



# MODELO DE AUTOMATIZAÇÃO PARA MONITORAMENTO DE DELIBERAÇÕES DE ATOS DE PESSOAL PARA O TRIBUNAL DE CONTAS DA UNIÃO

## AUTOMATION MODEL FOR MONITORING PERSONNEL ACTS DELIBERATIONS AT THE BRAZILIAN FEDERAL COURT OF ACCOUNTS

### Helton Garcia

Mestre em engenharia pela Universidade de Brasília (UnB). Especialista em análise de dados para o controle pelo Instituto Serzedello Corrêa (ISC). Especialista em criptografia e segurança de redes pela Universidade Federal Fluminense (UFF). Graduado em engenharia da computação pelo Instituto Militar de Engenharia (IME). Certified Information Systems Auditor/Information Systems Audit and Control Association (CISA/ISACA). Possui outras 20 certificações profissionais. Subject Matter Expert/ISACA, desde 2019. Auditor Federal de Controle Externo (AUFC) do Tribunal de Contas da União desde 2009.

**Lattes:** <http://lattes.cnpq.br/2906122684922795>

**Orcid:** <https://orcid.org/0009-0003-6916-9955>

**E-mail:** [heltonFG@tcu.gov.br](mailto:heltonFG@tcu.gov.br)

### Edans Flávio de Oliveira Sandes

Doutor em informática pela Universidade de Brasília (UnB). Auditor Federal de Controle Externo (AUFC) do Tribunal de Contas da União (TCU), onde atua com análise de dados. Em 2016, recebeu o prêmio nacional de melhor tese em ciência da computação no Prêmio Capes de Tese, o que lhe permitiu a realização de um pós-doutorado no ano seguinte.

**Lattes:** <http://lattes.cnpq.br/5132071359802549>

**Orcid:** <https://orcid.org/0000-0001-9502-7799>

**E-mail:** [edansFS@tcu.gov.br](mailto:edansFS@tcu.gov.br)

## RESUMO

Uma das competências constitucionais do Tribunal de Contas da União (TCU) é apreciar, para fins de registro, a legalidade de atos de pessoal (Brasil, 1988). O crescimento do estoque de deliberações a serem monitoradas constitui grande desafio, agravado pela necessidade de racionalização de recursos. Uma solução é empregar inteligência artificial (IA). Este trabalho tem como objetivo propor modelo baseado em aprendizagem de máquina para automatizar o monitoramento de deliberações relacionadas a atos



de pessoal, incluindo cessação de pagamentos envolvendo atos de admissão e de concessão de aposentadoria e reforma. A arquitetura do modelo é formada por cinco conjuntos de módulos e emprega algoritmos não supervisionados e supervisionados (Naive-Bayes), seguindo a metodologia CRISP-DM. Cenários sobre cessação de pagamentos foram submetidos a cruzamento de dados com informações extraídas do Sistema Integrado de Administração de Pessoal (Siape). O modelo atendeu aos critérios de sucesso de mineração de dados com acurácia de 93,89% e precisão de 96% e 97%, nas categorias de interesse. O uso de *oversampling* reduziu as divergências de 1,65% para 0,23%. Os resultados foram considerados satisfatórios pela unidade responsável por atos de pessoal e por áreas de coordenação de negócio do TCU. Trata-se de trabalho inovador, dado não haver no TCU, até então, solução especialista aplicável a tal objeto.

**Palavras-chave:** mineração de dados; aprendizagem de máquina; clusterização; classificador Naive-Bayes; atos de pessoal; monitoramento.

## ABSTRACT

*One of the constitutional responsibilities of the Brazilian Federal Court of Accounts (TCU) is to assess, for registration purposes, the legality of personnel acts (Brazil, 1988). The growing volume of deliberations to be monitored poses a major challenge, aggravated by the trend towards resource rationalization. A promising solution is to employ Artificial Intelligence (AI). This study presents a machine learning (ML) based model designed to automate the monitoring of deliberations related to personnel acts, including the cessation of payments involving admissions, retirements, and personnel reforms. The model architecture consists of five sets of modules, containing unsupervised and supervised (Naïve-Bayes) ML algorithms, following the CRISP-DM methodology. Scenarios involving cessation of payments were cross-referenced with data extracted from the Integrated Personnel Administration System (Siape). The model met the data mining success criteria, achieving 93.89% accuracy and 96% and 97% precision for the categories of interest. Additionally, the use of oversampling reduced the percentage of discrepancies from 1.65% to 0.23%. The results were considered satisfactory by the personnel management unit, as well as by the TCU's business coordination areas. This represents an innovative development, as no specialized solution of this nature had previously been implemented at the Court.*

**Keywords:** data mining; machine learning; clustering; Naïve-Bayes classifier; personnel acts; monitoring.

**Data de submissão:** 6 de maio de 2024  
**Data de aprovação:** 16 de junho de 2024



## 1. INTRODUÇÃO

O TCU, órgão de controle externo do Poder Legislativo brasileiro, tem como responsabilidade prevista na Constituição federal apreciar, para fins de registro, a legalidade dos atos de pessoal.

Constatou-se, por meio de estudos e interações realizadas na unidade técnica do TCU responsável pelos atos de pessoal, que havia necessidade de redução do estoque de processos de monitoramento de pessoal. Atualmente, a vazão de entrada tende a ser maior que a de saída, acarretando histórico deficitário. Apesar da substituição de sistemas com novas funcionalidades e das soluções informatizadas, que colaboravam na automatização de rotinas para apreciação de atos de pessoal, fazia-se necessário verificar se as deliberações e recomendações advindas da apreciação dos atos eram corretamente cumpridas. Aliada a essa necessidade, evidenciou-se urgência na racionalização de recursos, decorrente da escassez de pessoal.

O objetivo geral do presente trabalho foi propor modelo baseado em aprendizagem de máquina, para tornar mais eficiente o monitoramento de acórdãos de pessoal. A proposta visou a automatizar o acompanhamento de deliberações acerca da cessação de pagamentos referentes a atos de admissão, concessão de aposentadoria e reforma publicados entre 2014 e 2019.

Para tanto, foram estabelecidos os seguintes objetivos específicos: *i)* assegurar alto grau de acurácia na classificação dos acórdãos monitorados, conforme critérios de sucesso definidos pela unidade de negócio; *ii)* identificar padrões de recorrência, bem como obter indícios de possíveis inconsistências ou fragilidades nos procedimentos avaliados, por meio de análise exploratória de dados, de forma a subsidiar o planejamento de ações de controle. A aplicação imediata dos resultados no ambiente de produção da unidade técnica interessada foi considerada viável.

O estudo apresentado neste artigo está organizado da seguinte forma: na segunda seção, aborda-se a revisão de literatura; na terceira seção, descreve-se a metodologia; na quarta seção, que contém o desenvolvimento da pesquisa, apresentam-se e discutem-se os resultados obtidos; e, na quinta seção, expõem-se as considerações finais, bem como as sugestões para trabalhos futuros.

## 2. REVISÃO DE LITERATURA

Macedo (2004) esclarece que, até o final da década de 1980, o TCU não dispunha de solução sistematizada para apreciação de atos de pessoal. O controle era realizado por meio da análise de “processos convencionais”, ou seja, compostos por documentos em papel numerado, autuado, com páginas rubricadas e cópias dos fatos geradores do direito à inativação, à percepção de parcelas de proventos e à concessão de pensão, remetidos pelos diversos setores de recursos humanos dos órgãos jurisdicionados.

A partir da década de 1990, o referido controle passou a ser progressivamente feito por meio do Sistema de Apreciação e Registro de Atos de Concessão (Sisac). A substituição definitiva pela análise informatizada ocorreu a partir de 2001 (Brasil, 2018).



Branco (2014) dispõe que, em 2013, o Sisac continha mais de 4 milhões de atos de pessoal, tendo cada ato mais de 100 campos. No mesmo ano, o TCU havia alcançado a marca de 1 milhão de atos apreciados de forma automática, sem intervenção humana, por meio de instrumentos denominados tipologias. Entende-se como tipologia a descrição de situação ou sequência de atos que pode indicar diretamente potencial de ilicitude ou medir aspecto de risco (Balaniuk, 2010).

Em 2018, o Sisac foi substituído pelo e-Pessoal (Brasil, 2018), que trazia novas funcionalidades, a saber: *i)* realização de análises automáticas já na entrada dos dados; *ii)* detecção precoce de problemas com a documentação; e *iii)* possibilidade de inclusão de documentos, como laudos médicos, comprovantes de recolhimentos previdenciários ou decisões judiciais. Entre os benefícios esperados com a nova plataforma, citam-se os seguintes: *i)* ganhos de economicidade e eficiência; *ii)* encaminhamento direto dos processos aos auditores responsáveis; *iii)* agilidade no tratamento de problemas que, até então, exigiam diligências consideradas onerosas.

Por um lado, as soluções informatizadas contribuíram para automatizar as rotinas relacionadas à apreciação de atos de pessoal. Por outro lado, fazia-se necessário verificar se as deliberações resultantes dessas análises estavam sendo corretamente cumpridas e se resultados esperados estavam sendo alcançados.

Até o início da pesquisa, em 2019, o monitoramento das deliberações relacionadas a atos de pessoal ainda era realizado manualmente, conforme informações da unidade técnica de auditoria responsável pela análise desses procedimentos no TCU. Verificou-se que o acompanhamento dos acórdãos era tratado como prioridade, devido à tendência histórica de aumento no estoque de deliberações a serem monitoradas. No entanto, remanescia a oportunidade de automatização desse processo, visando a obter maior eficiência.

Outro aspecto importante é o cenário de escassez de recursos humanos, realidade que afeta não apenas o Tribunal, mas toda a Administração Pública federal. Atualmente, a reposição de pessoal é alternativa pouco viável, diante das restrições orçamentárias do setor público, mesmo com o crescimento das demandas. Ao mesmo tempo, os objetivos organizacionais tendem a crescer, tornando a eficiência fator crítico para o sucesso da instituição. Nesse contexto, uma possível solução é adotar tecnologias baseadas em IA, capazes de automatizar atividades que antes eram realizadas por humanos.

O termo IA foi cunhado em 1956, por John McCarthy, considerado o pai do tema, e definido como “a ciência e engenharia de fabricação de máquinas inteligentes” (Mondal, 2020; Paliwal, Bharti; Mishra, 2020). Na mesma década, surgiu o termo “aprendizagem de máquina” – *machine learning* (ML) – que aproveita conceitos de diversas áreas da ciência, como a probabilidade, a estatística e a ciência da computação (El Naqa; Murphy, 2015). Samuel (1959) definiu esse campo como o “estudo que oferece a máquinas a possibilidade de aprender sem serem explicitamente programadas”.

Outra definição do termo é proposta por Mitchell (1997, p. 2, tradução nossa), que conceitua aprendizagem de máquina como “um programa de computador que aprende com dada experiência E, em relação a alguma tarefa T e a uma medida de desempenho P, de modo que



seu desempenho em T, medido por P, melhora com a experiência E”. Essas tecnologias são capazes de transformar dados brutos em serviços úteis, aumentando significativamente o valor extraído das informações disponíveis.

Monitoramento é o instrumento de fiscalização utilizado pelo Tribunal para verificar o cumprimento de suas deliberações e os resultados delas advindos, segundo o art. 243 da Resolução-TCU 246, de 30 de novembro de 2011, que dispõe sobre o Regimento Interno do Tribunal de Contas da União (RITCU) (Brasil, 2011). É um dos instrumentos de fiscalização previstos no RITCU, além de auditoria, acompanhamento, inspeção e levantamento. Assim como a auditoria e o acompanhamento, o monitoramento deve integrar o plano de fiscalização elaborado pela Presidência do TCU, em consulta com os relatores das listas de unidades jurisdicionadas, e ser aprovado pelo Plenário, em sessão de caráter reservado (arts. 238 e 244 do RITCU).

Nesse contexto, a Resolução-TCU 353, de 22 de março de 2023, veio complementar o marco normativo, ao estabelecer procedimentos para exame, apreciação e registro dos atos de admissão de pessoal e de concessão de aposentadoria, reforma e pensão pelo TCU (Brasil, 2023b).

Ainda em 2023, foi publicada reestruturação interna que atribuiu à Unidade de Auditoria Especializada em Pessoal (AudPessoal) (Brasil, 2023a) a responsabilidade pelo monitoramento de acórdãos do TCU proferidos em processos relacionados a atos de pessoal. A atuação abrange admissão, concessão de aposentadoria, pensão civil, reforma, pensão militar e pensão especial de ex-combatente, conforme estabelecido no inciso II do art. 5º da Portaria 1, de 10 de junho de 2019, da Secretaria de Fiscalização de Pessoal (Sefip) (Brasil, 2019b).

**Figura 1 – Distribuição de acórdãos de pessoal por ano**



Fonte: elaborada pelos autores.

As atividades de monitoramento são geridas e controladas por sistema informatizado, mas o processamento e a análise são feitos manualmente. A conferência do cumprimento de deliberações é feita confrontando-se o disposto no acórdão com a situação presente do jurisdicionado afetado, o que inclui consultas às bases de dados do órgão. Trata-se, portanto, de procedimento oneroso e que demanda alocação de recursos que poderiam ser dedicados a outras atividades de fiscalização da unidade.



O estoque de deliberações a serem monitoradas apresenta tendência histórica de crescimento. Entre 2014 e 2017, o quantitativo era de até 16.000 acórdãos de pessoal por ano (Figura 1). De 2018 a 2021, houve variação significativa para mais de 20.000 por ano, com breve queda, em 2022, e retorno à tendência no ano seguinte. Cada acórdão pode apresentar uma ou mais deliberações.

### 3. METODOLOGIA

A metodologia aplicada ao trabalho foi a *Cross-Industry Standard Process for Data Mining* (CRISP-DM), concebida para guiar projetos de mineração de dados (Chapman, 1999; Wirth; Hipp, 2000) e aplicável a diversas áreas, tais como educação, saúde, engenharia e governo (Schröer; Kruse; Gómez, 2021). Como modelo de processo, fornece visão geral do ciclo de vida da mineração de dados. Como metodologia, inclui descrições das fases típicas de um projeto e das tarefas envolvidas em cada fase, bem como explicação dos relacionamentos entre as tarefas. A sequência das fases não é rigorosa.

No presente trabalho, decidiu-se simplificar o escopo inicial da pesquisa, direcionada, então, para os atos de pessoal envolvendo admissão e concessão de aposentadoria e de reforma de pessoal publicados entre 2014 e 2019.

#### 3.1 ENTENDIMENTO DO NEGÓCIO

O objetivo desta fase foi coletar as expectativas da organização quanto aos possíveis benefícios da aplicação da mineração de dados. Para isso, procedeu-se a estudo acerca dos instrumentos de fiscalização adotados pelo TCU, complementado por reuniões com a unidade técnica responsável pelo monitoramento e pela auditoria de atos de pessoal.

O art. 1º da Resolução TCU 255, de 26 de setembro de 1991 (Brasil, 1991), dispõe sobre a apreciação, pelo TCU, para fins de registro, da legalidade dos atos de admissão de pessoal na Administração direta e indireta, incluídas as fundações instituídas e mantidas pelo poder público, excetuadas as nomeações para cargo de provimento em comissão; bem como dos atos de concessão de aposentadorias, reformas e pensões, ressalvadas melhorias posteriores que não alteram o fundamento legal do ato concessório.

Para os atos de pessoal considerados legais, a apreciação para fins de registro consiste em reconhecer a legalidade do ato e registrar as informações pertinentes. No caso de admissões, são incluídos dados como o nome do servidor ou do empregado público, o código do órgão, o ano e o regime de ingresso. Para concessões de aposentadoria, reforma ou pensão, registram-se o nome do inativo ou do(s) beneficiário(s), a espécie de concessão, o código do órgão e o ano correspondente, conforme previsto no art. 10 da Resolução TCU 255/1991.

Para os atos considerados ilegais, o TCU fixa prazo para que o órgão de origem adote as medidas corretivas indicadas para o exato cumprimento da lei. Finalizado o prazo, o ato pode ser julgado ilegal, e o registro, negado. O julgamento de ilegalidade e a negativa de registro



podem implicar revogação do ato de admissão, devendo o órgão de origem promover a dispensa da pessoa ilegalmente admitida e cessar todo e qualquer pagamento feito a ela, a partir da publicação da decisão no Diário Oficial da União (DOU), sob pena de obrigação de ressarcimento, pelo responsável, das quantias pagas após a referida data.

Quando a ilegalidade do ato de concessão consiste na outorga de vantagens indevidas (art. 20 da Resolução TCU 255/1991), a recusa de registro obriga o órgão concedente a cessar, no todo ou em parte, o pagamento dos proventos ou dos benefícios, a partir da publicação da decisão do TCU. Caso o pagamento não seja suspenso ou na existência de indício de procedimento culposo ou doloso, o TCU pode converter o processo em tomada de contas especial, para apurar responsabilidades e promover ressarcimento das despesas ilegalmente efetuadas.

### 3.2 ENTENDIMENTO DOS DADOS

Esta fase envolveu acesso aos dados disponíveis e exploração deles, para fins de mineração, possibilitando determinar a qualidade das informações e descrever os resultados obtidos na documentação do projeto (Chapman, 1999; Wirth; Hipp, 2000). Neste contexto, foram realizadas as seguintes atividades: *i)* levantamento e estudo das bases e dos modelos de dados; *ii)* identificação dos campos disponíveis e sua formatação; *iii)* obtenção de autorização para acesso aos sistemas e aos dados mencionados, com permissão restrita à leitura; *iv)* levantamento de formatos disponíveis para entrega das informações.

Foram consideradas, como principais fontes as seguintes: *i)* busca textual de jurisprudência relacionada a acórdãos; *ii)* Sistema de Apoio ao Gerenciamento e Acompanhamento das Sessões (Sagas); *iii)* base de dados do Siape; e *iv)* dados do Registro de Apreciações, Deliberações e Acompanhamento de Resultados (Radar) e do Radar Expandido (Radex), os quais serão detalhados a seguir.

Busca textual - é a base oficial de pesquisa de acórdãos do TCU, com dados desde 1992. Qualquer pessoa pode acessar os registros, atualizados a partir do sistema Sagas, que apoia gabinetes e sessões dos colegiados.

Sagas - é o sistema responsável pela oficialização dos acórdãos após as deliberações ocorridas no plenário e nas câmaras. A busca textual permite a extração de informações tanto dos documentos quanto de seus metadados armazenados no banco de dados.

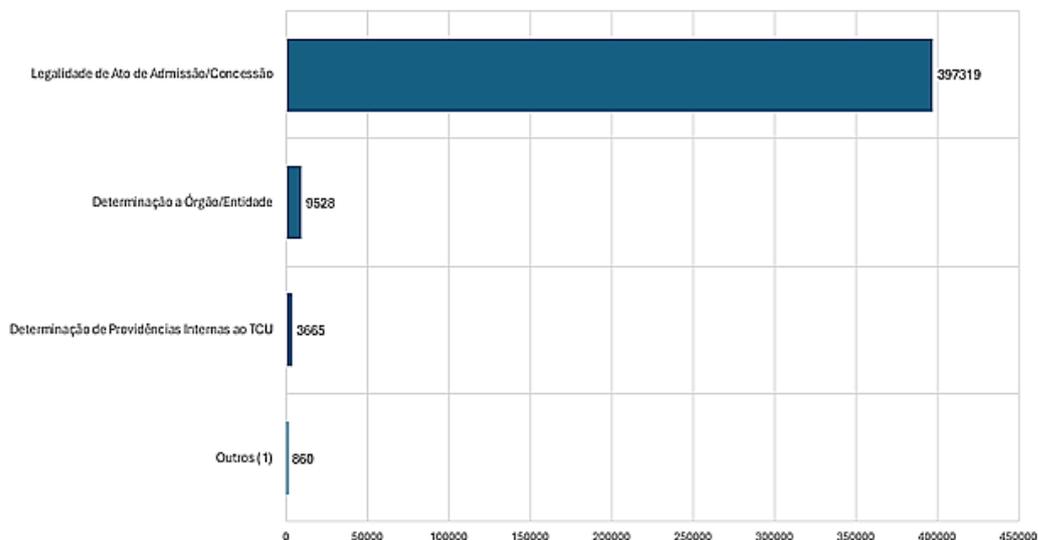
Siape - é o sistema que centraliza o processamento da folha de pagamento de servidores de órgãos federais vinculados ao Tesouro. Abrange servidores ativos, aposentados e pensionistas sob diversos regimes (RJU, CLT, temporários) (Brasil, 1990). Segundo relatório de 2019, 213 unidades jurisdicionadas estavam integradas ao Siape (Brasil, 2019a).

Radar/Radex - são sistemas utilizados para a gestão e o monitoramento de deliberações do TCU (Macedo, 2004). O Radar registra as apreciações e os respectivos resultados. O Radex, seu sucessor, introduziu melhorias no cadastro e na distribuição de acórdãos para as unidades técnicas. A coleta de dados considerou deliberações proferidas entre 2014 (início



da temporalidade relevante) e 2019 (corte metodológico), conforme informado pela unidade técnica responsável, considerando os prazos de cumprimento estabelecidos pela Resolução TCU 255/1991 (Brasil, 1991).

**Figura 2 – Tipos de deliberação<sup>1</sup>**



Fonte: elaborada pelos autores.

Além disso, deliberações associadas ao tipo 25 “Legalidade de ato de admissão/concessão” foram excluídas do escopo, pois não geram necessidade de monitoramento, por trivialidade. Entre as demais, o escopo foi limitado ao tipo 1 “Determinação a órgão/entidade”, por ser majoritário e figurar entre itens a serem monitorados, ou seja, por conter deliberações que versam sobre cessão de pagamento, aspecto de interesse deste trabalho. O tipo 1 representa 67,8% do escopo remanescente.

### 3.3 PREPARAÇÃO DE DADOS

Na etapa de preparação de dados, as bases selecionadas na fase anterior foram submetidas à identificação dos campos relevantes, à formatação e à obtenção de autorizações para acesso aos sistemas e aos dados mencionados, com permissões restritas à leitura. A etapa também incluiu o levantamento de formatos disponíveis para extração das informações. O CRISP-DM considera essa fase como uma das mais importantes e onerosas, podendo consumir até 70% do tempo do projeto (Chapman, 1999).

1 Outros: Abertura de novo processo/Apartado; Determinação de realização de fiscalização; Prorrogação de prazo de deliberação; Apensamento do atual processo a outro(s); Citação de responsável; Autorização de recolhimento parcelado de dívida; Aplicação de multa a responsável; Modificação da natureza do processo; Julgamento das contas do responsável; Imputação de débito a responsável; Audiência de responsável; Sobrestamento do julgamento; Recomendação a órgão/entidade; Diligência a órgão/entidade; Tornar deliberação sem efeito; Retirada da chancela de sigiloso; Arquivamento de processo; Conhecimento/Provimento de denúncia/Representação/Solicitação/Consulta.



Para a busca textual de acórdãos, foi realizada extração de dados diretamente dos arquivos textuais, por meio de expressões regulares. Importante destacar que a junção da busca textual com a base do Radar (seguinte) foi necessária por várias razões: *i)* a extração de partes interessadas diretamente do conteúdo textual dos itens e dos subitens mostrou-se onerosa para o tempo do projeto; *ii)* isso também foi observado para a extração do tipo do ato, pois, ao contrário do que se esperava, o domínio de arquivos analisados continha disposições não uniformes, que dificultavam a extração completa dos arquivos textuais; *iii)* em testes substantivos, verificaram-se registros com menção indireta de interessados, o que inviabilizou a extração de dados simplesmente por inexistência de menção explícita.

Um exemplo de extração textual encontra-se na Figura 3, acerca do Acórdão 10.974, de 24 de novembro de 2015 (Brasil, 2015), e tabulação de campos de interesse. Então, além da identificação dos interessados em cada deliberação, por meio de NOME, CPF e tipo do ato (TIPO\_ATO), os seguintes campos foram extraídos: “NUM\_PROCESSO”, “ANO\_PROCESSO”, “NUM\_ACORDAO”, “ANO\_DECISAO”, “APRECIADOR”.

Figura 3 – Exemplo de extração de dados textuais – Acórdão 10.974/2015-TCU-2a Câmara

The screenshot displays a document viewer interface. At the top, a table lists data for two records (220 and 221). Below the table, a document titled 'AC-10974-2015-2.txt' is open. The document content includes a header 'ACÓRDÃO Nº 10974/2015 - TCU - 2ª Câmara' and a list of items. Red boxes highlight specific fields in the table and corresponding text in the document, such as 'Processo TC 025.757/2015-6', 'Interessados: Al...ho (CPF 11...34) e Ed...ia (CPF 04...72)', and 'Unidade Técnica: Secretaria de Fiscalização de Pessoal - Sefip'. A blue box on the right side of the document highlights a list of file names, with 'AC-10974-2015-2.txt' at the bottom. A blue arrow points from this box to the document content.

	nome	cpf	num_processo	ano_processo	tipo_ato	num_acordao	ano_decisao	apreciador
220	lho 1	4	257	2015	APOSENTADORIA	974	2015	Segunda Câmara
221	rea 0	2	257	2015	APOSENTADORIA	974	2015	Segunda Câmara

ACÓRDÃO Nº 10974/2015 - TCU - 2ª Câmara

- Processo TC 025.757/2015-6.
- Grupo I - Classe V - Aposentadoria.
- Interessados: Al...ho (CPF 11...34) e Ed...ia (CPF 04...72).
- Unidade: Fundação Universid...
- Relatora: ministra Ana Arraes.
- Representante do Ministério Público: promotor Sergio Ricardo Costa Caribé.
- Unidade Técnica: Secretaria de Fiscalização de Pessoal - Sefip.
- Representação legal: não há.
- Acórdão:

VISTOS, relatados e discutidos estes atos de concessão de aposentadoria a ex-servidores da Uni... nos quais foi constatado o pagamento ilegal do percentual de 3,17%, a título de diferença de URV, sem incorporação de tal vantagem em reestruturações salariais posteriores.

ACORDAM os ministros do Tribunal de Contas da União, reunidos em sessão da 2ª Câmara, diante das razões expostas pela relatora e com fundamento no art. 71, inciso III, da Constituição Federal, c/c os arts. 1º, inciso V, 3º, inciso II, e 45 da Lei 8.443/1992, e na Súmula TCU 106, em:

- considerar ilegais e negar registro aos atos de Al...ho e Ed...ia;
- dispensar o recolhimento das quantias indevidamente recebidas de boa-fé pelos interessados até a data da notificação desta deliberação à unidade jurisdiccionada;
- determinar à U... que:
  - promova o ajuste das parcelas relativas ao percentual de 3,17%, mesmo que deferidas judicialmente, levando em conta a reestruturação da carreira de magistério superior promovida pela Lei 12.772/2012, nos termos do art. 10 da NR 2 225/2001 e do entendimento deste Tribunal, consubstanciado no acórdão 2.161/2005-Plenário;
  - reanalise o ato concessório de Ed...ia, tendo em vista o equívoco relativo à sua prevista no art. 62-A da Lei 8.112/1990;
  - emita novos atos, livres da irregularidade apontada, e os submeta ao TCU pelo Sistema de A... de Admissão e Concessões (SIAAC) no prazo de 30 (trinta) dias; e
  - comunique aos interessados o teor deste acórdão e os alerte que, no caso de não provimento de interposto junto ao TCU, deverão ser repostos os valores recebidos após a ciência do acórdão pela U... no prazo de 30 (trinta) dias, comprovantes das datas de ciência pelos interessados.

10. Ata nº 41/2015 - 2ª Câmara.

11. Data da Sessão: 24/11/2015 - Ordinária.

12. Código eletrônico para localização na página do TCU na Internet: AC-10974-41/15-2.

13. Especificação do quorum:

- Ministros presentes: Raimundo Carreiro (Presidente), Augusto Nardes e Ana Arraes (Relatora).
- Ministro-Substituto convocado: Augusto Sherman Cavalcanti.

13.3. Ministros-Substitutos presentes: Marcos Benquerer Costa e André Luís de Carvalho.

(Assinado Eletronicamente)  
 RAIMUNDO CARREIRO  
 (Assinado Eletronicamente)]  
 ANA ARRAES

Presidente  
 Relatora

busca\_textual

- AC-622-2015-1.txt
- AC-622-2015-2.txt
- AC-622-2015-P.txt
- AC-623-2015-1.txt
- AC-623-2015-2.txt
- AC-623-2015-P.txt
- AC-624-2015-1.txt
- AC-624-2015-2.txt
- AC-624-2015-P.txt
- AC-625-2015-1.txt
- AC-625-2015-2.txt
- AC-625-2015-P.txt
- AC-626-2015-1.txt
- AC-626-2015-2.txt
- AC-626-2015-P.txt
- AC-627-2015-1.txt
- AC-627-2015-2.txt
- AC-10974-2015-2.txt

Fonte: elaborada pelos autores.



Figura 4 – Extração de informações armazenadas em campos textuais

```

DESCR_1:
[['Determinação a Órgão/Entidade: Instituto Federal de Educação, Ciência e Tecnologia Catarinense: 9.3. determinar ao Instituto Federal de Educação, Ciência e Tecnologia Catarinense que:\r\n9.3.1. no prazo de 15 (quinze) dias a contar da ciência desta decisão, cesse pagamentos relativos ao ato impugnado e comunique ao Tribunal as providências adotadas, sob pena de solidariedade da autoridade administrativa no ressarcimento das quantias pagas após essa data, sem prejuízo das sanções previstas na Lei 8.443/1992;\r\n PRAZO PARA CUMPRIMENTO: 15 DIAS']]

ORGAO:
[['Instituto Federal de Educação, Ciência e Tecnologia Catarinense']]

ENC:
[['9.3. determinar ao Instituto Federal de Educação, Ciência e Tecnologia Catarinense que 9.3.1. no prazo de 15 (quinze) dias a contar da ciência desta decisão, cesse pagamentos relativos ao ato impugnado e comunique ao Tribunal as providências adotadas, sob pena de solidariedade da autoridade administrativa no ressarcimento das quantias pagas após essa data, sem prejuízo das sanções previstas na Lei 8.443/1992;']]

PRAZO:
[['15 DIAS']]

```

Fonte: elaborada pelos autores.

Os indicadores numéricos ordinais foram substituídos por suas formas por extenso. No campo “CPF”, o formato foi padronizado para representar textos fixos com onze caracteres, com inclusão de zeros à esquerda (“01234567890”), visando a manter equivalência com o formato utilizado na base de dados do Siape. Para o campo “APRECIADOR”, foi realizado ajuste de conteúdo, para preservar a convenção adotada nos registros do sistema Radar. Alguns dados também foram ajustados, a fim de que pudessem melhor se adaptar às fases posteriores do trabalho. Para isso, utilizou-se a linguagem de programação “Python”, com auxílio da biblioteca “Pandas”, para fazer as devidas transformações.

Sobre a base de dados do Radar, diversos campos foram convertidos para os formatos de número inteiro<sup>2</sup> e de cadeia de caracteres<sup>3</sup>. Além disso, a partir do campo descritivo das deliberações, foram extraídos atributos individualizados referentes a: *i)* órgão ou entidade jurisdicionada; *ii)* prazo para cumprimento da deliberação; e *iii)* encaminhamento do dispositivo (Figura 4). Para a base de dados do Siape, também foram realizadas conversões para o formato inteiro<sup>4</sup> e selecionados apenas os campos de interesse<sup>5</sup>, especificamente aqueles relacionados às rubricas vigentes. A arquitetura será detalhada na subseção a seguir.

2 Campos do sistema Radar convertidos para o tipo inteiro: COD\_PROCESSO, PROCESSO, SEQ\_DELIBERACAO, TIPO\_DELIBERACAO, ANO\_PROCESSO, NUM\_PROCESSO, COD\_APRECIACAO, COD\_TIPO\_DECISAO e ANO\_DECISAO.

3 Campos do sistema Radar convertidos para texto: DESCR, NUMDELIB, ORGAO, ENC, PRAZO, SIGLA e APRECIADOR.

4 Campos do sistema Siape convertidos para texto: NUM\_MATRICULA, COD\_ORGAO e COD\_TIPO\_RUBRICA.

5 Os campos de interesse foram extraídos das tabelas HIST\_SERVIDOR, RUBRICA e TIPO\_RUBRICA.



### 3.4 MODELAGEM

A fase de modelagem compreendeu o desenvolvimento de protótipo utilizando técnicas de mineração de dados, com o objetivo de identificar padrões relevantes nos dados analisados. Dentre as técnicas utilizadas, podem ser citados o agrupamento de registros com características semelhantes (clusterização) e a identificação dos perfis relacionados a cada agrupamento (classificação dos *clusters*).

Além disso, também foi realizada extração de informações (*parser*) a partir das deliberações contidas em um ou mais *cluster(s)* com grau satisfatório de homogeneidade. A validação do modelo seguiu critérios de sucesso da mineração de dados, como a utilização da matriz de confusão e o cálculo da acurácia.

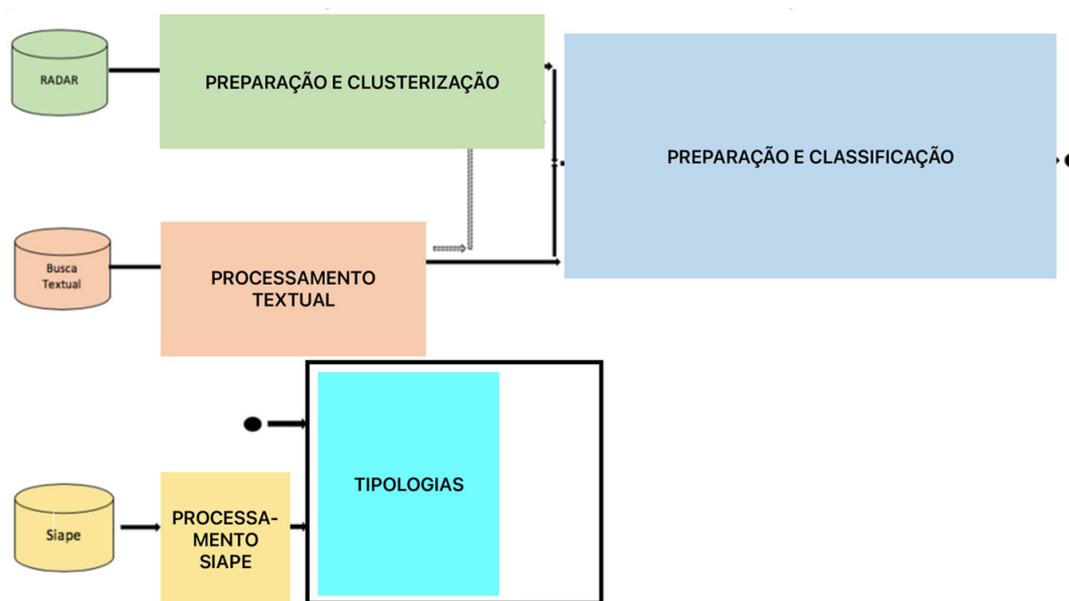
Diversas iterações foram necessárias até o desenvolvimento da versão final do modelo, com sucessivos ajustes de parâmetros, visando à melhoria dos resultados obtidos. O próprio modelo passou por diferentes reestruturações, particularmente impostas para minimizar problemas de desempenho. Por essa razão, alguns aspectos relacionados à fase anterior podem ser mencionados nesta subseção. Além disso, o modelo foi dividido em vários módulos ou arquivos (*scripts*), para facilitar o desenvolvimento incremental, a leitura do código e a depuração de eventuais erros.

A arquitetura da solução é proposta nesta subseção, como ilustrado na Figura 5. Para facilitar a visualização, conjuntos de módulos e elementos sequenciais foram identificados e organizados por cores.

- VERDE - associada ao Radar e à clusterização. É formada por módulos responsáveis pela coleta, limpeza, filtragem e aplicação de algoritmo de clusterização.
- VERMELHA - formada por módulos responsáveis pela filtragem, *parser* e limpeza de dados provenientes da busca textual de acórdãos. Por isso o nome “processamento textual”.
- AZUL - responsável por integrar os resultados obtidos nas duas primeiras etapas (verde e vermelha) e proceder à classificação das deliberações por meio de algoritmo supervisionado, particularmente para rotular deliberações que versam sobre pagamento ou não. Em seguida, os dados são preparados para a fase de cruzamento.
- AMARELA - constituída por módulo de consulta e limpeza de dados provenientes do Siape.
- AZUL CLARA - responsável pela busca de dados sobre cessação de pagamento no Siape.



Figura 5 – Visão geral da arquitetura



Fonte: elaborada pelos autores.

Dentre os 13 módulos constantes do modelo, são destacados, a seguir, os dois módulos principais, que contêm os algoritmos de aprendizado de máquina não supervisionados (módulo 5) e supervisionados (módulo 8), além de comentários gerais, para facilitar a contextualização.

O módulo 5 tem como função proceder ao agrupamento de deliberações, para facilitar a seleção de deliberações sintaticamente afins. É parte integrante do conjunto de módulos de cor verde. Foi escolhida essa técnica não supervisionada de aprendizagem de máquina por ser uma abordagem em que a inferência de resultados ocorre sem o emprego de rótulos de saída (*targets*) associados ao conjunto de dados (Mondal, 2020). Na clusterização/no agrupamento (*clustering*), ocorre o particionamento de elementos em grupos naturais chamados *clusters*, segundo critérios de similaridade previamente definidos (Zaki; Meira Jr., 2014).

A estratégia inicial foi estabelecer agrupamentos com alta similaridade sintática; por isso a necessidade de escolha restritiva de parâmetros<sup>6</sup>. Além disso, não foi definido o quantitativo de *clusters* (K), segundo justificativas a seguir.

Após várias iterações na execução do módulo, foi escolhida a função de clusterização hierárquica aglomerativa (Bouguettaya; Yu; Liu; Zhou, 2015), a fim de agregar elementos ou grupos similares (*clusters*) e organizá-los em uma estrutura hierárquica. Nesse método, a aglomeração ocorre de baixo para cima (*bottom-up*), ou seja, inicia-se a partir de elementos individuais que são progressivamente combinados, em contraste com a abordagem divisiva, de cima para baixo (*top-down*), que parte de um grupo único e o subdivide sucessivamente.

6 A versão final apresentou os seguintes parâmetros de configuração: *fcluster*, *single*, *metric: Jaccard*. Para a classe *CountVectorizer*: *ngram\_range = (2,5)*, *min\_df = 2*, *max\_df = 0,1*.



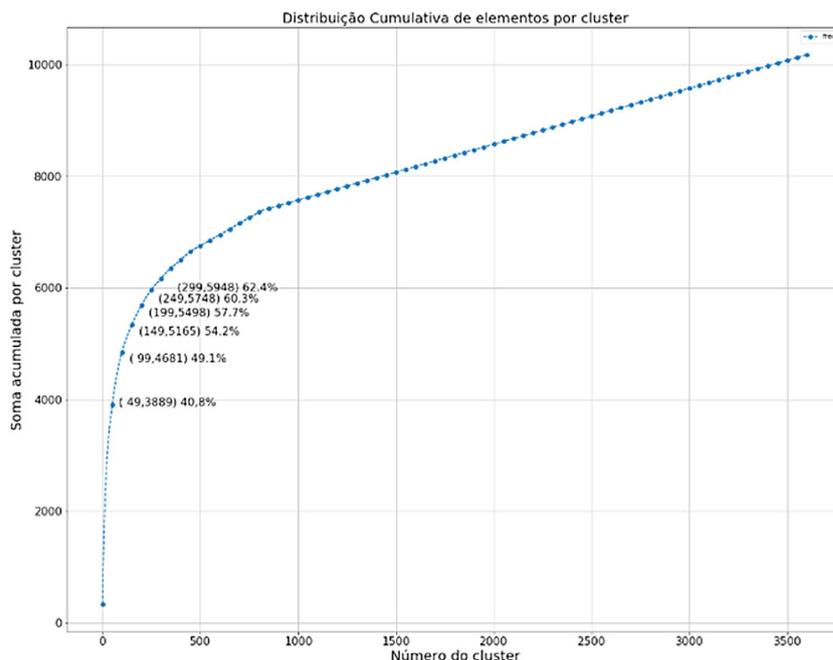
A métrica adotada para o processo de aglomeração foi a distância de *Jaccard*, cuja escolha se baseia na análise apresentada por Pandit e Gupta (2011), que recomendam seu uso para classificação textual de documentos, em detrimento da distância euclidiana, considerada padrão para problemas geométricos. A distância de *Jaccard* mensura a similaridade entre dois conjuntos de dados e é definida pela razão entre o número de elementos em comum (interseção) e o total de elementos distintos (união), conforme descrito por Irani, Pise e Phatak (2016).

O método escolhido na aglomeração foi o single, seguindo aderência à necessidade de parâmetros mais restritivos (Bouguettaya; Yu; Liu; Zhou, 2015). Para o método, a similaridade entre dois *clusters* C1 e C2 é dada pelo valor mínimo entre dois pontos mais próximos, tal que  $d(C1, C2) = \min(\text{dist}(C1(i), C2(j))), i \in C1, j \in C2$ .

Durante a fase de exploração do modelo, foi observado significativo tempo de execução<sup>7</sup>, o que foi ao encontro do observado na literatura (Bouguettaya; Yu; Liu; Zhou, 2015). Como solução de contorno, o dimensionamento de entrada foi minimizado com o emprego do módulo 2 (RADAR\_FILTER).

Um dos parâmetros restritivos empregados na aglomeração foi o de *ngrams*. Refere-se ao sequenciamento contíguo de itens de palavras a partir de determinado texto. Por exemplo, quando se utiliza a técnica de bigramas (*ngrams* = 2), na frase “Todos são iguais perante a lei”, os bigramas formados são “Todos são”, “são iguais”, “iguais perante”, “perante a” e “a lei”. Após algumas iterações, o parâmetro *ngrams* foi ajustado entre 2 e 5, em virtude dos resultados observados.

**Figura 6 – Distribuição de *clusters* - distribuição cumulativa por *cluster***



7 A complexidade computacional é na ordem  $O(n^2 \log(n))$ , em termos de tempo de execução, onde  $n$  é o número de elementos.



Fonte: elaborada pelos autores.

A partir de acórdãos publicados em 2014, foram coletados 638.606 registros de deliberações. Destes, foram filtrados 10.176 elementos, que redundaram em 3.606 *clusters*<sup>8</sup>. A configuração de parâmetros restritivos resultou em quantidade significativa de *clusters*, na razão aproximada de 1:3. Contudo, a distribuição foi acentuadamente assimétrica. A primeira centena de *clusters* abrangeu praticamente metade do quantitativo de deliberações (Figura 6).

O resultado assimétrico pode ser explicado pela aparente uniformidade textual de alguns tipos específicos de deliberação. Por outro lado, boa parte dos *clusters*, a partir de  $k = 300$ , apresentam quantidade mínima de elementos, onde  $k$  é identificador do  $k$ -ésimo *cluster*. A referência a órgãos, entidades e nomes de interessados pode ter provocado distanciamento suficiente para repelir a aglomeração de conjuntos e de um novo conjunto ou, até mesmo, a redação de tipo específico de deliberação.

Figura 7 – Amostragem de deliberações nos top 3 maiores *clusters*

k	Amostragem
0	9.3.2. comunique <u>à</u> interessad <u>a</u> a deliberação deste tribunal e <u>a</u> alerte de que o efeito suspensivo proveniente da eventual interposição de recurso junto ao tcu não <u>a</u> eximirá da devolução dos valores indevidamente recebidos após a notificação. <u>em caso de não provimento do apelo</u>
	9.3.2. comunique <u>ao</u> interessad <u>o</u> a deliberação deste tribunal e <u>o</u> alerte de que o efeito suspensivo proveniente de eventual interposição de recursos <u>s</u> junto ao tcu não <u>o</u> eximirá da devolução dos valores indevidamente recebidos após a notificação. <u>em caso de não provimento dos apelos; e</u>
	9.4.2. comunique <u>à</u> interessad <u>a</u> a deliberação deste tribunal e <u>a</u> alerte de que o efeito suspensivo proveniente da eventual interposição de recursos <u>s</u> , junto ao tcu, não <u>a</u> eximirá da devolução dos valores indevidamente recebidos após a notificação
1	<u>b)</u> determinar <u>ao órgão/entidade</u> de origem <u>fazendo-se acompanhar de cópia da instrução da unidade técnica</u> que, no prazo de trinta dias, submeta ao tcu, pelo sistema de apreciação e registro de atos de admissão e concessão (sisac), novos <u>s</u> atos <u>s</u> , livres <u>s</u> das falhas apontadas, com fundamento nos arts. 45, caput, da lei 8.443/1992, 260, § 6º, do regimento interno <u>do tcu</u> , 3º, §§ 6º e 7º, da resolução - tcu 206/2007 e 15, caput e § 1º, da instrução normativa - tcu 55/2007
	<u>1.8.</u> determinar <u>à unidade</u> de origem que, no prazo de trinta dias, submeta ao tcu, pelo sistema de apreciação e registro de atos de admissão e concessões (sisac), novo <u>s</u> ato <u>s</u> , livre <u>s</u> das falhas apontadas, com fundamento nos arts. 45, caput, da lei 8.443/1992, 260, § 6º, do regimento interno <u>do tcu</u> , 3º, §§ 6º e 7º, da resolução - tcu 206/2007 e 15, caput e § 1º, da instrução normativa - tcu 55/2007
	<u>1.8.</u> determinar <u>à unidade</u> de origem que, no prazo de trinta dias, submeta ao tcu, pelo sistema de apreciação e registro de atos de admissão e concessões (sisac), novo <u>s</u> ato <u>s</u> , livre <u>s</u> das falhas apontadas, com fundamento nos arts. 45, caput, da lei 8.443/1992, 260, § 6º, do regimento interno, 3º, §§ 6º e 7º, da resolução - tcu 206/2007 e 15, caput e § 1º, da instrução normativa - tcu 55/2007
2	9.3.3. encaminhe a este tribunal, no prazo de 30 (trinta) dias, a partir da ciência deste acórdão, por cópia, comprovante da data em que <u>o</u> interessad <u>o</u> tom <u>ou</u> conhecimento <u>desta deliberação</u>
	9.3.3. encaminhe a este tribunal, no prazo de 30 (trinta) dias, a partir da ciência deste acórdão, por cópia, comprovante da data em que <u>a</u> interessad <u>a</u> <u>dele</u> tom <u>ar</u> conhecimento
	9.3.3. encaminhe a este tribunal, no prazo de 30 (trinta) dias a partir da ciência deste acórdão, por cópia, comprovante da data em que <u>o</u> interessad <u>o</u> tom <u>ou</u> conhecimento <u>desta deliberação; e</u>

Fonte: elaborada pelos autores.

A Figura 7 apresenta amostragem de deliberações agrupadas nos três maiores *clusters*, identificados como  $k = 0$ ,  $k = 1$  e  $k = 2$ , que contêm, respectivamente, 332, 243 e 203

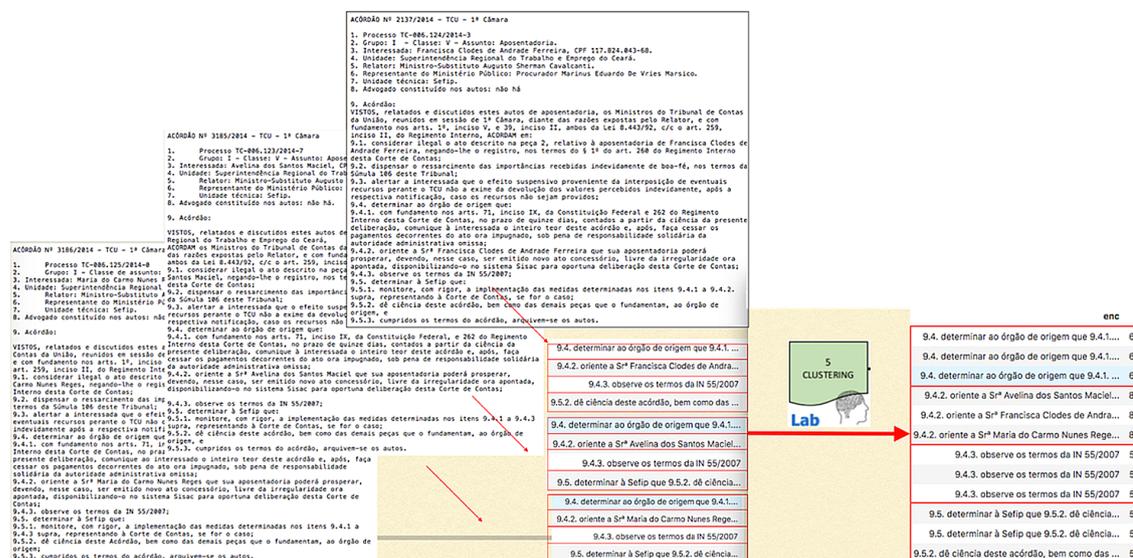
8 *Clusters* com mais elementos: [(0, 332), (1, 243), (2, 203), (3, 172), (4, 158), (5, 151), (6, 134), (7, 131), (8, 124), (9, 124), (10, 110)].



elementos. A capitalização minúscula decorre do tratamento e da limpeza de dados. Para fins didáticos, as diferenças textuais entre as deliberações e entre cada *cluster* foram assinaladas na cor vermelha.

Já a Figura 8 apresenta amostragem de deliberações resultantes da aplicação do algoritmo de clusterização. Foram selecionados, para fins ilustrativos, os Acórdãos 3.186/2014-1, 3.185/2014-1 e 2.137/2014-1. As deliberações cadastradas para monitoramento correspondem, respectivamente, às seções 9.4, 9.4.2, 9.4.3 e 9.5.2; e 9.4, 9.4.2, 9.4.3 e 9.5; 9.4, 9.4.2, 9.4.3 e 9.5. Após a execução do referido módulo, os primeiros registros de cada acórdão foram associados ao valor  $k = 62$ . Em seguida,  $k = 83$ ,  $k = 55$  e  $k = 52$ , respectivamente. A amostragem tem caráter didático, uma vez que a cada acórdão foi associada a mesma sequência de *clusters*.

Figura 8 – Amostragem de deliberações após clusterização



Fonte: elaborada pelos autores.

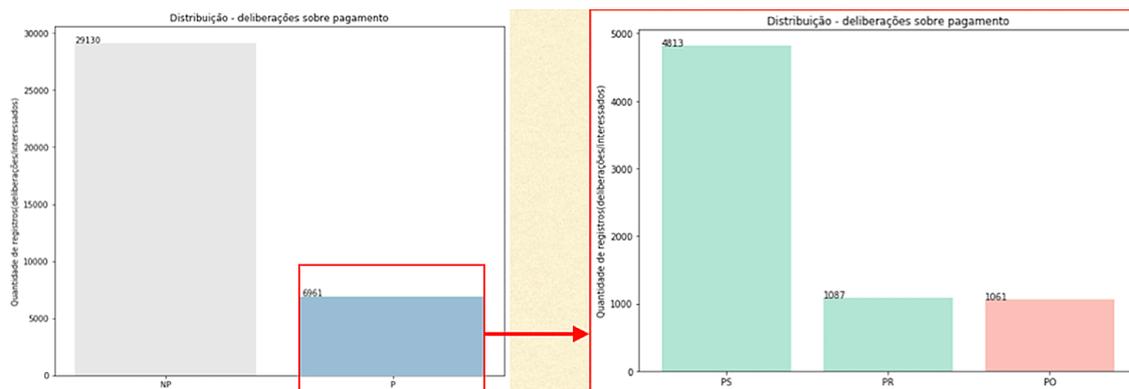
O Módulo 8 – CLASSIFIER – realiza a classificação por meio de algoritmo supervisionado, em particular sobre textos de deliberação acerca de cessação de pagamento. É parte integrante do conjunto de módulos de cor azul. A classificação do modelo (M), contendo conjunto de dados de entrada  $x$ , resulta em  $\hat{y} = M(x)$ , onde  $\hat{y} \in \{c_1, c_2, \dots, c_i, \dots, c_k\}$ , onde  $c_i$ , resulta em, onde  $\hat{y}$  é a  $i$ -ésima classe categórica. Para a construção de M, é necessário conjunto de treinamento que contenha elementos previamente rotulados (*training set*). O algoritmo Naive Bayes pode ser empregado para solucionar problemas de classificação multiclasse (Zaki; Meira Jr., 2014).

Durante as simulações, foi necessário ajustar o modelo, eliminando-se palavras irrelevantes (*stopwords*), ao contrário da estratégia de parametrização adotada no algoritmo de *clustering* (raia VERDE). Com a medida, o somatório de palavras de todos os textos de entrada caiu de 8.810.383 para 5.236.993.



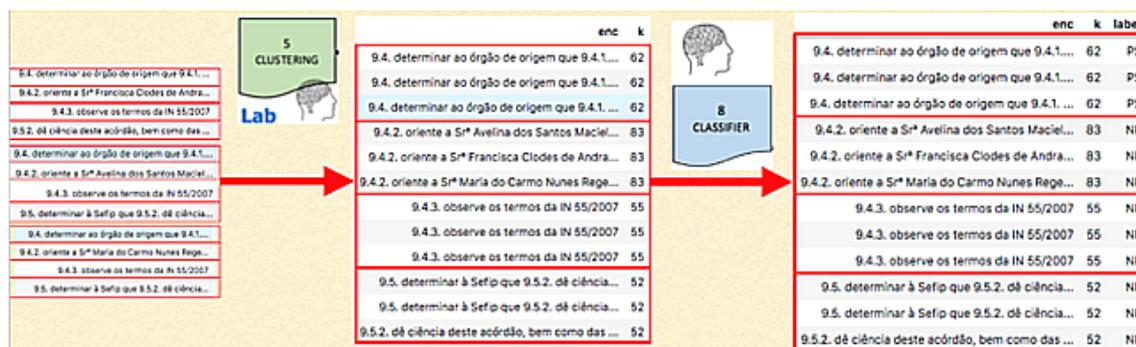
Na Figura 9, foram classificadas 29.130 deliberações como não relacionadas a pagamento (NP) e 6.961 relacionadas a pagamento (P). Dentro da categoria (P), 4.813 deliberações trataram de suspensão de pagamento (PS), 1.087 de suspensão de rubrica (PR) e 1.061 foram categorizadas em outros casos (PO), que demandavam tratamento manual.

**Figura 9 – Resultados do classificador Naive Bayes**



Fonte: elaborada pelos autores.

**Figura 10 – Resultados do classificador - amostragem de deliberações**



Fonte: elaborada pelos autores.

Para facilitar a compreensão, a Figura 10 representa continuação da Figura 8, já apresentada na seção anterior e que contém a mesma amostragem dos Acórdãos 3.186/2014-1, 3.185/2014-1 e 2137/2014-1 e respectivas deliberações, descrita na clusterização (VERDE). O passo relativo ao classificador produziu como resultados os rótulos os, para os elementos de k = 62, e NP, para os elementos de k = [52,55,83]. Cabe ressaltar que a classificação é orientada a elementos. A unanimidade de rotulação dentro de um mesmo cluster é desejável, caso a acurácia da clusterização tenha permitido o agrupamento de textos, de fato, afins.

Após destaque dos módulos 5 e 8, ressalta-se, também, a importância dos módulos responsáveis por identificar situações de cessação de pagamento. Eles estão diretamente ligados ao objetivo central da análise de dados: levantar casos de interrupção de pagamentos. As informações extraídas servem como insumo para o cruzamento com a base de dados-alvo, neste caso, o Siape.



Também merecem destaque os módulos da raia AMARELA, relativos ao cruzamento de dados com o Siape. Conforme já esclarecido, os demais módulos não foram detalhados, em razão do limite de páginas do artigo. A avaliação do modelo, com base em critérios de mineração de dados e de negócio, é apresentada a seguir.

### 3.5 AVALIAÇÃO

A fase de avaliação compreende a verificação dos resultados obtidos por meio das técnicas aplicadas, a partir de critérios de sucesso do negócio. Também contempla o levantamento de versões finais da modelagem e a identificação de oportunidades de melhoria, para trabalhos futuros.

Esta subseção está organizada nas seguintes partes: *i)* validação do modelo; *ii)* avaliação dos resultados; e *iii)* apresentação dos resultados consolidados. Cabe ressaltar que a validação do modelo (*assess model*) é normalmente apresentada na fase anterior, contudo optou-se por movê-la para esta seção, para facilitar a organização do presente trabalho.

A validação do modelo é baseada em critérios de sucesso de mineração de dados. A validação do classificador textual de deliberações adotou critérios de mensuração de desempenho enumerados por Zaki e Meira Jr. (2014), como acurácia, precisão e matriz de confusão, fornecendo visão detalhada do classificador em diversas situações.

A acurácia obtida foi de 93,89%. A precisão de resultados por categorias de interesse P variou entre 96% e 97%. Os valores obtidos pela acurácia e precisão estão alinhados com um dos objetivos de negócio, ou seja, estabelecer modelo de elevado grau de acerto. Ainda sobre precisão, ressalta-se prevalente equalização de valores, com pequeno desbalanceamento de resultados em relação à categoria NP. O resultado decorreu da necessidade de ajuste da quantidade de entradas do modelo.

Durante as interações que envolveram modelagem e validação, observou-se possibilidade de aplicação de técnicas de balanceamento, para melhoria de resultados, dada a elevada desproporcionalidade entre classes. A classe predominante era a de “não pagamento”, enquanto a de “pagamento” aparecia em menor proporção. Tal assimetria foi simples de entender, já que o fato estava diretamente relacionado à sistemática de elaboração das próprias deliberações.

Decisões que versam sobre cessação de pagamento normalmente são acompanhadas de vários dispositivos, dentre eles, comunicar ao órgão e ao interessado a decisão exarada, assinar prazo para cumprimento de ações e emitir novos atos. Contudo, as mesmas deliberações podem ocorrer com ausência de efeitos suspensivos, mesmo em caso de ilegalidade, conforme caso concreto.

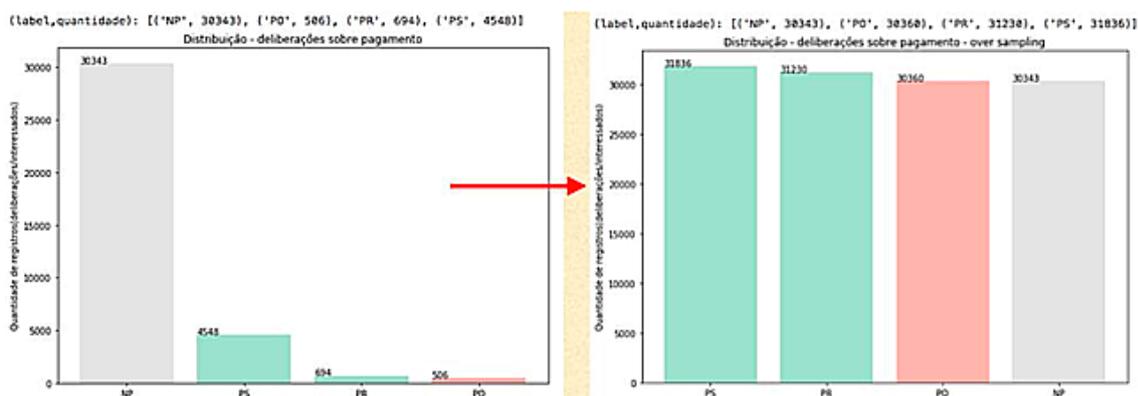
Existem várias técnicas na literatura para aplicar o rebalanceamento dos dados (Kaur, 2019). O objetivo dessas técnicas é mitigar riscos de vieses sobre a classe majoritária, que aumentam a acurácia do classificador. Dentre as técnicas de pré-processamento



por amostragem, têm-se as seguintes: o *oversampling* (OS), que objetiva aumentar o número de registros da classe minoritária; o *undersampling*, que visa a reduzir o número de registros da classe majoritária; e os métodos híbridos, que possuem características das duas anteriores (Chawla, 2005). Avaliações na literatura demonstram que as técnicas baseadas em *oversampling* tendem a ser mais satisfatórias que as baseadas em *undersampling* (Mohammed; Rawashdeh; Abdullah, 2020). Deste modo, optou-se por empregar a técnica de *oversampling* neste trabalho.

A Figura 11 apresenta uma comparação visual entre os dados antes e depois da aplicação da modelagem, com o intuito de transmitir o impacto prático do modelo proposto. Para isso, os resultados obtidos com e sem a aplicação de *oversampling* foram comparados, usando classificador Naive Bayes. As duas abordagens são identificadas como NB+OS (com *oversampling*) e NB-OS (sem *oversampling*). Observa-se desbalanceamento acentuado na classe NP (não pagamento), cuja distribuição é significativamente desigual em relação à classe oposta, o que reforça a relevância do balanceamento, para melhorar o desempenho do modelo<sup>9</sup>.

**Figura 11 – Comparação visual entre dados pré e pós-processamento (*oversampling*)**



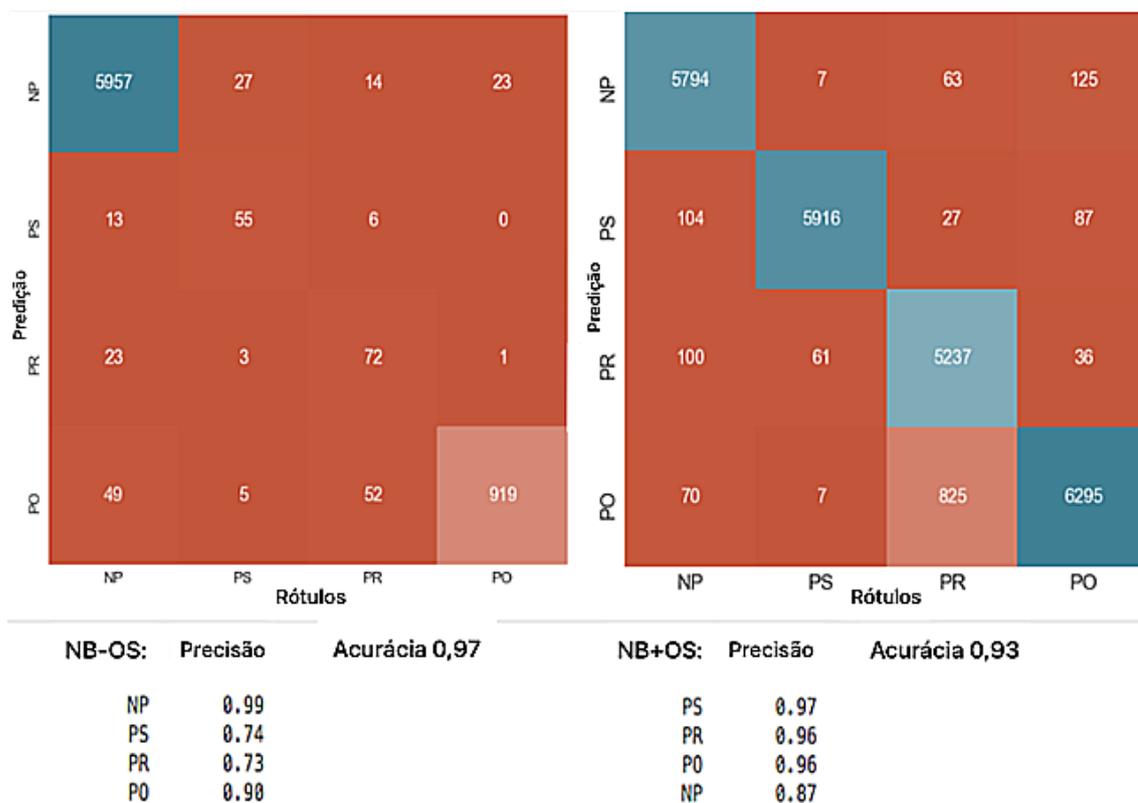
Fonte: elaborada pelos autores.

Cabe ressaltar que o critério de decisão, para aferir qualidade no algoritmo de clusterização, foi baseado no percentual de elementos em *clusters* com divergências de rotulação, porquanto o classificador é executado por deliberação, e não por *clusters*, e a expectativa é de que todos os elementos de mesmo *cluster* sejam rotulados com a mesma classe.

<sup>9</sup> O *dataset* inicial apresenta distribuição de entradas contendo deliberações de pagamento {rótulo, quantidade}: {(‘NP’, 30343), (‘PO’, 506), (‘PR’, 694), (‘PS’, 4548)}. Já o resultado foi o seguinte, após o emprego de *oversampling* {rótulo, quantidade}: {(‘NP’, 30343), (‘PO’, 30360), (‘PR’, 31230), (‘PS’, 31836)}.



**Figura 12 – Oversampling – matriz de confusão e accuracy score: NB-OS e NB+OS**



Fonte: elaborada pelos autores.

O percentual de **clusters** que apresentaram divergência de rotulação em pelo menos um de seus elementos foi de 1,65%, na execução com o modelo NB-OS (sem *oversampling*). Com aplicação do *oversampling* (NB+OS), o percentual melhorou para 0,23%. Os resultados consideram a proporção de elementos pertencentes a clusters nos quais todos os membros compartilham o mesmo rótulo (P ou NP), refletindo a consistência interna da rotulação após aplicação do modelo.

A Figura 12 apresenta comparativo entre os resultados obtidos em NB-OS e os obtidos em NB+OS, respectivamente, baseados na matriz de confusão e nas métricas de precisão e acurácia. Observa-se que a predição incorreta de valores para as classes minoritárias ocorre com frequência relativamente maior que a prevalente (matriz à esquerda). Ou seja, o quantitativo de falsos positivos é relativamente maior nas classes minoritárias. Com aplicação do *oversampling*, o quantitativo de falsos positivos torna-se mais equilibrado entre as categorias (matriz à direita). Efeitos similares podem ser observados na acurácia. Embora o resultado absoluto tenha sido um pouco melhor para NB-OS, no NB+OS, os resultados parciais por classe mostraram-se mais equânimes.

A avaliação foi conduzida com base em critérios de sucesso do negócio, estabelecidos no início do projeto, com o objetivo de assegurar que os produtos gerados fossem úteis à organização (Chapman, 1999; Wirth; Hipp, 2000). Nesse contexto, os resultados do trabalho foram apresentados à área de negócio responsável pelo tema no TCU, envolvendo a, então, Coordenação-Geral de Controle Externo, de Gestão de Processos e Informações (CGGINF),



vinculada à Secretaria-Geral de Controle Externo (Segecex); e as chefias da unidade técnica responsável por atos de pessoal. Nessas ocasiões, a iniciativa foi considerada satisfatória e aderente às necessidades do negócio, particularmente quanto ao potencial de ampliação do alcance e da efetividade das ações da unidade técnica.

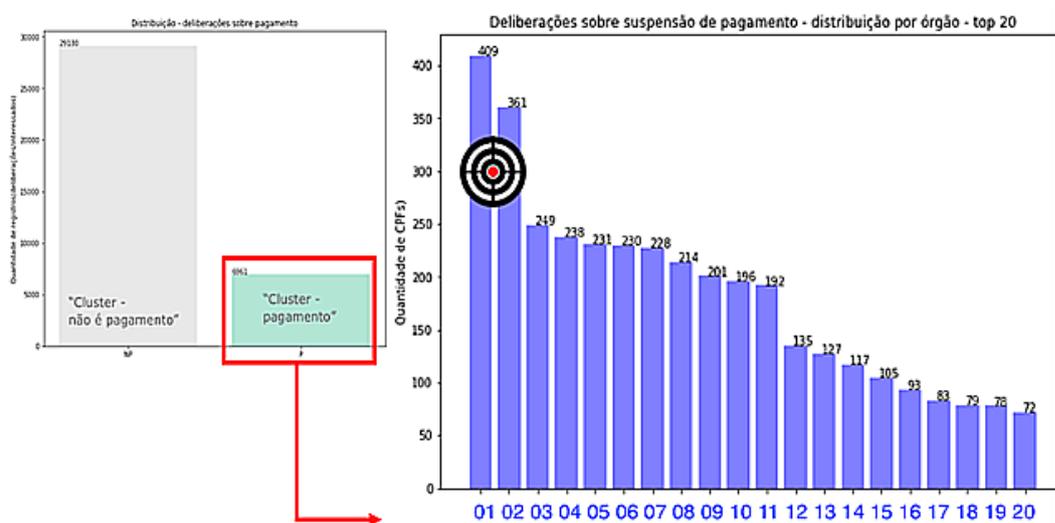
#### 4. RESULTADOS E DISCUSSÃO

Os dados apresentados nesta seção dizem respeito à verificação do correto cumprimento das determinações relacionadas a atos de pessoal. As informações consolidadas foram organizadas segundo as situações de cessação de pagamento adotadas pelo modelo, com a seleção de 1.120 deliberações de um total de 5.900 classificadas como relativas a pagamento. As tipologias propostas abrangeram 19% do domínio de deliberações rotuladas como pertinentes à interrupção de pagamentos, das quais 45% foram identificadas como contendo indícios que justificariam procedimento investigativo. As demais (55%) puderam ser aproveitadas para reduzir o estoque de deliberações a serem monitoradas, por terem sido conferidas segundo o modelo proposto.

A Figura 13 apresenta a distribuição de deliberações sobre suspensão de pagamento por órgão, com destaque para os 20 casos de maior prevalência. Observa-se que os dois primeiros colocados se destacam em relação aos demais órgãos, o que pode facilitar a concentração de esforços na redução do estoque de objetos pendentes de monitoramento.

Ainda com vistas a facilitar o planejamento de ações de controle, os dez órgãos mais prevalentes representam cerca de 3% do universo e concentram cerca de um terço (36,7%) das deliberações sobre cessação de pagamento. Ao ampliar o domínio de órgãos para vinte, ou seja, para cerca de 5% do universo, a concentração de deliberações ultrapassa a metade das deliberações (52,3%). O diagnóstico acerca da concentração de casos pode ser empregado para focar esforços de monitoramento e gerar eficiência para as equipes de fiscalização.

**Figura 13 – Resultados consolidados e as 20 maiores prevalências (top 20)**



Fonte: elaborada pelos autores.





justificariam apuração mais detalhada. As demais, sem indícios de irregularidades, foram sugeridas para redução do estoque de monitoramentos pendentes.

A análise exploratória revelou que dez órgãos concentram um terço das deliberações sobre cessação de pagamento, embora representem apenas 3% do domínio. Ao ampliar para 20 órgãos (5% do total), essa concentração atinge 52,3%. Esse diagnóstico pode apoiar a definição de alvos prioritários e gerar ganhos de escala para ações de controle.

O modelo atendeu aos critérios de sucesso estabelecidos para a mineração de dados. Obteve-se acurácia de 93,89% e precisão entre 96% e 97%. O uso do algoritmo de clusterização mostrou consistência na rotulagem das deliberações, enquanto a aplicação de *oversampling* mostrou-se vantajosa, visto que reduziu o percentual de divergências de 1,65% (NB-OS) para 0,23% (NB+OS).

Também foram atendidos os critérios de sucesso do negócio. Duas agendas realizadas com diferentes níveis hierárquicos da Segecex indicaram boa receptividade à proposta, e a área de negócio confirmou a recorrência de órgãos previamente identificados nas análises.

Ademais, foi observado que a simples aplicação das etapas de clusterização e classificação já seriam suficientes para gerar ganhos de produtividade. Apesar das limitações de tempo para incluir mais deliberações, os resultados foram satisfatórios, tanto na modelagem quanto na validação do modelo.

A base de conferência utilizada neste trabalho foi restrita ao Siape. As bases externas, como a extra-Siape e a Relação Anual de Informações Sociais (Rais), serão exploradas em etapas futuras. Além disso, propõem-se as seguintes direções para os próximos trabalhos: *i)* padronizar os textos das deliberações, com o objetivo de reduzir a complexidade da automação; *ii)* desenvolver novas tipologias de análise, para ampliar a capacidade de interpretação dos dados; *iii)* implementar a atualização contínua das bases e deliberações; *iv)* incluir códigos de rubricas, para facilitar o cruzamento com os dados do Siape; *v)* considerar, além da data de publicação, a data de comunicação, para cálculo de prazos; e *vi)* aprimorar tecnicamente os modelos, por meio da adoção de novas bibliotecas de Processamento de Linguagem Natural (PLN) e classificadores mais avançados.

## REFERÊNCIAS

BALANIUK, Remis. A Mineração de Dados como apoio ao Controle Externo. **Revista do TCU**, Brasília, v. 117, n. 1, p. 79-86, jan. 2010.

BOUGUETTAYA, Athman; YU, Qi; LIU, Xumin; ZHOU, Xiangmin. Efficient agglomerative hierarchical clustering. **Expert Systems with Applications**, [s. l.], v. 42, n. 5, p. 2785-2797, abr. 2015.



BRANCO, Claudio Souza Castello. Histórico sobre a obtenção e o tratamento de dados para o Controle Externo no TCU, de 1995 a 2014. **Revista do TCU**, Brasília, v. 131, n. 1, p. 12-21, jan. 2014.

BRASIL. [Constituição Federal (1988)]. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, 1988. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/constituicao/constituicao.htm](http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm). Acesso em: 8 nov. 2024.

BRASIL. **Lei n. 8.112, de 11 de dezembro de 1990**. Dispõe sobre o regime jurídico dos servidores públicos civis da União, das autarquias e das fundações públicas federais. Brasília, DF: Presidência da República, 1990. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/leis/l8112cons.htm](http://www.planalto.gov.br/ccivil_03/leis/l8112cons.htm). Acesso em: 8 nov. 2024.

BRASIL. Tribunal de Contas da União. Conheça a nova estrutura da Secretaria-Geral de Controle Externo, área-fim do TCU. **Portal TCU**, Brasília, 9 de fevereiro de 2023a. Notícias. Disponível em: <https://portal.tcu.gov.br/imprensa/noticias/conheca-a-nova-estrutura-da-secretaria-geral-de-controle-externo-area-fim-do-tcu.htm>. Acesso em: 8 nov. 2024.

BRASIL. Tribunal de Contas da União. Implementação do e-Pessoal chega à reta final. **Portal TCU**, Brasília, 15 de março de 2018. Notícias. Disponível em: <https://portal.tcu.gov.br/imprensa/noticias/implementacao-do-e-pessoal-chega-a-reta-final.htm>. Acesso em: 8 nov. 2024.

BRASIL. Tribunal de Contas da União (Plenário). **Acórdão n. 1.032/2019**. Processo: 24.000/2018-3. Relatório de Acompanhamento. Relatório de Acompanhamento autuado com o objetivo de realizar o 4º Ciclo da fiscalização contínua realizada nos dados cadastrais e nas folhas de pagamento de diversos órgãos da administração pública federal, referentes aos meses de março a setembro de 2018. Relator: Ministro Aroldo Cedraz, 8 de maio de 2019a. Disponível em: <https://pesquisa.apps.tcu.gov.br/redireciona/acordao-completo/ACORDAO-COMPLETO-2353039>. Acesso em: 8 nov. 2024.

BRASIL. Tribunal de Contas da União. **Portaria-Sefip n. 1, de 10 de junho de 2019**. Dispõe sobre a organização interna, as competências e atividades da Secretaria de Fiscalização de Pessoal (Sefip). Brasília, DF: TCU, 2019b. Disponível em: <https://pesquisa.apps.tcu.gov.br/redireciona/norma/NORMA-1613>. Acesso em: 8 nov. 2024.

BRASIL. Tribunal de Contas da União. **Resolução TCU n. 246, de 30 de novembro de 2011**. Altera o Regimento Interno do Tribunal de Contas da União, aprovado pela Resolução TCU nº 155, de 4 de dezembro de 2002. Brasília, DF: TCU, 2011. Disponível em: <https://pesquisa.apps.tcu.gov.br/redireciona/norma/NORMA-21217>. Acesso em: 8 nov. 2024.

BRASIL. Tribunal de Contas da União. **Resolução TCU n. 255, de 26 de setembro de 1991**. Dispõe sobre a apreciação, pelo Tribunal de Contas da União, para fins de registro da legalidade dos atos de admissão de pessoal e de concessão de aposentadoria, reformas e pensões. Brasília, DF: TCU, 1991. Disponível em: <https://pesquisa.apps.tcu.gov.br/redireciona/norma/NORMA-20028>. Acesso em: 8 nov. 2024.



BRASIL. Tribunal de Contas da União. **Resolução TCU n. 353, de 22 de março de 2023**. Estabelece procedimentos para exame, apreciação e registro dos atos de admissão de pessoal e de concessão de aposentadoria, reforma e pensão pelo TCU. Brasília, DF: TCU, 2023b. Disponível em: <https://pesquisa.apps.tcu.gov.br/redireciona/norma/NORMA-21630>. Acesso em: 8 nov. 2024.

BRASIL. Tribunal de Contas da União (Segunda Câmara). **Acórdão n. 10.974/2015**. Processo: 025.757/2015-6. Aposentadoria. Relatora: Ministra Ana Arraes, 24 de novembro de 2015. Disponível em: <https://pesquisa.apps.tcu.gov.br/redireciona/acordao-completo/ACORDAO-COMPLETO-1584431>. Acesso em: 8 nov. 2024.

CHAPMAN, Pete. **The CRISP-DM user guide**. 4. CRISP-DM, 1999.

CHAWLA, Nitesh V. Data mining for imbalanced datasets: an overview. *In*: MAIMON, Oded; ROKACH, Lior. (ed.). **Data mining and knowledge discovery handbook**. Boston: Springer, 2005. p. 853-867.

EL NAQA, Issam; MURPHY, Martin J. What is machine learning? *In*: EL NAQA, I.; LI, R.; MURPHY, M (ed.). **Machine learning in radiation oncology**. Boston: Springer, 2015.

IRANI, Jasmine; PISE, Nitin; PHATAK, Madhura. Clustering techniques and the similarity measures used in clustering: a survey. **International journal of computer applications**, [s. l.], v. 134, n. 7, p. 9-14, jan. 2016.

KAUR, Harsurinder; PANNU, Husanbir Singh; MALHI, Avleen Kaur. A systematic review on imbalanced data challenges in machine learning: Applications and solutions. **ACM computing surveys (CSUR)**, [s. l.], v. 52, n. 4, p. 1-36, ago. 2019.

MACEDO, Mônica de Lima. **Qualidade no Tribunal de Contas da União**: práticas adotadas pela Secretaria de Fiscalização de Pessoal para controlar os gastos de pessoal. 2004. Monografia (Especialização em Controle Externo) – Tribunal de Contas da União, Instituto Serzedello Corrêa, Brasília, DF, 2004.

MITCHELL, Tom M. **Machine learning**. Nova York: McGraw-Hill, 1997.

MOHAMMED, Roweida; RAWASHDEH, Jumanah; ABDULLAH, Malak A. Machine learning with oversampling and undersampling techniques: overview study and experimental results. **11th international conference on information and communication systems (ICICS)**, 2020. p. 243-248.

MONDAL, Bhaskar. Artificial intelligence: state of the art. *In*: BALAS, Valentina E.; KUMAR, Raghvendra; SRIVASTAVA, Rajshree (ed.). **Recent trends and advances in artificial intelligence and internet of things**. Berlim: Springer, 2020. p. 389-425.



PALIWAL, Shweta; BHARTI, Vishal; MISHRA, Amit Kumar. Ai chatbots: transforming the digital world. In: BALAS, Valentina E.; KUMAR, Raghvendra; SRIVASTAVA, Rajshree (ed.). **Recent trends and advances in artificial intelligence and internet of things**. Berlim: Springer, 2020. p. 455-482.

PANDIT, Shraddha; GUPTA, Suchita. A comparative study on distance measuring approaches for clustering. **International Journal of Research in Computer Science**, [s. l.], v. 2, n. 1, p. 29-31, 2011.

SAMUEL, Arthur L. Some studies in machine learning using the game of checkers. **IBM Journal of research and development**, [s. l.], v. 3, n. 3, p. 210-229, jul. 1959.

SCHRÖER, Christoph; KRUSE, Felix; GÓMEZ, Jorge Marx. A systematic literature review on applying CRISP-DM process model. **Procedia Computer Science**, [s. l.], v. 181, p. 526-534, fev. 2021.

WIRTH, Rüdiger; HIPPE, Jochen. CRISP-DM: Towards a standard process model for data mining. **4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining**, [s. l.], 2000. p. 29-39.

ZAKI, Mohammed J.; MEIRA JR., Wagner. **Data mining and analysis: fundamental concepts and algorithms**. Inglaterra: Cambridge University Press, 2014.

---

Os conceitos e as interpretações emitidos nos trabalhos assinados são de exclusiva responsabilidade de seus autores.

