



SENADO FEDERAL
Instituto Legislativo Brasileiro – ILB

Orlando de Sá Cavalcante Neto

**Desenvolvimento e aplicação de soluções de inteligência artificial
pela Consultoria de Orçamentos, Fiscalização e Controle do
Senado Federal: estudo de caso sobre a classificação do Resultado
Primário**

Brasília

2022



Orlando de Sá Cavalcante Neto

Desenvolvimento e aplicação de soluções de inteligência artificial pela Consultoria de Orçamentos, Fiscalização e Controle do Senado Federal: estudo de caso sobre a classificação do Resultado Primário

Monografia apresentada ao Instituto Legislativo Brasileiro – ILB como pré-requisito para a obtenção de certificado de conclusão de Curso de Pós-Graduação *Lato Sensu* em Orçamento Público.

Orientadora: Rita de Cássia Leal Fonseca dos Santos

Brasília

2022



Termo Geral de Autorização para Publicação Digital na BDSF

Como titular dos direitos autorais do conteúdo supracitado, autorizo a Biblioteca Digital do Senado Federal (BDSF) a disponibilizar este trabalho gratuitamente, de acordo com a licença pública Creative Commons – Atribuição - Uso Não Comercial – Compartilhamento pela mesma Licença 3.0 Brasil. Tal licença permite copiar, distribuir, exibir, executar a obra e criar obras derivadas, sob as seguintes condições: dar sempre crédito ao autor original, não utilizar a obra com finalidades comerciais e compartilhar a nova obra pela mesma licença no caso de criar obra derivada desta.

Assinatura do Autor / Titular dos direitos autorais

Cavalcante Neto, Orlando de Sá

Título: Desenvolvimento e aplicação de soluções de inteligência artificial pela Consultoria de Orçamentos, Fiscalização e Controle do Senado Federal: estudo de caso sobre a classificação do Resultado Primário / Orlando de Sá Cavalcante Neto. - Brasília, 2022 67p.

Orientadora: Dra. Rita de Cássia Leal Fonseca dos Santos.

Trabalho de conclusão de curso (Pós-Graduação Lato Sensu em Orçamento Público) - Instituto Legislativo Brasileiro, Senado Federal, 2022.

1. orçamento público. 2. resultado primário. 3. classificação orçamentária. 3. inteligência artificial voltada à área de negócio. 4. Inovação no setor público.



Orlando de Sá Cavalcante Neto

Desenvolvimento e aplicação de soluções de inteligência artificial pela Consultoria de Orçamentos, Fiscalização e Controle do Senado Federal: estudo de caso sobre a classificação do Resultado Primário

Monografia apresentada ao Instituto Legislativo Brasileiro – ILB como pré-requisito para a obtenção de certificado de conclusão de Curso de Pós-Graduação *Lato Sensu* em Orçamento Público.

Aprovada em Brasília, em 25 de outubro de 2022 por:

Banca Examinadora:

Profa. Dra. Rita de Cássia Leal Fonseca dos Santos
Senado Federal

Prof. Dr. Ricardo Suganuma
Senado Federal

Profa. Ma. Ana Cláudia Castro Silva Borges
Senado Federal



AGRADECIMENTOS

O autor agradece a Marco Túlio de Carvalho, amigo e parceiro de vinte anos de trabalho, fundamental para o desenvolvimento do SIGA Brasil e para o projeto-piloto de inteligência artificial relatado neste trabalho.

Agradece a Rita de Cássia Leal Fonseca dos Santos, amiga, orientadora deste TCC e parceira nos projetos SIGA Brasil e Orçamento Fácil.

Agradece, ainda, a Alessandra Pontes Roscoe pelo apoio e revisão deste TCC.



RESUMO

Em matéria orçamentária, a investigação e o tratamento manual dos dados têm se mostrado incapazes de dar respostas às demandas crescentes por transparência no gasto público e por tomadas de decisão baseadas em evidências, com a tempestividade e objetividade que o dinamismo da realidade impõe. Na atualidade, a fronteira tecnológica em contextos de grandes bases de dados se situa no desenvolvimento e na utilização de produtos baseados em inteligência artificial (IA), voltados a expandir a identificação de fenômenos, padrões e eventos não prontamente visíveis pelos métodos centrados nas habilidades humanas de percepção. Embora reconhecidos como necessários, o desenvolvimento e a aplicação de soluções de IA no setor público ainda são cercados de barreiras relacionadas principalmente à percepção de que se trata de área de conhecimento excessivamente complexa aos não-iniciados em ciência de dados e que os investimentos para viabilizar as soluções importam custos proibitivos. Trata-se muitas vezes de percepções equivocadas, oriundas do descompasso entre o ritmo de avanço das ferramentas e metodologias de IA aplicáveis à gestão pública e a capacidade dos órgãos de se apropriarem dessas possibilidades emergentes. Este trabalho analisa a prontidão da Consultoria de Orçamentos do Senado Federal para desenvolver projetos de IA a partir de estudo de caso sobre o desenvolvimento do projeto-piloto de aplicação de IA à classificação do Demonstrativo do Resultado Primário da União e com base nos princípios de inovação e recomendações da OCDE.



ABSTRACT

In budgetary matters, research and manual data processing have proved incapable of responding to the growing demands for transparency in public spending and evidence-based decision-making, with the timeliness and objectivity that the dynamism of reality imposes. Currently, the technological frontier in contexts of large databases lies in the development and use of products based on artificial intelligence (AI), aimed at expanding the identification of phenomena, patterns and events not readily visible by methods centered on human skills. Although recognized as necessary, the development and application of AI solutions in the public sector are still surrounded by barriers related mainly to the perception that this area of knowledge is excessively complex for those uninitiated in data science and that investments to enable solutions are cost prohibitive. These are often misperceptions, arising from the mismatch between the pace of advancement of AI tools and methodologies applicable to public management and the ability of agencies to appropriate these emerging possibilities. This paper analyzes the readiness of the Federal Senate's Budget Consultancy Office to develop AI projects, considering a case study on the development of a pilot project for applying AI to the classification of the Federal Government's Primary Outcome Statement and the OECD innovation principles and recommendations.



LISTA DE ILUSTRAÇÕES

Figura 1 – Diagrama de Venn da Ciência de Dados

Figura 2 – Representação gráfica da validação cruzada

Figura 3 – Score F1 das validações cruzadas da seleção automática de atributos por número de atributos preditores selecionados

Figura 4 – Importância relativa dos atributos para a predição do Subgrupo

Figura 5 – Representação gráfica da precisão e da revocação



LISTA DE TABELAS

- Tabela 1** – Amostras por classe na base de treinamento e de teste
- Tabela 2** – Classificação do imposto de renda de pessoa física 2014 e 2021
- Tabela 3** – Comparativo entre Desdobramento e Natureza da receita no relacionamento com o Subgrupo
- Tabela 4** – Resultados dos modelos na validação cruzada e estimativas MAP, IC e desvio padrão
- Tabela 5** – Teste t correlacionado bayesiano da validação cruzada dos modelos
- Tabela 6** – Resultados dos modelos na validação cruzada na base de teste e das estimativas MAP e IC
- Tabela 7** – Precisão, revocação, F1 e instâncias por classes de predição na base de teste do modelo *AdaBoost*
- Tabela 8** – PLOA 2021 - Demonstrativo do Resultado Primário - Comparativo do Autógrafo originário da predição com o Autógrafo publicado
- Tabela 9** – Matriz de Confusão Multiclasse da predição com modelo AdaBoost e engenharia de atributos - Predição Final do Projeto
- Tabela 10** – Matriz de Confusão Multiclasse da predição com modelo AdaBoost sem engenharia de atributos - Preditores iniciais adotados pela Consultoria de Orçamentos do Senado



Sumário

1	CONTEXTUALIZAÇÃO E JUSTIFICATIVA	11
2	OBJETIVOS	13
3	METODOLOGIA	13
4	FUNDAMENTAÇÃO TEÓRICA	14
4.1.	Inteligência artificial e aprendizado de máquina	14
4.2.	Habilidades para inovação segundo a OCDE	18
4.2.1	Alfabetização de dados	19
4.2.2	Iteração.....	21
4.2.3	Curiosidade	22
4.2.4	Insurgência.....	23
4.2.5	Contação de história (narrativa).....	24
4.2.6	Centricidade no usuário	24
4.3.	Desenvolvimento de sistemas de IA	25
4.3.1	O valor e a qualidade dos dados.....	27
4.3.2	Multidisciplinariedade e equipe	28
4.3.3	Desenvolvimento progressivo de capacidades.....	29
5	ESTUDO DE CASO: CLASSIFICAÇÃO DA RECEITA NA FORMA DO DEMONSTRATIVO DO RESULTADO PRIMÁRIO	30
5.1	Entendimento do negócio e definição do problema.....	30
5.2	Características do projeto-piloto de IA aplicada ao Demonstrativo de Resultado Primário.....	32
5.3	Conceitos básicos: algoritmo e método, biblioteca e modelo	33
5.4	Entendimento e preparação dos dados	34
5.4.1	Bases de Dados e classes do Resultado Primário.....	34
5.4.2	Seleção de atributos	37
5.5	Métrica F1 de avaliação de resultados de treinamento e de teste.....	45
5.6	Escolha do modelo.....	47
5.7	Resultados da predição na base de teste.....	50
5.8	Considerações finais sobre o projeto-piloto.....	53
6	PRONTIDÃO PARA O DESENVOLVIMENTO DE SOLUÇÕES DE IA: APRENDIZADOS SOBRE AS POSSIBILIDADES E DESAFIOS NA CONORF	54
7	CONSIDERAÇÕES FINAIS.....	59
	Referências.....	63
	Apêndice – Matriz de confusão multiclasse	65



1 CONTEXTUALIZAÇÃO E JUSTIFICATIVA

A Consultoria de Orçamentos, Fiscalização e Controle (CONORF) é órgão de assessoramento superior do Senado Federal, voltado a apoiar o processo legislativo das matérias orçamentárias e financeiras submetidas ao Congresso Nacional. Dentre as principais, destacam-se as leis orçamentárias anuais e as leis de créditos adicionais, que compreendem milhares de decisões expressas em códigos, dados e informações estruturadas. Para o desempenho de sua função, portanto, a CONORF utiliza como ferramenta de trabalho bases de dados de largo escopo e sistemas informatizados alimentados diariamente com grandes volumes de conteúdos. Nesse contexto, o órgão se depara continuamente com dois desafios correlatos: por um lado, converter em dados codificados as decisões tomadas no processo legislativo orçamentário federal, de modo a alimentar os sistemas informatizados de elaboração e execução orçamentárias; por outro lado, converter em informações e conhecimentos os dados orçamentários e financeiros que alimentam diariamente esses sistemas, a partir de ações e decisões oriundas de atores externos, especialmente do Poder Executivo.

Desde sua gênese, a CONORF tem se empenhado em trabalhar em parceria com o PRODASEN no desenvolvimento de soluções de tecnologia da informação para apoiar o duplo desafio acima mencionado. No primeiro caso, são exemplos relevantes a produção de sistemas de elaboração orçamentária (SELOR/LEXOR) e de sistema de emendas ao orçamento. No segundo caso, o destaque se deu na produção de um sistema de recuperação e transparência de dados da elaboração e execução do orçamento, denominado SIGA Brasil, que se tornou referência internacional¹. Tais sistemas demandam constante atualização para incorporarem novas funcionalidades e plataformas tecnológicas, configurando linha de trabalho continuada entre os processos da CONORF.

O impacto do desenvolvimento de soluções de tecnologia da informação sobre a produtividade e qualidade dos trabalhos da CONORF é, portanto, expressivo, embora ainda pendente de avaliação sistemática. Evidências robustas desse impacto vão desde a viabilização do assessoramento aos parlamentares e ao processo legislativo, impraticável sem os sistemas mencionados, até a capacitação dos diversos atores envolvidos ou interessados no processo orçamentário (autoridades, servidores públicos, mídia, sociedade civil e outros) para

¹ <https://publicadministration.un.org/publications/content/PDFs/E-Library%20Archives/World%20E-Parliament%20Report%20series/EParliament%20Report%202008.pdf>



enxergar as decisões e tomar parte no processo decisório. As ferramentas desenvolvidas proporcionam melhoria quantitativa e qualitativa dos processos e produtos orçamentários e de políticas públicas, tornando imperativo à CONORF o investimento na busca continuada de soluções de TI para a evolução institucional do parlamento.

Na atualidade, a fronteira tecnológica em contextos de grandes bases de dados se situa no desenvolvimento e na utilização de produtos baseados em inteligência artificial (IA), voltados a expandir a identificação de fenômenos, padrões e eventos não prontamente visíveis pelos métodos centrados nas habilidades humanas de percepção. Em matéria orçamentária, particularmente, a investigação e o tratamento manual dos dados têm se mostrado incapazes de dar respostas às demandas crescentes por transparência no gasto público e por tomadas de decisão baseadas em evidências, com a tempestividade e objetividade que o dinamismo da realidade impõe. Descritores genéricos utilizados nos demonstrativos orçamentários impedem a pronta identificação dos destinatários e dos objetos do gasto público, enquanto o desdobramento de classificadores de receitas e despesas em volumosos eventos contábeis torna penosos e, em muitos casos, inviáveis os trabalhos de mapeamento de efeitos, resultados e impactos das decisões orçamentárias. Avanços no emprego de máquinas para investigar grandes bases de dados e trazer à luz informações e conhecimentos relevantes se mostram inescapáveis à melhoria no desempenho da gestão pública.

Embora reconhecidos como necessários, o desenvolvimento e a aplicação de soluções de IA no setor público ainda são cercados de barreiras relacionadas principalmente à percepção de que se trata de área de conhecimento excessivamente complexa aos não-iniciados em ciência de dados e que os investimentos para viabilizar as soluções importam custos proibitivos. Trata-se muitas vezes de percepções equivocadas, oriundas do descompasso entre o ritmo de avanço das ferramentas e metodologias de IA aplicáveis à gestão pública e a capacidade dos órgãos de se apropriarem dessas possibilidades emergentes. Coloca-se, portanto, a necessidade de desmistificar e difundir as vertentes da IA nas áreas de negócio do setor público e oferecer evidências e metodologias para o desenvolvimento de soluções que levem as organizações a novos patamares de desempenho na criação de valor público.

Este trabalho está dividido em sete partes, incluindo esta seção de contextualização e justificativa. A seção 2 identifica os objetivos do trabalho. A metodologia adotada é o foco da seção 3. A seção 4 trata da fundamentação teórica. O estudo de caso sobre a classificação da receita na forma do Demonstrativo do Resultado Primário é apresentado na seção 5. A seção



6 aborda as possibilidades e desafios para a participação da CONORF no desenvolvimento de futuras soluções de IA. As considerações finais são descritas na seção 7.

2 OBJETIVOS

O objetivo geral do presente trabalho é mapear possibilidades e desafios no emprego de tecnologia baseada em IA em processos de trabalho da CONORF, tomando por base os aprendizados obtidos em projeto-piloto de inteligência artificial (IA) aplicada ao Demonstrativo do Resultado Primário, desenvolvido recentemente pela CONORF em parceria com o PRODASEN, e da análise de maturidade da organização em face dos preceitos teóricos atualmente aceitos no campo da IA, voltados a criar um ambiente propício à IA no âmbito da Consultoria e ao planejamento e execução de projetos de IA.

Para tanto, a pesquisa observará os seguintes objetivos específicos:

1. Revisar a literatura de IA aplicada ao setor público, incluindo os preceitos fundamentais sobre aprendizado de máquinas e as melhores práticas recomendadas por entidades internacionais de referência na matéria;
2. Revisar os princípios e recomendações da OCDE sobre inovação no setor público;
3. Descrever o desenvolvimento do projeto-piloto aplicado ao Demonstrativo do Resultado Primário, especialmente quanto aos requisitos de formulação do objetivo, preparação das bases de dados, seleção do modelo e apuração dos resultados obtidos;
4. Analisar as evidências que o projeto-piloto trouxe sobre a maturidade da CONORF para deflagrar e participar do desenvolvimento de soluções de IA aplicada aos problemas orçamentários, frente às recomendações de melhores práticas na área emitidas pela OCDE.
5. Refletir sobre os desafios postos à CONORF para se preparar e se colocar em posição de explorar e tirar proveito crescente dos avanços em IA aplicada ao processo orçamentário.

3 METODOLOGIA

O delineamento do presente trabalho tem duas componentes. A primeira descreve o projeto-piloto de inteligência artificial (IA) sobre a classificação da receita orçamentária na forma do Demonstrativo do Resultado Primário. A segunda utiliza os achados do projeto-



piloto e os preceitos sobre inovação para refletir sobre a prontidão da CONORF a inovações como a do caso relatado.

A metodologia adotada no trabalho baseia-se em pesquisa qualitativa, de natureza descritiva e normativa, fundamentada em estudo de caso holístico (unidade única de análise) a partir de informações documentais e das experiências do autor deste TCC como consultor integrante da CONORF e co-autor do projeto-piloto, figurando, nesta condição, como observador-participante do caso relatado.

Segundo Yin (2015, p. 15), um objeto de interesse pode ser investigado a partir de vários métodos não mutualmente exclusivos. Cabe ao pesquisador identificar situações em que um método específico tem vantagens sobre as alternativas. O método do estudo de caso, segundo o autor, é especialmente adequado quando se investiga uma questão de “como” ou “porque” e a pesquisa versa sobre um conjunto de eventos contemporâneos. É precisamente a situação do presente trabalho.

O uso da observação direta do pesquisador como fonte de informações, ademais, é legítimo sempre que a observação participante proporcione “oportunidades incomuns para a coleta dos dados do estudo de caso” (YIN, 2015, p. 121), não disponíveis de outro modo. Ele chega a ressaltar que “essa perspectiva é valiosa na produção de um retrato precioso do fenômeno do estudo de caso” (YIN, 2015, p. 121).

Além das informações oriundas do pesquisador como observador-participante, o presente trabalho utilizou informações obtidas em conversas frequentes com o outro co-autor do projeto-piloto relatado, o Analista de Sistemas do PRODASEN, Marco Túlio de Carvalho.

4 FUNDAMENTAÇÃO TEÓRICA

4.1. Inteligência artificial e aprendizado de máquina

O conceito de IA ainda não é um consenso, mesmo para especialistas. Isso porque a IA se expressa em vários campos, sendo empregada em múltiplas possibilidades e perspectivas. A Estratégia Brasileira de Inteligência Artificial-EBIA² define a IA como “um conjunto de técnicas destinadas a emular alguns aspectos da cognição de seres vivos usando máquinas” (BRASIL, 2021, p. 4). Embora as pessoas tenham a intuição de que o conceito se

² A Estratégia Brasileira de Inteligência Artificial é um documento elaborado pelo Ministério da Ciência, Tecnologia e Inovações, alinhado com as diretrizes da OCDE, que tem como objetivo nortear o desenvolvimento de ações voltadas à pesquisa, inovação e desenvolvimento de soluções em IA.



relaciona com a possibilidade de uma máquina desenvolver inteligência, não é simples entender como isso é possível na realidade.

Historicamente, o desenvolvimento da IA pode ser dividido em três períodos de florescimento. No primeiro, entre 1950 e meados da década de 1970, a IA seguia a abordagem baseada na lógica simbólica e os projetos eram voltados ao raciocínio lógico, reconhecimento de caracteres e processamento de linguagem natural. O segundo período, na década de 1980, teve como tendência o desenvolvimento de sistemas baseados em conhecimento. Em síntese, o esforço se concentrava na criação de programas de computador que respondiam ou resolviam problemas sobre um assunto específico, tendo como parâmetro regras lógicas derivadas do conhecimento de especialistas. As principais dificuldades dessa perspectiva eram as capacidades de aprendizado e de abstração (OECD, 2019a, p. 20).

O aprendizado estatístico caracteriza o terceiro período, que teve início na década de 1990. A partir do desenvolvimento tecnológico, da popularização dos computadores e da criação da Internet, a geração e disponibilização de grande quantidade de dados tornou-se um terreno profícuo para o emprego de técnicas estatísticas na análise de dados e possibilitou a superação dos principais obstáculos que marcaram o período anterior (OECD, 2019a, p. 20).

A Organização para a Cooperação e Desenvolvimento – OCDE tem se dedicado a discutir diversos aspectos da IA nos últimos anos. O resultado disso aparece na grande quantidade de publicações de referência sobre o assunto. A OCDE criou os princípios para uso inovador e confiável da tecnologia que respeite os direitos humanos e os valores democráticos (OECD, 2019b) e as recomendações para os formuladores de políticas relativas às políticas nacionais e cooperação internacional para IA confiável (OECD, 2021).

A OCDE define uma aplicação de IA como "um sistema baseado em máquina que pode, para um determinado conjunto de objetivos definidos pelo homem, fazer previsões, recomendações ou decisões que influenciam ambientes reais ou virtuais" (OECD, 2019a, p. 23, tradução nossa). Como essa ainda é uma definição ampla, este trabalho se restringirá às soluções de TI associadas a aprendizado de máquina.

Para Shah, o aprendizado de máquina é um ramo da IA e consiste “em habilitar computadores e outros sistemas artificiais para que aprendam sem programá-los explicitamente” (Shah, 2020, p.146, tradução e grifo nosso). O sistema deve acessar alguns dados, que chamamos preditores, e aprender com eles. Então, com base nesse conhecimento, inferir coisas, (resultados), a partir de dados ainda não vistos. Essa inferência caracteriza a abstração. Com o aprendizado e a abstração, torna-se possível transformar dados em informação. Em outro trecho ele explica que “as coisas aprendem quando mudam seu



comportamento de maneira que melhore seu desempenho no futuro” (Shah, 2020, p. 211, tradução nossa).

De acordo com a OCDE, o aprendizado de máquina “consiste num conjunto de técnicas que permitem que as máquinas aprendam de maneira automatizada, sem instruções explícitas de um ser humano, contando com padrões e inferências”. As máquinas são ensinadas a alcançar um resultado ao serem expostas a muitos exemplos de resultados corretos, um processo que é chamado de treinamento (OECD, 2019^a, p. 27). Em resumo, o processo de aprendizado de máquina consiste em gerar conhecimento diante de novos dados, com base no conhecimento aprendido de dados previamente observados, fundamentado na estatística. Os dados utilizados no aprendizado de máquinas podem ser de diversos tipos, como número, texto, som, imagem ou vídeo.

Olckers A. (2017) faz uma analogia entre aprendizado de máquina e a forma como uma criança aprende. Segundo ele, os pais não costumam utilizar definições científicas para ensinar aos filhos como diferenciar gatos e cachorros. Em vez disso, as crianças são apresentadas a muitas imagens desses animais até compreenderem os dois conceitos. A criança vê um animal e diz que é um cachorro. Os pais confirmam a resposta. A criança vê outro animal e diz que também é um cachorro, então os pais a corrigem, afirmando que não é um cachorro, é um gato. Assim, o aprendizado vai se reforçando.

Cães e gatos possuem algumas características em comum, são quadrúpedes e têm rabos, isso às vezes confunde as crianças. Mas, com o tempo, ela começa a entender intuitivamente as características específicas de cada um dos animais. Por exemplo, gatos têm orelhas pontudas e o focinho é, proporcionalmente à cabeça, muito menor que o do cachorro. Esse processo é retroalimentado com os acertos e a correção de erros. O conhecimento aumenta e a probabilidade de novas classificações equivocadas diminui.

Como seres humanos, computadores podem ser ensinados de forma similar por meio do chamado aprendizado de máquina. Para isso, primeiro é preciso uma base de dados com vários exemplos desses animais já classificados, o que constitui a base de aprendizagem. Apoiada nesses dados, a máquina apreende as características semelhantes dos indivíduos da mesma espécie. Então, novas imagens são apresentadas e formam a base de teste, para que o computador possa inferir a classificação dos indivíduos dessa base (OLCKERS, 2017). A inferência não é realizada por meio de definição de regras de classificação, mas fundamentada em conceitos matemáticos e estatísticos. Os erros cometidos pelo processo de inferência são ajustados e as novas classificações são incorporadas à base de aprendizado, enriquecendo o conhecimento da máquina.



Aqui cabe uma observação importante sobre a diferença entre o aprendizado humano e o de máquina. Quando a criança aprende a diferenciar gatos e cachorros, existe uma cadeia de conhecimentos subjacentes. Os dois são animais, o gato mia, o cachorro late, eles gostam de carinho, o cachorro pode morder e pode existir também um sentimento afetivo da criança com relação a esses animais. Já no aprendizado de máquina, o conhecimento ainda é restrito. A máquina analisa a distribuição dos pixels da imagem e a classifica, de acordo com as semelhanças, como cão ou gato. Nesse caso, o aprendizado é focado na resolução de um problema específico, que é de classificação, com baixa contextualização em relação ao aprendizado humano.

Quanto ao tipo, o aprendizado de máquina pode ser supervisionado ou não supervisionado. No primeiro caso, o aprendizado baseia-se em como os preditores e os resultados se relacionam. Preditores e resultados são atributos, também chamados de variáveis, previamente conhecidos na base de dados de treinamento. Imagine que se pretenda inferir a receita do Imposto de Renda Pessoa Física – IRPF do próximo ano – o atributo de resultado. Dados sobre a receita do IRPF de outros anos constam da base de treinamento. Também constam da base de treinamento os atributos que têm correlação com o resultado, os preditores, que, no caso do aprendizado de máquina supervisionado para inferir o IRPF, poderiam ser o PIB, a inflação e a taxa de desemprego em diversos anos. Na base de teste há apenas preditores, omitindo-se atributos de resultado, que são exatamente o que se deseja prever.

Então, no aprendizado supervisionado, primeiro é necessário conhecer a relação entre os preditores (PIB, inflação, taxa de desemprego em diversos anos) e os resultados (IRPF de cada ano) em uma base de dados de treinamento e, depois, aplicar essa relação para inferir novos resultados na base de teste, que contém amostras não vistas na fase de treinamento (PIB, inflação e taxa de desemprego projetados para o ano seguinte). Assim, na base de teste há os preditores, mas não estão disponíveis os atributos de resultado, e nela não há um conhecimento prévio da relação entre preditores e resultados. A inferência é gerada automaticamente na base de teste de acordo com as características e os padrões dos preditores e resultados aprendidos na base de treinamento (SHAH, 2020, p. 146). O modelo aprende com a base de treinamento e aplica o conhecimento adquirido para inferir o resultado na base de teste.³

³ Há aqui uma simplificação, para fins didáticos, do aprendizado de máquina supervisionado. Vale mencionar que existem várias técnicas para realizar esse aprendizado.



No caso do aprendizado de máquinas não supervisionado, não há atributos preditores nem atributos de resultado da forma do aprendizado supervisionado. Na base de treinamento há apenas atributos e o algoritmo agrupa os dados em clusters com base em algum tipo de semelhança por ele identificada. A quantidade de clusters é definida pela equipe, que, após o procedimento de criação dos clusters, analisa se o resultado obtido representa um conhecimento significativo. A quantidade de clusters pode ser alterada em uma nova rodada e a análise é realizada novamente. O aprendizado de máquinas não supervisionado é bastante utilizado para encontrar *outliers*, dados que se diferenciam significativamente da normalidade e, também, para encontrar padrões ocultos na base. Para exemplificar, o aprendizado não supervisionado poderia ser utilizado na avaliação de padrões a partir dos textos constantes do campo “observação” dos empenhos associados a uma subfunção orçamentária, a fim de destacar características da execução orçamentária de uma determinada política pública, invisíveis a análises baseadas nas classificações orçamentárias clássicas.

Uma das aplicações do aprendizado de máquina supervisionado é a de classificação. Nela, o objetivo consiste em classificar os dados da base de teste em classes apreendidas pela exposição do modelo preditivo a uma base de treinamento. A diferenciação entre gatos e cães descrita anteriormente é um exemplo de aprendizado de máquina supervisionado para tarefa de classificação, em que a tarefa é classificar imagens em uma das duas classes: cães ou gatos. Segundo Shac (2020, p. 173), a tarefa de classificação consiste em, a partir de um conjunto de dados e suas classes correspondentes, aprender as regras e padrões dessas classificações, para que seja possível inferir a classe correta em um novo conjunto de dados. O projeto de classificação da receita nos moldes do Demonstrativo do Resultado Primário, que será descrito e analisado neste trabalho, pertence à categoria de aprendizado de máquina supervisionado aplicado à classificação.

Um dos elementos constituintes dos processos de desenvolvimento da IA é o atributo da inovação, referido nas definições acima e fator essencial para uma organização que pretenda se destacar no uso dessa tecnologia. A próxima sessão discutirá esse requisito.

4.2 Habilidades para inovação segundo a OCDE

Fawcet e Provost (2013, p. 313) afirmam que dois fatores são importantes para uma empresa obter o máximo da riqueza dos dados. O primeiro é que a administração da empresa deve pensar em dados analiticamente. O segundo está relacionado à criação, pela



administração, de uma cultura em que a ciência de dados prospere e os cientistas de dados tenham espaço para atuação. Uma cultura empresarial particularmente adequada à implantação de soluções de ciência de dados deve ser entendida como um ativo intangível estratégico, que proporciona vantagens para a empresa em relação a competidores (FAWCET e PROVOST, 2013, 315 e 318). A ideia da cultura voltada à ciência de dados como um ativo intangível também se aplica ao setor público, no sentido de que o conhecimento é base para aperfeiçoamento dos serviços.

A OCDE publicou em 2017 a segunda versão das “Principais Habilidades para Inovação no Setor Público” (OECD, 2017, tradução nossa). Nesse documento são apresentadas seis principais habilidades para promover a inovação do setor público: alfabetização (ou literacia) de dados; iteração; curiosidade; insurgência; contação de história (narrativa); e centralidade no usuário.

Embora tenham sido elaboradas para a inovação do setor público em termos gerais, as habilidades identificadas possuem ampla aderência a projetos de IA. As sessões seguintes detalham cada habilidade de inovação descrita pela OCDE e apontam o alinhamento dessas habilidades com a IA.

4.2.1 Alfabetização de dados

De acordo com a OCDE (OECD, 2017, pp. 12-13), a habilidade de alfabetização de dados compreende o reconhecimento do valor e da importância dos dados, a capacidade de trabalhar com um especialista em dados e a consciência de que decisões devem ser fundamentadas em dados e evidências, e não em suposições. Um forte relacionamento entre a equipe de negócio e a equipe de especialista em dados pode conduzir ao uso dos dados no momento certo para tomar a decisão correta. A alfabetização de dados permite uma maior integração nesse relacionamento.

Outros autores argumentam sobre a importância da alfabetização de dados para o pessoal da área do negócio no processo de desenvolvimento de soluções de IA. Shah relata que o treinamento em alfabetização de dados para a área do negócio é uma prioridade (HARRIS, 2012 apud SHAH, 2020 p. 22). Para ele, a tomada de decisão baseada em dados é uma força motriz para a inovação nos negócios. A área do negócio deve ser treinada para entender quais dados são adequados e como usar a visualização e simulação para processar e



interpretar os dados. Shah (2020, p.22) define a alfabetização de dados como “a habilidade de extrair informações significativas de um conjunto de dados”.

Fawcet e Provost (2013, p. 313) defendem que é importante entender a ciência de dados, mas isso não significa que o pessoal da área de negócio precise ser cientista de dados. Afirmam que compreender conceitos fundamentais e dominar estruturas para organizar o pensamento analítico de dados permitem não apenas a interação com competência, mas também ajudam a melhorar a tomada de decisão baseada em dados. A área do negócio deve se apropriar desses conceitos para avaliar oportunidades de emprego da ciência de dados, fornecer recursos apropriados para as equipes de ciência de dados e investir em dados e experimentos. Além disso, eles sustentam que a área de negócio deve estar preparada para orientar a equipe de ciência de dados de modo a garantir que ela permaneça no caminho certo para uma solução de negócio efetivamente útil. Essa capacidade só é possível se a equipe de negócio realmente dominar os conceitos do pensamento analítico de dados.

Outro aspecto relevante defendido por eles é que o domínio de uma base sólida nos fundamentos da ciência de dados tem implicações estratégicas de longo prazo. À medida que a área do negócio aumenta sua exposição a projetos desse tipo, vislumbra, cada vez mais, outras oportunidades para o emprego de IA. Uma vez que certa capacidade de IA tenha sido desenvolvida, outras aplicações tornam-se óbvias. Eles citam a famosa frase de Louis Pasteur ‘o acaso favorece a mente preparada’ (FAWCET e PROVOST, 2013, p. 314).

A ciência de dados consiste na combinação de vários campos do conhecimento, inclusive a IA, com o objetivo de extrair valor dos dados. Shah C. (2020) define a ciência de dados como um campo de estudo e prática que envolve coleta, armazenamento e processamento de dados para obter percepções importantes sobre um problema ou fenômeno. Esses dados podem ser gerados por humanos ou máquinas e podem estar em diferentes formatos, como texto, áudio, vídeo, etc. Citando Frank Lo, ele afirma que a ciência de dados, em sua essência, envolve a descoberta de *insights* a partir da mineração de dados. Isso acontece por meio da exploração dos dados com a utilização de várias ferramentas e técnicas, testando hipóteses e criando conclusões a partir de evidência baseada em dados e análises (LO apud SHAH, 2020 p. 4).

A alfabetização de dados não deve ser vista somente sobre o ponto de vista da área de negócio. Em projetos de IA, a capacidade de especialistas em dados de se comunicar com os não-especialistas da área de negócio sobre os resultados da análise é tão importante quanto coletar e analisar os dados (OECD, 2017, p. 13).



4.2.2 Iteração

A habilidade de iteração está relacionada ao desenvolvimento incremental e progressivo de um projeto. Além da aplicação no gerenciamento e execução de projetos, a iteração também é associada à utilização de protótipos e experimentos. A OCDE define protótipo como aquilo que um produto, serviço, sistema, ou política possa parecer. Eles são utilizados na verificação se algo funciona e também como versões iniciais, que após incrementos e refinamentos, passam a constituir um produto final. São úteis para exercícios internos como prova de conceitos e também como testes. Testes e experimentos fornecem uma maneira robusta de avaliar se a abordagem funciona. Abordagens iterativas e incrementais limitam os riscos associados a testes de condutas ou de métodos que ainda não foram utilizados (OECD, 2017 p. 10 e 11).

Schmitt (2020b) propõe uma sistemática simples e objetiva para encontrar uma abordagem para a aplicação de IA que funcione: 1) pesquise artigos acadêmicos, 2) construa uma base de dados, 3) defina uma métrica de avaliação e 4) Experimente até achar um bom desempenho. Ele sugere que a experimentação comece a partir de uma abordagem mais promissora, seja avaliada com base nos resultados e aperfeiçoada. Então, deve-se repetir o procedimento (iterar) até encontrar uma abordagem que seja boa o suficiente.

De acordo com Ubaldi *et al* (2019, p.10), sistemas de IA incorporam várias fases do ciclo de vida comuns às dos softwares tradicionais. Porém, no desenvolvimento de projetos de IA, as fases ocorrem de forma iterativa e não são necessariamente sequenciais. “A experimentação e o aprendizado iterativo são cruciais para o desenvolvimento de capacidades em IA”. Se não houver liberdade para experimentar novas formas de desenvolver e fornecer serviços, é improvável que o potencial da IA seja alcançado (BERRYHILL *et al*, 2019A, p. 93, tradução nossa).

Fawcet e Provost (2013, p. 34) argumentam que a visão de um projeto de IA nos moldes dos processos de desenvolvimento de softwares é um erro. A aplicação em IA (eles citam a mineração de dados como exemplo) é um empreendimento exploratório mais próximo da pesquisa e desenvolvimento do que da engenharia de software. Os efeitos são muito menos certos e os resultados de uma determinada etapa podem mudar a compreensão fundamental do problema. A relevância do processo de iteração reside nesses fatos. Projetar uma solução de IA diretamente para implantação pode ser um compromisso prematuro e



caro. Em vez disso, eles defendem que projetos devem investir em informações para reduzir a incerteza.

Pequenos investimentos podem ser feitos por meio de projetos-piloto e protótipos. Para eles, é importante que as equipes sejam capazes de formular bem os problemas, prototipar soluções rapidamente, fazer suposições razoáveis diante de problemas mal estruturados, projetar experimentos e analisar os resultados (FAWCET e PROVOST, 2013, p. 35).

4.2.3 Curiosidade

Inovação envolve invenção, criar coisas novas e fazer algo de forma diferente. Curiosidade e pensamento criativo fazem parte da essência da inovação. É necessário entender que a forma atual de trabalhar é apenas uma abordagem possível e cada abordagem tem vantagens e desvantagens. Técnicas de reformulação permitem pensar um problema sobre perspectiva diferente ou desafiar suposições padronizadas. A habilidade de curiosidade também envolve descobrir como outras pessoas fazem um trabalho semelhante de forma diferente. Não existe uma maneira certa de abordar um problema ou situação. Além disso, investigar um problema somente por uma perspectiva pode ocultar questões e oportunidades (OECD, 2017 p. 16 e 17).

Para Fawcet e Provost (2013, p. 26, tradução nossa), “a mineração de dados é um trabalho artesanal. Ela envolve a aplicação de quantidade substancial de ciência e tecnologia, mas a aplicação adequada ainda envolve arte”. Também, no processo de desenvolvimento de uma solução de IA, o estágio do entendimento do negócio representa uma parte artesanal onde a criatividade desempenha um grande papel. Para esses autores, a chave para o sucesso está na formulação criativa de um problema em ciência de dados a partir do problema do negócio (FAWCET e PROVOST, 2013, p. 28).

Outra característica importante da curiosidade e criatividade é a adoção de uma mentalidade de aprendizado contínuo, ser capaz de absorver e usar novas ideias, identificar suas próprias limitações e aprender mais sobre elas. Novos conhecimentos são produzidos o tempo todo e podem nos desafiar e nos forçar a ‘desaprender’ coisas que já sabemos ou fazemos (OECD, 2017). Alinhado a esta característica da habilidade de curiosidade, Schmitt (2020b) sugere pesquisar a forma como outras equipes resolveram tarefas semelhantes,



envolvendo ou não o aprendizado de máquina. Também destaca a importância de consultar artigos acadêmicos.

Imaginar possibilidades para reaproveitar soluções já utilizadas em outros projetos também configura uma atitude de criatividade. Uma habilidade crítica em ciência de dados é a capacidade de decompor um problema de análise de dados em partes, de forma que cada parte corresponda a uma tarefa conhecida para a qual existem ferramentas disponíveis. Reconhecer problemas familiares e suas soluções evita perder tempo e recursos reinventando a roda. Esse reaproveitamento permite que as pessoas focalizem a atenção em partes mais interessantes do processo que requerem envolvimento humano. Para as partes que não foram automatizadas, a criatividade e a inteligência humanas devem entrar em ação (FAWCET e PROVOST, 2013, p. 20).

4.2.4 Insurgência

Fazer coisas novas acontecerem caracteriza a habilidade da insurgência. Para isso, é preciso que a equipe tenha liberdade para experimentar formas alternativas de fazer as coisas, inclusive aquelas que podem não funcionar. Esse comportamento pode levar a maneiras de trabalhar que de outra forma permaneceriam ocultas. Tentar algo novo e verificar que não funciona não deve ser considerado um fracasso, mas uma oportunidade de aprendizado, ninguém deve ser culpado pelo não-funcionamento de uma tentativa de resolver um problema (OECD, 2017 p. 20 e 21).

Olhar com estranheza o que é feito e como é feito no dia a dia é fundamental. O contentamento com os processos atuais sem que haja questionamentos conduz à manutenção do *status quo*. Essa atitude se aplica a projetos de IA. É preciso pensar diferente para conceber soluções eficientes a fim de criar novos processos ou automatizar os existentes. A visão comum dificilmente vislumbraria as oportunidades. A insurgência, nesse caso, está associada à percepção das possibilidades de melhorias que podem advir da automação. Também está associada à perseverança para a realização do projeto e à vontade de mudar. O projeto de classificação da receita na forma do Demonstrativo do resultado primário originou-se exatamente desse olhar de estranheza sobre um trabalho, que era realizado manualmente no processo legislativo orçamentário.

Outra vertente da insurgência é a consciência de que trabalhar sozinho raramente leva à resolução dos problemas. Os desafios demandam colaboração e alianças entre



parceiros internos e externos a fim de oferecer sinergias que favoreçam inovações (OECD, 2017, p. 20). O trabalho em ciência de dados é extremamente dependente de uma equipe multidisciplinar e do compartilhamento de informações.

4.2.5 Contação de história (narrativa)

Compartilhar experiências é um componente crucial para a inovação no setor público. Entre outras possibilidades, a habilidade de contação de história pode ser utilizada para delinear o futuro e para inspirar ação de outras pessoas. A mudança no setor público deve ser pensada como um ato contínuo e não simplesmente como uma transposição de um estado estático A para o estado estático B. A contação de história é um relevante fator de transformação contínua de ambientes operacionais, expectativas e necessidades do usuário. A inovação é impulsionada pelo intercâmbio de conhecimentos e práticas. As histórias são úteis para compartilhar experiências (OECD, 2017 p. 18 e 19).

Além disso, segundo a OCDE, é necessária uma narrativa clara para garantir e manter o apoio e minimizar a resistência para implementação da IA no setor público. A narrativa deve explicar “como a IA pode ajudar os servidores a prestar melhores serviços, reduzir a quantidade de tempo que gastam em tarefas de rotina e permitir que foquem em tarefas de maior valor” (BERRYHILL, J. *et al*, 2019, p. 92, tradução nossa).

Schmitt (2020c) relata sobre a motivação em um projeto de IA. Para ele, será difícil para a equipe fazer boas sugestões se não entender o motivo do projeto, por isso, é necessário mostrar porque o projeto é importante, qual problema deve ser resolvido, qual é objetivo estratégico. É preciso ser claro sobre o significado amplo do projeto. As histórias falam sobre a importância do projeto, as equipes têm que estar convictas disso.

4.2.6 Centricidade no usuário

Ser centrado no usuário é descobrir o que os usuários precisam e criar uma política ou um serviço que atendam essas necessidades. O governo não deve presumir a necessidade dos usuários, ele deve criar oportunidade para que eles se manifestem. Também é necessário considerar a diversidade dos usuários e garantir a facilidade da utilização dos produtos oferecidos (OECD, 2017 p. 14 e 15).



Na próxima seção, discutem-se os princípios aqui revisados ao desenvolvimento de sistemas de IA.

4.3 Desenvolvimento de sistemas de IA

Schmitt (2020a) argumenta que softwares tradicionais são construídos com base na dedução. Um conjunto de regras é criado e codificado, para depois ser aplicado aos dados. Já as soluções de IA que utilizam o aprendizado de máquina são elaboradas em torno da indução. O algoritmo descobre as regras automaticamente, tomando como base um grande número de exemplos, e as aplica em outros dados. Por isso, ele considera o aprendizado de máquina como ‘aprendizado de dados’.

Em outro artigo, ele explica que, no desenvolvimento de um software tradicional, primeiro é necessário responder à questão do que se quer implementar para, em seguida, implementar a solução. Nos projetos de aprendizado de máquinas, é necessário explorar primeiro o que é possível realizar com as bases de dados acessíveis. Então a pergunta é deslocada para o que se pode implementar a partir dessas bases (SCHMITT, 2020a).

Essas diferenças de concepções têm implicações no processo de criação de cada um desses tipos soluções e no escopo de inovações passíveis de serem desencadeadas. A Constituição de 1988 devolveu ao Congresso Nacional a possibilidade de emendar o orçamento. No início, havia dificuldade entre os parlamentares e gabinetes de apoio para indicar as classificações orçamentárias das emendas a serem apresentadas. Essa dificuldade acarretava problemas, por parte das Consultorias de Orçamentos do Senado Federal e da Câmara dos Deputados, na análise das mais de dez mil emendas propostas por ano. Por isso, as Consultorias, em parceria com o PRODASEN, criaram o sistema de emendas, com tecnologia de software tradicional. A aplicação serviria como um tutorial⁴ para classificar orçamentariamente as emendas. Essa definição seria a resposta inicial para o desenvolvimento do sistema. Os passos seguintes compreenderiam definir e colher as informações necessárias para atingir o objetivo. Como resultado, decidiu-se que essas informações seriam organizadas anual e manualmente pelas Consultorias e teria uma abrangência limitada. O sistema foi implementado e ainda é utilizado.

Em uma solução de IA, o objetivo seria o mesmo, a classificação das emendas. No entanto, após a definição do problema, a pergunta inicial para o desenvolvimento da aplicação seria: as informações constantes das bases de dados orçamentárias são suficientes

⁴ Definido como Geratriz.



para um algoritmo aprender a relação entre os textos das emendas dos anos anteriores e as classificações estabelecidas pelos pareceres e, com base nesse aprendizado, atribuir classificação a uma nova emenda apresentada? Se a resposta for negativa, o desenvolvimento da aplicação deve parar até quando novas bases de dados com as informações necessárias estejam disponíveis.

A OCDE (2019a, p. 25) reforça esse entendimento. Segundo a Organização, os sistemas de IA incorporam muitas fases do ciclo de vida dos softwares tradicionais. No entanto, as fases do ciclo de desenvolvimento de aplicações em IA geralmente não são necessariamente sequenciais e ocorrem de forma iterativa. Além disso, “a centralidade dos dados e dos modelos que dependem dos dados para treinamento e avaliação distingue o ciclo de vida de muitos sistemas de IA do ciclo de desenvolvimento de sistemas mais gerais. Alguns sistemas de IA baseados em aprendizado de máquinas podem iterar e evoluir com o tempo” (OCDE, 2019a, p. 26, tradução nossa).

Para a Organização, as fases do ciclo de vida dos sistemas de AI são: 1) desenho, dados e modelagem, que inclui: planejamento e desenho; coleta e processamento de dados; e construção do modelo e interpretação; 2) verificação e validação; 3) emprego; e 4) operação e monitoração (OCDE, 2019a, p. 26).

Embora as fases do ciclo de vida dos sistemas de IA definidas pela OCDE sejam mais recentes, o documento de referência sobre o processo de modelagem é o “CRISP-DM 1.0: Step-by-step data mining guide” (CHAPMAN, P. *et al*, 2000). Segundo o CRISP-DM, o ciclo de modelagem do processo de IA (tratado no documento como mineração de dados) tem seis fases: 1) conhecimento do negócio, 2) conhecimento dos dados, 3) preparação dos dados, 4) modelagem, 5) avaliação, e 6) emprego. Fawcet e Provost (2013, p. 34) argumentam que, embora o ciclo CRISP-DM possa parecer similar ao ciclo de desenvolvimento de softwares, é um erro pensar que são semelhantes. Para eles, o projeto de mineração de dados é um empreendimento baseado na exploração, iteração de abordagens e estratégias mais próximo da pesquisa e desenvolvimento do que da engenharia, que é própria de *designs* de software.

Apesar da diferença de nomenclatura das fases, o ciclo de desenvolvimento da OCDE e o CRISP-DM, na essência, são muito parecidos e podem ser ambos considerados como referências para definição de metodologias de planejamento e execução de projetos de IA em organizações públicas, como a CONORF e o PRODASEN. O ciclo de desenvolvimento do projeto-piloto de classificação da receita, adiante relatado, é muito próximo dessas duas metodologias.



A consciência da diferença entre processo de criação de soluções tradicionais de TI e aplicações IA é fundamental. No entanto, como se discutirá adiante, a experiência do projeto-piloto mostrou que vários requisitos devem ser considerados nos projetos de desenvolvimento de aplicações em IA, em adição àqueles preconizados pelos métodos mencionados anteriormente. Os mais relevantes, segundo preceitos internacionalmente aceitos, são apresentados a seguir.

4.3.1 O valor e a qualidade dos dados

Base de dados é o ponto central de um projeto de IA. Sem qualidade da base é impossível conseguir resultados consistentes. Por isso, muitos pesquisadores entendem que a base de dados merece ser tratada como um ativo intangível da organização, pois tem potencialidade para gerar valor. Mas, para isso, é necessária uma visão diferenciada com relação a esse bem intangível, especialmente a disposição de investir para aprimorá-lo continuamente.

Para Schmitt (2020a), a vantagem competitiva de uma empresa são os dados e não os algoritmos. Empresas como o Google e o Facebook abrem o código-fonte de seus algoritmos, mas mantêm seus dados em segredo. A fórmula do sucesso para produzir valor em ciência de dados é escolher o problema certo aliado à decisão de que dados devem ser utilizados e ao bom tratamento dos dados. Na mesma direção, Fawcett e Provost (2013, p. 9 e 315) entendem que os dados e a capacidade de extração de conhecimento útil dos dados devem ser considerados ativos estratégicos complementares.

Mas, para extrair valor dos dados, é preciso que eles tenham qualidade e isso requer, na maioria das vezes, muito trabalho. Nas duas metodologias sobre o ciclo de desenvolvimento da IA vistas anteriormente, há previsão para coleta, compreensão, tratamento e estruturação dos dados. Schmitt (2020a) entende que executar um projeto sem que os dados estejam limpos e organizados é uma armadilha.

Segundo a OCDE, os dados são o combustível da ciência de dados. Eles são o ponto crítico da maioria dos projetos de IA, especialmente para projetos de aprendizado de máquina, em que o objetivo é aprender com os dados. Uma estratégia de dados clara que permita acessar dados robustos e precisos é necessária para implantar IA de maneira eficaz (BERRYHILL, J. *et al*, 2019, p. 30 e 89).

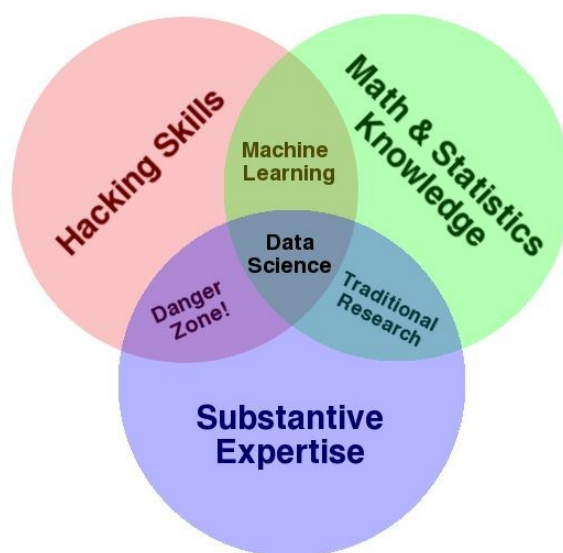


Guardadas as particularidades, o setor público deve ter a consciência do valor das bases de dados e dos projetos de IA sob ótica semelhante à do setor privado. Fawcet e Provost (2013, p. 316) explicam que o valor de um ativo para empresa depende de decisões estratégicas. Por isso, eles afirmam que, para obter uma vantagem competitiva, é estratégico pensar tanto os dados quanto a capacidade de ciência de dados como ativos.

4.3.2 Multidisciplinariedade e equipe

De acordo com Fawcet e Provost (2013, p. 9, tradução nossa) “A melhor equipe de ciência de dados pode render pouco valor sem os dados apropriados; os dados certos muitas vezes não podem melhorar substancialmente as decisões sem um talento adequado em ciência de dados”. A qualidade dos dados é importante, mas extrair valor dos dados está no mesmo nível de relevância e isso requer habilidades. Com o objetivo de descrever as habilidades necessárias de um de cientista de dados ideal, Conway (2010) apresentou o “Diagrama de Venn da Ciência de Dados”.

Figura 1 - Diagrama de Venn da Ciência de Dados



Fonte: Conway (2010)

Segundo o diagrama, a característica ideal de um cientista de dados é representada pela combinação de três habilidades necessárias para extrair valor dos dados: 1) expertise na área de negócio; 2) conhecimento matemático e estatístico; e 3) habilidades de hacker, que



Conway (2010) entende como conhecimentos sobre manipulação de arquivos de texto na linha de comando, operações vetorizadas e pensamento algorítmico. Em resumo, as habilidades de hacker estariam associadas à familiaridade com atividades operacionais típicas de TI.

No entanto, é raro encontrar alguém com características que a posicionem no centro da interseção do diagrama. Por isso, a literatura recomenda a formação de equipe multidisciplinar para trabalhar em projetos de IA, em que os membros tenham conhecimentos representados pelas áreas de interseção do diagrama. A OCDE vai além. Ela defende que a equipe de IA deve ser multidisciplinar, diversa e inclusiva e explica que esses atributos são, provavelmente, o maior fator de capacitação para se alcançar soluções em IA que sejam eficazes, éticas, bem-sucedidas e justas. Essa abordagem deve ser empregada, tanto para iniciativas de estratégias abrangentes de âmbito nacional, como para pequenos projetos de IA (BERRYHILL, J. *et al*, 2019, p. 101).

A multidisciplinariedade do time se caracteriza pelas diferenças em formações educacionais, experiências e níveis profissionais, conjunto de habilidades e outras. Os esforços em IA precisam ser tecnologicamente viáveis, mas também devem ser permitidos por lei e aceitáveis pelas partes interessadas, inclusive o público. Por isso, projetos de IA têm dimensão multidisciplinar, requerem considerações tecnológicas, éticas, legais, políticas, entre outras (BERRYHILL, J. *et al*, 2019, p. 101).

A diversidade e inclusão focam nas diferenças de gênero, raças, idades, origens socioeconômicas e outras, em um ambiente onde as opiniões são valorizadas. Equipes diversificadas e inclusivas podem prevenir ou eliminar vieses desde o início do projeto. Além disso, pesquisas demonstram que equipes com diversidade produzem resultados significativamente melhores (BERRYHILL, J. *et al*, 2019, p. 101 e p. 103 *apud* OCDE, 2019a).

A formação de uma equipe multidisciplinar, diversa e inclusiva não é simples nem rápida. Ela pode ser desenvolvida internamente ou criada com recursos externos, por meio de contratos ou parcerias com a iniciativa privada ou outros órgãos do governo. Ela pode também ter uma composição mista.

4.3.3 Desenvolvimento progressivo de capacidades

Alguns projetos de IA requerem grande aporte de recursos, outros não. A OCDE alerta sobre a necessidade de uma infraestrutura sólida, robusta e flexível para a utilização de



IA no setor público (BERRYHILL, J. *et al*, 2019, p. 131). Por isso, deve-se sempre estar atento para a magnitude e custo do projeto.

Projetos de IA não necessariamente são grandes e dispendiosos. Uma abordagem possível é o desenvolvimento progressivo de capacidades por meio de pequenos projetos. A experiência com esses projetos conduz ao amadurecimento do órgão para lidar com projetos maiores. Para isso, é necessário planejamento e dimensionamento corretos de um programa que contenha uma continuidade nos projetos de IA.

A próxima seção descreverá o desenvolvimento do projeto-piloto de IA aplicada ao Demonstrativo do Resultado Primário da União, que trará elementos de aprendizado sobre a prontidão da CONORF à inovação, discutida na sequência.

5 ESTUDO DE CASO: CLASSIFICAÇÃO DA RECEITA NA FORMA DO DEMONSTRATIVO DO RESULTADO PRIMÁRIO

Há alguns anos a CONORF começou a prospectar formas de inserção de soluções de IA em seus processos de trabalho. Em parceria com o PRODASEN, após avaliação de vários assuntos potenciais para aplicação da tecnologia, decidiu-se criar o projeto-piloto de IA sobre a classificação da receita na forma do Demonstrativo do Resultado Primário, que teve início em abril de 2020. O projeto-piloto serviria para a formação de conhecimento e experimentação e para se tornar berço de inúmeros outros projetos a serem desenvolvidos no futuro.

A escolha fundamentou-se na necessidade de resolver um problema existente no processo de elaboração orçamentária e em estudos prévios sobre as possibilidades de utilização de IA para solução do problema.

5.1 Entendimento do negócio e definição do problema

Após longo período de desequilíbrio nas contas públicas, elevado endividamento e crescimento acelerado da inflação e dos juros, o Brasil adotou, no final dos anos 90, novo regime fiscal e monetário conhecido como Plano Real. Um dos principais pilares do sistema de ajuste macroeconômico adotado foi o estabelecimento de metas anuais de resultado primário, que funcionariam como condicionante da gestão das receitas e despesas públicas. A partir da adoção do regime previsto na Lei de Responsabilidade Fiscal, o cálculo do resultado



primário passou a ser informação crítica de elevado impacto para o processo de tomada de decisões sobre as contas governamentais.

O cálculo do resultado primário consiste basicamente na diferença entre a receita e a despesa primária. No entanto, a classificação da receita, na forma que consta do Demonstrativo do Resultado Primário, não é simplória e exige um esforço da equipe de consultores de orçamento para a aferição no projeto de lei encaminhado pelo Executivo e para a elaboração do autógrafo, no caso de haver reestimativa da receita no processo orçamentário. Esses procedimentos têm sido, historicamente, realizados de forma assistemática e em base manual, com elevado consumo de tempo e trabalho humano. Por essa razão, o processo está sujeito à ocorrência de erros que podem ter consequências graves para a gestão orçamentária. Por fim, tem-se verificado também comprometimento da transparência orçamentária e limitações na participação democrática sobre as finanças públicas.

Os Demonstrativos do Resultado Primário, pelo critério acima da linha, tanto da elaboração quanto da execução do orçamento, utilizam uma classificação da receita cuja fórmula de agregação não é descrita em nenhum documento do governo e não consta das bases de dados dos principais sistemas orçamentários e financeiros do Governo Federal recebidas pelo Congresso Nacional, em especial do Sistema Integrado de Planejamento e Orçamento – SIOP (elaboração do orçamento) e do Sistema Integrado de Administração Financeira – SIAFI (execução do orçamento). Existe, para consulta no SIOP, a classificação NFGC, pela qual é possível derivar a classificação utilizada nos Demonstrativos de Resultado Primário. No entanto, essa classificação, realizada manualmente pela a Secretaria de Orçamento Federal – SOF, não acompanha a base dos projetos de leis orçamentárias (PLOAs). A classificação NFGC é utilizada na conferência do trabalho de classificação da receita no padrão do Demonstrativo do Resultado Primário realizado pelas Consultorias de Orçamentos. Ela era encaminhada, quando solicitada pelo Congresso, por meio de planilha e somente em 2020, por solicitação da equipe deste projeto, foi disponibilizada ao Congresso Nacional a possibilidade de consulta pelo SIOP.

No processo legislativo orçamentário, o Demonstrativo do Resultado Primário está incluído no projeto de lei e deve constar do autógrafo. Reestimativas da receita realizadas pelo Congresso alteram o Demonstrativo a ser encaminhado no autógrafo. Até o início desse projeto, para conferir os valores do projeto de lei, as Consultorias de Orçamento da Câmara dos Deputados e do Senado Federal derivavam a classificação da receita utilizada no Demonstrativo de forma manual, a partir das classificações orçamentárias tradicionais da receita, documentadas e disponíveis nas diversas bases de dados orçamentárias. Da mesma



forma, no momento da elaboração dos autógrafos do PLOA, quando havia modificação da receita durante o processo legislativo orçamentário, era necessário construir o mesmo Demonstrativo e a tradução manual entre as classificações tornava-se novamente necessária.

Diante desse problema, a Consultoria de Orçamentos e o PRODASEN iniciaram um projeto-piloto que visava, por meio da utilização da inteligência artificial (IA), inferir a classificação da receita da elaboração orçamentária nos moldes do Demonstrativo do Resultado Primário.

5.2 Características do projeto-piloto de IA aplicada ao Demonstrativo de Resultado Primário

O projeto de classificação da receita na forma do Demonstrativo do Resultado Primário pertence a categoria de aprendizado de máquina supervisionado para a tarefa de classificação. O problema inicial do projeto consiste em inferir, para a base de dados do Projeto de Lei Orçamentária para 2021 (PLOA 2021), chamada de base de teste, a classificação própria do Demonstrativo do Resultado Primário (classes de predição). Essa inferência é realizada a partir de algumas classificações orçamentárias da receita constantes do PLOA 2021 em função do relacionamento conhecido entre as classificações orçamentárias da receita (atributos preditores) e a classificação do Demonstrativo de Resultado Primário (classes), constantes das bases de dados dos autógrafos do PLOA 2019 e 2020, chamadas de base de treinamento (aprendizado). Em suma, a partir da base de treinamento (PLOA 2019 e 2020) a IA aprende o relacionamento entre as classificações orçamentárias da receita e as classes do Demonstrativo do Resultado Primário. Em seguida, aplica-se o conhecimento aprendido para inferir essas classes na base de teste (PLOA 2021). No ano seguinte, o PLOA 2021 incorpora-se à base de treinamento e a predição é efetivada na nova base de teste, a do PLOA 2022.

O projeto foi desenvolvido no ambiente Google Colab, na linguagem de programação Python, que é amplamente utilizada em aplicações de IA. No ambiente de desenvolvimento, os dados são extraídos do SIGA Brasil e alimentados na base de dados do projeto. No ambiente de produção, o código *Python* desenvolvido no Google Colab foi migrado para o repositório GitHub do Senado Federal. Na produção o acesso às bases de dados de treinamento e teste é realizado diretamente nos bancos de dados do SIGA Brasil.



As subseções seguintes apresentam as diversas fases do projeto de classificação da receita na forma do Demonstrativo do Resultado Primário. A modelagem das fases do projeto, embora tenha suas particularidades, são conexas às definidas pela OCDE e pelo CRISP-DM, abordados na seção 4.3. Porém, antes de avançar nas diversas fases, é apresentada uma subseção contendo alguns conceitos básicos necessários ao entendimento do projeto.

5.3 Conceitos básicos: algoritmo e método, biblioteca e modelo

Em IA, algoritmos, também chamados de métodos, são programas utilizados em diversas funções. Geralmente, os métodos são disponibilizados em bibliotecas para serem aplicados em linguagens de programação.

Biblioteca é um conjunto de funções úteis (programas) que eliminam a necessidade de escrever códigos do zero. O Python tem diversas bibliotecas especializadas em IA, que são utilizadas em tarefas múltiplas nas diversas fases do projeto. Por exemplo, na fase de análise exploratória de dados, as bibliotecas estatísticas são utilizadas para avaliar a distribuição de dados, plotar gráficos, identificar *outliers* e o grau de correlação entre variáveis. Outros tipos de bibliotecas são utilizados na etapa de pré-processamento e ajuste da base para tratamento de dados faltantes e de outliers, análise de multicolinearidade, criação de atributos derivados e operações de conversão da base em formatos possíveis de serem utilizados por outros algoritmos. Há também tipos de bibliotecas que auxiliam na escolha, treinamento e parametrização de algoritmos, além de bibliotecas que contém os algoritmos a serem utilizados finalisticamente na aplicação IA para solucionar o problema inicial do projeto, como por exemplo, as especializadas em classificação, regressão ou clusterização.

Após a definição do problema, escolha e tratamento da base de dados, o processo de desenvolvimento de uma aplicação de IA é concentrado na escolha, parametrização e aferição dos métodos a serem empregados para resolver o problema proposto.

Os métodos têm uma série de parâmetros internos que são empregados de acordo com a situação de uso. Muitas vezes, a parametrização do algoritmo é realizada por meio de procedimentos automáticos realizados na base de treinamento. Definido os parâmetros de um algoritmo, ele passa a ser chamado de modelo, que será utilizado na aplicação IA. Esses conceitos serão empregados nas seções seguintes.



5.4 Entendimento e preparação dos dados

5.4.1 Bases de Dados e classes do Resultado Primário

Na aplicação de classificação da receita de acordo com o Demonstrativo do Resultado Primário, a base de dados para treinamento, avaliação e seleção de parâmetros dos métodos utilizados no estudo corresponde ao conjunto das receitas primárias previstas nas esferas fiscal e da seguridade social constantes dos autógrafos dos PLOAs 2019 e 2020. A fonte de dados é o *data warehouse* do SIGA Brasil.

Agrupando-se os dois exercícios, a base de treinamento possui 3.751 amostras (linhas) de receitas, cada qual contendo os atributos: Esfera, Fonte de recursos, Natureza da receita, Unidade orçamentária e Subgrupo. As linhas de receitas foram filtradas pelo critério “receita primária” do atributo Identificador de resultado primário.

A base de dados de teste corresponde à receita primária prevista no PLOA 2021 e possui 1.684 amostras. Os atributos pertencentes a essa base são os mesmos da base de treinamento, excluído o atributo Subgrupo, cujas classes constituem o alvo da predição. O Demonstrativo do Resultado Primário é apresentado no Quadro 9A do Volume I do PLOA 2021 (BRASIL 2020, p. 249) e discrimina a receita em 20 classes. Não há qualquer relacionamento entre os dados do Demonstrativo e a base de dados da receita que acompanha o projeto de lei encaminhado pelo Poder Executivo ao Congresso Nacional.

É preciso esclarecer que o identificador de resultado primário atribui valor primário ou financeiro à receita e não se confunde com a classificação utilizada no Demonstrativo de Resultado Primário. Essa classificação contém as classes alvo da predição, objetivo da aplicação de IA do projeto-piloto. No SIGA Brasil, a classificação do Demonstrativo de Resultado Primário é chamada de Subgrupo. Essa denominação decorre da estrutura de classificação hierárquica no Demonstrativo de Resultado Primário que se convencionou chamar de Grupo e Subgrupo da Receita.

Neste trabalho, utiliza-se o Subgrupo em referência à classificação do Demonstrativo de Resultado Primário. O Subgrupo não pertence originalmente às bases dos PLOAs e foi classificado manualmente na base do SIGA Brasil, especialmente para o presente projeto. Iniciativas anteriores de derivação manual da classificação, realizadas pela Consultoria de Orçamentos do Senado, não haviam sido incorporadas à base e se perderam.



Consta ainda da base do SIGA Brasil o atributo Descrição da NFGC, extraído do Sistema Integrado de Orçamento e Planejamento - SIOP, definido no SIGA Brasil como Classificação SOF.

Como característica, a base de dados da elaboração da receita, por ser fonte do processo legislativo orçamentário, é estruturada e de ótima qualidade. Porém, vale salientar que há uma distribuição bastante desproporcional entre as classes do atributo Subgrupo, o que caracteriza o desbalanceamento do conjunto de dados. Esse desbalanceamento, devido à desproporcionalidade do número de amostras entre as classes e à baixa incidência de amostras em algumas classes, é um fator de dificuldade à predição, uma vez que “o número de instâncias de treinamento em determinadas classes é insuficiente para que os modelos consigam aprender os conceitos inerentes a essas classes” (CARVALHO, 2021, p. 29). Além disso, aplicações de IA para classificação geralmente são empregadas na classificação da amostra em bases de dados com apenas duas classes. No caso do projeto de classificação da receita na forma do Demonstrativo do Resultado Primário, o alvo da predição contém vinte classes (os subgrupos), por isso, esse tipo de classificação é denominada multiclasse. O desbalanceamento e a multiclasse são fatores que dificultam o processo de predição. Trata-se de características singulares da base que mereceram tratamentos especiais no projeto.

As classes desbalanceadas configuram uma particularidade da base de treinamento e têm implicações práticas. Suponha que a base de teste tenha as distribuições de amostras entre as classes semelhantes à de treinamento, as duas menores classes têm apenas duas ocorrências e a maior, 2.238. Isso significa que um erro de predição nas classes menores implica em 50% de erro com relação à quantidade de amostras da classe, mas representa apenas 0,026% do total. Por outro lado, um erro na maior classe equivale a apenas 0,044% da classe. Além de ser um elemento de dificuldade para predição, o desbalanceamento na distribuição das amostras entre as classes tem repercussões no método de aferir o resultado das predições.

O desbalanceamento e o fator multiclasse observado no projeto-piloto foram assuntos de um MBA em ciência de dados cursado pelo integrante do PRODASEN no projeto. O trabalho de conclusão de curso foi especificamente voltado à superação dessas dificuldades, o título do TCC é “Métodos de aprendizado de máquina para classificação das receitas na forma da apuração do Resultado Primário em um cenário de múltiplas classes desbalanceadas: Orçamento Geral da União” (Carvalho, 2021).

Uma das soluções encontrada para contornar os problemas do desbalanceamento e da multiclasse foi gerar novas observações por amostragem aleatória na base de treinamento.



Esse procedimento, chamado de reamostragem, teve como objetivo preservar a distribuição das classes e, ao mesmo tempo, gerar observações suficientes em classes raras de modo a assegurar um mínimo de 10 instâncias em cada classe. Para isso, foi utilizado o algoritmo *RandomOverSampler* da biblioteca *Imblearn*, aplicado nas classes da base de treinamento com menos de dez observações (CARVALHO, 2021, p. 30). Sem a reamostragem seria impossível a realização de etapas posteriores do projeto como as validações cruzadas, abordadas mais à frente. A Tabela 1 apresenta as amostras por Subgrupo nas bases de treinamento, de reamostragem e de teste.

Tabela 1 – Amostras por classe na base de treinamento e de teste

Subgrupo (Classe do Demonstrativo do Resultado Primário)	Treinamento		Treinamento reamostragem		Teste	
	Qtd*	%	Qtd*	%	Qtd*	%
RECEITA PRÓPRIA (FTS 50 & 81)	2.238	59,66	2.238	59,32	1.046	62,11
DEMAIS RECEITAS	626	16,69	626	16,59	312	18,53
IMPOSTO SOBRE A RENDA	205	5,47	205	5,43	50	2,97
IPI	125	3,33	125	3,31	45	2,67
EXPLORAÇÃO DE RECURSOS NATURAIS	116	3,09	116	3,07	60	3,56
OUTRAS ADMINISTRADAS PELA RFB	112	2,99	112	2,97	54	3,21
CONCESSÕES E PERMISSÕES	88	2,35	88	2,33	32	1,9
RECURSOS DO RGPS	48	1,28	48	1,27	17	1,01
PIS/PASEP	41	1,09	41	1,09	10	0,59
CIDE - COMBUSTÍVEIS	32	0,85	32	0,85	12	0,71
COFINS	22	0,59	22	0,58	6	0,36
CSLL	22	0,59	22	0,58	6	0,36
CONTRIBUIÇÃO PARA O PLANO DE SEGURIDADE SOCIAL DO SERVIDOR	19	0,51	19	0,5	9	0,53
CONTRIBUIÇÃO DO SALÁRIO EDUCAÇÃO	16	0,43	16	0,42	6	0,36
IOF	12	0,32	12	0,32	6	0,36
IMPOSTO DE IMPORTAÇÃO	11	0,29	11	0,29	5	0,3
DIVIDENDOS	8	0,21	10	0,27	4	0,24
COMPLEMENTO DO FGTS	6	0,16	10	0,27	2	0,12
OPERAÇÕES COM ATIVOS	2	0,05	10	0,27	1	0,06
RESSARCIMENTO DE DESONERAÇÕES PREVIDENCIÁRIAS	2	0,05	10	0,27	1	0,06
Total	3.751		3.773		1.684	

* Qtd=quantidade

Fonte: Carvalho, Marco. Métodos de aprendizado de máquina para classificação das receitas na forma da apuração do resultado primário em um cenário de múltiplas classes desbalanceadas: Orçamento Geral da União

As bases de dados do orçamento são ampliadas a cada exercício. O Poder Executivo não realiza o tratamento dessas bases do ponto de vista da plurianualidade, o que proporciona diversos problemas nas classificações orçamentárias. Outros problemas relacionados à plurianualidade são provocados pela própria dinâmica do processo político e administrativo.

Por isso, um aspecto importante da base de aprendizado do projeto-piloto envolve o atributo Unidade Orçamentária. O PLOA 2019 foi elaborado pelo governo do Presidente Temer. Houve reforma ministerial no governo do Presidente Bolsonaro em 2019 e o PLOA



2020 já incorpora as Unidades Orçamentárias (códigos) com alterações decorrentes dessa reforma. Apesar da dificuldade que o comportamento desse atributo pode gerar, reformas ministeriais são comuns em mudanças de governos. Para contornar esse problema, foi realizada uma compatibilização entre as UOs da base de treinamento de 2019 e 2020, principalmente as relacionadas às receitas da classe Recursos do RGPS.

Alterações do atributo natureza da receita ao longo dos anos também mereceram cuidado no projeto-piloto. A cada ano, novas naturezas da receita podem ser criadas, fundidas ou desmembradas.

Por esse motivo, é importante que o modelo aprenda a lidar com as alterações nas classificações. A experiência do projeto-piloto mostrou que problemas originários da plurianualidade orçamentária não devem ser negligenciados e correções da base de treinamento e dos modelos, em função de alterações nas classificações orçamentárias, são necessárias. Porém, os modelos, em alguns casos, comportam a autoadaptação. Isso é possível de ocorrer na medida em que o modelo entenda que há uma ordem de importância entre os preditores para inferir o resultado. Esse assunto será detalhado no próximo tópico.

De qualquer forma, é preciso avaliar e escolher uma solução. Essa é uma lição importante do projeto. Classificações orçamentárias não são estáticas, é necessário um olhar atento às mudanças no decorrer dos anos. Geralmente, não há um tratamento plurianual das bases de dados orçamentárias por parte do Poder Executivo.

5.4.2 Seleção de atributos

5.4.2.1 Etapa preliminar de seleção de atributos

A seleção de atributos para classificação em aprendizado de máquina supervisionado consiste na escolha do conjunto de atributos que possibilite a predição da classe com o melhor resultado. Esse procedimento é chamado de engenharia de atributos. Berryhil *et al* citam e completam o glossário do Google's Machine Learning para definir engenharia de atributos (feature engineering) como 'o processo de determinar qual atributo pode ser útil em um modelo de treinamento'. Em outras palavras, a engenharia de atributos consiste em selecionar as variáveis mais relevantes para atuarem como atributos preditores no processo de predição das classes. (BERRYHILL *et al*, 2019, p. 38).



A análise de atributo no projeto-piloto foi realizada pela combinação de duas metodologias complementares. A primeira, baseada no conhecimento do negócio, de forma a incluir no processo conhecimentos que não são possíveis serem absorvidos pelo algoritmo e a segunda baseada em procedimentos automáticos. As duas metodologias foram avaliadas por meio de medição de resultados.

O procedimento de avaliação do desempenho da análise de atributo é realizado na base de aprendizagem. Essa base contém o atributo de resultado e os outros atributos que, por meio de um processo, devem ser selecionados como preditores. O relacionamento entre o atributo de resultado e os outros atributos da base de treinamento é conhecido. A técnica utilizada para essa tarefa em conjunto com a avaliação do resultado será detalhada mais a frente, mas, de forma simplificada, consiste em separar uma parte dos dados da base de treinamento para a máquina aprender a relação entre um conjunto de preditores e o atributo de resultado. A outra parte da base serve como uma simulação da base de teste, cujo atributo de resultado é omitido do algoritmo preliminarmente. Então, é realizada a predição da classe nessa base. O resultado dessa predição é comparado com o atributo de resultado que já era conhecido, mas que havia sido omitido. Por fim, é realizada a mensuração dos erros e acertos da predição. O processo é realizado várias vezes para encontrar o melhor conjunto preditores.

A engenharia de atributos é um elemento essencial na tarefa de predição. Como ressalta Schmitt (2020c, tradução nossa), “um algoritmo não entende nosso mundo. É crucial que você dê ao cientista de dados algumas dicas sobre quais dados são realmente relevantes para que ele possa selecionar e dividir os dados de uma maneira que o algoritmo possa entender”.

O domínio do negócio e a análise de dados são fundamentais para o processo de seleção preliminar de atributos, que são utilizados para construir modelos. Para Fawcet e Provost, (2013, p. 25, 26 e 30) a mineração de dados (entendida IA) é uma arte. Envolve a aplicação substancial de ciência e tecnologia, mas também envolve arte. Os cientistas de dados podem gastar um tempo considerável no início do processo de aprendizado de máquina para definir variáveis que serão utilizadas posteriormente. Para eles, esse é um dos principais pontos em que a criatividade humana, o bom senso e o conhecimento do negócio entram no jogo.

No projeto de classificação da receita na forma da apuração do resultado primário, os atributos (variáveis) são as classificações orçamentárias da receita. Segundo Schmitt (2020a), um sinal comum de que a solução de aprendizado de máquina provavelmente será uma boa opção é a existência de tarefa, real ou imaginada, baseada em regras ou heurística



feitas à mão. Se for possível resolver parcialmente o problema usando heurística, o aprendizado de máquina ajudará a encontrar, automaticamente, regras mais complexas e precisas utilizando exemplos existentes. A Consultoria do Senado utilizava regras e intervenções manuais para marcar a classificação da receita no padrão do Demonstrativo do Resultado Primário nas bases do PLOA. Para isso, empregava, como atributos: Natureza da receita, Unidade orçamentária e Fonte de recursos.

Essas variáveis foram o ponto de partida da engenharia de atributos do projeto. A primeira parte do estudo focou na análise sobre a efetividade das variáveis já utilizadas pela Consultoria. Houve o entendimento de que o conhecimento do negócio poderia contribuir para aprimorar a predição. Dessa forma, considerou-se que a engenharia de atributos não deveria se restringir à forma na qual as variáveis estavam disponibilizadas na base e procurou-se analisar o sentido semântico da regra de formação das variáveis.

Alguns códigos da classificação orçamentária da receita são compostos, formados pela agregação de subclassificações, por isso são chamados de códigos estruturados. Os dígitos desses códigos têm significados próprios. Amparado pela estrutura do código da Fonte de recursos, composto pelo Grupo de fonte, primeiro dígito, e pela Especificação de fonte, segundo e terceiro dígitos, levantou-se a hipótese de que o Grupo de fonte poderia induzir erros na predição. A hipótese foi fundamentada no fato de que linhas com a mesma Especificação de fonte e natureza da receita, mas com Grupos de fonte diferentes poderiam ter predição em classes distintas, o que de fato ocorreu em predições preliminares.

Na elaboração da receita, o grupo de fontes com dígito igual a 9 indica Receita condicionada⁵, mas, em última análise, é a mesma receita do grupo com dígito 1, Recursos arrecadados no exercício corrente. Dessa forma decidiu-se substituir a Fonte de recursos pela Especificação de fonte como atributo preditor. Vale mencionar que essa substituição de atributos teria o mesmo impacto na predição da execução da receita, onde, além da receita condicionada, o Grupo de fonte define se a mesma receita é arrecadada no exercício em curso ou em exercício anterior.

A natureza da receita também é um código composto estruturado e a Portaria Interministerial STN/SOF nº 5/2015 (BRASIL, 2015) alterou as subclassificações que a compõem. Para o projeto-piloto, as mudanças mais significativas referem-se às subclassificações da natureza da receita Desdobramento e Tipo de receita. No domínio da subclassificação Tipo, que corresponde ao oitavo e último dígito do código da natureza da

⁵ Haverá mudanças na estrutura do código de fontes de recursos a partir do Orçamento para 2023.



receita, coexistem três categorias de informação. Na primeira categoria, definida com o dígito igual a zero, o Tipo representa a natureza da receita não valorizável ou agregadora. Valores do Tipo diferentes de zero se enquadram na segunda e terceira categorias de informação. A segunda categoria, descrita literalmente na Portaria Interministerial, define o tipo da receita como: Principal; Multas e juros; Dívida ativa; e Multas e juros de mora da dívida ativa da respectiva receita. A terceira categoria é, na verdade, uma permissão para que o ente público possa identificar peculiaridades ou necessidades gerenciais de cada natureza de receita. No PLOA 2021, o domínio do Tipo é definido conforme a segunda categoria. Na execução orçamentária da LOA 2021, o Tipo engloba sete instâncias, definidas com o respaldo da regra da terceira categoria.

Na estrutura anterior a 2015, as informações estabelecidas atualmente pelo Tipo encontravam-se dispersas na natureza da receita. Na taxionomia da classificação da receita até 2014, por exemplo, a receita do Imposto sobre a Renda das Pessoas Físicas - IRPF era distribuída em dez instâncias da natureza da receita. Na taxionomia a partir de 2015, o IRPF passou a ser descrito somente por três instâncias da subcategoria da natureza da receita Desdobramento, considerado aqui como o agrupamento do primeiro ao sétimo dígitos da natureza da receita. O detalhamento passou a ser informado pelo Tipo, o oitavo dígito. A Tabela 2 apresenta as classificações das receitas relativas ao Imposto de Renda das Pessoas Físicas - IRPF até 2014 e a partir de 2015.

Tabela 2 - Classificação do Imposto de Renda de Pessoa Física 2014 e 2021

Código e descrição da Natureza da Receita do IRPF 2014	
1.1.1.2.04.10	Pessoas Físicas
1.1.1.2.04.11	Receita de Parcelamentos - Imposto sobre a Renda - Pessoas Físicas
1.1.1.2.04.31	Retido nas Fontes – Trabalho
1.1.1.2.04.32	Retido nas Fontes – Capital
1.9.1.1.02.01	Multa e Juros de Mora do Imposto sobre a Renda das Pessoas Físicas
1.9.1.1.02.03	Multa e Juros de Mora do Imposto sobre a Renda Retido nas Fontes
1.9.1.1.02.06	Receita de Parcelamentos - Multas e Juros de Mora do Imposto sobre a Renda - Pessoas Físicas
1.9.1.3.02.01	Multas e Juros de Mora da Dívida Ativa do Imposto sobre a Renda das Pessoas Físicas
1.9.1.3.02.03	Multas e Juros de Mora da Dívida Ativa do Imposto sobre a Renda das Pessoas Físicas Retido nas Fontes
1.9.1.3.02.06	Receita de Parcelamentos - Multas e Juros de Mora da Dívida Ativa do Imposto sobre a Renda - Pessoas Físicas
Código e descrição do Desdobramento - Natureza da Receita do IRPF 2021 (Natureza da Receita até o 7º dígito)	
1.1.1.3.01.1	Imposto sobre a Renda - Pessoas Físicas - IRPF
1.1.1.3.03.2	Imposto sobre a Renda - Retido na Fonte - Capital
1.1.1.3.03.1	Imposto sobre a Renda - Retido na Fonte - Trabalho

Fonte: Ementários da receita 2014 e 2021

A análise visual da Tabela 2 evidencia que o modelo de classificação atual facilita a identificação humana do IRPF no universo das receitas. Se a nova taxionomia simplifica a



identificação do IRPF, então ela deveria ter melhor desempenho na predição baseada em aprendizado de máquina. Dessa constatação, surgiu a hipótese de que o Desdobramento seria um preditor mais preciso que a natureza da receita para classificar a receita entre as classes do Subgrupo.

Para comprovar essa tese, foi realizada a análise de diagnóstico, que se baseia na aplicação de técnicas como *drill down*, *data discovery*, análise e correlação de dados, a fim de avaliar o relacionamento entre Desdobramento e Subgrupo. Utilizou-se para isso a ferramenta Business Objects da SAP. A análise foi efetivada na base de treinamento do PLOA 2020. Vale mencionar que o Desdobramento é subdividido em três subclasses, Desdobramento 1, Desdobramento 2 e Desdobramento 3, que representam, respectivamente, o quarto; quinto e sexto; e sétimo dígitos na natureza da receita. O exame foi realizado em cada um dos três níveis do Desdobramento. De acordo com a análise, a classificação da natureza da receita até o Desdobramento 3 (primeiro ao sétimo dígitos da natureza da receita) foi a mais promissora e denominada no projeto como Desdobramento. A Tabela 3 evidencia o relacionamento do Desdobramento e da Natureza da Receita com o Subgrupo. Das 226 instâncias do Desdobramento constantes da base de treinamento PLOA 2020, 206 apresentaram relacionamento de 1 para 1 com o Subgrupo, ou seja, são suficientes para classificar sozinhas toda linha da receita. Por outro lado, 20 instâncias do Desdobramento necessitam de se associar a outros atributos preditores para classificar a receita no Subgrupo. Nesse contexto, confirmou-se a expectativa do Desdobramento ser mais eficaz como preditor de subgrupo que a natureza da receita e também ser considerado o atributo preditor mais forte para essa predição, o que foi referendado em predições preliminares e, numa etapa seguinte, pela seleção automática de atributos.

Tabela 3 - Comparativo entre Desdobramento e Natureza da Receita no relacionamento com o Subgrupo - Base de Treinamento PLOA 2020

Classificação	Instâncias	Instâncias ligadas a	
		Um Subgrupo	Mais de um Subgrupo
Desdobramento	226	206	20
Natureza da Receita	372	344	28

Fonte: próprio autor

A Secretaria de Orçamento Federal – SOF criou, manualmente, a classificação NFGC, que é um agrupamento de naturezas da receita. O acesso a essa classificação, constante do SIOP, foi disponibilizada às Consultorias do Congresso Nacional, por solicitação desse



projeto a partir de 2020. Essa classificação foi incorporada no SIGA Brasil como “Classificação SOF”. Até então, a classificação era recebida, quando solicitada pela Consultoria do Senado, em forma de planilha. Trata-se de uma classificação intermediária com 98 classes e era utilizada, pela Consultoria, como fonte de consulta na tarefa de derivar, manualmente, o Subgrupo a partir da base de dados da receita do PLOA.

Foi avaliada no projeto-piloto se a predição do Subgrupo deveria utilizar o processo indireto, caracterizado pela predição da NFGC e em um segundo momento, proceder a classificação da linha da receita com base no relacionamento entre a NFGC e o Subgrupo. Foram realizadas predições com base nessa metodologia e o resultado foi considerado pior em relação ao alcançado com a utilização do Desdobramento como um dos atributos. Assim, o método indireto foi descartado e o conjunto dos atributos preditores na fase preliminar foi definido como Desdobramento, Unidade Orçamentária e Especificação de Fonte. Havia ainda uma dúvida se o Tipo, separado do Desdobramento, deveria ser utilizado como preditor.

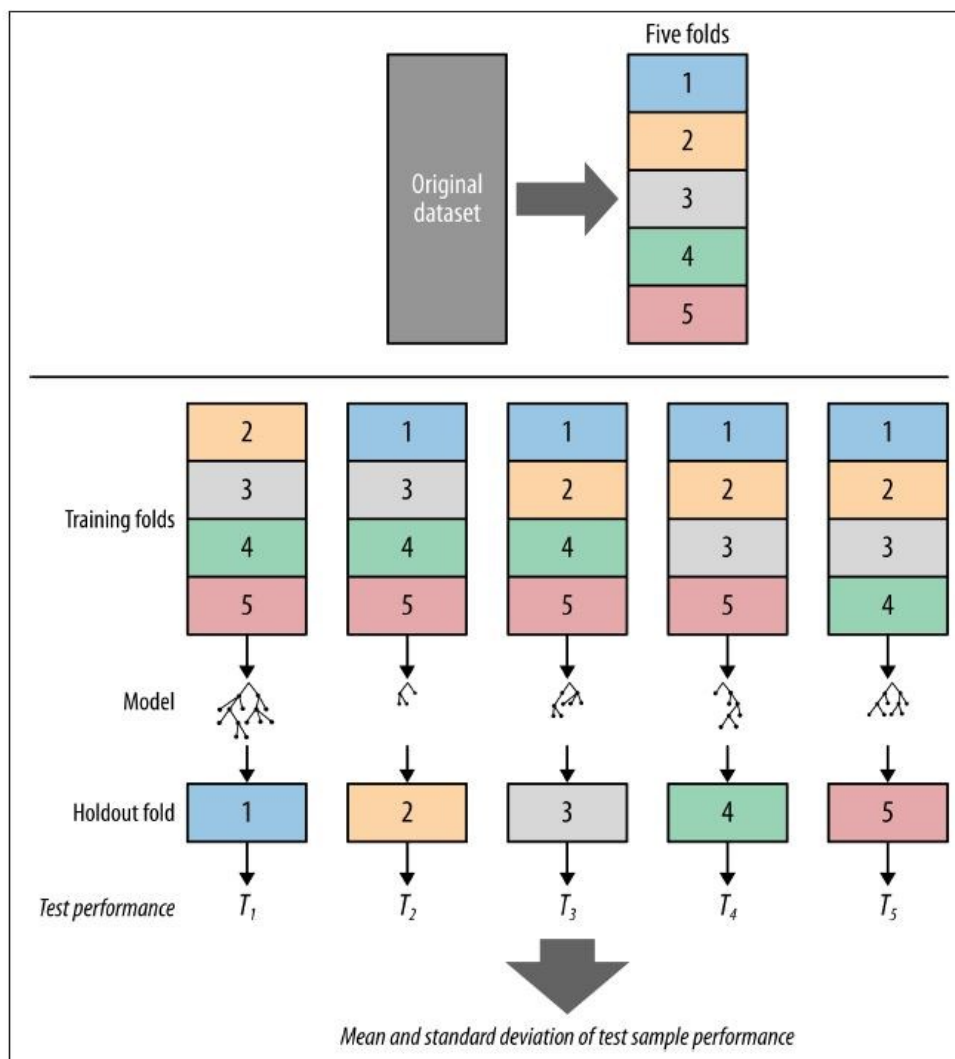
5.4.2.2 Seleção automática de atributos

No projeto-piloto, foi empregado, no segundo momento, o processo de seleção automática de atributos para validar a etapa preliminar e para verificar se algum outro atributo poderia aprimorar a capacidade preditiva dos modelos. Esse procedimento foi realizado na base de treinamento já com reamostragem, por meio do algoritmo *Recursive Feature Elimination – RFE*, utilizando o algoritmo *XGBoost* como estimador. Em outras palavras, o *XGBoost* é o algoritmo que aprende a relação entre o conjunto de atributos preditores e o atributo de resultado numa parte da base de aprendizagem e realiza a predição da classe na outra parte dessa base não vista no treinamento.

Essa repartição da base de teste para a avaliação de desempenho do conjunto de preditores foi realizada, no projeto, por meio de uma técnica mais sofisticada que a didaticamente explicada anteriormente, chamada de validação cruzada. Essa técnica diminui o risco de se avaliar parcialmente o algoritmo e consiste em fatiar a base de aprendizado, proporcionalmente em relação às classes, em partes, chamadas de *folds*. O número de partes, denominado de *K*, é o parâmetro da validação cruzada, por isso, ela também é chamada de validação cruzada *K-fold*. A figura 2 é uma representação gráfica de uma validação cruzada com *K* igual a cinco.



Figura 2 – Representação gráfica da validação cruzada



Fonte: Fawcett e Provost (2013, p. 128)

Nessa representação, a base de aprendizado é dividida aleatoriamente, respeitada a proporcionalidade das amostras entre as classes, em cinco *folds* (partes). São realizadas cinco rodadas de treinamento e predição das classes. Em cada uma das cinco rodadas, a base de aprendizado, onde o algoritmo é treinado e aprende a relação entre os atributos preditores e o de resultado, é composta pelo conjunto de quatro *folds*. O *fold* restante, chamado de *holdout fold* (base de validação), é utilizado como uma simulação da base de teste para a realização da predição da classe. Nesse *fold*, no primeiro momento, o atributo de resultado, a classe, é omitido do algoritmo. O algoritmo utiliza o conhecimento da relação aprendida no treinamento e o aplica com os atributos preditores desse *fold* para realizar a predição das classes. Depois, o algoritmo acessa o atributo que havia sido omitido, compara com o atributo



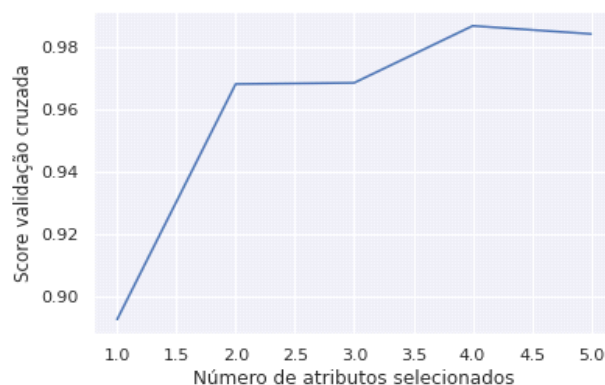
obtido na predição e produz um score do resultado da predição. Em cada uma das cinco rodadas é avaliado o resultado da predição. Por último, é realizada a média dos resultados para se chegar ao resultado final.

No projeto-piloto, a validação cruzada para a seleção automática de atributo foi realizada com dez *folds* (K=10), o que explica a necessidade de se aplicar a reamostragem na base de treinamento, de tal forma que cada classe tivesse, no mínimo, dez observações. A validação cruzada foi efetivada por meio do algoritmo RFE, que realiza a separação da base por *folds* e utiliza o algoritmo *XGBoost* para realizar a predição em cada rodada. O algoritmo RFE também é o provedor dos atributos a serem utilizados pelo *XGBoost*. Além dos atributos selecionados na etapa preliminar, Desdobramento, Especificação de Fonte e Unidade Orçamentária, foram acrescentados, para a avaliação automática, os atributos Tipo de Receita e Esfera. O RFE inicia o processo fornecendo esses cinco atributos para o *XGBoost* realizar a predição na validação cruzada, em seguida faz todas as combinações possíveis com quatro, três, dois e um atributo. O *XGBoost* executa a predição de todas essas combinações e avalia o resultado. Medido os resultados, é realizada a escolha do melhor conjunto de atributos preditores.

O resultado dos testes realizados pela seleção automática de atributos considerou o Desdobramento, Especificação de Fonte, Unidade Orçamentária e Esfera como o melhor conjunto de atributos para predição do Subgrupo. A Seleção automática acrescentou a esfera aos preditores definidos na etapa preliminar e eliminou a dúvida, rejeitando o Tipo de receita como atributo preditor. A métrica utilizada para a mensuração de resultados adotada no projeto foi a F1, que será descrita no próximo tópico. A Figura 3 apresenta o score da métrica F1 em relação ao número de atributos preditores. O valor mais alto ocorre com o número de quatro preditores. Ao incluir o quinto preditor, o Tipo de receita, o score da F1 cai. Os números do eixo horizontal identificam respectivamente os preditores Unidade Orçamentária, Desdobramento, Esfera, Especificação de fonte e Tipo. Pelas inclinações das retas que ligam os atributos, é possível verificar a maior relevância na predição dos atributos Desdobramento e Especificação de fonte. A inclinação negativa mostra que o Tipo é ineficaz e piora a predição.



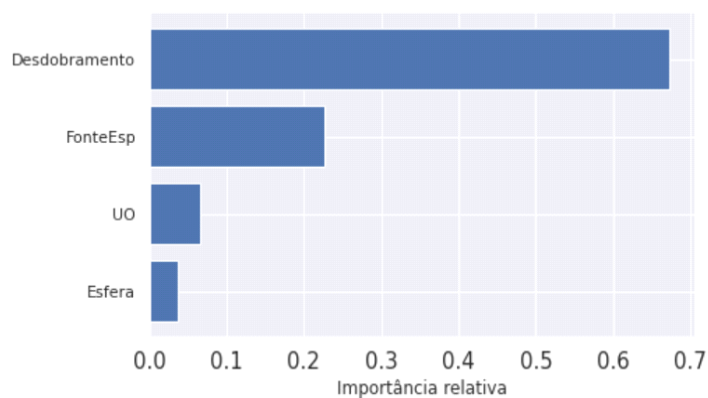
Figura 3 - Score F1 das validações cruzadas da seleção automática de atributos por número de atributos preditores selecionados



Fonte: Carvalho, Marco. Métodos de aprendizado de máquina para classificação das receitas na forma da apuração do resultado primário em um cenário de múltiplas classes desbalanceadas: Orçamento Geral da União.

A seleção automática de atributos também confirmou a hipótese levantada na etapa preliminar de seleção que considerava o Desdobramento como o atributo mais significativo na predição do Subgrupo. A Figura 4 identifica a importância relativa dos atributos para a predição do Subgrupo.

Figura 4 - Importância relativa dos atributos para a predição do Subgrupo



Fonte: próprio autor

5.5 Métrica F1 de avaliação de resultados de treinamento e de teste

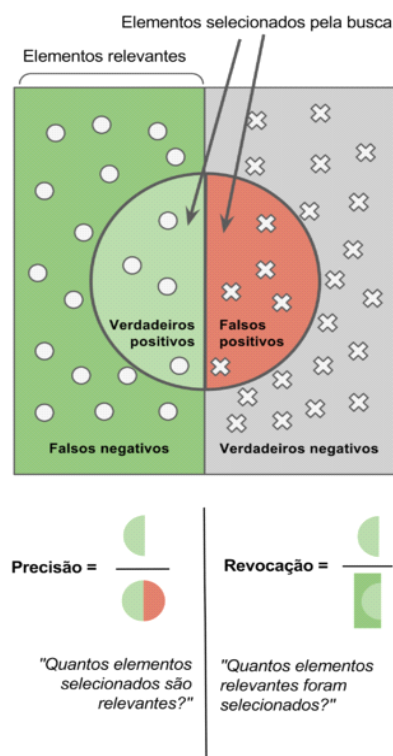
A verificação dos resultados é um fator essencial no desenvolvimento dos projetos de IA. É por meio da mensuração das predições que se torna possível avaliar o caminho escolhido, corrigir as rotas ou aprofundar na direção escolhida. Essas aferições estão presentes em quase todas as etapas do processo de criação de uma aplicação em IA. Os



administradores do projeto precisam ter em mente quando medir, o que medir e como medir. Com relação ao como medir, uma das questões que se apresenta é qual a métrica deve ser utilizada. O projeto-piloto utilizou métrica F1 ponderada para aferir os resultados.

Essa métrica mede os resultados das predições, ponderando o peso das classes de acordo com o número de instâncias de cada uma delas em relação ao número total de instâncias. O método de mensuração toma como base a média harmônica entre a precisão e a revocação em cada classe e em cada um dos *folds* testados na validação cruzada. Segundo Carvalho (2021, p. 31), por uma característica da média harmônica, a métrica F1 sofre maior influência do menor valor entre precisão e revocação. Ela assume valores entre zero e um. Quanto mais próximo de um for o valor da F1, melhor é o desempenho geral do algoritmo em todas as classes. A precisão é calculada pela razão entre as instâncias preditas corretamente na classe, chamadas de verdadeiros positivos, e o número total de predições na classe, que corresponde à soma das predições corretas com as instâncias preditas erroneamente na classe, chamadas de falsos positivos. A revocação é a fração entre os verdadeiros positivos na classe e todas as instâncias que deveriam ser classificadas na classe, mas que foram classificadas erradamente em outras classes, denominadas de falsos negativos. A figura 5 mostra a representação gráfica do conceito de precisão e de revocação.

Figura 5 – Representação gráfica da precisão e da revocação



Fonte: Wikipedia



5.6 Escolha do modelo

Diversas bibliotecas do *Python* possuem algoritmos capazes de realizar a classificação por meio do aprendizado de máquinas supervisionado. Uma das tarefas inerentes aos projetos de IA é definir o algoritmo que se tornará o modelo de predição da aplicação. Normalmente, essa definição é fundamentada numa série de exercícios de predições realizadas na base de treinamento após a seleção dos atributos preditores realizada na fase anterior.

Para a escolha do modelo de predição, foram avaliados cinco algoritmos: *Decision Tree*, *Bagging*, *Random Forest*, *AdaBoost* e *Support Vector Machine (SVM)*. Cada um desses algoritmos utiliza uma sistemática para realizar a predição das classes. Além disso, cada um emprega uma técnica específica para lidar com o problema de desbalanceamento em múltiplas classes. As características e ajustes dos algoritmos utilizados no projeto-piloto estão descritas em Carvalho (2021).

O processo de escolha do modelo de predição teve como ponto de partida a avaliação de predições na base de treinamento por meio de uma técnica chamada de validação cruzada encaixada ou aninhada (*nested cross validation*), que possui dois momentos. No primeiro momento, cada um dos algoritmos utiliza a técnica da validação cruzada para fazer uma série de predições preliminares, a fim de definir, automaticamente, seus parâmetros internos. Após a definição interna dos parâmetros, o algoritmo ou método passa a ser chamado de modelo. Então, no segundo momento, em uma nova validação cruzada, aplica-se o modelo para realizar a predição das classes em cada holdout fold (base de validação). A métrica F1 foi utilizada para medir o resultado da predição de cada modelo na validação cruzada encaixada. Como a validação cruzada encaixada foi configurada para dez folds, para cada método obteve-se dez resultados de F1.

Em seguida, ainda no âmbito da tarefa de comparação dos modelos, procedeu-se a análise bayesiana da métrica F1, que mediu os resultados da validação cruzada. O resultado da análise bayesiana é uma distribuição a posteriori de cada modelo. Para a comparação entre os modelos, são utilizados os parâmetros dessa posteriori: média, moda (*Maximum A Posteriori*-MAP) e desvio padrão. Também foi utilizada para a comparação entre os modelos, a amplitude entre os limites do intervalo de credibilidade (IC), que foi definido em 98%. Para conduzir essa análise, foi utilizada a biblioteca de programação probabilística PyMC3, que



computa as estimativas da posteriori numericamente, utilizando algoritmos conhecidos como *Markov Chain Monte Carlo* (MCMC) (CARVALHO, 2021, p. 34). Essa biblioteca, antes da análise bayesiana gera uma quantidade de amostras definidas no parâmetro da biblioteca. No projeto, ela gerou cinco mil amostras da F1 para cada modelo. A Tabela 4 apresenta, em cada um dos modelos, os valores dos parâmetros da distribuição da métrica F1 para a validação cruzada (segundo momento da validação cruzada encaixada), a estimativa bayesiana MAP, o intervalo de credibilidade (IC) e o desvio padrão .

Tabela 4 – Resultados dos modelos na validação cruzada e estimativas MAP, IC e desvio padrão

Modelo	Validação Cruzada	Estimativas F1: MAP, IC e Desvio Padrão			
		F1	MAP F1	IC (98%)	Desvio Padrão
Decision Tree		0,957	0,957	[0,947 . 0,967]	0,012
Bagging		0,995	0,995	[0,991 . 0,998]	0,003
Random Forest		0,996	0,996	[0,994 . 0,999]	0,004
AdaBoost		0,997	0,997	[0,994 . 1,000]	0,003
SVM		0,945	0,945	[0,933 . 0,956]	0,012

Fonte: próprio autor

A avaliação da Tabela 4 evidencia que os melhores resultados foram atingidos pelo *AdaBosst*, *Random Forest* e *Bagging*. Na estatística bayesiana, a incerteza sobre o parâmetro (F1 ponderada) é quantificada pela distribuição a posteriori. Foram utilizados quatro parâmetros da distribuição posteriori para mensurar essa incerteza: F1 da validação cruzada, MAP F1, diferença entre os limites superior e inferior do IC e desvio padrão da F1. A melhor F1 da validação cruzada é do *AdaBoost* (0,997), seguido pelo *Random Forest* (0,996). Esse resultado se repete com o MAP F1, que representa a moda da distribuição a posteriori. O limite superior do IC (98%) do *Adaboost* é o mais próximo de 1, porém o limite superior do IC do *Random Forest* é bem próximo (0,999) e esse modelo tem a menor diferença entre os limites do intervalo de credibilidade (0,999 - 0,994 = 0,005). O *AdaBosst* e o *Bagging* obtiveram os melhores resultados do desvio padrão com 0,003.

Para realizar uma comparação mais precisa dos modelos, foi empregado o teste t correlacionado bayesiano, na base de treinamento, por meio da biblioteca *Baycomp*, disponível em *Python*. Esse teste calcula a diferença, entre dois modelos, da média dos resultados da validação cruzada encaixada e estima numericamente a distribuição das



diferenças médias com o algoritmo da família MCMC. No resultado da aplicação do teste t, é especificada uma região de prática equivalência (ROPE). No projeto, o ROPE foi definido com o intervalo de $[-0,01, 0,01]$. A região da distribuição das diferenças das médias que recai nesse intervalo e corresponde à probabilidade de prática equivalência entre os modelos. Ou seja, se a diferença dessa distribuição estiver completamente dentro do intervalo, os modelos são considerados praticamente equivalentes. A análise permite, ainda, verificar a superioridade de um modelo sobre o outro a partir das diferenças médias da métrica F1 (CARVALHO, 2021, p. 35).

A Tabela 5 apresenta o resultado do teste t correlacionado bayesiano da validação cruzada. A comparação é realizada entre dois modelos. A coluna Esquerda indica a probabilidade do modelo 1 ser superior ao modelo 2. A coluna Direita mostra a probabilidade do modelo 2 ser superior ao modelo 1. A coluna ROPE (1%) é a região prática equivalente entre o modelo 1 e o modelo 2 (CARVALHO, 2021, p. 50).

Tabela 5 - Teste t correlacionado bayesiano da validação cruzada dos modelos

Modelo 1	Modelo 2	Esquerda	ROPE (1%)	Direita
Decision Tree	Bagging	0,0	0,0	1,0
	Random Forest	0,0	0,0	1,0
	AdaBoost	0,0	0,0	1,0
	SVM	0,592	0,399	0,009
Bagging	Random Forest	0,0	0,999	0,001
	AdaBoost	0,0	1,0	0,0
	SVM	1,0	0,0	0,0
Random Forest	AdaBoost	0,0	1,0	0,0
	SVM	1,0	0,0	0,0
AdaBoost	SVM	1,0	0,0	0,0

Fonte: Carvalho, Marco. Métodos de aprendizado de máquina para classificação das receitas na forma da apuração do resultado primário em um cenário de múltiplas classes desbalanceadas: Orçamento Geral da União.

Na comparação entre o *Random Forest* e o *AdaBoost*, os valores zeros das colunas Esquerda e Direita da Tabela 5 indicam que não há superioridade entre esses dois modelos. A ROPE igual a 1 indica que esses modelos são praticamente equivalentes. A comparação do *Bagging* com o *AdaBoost* apresenta esse mesmo resultado e a comparação entre o *Bagging* e o *Random Forest* evidencia uma mínima superioridade deste último, representada pelo valor 0,001 na coluna Direita. Em síntese, a diferença apontada pelo teste t correlacionado bayesiano entre o *AdaBoost*, o *Random Forest* e o *Bagging* na validação cruzada é inferior a 1% o que indica que os modelos obtiveram resultados, conforme o jargão da estatística bayesiana, praticamente equivalentes.



A correção de cem por cento da classificação da receita na forma do Demonstrativo do Resultado primário é um requisito do projeto, uma vez que o Demonstrativo faz parte do autógrafa do projeto de lei orçamentária. No entanto, modelos preditivos somente atingem cem por cento de predições corretas em raríssimos casos. Isso é uma das principais características inerentes aos modelos. Assim, algumas etapas do projeto-piloto diferenciam das geralmente utilizadas nas aplicações preditivas de classificação. Em princípio, nos projetos de IA relacionados à classificação, escolhido um modelo na fase de treinamento, somente esse modelo é aplicado na fase teste. Diferentemente, no projeto-piloto decidiu-se realizar a predição na base de teste (PL 2021) com todos os modelos. Para realizar a avaliação, a receita do PL 2021 foi classificada manualmente com auxílio da Classificação NFGC. A tabela 6 apresenta os resultados das predições dos modelos na base de teste.

Tabela 6 – Resultados dos modelos na validação cruzada na base de teste e das estimativas MAP e IC

Modelo	Estimativa MAP F1 e IC		Teste
	MAP F1	IC (98%)	
Decision Tree	0,957	[0,947 . 0,967]	0,966
Bagging	0,995	[0,991 . 0,998]	0,986
Random Forest	0,996	[0,994 . 0,999]	0,985
AdaBoost	0,997	[0,994 . 1,000]	0,992
SVM	0,945	[0,933 . 0,956]	0,984

Fonte: Carvalho, Marco. Métodos de aprendizado de máquina para classificação das receitas na forma da apuração do resultado primário em um cenário de múltiplas classes desbalanceadas: Orçamento Geral da União

O *AdaBoost* obteve a melhor F1 na predição na base de teste (0,992) e, além disso, o valor da F1 do teste foi o mais próximo do intervalo de credibilidade do modelo verificado na fase de treinamento. Assim, o modelo *AdaBoost* foi a escolhido para realizar a predição no projeto.

5.7 Resultados da predição na base de teste

A Tabela 7 apresenta os valores da precisão, revocação, F1 e instâncias por classes da predição do modelo *AdaBoost* na base de teste. Verifica-se que os erros na predição ocorreram somente nas duas classes com os maiores números de instâncias, o que mostra que o modelo é eficiente para lidar com o desbalanceamento multiclasse.



Tabela 7 - Precisão, revocação, F1 e instâncias por classes de predição na base de teste do modelo *AdaBoost*

Classe	Precisão	Revocação	F1	Instâncias
RECEITA PRÓPRIA (FTS 50 & 81)	1,000	0,987	0,993	1046
DEMAIS RECEITAS	0,957	1,000	0,978	312
EXPLORAÇÃO DE RECURSOS NATURAIS	1,000	1,000	1,000	60
OUTRAS ADMINISTRADAS PELA RFB	1,000	1,000	1,000	54
IMPOSTO SOBRE A RENDA	1,000	1,000	1,000	50
IPi	1,000	1,000	1,000	45
CONCESSÕES E PERMISSÕES	1,000	1,000	1,000	32
RECURSOS DO RGPS	1,000	1,000	1,000	17
CIDE - COMBUSTÍVEIS	1,000	1,000	1,000	12
PIS/PASEP	1,000	1,000	1,000	10
CONTRIBUIÇÃO PARA O PLANO DE SEGURIDADE SOCIAL DO SERVIDOR	1,000	1,000	1,000	9
COFINS	1,000	1,000	1,000	6
CONTRIBUIÇÃO DO SALÁRIO EDUCAÇÃO	1,000	1,000	1,000	6
CSLL	1,000	1,000	1,000	6
IOF	1,000	1,000	1,000	6
IMPOSTO DE IMPORTAÇÃO	1,000	1,000	1,000	5
DIVIDENDOS	1,000	1,000	1,000	4
COMPLEMENTO DO FGTS	1,000	1,000	1,000	2
OPERAÇÕES COM ATIVOS	1,000	1,000	1,000	1
RESSARCIMENTO DE DESONERAÇÕES PREVIDENCIÁRIAS	1,000	1,000	1,000	1
Média macro	0,998	0,999	0,999	1684
Média ponderada	0,992	0,992	0,992	1684

Fonte: Carvalho, Marco. Métodos de aprendizado de máquina para classificação das receitas na forma da apuração do resultado primário em um cenário de múltiplas classes desbalanceadas: Orçamento Geral da União

Em números absolutos, comparada com o Subgrupo derivado manualmente da classificação NFGC pela Consultoria de Orçamentos do Senado, a predição com modelo *AdaBoost* errou na classificação de 14 das 1684 instâncias, o que implica em 99,17% de linhas classificadas corretamente. Os erros de classificação ocorreram em apenas duas classes, Receita Própria (Fts 50 & 81) e Demais Receitas. Essas classes são as que apresentam as maiores quantidades de instâncias, 1046 e 312, respectivamente. É importante destacar que, das 1864 linhas (instâncias representadas pelo conjunto da Natureza da Receita, Unidade orçamentária, Fonte de Recursos e Esfera) preditas, 1006 não tinham correspondências na base de treinamento, eram novas na base do PL 2021.

Com o objetivo de avaliar o ganho no processo de seleção de atributos, foi realizada a predição com o modelo *AdaBoost*, aplicando-se os atributos preditores que eram utilizados anteriormente pela Consultoria de Orçamentos do Senado: Natureza da Receita, Unidade Orçamentária e Fonte de recursos. Essa predição incorreu em 28 erros de classificação de instâncias, distribuídas em 3 classes. Assim, é possível afirmar que a seleção de atributos reduziu em 50% os erros de predição das instâncias, além de ter diminuído, de três para um, o número de classes com erros na predição em relação à predição com o conjunto de atributos preditores anterior. Além disso, um dos erros da predição que utilizou os atributos iniciais ocorreu em uma classe de baixa frequência de instâncias (17), o que eleva o percentual de erro em uma determinada classe.



O Apêndice apresenta as matrizes de confusão multiclasse dessas duas predições com o modelo *AdaBoost*. A Matriz de confusão é uma representação gráfica utilizada para avaliar modelos de classificação em aprendizado de máquina, que mostra as frequências de verdadeiro positivo (VP), falso positivo (FP), verdadeiro negativo (VN) e falso negativo (FN) por classe das predições. A primeira matriz de confusão representa a predição efetivada com os atributos preditores escolhidos após a fase de seleção de atributos. A segunda matriz, mostra o resultado da classificação com os atributos preditores originalmente utilizados pela Consultoria do Senado. O Apêndice apresenta também os números para a F1 das duas matrizes de confusão. Além disso, consta do Apêndice a explicação de como interpretar a matriz de confusão multiclasse.

Também foi realizado a avaliação das divergências da predição do *AdaBoost* com a classificação manual com base na NFGC. A análise revelou que uma das divergências ocorreu porque o PL 2021 encaminhado pelo Poder Executivo continha uma natureza da receita classificada no subgrupo errado. Assim, o número de erros da predição do PLOA 2021, realizada no pelo *AdaBoost*, caiu para 13 instâncias, elevando o índice de acertos da classificação para 99,23%. As instâncias classificadas erradamente foram corrigidas manualmente e a classificação Subgrupo foi incluída no universo de Receita Elaboração do PLOA 2021.

Após a publicação do autógrafo do PLOA 2021, foi realizada uma nova predição, dessa vez, tendo como alvo a reestimativa da receita. A base de treinamento e o modelo foram os mesmos utilizados para a predição do PLOA 2021. A reestimativa da receita envolveu 1.093 linhas, compostas pela Natureza da Receita, Fonte, UO e Esfera. Dessas instâncias, 1.066 tratavam de acréscimos e cancelamentos das receitas em instâncias que já constavam da Proposta do Executivo e deveriam acompanhar as mesmas classificações de Subgrupos. A análise mais detalhada das outras 27 instâncias revelou que 23 possuíam fontes 150 e 163, também não pertenciam à UO 25917 – Fundo do Regime Geral da Previdência Social, portanto, deveriam ser classificadas no Subgrupo 3.08 – Receita Própria. As quatro instâncias restantes, com Fontes 136, 172 e 174, não estavam incluídas na Proposta do Executivo. Porém, sob a ótica do conjunto dos atributos preditores escolhidos no processo de seleção de atributos, formado pelo Desdobramento, Especificação de fonte, UO e Esfera, existiam no PLOA e estavam classificadas no subgrupo 3.09 - Demais Receitas. Nesse caso, o Tipo da receita e a Especificação da fonte diferenciavam os dois conjuntos de preditores. Como visto, segundo a engenharia de atributos do projeto, o Desdobramento é o atributo preditor mais efetivo para inferir a classificação na forma do Demonstrativo do Resultado



Primário. Por outro lado, a inserção do Tipo, último dígito da Natureza da Receita e que a diferencia do Desdobramento, piora a predição. Em síntese, essas quatro linhas do conjunto de dados da reestimativa, sob a ótica dos atributos preditores selecionados, já constavam do PLOA 2021, por isso deveriam ser classificadas no mesmo Subgrupo.

A análise da reestimativa sob a perspectiva da linha do Desdobramento, mostra que apenas 13 instâncias não estavam incluídas no PLOA 2021 e todas possuíam Especificações de fonte 150 ou 163, o que simplificou muito a tarefa de classificação. Por isso, se não houvesse passado pela engenharia de atributos, as linhas da reestimativa, baseadas na Natureza da Receita, que não existiam na base de treinamento, teriam grandes chances de serem classificadas em Subgrupos errados.

Assim, a predição da reestimativa obteve um acerto de 100%. No entanto, o Autógrafo do PLOA publicado possui quatro Subgrupos com valores divergentes dos alcançados no conjunto das predições do PLOA, já corrigidas, e da Reestimativa da receita. O demonstrativo do autógrafo possui os valores agregados e não se teve acesso aos dados detalhados por linha. De qualquer forma, há evidências de erros na parte da receita do Demonstrativo do Resultado Primário do Autógrafo do PLOA 2021 publicado. A Tabela 8 apresenta os dados da receita do Demonstrativo do Resultado Primário durante o processo orçamentário do PLOA 2021 com destaque para as divergências entre o Autógrafo publicado e o estimado pelo projeto-piloto.

Tabela 8 - PLOA 2021 - Demonstrativo do Resultado Primário - Comparativo do Autógrafo originário da predição com o Autógrafo publicado

SUBGRUPO	PLOA 2021	REESTIMATIVA	AUTÓGRAFO 2021 (Predição)	AUTÓGRAFO 2021 (Publicado)	AUTÓGRAFO 2021 (Diferença)
1.01 - IMPOSTO DE IMPORTAÇÃO	44.669,6	-687,1	43.982,5	43.982,5	0
1.02 - IPI	53.096,2	8.497,6	61.593,8	61.593,8	0
1.03 - IMPOSTO SOBRE A RENDA*	400.814,2	25.642,7	426.456,9	426.456,9	0
1.04 - IOF	47.040,3	384,4	47.424,7	47.424,7	0
1.05 - COFINS	246.518,4	884,0	247.402,3	247.402,3	0
1.06 - PIS/PASEP	67.567,1	4.505,9	72.073,0	72.073,0	0
1.07 - CSLL	81.928,3	6.761,7	88.690,0	88.690,0	0
1.09 - CIDE - COMBUSTÍVEIS	2.610,3	-617,1	1.993,1	1.993,1	0
1.10 - OUTRAS ADMINISTRADAS PELA RFB	30.758,1	-11.497,2	19.260,9	19.428,7	-167,7
2.01 - RECURSOS DO RGPS	413.838,2	-3.362,8	410.475,4	410.472,2	3,2
2.02 - RESSARCIMENTO DE DESONERAÇÕES PREVIDENCIÁRIAS	3.704,9	4.440,5	8.145,4	8.145,4	0
3.01 - CONCESSÕES E PERMISSÕES	5.240,0	0	5.240,0	5.240,0	0
3.02 - COMPLEMENTO DO FGTS	67,7	4,8	72,5	72,5	0
3.03 - CONTRIBUIÇÃO PARA O PLANO DE SEGURIDADE SOCIAL DO SERVIDOR	17.853,2	3,3	17.856,5	17.856,5	0
3.04 - CONTRIBUIÇÃO DO SALÁRIO EDUCAÇÃO	20.924,5	459,4	21.383,8	21.383,8	0
3.05 - EXPLORAÇÃO DE RECURSOS NATURAIS	60.034,7	-596,0	59.438,8	59.438,8	0
3.06 - DIVIDENDOS	9.736,9	0,0	9.736,9	9.736,9	0
3.07 - OPERAÇÕES COM ATIVOS	1.159,1	2,9	1.162,0	1.162,0	0
3.08 - RECEITA PRÓPRIA (FTS 50 & 81)	14.795,7	384,9	15.180,6	15.173,7	6,9
3.09 - DEMAIS RECEITAS	37.741,9	110,4	37.852,4	37.694,8	157,6
TOTAL	1.560.099	35.322	1.595.422	1.595.422	0

batidos os incentivos fiscais (R\$ 1,8 milhão)
ate: próprio autor



Havia dois objetivos principais que fundamentaram o projeto-piloto. O primeiro partia do reconhecimento da necessidade de superar os obstáculos causados pela dificuldade da classificação manual da receita para a avaliação do Demonstrativo do Resultado Primário encaminhado pelo Poder Executivo e para o ajuste do Demonstrativo do Autógrafo, no caso da ocorrência de reestimativa da receita. O segundo baseava na crença de que o domínio da IA poderia abrir novas perspectivas para o aprimoramento dos trabalhos da Consultoria de Orçamentos. Esses dois objetivos foram atingidos.

No entanto, modelos preditivos dificilmente atingem cem por cento de acertos e, em muitos casos, essa característica não é relevante. Porém, na tarefa objeto do projeto-piloto, a classificação de um demonstrativo que consta do projeto de lei e do autógrafo, a incerteza não é aceitável. Assim, é preciso reconhecer que, no caso do projeto, a IA não substituirá totalmente a ação humana. No entanto, não há dúvidas que seu emprego trouxe muitas vantagens em relação à metodologia anterior. Por isso, é necessário transformar o projeto-piloto em ferramenta de produção com a transmissão do conhecimento adquirido a atores das Consultorias de Orçamento da Câmara e do Senado que atuam nessa atividade.

6 PRONTIDÃO PARA O DESENVOLVIMENTO DE SOLUÇÕES DE IA: APRENDIZADOS SOBRE AS POSSIBILIDADES E DESAFIOS NA CONORF

A experiência adquirida ao longo do projeto permitiu refletir com mais clareza sobre novas implementações, organização e administração de projetos de IA, além do trabalho cooperativo entre a CONORF e PRODASEN. A partir dessa nova visão, foi possível perceber habilidades desejáveis aos integrantes da Consultoria e abordagens relevantes que podem subsidiar a CONORF na criação de soluções baseadas em IA.

As habilidades descritas pela OCDE, detalhadas no tópico 4.2, possuem grande compatibilidade com o projeto de IA para classificação da receita na forma do Demonstrativo do Resultado Primário e, conseqüentemente, com projetos de IA de maneira mais ampla. Elas podem servir de referência no desenvolvimento de uma cultura no âmbito da CONORF que reconheça a importância da IA na rotina de trabalho e que conduza à capacitação para o empreendimento de futuros projetos.

Considerando-se os dois fatores apontados por Fawcett para o desenvolvimento de IA em organizações (que os dados sejam pensados como recurso analítico e que se crie uma cultura de ciência de dados), o histórico de atuação da CONORF e a experiência relatada na



sessão anterior apontam para as seguintes evidências preliminares. O órgão, pela natureza de sua atuação institucional e pela robustez dos produtos informáticos que vem criando nos últimos 25 anos, já desenvolveu percepção coletiva da importância dos dados como recurso estruturante dos valores que entrega ao processo legislativo e à sociedade. Por outro lado, a necessidade de criação de uma cultura de ciência de dados na CONORF se revela no caráter pessoal das iniciativas, centradas na proatividade isolada de poucas pessoas, em contraposição a se ter linha de trabalho e plano de ação dedicados à abordagem analítica dos dados.

Já em relação à habilidade de alfabetização de dados, verifica-se que a CONORF, pelo histórico de uso de ferramentas como *Business Intelligence*, Processamento Analítico Online (OLAP) e *Business Discovery*, adquiriu extenso domínio das bases de dados sobre elaboração e execução orçamentária, já intrínseco à cultura de trabalho da Consultoria. Dois fatores, contudo, evidenciam a necessidade de maior investimento na área de dados. Primeiro, a cultura hoje presente no órgão se assenta mais no uso das funcionalidades prontas oferecidas pelas ferramentas do que na apropriação de oportunidades disponíveis para o desenvolvimento descentralizado de soluções, desde páginas temáticas no Painel Cidadão do sistema SIGA Brasil até o acoplamento de técnicas para a produção de indicadores, informações e conhecimentos. Segundo, os avanços rápidos na tecnologia da informação tornam imprescindível a atualização das ferramentas já em produção, além de abrirem frentes para a exploração de novas oportunidades. A mera “cultura de uso” de ferramentas se mostra, por isso, insuficiente ao pleno desenvolvimento da organização.

Integrar o conhecimento de novas tecnologias com o entendimento das possibilidades dos dados disponíveis deve ser visto como uma vertente do pensamento analítico de dados, incluindo-se, aí, o próprio reconhecimento de eventuais insuficiências ou limitações dos dados, a serem superadas antes de se iniciar projetos como os de IA. Foi o caso do projeto para classificação da receita na forma da apuração do resultado primário, amplamente fundamentado na alfabetização de dados e na interação entre o negócio e a área de ciência de dados, essenciais para o sucesso da iniciativa.

Com relação à habilidade de iteração, a aplicação de IA para classificar a receita na forma do Demonstrativo do Resultado Primário foi concebida como um projeto-piloto e executada incremental e iterativamente de forma a conhecer e aplicar tecnologias a partir de um problema real operacional da Consultoria de Orçamentos. Verificou-se que o aprendizado durante a ação de desenvolvimento é o modo natural de materializar as possibilidades informáticas para as peculiaridades de cada desafio institucional. No caso do projeto-piloto,



por exemplo, verificou-se a necessidade de testar várias vezes e de diversas formas as opções disponíveis e, assim, aprender com os acertos e os erros. Avanços e retrocessos fazem parte da dinâmica do empreendimento, sendo a consciência desse fato pela área de negócio um requisito para a inovação.

Esse requisito está estreitamente ligado à habilidade de curiosidade, também enfatizada pela OCDE. O projeto-piloto se deparou com vários desafios e se serviu da criatividade como recurso fundamental para a superação das dificuldades. A criatividade, aliada à curiosidade em aprender, são comportamentos imprescindíveis no desenvolvimento da cultura voltada à IA, requerendo que a organização estimule o aprendizado contínuo e a prospecção de experiências sobre problemas semelhantes tratados por outros atores ou em outras circunstâncias. Não menos importante é aproveitar soluções já maduras para a resolução de outros problemas, ainda que, a princípio, pareçam ser de diferentes naturezas. Trata-se, sempre, de exercitar a curiosidade e a criatividade, sem as quais, ao contrário do que normalmente se pensa, as chances de desperdício de tempo e recursos tendem a ser enormes.

Especialmente relevante para a cultura hoje instalada na CONORF é a habilidade de insurgência, ou seja, não se contentar com as soluções existentes. Mesmo que haja um modelo funcionando e atendendo a critérios satisfatórios de desempenho, é responsabilidade institucional de o Órgão manter-se atento aos avanços que possam elevar seu desempenho. Em seus primórdios, a CONORF já possuía uma solução de extração de dados orçamentários que utilizava o Microsoft Access®, ferramenta funcional, mas muito aquém do poder de tratamento de dados de outras opções disponíveis no mercado. Foi a inconformidade – pode-se dizer até a angústia – de um grupo de servidores do órgão por encontrar outra forma de extração que fosse mais simples e eficiente que levou ao nascimento do SIGA Brasil e instalou no contexto informacional da administração pública federal uma ferramenta disruptiva de acesso às contas públicas. Em que pese a extensão dos avanços obtidos, é característica da habilidade de insurgência se fortalecer com as conquistas e, por isso mesmo, colocá-las sob constante questionamento, em vez de cristalizá-las. Com o SIGA Brasil foi assim. Logo após o lançamento da primeira versão e o sucesso expresso nos indicadores de utilização, o sistema passou a ser objeto de novas prospecções para reduzir o peso do conhecimento especializado como requisito para uso do sistema e, desse modo, ampliar o universo de usuários.

Note-se que não havia uma necessidade “objetiva” por tal avanço, já que o sistema atendia às tarefas ordinárias do processo orçamentário, tais como postas. Foi, novamente, a habilidade de insurgência que alimentou o desenvolvimento de novas funcionalidades no



sistema e a oferta de novas versões da ferramenta, promovendo não apenas acessibilidade da informação a um público ampliado, como também abrindo novas perspectivas para o próprio processo orçamentário instalado. A possibilidade de se obter com facilidade séries históricas dos dados orçamentários por meio de consultas plurianuais, por exemplo, abriu todo um novo campo de análise da evolução dos gastos federais, propiciada pela versão do sistema denominada de SIGA Painéis, segunda geração do SIGA Brasil. Atualmente, as duas gerações se complementam e oferecem ao usuário possibilidades bem maiores de acesso às informações. Ainda assim, a equipe continua buscando novas formas de tornar o serviço mais eficiente e de apoiar a melhoria continuada do processo orçamentário.

Outra face da insurgência direciona o olhar não para dentro, mas expande a visão para fora. A CONORF atua em projetos de tecnologia da informação ao lado do PRODASEN e também por meio de parcerias externas com os outros Poderes e organizações não-governamentais. Foi o caso, durante quase uma década, da elaboração de orçamentos temáticos, com o objetivo de apoiar a transparência e participação social de segmentos interessados nas políticas para as mulheres, os indígenas, as crianças e adolescentes, os idosos, as pessoas com deficiência, as pessoas em insegurança alimentar e outros.⁶ Não menos importantes têm sido os relacionamentos externos criados no âmbito do projeto Orçamento Fácil – série de vídeos animados que promovem a democratização do conhecimento orçamentário ao público leigo e, dessa forma, capacita-o a participar do processo orçamentário e a fazer uso do sistema SIGA Brasil. No presente momento, está em vias de formulação a estratégia de levar esse conhecimento a crianças e jovens nas redes públicas e privadas de ensino.

O projeto-piloto de classificação da Receita também buscou parcerias para aprimorar o conhecimento do negócio, em diálogo com a Secretaria do Tesouro Nacional – STN e a Secretaria de Orçamento Federal – SOF. Esse caso evidencia como a busca pelo olhar externo, motivado pela insurgência, permite a exposição do trabalho a outros agentes e proporciona ganhos de conhecimento oriundos de uma inteligência coletiva, muito superior em resultados ao que seria possível a qualquer trabalho isolado e solitário. Problemas relacionados ao orçamento público não são privativos da CONORF e ouvir o que outros agentes têm a dizer sobre eles é primordial.

Já a habilidade de contação de história, no caso da CONORF, foi importante como mecanismo de engajamento e motivação dos próprios membros do órgão. Pouco

⁶ Na versão atual, os orçamentos temáticos foram substituídos pelas consultas publicadas no Painel Cidadão do SIGA Brasil.



conhecedores das tecnologias, seus conceitos e funcionalidades, os servidores poderiam se tornar ariscos a mudanças nas ferramentas de trabalho se não lhes fosse possível visualizar com clareza as novas possibilidades de ganhos no processo continuado de insurgência (ou seja, depois de supridas as necessidades mais prementes). Para tanto, utilizou-se depoimentos de usuários internos e externos para falar aos servidores da Consultoria e do PRODASEN sobre a relevância, o impacto e as dificuldades remanescentes dos projetos de *Business Intelligence* e do SIGA Brasil em suas realidades institucionais. A equipe da área de negócios também trouxe à CONORF as histórias obtidas em visitas a organizações do terceiro setor, destinadas a entender os trabalhos que desenvolvem a partir das informações colhidas do SIGA Brasil e o impacto difuso do sistema em contextos mais amplos do que a administração pública. Essas atitudes provocaram grande impacto na motivação das equipes.

Embora haja a necessidade de se estabelecer uma estratégia mais robusta de engajamento e motivação da CONORF para os projetos de IA, a narrativa sobre o desenvolvimento e os resultados do projeto-piloto de IA para a classificação da receita na forma do Demonstrativo do Resultado Primário tem sido veiculada a esses grupos de servidores, juntamente com histórias sobre possibilidades futuras de IA na CONORF. As histórias revelam o porquê dos projetos e tornam as equipes mais conscientes da importância de apoiarem e se envolverem nesses trabalhos. Essa convicção foi um fator decisivo para o sucesso dos projetos de TI da CONORF.

As soluções de TI da CONORF/PRODASEN têm sido voltadas a usuários e guardam características bem semelhantes às definidas pela habilidade de centralidade no usuário. Em cada aplicação, os projetos definiram quem seriam os usuários e os classificaram quanto à origem, domínio do assunto orçamentário e habilidade digital. Para cada grupo foi definida a abrangência das relações entre produto ofertado, conteúdo e facilidade de uso. De forma mais intensa nos projetos em que o público-alvo primário eram os participantes do processo legislativo, os usuários foram ouvidos no início, no desenvolvimento e na manutenção dos produtos, bem como palestras e cursos foram oferecidos. Com menor intensidade devido a dificuldades de pessoal, tempo e recursos, os usuários externos foram também abordados tanto na etapa de especificação dos produtos quanto na testagem dos protótipos desenvolvidos.

A grande lacuna a suprir nos projetos da CONORF/PRODASEN, quanto à habilidade de centralidade no usuário, é o desenvolvimento de uma estratégia de avaliação sistemática do alcance, usabilidade e impacto dos produtos gerados, como forma de se levantarem evidências que retroalimentem o processo de avanço continuado das ferramentas.



Isso é mais relevante quando se reconhece que há produtos que demandaram mais de uma década de estudos e desenvolvimento tecnológico para atingir certos grupos de usuários. Foi um longo processo para alcançar a usabilidade desconectada do conhecimento prévio de conceitos orçamentários proporcionada pelo Painel Cidadão do SIGA Brasil. Também demandou muito tempo para disponibilizar a visão plurianual detalhada do orçamento própria do Painel Especialista.

O projeto de classificação da receita na forma do Demonstrativo do Resultado primário é direcionado ao público interno das Consultorias de Orçamento do Senado Federal e da Câmara dos Deputados na fase operacional da elaboração orçamentária, mas alcançará o público externo quando essa informação gerencial for disponibilizada. Ainda assim, trata-se apenas da “ponta de um iceberg” quando se compreende o vasto potencial das aplicações de IA tanto para usuários do Congresso Nacional, na análise e apreciação da proposta orçamentária diante das políticas públicas ali propostas, quanto para o público externo, no acompanhamento das políticas públicas lastreadas pelo orçamento.

A esse propósito, vale enfatizar que, dependendo do horizonte de futuros projetos de IA gerenciados pela parceira CONORF/PRODASEN, como, por exemplo, os voltados à avaliação de políticas públicas, será necessária a prospecção de novas bases, que devem ser encaradas como acréscimo ao ativo informacional da Conorf e à sua fronteira de produção de valor. Isso posto, é imprescindível que essa prospecção leve em conta a relação custo-benefício e o retorno sobre o investimento.

7 CONSIDERAÇÕES FINAIS

O planejamento e a execução de aplicações de TI pelo Senado Federal, na área do orçamento da União, são realizados com base na ampla experiência acumulada pelo PRODASEN em conjunto com a CONORF. Trata-se de uma forma de trabalho já amadurecida por mais de duas décadas. No entanto, embora a gestão de projetos de IA deva ser amparada por metodologias já consagradas, como a CRISP, os projetos de IA têm características próprias que os diferenciam de projetos tradicionais de TI. Por isso é imprescindível que os gestores estejam atentos a questões e características intrínsecas ao universo CONORF/PRODASEN. Além da implementação de cultura voltada à inovação, a experiência do projeto-piloto mostra que os três aspectos apontados na seção 4.3 merecem atenção especial na gestão de projetos de IA pela CONORF: valor e a qualidade dos dados;



formação e a multidisciplinariedade da equipe; e desenvolvimento progressivo de capacidades.

É preciso ter consciência da complexidade e possibilidades de exploração das bases de dados orçamentárias. A Consultoria de Orçamentos, em conjunto com o PRODASEN, trabalha com as bases de elaboração e execução orçamentária há mais de duas décadas. Grande esforço tem sido realizado para melhorar a qualidade dos dados, mas ainda há muito trabalho a ser realizado. Dados recebidos do Sistema de Administração Financeira - SIAFI são constantemente homologados pela Consultoria e avaliados diariamente pelos usuários. Uma boa parte desses dados ainda não foi explorada com profundidade e aplicações em IA poderiam desvendar um mundo de riquezas de informações.

Um fator que impacta a qualidade dos dados está associado ao fato de que classificações orçamentárias não são estáticas. Elas são criadas para evidenciar perspectivas do orçamento, mas, além de alterações no domínio, ocorrem mudanças no sentido das classificações, que é o significado de seu conteúdo. A necessidade de avaliar o orçamento sob novos ângulos nem sempre é acompanhada com a mesma dinamicidade pelos sistemas orçamentários. Muitas vezes optou-se por adaptar classificações ao invés de criar novas com o objetivo de evitar impactos nos sistemas orçamentários existentes, como o Sistema Integrado de Administração Financeira – SIAFI. Com isso, perspectivas diferentes foram misturadas em uma mesma classificação. Exemplos dessas mudanças, nos últimos anos, aconteceram nas classificações da despesa Indicador de resultado primário e Identificador de Uso. Na primeira, foram incorporadas informações relacionadas a emendas ao orçamento de execução impositiva e as de Relator Geral e de Comissão. Na segunda, domínios relacionados à marcação da base de despesas destinadas à saúde e à educação. Essas adaptações acrescentam novas informações às classificações, mas acabam por deteriorar, em parte, as informações originais. Alterações nos domínios e nos sentidos das classificações orçamentárias devem ser consideradas nos projetos de aprendizagem de máquina.

A formação da equipe constitui outro fator que deve ser considerado nos projetos de IA. No caso específico da CONORF/PRODASEN, duas questões merecem ser ponderadas com relação à formação da equipe. Em primeiro lugar, o Orçamento da União não é um tema trivial, de rápida absorção. A experiência de trabalho com o PRODASEN nos projetos de sistemas operacionais e gerenciais mostra que o conhecimento do negócio por parte da equipe de TI é fundamental e leva anos para ser consolidado. Em 2020, um pouco antes do início do processo legislativo do Orçamento 2021, a equipe do PRODASEN responsável pelo LEXOR, sistema utilizado na elaboração do orçamento, era composta por cinco membros, quatro



servidores efetivos e um terceirizado. Uma semana antes do início do processo, uma servidora efetiva adoeceu e saiu da equipe e o membro terceirizado recebeu uma oferta de trabalho mais atraente e se desligou do Senado. Em consequência, a equipe foi reduzida quase à metade e todo o investimento de capacitação com o servidor terceirizado se perdeu. Soma-se a isso a dificuldade de contratação de pessoal terceirizado, tanto pelo PRODASEN, quanto pela CONORF. Por essas razões, o desenvolvimento da capacidade interna da CONORF e do PRODASEN é recomendável para projetos de IA. Evidentemente, o tamanho da equipe deve ser proporcional à dimensão do projeto e, com o passar dos anos, uma equipe mista com atores externos pode ser formada, desde que o domínio do conhecimento seja salvaguardado pelas equipes internas.

Existem várias abordagens para a administração de IA no setor público, mas o fato é que o Senado não possui uma política institucionalizada para o tema neste momento. Administradores de projetos apoiados nessa tecnologia no âmbito da CONORF devem ter consciência dessa limitação. Além disso, há restrição de pessoal no PRODASEN, mesmo para atender a CONORF no desenvolvimento de aplicações operacionais do processo orçamentário.

O projeto-piloto de classificação da receita foi realizado por um servidor da CONORF e um do PRODASEN, ambos sem dedicação exclusiva ao projeto. Todo projeto foi desenvolvido em um software de programação livre, o *Python*, em um ambiente também livre, o Google Colab. Portanto, não houve custos com aquisição de softwares e infraestrutura. O ambiente de produção utiliza o mesmo software, mas em infraestrutura suportada pelo PRODASEN, que já era existente antes do projeto. O impacto financeiro, portanto, tende a zero. A experiência do projeto-piloto mostra que é possível empreender soluções em IA sem a mobilização de grandes recursos em infraestrutura de TI ou por meio de parcerias com outros órgãos, como preconizado pela OCDE. A disponibilidade de pessoal capacitado, contudo, é mandatória.

O que parece ser mais importante como estratégia para absorção dessas tecnologias pela CONORF nesse primeiro momento é a capacitação e a difusão da cultura da IA. Apesar de o projeto-piloto focar em uma solução pontual – o aprendizado supervisionado de máquina voltado à classificação – ele possibilitou uma visão ampla sobre aspectos gerais do emprego da ciência de dados. Schmitt (2020c) argumenta que a experiência prática de um projeto real é a melhor forma de aprender IA. Para ele, isso deve ficar claro para os membros da equipe desde o início do projeto: aprendizado em ação.



Novos projetos exigirão outros conhecimentos específicos, como por exemplo, a aplicação de IA para processamento de linguagem natural. Trata-se de um processo iterativo de busca de conhecimento e ampliação de horizontes. O desenvolvimento progressivo de capacidades por meio de experiências com pequenos projetos pode alavancar a expertise da Consultoria e do PRODASEN e prepará-los para empreendimentos em IA mais robustos no futuro.



Referências

- BERRYHILL, J. *et al.* **Hello, World: Artificial Intelligence and its use in the Public Sector.** OECD Working Papers on Public Governance Nº. 36. 2019.
- BRASIL. [Portaria da Secretaria de Orçamento Federal nº 23, de 04 de maio de 2017.](#) **Manual Técnico do Orçamento – MTO 2021(7ª versão).** 2021.
- BRASIL. Portarias do Ministério da Ciência, Tecnologia e Inovações GM nº 4.617, de 6 de abril de 2021 e nº 4.979, de 13 de julho de 2021. **Estratégia Brasileira de Inteligência Artificial.** 2021.
- BRASIL. [Portaria Interministerial nº 5, de 25 de agosto de 2015.](#) **Altera o Anexo I e os arts. 2º e 4º da Portaria Interministerial STN/SOF no163, de 4 de maio de 2001.** 2015.
- BRASIL. Orçamentos da União Exercício Financeiro 2021: Projeto de Lei Orçamentária. Volume I. 2020.
- CARVALHO, T. **Métodos de aprendizado de máquina para classificação das receitas na forma da apuração do Resultado Primário em um cenário de múltiplas classes desbalanceadas: Orçamento Geral da União.** Trabalho de Conclusão de Curso - MBA em Ciência de Dados (CEMEAI), USP - São Carlos. 2021.
- CHAPMAN, P. *et al.* **CRISP-DM 1.0: Step-by-step data mining guide.** SPSS inc, v. 78, p. 1–78, 2000.
- CONWAY, D. **The Data Science Venn Diagram.** Site Drew Conway Data Consulting. 2010. Disponível em: <<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>>. Acesso em: 03 jan. 2021.
- FAWCETT, T e PROVOST, F. **Data Science for Business: what you need to know about data mining and data-analytic thinking.** Organizado por F. Edition. Sebastopol - CA: O’Reilly Media, Inc. 2013.
- OECD. **Artificial Intelligence in Society.** Paris: OECD Publishing. 2019a.
- OECD. **Core skills for public sector innovation: a beta modelo of skills to promote and enable innovation in public sector organizations.** 2017.
- OECD. **OECD Principles on AI.** 2019b. Disponível em: <<https://www.oecd.org/going-digital/ai/principles/>> Acesso em 2021/11/21.
- OECD. **Recommendation of the Council on Artificial Intelligence.** OECD Legal Instruments. 2021.



OLCKERS, A. How to teach a computer to distinguish cats from dogs. 2017. Disponível em: <<https://medium.com/@TheGeekiestOne/how-to-teach-a-computer-to-distinguish-cats-from-dogs-d66cc0679287>>>. Acesso em 20/11/2021.

SHAH C. **A Hands-on Introduction to Data Science**. Reino Unido: Cambridge University Press. 2020.

SCHMITT, M. **An Introduction to Machine Learning for Managers: Everything in your Machine Learning Project**. 2020a. Disponível em: < <https://www.datarevenue.com/en-blog/introduction-to-machine-learning-for-managers>>. Acesso em 23/03/2021.

SCHMITT, M. **The Machine Learning Workflow**. 2020b. Disponível em: <<https://www.dhttps://www.datarevenue.com/en-blog/machine-learning-workflow>>. Acesso em 23/03/2021.

SCHMITT, M. **The Machine Learning Project Checklist**. 2020c. Disponível em: <<https://www.dhttps://www.datarevenue.com/en-blog/machine-learning-project-checklist>>. Acesso em 23/03/2021.

UBALDI B. *et al.* **State of the art in the use of emerging technologies in the public sector**. OECD Working Papers on Public Governance No. 31. 2019.

YIN, R. **Estudo de Caso: Planejamento e Métodos**. Porto Alegre, Bookman, 2015.



Apêndice – Matriz de confusão multiclasse

Interpretação da matriz de confusão multiclasse

Na matriz de confusão multiclasse, as classes reais são representadas nas linhas e as classes preditas, nas colunas. Os valores constantes da interseção entre a linha e a coluna com mesmo nome para classe real e classe predita indicam os verdadeiros positivos (VP) e formam a diagonal principal na matriz de confusão. Os falsos positivos (FP) são os valores da coluna das classes preditas que cruzam as linhas de nomes diferentes das classes reais. Os verdadeiros negativos (VN) são representados pelos valores das células que não pertencem à linha e à coluna com mesmo nome de classes reais e preditas. Os falsos negativos (FN) são representados pelos valores das células da linha da classe real que não cruza a coluna da classe predita de mesmo nome. A figura X abaixo explica a interpretação da matriz de confusão multiclasse para a classe COFINS.

Figura 6 – Interpretação da Matriz de confusão multiclasse para a classe COFINS

Classe Real	Classe Predita				
CIDE - Combustíveis					
COFINS					
Complemento do FGTS					
Concessões e permissões					
Contribuição do salário educação					
	CIDE - Combustíveis	COFINS	Complemento do FGTS	Concessões e permissões	Contribuição do salário educação

Legenda

- Verdadeiro positivo (VP)
- Falso positivo (FP)
- Verdadeiro negativo (VN)
- Falso Negativo (FN)

Fonte: próprio autor



Tabela 10 - Matriz de Confusão Multiclasse da predição com modelo AdaBoost sem engenharia de atributos - Preditores iniciais adotados pela Consultoria de Orçamentos do Senado

Classe Real	Classe Predita																		VP	FP	FN	Precisão	Revocação	F1	Instâncias	Precisão	Revocação	F1		
	(A)	(B)	(C)	(D=A/(A+B))	(E=A/(A+C))	2*(D*E)/(D+E)	Ponderada	Ponderada	Ponderada																					
CIDE - Combustíveis	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	1	1	1	12	12	12	12	
COFINS	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	1	1	1	6	6	6	6	
Complemento do FGTS	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	1	1	2	2	2	2	
Concessões e permissões	0	0	0	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32	0	0	1	1	1	32	32	32	32	
Contribuição do salário educação	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	1	1	1	6	6	6	6	
Contribuição para o PSS do servidor	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	1	1	1	9	9	9	9	
CSLL	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	1	1	1	6	6	6	6	
Demais receitas	0	0	0	0	0	0	0	312	0	0	0	0	0	0	0	0	0	0	0	312	27	0	0,920	1	0,959	312	287,150	312	299,060	
Dividendos	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	4	0	0	1	1	1	4	4	4	4	
Exploração de recursos naturais	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	60	0	0	1	1	1	60	60	60	60	
Imposto de importação	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	5	0	0	1	1	1	5	5	5	5	
Imposto sobre a renda	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	50	0	0	1	1	1	50	50	50	50	
IOF	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	6	0	0	1	1	1	6	6	6	6	
IPI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	45	0	0	0	0	45	0	0	1	1	1	45	45	45	45	
Operações com ativos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	1	1	1	1	1	1	
Outras administradas pela RFB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	51	0	0	51	0	3	1	0,944	0,971	54	54	51	52,457	
PIS/PASEP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	10	0	0	1	1	1	10	10	10	10	
Receita própria (FTS 50 & 81)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1021	1	0	25	1	0,977	0,988	1.046	1.046	1.021,977	1.033,849	
Recursos do RGPS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	1	0	0,944	1	0,971	17	16	17	16,514	
Ressarcimento de desonerações Prev	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1	
																					1.656	28	28			1.684	1.658	1.657	1.657	
																								0,993	0,996	0,994				
																								0,985	0,984	0,984				

Fonte: próprio autor

