Instituto Serzedello Corrêa – ISC
Pós-Graduação em Análise de Dados para o Controle

HELTON FABIANO GARCIA

MINERAÇÃO DE DELIBERAÇÕES PARA MONITORAMENTO DE ATOS DE PESSOAL

Brasília
2019

HELTON FABIANO GARCIA

MINERAÇÃO DE DELIBERAÇÕES PARA MONITORAMENTO DE ATOS DE PESSOAL

Trabalho de conclusão do curso de pós-graduação lato sensu em Análise de Dados para o Controle realizado pela Escola Superior do Tribunal de Contas da União como requisito para a obtenção do título de especialista em Análise de Dados.

Orientador: Prof. Dr. Edans Flávius de Oliveira Sandes.

Brasília
2019

REFERÊNCIA BIBLIOGRÁFICA

GARCIA, Helton Fabiano. **Mineração de deliberações para monitoramento de atos de pessoal**. 2019. Trabalho de Conclusão de Curso (Especialização em Análise de Dados para o Controle) – Escola Superior do Tribunal de Contas da União, Instituto Serzedello Corrêa, Brasília DF.

CESSÃO DE DIREITOS

NOME DO AUTOR: Helton Fabiano Garcia
TÍTULO: Mineração de deliberações para monitoramento de atos de pessoal
GRAU/ANO: Especialista/2019

_____

Helton Fabiano Garcia
heltonFG@tcu.gov.br

Ficha catalográfica

Garcia, Helton Fabiano

Mineração de deliberações para monitoramento de atos de pessoal / Helton Fabiano Garcia; orientador, Edans Flávius de Oliveira Sandes, 2019.

108 p.

Monografia (especialização) - Escola Superior do Tribunal de Contas da União, Curso de Especialização em Análise de Dados para o Controle, Brasília, 2019.

Inclui referências.

1.Análise de Dados. 3. Mineração de Dados. 4. Classificação Textual. 5. Aprendizado de Máquina. I. Sandes, Edans Flávius de Oliveira. II. Escola Superior do Tribunal de Contas da União. especialização em Análise de Dados para o Controle. III. Título.

HELTON FABIANO GARCIA

MINERAÇÃO DE DELIBERAÇÕES PARA MONITORAMENTO DE ATOS DE PESSOAL

Trabalho de conclusão do curso de pós-graduação lato sensu em Análise de Dados para o Controle realizado pela Escola Superior do Tribunal de Contas da União como requisito para a obtenção do título de especialista em Análise de Dados.

Brasília, 13 de dezembro de 2019.

Banca examinadora:

_____

Prof. Dr. Edans Flávius de Oliveira Sandes

Orientador

Instituto Serzedello Corrêa - TCU

_____

Prof. MSc. Saul Campos Berardo

Instituto Serzedello Corrêa - TCU

AGRADECIMENTOS

Agradeço a Deus pela saúde e pela oportunidade de chegar até aqui.

À minha família, pelo amor, paciência, força e certeza de estarem sempre torcendo para prosseguir em frente, bem como pela compreensão pelos momentos de ausência.

Ao Professor Edans Sandes pela orientação objetiva, apoio oportuno e didática exemplar. Também por transmitir segurança e tranquilidade de estarmos no caminho certo.

Ao Professor Saul Berardo pela participação na banca examinadora e pelas valorosas e detalhadas contribuições para a melhoria do presente trabalho.

Ao TCU, pela organização de curso pioneiro na Administração Pública e, em particular, ao Secretário de Infraestrutura de TI, José Renato Affonso, por autorizar iniciativa de participar da seleção, nos termos do item 2.2, Edital-ISC nº 7, de 10/05/18.

Ao meu chefe imediato Geraldo Magela Lopes de Freitas pelo apoio à empreitada e pela ótima relação profissional e compreensão da importância da iniciativa.

Ao amigo Helton Onésio pela disponibilidade em sempre ajudar e em servir de facilitador para comunicação a partes envolvidas da Sefip e pelas sugestões e orientações construtivas.

Ao amigo Ricardo Santos pela oportunidade de compartilhar a jornada de estudo, com os inúmeros trabalhos ao longo do curso. A jornada foi longa e gratificante.

À equipe de segurança – Geraldo, André, Cristiane, Baldez, Wendell, Jônatas e Ana – pelo alto profissionalismo, dedicação e excelente clima organizacional.

Por fim, e não menos importante, gostaria de agradecer a todos os colegas que contribuíram para a viabilização do presente trabalho, compartilhando gentilmente precioso tempo ou informações. Como o trabalho demandou extensivo contato entre múltiplas áreas da Egrégia Casa, o agradecimento individual se faz devido e merecido. Sebastião Arantes e Rodrigo Bento (Sefip), pela visão geral das respectivas subunidades, em particular sobre o monitoramento de acórdãos de pessoal; Marcos Drach e Luciana Marinho (Seproc), pelo apoio com o RADAR; José Luiz Costa (Sefip), pelo didático apoio com o SIAPE; Andrea Ribeiro e Flávio Ferreira (Sefip), pelo apoio com o e-Pessoal/Sisac; Alessandra Requena (STI), pelo apoio com a busca textual; Reginaldo Fernandes, Alysson Paulista e Barnabé Pereira (Sefip) pelas informações sobre a Sefip; Marcos Paulo, Luciana Tsujiguchi e Raquel Zampietro (STI), por modelos de dados; Cláudio Queiroz e Patrícia Cursino (Setic), pelo apoio com a infraestrutura de TI; Marcus Borela e Luís André (STI) pelas sábias orientações sobre aprendizagem de máquina.

"*A persistência é o caminho do êxito.*"

Charles Spencer Chaplin (1889 – 1977)

"*Somos o que repetidamente fazemos.*"

Aristóteles (384 a.C. – 322 a.C.)

"*Trabalhe duro e em silêncio.*"

Dale Carnegie (1888 – 1955)

RESUMO

O crescimento do estoque de deliberações a serem monitoradas pela Secretaria de Fiscalização de Pessoal (Sefip) do Tribunal de Contas da União (TCU) mostra-se enorme desafio, agravado pela tendência de racionalização de recursos. Este trabalho tem como objetivo fornecer insumos para tornar mais eficiente o monitoramento de acórdãos de pessoal, a cargo da Sefip, permitindo assim a redução do estoque de deliberações monitoradas. Para isso, propõe-se um modelo baseado em aprendizagem de máquina, com fins de automatizar o monitoramento de deliberações acerca de cessação de pagamentos de atos de admissão e de concessão de aposentadoria e de reforma de pessoal. Durante o desenvolvimento do trabalho, realizou-se a coleta, interpretação e verificação automatizada em base de dados de pagamento, utilizando tipologias definidas. Também são apresentadas informações decorrentes de análise exploratória. A arquitetura do modelo é formada por treze módulos, contendo em dois deles algoritmos não supervisionados e supervisionados de aprendizagem de máquina. O protótipo desenvolvido neste trabalho apresentou resultados considerados satisfatórios pela Sefip. Como trabalhos futuros, vislumbra-se a criação de projeto para evolução do protótipo e implantação no ambiente do TCU, sob a denominação sugerida de AMANDA (**A**utomatização de **Mon**itoramento de **D**eliberações de **A**tos).

Palavras-chave: Mineração de dados. Análise de dados. Ciência de dados. Aprendizado de máquina. Clusterização. Classificador Naive Bayes. Monitoramento de deliberações de acórdãos de pessoal. Atos de Pessoal. Classificação textual.

ABSTRACT

The growth of deliberations inventory monitored by the Personnel Audit Secretariat (Sefip) of the Federal Court of Accounts (TCU) shows a huge challenge, compounded by the trend towards resource rationalization. This work aims to present a proposal to make the effort regarding the decision monitoring related to Personnel Acts more efficient by Sefip, thus reducing the deliberations inventory. To this end, a machine learning-based model is proposed to automate the deliberations monitoring about paycheck stub cessation from admissible and retirement personnel acts. During the development of the work, it was performed the collection, interpretation and automated verification in the payroll database using defined typologies. It is also proposed detailed information arising from exploratory analysis. The prototype architecture of the model consists of thirteen modules, containing in two of them, algorithms unsupervised and supervised machine learning algorithms. The results of the prototype developed in this work are considered satisfactory by Sefip. As future work, we envision the creation of a project as an evolution and deployment in the TCU production environment, under the suggested denomination AMANDA (Automation of Personnel Act Deliberation Monitoring).

***Keywords*:** *Data mining. Data analysis. Data science. Machine learning. Clustering. Naive Bayes classifier. Monitoring of personnel Acts deliberations. Personnel Acts. Textual classification.*

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

1. INTRODUÇÃO

O artigo publicado na revista *The Economist*, intitulado *Fuel of the future* (THE ECONOMIST, 2017) compara os dados produzidos na era digital moderna com petróleo. “Os dados são para o presente século o que o petróleo foi para o anterior: a força motriz para crescimento e mudanças”.

O artigo esclarece ainda que o valor dos dados tem aumentado progressivamente. Em particular, com a transformação de dados em serviços, a partir de soluções baseadas em inteligência artificial (IA). Empresas (ou entidades ou órgãos) capazes de analisar dados e empregar técnicas de IA, como aprendizagem de máquina, tornaram-se as novas exploradoras de “petróleo digital”. Menciona ainda que tais empresas se tornaram mais valiosas no mercado de ações do que as tradicionais atuantes no ramo petrolífero.

Preliminarmente, cabe considerar a definição de alguns termos. O que é IA? Aprendizagem de máquina? Mineração de dados? Inteligência artificial e aprendizagem de máquina podem ser consideradas palavras populares do momento e que podem ser interpretadas (erroneamente) como tendo a mesma semântica (OPPERMANN, 2019).

O termo inteligência artificial foi inicialmente cunhado, em 1956, por John McCarthy, considerado o pai da IA, além de idealizador da linguagem LISP, como sendo “a ciência e engenharia de fabricação de máquinas inteligentes” (PEART, 2017).

O termo aprendizagem de máquina (*machine learning* – ML) também foi criado na mesma década. Em 1959, Artur Samuel, à época, pesquisador da IBM, definiu o termo como sendo “campo de estudo que oferece a máquinas a possibilidade de aprender sem serem explicitamente programadas”. Outra definição foi proposta por Tom Mitchell “um programa de computador aprende com dada experiência E em relação a alguma tarefa T e alguma medida de desempenho P, se seu desempenho T, medido por P, melhorar com a experiência E” (PUGET, 2016).

AMATRIAIN (2016) define mineração de dados (MD) como sendo “campo interdisciplinar objetivando a descoberta de propriedades de conjuntos de dados”. Para o autor, a MD pode empregar vários instrumentos para atingir o seu objetivo. ML é um deles. No mesmo sentido ZAKI e MEIRA JR. (2014) esclarecem que MD e ML podem oferecer percepções e conhecimentos fundamentais dos dados.

MAYES (2017) apresenta análise entre cada um dos termos supracitados IA, ML e DL.

Outros termos correlatos costumam ser também mencionados. Uma definição é apresentada em cada respectiva referência: *Big data* (DONTHA, 2017); uma definição sobre ciência de dados e o papel do cientista de dados (ESTEVES, 2019); um breve histórico sobre ciência de dados (GANGANE, 2019); *Deep learning*, bem como considerações acerca de início de nova era em relação a ML (OPPERMANN, 2019).

Uma vez que os dados têm sido considerados o novo petróleo da economia mundial e que instrumentos de IA podem ser capazes de agregar valor a serviços baseados em dados, qual o contexto do TCU?

Uma das competências constitucionais do TCU é apreciar a legalidade de atos de pessoal (BRASIL, 1988). MACEDO (2004) esclarece que, no TCU, até o final da década de 80, não havia solução sistematizada para apreciação de atos de pessoal. O controle era realizado pela análise de "processos convencionais", ou seja, por meio de conjunto de papéis, muitas vezes volumoso, numerados, autuados, com folhas rubricadas, contendo cópias dos fatos geradores do direito à inativação, à percepção de parcelas de proventos e ao recebimento de pensão, remetidos pelos diversos setores de recursos humanos dos órgãos jurisdicionados.

Esclarece ainda que, a partir da década de 90, o referido controle passou a ser feito por meio de sistema informatizado, denominado Sisac (Sistema de Apreciação e Registro de Atos de Concessão), o qual foi progressivamente implantado. A substituição definitiva pela análise informatizada ocorreu definitivamente, a partir de 2001.

BRANCO (2014) dispõe que, em 2013, o Sisac continha mais de 4 milhões de atos de pessoal, cada ato com mais de 100 campos cada. 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 um aspecto de risco (BALANIUK, 2010). O termo também pode estar associado a trilhas de auditoria.

TCU (2019b) anunciou que o Sisac seria substituído pelo e-Pessoal. O objetivo seria de incrementar novas funcionalidades ao sistema de atos de pessoal (PORTAL TCU, 2018).

Por um lado, a presença de soluções informatizadas foi capaz de colaborar para a automatização de rotinas para a apreciação de atos de pessoal. Por outro lado, o monitoramento de deliberações versando sobre atos de pessoal, permaneceu sob processamento manual, conforme entrevista realizada com equipes de fiscalização da Secretaria de Fiscalização de Pessoal (Sefip).

RIBEIRO (2017), em apresentação elaborada pela Sefip, em 2017, esclareceu que o monitoramento de acórdãos foi selecionado como uma das “diretrizes de prioridades da Sefip”. Essa necessidade se torna cada vez mais prioritária visto que o estoque de deliberações a serem monitoradas apresenta tendência histórica de crescimento. Contudo, remanesce a oportunidade de automatização do processo de monitoramento.

Outro aspecto importante é o cenário de racionalização de recursos humanos, cujo movimento não é apenas tendência no Tribunal, mas também em toda Administração Pública Federal. A reposição de quadros e até mesmo o seu reforço pode ser considerado, dentro de breve, como opção escassa ou inviável para fazer frente ao aumento de demandas. De outra sorte, os objetivos de negócio tendem a crescer. A eficiência torna-se fator crítico para o sucesso da organização. Fazer mais com menos. Adotar soluções inteligentes capazes de automatizar processos até então realizados por pessoas.

Diante do exposto, não há dúvida acerca da existência de grande volume de dados, da necessidade de colaborar para objetivos de negócio e de considerar o recurso humano como fator de escassez. Urge a necessidade de contribuir para objetivos da organização por meio de serviços baseados em inteligência artificial capazes de agregar valor a partir da exploração de dados. Em suma, o desafio se resume em como prospectar o “petróleo digital” disponível no Tribunal.

O documento está organizado da seguinte forma. Seção 2: Apresentação do problema relacionado a atos de pessoal e da justificativa para proposição do presente trabalho. Seção 3: Apresentação de objetivos gerais e específicos. Seção 4: Descrição da metodologia de trabalho adotada, com o emprego do CRISP-DM, modelo amplamente reconhecido para mineração de dados. Seção 5: Descrição do entendimento do negócio, versando sobre coleta de expectativas sobre o que a organização espera da mineração de dados. Seção 6: Entendimento dos Dados, compreendendo o acesso e a exploração de dados disponíveis para mineração. Seção 7: Preparação de Dados, com a seleção de dados considerados relevantes para as metas de mineração. Seção 8: Modelagem, com o desenvolvimento de protótipo. Seção 9: Avaliação, compreendendo a verificação dos resultados obtidos por meio das técnicas elencadas na modelagem, a partir de critérios de sucesso de negócio, bem como levantamento de versões finais da modelagem e oportunidades de melhoria para trabalhos futuros. Seção 10: Implantação, compreendendo plano de implantação e de sustentação, juntamente com relatório final contendo resultados obtidos. Seção 11: Considerações Finais. Por fim, ao final dos documentos foram incluídos anexos, em complemento ao trabalho apresentado.

2. PROBLEMA E JUSTIFICATIVA

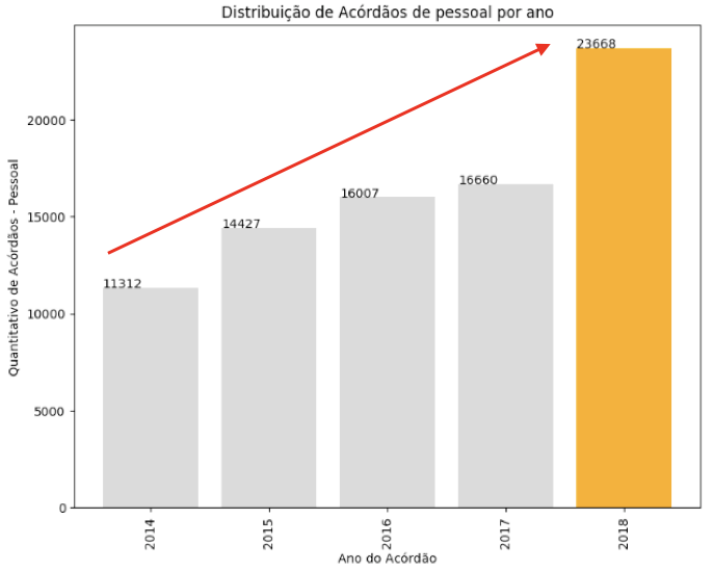
Monitoramento é o instrumento de fiscalização utilizado pelo Tribunal para verificar o cumprimento de suas deliberações e os resultados delas advindos, conforme o art. 243 da Resolução-TCU nº 246, de 30 de novembro de 2011, que dispõe sobre Regimento Interno do Tribunal de Contas da União (RITCU) (TCU, 2011).

Nos termos da Portaria-Sefip n°1, de 10 de junho de 2019, que dispõe sobre a organização interna, as competências e atividades da Sefip, a unidade é responsável pelo monitoramento de acórdãos do TCU proferidos em processos relacionados a atos de pessoal, abrangendo admissão, concessão de aposentadoria, pensão civil, reforma, pensão militar e pensão especial de ex-combatente (art. 5º, II), bem como em processos de fiscalização e em processos cujo objeto seja relacionado a pessoal, relacionados a: Tomada de Contas Especial (TCE), denúncia, representação, consulta, solicitação do Congresso Nacional (art. 10, III) (SEFIP, 2019).

Atualmente, o monitoramento previsto no art. 5º, II encontra-se a cargo da 2ª Diat. O previsto no art. 10, III, do Sinfip.

Ocorre que, em entrevista com as áreas supracitadas, foi verificado que as atividades são processadas **manualmente**. A conferência do cumprimento de deliberações é feita confrontando o disposto no acórdão com a situação presente do jurisdicionado afetado, a exemplo de consulta às bases de dados do órgão. Ou seja, trata-se rito oneroso e demanda alocação de recursos que poderiam ser dedicados a outras atividades da unidade, como a fiscalização.

Figura 1: Distribuição de acórdãos de pessoal por ano



Fonte: Elaborada pelo autor (2019).

O estoque de deliberações a serem monitorados apresenta tendência histórica de crescimento. É o que pode ser observado a partir dos dados obtidos do sistema RADAR, que gerencia informações estruturadas das deliberações dos colegiados TCU. Entre 2014 e 2018, foram registrados 82.074 acórdãos de pessoal, com média de 16.414,8 por ano (Figura 1). Em 2019, até 20/09, a demanda do ano já se encontrava em 15.863, valor próximo à média anual. Cada acórdão pode apresentar uma ou mais deliberações.

Em relação às deliberações, foram observadas 22.059 do tipo "Determinação a Órgão/Entidade", associadas à unidade técnica Sefip. O quantitativo contribui para o aumento do estoque de itens passíveis de monitoramento da unidade. Essa quantidade representa parcela significativa de deliberações, sendo esse o escopo da monografia.

Por fim, a justificativa do presente trabalho é a 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, conforme relatado pela unidade. Logo, a possibilidade de aplicação no âmbito do ambiente de produção da unidade técnica é **imediata**.

3. OBJETIVOS

A partir da definição do problema a ser tratado e do escopo de pesquisa, são definidos os seguintes objetivos, classificando-os em:

3.1. Objetivo geral

Fornecer insumos para tornar mais eficiente o monitoramento[1] de acórdãos de pessoal, a cargo da Sefip (Secretaria de Fiscalização de Pessoal), conforme previsto no inciso I do art. 1º da Portaria-Sefip nº 2, de 2 janeiro de 2015 (grifo próprio): "*Art. 1º Fica delegada competência aos titulares da 1ª e 2ª Diretorias de Instrução de Atos de Pessoal (1ª e 2ª Diat), Diretoria de Tecnologia de Informação em Pessoal (Ditip), e ao chefe do Serviço de Instrução de Processos de Pessoal (Sinfip), no âmbito de suas atribuições, e, em seus impedimentos eventuais, aos*

[1] Conforme art. 243 da Resolução-TCU nº 246, de 30 de novembro de 2011, que dispõe sobre Regimento Interno do Tribunal de Contas da União (RITCU) (grifo próprio), "**Monitoramento** é o instrumento de fiscalização utilizado pelo Tribunal para verificar o cumprimento de suas deliberações e os resultados delas advindos".

respectivos substitutos, para a prática dos seguintes atos: I – emitir parecer conclusivo, em nome da Secretaria, sobre os atos sujeitos a registro, bem como os respectivos ***monitoramentos****, desde que as propostas sejam uniformes".*

3.2. Objetivos específicos

- Classificar deliberações versando sobre cessação de pagamento relacionados a apreciações de atos de pessoal de admissão e de concessão;
- Assegurar alto grau de acurácia, conforme critérios de sucesso definidos pelo negócio durante a fase de entrevistas, mesmo que haja alguma redução de escopo;
- Obter indícios a partir de informações decorrentes de análise exploratória, a fim de contribuir para o planejamento de ações de controle;
- Identificar tipologias capazes de permitir o monitoramento de deliberações versando sobre cessação de pagamento de atos de pessoal; e
- Estabelecer arquitetura modular, de forma a assegurar evolução incremental e iterativa do modelo proposto, bem como a facilitar a conversão do protótipo em solução, a ser definida por meio de projeto futuro.

4. METODOLOGIA

A metodologia aplicada no presente trabalho foi a *Cross-Industry Standard Process for Data Mining* (CRISP-DM), concebida originalmente para mineração de dados (IBM, 2014).

O CRISP-DM, como modelo de processo, fornece uma visão geral do ciclo de vida da mineração de dados. Por outro lado, como metodologia, inclui descrições das fases típicas de um projeto, tarefas envolvidas em cada fase e explicação dos relacionamentos entre essas tarefas.

O ciclo de vida é composto de seis fases. As transições entre cada fase são representadas por setas, indicando as dependências mais importantes. A sequência das fases não é rigorosa. Projetos podem avançar ou regredir entre as fases, conforme necessário.

A fase de entendimento de negócio visa coletar expectativas sobre o que a organização espera obter da mineração de dados. Trata-se de fator crítico para o direcionamento de ações a consulta a partes envolvidas no negócio da organização.

Figura 2: CRISP-DM - ciclo de vida

Fonte: (IBM, 2014).

A fase de entendimento dos dados compreende o acesso e a exploração de dados disponíveis para mineração usando tabelas e gráficos, permitindo determinar qualidade e descrever resultados obtidos na documentação do projeto.

A fase seguinte, preparação de dados compreende procedimentos para derivar novos atributos, remover ou substituir valores em branco ou omissos, agregar registros, mesclar conjuntos de dados, classificar dados para modelagem, selecionar amostragem de dados, dividir em conjuntos de dados de treinamento e de teste.

A fase de modelagem compreende o desenvolvimento de protótipo, com o emprego de técnicas de mineração de dados, como por exemplo *clusterização*, bem como validação do modelo, por meio de critérios como matriz de confusão e acurácia.

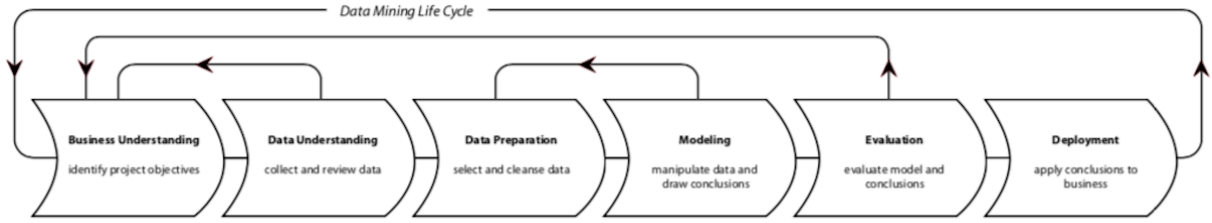
A fase de avaliação compreende a verificação dos resultados obtidos por meio das técnicas elencadas anteriormente com propósito estabelecido, a partir de premissas de negócio, bem como o levantamento de oportunidades de melhoria para trabalhos futuros.

Por fim, a fase de implantação compreende a implantação de solução, juntamente com relatório final contendo resultados obtidos.

Adicionalmente às fases supracitadas, cumpre ressaltar também a necessidade de levantamento bibliográfico e estudo da própria metodologia, bem como das técnicas de mineração de dados envolvidas.

O desenvolvimento deste trabalho será descrito nas seções 5 a 9, relacionados individualmente a cada fase do CRISP-DM, juntamente com as respectivas atividades selecionadas e consideradas relevantes para o presente trabalho. Em suma, será feita breve apresentação sobre a unidade técnica responsável por fiscalização de pessoal no TCU (Sefip) e por gestão de sistemas, Seproc (Secretaria de Gestão de Processos) (fase de entendimento do negócio), bem como sobre as bases de dados e sistemas envolvidos (fase de entendimento dos dados). Técnicas empregadas na limpeza e formatação de dados serão descritas na fase de preparação de dados. O desenvolvimento do protótipo será descrito na fase de modelagem. A verificação dos resultados obtidos será descrita na fase de avaliação. A implantação da solução e a elaboração do relatório final serão descritos na fase de implantação. Para facilitar a identificação de cada fase, o início de cada seção será identificado por meio do esquema apresentado na Figura 3, com destaque à fase em pauta.

Figura 3: CRISP-DM - ciclo de vida

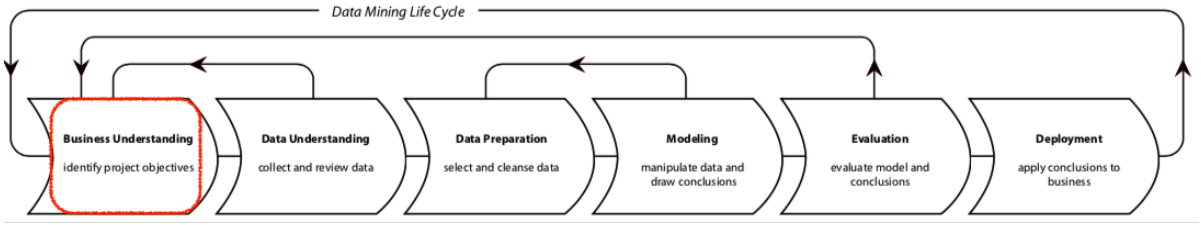


Fonte: (LEAPER, 2009).

5. ENTENDIMENTO DO NEGÓCIO

A fase de entendimento de negócio visa coletar expectativas sobre o que a organização espera obter da mineração de dados. Trata-se de fator crítico para o direcionamento de ações, a consulta a partes envolvidas no negócio da organização.

Figura 4: Ciclo de vida - mineração de dados - Entendimento do negócio



Fonte: (LEAPER, 2009).

Nesse sentido, foi procedido estudo acerca dos instrumentos de fiscalização realizados pelo TCU e foram realizadas entrevistas junto à unidade técnica responsável pelo monitoramento e auditoria de pessoal, com vistas a entender o negócio e proceder levantamento de situações-problema, no contexto do conhecimento adquirido no curso.

5.1. Sefip

A Secretaria de Fiscalização de Pessoal (Sefip) é unidade responsável por examinar e fiscalizar as despesas de pessoal dos órgãos e entidades integrantes da Administração Pública Federal e os atos de admissão e de concessão de aposentadoria, reforma e pensão. Encontra-se vinculada à Coordenação-Geral de Controle Externo de Gestão de Processos e Informações (Copin), parte integrante do Núcleo Estratégico de Controle Externo da Secretaria-Geral de Controle Externo (Segecex), conforme previsto na Portaria-Sefip nº 1, de 10 de junho de 2019 (SEFIP, 2019).

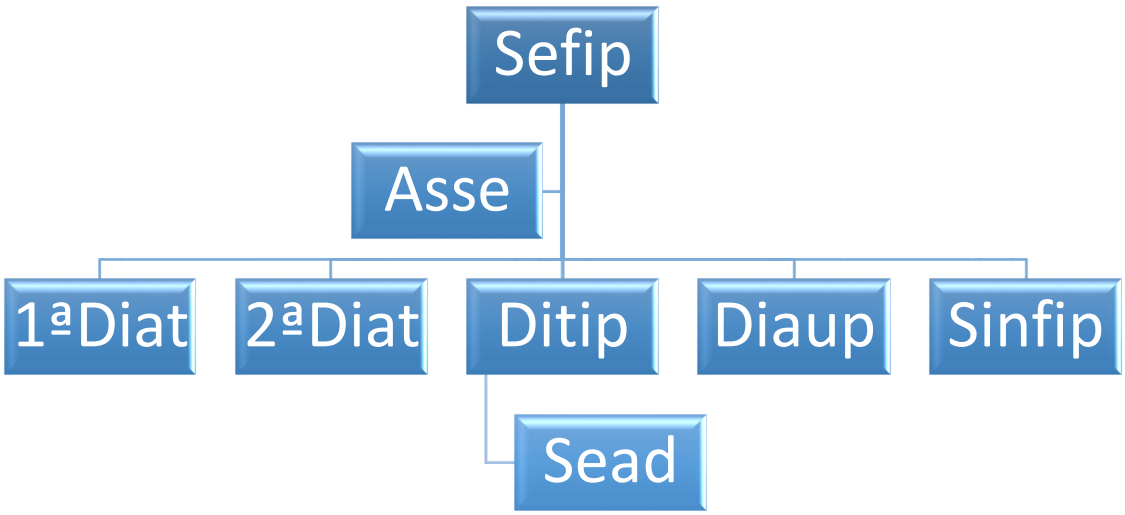
As competências da Sefip decorrem de previsão normativa associada ao controle externo, exercido com auxílio do TCU, nos termos da Constituição Federativa de 1988:

> "*Art. 71. O controle externo, a cargo do Congresso Nacional, será exercido com o auxílio do Tribunal de Contas da União, ao qual compete: [...] III - apreciar, para fins de registro, a legalidade dos atos de admissão de pessoal, a qualquer título, 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 a das concessões de aposentadorias, reformas e pensões, ressalvadas as melhorias posteriores que não alterem o fundamento legal do ato concessório*;" (BRASIL, 1988)

A Sefip dispõe da seguinte estrutura: 1ª Diretoria de Instrução de Atos de Pessoal (1ª Diat); 2ª Diretoria de Instrução de Atos de Pessoal (2ª Diat); Diretoria de Tecnologia de Informação em Pessoal (Ditip); Diretoria de Auditoria em Pessoal (Diaup); Assessoria (Asse); Serviço de Análise de Dados (Sead); Serviço de Instrução de Processos de Pessoal (Sinfip).

Segundo o Relatório Anual de Atividades do TCU, relativo a 2018 (TCU, 2019c), dos 143.006 atos apreciados no ano, 1.287 tiveram registro negado em razão de ilegalidades. Nesses casos, o Tribunal determina ao órgão de origem que adote as medidas cabíveis, fazendo cessar todo e qualquer pagamento decorrente do ato impugnado.

Figura 5: Organograma da Sefip



Fonte: Elaborada pelo autor (2019).

Figura 6: Relatório Anual de Atividades do TCU:2018. Atos de Pessoal

Atos de pessoal	2014*	2015*	2016	2017	2018
Apreciados conclusivamente:	105.035	83.007	80.997	76.442	143.006
a) ilegais	1.352	805	1.898	1.113	1.287
b) legais	92.775	69.268	59.406	60.119	97.177
c) prejudicados por perda de objeto e por inépcia do ato	---	---	19.693	15.210	44.542

Fonte: (TCU, 2019c).

5.2. Monitoramento de deliberações

Conforme art. 243 da Resolução-TCU nº 246, de 30 de novembro de 2011, que dispõe sobre Regimento Interno do Tribunal de Contas da União (RITCU), "Monitoramento é o instrumento de fiscalização utilizado pelo Tribunal para verificar o cumprimento de suas deliberações e os resultados delas advindos"(TCU, 2011).

Trata-se de um dos instrumentos de fiscalização previsto no RITCU: Levantamento; auditoria; inspeção; acompanhamento; e monitoramento. Monitoramentos, assim como auditorias e acompanhamentos, devem atender a plano de fiscalização elaborado pela

Presidência, em consulta com os relatores das listas de unidades jurisdicionadas, e aprovado pelo Plenário em sessão de caráter reservado (art. 238 e 244).

A Sefip realiza monitoramento acerca do cumprimento de acórdãos do TCU proferidos em processos relacionados a atos de pessoal, abrangendo admissão, concessão de aposentadoria, pensão civil, reforma, pensão militar e pensão especial de ex-combatente (art. 5º, II), bem como em processos de fiscalização e em processos cujo objeto seja relacionado a pessoal, relacionados a: Tomada de Contas Especial (TCE)[2], Denúncia, Representação, Consulta, Solicitação do Congresso Nacional (art. 10, III) (SEFIP, 2019) (grifo próprio):

> *"Art 5º [...] II - monitorar o cumprimento de acórdãos do TCU proferidos em processos relacionados aos **atos de pessoal** (**admissão, concessão de aposentadoria, pensão civil, reforma, pensão militar e pensão especial de ex-combatente**), podendo para tanto, promover diligências, na forma do Regimento Interno do TCU, para sanear os processos;"*
>
> *"Art 10. [...] III - monitorar o cumprimento de acórdãos do TCU proferidos em processos da sua competência[3] e em processos de **fiscalização** ou naqueles cujos monitoramentos tenham sido determinados por Colegiado ou Relator, com exceção dos processos referentes a atos de pessoal;"*

Atualmente, o monitoramento do previsto no art. 5º, II encontra-se a cargo da 2ª Diat. O previsto no art. 10, III, do Sinfip[4].

Em entrevista com as áreas supracitadas, foi esclarecido que o monitoramento é processado manualmente. É procedida a conferência do cumprimento de deliberações confrontando o disposto no acórdão com a situação presente do jurisdicionado afetado, a exemplo da consulta às bases de dados do órgão. Foi esclarecido também o emprego do sistema SisMonitoramento (TCU, 2019g).

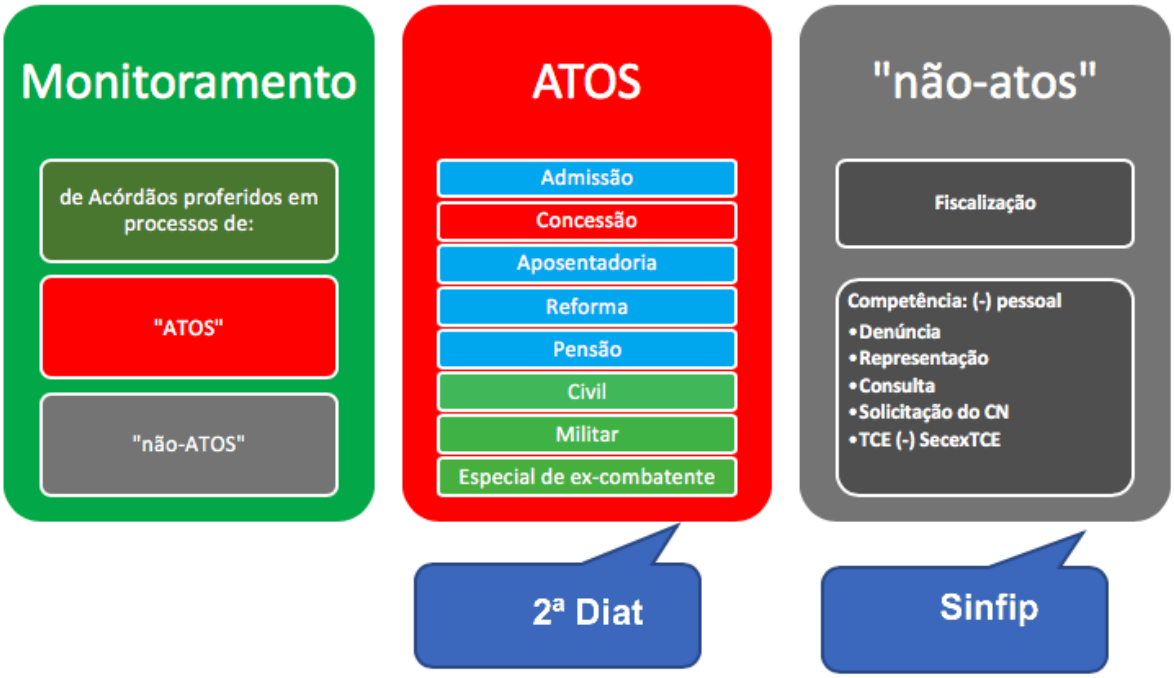
---

[2] Com exceção daqueles de competência originária da Secretaria de Controle Externo de Tomada de Contas Especial.

[3] Processos cujo objeto esteja relacionado a pessoal, a exemplo de **Denúncia, Representação, Consulta, Solicitação** do Congresso Nacional, entre outros, desde que esses processos não possuam por objeto atos de pessoal, bem como processos de Tomada de Contas Especial (**TCE**), cujo objeto esteja relacionado a pessoal, com exceção daqueles de competência originária da Secretaria de Controle Externo de Tomada de Contas Especial.

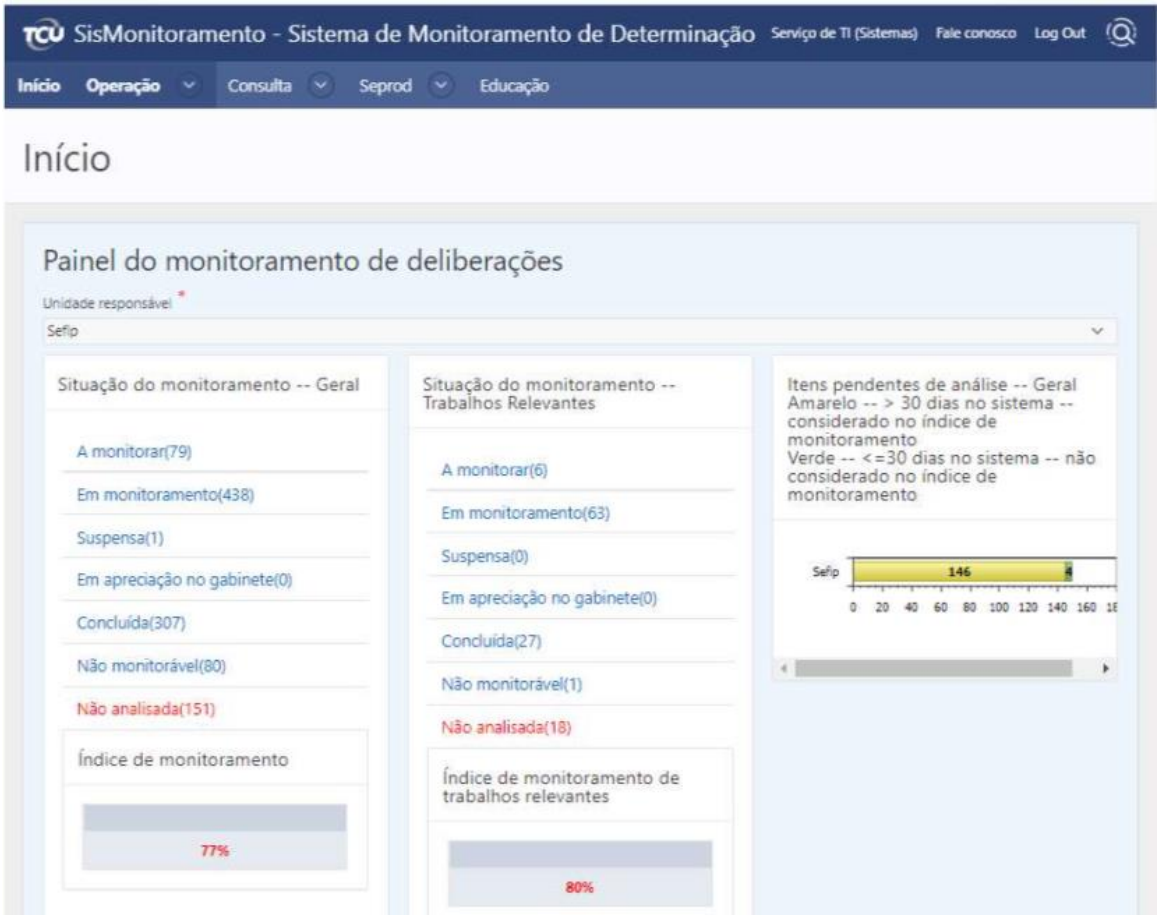
[4] Para o escopo do trabalho, convencionou-se o termo "não-atos" para processos de pessoal não relacionados a atos, por se tratar de sintaxe empregada informalmente pela unidade.

Figura 7: Monitoramentos de competência da Sefip



Fonte: Elaborada pelo autor (2019).

Figura 8: SisMonitoramento



Fonte: (TCU, 2019g).

O SisMonitoramento é sistema proveniente de desenvolvimento descentralizado, empregando tecnologia *Oracle Application Express*® (APEX)[5], visando acompanhamento sistemático pelas unidades técnicas das deliberações do tipo determinação e recomendação exaradas pelo Tribunal às unidades jurisdicionadas. O ambiente permite identificar deliberações a serem monitoradas, selecionar itens para tratamento e registrar tratativas. Contudo, o processo é manual. O SisMonitoramento emprega dados provenientes de outro sistema, denominado RADAR, a ser tratado a seguir.

5.3. Atos de Pessoal

A Resolução-TCU n° 255, de 26 de setembro de 1991 (TCU, 1991), dispõe sobre a apreciação, pelo TCU, para fins de registro da legalidade dos atos de admissão de pessoal, a qualquer título, 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 as melhorias posteriores que não alteram o fundamento legal do ato concessório.

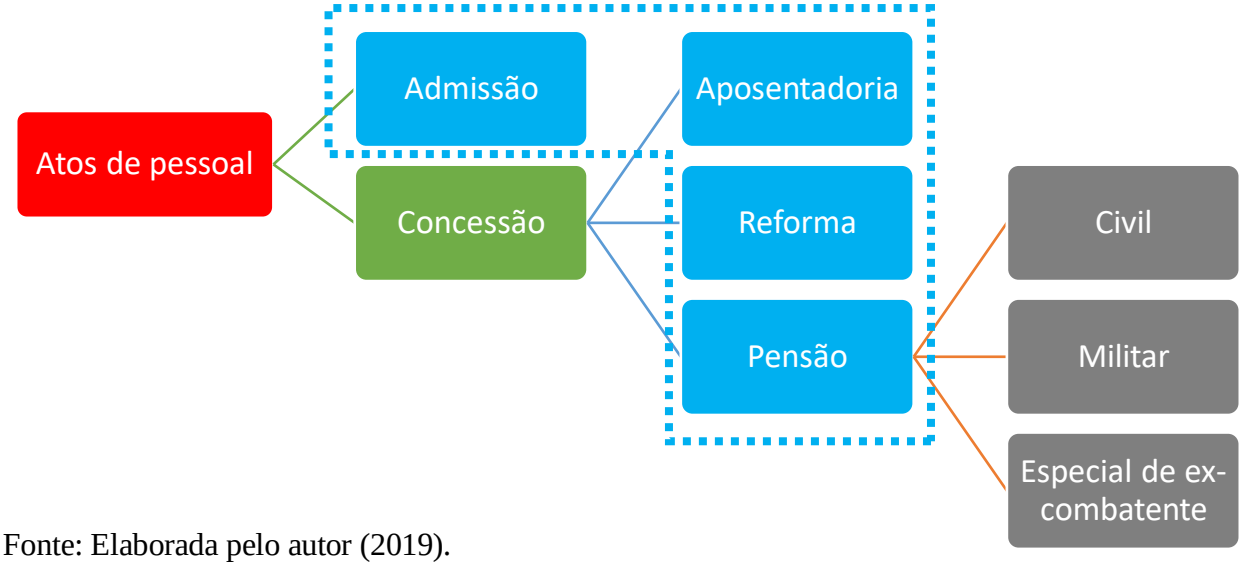
Para os atos de pessoal considerados legais, a apreciação para fins de registro consiste em identificar a legalidade do ato, com o registro das informações incluindo nome do servidor ou do empregado público, código do órgão, ano e regime de admissão, no caso de admissão. Para concessão de aposentadoria, reforma ou pensão, nome do inativo ou do(s) beneficiário(s), espécie de concessão, código do órgão e ano, conforme previsto no artigo 10 da referida norma.

Para os atos considerados ilegais, o Tribunal fixará prazo para que o órgão de origem adote as medidas corretivas que indicar para o exato cumprimento da lei. Finalizado o prazo, o ato poderá 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 à mesma, a partir da publicação da decisão do Tribunal no Diário Oficial da União (DOU), sob pena de obrigação de ressarcimento, pelo responsável, das quantias pagas após a referida data.

---

[5] Disponível em *https://apex.oracle.com*.

Figura 9: Atos de pessoal



Fonte: Elaborada pelo autor (2019).

Quando a ilegalidade do ato de concessão consistir na outorga de vantagens indevidas (art. 20), a recusa de registro obriga o órgão concedente a cessar o pagamento dos proventos ou benefícios a partir da publicação da decisão do Tribunal no DOU, no todo ou na parcela impugnada pelo Tribunal, sob pena de responsabilidade do respectivo ordenador de despesa. Caso o pagamento não seja suspenso ou na existência de indício de procedimento culposo ou doloso, o Tribunal poderá converter o processo em tomada de contas especial para apurar responsabilidades e promover o ressarcimento aos cofres públicos das despesas ilegalmente efetuadas.

As relações dos atos de admissão e de concessão e as respectivas decisões do Tribunal são publicados no DOU, constituindo-se título de legalidade para fins de direito (arts. 11 e 12).

5.4. RADAR/RADEX

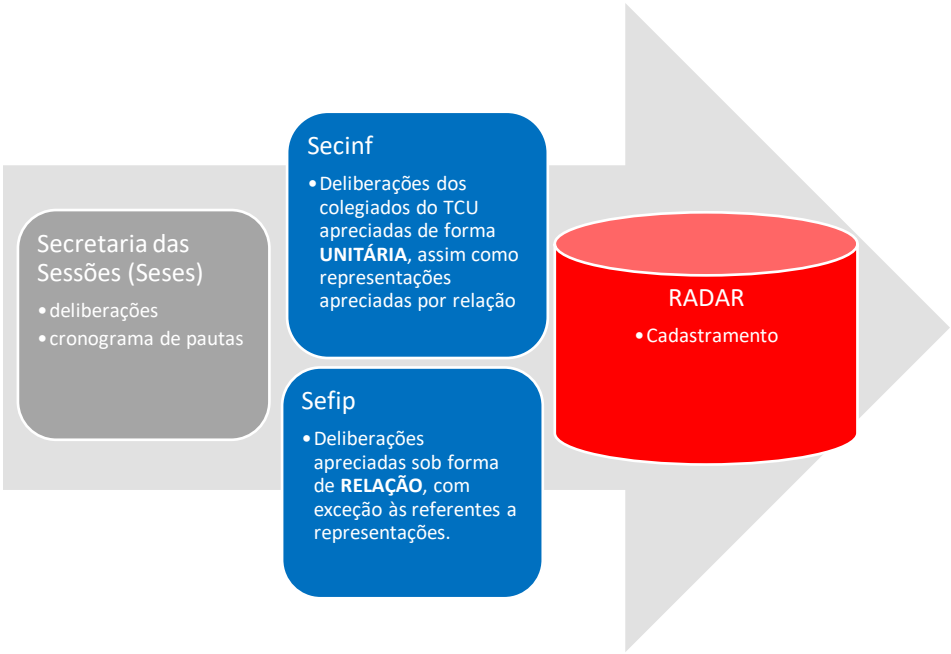
O RADAR é sistema corporativo desenvolvido para registro de apreciações, deliberações e acompanhamento de resultados do TCU. O objetivo é gerenciar informações estruturadas das deliberações dos colegiados TCU.

O monitoramento permite que as unidades técnicas do Tribunal acompanhem os resultados das deliberações por meio da utilização e do preenchimento de informações do Sistema RADAR (MACEDO, 2004).

As informações do sistema são atualizadas pelo Secinf/Seproc (Serviço de Cadastramento de Informações/Secretaria de Gestão de Processos), a quem compete registrar informações decorrentes de deliberações dos colegiados, bem como gerenciar e zelar pela atualização de

cadastros e bases de dados em função das deliberações do TCU. Tais competências derivam das originárias da unidade, previstas nos incisos I e VII do art. 46, da Resolução n° 305, de 28 de dezembro de 2018 (TCU, 2018).

Figura 10: RADAR



Fonte: Elaborada pelo autor (2019).

Segundo o Manual de Procedimento editado pelo referido Serviço (SEPROC, 2019a), em minuta, as deliberações relativas à área de **fiscalização de pessoal** são registradas no RADAR, conforme os seguintes critérios:

- Deliberações dos colegiados do TCU apreciadas de forma **unitária**, assim como as relativas a representações apreciadas por meio de relação. O cadastramento é de responsabilidade do Secinf, desde 10/06/2008; e
- Deliberações apreciadas sob forma de **relação**, com exceção às relacionadas a representações. O cadastramento é de responsabilidade da Sefip.

O cadastramento de itens de acórdãos no RADAR é feito na condição de "proposta", tornando-se disponível para uso pelas unidades do TCU após revisão, situação em que o estado muda para "homologação". Em caso de eventual divergência ou dúvida, o manual estabelece prevalência do texto do acórdão, porquanto tenha sido aprovada pelo Colegiado. Por essa razão, o cadastramento necessita ser pautado pela fidelidade aos termos deliberados pela Egrégia Corte.

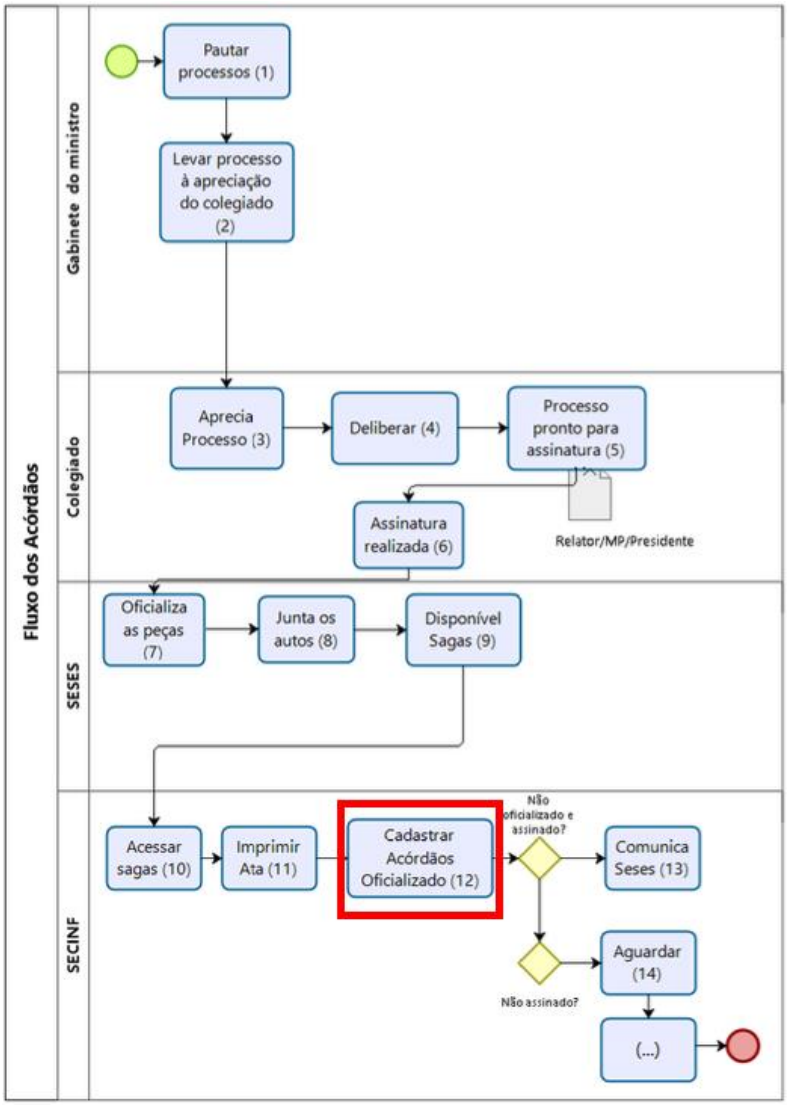
O referido manual destaca ainda que determinações dirigidas a órgãos ou entidades são cadastradas separadamente itens ou subitens, conforme o caso, tendo em vista preservar comandos específicos a cada jurisdicionado e facilitar o monitoramento individual por parte da

unidade técnica. O mesmo raciocínio é aplicado ao registro de audiências, feito separadamente para cada responsável. Quando houver necessidade de registro de vários itens para cada responsável, para facilitar a organização, o cadastro é feito de forma agrupada.

O fluxograma (Figura 11) resume o fluxo de trabalho relacionado às deliberações sob responsabilidade do Secinf, com destaque ao cadastro de acórdãos, em vermelho.

O RADEX é sistema corporativo que visa substituir o RADAR, uma vez que emprega tecnologias legadas (*Borland Delphi*®) para a aplicação de cadastro, dificultando o processo de sustentação do sistema. O novo sistema tem como objetivo oferecer funcionalidades para agilizar o cadastro de informações, melhorar o monitoramento de operações no sistema e a distribuição de acórdãos para as unidades (SEPROC, 2019b).

Figura 11: Fluxograma de acórdãos



Fonte: (SEPROC, 2019a).

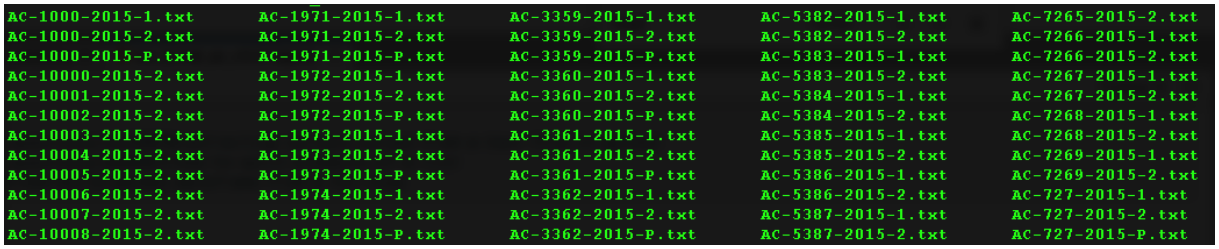
5.5. Busca Textual de Acórdãos

A Busca Textual de Acórdãos é a base oficial do TCU para disponibilização de pesquisa de inteiro teor de acórdãos, contendo informações desde 1992 (TCU, 2019a). O responsável por administrar a solução é o Seint/STI (Serviço de Integração e Métricas de Sistemas).

Trata-se de instrumento de transparência. Por meio da pesquisa, qualquer pessoa pode obter registros de deliberações do Tribunal.

O processo de atualização é feito a partir dos dados produzidos por outra solução corporativa, denominada Sagas - Sistema de apoio ao gerenciamento e acompanhamento das sessões - (TCU, 2019e), para apoio a gabinetes de autoridade, gerenciamento e acompanhamento das sessões de colegiado. A unidade gestora é a Seses/Segepres (Secretaria das Sessões).

Figura 12: Arquivos textuais de acórdãos



Fonte: Elaborada pelo autor (2019).

A partir do Sagas, documentos relativamente a acórdãos são oficializados, após deliberação em sessões dos Colegiados (Plenário e Primeira e Segunda Câmara). A solução de Busca Textual realiza a extração de dados, basicamente a partir da extração textual de documentos em formato de texto enriquecido (extensão .rtf), bem como a extração de metadados armazenados em bancos de dados.

O resultado é armazenado em banco de dados textual Apache Solr (THE APACHE SOFTWARE FOUNDATION, 2019), do tipo “noSQL” (banco de dados não relacional). Para deliberações oficializadas e públicas, o armazenamento ocorre em inteiro teor. Para oficializadas e classificadas, apenas parcialmente.

5.6. Base de dados SIAPE

O SIAPE (MPOG, 1989), Sistema Integrado de Administração de Pessoal, visa centralizar o processamento da folha de pagamentos por meio da alimentação descentralizada de informações dos órgãos que dependem do tesouro para fazer frente a suas despesas de pessoal, envolvendo as administrações direta, fundacional e autárquica do Poder Executivo.

O sistema processa o pagamento de servidores ativos, aposentados e pensionistas regidos tanto pelo Regime Jurídico Único Federal (BRASIL, 1990) quanto pela CLT (Consolidação das Leis do Trabalho) e por outros regimes (contratos temporários, estágios, residência médica).

A base denominada "extra-SIAPE" é constituída por UJs que não utilizam o SIAPE, cujos dados são remetidos mensalmente para a Sefip. Abrange servidores dos Poderes Judiciário e Legislativo, do Ministério Público da União, do Tribunal de Contas da União, do Banco Central, das Forças Armadas (apenas os militares) e das seguintes empresas estatais: Banco do Nordeste, Banco Nacional de Desenvolvimento Econômico e Social, Caixa Econômica Federal, Empresa Brasileira de Correios e Telégrafos, Centrais Elétricas Brasileiras S.A e Petróleo Brasileiro S.A.

Em Relatório de Acompanhamento (TCU, 2019d), TC 024.000/2018-3, de 29 de março de 2019, abrangendo a fiscalização de dados cadastrais e financeiros referentes aos meses de março a setembro de 2018 de servidores e pensionistas, foram observadas 213 unidades jurisdicionadas (UJ) integradas ao SIAPE. Ao "extra-SIAPE", 84.

A tabela a seguir apresenta quantitativo de servidores, instituidores de pensão e pensionistas identificados no referido relatório:

Tabela 1: Quantitativo de servidores, inativos, instituidores de pensão e pensionistas

Base	Ativos	Inativos	Instituidores de pensão	Pensionistas	Totais
Extra-siape (9/2018)	807.985	80.539	28.066	167.566	1.819.462
Siape (9/2018)	766.047	413.025	384.853	605.466	2.169.391
Totais	**1.574.032**	**493.564**	**412.919**	**773.032**	**3.988.853**

Fonte: (TCU, 2019d)

5.7. Sisac/e-Pessoal

No contexto do processamento eletrônico de informações relativas às admissões e às concessões de aposentadoria, reforma e pensão, sujeitas a registro pelo TCU, dois sistemas corporativos podem ser mencionados: Sisac e e-Pessoal.

O **Sisac** (Sistema Integrado de Apreciação de Atos de Pessoal) (TCU, 2019f) foi desenvolvido em 1992, permitindo, à época, a emissão de pareceres de legalidade de atos de pessoal (PORTAL TCU, 2018). Além disso, o sistema permite efetuar crítica com base em parâmetros previamente definidos no sistema, baseados na legislação e na jurisprudência, realizar diligência dos atos rejeitados pela crítica aos respectivos órgãos de controle interno.

Segundo informações encontradas na monografia de especialização promovida pelo ISC/TCU (MACEDO, 2004), até então, a sistemática era adotar abordagem convencional. Os órgãos enviavam fichas-resumo ao Tribunal para apreciação de atos.

Quanto à produção, o artigo publicado na Revista TCU, intitulado "Histórico sobre a obtenção e o tratamento de dados para o Controle Externo no TCU, de 1995 a 2014" (BRANCO, 2014) apresenta valores quantitativos. Até o ano 2000, havia mais de 400 mil atos de pessoal em estoque. Com a nova sistemática de análise automática de atos de admissão, incluindo o desenvolvimento de uma série de críticas eletrônicas que permitiam ao TCU concluir pela legalidade ou ilegalidade desses atos, a produção anual passou de 25 mil para mais de 70 mil atos. Em 2005, a produção anual atingiu patamar superior a 90 mil.

A partir de 2009, novas críticas eletrônicas passaram a acessar campos de vários sistemas informatizados da Administração Pública Federal, como o Sistema Integrado de Recursos Humanos (Siape), o Sistema Informatizado de Controle de Óbitos (Sisobi); Relação Anual de Informações Sociais (Rais), Cadastro de Pessoa Física (CPF), elevando a produção anual de apreciação para mais de 120 mil atos, com incremento da qualidade das decisões em função da integração de sistemas.

Em 2013, o Sisac continha mais de 4 milhões de atos de pessoal, cada ato com mais de 100 campos cada. No mesmo ano, o TCU alcançou a marca de 1 milhão de atos apreciados de forma automática, sem intervenção humana.

Contudo, o sistema demandou melhorias estruturantes. Conforme relatório apresentado pela Sefip, em 2017, "informações que chegam ao TCU por meio do Sisac, por vezes, são insuficientes para análises automatizadas mais precisas" (RIBEIRO, 2017), acarretando realização de novas diligências ao jurisdicionado para sanear inconsistências e descumprimento de os prazos processuais.

O **e-Pessoal** é sistema concebido para substituir o Sisac (TCU, 2019b). Conforme veiculado em notícia publicada no Portal do TCU (PORTAL TCU, 2018), tem como objetivo incrementar novas funcionalidades na apreciação automática de atos de pessoal. Como benefícios esperados visa: Agilizar a tramitação de processos; facilitar o controle de erros, reduzir o esforço dos órgãos envolvidos nas atividades de controle interno e de gestão de pessoal; aumentar a qualidade das avaliações processuais; e diminuir o tempo entre a emissão e o julgamento do ato pelo TCU.

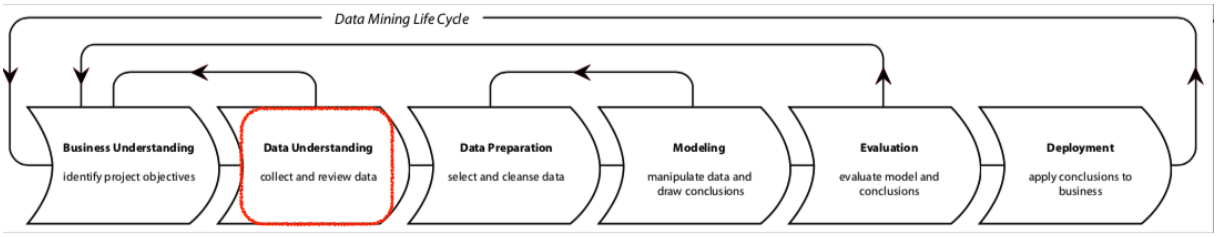
Ainda segundo a notícia, com a implementação do e-Pessoal, o Sisac seria desativado no dia 05 de março de 2018, condicionando os órgãos da Administração Pública Federal a adotar a

nova plataforma. Ainda no contexto de inovações, o sistema permite anexação de documentos, como laudos médicos e decisões judiciais, e permite ao controle interno dos órgãos a emissão de pareceres sobre aspectos específicos de atos, como avaliação da justificativa do gestor sobre algum apontamento detectado pelo sistema de forma automática.

6. ENTENDIMENTO DOS DADOS

A fase de entendimento dos dados compreende o acesso e a exploração de dados disponíveis para mineração usando tabelas e gráficos, permitindo determinar qualidade e descrever resultados obtidos na documentação do projeto (IBM, 2014).

Figura 13: Ciclo de vida - mineração de dados - Entendimento dos dados



Fonte: (LEAPER, 2009).

Nesta fase, foi realizado: a) levantamento e estudo de bases e de modelos de dados considerados úteis para o corrente trabalho; b) levantamento de campos e formatação; c) autorização para acesso aos sistemas e aos dados mencionados, apenas como leitura; d) levantamento de formatos disponíveis para entrega. Foram consideradas as seguintes fontes:

- Base de dados do RADAR/RADEX;
- Base de dados do SIAPE; e
- Busca textual de jurisprudência sobre acórdãos.

O objetivo inicial é estabelecer dois elos que ao final são ligados. De um lado, os dados relacionados às decisões do Tribunal acerca de monitoramento de acórdãos de pessoal (RADAR e Busca Textual). De outro, o critério, formado por bases de dados relacionadas ao jurisdicionado (SIAPE).

Além disso, conforme sinalizado na metodologia de referência, foi demandado tempo significativo com entrevistas e consultas a áreas responsáveis pelos dados, a fim de compreender cada base. O entendimento prático acabou se tornando essencial, inclusive, em situações de documentação escassa ou desatualizada.

Outro aspecto relevante foi a necessidade de contatar mais de um responsável, particularmente em situações de fragmentação de conhecimento. As duas situações acabaram se tornando comuns, como estimado no início do projeto.

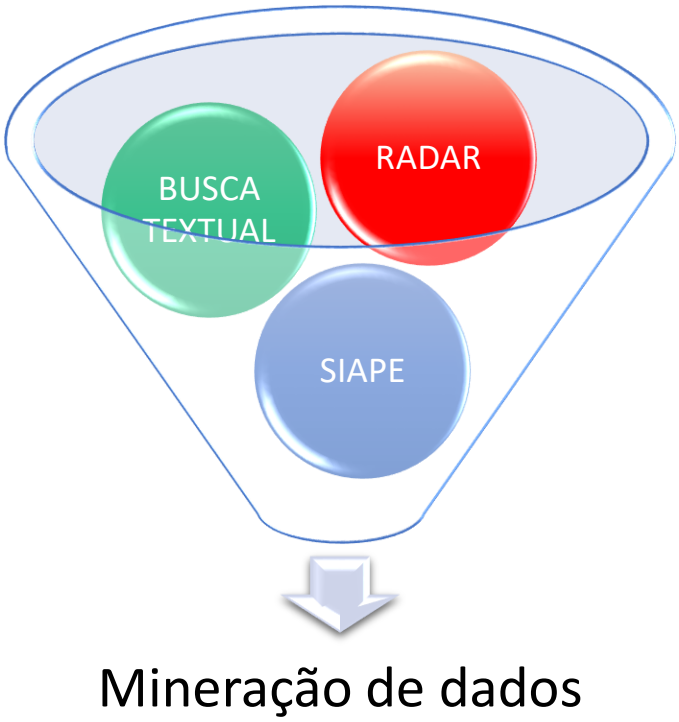
6.1. RADAR

Para facilitar o entendimento do ambiente de produção do TCU, foi solicitada a disponibilização de modelo dados do **RADAR**, a fim de facilitar a identificação de ligações do modelo relacional, bem como permissões de acesso às tabelas consideradas necessárias.

O acesso somente de leitura às informações registradas no RADAR foi autorizado pelo Secretaria de Soluções de Tecnologia da Informação (STI), manifestada concordância prévia da subunidade responsável pela gestão da informação, Serviço de Cadastramento de Informações (Secinf/Seproc). O acesso foi franqueado por tempo determinado, enquanto perdurar o presente trabalho (chamado SD 895674, de 06 de maio de 2019).

O modelo de dados contendo inventário de tabelas foi disponibilizado em forma de links (STI, 2019a) e (STI, 2019b). Foram selecionadas as seguintes tabelas/visões - Tabela 2, conforme extração realizada por consulta SQL, especificada no ANEXO A.

Figura 14: Bases de dados envolvidas



Fonte: Elaborada pelo autor (2019).

Tabela 2: Inventário de tabelas - RADAR

Inventário de Tabelas – RADAR	
Tabela/Visão	**Descrição**
VW_ESC_DEL_EFET_COD_APREC (VW)	Trata-se de visão contendo principais campos de interesse relacionados ao RADAR: COD_PROCESSO: código do processo, SEQ_DELIBERACAO: número de sequência de deliberação, TIPO_DELIBERACAO: tipo de deliberação (conforme especificado na tabela TIPO_DELIBERACAO), NUMDELIB: número de deliberação, ANO_ACORDAO e NUM_ACORDAO: informações derivadas de NUMDELIB para obtenção de ano e número do acórdão, DESCR: trata-se de campo importante, pois contém descritivo da proposta de encaminhamento do acórdão, bem como prazo e órgãos jurisdicionados envolvidos, ANO_PROCESSO e NUM_PROCESSO: ano e número do processo, COD_TIPO_DECISAO: tipo de decisão relacionada ao acórdão (conforme especificado na visão VW_SF_APRECIACAO), ANO_DECISAO, APRECIADOR: 1° ou 2° Câmara ou Plenário do TCU, DATA: data de publicação.
PROCESSO_GESTAO (PG)	Tabela contendo informações sobre processo registrado na visão anteriormente descrita. Relaciona-se pelos campos: VW.COD_PROCESSO = PG.COD.
UNIDADE_ORGANIZACIONAL_TCU (UOT)	Tabela contendo descritivo de unidades técnicas do TCU. A sigla da unidade técnica é identificável pelo campo UOT.SIGLA. Para a Sefip, o código relacionado à pesquisa de substring "SEFIP" resulta em 191200 (*upper(uot.sigla) like '%SEFIP%'*). Relaciona-se pelos campos: PG.COD_UNID_RESPONSAVEL_TECNICA = UOT.COD.
TIPO_DELIBERACAO (TD)	Tabela contendo descritivo de tipos de deliberação especificado no campo VW.TIPO_DELIBERACAO. Trata-se de tabela importante na determinação de escopo do trabalho. Sendo assim, a lista completa de tipos encontra-se especificada no ANEXO A). Relaciona-se pelos campos: VW.TIPO_DELIBERACAO = TD.COD.
VW_SF_APRECIACAO (VSA)	Visão contendo descritivo de tipos de deliberação especificados no campo VW. COD_TIPO_DECISAO. Relacionamento: VW.COD_TIPO_DECISAO = VSA.COD.

Fonte: Elaborada pelo autor (2019).

Ainda sobre inventário de dados do RADAR, foram contabilizados no dia da coleta, 639.606 registros de deliberações armazenados na visão VW, associados à Sefip (código 191200), considerando acórdãos publicados entre 01/01/2014 e 20/09/2019.

6.2. Busca Textual

O levantamento de informações da busca textual de Jurisprudência foi disponibilizado pelo Seint/STI na forma de arquivos textuais individualizados. Ou seja, cada conteúdo de acórdão armazenado documento (.txt), sendo feita distinção do tipo de apreciador, seja Primeira ou Segunda Câmara ou Plenário. Ao total, foram disponibilizados 109.776 arquivos, compreendidos entre 2014 e 2019. Foram selecionadas as tabelas/visões especificadas a seguir:

Tabela 3: Inventário de dados - busca textual

Inventário de dados – Busca textual	
Tabela/Visão	Descrição

NOME	Nome do interessado. Trata-se de dado nem sempre disponível no conteúdo textual do campo de deliberações do RADAR (VW.DESCR), porquanto a menção a interessados pode ser feita de forma indireta, a exemplo de "os referidos".
CPF	CPF do interessado. Trata-se de dado nem sempre disponível no conteúdo textual do campo de deliberações do RADAR (VW.DESCR), porquanto a menção a interessados pode ser feita de forma indireta, a exemplo de "os referidos". Além disso, para facilitar a realização de operações de junção com dados do SIAPE, a formatação foi atribuída como sendo sequência de 11 (onze) caracteres.
NUM_PROCESSO	Número do processo. Formatação do tipo numérico inteiro.
ANO_PROCESSO	Ano do processo. Formatação do tipo numérico inteiro.
TIPO_ATO	O tipo do ato pode ser de admissão (ATO(S) DE ADMISSÃO) e de concessão de aposentadoria (APOSENTADORIA), reforma (REFORMA) e pensão (PENSÃO). Trata-se de dado nem sempre disponível ou mencionado de forma não uniformizada no conteúdo textual do campo de deliberações do RADAR, dificultando a extração do dado no conteúdo textual (VW.DESCR).
NUM_ACORDAO	Número do acórdão. Formatação do tipo numérico inteiro.
ANO_DECISAO	Ano da decisão. Formatação do tipo numérico inteiro.
APRECIADOR	O apreciador pode ser Plenário, Primeira Câmara ou Segunda Câmara.

Fonte: Elaborada pelo autor (2019).

6.3. SIAPE

O acesso somente de leitura às informações registradas no **SIAPE** foi autorizado pela Sefip. O acesso foi franqueado por tempo determinado, enquanto perdurar o presente trabalho.

O acesso à base de dados BD_SIAPE foi realizada por meio do LabContas (servidor SRV-BD-INT-2). O entendimento do modelo de dados foi compartilhado pela equipe de auditoria da Sefip. Foram selecionadas as seguintes tabelas/visões a seguir:

Tabela 4: Inventário de tabelas - SIAPE

	Inventário de tabelas - SIAPE
Tabela/Visão	**Descrição**
HIST_SERVIDOR (HS)	Trata-se de tabela contendo informações de servidores. Armazena registros de Cadastro de Pessoa Física (NUM_CPF, formado por 11 caracteres), viabilizando operações de junção de registros de mesma natureza obtidos na extração textual de acórdãos, relativos a **servidores**. Na mesma tabela, há registros que viabilizam operações de junção com outras tabelas do SIAPE, como identificador de matrícula do servidor (NUM_MATRICULA, do tipo numérico, contendo 7 dígitos, ou simplesmente numérico**(7)**), de órgão de origem (COD_ORGAO, do tipo numérico**(5)**) e mês de referência (ANOMES_FOLHA, do tipo data com representação reversa, YYYY-MM-DD), no caso de pagamento. Os referidos campos são, em suma, os principais elementos da tabela.
HIST_PENSIONISTA (HP)	Trata-se de tabela contendo estrutura análoga à HS. Contudo, com informações de pensionistas. Armazena registros de Cadastro de Pessoa Física (NUM_CPF, formado por 11 caracteres), viabilizando operações de

	junção de registros de mesma natureza obtidos na extração textual de acórdãos, relativos a **pensionistas** [6]. Na mesma tabela, há registros que viabilizam operações de junção com outras tabelas do SIAPE, como identificador de matrícula do **pensionista** (NUM_MATRICULA, do tipo numérico**(8)**[7] e mês de referência (ANOMES_FOLHA, do tipo data com representação reversa, YYYY-MM-DD), no caso de pagamento. Os referidos campos são, em suma, os principais elementos da tabela.
RUBRICA (R)	Tabela contendo descritivo de valores consolidados de pagamentos, associados a servidores, conforme alocação em órgãos e periodicidade, conforme registros respectivos obtidos pelos campos NUM_MATRICULA, COD_ORGAO e ANOMES_FOLHA, cujas operações de junção resultantes são: HS.NUM_MATRICULA =R.NUM_MATRICULA; HS.COD_ORGAO =R.COD_ORGAO e HS.ANOMES_FOLHA=R.ANOMES_FOLHA.
DADOS_PENSAO (DP)	Tabela contendo descritivo acerca de pensões, incluindo a associação entre **beneficiário** (NUM_MATRICULA_BENEFICIARIO, do tipo numérico**(8)**[8]), respectivo mês de pagamento de referência (ANOMES_FOLHA) e **instituidor** (ou seja, o servidor ao qual o beneficiário está relacionado; NUM_MATRICULA_INSTITUIDOR, do tipo numérico**(7)**) de determinado órgão (COD_ORGAO). Uma possível operação de junção com dados de pensionistas (HP), relativamente a determinado período de pagamento é HP.NUM_PENSIONISTA = DP. NUM_MATRICULA_BENEFICIARIO; HP.ANOMES_FOLHA = DP.ANOMES_FOLHA. Adicionalmente, a referida junção pode fornecer informações do instituidor e do órgão ao qual encontra-se associado: DP.NUM_MATRICULA_INSTITUIDOR e DP.COD_ORGAO.
RUBRICA_PENSIONISTA (RP)	Trata-se de tabela contendo estrutura análoga à R. Contém descritivo de valores consolidados de pagamentos (rubricas), associados a **pensionistas**. Uma possível operação de junção com a tabela DADOS_PENSAO (DP), relativamente a determinado período de pagamento é: RP.NUM_MATRICULA_PENSIONISTA = DP. NUM_MATRICULA_ BENEFICIARIO; RP.ANOMES_FOLHA = RP.ANOMES_FOLHA; RP. COD_ORGAO=DP.COD_ORGAO; RP.NUM_MATRICULA_ INSTITUIDOR = DP.NUM_MATRICULA_INSTITUIDOR.
TIPO_RUBRICA (TR)	Tabela contendo detalhamento de rubricas registradas em R e RP. Uma possível operação de junção é feita com o código da rubrica:

[6] Trata-se de ponto de atenção entre campos de mesma sintaxe observado em tabelas distintas. O campo NUM_CPF, da tabela HS, está associada à identificação de servidores. O campo de mesma sintaxe encontrada na tabela HP, à de pensionistas. Por esta razão, a segunda tabela não associa o identificador de órgão ao pensionista registrado na tabela. Fá-lo-á em tabelas específicas para associar o órgão do servidor ao qual o pensionista está associado.

[7] Trata-se de ponto de atenção entre campos de mesma sintaxe observado em tabelas distintas. O campo NUM_MATRICULA, da tabela HS, possui 7 dígitos. Na tabela, HP, o campo de mesma sintaxe possui 8.

[8] Trata-se de reforço ao ponto de atenção entre campos de mesma sintaxe observado em tabelas distintas. O campo NUM_MATRICULA, da tabela HS, possui 7 dígitos. Na tabela, HP, o campo de mesma sintaxe possui 8.

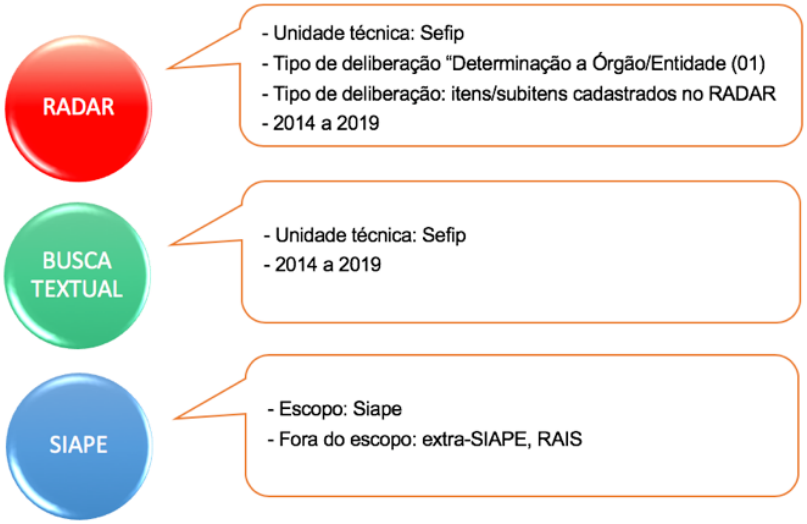
	TR.COD_RUBRICA=R.COD_TIPO_RUBRICA; TR.COD_RUBRICA=RP.COD_TIPO_RUBRICA. São campos relevantes o identificador da rubrica (NOME_RUBRICA) e a vigência (SE_VIGENTE), com valores possível 's' e 'n', sim e não, respectivamente. Da sintaxe, infere-se semântica acerca da vigência da rubrica.

Fonte: Elaborada pelo autor (2019).

6.4. Seleção de registros (escopo)

Quanto ao **escopo**, os seguintes critérios foram adotados, inclusive para viabilizar a execução do trabalho (Figura 15). O levantamento foi delimitado às deliberações de acórdãos, cuja unidade técnica consta como sendo a Sefip. Convenção aplicável às bases de dados do RADAR e Busca Textual.

Figura 15: Delimitação de escopo



Fonte: Elaborada pelo autor (2019).

Em relação ao período selecionado, o limite inferior de 2014 foi convencionado porque a temporalidade de apreciação de atos, conforme informado em entrevista, encontra-se no referido ano. O limite superior de 2019 foi convencionado por simplificação, uma vez que deliberações podem estar associadas a prazos de cumprimento, nos termos do Resolução-TCU n° 255, de 26 de setembro de 1991 (TCU, 1991).

No RADAR, as deliberações associadas ao tipo 25 - "Legalidade de Atos de Admissão e Concessão" - foram excluídas do escopo, pois não geram necessidade de monitoramento (Figura 16). Entre as deliberações remanescentes, o escopo foi limitado ao tipo 1 "Determinação ao Órgão/Entidade", por ser majoritário e figurar entre itens a serem monitorados, ou seja, contém deliberações versando sobre cessão de pagamento, tema de interesse do trabalho.

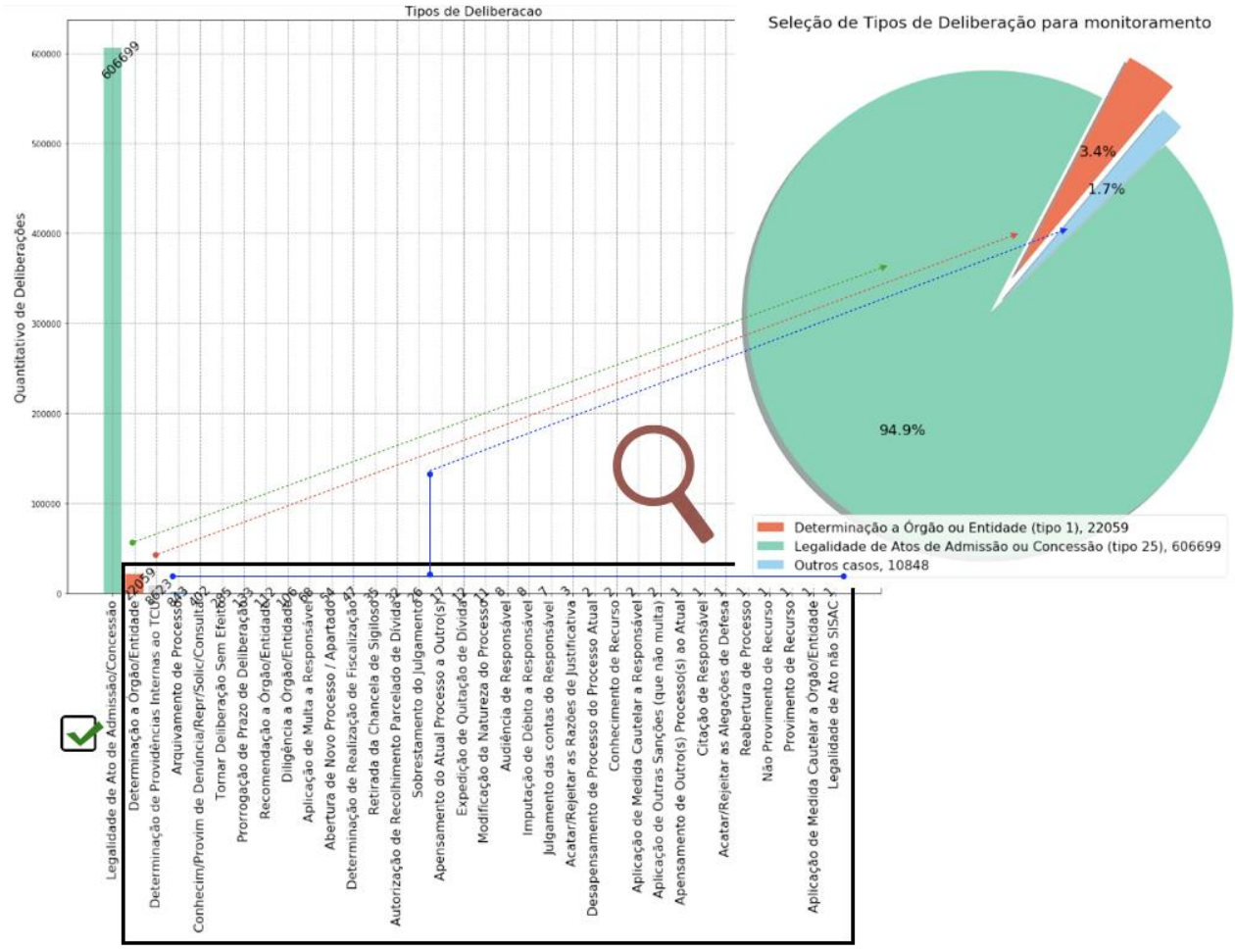
Figura 16: Tipos de deliberação

TIPO_DELIBERACAO	DESCR	
25	Legalidade de Ato de Admissão/Concessão	397319
1	Determinação a Órgão/Entidade	9528
15	Determinação de Providências Internas ao TCU	3665
5	Arquivamento de Processo	314
3	Conhecim/Provim de Denúncia/Repr/Solic/Consulta	215
8	Sobrestamento do Julgamento	79
45	Recomendação a Órgão/Entidade	60
2	Diligência a Órgão/Entidade	43
39	Tornar Deliberação Sem Efeito	41
11	Retirada da Chancela de Sigiloso	23
6	Abertura de Novo Processo / Apartado	20
26	Determinação de Realização de Fiscalização	14
41	Prorrogação de Prazo de Deliberação	12
12	Apensamento do Atual Processo a Outro(s)	10
17	Citação de Responsável	6
29	Autorização de Recolhimento Parcelado de Dívida	6
16	Aplicação de Multa a Responsável	6
7	Modificação da Natureza do Processo	5
22	Julgamento das contas do Responsável	2
20	Imputação de Débito a Responsável	2
23	Audiência de Responsável	2

Name: COD_PROCESSO, dtype: int64

Fonte: Elaborada pelo autor (2019).

Figura 17: Distribuição de tipos de deliberação de acórdãos de pessoal



Fonte: Elaborada pelo autor (2019).

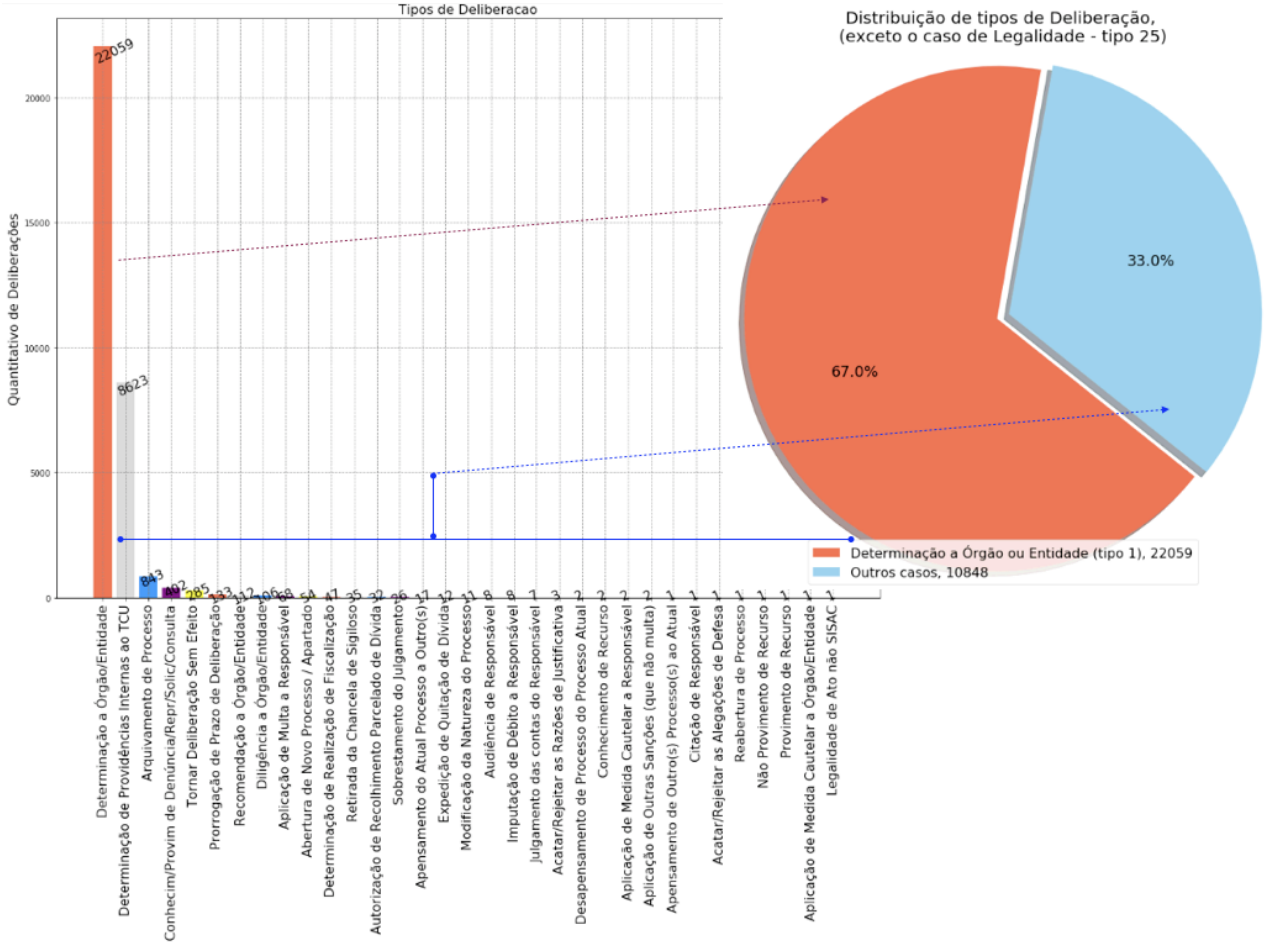
A Figura 17, apresenta distribuição de tipos de deliberação de acórdãos de pessoal. Observa-se prevalência de casos relacionados a legalidade de apreciação de atos de admissão e de concessão (606.699 casos), tipo 25, em relação aos demais tipos.

Ocorre que, ao excluir o tipo mais comum, outra discrepância pode ser observada. O tipo 1 - determinação a Órgão/Entidade torna-se discrepante em relação aos demais (22.059 casos) (Figura 18).

Então, excluídos os casos de legalidade de atos de admissão e de concessão (tipo 25), 67,8% representam casos do tipo 1, o qual comumente figuram itens acerca de cessação de pagamentos a serem monitorados.

Diante do exposto, deliberações associadas a "Determinação a Órgão/Entidade" representam parcela significativa de itens a serem monitorados, conforme período de apuração, sendo esse o escopo da monografia.

Figura 18: Distribuição de tipos de deliberação de acórdãos de pessoal, excluindo casos de legalidade



Fonte: Elaborada pelo autor (2019).

Acerca da base de dados empregada como conferência, o escopo foi limitado ao SIAPE. A base de dados do extra-SIAPE e da RAIS foram selecionados como trabalhos futuros, a fim de dimensionar o escopo do trabalho ao tempo disponível.

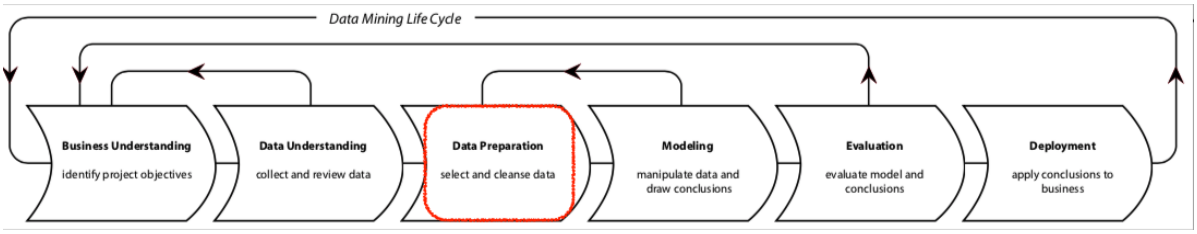
Por fim, escopo do trabalho foi limitado ao monitoramento de atos de pessoal. As deliberações decorrentes dos convencionados "não-atos" foram excluídas do trabalho, tendo em vista viabilizar o projeto dentro do tempo e dos recursos disponíveis.

7. PREPARAÇÃO DE DADOS

Baseado na coleta inicial de dados conduzida na fase anterior, a próxima etapa é a seleção de dados relevantes às metas de mineração de dados. A fase de preparação de dados compreende procedimentos para derivar novos atributos, remover ou substituir valores em branco ou omissos, agregar registros, mesclar conjuntos de dados, classificar dados para modelagem, selecionar amostragem de dados, dividir em conjuntos de dados de treinamento e de teste.

O CRISP-DM considera um dos aspectos mais importantes e frequentemente mais demorados da mineração de dados, sendo responsável por consumir 50 a 70% do projeto. A dedicação às primeiras fases de entendimento de negócio e de dados pode minimizar o impacto, mas pode remanescer demanda significativa de na preparação de dados para mineração (IBM, 2014).

Figura 19: Ciclo de vida - mineração de dados - Preparação dos dados



Fonte: (LEAPER, 2009).

Na preparação de dados, as bases e modelos de dados selecionados na fase anterior passaram por levantamento de campos úteis, formatação, autorização para acesso aos sistemas e aos dados mencionados, apenas como leitura, levantamento de formatos disponíveis para entrega.

Além disso, conforme sinalizado na metodologia de referência, foi demandado tempo significativo com entrevistas e consultas a áreas responsáveis pelos dados, a fim de compreender cada base. O entendimento prático acabou se tornando essencial, inclusive, em situações de documentação escassa ou desatualizada. Outro aspecto relevante foi a necessidade de contatar

mais de um responsável, particularmente em situações de fragmentação de conhecimento. As duas situações não foram consideradas incomuns como se estimou no início do projeto. Para facilitar a organização, a seção foi dividida conforme fontes selecionadas na fase anterior, a citar:

- Base de dados do RADAR/RADEX;
- Base de dados do SIAPE; e
- Busca textual de jurisprudência sobre acórdãos.

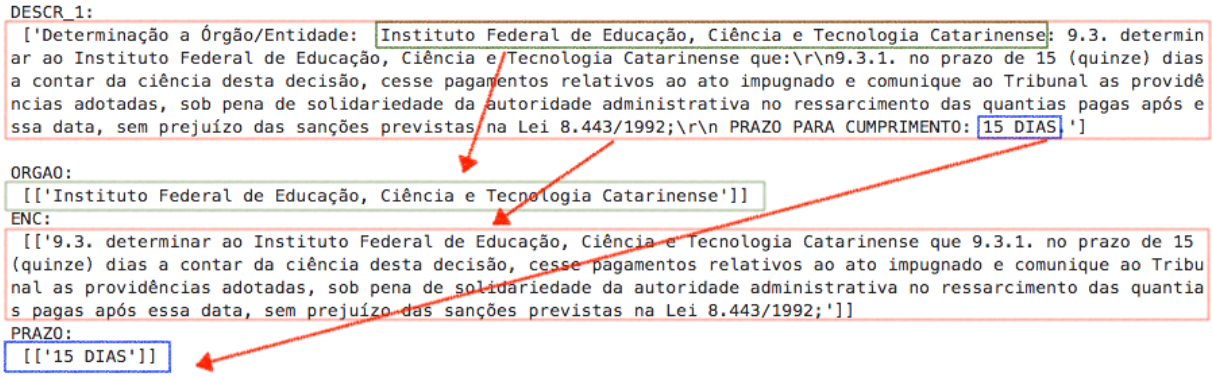
7.1. RADAR

Sobre a base de dados do RADAR os seguintes procedimentos de limpeza foram executados. Carregamento de dados coletados da base de dados para dataframe no ambiente *Jupyter Notebook* (PROJECT JUPYTER, 2019), empregando linguagem *Python* (PYTHON, 2001) e a biblioteca Pandas (THE PANDAS PROJECT, 2019) .

Foi retificada ou ratificada, conforme o caso, a conversão de tipo para formato inteiro para os seguintes campos, uma vez que a biblioteca Pandas é capaz de realizar tipagem automática de variáveis: COD_PROCESSO, PROCESSO, SEQ_DELIBERACAO, TIPO_DELIBERACAO, ANO_PROCESSO, NUM_PROCESSO, COD_APRECIACAO, COD_TIPO_DECISAO e ANO_DECISAO. Para o formato de cadeias de caracteres (“object”): DESCR, NUMDELIB, ORGAO, ENC, PRAZO, SIGLA e APRECIADOR.

Foram derivados novos atributos a partir de campo descritivo de itens/subitens de deliberações (DESCR_1). Em cada registro, foram mescladas informações acerca de Órgão ou Entidade jurisdicionada, prazo para cumprimento da deliberação e o dispositivo em si. Os dados foram extraídos e segmentados em campos distintos, a citar: ORGAO, PRAZO, ENC, respectivamente. O atributo de origem foi descartado.

Figura 20: Extração de informações armazenadas em campos textuais do RADAR



Fonte: Elaborada pelo autor (2019).

Por outro lado, para o processamento de itens e subitens de encaminhamentos, optou-se inicialmente por não aplicar técnicas de processamento textual, como remoção de *stopwords*, *stemização* e *lematização.* Optou-se por empregar tokenização por "n-gramas" com parâmetros restritivos. A justificativa será detalhada na fase de modelagem.

Finalmente, em novo ciclo de execução da metodologia, foi observada a possibilidade de emprego do registro DATA, relacionado à data de publicação do acórdão. No escopo dos dados recebidos, não foram encontrados registros de tempo que permitissem associação ao prazo processual de cada deliberação. Sendo assim, foi estabelecida condição de contorno, a ser descrita em detalhes na modelagem.

7.2. Busca Textual

Para a busca textual de acórdãos foi realizada extração de dados diretamente dos arquivos textuais. O carregamento de dados foi feito para o mesmo ambiente descrito para o RADAR.

A junção da busca textual com a base do Radar foi necessária por várias motivações. Primeira: A extração de partes interessadas diretamente do conteúdo textual dos itens e subitens cadastrados no RADAR mostrou-se onerosa para o tempo do projeto. Segunda: Em testes substantivos, verificou-se registros contendo menção indireta de interessados, inviabilizando a extração de dados simplesmente por inexistência de menção explícita.

Terceira: Dado que o escopo do projeto foi redefinido para atos de pessoal, durante a preparação dos dados, observou-se a conveniência e a oportunidade de extrair o identificador do tipo do ato. No mesmo sentido da primeira motivação, a extração do tipo do ato diretamente do RADAR mostrou-se onerosa para o tempo do projeto, pois estava contido de maneira não uniforme no texto das deliberações. Em algumas situações verificadas em amostras, a informação acerca do tipo do ato era somente interpretada, inviabilizando a extração por ausência de dados.

É importante destacar a redundância premeditada de dados extraídos dos acórdãos em relação aos obtidos na base anterior. A decisão foi tomada para facilitar a validação de registros, sem comprometer a viabilidade do projeto. Então, além da identificação dos interessados em cada deliberação, por meio de NOME e CPF, e do tipo do ato (TIPO_ATO), os seguintes campos redundantes foram extraídos: NUM_PROCESSO, ANO_PROCESSO, NUM_ACORDAO, ANO_DECISAO, APRECIADOR.

Foi retificada ou ratificada, conforme o caso, a conversão de tipo para formato inteiro para os seguintes campos, uma vez que a linguagem é capaz de realizar tipagem automática de

variáveis: NUM_PROCESSO, ANO_PROCESSO, NUM_ACORDAO, ANO_DECISAO, APRECIADOR. Para o formato de cadeias de caracteres (“object”): NOME, CPF, TIPO_ATO.

Para o campo CPF, o formato numérico foi dispensado para manter alinhamento com a base de dados do SIAPE, que adota tipagem de cadeia de caracteres de tamanho fixo 11 (VARCHAR(11)), facilitando operações de junção durante a fase de modelagem. Para o campo APRECIADOR, foi procedido tratamento de conteúdo, apenas para preservar convenção adotada nos dados cadastrados no RADAR. Ou seja, indicadores numéricos ordinais foram substituídos por correspondentes por extenso.

Figura 21: Extração de dados textuais - amostragem - Acórdão n°10.974/2015 (2ª Câmara)

Fonte: Elaborada pelo autor (2019).

Para a base de dados do RADAR e a busca textual de Jurisprudência foi necessário proceder mescla de conjuntos de dados, a fim de consolidar os registros obtidos nas duas bases.

Para ambas as bases também foi procedida a remoção e a substituição de valores. Foram excluídos caracteres, inclusive de tabulação, quebra de linha, em branco, caracteres alfanuméricos e termos considerados não relevantes para o objetivo da mineração.

Um exemplo de extração textual encontra-se em destaque na amostragem contendo informações do Acórdão n°10974/2015 – 2ª Câmara, contendo tabulação de campos de interesse, segundo o inventário supracitado (Figura 21).

7.3. SIAPE

Para a base de dados do SIAPE, os seguintes procedimentos de limpeza foram executados: seleção de campos de interesse, associados a tipos de rubrica vigentes ([tipo_rubrica] .SE_VIGENTE='S'); conversão para tipo inteiro dos campos NUM_MATRICULA, COD_ORGAO e COD_TIPO_RUBRICA. Os campos foram extraídos das tabelas HIST_SERVIDOR, RUBRICA e TIPO_RUBRICA.

A conexão ao banco de dados BD_SIAPE foi feita no ambiente Python empregando biblioteca PYODBC. Para viabilizar o carregamento e a análise no referido ambiente, a consulta foi limitada aos registros de CPFs contidos nos registros classificados no módulo de levantamento de tipologias, como parte integrante do espaço de busca. Maior detalhamento sobre arquitetura será descrito no capítulo sobre modelagem.

Para todas as bases empregadas, foi convencionada amostragem de dados, a fim de agilizar o processo de execução de código. Para facilitar a organização de códigos e arquivos de entrada e saída, as amostragens foram identificadas com o sufixo “demo”. Não houve necessidade de divisão de conjunto de dados de treinamento e de teste, devido à escolha de modelos não supervisionados de clusterização.

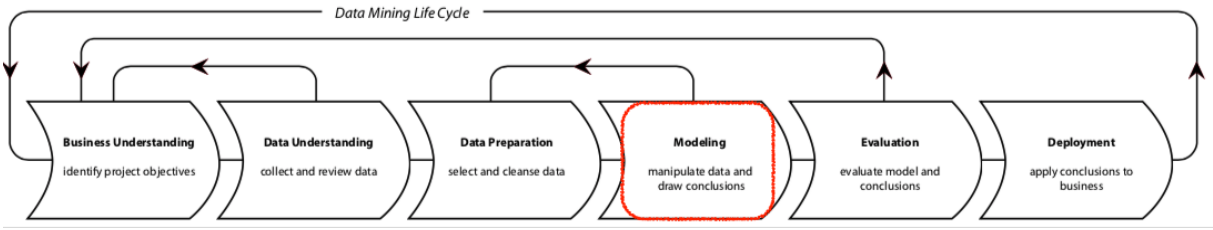
8. MODELAGEM

A **fase** de **modelagem** compreende o desenvolvimento de protótipo, com o emprego de técnicas de mineração de dados, *clusterização* e classificação de *clusters, parser* de deliberações de um (ou mais) *clusters* que apresentarem razoável homogeneidade e transformação de deliberações em *scripts* de tipologias, bem como validação do modelo segundo critérios de sucesso de mineração de dados, por meio de critérios como matriz de confusão e acurácia.

Diversas iterações foram necessárias até o desenvolvimento da versão final do modelo. Ajustes de parâmetros foram necessários para a melhoria de resultados. O próprio modelo passou

por várias disposições, particularmente impostas para minimizar problemas de desempenho. Além disso, o modelo foi dividido em vários módulos ou arquivos (*scripts*) para facilitar o desenvolvimento e a depuração de problemas. A ligação entre cada módulo é feita basicamente por arquivos de entrada/saída. Além disso, esta seção abordará tanto aspectos da fase de modelagem como da fase de preparação de dados, visto que essas duas fases interagem fortemente entre si.

Figura 22: Ciclo de vida - mineração de dados – Modelagem



Fonte: (LEAPER, 2009).

8.1. Visão Geral

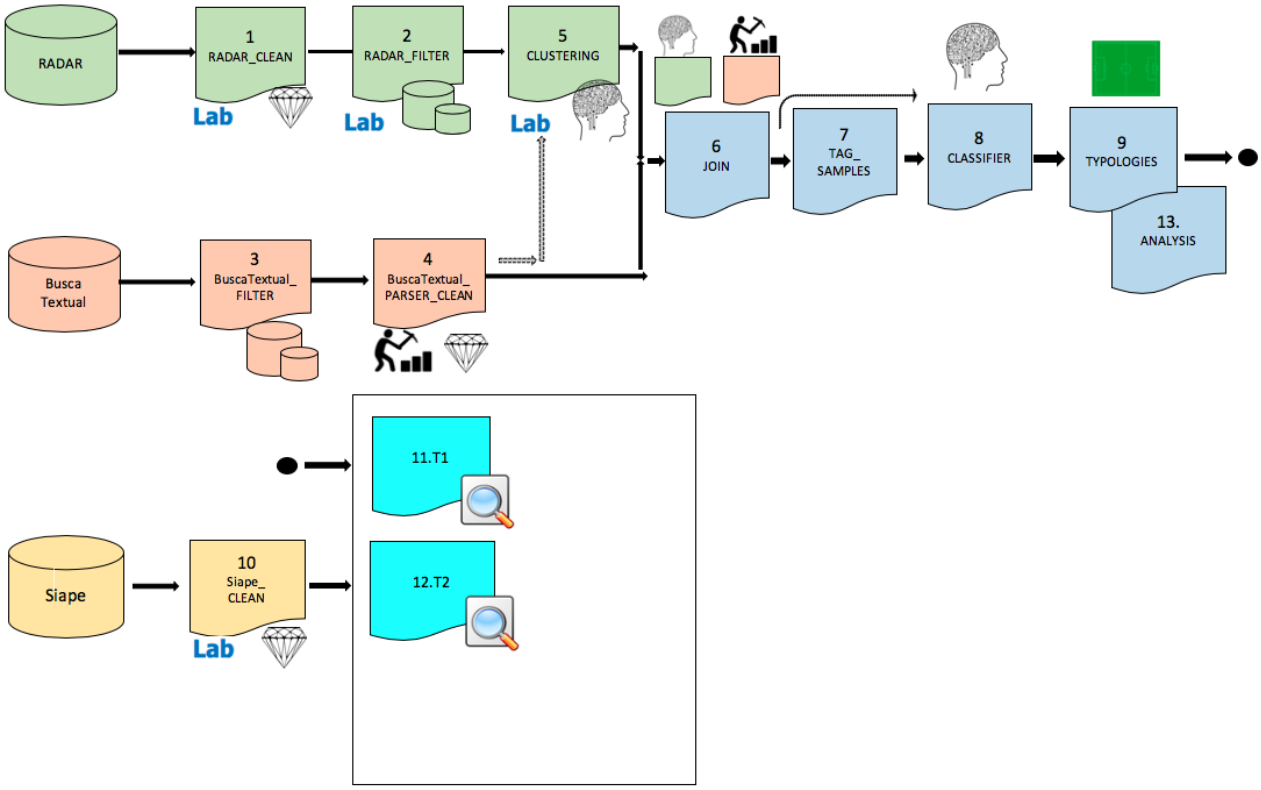
A arquitetura proposta no presente trabalho encontra-se na Figura 23. Embora a metodologia CRISP-DM (IBM, 2014) não estabeleça explicitamente a qual capítulo pertence o modelo, tomou-se a iniciativa de apresentá-lo na fase da modelagem.

Ainda conforme a referida metodologia, diversas iterações foram necessárias até o desenvolvimento da versão final. Ajustes de parâmetros foram necessários para a melhoria de resultados. O próprio modelo passou por várias modificações, particularmente impostas para minimizar problemas de desempenho. Além disso, o modelo foi dividido em vários módulos ou arquivos (*scripts*) para facilitar o desenvolvimento e a depuração de problemas. A ligação entre cada módulo é feita basicamente por arquivos de entrada/saída.

Para facilitar a visualização do modelo proposto, conjuntos de módulos sequenciais foram identificados por cor distinta. Desta forma, 05 (cinco) raias principais foram criadas: VERDE, VERMELHA, AMARELA, AZUL, AZUL CLARA.

Primeira: Raia VERDE, formada por módulos responsáveis pela coleta, limpeza, filtragem e aplicação de algoritmo de clusterização. Segunda: Raia VERMELHA, formada por módulos responsáveis pela filtragem, *parser* e limpeza de dados provenientes da Busca Textual de acórdãos. Terceira: Raia AMARELA, constituída por módulo de consulta e limpeza de dados provenientes do SIAPE.

Figura 23: Arquitetura – visão geral



Fonte: Elaborada pelo autor (2019).

Os três primeiros conjuntos podem ser executados em paralelo. Uma exceção é feita ao módulo de clusterização, conforme detalhado na RAIA VERDE.

A quarta raia, AZUL, é responsável por fazer a junção dos resultados obtidos na primeira e segunda raias, por realizar a classificação por meio de algoritmo supervisionado, particularmente para rotular deliberações versando sobre pagamento. Em seguida, prepara os dados para distribuição para as diferentes tipologias propostas.

A última raia é a AZUL CLARA, responsável pela implementação das tipologias propriamente ditas.

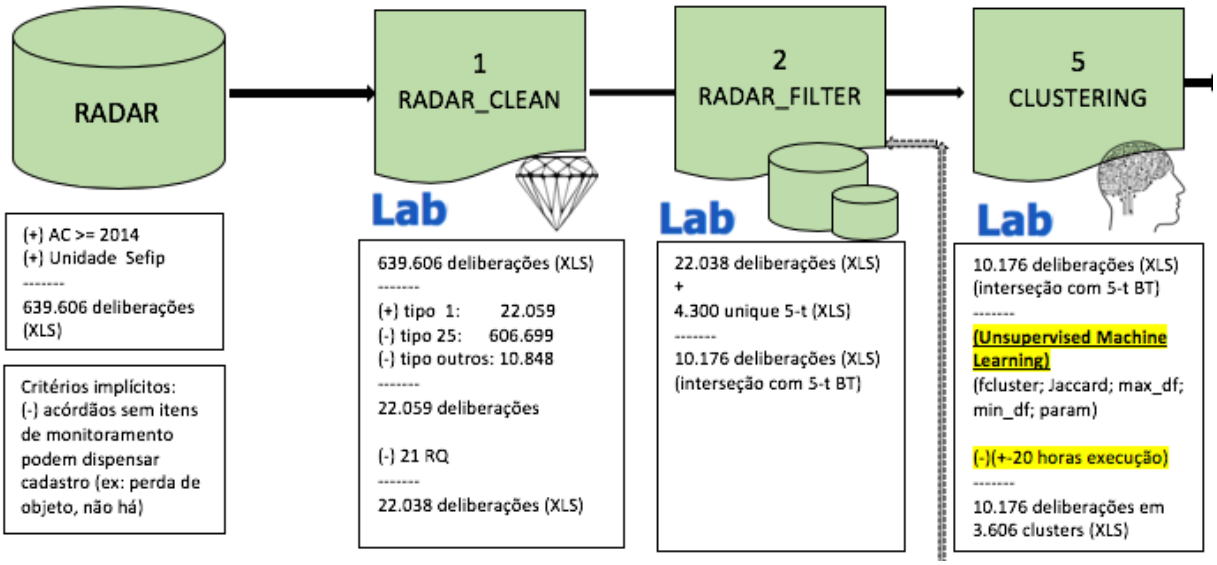
Para facilitar a identificação de características gerais de cada módulo, foram associados ícones acessórios como mnemônico. O ícone “Lab” indica a necessidade de execução no LabContas. O ícone contendo uma cabeça em perfil, indica a execução de algoritmos de aprendizagem de máquina. O contendo base maior e outra menor sinaliza a redução de domínio de dados de entrada, a fim de melhorar desempenho do modelo. O contendo uma pedra está associada a procedimentos de limpeza e formatação de dados. O contendo uma picareta está associado a procedimentos de *parser* de dados. Por fim, o meio-campo indica preparação de dados a serem distribuídos e a lupa, apresentação de resultados resultantes das tipologias. Nas próximas seções, cada raia será detalhada.

8.2. Raia VERDE (RADAR)

A raia VERDE é formada por módulos responsáveis pela coleta, limpeza, filtragem e aplicação de algoritmo de clusterização de dados oriundos do RADAR. Trata-se da vertente inicial do projeto, quando a intenção era aproveitar o caminho pavimentado pelo processo de catalogação de deliberações do referido sistema.

Em relação à obtenção dos mesmos dados via *parser* de arquivos textuais, o esforço, de plano, mostrava-se mais viável. Por outro lado, ao longo do projeto, observou-se que o caminho para a obtenção de dados por meio de arquivos textuais mostrou-se menos oneroso do que o inicialmente estimado. No mesmo sentido, alguns dados puderam ser obtidos apenas via busca textual, a exemplo do tipo do ato.

Figura 24: Raia Verde (RADAR)



Fonte: Elaborada pelo autor (2019).

O **módulo 1** – RADAR_CLEAN – procedeu a coleta de 638.606 deliberações cadastradas na base de dados do RADAR, a partir de acórdãos publicados a partir de 2014 até 20/09/2019. A data foi convencionada como linha estática de corte, cuja unidade técnica está relacionada à Sefip. Do montante, foram selecionadas 22.038 relacionadas ao tipo 1 – Determinação a órgão ou entidade, excluídas as deliberações não constantes em acórdão, mas sim em RQ (requerimento).

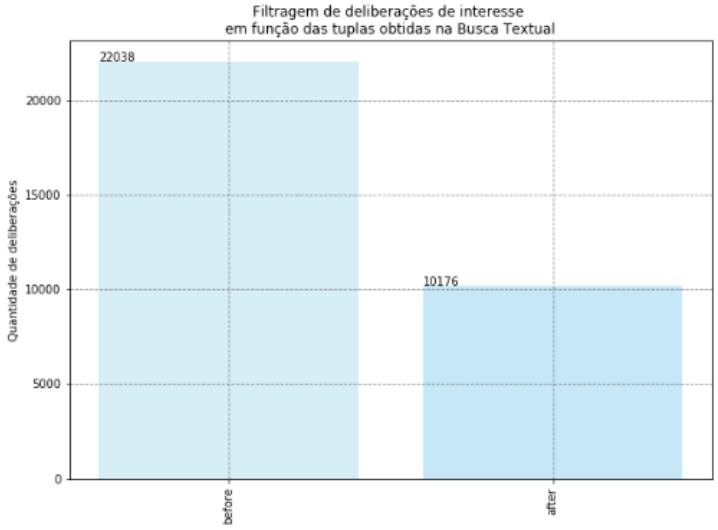
O **módulo 2** – RADAR_FILTER – tem como função reduzir o quantitativo de entradas para a clusterização (Figura 24). Verificou-se ao longo das iterações a necessidade de diminuir o domínio entrada do módulo seguinte (módulo 05) para ordem de grandeza que permitisse a sua

execução em tempo viável. Para cerca de 10.000 registros, verificou-se tempo de execução de cerca de 20 horas no ambiente LabContas. Para 30.000 registros, o tempo de execução foi abortado após o quarto dia de execução, sugerindo crescimento superior ao linear.

Adicionalmente, o espaço de entrada do módulo foi limitado ao domínio de acórdãos obtidos na RAIA VERMELHA, o qual viabilizou a coleta e o *parser* do tipo do ato de pessoal. Conforme descrito na fase de Preparação dos Dados, o tipo do ato pode até ser extraído a partir da base de dados do RADAR. Contudo, a sua menção pode estar dispersa em deliberações cadastradas no sistema ou inferida por interpretação textual, inviabilizando a extração no tempo do projeto. Sendo assim, a execução da RAIA VERDE é parcialmente paralela à VERMELHA, dada a dependência da entrada de resultados da referida raia no módulo 2.

A estratégia de redução de entradas foi baseada em combinações de registros obtidos a partir da raia VERMELHA, após *parser* da busca textual de acórdãos. O objetivo era selecionar apenas os registros de interesse selecionados na referida raia. Ao contrário do inicialmente previsto, constatou-se a possibilidade de obter dados a partir de *parser*, inclusive capazes de minimizar o domínio de entradas obtidas na raia VERDE, ou seja, a partir de dados do RADAR.

Figura 25: Filtragem de deliberações de interesse



Fonte: Elaborada pelo autor (2019).

Cabe ressaltar que os registros selecionados foram tuplas formadas por número e ano do acórdão, apreciador e número e ano de Processo – denominada **tupla 5-T**. Embora cada acórdão esteja relacionado a um Processo, o critério de lançamento de registros no RADAR é feito por itens ou subitens de deliberação. Sendo assim, a relação passa a ter cardinalidade múltipla (*m:n*). Ressalta-se também a tupla de cinco registros não representa chave primária, dado que um mesmo acórdão pode conter mais de uma deliberação. A intenção é simplesmente garantir

corretude e unicidade de caminho na associação reversa entre subitens e itens de deliberação e o respectivo documento que o originou.

Sendo assim, foram obtidas 4.300 tuplas 5-T (únicas), a partir da raia VERMELHA, possibilitando a redução de 22.038 registros para 10.176 (Figura 25).

O **módulo 5** – CLUSTERING – tem como função proceder o agrupamento de deliberações. (ZAKI e MEIRA JR., 2014) definem *clustering* como sendo o particionamento de elementos em grupos naturais chamados *clusters*, segundo critérios de similaridade previamente definidos.

Os dados de entrada deste módulo são provenientes da coluna de encaminhamentos (ENC), processados no módulo anterior (2 – RADAR_FILTER), conforme detalhamento descrito no inventário da Tabela 2 (item 6.1).

Foi escolhido modelo não supervisionado de aprendizagem de máquina. SONI (2019) analisa distinção entre algoritmos supervisionados e não supervisionados. Esclarece ainda que, para esses, a inferência de resultados ocorre sem o emprego de rótulos de saída (*targets*) associados a conjunto de dados (*data*) e que a clusterização é problema típico para emprego de modelos não supervisionados. Por isso, a escolha.

A estratégia inicial foi estabelecer agrupamentos contendo alta similaridade sintática. Por isso, a necessidade de escolha restritiva de parâmetros. 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. Além disso, não foi definido o quantitativo de *clusters* (K), conforme justificativas abaixo.

Após várias iterações na execução do módulo, foi escolhida função de clusterização hierárquica aglomerativa, a fim de agregar elementos ou grupos similares (*clusters*) e organizá-los de forma hierárquica, a partir de conjunto de parâmetros previamente definidos. (JAIN, 2019) estabelece que a aglomeração ocorre de baixo para cima (*botton-up*), ao contrário da clusterização por divisão, de cima para baixo (*top-down*).

Uma vantagem observada na ausência de parametrização de K foi a flexibilidade de resultados, conforme analisado por (BOEHMKE e GREENWELL, 2019). Mesmo assim, várias tentativas foram realizadas empregando algoritmo K-MEANS (MACQUEEN, 1968). Contudo a busca manual pelo valor adequado mostrou-se onerosa ao projeto. Foi descartada.

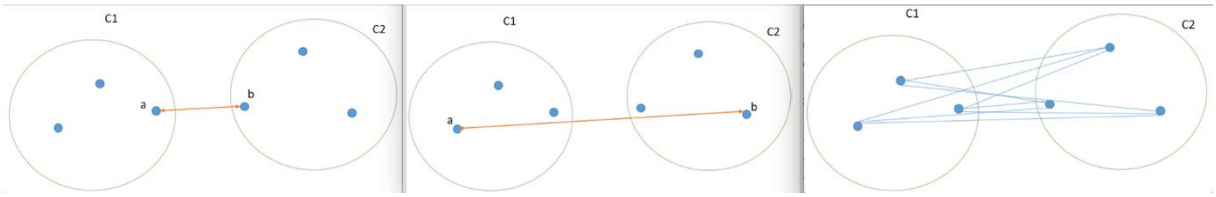
Para a implementação, a versão final utilizou função *fcluster*, disponível na biblioteca *SciPy* (THE SCIPY COMMUNITY, 2019). Como parâmetros da função de clusterização, foi escolhida a distância de *Jaccard*, associada ao método *single*, conforme justificativas a seguir.

A escolha da métrica associada à distância de *Jaccard* é justificada pela análise apresentada por (PANDIT e GUPTA, 2011), o qual recomenda o uso da métrica 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 definido pela razão entre a intersecção e a união de elementos (IRANI, 2016):

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}, J(A,B) = [0;1]$$

A escolha do método *single* teve como base aderência à necessidade de parâmetros mais restritivos, conforme apresentado por (ALTO, 2019) (figura abaixo). Para o método, a similaridade entre dois *clusters* C1 e C2 é dada pelo valor mínimo entre dois pontos mais próximos, tal que $d(C1, C2) = \min\left(dist\big(C1(i), C2(j)\big)\right), i \in C1, j \in C2$.

Figura 26: Parametrização de métodos *single*, *complete*, *average*



Fonte: (ALTO, 2019).

Durante a fase de exploração do modelo, foi observado significativo tempo de execução, como já descrito na seção anterior, o que foi ao encontro do apontado por (JAIN, 2019). Segundo o autor, apesar das vantagens da adoção do modelo, a complexidade em espaço e tempo tornam-se desafios, na ordem de $O(n^2)$ e $O(n^2 log(n))$, respectivamente, onde n é o número de elementos. Como solução de contorno, o dimensionamento de entrada foi minimizado, conforme descrito no módulo 2 (RADAR_FILTER).

Para a classe *CountVectorizer* foram definidos os seguintes parâmetros para a versão final do modelo: *ngram_range* = (2,5), *min_df* = 2, *max_df* = 0,1.

A classe *CountVectorizer* está disponível na popular biblioteca de aprendizagem de máquina *scikit-learn* (PEDREGOSA, 2011). A classe tem como função converter conjunto de dados textuais de entrada (*corpus*) em matriz contendo a contagem de elementos (*tokens*).

Dado o uso de parâmetros restritivos na verificação de similaridade, optou-se por não excluir palavras de parada, comumente consideradas irrelevantes para análises de similaridade textual. A biblioteca NLTK as denomina de *stopwords* (NLTK PROJECT, 2019).

Um dos parâmetros restritivos foi o emprego de *ngrams*. Em mineração textual refere-se ao sequenciamento contíguo de itens de palavras a partir de determinado texto. A escolha de

sequencias a partir de bigramas teve como objetivo a comparação de sequencias relevantes de vetores multidimensionais formados por, pelo menos, duas palavras. A limitação superior a cinco foi ajustada após algumas iterações (*tuning*), por ter apresentado resultados relevantes, mantendo custo aceitável em tempo de execução.

Os parâmetros *max_df* e *min_df* também foram ajustados de forma restritiva. O *max_df* é empregado para remover termos que aparecem com muita frequência. O valor 0.1 significa "ignorar termos que aparecem em mais de 10% dos registros". O valor padrão é 1.0, ou seja, ignorar termos que aparecem em mais de 100% dos registros. Isto é, nenhum valor é descartado.

O parâmetro *min_df*, por outro lado, é empregado para remover termos que aparecem com pouca frequência. O valor 2 significa "ignorar termos que aparecem em menos de 2 registros". O valor padrão é 1, ou seja, ignorar termos que aparecem em menos de 1 documento. Isto é, nenhum valor é descartado.

$$CountVectorizer\begin{cases} ngram\_range = (2,5) \\ min\_df = 2 \\ max\_df = 0.1 \end{cases}, fcluster\begin{cases} single \\ metric = jaccard \\ criterion = 'distance' \end{cases}$$

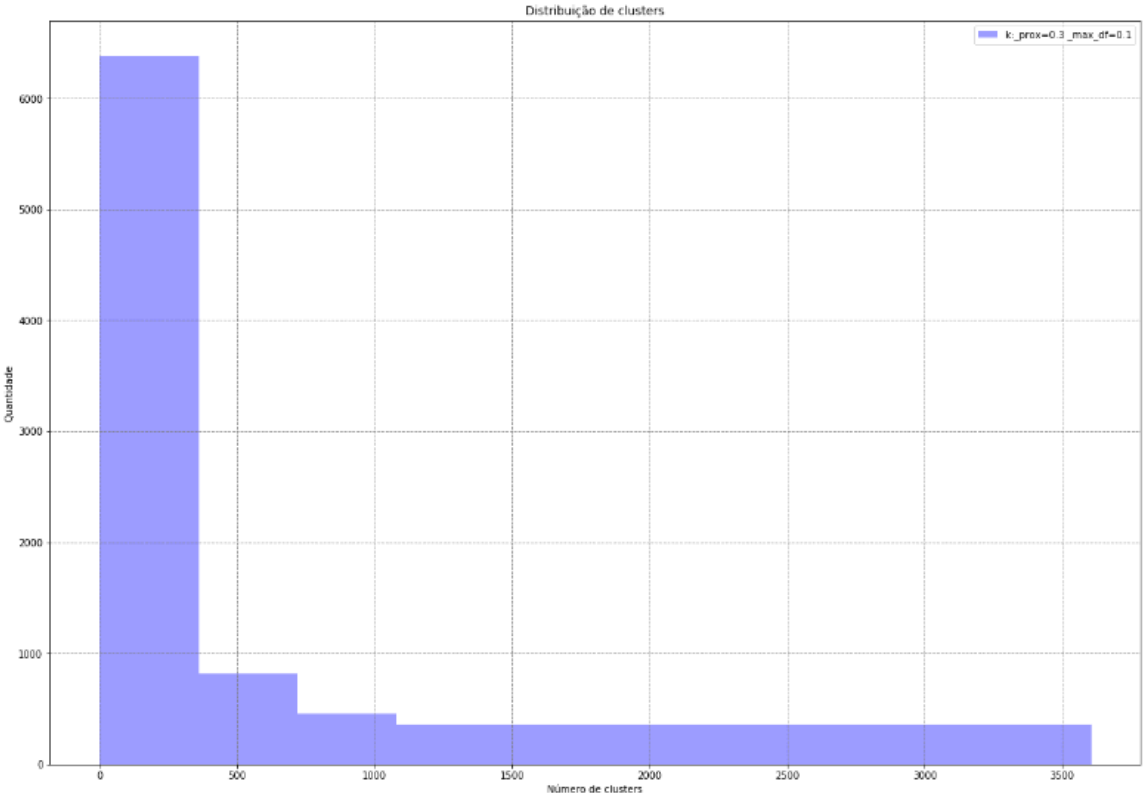
Foram obtidos os seguintes resultados, a partir de seleção dos parâmetros supracitados e do *corpus* de entrada obtido do módulo anterior, após filtragem preliminar de deliberações $inicio: 2019-10-17, 22:58:12, total: 20:29:23$ $(h:m:s)$.

Resultados: Quantidade de elementos: 10.176. Quantidade de *clusters*: 3.606. *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)]. Para mais detalhes, vide Figura 27 e Figura 28.

Observa-se que 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* ([0;99]) abrange praticamente metade do quantitativo de deliberações, como pode ser visto na Figura 27.

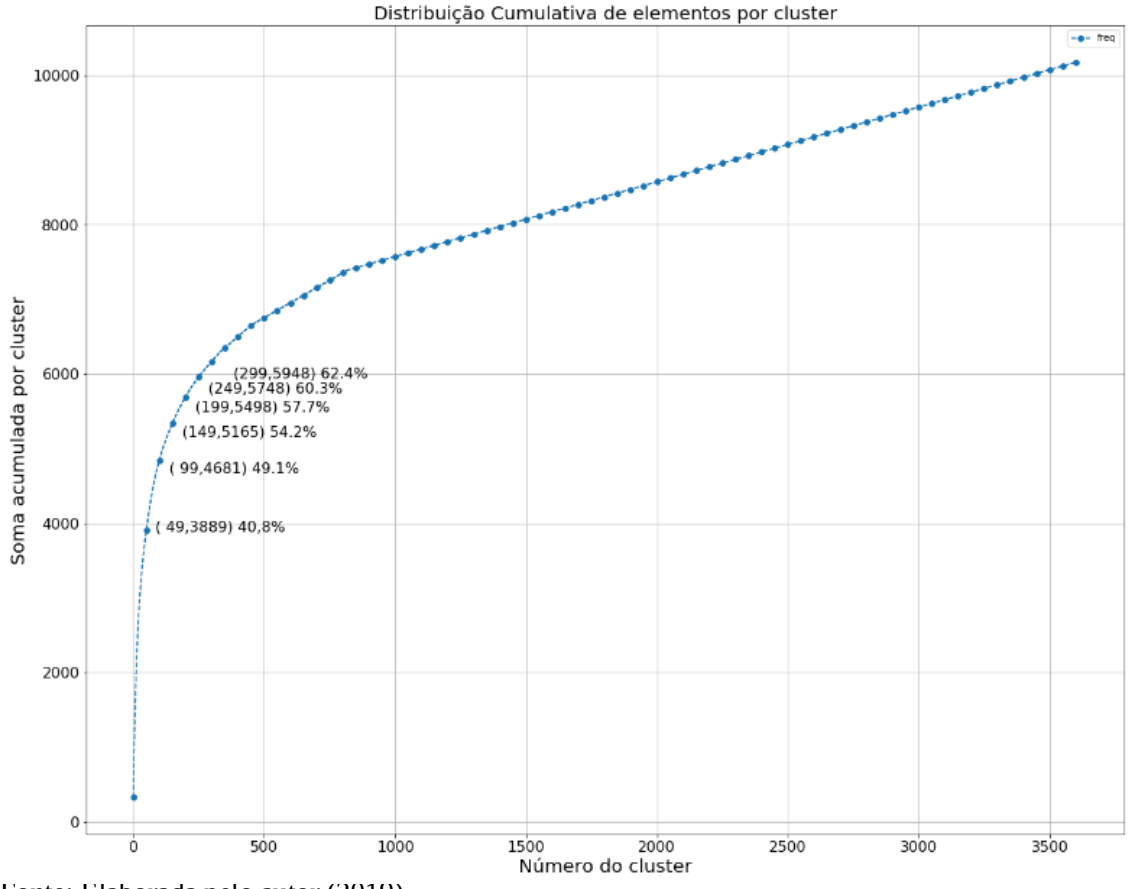
O resultado assimétrico pode ser explicado pela aparente uniformidade textual de alguns tipos específicos de deliberação. Por outro lado, boa parte dos *clusters* a partir de k=300 apresentam quantidade mínima de elementos, onde k é identificador do k-ésimo cluster. A referência a órgãos, entidades e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

Figura 27: Distribuição de *clusters*



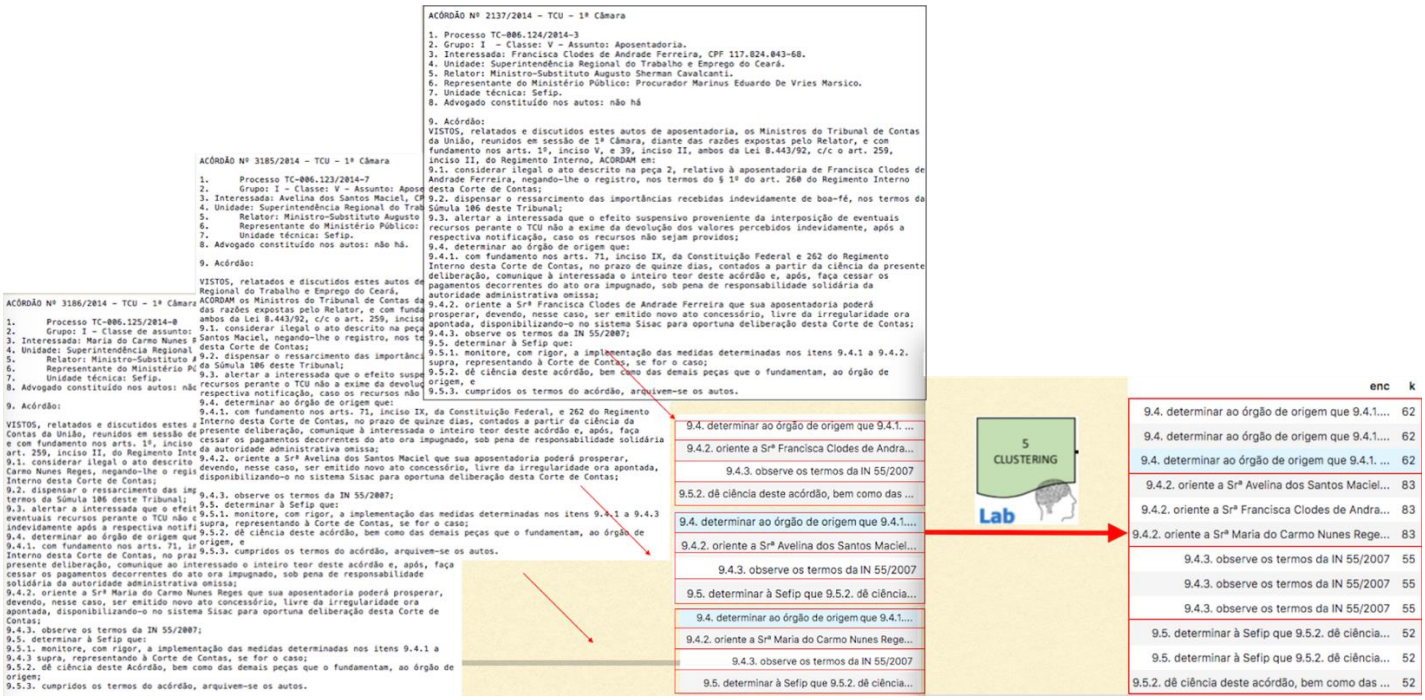
Fonte: Elaborada pelo autor (2019).

Figura 28: Distribuição cumulativa por *cluster*



Fonte: Elaborada pelo autor (2019).

Figura 29: Amostragem de deliberações após clusterização



Fonte: Elaborada pelo autor (2019).

A Figura 28 apresenta distribuição cumulativa por *cluster*: k=[0;49], 3.876, 38,1%; k=[50;99], 4.834, 47,5 %; k=[100;149], 5.332, 52,4 %; k=[150;199], 5.687, 55,9%; k=[200;249], 5.957, 58,5 %; k=[250;299], 6.164, 60,6%. Cabe ressaltar que a rotulação de *clusters* foi ordenada de forma crescente em relação ao número de elementos para facilitar a análise.

A Figura 29 apresenta amostragem de deliberações resultantes do algoritmo de clusterização. Foram selecionados os Acórdãos n° 3.186/2014-1, 3.185/2014-1 e 2137/2014-1. As deliberações 9.4, 9.4.2, 9.4.3 e 9.5.2; 9.4, 9.4.2, 9.4.3 e 9.5; e 9.4, 9.4.2, 9.4.3 e 9.5, respectivamente, foram cadastradas para monitoramento. Após a execução do referido módulo, aos primeiros registros de cada acórdão foram associados o 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.*

A Tabela 5 apresenta amostragem de deliberações agrupadas nos três maiores *clusters* [(0, 332), (1, 243), (2, 203)]. A capitulação minúscula decorre do tratamento e limpeza de dados. Para fins didáticos, as diferenças textuais entre as deliberações entre cada *cluster* foram assinaladas em cor **vermelha**.

Tabela 5: Amostragem de deliberações nos top 3 maiores *clusters*

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

Fonte: Elaborada pelo autor (2019).

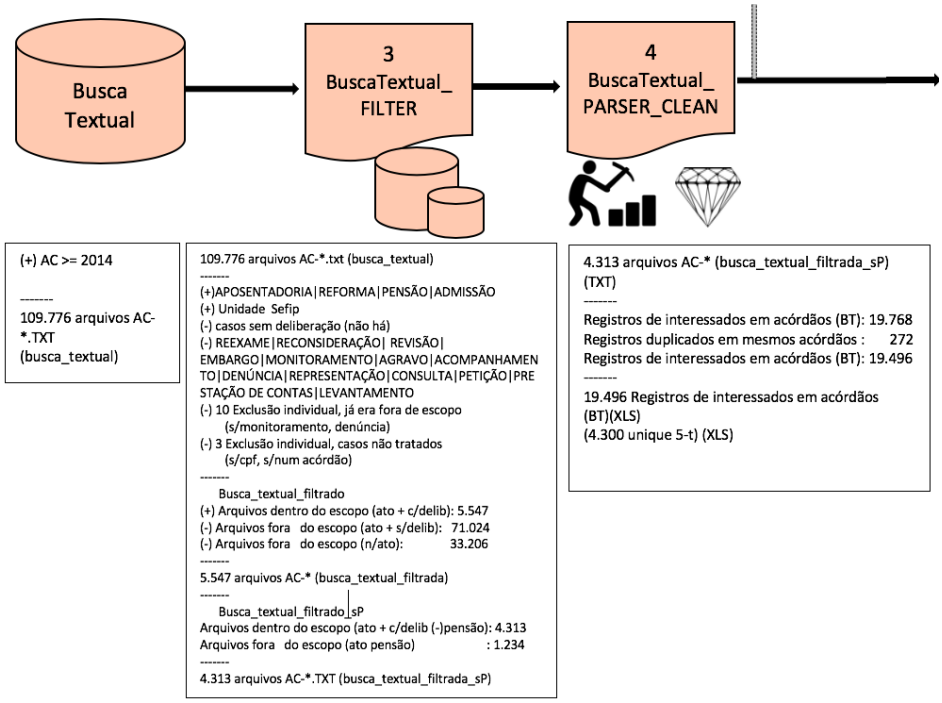
8.3. Raia VERMELHA (BUSCA TEXTUAL)

A raia VERMELHA é formada por módulos responsáveis pela coleta, limpeza, filtragem e *parser* de dados provenientes da busca textual de Acórdãos.

Conforme descrito no item anterior, a obtenção de dados via *parser* de arquivos textuais foi considerada inicialmente fonte alternativa para obtenção de dados, em função do pouco tempo de projeto. Contudo, a implementação mostrou-se viável. Por outro lado, houve oportunidade de reduzir o domínio de dados da raia VERDE em função de dados obtidos no *parser*, a exemplo do tipo do ato e do tipo de acórdão.

O **módulo 3** – BT_FILTER – procedeu a filtragem de 109.776 arquivos textuais de entrada contendo acórdãos, podendo incluir, no mesmo arquivo, conteúdo dos respectivos Relatório e Voto. Cabe repisar que a coleta foi procedida a partir de decisões datadas a partir de 2014.

Figura 30: Raia VERMELHA (busca textual)



Fonte: Elaborada pelo autor (2019).

A filtragem tem como finalidade diminuir a complexidade do domínio de entrada, mantendo apenas os arquivos de interesse, com vistas a desonerar o módulo seguinte, responsável pela limpeza e *parser* de dados. Sendo assim, a saída do módulo é formando diretório de saída, contendo acórdãos de interesse.

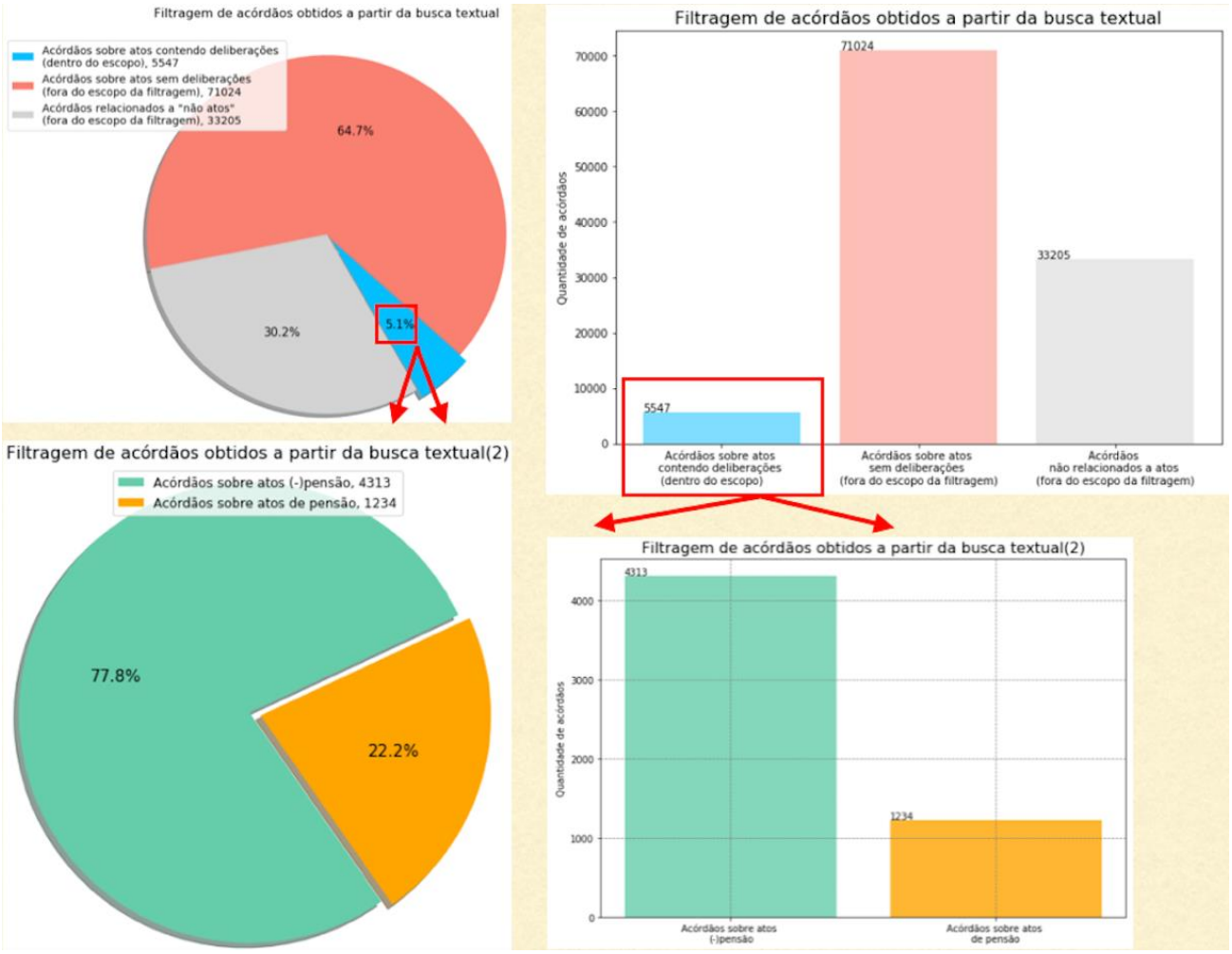
Foram considerados de interesse, as decisões contendo a Sefip como unidade técnica, relacionados a atos de pessoal. Foram excluídas situações de trivialidade, como acórdãos sem deliberação para monitoramento, situação prevalente em apreciação por legalidade de ato de admissão. acórdãos versando sobre "não-atos" também foram excluídos por estarem fora de escopo do trabalho, bem como recursos, levantamento, acompanhamento. Também foram excluídas manualmente algumas situações anômalas, a exemplo de acórdãos sem registro de CPF no rol de interessados, inviabilizando análise no Siape, a ser detalhando em seção própria.

Como vistas a ajustar o escopo do trabalho ao tempo disponível para o projeto, houve também necessidade de nova iteração no ciclo do CRISP-DM, com vistas de reduzir o escopo de atos de pessoal. Sendo assim, atos relacionados à concessão de pensão (civil, militar e ex-combatente) poderão ser analisados oportunamente como trabalhos futuros.

Como saída do módulo, foram registrados 4.313 arquivos identificados como sendo associados a atos de pessoal (admissão e concessão de aposentadoria e reforma). Foram excluídos 71.024 arquivos, considerados casos de trivialidade, ou seja, atos de pessoal sem deliberação, e

33.205 arquivos associados a casos fora de escopo, como “não-atos”, Recursos e casos anômalos. Finalmente, foram excluídos 1.234 arquivos identificados como sendo associados a apreciação de atos de concessão de pensão por ajuste ao tempo do projeto (Figura 31).

Figura 31: Resultados - módulo 3 - BT_FILTER



Fonte: Elaborada pelo autor (2019).

O **módulo 4** – BT_PARSER – é responsável pelo *parser* e limpeza de dados. Trata-se do módulo principal da raia, concentrando a inteligência para localização, coleta e atribuição semântica de segmentos de texto de interesse. Em analogia coloquial, os segmentos podem ser comparados a peças de quebra-cabeça dispersos em um tabuleiro. O papel do módulo é de localizá-los e dar sentido a eles (Figura 32).

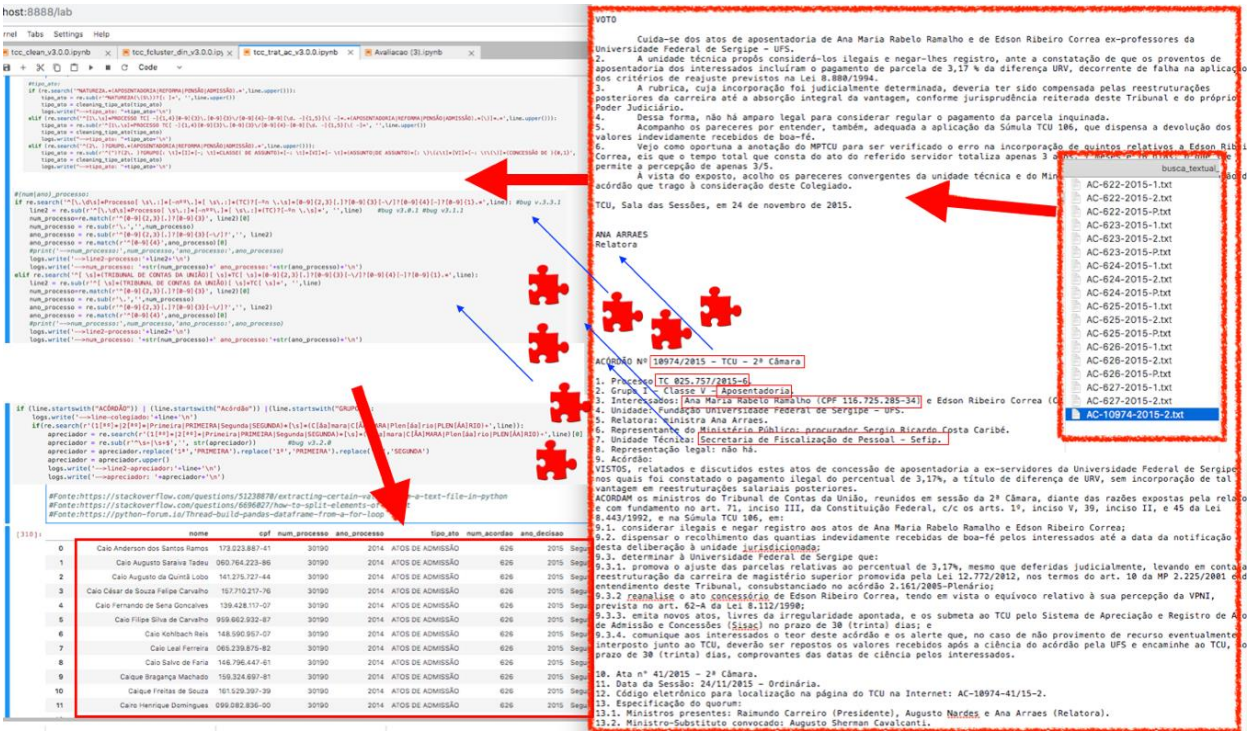
O módulo foi desenvolvido inicialmente com o objetivo principal de identificar os interessados em cada acórdão. Os dados disponibilizados do RADAR careciam de tal informação. Tendo em vista assegurar o cumprimento do cronograma do projeto, decidiu-se por

estabelecer linha de corte temporal na obtenção de dados por meio de sistemas e estabelecer caminhos alternativos com base em dados já disponíveis, com o uso da busca textual de acórdãos.

Sendo assim, os registros foram obtidos da busca textual contendo o interessado como referência principal. Ou seja, cada registro é referenciado por um interessado.

Foi realizado o tratamento (*parser*) dos seguintes campos: nome e CPF de interessados, tipo de ato, número e ano do acórdão, apreciador, número e ano do Processo.

Figura 32: *Parser* - visão geral



Fonte: Elaborada pelo autor (2019).

Uma das limitações encontradas nas tabelas disponibilizadas do RADAR foi justamente relacionada à obtenção dos interessados. Algumas tentativas foram preliminarmente realizadas para extrair textos de deliberações. Contudo, a menção de interessados não pôde ser tratada como regra. Pelo contrário, possivelmente motivado pelo reuso de textos com estrutura sintática similar em diferentes decisões. Por isso, a opção de extrair dados diretamente da busca textual.

Da parte da extração de dados da busca textual, a aparente uniformidade sintática, na prática, mostrou-se grande desafio. Alguns exemplos obtidos pela saída do módulo

Primeiro: O tipo de ato necessitou ser localizado em três possíveis posições do acórdão. Embora tenha sido demandado maior quantidade de iterações no modelo, foi possível realizar a

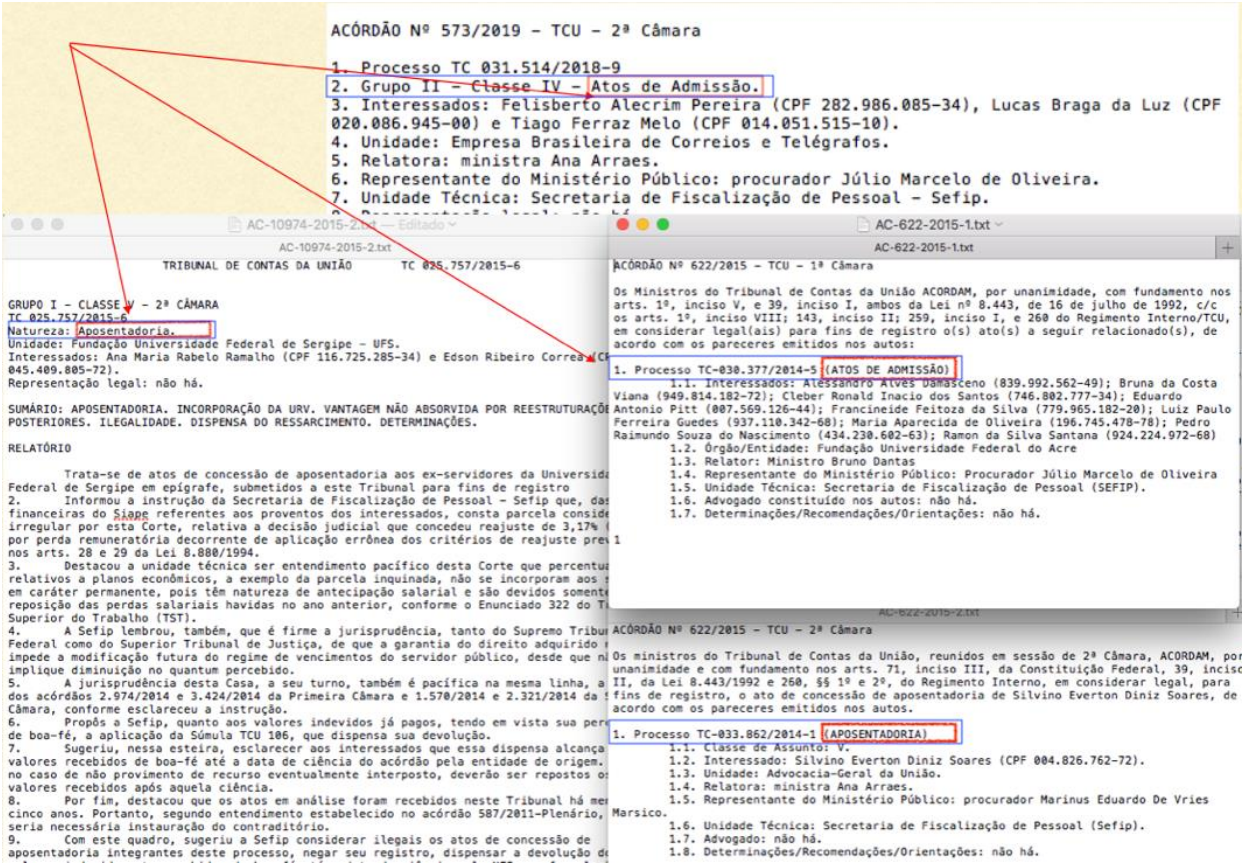
extração das diferentes disposições. A figura abaixo contém inventário de principais formatos encontrados, juntamente com amostragem extraídas de acórdãos:

Tabela 6: Expressão regular

Linha	Expressão Regular	qtd
Natureza	*egrep -riH "^NATUREZA.*(APOSENTADORIA|REFORMA|ADMISSÃO).*" . | wc -l*	2.411
Grupo	*egrep -riH "^(2\.)?GRUPO.*(APOSENTADORIA|REFORMA|ADMISSÃO).*" . | wc -l*	2.279
Processo	*egrep -riH "^[1\.\s]*PROCESSO TC[-]{1,4}[0-9]{3}\.[0-9]{3}\/[0-9]{4}-[0-9][\d. -]{1,5}[\(-]*.*(APOSENTADORIA|REFORMA|ADMISSÃO).*[\)]*.*" . | wc -l*	4

Fonte: Elaborada pelo autor (2019).

Figura 33: *Parser* de tipos de ato



Fonte: Elaborada pelo autor (2019).

Segundo: A listagem de CPFs na relação de interessados seguiu diferentes disposições. Embora tenha sido demandado maior quantidade de iterações no modelo, foi possível realizar a extração das diferentes disposições. A figura abaixo contém inventário de principais formatos encontrados, juntamente com amostragem extraídas de acórdãos.

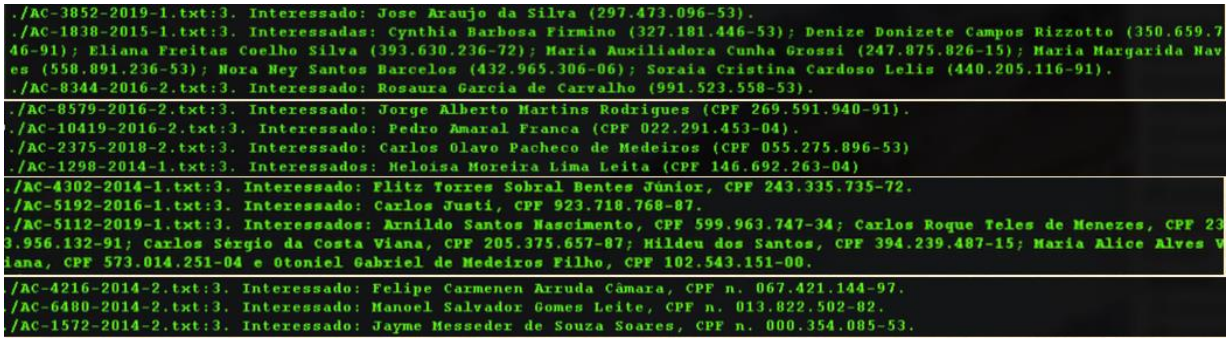
Tabela 7: Expressão regular

Principais formatos	Expressão Regular	qtd
(000.000.000-00)	*egrep -rH "^[- \s\.0-9]+Interessad(o|a)[s]{0,1}:.*\([0-9]{3}\.[0-9]{3}\.[0-9]{3}-[0-9]{2}\).*" . | wc -l*	1.412
(CPF 000.000.000-00)	*egrep -rH "^[- \s\.0-9]+Interessad(o|a)[s]{0,1}:.*\(CPF.*[0-9]{3}\.[0-9]{3}\.[0-9]{3}-[0-9]{2}\).*" . | wc -l*	899
CPF 000.000.000-00	*egrep -rH "^[- \s\.0-9]+Interessad(o|a)[s]{0,1}:.*CPF [0-9].*" . | wc -l*	933
000.000.000-00	*egrep -rH "^[- \s\.0-9]+Interessad(o|a)[s]{0,1}:.*[^\(][0-9]{3}\.[0-9]{3}\.[0-9]{3}-[0-9]{2}[^\)].*" . | wc -l*	44

CPF n 000.000.000-00	*egrep -rH "^[- \s\.0-9]+Interessad(o|a)[s]{0,1}:.*CPF (n|N).*" . | wc -l*	9
(CPF n 000.000.000-00)	*egrep -rH "^[- \s\.0-9]+Interessad(o|a)[s]{0,1}:.*\(CPF (n|N).*" . | wc -l*	0

Fonte: Elaborada pelo autor (2019).

Figura 34: *Parser* de CPFs

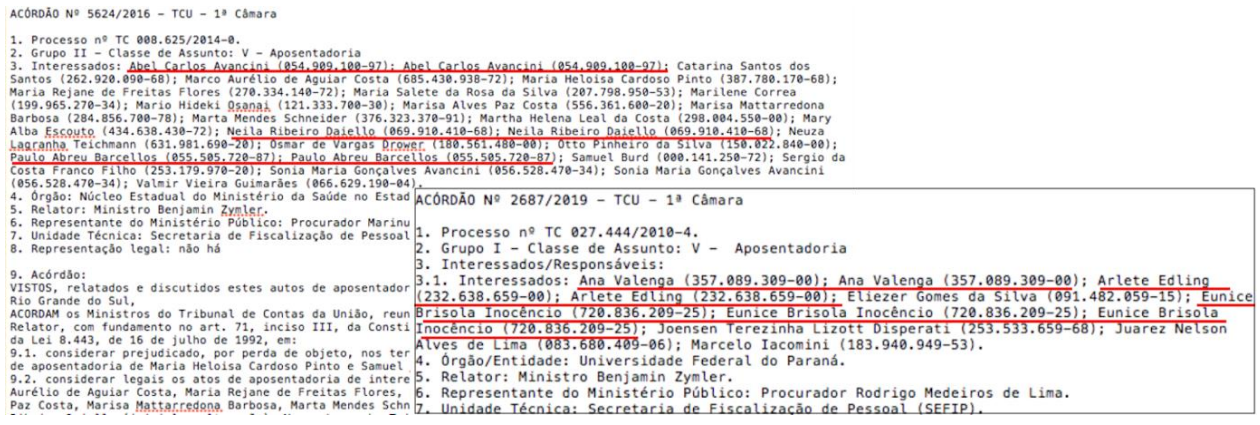


Fonte: Elaborada pelo autor (2019).

Terceiro: A identificação de duplicidades na listagem de interessados em um mesmo acórdão. Cabe frisar que a raia VERMELHA foi desenhada para levantar interessados em deliberações. Por isso, os registros foram orientados a discriminar um interessado por registro.

Durante as iterações realizadas na execução do módulo, foram identificadas duplicidades de interessados por acórdão, totalizando 272 (duzentos e setenta e dois) casos, a exemplo da amostragem abaixo contendo excerto dos Acórdãos 5.624/2016-1 e 2.687/2019-1. O esforço na implementação de condições de contorno demandou poucas iterações no módulo. As duplicidades foram eliminadas.

Figura 35: Amostragem de acórdãos contendo duplicidade de interessados



Fonte: Elaborada pelo autor (2019).

Cabe ressaltar que, em apresentação realizada na unidade técnica interessada, foi esclarecido que a duplicidade pode ter sido ocasionada por lançamento automático de sistema no caso de envolvimento de dois atos de um mesmo interessado.

Por fim, o rol de situações anômalas apresentados não é exaustivo. Outras situações decorrentes de análise exploratória foram enumeradas no ANEXO B.

Como resultado, foram extraídos e tratados os dados de 19.496 registros de interessados em acórdãos, já considerando a exclusão de 272 duplicidades acima identificadas, totalizando 19.768 registros preliminares.

Outro resultado produzido pelo módulo está relacionado à geração de tuplas 5-T (número e ano do acórdão, apreciador e número e ano de Processo). Conforme já mencionado na raia VERDE, verificou-se a necessidade de reduzir o quantitativo de entradas para a clusterização, selecionando apenas os registros de interesse selecionados na referida raia. Ao contrário do inicialmente previsto, o *parser* propiciou a possibilidade de extrair elementos capazes de direcionar o domínio de documentos de interesse. Então, a intenção foi aproveitar os resultados obtidos para viabilizar o tempo de execução do algoritmo de clusterização.

Uma possibilidade de trabalho futuro é a extensão da capacidade de *parser* de dados, particularmente para extração de deliberações. Sabe-se que nem todas as deliberações são cadastradas no RADAR. Se houver a devida implementação dos mesmos critérios de registro de itens/subitens, particularmente os passíveis de monitoramento, pode-se aventar a dispensa da referida base de dados.

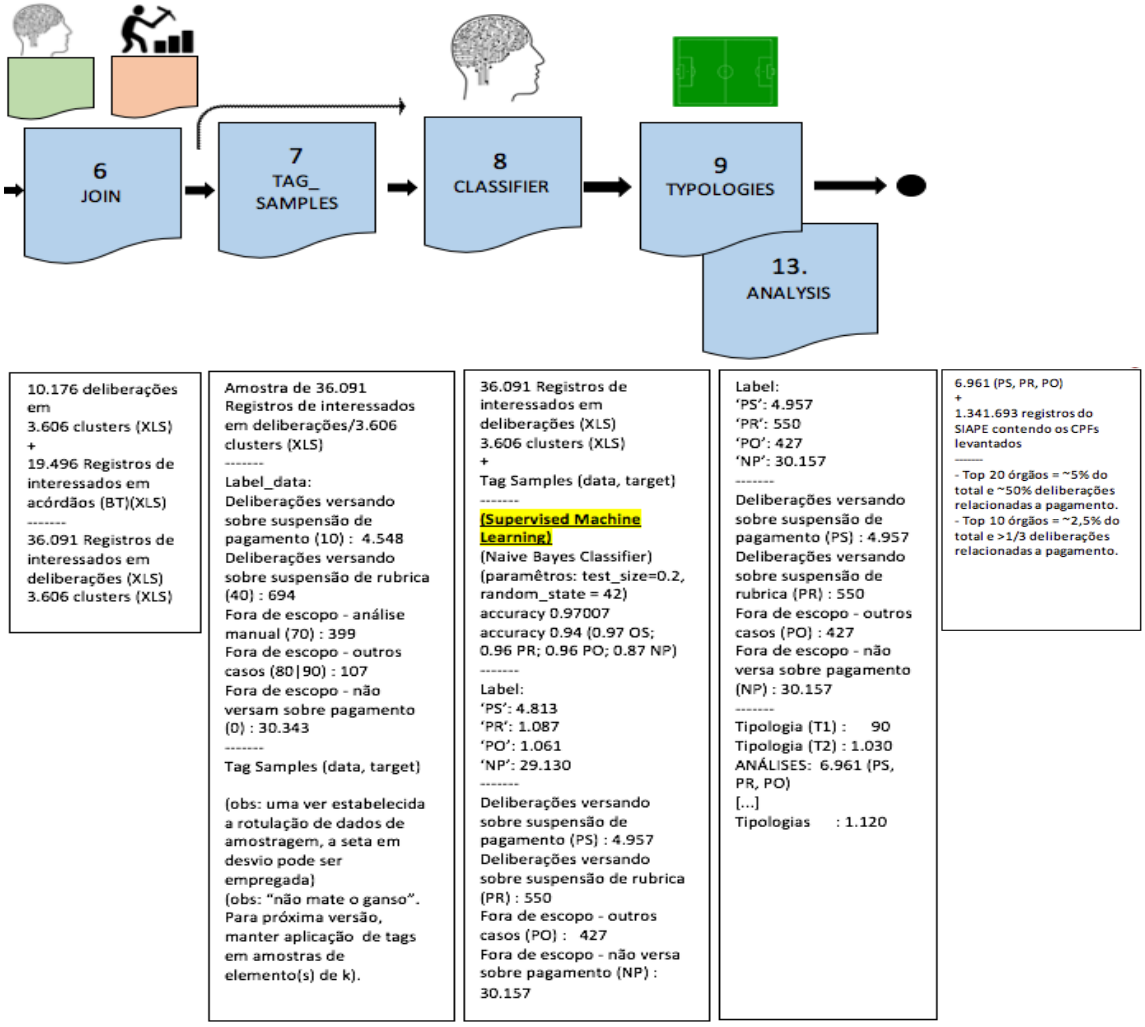
Repisa-se novamente que as tuplas 5-T não representam chave primárias, dado que um mesmo acórdão pode conter mais de uma deliberação. Apesar de um acórdão corresponder a um processo (*1:n*), o critério de lançamento de registros no RADAR é por subitem/item de deliberação, transformando a relação em cardinalidade múltipla (*m:n*). A intenção é simplesmente garantir corretude e unicidade de caminho na associação reversa entre subitens e itens de deliberação e o respectivo documento que o originou.

Então, o segundo resultado foi a obtenção de 4.300 tuplas 5-T (únicas), possibilitando a redução de registros de entrada da raia VERDE (módulo 5) de 22.038 registros para 10.176.

8.4. Raia AZUL

A raia AZUL é responsável por fazer a junção dos resultados obtidos na primeira e segunda raias, por realizar a classificação por meio de algoritmo supervisionado, por rotular deliberações versando sobre cessação de pagamento. Em seguida, prepara os dados para distribuição para as diferentes tipologias propostas.

Figura 36: Raia AZUL



Fonte: Elaborada pelo autor (2019).

O **módulo 6** – JOIN – realiza a junção dos resultados obtidos nas raias anteriores. A primeira parte contém registros organizados por deliberação, conforme critério de armazenamento do RADAR. A segunda foi organizada por interessados, a partir de *parser* da busca textual de Acórdãos, com o objetivo de complementar as informações da raia anterior.

O critério de junção foi baseado nas combinações de tuplas 5-T, descrito nas raias anteriores, a fim de associar deliberações e interessados. O objetivo é assegurar corretude e unicidade na operação algébrica relacional. Para facilitar entendimento, vide Figura 37.

Como resultado, a entrada formada por 10.176 registros de deliberações e 19.496 de interessados em acórdãos, foram gerados 36.091 de registros de interessados em deliberações, dispostos em 3.606 *clusters*. O valor resultante não é múltiplo das entradas devido à relação *m:n* de acórdãos e processos, descrito na raia VERDE.

Figura 37: Emulando operação algébrica relacional de junção (tuplas 5-T)

Fonte: Elaborada pelo autor (2019).

O **<u>módulo 7</u>** – TAG_SAMPLES – é responsável pela preparação do modelo de classificação, com a rotulação de amostragem de dados de entrada (*tag samples*) para o treinamento do classificador, implemenado no módulo seguinte.

O objetivo é preparar amostra de registros obtidos na massa de dados de entrada e realizar a devida rotulação. O critério de rotulação precisa estar alinhado com o alvo a ser almejado pelo classificador. Ou seja, classificar deliberações versando sobre cessação de pagamento, a fim de servir de insumo para a definição de tipologias capazes de realizar o cruzamento com o SIAPE.

A amostragem de dados e a "correta" rotulação são insumos para o treinamento de algoritmos de aprendizagem de máquina supervisionados (SONI, 2019). O autor esclarece ainda que o termo "correto" não significa que todas as correspondências da amostra sejam verdadeiras. Contudo, o grau de incorretude pode afetar a acurácia do modelo.

Na versão final, foram definidos os seguintes rótulos. NP e P* (união dos subtipos "PR", "PS" e "PO", como acrônimos de "não pagamento" e "pagamento", respectivamente).

O rótulo "NP" refere-se a deliberações não associadas à cessação de pagamento. Cita-se como exemplo textos associados à determinação ao órgão para submeter novos atos – deliberação bastante comum em apreciação de atos, para comunicar interessados acerca do inteiro teor da decisão e para encaminhar ao Tribunal comprovação de que os interessados tomaram conhecimento das decisões.

O rótulo "P*", por sua vez, refere-se a casos de cessação de pagamento. Trata-se de rótulo alinhado como o objetivo de levantamento de possíveis tipologias para verificação no SIAPE. Há três subdivisões. As duas primeiras subdivisões são "PS" e "PR" para casos de suspensão total e parcial de vencimentos, isto é, "pagamento suspensão" e "pagamento rubrica",

respectivamente. A terceira é a "PO" para outros casos, ou seja, para textos versando sobre pagamento, mas que, por alguma razão não podem ser rotulados como passíveis de verificação por tipologias. Cita-se como exemplo deliberações versando sobre a possibilidade de cessação de pagamentos a depender de resultado de decisão judicial. A automatização pode ser considerada onerosa ou inviável, demandando tratamento manual. Então, na prática, os rótulos de interesse são as duas primeiras subdivisões de "P*".

Diversas iterações do modelo foram necessárias para definir a versão final dos rótulos. Parte do desafio decorreu da necessidade de entender os textos produzidos em deliberações de pessoal. Vivência pretérita à unidade poderia suprir em parte a necessidade.

Por outro lado, a necessidade de melhor entender os textos estruturados em deliberações de pessoal acabou ampliando o espaço amostral de rotulação. A intenção inicial era de estratificar amostras a partir de diferentes critérios, como a seleção de subconjuntos a partir de *clusters* e de temporalidade (anual). O primeiro critério teria como objetivo minimizar a seleção de textos similares, já que o papel da clusterização era de justamente agrupar textos afins. O segundo, para explorar o viés de mudança de estilos de escrita com o passar do tempo.

Contudo, o manuseio textual empregando expressões regulares, somado à avidez do autor por correto entendimento acerca do assunto, acabou conduzindo à varredura de todo o domínio de dados. Em tese, a amostra de dados rotulados serviria de insumo para treinamento do algoritmo supervisionado de classificação, a fim de produzir modelo de maior acurácia. Na prática, porém, dado o tamanho da amostragem, o efeito foi ligeiramente reverso.

Cumpre ressaltar que o excesso de precaução não acarretou prejuízos ao trabalho. Pelo contrário, proporcionou melhor entendimento dos dados. Além disso, preservou-se alinhamento às boas práticas, com a adoção de modelo de rotulação seguido de classificação. Julga-se cabível apenas a ressalva de ajustar o tamanho da amostra para futura execuçã do módulo, uma vez que o entendimento dos dados já foi etapa superada.

A seta em desvio na Figura 36 indica que a execução do modelo pode dispensar a rotulação de novas amostras, particularmente se as deliberações supervenientes mantiverem, de forma geral, padrão similar de escrita. Com o passar do tempo e com a natural rotação de integrantes da Unidade Técnica de Pessoal e de Gabinetes, o estilo textual pode sofrer mudanças e ensejar nova rotulação de amostras.

O **Módulo 8** – CLASSIFIER – realiza a classificação por meio de algoritmo supervisionado, em particular sobre textos de deliberação acerca de cessação de pagamento.

ZAKI e MEIRA JR. (2014) explicam que a classificação é modelo (M) capaz de realizar a predição de classes $\hat{y} \in \{c_1, c_2, \ldots, c_i \ldots, c_k\}$, onde c_i é a i-ésima classe categórica, de conjunto de dados de entrada x, resultando em $\hat{y} = M(x)$. Para a construção de M, é necessário conjunto de treinamento, contendo elementos previamente rotulados (*training set*). O conjunto foi definido no módulo anterior. Após a aprendizagem do modelo, a predição de classes pode ser aplicada para o conjunto de entrada, com o objetivo de maximizar a probabilidade posterior de acerto (P) atrubuindo à entrada x, a classe $c_i, onde\ P(c_i\ |\text{x})$.

Os referidos autores esclarecem ainda que o algoritmo *Naive Bayes* pode ser empregado para solucionar problemas de classificação multiclasse. Ele é baseado na teoria de *Bayes*, que estabelece independência entre os preditores. Assume-se, então, que a presença de uma característica particular em uma classe não está relacionada com a de qualquer outro recurso. Todas as propriedades contribuem de forma independente para a probabilidade. Por isso, o uso do termo *naive* (ingênuo). BELLHOUSE (2004) apresenta uma biografia sobre Thomas Bayes.

Pelo Teorema de *Bayes*, a probabilidade posterior $P(c_i|x)$, ou seja, a probabilidade de hipótese c_i dada a observação de x, é dada por $P(c_i|x) = \frac{P(x|c_i)P(c_i)}{P(x)}$, onde $P(x|c_i)$ é a probabilidade de obtenção de x dado que a classe c_i é verdadeira; $P(c_i)$ é a probilidade original da classe (ou probabilidade anterior); $P(x)$ é a probabilidade original do preditor, independente da hipótese (c_i).

Neste módulo optou-se por empregar o algoritmo *Naive Bayes*. O emprego é justificado pelas considerações apresentadas por ZAKI e MEIRA JR. (2014), acerca das vantagens obtidas em função da efetividade de resultados em trabalhos acadêmicos e da simplicidade do modelo frente a outros de mesma natureza, mesmo considerando que a premissa de independência é dificilmente mantida no mundo real.

Os dados de entrada deste módulo são provenientes das colunas de encaminhamento (ENC) e de rotulação (LABEL) totalizando 36.091 registros de saída do módulo anterior (7 – TAG_SAMPLES).

O balanceamento de classes foi necessário para melhoria do modelo. Na fase de *tag_samples* foi aventada a hipótese de relevante desproporção entre classes. Mesmo com a definição de categorias entre versões distintas do modelo, a proporção de textos versando sobre não pagamento e pagamento era notadamente desbalanceada em favor da primeira. A razão foi simples de entender e está diretamente relacionada à sistemática de elaboração das próprias deliberações. A presença de decisões versando sobre cessação de pagamento normalmente é

acompanhada de vários outros dispositivos, entre eles, acerca de comunicar ao órgão e ao interessado sobre a decisão exarada, de assinar prazo para o cumprimento de ações e de emitir novos atos. Contudo, as mesmas deliberações podem ocorrer com a ausência de dispositivos com efeito suspensivo, mesmo em caso de ilegalidade, conforme o caso concreto.

Sendo assim, foi empregada a técnica de *oversampling* (DERNONCOURT, 2018) (DATAMAN, 2018a), cuja descrição de resultados será feita na fase seguinte, relacionada à validação do modelo, juntamente com critérios de validação do classificador.

Outro parâmetro adotado na versão final do modelo foi a eliminação de palavras de parada (*stopwords*) (NLTK PROJECT, 2019), ao contrário da estratégia de parametrização adotada no algoritmo de clustering (raia VERDE). Após a eliminação de *stopwords*, o somatório de palavras de todos os textos de entrada caiu de 8.810.383 para 5.236.993. Para facilitar o entendimento, uma amostra textual encontra-se na Figura 38.

Figura 38: Parametrização com NLTK - remoção de *stopwords*

```
['dar ciência desta decisão interessadas superintendência estadual funasa estado ceará',
 '9.4. dar ciência desta decisão às interessadas e à Superintendência Estadual da Funasa no Estado do Ceará',
 'NP'],
```

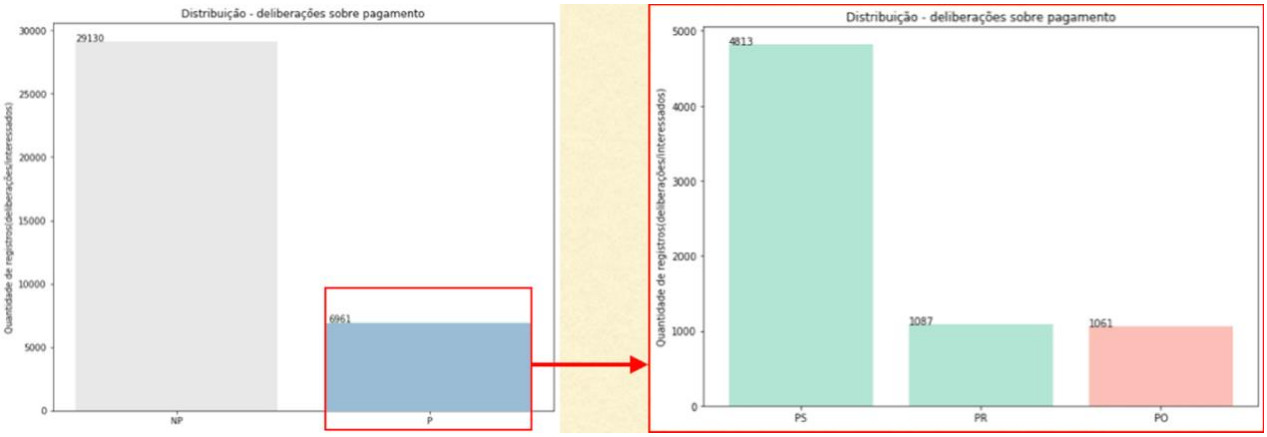
Fonte: Elaborada pelo autor (2019).

Para a execução do módulo, os dados pré-processados foram divididos em conjuntos de treinamento e teste na proporção de 80:20. A validação do modelo, considerando os dados de teste, obteve acurácia de 93,89%. A precisão de resultados por categorias de interesse (P*), variou entre 96% e 97%. Maior detalhamento pode ser observado na seção 9.1.

Os 36.091 registros de entrada foram reclassificados utilizando o modelo treinado. Os seguintes resultados foram obtidos: Deliberações que não versam sobre pagamento (“NP”) : 29.130. Deliberações sobre pagamento (“P”), 6.961, sendo deliberações versando sobre suspensão de pagamento (PS), 4.813; deliberações versando sobre suspensão de rubrica (PR), 1.087 e deliberações sobre pagamento, mas envolvendo outros casos, os quais demandam tratamento manual (PO), como decisões dependentes de ações judiciais, 1.061. Detalhes na Figura 39 e na Figura 40.

Para facilitar a compreensão do módulo, a Figura 41 representa continuação da Figura 29, contendo a mesma amostragem de Acórdãos 3.186/2014-1, 3.185/2014-1 e 2137/2014-1 e respectivas deliberações, descrita na clusterização (raia VERDE). O passo relativo ao classificador produziu como resutados os rótulos (campo *label*) “PS” para os elementos de k=62 e “NP” para os elementos de k=[52,55,83].

Figura 39: Resultados do classificador *Naive Bayes*



Fonte: Elaborada pelo autor (2019).

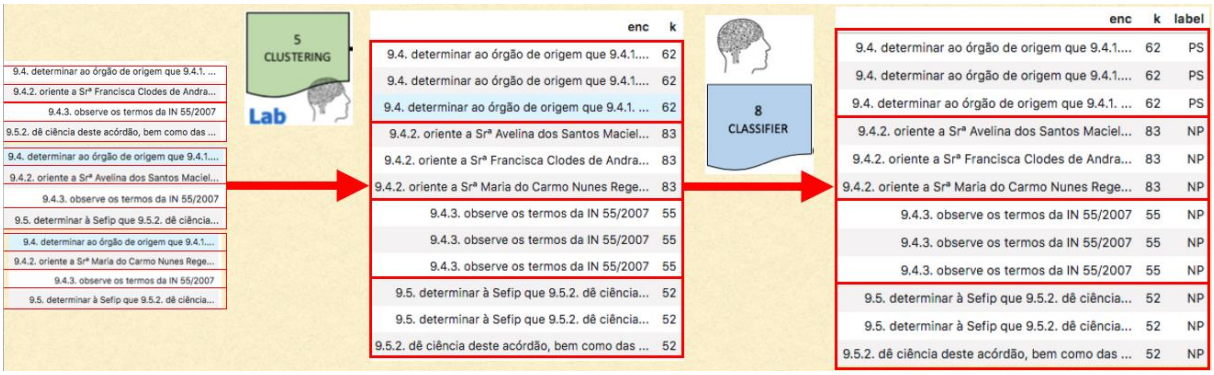
Figura 40: Resultados do classificador Naive Bayes (2)

```
------------------------------------------------------------
Deliberações versando sobre suspensão de pagamento ou rubrica:  5900
Deliberações versando sobre suspensão de pagamento      (PS) :  4813
Deliberações versando sobre suspensão de rubrica        (PR) :  1087
Fora de escopo - outros casos                           (PO) :  1061
Fora de escopo - não versa sobre pagamento              (NP) : 29130
------------------------------------------------------------
Deliberações - total                                         : 36091

------------------------------------------------------------
Clusters versando sobre suspensão de pagamento               :  550
Clusters versando sobre suspensão de rubrica                 :  247
Clusters versando sobre suspensão pagamento ou rubrica       :  794
Fora de escopo                                               : 2814
------------------------------------------------------------
Clusters                                                     : 3606
```

Fonte: Elaborada pelo autor (2019).

Figura 41: Amostragem de deliberações após classificação



Fonte: Elaborada pelo autor (2019).

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.

Como trabalhos futuros pode ser considerada a comparação com outros algoritmos de aprendizagem supervisionada destinados a resolver problemas de classificação, a exemplo de

kNN (*K Nearest Classifier*) e árvore de decisão (ZAKI e MEIRA JR., 2014), porquanto a comparação não pôde ser realizada em função do dimensionamento do cronograma do projeto.

O **Módulo 9** – TYPOLOGIES – realiza a preparação de dados para distribuição para as tipologias propostas no presente trabalho, segundo critérios de seleção previamente definidos. Na arquitetura apresentada no início da seção (Figura 23), o mnemônico associado ao módulo foi o de campo de futebol, particularmente ao meio-campo, local em que são organizadas as jogadas. Por isso, optou-se pelo emprego para facilitar o entendimento.

A definição de tipologias manteve alinhamento ao objetivo na análise de dados, que é levantar casos de cessação de pagamento, para servir de insumo à realização do cruzamento com a base de dados alvo, no caso, SIAPE.

Cabe registro preliminar de dificuldade encontrada durante iterações envolvendo o presente módulo. Para o exercício laboral na produção de textos, o reuso de conteúdo é considerado boa prática. E a disposição genérica de termos pode facilitar emprego em trabalhos supervenientes de mesma natureza.

Contudo, para a mineração de dados torna-se grande desafio. O entendimento de subitens/itens de deliberação passa a depender da análise de todo o acórdão. Conforme o caso, pode depender também da interpretação do conteúdo textual do Relatório ou, até mesmo, do Voto. Para facilitar o entendimento, cita-se como exemplo rol de excertos de diferentes deliberações versando sobre pagamento (Figura 42).

Além disso, a disposição de conteúdo textual genérico também tornou-se desafio para elencar a decisão aos correspondentes afetados no acórdão, mencionados, neste caso, apenas no rol de interessados, campo normalmente localizado no cabeçalho do documento. Em decisões envolvendo atos de pessoal, a enumeração de lista de interessados em um mesmo acórdão permite ganho de escala na tomada de decisão da Egrégia Corte. Por outro lado, para fins de mineração textual, torna-se grande desafio a automatização da associação semântica de conteúdo genérico com a devida inferência de interessados afetados, quando a deliberação não menciona explicitamente quem são os envolvidos.

O desafio torna-se maior quando há várias deliberações genéricas acerca de diferentes tipos de decisão, ora por legalidade, ora por ilegalidade, sem qualquer menção direta ao subconjuntos de jurisdicionados afetados, contidos no conjunto de interessados.

O Acórdão 11.234/2017-1 é exemplo emblemático acerca do argumento exposto (Figura 43). O subitem 9.3.1. dispõe sobre a deliberação de cessar pagamentos decorrentes dos atos impugnados. Caso o questionamento seja na direção de entender a quais atos a deliberação se

refere, é necessário analisar todo o contexto do acórdão para então inferir quais são os interessados afetados.

Figura 42: Deliberações genéricas - cessação de pagamentos

determinar que, no prazo de 15 (quinze) dias contados da ciência desta deliberação, a Gerência Executiva do INSS em Chapecó '' SC dê ciência do inteiro teor desta deliberação à interessada
ACORDAM os Ministros do Tribunal de Contas da União, reunidos em sessão da Segunda Câmara, diante das razões expostas pelo relator, e com fundamento no art. 71, incisos III e IX, da Constituição Federal de 1988, c/c os arts. 1°, inciso V, 39, inciso II, e 45 da Lei 8.443/1992, e ainda com os arts. 261, caput e § 1°, e 262, caput e § 2°, do Regimento Interno/TCU, em 9.3. determinar ao Tribunal Superior do Trabalho que 9.3.1. faça cessar os pagamentos decorrentes do ato considerado ilegal, no prazo de 15 (quinze) dias, contado da ciência da presente deliberação, sob pena de responsabilidade solidária da autoridade administrativa omissa, até a emissão de novo ato escoimado das irregularidades verificadas, a ser submetido à apreciação do TCU
9.4.3. faça cessar o pagamento decorrente do ato considerado ilegal, sob pena de responsabilidade solidária da autoridade administrativa omissa, até a emissão de novo ato, livre das irregularidades apontadas no presente processo e, agora, fundamentado no art. 40, §1°, inciso III, alínea "b", da Constituição Federal, com a redação dada pelas Emendas Constitucionais n°s 20/1998 e 41/2003, a ser submetido à apreciação do TCU
9.4.22. suspenda os pagamentos efetuados com base nos atos ora impugnados, sob pena de responsabilidade solidária da autoridade administrativa omissa
9.4.2. suspenda os pagamentos efetuados com base no ato ora impugnado, sob pena de responsabilidade solidária da autoridade administrativa omissa
9.4.2. fazer cessar, no prazo de 15 (quinze) dias, o pagamento decorrente do ato considerado ilegal, sob pena de responsabilidade solidária da autoridade administrativa omissa
9.4.2. fazer cessar, no prazo de 15 (quinze) dias, o pagamento decorrente do ato considerado ilegal, sob pena de responsabilidade solidária da autoridade administrativa omissa
9.4.2. fazer cessar, no prazo de 15 (quinze) dias, o pagamento decorrente do ato considerado ilegal, sob pena de responsabilidade solidária da autoridade administrativa omissa
9.4.2. fazer cessar, no prazo de 15 (quinze) dias, o pagamento decorrente do ato considerado ilegal, sob pena de responsabilidade solidária da autoridade administrativa omissa
9.4.2. fazer cessar, no prazo de 15 (quinze) dias, o pagamento decorrente do ato considerado ilegal, sob pena de responsabilidade solidária da autoridade administrativa omissa
9.3.3. suspenda, no prazo de trinta dias, os pagamentos efetuados com base no ato ora impugnado
9.3.3. suspenda, no prazo de trinta dias, os pagamentos efetuados com base no ato ora impugnado
9.3.3. suspenda, no prazo de trinta dias, os pagamentos efetuados com base no ato ora impugnado
9.3.3. faça cessar, no prazo de 15 (quinze) dias, os pagamentos decorrentes do ato impugnado, sob pena de responsabilidade solidária da autoridade administrativa omissa
9.3.3. faça cessar, após a devida notificação, os pagamentos decorrentes do ato ora impugnado, sob pena de responsabilidade solidária da autoridade administrativa omissa
9.3.3. faça cessar, após a devida notificação, os pagamentos decorrentes do ato ora impugnado, sob pena de responsabilidade solidária da autoridade administrativa omissa
9.3.3. faça cessar, após a devida notificação, os pagamentos decorrentes do ato considerado ilegal, sob pena de responsabilidade solidária da autoridade administrativa omissa
9.3.3. faça cessar, a partir da ciência do presente acórdão, os pagamentos decorrentes do ato considerado ilegal, sob pena de responsabilidade solidária da autoridade administrativa omissa
9.3.3. faça cessar os pagamentos decorrentes do ato ora impugnado, sob pena de responsabilidade solidária da autoridade administrativa omissa, nos termos do inciso IX do art. 71 da Constituição Federal
9.3.3. faça cessar os pagamentos decorrentes do ato ora considerado ilegal, até a emissão de novo ato, livre das irregularidades identificadas no presente processo, a ser submetido à apreciação do TCU, sob pena de responsabilidade solidária da autoridade administrativa omissa
9.3.3. faça cessar os pagamentos decorrentes do ato considerado ilegal, sob pena de responsabilidade solidária da autoridade administrativa omissa, até a eventual emissão de novo ato, livre das irregularidades apontadas no presente processo, a ser submetido à apreciação do TCU

Fonte: Elaborada pelo autor (2019).

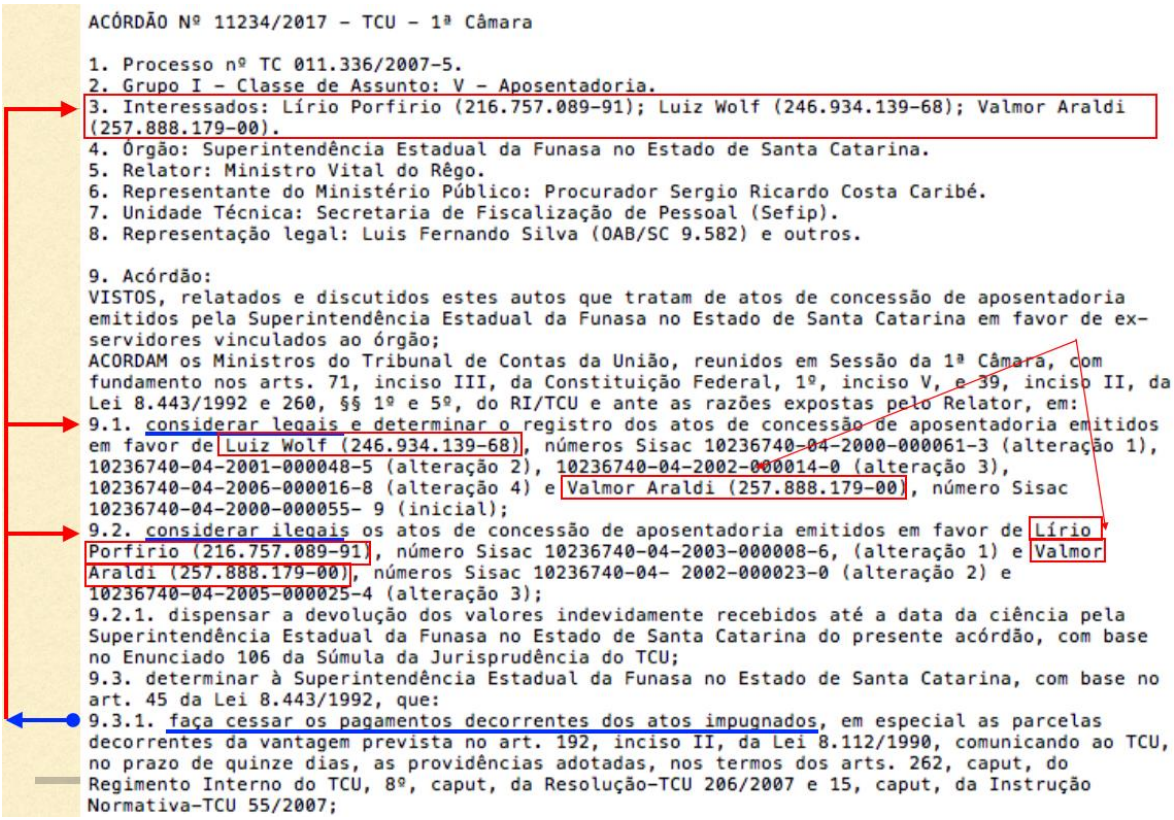
No caso concreto, há dois outros itens de deliberação – 9.1. e 9.2. – um versando sobre legalidade e ilegalidade de atos, respectivamente, elencando os interessados afetados. Caso a solução se restrinja a identificação dos interessados enumerados em cada caso, o problema estaria esclarecido.

Um desafio adicional foi identificado. No mesmo acórdão, um mesmo interessado é mencionado nos dois itens de deliberação. Como a única diferença entre os dois contéudos são os identificadores do Sisac associados ao nome do interessado, pode-se inferir que há mais de um ato sendo apreciado.

Outro desafio adicional também foi identificado. O item 9.2 menciona dois números de Sisac, e não apenas um. Então, no caso concreto, um mesmo interessado é enumerado em

deliberação versando sobre legalidade e também em outra versando sobre ilegalidade. E neste caso, em dois atos distintos.

Figura 43: Deliberações genéricas - caso do Acórdão 11.234/2017-1



ACÓRDÃO Nº 11234/2017 - TCU - 1ª Câmara

1. Processo nº TC 011.336/2007-5.
2. Grupo I - Classe de Assunto: V - Aposentadoria.
3. Interessados: Lírio Porfirio (216.757.089-91); Luiz Wolf (246.934.139-68); Valmor Araldi (257.888.179-00).
4. Órgão: Superintendência Estadual da Funasa no Estado de Santa Catarina.
5. Relator: Ministro Vital do Rêgo.
6. Representante do Ministério Público: Procurador Sergio Ricardo Costa Caribé.
7. Unidade Técnica: Secretaria de Fiscalização de Pessoal (Sefip).
8. Representação legal: Luis Fernando Silva (OAB/SC 9.582) e outros.

9. Acórdão:
VISTOS, relatados e discutidos estes autos que tratam de atos de concessão de aposentadoria emitidos pela Superintendência Estadual da Funasa no Estado de Santa Catarina em favor de ex-servidores vinculados ao órgão;
ACORDAM os Ministros do Tribunal de Contas da União, reunidos em Sessão da 1ª Câmara, com fundamento nos arts. 71, inciso III, da Constituição Federal, 1º, inciso V, e 39, inciso II, da Lei 8.443/1992 e 260, §§ 1º e 5º, do RI/TCU e ante as razões expostas pelo Relator, em:
9.1. considerar legais e determinar o registro dos atos de concessão de aposentadoria emitidos em favor de Luiz Wolf (246.934.139-68), números Sisac 10236740-04-2000-000061-3 (alteração 1), 10236740-04-2001-000048-5 (alteração 2), 10236740-04-2002-000014-0 (alteração 3), 10236740-04-2006-000016-8 (alteração 4) e Valmor Araldi (257.888.179-00), número Sisac 10236740-04-2000-000055- 9 (inicial);
9.2. considerar ilegais os atos de concessão de aposentadoria emitidos em favor de Lírio Porfirio (216.757.089-91), número Sisac 10236740-04-2003-000008-6, (alteração 1) e Valmor Araldi (257.888.179-00), números Sisac 10236740-04- 2002-000023-0 (alteração 2) e 10236740-04-2005-000025-4 (alteração 3);
9.2.1. dispensar a devolução dos valores indevidamente recebidos até a data da ciência pela Superintendência Estadual da Funasa no Estado de Santa Catarina do presente acórdão, com base no Enunciado 106 da Súmula da Jurisprudência do TCU;
9.3. determinar à Superintendência Estadual da Funasa no Estado de Santa Catarina, com base no art. 45 da Lei 8.443/1992, que:
9.3.1. faça cessar os pagamentos decorrentes dos atos impugnados, em especial as parcelas decorrentes da vantagem prevista no art. 192, inciso II, da Lei 8.112/1990, comunicando ao TCU, no prazo de quinze dias, as providências adotadas, nos termos dos arts. 262, caput, do Regimento Interno do TCU, 8º, caput, da Resolução-TCU 206/2007 e 15, caput, da Instrução Normativa-TCU 55/2007;

Fonte: Elaborada pelo autor (2019).

Portanto, a análise depende da compreensão do contexto do acórdão, dificultando sobremaneira qualquer estratégia de automação, a menos que padrões de elaboração textual possam ser criados para facilitar a mineração textual e, ao mesmo texto, propiciar igual ou superior ganho no exercício laboral de produção de documentos versando sobre pessoal. A proposta será elencada no rol de oportunidades de melhoria do presente trabalho.

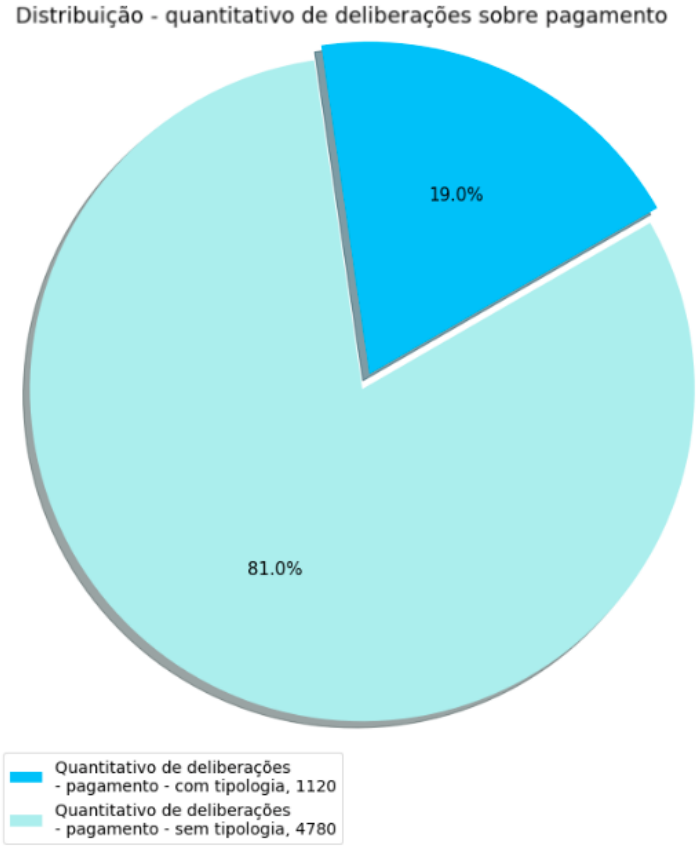
Diante da dificuldade exposta, a estratégia foi mapear tipologias capazes de criar alguma condição de contorno vencedora, versando sobre cessação de pagamentos e englobando o maior número possível de casos rotulados como P* (pagamento).

Sendo assim, duas tipologias foram propostas, T1 e T2. Também foram apresentadas informações decorrentes de análise exploratória, com o intuito de igualmente facilitar o levantamento de indícios e direcionar o planejamento de ações de controle.

Como resultados, as tipologias permitiram selecionar 1.120 deliberações pelas tipologias, do total de 5.900 classificadas como versando sobre pagamento (Figura 44).

Outro resultado foi a seleção de interessados abrangidos pelas tipologias para redução de espaço de busca no SIAPE.

Figura 44: Distribuição - quantitativo de deliberações sobre pagamento



Fonte: Elaborada pelo autor (2019).

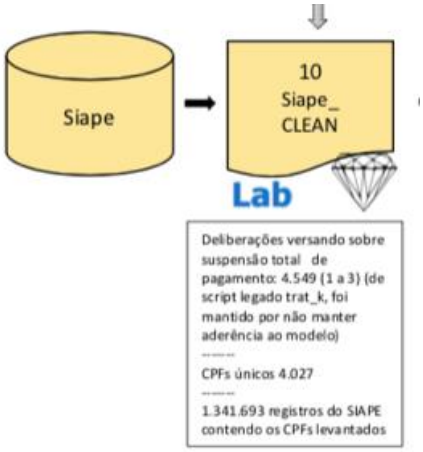
Cumpre ressaltar novamente a prevalência de deliberações genéricas no espaço remanescente, consubstanciando-se em desafio a implementação de novas tipologias capazes de propiciar a adequada identificação de interessados, a ser proposto como trabalho futuro, como continuidade à seleção das demais deliberações versando sobre cessação de pagamento.

Por fim, o módulo 13 será analisado na raia AZUL CLARA.

8.5. Raia AMARELA

A raia AMARELA é constituída de módulo de consulta e limpeza de dados provenientes do SIAPE – **módulo 10** – Siape_CLEAN. A consulta foi realizada no ambiente LabContas, uma vez que cópia da base de dados encontra-se disponível naquele espaço. Para minimizar o espaço de busca, a consulta foi limitada ao escopo de interessados identificados nas tipologias. Como resultados, foram selecionados 1.341.693 registros do SIAPE contendo os CPFs empregados.

Figura 45: Raia AMARELA (SIAPE)

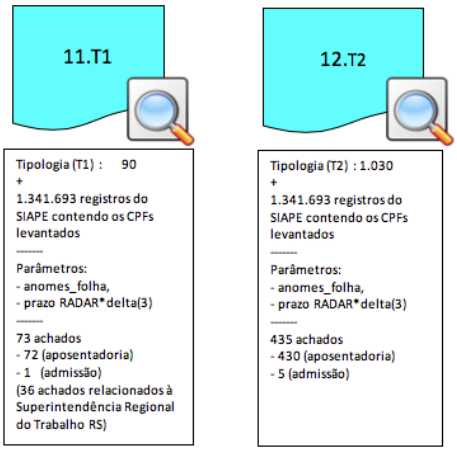


Fonte: Elaborada pelo autor (2019).

8.6. Raia AZUL CLARA (tipologias)

A raia AZUL CLARA é responsável pela implementação das tipologias propriamente ditas. A entrada é formada por conjuntos. O primeiro é formado pelas entradas selecionadas no módulo 9 (TYPOLOGIES). O segundo, pelo resultado da busca no SIAPE – módulo 10.

Figura 46: Raia AZUL CLARA (tipologias)



Fonte: Elaborada pelo autor (2019).

O **módulo 11** – T1 – contém implementação da tipologia 1. Foi baseada em deliberações "não genéricas", contendo alguma identificação de interessados no conteúdo textual da decisão acerca de suspensão de pagamento.

A estratégia de desenho da tipologia baseou-se na possibilidade de identificar o afetado pela deliberação rotulada como sendo sobre cessação de pagamento, por meio de alguma remissão aos afetados na própria decisão. A identificação seria feita por meio de partes do nome ou do CPF, registrados no cabeçalho do arquivo textual do acórdão. A estratégia leva em

consideração a necessidade de estabelecer condições de contorno para as deliberações genéricas discutidas no módulo 9. Para tanto, quatro critérios de associação foram implementados.

Primeiro: Identificação a partir de menção ao CPF no texto da decisão. É o caminho capaz de assegurar vitória rápida, empregando a limpeza e seleção de campos, sem pontos ou traços, para posterior comparação com o obtido no rol de interessados: *CPF (interessado) df2['enc_temp'] = [re.sub(r'(?<=[0-9]{3})(\.|-)(?=[0-9]{2,3})','',i) for i in df2['enc_temp']]*.

Segundo e terceiro: Estratégia similiar ao anterior, mas por meio do nome completo ou primeiro e último nome, prática comum no trabalho de elaboração de textos.

Quarto: Por similaridade de nomes. Outra prática comum na elaboração de textos é o emprego de variações de nomes, a exemplo da abreviação de prenomes. A estratégia visa também abranger eventual possibilidade de incorreções materiais e supressão de partes do nome, inclusive do sobrenome.

Para o quarto critério, a seleção foi baseada em similaridade (*similarity*) de nomes, por meio da biblioteca *fuzzywuzzy* (SOMA, 2017). Após várias iterações, a versão final do modelo foi configurada com os seguintes paramêtros: *fuzzymatch_min_rate* = 64; *fuzzymatch_threshold* = 58; *fuzz.ratio*. Ou seja, o primeiro parâmetro indica o percentual mínimo aceitável para aceitação de similaridade entre o termo testado e o critério adotado, ou seja 64% de similaridade (*match*). O segundo é o limiar mínimo para depuração. O valor foi mantido em cerca de 10% abaixo do primeiro, a fim de propiciar análise de resultados logo abaixo do mínimo aceitável. Se o modelo estiver acertando em associações, mesmo que abaixo do mínimo, o critério de similaridade poderia ser ajustado a menor, durante o *tuning*. Caso as similaridades observadas abaixo do mínimo aceitável e acima do limiar apontassem para termos não correspondentes, a configuração seria mantida e o critério de parada atingido. O terceiro parâmetro indica que a ordenação e a repetição de termos não pode ser ignorada.

O critério de obtenção de nomes no conteúdo textual foi baseado no reconhecimento de entidades mencionadas ou *Named-Entity Recognition* (NER) (BHANDARI, 2018; NLTK PROJECT, 2019), para nomes próprios (*NNP*) de pessoa (*PERSON*). Os nomes próprios de pessoas foram extraídos em forma de lista (*choices*) e comparados ao nome associado ao interessado conforme critério de similaridade acima descrito. Foi definida a função *get_human_names(text)*.

A extração de nomes mostrou-se satisfatória para os objetivos do projeto, mesmo com a biblioteca em idioma original (inglês). Como trabalho futuro, convém levantar a existência e avaliar versão em português. Em particular, com vistas à melhoria da pontuação para nomes

contendo termos, tais como “de”, “da”, “do”, a exemplo de “Maria Silva Sousa do Carmo”. Observou-se tendência de obtenção de parte do nome para verificação de similaridade. Mesmo com a tendência de parte do nome assegurar *match* na comparação com o nome completo, é uma caminho incremental para aprimoramento do modelo.

Para facilitar o entendimento, os acórdãos abaixo servem como exemplo de aplicação de diferentes estratégias para reconhecimento de interessados em deliberações consideradas versando sobre cessação de pagamento, seja por CPF, nome completo/primeiro-último nome e similaridade de nomes (Figura 47).

Figura 47: Tipologia 1 - amostragem de Acórdãos

ACÓRDÃO Nº 10509/2017 - TCU - 1ª Câmara

1. Processo nº TC 028.475/2017-8.
2. Grupo I - Classe de Assunto: V - Aposentadoria
3. Interessados/Responsáveis:
3.1. Interessados: Beatriz Pasa (271.095.260-20); Carmen Berta Tréz Rodrigues (118.240.930-04); Catarina Goggia (310.124.349-15); Claudio Cezar Peres (316.504.300-00); Claudio José Diettrich (017.938.610-72).
4. Órgão/Entidade: Superintendência Regional do Trabalho e Emprego no Estado do Rio Grande do Sul.
5. Relator: Ministro Benjamin Zymler.
6. Representante do Ministério Público: Procurador Marinus Eduardo De Vries Marsico.
7. Unidade Técnica: Secretaria de Fiscalização de Pessoal (SEFIP).
8. Representação legal: não há

9. Acórdão:
VISTOS, relatados e discutidos estes autos de aposentadorias emitidos no âmbito da Superintendência Regional do Trabalho e Emprego no Estado do Rio Grande do Sul,
ACORDAM os Ministros do Tribunal de Contas da União, reunidos em sessão da 1ª Câmara, diante das razões expostas pelo Relator, com fundamento no art. 71, inciso III, da Constituição Federal e nos arts. 1º, inciso V, 39, inciso II, e 45 da Lei 8.443, de 16 de julho de 1992, em:
9.1. considerar legais os atos de aposentadoria de Beatriz Pasa (271.095.260-20), Carmen Berta Tréz Rodrigues (118.240.930-04), Catarina Goggia (310.124.349-15), Claudio Cezar Peres (316.504.300-00) e Claudio José Diettrich (017.938.610-72), determinando-se o correspondente registro;
9.2. determinar à Superintendência Regional do Trabalho e Emprego no Estado do Rio Grande do Sul, com fulcro nos arts. 71, inciso IX, da Constituição Federal e 6º, § 2º, da Resolução TCU 206/2007, que:
9.2.1. faça cessar, no prazo de 15 (quinze) dias contado a partir da ciência desta deliberação, sob pena de responsabilidade solidária da autoridade administrativa omissa, o pagamento aos inativos Beatriz Pasa (271.095.260-20), Carmen Berta Tréz Rodrigues (118.240.930-04), Catarina Goggia (310.124.349-15), Claudio Cezar Peres (316.504.300-00) e Claudio José Diettrich (017.938.610-72) do Bônus de Eficiência e Produtividade, previsto na Lei 13.464/2017, por incompatível com o art. 40, caput e §§ 1º, 3º, 4º e 18, da Constituição Federal, dada a expressa exclusão da vantagem, de caráter pro labore faciendo, da base de cálculo de contribuição previdenciária;
9.2.2. dê ciência do inteiro teor desta deliberação aos interessados, alertando-os de que o efeito suspensivo proveniente da interposição de eventuais recursos, caso não providos, não os exime da devolução dos valores indevidamente percebidos após a notificação;
9.2.3. envie a esta Corte de Contas, no prazo de 30 (trinta) dias, por cópia, comprovante de que

ACÓRDÃO Nº 5624/2016 - TCU - 1ª Câmara

1. Processo nº TC 008.625/2014-0.
2. Grupo II - Classe de Assunto: V - Aposentadoria
3. Interessados: Abel Carlos Avancini (054.909.100-97); Abel Carlos Avancini (054.909.100-97); Catarina Santos dos Santos (262.920.090-68); Marco Aurélio de Aguiar Costa (685.430.938-72); Maria Heloisa Cardoso Pinto (387.780.170-68); Maria Rejane de Freitas Flores (270.334.140-72); Maria Salete da Rosa da Silva (207.760.950-53); Marilene Correa (199.965.270-34); Mario Hideki Osanai (121.333.700-30); Marisa Alves Paz Costa (556.361.600-20); Marisa Mattarredona Barbosa (284.856.700-78); Marta Mendes Schneider (376.323.370-91); Martha Helena Leal da Costa (298.004.550-00); Mary Alba Escouto (434.638.430-72); Neila Ribeiro Daiello (069.910.410-68); Neila Ribeiro Daiello (069.910.410-68); Neuza Lagranha Teichmann (631.981.690-20); Osmar de Vargas Drower (180.561.480-00); Otto Pinheiro da Silva (150.022.840-00); Paulo Abreu Barcellos (055.505.720-87); Paulo Abreu Barcellos (055.505.720-87); Samuel Burd (000.141.250-72); Sergio da Costa Franco Filho (253.179.970-20); Sonia Maria Gonçalves Avancini (056.528.470-34); Sonia Maria Gonçalves Avancini (056.528.470-34); Valmir Vieira Guimarães (066.629.190-04).
4. Órgão: Núcleo Estadual do Ministério da Saúde no Estado do Rio Grande do Sul.
5. Relator: Ministro Benjamin Zymler.
6. Representante do Ministério Público: Procurador Marinus Eduardo De Vries Marsico.
7. Unidade Técnica: Secretaria de Fiscalização de Pessoal (SEFIP).
8. Representação legal: não há

9. Acórdão:
VISTOS, relatados e discutidos estes autos de aposentadorias concedidas pelo Núcleo Estadual do Ministério da Saúde no Rio Grande do Sul,
ACORDAM os Ministros do Tribunal de Contas da União, reunidos em sessão da 1ª Câmara, diante das razões expostas pelo Relator, com fundamento no art. 71, inciso III, da Constituição Federal e nos arts. 1º, inciso V, 39, inciso II, e 45 da Lei 8.443, de 16 de julho de 1992, em:
9.1. considerar prejudicado, por perda de objeto, nos termos do art. 260, § 5º, do Regimento Interno, o exame dos atos de aposentadoria de Maria Heloisa Cardoso Pinto e Samuel Burd;
9.2. considerar legais os atos de aposentadoria de interesse de Abel Carlos Avancini (inicial e alteração), Marco Aurélio de Aguiar Costa, Maria Rejane de Freitas Flores, Maria Salete da Rosa da Silva, Marilene Correa, Marisa Alves Paz Costa, Marisa Mattarredona Barbosa, Marta Mendes Schneider, Martha Helena Leal da Costa, Mary Alba Escouto, Neila Ribeiro Daiello (inicial e alteração), Neuza Lagranha Teichmann, Osmar de Vargas Drower, Otto Pinheiro da Silva, Paulo Abreu Barcellos (inicial e alteração), Sergio da Costa Franco Filho, Sonia Maria Gonçalves Avancini (inicial) e Valmir Vieira Guimarães, ordenando seu registro;
9.3. considerar ilegais os atos de aposentadoria de interesse de Catarina Santos dos Santos, Mario Hideki Osanai e Sonia Maria Gonçalves Avancini (alteração), recusando seu registro;
9.4. dispensar o ressarcimento das quantias indevidamente recebidas, em boa-fé, pelos inativos mencionados no subitem anterior, consoante o Enunciado 106 da Súmula de Jurisprudência deste Tribunal;
9.5. determinar ao Núcleo Estadual do Ministério da Saúde no Rio Grande do Sul, com fulcro nos arts. 71, inciso IX, da Constituição Federal e 262 do Regimento Interno desta Corte, que:
9.5.1. faça cessar, no prazo de 15 (quinze) dias contado a partir da ciência desta deliberação, sob pena de responsabilidade solidária da autoridade administrativa omissa, os pagamentos decorrentes dos atos de aposentadoria de Catarina Santos dos Santos e Mario Hideki Osanai;
9.5.2. corrija, no sistema Siape, a fundamentação legal da aposentadoria de Sonia Maria Gonçalves Avancini, excluindo a

ACÓRDÃO Nº 2687/2019 - TCU - 1ª Câmara

1. Processo nº TC 027.444/2010-4.
2. Grupo I - Classe de Assunto: V - Aposentadoria
3. Interessados/Responsáveis:
3.1. Interessados: Ana Valenga (357.089.309-00); Ana Valenga (357.089.309-00); Arlete Edling (232.638.659-00); Arlete Edling (232.638.659-00); Eliezer Gomes da Silva (091.482.059-15); Eunice Brisola Inocêncio (720.836.209-25); Eunice Brisola Inocêncio (720.836.209-25); Eunice Brisola Inocêncio (720.836.209-25); Joensen Terezinha Lizott Disperati (253.533.659-68); Juarez Nelson Alves de Lima (083.680.409-06); Marcelo Iacomini (183.940.949-53).
4. Órgão/Entidade: Universidade Federal do Paraná.
5. Relator: Ministro Benjamin Zymler.
6. Representante do Ministério Público: Procurador Rodrigo Medeiros de Lima.
7. Unidade Técnica: Secretaria de Fiscalização de Pessoal (SEFIP).
8. Representação legal:
8.1. Vinicius Rafael Presente (66052/OAB-PR) e outros, representando Eunice Brisola Inocêncio.
8.2. Ernani Moreno Silva (38.050/OAB-PR) e outros, representando Ana Valenga.
8.3. Mauricio de Jesus Tozetti (38.229/OAB-PR) e outros, representando Eliezer Gomes da Silva.

9. Acórdão:
VISTOS, relatados e discutidos estes autos de processo de aposentadoria de ex-servidores da Universidade Federal do Paraná,
ACORDAM os Ministros do Tribunal de Contas da União, reunidos em sessão da Primeira Câmara, diante das razões expostas pelo relator e com fundamento na Constituição Federal, art. 71, III e IX, e na Lei 8.443/1992, arts. 1º, V, e 39, II, em:
9.1. considerar legais os atos de concessão de aposentadoria aos ex-servidores Arlete Edling, Juarez Nelson Alves de Lima e Eunice Brisola Inocêncio (número de controle 10792600-04-2015-000133-0) e determinar seus registros;
9.2. considerar ilegais os atos de concessão de aposentadoria aos srs. Ana Valenga, Eliezer Gomes da Silva, Eunice Brisola Inocêncio (números de controle 10792600-04-2016-000113-8 e 10792600-04-2005-000319-5), Joensen Terezinha Lizott e Marcelo Iacomini e a eles negar registro;
9.3. dispensar a devolução dos valores indevidamente recebidos de boa-fé pelos interessados citados no subitem anterior, nos termos do Enunciado 106 da Súmula de Jurisprudência deste Tribunal;

ral do Paraná que adote as seguintes medidas, sob pena de
idade administrativa omissa:
esta deliberação aos srs. Ana Valenga, Eliezer Gomes da Silva,
erezinha Lizott e Marcelo Iacomini no prazo de quinze dias e
vantes de notificação nos quinze dias subsequentes;
nta dias, os pagamentos efetuados com base nos atos ora
Valenga, Eliezer Gomes da Silva, Joensen Terezinha Lizott e

liezer Gomes da Silva, Joensen Terezinha Lizott Disperati e
poderão prosperar, nos moldes em que foram concedidas,
ição, de forma indenizada, sobre os períodos de atividade
ciado 268 da Súmula da Jurisprudência do TCU;
da Silva que, em caso de não recolhimento da contribuição
do por idade, com proventos proporcionais ao tempo de

Fonte: Elaborada pelo autor (2019).

O cruzamento de dados com a base do SIAPE foi realizado empregando os seguintes paramêtros: *DELTA=5*; PRAZO_LIMITE = DATA*PRAZO*DELTA. Além disso, as rubricas de pagamento foram agrupadas para obtenção de somatórios, a fim de facilitar a verificação de pagamentos por mês (*anomes_folha*).

O DELTA foi definido como sendo multiplicador sobre o prazo em dias definido na deliberação, extraída do conteúdo textual do cadastro no RADAR (PRAZO). A data de início da contagem foi convencionada como sendo a data de publicação do acórdão, obtida em tabela do mesmo sistema (DATA). A convenção foi necessária porque a data de comunicação ao órgão, possivelmente associada ao prazo oficial de início de contagem de prazos, não pôde ser obtida no prazo do projeto. Por isso, o DELTA foi estabelecido com margem a maior, a fim de minimizar riscos de convencionar prazo abaixo do realizado na prática (PRAZO_LIMITE).

Então, se houver algum pagamento identificado após PRAZO_LIMITE, considera-se um indício de irregularidade.

Sendo assim, uma oportunidade de melhoria é ajustar o modelo para que o prazo de início seja contado a partir da data de comunicação ao órgão jurisdicionado, dispensando o ajuste realizado em função dos dados disponibilizados, como a data de publicação do acórdão.

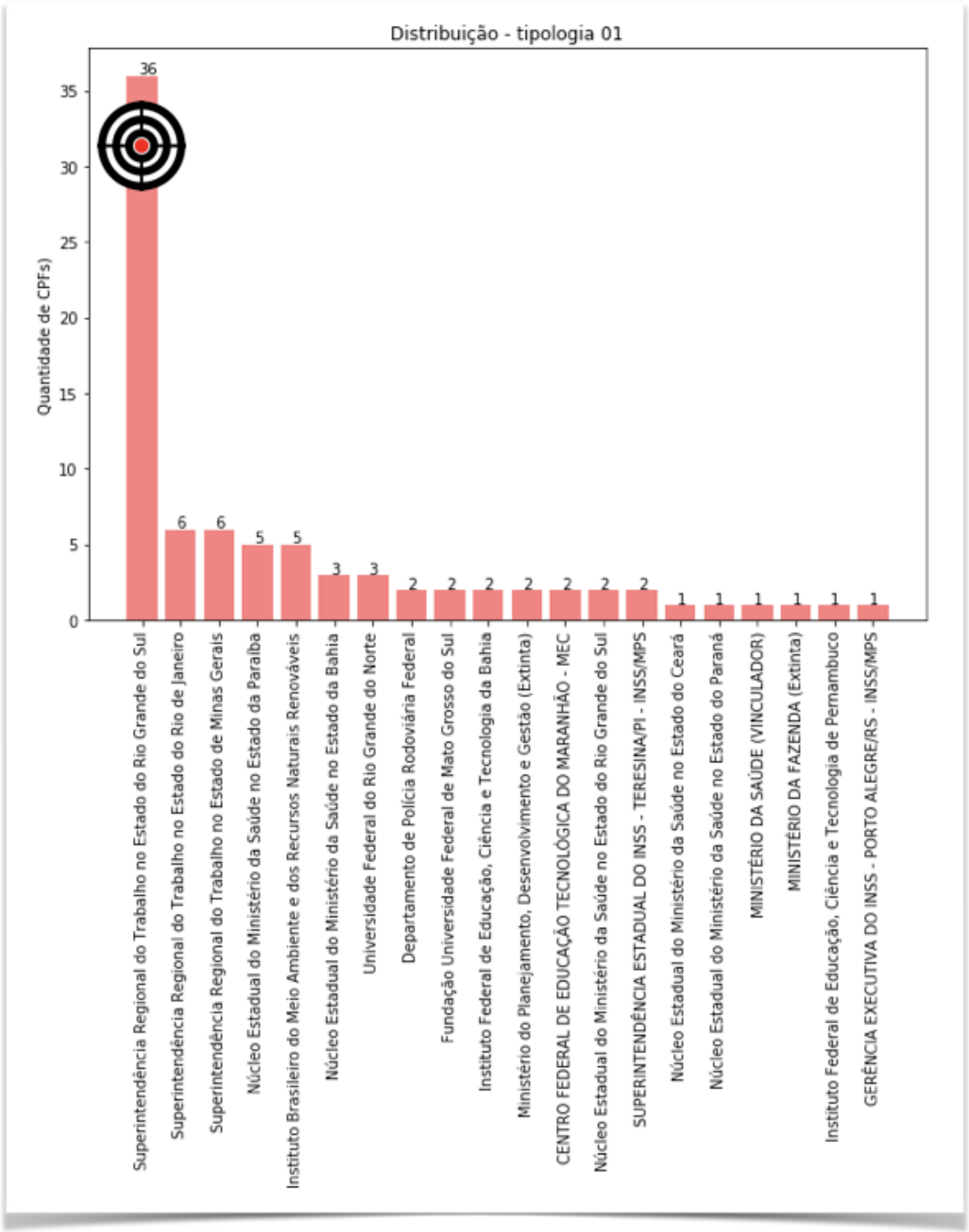
Como resultados, foram selecionadas 87 deliberações, das quais 73 apresentaram indícios de não cessação de pagamentos após o prazo limite convencionado, sendo um caso relacionado a ato de admissão de pessoal e 72 relacionados a atos de concessão de aposentadorias.

A Figura 48 apresenta resultados distribuídos por órgãos jurisdicionados. Ressalta-se posição de visual discrepância observada pela primeira posição, a qual pode servir para facilitar o planejamento de ações de controle, dada a concentração de eventos. A Figura 49 apresenta distribuição por atos de pessoal e de acórdãos.

O **módulo 12** – T2 – contém implementação da tipologia 2. Foi baseada em deliberações genéricas acerca de suspensão de pagamento, mas contidas em acórdãos contendo apenas um interessado. Cabe ressaltar que a tipologia levou em consideração as deliberações já selecionadas anteriormente.

A estratégia de desenho da tipologia, tal como a anterior, levou em consideração o problema da deliberação genérica discutido no módulo 9. Para o corrente caso, a tipologia foi desenhada para tratar conteúdo textual genérico. Em específico, para selecionar deliberações associadas a acórdãos contendo apenas um interessado. Sendo assim, a inferência lógica acerca do afetado pela decisão seria trivial.

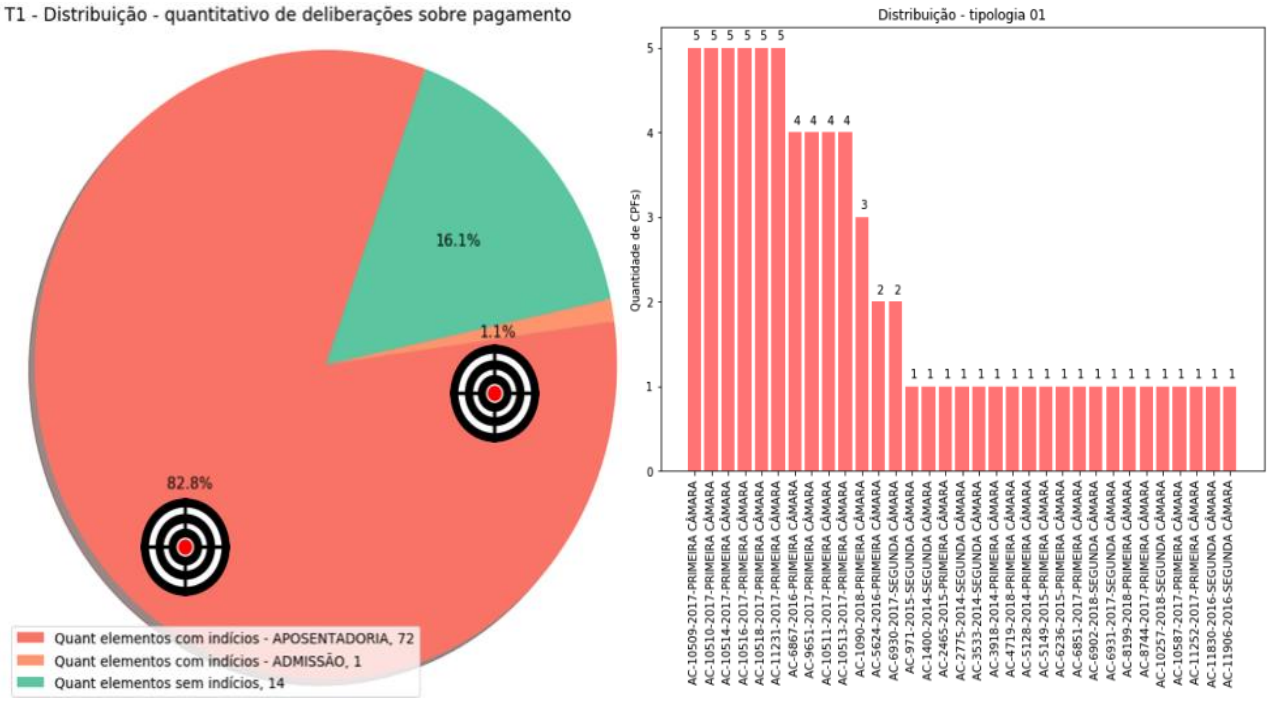
Figura 48: Tipologia 1 - resultados - distribuição por órgãos



Fonte: Elaborada pelo autor (2019).

O critério de levantamento de acórdãos contendo apenas um interessado foi o quantitativo de CPFs identificados no cabeçalho do documento. Isso reforça a necessidade da exclusão de decisões sem registro de CPF de interessados, conforme caso registrados na raia VERMELHA. Como o CPF é empregado também no cruzamento de dados com o SIAPE e como o número de casos é pouco relevante em relação ao domínio, a hipótese de relevar a ausência do referido registro está descartada da versão atual.

Figura 49: Tipologia 1 - resultados - distribuição por atos e acórdãos



Fonte: Elaborada pelo autor (2019).

Figura 50: Tipologia 2 - amostragem de Acórdão

ACÓRDÃO Nº 3556/2019 - TCU - 2ª Câmara

1. Processo nº TC 012.901/2009-3.
2. Grupo I - Classe de Assunto: V - Aposentadoria.
3. Interessados/Responsáveis:
3.1. Interessado: Xisto Moreira (070.191.067-49).
4. Órgão/Entidade: Departamento de Polícia Rodoviária Federal.
5. Relator: Ministro Augusto Nardes.
6. Representante do Ministério Público: Procurador Sergio Ricardo Costa Caribé.
7. Unidade Técnica: Secretaria de Fiscalização de Pessoal (SEFIP).
8. Representação legal: não há.

9. Acórdão:
VISTOS, relatados e discutidos estes autos de aposentadoria concedida no âmbito do Departamento de Polícia Rodoviária Federal.
ACORDAM os Ministros do Tribunal de Contas da União, reunidos em sessão da Segunda Câmara, diante das razões expostas pelo Relator, e com fundamento no art. 71, incisos III e IX, da Constituição Federal de 1988, c/c os arts. 1º, inciso V, 39, inciso II, e 45 da Lei 8.443/1992, e ainda com os arts. 260, § 1º, 261, caput e § 1º, e 262, caput e § 2º, do Regimento Interno do TCU, em:
9.1. considerar ilegal e negar registro ao ato de aposentadoria de Xisto Moreira (070.191.067-49);
9.2. dispensar o ressarcimento das quantias indevidamente recebidas de boa-fé pelo interessado até a data da ciência pelo Departamento de Polícia Rodoviária Federal do presente acórdão, com base no Enunciado 106 da Súmula da Jurisprudência do TCU;
9.3. determinar ao Departamento de Polícia Rodoviária Federal que:
9.3.1. no prazo de 15 (quinze) dias, contados da ciência da presente deliberação, faça cessar os pagamentos decorrente do ato ora considerado ilegal, sob pena de responsabilidade solidária da autoridade administrativa omissa;
9.3.2. no prazo de 15 (quinze) dias, contados do conhecimento da presente deliberação, dê ciência do inteiro teor deste acórdão ao interessado, esclarecendo-lhe que o efeito suspensivo proveniente da interposição de recurso não o exime da devolução dos valores percebidos indevidamente após a notificação, em caso de não provimento do recurso;
9.3.3. no prazo de 30 (trinta) dias, informe ao TCU as medidas adotadas e encaminhe comprovante sobre a data em que o interessado tomou conhecimento do contido no subitem anterior;
9.4. dar ciência desta deliberação ao interessado e ao Departamento de Polícia Rodoviária Federal.

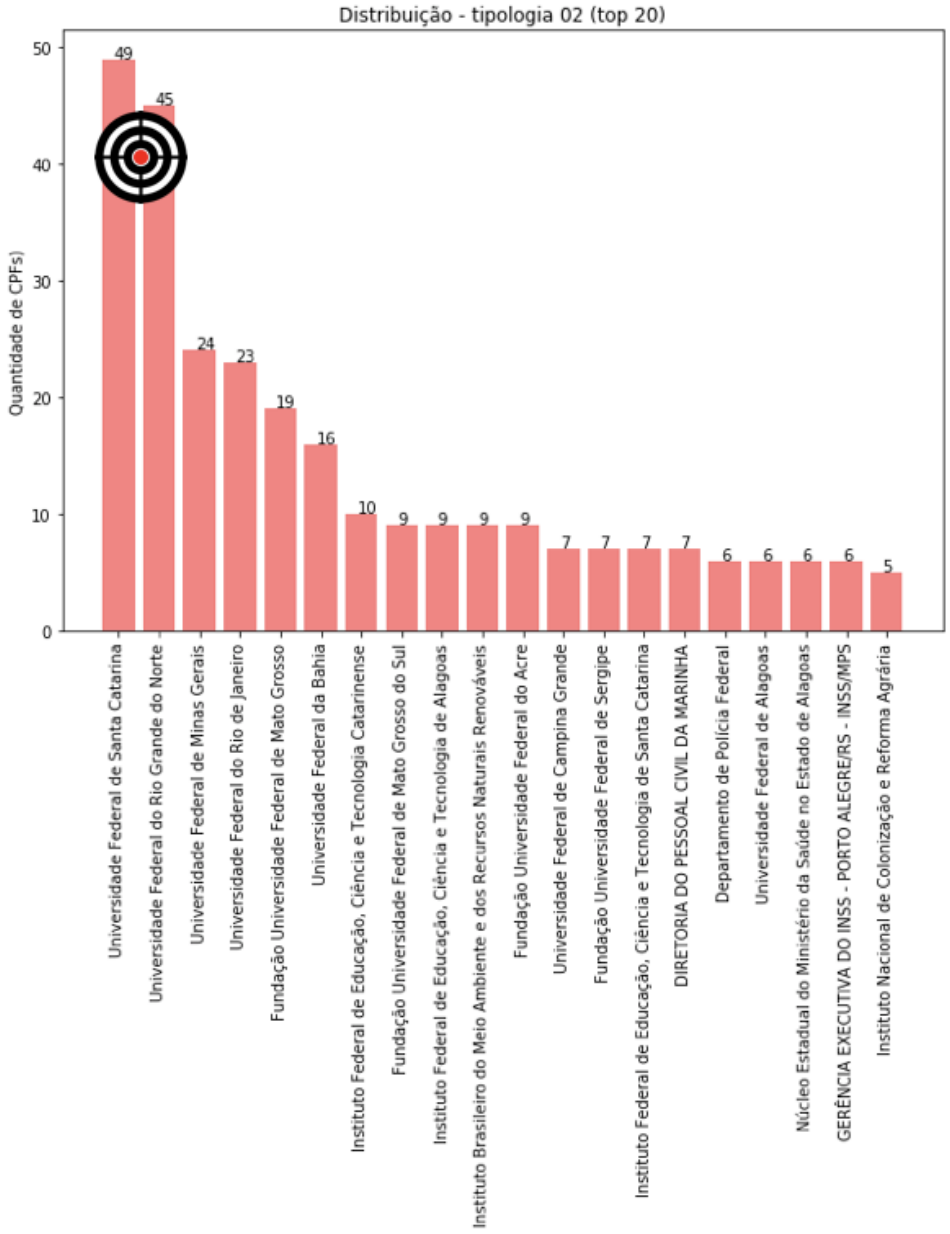
Fonte: Elaborada pelo autor (2019).

Para facilitar o entendimento, os acórdãos abaixo servem como exemplo de aplicação de diferentes estratégias para reconhecimento de interessados em deliberações consideradas

versando sobre cessação de pagamento, seja por CPF, nome completo/primeiro-último nome e similaridade de nomes (Figura 50).

Como resultados, foram selecionadas 1.030 deliberações, das quais 435 apresentaram indícios de não cessação de pagamentos após o prazo limite convencionado, sendo 5 casos relacionados a atos de admissão de pessoal e 430 relacionados a atos de concessão de aposentadorias.

Figura 51: Tipologia 2 - resultados - distribuição por órgãos

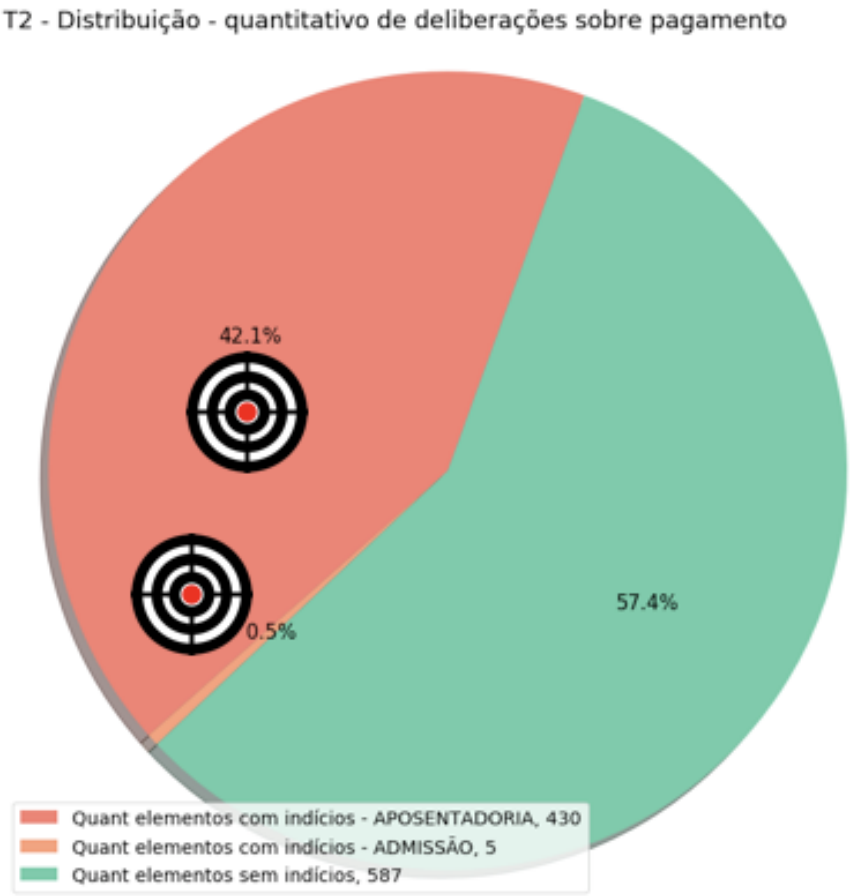


Fonte: Elaborada pelo autor (2019).

A Figura 51 apresenta resultados distribuídos por órgãos jurisdicionados. Ressalta-se posição de visual discrepância observada pelas duas primeiras posições, as quais pode servir para

facilitar o planejamento de ações de controle, dada a concentração de eventos. A Figura 52 apresenta distribuição por atos de pessoal.

Figura 52: Tipologia 2 - resultados - distribuição por atos



Fonte: Elaborada pelo autor (2019).

8.7. Conclusão

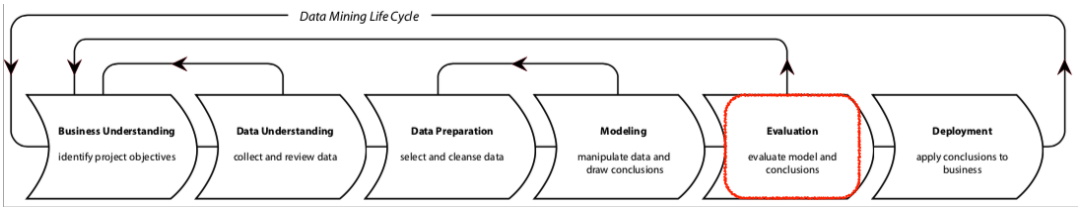
Nessa parte, foi realizado o detalhamento relativo à fase de modelagem previsto no CRISP-DM (IBM, 2014). Optou-se por descrever a validação do modelo (*Assess Model*), segundo critérios de sucesso de mineração de dados, na próxima parte, juntamente com a validação de resultados segundo critérios de sucesso de negócio (*Evaluate Results*). Sendo assim, os resultados consolidados pelas tipologias propostas, juntamente com informações decorrentes de análise exploratória também serão apresentados na fase seguinte.

9. AVALIAÇÃO

A **fase** de **avaliação** compreende a verificação dos resultados obtidos por meio das técnicas elencadas anteriormente, a partir de critérios de sucesso de negócio, bem como

levantamento de versões finais da modelagem e oportunidades de melhoria para trabalhos futuros. Adicionalmente, serão apresentados itens previstos pelo CRISP-DM para exposição no anterior, versando sobre validação do modelo e resultados consolidados do modelo.

Figura 53: Ciclo de vida - mineração de dados – Avaliação



Fonte: (LEAPER, 2009).

9.1. Validação do modelo

Conforme sinalizado na conclusão da fase anterior, optou-se por apresentar a validação do modelo (*Assess Model*), segundo critérios de sucesso de mineração de dados, na fase de avaliação.

A validação do classificador textual de deliberações adotou critérios de mensuração de desempenho enumerados por (ZAKI e MEIRA JR., 2014): *accuracy score/precision* e matriz de confusão (*confusion matrix)* - Figura 54.

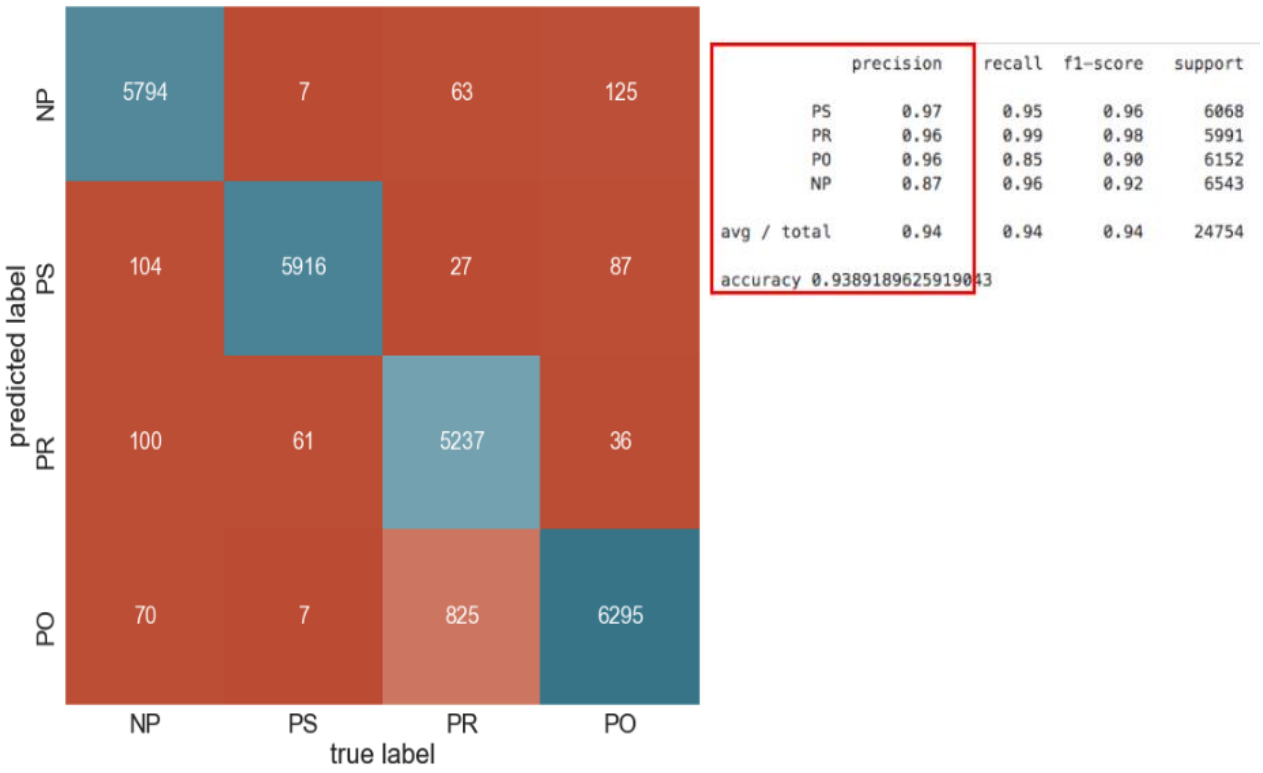
A acurácia obtida foi de 93,89%. A precisão de resultados por categorias de interesse (P*), variou entre 96% e 97%. Adianta-se então, que os valores obtidos pela acurácia e precisão estão alinhados com os objetivos de negócio, ou seja, estabelecer modelo que elevado grau de acerto.

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.

Conforme descrito no módulo 8, o balanceamento de classes foi necessário para melhoria do modelo. Na fase de *tag_samples* foi aventada a hipótese de elevada desproporcionalidade entre classes. Mesmo com a definição de categorias entre versões distintas do modelo, a proporção de textos versando sobre não pagamento e pagamento era notadamente desbalanceada em favor da primeira. A razão foi simples de entender e está diretamente relacionada à sistemática de elaboração das próprias deliberações. A presença de decisões versando sobre cessação de pagamento normalmente é acompanhada de vários outros dispositivos, entre eles, acerca de comunicar ao órgão e ao interessado sobre a decisão exarada, de assinar prazo para o

cumprimento de ações e de emitir novos atos. Contudo, as mesmas deliberações podem ocorrer com a ausência de dispositivos com efeito suspensivo, mesmo em caso de ilegalidade, conforme o caso concreto.

Figura 54: Validação do modelo - *confusion matrix/accuracy score*



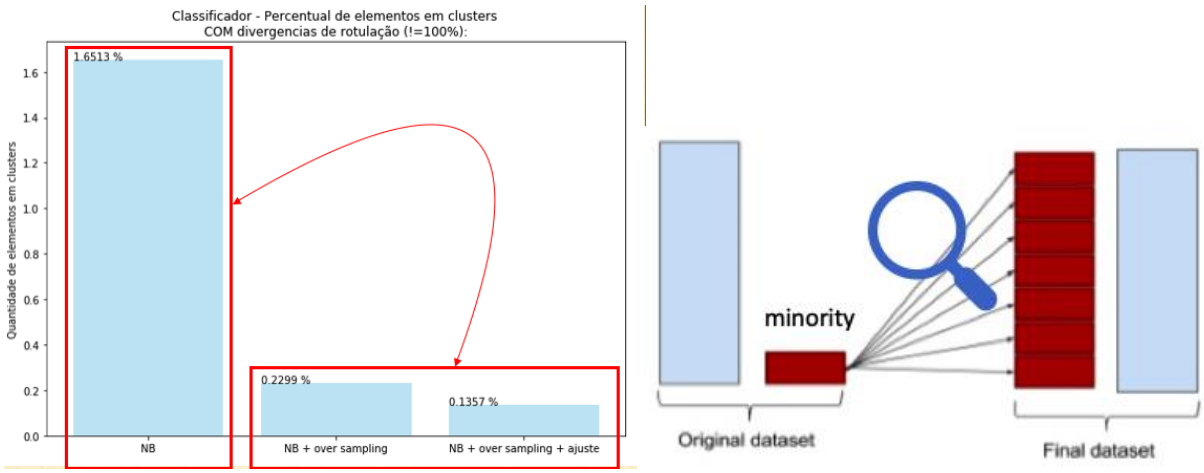
Fonte: Elaborada pelo autor (2019).

Sendo assim, foi empregada a técnica de *oversampling* (OS). (DATAMAN, 2018b) esclarece que a presença de classes desbalanceadas pode acarretar viés para a majoritária, reduzindo a acurácia do classificador e aumento a possibilidade de falsos positivos. Esclarece ainda que técnica de *undersampling* ou *oversampling* são opções possíveis para melhorar resultados. Ainda segundo o autor, o desbalanceamento ocorre quando ao menos uma das classes é significativamente menor que as demais, como observado nas primeiras iterações do modelo. A Figura 55 (à direita) apresenta visão conceitual da técnica. (DERNONCOURT, 2018) apresenta proposta de codificação por meio de *oversampling*. Adicionalmente, foi procedido ajuste manual de algumas classes, com vista a obter algum ganho de desempenho.

Os resultados foram satisfatórios e baseados no percentual de elementos em *clusters* com divergências de rotulação. A expectativa é de que a todos os elementos de um mesmo *cluster* seja atribuído o mesmo rótulo (P* ou PN), durante a etapa de classificação. Com a execução do

classificador *Naive Bayes*, sem a aplicação de *oversampling* (NB-OS), o percentual de *clusters* com divergências de rotulação em pelo menos um elemento foi de 1,65% (NB-OS). Após a aplicação de OS (NB+OS), o percentual caiu para 0,23%. Com o ajuste manual, o resultado teve ligeira melhora para 0,14% - Figura 55 (à esquerda).

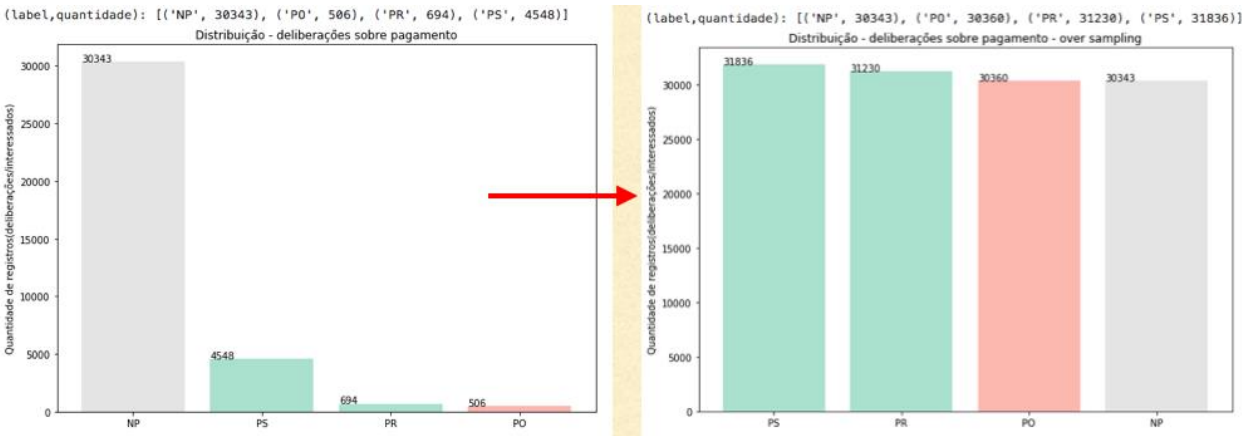
Figura 55: *Oversampling* - resultados e conceito



Fonte: Elaborada pelo autor (2019).

O *dataset* original, evidenciando o visível desbalanceamento da classe NP em relação às demais, pode ser visto na Figura 56 (à esquerda). O *dataset* apresenta distribuição de entradas contendo deliberações de pagamento. À direita da mesma figura, o resultado final, após o emprego de *oversampling*.

Figura 56: *Oversampling* - dados originais e após OS

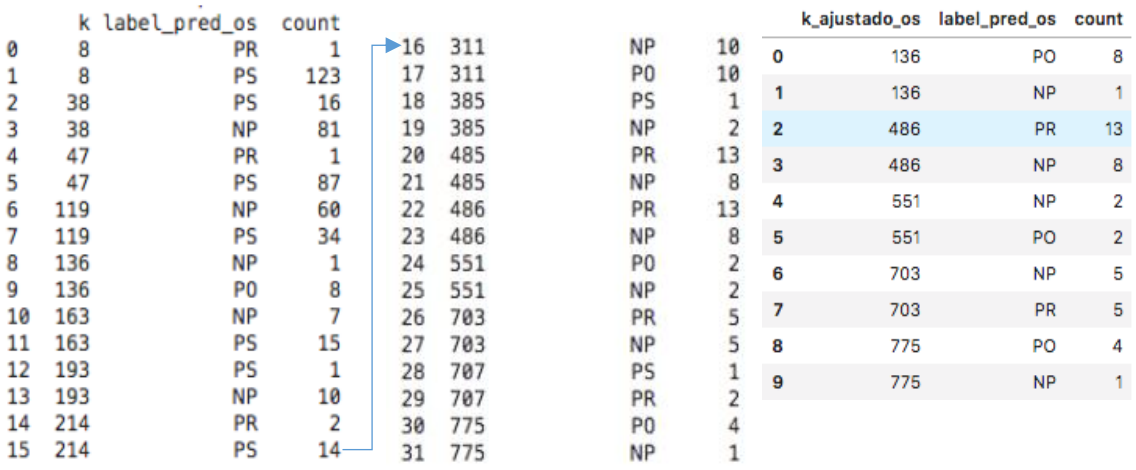


Fonte: Elaborada pelo autor (2019).

A Figura 57 apresenta inventário de *clusters* ajustados. Observa-se, por exemplo, o ajuste de k=8, com apenas um elemento divergente (PR,1) em relação aos demais (PS,123), onde (*label_pred_os*, *count*). O ajuste manual, mencionado na parte de resultados, diz respeito à retificação de rótulos em elementos de *clusters* divergentes. Em particular, o ajuste foi aplicado em classes P* (união das subclasses PR, PS e PO), a fim de possibilitar melhoria de resultados do classificador para deliberações versando sobre pagamento.

Cabe ressaltar ainda que o critério baseado no percentual de elementos em *clusters* com divergências de rotulação pode ser empregado para avaliar a qualidade no algoritmo de clusterização, 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.

Figura 57: *Oversampling* - resultados após OS e ajuste



	k	label_pred_os	count
0	8	PR	1
1	8	PS	123
2	38	PS	16
3	38	NP	81
4	47	PR	1
5	47	PS	87
6	119	NP	60
7	119	PS	34
8	136	NP	1
9	136	PO	8
10	163	NP	7
11	163	PS	15
12	193	PS	1
13	193	NP	10
14	214	PR	2
15	214	PS	14

16	311
17	311
18	385
19	385
20	485
21	485
22	486
23	486
24	551
25	551
26	703
27	703
28	707
29	707
30	775
31	775

NP	10
PO	10
PS	1
NP	2
PR	13
NP	8
PR	13
NP	8
PO	2
NP	2
PR	5
NP	5
PS	1
PR	2
PO	4
NP	1

	k_ajustado_os	label_pred_os	count
0	136	PO	8
1	136	NP	1
2	486	PR	13
3	486	NP	8
4	551	NP	2
5	551	PO	2
6	703	NP	5
7	703	PR	5
8	775	PO	4
9	775	NP	1

Fonte: Elaborada pelo autor (2019).

Como trabalho futuro, o ajuste manual pode ser automatizado, com a reclassificação de elementos quando a proporção majoritária for maior que determinado percentual limiar mínimo de aceitação, por exemplo 90%.

A Figura 58 apresenta comparativo da validação do classificador NB sem e com aplicação de OS, respectivamente. A parametrização NB-OS foi obtida em versões preliminares do modelo. Observa-se que a acurácia de NB-OS foi de 97,00%. Para NB+OS, de 93,89%. A aparente superioridade do primeiro resultado não necessariamente representa superioridade no classificador. Nota-se que a precisão do classificador por classe é concentrada na classe NP (não pagamento), 0,99. Para as demais, chega a 0,73. Por outro lado, para NB+OS, a precisão da classe P* (pagamento) varia entre 0,96 e 0,97. Sendo assim, para a classe P*, para NB+OS mostra-se mais adequado à classificação de deliberações sobre pagamento.

Após várias iterações do modelo, optou-se por criar ligeiro desbalanceamento na configuração de OS no dataset de entrada para a classe NP, como pode ser visto na Figura 56 (à direita). O objetivo foi proceder pequeno ajuste na precisão da classe P*. O resultado pode ser visto Figura 58 (à direita), para NB+OS, válida para a versão final.

O *dataset* original, evidenciando o visível desbalanceamento da classe NP em relação às demais, pode ser visto na Figura 56 (à esquerda). O *dataset* apresenta distribuição de entradas contendo deliberações de pagamento. À direita da mesma figura, o resultado final, após o emprego de oversampling.

Figura 58: *Oversampling - Accuracy score*: NB-OS (esquerda) e NB+OS (direita)

	precision	recall	f1-score	support
NP	0.99	0.99	0.99	6042
PS	0.74	0.61	0.67	90
PR	0.73	0.50	0.59	144
PO	0.90	0.97	0.93	943
avg / total	0.97	0.97	0.97	7219

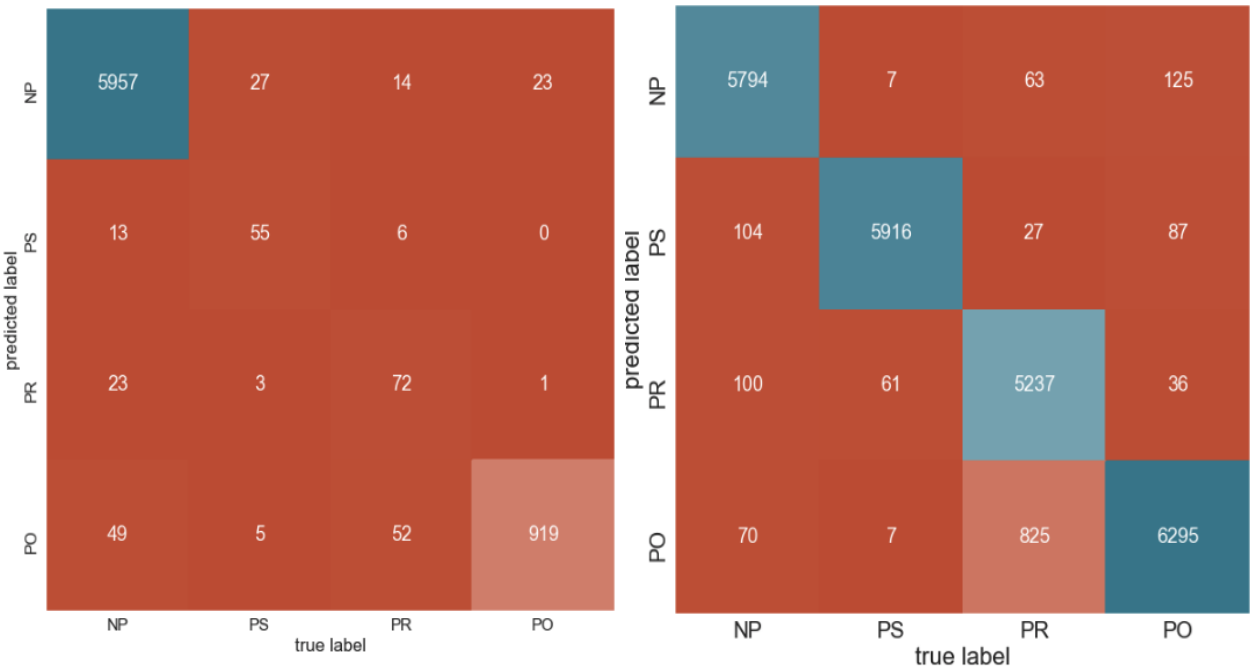
accuracy 0.9700789583044743

	precision	recall	f1-score	support
PS	0.97	0.95	0.96	6068
PR	0.96	0.99	0.98	5991
PO	0.96	0.85	0.90	6152
NP	0.87	0.96	0.92	6543
avg / total	0.94	0.94	0.94	24754

accuracy 0.9389189625919043

Fonte: Elaborada pelo autor (2019).

Figura 59: Matriz de Confusão - NB-OS e NB+OS



Fonte: Elaborada pelo autor (2019).

O emprego da técnica de OS também produz decorrências na matriz de confusão. A Figura 59 apresenta comparativo entre os resultados obtidos em NB-OS e NB+OS, respectivamente. Observa-se que a predição errada de valores para as classes não prevalentes 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 a aplicação de OS, o quantitativo de falsos positivos e verdadeiros positivos torna-se mais equilibrado entre as categorias (matriz à direita).

Diante do exposto, conclui-se que o modelo atende critérios de sucesso da mineração de dados. A validação do classificador textual de deliberações adotou critérios enumerados por (ZAKI e MEIRA JR., 2014): *accuracy score/precision* e matriz de confusão (*confusion matrix)*. Optou-se por apresentar a validação do modelo (*Assess Model*) na fase de avaliação. A acurácia obtida foi de 93,89%. A precisão de resultados por categorias de interesse (P*), variou entre 96% e 97%. Adianta-se então, que os valores obtidos pela acurácia e precisão estão alinhados com os objetivos de negócio, ou seja, estabelecer modelo que elevado grau de acerto. Além disso, para o algoritmo de clusterização, o percentual de elementos em *clusters* com divergências de rotulação pode ser empregado para validar o algoritmo não supervisionado, uma vez que o classificador é executado por deliberação e não por *clusters* e a expectativa é de que todos os elementos de um mesmo *cluster* sejam rotulados com a mesma classe. Por fim, o emprego de *oversampling* mostrou-se vantajoso, reduzindo o percentual de divergências de 1,6513% (NB-OS), para 0,2299% (NB+OS) e para 0,1357% após ajuste manual.

9.2. Avaliação de resultados

Trata-se da avaliação de resultados (*Evaluation Results*), segundo critérios de sucesso de negócio, estabelecidos no início do projeto. É a chave para garantir que a organização pode usar os resultados obtidos (IBM, 2014). Para proceder avaliação, o mesmo autor sugere a enumeração de questionamentos:

Primeiro: “Em geral, como esses resultados respondem às metas de negócios de sua organização?”. Segundo: “Os seus resultados estão indicados claramente e de uma forma que possam ser facilmente apresentados?”. Terceiro: “Há descobertas particularmente novas ou exclusivas que devem ser destacadas?”.

Para responder aos questionamentos, cumpre destacar os resultados de duas apresentações ocorridas junto à área de negócio, ocorridas nos dias **7/11/19** e **27/11/19**.

Na primeira, a área de negócio foi representada no âmbito da Unidade Técnica diretamente interessada (SEFIP). Estiveram presentes quase a totalidade de chefes e ocupantes de função da unidade: Secretário (em substituição), Reginaldo Fernandes; Diretores, Herbert Souza (1ª DIAT), Sebastião Arantes Júnior (2ª DIAT), Helton Onésio (DIAUP); Especialista Sênior II, José Luiz Costa; Chefe de Serviço, Andrea Ribeiro (SEAD); integrantes da Assessoria, Allysson Paulista e Barnabé Pereira, bem como demais servidores lotados na referida Unidade. Esteve presente também o orientador Prof. Dr. Edans Sandes. O objetivo foi de apresentar a proposta a área de negócio, juntamente com os resultados obtidos.

A segunda reunião foi agendada pela Sefip, em pauta destinada a apresentar soluções para automatização de processos da unidade. Por isso, o trabalho foi proposto para ser incluído na agenda. Estiveram presentes: Paulo Wiechers Martins, Secretário-Geral de Controle Externo (SEGECEX); Marcelo Eira, Secretário-Geral Adjunto de Controle Externo; Felipe Peñaloza, Coordenador-Geral de Controle Externo, de Gestão de Processos e Informações (COPIN), em substituição; Reginaldo Fernandes, Secretário da SEFIP, em substituição; Helton Onésio, Diretor da DIAUP/SEFIP; José Luiz Costa, Especialista Sênior II/SEFIP; Fernando Moreira, Assessor da COPIN.

Em ambas as ocasiões, os resultados apresentados foram considerados satisfatórios. A proposta foi considerada aderente às necessidades de negócio, particularmente para ganho de escala em alcance de resultados para a unidade técnica. Também foi discutido o cenário atual de racionalização de recursos humanos e o estímulo a iniciativas capazes de tornar mais eficiente o esforço para realização de ações de controle. Sendo assim, infere-se que o primeiro questionamento tenha sido atendido, sobre se resultados respondem às metas de negócio da organização.

Acerca do segundo questionamento “Os resultados estão indicados claramente e de forma que possam ser facilmente apresentados?”, esclarece-se que, em ambas as reuniões, os comentários apresentados durante cada oportunidade foram positivos. Na segunda agenda, houve necessidade de ajuste do tempo de apresentação para aproveitar o horário disponível. Mesmo assim, foi recebida avaliação positiva quanto à clareza na explanação. Sendo assim, infere-se que o segundo questionamento tenha sido atendido.

Sobre o terceiro questionamento “há descobertas particularmente novas ou exclusivas que devem ser destacadas?”, esclarece-se que na primeira apresentação, a área de negócio confirmou que os órgãos identificados nos resultados das tipologias eram, de fato, recorrentes no histórico de alterações relativas a atos de pessoal. Portanto, as principais descobertas não podem ser

consideradas novas. O caráter inovador pode ser aplicado à forma de trabalho proposta pelo modelo, com funcionalidades que podem propiciar ganho de produção com a automatização, por meio de tipologias, do monitoramento de deliberações versando sobre atos de pessoal acerca de cessação de pagamento. Infere-se, então, que a terceira questão também tenha sido atendida.

Além disso, foram registrados os seguintes comentários acerca do trabalho ou oportunidade de melhoria para trabalhos futuros.

Primeiro: Em ambas as ocasiões foi sugerida a transformação do modelo proposto em solução implantada no ambiente de produção, o que poderia ser feito por meio de projeto, a ser previsto como trabalho futuro.

Segundo: Conforme já mencionado, os órgãos apontados como recorrentes, dentro do escopo de resultados obtidos na mineração de dados, foram confirmados pela Unidade Técnica, o que reforça não somente a avaliação com objetivos de negócio, mas também a validação do modelo de mineração de dados.

Terceiro: Por ocasião da primeira agenda, foi declarado que apenas a clusterização e a classificação entre deliberações versando sobre pagamento e não pagamento representariam potencial de ganho na capacidade de produção. Isso reforça a suficiência na definição de escopo e dos resultados obtidos pelo projeto.

Quarto: Também por ocasião da primeira agenda, foi sugerida a necessidade de aprimorar a padronização de modelos textuais adotados na unidade técnica, a fim de estimular a uniformização de textos de deliberações versando sobre atos de pessoal, de forma a facilitar o esforço de automatização, por meio de mineração textual. Adicionalmente, foi sugerida também a separação de apreciações versando sobre legalidade e ilegalidade, para que possam ser publicados em acórdãos distintos, também com vistas a facilitar a mineração textual.

Quinto: Foi sugerida a possibilidade de registro de rubricas, pela unidade técnica, para assistir o esforço de mineração textual, viabilizar o cruzamento de dados no SIAPE, a partir do código correspondente, e facilitar o levantamento de novas tipologias.

Sexto: Foi compartilhada preocupação acerca de possível tarja ou mascaramento de CPF, a qual apenas sob o ponto de vista de funcionalidade, poderia afetar a mineração de dados oriundos da busca textual de acórdãos.

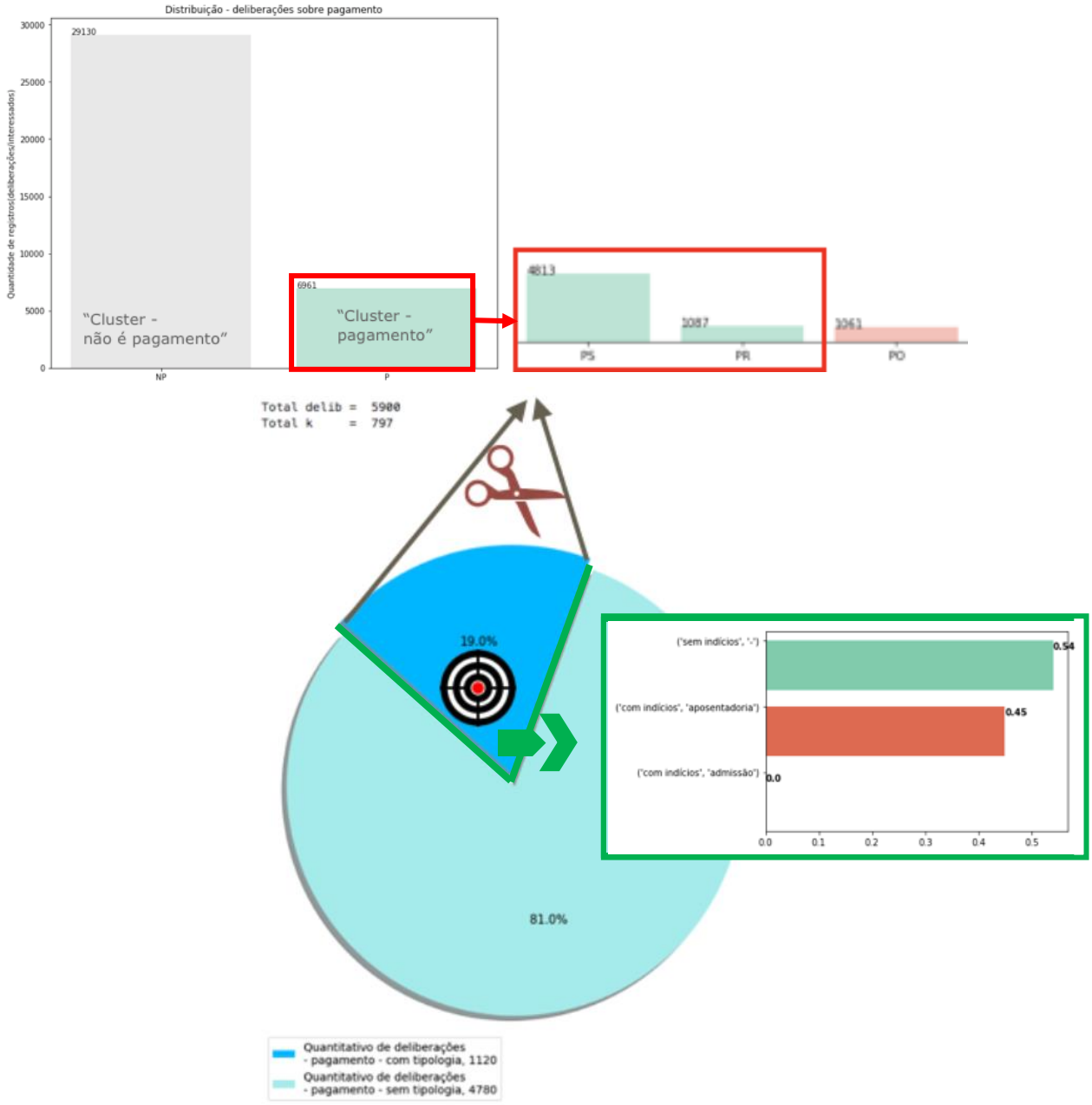
Diante do exposto, conclui-se que o modelo atende critérios de sucesso de negócio, justificados pela realização de duas agendas envolvendo diferentes escalões do Controle Externo. Em ambos os encontros, as avaliações foram positivas.

9.3. Resultados consolidados

Conforme sinalizado na conclusão da fase anterior, optou-se por apresentar os resultados consolidados pelas tipologias propostas, juntamente com informações decorrentes de análise exploratória, na fase de avaliação.

Os resultados abrangem os consolidados nas tipologias apresentas, totalizando a seleção de 1.120 deliberações, do total de 5.900 classificadas como versando sobre pagamento (PR, PS), conforme Figura 60.

Figura 60: Resultados consolidados



Fonte: Elaborada pelo autor (2019).

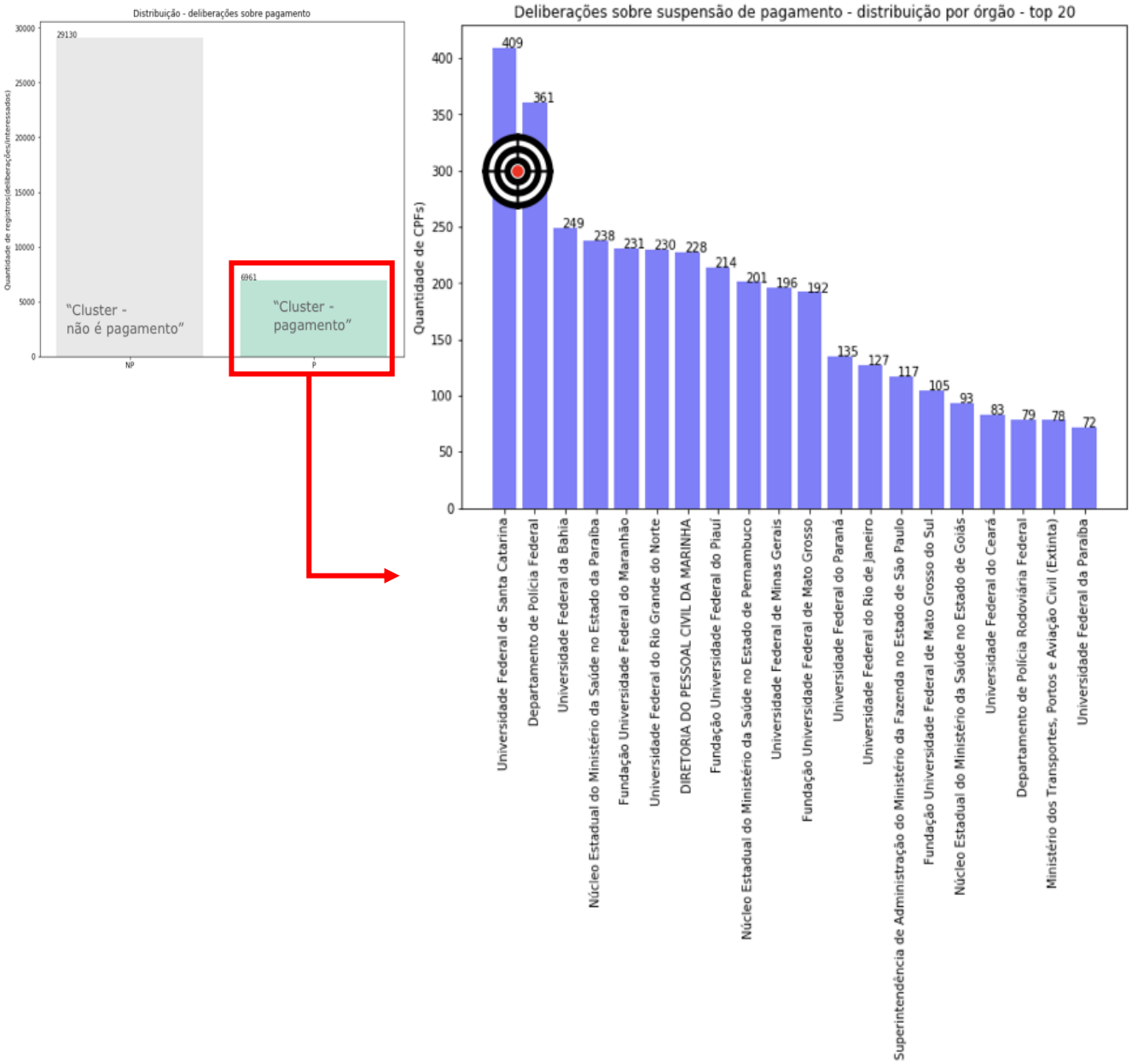
As tipologias propostas abrangeram 19% do domínio de deliberações rotuladas como sendo acerca de cessação de pagamento, das quais 45% foram selecionadas como indícios para procedimento investigativo. Os demais (55%) também podem ser aproveitados para minimizar o estoque de deliberações a serem monitoradas, uma vez terem sido conferidas pelo modelo proposto.

Cumpre ressaltar novamente a prevalência de deliberações genéricas no espaço remanescente, consubstanciando-se em desafio a implementação de novas tipologias capazes de propiciar a adequada identificação de interessados, a ser proposto como trabalho futuro, como continuidade à seleção das demais deliberações versando sobre cessação de pagamento. Seja pelo levantamento de novas tipologias a partir da análise por deliberação, seja pela possibilidade de análise de contexto. Nesse caso, podendo haver custo adicional de implementação pela revisão de arquitetura do modelo proposto.

As informações decorrentes de análise exploratória possuem o intuito de igualmente facilitar o levantamento de indícios e direcionar o planejamento de ações de controle. A Figura 61 apresenta distribuição de deliberações sobre suspensão de pagamento por órgão. Em destaque, as vinte maiores prevalências. Observa-se que as duas primeiras posições se destacam em relação aos demais órgãos, o que pode facilitar a concentração de esforços para minimização de estoque de objetos de monitoramento.

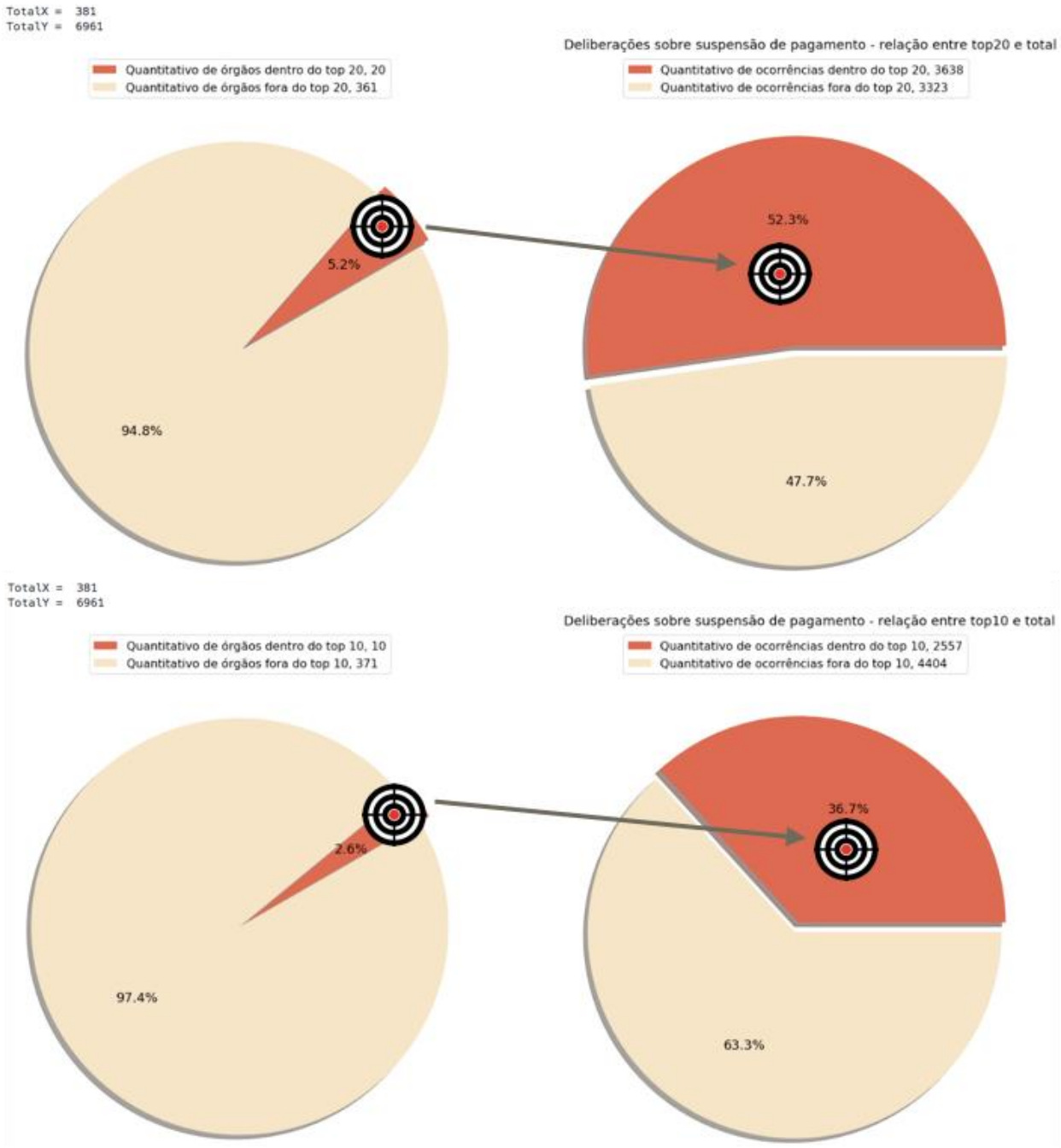
Ainda com vistas a facilitar o planejamento de ações de controle, outra análise encontra-se na Figura 62. É apresentado comparativo entre órgãos recorrentes e deliberações sobre pagamento. Observa-se que os dez órgãos mais prevalentes representam cerca de 3% do universo e concentram cerca de 1/3 das deliberações sobre cessação de pagamento. Ou seja, top 10(~3%)= ~1/3 dos casos. Ao ampliar o domínio de órgãos para 20, ou seja, para cerca de 5% do domínio, a concentração de deliberações ultrapassa a metade do domínio de deliberações (52,3%). Em suma, top 20(~5%)=~1/2 dos casos. O diagnóstico acerca da aglomeração de casos mais recorrentes pode consubstanciar-se em alvos compensadores para esforços concentrados de monitoramento e, até mesmo, representar ganho de escala na melhoria de alocação de tarefas de similar natureza às equipes de fiscalização.

Figura 61: Análise - distribuição de deliberações versando sobre pagamento por órgão



Fonte: Elaborada pelo autor (2019).

Figura 62: Análise – comparativo entre órgãos recorrentes e deliberações sobre pagamento



Fonte: Elaborada pelo autor (2019).

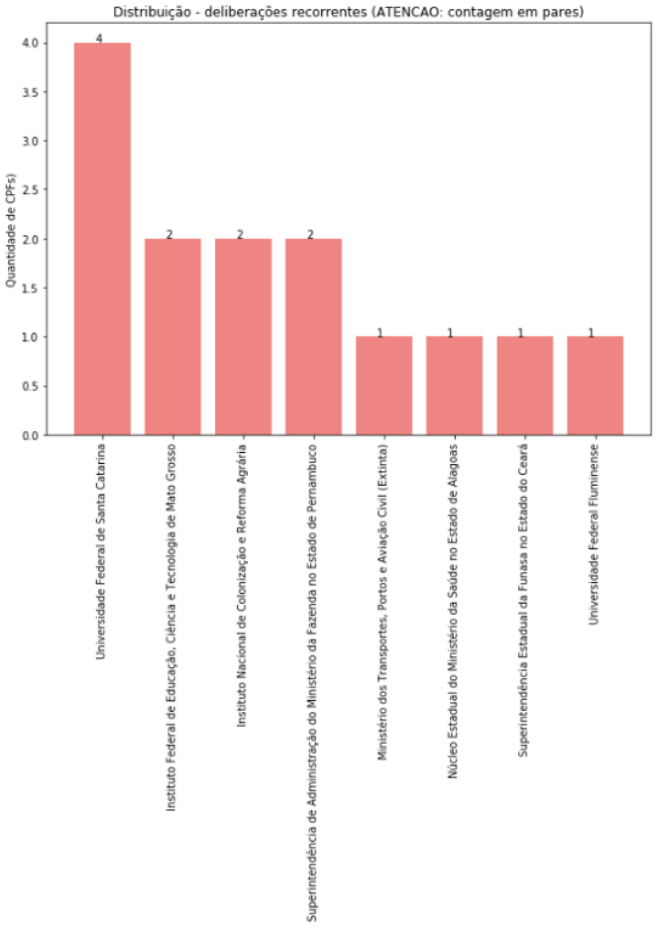
Outra análise encontra-se na Figura 63, versando sobre recorrência de deliberações versando sobre suspensão de pagamento a mesmos interessados. Trata-se de descoberta ocorrida durante as diversas iterações do modelo. Observou-se 07 casos de interessados relacionados a deliberações acerca de cessação de pagamento registrados em acórdãos distintos.

Figura 63: Recorrência de deliberações sobre suspensão de pagamento a mesmos interessados

	cpf	nome	tipo_ato	enc	data	prazo	orgao	num_processo	ano_processo	num_acordao	ano_acordao	apreciador
0	01275372791	José Fernandes Senna	APOSENTADORIA	9.3.2. fazer cessar, no prazo de 15 (quinze) d...	2014-02-04	15 DIAS	Universidade Federal Fluminense	22980	2010	266	2014	PRIMEIRA CÂMARA
1	01275372791	José Fernandes Senna	APOSENTADORIA	9.3. determinar ao Ministério dos Transportes,...	2018-06-26	120 DIAS	Ministério dos Transportes, Portos e Aviação C...	17553	2011	5110	2018	SEGUNDA CÂMARA
2	12324337487	José Campos Dias	APOSENTADORIA	9.3.2. fazer cessar, no prazo de 15 (quinze) d...	2017-10-10	15 DIAS	Superintendência de Administração do Ministéri...	24244	2017	9212	2017	SEGUNDA CÂMARA
3	12324337487	José Campos Dias	APOSENTADORIA	9.3.2. fazer cessar, no prazo de 15 (quinze) d...	2018-07-24	15 DIAS	Superintendência de Administração do Ministéri...	15529	2018	6333	2018	SEGUNDA CÂMARA
4	00365840378	Clovis Acario Maciel	APOSENTADORIA	9.3. determinar ao Instituto Nacional de Colon...	2014-11-11	15 DIAS	Instituto Nacional de Colonização e Reforma Ag...	24015	2014	7158	2014	PRIMEIRA CÂMARA
5	00365840378	Clovis Acario Maciel	APOSENTADORIA	9.3. determinar ao Instituto Nacional de Colon...	2014-07-15	15 DIAS	Instituto Nacional de Colonização e Reforma Ag...	12319	2014	3942	2014	PRIMEIRA CÂMARA
6	43298613991	Newton Valladão Panizzi	APOSENTADORIA	9.3. determinar à Universidade Federal de Sant...	2014-11-11	15 DIAS	Universidade Federal de Santa Catarina	24268	2014	7125	2014	PRIMEIRA CÂMARA
7	43298613991	Newton Valladão Panizzi	APOSENTADORIA	1.7.3. faça cessar os pagamentos decorrentes d...	2014-05-27	15 DIAS	Universidade Federal de Santa Catarina	6627	2014	2170	2014	PRIMEIRA CÂMARA
8	07898282434	Genesivan Bonaparte Santos	APOSENTADORIA	9.3. determinar ao Núcleo Estadual do Ministér...	2014-11-11	15 DIAS	Núcleo Estadual do Ministério da Saúde no Esta...	27676	2014	7137	2014	PRIMEIRA CÂMARA
9	07898282434	Genesivan Bonaparte Santos	APOSENTADORIA	9.3.2. abstenha-se de realizar pagamentos com ...	2014-12-02	15 DIAS	Superintendência Estadual da Funasa no Estado ...	24049	2014	7869	2014	PRIMEIRA CÂMARA
10	22248374020	Nelson Blank	APOSENTADORIA	9.3. determinar à Universidade Federal de Sant...	2014-11-04	15 DIAS	Universidade Federal de Santa Catarina	24260	2014	6970	2014	PRIMEIRA CÂMARA
11	22248374020	Nelson Blank	APOSENTADORIA	1.7.3. faça cessar os pagamentos decorrentes d...	2014-05-27	15 DIAS	Universidade Federal de Santa Catarina	6651	2014	2173	2014	PRIMEIRA CÂMARA
12	05189985253	Divanir Gonçalves da Costa	APOSENTADORIA	1.7. Determinar ao Instituto Federal de Educaç...	2019-02-19	15 DIAS	Instituto Federal de Educação, Ciência e Tecno...	33207	2018	1440	2019	PRIMEIRA CÂMARA
13	05189985253	Divanir Gonçalves da Costa	APOSENTADORIA	9.3. determinar ao Instituto Federal de Educaç...	2016-11-08	15 DIAS	Instituto Federal de Educação, Ciência e Tecno...	20189	2016	11869	2016	SEGUNDA CÂMARA

Fonte: Elaborada pelo autor (2019).

Figura 64: Recorrência de deliberações sobre suspensão de pagamento a mesmos interessados(2)



Fonte: Elaborada pelo autor (2019).

Na primeira apresentação realizada à área de negócio (7/11/19) foi aventada a possibilidade de terem sido motivados por atos distintos de um mesmo interessado, cabendo verificação. A Figura 64 apresenta distribuição alternativa por órgão, evidenciando concentração de casos no mesmo órgão destacado nos resultados consolidados.

Diante do exposto, foram apresentados resultados consolidados a partir das tipologias propostas e informações decorrentes de análise exploratória, com o intuito de levantar indícios para posterior investigação e direcionar o planejamento de ações de controle, particularmente ao monitoramento de deliberações de atos de pessoal versando sobre cessação de pagamento.

9.4. Próximos passos – trabalhos futuros

Como próximos passos, são enumeradas as seguintes propostas de trabalhos futuros e de oportunidades de melhoria. Boa parte já foi mencionada ao longo dos itens anteriores, e consolidada no presente espaço.

Propostas de trabalhos futuros:

- Projeto: Transformação do modelo proposto em solução a ser implantada no ambiente de produção, o que poderia ser feito por meio de projeto, conforme sugerido em reuniões com realizadas com áreas de negócio;
- Denominação: Nomeação de projeto seguindo práticas de uso relacionado a nomes femininos para soluções baseadas em aprendizagem de máquina no TCU. Uma sugestão é AMANDA (**A**utomatização de **Mon**itoramento de **D**eliberações de **A**tos);
- Padronização textual: Conforme registrado na primeira reunião realizada com a área de negócio, foi sugerida a necessidade de aprimorar padronização de modelos textuais adotados na unidade técnica, a fim de estimular a uniformização de textos de deliberações versando sobre atos de pessoal, para facilitar o esforço de automatização, por meio de mineração textual. Adicionalmente, foi sugerida também a separação de apreciações versando sobre legalidade e ilegalidade, para que possam ser publicados em acórdãos distintos, também com vistas a facilitar a mineração textual;
- Novas tipologias: Previsão de novas tipologias para minimização de estoque de deliberações rotuladas como sendo de cessação de pagamentos, a partir da arquitetura baseada em análise por deliberação, adotada no modelo;
- Análise contínua: Implementação de funcionalidades que permitam a atualização contínua de deliberações e das bases de dados envolvidas;

- Interação: Implementação de funcionalidades que permitam interação com usuário, permitindo, a possibilidade de dar baixa em itens já monitorados e retificar ajustes em resultados obtidos em aprendizagem de máquina, como rótulos;
- Rubricas: Conforme registrado na primeira reunião realizada com a área de negócio, foi sugerida a possibilidade de registro de rubricas, pela unidade técnica, para assistir o esforço de mineração textual, viabilizar o cruzamento de dados no SIAPE, a partir do código correspondente, e facilitar o levantamento de novas tipologias;
- Data de comunicação: Ajuste do modelo para que o prazo de início seja contado a partir da data de comunicação ao órgão jurisdicionado, dispensando o ajuste realizado em função dos dados disponibilizados, a exemplo da data de publicação do Acordão;
- Clusterização: Automatização de procedimento de ajuste de rótulos de elementos de um mesmo *cluster*, com a possiblidade de reclassificação quando a proporção majoritária for maior que determinado percentual limiar mínimo de aceitação, por exemplo 90%;
- NLTK: Avaliação de versões em português, se houver, de biblioteca de linguagem natural para reconhecimento de entidades mencionadas (NER)(NLTK PROJECT, 2019). Em particular, com vistas à melhoria da pontuação para nomes contendo termos, tais como "de", "da", "do", a exemplo de "Maria Silva Sousa do Carmo". Observou-se tendência de obtenção de parte do nome para verificação de similaridade. Mesmo com a tendência de parte do nome assegurar *match* na comparação com o nome completo, é um caminho incremental para aprimoramento do modelo;
- *Parser*: Extensão da capacidade de *parser* de dados, particularmente para extração de deliberações. Sabe-se que nem todas as deliberações são cadastradas no RADAR. Se houver a devida implementação dos mesmos critérios de registro de itens, particularmente os passíveis de monitoramento, pode-se aventar a dispensa da referida base de dados;
- Classificador: Comparação com outros algoritmos de aprendizagem supervisionada destinados a resolver problemas de classificação, a exemplo de kNN (*K Nearest Classifier*) e árvore de decisão (ZAKI e MEIRA JR., 2014), porquanto a comparação não pôde ser realizada em função do dimensionamento do cronograma do projeto; e
- Análise de contexto: Estudar a viabilidade de melhoria da capacidade de análise de deliberações, ora feita por item de deliberação para contextual. Para isso, será necessário estudar mudanças na arquitetura necessárias para a implementação da funcionalidade. Pode demandar custo adicional pela possibilidade de demanda de alterações mais profundas no desenho do modelo. Sendo assim, convém avaliar a conveniência e

oportunidade de percorrer o caminho ou buscar o levantamento de novas tipologias a partir de rotas alternativas, em função da prevalência de deliberações genéricas remanescentes.

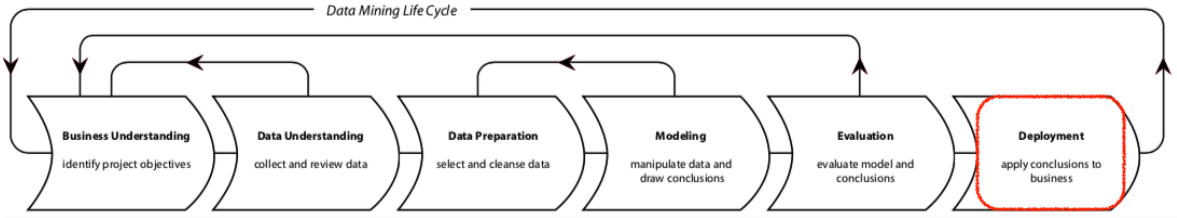
Também foram elencadas oportunidades de melhoria, não somente para o escopo do presente trabalho, mas também para melhoria de outros processos da unidade técnica de pessoal:

- Classificador de processos: Proposição de classificador de processos relacionados a apreciação de atos. Na verdade, era uma das propostas inicialmente aventadas para monografia. Durante a fase de definição de proposta de monografia, foram realizadas entrevistas com servidores da unidade para entendimento do negócio e levantamento e necessidades. Um problema levantado foi relacionado à distribuição de processos de apreciação de atos para as equipes de auditoria. A distribuição é feita a partir de vários critérios, como temporalidade, materialidade, recorrência. Contudo, é feita manualmente. Foi mencionado também é desejável a distribuição de processos afins para mesmas pessoas, para possibilitar ganho de escala. Sendo assim, a implementação de um classificador poderia facilitar a identificação e priorização de processos e tornar mais eficiente a produção na unidade; e
- Análise comportamental sobre folha. Trata-se de outra proposta externa ao escopo do trabalho. Foi aventada durante a fase de modelagem. O objetivo é estabelecer nova frente de trabalho, mais próxima da linha de frente. Ou seja, implementar recursos de aprendizagem de máquina para analisar lançamentos em folha de pagamento, inclusive rubricas variáveis. A implementação poderia antecipar a identificação de problemas e, até mesmo, minimizar o fluxo de entrada de processos na unidade.

10. IMPLANTAÇÃO

A **fase** de **implantação** compreende o plano de implantação e de sustentação da solução, juntamente com relatório final contendo resultados obtidos. Revisão do projeto, como o registro da experiência da jornada. Adicionalmente às fases supracitadas, cumpre ressaltar também a necessidade de levantamento bibliográfico e estudo da própria metodologia, bem como das técnicas de mineração de dados envolvidas.

Figura 65: Ciclo de vida - mineração de dados – Implantação



Fonte: (LEAPER, 2009).

O plano de implantação compreende desenho de projeto para transformação do protótipo apresentado em solução corporativa, a ser implantada no ambiente de produção do Tribunal. O objetivo é incluir serviços capazes de permitir interação com equipes da unidade técnica de pessoal. Entre as funcionalidades, a possibilidade de dar baixa em itens já monitorados e retificar ajustes em resultados obtidos em aprendizagem de máquina, como rótulos, como parte integrante da melhoria do próprio modelo.

A proposta de projeto para transformação do protótipo apresentado em solução corporativa já foi encaminhado para a Sefip para análise.

O plano de sustentação compreende o ciclo de evolução incremental e iterativo, particularmente para aprimoramento do desempenho, da acurácia e da identificação de novas tipologias.

O relatório final é o arcabouço de conhecimento documentado do projeto. A começar pela presente monografia, bem como cadernos de codificação, apresentações feitas à área de negócio e à banca examinadora. Toda a documentação será disponibilizada para consulta em repositório e momento oportunos, seguindo critérios de classificação da informação quanto à confidencialidade estabelecidos por normativos de segurança no Tribunal.

Por fim, experiência relacionada à jornada do projeto pode ser resumida em poucas palavras: extremamente gratificante e enriquecedora. Em particular, acerca da oportunidade de encarar desafio para contribuir diretamente para alcance de resultados de fiscalização do Tribunal, mesmo com pouco tempo para a execução do projeto. Espera-se que o protótipo possa, de fato, transformar-se em solução implantada no ambiente de produção e sirva de força motriz disruptiva para redução de estoques de monitoramento de deliberações versando sobre pagamentos acerca apreciações de atos de pessoal.

11. CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo propor protótipo de modelo capaz de realizar o monitoramento de deliberações de atos de admissão e de concessão de aposentadoria e reforma de pessoal, por meio de tipologias relacionadas a determinações sobre cessação de pagamento. Também são apresentadas informações decorrentes de análise exploratória. Para ambos os casos, os resultados visam o levantamento de indícios e direcionar o planejamento de ações de controle.

Foram obtidos os seguintes resultados: 1.120 deliberações foram selecionadas por meio de tipologia, do total de 5.900 classificadas como versando sobre pagamento (PR e PS). As tipologias propostas foram capazes de abranger 19% do domínio de deliberações rotuladas como sendo acerca de cessação de pagamento, das quais 45% foram selecionadas como indícios para procedimento investigativo. Os demais serviram para minimizar o estoque de monitoramentos pendentes, uma vez terem sido conferidas pelo modelo proposto.

Também foram apresentados resultados relacionados a informações decorrentes de análise exploratória. Verificou-se que a distribuição de deliberações sobre suspensão de pagamento aponta que os 10 órgãos prevalentes, os quais representam cerca de 3% do domínio analisado, concentram cerca de 1/3 das deliberações sobre cessação de pagamento. Ao ampliar o domínio para 20 (5% do domínio), a concentração de deliberações ultrapassa a metade do escopo de deliberações (52,3%). O diagnóstico acerca da aglomeração de casos mais recorrentes pode consubstanciar-se em alvos compensadores para ações de controle e, até mesmo, representar ganho de escala na melhoria de alocação de tarefas de similar natureza às equipes de fiscalização.

O modelo atendeu critérios de sucesso de mineração de dados. A acurácia obtida no classificador foi de 93,89%. A precisão de resultados por categorias de interesse (P*), variou entre 96% e 97%. Além disso, para o algoritmo de clusterização, o percentual de elementos em *clusters* com divergências de rotulação pode ser empregado para validar o algoritmo não supervisionado, uma vez que o classificador é executado por deliberação e não por *clusters* e a expectativa era de que todos os elementos de um mesmo *cluster* fossem rotulados com a mesma classe. Por fim, o emprego de *oversampling* mostrou-se vantajoso, reduzindo o percentual de divergências de 1,6513% (NB-OS), para 0,2299% (NB+OS) e para 0,1357% após ajuste manual.

O modelo atendeu critérios de sucesso de negócio, justificados pela realização de duas agendas envolvendo diferentes escalões da Secretaria Geral de Controle Externo. Em ambos os encontros, as avaliações foram positivas. Isso assegurou tranquilidade de estar no caminho certo. Ao mesmo, reforçou a responsabilidade de melhorar a qualidade do trabalho. Foram experiências das quais recomenda-se fortemente a recorrência em demais trabalhos promovidos pelo ISC/TCU, como boa prática para acompanhamento e validação de trabalhos junto à área de negócio, não se limitando à banca acadêmica. Sendo assim, externa-se gratidão pelas oportunidades.

Ainda sobre resultados, durante as reuniões com a área de negócio, foram confirmados resultados relacionados a órgão recorrentes, identificados nas tipologias e nas análises. Também foi declarado que apenas a clusterização e a classificação entre deliberações versando sobre

pagamento e não pagamento já seria suficiente para representar potencial de ganho na capacidade de produção. Isso reforçou a suficiência na definição de escopo e dos resultados obtidos pelo projeto.

Embora o tempo do projeto tenha limitado o levantamento de novas tipologias, foram consideradas satisfatórias do ponto de vista de critérios de sucesso do negócio e da validação da modelagem de mineração de dados.

Foi registrado óbice relacionado a prevalência de deliberações genéricas no espaço remanescente, consubstanciando-se em desafio a implementação de novas tipologias capazes de propiciar a adequada identificação de interessados.

Há muito para ser feito ainda. Entre as avaliações recebidas da área de negócio, ressalta-se a necessidade de propor projeto para transformação do protótipo acadêmico em solução. Necessidade de aprimorar a padronização textual para facilitar a mineração de dados. Proposição de novas tipologias. Também foram elencadas sugestões para trabalhos futuros e sugestões de trabalhos para atendimento de outras demandas da unidade de pessoal, relacionadas questões levantadas durante o trabalho realizado.

Por fim, a finalização da jornada acadêmica pode apenas representar o início de uma jornada ainda maior. Desafiadora, gratificante e benéfica.

REFERÊNCIAS

ALTO, Valentina. **Unsupervised Learning: K-means vs Hierarchical Clustering**. Disponível em: <https://towardsdatascience.com/unsupervised-learning-k-means-vs-hierarchical-clustering-5fe2da7c9554>. Acesso em: 20 set 2019.

BALANIUK, Remis. **Gestão do Conhecimento no TCU**. . [S.l: s.n.]. Disponível em: <https://portal.tcu.gov.br/data/files/BF/80/06/85/2C75D410F10055D41A2818A8/1652934.PPT>. Acesso em: 16 jan 2020. , 22 Nov 2010

BHANDARI, Shivansh. **Improving the extraction of human names with NLTK**. Disponível em: <https://stackoverflow.com/questions/20290870/improving-the-extraction-of-human-names-with-nltk>. Acesso em: 1 out 2019.

BOEHMKE, Bradley e GREENWELL, Brandon. **Hands-On Machine Learning with R**. Disponível em: <https://bradleyboehmke.github.io/HOML/hierarchical.html>. Acesso em: 21 nov 2019.

BRANCO, Cláudio Souza Castello. **Histórico sobre a obtenção e o tratamento de dados para o Controle Externo no TCU, de 1995 a 2014**. Revista do TCU, v. 131, Dez 2014. Disponível em: <https://revista.tcu.gov.br/ojs/index.php/RTCU/article/view/58>.

BRASIL. **Constituição da República Federativa do Brasil de 1988. Promulgada em 05 de outubro de 1988**. Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm>. Acesso em: 15 ago 2019.

BRASIL. **Lei nº 8.112, de 11 de dezembro de 1990, que dispõe sobre o regime jurídico dos servidores públicos civis da União, das autarquias e das fundações públicas federais**. , 11 Dez 1990. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/l8112cons.htm>. Acesso em: 20 ago 2019.

DATAMAN, _. **Using Over-Sampling Techniques for Extremely Imbalanced Data**. Disponível em: <https://towardsdatascience.com/sampling-techniques-for-extremely-imbalanced-data-part-ii-over-sampling-d61b43bc4879>. Acesso em: 1 nov 2019a.

DATAMAN, _. **Using Under-Sampling Techniques for Extremely Imbalanced Data**. Disponível em: <https://towardsdatascience.com/sampling-techniques-for-extremely-imbalanced-data-part-i-under-sampling-a8dbc3d8d6d8>. Acesso em: 27 nov 2019b.

DERNONCOURT, Franck. **Duplicating training examples to handle class imbalance**

in a pandas data frame. Disponível em: <https://stackoverflow.com/questions/48373088/duplicating-training-examples-to-handle-class-imbalance-in-a-pandas-data-frame>. Acesso em: 1 nov 2019.

DONTHA, Ramesh. **What I Always Wanted To Know About Big Data**. Disponível em: <https://www.linkedin.com/pulse/what-i-always-wanted-know-big-data-afraid-ask-ramesh-dontha?trk=mp-author-card>. Acesso em: 30 nov 2019.

ESTEVES, Hugo. **Data Science What?** Disponível em: <https://towardsdatascience.com/data-science-what-ae9b5c7ffc22>. Acesso em: 1 dez 2019.

GANGANE, Tejali. **WW(What and When) of Data Science**. Disponível em: <https://towardsdatascience.com/ww-what-and-when-of-data-science-dc4cc0bfd7b>. Acesso em: 1 dez 2019.

IBM. **IBM® SPSS® Modeler CRISP-DM (Cross-Industry Standard Process for Data Mining) Guide**. . [S.l: s.n.]. Disponível em: <https://www.ibm.com/support/knowledgecenter/en/SS3RA7_15.0.0/com.ibm.spss.crispdm.help/crisp_overview.htm>. , 2014

IRANI, Jasmine. **Clustering Techniques and the Similarity Measures used in Clustering: A Survey**. International Journal of Computer Applications, Jan 2016. Volume 134, number 7Disponível em: <https://www.ijcaonline.org/research/volume134/number7/irani-2016-ijca-907841.pdf>. Acesso em: 20 set 2019.

JAIN, Pawan. **Hierarchical clustering Clearly Explained**. Disponível em: <https://towardsdatascience.com/https-towardsdatascience-com-hierarchical-clustering-6f3c98c9d0ca>. Acesso em: 1 set 2019.

LEAPER, Nicole. **A Visual Guide to CRISP-DM Methodology**. . [S.l: s.n.]. Disponível em: <http://www.crisp-dm.org/download.htm>. Acesso em: 16 ago 2019. , 2009

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. 170 f. Monografia – Instituto Serzedello Corrêa/Tribunal de Contas da União, 2004. Disponível em: <https://portal.tcu.gov.br/biblioteca-digital/qualidade-no-tribunal-de-contas-da-uniao-praticas-adotadas-pela-secretaria-de-fiscalizacao-de-pessoal-para-controlar-os-gastos-de-pessoal.htm>. Acesso em: 7 set 2019.

MACQUEEN, J. **Some methods for classification and analysis of multivariate observations.** Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and

Probability, v. 1, p. 281–297, 1968.

MPOG. **SIAPE, Sistema Integrado de Administração de Pessoal**. Disponível em: <http://www.siapenet.gov.br/Portal/Servico/Apresentacao.asp>. Acesso em: 20 ago 2019.

NLTK PROJECT. **Natural Language Toolkit (NLTK)**. Disponível em: <http://www.nltk.org>. Acesso em: 20 set 2019.

OPPERMANN, Artem. **Artificial Intelligence vs. Machine Learning vs. Deep Learning**. Disponível em: <https://towardsdatascience.com/artificial-intelligence-vs-machine-learning-vs-deep-learning-2210ba8cc4ac>. Acesso em: 30 nov 2019.

PANDIT, Shraddha e GUPTA, Suchita. **A Comparative Study on Distance Measuring Approaches for Clustering**. International Journal of Research in Computer Science, White Globe Publications, 2011. Volume 2, Issue 1, p. 29–31.

PEART, Andy. **Homage to John McCarthy, the Father of Artificial Intelligence (AI)**. Disponível em: <https://www.artificial-solutions.com/blog/homage-to-john-mccarthy-the-father-of-artificial-intelligence>. Acesso em: 30 nov 2019.

PEDREGOSA, F. Et all. **Scikit-learn: Machine Learning in Python**. Journal of Machine Learning Research, 2011. volume 12, p. 2825–2830.

PORTAL TCU. **Notícia: Implementação do e-Pessoal chega à reta final**. Disponível em: <https://portal.tcu.gov.br/imprensa/noticias/implementacao-do-e-pessoal-chega-a-reta-final.htm>. Acesso em: 30 ago 2019.

PROJECT JUPYTER. **Project Jupyter®**. Disponível em: <https://jupyter.org>. Acesso em: 1 jun 2018.

PUGET, Jean Francois. **What Is Machine Learning?** Disponível em: <https://www.ibm.com/developerworks/community/blogs/jfp/entry/What_Is_Machine_Learning?lang=en>. Acesso em: 30 nov 2019.

PYTHON, Software Foundation. **Python**. Disponível em: <https://www.python.org/>. Acesso em: 1 jun 2018.

RIBEIRO, Andrea. **Atos de Pessoal – e-Pessoal (apresentação para Gabinetes e MPJTCU)**. . [S.l: s.n.]. . Acesso em: 7 set 2019. , 15 Mar 2017

SEFIP. **Portaria-Sefip nº 1, de 10 de junho de 2019, que dispõe sobre a organização interna, as competências e atividades da Secretaria de Fiscalização de Pessoal (Sefip)**. , 10 Jun 2019. Disponível em: <www.tcu.gov.br>.

SEPROC. **Manual de Procedimento - Secinf (minuta)**. . [S.l: s.n.]. . Acesso em: 2 set 2019a. , 2019

SEPROC. **Projeto RADEX (minuta)**. . [S.l: s.n.]. . Acesso em: 3 set 2019b. , 2019

SOMA, Jonathan. **Fuzzing matching in pandas with fuzzywuzzy**. Disponível em: <http://jonathansoma.com/lede/algorithms-2017/classes/fuzziness-matplotlib/fuzzing-matching-in-pandas-with-fuzzywuzzy/>. Acesso em: 1 out 2019.

SONI, Devin. **Supervised vs. Unsupervised Learning**. Disponível em: <https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d>. Acesso em: 1 out 2019.

STI. **Modelo de dados - ambiente de produção TCU**. Disponível em: <http://srv-pdesigner1a:3030/cmr/pages/loginFormProfile.faces>. Acesso em: 15 ago 2019a.

STI. **Modelo de dados - ambiente de produção TCU - RADAR**. Disponível em: <http://srv-%20svn.tcu.gov.br/disin/DocumentacaoAD/Diagramas/Arquivos_htm/dados.htm>. Acesso em: 15 ago 2019b.

TCU. **Busca Textual de Acórdãos**. Disponível em: <https://pesquisa.apps.tcu.gov.br/#/pesquisa/acordao-completo>. Acesso em: 16 ago 2019a.

TCU. **e-Pessoal**. Disponível em: <https://portal.tcu.gov.br/pessoal/>. Acesso em: 15 ago 2019b.

TCU. **Relatório Anual de Atividades do TCU:2018**. . [S.l: s.n.], 2019c. Disponível em: <https://portal.tcu.gov.br/transparencia/relatorios/relatorios-de-atividades/relatorios-de-atividades.htm>.

TCU. **Relatório de Acompanhamento - TC 024.000/2018-3**. . [S.l: s.n.]. Disponível em: <https://pesquisa.apps.tcu.gov.br/#/documento/processo/024.000%252F2018-3/%20/ANO%20desc,%20NUMEROPROCESSOCOMZEROS%20desc/1/%20>. Acesso em: 20 ago 2019d. , 29 Mar 2019

TCU. **Resolução nº 255, de 26 de setembro de 1991, que 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.** , 26 Nov 1991. Disponível em: <www.tcu.gov.br>. Acesso em: 3 set 1991.

TCU. **Resolução-TCU n° 305, de 28 de dezembro de 2018, que define a estrutura, as competências e a distribuição das funções de confiança das unidades da Secretaria do Tribunal de Contas da União**. , 28 Dez 2018. Disponível em:

<https://portal.tcu.gov.br/biblioteca-digital/resolucao-305-2018.htm>. Acesso em: 20 ago 2019.

TCU. Resolução-TCU número 246. . **Resolução-TCU nº 246, de 30 de novembro de 2011, que altera o Regimento Interno do Tribunal de Contas da União (RITCU), aprovado pela Resolução TCU nº 155, de 4 de dezembro de 2002, republicado no Boletim do Tribunal de Contas da União Especial, de 02 de janeiro de 2015.** , 30 Nov 2011. Disponível em: <https://portal.tcu.gov.br/normativos/regimentos-internos/>. Acesso em: 15 ago 2019.

TCU. **Sagas**. Disponível em: <https://contas.tcu.gov.br/sagas/Web/Sagas/PaginaInicialSagas.jspx>. Acesso em: 16 ago 2019e.

TCU. **Sisac**. Disponível em: <https://www.google.com/?client=safari>. Acesso em: 15 ago 2019f.

TCU. **SisMonitoramento v.2.1**. Disponível em: <https://contas.tcu.gov.br/ords/f?p=1538:1:2547936168102:::::>. Acesso em: 16 ago 2019g.

THE APACHE SOFTWARE FOUNDATION. **Apache Solr**. Disponível em: <http://lucene.apache.org/solr/>. Acesso em: 16 ago 2019.

THE ECONOMIST. **Fuel of the future**. The Economist, Maio 2017. , p. 14–17.

THE PANDAS PROJECT. **Pandas - Python Data Analysis Library**. Disponível em: <https://pandas.pydata.org>. Acesso em: 5 dez 2019.

THE SCIPY COMMUNITY. **Hierarchical clustering (scipy.cluster.hierarchy) - (scipy.cluster.hierarchy.fcluster)**. Disponível em: <https://docs.scipy.org/doc/scipy/reference/cluster.hierarchy.html>. Acesso em: 20 set 2019.

ZAKI, Mohammed J. e MEIRA JR., Wagner. **Data Mining and Analysis: Fundamental Concepts and Algorithms**. [S.l.]: Cambridge University Press, 2014. Disponível em: <http://www.dataminingbook.info/pmwiki.php/Main/BookDownload>. Acesso em: 1 fev 2019.

ANEXO A – CONSULTAS

O anexo apresenta consultas/códigos-fonte considerados relevantes para a consecução dos trabalhos, seja em linguagem SQL ou *Python*. O objetivo é facilitar o entendimento dos dados coletados e facilitar eventual posterior verificação. Foram utilizados dois ambientes. LabContas e local. Para o local, foram instaladas as seguintes ferramentas: ANACONDA NAVIGATOR v1.9.7, JUPITER LAB v0.35.5, bem com as seguintes bibliotecas:

PackageVersion : absl-py0.7.1, alabaster0.7.10, anaconda-client1.6.14, anaconda-navigator 1.9.7, anaconda-project 0.8.2, appnope0.1.0, appscript1.0.1, asn1crypto 0.24.0, astor0.7.1, astroid1.6.1, astropy2.0.3, attrs17.4.0, Babel2.5.3, backports.shutil-get-terminal-size 1.0.0, basemap1.2.0, beautifulsoup4 4.6.0, bitarray 0.8.1, bkcharts 0.2, blaze0.11.3, bleach 2.1.2, bokeh0.12.13 , boto 2.48.0, Bottleneck 1.2.1, cachetools 3.1.0, certifi2019.3.9, cffi 1.11.4, chardet3.0.4, click6.7, cloudpickle0.5.2, clyent 1.2.2, colorama 0.3.9, conda4.6.14, conda-build3.4.1, conda-verify 2.0.0, contextlib20.5.5, cryptography 2.6.1, cycler 0.10.0, Cython 0.27.3, cytoolz0.9.0, d2l0.9.2, dask 0.16.1, datashape0.5.4, decorator4.2.1, distributed1.20.2, docutils 0.14 , entrypoints0.2.3, et-xmlfile 1.0.1, fastcache1.0.2, filelock 2.0.13, Flask0.12.2, Flask-Cors 3.0.3, gast 0.2.2, geos 0.2.1, gevent 1.2.2, glob20.6, gmpy22.0.8, google-api-python-client 1.7.8, google-auth1.6.3, google-auth-httplib2 0.0.3, graphviz 0.8.4, greenlet 0.4.12, grpcio 1.20.1, h5py 2.7.1, heapdict 1.0.0, html5lib 1.0.1, httplib2 0.12.1, idna 2.6, imageio2.2.0, imagesize0.7.1, ipykernel4.8.0, ipython6.2.1, ipython-genutils 0.2.0, ipywidgets 7.1.1, isort4.2.15, itsdangerous 0.24 , jdcal1.3, jedi 0.11.1, Jinja2 2.10 , jsonschema 2.6.0, jupyter1.0.0, jupyter-client 5.2.2, jupyter-console5.2.0, jupyter-core 4.4.0, jupyterlab 0.35.5, jupyterlab-launcher0.10.2, jupyterlab-server0.2.0, Keras2.2.4, Keras-Applications 1.0.7, Keras-Preprocessing1.0.9, kiwisolver 1.0.1, lazy-object-proxy1.3.1, LinearAlgebra1.4.1, llvmlite 0.21.0, locket 0.2.0, lxml 4.1.1, Markdown 3.1, MarkupSafe 1.0, matplotlib 2.2.2, mccabe 0.6.1, mglearn0.1.7, mistune0.8.3, mock 2.0.0, mpmath 1.0.0, msgpack-python 0.5.1, multipledispatch 0.4.9, mxnet1.4.0, navigator-updater0.1.0, nbconvert5.3.1, nbformat 4.4.0, networkx 2.1, nltk 3.2.5, nose 1.3.7, notebook 5.4.0, numba0.36.2, numexpr2.6.4, numpy1.16.3, numpydoc 0.7.0, odo0.5.1, olefile0.45.1, openpyxl 2.4.10, packaging16.8 , pandas 0.22.0, pandas-profiling 1.4.1, pandocfilters1.4.2, parso0.1.1, partd0.3.8, path.py10.5 , pathlib2 2.3.0, patsy0.5.0, pbr5.2.0, pep8 1.7.1, pexpect4.3.1, pickleshare0.7.4, Pillow 5.0.0, pip19.1 , pkginfo1.4.1, pluggy 0.6.0, ply3.10 , prompt-toolkit 1.0.15, protobuf 3.7.1, psutil 5.4.3, ptyprocess 0.5.2, py 1.5.2, pyasn1 0.4.5, pyasn1-modules 0.2.4, pycodestyle2.3.1, pycosat0.6.3, pycparser2.18 , pycrypto 2.6.1, pycurl 7.43.0.2, pyflakes 1.6.0, Pygments 2.2.0, pylint 1.8.2, pyodbc 4.0.22, pyOpenSSL17.5.0, pyparsing2.2.0, pyproj 1.9.5.1 , pyshp1.2.12, PySocks1.6.7, pystan 2.17.1.0, pytest 3.3.2, python-dateutil2.6.1, pytz 2017.3, PyWavelets 0.5.2, PyYAML 3.12 , pyzmq16.0.3, QtAwesome0.4.4, qtconsole4.3.1, QtPy 1.4.2, requests 2.21.0, rope 0.10.7, rsa4.0, ruamel-yaml0.15.35 , scikit-image 0.15.0, scikit-learn 0.19.1, scipy1.0.0, seaborn0.8.1, selenium 3.14.0, Send2Trash 1.4.2, setuptools 40.4.3, simplegeneric0.8.1, singledispatch 3.4.0.3 , six1.11.0, snowballstemmer1.2.1, sortedcollections0.5.3, sortedcontainers 1.5.9, Sphinx 1.6.6, sphinxcontrib-websupport 1.0.1, SQLAlchemy 1.2.1, statsmodels0.8.0, sympy1.1.1, tables 3.4.2, tblib1.3.2, tensorboard1.13.1, tensorflow 1.13.1, tensorflow-estimator 1.13.0, termcolor1.1.0, terminado0.8.1, testpath 0.3.1, toolz0.9.0, tornado4.5.3, traitlets4.3.2, typing 3.6.2, unicodecsv 0.14.1, Unidecode1.0.22, uritemplate3.0.0, urllib31.22 , watermark1.6.1, wcwidth0.1.7, webencodings 0.5.1, Werkzeug 0.14.1, wheel0.30.0, widgetsnbextension 3.1.0, wkhtmltopdf0.2, wordcloud1.5.0, wrapt1.10.11 , xlrd 1.1.0, XlsxWriter 1.0.2, xlwings0.11.5, xlwt 1.2.0, zict 0.1.3

Código SQL (1)

```
select   vw.cod_processo
      , vw.processo
      , vw.seq_deliberacao
      , VW.TIPO_DELIBERACAO
      , TD.DESCR
      , vw.numdelib
      , to_number(regexp_substr(regexp_substr(vw.numdelib,'\/[0-9]{4}-'),'[0-9]{4}') ) as ANO_ACORDAO
      , to_number(regexp_replace(regexp_replace(regexp_replace(regexp_substr(vw.numdelib,'C-[0-9]*[\.]*[0-9]*[\.]*[0-9]*-'),'C-',''),'-',''),'\.','')) as NUM_ACORDAO
      , vw.descr
      , uot.sigla
      , VW.ANO_PROCESSO
      , VW.NUM_PROCESSO
      , vw.cod_apreciacao
      , VW.COD_TIPO_DECISAO
      , vw.ano_decisao
      , vw.apreciador
      , VW.DATA as data
from VW_ESC_DEL_EFET_COD_APREC vw, processo_gestao pg, unidade_organizacional_tcu uot, TIPO_DELIBERACAO td --, VW_SF_APRECIACAO vsa
where  vw.cod_processo = pg.cod
  and  pg.cod_unid_responsavel_tecnica = 191200
  and  pg.cod_unid_responsavel_tecnica = uot.cod
  -- and  vw.ano_processo >= 2015
  -- and  vw.ano_processo <= 2018
  and  VW.TIPO_DELIBERACAO (+)= TD.COD
  and  to_number(regexp_substr(regexp_substr(vw.numdelib,'\/[0-9]{4}-'),'[0-9]{4}') ) > 2013
  -- fonte https://stackoverflow.com/questions/46343120/regex-lookahead-in-oracle-sql
```

Código SQL (2)

```
-- Tipos de deliberação
select cod, descr from tipo_deliberacao order by cod

COD|DESCR
1|Determinação a Órgão/Entidade
2|Diligência a Órgão/Entidade
3|Conhecim/Provim de Denúncia/Repr/Solic/Consulta
4|Imputação de Débito a Órgão/Entidade
5|Arquivamento de Processo
6|Abertura de Novo Processo / Apartado
7|Modificação da Natureza do Processo
8|Sobrestamento do Julgamento
9|Reabertura de Processo
10|Aplicação da Chancela de Sigiloso
11|Retirada da Chancela de Sigiloso
12|Apensamento do Atual Processo a Outro(s)
13|Desapensamento de Processo do Processo Atual
14|Apensamento de Outro(s) Processo(s) ao Atual
15|Determinação de Providências Internas ao TCU
16|Aplicação de Multa a Responsável
17|Citação de Responsável
18|Fixação de novo prazo de recolhimento
19|Arquivamento por Economia Processual
20|Imputação de Débito a Responsável
21|Aplicação de Outras Sanções (que não multa)
22|Julgamento das contas do Responsável
23|Audiência de Responsável
24|Expedição de Quitação a Responsável
25|Legalidade de Ato de Admissão/Concessão
26|Determinação de Realização de Fiscalização
27|Cancelamento de Fiscalização
28|Expedição de Quitação de Dívida
29|Autorização de Recolhimento Parcelado de Dívida
30|Conhecimento de Recurso
31|Não Conhecimento de Recurso
33|Não Provimento de Recurso
35|Provimento de Recurso
36|Encaminhamento de Cobrança Executiva
37|Aplicação de Medida Cautelar a Responsável
38|Aplicação de Medida Cautelar a Órgão/Entidade
39|Tornar Deliberação Sem Efeito
40|Trancamento de Contas Iliquidáveis
41|Prorrogação de Prazo de Deliberação
42|Julgamento de Estágio de Desestatização
43|Expedição de Quitação a Órgão/Entidade
44|Legalidade de Ato não SISAC
45|Recomendação a Órgão/Entidade
46|Acatar/Rejeitar as Alegações de Defesa
47|Saneamento de Irregularidades Graves
48|Prosseguimento da Execução da Obra
49|Acatar/Rejeitar as Razões de Justificativa
50|Prorrogação de Prazo de Deliberação a responsável
51|Requisição de Serviços Técnicos Especializados
52|Comunicação ao Congresso não cumprimento contrato
53|Diligência a Responsável
56|Dar ciência
57|Inspeção
58|Diligência
59|Oitiva
60|Recebimento como Mera Petição
61|Provimento/Não Provimento de Recurso
63|Conhecimento/Não Conhecimento de Recurso
64|Recebimento como Mera Petição (Inv. Jurídica)
65|Provimento Parcial de Recurso
66|Adotar medida saneadora
67|Negar seguimento
68|Negar seguimento (Inv. Jurídica)
```

ANEXO B – INFORMAÇÕES DECORRENTES DE ANÁLISE EXPLORATÓRIA

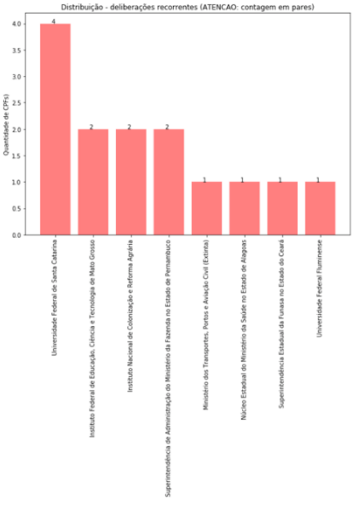
O anexo apresenta outras situações decorrentes de análise exploratória não elencadas na Modelagem. Embora tenha sido demandado maior quantidade de iterações no modelo, foi possível realizar a extração das diferentes situações, pelo aprimoramento do módulo de *parser*.

Descrição	qtd
Recorrência de deliberações versando sobre suspensão de pagamento a mesmos interessados	07
Acórdãos com ausência de CPF na identificação de interessados	34
Acórdãos contendo CPF "0" na identificação de interessados	01
Acórdãos contendo informações diversas na identificação do interessado	02(*)
Acórdãos contendo CPF de 8 dígitos na identificação de interessados	02
Acórdãos "duplicados"	01(*)
Acórdãos contendo registros de interessados separados por "_"	01
Acórdãos com ausência de registro de número ou ano (do Acórdão)	02
Distribuição de Acórdãos de atos x "não atos"	-
Acórdãos contendo registros duplicados de interessados(**)	272
Localização de CPF(**)	-
Localização do tipo de ato no Acórdão(**)	-

(*) Amostragem. (**) Vide Modelagem.

Recorrência de deliberações versando sobre suspensão de pagamento a mesmos interessados:

	cpf	nome	tipo_ato	enc	data	prazo	orgao	num_processo	ano_processo	num_acordao	ano_acordao	apreciador
0	01275372791	José Fernandes Senna	APOSENTADORIA	9.3.2. fazer cessar, no prazo de 15 (quinze) d...	2014-02-04	15 DIAS	Universidade Federal Fluminense	22980	2010	266	2014	PRIMEIRA CÂMARA
1	01275372791	José Fernandes Senna	APOSENTADORIA	9.3. determinar ao Ministério dos Transportes,...	2018-06-26	120 DIAS	Ministério dos Transportes, Portos e Aviação C...	17553	2011	5110	2018	SEGUNDA CÂMARA
2	12324337487	José Campos Dias	APOSENTADORIA	9.3.2. fazer cessar, no prazo de 15 (quinze) d...	2017-10-10	15 DIAS	Superintendência de Administração do Ministéri...	24244	2017	9212	2017	SEGUNDA CÂMARA
3	12324337487	José Campos Dias	APOSENTADORIA	9.3.2. fazer cessar, no prazo de 15 (quinze) d...	2018-07-24	15 DIAS	Superintendência de Administração do Ministéri...	15529	2018	6333	2018	SEGUNDA CÂMARA
4	00365840378	Clovis Acario Maciel	APOSENTADORIA	9.3. determinar ao Instituto Nacional de Colon...	2014-11-11	15 DIAS	Instituto Nacional de Colonização e Reforma Ag...	24015	2014	7158	2014	PRIMEIRA CÂMARA
5	00365840378	Clovis Acario Maciel	APOSENTADORIA	9.3. determinar ao Instituto Nacional de Colon...	2014-07-15	15 DIAS	Instituto Nacional de Colonização e Reforma Ag...	12319	2014	3942	2014	PRIMEIRA CÂMARA
6	43298613991	Newton Valladão Panizzi	APOSENTADORIA	9.3. determinar à Universidade Federal de Sant...	2014-11-11	15 DIAS	Universidade Federal de Santa Catarina	24268	2014	7125	2014	PRIMEIRA CÂMARA
7	43298613991	Newton Valladão Panizzi	APOSENTADORIA	1.7.3. faça cessar os pagamentos decorrentes d...	2014-05-27	15 DIAS	Universidade Federal de Santa Catarina	6627	2014	2170	2014	PRIMEIRA CÂMARA
8	07898282434	Genesivan Bonaparte Santos	APOSENTADORIA	9.3. determinar ao Núcleo Estadual do Ministér...	2014-11-11	15 DIAS	Núcleo Estadual do Ministério da Saúde no Esta...	27676	2014	7137	2014	PRIMEIRA CÂMARA
9	07898282434	Genesivan Bonaparte Santos	APOSENTADORIA	9.3.2. abstenha-se de realizar pagamentos com ...	2014-12-02	15 DIAS	Superintendência Estadual da Funasa no Estado ...	24049	2014	7869	2014	PRIMEIRA CÂMARA
10	22248374020	Nelson Blank	APOSENTADORIA	9.3. determinar à Universidade Federal de Sant...	2014-11-04	15 DIAS	Universidade Federal de Santa Catarina	24260	2014	6970	2014	PRIMEIRA CÂMARA
11	22248374020	Nelson Blank	APOSENTADORIA	1.7.3. faça cessar os pagamentos decorrentes d...	2014-05-27	15 DIAS	Universidade Federal de Santa Catarina	6651	2014	2173	2014	PRIMEIRA CÂMARA
12	05189985253	Divanir Gonçalves da Costa	APOSENTADORIA	1.7. Determinar ao Instituto Federal de Educaç...	2019-02-19	15 DIAS	Instituto Federal de Educação, Ciência e Tecno...	33207	2018	1440	2019	PRIMEIRA CÂMARA
13	05189985253	Divanir Gonçalves da Costa	APOSENTADORIA	9.3. determinar ao Instituto Federal de Educaç...	2016-11-08	15 DIAS	Instituto Federal de Educação, Ciência e Tecno...	20189	2016	11869	2016	SEGUNDA CÂMARA



Acórdãos com ausência de CPF na identificação de interessados:

ACÓRDÃO Nº 7951/2014 – TCU – 2ª Câmara

1. Processo nº TC 006.292/2013-5.
2. Grupo I – Classe de Assunto: V – Aposentadoria.
3. Interessado: Joaquim Torres Araújo.
4. Órgão: Tribunal Regional do Trabalho da 1ª Região/RJ.
5. Relator: Ministro Aroldo Cedraz.
6. Representante do Ministério Público: Procurador Marinus Eduardo De Vries Marsico.
7. Unidade Técnica: Secretaria de Fiscalização de Pessoal (Sefip).
8. Advogado constituído nos autos: não há.

9. Acórdão:
VISTOS, relatados e discutidos estes autos de análise de ato de aposentadoria de ex-servidores do Tribunal Regional do Trabalho da 1ª Região/RJ.
ACORDAM os Ministros do Tribunal de Contas da União, reunidos em Câmara, diante das razões expostas pelo relator, com fundamento nos arts Constituição Federal, 39, inciso II, da Lei nº 8.443, de 16 de julho de 2º, do Regimento Interno, em considerar ilegais o ato de aposentadoria Araújo, negando o respectivo registro.
9.1. determinar à Secretaria de Fiscalização de Pessoal que aut Representação com objetivo de:
9.1.1. aferir a boa-fé ou não do ex-servidor Joaquim Torres Ara verificar se é cabível ao caso a aplicação da Súmula TCU nº 106, ou se proceder ao ressarcimento ao Erário de forma solidária com o gestor que prejuízo;
9.1.2. identificar a responsabilidade do gestor à época dos fat encaminhamento do ato em análise, bem como se tal conduta poderá culmin previstas na Lei nº 8.443, de 16 de julho de 1992.

ACÓRDÃO Nº 758/2015 – TCU – 2ª Câmara

Os Ministros do Tribunal de Contas da União ACORDAM, por unanimidade, com fundamento nos arts. 143, inciso V, alínea e, e 183, inciso I, alínea d, do Regimento Interno/TCU, em prorrogar o prazo, por mais 15 (quinze) dias, a contar da notificação desta deliberação, para que a Superintendência Regional do Trabalho e Emprego no Estado do Amapá cumpra a determinação constante do Acórdão n. 7.607/2014 – 2ª Câmara:

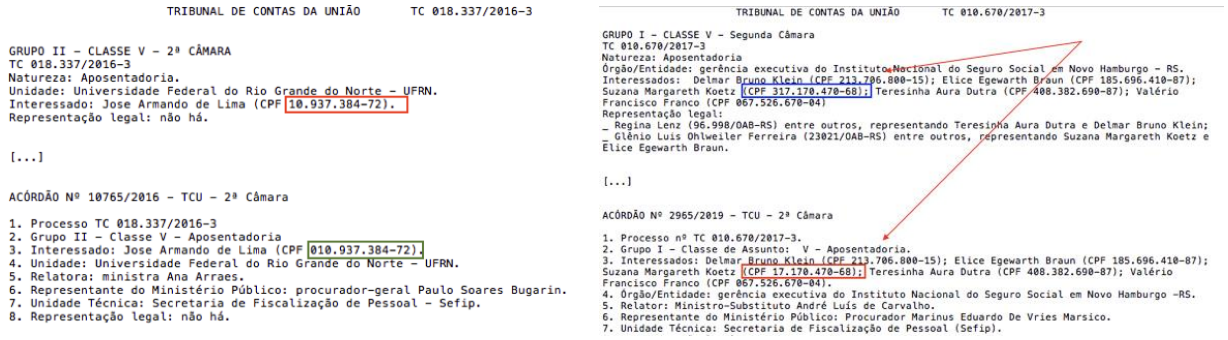
1. Processo TC-011.144/2012-2 (APOSENTADORIA)
1.1. Interessado: Joelma de Morais Santos, Superintendente Regional do Trabalho e Emprego no Estado do Amapá.
1.2. Órgão/Entidade: Superintendência Regional do Trabalho e Emprego no Estado do

ACÓRDÃO Nº 5705/2015 – TCU – 2ª Câmara

Os Ministros do Tribunal de Contas da União, quanto ao processo a seguir relacionado, ACORDAM, por unanimidade, com fundamento nos arts. 71, inciso III, da Constituição Federal de 1998, 1º, inciso V, e 39, inciso II, da Lei nº 8.443, de 16 de julho de 1992, 1º, inciso VIII, 143, inciso II, 259, inciso II, e 260, § 5º do Regimento Interno do TCU, e 7º da Resolução 206/2007-TCU, em considerar prejudicada, por perda de objeto, a apreciação do ato de concessão de pensão civil de Geraldo Arcanjo Vieira, de acordo com os pareceres emitidos nos autos.

1. Processo TC-007.491/2004-1 (PENSÃO CIVIL)
1.1. Interessado: Geraldo Arcanjo Vieira
1.2. Órgão/Entidade: Universidade Federal de Minas Gerais
1.3. Relator: Ministro Augusto Nardes
1.4. Representante do Ministério Público: Procurador-Geral Paulo Soares Bugarin
1.5. Unidade Técnica: Secretaria de Fiscalização de Pessoal (SEFIP).
1.6. Advogado constituído nos autos: não há.
1.7. Determinações/Recomendações/Orientações: não há.

Acórdãos contendo CPF de 8 dígitos na identificação de interessados:



TRIBUNAL DE CONTAS DA UNIÃO TC 018.337/2016-3

GRUPO II – CLASSE V – 2ª CÂMARA
TC 018.337/2016-3
Natureza: Aposentadoria.
Unidade: Universidade Federal do Rio Grande do Norte – UFRN.
Interessado: Jose Armando de Lima (CPF 10.937.384-72).
Representação legal: não há.

[...]

ACÓRDÃO Nº 10765/2016 – TCU – 2ª Câmara

1. Processo TC 018.337/2016-3
2. Grupo II – Classe V – Aposentadoria
3. Interessado: Jose Armando de Lima (CPF 010.937.384-72).
4. Unidade: Universidade Federal do Rio Grande do Norte – UFRN.
5. Relatora: ministra Ana Arraes.
6. Representante do Ministério Público: procurador-geral Paulo Soares Bugarin.
7. Unidade Técnica: Secretaria de Fiscalização de Pessoal – Sefip.
8. Representação legal: não há.

TRIBUNAL DE CONTAS DA UNIÃO TC 010.670/2017-3

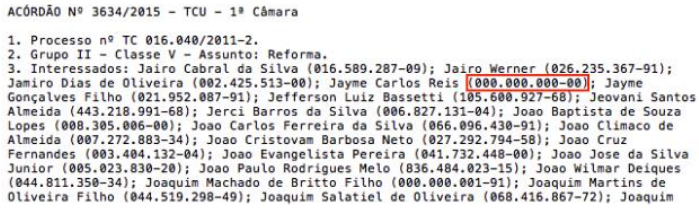
GRUPO I – CLASSE V – Segunda Câmara
TC 010.670/2017-3
Natureza: Aposentadoria
Órgão/Entidade: gerência executiva do Instituto Nacional do Seguro Social em Novo Hamburgo – RS.
Interessados: Delmar Bruno Klein (CPF 213.706.800-15); Elice Egewarth Braun (CPF 185.696.410-87); Suzana Margareth Koetz (CPF 317.170.470-68); Teresinha Aura Dutra (CPF 408.382.690-87); Valério Francisco Franco (CPF 067.526.670-04)
Representação legal:
– Regina Lenz (96.998/OAB-RS) entre outros, representando Teresinha Aura Dutra e Delmar Bruno Klein;
– Glênio Luis Ohlweiler Ferreira (23021/OAB-RS) entre outros, representando Suzana Margareth Koetz e Elice Egewarth Braun.

[...]

ACÓRDÃO Nº 2965/2019 – TCU – 2ª Câmara

1. Processo nº TC 010.670/2017-3.
2. Grupo I – Classe de Assunto: V – Aposentadoria.
3. Interessados: Delmar Bruno Klein (CPF 213.706.800-15); Elice Egewarth Braun (CPF 185.696.410-87); Suzana Margareth Koetz (CPF 17.170.470-68); Teresinha Aura Dutra (CPF 408.382.690-87); Valério Francisco Franco (CPF 067.526.670-04).
4. Órgão/Entidade: gerência executiva do Instituto Nacional do Seguro Social em Novo Hamburgo –RS.
5. Relator: Ministro-Substituto André Luís de Carvalho.
6. Representante do Ministério Público: Procurador Marinus Eduardo De Vries Marsico.
7. Unidade Técnica: Secretaria de Fiscalização de Pessoal (Sefip).

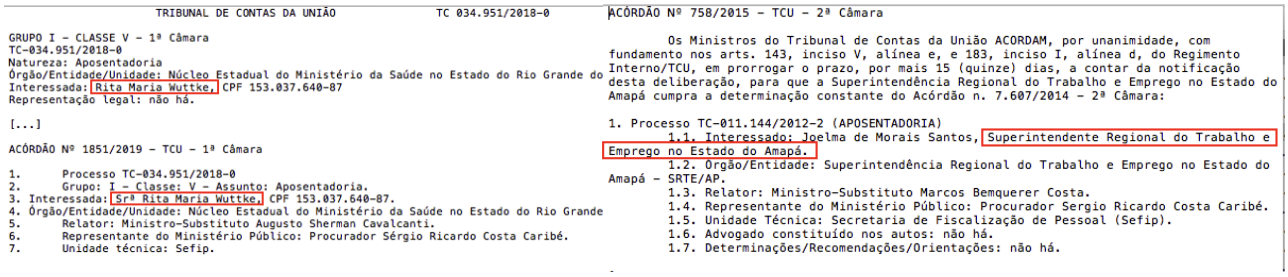
Acórdãos contendo CPF “0” na identificação de interessados:



ACÓRDÃO Nº 3634/2015 – TCU – 1ª Câmara

1. Processo nº TC 016.040/2011-2.
2. Grupo II – Classe V – Assunto: Reforma.
3. Interessados: Jairo Cabral da Silva (016.589.287-09); Jairo Werner (026.235.367-91); Jamiro Dias de Oliveira (002.425.513-00); Jayme Carlos Reis (000.000.000-00); Jayme Gonçalves Filho (021.952.087-91); Jefferson Luiz Bassetti (105.600.927-68); Jeovani Santos Almeida (443.218.991-68); Jerci Barros da Silva (006.827.131-04); Joao Baptista de Souza Lopes (008.305.006-00); Joao Carlos Ferreira da Silva (066.096.430-91); Joao Climaco de Almeida (007.272.883-34); Joao Cristovam Barbosa Neto (027.292.794-58); Joao Cruz Fernandes (003.404.132-04); Joao Evangelista Pereira (041.732.448-00); Joao Jose da Silva Junior (005.023.830-20); Joao Paulo Rodrigues Melo (836.484.023-15); Joao Wilmar Deiques (044.811.350-34); Joaquim Machado de Britto Filho (000.000.001-91); Joaquim Martins de Oliveira Filho (044.519.298-49); Joaquim Salatiel de Oliveira (068.416.867-72); Joaquim

Acórdãos contendo informações diversas na identificação do interessado (cargo, “sr.”):



TRIBUNAL DE CONTAS DA UNIÃO TC 034.951/2018-0

GRUPO I – CLASSE V – 1ª Câmara
TC-034.951/2018-0
Natureza: Aposentadoria
Órgão/Entidade/Unidade: Núcleo Estadual do Ministério da Saúde no Estado do Rio Grande do
Interessada: Rita Maria Wuttke, CPF 153.037.640-87
Representação legal: não há.

[...]

ACÓRDÃO Nº 1851/2019 – TCU – 1ª Câmara

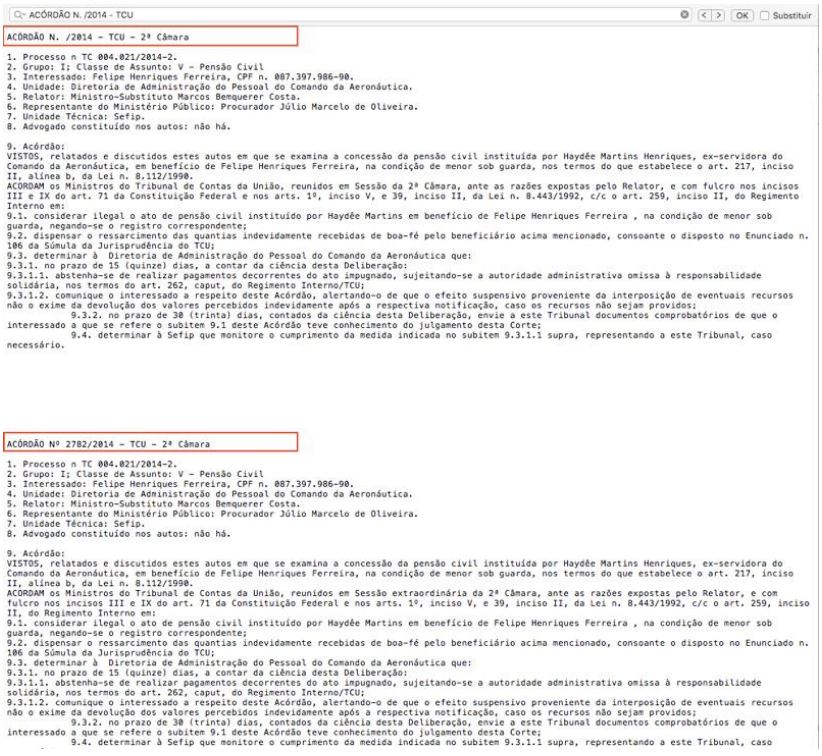
1. Processo TC-034.951/2018-0
2. Grupo: I – Classe: V – Assunto: Aposentadoria.
3. Interessada: Srª Rita Maria Wuttke, CPF 153.037.640-87.
4. Órgão/Entidade/Unidade: Núcleo Estadual do Ministério da Saúde no Estado do Rio Grande
5. Relator: Ministro-Substituto Augusto Sherman Cavalcanti.
6. Representante do Ministério Público: Procurador Sérgio Ricardo Costa Caribé.
7. Unidade técnica: Sefip.

ACÓRDÃO Nº 758/2015 – TCU – 2ª Câmara

Os Ministros do Tribunal de Contas da União ACORDAM, por unanimidade, com fundamento nos arts. 143, inciso V, alínea e, e 183, inciso I, alínea d, do Regimento Interno/TCU, em prorrogar o prazo, por mais 15 (quinze) dias, a contar da notificação desta deliberação, para que a Superintendência Regional do Trabalho e Emprego no Estado do Amapá cumpra a determinação constante do Acórdão n. 7.607/2014 – 2ª Câmara:

1. Processo TC-011.144/2012-2 (APOSENTADORIA)
1.1. Interessado: Joelma de Morais Santos, Superintendente Regional do Trabalho e Emprego no Estado do Amapá.
1.2. Órgão/Entidade: Superintendência Regional do Trabalho e Emprego no Estado do Amapá – SRTE/AP.
1.3. Relator: Ministro-Substituto Marcos Bemquerer Costa.
1.4. Representante do Ministério Público: Procurador Sergio Ricardo Costa Caribé.
1.5. Unidade Técnica: Secretaria de Fiscalização de Pessoal (Sefip).
1.6. Advogado constituído nos autos: não há.
1.7. Determinações/Recomendações/Orientações: não há.

Acórdãos “duplicados":



ACÓRDÃO N. /2014 – TCU

ACÓRDÃO N. /2014 – TCU – 2ª Câmara

1. Processo n TC 004.021/2014-2.
2. Grupo: I; Classe de Assunto: V – Pensão Civil
3. Interessado: Felipe Henriques Ferreira, CPF n. 087.397.986-90.
4. Unidade: Diretoria de Administração do Pessoal do Comando da Aeronáutica.
5. Relator: Ministro-Substituto Marcos Bemquerer Costa.
6. Representante do Ministério Público: Procurador Júlio Marcelo de Oliveira.
7. Unidade Técnica: Sefip.
8. Advogado constituído nos autos: não há.

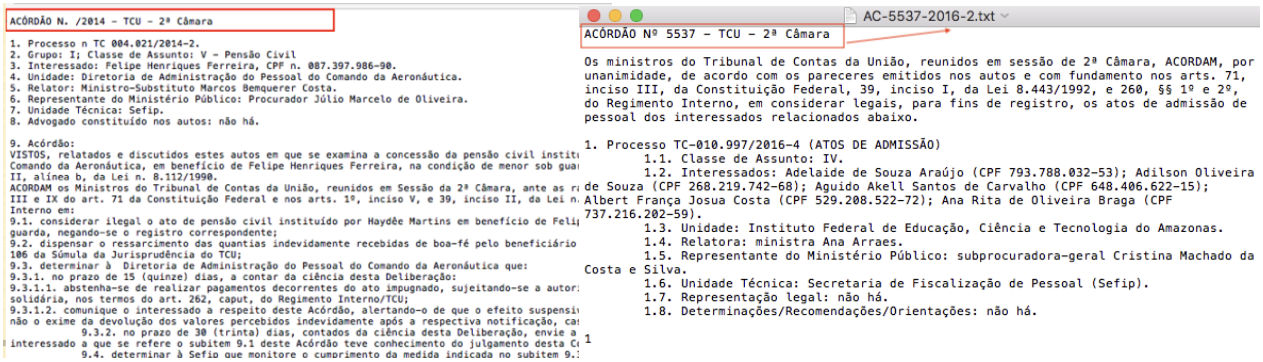
9. Acórdão:
VISTOS, relatados e discutidos estes autos em que se examina a concessão da pensão civil instituída por Haydée Martins Henriques, ex-servidora do Comando da Aeronáutica, em benefício de Felipe Henriques Ferreira, na condição de menor sob guarda, nos termos do que estabelece o art. 217, inciso II, alínea b, da Lei n. 8.112/1990.
ACORDAM os Ministros do Tribunal de Contas da União, reunidos em Sessão da 2ª Câmara, ante as razões expostas pelo Relator, e com fulcro nos incisos III e IX do art. 71 da Constituição Federal e nos arts. 1º, inciso V, e 39, inciso II, da Lei n. 8.443/1992, c/c o art. 259, inciso II, do Regimento Interno em:
9.1. considerar ilegal o ato de pensão civil instituído por Haydée Martins em benefício de Felipe Henriques Ferreira , na condição de menor sob guarda, negando-se o registro correspondente;
9.2. dispensar o ressarcimento das quantias indevidamente recebidas de boa-fé pelo beneficiário acima mencionado, consoante o disposto no Enunciado n. 106 da Súmula da Jurisprudência do TCU;
9.3. determinar à Diretoria de Administração do Pessoal do Comando da Aeronáutica que:
9.3.1. no prazo de 15 (quinze) dias, a contar da ciência desta Deliberação:
9.3.1.1. abstenha-se de realizar pagamentos decorrentes do ato impugnado, sujeitando-se a autoridade administrativa omissa à responsabilidade solidária, nos termos do art. 262, caput, do Regimento Interno/TCU;
9.3.1.2. comunique o interessado a respeito deste Acórdão, alertando-o de que o efeito suspensivo proveniente da interposição de eventuais recursos não o exime da devolução dos valores percebidos indevidamente após a respectiva notificação, caso os recursos não sejam providos;
9.3.2. no prazo de 30 (trinta) dias, contados da ciência desta Deliberação, envie a este Tribunal documentos comprobatórios de que o interessado a que se refere o subitem 9.1 deste Acórdão teve conhecimento do julgamento desta Corte;
9.4. determinar à Sefip que monitore o cumprimento da medida indicada no subitem 9.3.1.1 supra, representando a este Tribunal, caso necessário.

ACÓRDÃO Nº 2782/2014 – TCU – 2ª Câmara

1. Processo n TC 004.021/2014-2.
2. Grupo: I; Classe de Assunto: V – Pensão Civil
3. Interessado: Felipe Henriques Ferreira, CPF n. 087.397.986-90.
4. Unidade: Diretoria de Administração do Pessoal do Comando da Aeronáutica.
5. Relator: Ministro-Substituto Marcos Bemquerer Costa.
6. Representante do Ministério Público: Procurador Júlio Marcelo de Oliveira.
7. Unidade Técnica: Sefip.
8. Advogado constituído nos autos: não há.

9. Acórdão:
VISTOS, relatados e discutidos estes autos em que se examina a concessão da pensão civil instituída por Haydée Martins Henriques, ex-servidora do Comando da Aeronáutica, em benefício de Felipe Henriques Ferreira, na condição de menor sob guarda, nos termos do que estabelece o art. 217, inciso II, alínea b, da Lei n. 8.112/1990.
ACORDAM os Ministros do Tribunal de Contas da União, reunidos em Sessão extraordinária da 2ª Câmara, ante as razões expostas pelo Relator, e com fulcro nos incisos III e IX do art. 71 da Constituição Federal e nos arts. 1º, inciso V, e 39, inciso II, da Lei n. 8.443/1992, c/c o art. 259, inciso II, do Regimento Interno em:
9.1. considerar ilegal o ato de pensão civil instituído por Haydée Martins em benefício de Felipe Henriques Ferreira , na condição de menor sob guarda, negando-se o registro correspondente;
9.2. dispensar o ressarcimento das quantias indevidamente recebidas de boa-fé pelo beneficiário acima mencionado, consoante o disposto no Enunciado n. 106 da Súmula da Jurisprudência do TCU;
9.3. determinar à Diretoria de Administração do Pessoal do Comando da Aeronáutica que:
9.3.1. no prazo de 15 (quinze) dias, a contar da ciência desta Deliberação:
9.3.1.1. abstenha-se de realizar pagamentos decorrentes do ato impugnado, sujeitando-se a autoridade administrativa omissa à responsabilidade solidária, nos termos do art. 262, caput, do Regimento Interno/TCU;
9.3.1.2. comunique o interessado a respeito deste Acórdão, alertando-o de que o efeito suspensivo proveniente da interposição de eventuais recursos não o exime da devolução dos valores percebidos indevidamente após a respectiva notificação, caso os recursos não sejam providos;
9.3.2. no prazo de 30 (trinta) dias, contados da ciência desta Deliberação, envie a este Tribunal documentos comprobatórios de que o interessado a que se refere o subitem 9.1 deste Acórdão teve conhecimento do julgamento desta Corte;
9.4. determinar à Sefip que monitore o cumprimento da medida indicada no subitem 9.3.1.1 supra, representando a este Tribunal, caso

Acórdãos com ausência de registro de número ou ano (do Acórdão):



ACÓRDÃO N. /2014 - TCU - 2ª Câmara

1. Processo n TC 004.021/2014-2.
2. Grupo: I; Classe de Assunto: V - Pensão Civil
3. Interessado: Felipe Henriques Ferreira, CPF n. 087.397.986-90.
4. Unidade: Diretoria de Administração do Pessoal do Comando da Aeronáutica.
5. Relator: Ministro-Substituto Marcos Bemquerer Costa.
6. Representante do Ministério Público: Procurador Júlio Marcelo de Oliveira.
7. Unidade Técnica: Sefip.
8. Advogado constituído nos autos: não há.

9. Acórdão:
VISTOS, relatados e discutidos estes autos em que se examina a concessão da pensão civil institu
Comando da Aeronáutica, em benefício de Felipe Henriques Ferreira, na condição de menor sob gua
II, alínea b, da Lei n. 8.112/1990.
ACORDAM os Ministros do Tribunal de Contas da União, reunidos em Sessão da 2ª Câmara, ante as r
III e IX do art. 71 da Constituição Federal e nos arts. 1º, inciso V, e 39, inciso II, da Lei n.
Interno em:
9.1. considerar ilegal o ato de pensão civil instituído por Haydêe Martins em benefício de Felip
guarda, negando-se o registro correspondente;
9.2. dispensar o ressarcimento das quantias indevidamente recebidas de boa-fé pelo beneficiário
106 da Súmula da Jurisprudência do TCU;
9.3. determinar à Diretoria de Administração do Pessoal do Comando da Aeronáutica que:
9.3.1. no prazo de 15 (quinze) dias, a contar da ciência desta Deliberação:
9.3.1.1. abstenha-se de realizar pagamentos decorrentes do ato impugnado, sujeitando-se a autor
solidária, nos termos do art. 262, caput, do Regimento Interno/TCU;
9.3.1.2. comunique o interessado a respeito deste Acórdão, alertando-o de que o efeito suspensi
não o exime da devolução dos valores percebidos indevidamente após a respectiva notificação, ca
9.3.2. no prazo de 30 (trinta) dias, contados da ciência desta Deliberação, envie a
interessado a que se refere o subitem 9.1 deste Acórdão teve conhecimento do julgamento desta C
9.4. determinar à Sefip que monitore o cumprimento da medida indicada no subitem 9.

AC-5537-2016-2.txt

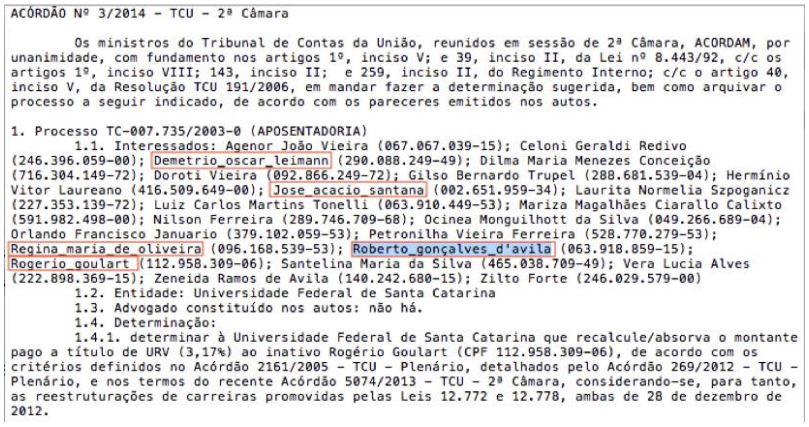
ACÓRDÃO Nº 5537 - TCU - 2ª Câmara

Os ministros do Tribunal de Contas da União, reunidos em sessão de 2ª Câmara, ACORDAM, por unanimidade, de acordo com os pareceres emitidos nos autos e com fundamento nos arts. 71, inciso III, da Constituição Federal, 39, inciso I, da Lei 8.443/1992, e 260, §§ 1º e 2º, do Regimento Interno, em considerar legais, para fins de registro, os atos de admissão de pessoal dos interessados relacionados abaixo.

1. Processo TC-010.997/2016-4 (ATOS DE ADMISSÃO)
1.1. Classe de Assunto: IV.
1.2. Interessados: Adelaide de Souza Araújo (CPF 793.788.032-53); Adilson Oliveira de Souza (CPF 268.219.742-68); Aguido Akell Santos de Carvalho (CPF 648.406.622-15); Albert França Josua Costa (CPF 529.208.522-72); Ana Rita de Oliveira Braga (CPF 737.216.202-59).
1.3. Unidade: Instituto Federal de Educação, Ciência e Tecnologia do Amazonas.
1.4. Relatora: ministra Ana Arraes.
1.5. Representante do Ministério Público: subprocuradora-geral Cristina Machado da Costa e Silva.
1.6. Unidade Técnica: Secretaria de Fiscalização de Pessoal (Sefip).
1.7. Representação legal: não há.
1.8. Determinações/Recomendações/Orientações: não há.

1

Acórdãos contendo registros de interessados separa dos por "_":



ACÓRDÃO Nº 3/2014 - TCU - 2ª Câmara

Os ministros do Tribunal de Contas da União, reunidos em sessão de 2ª Câmara, ACORDAM, por unanimidade, com fundamento nos artigos 1º, inciso V; e 39, inciso II, da Lei nº 8.443/92, c/c os artigos 1º, inciso VIII; 143, inciso II; e 259, inciso II, do Regimento Interno; c/c o artigo 40, inciso V, da Resolução TCU 191/2006, em mandar fazer a determinação sugerida, bem como arquivar o processo a seguir indicado, de acordo com os pareceres emitidos nos autos.

1. Processo TC-007.735/2003-0 (APOSENTADORIA)
1.1. Interessados: Agenor João Vieira (067.067.039-15); Celoni Geraldi Redivo (246.396.059-00); Demetrio_oscar_leimann (290.088.249-49); Dilma Maria Menezes Conceição (716.304.149-72); Doroti Vieira (092.866.249-72); Gilso Bernardo Trupel (288.681.539-04); Hermínio Vitor Laureano (416.509.649-00); Jose_acacio_santana (002.651.959-34); Laurita Normelia Szpoganicz (227.353.139-72); Luiz Carlos Martins Tonelli (063.910.449-53); Mariza Magalhães Ciarallo Calixto (591.982.498-00); Nilson Ferreira (289.746.709-68); Ocinea Monguilhott da Silva (049.266.689-04); Orlando Francisco Januario (379.102.059-53); Petronilha Vieira Ferreira (528.770.279-53); Regina_maria_de_oliveira (096.168.539-53); Roberto_gonçalves_d'avila (063.918.859-15); Rogerio_goulart (112.958.309-06); Santelina Maria da Silva (465.038.709-49); Vera Lucia Alves (222.898.369-15); Zeneida Ramos de Avila (140.242.680-15); Zilto Forte (246.029.579-00)
1.2. Entidade: Universidade Federal de Santa Catarina
1.3. Advogado constituído nos autos: não há.
1.4. Determinação:
1.4.1. determinar à Universidade Federal de Santa Catarina que recalcule/absorva o montante pago a título de URV (3,17%) ao inativo Rogério Goulart (CPF 112.958.309-06), de acordo com os critérios definidos no Acórdão 2161/2005 - TCU - Plenário, detalhados pelo Acórdão 269/2012 - TCU - Plenário, e nos termos do recente Acórdão 5074/2013 - TCU - 2ª Câmara, considerando-se, para tanto, as reestruturações de carreiras promovidas pelas Leis 12.772 e 12.778, ambas de 28 de dezembro de 2012.