy (ε), the AI algorithms trained from the covenants of each granting agency separately perform much lower than the algorithm trained with the complete population of covenants.

Training Sequence	**Accumulated *Score* Interval**						
	[0.0 to 0.4)	**[0.0 to 0.5)**	**[0.0 to 0.6)**	**[0.0 to 0.7)**	**[0.0 to 0.8)**	**[0.0 to 0.9)**	**[0.0 to 1.0]**
All Bodies	0.000%	0.067%	0.738%	1.332%	3.557%	9.463%	100,000%
Ministry of Tourism	5.570%	12.953%	19.329%	27.181%	38.054%	46.107%	100,000%
Ministry of Citizenship	37.785%	49.060%	59.463%	70.537%	82.013%	91.477%	100,000%
defense Ministry	85.772%	89.933%	91.477%	92.349%	92.819%	93.826%	100,000%
Ministry of Health	90.470%	94.362%	95.302%	95.570%	95.705%	95.839%	100,000%
Ministry of Agriculture	17.114%	34.698%	53.020%	73.826%	90.537%	97.315%	100,000%
Sport Ministery	14.094%	24.430%	50.134%	74.832%	92.349%	98.054%	100,000%
Funasa	99.262%	99.329%	99.329%	99.329%	99.329%	99.329%	100,000%
Justice ministry	99.530%	99.597%	99.597%	99.664%	99.664%	99.664%	100,000%
Ministry of Cities	99.933%	99.933%	99.933%	99.933%	99.933%	99.933%	100,000%
Agricultural Development	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 8 – Metric inaccuracy (ε) of s Algorithm s IA trained s from each agreements granting body. (False Positive Probability).

This finding is evidenced by means of Figure 18, below. The analysis to be made resides in the comparison of the metric inaccuracy (ε) between (a) the algorithm trained from the complete population, curve “All organs”, and (b) the algorithms trained from the agreements of each granting organ. Figure 18 illustrates the behavior of the metric inaccuracy (ε) among the 11 trained algorithms, an opportunity in which it is possible to observe the number of agreements categorized in each accumulated score interval between [0 to 1]. Thus, the algorithm that manages to concentrate a greater number of agreements in accumulated intervals closer to 1 has better performance. It is inferred, therefore, that the algorithm represented by the curve “All organs” is the one with the best metric inaccuracy (ε).

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

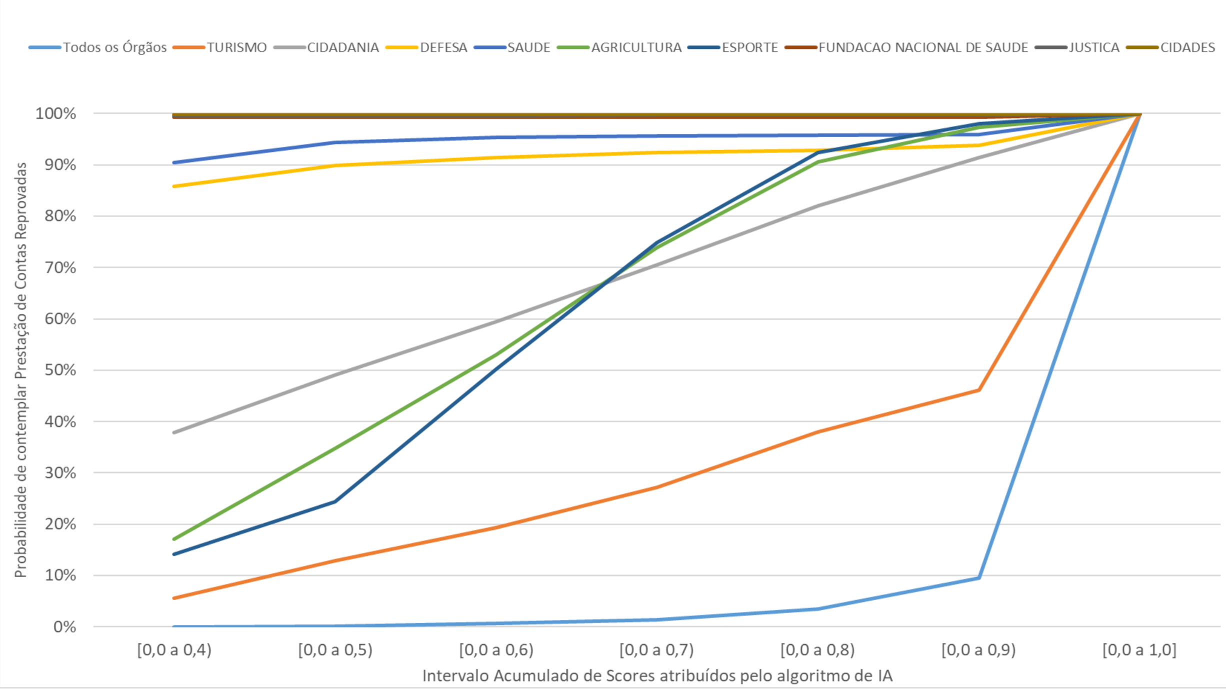

Figura 17 – Curvas da Métrica Inexatidão (ε) dos Algoritmos de IA treinados a partir dos convênios de cada órgão concedente.

Additionally, it is observed, from the point of view of the metric inaccuracy (ε), that the algorithm trained by the Ministry of Tourism agreements, curve "Tourism", performs better than the other algorithms, however without surpassing that trained with the population complete. This is due to the fact that the Ministry of Tourism has the longest history of disapproval of covenant accounts among all agencies, 779 covenants disapproved between 2008 and 2017, making the training sequence the least unbalanced (Bao et al. 2020; Zhang 2019a). The algorithms stamped on the “Cities” and “Agrarian Development” curves endorse this argument. The training sequences of these algorithms have, respectively, 2 and 0 agreements with rejected accountability, providing training sequences with intense imbalance and poor performance.

However, it is emphasized that a balanced training sequence is a necessary condition, but not enough to provide good performance. It is observed that the performance of the curve “Tourism” stands out due to the moderate imbalance of the training sequence coupled with its high volume, since the relative number of agreements with accountability concluded at the Ministry of Tourism is 13.19% of the total (Table 6), making it the third largest. This reinforces

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

the claims of Alpaydin (2020), Breiman (2001) and Domingos (2012) that longer training sequences provide better performance.

The ambiguity between volumetry and balance of the training sequence is portrayed in the "Agriculture", "Sport" and "Citizenship" curves. Although the Ministry of Agriculture has the largest number of agreements signed among all agencies (24.74% of the total), this was not enough to generate an algorithm with good performance. The Ministry of Agriculture lacks balance in its training sequence, since it has only 46 agreements with rejected accounts. The same situation occurs with the Ministry of Sport, which has a large relative volume (7.49% of the total) of the training sequence, but also a severe imbalance because it only failed 29 agreements.

On the other hand, there are some situations in which the algorithms have low performance, although they have been generated with relatively balanced training sequences. In these situations, the low performance of the metric inaccuracy (ε) is caused by the low volumetry of the training sequence. The algorithm represented by the "Defense" curve was generated with a sequence of 134 failed agreements, providing a relatively high balance, but the volume represents only 2.79% of the total.

In addition, there are those algorithms with very degraded performance because their training sequences combine nuances of low volume and high imbalance. The algorithms represented by the "Funasa", "Saude", "Justica" curves fit into this context. In turn, there was no possibility of producing an algorithm for the Secretariat for Agrarian Development, as this agency never failed to render accounts.

5.1.2. AUC metrics and inaccuracy (ε) according to the volume of the training sequence.

The evaluation of the performance of the accuracy of the trained algorithms by gradually increasing the training sequence, from the perspective of the AUC metric, allows us to conclude that the performance improves as the volume of data for training increases. However, the improvement reaches a level where any increase in the size of the training sequence does not

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

imply a performance improvement. Likewise, the evaluation, from the point of view of the inaccuracy metric (ε), allows to infer the same conclusions obtained under the prism of the AUC metric: the more volumetry, the better the performance, and there is a limit in which the increase in volumetry produces no improvement in performance. Therefore, the H3 of the Data training size construct is confirmed, as well as the H4, which deals with the saturation point, although the saturation in the AUC metric (year 2012) was slightly different from the inaccuracy metric (ε) (year 2014).

From the perspective of the AUC metric, as illustrated in Figure 19, the performances of the trained algorithms from the gradual increase in the size of the training sequence show continuous improvement behavior until they reach the same level of performance as the algorithm trained by the universe, the latter agreed upon. as benchmarking, whose AUC assumed a value of 0.96. This borderline level was obtained with the algorithm trained with the agreements signed up to 2012, as the AUC of this instance assumed the value of 0.96. As of the 2012 algorithm, all subsequent ones obtained the same performance, that is, a saturation point occurred when the training sequence reached approximately 72% of the population size (Table 7). This finding is in line with what Alpaydin (2020) and Zhang (2019a) claim about the ideal size of 75% of the total population to be used as a sample for training the algorithm.

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

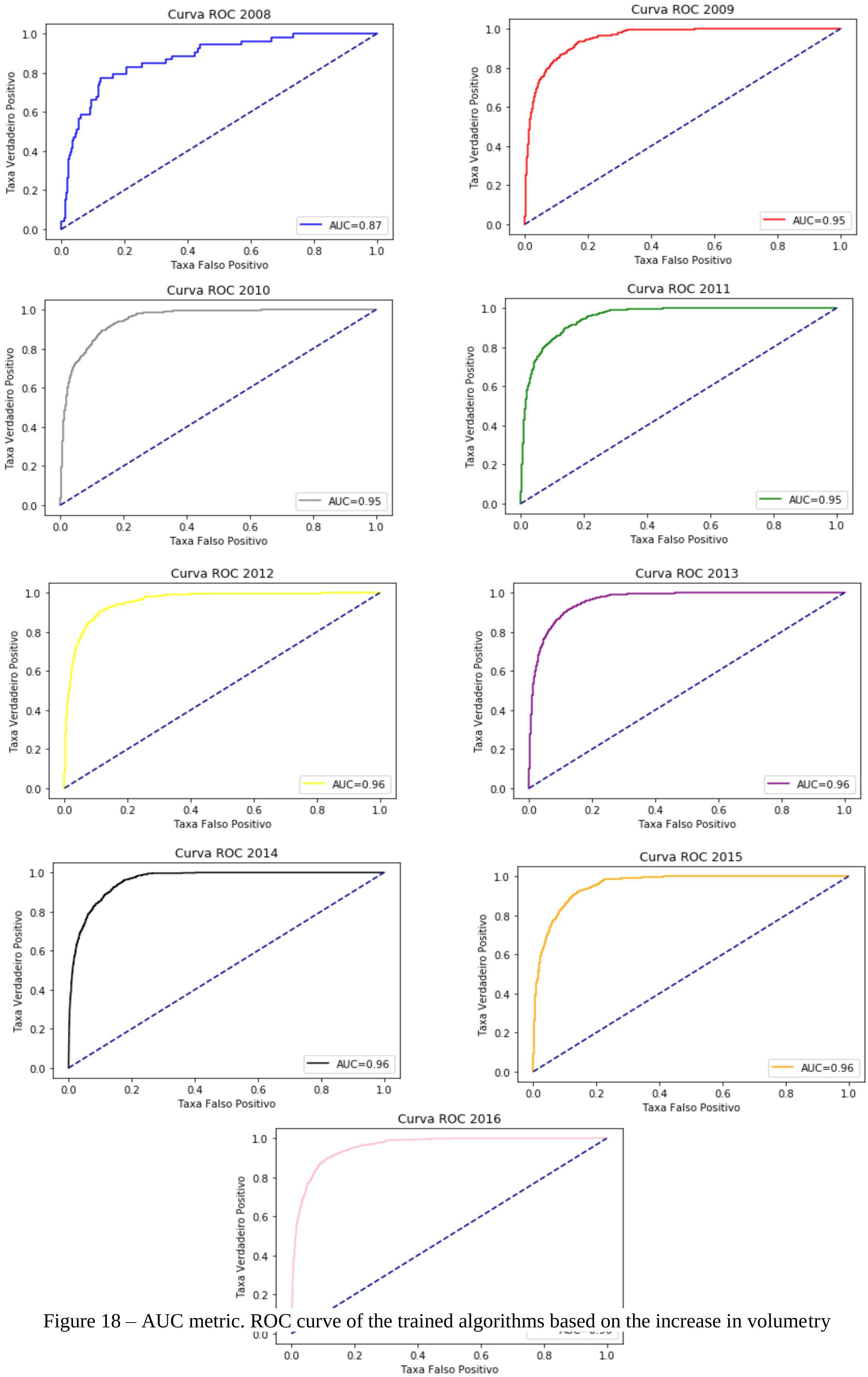

Figure 18 – AUC metric. ROC curve of the trained algorithms based on the increase in volumetry

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

On the other hand, Table 10 shows the performance from the perspective of the metric inaccuracy (ε) through the variation of the volumetry of the training sequence, incremented sequentially every year from 2008. The observation of this table allows to infer that the algorithm trained with the agreements signed in 2008 have a very unsatisfactory performance from the perspective of inaccuracy (ε).

Year Range	Intervalo Acumulado de *Score*						
	[0.0 to 0.4)	**[0.0 to 0.5)**	**[0.0 to 0.6)**	**[0.0 to 0.7)**	**[0.0 to 0.8)**	**[0.0 to 0.9)**	**[0.0 to 1.0]**
2008	85,55%	87,05%	88,26%	89,61%	90,68%	91,74%	100,00%
2009	0,93%	3,56%	12,53%	24,98%	36,01%	45,34%	100,00%
2010	0,07%	0,93%	2,21%	8,75%	17,94%	22,28%	100,00%
2011	0,00%	0,93%	3,56%	9,61%	11,74%	17,37%	100,00%
2012	0,00%	0,21%	1,14%	4,27%	9,82%	16,73%	100,00%
2013	0,00%	0,14%	1,28%	2,57%	5,49%	11,90%	100,00%
2014	0,00%	0,14%	0,36%	1,85%	4,13%	9,54%	100,00%
2015	0,00%	0,14%	0,57%	1,71%	4,20%	11,32%	100,00%
2016	0,00%	0,21%	0,57%	1,28%	4,13%	9,40%	100,00%

Table 9 – Inaccuracy (ε) metric of AI Algorithms trained from separate agreements by year of celebration . (False Positive Probability).

Then, the algorithm trained with the agreements signed until 2009 produced a better model than the previous one, but still unsatisfactory because the metric inaccuracy (ε) remains very degraded. Ongoing, it appears that the algorithm trained by the celebrated agreements until 2010 is worse than that of 2011, which in turn is worse than that of 2012, which in turn is worse than that of 2013. Consequently, this inference endorses hypothesis raised that the performance of the machine learning algorithm increases as there is more data in the learning universe, which is in line with recent literature (Bao et al. 2020; Cecchini et al. 2010; Larcker and Zakolyukina, 2012 ; Hooda, Bawa and Rana 2020). However, it is worth noting the fact that the 2010 algorithm is better than the 2011 algorithm up to the accumulated interval [0 to 0.6), although this algorithm was generated with a higher training sequence volume than that, the which in a way contradicts what Alpaydin (2020) asserts. However, Breiman (2001) asserts that those situations that escape the more volumetric and better performance dichotomy can occur when

TESOURO NACIONAL

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

some characteristic of the training sequence causes the algorithm to identify the classes explicitly, in this case disapproved and approved agreements.

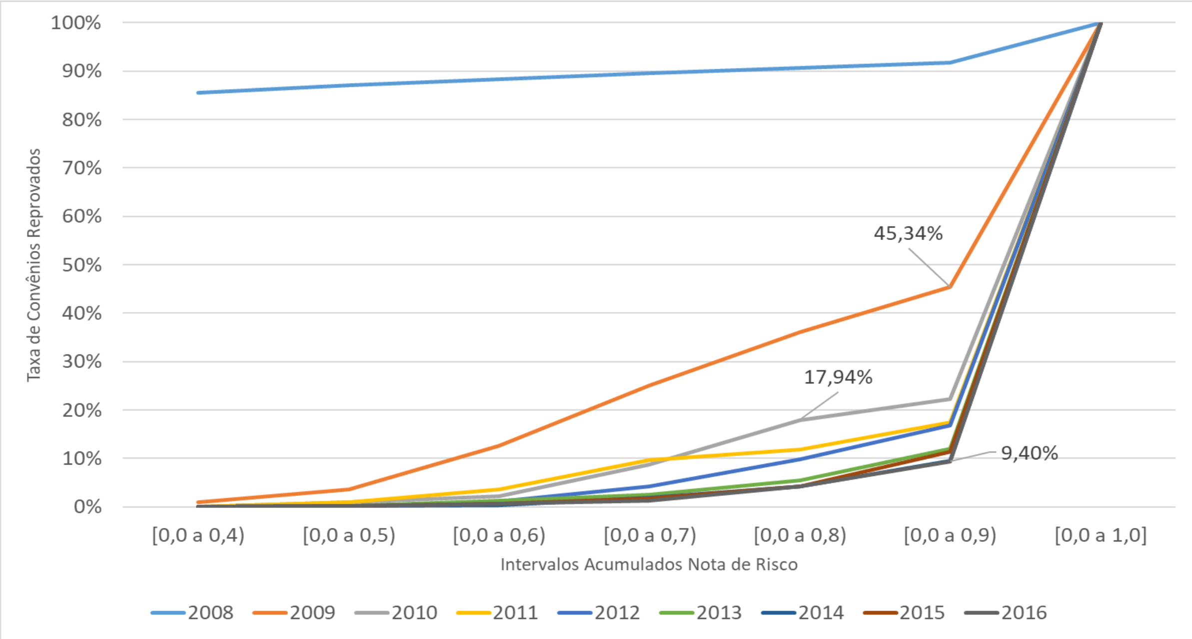

Figure 19 - Inaccuracy Metric Curve (ε) of AI Algorithms trained from separate agreements by year of celebration. (False Positive Probability).

Furthermore, it is inferred that the performance improvement of the algorithms trained with incremental volumetry each year, has stagnated from the algorithm trained with the agreements concluded until the year 2013. Figure 20, generated from the data in Table 10 , show a timid improvement in the metric inaccuracy (ε) of the algorithms trained since 2014, almost insignificant.

This fact allows us to answer hypothesis 4, to the extent that there is a saturation point and that it was possible to establish the size of the training sequence for machine learning in which the increase in volumetry is no longer significant for improving performance. In turn, the occurrence of a saturation point in supervised machine learning is corroborated by Alpaydin (2020) and Breiman (2001), Zhang (2019a). Additionally, the algorithms trained with saturated volumetry, that is, from 2013 onwards, establish an optimal threshold of false positive rates from the score 0.8. Figure 20 allows us to state that the curves of these algorithms have an inflection point precisely from the score 0.8, since the probability of occurrence of false positives remains below 10% up to that threshold, increasing dramatically after this limit.

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

6. CONCLUSIONS

The life cycle of the transfer of discretionary resources ends with its rendering of accounts and consequent analysis by the transferring body, which opines for the approval or rejection of the accounts. Accountability analysis is a lengthy process and encourages the use of resources for its realization, in addition to trained public servants (Meireles 2019; Ferreira and Bugarin 2008). In turn, the "Malha Fina de Convênios" system presents a quick and rational alternative for the analysis of accountability, configuring itself in innovation.

The objective of this research was, from the comparison between results generated by Artificial Intelligence algorithms and the results of the provision of covenants accounts, to identify how the identity of the transferring body of resources and the size of the training sequences are determinants in the precision classification of the machine learning algorithm. The result of the analysis of the rendering of accounts of an agreement can be classified as approved or rejected. Thus, it was possible to clarify the circumstances in which the adoption of the "Malha Fina de Convênios" is more assertive through the constraints Identity of the resource transferor (H1 and H2) and Size of data training (H3 and H4), dismissing objections against its implementation and validating the method of this innovative approach.

In Random Forest supervised machine learning, algorithms need to be trained first on the training data set and then tested on the test data set. The general rule in the division between training set and test is to choose about 75% of the sample as training data and the rest as test data (Alpaydin, 2020). However, this is not an established rule, since the guarantee of machine learning performance depends on the availability of data labeling (Issa, Sun and Vasarhelyi 2016; Sun 2019) and the balancing of labels within the training sequence (Bao et al. 2020; Zhang 2019a). In view of these precepts, the tests of this research were carried out in two parts, contemplating the intrinsic properties of each transferring organ and the reliability of learning according to the aggregation of data in the training sequence.

First, the precision performance of the algorithms was investigated through the parameterization of 10 training sequences according to the identity of the granting agency. It has been empirically demonstrated that, under certain circumstances, some algorithms trained

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

exclusively with covenants from a given granting body perform better than the algorithm trained with the entire population of the covenants (Bao et al. 2020; Cecchini et al. 2010; Tiwari and Hooda 2018). In turn, these circumstances materialize under the metric to be used to assess performance, AUC or inaccuracy (ε). While the AUC metric points to 4 algorithms with superior performance in relation to the one trained with the total population of covenants, the metric inaccuracy (ε) does not indicate any algorithm trained only with covenants of a certain organ with superior performance in relation to benchmarking. However, the superior performance of the 4 algorithms from the point of view of the AUC metric is residual, almost insignificant, and under the burden of resulting in 6 algorithms with inferior performance in the perspective of the metric inaccuracy (ε), being 2 of these very unsatisfactory. Therefore, the findings suggest that the adoption of the strategy of training algorithms separately according to the resource transfer agency is not advantageous.

Then, the precision behavior of the algorithms was investigated by increasing the size of the training sequence for machine learning. Thus, the possible occurrence of discrepancy in the 9 distinct algorithms generated was evaluated. It was found that as the size of the training sequence grows, the precision of the produced algorithm increases. Furthermore, the diagnosis of the results obtained indicates that the adoption of the predictive model to estimate the results of an agreement will not be problematic if the size of the training sequence is greater than or equal to the saturation point (Breiman 2001; Domingos 2012). In this research, two slightly different saturation points were observed for the metrics AUC and inaccuracy (ε). On the one hand, the AUC metric indicates that saturation occurs with a training sequence with agreements signed up to 2012, consisting of 72.18% of the total (Figure 19 and Table 7), while the metric inaccuracy (ε) indicates the composite sequence up to the year 2013, consisting of 84.77% of the total (Figure 20 and Table 7).

Indeed, the findings of this research validate the current methodology of the “Malha Fina de Convênios”, since it contains satisfactory results based on empirical tests of the predictive model. Malha Fina presents itself as a quick and rational alternative for accountability analysis. For Barzelay (1997), Bresser-Pereira (1998, 2008), Fukuyama (2004) and Power (1997, 2003a, 2003b), decision-making in a bureaucratic organization must be guided by the transaction cost of the processes. In this line, this system rationalizes the use of the workforce in the analysis of

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

rendering of accounts by Organs granting agencies through the adoption of a risk appetite threshold in which the probable agreements with rejectable accounts would be inadvertently approved (false positives). The present study shows an inflection in this risk threshold at 0.8 (Figure 16).

Furthermore, future research may advance the applicability of the “Malha Fina de Convênios” methodology in two ways. The first aspect consists of using other metrics, in addition to AUC and inaccuracy (ε), to evaluate the performance of the Artificial Intelligence algorithm. The second consists in validating the implementation of the “Malha Fina de Convênios” methodology in other modalities of transfers of funds from the Union to subnational entities that are not operationalized in the Platform + Brazil. This field of research proves to be fertile, as voluntary transfers are quantitatively lower than the financial volume that involves mandatory transfers (Amorim Neto and Simonassi 2013). Data extracted from SIAFI in May 2020 show that, of the total resources transferred by the Union, only about 3% correspond to voluntary transfers.

In due course, future research will be able to take advantage of the limitations faced in this research, adopting a strategy to deal with the low availability of data labeled in classes in a balanced way (Alpaydin 2020; Sun 2019), in this case, approved agreement account installments. or rejected. The other major limitation was the configuration of infrastructure with large computational capacity to run the simulations.

Like any discipline, there is a lot of "popular wisdom" around Artificial Intelligence and how its use can generate practical benefits. This research demonstrated that it is possible to extract results within the rational-legal Weberian framework of the Brazilian Public Administration through tools of Computer Science (Newell and Simon 1976). This goes back to a Renaissance situation that combines technology, cognitive science and human need to produce something that the world did not know was missing, constituting a new field of knowledge.

Voluntary transfers from the union do not appear as an isolated public policy, much less an end in themselves, since its primary purpose is to enable, as a cooperation, the financing of public services in subnational entities. It becomes essential that the equity criteria overlap with the

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

political objectives in the transfers resulting from voluntary transfers, as this resource would contribute to the reduction of interregional inequalities in the Brazilian federation, as pointed out by Abrucio (2005), Arretche (2010) and Fajardo (2016). Thus, the control and inspection of the resources transferred must be timely and agile, concomitant with the implementation of public policy, in order to guarantee equity and accountability (Schedler 1999).

The “Malha Fina de Convênios”, presents itself as a powerful instrument to modify the paradigm in which subnational entities with political alignment with the Union representative, or with greater representation in the National Congress, receive more resources, regardless of other criteria (Amorim Neto and Simonassi 2013; Brollo and Nannicini 2012; Ferreira and Bugarin 2008; Limongi and Figueiredo 2005; Meireles 2019; Soares and Melo 2016).

In line with the results already presented in the private sector, the spread of methods applied to risk control and mitigation in the Public Administration has a great chance of positively impacting management (Al-Qudah, Baniahmad and Al-Fawaerah 2013; Huang and Vasarhelyi 2019). The digitization of public services, with the intensive use of Artificial Intelligence, may become a current practice in Public Administration as a means to face the increasing demands of increasing efficiency, allowing qualified employees to handle more sophisticated tasks (Kim, Mannino and Nieschwietz 2009). It is noteworthy that the results of this research may be incorporated into the work process of the Union's voluntary transfers to subnational entities, including the possibility of normative changes that govern the process of the Union's voluntary transfers, in addition to removing fear in adopting the “Malha Fina de Convênios”. In this sense lies the main collaboration of this work.

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

REFERENCES

Abrucio, Fernando Luiz. 2005. "Reforma do Estado no federalismo brasileiro: a situação das administrações públicas estaduais". *Revista de Administração Pública*, v. 39, n.2, (2005): p. 401-420.

Abrucio, Fernando Luiz, e Cibele Franzese. "Federalismo e políticas públicas: o impacto das relações intergovernamentais no Brasil". In *Tópicos da Economia Paulista para Gestores Públicos,* editado por Araújo e Beira, v. 1, p. 13-31. São Paulo: Fundap, 2007.

ALPAYDIN, Ethem. *Introduction to machine learning*. MIT press, 2020.

Afonso, José Roberto, Eliane Cristina Araújo, e Bernardo Guelber Fajardo. "The Role of Fiscal and Monetary Policies in the Brazilian Economy: Understanding Recent Institutional Reforms and Economic Changes." *Quarterly Review of Economics and Finance* (2016): 41–55. https://doi.org/10.1016/j.qref.2016.07.005.

Al-qudah, Anas Ali, Ahmad Yahiya Baniahmad, e Nawwaf Al-fawaerah. "The Impact of Information Technology on the Auditing Profession." *Management and Administrative Sciences Review* (2013) 430: 423.

Alves, Maria Fernanda Colaço Alves, e Paulo Carlos Du Pin Calmon. "Múltiplas Chibatas? – Governança Da Política de Controle Da Gestão Pública Federal." *ENAPG - Encontro de Administração Pública e Governança*, (2008): 1–16.

Amorim Neto, Octavio. "Gabinetes presidenciais, ciclos eleitorais e disciplina legislativa no Brasil". *Dados*, v. 43, n. 3, (2000): p. 479-519.

Amorim Neto, Octavio, e Andrei Gomes Simonassi. "Bases Políticas Das Transferências Intergovernamentais No Brasil (1985-2004)." *Revista de Economia Política* (2013): 704–25.

Arretche, Marta. "Federalismo e Igualdade Territorial: Uma Contradição em Termos?" *Dados*

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

(2010): 587–620. https://doi.org/10.1590/S0011-52582010000300003.

Bao, Yang, Bin Ke, Bin Li, Y. Julia Yu, e Jie Zhang. 2020. "Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach." *Journal of Accounting Research*, (2020): 199–235. https://doi.org/10.1111/1475-679X.12292.

Barzelay, Michael. "Central Audit Institutions and Performance Auditing: A Comparative Analysis of Organizational Strategies in the OECD". *Governance: An International Journal of Policy, Administration, and Institutions*, (1997): 235-260.

Barzelay, Michael. *Public Management as a Design-Oriented Professional Discipline*. Cheltenham, Reino Unido. Edward Elgar Publishing Limited The Lypiatts. 2019.

Barzelay, Michael. 2018. "The Study of Public Management : Conceptualizing a Design-Oriented Social Science." *Revista Do Serviço Público* (2018): 33–56.

Barzelay, Michael, e Raquel Gallego. "From 'New Institutionalism' to 'Institutional Processualism': Advancing Knowledge about Public Management Policy Change." *Governance* (2006): 531–57. https://doi.org/10.1111/j.1468-0491.2006.00339.x.

Brasil. Controladoria-Geral da União, Secretaria Federal de Controle Interno. *Instrução Normativa nº 03, de 09.06.2017*. Aprova o Referencial Técnico da Atividade de Auditoria Interna Governamental do Poder Executivo Federal. 2017a. Disponível em https://www.cgu.gov.br/sobre/legislacao/arquivos/instrucoes-normativas/in_cgu_03_2017.pdf. Acesso em 25 mai. 2020.

Brasil. Controladoria-Geral da União, Secretaria Federal de Controle Interno. *Portaria nº 378, de 07.02.2017*. Constitui Grupo de Trabalho para avaliar as transferências voluntárias da união. 2017b.

Brasil. Controladoria-Geral da União, Secretaria Federal de Controle Interno. *Relatório de Avaliação da Gestão das Transferências Voluntárias da União*. 2018a. Disponível em

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

https://auditoria.cgu.gov.br/download/11014.pdf. Acesso em 25 mai. 2020.

Brasil. Controladoria-Geral da União e Ministério da Economia. *Instrução Normativa Interministerial nº 1, de 14 de fevereiro de 2019.* Estabelece diretrizes e parâmetros para a análise de prestação de contas simplificada. 2019a. Disponível em http://portal.convenios.gov.br/analise-informatizada. Acesso em 25 mai. 2020.

Brasil. Controladoria-Geral da União e Ministério do Planejamento, Desenvolvimento e Gestão. *Instrução Normativa Interministerial nº 5, de 6 de novembro de 2018*. Estabelece diretrizes e parâmetros para a análise de prestação de contas simplificada. 2018b. Disponível em http://portal.convenios.gov.br/analise-informatizada. Acesso em 25 mai. 2020.

Brasil. Ministério da Economia. *Plataforma +Brasil*. 2019b. Disponível em http://plataformamaisbrasil.gov.br/. Acesso em 25 mai. 2020.

Brasil. Tribunal de Contas da União. *Transferências Governamentais Constitucionais*. 2008. Disponível em: <https://portal.tcu.gov.br/>. Acesso em: 25 mai. 2020.

Breiman, Leo.. “Random Forests.” *Machine Learning*, (2001): 1–122. https://doi.org/10.1201/9780367816377-11.

Bresser Pereira, Luiz Carlos. “Da administração burocrática à gerencial”. *Revista do Serviço Público*, v. 47, n. 1, (1996).

Bresser Pereira, Luiz Carlos. “A reforma do estado dos anos 90: lógica e mecanismos de controle”. *Lua Nova: Revista de Cultura e Política*, (1998): 49-95.

Bresser-Pereira, Luiz Carlos. 2008. “O Modelo Estrutural de Gerência Pública.” *Revista de Administração Pública*, (2008): 391–410. https://doi.org/10.1590/s0034-76122008000200009.

Brollo, Fernanda, e Tommaso Nannicini. “Tying Your Enemy ’ s Hands in Close Races : The

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Politics of Federal Transfers in Brazil Tying Your Enemy ' s Hands in Close Races : The Politics of Federal Transfers in Brazil". *American Political Science Review*, (2012): 742–761. https://doi.org/10.1017/S0003055412000433.

Brown-Liburd, Helen, e Miklos A. Vasarhelyi. "Big Data and Audit Evidence." *Journal of Emerging Technologies in Accounting* (2015): 1–16. https://doi.org/10.2308/jeta-10468.

Campana, Priscilla de Souza Pestana. "A Cultura Do Medo Na Administração Pública e a Ineficiência Gerada Pelo Atual Sistema de Controle." *Revista de Direito*, (2017): 189–216.

Cecchini, Mark, Haldun Aytug, Gary J. Koehler, e Praveen Pathak. "Detecting Management Fraud in Public Companies." *Management Science*, (2010): 1146–1160. https://doi.org/10.1287/mnsc.1100.1174.

Committee Of Sponsoring Organizations Of The Treadway Commission – COSO. *Controle Interno – Estrutura Integrada. Sumário Executivo*. Traduzido por PwC. Instituto dos Auditores Internos do Brasil, 2013.

Creswell, John W. "Projeto de Pesquisa: métodos qualitativo, quantitativo e Misto**"**. Porto Alegre: Artmed, 2007.

Dallaverde, Alexsandra Katia. "As transferências voluntárias no modelo constitucional brasileiro". São Paulo: Blucher, 2016.

Domingos, Pedro.. "A Few Useful Things to Know About Machine Learning." *Communications of the ACM*, (2012): 9–48. https://doi.org/10.1007/978-3-642-35289-8-3.

Dunleavy, Patrick, Helen Margetts, Simon Bastow, e Jane Tinkler. 2006. "New Public Management Is Dead - Long Live Digital-Era Governance." *Journal of Public Administration Research and Theory*, (2006): 467–94. https://doi.org/10.1093/jopart/mui057.

Fajardo, Bernardo Guelber. "Vieses orçamentários em entes subnacionais: uma análise sob a ótica da estimação das receitas estaduais." Tese de Doutorado em Administração Pública,

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Fundação Getúlio Vargas, Rio de Janeiro, Brasil, 2016.

Fayol, Henry. “General and industrial management”. London, Pitman, 1949.

Fawcett, Tom.. “An Introduction to ROC Analysis.” *Pattern Recognition Letters,* (2006): 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.

Ferreira, Ivan F. S., e Mauricio S. Bugarin. “Transferências Voluntárias e Ciclo Político-Orçamentário No Federalismo Fiscal Brasileiro.” *Revista Brasileira de Economia*, (2018): 271–300. https://doi.org/10.1590/s0034-71402007000300001.

Fukuyama, Francis. “*A Construção de Estados. Governo e Organização no Século XXI*”. Rio de Janeiro, Rocco, 2004.

Grover, Dhruv, Sebastian Bauhoff, e Jed Friedman. “Using Supervised Learning to Select Audit Targets in Performance-Based Financing in Health: An Example from Zambia.” *PLoS ONE*, (2019): 1–13. https://doi.org/10.1371/journal.pone.0211262.

Heber, Florence. “O Diálogo Entre Burocracia e Inovação.” *ANPAD*, (2014) 1–10.

Hood, Christopher, Oliver James e Colin Scott. “Regulation of government: Has it increased, is it increasing, should it be diminished?” *Public Administration*, v. 78, n. 2, (2000): p. 283–304.

Hooda, Nishtha, Seema Bawa, e Prashant Singh Rana. “Optimizing Fraudulent Firm Prediction Using Ensemble Machine Learning: A Case Study of an External Audit.” *Applied Artificial Intelligence*, (2020): p. 20–30. https://doi.org/10.1080/08839514.2019.1680182.

Huang, Feiqi, e Miklos A. Vasarhelyi. “Applying Robotic Process Automation (RPA) in Auditing: A Framework.” *International Journal of Accounting Information Systems* (2019): 100433. https://doi.org/10.1016/j.accinf.2019.100433.

Instituto Brasileiro de Geografia Estatística – IBGE – (2020). *Código dos Municípios IBGE*.

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Disponível em https://www.ibge.gov.br/explica/codigos-dos-municipios.php. Acesso em: 12 abr. 2020

Issa, Hussein, Ting Sun, e Miklos A. Vasarhelyi. "Research Ideas for Artificial Intelligence in Auditing: The Formalization of Audit and Workforce Supplementation." *Journal of Emerging Technologies in Accounting*, (2016): 1–20. https://doi.org/10.2308/jeta-10511.

Jackson, Bill. "Designing Projects and Project Evaluations Using the Logical Framework Approach." *Monitoring and Evaluation Iniciative: IUCN*, (1997): 1–11.

Kim, Hyo Jeong, Michael Mannino, e Robert J. Nieschwietz. "Information Technology Acceptance in the Internal Audit Profession: Impact of Technology Features and Complexity." *International Journal of Accounting Information Systems*, (2009): 214–28. https://doi.org/10.1016/j.accinf.2009.09.001.

Kohavi, Ron. "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." *International Joint Conference on Artificial Intelligence*, (1995): 1137–1145. https://doi.org/10.1067/mod.2000.109032.

Larcker, David F, e Anastasia A. Zakolyukina. "Detecting Deceptive Discussions in Conference Calls." *Journal of Accounting Research*, (2012): 495–540.

Limongi, Fernando, e Argelina Figueiredo. "Processo Orçamentário e Comportamento Legislativo: Emendas Individuais, Apoio ao Executivo e Programas de Governo." *Dados*, (2005): 737–76. https://doi.org/10.1590/s0011-52582005000400002.

Loureiro, Maria, e Fernando Abrucio. 2004. "Política e Reformas Fiscais No Brasil Recente." *Brazilian Journal of Political Economy*, v. 24, (2004): 50–72.

Maciejewski, Mariusz. "To do more, better, faster and more cheaply: using big data in public administration". *International Review of Administrative Sciences*, (2017): 120–135.

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Margetts, Helen, e Patrick Dunleavy. "Better Public Services through E-Government: Academic Article in Support of Better Public Services through e-Government". *Report by the Comptroller and Auditor General HC 704-III Session*, (2002). April: 1–21.

Meireles, Fernando. "Alinhamento Partidário e Demanda Por Transferências Federais No Brasil." *Revista de Administração Pública*, (2019): 173–94. https://doi.org/10.1590/0034-761220170282.

Mintzberg, Henry. "*The structuring of organizations: A synthesis of the research*." Englewood Cliffs, N.J: Prentice-Hall, 1979

Monteiro, Renato Pereira. "Análise do sistema de controle interno no Brasil: objetivos, importância e barreiras para sua implantação". *Revista Contemporânea de Contabilidade*, (2015): 159-181.

Newell, Allen, e Hebert Simon. "Computer science as empirical inquiry: symbols and search." *Communications of the ACM* . (1976): 113–126.

Oliveira, Antonio. "Burocratas da linha de frente: executores e fazedores das políticas públicas". *Revista de Administração Pública*, (2012): 1551 - 1573.

Olivieri, Cecília. "Política e burocracia no Brasil: o controle sobre a execução das políticas públicas". Tese de Doutorado em Administração Pública e Governo, Fundação Getúlio Vargas, São Paulo, Brasil, 2008.

Olivieri, Cecília. "Os controles políticos sobre a burocracia". *Revista de Administração Pública*, (2011): 1395 – 1424.

Parker, Lee D., Kerry Jacobs, e Jana Schmitz. "New Public Management and the Rise of Public Sector Performance Audit." *Accounting, Auditing & Accountability Journal*, (2018): 280–306. https://doi.org/10.1108/aaaj-06-2017-2964.

Power, Michael.. "Evaluating the Audit Explosion." *Law and Policy*, 25, (2003a): 185–202. https://doi.org/10.1111/j.1467-9930.2003.00147.x.

Power, Michael . "The Audit Society - Second Thoughts." *International Journal of Auditing*, 4, (2003b): 111–119. https://doi.org/10.1111/1099-1123.00306.

Power, Michael. "The Risk Management of Nothing." *Accounting, Organizations and Society*, (2009): 849–855. https://doi.org/10.1016/j.aos.2009.06.001.

Ramos, Alberto Guerreiro. 1983. "Administração e contexto brasileiro: esboço de uma teoria geral da administração". Rio de Janeiro: FGV, 1983.

Schedler, Andreas. "Conceptualizing Accountability". In *The Self-Restraining State: Power and Accountability in New Democracies*, edited by Diamond, L. J.; Plattner, M. F. Boulder Colorado: Lynne Rienner Publishers, 1999.

Soares, Márcia Miranda, e Bruno Guimarães De Melo. "Condicionantes Políticos e Técnicos Das Transferências Voluntárias Da União Aos Municípios Brasileiros". *Revista de Administração Pública*, (2016): 539–61.

Stanisic, Nemanja, Tijana Radojevic, and Nenad Stanic."Predicting the Type of Auditor Opinion: Statistics, Machine Learning, or a Combination of the Two?" *The European Journal of Applied Economics*, (2019): 1–58. https://doi.org/10.5937/ejae16-21832.

Sun, Ting Sophia. "Applying Deep Learning to Audit Procedures: An Illustrative Framework." *Accounting Horizons* (2019): 89–109. https://doi.org/10.2308/acch-52455.

Tenorio, Fernando Guilherme. "Weber e a burocracia." *Revista de Serviço Público*, vol. 109, no 4, (1981): p. 79-88.

Tiwari, Arpit.; Hooda, Nishtha. "Machine Learning Framework for Audit Fraud Data Prediction". *ResearchGate*, v. 7, n. 6, (2018): 164–167.

TesouroNacional

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

Taylor, Frederick Winslow. "*The principles of scientific management*". New York: Harper & Brothers, 1911.

The Institute of Internal Auditors – IIA (2012). *International Standards for The Professional Practice of Internal Auditing (Standards*). Disponível em de https://na.theiia.org/standards-guidance/Public%20Documents/IPPF%202013%20English.pdf. Acesso em: 10 jan. 2020.

Trevisan, Andrei Pittol, e Hans Michael Van Bellen. "Avaliação de políticas públicas: uma revisão teórica de um campo em construção". *Revista de Administração Pública*, v. 42, n. 3, (2008), p. 529–550.

United States Government Accountability Office – GAO (2011). *Government Auditing Standards*. Disponível em http://gao.gov/assets/590/587281.pdf. Acesso em 25 abr. 2020.

Vasarhelyi, Miklos A., e Fern B. Halper. "The Continuous Audit of Online Systems." *Auditing: A Journal of Practice and Theory*. (1991): 110-125.

Vasarhelyi, Miklos A., Alexander Kogan, e Brad Tuttle. "Big data in accounting: An overview". *Accounting Horizons*, v. 29, n. 2, (2015): p. 381–396.

Wirtz, Bernd W., Jan C. Weyerer, e Carolin Geyer. 2019. "Artificial Intelligence and the Public Sector—Applications and Challenges." *International Journal of Public Administration* (2019): 596–615. ttps://doi.org/10.1080/01900692.2018.1498103.

Zhang, Abigail Chanyuan. "Predict Audit Quality Using Machine Learning Algorithms." *SSRN Electronic Journal*, (2019): 1–40. https://doi.org/10.2139/ssrn.3449848.

Zhang, Abigail Chanyuan. "Intelligent Process Automation in Audit." *Journal of Emerging Technologies in Accounting* (2019): 69–88. https://doi.org/10.2308/jeta-52653.

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

ANNEX I - PHYTON CODE USED IN SIMULATIONS

Code for training the algorithm Machine Learning:

```
import pyodbc
from math import floor
from sklearn.tree import DecisionTreeClassifier
import numpy as np
import functools
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.metrics import f1_score
from sklearn.metrics import recall_score
import matplotlib.mlab as mlab
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.ensemble import ExtraTreesClassifier
from sklearn.model_selection import cross_validate
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import cross_val_predict
import pickle
import importlib
# Connection to database to retrieve the covenants
conn = pyodbc.connect(
    r'DRIVER={ODBC Driver 13 for SQL Server};'
    r'SERVER=sdh-die-bd.cgu.local;'
    r'DATABASE=mdl_prestacao_siconv;'
    r'Trusted_Connection=yes;'
    #r'UID=;'
    #r'PWD=;'
    )
query = """
        SEQUENCIA DE TREINAMENTO BANCO DE DADOS PLATAFORMA +BRASIL
                PARAMETRIZADA CONFORME O TAMANHO DA SEQUENCIA DE TREINAMENTO E A
IDENTIDADE DO ORGAO CONCEDENTE
        """
data = pd.read_sql(query, conn)
conn.close()
columns = [column for column in data.columns if column not in ['classes_num', 'num_convenio']]
labels = data["classes_num"].values
features = data[list(columns)].values
sample_weight = np.random.RandomState(1234).rand(labels.shape[0])
```

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

```
f_train, f_test, c_train, c_test, sw_train, sw_test = train_test_split(features, labels, sample_weight, test_size=0.40, random_state=1234)
#Recurso Computacional - Esforço
forest = ExtraTreesClassifier(n_estimators=1000, max_depth=None, min_samples_split=0.02, random_state=1234, n_jobs=3, class_weight={0: 1, 1: 1000}, max_features='log2')
#A variavel et_pred é o modelo criado a partir da sequencia de treinamento da query anterior
et_pred = forest.fit(f_train, c_train, sample_weight=sw_train)
----
#Matrizes de Confusao para geracao das notas de predicao --> risco
# Matris 2 X 2 -> linha sequencia de treinamento /coluna nota atribuida pelo sistema
predictions = et_pred.predict(f_test)
print(f1_score(c_test, predictions, average=None))
confusion_matrix(c_test, predictions, labels=[0,1])
predictions_prob = et_pred.predict_proba(f_test)

# Connection to database to retrieve the covenants
conn = pyodbc.connect(
   r'DRIVER={ODBC Driver 13 for SQL Server};'
   r'SERVER=sdh-die-bd.cgu.local;'
   r'DATABASE=mdl_prestacao_siconv;'
   r'Trusted_Connection=yes;'
   #r'UID=Trusted_Connection;'
   #r'PWD=Trusted_Connection;'
   )
query = """
       SEQUENCIA DE TREINAMENTO BANCO DE DADOS PLATAFORMA +BRASIL
           """
data2 = pd.read_sql(query, conn)
conn.close()
----
columns2 = [column for column in data.columns if column not in ['classes_num', 'num_convenio']]
labels2 = data2["classes_num"].values
features2 = data2[list(columns)].values
----
#Matrizes de Confusao para geracao das notas de predicao --> risco
# A Variavel et_pred eh aquela que contem o modelo criado a partir do bloco da primeira query
predictions2 = et_pred.predict(features2)
print(f1_score(labels2, predictions2, average=None))
confusion_matrix(labels2, predictions2, labels=[0,1])
----
#Imprimindo as variaveis do teste feito no modelo
predictions_prob2 = et_pred.predict_proba(features2)
----
```

Public Finance Notebooks – Special Edition, Brasília, v. 02, n. 1, p. 1-80, apr. 2021

```
data2['scores0']=predictions_prob2[:,0]
data2['scores1']=predictions_prob2[:,1]
data2.to_csv('Modelo_Todos_Orgaos_resultado_para_todos_orgaos.csv')-
```