Working Papers

ISSN 2219-780X

4

NOVEMBER
2021

MINISTÉRIO DA
ECONOMIA

Receita Federal

CIAT

ARTIFICIAL INTELLIGENCE FOR EXPEDITED DECISION AND RISK ANALYSIS IN THE BRAZILIAN FEDERAL TAX ADMINISTRATIVE LITIGATION

Artificial Intelligence for Expedited Decision and Risk Analysis in the Brazilian Federal Tax Administrative Litigation

Tiago Monteiro Cardoso
Willian Darwin Junior
Patrick Moreira Nogali
Sérgio Adriano Mazzetti
Jorge Eduardo de Schoucair Jambeiro Filho
Cláudia Maria de Andrade
Verônica Maria Perrotta de Seixas
Daniela Spinola Pereira Caldas
Ney Araújo Nogueira

Serie: *Working Papers*
ISSN 2219-780X

Artificial Intelligence for Expedited Decision and Risk Analysis in the Brazilian Federal Tax Administrative Litigation

WP-04-2021

Diagramming: CIAT Communication and Publications Coordination

Content

Synopsis

The Special Secretariat of the Federal Revenue of Brazil is employing different artificial intelligence (AI) approaches to analyze risk in Administrative Litigation and speed up litigation decisions. Employed strategies are in different stages of development and include supervised learning, unsupervised learning, semi-supervised learning, and active learning. Besides using AI to solve specific problems, to facilitate the flow of information from AI to users and vice versa, a systemic change to how administrative litigation is handled is planned. This article intends to provide a general overview of how this is being conducted in the Brazilian Tax Administration and what are the expectations and visions for the upcoming future.

Keywords:

Tax litigation, administrative process, risk analysis, artificial intelligence.

ntroduction

Brazilian Tax Administrative Litigation is handled by Offices of Appeals within the Special Secretariat of the Federal Revenue of Brazil (*RFB*) and, as a higher administrative court, by the Administrative Council of Tax Appeals (*CARF*).

Two issues related to this Litigation process worry RFB administration:

a. the time it takes to process tax appeals and issue decisions; and

b. the absence of risk control of litigious tax credits.

As far as the first issue is concerned, it takes a long time for a tax appeal to be treated by the administration. Every year thousands of tax assessments are the subject of appeals filed by the taxpayers. A proportion of RFB’s workforce is assigned to offices of appeal: they constitute the officers responsible for issuing decisions, typically on very different matters and sometimes different kinds of taxes. Since the number of officers in charge tends to remain constant or slightly decrease (due to retirements) and their ability to process extensive files is naturally limited, the total inventory of appeals awaiting decision has increased considerably.

As to the second issue, risk analysis is paramount for Tax Administration, and it must be applied and considered in all parts of its business processes, including administrative litigation. Its baseline is identifying risks and measuring their impact and probability. Whenever any of the two administrative courts agree with the taxpayer, a mistake must have been made within RFB. Maybe the taxpayer should not have been penalized from the very beginning, maybe the evidential elements should have been better prepared, maybe legislation itself is not clear or maybe something unforeseen has happened, but something should be changed.

Modern risk analysis makes use of complex statistics tools, which approximates it to another exponentially growing field with similar roots: Artificial Intelligence (AI). Therefore, it is natural for those fields to come together to provide integrated solutions.

Although Brazil has started long ago detecting and monitoring risk in the litigation process, a more systemic approach has become more intense in the last decade, by changing its focus from following up relevant cases to more statistic-based management of all federal tax litigation universe.

In the last two years, many improvements were made or are in progress. Risks mitigation is surely among their main concerns, but there are also many other goals that contribute to it indirectly, like speeding up business processes and providing better user interfaces and ready-to-use information so qualified process operators can concentrate on their expertise, avoiding overloads of repetitive data gathering.

In parallel, a detached project was created to identify the most evident and relevant losses both in tax recovery and tax credit issues regarding non-compliances and divergences between tax auditing and administrative judgment business processes. In parallel, a systemic review of the complete administrative litigation process is underway within a perspective of both process integration and artificial intelligence. Business analytics, document labeling, natural language processing, unsupervised, semi-supervised, and supervised learning, interoperability among systems, and data governance are all tools that are being applied in each stage and globally to tax litigation business processes.

Recently, RFB created the *Center of Excellence in Artificial Intelligence (CEIA)* with a small group of AI experienced members of its staff, including masters and Ph.D.'s in the field. A multidisciplinary team was assembled to introduce AI in the area of administrative litigation including members of CEIA and specialists in litigation.

In this article, we are going to describe the ongoing projects to speed up tax appeals handling and decision and to implement credit risk control in the litigation life cycle.

1 Tax appeals inventory management and distribution

Right after a taxpayer files an appeal, the document is received and recorded in a transaction system together with all other associated documents or data files. The appeal files integrate a large inventory which easily overcomes the litigation team capacity of manual treatment. To cope with this increasing demand, the administration developed a system of thematic classification of tax appeals to distribute similar files to the same officers of appeal. Thus, the natural goal of the first efforts towards the use of artificial intelligence has been the classification of tax appeal files in clusters. Thematic labeling directly improves the fluency of the decision process in both stages – distribution of files and decision making, while at the same time being a necessary prerequisite for the implementation of risk control practices.

1.1 Supervised machine learning for clustering tax appeals

Our first artificial intelligence goal was the development of a supervised machine learning algorithm for the thematic classification of tax appeal files whose subject was the personal income tax. RFB currently holds an inventory of over 100.000 files of this type waiting for a decision. The project started with the selection of a sample of 2,000 tax appeal files for manual labeling by a team of experienced appeals officers. Labels were expected to be nonexclusive, and the set of labels assigned to a specific file was expected to convey information enough to prescribe the best processual treatment for it.

The appeal officers stressed that the distribution of files previously grouped in thematic classes implied processual economy, for several administrative proceedings could be prepared in batches if the files to be processed were similar in some way. However, two years ago there were still no predefined classes. In addition, when testing clustering based on certain documents (such as the taxpayer's appeal request), we ended up with many nonsense clusters, something that is observed in the word clouds that will be discussed in section 1.2. So we concluded that the problem of grouping similar appeal files should not be handled as a clustering problem in a technical sense, but as another well known machine learning task: classification. While clustering is an unsupervised type of problem, classification is of a supervised type and its handling is thus considerably easier.

The team of experts eventually decided on the use of 94 labels and the AI team modeled the problem as a multilabel supervised learning task. The AI approach was divided into three phases: preprocessing, feature engineering, and training. In the preprocessing phase, we needed to extract from each of the 2000 labeled files only the documents that would be useful for the training of the classifiers. Each tax appeal file consists of multiple documents (notifications, taxpayer's requests, bills, and so on) in PDF format, of which many are not significant for the task of classification. Many sections within relevant documents can also be discarded. We thus wrote routines to extract the relevant documents from the file and to extract the relevant sections of text from those documents.

In the next phase (feature engineering) we defined a working set of features. These came from two sources: 1) unstructured data extracted through a bag-of-words approach from two kinds of relevant documents extracted in the preprocessing phase: the tax fine notification and the taxpayer's appeal request, and 2) structured data extracted from the database of tax appeals.

The third phase (training) involved the testing of 4 classifiers: Logistic Regression, XGBoost, Support Vector Machine, and Complement Naïve Bayes. We worked in a Python 3[1] environment and used Numpy[2], Pandas[3] and Scikit-Learn[4] libraries.

The trials conducted with the sample of 2000 files have demonstrated that the best algorithms can attain over 80% sensitivity and specificity for most labels (52 so far). This result was considered fair enough for the first cycle of tests and we decided to try the classifiers in a production environment. Thus, in the first semester of 2021 circa 6800 taxpayers' tax appeal files were distributed to offices of appeal after being classified and thematically grouped.

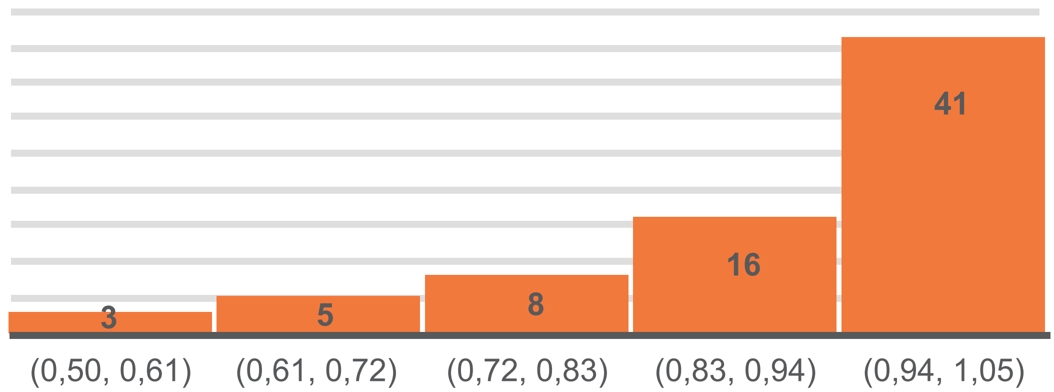
1 Van Rossum G, Drake FL. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009.

2 Harris, C.R., Millman, K.J., van der Walt, S.J. et al. Array programming with NumPy. Nature 585, 357–362 (2020). DOI: 10.1038/s41586-020-2649-2. (Publisher link).

3 The Pandas Development Team.pandas-dev/pandas: Pandas. Zenodo,2020. Disponível em: https://doi.org/10.5281/zenodo.3509134.

4 Pedregosa et al. Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830, 2011.

Figure 1. **Number of labels in each algorithm labeling sensibility intervals**



Source: Provided by the authors.

As a final remark, we stress that there will be a significant distinction in treatment between recently filed tax appeals and the ones already in inventory (which like the above-mentioned 6800 files are being currently processed by the administration). A new interface for the reception of taxpayers' complaints – "e-Defesa" - has recently been launched with the promise of recording as structured data all the information which had been available so far only in natural language.

e-Defesa facilitates the drawing up of a tax appeal request, by presenting the taxpayer with suggestions of relevant allegations for each type of violation included in the notification of tax assessment, and by indicating the documents that must be delivered to the RFB. Tax violations along with the corresponding counter-allegations are comprehensibly codified so that minimum use of natural language is necessary, and the key points of the litigation process can be stored as structured data.

The classification tasks will be significantly made easier with the full incorporation of e-Defesa resources and will be thus naturally reformulated. However, the inventory of old files still presents us with the necessity of facing the issues arising from the processing of long text sections. This is the raison d'être of the present project and why we keep investigating further improvements in natural language processing methods.

1.2 Unsupervised machine learning for clustering tax appeals

We cannot always expect to count on experts to label documents manually. The manual labeling of tax appeal files is a recent and time-consuming activity. Therefore, there will be situations where there are no previous labels. In those cases, unsupervised approaches are a good tool to be applied, especially considering that the tax appeal files contain parts which summarize their contents. The "ementa" ("decision summary") is a short, clear list of the key juridical points under discussion and reflects the point of view of the appeals officer.

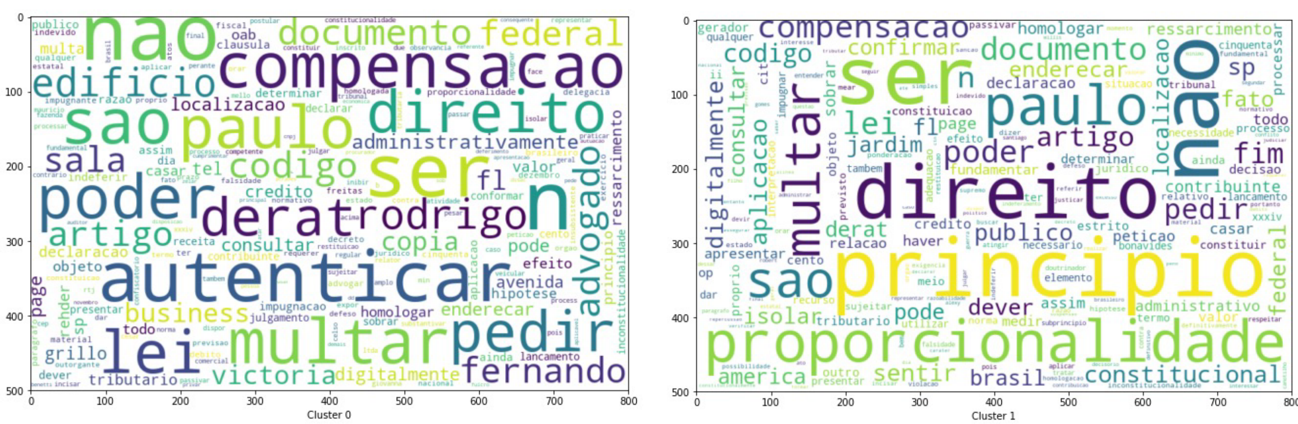
For that reason, a first stage was conducted using not the complete texts but only the decision summary, because then it would be easier to cluster the corresponding files. A corpus with over 120,000 decisions was constructed. A sample of 2,279 tax appeal files was taken and two clustering experiments were done, the first one to separate Provisional Contribution on Financial Transactions (CPMF) files from other kinds of appeals and the second one to identify topics within those documents. The word clouds generated in this manner already made some sense for specialists. Although the decision summary is only produced after the decision has been issued and our goal is to group the files before distribution for analysis, the procedure is justified: by clustering the appeal files based only on natural language processing of the decision summaries (which are much shorter than the taxpayers' original requests) we have been able to achieve relevant groupings of the files. Next, we can use those groups' identifiers as labels in the much easier task of supervised learning. In this way the goals of this phase of the project naturally connect with the previous one.

Considering that, first some logical infrastructure was developed in terms of collecting a minimal corpus and conforming a Portuguese language lexical and synonym database for future use. Next, a sample of texts taken from appeal files was treated for identifying thematic similarities using those texts as bags-of-words, after a time-consuming although non-exhaustive cleaning previous step. Cleaned texts were tokenized and lemmatized, stop-words removed, named entities identified, but still some noise was detected, like typos and residual characters.

For grouping those tax appeals, one thing to consider is that they are not one-topic texts, because a typical document may have many topics, each one revealing a different taxpayer argument, although even related among themselves. An initial approach was taken by disregarding that multidisciplinary nature and applying a sparse word matrix based on the ITF-DF metric for each word and each text and k-means clustering algorithms. Further, an improvement in that metric was tried by also considering the probability mass associated with each word taking as reference an all-texts corpus. Finally, for dealing with the many-topics nature, a Latent Dirichlet Allocation (LDA) was applied. As it is an essentially empirical method and whose results must be assessed by the experts supporting the project, the output visualization was chosen to be the detected clusters word-clouds, where those experts can check if the most prominent words define a relevant semantic field easily associated to a usual theme for that business process.

Clusterization was somewhat consistent, but it was not enough, hence the option to step further trying LDA. In this project, a supervised approach was not yet applied, but it is foreseeable for the next stage. This project is very recent, and, despite some promising perspectives and partial results, there are no definitive results to be registered and made public so far.

Figure 2. **Examples of word clouds for two of the identified clusters by the algorithm**



Source: Provided by the authors.

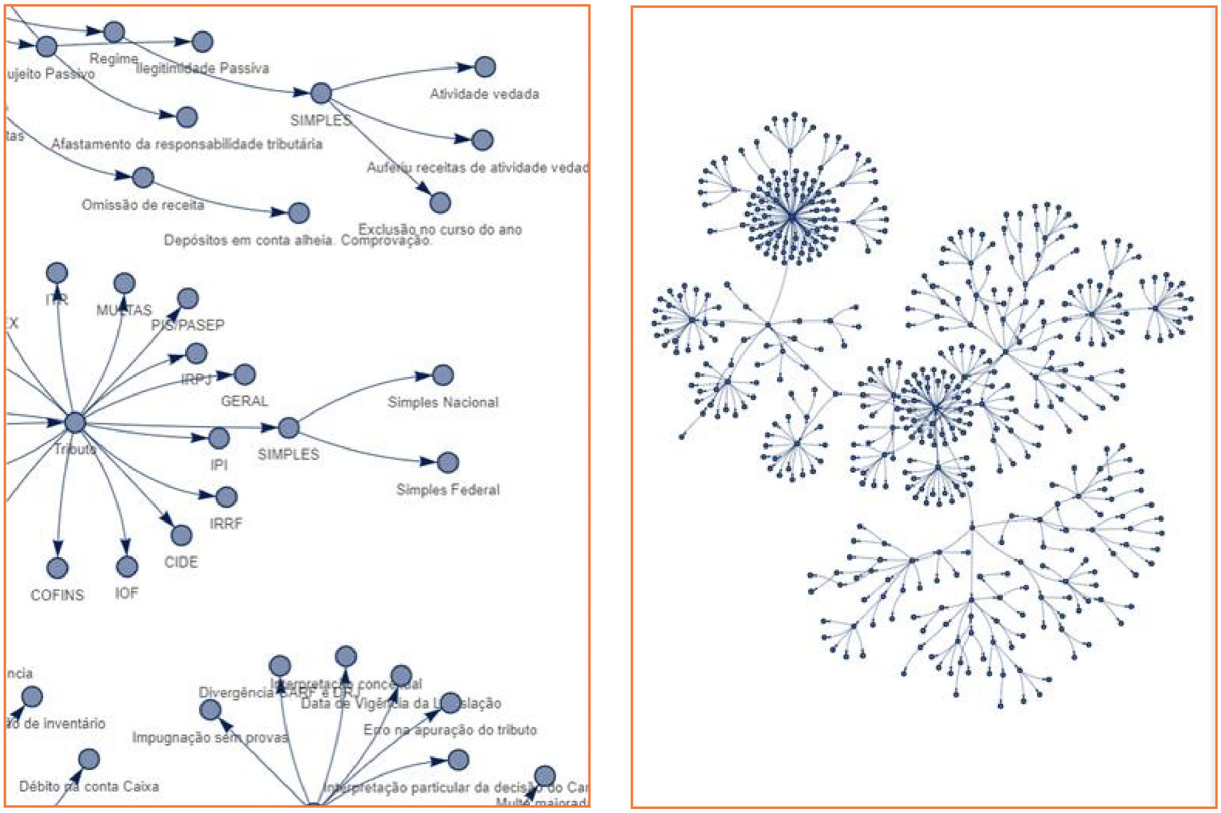
1.3 Graph tool interface for tax appeal clustering and distribution supervision

Until here, the focus was mainly on developing algorithms for automatic establishing groups of tax appeals to be distributed, but no attention was driven to how people would interact with those clustering definitions. Therefore, another frontline is developing a graphical interface to allow users to correct the algorithms' suggestions.

For this human action on the many labels to be efficient, the visual organization of information was necessary. A tree-oriented model was tried, but as labels appear in many branches, we opted for the acyclic directed graph model (DAG).

This tool is being developed as a graph-based interface oriented by usability and complex network algorithms for detecting patterns and improve visualization. A draft of what could be visualizations provided by that tool is presented in Figure 3.

Figure 3. **Draft of the graph tool user interface diagrams (under development)**



Source: Provided by the authors.

Users can rearrange the labels and also associate them with files, through the interface that is under development.

Using the structure of DAG, the algorithms can act simply: knowing the leaves, knowing all the ascendants. As each label is associated with a binary classification problem, we can predict all labels even if only leaves are indicated.

2 Tax appeal analysis

While all activities described so far refer to receiving and treating tax appeals selection and distribution, another very important group of activities must be considered which are related to the analysis of the appeal files themselves.

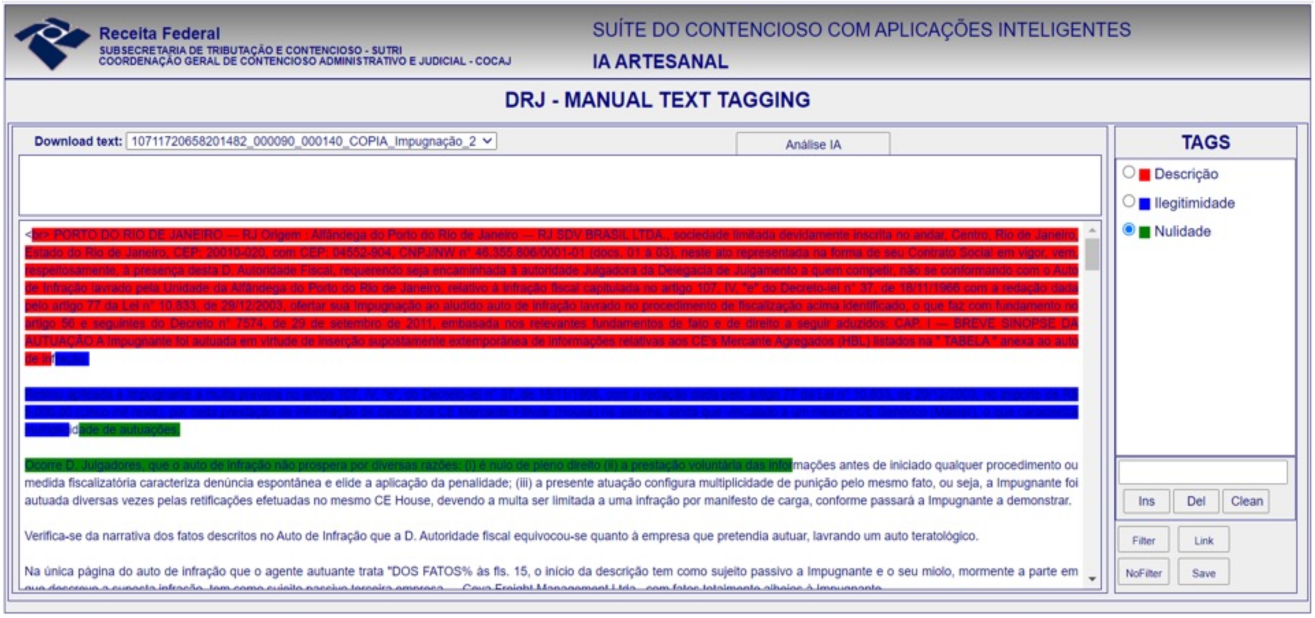
From the appeals officer's perspective, each appeal file must be prepared with all the data, information, and knowledge necessary for its complete analysis and decision making, and all that in a ready-to-use format. Besides, it is much more effective if process workloads are as similar as they can be, allowing the knowledge acquired and consolidated to analyze one appeal file to be promptly applied by the judge to others in the same group.

Hence, two different frontlines were designed to help in administrative judge work: one for developing a user interface-based tool for identifying thematic topics within texts and similarities among them for making more homogenous sets of appeal files to be distributed to each appeal officer, and another one for collecting information and knowledge subsides (legislation, similar decisions, etc.) based on the identified topics in each file.

2.1 Supporting tools for tax appeal analysis

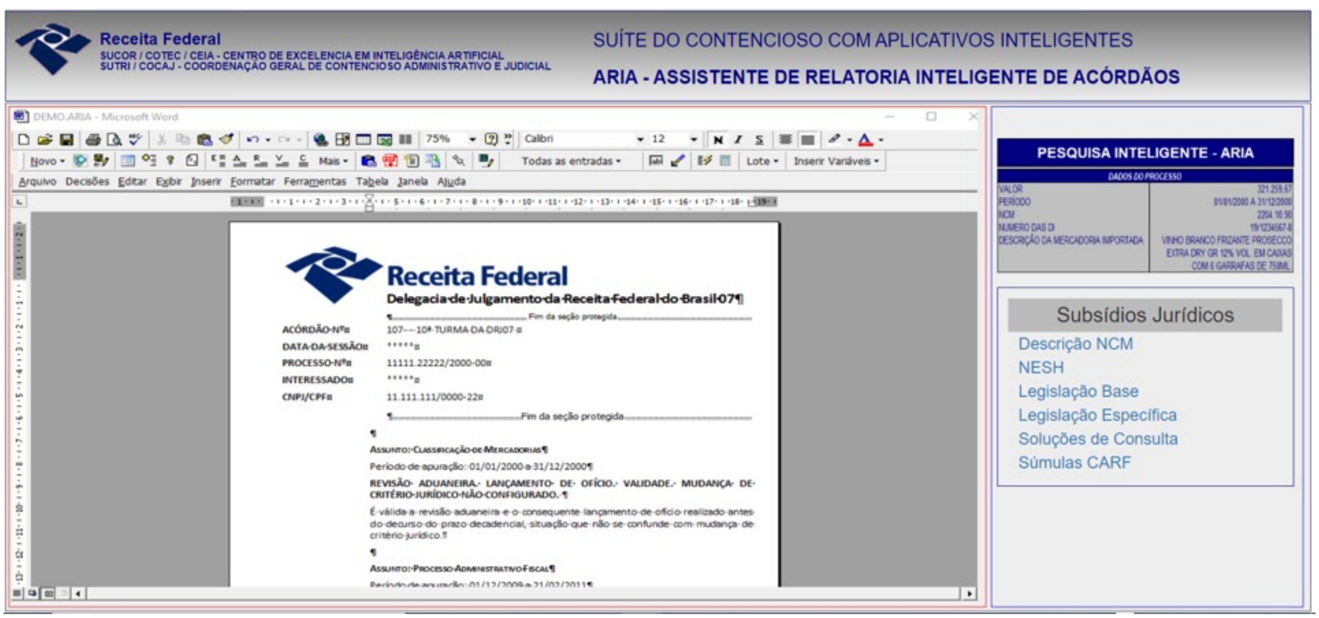
We are currently investigating new AI tools and correlated user interfaces applied to the administrative litigation proceedings in the case of customs fees. Our focus is on improving the user experience of the officers in charge of appeals analyses and decisions. Functionalities will include text similarity detection and thematic prospection of similar decisions. Two tools are under development: 1) one for tagging texts, which intends to provide material for future thematic AI learning, and 2) another one for supporting the entire administrative process of analysis and decision. Drafts of those tools are presented in Figure 4 and Figure 5, respectively.

Figure 4. **Text tagging tool user interface (in development)**



Source: Provided by the authors.

Figure 5. **Draft of the tax appeal analysis supporting tool user interface (in development)**



Source: Provided by the authors.

In subsections 1.1 to 1.3, we have explained the current efforts to thematically classify tax appeal files. The tagging tool described in the present section will help us in the identification and clustering of parts of files so that we can attain an ever-micro-grained knowledge of single documents, subsections of single documents, and even juridical arguments occurring within a given subsection. This new data will provide us with a wealth of new features to be used in classification tasks while constituting the training set of a recommendation system that will support the appeal officers by suggesting them specific pieces of legislation and typical juridical responses. The recommendation system will be integrated with an award-winning solution already in production. “Decisões”[5] is a user interface built on Word that facilitates the drawing up of appeal decisions by suggesting proper layouts, legal citations, and texts.

2.2 Jurisprudence analisys

Brazilian experience with artificial intelligence in administrative tax litigation includes the creation of a model to predict the use of CARF’s BINDING LEGAL PRECEDENTS (BLPs)[6].

BLPs speed up the jurisdictional activity through the standardization of jurisprudence, that is, it summarizes the peaceful and dominant decisions of a court and makes them mandatory. In addition, the use of BLPs reduces the number of appeals and allows for simpler and faster judgment sessions (non-face-to-face and virtual).

The inventory of administrative tax appeals awaiting a decision is in the order of US$ hundreds of billions[7], with an average processing time of about 3 years. Therefore, the demand for faster decisions is critical.

---

5 Innovation Creativity Award from the Federal Revenue of Brazil (3rd place) - 2006.

6 Nogali, Patrick Moreira. Julgamento tributário agilizado com inteligência artificial: criação de modelo de predição de uso de súmulas do Conselho Administrativo de Recursos Fiscais (CARF). Course monograph (Specialization in data Science and Big Data) - Pontifícia Universidade Católica de Minas Gerais. Belo Horizonte. 2021.

7 See: http://carf.economia.gov.br/dados-abertos/relatorios-gerenciais/2020/dados-abertos-202012-dezembro.pdf

The use of the BLPs, however, consumes appeals officers' time, as it requires individual research work. When creating a BLP prediction model, we seek to save judgment time through machine learning. The model predicts, based on documents before the trial (report of the judgment) which of the 161 BLPs in force will likely be added by the appeal officer later.

The BLPs applied in decisions are not identified in a structured way, but we can confidently generate this information by applying regular expressions to their text.

The application of each BLP was treated as a separate binary prediction problem. Models were created for each of these problems in two stages of machine learning: Natural Language Processing (NLP) and Classification.

This time, to make our work easier, we employed Pycaret, a tool that tests several NLP and classification models simultaneously[8].

The best models for each BLP were produced by different algorithms, for example, BLP #1 had better indexes with Random Forest Classifier, while BLP #2[9], the most used in the dataset, had better precision indices with Ridge and Logistic Regression and accuracy varied, being satisfying in some cases, but not in others, as seen in Table 1.

8 ALI, Moez. PyCaret: An open-source, low-code machine learning library in Python. 2020. https://www.pycaret.org. Acesso em 24/09/2021.

9 CARF BLP #2: "The CARF is not competent to rule on the unconstitutionality of a tax law"

Table 1. **Best results for some BLPs.**

BLP #	Ocurrences	Best Model	Accuracy	Sensibility	Specificity
2	22,14%	Ridge	0,8771	1,0000	0,8502
1	2,01%	Random Forest	0,9980	0,8333	0,9816
157	0,06%	Random Forest + Boost	0,9881	0,6000	0,9960
17	0,03%	Naive Bayes + Boost	0,9542	0,0435	0,9979

Source: Provided by the authors.

Note: In particular, predictions of rarely used BLPs are still weak and in the case of the rare ones, algorithms tend to just always predict that they will not be used, that is, they estimate that the probability of them being used is below 50% for all appeal files. However, even in low ranges, probabilities vary and there is a big stock of appeal files that pre-date the existence of the BLPs.

We plan to use start an active learning procedure where an algorithm will ask appeals officers to, among the old stock, label those files whose probabilities of being related to a rare label is highest.

A planned evolution for this model is to start making forecasts based on other documents before the judgment, such as, for example, the challenge of a tax infraction notice. Another evolution will be the forecast of citations of other types of legislation that are not CARF BLPs.

It is necessary to recognize that the algorithm works worse for the rare cases that may be being forgotten, however as the appeals officers receive the suggestions and confirm or not, the performance tends to improve

.

3 Strategic management and risk analysis

At tactical and strategic levels, the presence of risk analysis becomes more evident. One reason for that is the proper statistical nature of risk, which is better quantified the more aggregated the level of perspective.

Therefore, management tools and business analytics are very powerful means for dealing with risk analysis and management.

3.1 Management tools and business analytics

Strong risk analysis demands strong information-based management. Tax appeals file management tools are being reinforced in the last years in the Brazilian Tax Administration (RFB). Two great drivers for that are the acquisition of an integrated solutions platform (office, communications, storage, analytics, etc.) from a market-leading IT provider and the boost of internal software development including AI algorithms, servers, and web-based intranet solutions.

Another extremely relevant dimension in that front is the data and information challenge. Although RFB has a lot of data, just as any other big organization in the world, that data was produced along decades of transactional systems operation and so they are transactional system-oriented, not business-process or user-oriented. That makes it much more difficult to produce management useful and reliable information. In the past, it was not unusual that the same report on the same information taken from two different systems provides two different numbers.

Regarding that issue, RFB started a Data Governance initiative and created sectorial groups and a committee to address the problem. One of the results of that work is the ongoing creation of information bases that are permanently updated business and user-oriented secondary databases. Those information bases will be the primary information source for management tools and business analytics, granting readiness, consistency, and reliability for managers and, consequently, for risk analysis.

Some results of those initiatives can be seen in the illustrative case described in the next topic.

3.2 A case of risk analysis and treatment

As a case report in terms of risk analysis, there is one project conducted by the RFB head office which focused on the relatively high risk of tax appeals to be decided in Tax Administration disfavor, representing a great loss of time and work and a strong need for internal compliance. That risk was partially quantified by considering a slice of more than 17,000 appeals with Tax Administration's unfavorable unanimous decision as a first stage for dealing with it.

The next stage was diagnosing which were the problems with those appeals for further planning and adoption of correction measures. As analyzing more than 17,000 would consume too much effort, it was mandatory to seek help in technology. Thus, the project team decided on an AI semi-supervised approach in which a first small sample would be manually labeled by business specialists and then that sample would be submitted to classification algorithms in subsequent improvement cycles supervised by specialists interacting with the algorithm by an appropriate user interface.

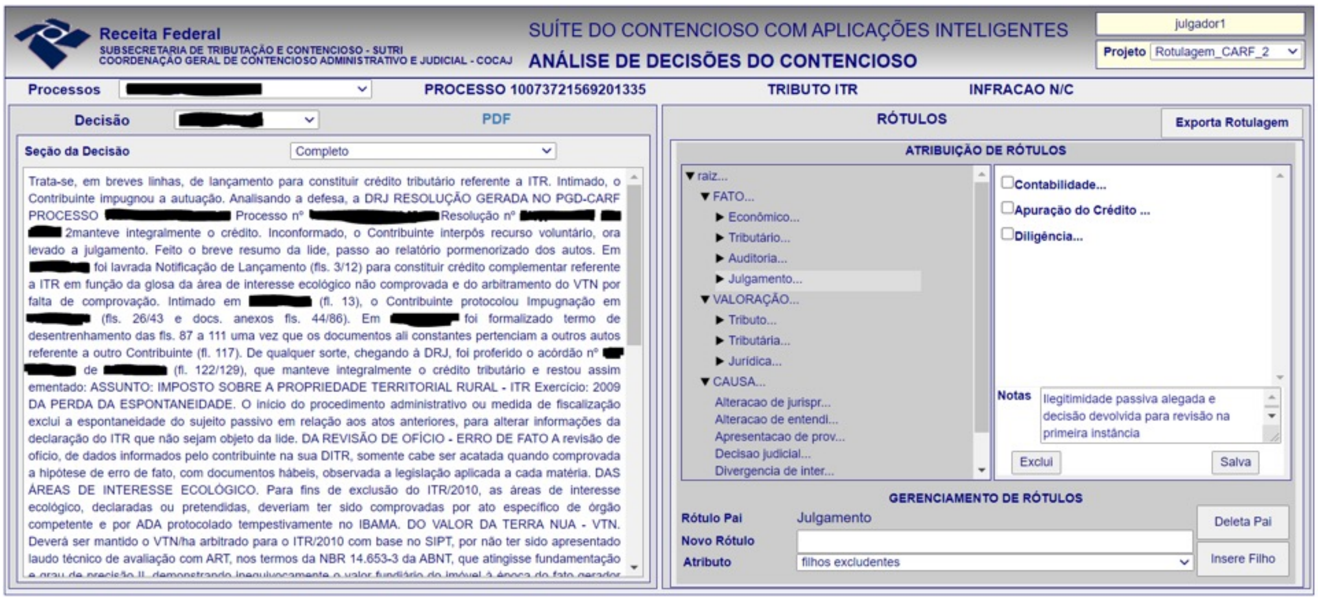
For that matter, a sample of 800 tax appeals files was taken and submitted to a group of 11 expert judges, divided by tax specialty, and they were asked to analyze and mark those files with appropriate labels to indicate the main causes of the loss and possible solutions. In parallel, a web-based intranet user interface labeling tool was developed, and its first version became ready in the middle of the labeling work. Then, all the labeling done so far was inserted into that tool and labelers started to use it.

Besides the immediate benefit of organizing and registering the labeling in a database, there were other benefits in providing a unique labeling tool for the entire labeling team, such as unifying language and labels and forcing online consolidation for the labeling structure (at first a tree-shaped structure that further migrated to a more representative directed acyclic graph - DAG).

After the manual labeling was completed, the team delivered as a result the final labeling of the 800 files and a report describing relevant situations encountered and difficulties in the labeling process itself for further improvements in the tool and in the labeling methodology. Essentially, they pointed out a few issues in tax administration that could be treat immediately regardless of deeper analysis and that the tool needed to be more user-friendly in interacting with the label's structure. Both considerations generated the expected correction actions.

A second stage is now under execution treating another slice of tax appeals on credit recognition.

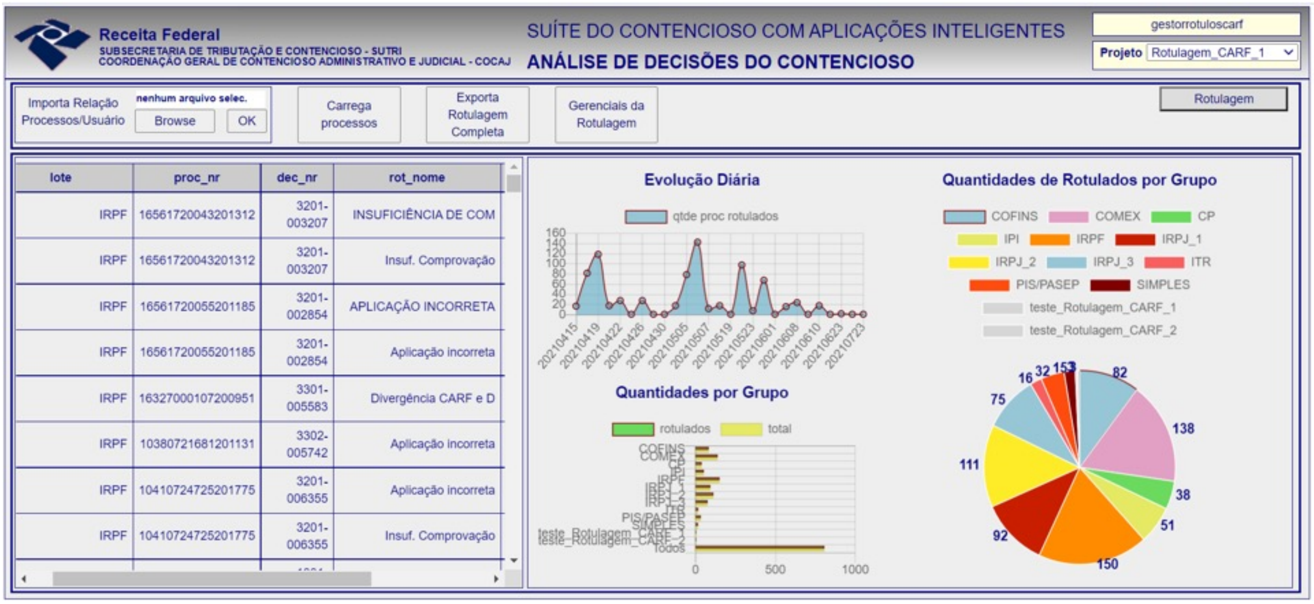
Figure 6. Tax appeal file labeling tool user interface (version 1.0)



Source: Provided by the authors.

Note: Labeling screen. The left panel shows the entire text extracted from the pdf decision file, while the left panel shows tree-structured labels and the selected labels to apply to that specific file. Users can also include new labels whenever needed.

Figure 7. **Tax appeal file labeling tool user interface (version 1.0)**

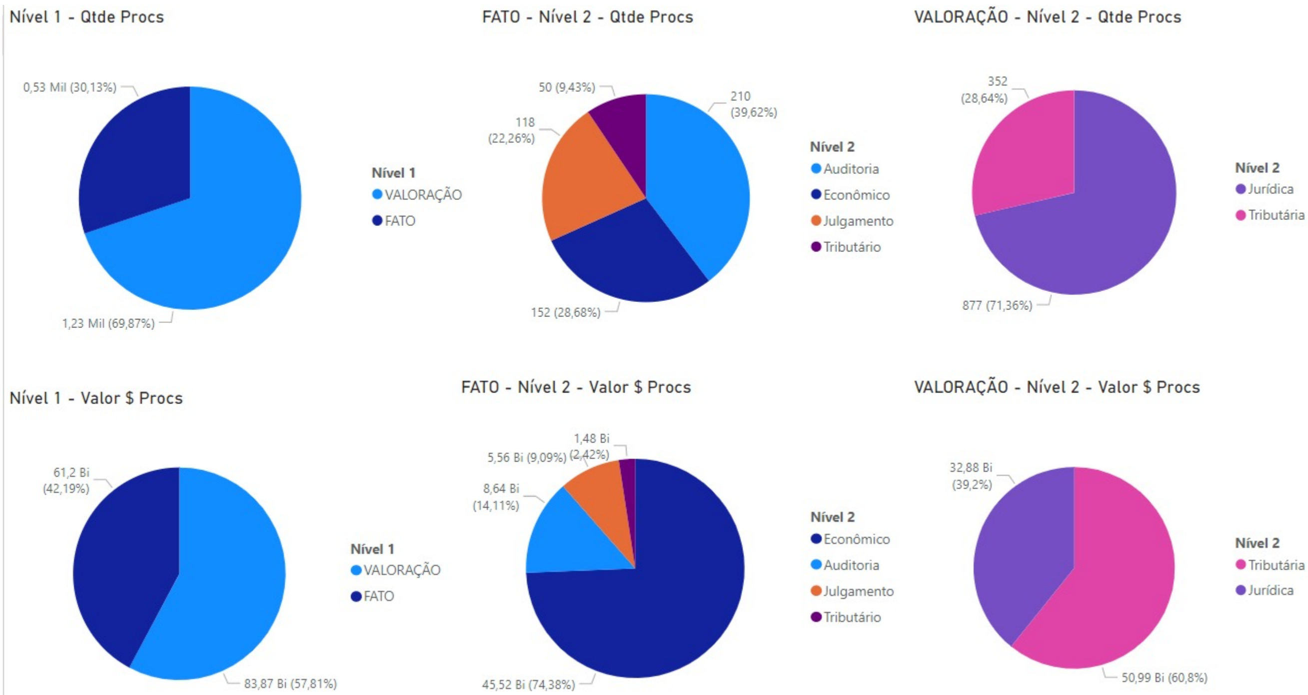


Source: Provided by the authors.

Note: Management screen. Shows the evolution of the labeling activity by each subgroup of files.

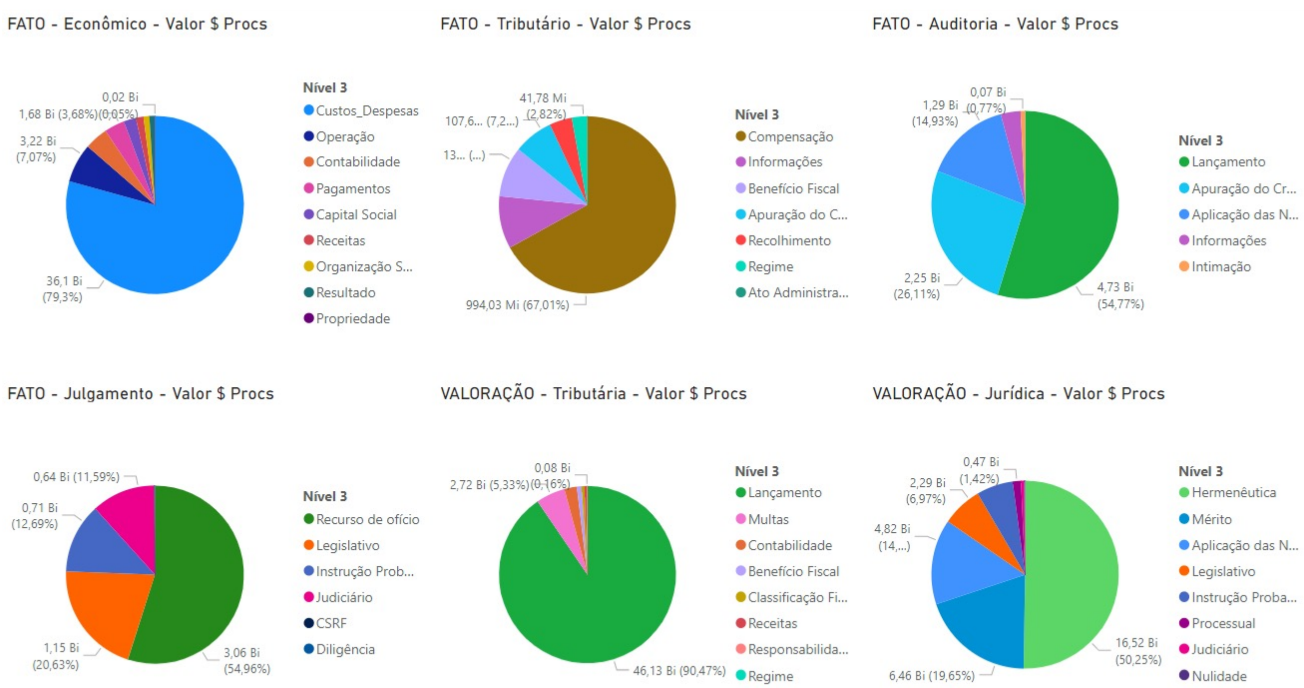
The results obtained in that first stage of manual labeling can be seen in Figure 8 and Figure 9. In summary, it was observed that the causes of loss were concentrated in some categories, such as issues related to company costs, tax compensations, and errors in the auditing process. It was also detected a relevant loss in mandatory appeals which reveals law application divergences between instances. The adequate correction measures were addressed for each one of those categories.

Figure 8. **Labels distribution among the 800 appeals analysis – Levels 1 and 2 (tax administration process stage)**



Source: Provided by the authors.

Figure 9. **Labels distribution among the 800 files analysis – Level 3 (total BRL$ value)**



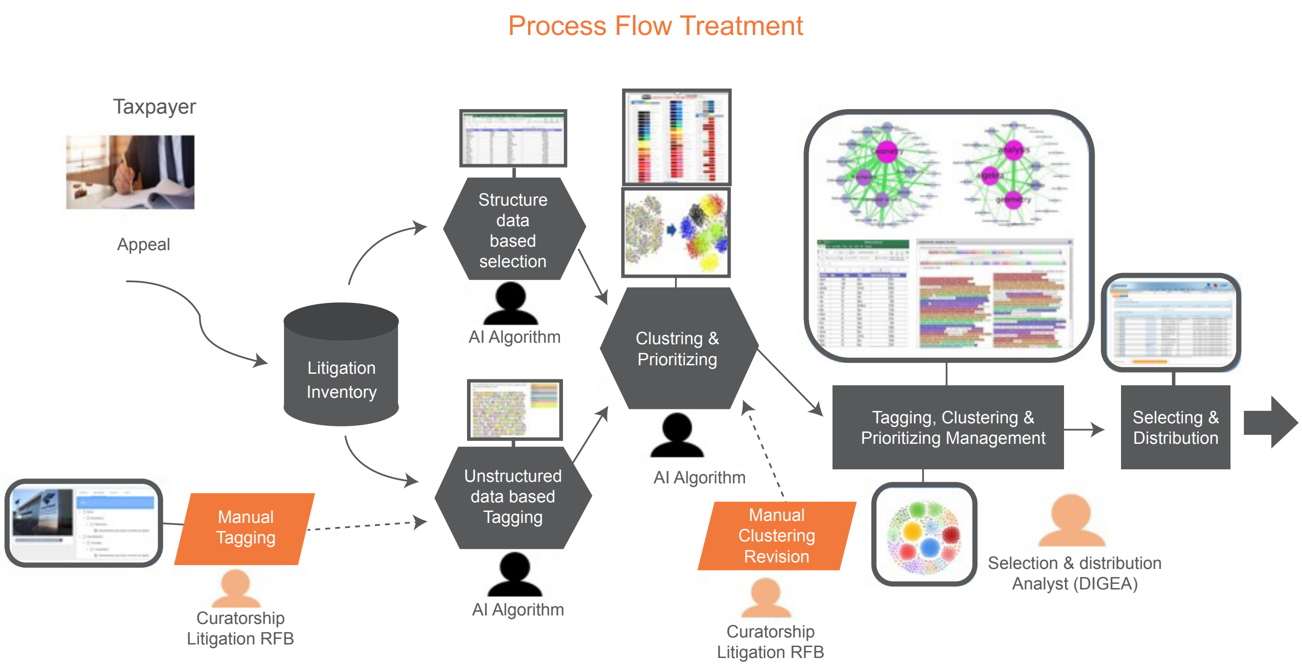
Source: Provided by the authors.

4 Future

The risk analysis and artificial intelligence approach in *SUTRI* is being conducted from a systemic perspective. About a year ago a program was created and designated Litigation Suite with Intelligent Applications - *Suíte do Contencioso com Aplicações Inteligentes* (SCAI) which consists not only in applying risk analysis and AI to the existing tax appeals punctually but to gradually remodel the entire litigation macro-process to absorb the best from those areas.

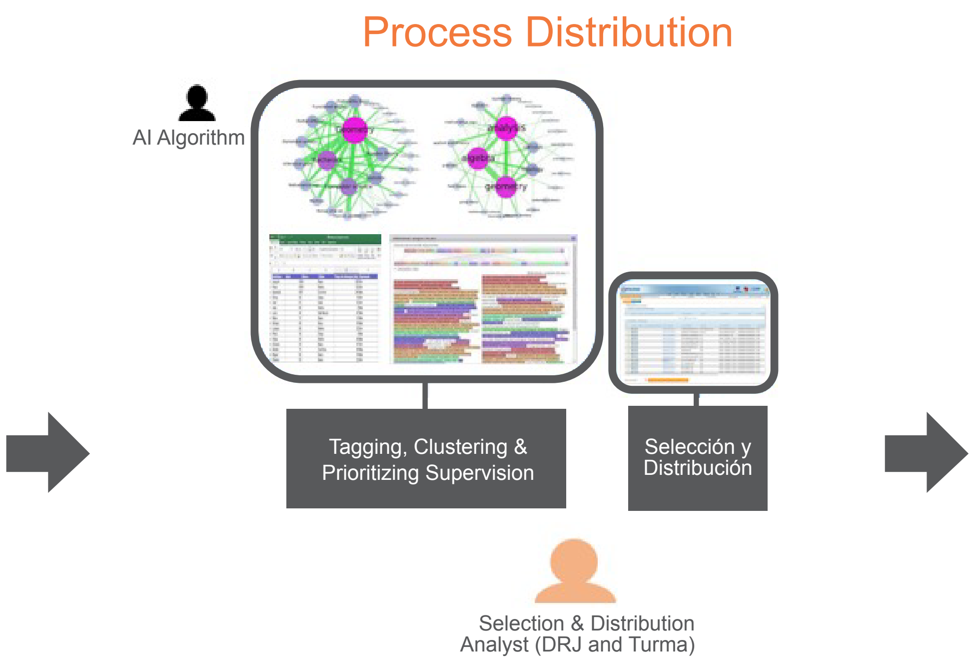
The program contains all initiatives already described and some others to be developed soon, all from an integrative perspective. The vision of this program can be seen in the next figures.

Figure 10. Overview of the litigation process from start to distribution



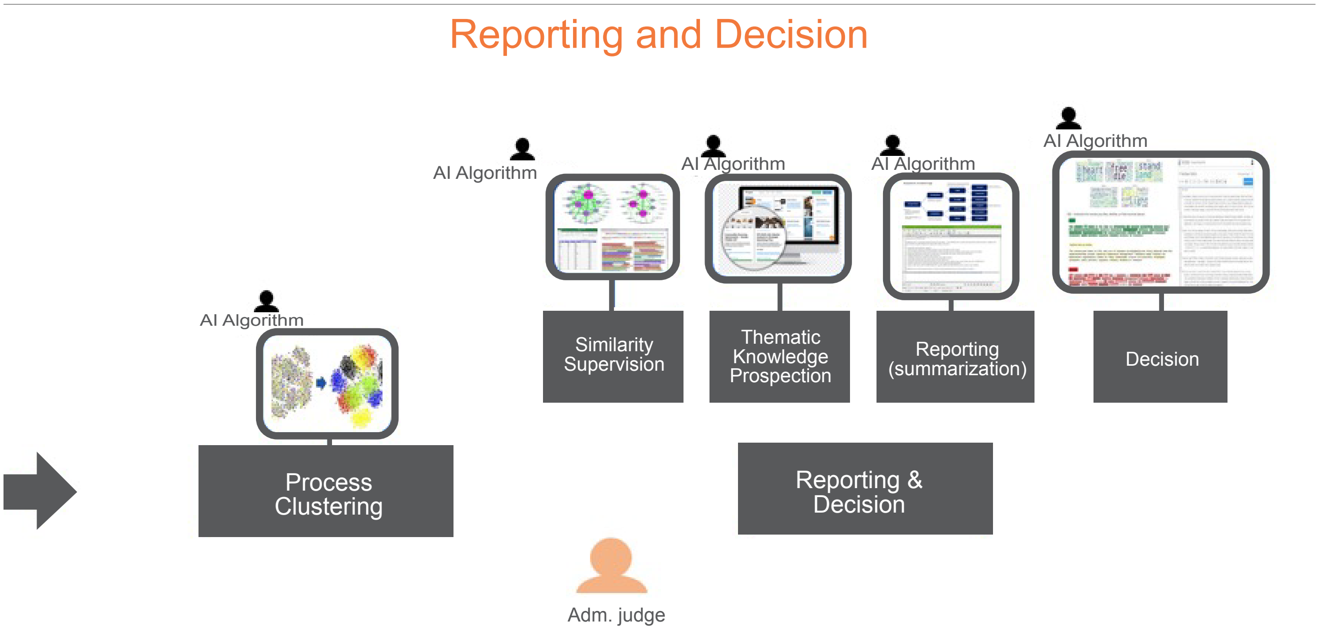
Source: Provided by the authors.

Figure 11. Litigation process – Tax appeal file distribution



Source: Provided by the authors.

Figure 12. Litigation process



Source: Provided by the authors.

Note: Report elaboration and decision (vote) by administrative judges using judgment support tool (Assistente de Relatoria com uso de Intelligência Artificial - ARiA).

Figure 13. **Litigation process– management and risk analysis by operational, tactical, and strategical managers**

Source: Provided by the authors.

Conclusion

The option for dealing with risk analysis and artificial intelligence as two sides of the same subject seems so far to have been a good choice.

Although the initiatives have produced until now more local benefits, we can see the potential for overwhelming results in the future. One example derives from the debated court losses case, where it was easier to achieve global results adopting technological approaches to the problem.

The Brazilian Tax Administration intends to persevere in that direction because it seems to be one with excellent expectations.

References

Buitinck et al. API design for machine learning software: experiences from the scikit-learn project, 2013.

Harris, C.R., Millman, K.J., van der Walt, S.J. et al. (2020). Array programming with NumPy. Nature 585, 357–362 DOI: 10.1038/s41586-020-2649-2. (Publisher link).

Nogali, P. M. (2021). Julgamento tributário agilizado com inteligência artificial: criação de modelo de predição de uso de súmulas do Conselho Administrativo de Recursos Fiscais (CARF). Course monograph (Specialization in data Science and Big Data) - Pontifícia Universidade Católica de Minas Gerais. Belo Horizonte.

The Pandas Development Team. pandas-dev/pandas: Pandas. Zenodo, 2020. https://doi.org/10.5281/zenodo.3509134.

Pedregosa et al. Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830, 2011.

Van Rossum G, Drake FL. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009.

Working Papers *Series*

CIAT Executive Secretary
P.O. Box: 0834-02129, Panama, Republic of Panama
Phone: (507) 307.2428
Fax: (507) 264.4926
E-mail: ciat@ciat.org
Web: www.ciat.org