

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

RHUAN PAULO LOPES BARROS

**Análise de Decisões Judiciais Utilizando
Aprendizado de Máquina com Supervisão
Fraca**

Dissertação apresentada como requisito parcial
para a obtenção do grau de Mestre em Ciência da
Computação

Orientador: Prof. Dr. Leandro Krug Wives

Porto Alegre
2022

CIP — CATALOGAÇÃO NA PUBLICAÇÃO

, Rhuan Paulo Lopes Barros

Análise de Decisões Judiciais Utilizando Aprendizado de Máquina com Supervisão Fraca / Rhuan Paulo Lopes Barros . – Porto Alegre: PPGC da UFRGS, 2022.

142 f.: il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2022. Orientador: Leandro Krug Wives.

1. Mineração de Dados. 2. Aprendizado de Máquina. 3. Supervisão Fraca. 4. Classificação de Textos. 5. Jurimetria. 6. Texto Jurídico. 7. Texto Legal. I. Krug Wives, Leandro. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos André Bulhões Mendes

Vice-Reitora: Prof^a. Patricia Pranke

Vice-Pró-Reitor de Pós-Graduação: Prof^a. Júlio Otávio Jardim

Diretora do Instituto de Informática: Prof^a. Carla Maria Dal Sasso Freitas

Coordenador do PPGC: Prof. Cláudio Jung

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

AGRADECIMENTOS

Em momentos de devaneios, a mente egoica se permite pensar que a obtenção de conquistas pessoais ou de certo estágio de desenvolvimento acadêmico ocorrem por meio de esforço próprio e que ela mesma seria a única responsável pelos seus próprios resultados, enfim, que tudo seria em função de seu próprio mérito. Entretanto não há falácia maior do que pensar que chegamos sozinhos onde estamos atualmente. Assim, todos usufruímos de benefícios apenas por crescermos em conjunto ou por vivermos em sociedade.

Desse modo, inicio os agradecimentos honrando meus pais pois acredito que iniciamos nossa vida pelo menos a partir do ponto de evolução que eles alcançaram, neste caso arduamente, e me ajudaram a me formar como pessoa. Também preciso agradecer imensamente aos meus orientadores, primeiramente na faculdade, o professor André Peres e depois no mestrado, o professor Leandro Krug Wives, que me ajudaram a trilhar o caminho acadêmico. Ambos tiveram paciência para me ouvir em um momento que eu tinha apenas pensamentos vagos sobre certo estudo que eu queria desenvolver, e depois, imprimiram grande confiança de que eu conseguiria desenvolver o projeto.

Eu recebi ajuda diretamente no meu trabalho também de uma pessoa que já deixou este mundo, Etiene Hubert da Silva Jaccottet. Passamos incansáveis horas idealizando o desenvolvimento da pesquisa e as possíveis aplicações futuras. Também tive a graça de receber o apoio de duas pessoas muito importantes na minha vida na fase de anotação da base de dados padrão-ouro, meu pai, José Paulo de Oliveira Barros, e minha namorada Laura Razzolini Ávila. Ambos utilizaram seu conhecimento do domínio do Direito para anotar centenas de documentos. E por fim, ao Instituto de Informática da UFRGS por me dar acesso a essa grande instituição de ensino e a essa nova formação acadêmica.

RESUMO

Existe uma preocupação crescente de que a Justiça do Trabalho brasileira esteja demasiadamente inclinada a proteger empregados em relação a empregadores. Além disso, as próprias empresas nas suas relações empregado-empregador encontram-se inseguras em relação a quais decisões serão tomadas pelos magistrados em seus processos judiciais. Por outro lado, novas soluções tecnológicas estão sendo implementadas com intuito de aumentar a eficiência judiciária brasileira, como, por exemplo, Processo Judicial Eletrônico (PJe). Esse sistema permite a tramitação completa do processo judicial de maneira digital, contendo atualmente milhões de ações em tramitação. Entretanto, a exploração dessa massa de documentos judiciais não é trivial, pois tais documentos encontram-se disponibilizados em texto puro sem o enriquecimento necessário para a extração de conhecimento de valor. Dessa maneira, a presente pesquisa empregou técnicas de Aprendizado de Máquina Supervisionado a fim de verificar se seria possível observar eventual tendência de julgamento de determinados tribunais utilizando métodos computacionais. Desse modo, foi desenvolvida uma base de dados padrão-ouro a qual foi utilizada para a realização de testes, e também foi desenvolvida automaticamente uma base de treinamento por meio de técnica de Supervisão Fraca. Após, uma base de decisões judiciais de mais de 20 mil documentos foi classificada utilizando um modelo treinado com o algoritmo *Gradient Boosting* o qual obteve 92% na métrica F1 macro. Assim, foi possível observar diferença estatística na proporção de julgamentos a favor dos empregados em ambos os tribunais.

Palavras-chave: Mineração de Dados. Aprendizado de Máquina. Supervisão Fraca. Classificação de Textos. Jurimetria. Texto Jurídico. Texto Legal.

Analysis of Court Decisions using Machine Learning with Weak Supervision

ABSTRACT

There is growing concern that the Brazilian Labor Court is too inclined to protect employees over employers. In addition, in their employee-employer relations, the companies themselves are unsure of what decisions will be taken by magistrates in their legal proceedings. On the other hand, new technological solutions are being implemented in order to increase the Brazilian judicial efficiency, such as, for example, the Electronic Judicial Process (PJe). This system allows the complete processing of the judicial process digitally, currently containing millions of lawsuits in process. However, the exploration of this mass of court documents is not trivial as such documents are available in plain text without the necessary enrichment to extract valuable knowledge. Thus, this research used Supervised Machine Learning techniques in order to verify if it would be possible to observe a possible tendency to judge of certain courts using computational methods. Thus, a gold standard database was developed, which was used for testing, and a training database was also automatically developed using the Weak Supervision technique. Afterward, a court decision base of more than 20 thousand documents was classified using a model trained with the Gradient Boosting algorithm that obtained 92% in the F1 macro metric. Thus, it was possible to observe a statistical difference in the proportion of judgments in favor of employees in both courts.

Keywords: Data Mining, Machine Learning, Weak Supervision, Text Classification, Jurimetry, Law text, Legal Text.

LISTA DE FIGURAS

Figura 2.1 Organização judiciária brasileira conforme a Constituição Federal do Brasil, Capítulo III	19
Figura 2.2 Organização judiciária dos Tribunais Regionais do Trabalho conforme a Constituição Federal do Brasil, Capítulo III	20
Figura 2.3 Metodologia de desenvolvimento de bases de texto anotadas chamada ciclo <i>MATTER</i>	25
Figura 2.4 Ilustração sobre o funcionamento do algoritmo de classificação Rocchio. É possível observar os círculos preenchidos pretos que representam os centroides das classes e uma estrela que representa uma observação a ser classificada, nesse caso, com o rótulo <i>China</i>	28
Figura 2.5 Ilustração sobre o funcionamento do algoritmo de Aprendizado Ativo.....	31
Figura 2.6 Ilustração sobre o funcionamento do algoritmo de Aprendizado Ativo.....	34
Figura 2.7 Ilustração da utilização de matriz de covariância para cálculo de acurácias de fontes de rotulação sem a utilização de base de dados padrão-ouro.....	36
Figura 2.8 Ilustração da utilização de complementação de matriz para cálculo de acurácias de fontes de rotulação sem a utilização de base de dados padrão-ouro.	36
Figura 4.1 Ciclo de desenvolvimento do processo metodológico CRISP-DM Guide 1.0.....	45
Figura 5.1 Ilustração explicativa do <i>pipeline</i> da Validação Experimental da pesquisa incluindo visão geral do processo.....	55
Figura 5.2 Ilustração explicativa do fluxo de pré-processamento para modelagem por algoritmos de Aprendizado de Máquina.....	58
Figura 6.1 Trecho do documento Diretrizes para anotação manual de documentos jurídicos para pesquisa de mestrado de Rhuan Barros.....	74
Figura 6.2 Histograma apresentando o balanceamento das classes da base de dados anotada em relação ao tipo de reclamante.	76
Figura 6.3 Histograma apresentando o balanceamento das classes da base de dados anotada em relação ao deferimento ou não da decisão.	77
Figura 6.4 Histograma quantidade de palavras no dispositivo por tribunal.....	77
Figura 6.5 Gráfico que apresenta os termos mais associados com cada uma das classes anotadas. Mais próximas do canto superior esquerdo encontram-se palavras mais associadas com o rótulo <i>Deferimento</i> . Já próximas do canto inferior direito encontram-se palavras mais associadas com o rótulo <i>Indeferimento</i>	79
Figura 6.6 Gráfico de histograma que apresenta as porcentagens dos rótulos aplicados pela Funções de Rotulação.....	83
Figura 6.7 Gráfico de histograma que apresenta as porcentagens dos rótulos finais aplicados pelo modelo desenvolvido.....	84
Figura 6.8 Ilustração do projeto de teste do classificador de tipo de requerente informando as quantidades específicas de instâncias da base de dados.....	85
Figura 6.9 Métricas calculadas utilizando a base de 30% reservada previamente para testes do modelo de classificação do tipo de requerente.....	86
Figura 6.10 Ilustração do projeto de teste do classificador de tipo de requerente informando as quantidades específicas de instâncias da base de dados.....	89

Figura 6.11 Modelos treinados com base padrão-ouro e métricas calculadas utilizando a base de 30% reservada previamente para testes do modelo de classificação do deferimento ou não da decisão.	90
Figura 6.12 Modelos treinados utilizando base criada programaticamente balanceada e métricas calculadas utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.....	93
Figura 6.13 Modelos treinados utilizando base criada programaticamente completa e métricas calculadas utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.....	95
Figura 6.14 Porcentagem de deferimento de recursos em relação ao recorrente ser empresa e empregado no Tribunal Regional do Trabalho da 4ª Região.	97
Figura 6.15 Porcentagem de deferimento de recursos em relação ao recorrente ser empresa e empregado das Turmas Recursais do Tribunal Regional do Trabalho da 4ª Região.	98
Figura 6.16 Distribuição da porcentagem de deferimento de recursos em relação ao recorrente ser empresa e empregado das Turmas Recursais do Tribunal Regional do Trabalho da 4ª Região.....	99
Figura 6.17 Porcentagem de deferimento de recursos em relação ao recorrente ser empresa e empregado no Tribunal Regional do Trabalho da 3ª Região.	99
Figura 6.18 Porcentagem de deferimento de recursos em relação ao recorrente ser empresa e empregado das Turmas Recursais do Tribunal Regional do Trabalho da 3ª Região.	101
Figura 6.19 Distribuição da porcentagem de deferimento de recursos em relação ao recorrente ser empresa e empregado das Turmas Recursais do Tribunal Regional do Trabalho da 3ª Região.....	102
Figura A.1 O quadro A apresenta radiografia de tórax (<i>CXR</i>), o quadro B apresenta radiografia de extremidade (<i>EXR</i>), o quadro C, radiografia de cabeça (<i>HCT</i>) e o quadro D apresenta eletroencefalografia (<i>EEG</i>). DP significa <i>Data-Programming</i> , ou seja, técnica de <i>Weak Supervision</i> . <i>FS</i> significa <i>Full hand-labeled supervision</i> , ou seja, técnica de <i>Aprendizado Supervisionado</i> com base de dados anotada manualmente. A linha tracejada representa intervalo de confiança para <i>FS</i> e a área sombreada para <i>DP</i> . Ambas em relação a cinco treinamentos com sementes aleatórias.	119
Figura A.2 Distribuições de unigramas da base de dados do TRT da 3ª Região.	120
Figura A.3 Distribuições de bigramas da base de dados do TRT da 3ª Região.....	120
Figura A.4 Distribuições de trigramas da base de dados do TRT da 3ª Região.....	121
Figura A.5 Distribuições de unigramas da base de dados do TRT da 4ª Região.....	121
Figura A.6 Distribuições de bigramas da base de dados do TRT da 4ª Região.....	121
Figura A.7 Distribuições de trigramas da base de dados do TRT da 4ª Região.....	122
Figura A.8 Planilha contendo decisões judiciais com alguns rótulos de exemplo.....	124
Figura A.9 Matriz de confusão construída utilizando a base de 30% reservada previamente para testes do modelo de classificação do deferimento ou não da decisão.....	131
Figura A.10 Matriz de confusão construída utilizando a base de 30% reservada previamente para testes do modelo de classificação do deferimento ou não da decisão.....	132
Figura A.11 Matriz de confusão construída utilizando a base de 30% reservada previamente para testes do modelo de classificação do deferimento ou não da decisão.....	132

LISTA DE TABELAS

Tabela 3.1 Quadro comparativo dos trabalhos relacionados em relação às principais características analisadas.	42
Tabela 5.1 Resumo da classificação metodológica da pesquisa.	52
Tabela 6.1 Dados do Justiça em Números (CNJ, 2020) que apresentam a quantidade de casos novos ajuizados no ano de 2019.	64
Tabela 6.2 Quantidade de decisões disponibilizadas e a quantidade da amostra inicial extraída para cada tribunal.	67
Tabela 6.3 Relação dos arquivos gerados a partir da extração dos dados dos tribunais.	68
Tabela 6.4 Descrição da base de dados extraída do TRT da 3ª Região em relação à quantidade de palavras.	68
Tabela 6.5 Descrição da base de dados extraída do TRT da 4ª Região em relação à quantidade de palavras.	68
Tabela 6.6 Matriz de confusão em relação aos rótulos aplicados.	75
Tabela 6.7 Balanceamento das classes da base de dados anotada em relação ao tipo de requerente.	75
Tabela 6.8 Balanceamento das classes da base de dados anotada em relação ao deferimento ou não da decisão.	76
Tabela 6.9 Descrição da base de dados anotada manualmente em função da quantidade de palavras no campo <i>Dispositivo</i> em relação a cada rótulo aplicado em cada tribunal.	78
Tabela 6.10 Lista de funções de rotulação criadas para aplicação automática de rótulos à base de dados.	80
Tabela 6.11 Quadro resumo contendo as bases de dados criadas na presente pesquisa incluindo as quantidades de instâncias por classe.	82
Tabela 6.12 Apresenta lista com as Funções de Rotulação aplicadas à base de dados e respectivos dados de cobertura, sobreposições e conflitos.	83
Tabela 6.13 Métricas calculadas utilizando a base de 30% reservada previamente para testes do modelo de classificação do tipo de requerente.	87
Tabela 6.14 Parâmetros e argumentos do modelo de aprendizado de máquina treinado utilizando o algoritmo SVM LinearSVC da biblioteca Scikit Learning.	87
Tabela 6.15 Modelos treinados com base padrão-ouro e métricas calculadas utilizando a base de 30% reservada previamente para testes do modelo de classificação do deferimento ou não da decisão.	90
Tabela 6.16 Modelos treinados utilizando base criada programaticamente balanceada e métricas calculadas utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.	92
Tabela 6.17 Modelos treinados utilizando base criada programaticamente completa e métricas calculadas utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.	94
Tabela 6.18 Parâmetros e argumentos do modelo de aprendizado de máquina treinado utilizando o algoritmo Gradient Boosting da biblioteca Scikit Learning.	96
Tabela 6.19 Porcentagem de deferimento de recursos em relação ao recorrente ser empresa e empregado das Turmas Recursais do Tribunal Regional do Trabalho da 4ª Região.	97

Tabela 6.20 Porcentagem de deferimento de recursos em relação ao recorrente ser empresa e empregado das Turmas Recursais do Tribunal Regional do Trabalho da 3ª Região.	100
--	-----

LISTA DE ABREVIATURAS E SIGLAS

DM *Data Mining*

IE *Information Extraction*

KNN *K Nearest Neighbors*

PLN *Processamento de Linguagem Natural*

TF-IDF *Term Frequency-Inverse Document Frequency*

WWW *World Wide Web*

TRT *Tribunal Regional do Trabalho*

SUMÁRIO

1 INTRODUÇÃO	15
2 FUNDAMENTAÇÃO TEÓRICA	18
2.1 Fundamentos teóricos relacionados ao Direito	18
2.1.1 Organização Judiciária	18
2.1.2 Acórdão	21
2.1.3 Magistrados e discricionariedade	22
2.2 Mineração de dados	23
2.3 Aprendizado de máquina	23
2.3.1 Aprendizado de máquina supervisionado	24
2.4 Anotação de bases de documentos de texto	24
2.4.1 Índice para avaliação de acordo entre anotadores	26
2.5 Algoritmos de aprendizagem de máquina supervisionada	27
2.5.1 <i>K</i> -Vizinhos-mais-próximos (<i>K-nearest Neighbor</i>)	27
2.5.2 Classificador Rocchio ou Centroide mais próximo	27
2.5.3 Classificador Naive Bayes	28
2.5.4 Árvore de decisões (<i>Decision Tree</i>)	28
2.5.5 Florestas aleatórias (<i>Random Forest</i>)	29
2.5.6 Classificador Gradient Tree Boosting	29
2.5.7 Máquina de Vetores de Suporte (<i>Support Vector Machine</i>)	29
2.6 Aprendizado de Máquina com dados limitados	30
2.6.1 Aprendizado Ativo (<i>Active Learning</i>)	31
2.6.2 Aprendizado Semi-supervisionado (<i>Semi-supervised Learning</i>)	32
2.6.3 Transferência de Aprendizado (<i>Transfer Learning</i>)	32
2.6.4 <i>Few-Shot Learning</i>	32
2.6.5 Supervisão Fraca (<i>Weak Supervision</i>)	33
2.6.5.1 Algoritmo do Modelo Generativo do <i>Snorkel Framework</i>	34
2.7 Resumo do Capítulo	37
3 TRABALHOS RELACIONADOS	39
3.1 Resumo do Capítulo	42
4 METODOLOGIA	45
4.1 Processo de descoberta de conhecimento em base de dados	45
4.1.1 Compreensão do Negócio	46
4.1.2 Compreensão dos dados	46
4.1.3 Preparação dos dados	47
4.1.4 Modelagem	48
4.1.5 Avaliação	48
4.1.6 Aplicação	49
4.2 Teste de hipótese estatística	49
4.3 Resumo do Capítulo	51
5 MATERIAIS E MÉTODOS	52
5.1 Classificação da pesquisa	52
5.2 Método de Validação Experimental	53
5.2.1 Coleta dos Dados	54
5.2.2 Preparação dos Dados	55
5.2.3 Modelagem	57
5.2.4 Aplicação	59
5.3 Resumo do Capítulo	59

6 VALIDAÇÃO EXPERIMENTAL.....	60
6.1 Compreensão do negócio	60
6.1.1 Objetivo de negócio	61
6.1.1.1 Teste estatístico	61
6.1.1.2 Critério de sucesso do objetivo de negócio.....	62
6.1.2 Requisitos e restrições.....	62
6.1.3 Custo-benefício	62
6.1.4 Objetivo de mineração de dados	63
6.2 Compreensão dos dados	63
6.2.1 Coleta de dados	65
6.2.1.1 Métodos de extração	66
6.2.1.2 Tamanho da amostra	66
6.2.1.3 Dificuldades encontradas	67
6.2.2 Descrição dos dados.....	67
6.2.3 Exploração dos dados	68
6.2.4 Qualidade dos dados coletados	68
6.3 Preparação dos dados	69
6.3.1 Limpeza dos dados.....	69
6.3.2 Transformação dos dados.....	69
6.3.2.1 Enriquecimento com metadados das decisões judiciais.....	70
6.3.2.2 Extração do dispositivo da sentença	70
6.3.2.3 Remoção de documentos com mais de um recorrente.....	70
6.3.3 Anotação manual da base de documentos	71
6.3.3.1 Anotação para desenvolvimento do modelo de classificação do tipo de re- corrente	72
6.3.3.2 Execução do processo de anotação manual em relação ao tipo de recorrente....	72
6.3.3.3 Anotação para desenvolvimento do modelo de classificação em relação ao deferimento ou indeferimento da decisão	72
6.3.3.4 Execução do processo de anotação manual em relação ao deferimento ou indeferimento da decisão	73
6.3.3.5 Avaliação da tarefa de anotação e criação do padrão-ouro	74
6.3.4 Exploração das bases de dados anotadas manualmente.....	75
6.3.4.1 Base de dados anotada manualmente quanto ao tipo de requerente	75
6.3.4.2 Base de dados anotada manualmente quanto ao deferimento ou não da decisão	76
6.3.5 Anotação automática da base de documentos por meio de Supervisão Fraca.....	79
6.4 Modelagem.....	82
6.4.1 Modelo: classificação do tipo de requerente como empregado ou empresa.....	83
6.4.2 Modelo: classificar a decisão em deferimento ou indeferimento	86
6.4.2.1 Experimento com base de dados criada manualmente padrão-ouro	88
6.4.2.2 Experimento com base de dados criada programaticamente balanceada	91
6.4.2.3 Experimento com base de dados criada programaticamente completa	92
6.5 Avaliação de resultados.....	94
6.5.1 TRT da 4ª Região - avaliação geral.....	96
6.5.2 TRT da 4ª Região - avaliação das Turmas Recursais.....	97
6.5.3 TRT da 3ª Região - avaliação geral.....	98
6.5.4 TRT da 3ª Região - avaliação das Turmas Recursais.....	100
6.6 Aplicação.....	100
6.7 Limitações.....	101
6.8 Resumo do Capítulo.....	103
7 DISCUSSÃO DE RESULTADOS.....	105
8 CONCLUSÕES	107

REFERÊNCIAS.....	109
APÊNDICE A — APÊNDICE.....	113
A.1 Exemplo de acórdão do Tribunal Regional do Trabalho da 4ª Região	113
A.2 Aplicações práticas da técnica de Supervisão Fraca utilizando Snorkel Framework	117
A.3 Distribuições dos unigramas, bigramas e trigramas na fase de exploração inicial dos documentos extraídos da internet	120
A.4 Diretrizes para anotação manual de documentos jurídicos para pesquisa de mestrado de Rhuan Barros	122
A.5 Exemplos de instâncias em que houve divergência de anotação.....	129
A.6 Figuras das matrizes de confusão do classificador quanto ao deferimento ou não da decisão treinado com base de dados criada manualmente	131
A.7 Figuras das matrizes de confusão do classificador quanto ao deferimento ou não da decisão treinado com base de dados balanceada criada programaticamente.....	135
A.8 Figuras das matrizes de confusão do classificador quanto ao deferimento ou não da decisão treinado com base de dados completa criada programaticamente.....	135

1 INTRODUÇÃO

Existe uma preocupação crescente de que a Justiça do Trabalho brasileira esteja demasiadamente inclinada a proteger empregados em relação a empregadores (SALAMA; CARLOTTI; YEUNG, 2018). Além disso, as próprias empresas nas suas relações empregado-empregador encontram-se inseguras em relação a quais decisões serão tomadas pelos magistrados em seus processos judiciais (MACIEL, 2013). Por outro lado, advogados buscam conhecer as tendências de julgamentos de tribunais e juízes de modo a melhor argumentar em suas petições. Assim, a partir da análise de decisões judiciais, seria possível observar tendência de julgamento de determinado Juiz ou Tribunal utilizando métodos computacionais?

Uma característica especial do processo trabalhista no Brasil, e amplamente incentivada pelos magistrados, é a possibilidade de realização de acordo em qualquer fase do processo antes de haver sentença transitada em julgado. Assim, eventualmente advogados deparam-se em situações em que podem oferecer ou devem decidir se aceitam um acordo que pode dar fim definitivo ao processo. Por exemplo, há processos trabalhistas¹ que o montante envolvido chega a milhões de Reais, desse modo, o operador do Direito busca cercar-se de informações sobre decisões passadas do magistrado designado ao caso e sobre como ele decidiu casos anteriores similares antes de tomar uma decisão em um acordo de milhões de Reais.

Nesse contexto, o problema de negócio em questão que pressiona advogados é como saber com confiança qual a tendência de opinião de um Tribunal, Turma Recursal ou magistrado que vai julgar determinado processo. Não seria interessante tomar uma decisão desse patamar baseado somente em uma opinião subjetiva. Portanto, nessa situação, o analista jurídico procura se cercar do máximo de evidências para fundamentar sua decisão em aceitar ou não um acordo de milhões de reais, por exemplo. Desse modo, para realizar essa busca por evidências, advogados gastam horas realizando pesquisas jurisprudenciais para fundamentar suas petições utilizando páginas de busca na internet.

Dessa forma, no Brasil novas soluções tecnológicas estão sendo implementadas com intuito de aumentar a eficiência judiciária brasileira. Assim, em 2011 o Conselho Nacional de Justiça iniciou a implantação do Processo Judicial Eletrônico (PJe), o qual permite a tramitação do processo de maneira digital em sistema computacional (MONTE-NEGRO, 2016). Atualmente, mais de 8 milhões de processos estão em tramitação nesse

¹ <https://www.trt4.jus.br/portais/trt4/modulos/noticias/149267>

sistema, e mais de 100 milhões de processos estão em andamento na justiça, de maneira parcialmente digital. Desse modo, é possível observar como o corpus de documentos digitais mantidos já demanda grande quantidade de armazenamento, sem contar todos os novos documentos e decisões judiciais publicadas diariamente. Entretanto, a exploração dessa massa de documentos judiciais não é trivial pois tais documentos encontram-se disponibilizados em texto puro sem o enriquecimento necessário para a extração de conhecimento de valor. Desse modo, é possível verificar esforços de pesquisadores como Salama et al. (2011), que realizaram anotação e extração de informações de milhares de decisões judiciais de maneira manual. Além disso, Salama, Carlotti e Yeung (2018) estão desenvolvendo pesquisa semelhante a esta, explorando apenas a localização de palavras-chave nas decisões judiciais, entretanto com foco na previsibilidade das decisões. Por outro lado, é possível observar no contexto internacional movimento semelhante de divulgação de documentos judiciais a fim de permitir o processamento computacional e análise dos dados. Por exemplo, na China a partir de 2016, o governo iniciou plano de uso de tecnologias de *Big Data* e *Blockchain* a fim de otimizar o processo judicial, conhecido como *Smart Court* (SHI; SOURDIN; LI, 2021). Assim, o acesso a esses dados permitiu a realização de pesquisas (SONG et al., 2019; FANG et al., 2020; LEI et al., 2017) com objetivo de extrair conhecimento oculto dos corpus de documentos, entretanto, os pesquisadores esbarraram em dificuldades devido à pouca quantidade de dados rotulados.

Assim, considerando o problema de negócio da área do Direito em relação à dificuldade de análise manual de decisões judiciais em busca de conhecimento oculto sobre a tendência de opinião de magistrados, foi aplicada abordagem metodológica de acordo com CRISP-DM Guide 1.0 (CHAPMAN et al., 2000) para a descoberta de conhecimento em base de dados para o processamento de milhares de decisões judiciais de dois Tribunais Regionais do Trabalho brasileiros utilizando técnicas de Supervisão Fraca *Weak Supervision* de forma a realizar a classificação automática dos documentos por meio de Aprendizagem de Máquina Supervisionada. Dessa maneira, foi possível realizar análise estatística para responder a questão de pesquisa estabelecida: **Seria possível que os tribunais avaliados e suas turmas recursais julguem favoravelmente proporção significativamente maior de recursos para uma das partes do que para outra em média?**

Dessa maneira, este trabalho constitui na execução de pesquisa de natureza aplicada, pois busca desenvolver conhecimento que pode ser utilizado para aprimorar processo de tomada de decisões em empresas e em escritórios de advocacia e tem como objetivo geral responder a hipótese estabelecida. Além disso, em função de que as bases

de dados analisadas não apresentam dados rotulados disponíveis ao público, considera-se como diferencial a construção de uma base de dados padrão-ouro e a realização de experimento de modo a avaliar a performance de técnica de Supervisão Fraca com documentos jurídicos. Enfim, a análise estatística das decisões proporcionará conhecimento em relação à proporção de julgamentos favoráveis às partes empregados e empresas de dois tribunais brasileiros, o que permitirá a avaliação estratégica da melhor forma de se levar a pretensão ao órgão judicial de acordo com os dados obtidos.

Os objetivos específicos são os seguintes:

- realizar análise exploratória dos documentos coletados na Web;
- realizar anotação manual de base de decisões judiciais em relação à parte favorecida a fim de criar uma base de dados padrão-ouro;
- avaliar a performance de algoritmos de classificação baseado em aprendizagem de máquina supervisionado;
- avaliar a performance de técnica de Supervisão Fraca com documentos jurídicos;
- classificar quantidade significativa de decisões judiciais de modo a realizar análise quantitativa;
- aplicar testes estatísticos a fim de verificar a existência de diferenças de proporção de julgados, como também o nível de força e o nível de efeito.

Assim, o trabalho está estruturado da seguinte forma: primeiramente, a Seção 2 apresenta a base teórica deste trabalho por meio da discussão dos principais conceitos relacionados ao Direito e também apresenta conceitos relacionados a Aprendizado de Máquina com dados limitados e como essas técnicas podem ser usadas para trabalhar com dados não rotulados. A seguir, a Seção 3 expõe os trabalhos relacionados que apresentam técnicas de processamento de documentos jurídicos. Após, a Seção 4 introduz os conceitos metodológicos utilizados neste trabalho. Então, a Seção 6 detalha e experimentação realizada bem como os resultados encontrados. Finalmente, a Seção 8 apresenta as considerações finais, incluindo limitações e discussão de trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados os principais conceitos relacionados à pesquisa de modo a auxiliar a sua compreensão. Assim, primeiramente é realizada uma breve explanação quanto à estrutura do judiciário de modo a contextualizar o leitor em relação à posição hierárquica dos tribunais foco da pesquisa, como também é levantada breve consideração quanto à questão da livre discricionariedade para julgar outorgada pela lei aos magistrados. Após, são apresentados conceitos relacionados aos principais algoritmos e técnicas de aprendizagem de máquina utilizados na pesquisa.

2.1 Fundamentos teóricos relacionados ao Direito

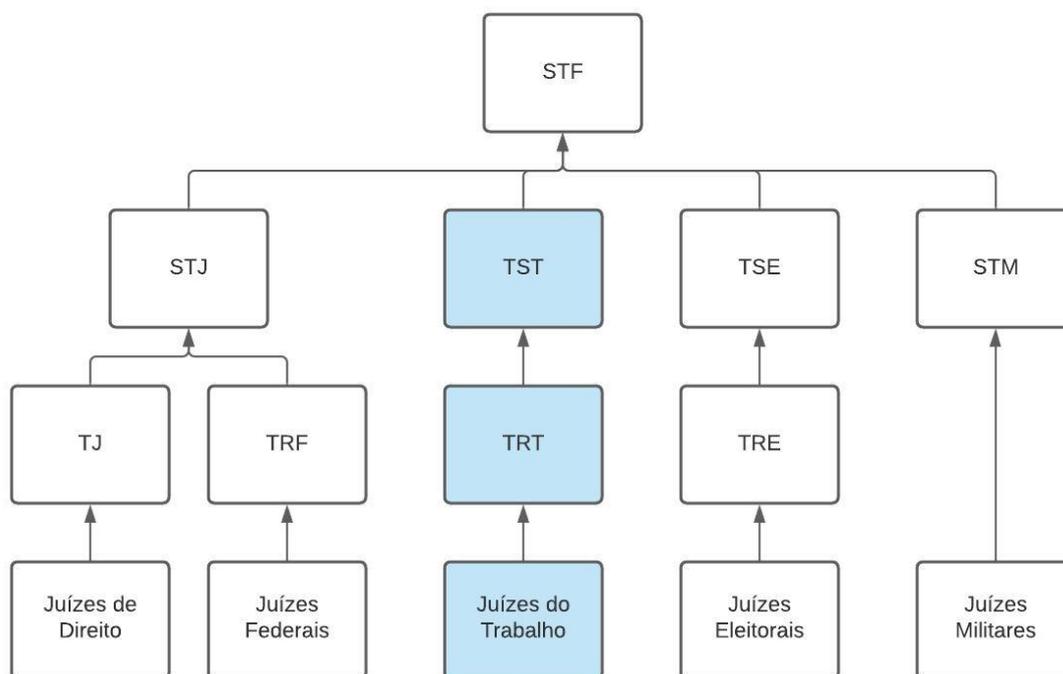
Pelo fato deste trabalho ser aplicado à uma área não muito tradicional da Computação, nesta subseção são apresentados alguns conceitos relevantes para a compreensão do contexto e de sua aplicação.

2.1.1 Organização Judiciária

O Poder Judiciário brasileiro foi organizado pela Constituição Federal de 1988, Capítulo III (BRASIL, 1988), de modo a conter estruturas verticais e horizontais especializadas em cada matéria de Direito, a saber, Justiça do Trabalho, Justiça Eleitoral e Justiça Militar. Também há a Justiça Federal, especializada em causas cíveis, criminais e trabalhistas em que a União é parte, e a Justiça dos Estados, que tem competência residual, ou seja, tem competência para julgar causas que não forem explicitamente de competência de outras justiças especializadas (Figura 2.1). Assim a Justiça do Trabalho tem por função julgar especialmente as ações que envolvam relações de trabalho entre empregados e empregadores. Desse modo, esse poder é composto de diversos órgãos que auxiliam a alcançar seu objetivo de promover a Justiça na sociedade brasileira.

Tal estrutura é formada de modo que o processo legal seja executado de maneira a prover diversos graus de verificação da aplicação das leis. Assim, quando o processo inicia, ele deve ser apresentado a um único juiz o qual analisará os documentos apresentados e poderá exigir a execução de diligências de modo a permitir a aferição da verdade na situação em questão. Essa fase é considerada o primeiro grau de jurisdição. Por fim,

Figura 2.1: Organização judiciária brasileira conforme a Constituição Federal do Brasil, Capítulo III



Fonte da imagem: o autor.

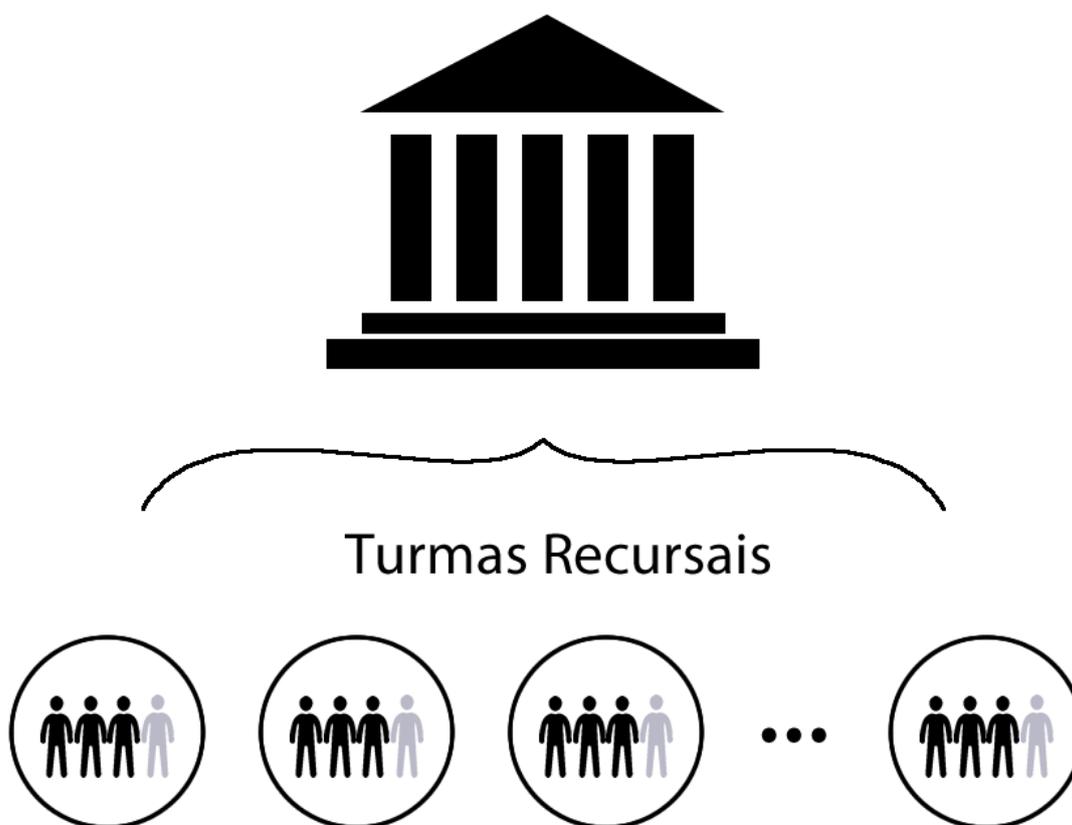
esse magistrado proferirá um julgamento chamado de sentença.

Por outro lado, é possível que os cidadãos e empresas envolvidos no processo legal, chamados de partes, possam não estar satisfeitos com a decisão do juiz de primeira instância. Assim, existe a possibilidade de que o empregado, conhecido como reclamante, ou a empresa, conhecida como reclamada, promovam um recurso ao tribunal competente. Nesse Tribunal, diferentemente, a causa será analisada e julgada por não apenas um magistrado, mas por três. Essa decisão é chamada de acórdão, a qual representa a construção coletiva (GUIMARÃES, 2011) dos magistrados do tribunal, chamados de desembargadores.

A nível de segundo grau de jurisdição, cada Tribunal do Trabalho é composto de diversas Turmas Recursais as quais são compostas por 4 desembargadores. Tais Turmas recebem aleatoriamente os processos em que as partes fazem recursos ao Tribunal de modo que não seja possível para uma parte escolher quais magistrados irão julgar seu processo. Além disso, os 4 desembargadores trabalham em regime de revezamento de modo que em cada processo apenas 3 façam parte do julgamento, Figura 2.2. Além disso, existem casos em que é possível impetrar recursos dirigidos ao TST ou ao STF, entretanto, de

modo geral, esses casos fazem parte de um fluxo extraordinário de processos judiciais¹. Por exemplo, processos de dissídios coletivos devem ter o seu primeiro julgamento diretamente a nível de tribunal por uma seção especializada e seus recursos são julgados pelo TST. Esses tipos de processos com fluxos especiais não são analisados em detalhes nessa pesquisa em vista de não fazerem parte do escopo estipulado pela questão de pesquisa 6.1.1.

Figura 2.2: Organização judiciária dos Tribunais Regionais do Trabalho conforme a Constituição Federal do Brasil, Capítulo III



Fonte da imagem: o autor.

Assim, essa estrutura permite que o cidadão trabalhador que sentir que teve seus direitos lesados possa recorrer à Justiça do Trabalho de modo a procurar a correta aplicação das normas trabalhistas, como, por exemplo, pagamento de valores de férias em atraso, décimo terceiro salário ou até mesmo adicionais de insalubridade ou periculosidade. Por outro lado, as empresas também podem iniciar a ação judicial, entretanto, nesse caso, de modo geral o objetivo é depositar em juízo previamente valores devidos ao funcionário que não os recolheu previamente por qualquer motivo.

¹<<https://www.trt4.jus.br/portais/trt4/estrutura>>.

2.1.2 Acórdão

Acórdão é o nome do documento em que os magistrados de certo tribunal proferem à sua decisão de maneira colegiada (GUIMARÃES, 2011). Ou seja, três magistrados apresentam a sua opinião sobre determinado assunto trazido ao tribunal por meio de um recurso, na maioria dos casos, e condensam o seu julgamento na forma de um acordo de caráter decisório. Assim é proferida a jurisdição em nível de segundo grau pelos tribunais especializados, como é o caso da Justiça do Trabalho.

Entretanto, há situações em que a decisão não analisa efetivamente os pedidos contidos nos recursos, nesse caso é considerado que não houve "resolução de mérito", conforme Artigo 487 do Código de Processo Civil (BRASIL, 2015). Esses casos em que não houve resolução de mérito podem ocorrer devido a, por exemplo, anulação da sentença a que as partes fazem recurso devido a falha grave cometida a nível de primeiro grau ou então houve declaração de incompetência da Justiça do Trabalho em função de que o objeto da ação deve ser julgado por outro ramo da Justiça.

Quanto à estrutura do documento, o acórdão é composto de diversos elementos que são requisitos essenciais: protocolo inicial, relatório, fundamentação, dispositivo e protocolo final, conforme Artigo 489 do Código de Processo Civil (BRASIL, 2015; GUIMARÃES, 2011). Além disso, é comum os magistrados seguirem um padrão de vocabulário comum e utilizarem as mesmas palavras para representar o início de cada seção e recursos de estilo de texto como letras maiúsculas de modo a facilitar a leitura pelos advogados e pelas partes dos documentos. Por exemplo, geralmente os magistrados adotam a palavra *RELATÓRIO* e *FUNDAMENTAÇÃO* para representar o início dessas seções no texto.

Para fins de análise documentária, observa-se que a praxe judiciária fornece algumas expressões-chave para a identificação específica do relatório no corpo do acórdão, tais como as expressões iniciais: Vistos, relatados e discutidos, ou simplesmente, Vistos etc. e, como fecho, expressões como É o relatório, É como relato, Era o que cumpria relatar etc. (GUIMARÃES, 2011).

De modo a permitir a visualização do formato real desse documento, foi disponibilizado um acórdão de exemplo que pode ser encontrado no Apêndice A.1. A seguir são descritos em detalhes os elementos que o compõem:

- Protocolo inicial: diz respeito à descrição e identificação do documento com dados, como, por exemplo, a identificação do Tribunal, Turma Recursal, número do processo.

- **Relatório:** é um resumo do processo judicial em que o magistrado descreve os acontecimentos dos principais fatos que ocorreram relacionados com a questão trazida a julgamento como também os principais elementos do processo que serão utilizados para a construção da lógica de argumentação para definição do Direito em questão na seção de Fundamentação e Dispositivo.
- **Fundamentação:** é composta dos argumentos utilizados para se chegar à conclusão lógica da decisão final que será proferida no Dispositivo. Além disso, são analisados todos os argumentos relacionados a cada um dos pedidos de modo a trazer uma resposta e uma decisão lógica baseada nos fatos relatados no relatório.
- **Dispositivo:** é a conclusão lógica do disposto no Relatório e na Fundamentação em poucos parágrafos, de caráter sucinto, de modo a acolher ou não o recurso. Assim, o magistrado resolve as questões levantadas e apresenta sua opinião final em relação ao recurso inteiro de modo geral, utilizando tradicionalmente as palavras “dar provimento”, “dar provimento parcial” ou “negar provimento”.

[...] o Dispositivo inicia-se por locuções como: O que posto..., Isto posto, decidem..., Ante o exposto, acordam..., Pelo que acordam..., Fundamentos pelos quais acordam... e revela a essência do acórdão [...] (GUIMARÃES, 2011).

2.1.3 Magistrados e discricionariedade

Conforme Vieira (2015), o governo tem para si a autoridade de julgar os conflitos sociais e investiu os magistrados com o poder suficiente para a tomada de decisões no formato de processos judiciais. Além disso, os deputados e senadores têm a responsabilidade de redigir as normas e leis que regem o trabalho dos juízes e também instruem como as leis devem ser aplicadas a cada caso. Dessa maneira, os juízes têm o seu poder determinado e limitado pelas normas legais. Entretanto, há situações em que não há no ordenamento jurídico legislação que seja aplicável ao caso trazido a julgamento ou pode haver até mesmo lacunas na lei que a tornem de difícil aplicação devido à pouca clareza. Nesses casos, os magistrados acabam por ter que analisar o caso concreto sob diversos ângulos subjetivos éticos e morais e a sentenciar utilizando o poder discricionário investido, entretanto não necessariamente de arbitrariedade, mas de valoração e balanceamento.

Logo, diante da crescente complexidade social e da velocidade da sua transformação, o legislador vai sempre conceder um grau de discricionariedade às concretizações da realidade, para garantir uma decisão correta no caso concreto. Não há como negar, também, que há situações submetidas ao pleito do

Judiciário para as quais existem alternativas de decisões válidas e legítimas, de maneira que o julgador, baseado nas peculiaridades do caso concreto, bem como nas suas convicções político-sociais, poderá optar livremente pela hipótese que lhe parece ser a mais coerente, vale dizer, que melhor lhe convencer, considerando que a sua escolha será feita sempre de forma motivada. Este instrumental mais flexível, e, portanto, adjudicador de maior poder aos juízes, vem-se mostrando crescentemente como uma necessidade nos dias atuais. É importante realçar que a discricionariedade conferida ao magistrado possui uma zona de abrangência, limite, ou seja, sua liberdade de convencimento não poderá extrapolar os limites do razoável, do proporcional, do exigível, do adequado. Na verdade, a valoração feita pelo juiz para a solução de questões não prescritas expressamente no ordenamento jurídico pátrio deve coadunar-se com as concepções sociais vigentes e dominantes, sendo que seus critérios pessoais não haverão de conflitar com o que se considera padrão na sociedade em que se vive (VIEIRA, 2015).

2.2 Mineração de dados

Mineração de dados é o estudo e a aplicação de técnicas e métodos que compreendem a coleta, limpeza, processamento e análise de dados com o objetivo de extrair conhecimento oculto, conforme definição de Aggarwal (2015). Assim, podem ser empregadas diversas técnicas computacionais para extração e preparação de dados de modo a serem realizadas análises estatísticas como também a serem realizados processamentos com algoritmos de Aprendizado de Máquina. Enfim, a Mineração de Dados é um termo usado para identificar um grande conjunto de técnicas cujo objetivo é extrair novos conhecimentos de dados que tragam valor aos negócios e a pesquisas.

2.3 Aprendizado de máquina

De acordo com Ray (2019), o Aprendizado de Máquina é uma subárea da Inteligência Artificial em que um algoritmo computacional realiza uma tarefa e sua performance melhora de acordo com a sua experiência por meio do uso de dados. Assim, de acordo com os tipos de dados disponíveis, existem diferentes categorias de algoritmos que podem explorá-los. Aprendizado de máquina supervisionado explora bases de dados em que há um conjunto de observações contendo dados de entrada com características e dados de saída de modo a gerar um modelo representativo ou preditivo das observações. Por outro lado, algoritmos de Aprendizado de Máquina Não-supervisionado exploram bases de dados que não contém rótulo, ou seja, não contém os dados de saída, conforme descrito previamente. Assim, esse tipo de algoritmo busca encontrar estruturas latentes

aos dados, como características identificadoras e subgrupos (RAY, 2019).

2.3.1 Aprendizado de máquina supervisionado

Aprendizado de máquina supervisionado é uma tarefa em que um algoritmo computacional é utilizado para realizar o processamento de uma base de dados composta de instâncias ou de observações de um fenômeno formadas em pares. Tais pares são compostos por características das instancias e um rótulo no formato de um metadado. Desse modo, um algoritmo pode analisar os atributos das instancias e relacionar ao rótulo apresentado de maneira conjunta em todo o *dataset* de modo a identificar padrões e a abstrair em um modelo computacional. Enfim, tal modelo pode ser utilizado para realizar previsões de metadado em instancias similares (BRAMER, 2007; RUSSELL; NORVIG, 2002; PUSTEJOVSKY; STUBBS, 2012).

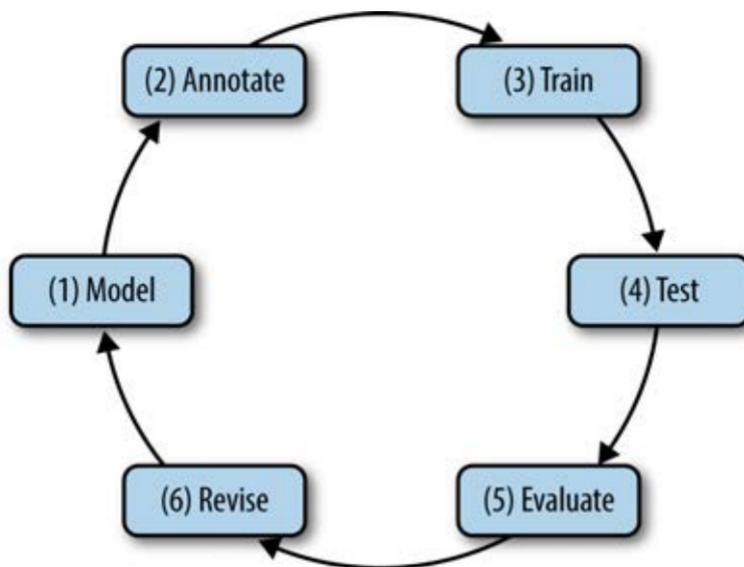
2.4 Anotação de bases de documentos de texto

De acordo com Pustejovsky e Stubbs (2012) e Ide e Pustejovsky (2017), a adição de metadados a documentos de texto de modo a permitir a sua utilização por meio de algoritmos de aprendizado de máquina supervisionado é chamada de anotação de documentos. Assim, esses autores expõem uma metodologia chamada de ciclo de desenvolvimento “MATTER”, a qual apresenta diversas fases iterativas cujo objetivo é a aplicação de metadados a documentos de texto. Tal metodologia é composta das seguintes fases: *Model, Annotate, Train, Test, Evaluate, Revise*, Figura 2.3. O processo é executado fase após fase, na ordem apresentada, entretanto, segundo os proponentes, é comum haver retorno aos passos anteriores para a realização de revisões, pois, conforme os pesquisadores executam a tarefa de modelagem e de anotação, eles vão criando novos entendimentos sobre o assunto, os quais permitem uma elaboração mais aprofundada das fases anteriores. Enfim, a seguir são apresentadas brevemente cada uma das fases.

- *Model*: apresenta descrição dos elementos que compõem o metamodelo que será utilizado para a realização da anotação dos documentos. Assim, Pustejovsky e Stubbs (2012) definem que o modelo consiste de um vocabulário de termos T , do relacionamento entre esses termos R , e da interpretação I :

[...] we will define a model as consisting of a vocabulary of terms, T ,

Figura 2.3: Metodologia de desenvolvimento de bases de texto anotadas chamada ciclo *MATTER*



Fonte: Pustejovsky e Stubbs (2012)

the relations between these terms, R, and their interpretation, I. So, a model, M, can be seen as a triple, $M = \langle T, R, I \rangle$ [...] (PUSTEJOVSKY; STUBBS, 2012).

Além disso, é importante a construção de um documento chamado *Diretrizes de Anotação* o qual apresenta de maneira didática como os rótulos devem ser aplicados aos documentos, além de exibir, por exemplo, casos controversos de anotação e a respectiva anotação correta conforme interesse da pesquisa.

- *Annotate*: consiste na aplicação dos rótulos aos documentos de acordo com o modelo estipulado e com suas Diretrizes de Anotação. Tal atividade, de modo geral, precisa ser realizada por humanos pois o objetivo geral sob a perspectiva da Inteligência Artificial é que o computador *aprenda* a realizar a mesma tarefa automaticamente como pessoas fariam por meio do algoritmo de aprendizado de máquina supervisionado. Assim, recomenda-se a realização da anotação de cada documento por pelo menos duas pessoas. Entretanto, as pessoas selecionadas para realizar a anotação podem discordar em relação ao rótulo correto de determinado documento. Dessa maneira, Pustejovsky e Stubbs (2012) sugerem o cálculo de um indicador que represente o nível de acordo entre anotadores. Esse indicador permite a verificação da consistência dos rótulos assinalados. Nesse caso, um valor alto desse índice indica que os anotadores concordaram na maioria dos rótulos. Finalmente, o pesquisador deve analisar cada instância anotada que teve discordância no rótulo

aplicado e fazer a anotação que julgar correta manualmente, esse processo é chamado de *Adjudicação*. Após essa fase, a base de documentos anotada está pronta para ser utilizada com algoritmos de aprendizado de máquina supervisionado e é chamada de *Gold Standard*.

- *Train*: Consiste na utilização da base de documentos de texto adjudicada junto a algoritmos de aprendizado de máquina supervisionado para a realização de treinamento. Nessa fase, é utilizada apenas uma amostra da quantidade total da base, deixando uma parte para a realização de testes na fase seguinte.
- *Test*: Consiste na verificação do modelo de algoritmo de aprendizado de máquina criado na fase anterior com a amostra selecionada para testes.
- *Evaluate*: Os documentos anotados na fase anterior são então avaliados com a utilização de métricas-padrão comumente utilizadas em tarefas de aprendizado de máquina, como, por exemplo, acurácia, precisão e revocação.
- *Revise*: Por fim é realizada uma revisão geral dos modelos criados e das diretrizes de anotação de modo a anotação final da base de documentos de maneira mais confiável possível.

2.4.1 Índice para avaliação de acordo entre anotadores

Conforme exposto, a fase de anotação pode gerar uma base de documentos de texto rotulados inconsistente, ou seja, pode haver instâncias anotadas com rótulos diferentes por cada um dos anotadores. Desse modo, Pustejovsky e Stubbs (2012) propõem a utilização do índice *Cohen's Kappa* (k) para a verificação do nível de concordância entre anotadores, de acordo com a Fórmula 2.1. Além disso, apresentam breve orientação sobre a interpretação do valor encontrado, indicando que valores acima de 0,8 podem ser considerados com excelente nível de concordância.

$$K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \quad (2.1)$$

- $Pr(a)$: é a concordância relativa observada entre os anotadores.
- $Pr(e)$: é a expectativa de concordância aleatória entre os anotadores.

No Capítulo 6.3.3 é apresentada uma demonstração detalhada do cálculo do índice *Cohen's Kappa* (k) em função da base de textos anotada criada para a presente pesquisa. Além disso, existe a possibilidade de se empregar metodologia a qual permite a anotação

automática de documentos por meio de algoritmo computacional, a qual é explorada no Capítulo 2.6.

2.5 Algoritmos de aprendizagem de máquina supervisionada

Nesta seção serão descritos os principais algoritmos de aprendizado de máquina supervisionado.

2.5.1 *K*-Vizinhos-mais-próximos (*K-nearest Neighbor*)

Este classificador analisa k observações mais próximas do caso a ser analisado e assinala a classe que a maioria dessas observações pertence. O cálculo da distância pode ser operacionalizado computacionalmente com diversas medidas de distância, como, por exemplo, a distância Euclidiana, Manhattan, Minkowski ou Ponderada (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

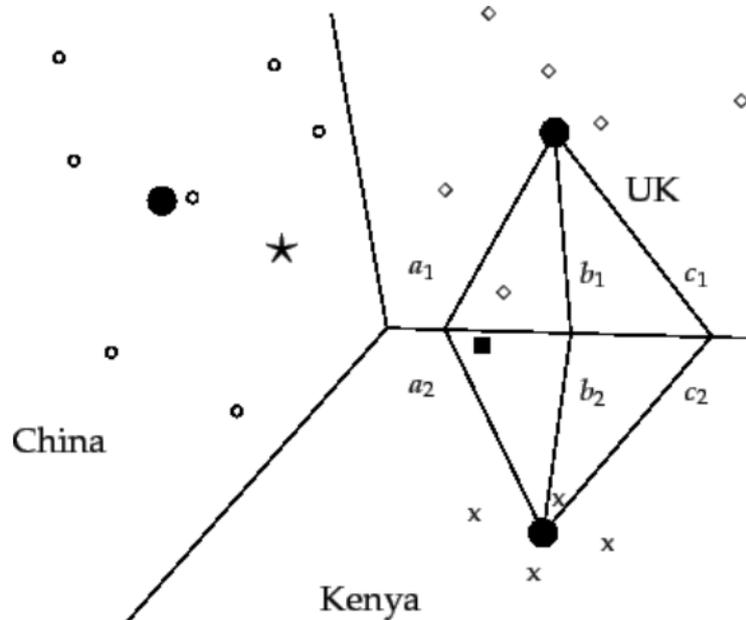
2.5.2 Classificador Rocchio ou Centroide mais próximo

A implementação do algoritmo do Centroide mais próximo pela biblioteca Scikit Learning² tem funcionamento similar ao *K*-Vizinhos-mais-próximos, entretanto, a observação a ser classificada recebe o rótulo do centroide mais próximo. O centroide, por sua vez, é calculado em função da média dos vetores da base de treinamento para cada classe. Assim, os documentos podem ser vetorizados por meio de $TF - IDF$ ³ e a sua distância do centroide mais próximo pode ser calculada por meio da distância Euclideana. Na Figura 2.4, é possível observar os círculos preenchidos pretos que representam os centroides das classes e uma estrela que representa uma observação a ser classificada, nesse caso, com o rótulo *China*.

²<<https://scikit-learn.org/stable/modules/neighbors.html>>

³<https://scikit-learn.org/stable/modules/feature_extraction.html>

Figura 2.4: Ilustração sobre o funcionamento do algoritmo de classificação Rocchio. É possível observar os círculos preenchidos pretos que representam os centroides das classes e uma estrela que representa uma observação a ser classificada, nesse caso, com o rótulo *China*



Fonte: Schütze, Manning e Raghavan (2008)

2.5.3 Classificador Naive Bayes

Classificador probabilístico que utiliza o teorema de Bayes assumindo que todas as palavras são independentes entre si, ou seja, não há correlação entre as palavras presentes nos documentos. Isso permite o cálculo da probabilidade de um documento ser de determinada classe baseado na frequência das palavras presentes nos documentos em determinada classe na base de treinamento (PEDREGOSA et al., 2011; SCHÜTZE; MANNING; RAGHAVAN, 2008). Ou seja, as palavras nos documentos de determinada classe servem de evidência para o cálculo da probabilidade de um novo documento corresponder a certa classe.

2.5.4 Árvore de decisões (*Decision Tree*)

Classificador que utiliza organização no formato de árvore para representar hierarquicamente uma estrutura de nodos contendo decisões baseadas em questões booleanas (GRUS, 2019). As questões booleanas contidas nos nodos são geradas a partir de algoritmos de otimização cujo objetivo é encontrar *features* que discriminem da melhor maneira

possível as observações⁴.

2.5.5 Florestas aleatórias (*Random Forest*)

Este classificador utiliza o algoritmo de Árvore de decisões de modo a construir diversas árvores diferentes compostas de *features* selecionadas aleatoriamente a partir de uma subparte da base de treinamento. Assim, são geradas diversas classificações diferentes as quais são utilizadas para determinar a saída final do algoritmo, por meio de um *ensemble* de árvores de decisões profundas. Tem por principal característica apresentar generalização melhor que o algoritmo Árvore de decisões (PEDREGOSA et al., 2011; LOUPPE, 2014; HORNING, 2013).

2.5.6 Classificador Gradient Tree Boosting

Este classificador utiliza internamente o algoritmo de Árvore de Decisões múltiplas vezes, cada uma com *features* diferentes de maneira combinada, de forma a montar um *ensemble* de modelos a fim de alcançar uma maior performance do que apenas uma Árvore de decisões sozinha. Assim, o algoritmo usa o *Gradiente Decendente* de forma a minimizar uma função de perda que leva em consideração todos os modelos de Árvores de Decisões sendo otimizados. Ao mesmo tempo, o algoritmo dá mais importância a *features* que são mais utilizadas e vai inserindo novas árvores no *ensemble*. Por outro lado, as novas árvores inseridas buscam superar deficiências encontradas nos modelos adicionados anteriormente, de modo que elas desempenhem melhor em instâncias que foram classificadas incorretamente pelos modelos anteriores.

2.5.7 Máquina de Vetores de Suporte (*Support Vector Machine*)

Classificador que utiliza a projeção dos dados em um espaço ortogonal de modo a encontrar um hiperplano que separe os dados considerando a distância máxima possível dos pontos mais próximos de cada classe. Além disso, este classificador pode apresentar boa performance mesmo contendo alto viés durante a fase de treinamento, pois as margens maximizadas podem incluir classes incorretamente classificadas, o que diminui a

⁴São utilizadas técnicas como *Gini Impurity*, cálculo de entropia e redução de variância.

variância na fase de treinamento (SCHÜTZE; MANNING; RAGHAVAN, 2008; GRUS, 2019). Dessa maneira, o algoritmo vai realizando testes em dimensões cada vez maiores de modo a encontrar um plano que se adéque aos dados apresentados, entretanto, em dimensões muito altas o algoritmo pode ter performance similar ao algoritmo de *K*-Vizinhos-mais-próximos.

2.6 Aprendizado de Máquina com dados limitados

De acordo com Ratner et al. (2019) e Kurzweil (2013), antes do advento da popularização do Aprendizado de Máquina nas últimas décadas, muitos sistemas de Inteligência Artificial eram desenvolvidos utilizando bases de dados de conhecimento e com heurísticas ou regras programadas em conjunto com especialistas das áreas de domínio de modo a realizar o desenvolvimento manual de *features* ou a programar sistemas especialistas.

Com o desenvolvimento da computação e da internet, houve uma explosão na quantidade de conteúdos e bases de dados que puderam ser anotadas por especialistas e que puderam ser utilizadas em conjunto com algoritmos de Aprendizado de Máquina supervisionado para a produção de modelos de Inteligência Artificial. Tais sistemas utilizam modelos de aprendizado automático que permitem a extração de *features* dos documentos anotados manualmente sem a necessidade de intervenção de especialistas na fase de treinamento (KURZWEIL, 2013; RATNER et al., 2019).

Desse modo, o estudo realizado por Alexander Wissner-Gross (2016) afirma que há evidências de que a maior disponibilidade de bases de dados anotadas foi responsável pela grande evolução observada na área de inteligência artificial. Assim, em seu estudo, são relacionados diversos casos. Por exemplo, há uma pesquisa realizada em 1994 que alcançou performance comparada a humana em reconhecimento de fala utilizando um banco de dados de artigos do *Wall Street Journal*. Outro exemplo foi o caso do tradutor do Google que em 2005 alcançou alta performance tirando vantagem de uma base de documentos de 1,8 trilhões de *tokens*.

Entretanto, Pustejovsky e Stubbs (2012) alegam que a anotação de bases de documentos pode consumir muito tempo, além de que, dependendo da tarefa, pode necessitar de especialistas de alto nível de escolaridade. Por exemplo, anotação de documentos de bases de dados de prontuários médicos pode necessitar a alocação de diversos especialistas. Por outro lado, com o tempo, diversas técnicas começaram a ser desenvolvidas de modo a superar esses problemas e a alavancar dados não anotados nas tarefas de treina-

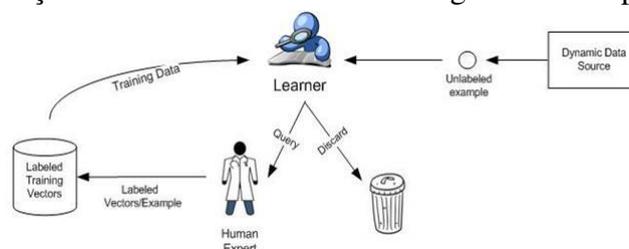
mento de modelos de Aprendizado de Máquina (RATNER et al., 2019).

Assim, pesquisadores começaram a desenvolver diversas técnicas que permitem aos software *aprender* utilizando dados já disponíveis, sejam anotados ou não. Por exemplo, no caso de sistemas que utilizam imagens, uma técnica comum usada é a de rotacionar, inserir ruído ou distorções de modo a simular as condições em que as fotos podem ser encontradas em tarefas reais. Essa técnica permite utilizar uma base de dados já anotadas e a multiplicá-la de acordo com os efeitos aplicados. Enfim, nas seções a seguir são apresentadas diversas técnicas desenvolvidas que permitem a alavancagem de dados anotados e não anotados de modo a gerar mais dados anotados a baixo custo (RATNER et al., 2019).

2.6.1 Aprendizado Ativo (*Active Learning*)

A técnica de Aprendizado Ativo é utilizada para realizar a anotação de dados por meio de especialistas, entretanto, a seleção das observações que serão anotadas é controlada por meio de algoritmo. Assim, esse algoritmo estima quais são as observações que sejam mais importantes para a construção do modelo de Aprendizado de Máquina e as apresenta aos anotadores. Desse modo, é possível tirar o maior benefício possível do tempo despendido para a anotação dos documentos, de forma que, mesmo com poucos dados anotados já é possível ter noção dos níveis de acurácia do modelo criado (SHARF; RAZZAK, 2017; RATNER et al., 2019) – ver Figura 2.5.

Figura 2.5: Ilustração sobre o funcionamento do algoritmo de Aprendizado Ativo.



Fonte: Sharf e Razzak (2017)

2.6.2 Aprendizado Semi-supervisionado (*Semi-supervised Learning*)

Aprendizado semi-supervisionado é uma técnica que utiliza uma pequena base de dados anotada em conjunto com uma grande base de dados não anotada. Por meio de técnicas que geram suposições em cima dos dados anotados em relação a toda a base de dados não anotada, é possível considerar que determinadas observações não anotadas sejam anotadas com um rótulo suposto. Assim, por exemplo, no caso de técnicas que utilizem um espaço vetorial e utilizem técnicas de clusterização, observações concentradas em um aglomerado tendem a compartilhar o mesmo rótulo (ENGELEN; HOOS, 2020; RATNER et al., 2019).

2.6.3 Transferência de Aprendizado (*Transfer Learning*)

A Transferência de Aprendizado tem por objetivo utilizar modelos já treinados em uma tarefa relacionada em outra base de dados para o desenvolvimento de um novo modelo. Essa técnica é especialmente utilizada com Redes Neurais de modo que o treinamento do novo modelo não inicie com *features* zeradas ou geradas de maneira aleatória, (RATNER et al., 2019).

2.6.4 *Few-Shot Learning*

De acordo com Wang et al. (2020), esta técnica busca alcançar bons níveis de performance de aprendizado de máquina utilizando base de treinamento com muito poucas instâncias, ou, até mesmo, nenhuma. Diferentemente dos algoritmos tradicionais que criam modelos computacionais com vistas a realização de generalizações além das instâncias usadas para treinamento, os principais algoritmos de *Few-Shot Learning* buscam computar as diferenças e similaridades entre as instâncias usadas para treinamento e para testes. Assim, são construídos, por exemplo, *Embeddings* por meio de redes neurais específicas para cada classe as quais são utilizadas para cálculo da distância Euclidiana com uma nova instância.

2.6.5 Supervisão Fraca (*Weak Supervision*)

De acordo com Varma et al. (2019), Supervisão Fraca é uma técnica de Aprendizado de Máquina que combina instâncias rotuladas por meio de fontes de *baixa qualidade*, como por exemplo, *Crowdsourcing* e *Boosting*⁵, de modo a estimar automaticamente suas acurácias, utilizando um modelo generativo, e gerar assim uma base de instâncias anotadas para ser utilizada para a criação de modelo discriminativo a partir de algoritmos de Aprendizado de Máquina Supervisionado. Assim, esta técnica exige primeiramente a anotação da base de dados a partir de fontes de baixa qualidade e após o seu processamento por um modelo generativo de modo a combinar rótulos sobrepostos, correlacionados ou conflitantes e medir sua acurácia.

Recentemente, conforme observado por Wang, Hoang e Kan (2013), o desenvolvimento da Web 2.0 permitiu a alavancagem de grandes quantidades de pessoas para a realização de tarefas conjuntas por meio de *Crowdsourcing*. Assim, é possível gerar grandes quantidades de documentos anotados, até mesmo de maneira redundante, com custo menor do que o de alocar profissionais especializados para a realização da mesma tarefa. Desse modo, é possível utilizar técnicas de Supervisão Fraca em conjunto com a base de dados criada com anotações redundantes a fim de filtrar automaticamente rótulos conflitantes.

Além disso, outra fonte de *baixa qualidade* é a introduzida pelo pacote Python Snorkel Framework⁶ chamada de *Labels Function*. Por meio dessa técnica são codificadas Funções de Rotulação que utilizam conhecimento de profissionais da área de domínio, por exemplo, de modo a aplicar determinados rótulos a base de dados automaticamente (RATNER et al., 2019). Funções de Rotulação recebem uma observação como argumento e retornam um rótulo correto para aquela instância ou retornam um rótulo que indica que a Função de Rotulação não foi capaz de rotular aquela instância, chamado de *ABSTAIN*. Para realizar o processamento, as funções podem utilizar dados externos, como, por exemplo, um dicionário de dados de modo a inserir o rótulo correto em relação a cada entrada do dicionário encontrada. A Figura 2.6 apresenta ilustração com parte de código Python de Funções de Rotulação.

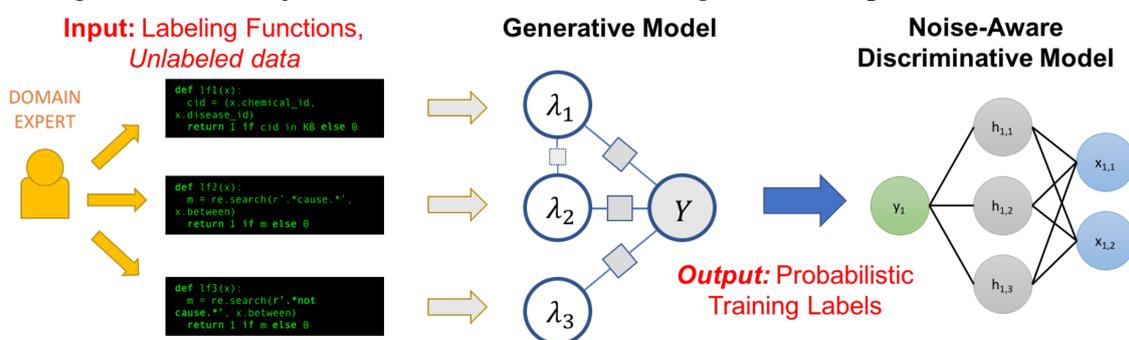
Outros tipos de Funções de Rotulação também podem realizar transformações a

⁵*Boosting* é uma técnica que busca utilizar modelos com acurácia fraca (ou seja, acurácia parecida com suposições aleatórias) a fim de criar um modelo de *Ensemble* com melhor acurácia (KEARNS, 1988). O assunto foi brevemente discutido na Seção 2.5.6.

⁶<<https://www.snorkel.org/>>

dados de entrada já anotados de modo a gerar observações levemente alteradas, mas já anotadas. Essa técnica é muito utilizada para o processamento de imagens como a aplicação de distorções e ruídos, mas também é possível realizar transformações de bases de dados textuais. Por exemplo, há estudos que realizam a substituição de certas palavras do texto pelo seu sinônimo mais próximo a partir de um dicionário, assim, é gerada uma base de dados levemente alterada, mas já anotada (RATNER et al., 2019). No Apêndice A.2 é realizada breve explanação de trabalhos em que houve a aplicação prática da técnica de Supervisão Fraca utilizando o Snorkel Framework em diversas áreas de domínios diferentes, como, por exemplo, médica e análise de clientes, sendo possível observar ganhos significativos em diversas métricas de Aprendizado de Máquina.

Figura 2.6: Ilustração sobre o funcionamento do algoritmo de Aprendizado Ativo.



Fonte: Sharf e Razzak (2017)

2.6.5.1 Algoritmo do Modelo Generativo do Snorkel Framework

De acordo com a documentação do Snorkel⁷ (RATNER et al., 2019; RATNER, 2019), o algoritmo do Modelo Generativo do Snorkel Framework é responsável por processar a matriz contendo o rótulo aplicado por cada uma das Funções de Rotulação a cada uma das instâncias a fim de medir suas acurácias em vista de que tais rótulos podem estar redundantes, sobrepostos, correlacionados ou até mesmo conflitantes. Esse algoritmo não utiliza uma base de dados padrão-ouro para a medição da acurácia, pois o fundamento da técnica de Supervisão Fraca é a de ser aplicada em situações em que essa base não existe ou não está disponível.

De modo a demonstrar a intuição subjacente ao algoritmo generativo desenvolvido pelos autores citados, consideremos ilustrativamente que cada função de rotulação contém o cerne do conhecimento obtido a partir de especialistas da área de domínio e que elas

⁷<https://snorkel.readthedocs.io/en/v0.9.6/packages/_autosummary/labeling/snorkel.labeling.model.label_model.LabelModel.html>

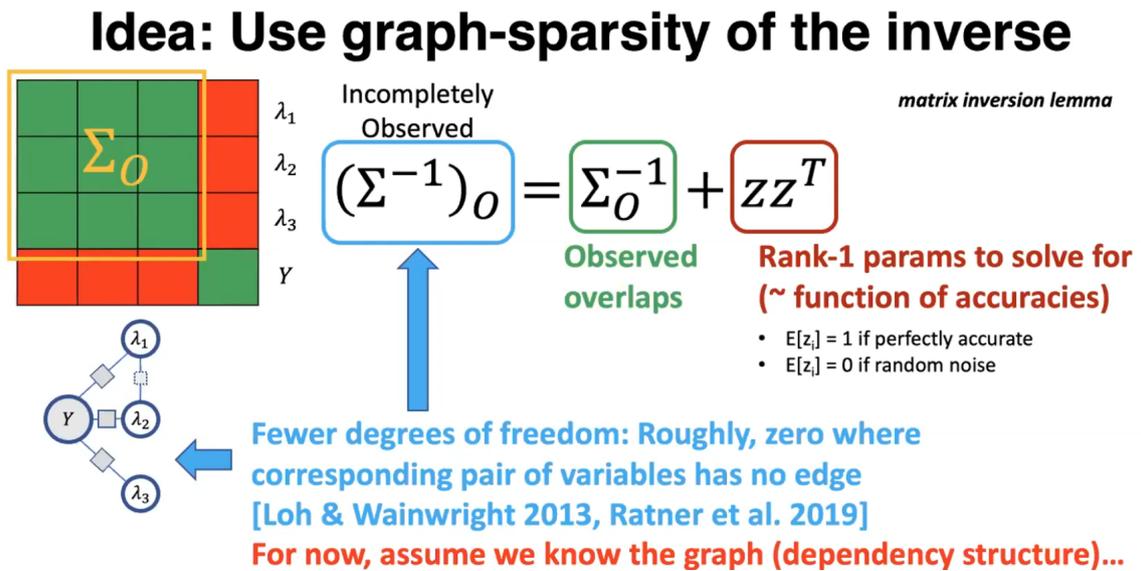
funcionam como se fossem votos. Assim, cada fonte de rotulação acrescenta ao modelo generativo uma parte do conhecimento da área de domínio. Podemos considerar também que cada fonte de rotulação apresenta o seu voto em função de uma perspectiva única da área de conhecimento em relação a instância sendo analisada, ou seja, mesmo que funções de rotulação apresentem votos contrastantes, cada uma delas precisa ser analisada independentemente e assinalada um peso específico no modelo generativo final. Por exemplo, em uma situação hipotética, podem existir funções de rotulação que verifiquem a existência de uma palavra positiva em uma instância, e outra função de rotulação que verifique a existência de uma palavra negativa. Apesar de cada uma aplicar um rótulo inverso a instância, não significa que ambas estejam erradas. Significa que cada uma está inserindo um conhecimento específico no modelo generativo, o qual irá analisar o conjunto da base de dados de modo a aplicar peso ideal a cada fonte de rotulação.

Assim, os autores consideram positivo ao modelo generativo que sejam criadas o maior número possível de funções de rotulação e que cada uma insira uma parte específica de conhecimento da área de domínio. Além disso, o fato de fontes de rotulação diferentes apresentarem o mesmo rótulo em uma certa quantidade de instâncias, ou, até mesmo, apresentarem rótulos contrastantes, acrescenta positivamente ao modelo generativo, pois, conforme as funções de rotulação são aplicadas a cada vez mais dados não rotulados, o erro de generalização do modelo generativo converge assintoticamente na mesma taxa que modelos tradicionais de aprendizado de máquina supervisionado.

Enfim, o objetivo do modelo generativo é calcular a probabilidade de cada função de rotulação aplicar o rótulo correto latente. Assim, a solução baseia no processamento da matriz de covariância das funções de rotulação incluindo também a variável aleatória não observada que contém o rótulo verdadeiro. Assim, de modo simplificado, calculando o inverso da matriz de covariância seria possível estimar os valores de covariância da variável latente, entretanto, isso não é possível de ser calculado diretamente exatamente porque esses valores não são conhecidos. Por outro lado, é possível decompor essa matriz em dois termos, um deles sendo a inversa da matriz de covariância das variáveis aleatórias que conhecemos e o outro é composto de parâmetros desconhecidos, conforme Figura 2.7. A partir disso, é possível calcular uma aproximação dos parâmetros desconhecidos considerando que uma parte suficiente dos valores de covariância entre as variáveis aleatórias das funções de rotulação vão ser zero, conforme Figura 2.8. Assim, é possível calcular algebricamente as acurácias de cada função de rotulação sem a utilização de base de dados padrão-ouro. É importante considerar que foi apresentada uma versão simplificada do

algoritmo generativo do Snorkel Framework que considera que as variáveis aleatórias que representam as funções de rotulação apresentam independência suficiente além de outras suposições consideradas pelos autores em seus estudos.

Figura 2.7: Ilustração da utilização de matriz de covariância para cálculo de acurácias de fontes de rotulação sem a utilização de base de dados padrão-ouro.

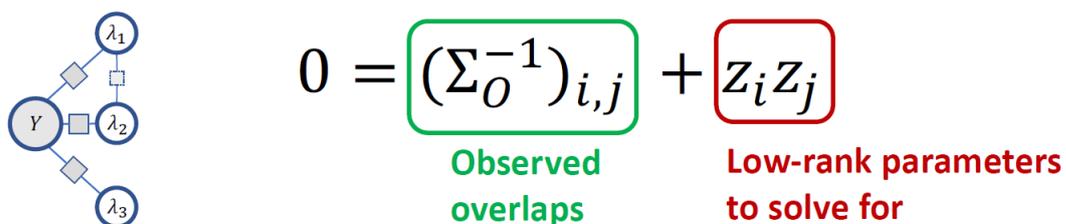


Fonte: Apresentação *Theory and Systems for Weak Supervision* de Christopher Ré no evento The Web Conference 2020.

Figura 2.8: Ilustração da utilização de complementação de matriz para cálculo de acurácias de fontes de rotulação sem a utilização de base de dados padrão-ouro.

Result: A matrix completion problem?

We get a set of equations. For any pair $i \neq j$ with no edge in graph—the lhs is 0



Fonte: Apresentação *Theory and Systems for Weak Supervision* de Christopher Ré no evento The Web Conference 2020.

2.7 Resumo do Capítulo

Nesse capítulo, primeiramente foram apresentados conceitos relacionados ao Direito em virtude de que a presente pesquisa apresenta carácter multidisciplinar, pois o objetivo de negócio proposto diz respeito a criar conhecimento novo a partir de documentos judiciais por meio da utilização de técnicas computacionais. Desse modo, são apresentadas noções sobre organização judiciária e documentos jurídicos a fim de que o leitor compreenda a origem dos documentos sendo processados e a maneira como eles são redigidos pelos magistrados.

Na sequência, são apresentados conceitos fundamentais relacionados a Inteligência Artificial e Aprendizado de Máquina. Nesse sentido, são abordados os principais tipos de algoritmos de Aprendizado de Máquina existentes e é colocado em foco o aprendizado supervisionado. Esse tipo de algoritmo foi considerado o ideal para atender ao objetivo de mineração de dados proposto de classificar decisões judiciais e tem por fundamento a existência de uma base de dados rotulada a ser utilizada para treinamento dos algoritmos. Assim, em vista de não haver previamente a esta pesquisa uma base de decisões judiciais anotadas de acordo com as necessidades do objetivo de mineração, há uma digressão em conceitos metodológicos e em aspectos inerentes ao desenvolvimento de uma base de dados anotada, a qual foi efetivamente desenvolvida na Seção 6.3.3. Além disso, são apresentados conceitos fundamentais em relação aos algoritmos de classificação de texto utilizados de modo a alcançar o objetivo de mineração.

Por outro lado, apesar de a presente pesquisa realizar efetivamente o desenvolvimento de uma base de dados padrão-ouro anotada, também foi explorada outra técnica que busca expandir bases de dados programaticamente. Assim, é apresentado o contexto atual de pesquisa chamado de *Aprendizado de Máquina com dados limitados* no qual são relacionadas diversas técnicas que buscam superar a falta de bases de dados anotadas para serem utilizadas em algoritmos de Aprendizado de Máquina Supervisionado, como também, explorar bases já existentes para expandi-las. Também é apresentado que um fator chave para esse tipo de algoritmo é o tamanho das bases de dados de treinamento.

Desse modo, é realizada uma digressão na técnica de Supervisão Fraca utilizando o Snorkel Framework, a qual é utilizada na fase de experimentação. Assim, Ratner et al. (2019) demonstraram que é possível montar um conjunto de instruções programáticas que compilam o conhecimento da área de domínio de profissionais, por exemplo, médicos ou advogados, e inserem esse conhecimento diretamente dentro do modelo de aprendizado

de máquina em uma fase prévia ao desenvolvimento da base de treinamento. Então esse modelo é aplicado para a montagem de uma nova base de documentos que corresponde ao conhecimento da área de domínio inserido. Importante frisar que a construção da base de documentos programaticamente pode ser realizada automaticamente o que permite a construção em muito menos tempo de bases de dados de tamanhos que levariam muitos meses para serem construídas manualmente.

3 TRABALHOS RELACIONADOS

Historicamente, pesquisas quantitativas ou qualitativas associadas a decisões judiciais ou documentos jurídicos eram realizadas manualmente utilizando uma pequena amostra selecionada conforme exposto no trabalho de Borden e Baron (2014) o qual traçou panorama histórico de pesquisas jurisprudenciais e legais utilizando métodos estatísticos. De maneira similar, o trabalho de Gabardo e Morettini (2013) relata que o Direito experienciou certo distanciamento de outras ciências, especialmente em relação a Ciência da Computação, pois até recentemente muitas pesquisas de análise jurisprudencial eram realizadas de maneira manual, como, por exemplo, o trabalho de Salama et al. (2011) o qual realizou investigação estatística de base de documentos jurídicos sem a utilização de métodos automáticos.

Em contrapartida, a utilização de métodos computacionais analíticos e preditivos para o desenvolvimento de soluções que possam extrair conhecimento oculto de documentos jurídicos esbarra em dificuldade de acesso a bases de dados especialmente desenvolvidas para estes fins. Entretanto, nos últimos anos foi possível observar um movimento crescente de disponibilização de dados jurídicos, não apenas no Brasil, mas também em países como a China (FANG et al., 2020). Mas infelizmente essas bases de documentos não foram especialmente criadas para o processamento computacional e assim carecem de quantidade de dados ou de metadados rotulados. Dessa maneira, diversos estudos recentes buscam superar essas dificuldades por meio da aplicação de técnicas que alavancam os conteúdos existentes ou que apresentam desempenho favorável utilizando bases de dados limitadas.

Nesse sentido, a pesquisa de Fang et al. (2020) explora uma base de processos judiciais trabalhistas composta de 3.795 instâncias classificadas com 111 rótulos e outra base de causas cíveis composta de 2.249 instâncias com 61 classes com distribuição exponencial, ou seja, a cauda dessas distribuições apresenta classes com poucas instâncias. Assim, os autores examinam a performance de técnica de *Few-shot Learning* por meio de estruturas de redes neurais adaptadas com a utilização de *Embedding* pré-treinado na língua chinesa. Nessa pesquisa foram extraídas por meio de expressões regulares as frases dos documentos que continham menções a disputas relacionadas a responsabilidade civil das partes que foram utilizadas como entrada nos algoritmos de Aprendizado de Máquina, entretanto, não são fornecidos mais detalhes desse processo. Essa abordagem apresentou pontuação de 93.13 na métrica F1-Macro nesse estudo. É importante conside-

rar que os autores também realizaram testes com algoritmos de *Support Vector Machines* e *Convolutional Neural Networks* sem a aplicação de *Embeddings* pré-treinado, os quais apresentaram respectivamente 76.24 e 69.44 pontos F1-Macro. Isso reforça a vantagem da utilização da técnica de *Few-shot Learning* em conjunto a *Embedding* pré-treinado.

Dificuldade similar foi encontrada no contexto europeu com a utilização de base de documentos rotulada, mas muito desbalanceada. Entretanto, nesse caso os autores Elnaggar et al. (2018) puderam experimentar a aplicação de técnica de *Transfer Learning* em vista de haver também disponibilidade de documentos de outras modalidades. Dessa maneira, foi possível explorar satisfatoriamente um corpus contendo de 4 a 8 milhões de documentos legais em 7 idiomas diferentes, além de outra base de documentos contendo de 18 a 22 mil documentos jurídicos incluindo um breve resumo. Assim, foi desenvolvido um modelo geral de aprendizagem profunda baseado na técnica proposta por (KAISER et al., 2017) treinado concomitantemente nas tarefas de tradução, sumarização e classificação de documentos. Em relação ao *pipeline* de treinamento, não foram fornecidos maiores detalhes de pré-processamento das instâncias de documentos e como esses documentos foram inseridos no algoritmo de Redes Neurais. Enfim, foi possível superar a performance na tarefa de classificação em aproximadamente 10 pontos F1 em relação ao *baseline* proposto. Além disso, é importante frisar a relevância que a utilização de parâmetros devidamente treinados apresenta na performance dos algoritmos. No caso da pesquisa de Fang et al. (2020), foram utilizadas *Embeddings* pré-treinadas no idioma chinês. Por outro lado, no trabalho de Elnaggar et al. (2018) o treinamento conjunto com o corpus de milhões de textos jurídicos permitiu a criação de *Embeddings* durante a própria fase de treinamento.

Abordando de maneira diferente o problema da pouca disponibilidade de dados desenvolvidos especialmente para tarefas de aprendizado de máquina, Boehm (2018) e Song et al. (2019) empregam técnicas que permitem a criação da base de dados de maneira automática. Assim, Boehm (2018) promoveu o desenvolvimento de base de decisões anotadas de maneira similar à presente pesquisa, entretanto, utilizando técnica de *Active Learning*. Nesse caso, não foi possível alcançar altas pontuações nas métricas escolhidas devido ao grande desbalanceamento das classes, pois se tratava de classificação binária em relação à importância ou não de uma frase em decisões judiciais, sendo que apenas 2,26% das instâncias foram consideradas verdadeiras. Outro fator sugerido pelos autores que influenciou na qualidade final dos resultados foi a complexidade da tarefa, pois a anotação com rótulo verdadeiro exigia a apreciação intelectual de 3 conceitos técnicos

do Direito em cada frase analisada, assim foi ponderado que em trabalhos futuros seria interessante dividir a tarefa de classificação em duas fases diferentes. Por outro lado, durante a fase de preparação da base de dados para anotação, o autor referiu que realizou a extração das frases necessárias a partir de seções específicas dos documentos que faziam menção aos detalhes legais interessantes aos objetivos de mineração. Ou seja, o autor explorou a estrutura comum de redação dos documentos jurídicos de maneira semelhante à realizada na presente pesquisa, o que reforça uma característica predominante nos documentos dessa área de domínio e que pode ser explorada para, como no caso de Boehm (2018), ajudar a diminuir o desbalanceamento de classes na tarefa de classificação de texto. Além disso, é possível observar menções à exploração da forma de redação comum dos documentos jurídicos também em Lei et al. (2017), Strickson e Iglesia (2020).

De maneira similar à presente pesquisa, Lei et al. (2017) realizou a anotação manual de 6.735 decisões em relação a 13 categorias. Desse modo, foi possível realizar a avaliação da performance dos algoritmos *Naive Bayes*, *Decision Tree*, *Random Forest* e *Support Vector Machine* sendo que este último alcançou a melhor performance em torno de 90% F1. Além disso, os autores fazem menção a um fator determinante, que também foi observado de maneira semelhante na presente pesquisa, em relação ao fato de que em todos os tipos de decisões judiciais é possível observar uma formatação padrão dos documentos como também um padrão de escolha de palavras, títulos e seções dos documentos. Assim, apenas as seções que seriam úteis para a tarefa de classificação foram extraídas dos documentos e compiladas na base de dados para treinamento e testes.

Outra pesquisa que explorou essas características dos documentos jurídicos foi a de Strickson e Iglesia (2020), em que a base de dados foi desenvolvida programaticamente, entretanto, diferentemente da presente pesquisa, nenhum *framework* foi usado para esta tarefa. Assim, os autores observaram, por meio de análise exploratória dos documentos, que os juízes seguem certo padrão de escolhas de palavras para indicar o deferimento ou não dos julgamentos. Desse modo, documentos que continham palavras como *allow* foram rotulados com a classe positiva e documentos que continham palavras como *dismiss* foram rotulados com a classe negativa. Como ponto negativo dessa pesquisa, pode ser notado que não foi calculada nenhuma métrica em relação à acurácia dos rótulos aplicados automaticamente, tampouco foi desenvolvida base de dados padrão-ouro para a realização de testes.

Já o trabalho de Song et al. (2019) propõe um método para a criação automática de uma base de dados por meio de técnica de recuperação de informações. Assim, esse

Tabela 3.1: Quadro comparativo dos trabalhos relacionados em relação às principais características analisadas.

TÍTULO	AUTORES	ANO	TAMANHO DO CORPUS	QUANTIDADE DE CLASSES	ALGORITMOS	TIPO DE CLASSIFICAÇÃO
Automatically classify Chinese judgment documents utilizing machine learning algorithms	M. Lei, J. Ge, Z. Li, C. Li, Y. Zhou, X. Zhou	2017	6735	13	KNN, Naive Bayes, Random Forest, SVM	Multiclasse
Few-Shot Learning for Chinese Legal Controversial Issues Classification	Y. Fang, X. Tian, H. Wu, S. Gu, Z. Wang, F. Wang	2020	5521	1295	Few-shot learning	Multiclasse
Supporting the legal reasoning process by classification of judgments applying active machine learning	Glaser I.; Landthaler J.; Matthes F.	2019	800	2	Active Learning	Binária
Employing Auto-Annotated Data for Government Document Classification	Y. Song, Z. Li, J. He, Z. Li, X. Fang, D. Chen	2019	4850	10	CNN, XGBoost	Multiclasse
Multi-Task Deep Learning for Legal Document Translation, Summarization and Multi-Label Classification	A. Elnaggar, C. Gebendorfer, I. Glaser	2018	4 milhões; 18 mil; 11 mil	6000	Transfer Learning MultiModel NN by Google	Multi-label
Legal Judgement Prediction for UK Courts	B. Strickson, B. De La Iglesia	2020	4959	2	SVM, Logistic Regression, Random Forest, k-Nearest Neighbour, Neural Networks	Binária

estudo selecionou manualmente uma série de palavras-chave que melhor representassem as classes e as utilizou para realizar a recuperação de documentos. Entretanto, não foi especificado qual algoritmo de ordenação foi utilizado. Após, foram selecionados os 30% primeiros documentos para cada classe e assim formada a base de dados para treinamento. Foi reportado aumento de 15% na métrica F1 em relação ao *baseline* proposto. Na Tabela 3.1 é possível observar quadro comparativo dos trabalhos relacionados em relação às principais características analisadas.

3.1 Resumo do Capítulo

Este capítulo inicia apresentando o contexto histórico de pesquisas quantitativas utilizando documentos jurídicos, além de verificar que próprios autores da área do Direito reconhecem a necessidade de serem realizados novos estudos utilizando técnicas computacionais, havendo, assim, uma grande lacuna a ser preenchida. Por outro lado, é possível observar que até poucos anos não havia bases de dados digitalizadas contendo os documentos jurídicos, no contexto brasileiro e também no contexto mundial, tornando assim, quase impossível a utilização de técnicas de Aprendizado de Máquina com decisões

judiciais, por exemplo.

Desse modo, foi realizada busca por pesquisas na literatura mundial que houvesse o emprego de técnicas de Aprendizado de Máquina para classificação de texto utilizando documentos jurídicos. Assim, devido às dificuldades inerentes a esta pesquisa em função da inexistência de uma base de dados anotada, foram filtradas as pesquisas que faziam menção a utilização de técnicas de Inteligência Artificial que permitissem a superação da falta de uma base de dados anotada padrão-ouro previamente disponível. Por outro lado, infelizmente, após a aplicação dos filtros citados, não houve nenhum trabalho que fizesse menção à utilização de documentos na língua portuguesa.

A pesquisa de FANG et al. apresenta os resultados da aplicação de técnicas de *Few-shot Learning* em conjunto a Redes Neurais e *Embedding* pré-treinado e relata que foram obtidos resultados muito satisfatórios. Ou seja, nesse caso, os autores utilizaram um tipo de algoritmo que permite o treinamento de um modelo de Aprendizado de Máquina Supervisionado mesmo contendo poucas instâncias. Por outro lado, a pesquisa de ELNAGGAR et al. tinha a tarefa de realizar classificações de documentos, mas a dificuldade eram as classes muito desbalanceadas. Nesse caso, foi possível obter melhora na performance por meio da utilização de técnica de *Transfer Learning*, pois também estavam disponíveis uma grande quantidade de documentos relacionados ao assunto, ou seja, contendo *features* parecidas. A pesquisa de Boehm (2018) por sua vez focou na construção otimizada de uma base de dados padrão-ouro utilizando técnica de *Active Learning*, entretanto foi relatado que não foram alcançados bons resultados de classificação.

Além das técnicas de Aprendizado de Máquina que buscam superar as dificuldades de falta de uma base de dados padrão-ouro previamente desenvolvida, foi possível verificar nos trabalhos relacionados diversas técnicas que exploravam a forma padronizada de redação dos documentos jurídicos. Nas pesquisas de Boehm (2018), Lei et al. (2017) foi possível verificar que houve um pré-processamento dos documentos de modo a extrair apenas certa parte dos documentos onde continham as informações necessárias aos modelos. Já a pesquisa de Strickson e Iglesia (2020) realizou a classificação direta dos documentos utilizando palavras-chave, entretanto sem a utilização de nenhum *framework* que permitisse aferir níveis de acurácia da técnica. Além disso, a presente pesquisa também emprega técnicas de pré-processamento de texto de modo a extrair partes e seções dos documentos, como também utiliza palavras-chave presentes no texto, entretanto, neste caso, para serem utilizadas em conjunto ao *Snorkel Framework*, e assim são calculadas métricas de acurácia entre outras. Dessa maneira, importante frisar a semelhança de al-

gumas técnicas utilizadas para extração de partes do texto pelos autores em ambos os trabalhos, mesmo havendo diferença de idiomas sendo português ou chinês, como também a dificuldade em encontrar bases de dados previamente prontas para trabalhos de classificação de texto.

Por fim, é apresentada a Tabela 3.1 a qual sumariza os trabalhos relacionados em termos das suas principais características. Assim, é possível observar que, das técnicas de Aprendizado de Máquina com dados limitados apresentadas na Seção 2.6, foram encontrados apenas trabalhos experimentais relacionados a documentos jurídicos utilizando *Active Learning*, *Transfer Learning* e *Few-shot Learning*, não tendo sido encontrado nenhum trabalho que utilize Supervisão Fraca. Portanto, esta pesquisa pretendeu, além de atender ao objetivo de negócio e à questão de pesquisa, avaliar a utilização de técnica de Supervisão Fraca aplicada a documentos jurídicos na língua portuguesa.

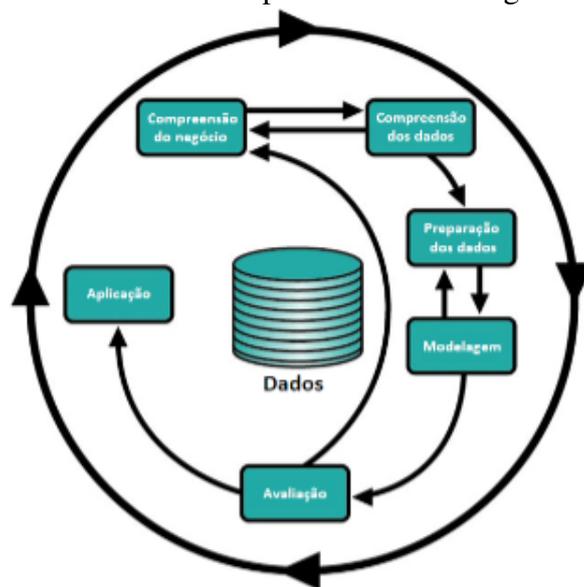
4 METODOLOGIA

A presente pesquisa adotou a metodologia CRISP-DM Guide 1.0 (CHAPMAN et al., 2000) para validação experimental, assim, considerou-se interessante apresentar a seguir os pontos-chave dessa metodologia. Além disso, devido à utilização de testes estatísticos em alguns experimentos, também se considerou interessante apresentar os pontos-chave da metodologia adotada apresentada no trabalho de Snijders (2002).

4.1 Processo de descoberta de conhecimento em base de dados

Este trabalho segue a abordagem metodológica descrita em CRISP-DM Guide 1.0 (CHAPMAN et al., 2000) para a descoberta de conhecimento em base de dados que é utilizada para extrair conhecimento útil de grandes coleções de dados. Essa metodologia apresenta um processo iterativo composto de diversas fases, que compõe desde a compreensão e as necessidades de negócio até a modelagem dos dados e sua aplicação, as quais são descritas em mais detalhes nas seções seguintes. A Figura 4.1 apresenta o ciclo do processo.

Figura 4.1: Ciclo de desenvolvimento do processo metodológico CRISP-DM Guide 1.0.



4.1.1 Compreensão do Negócio

Essa fase consiste em compreender o valor do conhecimento a ser gerado pela perspectiva do negócio, de modo a alinhar o projeto com os objetivos estratégicos da organização. Assim, esse estágio compreende a análise do contexto em que o negócio se encontra inserido de modo a compreender o que o cliente realmente necessita e seu objetivo principal. Essa análise vai guiar todas as fases do processo de descoberta de conhecimento, pois são definidos pontos fundamentais, como, por exemplo, o objetivo de mineração. Esse estágio é essencial pois uma possível definição incorreta dos problemas de negócio levaria a pesquisa invariavelmente a trazer resultados inúteis. A seguir são apresentados os principais elementos que compõem essa fase:

- **Objetivos de negócio:** compreensão do contexto em relação ao mercado em que o negócio se encontra; definição dos objetivos do negócio em relação ao projeto de descoberta de conhecimento em base de dados; definição de critérios de sucesso do projeto.
- **Recursos do projeto:** inclui inventário de recursos disponíveis para a realização do projeto; listagem de requisitos e restrições que podem haver, como, por exemplo, data limite para realização, níveis de qualidade, segurança e disposições legais em relação aos dados; além de análise da relação custo-benefício esperado obter com o projeto.
- **Objetivo de mineração:** definição clara do objetivo de mineração é um passo fundamental no processo que permite a execução satisfatória da pesquisa, além dos critérios de sucesso os quais vão permitir a verificação da eficiência dos modelos desenvolvidos.
- **Planejamento do Projeto:** inclui a listagem dos passos necessários para alcançar os objetivos, como também lista de ferramentas necessárias para a execução e prazos de execução.

4.1.2 Compreensão dos dados

Inicia com a coleta dos dados e com a exploração inicial, o que permite a identificação de problemas de qualidade e a aferição de conhecimentos estatísticos sobre a massa de dados. Essa fase pode identificar se realmente os dados podem responder às

perguntas do negócio e identificar as variáveis significativas. A seguir são apresentados os principais elementos que compõem essa fase:

- Coleta de dados: inclui a extração de dados iniciais ao projeto e a sua descrição de modo que seja especificada sua localização, técnica utilizada para extração e resolução de problemas encontrados.
- Descrição dos dados: inclui análise dos dados de maneira a especificar os tipos de dados disponíveis, seu formato e quantidade. Também permite inferir se os dados disponíveis satisfazem aos requisitos especificados.
- Exploração dos dados: análise da maneira como os dados estão distribuídos no banco de dados e de seus relacionamentos por meio da apresentação de gráficos e relatórios.
- Qualidade: verificação da qualidade dos dados extraídos por meio da verificação de possíveis erros na extração, além verificação se há valores que estejam faltando em determinados campos.

4.1.3 Preparação dos dados

O objetivo é o pré-processamento dos dados para torná-los relevantes e consistentes com respeito à tarefa de busca de conhecimento. Essa fase é extremamente necessária, pois os dados muitas vezes podem estar incompletos, inconsistentes ou podem, até mesmo, conter erros. A seguir são apresentados os principais elementos que compõem essa fase:

- Seleção e integração dos dados: se os dados estiverem distribuídos em diversas bases, será necessário realizar procedimentos para uni-los de modo a permitir posteriormente a seleção das melhores observações coletadas para análise e processamento.
- Limpeza dos dados: realizar o tratamento dos dados de modo a remover dados ou caracteres que podem reduzir os níveis de acurácia de certos modelos utilizados.
- Construção e formatação dos dados: certas técnicas de modelagem exigem que os dados estejam em determinado formato para a sua correta utilização.

4.1.4 Modelagem

Consiste na tarefa de escolha de métodos e parametrização para a extração de padrões, classificação, segmentação, regressão ou associação de itens, os quais gerarão novos conhecimentos sobre a importância de cada uma das variáveis em função do resultado esperado. A seguir são apresentados os principais elementos que compõem essa fase:

- Escolha da técnica de modelagem: Análise e escolha da melhor técnica de modelagem que se aplique ao caso.
- Testes: desenvolver uma técnica que permita a realização da avaliação do modelo após a sua construção. Muitas vezes, é construída uma base de dados anotada no estágio anterior de modo que seja possível separar uma parte para testes.
- Construção do modelo: referente ao desenvolvimento do modelo por meio de técnicas de treinamento por aprendizado de máquina, por exemplo, incluindo a escolha de parâmetros.
- Avaliação técnica: consiste em analisar o modelo construído em função de diversos parâmetros em busca da melhor combinação possível.

4.1.5 Avaliação

Fase em que os padrões reconhecidos, regras de associação e todo conhecimento gerado é analisado para verificação da sua real utilidade. Podem ser utilizadas medidas estatísticas, como também visualizações, para ajudar a perceber a utilidade dos dados. A seguir são apresentados os principais elementos que compõem essa fase:

- Análise de resultados: diferentemente da avaliação técnica do modelo, neste caso o importante é avaliar se o modelo atende aos requisitos de negócio, desse modo, é possível testar o modelo em um protótipo de aplicação real com usuários finais, por exemplo.
- Revisão e próximos passos: análise de todas as atividades realizadas e sua eficácia, além de descrever futuras ações.

4.1.6 Aplicação

Consiste na consolidação de todo processo na forma de relatório e publicação do conhecimento ou na incorporação da modelagem a um sistema computacional. A seguir são apresentados os principais elementos que compõem essa fase:

- Plano de implementação: certos casos exigem que os dados sejam transformados antes de serem processados por um modelo, assim, todo o processo precisa ser documentado.
- Plano de monitoramento: pode ser necessário a verificação da qualidade dos dados recebidos em um fluxo de processamento para alimentação de um modelo de dados.
- Relatório final e revisão: sumário de todo o projeto incluindo uma apresentação final para os clientes.

4.2 Teste de hipótese estatística

A presente pesquisa busca realizar análise quantitativa em relação à proporção de acórdãos julgados favoravelmente para as partes empregado e empresa de modo a responder à hipótese levantada na Seção 6.1.1. Assim, de modo a aferir se realmente existe diferença estatística entre as proporções obtidas, julga-se necessária a aplicação de testes estatísticos, como também o cálculo do nível de força e efeito.

O teste de hipótese estatística desenvolvido neste trabalho segue a proposta de metodologia descrita por SNIJDERS no trabalho Snijders (2002). De modo geral, o pesquisador define uma questão de pesquisa cuja resposta tem por base a análise de propriedades de um conjunto de observações obtidas a partir de um processo estocástico representado por meio de uma distribuição estatística, ou seja, dados quantitativos. Assim, dados são coletadas e resumidos de acordo com as propriedades sendo analisadas, que podem ser, por exemplo, valores médios para distribuições contínuas ou proporções para distribuições binomiais. Após, a questão de pesquisa é reescrita de forma a considerar a existência ou não de um efeito que pode ser observado considerando a propriedade medida por meio da comparação desse valor com outro valor específico ou com a mesma propriedade observada de outro grupo estatístico.

Desse modo, é definida a hipótese nula a qual verifica a inexistência de efeito, ou seja, que não há diferença nas observações em relação às propriedades em análise. Já em

relação à comparação entre dois grupos, a hipótese nula pode implicar a verificação da igualdade da propriedade entre dois grupos. Além disso, também é definida a hipótese alternativa a qual define a existência de algum efeito ou define a existência de diferença entre grupos em relação a propriedade sendo analisada. Ao mesmo tempo, é necessário definir o nível de significância que deve ser usado para comparar com o resultado do teste estatístico aplicado, normalmente definido como 5%.

Assim, o teste estatístico a ser escolhido computa os dados e apresenta a *estatística* do teste e o *P-value*. A *estatística* do teste apresenta o quanto os dados diferem da hipótese nula de acordo com a distribuição normal. Já o *P-value* apresenta a probabilidade de obter os dados analisados se a hipótese nula é na verdade verdadeira na população. A fins de interpretação de resultados, optou-se por usar o *P-value* nesta pesquisa em vista da sua objetividade, pois a interpretação depende da comparação do *P-value* com o nível de significância definido previamente. Se o *P-value* for menor que o nível de significância, rejeita-se a hipótese nula. Se o *P-value* for maior, há falha em rejeitar a hipótese nula.

Ademais, o teste estatístico a ser aplicado varia de acordo com o tipo de distribuição de dados e do tipo de propriedade da distribuição sendo analisada. Testes conhecidos como *paramétricos* basicamente assumem que a distribuição dos dados segue padrões de normalidade. Já os testes conhecidos como *não-paramétricos* não assumem que os dados seguem padrões de normalidade. Em relação aos testes *paramétricos*, existem testes de correlação que verificam o relacionamento entre variáveis sem a noção de causa e efeito, testes de regressão que verificam o relacionamento entre variáveis incluindo a noção de causa e efeito e testes de comparação que verificam a diferença entre médias e proporções.

Após a escolha do teste estatístico, é necessário verificar se as condições de validade do teste estão satisfeitas em relação aos dados da amostra. Assim, por exemplo, o teste estatístico que realiza a comparação entre proporções de duas distribuições binomiais necessita que as seguintes suposições estejam satisfeitas:

- Amostra deve conter observações extraídas aleatoriamente com probabilidades iguais.
- As observações devem ser independentes.
- $n1 * p1 \geq 5$ e $n1 * (1 - p1) \geq 5$, sendo $n1$ = quantidade de observações no grupo 1, $p1$ = proporção no grupo 1.

Por fim, o teste estatístico é aplicado e o *P-value* é comparado com o nível de significância definido previamente. Além disso, é necessário apresentar a conclusão do

teste informando a rejeição da hipótese nula ou falha da sua rejeição, como também, a interpretação dos resultados em relação aos objetivos de negócio estabelecidos.

4.3 Resumo do Capítulo

Neste capítulo, iniciou-se apresentando a contextualização metodológica da pesquisa. Após, foi apresentada a metodologia base utilizada para desenvolvimento de todo o processo de validação experimental. Em função do carácter eminentemente prático do estudo, optou-se pela metodologia CRISP-DM Guide 1.0 (CHAPMAN et al., 2000), a qual apresenta-se amplamente difundida na área profissional. Por outro lado, em vista de que a validação experimental inclui análise quantitativa de dados e análise de proporções de distribuições, optou-se por utilizar também metodologia de pesquisa estatística Snijders (2002) para esta fase dos experimentos.

5 MATERIAIS E MÉTODOS

Neste capítulo é apresentada visão geral da pesquisa por meio da classificação taxonômica e também por meio da descrição do método de validação experimental.

5.1 Classificação da pesquisa

A seguir é realizada a classificação da presente pesquisa de acordo com a taxonomia descrita em Wazlawick (2017). Por se tratar de um tema recente e seus resultados dependerem da aplicação prática de técnicas de mineração de texto no contexto de documentos jurídicos, este trabalho pode ser classificado como uma pesquisa de natureza aplicada, pois busca desenvolver conhecimento que pode ser utilizado para aprimorar processo de tomada de decisões em empresas e em escritórios de advocacia.

Os objetivos do trabalho o caracterizam como uma pesquisa descritiva, uma vez que os documentos jurídicos são classificados para realização de análise quantitativa.

Em relação aos procedimentos técnicos, foi adotada a pesquisa experimental para alcançar o propósito do trabalho, coletando dados de fontes públicas de sites governamentais e desenvolvimento de modelos de aprendizado de máquina. Quanto à abordagem, foi realizada pesquisa quantitativa em relação aos dados classificados por meio dos modelos de Aprendizado de Máquina.

A avaliação dos resultados da pesquisa quantitativa foi feita por meio de testes estatísticos de significância que permitem a avaliação dos resultados.

Na Tabela 5.1 é possível observar resumo da classificação da pesquisa desenvolvida com base na taxonomia descrita em Wazlawick (2017).

Tabela 5.1: Resumo da classificação metodológica da pesquisa.

Tipo	Classificação	Justificativa
Natureza	Aplicada	Aprimoramento do processo de tomada de decisões.
Objetivos	Descritiva	Descrição de tendência estatística de julgamento.
Procedimentos	Experimental	Coleta de dados, classificação, análise e resposta à hipótese estatística.
Abordagem	Quantitativa	Classificação dos dados por meio de modelo de Aprendizado de Máquina.

5.2 Método de Validação Experimental

A execução dos experimentos foi realizada utilizando a metodologia descrita em CRISP-DM Guide 1.0 (CHAPMAN et al., 2000) a qual apresenta um método iterativo composto de diversas fases cujo objetivo é alinhar o processo de Mineração de Dados às expectativas do negócio. Por outro lado, essa mesma metodologia prevê o retorno a fases iniciais de modo a alinhar os requisitos e objetivos de pesquisa de acordo com conhecimento novo obtido nas fases de execução de experimentos. Isso se deve principalmente à natureza da pesquisa de Mineração de Dados que busca extrair conhecimento oculto de uma base de documentos. Desse modo, a seguir será apresentado o resumo do método sequencial de execução da validação experimental após a execução iterativa da metodologia descrita no Capítulo 4.

Assim, primeiramente foi definida a questão de pesquisa como **Seria possível que os tribunais avaliados e suas turmas recursais julguem favoravelmente proporção significativamente maior de recursos para uma das partes do que para outra em média?** A seguir, foi definido que seria utilizado teste estatístico para verificação da questão de pesquisa. Então, foi definido como hipótese estatística o seguinte: **Os tribunais avaliados e suas turmas recursais julgam favoravelmente proporção significativamente maior de recursos para uma das partes do que para outra em média.** Assim, definiu-se como P_a a proporção de recursos de empregados deferidos e P_b a proporção de recursos de empresas deferidos. Então, foi definida como hipótese nula $H_0 P_a = P_b$. Já a hipótese alternativa como $P_a \neq P_b$. A seguir, é apresentado resumo do teste de hipótese estatística:

- P_a = proporção de recursos de empregados deferidos.
- P_b = proporção de recursos de empresas deferidos.
- H_0 : $P_a = P_b$.
- H_a : $P_a \neq P_b$.

De modo a realizar a análise quantitativa das proporções de julgados favoráveis às partes empregados e empresas era necessário obter instâncias de observações contendo essas informações. Assim, procedeu-se ao desenvolvimento da base de dados pela utilização de técnicas de Mineração de Dados. Portanto, foi definido como objetivo de Mineração de Dados **classificar automaticamente acórdãos judiciais em relação ao deferimento ou não e classificar automaticamente o requerente do recurso em relação a ser empresa ou a ser empregado. Além disso, classificar com os modelos desenvolvidos quanti-**

dade significativa de decisões judiciais e apresentar relatórios contendo as proporções de julgados positivamente para cada uma das partes.

A seguir, passou a execução do processo de Mineração de Dados o qual compreende as seguintes fases: Coleta de Dados, que inclui a extração dos documentos da internet; Preparação dos Dados, que inclui a limpeza dos dados, a anotação manual e anotação automática das bases de dados extraídas; Modelagem que inclui o desenvolvimento de modelos de Aprendizado de Máquina para a classificação automática das instâncias preparadas na fase anterior; Aplicação que inclui a efetiva classificação das instâncias, análise quantitativa, aplicação de testes estatísticos e desenvolvimento de gráficos. A Figura 5.1 apresenta ilustração explicativa do *pipeline* da Validação Experimental da pesquisa. Nas seções seguintes são apresentados pontos-chave das fases do processo de Mineração de Dados. Além disso, foi disponibilizado também por meio do repositório¹ o código-fonte de todo processamento da pesquisa, que pode ser usado para consulta dos detalhes de implementação e reprodução dos experimentos.

5.2.1 Coleta dos Dados

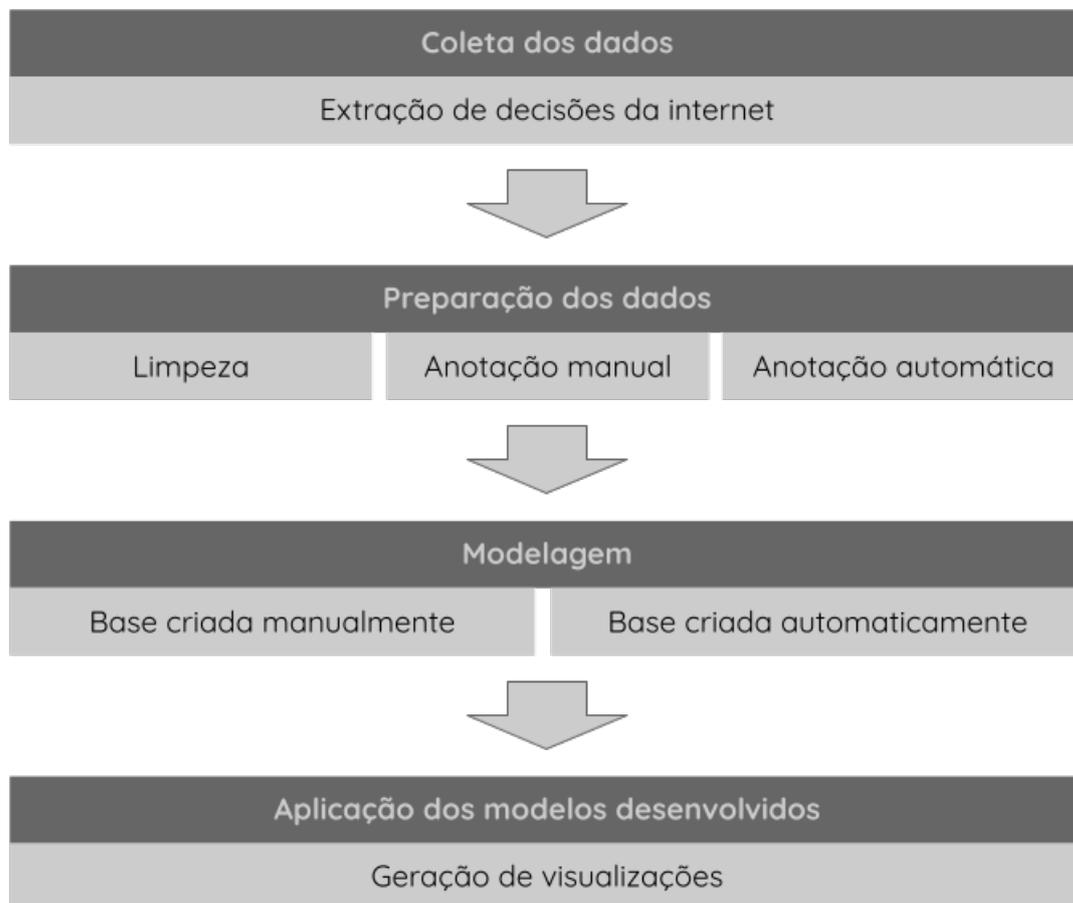
Os documentos judiciais que contém a informação sobre o julgamento de determinado processo judicial são chamados de **acórdãos**, o quais são disponibilizados via internet. Esses documentos contém informação quanto ao vencedor da causa. Assim, esses documentos são as instâncias que precisam ser coletadas e processadas na fase seguinte da Mineração de Dados. Assim, foi definido que seriam analisados quantitativamente acórdãos de dois tribunais diferentes, a saber Tribunal Regional do Trabalho da 3ª Região e Tribunal Regional do Trabalho da 4ª Região.

Os documentos do TRT da 3ª Região foram extraídos a partir do site <<http://lexml.gov.br/>> por meio dos arquivos sitemap.xml na raiz do site. Quanto ao site <<http://trt4.jus.br/>>, foi utilizada a página de busca disponibilizada que permite a filtragem de documentos por meio de palavras-chave. Assim, foram desenvolvidos robôs de extração de dados utilizando a linguagem Python e a biblioteca Scrapy². Na época da execução dos experimentos estavam disponibilizados nas páginas os montantes de acórdãos descritos na Tabela 6.2, essa tabela também apresenta a quantidade extraída efetivamente pelos robôs de extração. Diversas requisições HTTP foram negadas pelos servidores o

¹<https://github.com/rhuanbarros/jurimetria_justica_do_trabalho>

²<<https://scrapy.org/>>

Figura 5.1: Ilustração explicativa do *pipeline* da Validação Experimental da pesquisa incluindo visão geral do processo.



que resultou na impossibilidade de extração de alguns documentos.

5.2.2 Preparação dos Dados

A Preparação dos Dados consiste em realizar ajustes para a utilização dos dados por algoritmos de Aprendizado de Máquina. Assim, primeiramente é realizada uma limpeza para a remoção de instâncias que apresentem mensagens de erro ou que contenham caracteres truncados. Portanto foi aplicado um filtro que removeu todas as instâncias com menos de 150 palavras. Importante salientar que a *tokenização* foi realizada utilizando algoritmo que processa características específicas da língua portuguesa disponibilizado no repositório da pesquisa³.

Após a limpeza, passou-se a transformação dos dados. Primeiramente foram extraídos dados dos documentos por meio de expressões regulares o quais foram inseridos

³<https://github.com/rhuanbarros/jurimetria_justica_do_trabalho>

na base de dados como metadados. O código-fonte Python e Regex pode ser encontrado no repositório da pesquisa. Também foi realizada a extração de uma seção dos documentos de texto chamada **Dispositivo**. Essa seção contém informação suficiente para que os algoritmos de Aprendizado de Máquina extraíam *features* e classifiquem corretamente as instâncias. Além disso, todos os documentos mais de um recorrente também foram removidos. Nesse ponto dos experimentos, a base de dados contém os metadados extraídos juntamente com o dispositivo dos acórdãos e apenas decisões onde houve apenas um recorrente, resultando o **total de 22.946 instâncias**.

A seguir, prosseguiu-se a criação das bases de dados anotadas para utilização com os algoritmos de Aprendizado de Máquina. Como o objetivo de Mineração de Dados exige a realização de dois tipos de classificações, são necessárias duas bases de treinamento, uma que viabilize a classificação em relação ao tipo de recorrente e outra em relação ao deferimento ou não da causa judicial.

A criação da base de dados para classificação em relação ao tipo de recorrente foi realizada por meio da extração aleatória de 270 instâncias de cada um dos tribunais em formato CSV, totalizando 540 observações. Essas anotações foram realizadas exclusivamente pelo próprio autor apenas, pois tal tarefa não exige nenhum conhecimento específico relacionado ao Direito. Os rótulos aplicados foram os seguintes: RECLAMANTE ou RECLAMADA.

Após, passou-se a criação de uma base de dados padrão-ouro utilizada para classificação em relação ao deferimento ou não da causa. Essa base foi utilizada para realização de experimentos de modelagem de Algoritmo de Aprendizado de Máquina como também foi utilizada para testes de eficácia dos modelos desenvolvidos. Essa fase foi desenvolvida seguindo critérios rigorosos de qualidade de acordo com a metodologia descrita em (PUSTEJOVSKY; STUBBS, 2012). Foram selecionadas aleatoriamente 1.000 instâncias da base de dados preparada anteriormente que foram anotadas pelo próprio autor da pesquisa, o qual detém conhecimentos jurídicos suficientes para esta tarefa. Além disso, dessas 1.000, 500 foram anotadas também por uma pessoa bacharel em Direito e as outras 500 também foram anotadas por uma pessoa bacharel em Direito. Resultando assim, em 1.000 instâncias anotadas por duas pessoas. Importante considerar que foram utilizadas para anotação apenas a parte do **Dispositivo** dos acórdãos. Os rótulos aplicados foram os seguintes: DEFERIMENTO, INDEFERIMENTO ou SEM_ANALISE_MERITO.

Também foi realizada a anotação automática da base de documentos por meio de técnica de Supervisão Fraca utilizando o Framework Snorkel. Foram desenvolvidas 13

Funções de Rotulação que apenas verificam a existência de uma palavra na instância e aplicam um rótulo específico, sendo possível verificar detalhes na Tabela 6.10. Então, o algoritmo do Snorkel foi aplicado a totalidade de instâncias disponíveis de 22.946, resultando, assim, numa base de treinamento criada programaticamente totalizando 22.471 instâncias. Por outro lado, observou-se que a base de dados criada programaticamente apresentava grande desbalanceamento. Então procedeu-se ao tratamento dessa questão por meio de *Under-sampling* criando-se assim outra base de treinamento, mas nesse caso contendo o total de 1.644 instâncias, sendo 548 instâncias para cada classe.

Enfim, na fase de Preparação dos Dados foram criadas 4 bases de treinamento. Uma base para o treinamento em relação ao tipo de recorrente, contendo 540 instâncias. Três bases para o treinamento em relação ao deferimento ou não da causa. Sendo uma delas a base padrão-ouro anotada por especialistas, contendo 1.000 instâncias e duas delas criadas programaticamente contendo 22.471 instâncias e 1.644 instâncias, mas nesse caso balanceada.

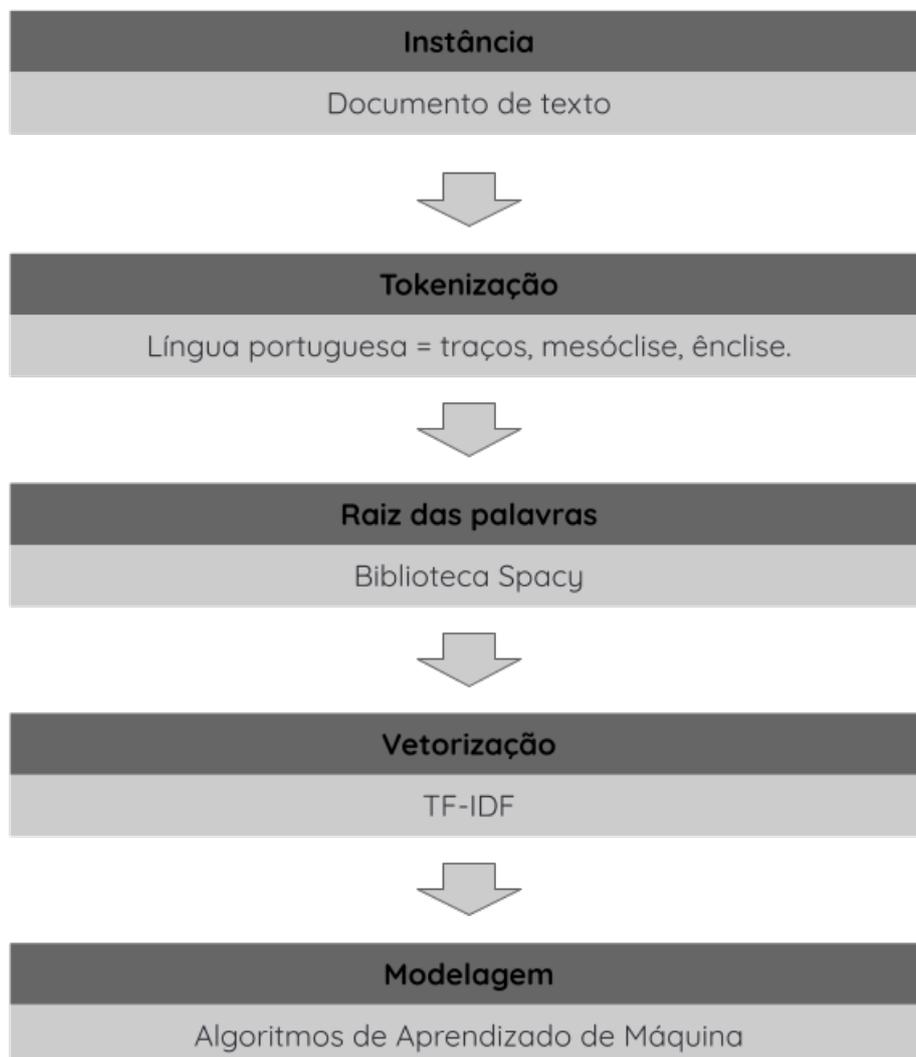
5.2.3 Modelagem

A seguir prosseguiu-se a fase de modelagem utilizando as bases de dados criadas nas fases anteriores. Todos os experimentos nessa fase foram desenvolvidos de acordo com o seguinte fluxo de pré-processamento: as instâncias são *tokenizadas* com algoritmo específico que trata detalhes da língua portuguesa, em seguida é extraída a raiz de cada palavra usando a biblioteca Spacy, após é realizada a vetorização por meio de *TF-IDF* e por fim é realizada a modelagem com algoritmos de Aprendizado de Máquina, conforme ilustrado na Figura 5.2. Além disso todos os experimentos foram modelados utilizando os seguintes algoritmos da biblioteca Scikit Learn: Rocchio classifier, Gradient Boosting Classifier, Naive Bayes Classifier, K-nearest Neighbor, Support Vector Machine (SVM), Decision Tree, Random Forest. Também foi definido como padrão a utilização da métrica F1 como parâmetro de avaliação.

Primeiramente foi desenvolvido um experimento para classificação do tipo de requerente como empregado ou empresa sendo utilizado 70% da base para treinamento e 30% para testes. Enfim, o algoritmo *Support Vector Machine* (SVM) atingiu a métrica mais alta de 96,91%.

O desenvolvimento do modelo necessário para classificar as decisões em relação ao deferimento ou não da causa judicial foi realizado por meio de três experimentos. Um

Figura 5.2: Ilustração explicativa do fluxo de pré-processamento para modelagem por algoritmos de Aprendizado de Máquina.



utilizando a base de dados padrão-ouro, outro utilizando a base criada programaticamente balanceada e outro utilizando a base de dados criada programaticamente com todas as instâncias.

O experimento utilizando a base de dados padrão-ouro foi realizado utilizando 30% da base para testes e 70% para treinamento. Sendo que também foi realizada a validação-cruzada em 7 camadas com as instâncias de treinamento. Já quanto aos dois experimentos utilizando as bases de dados criadas programaticamente, sendo uma balanceada e a outra contendo a totalidade de instâncias, não houve a separação em 70/30. Nesse caso, foi utilizada a totalidade de cada base criada programaticamente para treinamento e usada a totalidade da base padrão-ouro para testes. Assim, foi selecionado o algoritmo *Gradient Boosting* pois foi o que atingiu a maior média na métrica F1 sendo de

92,48%.

5.2.4 Aplicação

Nessa fase foi realizada a aplicação dos dois modelos de Aprendizado de Máquina selecionados a totalidade de 22.946 instâncias extraídas e processadas nas fases anteriores. Nesse ponto, foi realizada a filtragem e remoção de todos os documentos que continham o rótulo SEM_ANALISE_MERITO em virtude de que os objetivos de negócio exigem apenas a verificação das instâncias rotuladas com DEFERIMENTO ou INDEFERIMENTO. A seguir foi realizado cálculo de proporções de julgamentos deferidos e indeferidos e aplicados testes estatísticos de proporção utilizando a biblioteca Python Statsmodel. Por fim, foram desenvolvidos gráficos que apresentam as proporções de julgamentos de maneira visualmente interessante.

5.3 Resumo do Capítulo

Este capítulo apresentou resumo do método de validação experimental executado na presente pesquisa e detalhado no Capítulo 6. Assim, primeiramente é realizada a classificação da pesquisa de modo a nortear o leitor sobre o tipo de estudo realizado. Na sequência é salientada o tipo de metodologia utilizada para execução do trabalho e como ela foi executada neste trabalho. Após é descrito método de execução da Validação Experimental de modo a permitir a fácil reprodução da pesquisa. Também é disponibilizado *link* para acesso ao código-fonte do projeto o que permite verificar detalhes de implementação.

6 VALIDAÇÃO EXPERIMENTAL

A validação experimental da presente pesquisa segue metodologia CRISP-DM Guide 1.0 (CHAPMAN et al., 2000) conforme descrito no Capítulo 4. Desse modo, considerou-se conveniente apresentar os resultados em seções com mesmos nomes das fases da metodologia proposta. Assim, o Capítulo Validação Experimental está subdividido nas seguintes Seções: Compreensão do negócio, Preparação dos dados, Modelagem, Avaliação de resultados e Aplicação. Também foi incluída ao final do capítulo a Seção Limitações que trata detalhadamente das limitações dos experimentos.

6.1 Compreensão do negócio

Tradicionalmente, pesquisas jurisprudenciais são realizadas para a verificação de como as decisões judiciais são feitas, dos motivos que levam um magistrado a decidir de determinada forma em cada caso, além dos argumentos em questão utilizados como também as consequências dessas decisões em relação a cada assunto. Com tais pesquisas busca-se diminuir a insegurança e conhecer os caminhos que os processos judiciais podem tomar de acordo com o entendimento dominante de um tribunal, turma recursal ou magistrado.

Entretanto, devido ao distanciamento do Direito de outras ciências (GABARDO; MORETTINI, 2013), como a Ciência da Computação, tais pesquisas eram realizadas sem a utilização de recursos automáticos, aliás, eram realizadas manualmente (SALAMA et al., 2011) com uma pequena amostra selecionada. Com a evolução das tecnologias de *Big Data* e mineração de texto, podemos processar e analisar a grande maioria dos documentos judiciais em busca de padrões desconhecidos, confirmando ou negando hipóteses supostas devido ao conhecimento geral adquirido na experiência de trabalho, e indo além, desenvolvendo sistemas preditivos baseados em uma grande gama de dados.

Além disso, o conhecimento da tendência de julgamento de uma turma recursal da Justiça do Trabalho em relação à determinada matéria pode ser um fator de vantagem no momento de realizar um acordo entre as partes, por exemplo. Especificamente nesse ramo do Judiciário brasileiro, o acordo entre as partes é incentivado em todos os níveis de jurisdição (TRT4, 2020) de forma que as partes entrem em consenso sozinhas em relação a um valor justo que leve o processo ao fim. Além disso, é importante notar que um processo judicial que tenha um acordo homologado por magistrado não tem mais

direito a recursos, dando, assim, fim definitivo ao caso, o que acarreta grande economia processual a Justiça. Então, quando um processo é dirigido a nível de segundo grau a uma turma recursal, o advogado que estiver mais bem informado sobre as tendências de opinião daquele órgão julgador, será mais capaz de tomar uma decisão de aceitar ou oferecer um acordo judicial que pode chegar a milhões de reais (TRT4, 2017).

6.1.1 Objetivo de negócio

De acordo com Salama, Carlotti e Yeung (2018), o conhecimento popular e a experiência de anos de operadores do Direito parecem encaminhar para a consolidação de um senso comum de que existem certos magistrados ou turmas recursais na Justiça do Trabalho inclinadas a proteger mais empregados do que empresas e vice versa. Isso levanta questionamentos éticos quanto à parcialidade dos magistrados, conforme abordado no Capítulo 2.1.3. Entretanto, essa situação é perfeitamente possível considerando que a legislação brasileira deixa muitos aspectos em aberto permitindo a discricionariedade dos magistrados em relação a que lado tomar para determinados assuntos a serem decididos. É importante considerar também que a dinâmica do Direito permite aos magistrados tomarem posições diferentes em relação a mesma matéria e a argumentar de acordo. Entretanto, o ponto interessante para os advogados e empresas na prática jurídica é conhecer quem são os magistrados ou turmas recursais mais inclinados para empresas ou empregados. Assim, deseja-se confirmar por meio de técnicas computacionais e estatísticas se a seguinte questão de pesquisa estabelecida se configura verdadeira: **Seria possível que os tribunais avaliados e suas turmas recursais julguem favoravelmente proporção significativamente maior de recursos para uma das partes do que para outra em média?**

6.1.1.1 Teste estatístico

A hipótese foi definida como **Os tribunais avaliados e suas turmas recursais julgam favoravelmente proporção significativamente maior de recursos para uma das partes do que para outra em média**. Assim a hipótese a ser validada exige a comparação de duas proporções em relação à existência de diferença estatística. Assim, definiu-se como *Pa* a proporção de recursos de empregados deferidos e *Pb* a proporção de recursos de empresas deferidos. Então, foi definida como hipótese nula $H_0 Pa = Pb$. Já a hipótese

alternativa como $P_a \neq P_b$. A seguir, é apresentado resumo do teste de hipótese estatística:

- P_a = proporção de recursos de empregados deferidos.
- P_b = proporção de recursos de empresas deferidos.
- H_0 : $P_a = P_b$.
- H_a : $P_a \neq P_b$.

6.1.1.2 Critério de sucesso do objetivo de negócio

Foi definido como critério de sucesso do projeto o processamento e análise de decisões judiciais por meio de testes estatísticos e o desenvolvimento de relatório contendo a proporção de decisões para cada uma das partes em cada dos tribunais e turmas recursais estudados, incluindo também dados dos testes estatísticos, como o nível de efeito e nível de força estatística.

6.1.2 Requisitos e restrições

O processamento e análise dos dados precisa ser realizado utilizando amostra extraída significativa para representar o total de decisões disponibilizadas publicamente. Além disso, é necessário manter o sigilo quanto a qualquer informação de nomes das partes que sejam citados nos processos judiciais. Por outro lado, a extração automatizada de dados de páginas de internet pode ser proibida, ou até mesmo, o processo pode causar certos danos aos servidores. Assim, foi definido também que seria respeitada qualquer limitação imposta pelos sites a robôs de busca e que a extração dos dados seria realizada da maneira mais preservada possível. Considerou-se necessário também a análise quantitativa de ao menos dois tribunais diferentes para permitir a verificação da performance dos modelos de Aprendizado de Máquina em documentos de origens diversas, como também permitir a comparação das proporções de julgados entre tribunais diferentes.

6.1.3 Custo-benefício

Grandes escritórios de advocacia lidam diariamente com milhares de processos trabalhistas de uma única empresa. De acordo com Salama, Carlotti e Yeung (2019), a condenação média no Tribunal Regional do Trabalho da 2ª Região está em torno de

R\$28.493,54 e o tempo médio de execução de dívida trabalhista de 4 anos e 10 meses. Considerando o caso de um processo judicial hipotético nessa média de valor de condenação, valeria a pena para a empresa efetivar acordo no valor até R\$21.340,00 e para empregado, acordo acima desse valor. Isso porque uma empresa poderia aplicar o dinheiro durante os 4 anos e 10 meses a uma taxa média de 1% ao mês e no final a aplicação poderia ter rendido R\$38,006,00. Quanto ao empregado, ele poderia aplicar o mesmo valor a uma taxa de 0,5% ao mês e ainda sair lucrando. Entretanto, o advogado que tiver posse de relatório contendo as tendências médias estatísticas de julgamento em relação a empregado e a empregador da turma recursal que recebeu o processo para julgamento pode tomar decisão diversa baseado nesses dados. Além disso, deve ser considerado que a turma recursal pode aumentar o valor de condenação (o que seria negativo para a empresa), ou diminuir o valor de condenação (o que seria negativo para o empregado). Assim, um empregado que receba oferta de acordo pouco abaixo do valor ótimo, poderia escolher aceitar considerando que a turma recursal julgando o processo tenha tendência de favorecer empresas e vice-versa.

6.1.4 Objetivo de mineração de dados

O objetivo de Mineração de Dados é **classificar automaticamente acórdãos judiciais em relação ao deferimento ou não e classificar automaticamente o requerente do recurso em relação a ser empresa ou a ser empregado**. Além disso, **classificar com os modelos desenvolvidos quantidade significativa de decisões judiciais e apresentar relatórios contendo as proporções de julgados positivamente para cada uma das partes**.

6.2 Compreensão dos dados

Primeiramente, procurou-se reunir a matéria prima da pesquisa, os acórdãos judiciais. Nesse caso, os acórdãos são os documentos judiciais que publicam a decisão dos magistrados indicando o deferimento ou não dos pedidos dos empregados e empregadores. Assim, esses documentos extraídos da internet vão compor a amostra para a execução da presente pesquisa. Ou seja, cada instância de observação da amostra é na verdade um acórdão judicial.

Tabela 6.1: Dados do Justiça em Números (CNJ, 2020) que apresentam a quantidade de casos novos ajuizados no ano de 2019.

Quantidade de casos novos em 2019	
TRT 4ª Região	267.036
TJ RS	1.413.893

Tais documentos são disponibilizados publicamente por todos os tribunais brasileiros em suas páginas de internet por meio de site de pesquisa em que os usuários podem informar palavras-chave para busca. Assim, por meio dessas páginas de busca, foi realizada uma exploração prévia da base de dados para reunir conhecimento sobre que tipos de documentos estão disponibilizados, qual a frequência e quantidade de publicação dos Tribunais, como os documentos são redigidos, que tipos de palavras são utilizadas, que tipos de metadados são disponibilizados em conjunto. Durante essa fase de exploração inicial, foram analisadas decisões judiciais da Justiça do Trabalho e da Justiça Comum do Rio Grande do Sul. Durante essa avaliação, foram considerados diversos fatores como a legibilidade dos documentos, quantidade de assuntos tratados nas decisões, estruturas comuns aos documentos, padrões seguidos pelos magistrados, além do conhecimento do pesquisador nas matérias do domínio do Direito analisadas.

Desse modo, decidiu-se por realizar a pesquisa utilizando documentos da Justiça do Trabalho. Tal decisão se baseia principalmente na experiência adquirida pelo pesquisador após anos de trabalho como servidor público da Justiça do Trabalho. Por outro lado, a Justiça do Trabalho trata de processos de causas estritamente trabalhistas, por ser uma Justiça especializada nesse tipo de causa, ao contrário do que ocorre na Justiça Comum, que trata de uma gama muito ampla de assuntos, desde cíveis até mesmo penais. Desse modo, supôs-se que os documentos criados abrangem uma gama menor de assuntos e que assim permitiriam uma análise mais acessível, objetiva e especializada. Apenas para título de comparação, podemos observar que a Justiça do Trabalho é muito menor que a Justiça Comum averiguando os dados de abertura de processos novos disponibilizados pela CNJ no relatório Justiça em Números (CNJ, 2020) e verificando a Tabela 6.1 com as informações resumidas. A partir desses dados, é possível verificar que a quantidade de processos abertos na Justiça do Estado do Rio Grande do Sul é aproximadamente 5 vezes maior.

6.2.1 Coleta de dados

Após a decisão de qual Justiça seria foco na pesquisa, passou-se a análise de quais páginas de internet seriam utilizadas para realizar a extração dos dados. Como os dados precisariam ser extraídos automaticamente, foram avaliados diversos sites para escolher os que não implementassem nenhum tipo de bloqueio a robôs de busca. Assim, foram escolhidos dois sites para extração de dados, o <<http://lexml.gov.br/>> e o <<http://trt4.jus.br/>>.

Por outro lado, de modo a realizar a extração de dados para evitar qualquer prejuízo aos órgãos públicos e seus domínios na internet, foram implementados métodos para a redução de qualquer possível dano gerado. Assim, foi colocado como diretriz da pesquisa o respeito a qualquer limitação imposta pelos próprios sites a robôs de busca. Tais limitações geralmente são encontradas nos termos de uso dos sites e em um arquivo chamado robots.txt na raiz do domínio. Além disso, não foram encontrados outros tipos de limitações, como por exemplo, a necessidade de o usuário provar que não é um robô pela leitura de caracteres em uma imagem, ou resolução de um problema de lógica que apenas um ser humano conseguiria resolver.

Ademais, nenhum dos sites apresenta qualquer documento jurídico indicando os termos de uso, tampouco apresentavam, à época da realização da atividade, qualquer limitação a robôs de busca de escanear todos os documentos públicos dos sites. Havia apenas bloqueios em relação a acessos às áreas privadas, como, por exemplo, páginas de intranet. Por outro lado, todos os documentos disponibilizados nos sites são de acesso público, sem nem a necessidade de utilizar login.

Além disso, todos os dados pessoais que possam identificar as partes não são publicados em nenhuma parte da presente pesquisa. Desse modo, manteve-se a anonimização das partes. Isso porque o mais importante para a pesquisa é analisar a massa de dados agregada, e não os indivíduos em si. Ademais, foi obtido por e-mail autorização para a utilização de decisões judiciais para a realização de pesquisa quantitativa com técnicas de inteligência artificial e jurimetria.

Enfim, o site do Tribunal Regional do Trabalho da 4ª Região foi escolhido por se tratar da Justiça do Trabalho com jurisdição em todo o território do Estado do Rio Grande do Sul o qual se torna mais interessante para a comunidade acadêmica da nossa Universidade como também para os cidadãos que usufruem das pesquisas realizadas na UFRGS. Por outro lado, foi decidido realizar a pesquisa também em outro Estado para permitir a

comparação e avaliação das possíveis diferenças nos dados estatísticos produzidos, como também observar possíveis diferenças de escritas nos documentos que poderiam afetar a performance de algoritmos desenvolvidos. Assim, as decisões do TRT da 4ª Região foram extraídas de seu próprio site e do site LexML foram extraídas as decisões do Tribunal Regional do Trabalho da 3ª Região. Tal Tribunal foi escolhido por ser o tribunal com mais decisões da Justiça do Trabalho disponibilizadas no site. Para ambos os tribunais optou-se por realizar a extração de decisões publicadas a partir do ano de 2017 até 2019.

6.2.1.1 Métodos de extração

Os robôs de extração de dados foram desenvolvidos utilizando a linguagem Python e a biblioteca Scrapy¹. Quanto ao site <lexml.gov.br>, eram disponibilizados diversos arquivos sitemap.xml na raiz do site para facilitar a extração de dados. Tais arquivos continham uma URL para cada documento disponibilizado. Além disso, por meio de cada URL era possível saber do que o documento tratava, permitindo assim a seleção de todos os documentos necessários a serem extraídos diretamente pelas URLs. Então, o conteúdo completo das decisões era disponibilizado em páginas HTML, as quais foram acessadas diretamente pelos robôs de extração de dados. Quanto ao site <trt4.jus.br>, há uma página de busca que permite a filtragem de documentos por meio de palavras-chave. Cada resultado é apresentado na mesma página com um link para o documento HTML contendo a decisão completa. Assim, foi desenvolvido um robô de extração de dados para simular um humano realizando a pesquisa no site, filtrar os resultados de acordo com os critérios necessários e acessar os links contendo os acórdãos judiciais completos.

6.2.1.2 Tamanho da amostra

A hipótese analisada exige a comparação da proporção de recursos de empregados deferidos com a proporção de recursos de empresas deferidos, conforme a Seção 6.1.1.1. Por outro lado, utilizou-se a biblioteca Statsmodels² para realização do teste estatístico, para cálculo das propriedades de efeito e força estatística, como também para cálculo do tamanho da amostra. Desse modo, o cálculo do tamanho da amostra exige como variáveis o tamanho das proporções sendo avaliadas, o tamanho da diferença entre as proporções que se deseja verificar, o nível de força estatística e o nível de confiança desejados. Entretanto, os valores dessas variáveis não são conhecidos antes da efetiva verificação dos

¹<<https://scrapy.org/>>

²<<https://www.statsmodels.org/stable/index.html>>

Tabela 6.2: Quantidade de decisões disponibilizadas e a quantidade da amostra inicial extraída para cada tribunal.

Tribunal	População	Amostra
TRT da 3ª Região	268.691	26.634
TRT da 4ª Região	140.545	29.894

dados.

Assim, buscou-se realizar a extração de decisões judiciais em certa quantidade que permitisse o processamento inicial e sua análise exploratória para permitir a verificação aproximada das variáveis necessárias. Por outro lado, importante considerar que o objetivo de negócio exige a realização de teste estatístico em relação às proporções em função de cada tribunal como também em relação a cada turma recursal. Desse modo, existe também a necessidade de se definir as quantidades mínimas de observações necessárias em relação a cada turma recursal, o que impõe outra dificuldade, pois não se tem conhecimento prévio das quantidades de processos deferidos para cada uma das partes em cada turma recursal. Enfim, a Tabela 6.2 apresenta o tamanho das amostras extraídas para análise inicial dos dados.

6.2.1.3 Dificuldades encontradas

Os robôs de extração de dados desenvolvidos não conseguiram extrair a totalidade dos documentos disponíveis em virtude de que houve problemas de requisições HTTP não respondidas pelos servidores, como também certa quantidade dos documentos foram disponibilizadas apenas em formato PDF. Enfim, foi possível extrair diretamente o texto puro da maioria dos arquivos, entretanto, os arquivos PDF precisaram ser processados em uma segunda etapa na fase de coleta.

6.2.2 Descrição dos dados

A fase de coleta dos dados resultou na compilação de 2 arquivos CSV em relação ao TRT da 3ª Região e 1 arquivo em relação ao TRT da 4ª Região, conforme Tabela 6.3. No caso do TRT da 3ª Região, foram gerados dois arquivos em virtude de dificuldades técnicas de processamento da carga de dados no momento da extração. Os arquivos CSV são compostos de um campo com o nome *INTEIRO_TEOR* contendo o texto puro dos acórdãos judiciais. Foi utilizada a biblioteca Pandas para manipulação dos arquivos nesse formato em virtude da facilidade de uso.

Tabela 6.3: Relação dos arquivos gerados a partir da extração dos dados dos tribunais.

Tribunal	Nome do arquivo	Quantidade de amostras efetivamente extraídas	Tamanho
TRT da 3ª Região	TRT3_inteiro_teor_2017_2018_2019_amostra_1.csv	17.846	245MB
TRT da 3ª Região	TRT3_inteiro_teor_2017_2018_2019_amostra_2.csv	8.788	120MB
TRT da 4ª Região	TRT4_inteiro_teor_2017_2018_2019_amostra.csv	29.894	541MB

Tabela 6.4: Descrição da base de dados extraída do TRT da 3ª Região em relação à quantidade de palavras.

	Palavras
Quantidade média	2.093
Desvio padrão	1.936
Quantidade mínima	1
25% percentil	810
50% percentil	1.457
75% percentil	2.690
Quantidade máxima	23.625

6.2.3 Exploração dos dados

As bases de dados foram analisadas separadamente, assim, o conjunto do TRT da 3ª Região apresentou o total de 26.634 instâncias, já o TRT da 4ª Região apresentou 29.894. A descrição da quantidade de palavras contidas nos documentos pode ser observada nas Tabelas 6.4 e 6.5. Os Gráficos contendo as distribuições dos unigramas, bigramas e trigramas é apresentada no Apêndice na Seção A.3.

6.2.4 Qualidade dos dados coletados

A qualidade das instâncias extraídas a partir dos arquivos HTML foi satisfatória na maioria dos casos, entretanto, houve algumas instâncias que continham apenas mensagens de erros com servidores, assim, tais documentos foram removidos na fase de Preparação

Tabela 6.5: Descrição da base de dados extraída do TRT da 4ª Região em relação à quantidade de palavras.

	Palavras
Quantidade média	2.738
Desvio padrão	2.357
Quantidade mínima	0
25% percentil	1.051
50% percentil	2.065
75% percentil	3.654
Quantidade máxima	23.391

dos Dados. Já em relação aos arquivos PDF, não foi possível utilizar nenhuma dessas instâncias. A extração do texto puro foi realizada por meio do aplicativo gratuito XpdfReader³, entretanto o processamento gerou arquivos com caracteres inseridos em partes incorretas do texto, especialmente caracteres referentes a cabeçalhos e rodapés.

6.3 Preparação dos dados

Nesta fase da pesquisa são realizados diversos processamentos com os documentos para prepará-los para serem usados na fase seguinte de modelagem de dados. Assim, são realizadas limpeza dos dados como também transformações para extrair e ressaltar as *features* mais importantes para a tarefa de classificação. Além disso, também é realizada a construção da base padrão-ouro, a qual é utilizada para os testes dos modelos desenvolvidos.

6.3.1 Limpeza dos dados

A limpeza dos documentos de texto foi realizada para garantir a qualidade mínima das instâncias para serem utilizadas nas fases subsequentes de transformação de dados e modelagem. Conforme as Tabelas 6.4 e 6.5, é possível observar que há instâncias contendo nenhuma ou apenas uma palavra. Assim, foi realizada a análise visual das instâncias contendo menos de 150 palavras e verificou-se que havia muitas mensagens de erros nos documentos. Algumas instâncias continham palavras irreconhecíveis. Enfim, considerou-se prudente remover todas as instâncias que continham menos de 150 palavras.

6.3.2 Transformação dos dados

Essa fase tem por objetivo extrair e ressaltar as *features* mais importantes para os algoritmos de modelagem de dados. Assim, são adicionados metadados às decisões, como também são removidos documentos e partes dos documentos que não são relevantes à pesquisa.

³<<https://www.xpdfreader.com/pdfotext-man.html>>

6.3.2.1 *Enriquecimento com metadados das decisões judiciais*

Os documentos coletados continham de maneira não estruturada em seu conteúdo de texto informações úteis importantes para a realização de agregações e análises em fases posteriores. Assim, foram extraídos por meio de expressões regulares os seguintes dados: data de publicação, nome do relator, órgão julgador, nome dos recorrentes, dispositivo do acórdão e a quantidade de recorrentes. Esses dados foram extraídos por meio da utilização de palavras-chave e strings Regex. O código-fonte Python e Regex pode ser encontrado no repositório da pesquisa⁴.

Entretanto houve dificuldades na aplicação dessa técnica pois certos documentos utilizavam um padrão para escrever essas informações e outros documentos usavam outro padrão. Assim, procurou-se criar expressões regulares que abrangessem o maior número de casos possíveis. Infelizmente, um pequeno número de instâncias não teve esses dados extraídos devido a um modo de escrita diverso e acabaram por ser removidas da base de dados. Enfim, todas as amostras em que as expressões regulares foram efetivas na extração foram enriquecidas com esses dados inseridos como colunas adjacentes.

6.3.2.2 *Extração do dispositivo da sentença*

De acordo com a Seção 2.1, as decisões proferidas pelos magistrados brasileiros devem ser redigidas seguindo certos padrões e conter determinados elementos. Um desses elementos é chamado de dispositivo da sentença, o qual é de caráter obrigatório e deve conter em todas as decisões e acórdãos publicados. Essa parte do texto deve conter particularmente a informação sobre se o recorrente teve seu pedido deferido ou não em poucos parágrafos, direto e objetivamente. Assim, essa parte do texto foi extraída, pois é nesse trecho que está a informação necessária a ser modelada pelo algoritmo de Aprendizado de Máquina a ser desenvolvido.

6.3.2.3 *Remoção de documentos com mais de um recorrente*

Muitos dos acórdãos publicados pelos tribunais dizem respeito a mais de um recurso impetrado, nesses casos, ambas as partes recorreram da decisão. Assim, o acórdão trata dos pedidos de recurso de ambos os recorrentes e o dispositivo da decisão precisa ser objetivo e dar uma resposta a cada um dos recursos.

Assim, esse tipo de documento, em que ambas as partes recorreram, exigiria uma

⁴<https://github.com/rhuanbarros/jurimetria_justica_do_trabalho>

anotação diferenciada para ser possível abranger por completo a semântica necessária para alcançar o objetivo de mineração. Nesse caso, seria necessário anotar individualmente os *tokens* do documento para ser possível identificar os elementos distintivos que representam cada uma das partes e o seu deferimento ou não.

Entretanto, construir um modelo que consiga identificar a resposta dos magistrados em relação a cada um dos recursos aumentaria o grau de dificuldade do modelo proposto a ser construído pois necessitaria da aplicação de demasiadas técnicas de aprendizado de máquina e processamento de linguagem natural. Assim, para não aumentar o nível de complexidade da pesquisa num primeiro momento, decidiu-se por não processar tais documentos. Além disso, o objetivo de negócio não exige especificamente que esse tipo de documento seja analisado pelo algoritmo, tornando, assim, fora do escopo do projeto analisar recursos com mais de um recorrente.

Desse modo, os documentos foram processados em busca de todos que continham mais de um recorrente por meio de análise do texto do campo *Recorrente* dos documentos contendo uma vírgula ou um *e*. Se houvesse, tal documento se tratava de mais de um recorrente e foram removidos da base. Enfim, após todo o processamento e remoção de instâncias desnecessárias, a base de documentos do Tribunal Regional do Trabalho da 3ª Região restou com o total de 10.875 instâncias, já o Tribunal Regional do Trabalho da 4ª Região com 12.071. Foram removidas 33.582 instâncias.

A remoção dessas instâncias da base de dados pode ter enfraquecido o nível de força do teste estatístico aplicado nas fases finais da pesquisa em vista de que houve diminuição considerável de documentos disponíveis para análise. Entretanto, neste ponto da pesquisa não é possível estimar o impacto visto que só é possível calcular o nível de força do teste estatístico após a realização do teste num momento em que a base de dados já foi totalmente processada em relação às proporções de julgados favoráveis a cada parte.

6.3.3 Anotação manual da base de documentos

Os documentos extraídos e processados nos capítulos anteriores foram salvos em formato CSV em ordem aleatória e analisados diretamente em aplicativo editor de planilha de texto. A anotação manual da base de documentos foi desenvolvida seguindo a metodologia apresentada no Capítulo 2.4 a qual inicia pela fase de modelagem do metamodelo de rótulos a serem aplicados. Assim, foi necessário que cada documento contivesse uma marcação em relação a se o requerente é empresa ou empregado e também outra marcação

indicando se o recurso foi deferido ou não.

6.3.3.1 Anotação para desenvolvimento do modelo de classificação do tipo de recorrente

Foram modelados dois rótulos utilizando a nomenclatura própria do Direito do Trabalho, conforme exposto na Seção 2.1.1: *RECLAMANTE*, que indica que a parte é empregado, e *RECLAMADA*, que indica que a parte é empresa. O modelo consiste de um vocabulário de termos T, do relacionamento entre esses termos R, e da interpretação I, conforme demonstrado abaixo.

- T = {Tipo_documento, RECLAMANTE, RECLAMADA}
- R = {Tipo_documento ::= RECLAMANTE | RECLAMADA}
- I = {RECLAMANTE = *Parte recorrente é empregado*, RECLAMADA = *Parte recorrente é empresa*}

6.3.3.2 Execução do processo de anotação manual em relação ao tipo de recorrente

A anotação dos documentos em relação ao tipo de recorrente foi realizada pelo próprio autor apenas, pois tal tarefa não exige nenhum conhecimento específico do Direito. A rotulação consistiu apenas na verificação do nome do recorrente ser nome de pessoa física ou jurídica. Assim, foi extraído aleatoriamente 270 instâncias de cada um dos tribunais em formato CSV, totalizando 540 observações. Essa quantidade foi definida arbitrariamente pelo autor após verificação da performance em testes de modelagem.

6.3.3.3 Anotação para desenvolvimento do modelo de classificação em relação ao deferimento ou indeferimento da decisão

A princípio foram modelados brevemente dois rótulos: *DEFERIMENTO* para indicar uma decisão que houve julgamento favorável e *INDEFERIMENTO* para indicar uma decisão que não houve julgamento favorável. Após, foi iniciada a fase de anotação pelo próprio autor, o qual observou que seria necessária a criação de outro rótulo pois havia documentos que não se encaixavam em nenhum dos casos. Assim, foi criado o rótulo *SEM ANÁLISE DE MÉRITO*.

Foi necessária a criação desse rótulo pois há acórdãos em que não há análise quanto aos pedidos contidos no recurso devido a algum elemento que prejudique sua apreciação pelos magistrados, conforme esclarecido no Capítulo 2.1.2. Entretanto, es-

ses acórdãos sem julgamento de mérito ocorrem em uma minoria de casos. Enfim, foi inevitável o retorno à fase de modelagem para adequação do modelo com o novo rótulo.

Além disso, considerou-se importante a criação do rótulo *SEM ANÁLISE DE MÉRITO* para ser possível verificar o tamanho da distribuição desse tipo de documento na base de dados. Em fases finais de análise quantitativa e cálculo de proporções de julgados deferidos e indeferidos, poderia haver a dúvida em relação a qual classe esse tipo de documento haveria sido classificada, em virtude de não haver sido treinado o classificador com esse tipo de documento previamente. Assim, poderia ser levantado o questionamento se esse tipo de documento não estaria inflando umas das proporções de deferidos ou indeferidos. Enfim, abaixo é apresentado o modelo criado para essa tarefa de anotação.

- T = {Tipo_documento, DEFERIMENTO, INDEFERIMENTO, SEM_ANALISE_MERITO}
- R = {Tipo_documento ::= DEFERIMENTO | INDEFERIMENTO | SEM_ANALISE_MERITO}
- I = {DEFERIMENTO = *Julgamento favorável ao recorrente*, INDEFERIMENTO = *Julgamento desfavorável ao recorrente*, SEM_ANALISE_MERITO = *Julgamento não é favorável nem desfavorável ao recorrente*}

6.3.3.4 Execução do processo de anotação manual em relação ao deferimento ou indeferimento da decisão

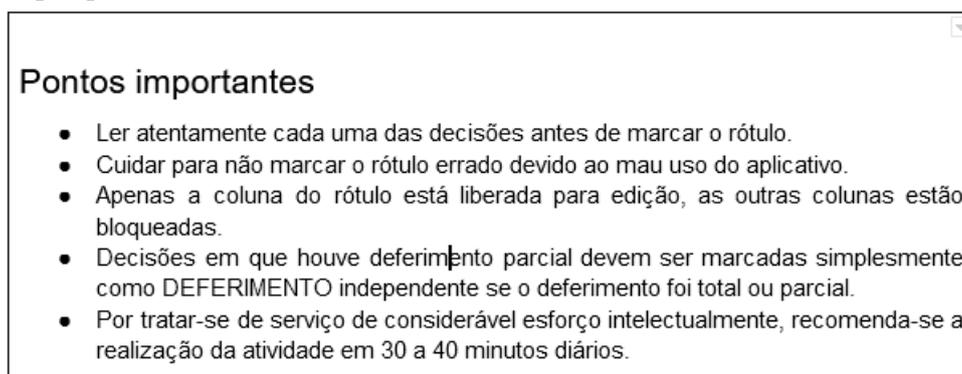
A rotulação dos documentos foi realizada por três pessoas, sendo o autor e dois voluntários bacharéis em Direito. Assim, foram elaboradas planilhas online no aplicativo Google Suite com as decisões para anotação. Constaram uma coluna com as decisões e outra para inserir o rótulo em cada linha respectivamente. A Figura A.8 apresenta foto da tela da planilha com algumas decisões de exemplo. Assim, foram criadas duas planilhas, uma para cada anotador. Em cada planilha foi inserido 250 acórdãos do TRT da 3ª Região e 250 acórdãos do TRT da 4ª Região. Ou seja, cada voluntário realizou a anotação de 500 documentos. Além disso, foi feita cópia de ambas as planilhas para serem anotadas pelo autor da pesquisa. Enfim, houve o total de 1000 documentos anotados por duas pessoas. Cada um dos documentos foi anotado por um dos anotadores voluntários e pelo autor da pesquisa. Cabe salientar que o autor da pesquisa detém conhecimentos jurídicos de nível técnico e mais de 10 anos de experiência profissional na Justiça brasileira, incluindo tempo de trabalho especificamente em varas do trabalho.

Vale ressaltar que a amostra selecionada para a tarefa de anotação inclui apenas decisões em que houve apenas um recorrente, o que tornou a tarefa de anotação mais simples porque havia apenas um recurso para ser apreciado, ou seja, apenas um recurso para ser deferido ou indeferido. Além disso, o processamento final da base de documentos para estimar a tendência de opinião dos magistrados também foi realizado com decisões em que houve apenas um recorrente, conforme exposto na Seção 6.3.2.3.

Por outro lado, houve a preocupação de criar a planilha para os anotadores de modo a facilitar a correta edição do documento. Assim, as áreas do documento que continham as decisões foram bloqueadas para edição, deixando livre para edição apenas a coluna *Rótulo*. Também houve a aplicação de cores alternadas para o fundo das linhas para facilitar a leitura, visualização e edição.

A seguir foi realizado o desenvolvimento das diretrizes de anotação a qual contém orientações gerais aos anotadores. Tal documento contém uma seção de introdução que expõe resumidamente o objetivo da pesquisa e como funciona o trabalho de anotação de documentos por anotadores voluntários. Após, são apresentados os rótulos com a devida explicação detalhada de como cada um deve ser usado. Além disso, são apresentados exemplos de decisões e o respectivo rótulo considerado correto pelo autor da pesquisa. Por fim, é apresentada uma lista de pontos importantes a serem observados pelos anotadores. A Figura 6.1 apresenta trecho do documento que pode ser encontrado completo no Apêndice A.4.

Figura 6.1: Trecho do documento Diretrizes para anotação manual de documentos jurídicos para pesquisa de mestrado de Rhuan Barros



6.3.3.5 Avaliação da tarefa de anotação e criação do padrão-ouro

Conforme exposto na Seção 2.4.1, foi realizada a avaliação do nível de concordância entre os rótulos aplicados pelos anotadores. Assim, a Tabela 6.6 apresenta a matriz

Tabela 6.6: Matriz de confusão em relação aos rótulos aplicados

		P2	P2	P2	
		DEFERIMENTO	INDEFERIMENTO	SEM_ANALISE_MERITO	TOTAL
P1	DEFERIMENTO	435	6	1	442
P1	INDEFERIMENTO	4	534	1	539
P1	SEM_ANALISE_MERITO	1	6	12	19
	TOTAL	440	546	14	1000

Tabela 6.7: Balanceamento das classes da base de dados anotada em relação ao tipo de requerente.

CLASSE	QUANTIDADE
RECLAMANTE	312
RECLAMADA	226

de confusão em relação aos rótulos aplicados. Aplicando a fórmula Cohen's Kappa é obtido o valor de 0,963 que é considerado alto nível de concordância. Após, foi realizada a adjudicação da base de dados em que o próprio autor resolveu as discordâncias, sendo possível observar algumas instâncias de exemplo no Apêndice A.5.

6.3.4 Exploração das bases de dados anotadas manualmente

Nesta fase da pesquisa é realizada novamente uma exploração da base de documentos, entretanto, neste caso é possível obter informações essenciais para as fases seguintes, pois os documentos receberam anotações. Assim, é possível analisar a frequência das palavras correlacionadas a cada categoria.

6.3.4.1 Base de dados anotada manualmente quanto ao tipo de requerente

O processamento da base de dados anotada manualmente quanto ao tipo de requerente gerou análises quanto ao balanceamento das classes e também em relação às palavras mais comuns. Assim, na Tabela 6.7 e na Figura 6.2 é possível observar o balanceamento das classes, sendo que a classe RECLAMADA contém pouco menos instâncias anotadas. Quanto à análise dos unigramas mais comuns em cada classe, é possível verificar que, para a classe RECLAMANTE, sobrenomes comuns no Brasil, como *Silva* e *Santos* figuram como mais frequentes. Já para a classe RECLAMADA, as palavras *Ltda*, *Brasil* e *SA* ficam respectivamente em primeiro, segundo e terceiro lugar na lista de mais frequentes.

Figura 6.2: Histograma apresentando o balanceamento das classes da base de dados anotada em relação ao tipo de reclamante.

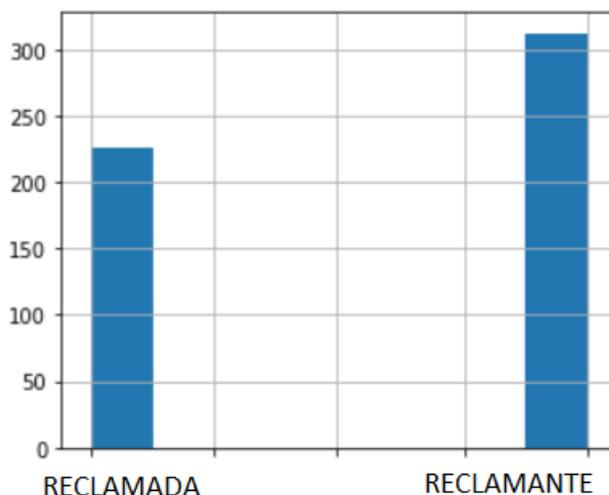


Tabela 6.8: Balanceamento das classes da base de dados anotada em relação ao deferimento ou não da decisão.

CLASSE	QUANTIDADE
INDEFERIMENTO	539
DEFERIMENTO	441
SEM_ANALISE_MERITO	20

6.3.4.2 Base de dados anotada manualmente quanto ao deferimento ou não da decisão

Foi realizado o processamento e análise da base de dados anotada manualmente para explorar o balanceamento das classes e as palavras mais comuns usadas no campo *Dispositivo* em função do rótulo recebido separadamente para cada tribunal. Assim, na Tabela 6.8 e na Figura 6.3 é possível verificar que há desbalanceamento da classe SEM_ANALISE_MERITO, a qual contém apenas 20 instâncias de um total de 1000. Quanto ao texto do campo *Dispositivo*, foi verificado que a maioria das instâncias contém até 250 palavras nesse campo, conforme Figura 6.4. Além disso, é possível observar na Tabela 6.9 que magistrados escrevem menos quando indeferem um recurso em comparação com os recursos deferidos.

Por outro lado, também foi elaborada análise de termos e suas associações em relação a cada classe anotada por meio da biblioteca Python Scattertext (KESSLER, 2017). Assim, é possível observar os termos mais associados com cada uma das classes anotadas para ambos os tribunais na Figura 6.5 e nas listas abaixo os termos mais associados para cada uma das classes. Os termos foram selecionados por meio de *tokenização* realizada utilizando expressões regulares que abrangem características da língua portuguesa.

Figura 6.3: Histograma apresentando o balanceamento das classes da base de dados anotada em relação ao deferimento ou não da decisão.

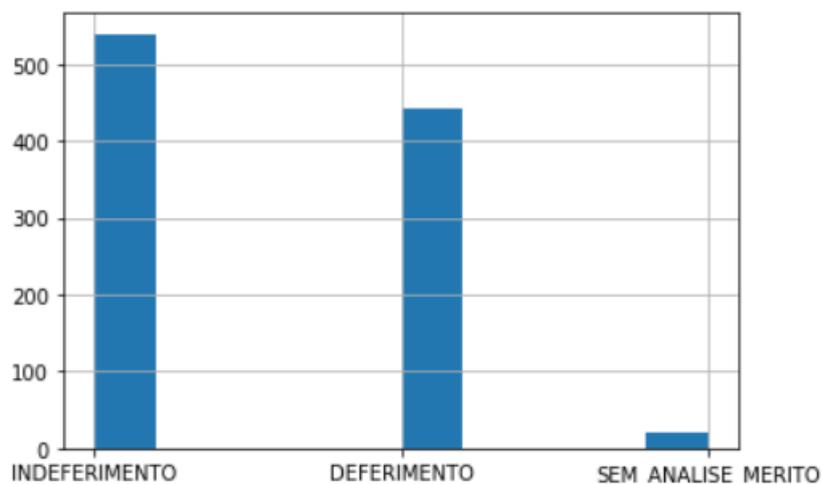
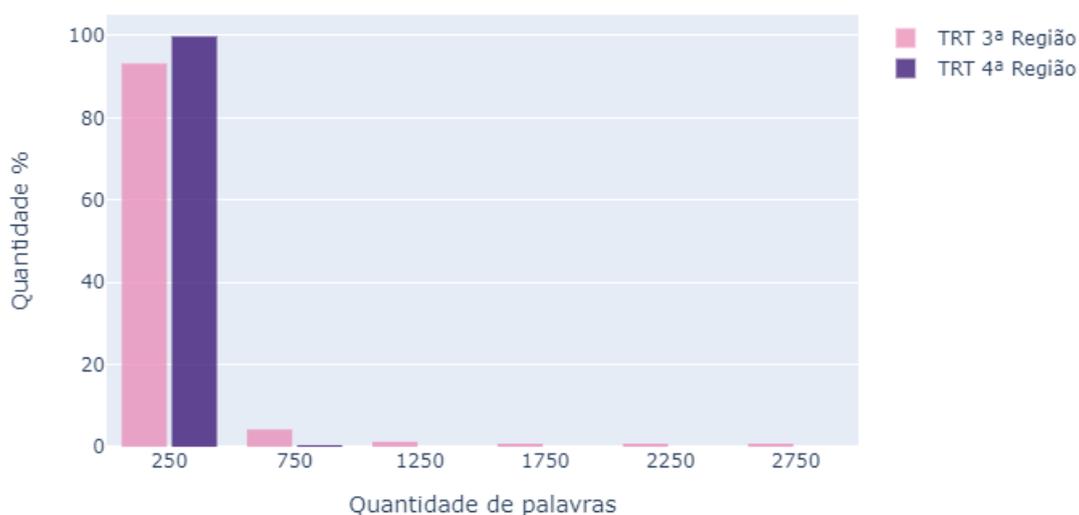


Figura 6.4: Histograma quantidade de palavras no dispositivo por tribunal



- Termos mais associados com os acórdãos anotados como DEFERIMENTO.
 - 'deu lhe',
 - 'dar provimento',
 - 'divergência deu',
 - 'dar',
 - 'unanimidade dar',
 - 'à condenação',
 - 'parcial',
 - 'condenar',

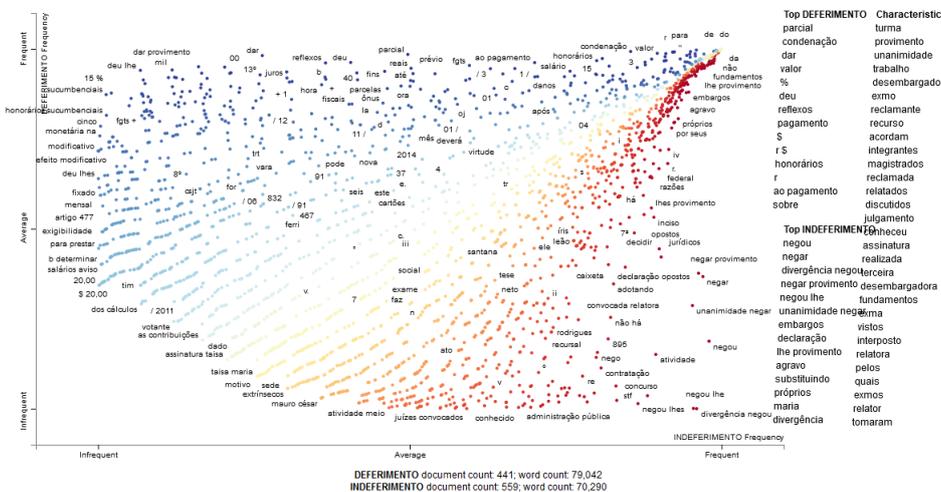
Tabela 6.9: Descrição da base de dados anotada manualmente em função da quantidade de palavras no campo *Dispositivo* em relação a cada rótulo aplicado em cada tribunal.

Tribunal	Rótulo	Palavras						
		Contagem de instâncias	Média	Mínimo	25%	50%	75%	Máximo
TRT3	DEFERIMENTO	197	248	75	133	161	219	2540
	INDEFERIMENTO	288	182	43	98	112	133	2743
	SEM_ANALISE_MERITO	15	290	95	103	123	154	1572
TRT4	DEFERIMENTO	244	124	34	83	113	149	494
	INDEFERIMENTO	251	53	35	44	47	50	995
	SEM_ANALISE_MERITO	5	84	64	69	84	90	117

- 'reflexos',
- 'provimento para'
- Termos mais associados com os acórdãos anotados como INDEFERIMENTO.
 - 'divergência negou',
 - 'negou lhe',
 - 'negou',
 - 'negar',
 - 'negar provimento',
 - 'unanimidade negar',
 - 'embargos de',
 - 'de declaração',
 - 'embargos',
 - 'declaração'
- Termos mais associados com os acórdãos anotados como SEM_ANALISE_MERITO.
 - 'ação de',
 - 'de cobrança',
 - 'cobrança de',
 - 'cobrança',
 - 'comum',
 - 'competência',
 - 'não conheceu',
 - 'estadual',
 - 'justiça comum',
 - 'rel',

A partir da anotação manual dos documentos e também de sua exploração por

Figura 6.5: Gráfico que apresenta os termos mais associados com cada uma das classes anotadas. Mais próximas do canto superior esquerdo encontram-se palavras mais associadas com o rótulo *Deferimento*. Já próximas do canto inferior direito encontram-se palavras mais associadas com o rótulo *Indeferimento*.



meio de técnicas de mineração de texto, foi possível observar diversos padrões de escrita dos magistrados particulares a sua área de domínio. Especialmente em relação aos recursos negados, é comum os magistrados utilizarem as mesmas palavras, como, por exemplo, “negar” e suas variações. Por outro lado, é possível observar que os magistrados têm o hábito de utilizar a palavra “dar” e suas variações para representar o deferimento de um recurso.

6.3.5 Anotação automática da base de documentos por meio de Supervisão Fraca

A anotação automática de documentos por meio de Supervisão Fraca permite o desenvolvimento rápido de bases de dados para a utilização em algoritmos de classificação de documentos por meio de Aprendizagem de Máquina Supervisionada, além de que a técnica permite a anotação de grande quantidade de instâncias diretamente. Assim, supôs-se que tal técnica poderia ajudar a mitigar ponto negativo da pesquisa em relação a pequena base de documentos anotada manualmente na Seção 6.3.3. Além disso, a inspiração para utilização dessa técnica foi obtida em vista de que o autor percebeu durante a fase de anotação manual que muitos termos se repetiam e que isso poderia ser usado a favor da pesquisa. Desse modo, optou-se por avaliar a performance dessa técnica em documentos jurídicos. Enfim, foi utilizado o Snorkel Framework (RATNER et al., 2017) para execução dessa etapa.

Tabela 6.10: Lista de funções de rotulação criadas para aplicação automática de rótulos à base de dados.

Função de Rotulação	Palavra-chave	Classificação
lf_negar_provimento	negar provimento	INDEFERIMENTO
lf_rejeitar	rejeitar	INDEFERIMENTO
lf_nao_acolher	não acolher	INDEFERIMENTO
lf_manter_decisao	manter decisão	INDEFERIMENTO
lf_julgou_improcedente	julgou improcedente	INDEFERIMENTO
lf_dar_provimento	dar provimento	DEFERIMENTO
lf_dar_parcial_provimento	dar parcial provimento	DEFERIMENTO
lf_proveu	proveu	DEFERIMENTO
lf_parcialmente_proveu	parcialmente proveu	DEFERIMENTO
lf_nulidade_sentenca	nulidade sentença	SEM_ANALISE_MERITO
lf_nao_conhecer	não conhecer	SEM_ANALISE_MERITO
lf_deixar_conhecer	deixar conhecer	SEM_ANALISE_MERITO
lf_prejudicada	prejudicada	SEM_ANALISE_MERITO

Assim, prosseguiu-se com a análise dos termos mais associados às classes anotadas em relação ao deferimento do acórdão na Seção 6.3.4. Desse modo, foi possível observar que existem certos padrões na redação dos documentos pelos magistrados, os quais podem ser explorados por meio da criação de *Label Functions* do Snorkel. Dessa maneira, a análise dessas informações gerou o desenvolvimento de heurísticas que processam por meio de Regex a presença ou não de certas palavras identificadoras do deferimento ou não do recurso pelo magistrado, conforme pode ser observado pela Tabela 6.10. Importante ressaltar que as palavras-chave são testadas utilizando o seu *lemma* processadas por meio da biblioteca Spacy ⁵. Além disso, há a remoção de *stopwords* que incluem pronomes oblíquos átonos os quais muitas vezes se encontram no meio de locuções verbais e poderiam interferir na localização das palavras-chave.

Em vista de que foram desenvolvidas diversas Funções de Rotulação para realizar a tarefa de classificação dos documentos, houve a sobreposição de classificação em diversas instâncias, como é possível observar pela coluna *Sobreposições* da Tabela 6.12, entretanto a maioria delas obteve porcentagens abaixo de 10% o que indica baixa sobreposição. Além disso, podemos observar que houve aproximadamente 10% de rotulações conflitantes em relação às classes INDEFERIMENTO e DEFERIMENTO, e menos de 5% em relação à classe SEM_ANALISE_MERITO, ou seja, nesses casos houve uma anotação indicando ao menos duas classes diferentes. Importante ressaltar que as instâncias que receberam o rótulo ABSTAIN não entram na contagem de conflitantes. O rótulo ABSTAIN é aplicado quando a função de rotulação não tem conhecimento suficiente para

⁵<<https://spacy.io/>>

aplicar um rótulo, conforme explicado na Seção 2.6.5. Por outro lado, é possível observar pela Figura 6.6 que aproximadamente 80% das instâncias receberam um rótulo e que em torno de 15% receberam dois ou mais rótulos.

Após, foi utilizado o algoritmo LabelModel do Snorkel para realizar o processamento da matriz contendo todas as classificações pelas Funções de Rotulação de modo gerar um modelo probabilístico a fim de ser usado para processar a base de dados completa e finalmente assinalar um rótulo final a cada instância. Cabe salientar que esse modelo probabilístico desenvolvido não tem capacidade de generalização ao realizar o processamento de instâncias que as Funções de Rotulação também não tiveram capacidade de aplicar um rótulo, em vista disso, instâncias que receberam o rótulo ABSTAIN, continuaram recebendo esse rótulo nessa fase de processamento.

Desse modo, o modelo treinado foi aplicado à totalidade 22.946 instâncias da base de documentos disponível, tendo efetivamente aplicado rótulo de INDEFERIMENTO, DEFERIMENTO ou SEM_ANALISE_MERITO a 97,91%, como é possível verificar pela Figura 6.7. Também é possível observar que aproximadamente 2% das instâncias receberam o rótulo ABSTAIN, ou seja, o modelo não aplicou nenhum rótulo. Portanto, a tarefa de anotação automática de documentos por meio de técnica de Supervisão Fraca gerou uma nova base de dados contendo o total de 22.471 instâncias.

Importante considerar que os tribunais avaliados disponibilizam centenas de milhares de decisões online de dezenas de anos passados. Entretanto, por dificuldades técnicas, não foram possíveis de serem extraídos. Dessa maneira, em trabalhos futuros, essa totalidade de documentos poderiam ser extraídos e processados por meio do Snorkel Framework o que tenderia a aumentar o nível de acurácia alcançado.

Além disso, é possível observar que o modelo generativo do Snorkel Framework não contém capacidade de generalização o que se traduz na sua incapacidade de rotular instâncias que as Funções de Rotulação não continham conhecimento de domínio suficiente para aplicar um rótulo correto e por fim aplicaram o rótulo ABSTAIN, nesse caso, em 2% das instâncias disponíveis. Entretanto, a capacidade de generalização é uma característica desejada em aplicações prática de modo que o modelo tenha habilidade de se adaptar e rotular instâncias previamente não vistas e oriundas da mesma distribuição.

Assim, essa incapacidade de generalização impede a utilização do modelo generativo do Snorkel Framework para a classificação final da base de dados da presente pesquisa. Desse modo, na Seção 6.4, foram realizados experimentos para o desenvolvimento de modelo de Aprendizado de Máquina utilizando algoritmos que contenham a

Tabela 6.11: Quadro resumo contendo as bases de dados criadas na presente pesquisa incluindo as quantidades de instâncias por classe.

Base de dados	Quantidade por classe			
	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO	Total
Padrão-ouro	539	441	20	1.000
Anotada automaticamente	12.000	9.923	548	22.471
Anotada automaticamente balanceada	548	548	548	1.644

capacidade de generalização. E por fim, na Seção 6.5 foi aplicado o modelo desenvolvido selecionado a totalidade de 22.946 instâncias. O que permitiu a anotação da totalidade de documentos disponíveis.

Ademais, da mesma maneira que ocorreu com a base de dados anotada manualmente, houve grande desbalanceamento da classe SEM_ANALISE_MERITO. Desse modo, foi aplicada técnica de *Under-sampling* a fim de serem removidas aleatoriamente instâncias das classes INDEFERIMENTO e DEFERIMENTO para conterem finalmente a mesma quantidade que a classe SEM_ANALISE_MERITO, totalizando 548 instâncias para cada classe. Portanto, a base de dados final balanceada criada programaticamente foi gerada contendo o total de 1.644 instâncias. Enfim, por meio da Tabela 6.11, é possível observar a comparação entre todas as bases de dados criadas na presente pesquisa incluindo as quantidades de instâncias por classe.

A checagem dos documentos anotados automaticamente foi realizada pelo próprio autor pela simples observação de algumas instâncias de modo a verificar erros grotescos. Por outro lado, seria inviável realizar a checagem total, além disso, não faz parte da técnica de Supervisão fraca a realização dessa checagem visto que o objetivo é construir a base automaticamente de maneira menos dispendiosa possível em relação ao tempo necessário e em relação ao custo financeiro. Além disso, a checagem é, de certo modo, realizada ao final da pesquisa por meio do treinamento do modelo utilizando a base de dados criada programaticamente e por fim aplicação desse modelo a base de dados padrão-ouro.

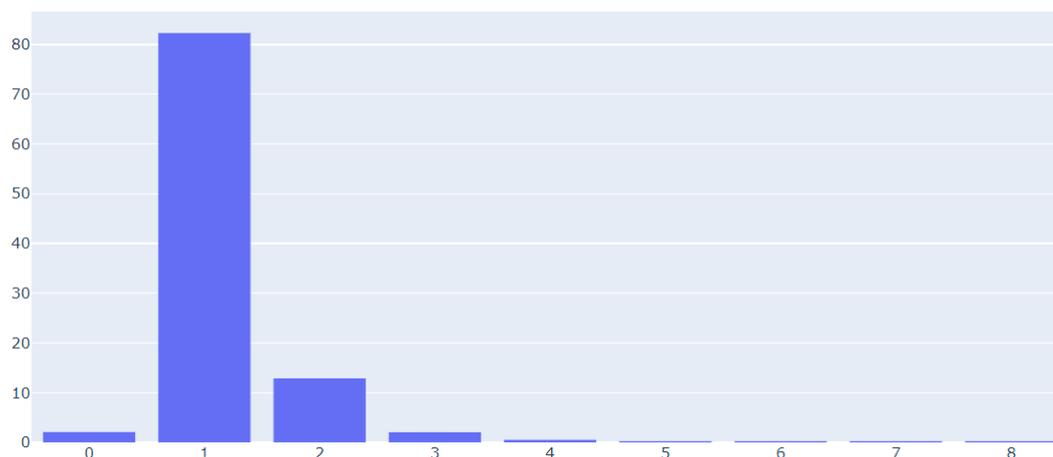
6.4 Modelagem

Para a atender ao objetivo de mineração de dados contido na Seção 6.1.4, primeiramente foi desenvolvido um modelo para classificação dos documentos em relação ao tipo de recorrente. Após, foi desenvolvido um modelo para classificação quanto ao deferimento ou indeferimento do dispositivo do acórdão. Em ambos os casos foram utilizadas as respectivas bases de documentos anotadas manualmente conforme exposto na Seção 6.3.3. Além disso, para o segundo modelo, foi utilizada também base de documentos

Tabela 6.12: Apresenta lista com as Funções de Rotulação aplicadas à base de dados e respectivos dados de cobertura, sobreposições e conflitos.

Função de rotulação	Polaridade	Cobertura	Sobreposições	Conflitos
If_negar_provimento	INDEFERIMENTO	0,489410	0,090125	0,063192
If_rejeitar	INDEFERIMENTO	0,072823	0,055173	0,033078
If_nao_acolher	INDEFERIMENTO	0,016691	0,000959	0,000567
If_manter_decisao	INDEFERIMENTO	0,007060	0,006755	0,002876
If_julgou_improcedente	INDEFERIMENTO	0,013902	0,011418	0,010198
If_dar_provimento	DEFERIMENTO	0,307548	0,073651	0,068160
If_dar_parcial_provimento	DEFERIMENTO	0,168483	0,037741	0,033819
If_proveu	DEFERIMENTO	0,019873	0,018260	0,014512
If_parcialmente_proveu	DEFERIMENTO	0,001700	0,001700	0,001525
If_nulidade_sentenca	SEM_ANALISE_MERITO	0,007757	0,007104	0,006406
If_nao_conhecer	SEM_ANALISE_MERITO	0,047372	0,030114	0,029766
If_deixar_conhecer	SEM_ANALISE_MERITO	0,003486	0,002310	0,002048
If_prejudicada	SEM_ANALISE_MERITO	0,020047	0,017999	0,017214

Figura 6.6: Gráfico de histograma que apresenta as porcentagens dos rótulos aplicados pela Funções de Rotulação.



anotada programaticamente conforme Seção 6.3.5.

6.4.1 Modelo: classificação do tipo de requerente como empregado ou empresa

A fase de preparação e exploração dos dados das seções anteriores resultou em uma lista contendo nomes de partes recorrentes a qual foi anotada manualmente com um rótulo indicando se a parte é empregado ou empresa. Assim, nesta seção são realizados experimentos para avaliar algoritmos de classificação de texto utilizando a base anotada construída.

A fase de avaliação técnica dos modelos desenvolvidos em relação aos níveis dos resultados apresentados foi realizada por meio das bases de dados desenvolvidas na fase

Figura 6.7: Gráfico de histograma que apresenta as porcentagens dos rótulos finais aplicados pelo modelo desenvolvido.



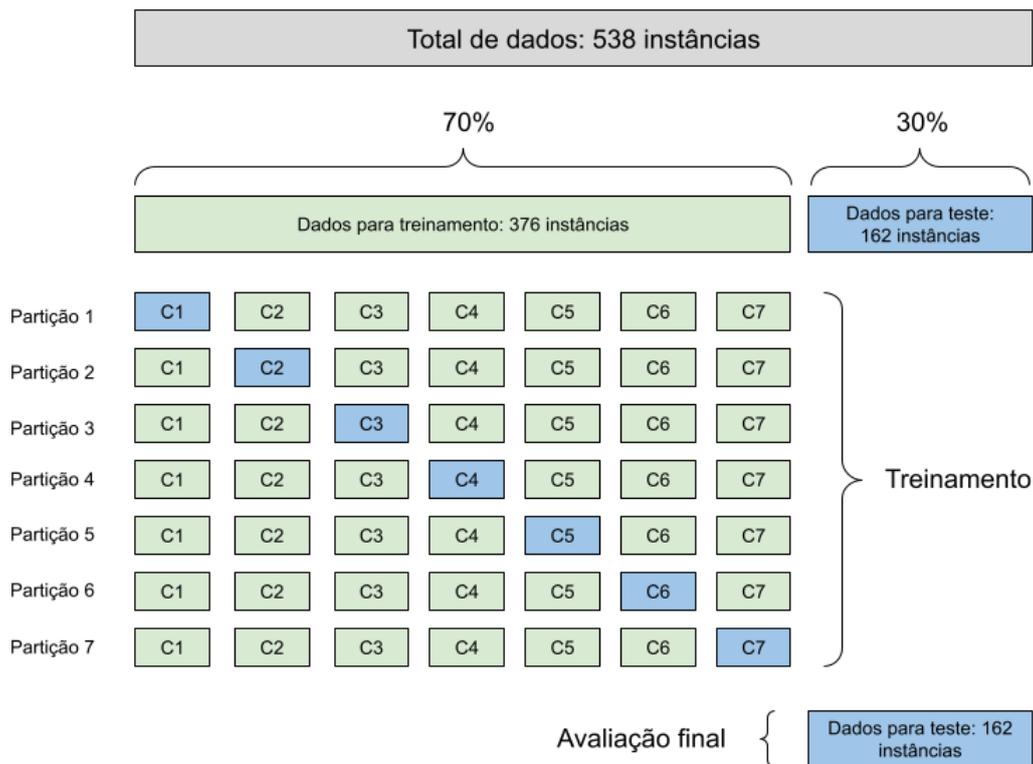
de anotação manual. Assim, a base de dados foi dividida em 2 partes, 30% para testes e 70% para treinamento dos modelos. Já essa parte de 70% dos dados foi utilizada para treinamento utilizando técnica de Validação Cruzada em 7 camadas, conforme Figura 6.8. As bases de documentos do TRT da 3ª e 4ª Região foram utilizadas em conjunto para treinamento e testes. Além disso, foram escolhidas as métricas acurácia, F1, revocação e precisão para comparação dos resultados.

O treinamento do modelo necessário para classificar o tipo de requerente sendo empregado ou empresa foi desenvolvido a partir de experimentos realizados utilizando os algoritmos de classificação da biblioteca Scikit Learn. Foram testados os seguintes algoritmos: Rocchio classifier, Gradient Boosting Classifier, Naive Bayes Classifier, K-nearest Neighbor, Support Vector Machine (SVM), Decision Tree, Random Forest.

Não foram utilizados algoritmos de Redes Neurais em vista de que esses algoritmos necessitam de bases de dados contendo uma grande quantidade de instâncias e muita diversidade de *features*, o que não seria o caso desta base de treinamento. Por outro lado, de modo a garantir a diversidade de métodos de processamento, foram selecionados algoritmos que utilizam técnicas probabilísticas, técnicas de distância no espaço vetorial e técnicas de *ensemble* de árvores de decisões.

Desse modo, os experimentos foram realizados utilizando os dados de requerentes anotados manualmente de ambos os tribunais de maneira conjunta, ou seja, a lista de requerentes anotada de ambos os tribunais foi unida em uma estrutura de dados única e preparada para o treinamento. Em relação à extração de *features* foi utilizada a classe CountVectorizer da biblioteca Scikit Learn para realizar a transformação de todas as pa-

Figura 6.8: Ilustração do projeto de teste do classificador de tipo de requerente informando as quantidades específicas de instâncias da base de dados.



lavras para minúsculas e também o algoritmo padrão dessa classe para realização de *tokenização* o qual extrai todas as palavras com dois ou mais caracteres, além de tratar caracteres de pontuação como separadores de palavras.

Após, foi aplicada a classe `TfidfTransformer` a qual gera uma matriz contendo a representação TF-IDF. Não houve a aplicação de nenhuma técnica de redução de dimensionalidade específica, tampouco foi realizado tratamento de stop-words ou extração de *Lemma* das palavras. Quanto aos algoritmos de classificação, foram utilizados os parâmetros padrão. Essa configuração mostrou-se satisfatória na fase de avaliação de resultados e por isso foi mantida.

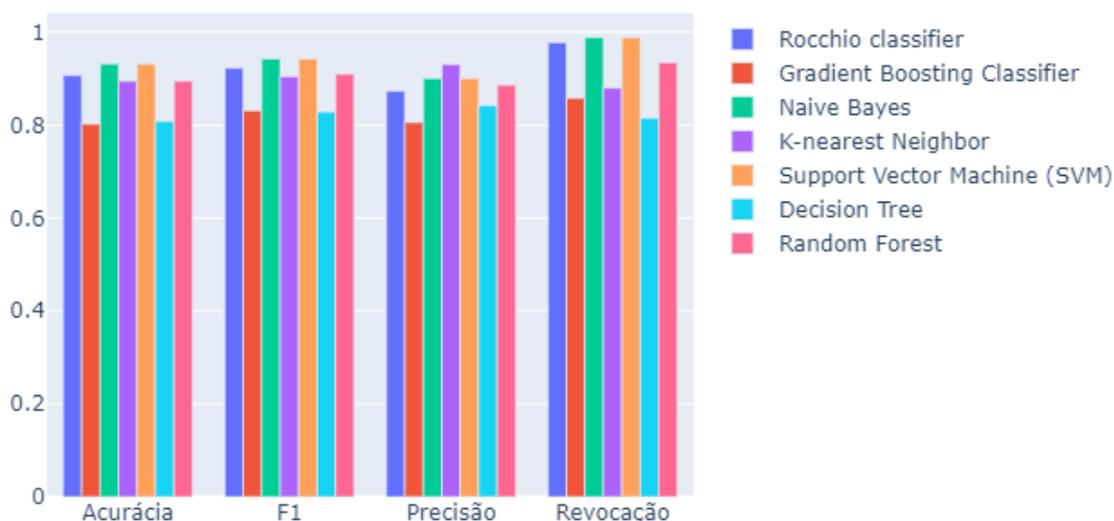
Durante a fase de desenvolvimento, foi utilizada apenas a base reservada de 70% dos dados para a realização de experimentos. Assim, foi utilizada técnica de Validação Cruzada em 7 camadas, sendo calculada a média das sete execuções para avaliar a performance dos algoritmos. Importante ressaltar que também foram utilizadas as matrizes de confusão para verificação de desempenho.

Após o desenvolvimento, foi utilizada a base de 30% para verificação final de desempenho dos algoritmos. Assim, foi selecionado o algoritmo com melhor valor de

F1 para a construção do modelo final a ser utilizado na Seção 6.5 para realização de processamento da base de documentos a fim de se atingir o objetivo de negócio e de mineração de texto. Além disso, a construção desse modelo foi realizada utilizando a totalidade da base de documentos anotada manualmente de modo a utilizar a integralidade de instâncias e de *features* disponíveis.

A avaliação técnica dos modelos desenvolvidos foi realizada conforme o projeto de teste. Na Tabela 6.13 e na Figura 6.9 são apresentados os valores de acurácia, F1, revocação e precisão, sendo possível observar que o algoritmo Support Vector Machine (SVM) obteve o maior valor de F1 com 96,91%, seguido pelo algoritmo Naive Bayes com 95,48%. A Tabela 6.14 apresenta a lista de parâmetros escolhidos para o modelo treinado com o algoritmo Support Vector Machine (SVM). A descrição completa dos parâmetros pode ser encontrada no site oficial ⁶.

Figura 6.9: Métricas calculadas utilizando a base de 30% reservada previamente para testes do modelo de classificação do tipo de requerente.



6.4.2 Modelo: classificar a decisão em deferimento ou indeferimento

A fase de preparação e exploração dos dados das seções anteriores resultou em uma base de documentos contendo milhares de decisões judiciais. Assim, foi realizada a anotação manual de 1.000 instâncias desses documentos para compor a base padrão-ouro. Assim, nesta seção são realizados experimentos para avaliar algoritmos de classificação de

⁶<https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html>

Tabela 6.13: Métricas calculadas utilizando a base de 30% reservada previamente para testes do modelo de classificação do tipo de requerente.

Classificador	Acurácia	F1	Precisão	Revocação
Rocchio	93,83%	94,95%	93,07%	96,91%
Gradient Boosting	87,65%	90,29%	85,32%	95,88%
Naive Bayes	94,44%	95,48%	93,14%	97,94%
K-nearest Neighbor	90,74%	92,15%	93,62%	90,72%
Support Vector Machine (SVM)	96,30%	96,91%	96,91%	96,91%
Decision Tree	85,80%	87,43%	93,02%	82,47%
Random Forest	91,36%	93,07%	89,52%	96,91%

Tabela 6.14: Parâmetros e argumentos do modelo de aprendizado de máquina treinado utilizando o algoritmo SVM LinearSVC da biblioteca Scikit Learning.

Parâmetro	Argumento
C	1.0
class_weight	None
dual	True
fit_intercept	True
intercept_scaling	1
loss	squared_hinge
max_iter	1000
multi_class	ovr
penalty	l2
random_state	None
tol	1
verbose	0

texto utilizando a base anotada construída. Também foi construída base de documentos anotada automaticamente para avaliação de técnica de Supervisão Fraca, de técnica de balanceamento de classes e de algoritmos de classificação.

6.4.2.1 Experimento com base de dados criada manualmente padrão-ouro

Inicialmente são realizados experimentos de modelagem de dados utilizando a base padrão-ouro. Assim é utilizada a base contendo 1.000 instâncias, sendo destas 539 da classe INDEFERIMENTO, 441 da classe DEFERIMENTO e 20 da classe SEM_ANALISE_MERITO. Desse modo, é possível observar que há grande desbalanceamento das classes.

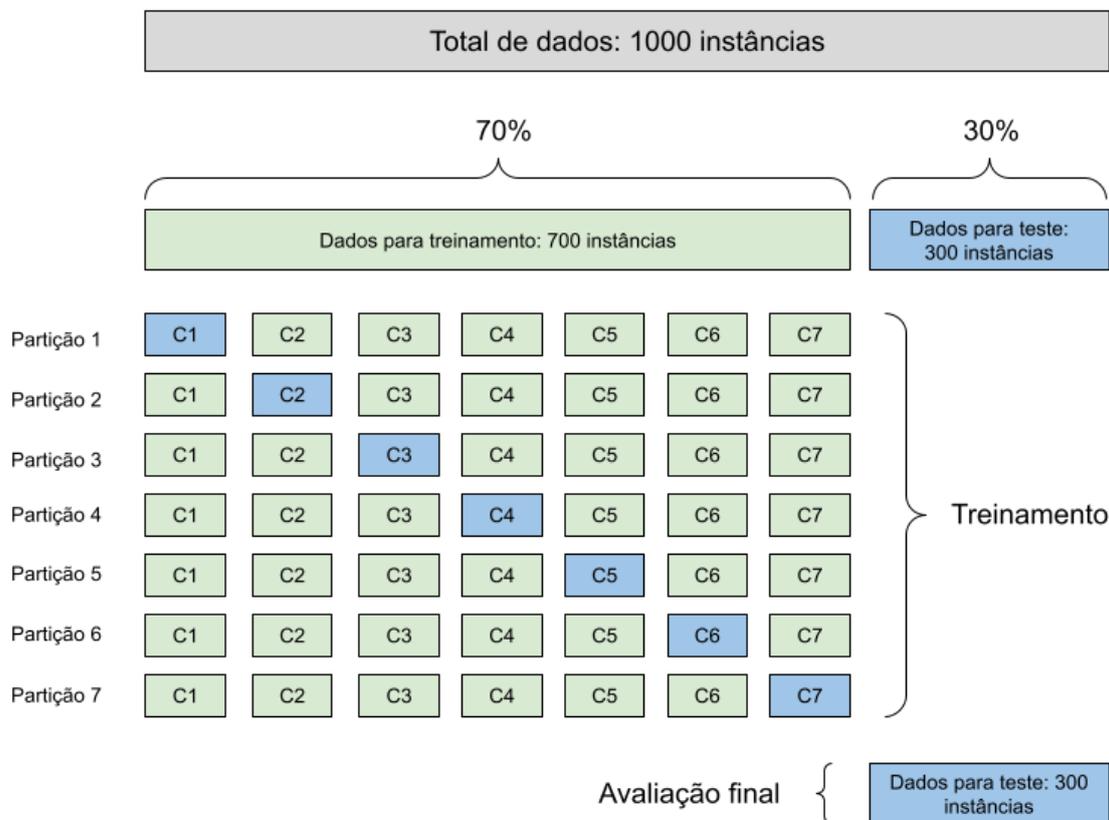
A avaliação técnica dos modelos desenvolvidos em relação aos níveis dos resultados apresentados foi realizada por meio da base de dados padrão-ouro desenvolvida conforme exposto na Seção 6.3.3. Assim, a base de dados foi dividida em 2 partes, 30% para testes e 70% para treinamento dos modelos. Já essa parte de 70% dos dados foi utilizada para treinamento utilizando técnica de Validação Cruzada em 7 camadas, conforme Figura 6.10. As bases de documentos do TRT da 3^a e 4^a Região foram utilizadas em conjunto para treinamento e testes. Além disso, foram escolhidas as métricas acurácia, f1, revocação e precisão para comparação dos resultados, sendo que, para as três últimas, foram calculadas suas métricas macro e micro por tratar-se de uma tarefa com três classes. Importante ressaltar que também foram utilizadas as matrizes de confusão para verificação de desempenho.

O treinamento dos modelos necessários para classificar o dispositivo da decisão foi desenvolvido a partir de experimentos realizados utilizando os algoritmos de classificação da biblioteca Scikit Learn. Foram testados os seguintes algoritmos: Rocchio classifier, Gradient Boosting Classifier, Naive Bayes Classifier, K-nearest Neighbor, Support Vector Machine (SVM), Decision Tree, Random Forest.

Não foram utilizados algoritmos de Redes Neurais em vista de que esses algoritmos necessitam de bases de dados contendo uma grande quantidade de instâncias e muita diversidade de *features*, o que não seria o caso desta base de treinamento. Por outro lado, de modo a garantir a diversidade de métodos de processamento, foram selecionados algoritmos que utilizam técnicas probabilísticas, técnicas de distância no espaço vetorial e técnicas de *ensemble* de árvores de decisões.

Desse modo, os experimentos foram realizados utilizando os dados da base padrão-ouro de ambos os tribunais de maneira conjunta, ou seja, a lista de decisões ano-

Figura 6.10: Ilustração do projeto de teste do classificador de tipo de requerente informando as quantidades específicas de instâncias da base de dados.



tada de ambos os tribunais foi unida em uma estrutura de dados única e preparada para o treinamento. Em relação à extração de *features* foi utilizada a classe `CountVectorizer` da biblioteca `Scikit Learn` para gerar a matriz de termos e documentos. Em conjunto a essa classe, foi utilizada a biblioteca `Spacy`⁷ para processamento do *lemma* dos termos. Também foi utilizado código `Regex` de *tokenização* personalizado para a língua portuguesa que abrangesse palavras com traços, além da remoção de *stop-words*.

Após, foi aplicada a classe `TfidfTransformer` a qual gera uma matriz contendo a representação TF-IDF. Não houve a aplicação de nenhuma técnica de redução de dimensionalidade específica. Quanto aos algoritmos de classificação, foram utilizados os parâmetros padrão. Essa configuração mostrou-se satisfatória na fase de avaliação de resultados e por isso foi mantida.

Durante a fase de desenvolvimento, foi utilizada apenas a base reservada de 70% dos dados para a realização de experimentos. Assim, foi utilizada técnica de Validação Cruzada em 7 camadas, sendo calculada a média das sete execuções para avaliar a performance dos algoritmos. Importante ressaltar que também foram utilizadas as matrizes de

⁷<<https://spacy.io/>>

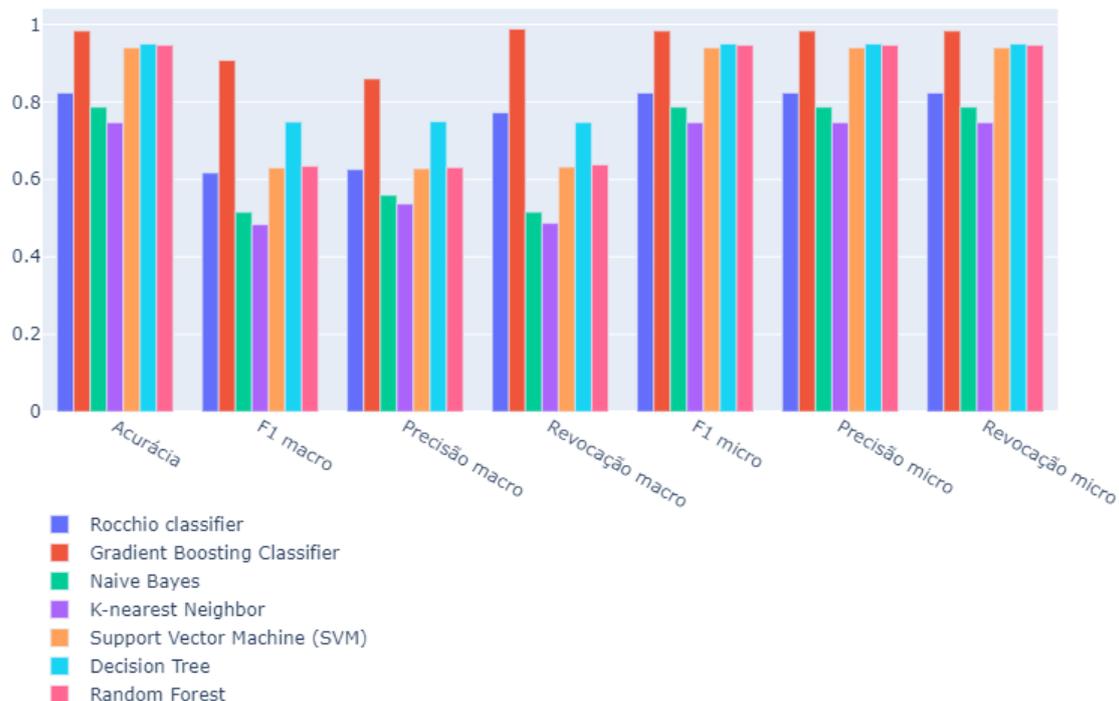
Tabela 6.15: Modelos treinados com base padrão-ouro e métricas calculadas utilizando a base de 30% reservada previamente para testes do modelo de classificação do deferimento ou não da decisão.

Classificador	Acurácia	F1 macro	Precisão macro	Revocação macro	F1 micro	Precisão micro	Revocação micro
Rocchio	82,33%	61,68%	62,55%	77,23%	82,33%	82,33%	82,33%
Gradient Boosting	98,33%	90,73%	85,96%	98,84%	98,33%	98,33%	98,33%
Naive Bayes	78,67%	51,56%	55,91%	51,57%	78,67%	78,67%	78,67%
K-nearest Neighbor	74,67%	48,33%	53,65%	48,66%	74,67%	74,67%	74,67%
SVM	94,00%	62,97%	62,78%	63,19%	94,00%	94,00%	94,00%
Decision Tree	95,00%	74,82%	74,96%	74,71%	95,00%	95,00%	95,00%
Random Forest	94,67%	63,43%	63,08%	63,79%	94,67%	94,67%	94,67%

confusão para verificação de desempenho. Após o desenvolvimento, foi utilizada a base de 30% para verificação final de desempenho dos algoritmos.

Os modelos desenvolvidos foram avaliados conforme o projeto de teste. É possível observar por meio da Tabela 6.15 e da Figura 6.11 que há grande discrepância entre os valores macro e micro. Por outro lado, verificando as matrizes de confusão apresentadas na Seção A.6, é visto que a maioria das instâncias da classe SEM_ANALISE_MERITO não foram classificadas corretamente pelos algoritmos que tiveram as piores performances em relação às métricas macro. Isso se deve ao fato do grande desbalanceamento dessa classe, conforme Figura 6.3.

Figura 6.11: Modelos treinados com base padrão-ouro e métricas calculadas utilizando a base de 30% reservada previamente para testes do modelo de classificação do deferimento ou não da decisão.



6.4.2.2 Experimento com base de dados criada programaticamente balanceada

Primeiramente é realizada a modelagem dos dados utilizando a base de dados criada automaticamente balanceada. Nesse caso, a base contém 1.644 instâncias, sendo destas 548 da classe INDEFERIMENTO, 548 da classe DEFERIMENTO e 548 da classe SEM_ANALISE_MERITO, estando assim, as classes totalmente niveladas.

A avaliação técnica dos modelos desenvolvidos em relação aos níveis dos resultados apresentados foi realizada por meio da base de dados padrão-ouro desenvolvida conforme exposto na Seção 6.3.3. Desse modo, houve o treinamento com a base de dados balanceada criada programaticamente e realizado o teste cruzado com a base anotada manualmente. Da mesma maneira, neste caso foram escolhidas as métricas acurácia, f1, revocação e precisão para comparação dos resultados, sendo que, para as três últimas, foram calculadas suas métricas macro e micro por tratar-se de uma tarefa com três classes. Importante ressaltar que também foram utilizadas as matrizes de confusão para verificação de desempenho.

O treinamento dos modelos foi realizado utilizando os algoritmos de classificação da biblioteca Scikit Learn. Foram testados os seguintes algoritmos: Rocchio classifier, Gradient Boosting Classifier, Naive Bayes Classifier, K-nearest Neighbor, Support Vector Machine (SVM), Decision Tree, Random Forest.

Não foram utilizados algoritmos de Redes Neurais em vista de que esses algoritmos necessitam de bases de dados contendo uma grande quantidade de instâncias e muita diversidade de *features*, o que não seria o caso desta base de treinamento. Por outro lado, de modo a garantir a diversidade de métodos de processamento, foram selecionados algoritmos que utilizam técnicas probabilísticas, técnicas de distância no espaço vetorial e técnicas de *ensemble* de árvores de decisões.

Desse modo, os experimentos foram realizados utilizando a base de dados balanceada criada programaticamente, conforme Seção 6.3.5, incluindo os dados de ambos os tribunais de maneira conjunta, ou seja, a lista de decisões anotada de ambos os tribunais foi unida em uma estrutura de dados única e preparada para o treinamento. Em relação à extração de *features* foi utilizada a classe CountVectorizer da biblioteca Scikit Learn para gerar a matriz de termos e documentos. Em conjunto a essa classe, foi utilizada a biblioteca Spacy⁸ para processamento do *lemma* dos termos. Também foi utilizado código Regex de *tokenização* personalizado para a língua portuguesa que abrangesse palavras com traços, além da remoção de *stop-words*. Após, foi aplicada a classe TfidfTransfor-

⁸<<https://spacy.io/>>

Tabela 6.16: Modelos treinados utilizando base criada programaticamente balanceada e métricas calculadas utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.

Classificador	Acurácia	F1 macro	Precisão macro	Revocação macro	F1 micro	Precisão micro	Revocação micro
Rocchio classifier	0,689	0,5658	0,6361	0,7606	0,689	0,689	0,689
Gradient Boosting Classifier	0,946	0,8475	0,8049	0,9446	0,946	0,946	0,946
Naive Bayes	0,748	0,6009	0,6401	0,7957	0,748	0,748	0,748
K-nearest Neighbor	0,741	0,5826	0,5963	0,6705	0,741	0,741	0,741
Support Vector Machine (SVM)	0,942	0,8179	0,776	0,9271	0,942	0,942	0,942
Decision Tree	0,944	0,8441	0,7977	0,9599	0,944	0,944	0,944
Random Forest	0,956	0,8684	0,8201	0,9692	0,956	0,956	0,956

mer a qual gera uma matriz contendo a representação TF-IDF. Não houve a aplicação de nenhuma técnica de redução de dimensionalidade específica. Quanto aos algoritmos de classificação, foram utilizados os parâmetros padrão.

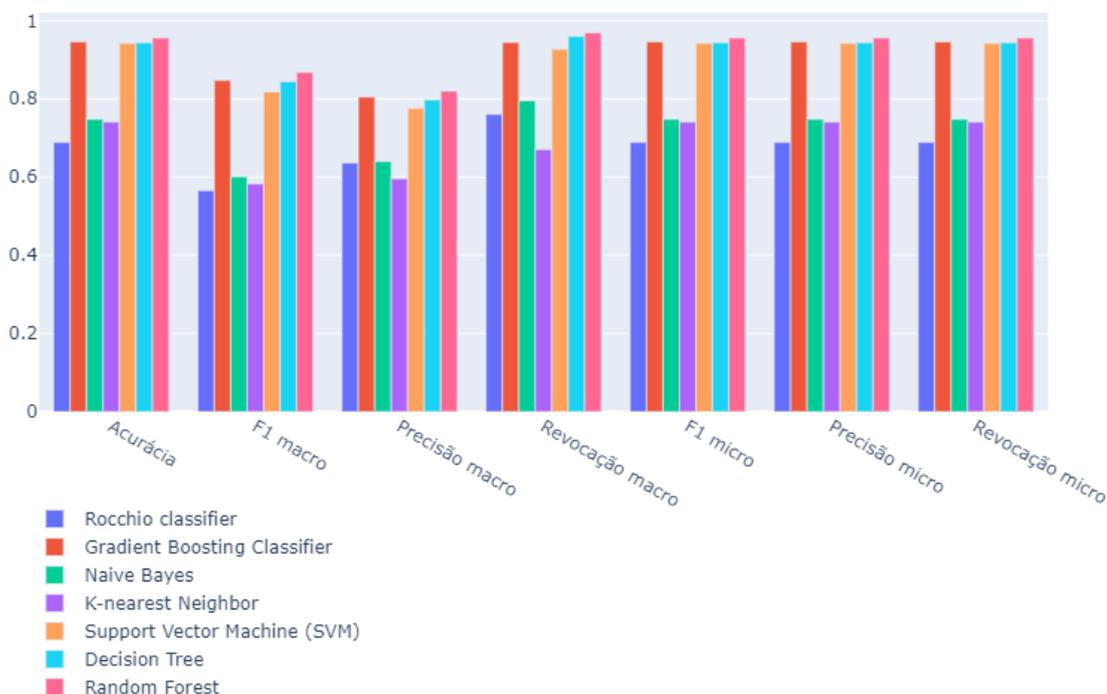
Os modelos desenvolvidos foram avaliados conforme o projeto de teste. Desse modo, os modelos foram treinados utilizando toda a base de dados criada programaticamente balanceada e foram usados para classificar toda a base padrão-ouro. A Tabela 6.16 e a Figura 6.12 apresentam as métricas de avaliação. É possível verificar que a performance máxima dos modelos em relação a métrica f1 macro foi inferior ao valor reportado na Seção 6.4.2.1 tendo nenhum classificador alcançado valor maior que 90%. Por outro lado, houve a concentração de algoritmos com média f1 macro entre 80% e 90%. Além disso, avaliando as matrizes de confusão presentes na Seção A.7 é possível verificar que todos os algoritmos tiveram classificações corretas em relação à classe SEM_ANALISE_MERITO, diferente do que ocorreu no experimento da Seção 6.4.2.1, o que pode ser considerado um efeito positivo da base de dados balanceada com número maior de instâncias.

6.4.2.3 Experimento com base de dados criada programaticamente completa

Neste experimento foi realizada a modelagem dos dados utilizando a base de dados criada automaticamente completa. Assim, foi utilizada a base anotada automaticamente completa contendo 22.471 instâncias, sendo destas 12.000 da classe INDEFERIMENTO, 9.923 da classe DEFERIMENTO e 548 da classe SEM_ANALISE_MERITO. Desse modo, é possível observar que também há grande desbalanceamento das classes.

Os experimentos realizados para desenvolvimento do modelo para classificar o dispositivo da decisão foram realizados utilizando os algoritmos de classificação da biblioteca Scikit Learn. Foram testados os seguintes algoritmos: Rocchio classifier, Gradient Boosting Classifier, Naive Bayes Classifier, K-nearest Neighbor, Support Vector Machine (SVM), Decision Tree, Random Forest.

Figura 6.12: Modelos treinados utilizando base criada programaticamente balanceada e métricas calculadas utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.



Não foram utilizados algoritmos de Redes Neurais em vista de que esses algoritmos necessitam de bases de dados contendo uma grande quantidade de instâncias e muita diversidade de *features*, o que não seria o caso desta base de treinamento. Por outro lado, de modo a garantir a diversidade de métodos de processamento, foram selecionados algoritmos que utilizam técnicas probabilísticas, técnicas de distância no espaço vetorial e técnicas de *ensemble* de árvores de decisões.

Desse modo, os experimentos foram realizados utilizando a base de dados completa criada programaticamente, conforme Seção 6.3.5, incluindo os dados de ambos os tribunais de maneira conjunta, ou seja, a lista de decisões anotada de ambos os tribunais foi unida em uma estrutura de dados única e preparada para o treinamento. Em relação a extração de *features* foi utilizada a classe `CountVectorizer` da biblioteca Scikit Learn para gerar a matriz de termos e documentos. Em conjunto a essa classe, foi utilizada a biblioteca Spacy⁹ para processamento do *lemma* dos termos. Também foi utilizado código Regex de *tokenização* personalizado para a língua portuguesa que abrangesse palavras com traços, além da remoção de *stop-words*.

Após, foi aplicada a classe `TfidfTransformer` a qual gera uma matriz contendo

⁹<https://spacy.io/>

Tabela 6.17: Modelos treinados utilizando base criada programaticamente completa e métricas calculadas utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.

Classificador	Acurácia	F1 macro	Precisão macro	Revocação macro	F1 micro	Precisão micro	Revocação micro
Rocchio classifier	0,721	0,584	0,6306	0,7813	0,721	0,721	0,721
Gradient Boosting Classifier	0,954	0,9248	0,9053	0,9504	0,954	0,954	0,954
Naive Bayes	0,846	0,5673	0,5721	0,5696	0,846	0,846	0,846
K-nearest Neighbor	0,806	0,5648	0,7378	0,552	0,806	0,806	0,806
Support Vector Machine (SVM)	0,961	0,9157	0,9237	0,9083	0,961	0,961	0,961
Decision Tree	0,951	0,9168	0,8936	0,9484	0,951	0,951	0,951
Random Forest	0,96	0,9179	0,9137	0,9231	0,96	0,96	0,96

a representação TF-IDF. Não houve a aplicação de nenhuma técnica de redução de dimensionalidade específica. Quanto aos algoritmos de classificação, foram utilizados os parâmetros padrão. Essa configuração mostrou-se satisfatória na fase de avaliação de resultados e por isso foi mantida.

Os modelos desenvolvidos foram avaliados conforme o projeto de teste. Desse modo, os modelos foram treinados utilizando toda a base de dados criada programaticamente e foram usados para classificar toda a base padrão-ouro. A Tabela 6.17 e a Figura 6.13 apresentam as métricas de avaliação. É possível verificar que houve quatro modelos que obtiveram performance máxima em relação a métrica f1 macro superiores aos valores reportados na Seção 6.4.2.1 e na Seção 6.4.2.2.

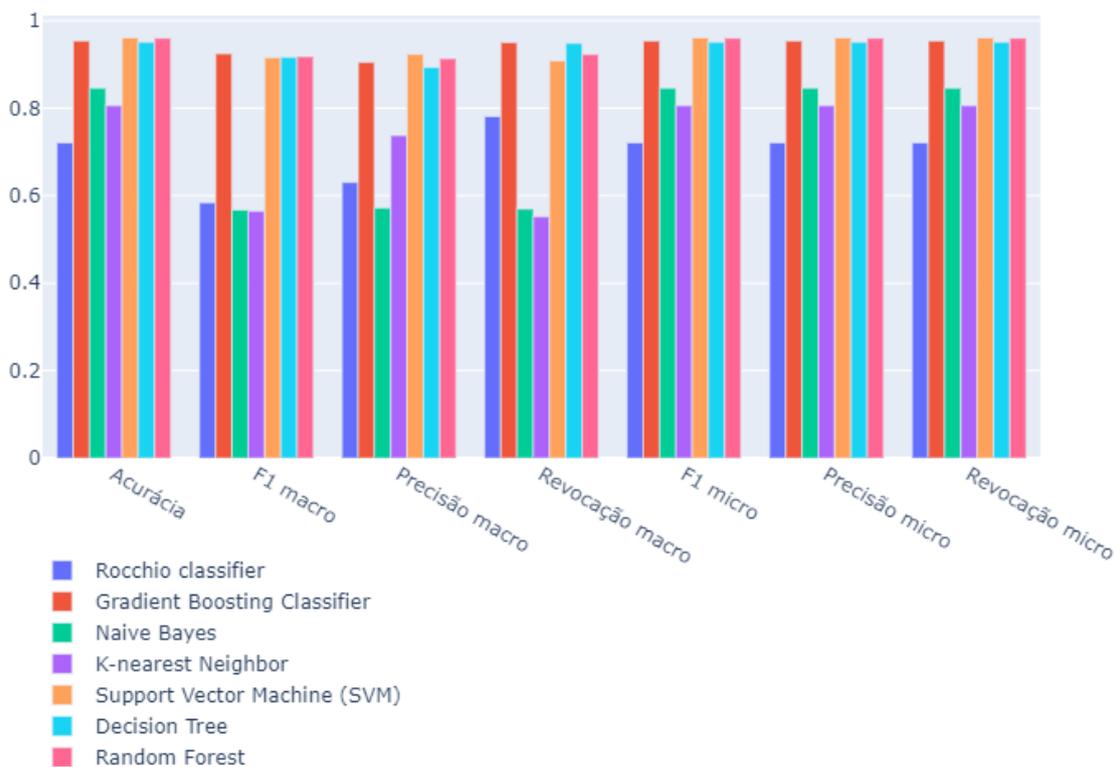
Além disso, avaliando as matrizes de confusão presentes na Seção 6.4.2.1 é possível verificar que alguns algoritmos não obtiveram classificações em relação à classe SEM_ANALISE_MERITO o que resultou em métricas f1 macro em torno de 60%. Enfim, o modelo que obteve a melhor média f1 macro, a saber, Gradient Boosting, foi selecionado para realizar a classificação final de toda a base de documentos para análise na Seção 6.5. A Tabela 6.18 apresenta a lista de parâmetros escolhidos para o modelo treinado com o algoritmo Gradient Boosting. A descrição completa dos parâmetros pode ser encontrada no site oficial ¹⁰.

6.5 Avaliação de resultados

O processo de mineração de dados textuais de decisões judiciais elaborado resultou no desenvolvimento de dois modelos baseados em aprendizado de máquina supervisionado e na sua utilização para o processamento de 22.946 decisões judiciais, sendo 12.071 do TRT da 4ª Região e 10.875 do TRT da 3ª Região. Com a utilização desses modelos, os objetivos de mineração foram atingidos que eram classificar automaticamente

¹⁰<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html>

Figura 6.13: Modelos treinados utilizando base criada programaticamente completa e métricas calculadas utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.



os acórdãos judiciais em relação ao deferimento ou não e classificar automaticamente o requerente do recurso em relação a ser empresa ou a ser empregado. Cabe salientar que, após a aplicação dos modelos de classificação, todas as instâncias que receberam o rótulo SEM_ANALISE_MERITO foram removidas da análise, pois não interessavam aos objetivos de negócio e de mineração de texto.

Desse modo, foi possível atender também ao objetivo de negócio para verificar a validade da hipótese por meio de teste estatístico se seria possível que os tribunais aliados e suas turmas recursais julguem favoravelmente proporção maior de recursos para uma das partes do que para outra. Assim, também são apresentadas a conclusão e a interpretação dos testes estatísticos de proporção, força e efeito. Após, foram desenvolvidos relatórios com gráficos para apresentar os resultados obtidos com o processamento dos documentos.

Tabela 6.18: Parâmetros e argumentos do modelo de aprendizado de máquina treinado utilizando o algoritmo Gradient Boosting da biblioteca Scikit Learning.

Parâmetro	Argumento
ccp_alpha	0.0
criterion	friedman_mse
init	None
learning_rate	0.1
loss	deviance
max_depth	3
max_features	None
max_leaf_nodes	None
min_impurity_decrease	0.0
min_samples_leaf	1
min_samples_split	2
min_weight_fraction_leaf	0.0
n_estimators	100
n_iter_no_change	None
random_state	1.0
subsample	0.0001
tol	0.1
validation_fraction	0.1
verbose	0
warm_start	False

6.5.1 TRT da 4ª Região - avaliação geral

Na Figura 6.14 é possível observar a porcentagem da quantidade de recursos deferidos e indeferidos no TRT da 4ª Região em relação ao recorrente ser empresa ou empregado. Assim, é possível verificar que 50% dos recursos impetrados pelos empregados foram deferidos ou parcialmente deferidos. Além disso, também é possível verificar que, dos recursos impetrados pelas empresas, 39% foram julgados deferidos ou parcialmente deferidos. Por outro lado, fica evidente a diferença de 12% na quantidade de recursos deferidos ou parcialmente deferidos quando o recorrente são os empregados. Importante frisar que essa média leva em consideração amostra dos acórdãos proferidos por todas as Turmas Recursais do tribunal em conjunto.

Nessa amostra foram processados 11.351 acórdãos, sendo 9.398 recursos da parte empregado e 1.953 recursos da parte empresa. A parte empresa recorreu aproximadamente 4 vezes menos do que a parte empregado. Além disso, em relação aos testes estatísticos aplicados, foi obtido *P-value* de 0%, ou seja, há grande confiança de que existe diferença estatística entre a proporção de deferimentos para empregado e para empresa.

Figura 6.14: Porcentagem de deferimento de recursos em relação ao recorrente ser empresa e empregado no Tribunal Regional do Trabalho da 4ª Região.

TRT da 4ª Região. Porcentagem de julgamentos favoráveis para cada uma das partes.



Tabela 6.19: Porcentagem de deferimento de recursos em relação ao recorrente ser empresa e empregado das Turmas Recursais do Tribunal Regional do Trabalho da 4ª Região.

TRIBUNAL	TURMA RECURSAL	TOTAL DE PROCESSOS	PVALUE	PVALUE INTERP.	POWER	POWER INTERP.	EFFECT	EFFECT INTERP.
TRT4	1ª Turma	1.639	0,7172	NÃO HÁ DIFERENÇA	0,0863	INACEITÁVEL	0,0217	INSIGNIFICANTE
TRT4	2ª Turma	1.704	0	HÁ DIFERENÇA	1	ACEITÁVEL	0,5267	MÉDIA
TRT4	3ª Turma	1.832	0	HÁ DIFERENÇA	1	ACEITÁVEL	0,4853	PEQUENA
TRT4	4ª Turma	1.800	0,0343	HÁ DIFERENÇA	0,9575	ACEITÁVEL	0,1343	INSIGNIFICANTE
TRT4	5ª Turma	1.504	0,0003	HÁ DIFERENÇA	0,9999	ACEITÁVEL	0,2355	PEQUENA
TRT4	6ª Turma	1.171	0,1741	NÃO HÁ DIFERENÇA	0,638	INACEITÁVEL	0,1051	INSIGNIFICANTE
TRT4	7ª Turma	825	0,1687	NÃO HÁ DIFERENÇA	0,6817	INACEITÁVEL	0,1307	INSIGNIFICANTE
TRT4	8ª Turma	448	0,0032	HÁ DIFERENÇA	0,9988	ACEITÁVEL	0,366	PEQUENA
TRT4	9ª Turma	131	0,1106	NÃO HÁ DIFERENÇA	0,8029	ACEITÁVEL	0,3813	PEQUENA
TRT4	10ª Turma	115	0,0752	NÃO HÁ DIFERENÇA	0,7266	INACEITÁVEL	0,3892	PEQUENA
TRT4	11ª Turma	182	0,3001	NÃO HÁ DIFERENÇA	0,4141	INACEITÁVEL	0,2038	PEQUENA

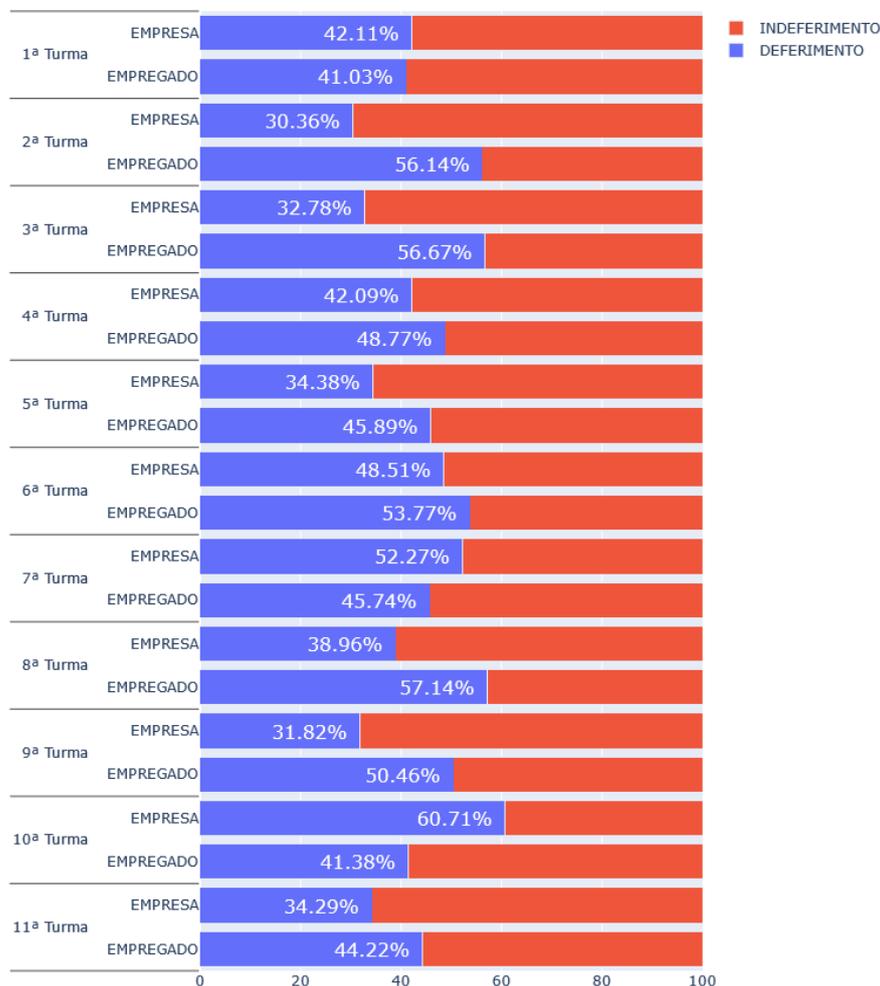
Em relação ao nível de efeito estatístico em função da tabela Cohen's, foi obtido o índice de 22,11%, ou seja, o nível de efeito estatístico da diferença de proporções é pequeno. Quanto à força do teste estatístico, foi obtido o 100%, ou seja, foi alcançado nível aceitável de força estatística e há risco baixo de não haver diferença entre as proporções.

6.5.2 TRT da 4ª Região - avaliação das Turmas Recursais

Em relação a dados individuais de cada uma das Turmas Recursais do TRT da 4ª Região, Figura 6.15, é possível verificar que as Turmas Recursais 2ª, 3ª, 5ª, 8ª, 9ª e 11ª apresentaram proporção parecida com a do tribunal para empregados. Por outro lado, a Figura 6.16 apresenta a distribuição das proporções de deferimentos das turmas recursais. Além disso, na Tabela 6.19 são apresentados os dados em relação aos testes estatísticos aplicados, como o *P-Value*, *Power* e *Effect*. Assim, é possível verificar que há diversas turmas recursais que o *Power* está abaixo de 0,8, o que ocorre devido à pouca quantidade de dados para análise, entretanto é possível superar essa limitação pela ingestão de mais observações em trabalhos futuros.

Figura 6.15: Porcentagem de deferimento de recursos em relação ao recorrente ser empresa e empregado das Turmas Recursais do Tribunal Regional do Trabalho da 4ª Região.

TRT da 4ª Região. Porcentagem de julgamentos favoráveis para cada uma das partes em relação a cada Turma Recursal.



6.5.3 TRT da 3ª Região - avaliação geral

Na Figura 6.17 é possível observar a porcentagem da quantidade de recursos deferidos e indeferidos no TRT da 3ª Região em relação ao recorrente ser empresa ou empregado. Assim, é possível verificar que 41,31% dos recursos impetrados pelos empregados foram deferidos ou parcialmente deferidos. Além disso, também é possível verificar que, dos recursos impetrados pelas empresas, 39,07% foram julgados deferidos ou parcialmente deferidos. Fica evidente a diferença de 2,24% na quantidade de recursos deferidos ou parcialmente deferidos quando o recorrente são os empregados. Importante frisar que essa média leva em consideração amostra dos acórdãos proferidos por todas as Turmas Recursais do tribunal em conjunto.

Nessa amostra foram processados 10.477 acórdãos, sendo 7.349 recursos da parte

Figura 6.16: Distribuição da porcentagem de deferimento de recursos em relação ao recorrente ser empresa e empregado das Turmas Recursais do Tribunal Regional do Trabalho da 4ª Região.

Distribuição do percentual de DEFERIMENTOS de acórdãos pelas 11 TURMAS RECURSAIS do TRT da 4ª Região

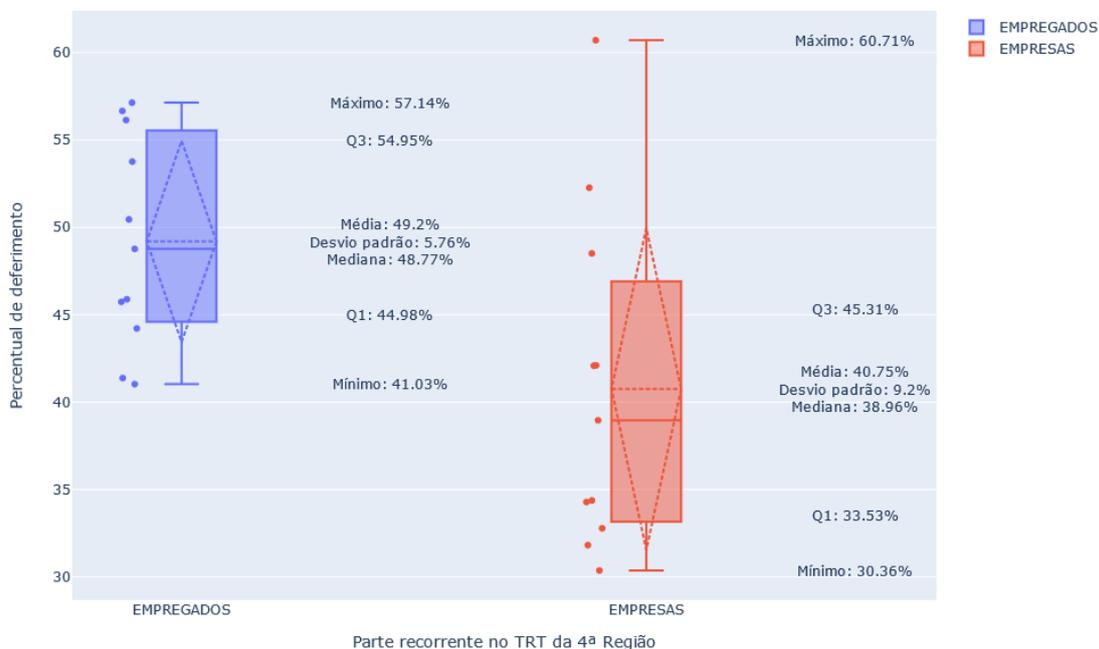
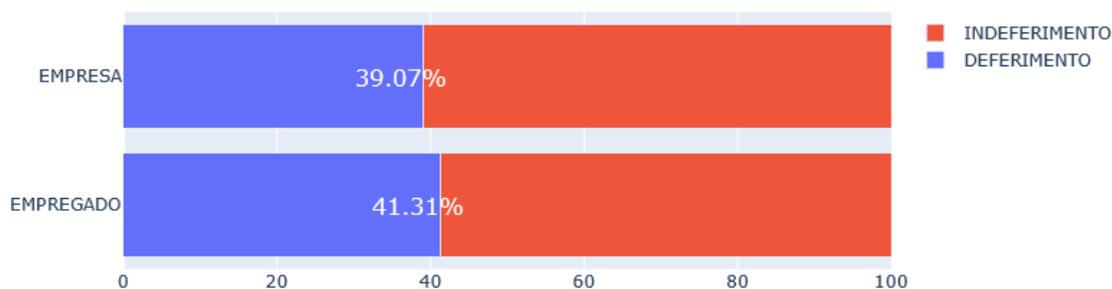


Figura 6.17: Porcentagem de deferimento de recursos em relação ao recorrente ser empresa e empregado no Tribunal Regional do Trabalho da 3ª Região.

TRT da 3ª Região. Porcentagem de julgamentos favoráveis para cada uma das partes.



empregado e 3.128 recursos da parte empresa. Nessa amostra a parte empresa recorreu aproximadamente 2 vezes menos do que a parte empregado. Além disso, em relação aos testes estatísticos aplicados, foi obtido *P-value* de 3,19%, ou seja, há grande confiança de que existe diferença estatística entre a proporção de deferimentos para empregado e para empresa. Em relação ao nível de efeito estatístico em função da tabela Cohen's, foi obtido o índice de 4,58%, ou seja, o nível de efeito estatístico da diferença de proporções é insignificante. Quanto à força do teste estatístico, foi obtido o 79,28%, ou seja, foi alcançado nível abaixo do aceitável de força estatística e há risco de não haver diferença

Tabela 6.20: Porcentagem de deferimento de recursos em relação ao recorrente ser empresa e empregado das Turmas Recursais do Tribunal Regional do Trabalho da 3ª Região.

TRIBUNAL	TURMA RECURSAL	TOTAL DE PROCESSOS	PVALUE	PVALUE INTERPRETACAO	POWER	POWER INTERPRETAÇÃO	EFFECT	EFFECT INTERPRETACAO
TRT3	1ª Turma	1058	0	HÁ DIFERENÇA	1	ACEITÁVEL	0,3749	PEQUENA
TRT3	2ª Turma	906	0,4009	NÃO HÁ DIFERENÇA	0,1789	INACEITÁVEL	0,0591	INSIGNIFICANTE
TRT3	3ª Turma	938	0,1047	NÃO HÁ DIFERENÇA	0,5978	INACEITÁVEL	0,1193	INSIGNIFICANTE
TRT3	4ª Turma	931	0,0014	HÁ DIFERENÇA	0,9803	ACEITÁVEL	0,2245	PEQUENA
TRT3	5ª Turma	889	0,3495	NÃO HÁ DIFERENÇA	0,2372	INACEITÁVEL	0,0692	INSIGNIFICANTE
TRT3	6ª Turma	973	0,0116	HÁ DIFERENÇA	0,9142	ACEITÁVEL	0,1788	INSIGNIFICANTE
TRT3	7ª Turma	1025	0	HÁ DIFERENÇA	1	ACEITÁVEL	0,4839	PEQUENA
TRT3	8ª Turma	850	0,0762	NÃO HÁ DIFERENÇA	0,6205	INACEITÁVEL	0,1321	INSIGNIFICANTE
TRT3	9ª Turma	1016	0	HÁ DIFERENÇA	1	ACEITÁVEL	0,5803	MÉDIA
TRT3	10ª Turma	957	0,0394	HÁ DIFERENÇA	0,747	INACEITÁVEL	0,1443	INSIGNIFICANTE
TRT3	11ª Turma	934	0	HÁ DIFERENÇA	1	ACEITÁVEL	0,3518	PEQUENA

entre as proporções. A saber, o nível de força estatística considerado aceitável é de 80%.

6.5.4 TRT da 3ª Região - avaliação das Turmas Recursais

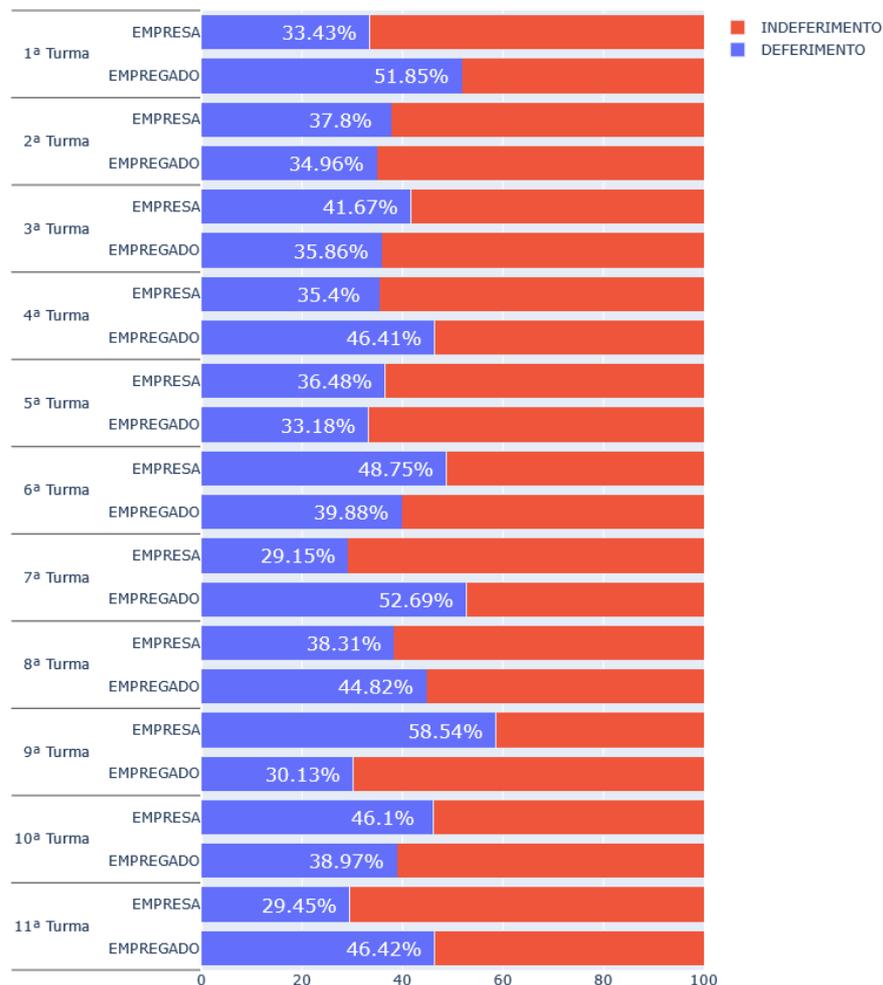
Em relação a dados individuais de cada uma das Turmas Recursais do TRT da 3ª Região, Figura 6.18, é possível verificar que as Turmas Recursais 1ª, 4ª, 7ª, 8ª e 11ª apresentaram proporção de deferimentos maior para empregados. Por outro lado, 6 das 11 Turmas Recursais apresentaram proporção de deferimentos maior para empresas. A Figura 6.19 apresenta a distribuição das proporções de deferimentos das turmas recursais. Além disso, na Tabela 6.20 são apresentados os dados em relação aos testes estatísticos aplicados, como o *P-Value*, *Power* e *Effect*. Assim, é possível verificar que há diversas turmas recursais que o *Power* está abaixo de 0,8, o que ocorre devido à pouca quantidade de dados para análise, entretanto é possível superar essa limitação pela ingestão de mais observações em trabalhos futuros.

6.6 Aplicação

A presente pesquisa não necessita da realização de *deployment* dos modelos desenvolvidos em nenhuma infraestrutura de processamento de dados para serem utilizados. Entretanto, é importante ressaltar algumas considerações em relação a futuras utilizações dos modelos desenvolvidos. Assim, ambos os modelos desenvolvidos assumem que os documentos estejam redigidos em língua portuguesa formal, ou seja, sem erros comuns de digitação. Além disso, seria interessante o monitoramento dos textos das decisões de modo à manutenção da qualidade das previsões, pois os níveis de acurácia dos modelos desenvolvidos dependem necessariamente de que os dados ingeridos estejam em níveis

Figura 6.18: Porcentagem de deferimento de recursos em relação ao recorrente ser empresa e empregado das Turmas Recursais do Tribunal Regional do Trabalho da 3ª Região.

TRT da 3ª Região. Porcentagem de julgamentos favoráveis para cada uma das partes em relação a cada Turma Recursal.

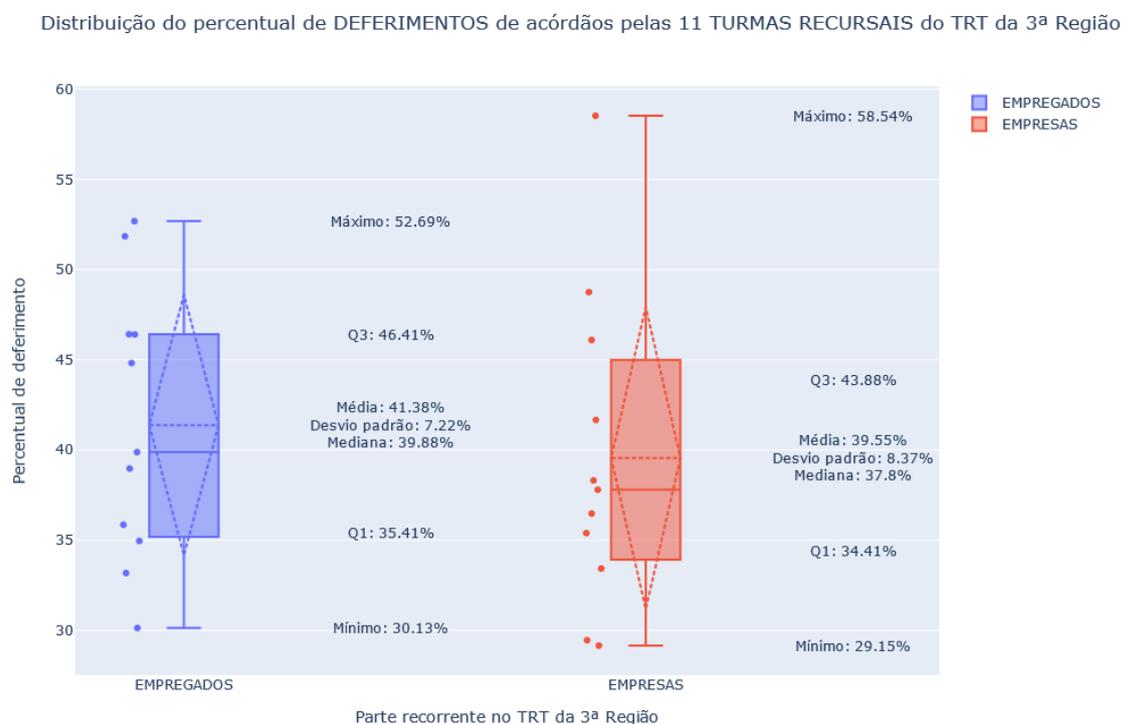


de qualidade previstos.

6.7 Limitações

Os experimentos da presente pesquisa apresentaram certas limitações, como por exemplo, a utilização do formato de redação dos documentos processados. Assim, o algoritmo de aprendizado de máquina desenvolvido levou em consideração padrões de formatação e redação específicos ao corpus jurídico dos Tribunais avaliados. Além disso, não é possível afirmar com certeza que outros tribunais sigam os mesmos padrões, pois não foi realizado experimento de análise de documentos dos demais tribunais brasileiros. Entretanto, essas informações contextuais dos documentos podem ser consideradas úteis

Figura 6.19: Distribuição da porcentagem de deferimento de recursos em relação ao recorrente ser empresa e empregado das Turmas Recursais do Tribunal Regional do Trabalho da 3ª Região.



para os algoritmos, pois elas mimicom a forma como os usuários, advogados e praticantes da atividade jurídica, leem os documentos.

Entretanto, não era objetivo da pesquisa, em um primeiro momento, o desenvolvimento de um algoritmo que processasse documentos de diversos formatos de redação de diversos tribunais diferentes. Desse modo, a padronização de redação dos documentos foi explorada a fim de se alcançar maior otimização do algoritmo. Por outro lado, mesmo em áreas devidamente estabelecidas, como a mineração de sentimentos, é possível encontrar algoritmos desenvolvidos especificamente para certos nichos, como é o caso dos micro-blogs, como o Twitter, que implicam desafios particulares devido a especificidades dos documentos encontrados (LIPPI; TORRONI, 2016).

Além disso, foi realizado o processamento de acórdãos judiciais em que houve apenas uma das partes recorrente. Desse modo, todos os documentos em que havia a apreciação de mais de um recurso foram removidos da base de dados. Entretanto, foi considerado que tal filtragem não apresentou prejuízos aos experimentos, visto que havia grande quantidade de documentos disponíveis para análise de acordo com técnicas de definição de amostra estatística, conforme Seção 6.2.1.2.

Outro fator que pode ser considerado uma limitação da pesquisa é o tamanho da

base de dados padrão-ouro desenvolvido na Seção 6.3.3. Entretanto, essa questão foi mitigada por meio da aplicação de técnica de Supervisão Fraca a qual permitiu explorar a grande quantidade de documentos disponíveis para o treinamento dos modelos desenvolvidos na pesquisa, conforme Seção 6.3.5. Cabe salientar também as limitações impostas pelas próprias técnicas de Aprendizado de Máquinas utilizadas na pesquisa, como, por exemplo, o *viés* ou a tendência inserida pelo algoritmo, que, para desenvolver um modelo, precisa realizar suposições e generalizações.

6.8 Resumo do Capítulo

Neste capítulo é apresentada a validação experimental completa da pesquisa de acordo com as metodologias propostas. Assim, iniciou-se pela elucidação dos elementos fundamentais do estudo que nortearam todas as fases seguintes por meio da Seção Compreensão do negócio. Nessa seção foram definidos os objetivos principais da validação experimental no contexto de negócio e de mineração de dados.

Após, na Seção Compreensão dos dados, iniciou-se a execução prática da pesquisa por meio da coleta dos dados da internet e da exploração inicial dos documentos obtidos. Essa fase trouxe resultados positivos para a pesquisa pois permitiu aferir a qualidade dos dados como também permitiu analisar o conteúdo dos documentos. Essa análise do conteúdo proporcionou inspiração positiva para o desenvolvimento de uma técnica para reduzir o tamanho dos documentos, a qual foi implementada na fase seguinte da pesquisa.

Já na fase da Seção Preparação dos dados houve a impressão de grande esforço para a preparação da melhor maneira possível de toda base de documentos para ser utilizada na fase de modelagem de dados. Assim, primeiramente foi realizada a devida limpeza de caracteres e documentos que não atingiam a qualidade mínima para a pesquisa. Após, foram aplicadas as transformações nos documentos idealizadas na fase anterior da pesquisa. Assim, foram extraídos apenas os dispositivos das decisões judiciais. É nos dispositivos do documento que contém as *features* necessárias para que os algoritmos de modelagem possam classificar corretamente as instâncias.

Ainda na fase de preparação dos dados, na Seção 6.3.3, houve a criação da base de documentos padrão-ouro, a qual foi desenvolvida não apenas com esforço do autor, como também de duas pessoas bacharéis em Direito. Dessa maneira, a base foi utilizada para a realização de testes dos modelos de Aprendizado de Máquina Supervisionado desenvolvidos, como também para análise exploratório dos documentos, mas neste caso, fazendo

a análise cruzada com os rótulos das instâncias. Essa análise permitiu-se inferir *features* que viabilizaram a criação de Funções de Rotulação utilizando o *Snorkel Framework* para aplicação de técnica de Supervisão Fraca, para criar uma base de documentos para treinamento muito maior do que a base padrão-ouro.

Na sequência, a Seção Modelagem apresenta a execução de quatro experimentos de modelagem de dados por meio de algoritmos de Aprendizado de Máquina Supervisionado. Assim, primeiramente foram avaliados diversos algoritmos para o desenvolvimento de um modelo para realizar a classificação dos documentos em relação ao tipo de parte requerente no processo, sendo ela empregado ou empresa. Nesse caso, o algoritmo que apresentou a melhor performance e foi escolhido foi o Support Vector Machine (SVM). Em relação aos outros três experimentos realizados nessa seção, em ambos foram avaliados diversos algoritmos para o desenvolvimento de um modelo para realizar a classificação dos documentos em relação ao deferimento ou não do recurso impetrado. Entretanto, o que diferencia é a base de documentos utilizada. Desse modo, no segundo experimento os algoritmos foram treinados utilizando a base de dados padrão-ouro a qual continha 1.000 instâncias para treinamento, mas as suas classes estavam muito desbalanceadas. No terceiro experimento foi avaliado o treinamento dos mesmos algoritmos utilizando a base de documentos criada automaticamente contendo 1.644 instâncias, mas nesse caso totalmente balanceada. Por último, foi realizado experimento de treinamento utilizando a base de documentos criada automaticamente contendo a totalidade das instâncias anotadas de 22.471, mas nesse caso muito desbalanceada para uma das classes. Enfim, foi realizada a avaliação técnica dos modelos desenvolvidos e selecionado o modelo criado com o algoritmo Gradient Boosting para a classificação efetiva dos documentos.

Por fim, na Seção Avaliação de resultados, foi realizada a classificação de 22.946 documentos com os modelos desenvolvidos, foram aplicados os testes estatísticos de acordo com a metodologia escolhida e por fim foram gerados diversos gráficos demonstrando as proporções de julgamentos encontrados na amostra selecionada. Enfim, foi possível identificar diferença estatística em ambos os tribunais com proporção maior de julgamentos para empregados, por outro lado, também foi observado que há turmas recursais que divergem da média geral de cada tribunal apresentando, assim, proporção maior de julgamentos para empresas. Também é possível verificar na Seção 6.7 o conjunto de limitações encontradas para a presente pesquisa de maneira detalhada.

7 DISCUSSÃO DE RESULTADOS

A presente pesquisa desenvolveu processo de Ciência de Dados para a classificação automática de decisões judiciais em relação ao beneficiário do julgamento ser empregado ou empresa de modo a quantificar a proporção de julgamentos favorável a cada parte. Assim, foi possível responder a questão de pesquisa: **Seria possível que os tribunais avaliados e suas turmas recursais julguem favoravelmente proporção significativamente maior de recursos para uma das partes do que para outra em média?** Enfim, foi possível verificar diferença de proporção estatística significativa em relação ao Tribunal Regional do Trabalho da 3 e 4ª Região. Cabe ressaltar que a diferença de proporção encontrada nos dados do TRT da 4ª Região foi de 12% a mais para empregados e, nos dados do TRT da 3ª Região, a diferença foi de 2% a mais para empregados.

Por outro lado, os dados sugerem haver também diferenças estatísticas de proporção se forem analisadas individualmente as Turmas Recursais que compõem cada tribunal. Em relação ao TRT da 4ª Região, é possível verificar que 6 das 11 Turmas Recursais acompanham a média geral do tribunal tendo maior proporção de julgados favoráveis a empregados, sendo que a 2ª, 3ª e 8ª Turmas Recursais apresentam mais de 2 desvios padrão de diferença. Por outro lado, em relação ao TRT da 3ª Região, também é possível verificar diferenças estatísticas de proporção. Entretanto, neste caso, menos da metade das Turmas Recursais acompanharam a média geral do tribunal. Assim, 5 das 11 Turmas Recursais apresentaram proporção de deferimentos maior para empregados, tendo 3 delas diferença maior que 2 desvios-padrão. Em contrapartida, 6 das 11 Turmas Recursais apresentaram proporção de deferimentos maior para empresas, tendo apenas 1 delas diferença maior que 3 desvios-padrão.

Enfim, as médias gerais dos tribunais analisados e também as médias específicas de diversas Turmas Recursais contribuíram para suportar a hipótese levantada de que há órgãos judiciais que julgam proporção consideravelmente maior para uma das partes do que para outra. Desse modo, a quantidade de dados extraídos em relação aos tribunais como um todo se mostrou satisfatória para alcançar o nível de força desejado de 80%. Entretanto, em relação às Turmas Recursais individualmente, menos da metade das análises obtiveram nível de força abaixo de 80%, como é possível ser observado pela Tabela 6.19 e 6.20, constando, assim, como uma limitação da presente pesquisa.

Além da análise quantitativa das decisões judiciais processadas automaticamente, também foi realizado experimento para avaliar a efetividade de técnica de Supervisão

Fracamente de modo a aumentar a quantidade de documentos anotados utilizados no treinamento dos modelos de Aprendizado de Máquina. Assim, a técnica se mostrou satisfatória, pois permitiu a anotação automática de mais de 22 mil documentos, ou seja, aumentando a base de treinamento em mais de 22 vezes. O presente resultado foi alcançado considerando que os magistrados empregam no dia-a-dia estilo de escrita padronizado e preferencialmente as mesmas palavras-chave em certos trechos dos documentos, ou, até mesmo, reutilizam modelos de documentos alterando apenas nomes de pessoas sendo processadas por exemplo. Por um lado, essa característica específica da base de documentos processada pode ser considerada positiva, pois permitiu a construção rápida de funções de rotulação as quais possibilitaram a anotação automática da base de documentos para treinamento. Entretanto, por outro lado, essa característica de padronização dos documentos pode imprimir desafios, pois muitas das instâncias anotadas automaticamente não apresentam grande variação de *features* o que pode impactar determinados algoritmos.

8 CONCLUSÕES

Esta pesquisa teve por objetivo principal utilizar métodos computacionais automáticos para responder a questão de pesquisa se **Seria possível que os tribunais avaliados e suas turmas recursais julguem favoravelmente proporção significativamente maior de recursos para uma das partes do que para outra em média?** Assim, foi executada pesquisa de descoberta de conhecimento em base de dados de acordo com a metodologia Crisp-dm (CHAPMAN et al., 2000) que incluiu o processamento por meio de Algoritmo de Aprendizado de Máquina e PLN de mais de 20 mil decisões judiciais de modo a realizar estudo quantitativo e estatístico da base de decisões. Os resultados indicaram que realmente há tribunais e turmas recursais que apresentam diferença significativa de proporção de julgamentos favorável para empregados e empresas e vice-versa.

Além disso, também foram atingidos os objetivos específicos que constituíam na análise exploratória dos documentos coletadas na Web, o que resultou em conhecimento sobre a frequência das palavras utilizadas pelos magistrados, como também no conhecimento sobre a padronização dos documentos. Igualmente foi desenvolvida base padrão-ouro para treinamento de algoritmos de Aprendizado de Máquina o que pode ser considerado um dos produtos da pesquisa.

Por outro lado, foi realizada também a experimentação da utilização de técnica de Supervisão Fraca para o desenvolvimento de base de treinamento criada automaticamente por meio da expansão dos dados, a qual apresentou resultados satisfatórios. Além disso, como trabalho futuro poderia ser considerada a elaboração de funções de rotulação que sejam mais flexíveis as palavras-chave utilizadas para classificar os documentos. Nesse caso, poderiam ser utilizadas *Embeddings* de modo a localizar expressões semelhantes às utilizadas na presente pesquisa para classificar documentos. Entretanto, considera-se satisfatório o experimento pois preencheu uma lacuna dos trabalhos relacionados pois não havia sido encontrado nenhum experimento utilizando técnica de Supervisão Fraca aplicada a documentos jurídicos até o momento.

De modo a permitir uma análise mais aprofundada e confiável dos resultados, propõe-se como sugestão que estudos futuros obtenham mais dados na fase de extração da amostra de documentos para considerar também a proporção de julgamento em relação a cada turma recursal, como também cada magistrado na função de relator. Essa consideração poderia permitir análises segmentadas por magistrado na função de relator de modo a viabilizar a análise da seguinte hipótese: **Relatores diferentes (dentro de uma mesma**

turma recursal) influenciam a proporção de julgamentos favoráveis a cada uma das partes, ou seja, há correlação entre o relator e a proporção de julgamentos de uma turma recursal.

Enfim, a presente pesquisa buscou trazer novos dados a discussões presentes na sociedade sobre a questão do viés na Justiça, principalmente, na Justiça do Trabalho. Assim, considerando que a pesquisa de Salama, Carlotti e Yeung (2018) encontrou resultados que sugerem que, a nível de primeiro grau de jurisdição na Justiça do Trabalho, há uma proporção muito maior de julgamentos favoráveis a empregados, os dados da presente pesquisa sugerem a confirmação dessa proporção a nível de segundo grau, apesar de haver algumas turmas recursais que realizam a correção dessa distorção. A fim de facilitar a reprodução da pesquisa, foi disponibilizado online na plataforma Github o código-fonte do projeto¹.

¹<https://github.com/rhuanbarros/jurimetria_justica_do_trabalho>

REFERÊNCIAS

- AGGARWAL, C. C. **Data mining: the textbook**. [S.l.]: Springer, 2015.
- Alexander Wissner-Gross. What Do You Consider the Most Interesting Recent [Scientific] News? What Makes It Important? **Edge**, p. 1–8, 2016. Available from Internet: <<http://www.edge.org/response-detail/26587>>.
- BOEHM, L. Supporting the legal reasoning process by classification of judgments applying active machine learning. 2018.
- BORDEN, B. B.; BARON, J. R. Finding the signal in the noise: Information governance, analytics, and the future of legal practice. **Richmond Journal of Law & Technology**, v. 20, n. 2, p. 7, 2014.
- BRAMER, M. **Principles of data mining**. [S.l.]: Springer, 2007.
- BRASIL. **Código de Processo Civil**. 2015. Available from Internet: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/113105.htm>.
- BRASIL, S. F. do. Constituição da república federativa do brasil. **Brasília: Senado Federal, Centro Gráfico**, 1988.
- BRINGER, E. et al. Osprey: Weak supervision of imbalanced extraction problems without code. In: **Proceedings of the 3rd International Workshop on Data Management for End-to-End Machine Learning**. [S.l.: s.n.], 2019. p. 1–11.
- CHAPMAN, P. et al. Crisp-dm 1.0: Step-by-step data mining guide. **SPSS inc**, v. 9, p. 13, 2000.
- CNJ. **Justiça em Números 2020**. 2020. Available from Internet: <<https://www.cnj.jus.br/pesquisas-judiciarias/justica-em-numeros/>>.
- DUNNMON, J. A. et al. Cross-modal data programming enables rapid medical machine learning. **Patterns**, Elsevier, v. 1, n. 2, p. 100019, 2020.
- ELNAGGAR, A. et al. Multi-task deep learning for legal document translation, summarization and multi-label classification. In: **Proceedings of the 2018 Artificial Intelligence and Cloud Computing Conference**. [S.l.: s.n.], 2018. p. 9–15.
- ENGELEN, J. E. V.; HOOS, H. H. A survey on semi-supervised learning. **Machine Learning**, Springer, v. 109, n. 2, p. 373–440, 2020.
- FANG, Y. et al. Few-shot learning for chinese legal controversial issues classification. **IEEE Access**, IEEE, v. 8, p. 75022–75034, 2020.
- GABARDO, E.; MORETTINI, F. T. R. Institucionalismo e pesquisa quantitativa como metodologia de análise de decisões judiciais. **REVISTA DA FACULDADE DE DIREITO DA UFMG**, n. 63, p. 151–180, 2013.
- GRUS, J. **Data science from scratch: first principles with python**. [S.l.]: O’Reilly Media, 2019.

GUIMARÃES, J. A. C. Elaboração de ementas jurisprudenciais: elementos teórico-metodológicos. **Série Monografias do CEJ**, v. 9, 2011.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference, and prediction**. [S.l.]: Springer Science & Business Media, 2009.

HORNING, N. Introduction to decision trees and random forests. **Am. Mus. Nat. Hist.**, v. 2, p. 1–27, 2013.

IDE, N.; PUSTEJOVSKY, J. **Handbook of linguistic annotation**. [S.l.]: Springer, 2017.

KAISER, L. et al. One model to learn them all. **arXiv preprint arXiv:1706.05137**, 2017.

KEARNS, M. Thoughts on hypothesis boosting. **Unpublished manuscript**, v. 45, p. 105, 1988.

KESSLER, J. S. Scattertext: a browser-based tool for visualizing how corpora differ. Association for Computational Linguistics, Vancouver, Canada, 2017.

KURZWEIL, R. **How to create a mind: The secret of human thought revealed**. [S.l.]: Penguin, 2013.

LEI, M. et al. Automatically classify chinese judgment documents utilizing machine learning algorithms. In: SPRINGER. **International Conference on Database Systems for Advanced Applications**. [S.l.], 2017. p. 3–17.

LIPPI, M.; TORRONI, P. Argumentation mining: State of the art and emerging trends. **ACM Transactions on Internet Technology (TOIT)**, ACM New York, NY, USA, v. 16, n. 2, p. 1–25, 2016.

LOUPPE, G. Understanding random forests: From theory to practice. **arXiv preprint arXiv:1407.7502**, 2014.

MACIEL, J. A. A insegurança jurídica e as súmulas do tribunal superior do trabalho. **Direito UNIFACS–Debate Virtual**, n. 156, 2013.

MONTENEGRO, M. C. **PJe atinge a marca de 7,4 mi de processos judiciais**. 2016. Available from Internet: <<http://www.cnj.jus.br/noticias/cnj/81864-pje-atinge-a-marca-de-7-4-mi-de-processos-judiciais>>.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. **the Journal of machine Learning research**, JMLR. org, v. 12, p. 2825–2830, 2011.

PUSTEJOVSKY, J.; STUBBS, A. **Natural Language Annotation for Machine Learning: A guide to corpus-building for applications**. [S.l.]: "O'Reilly Media, Inc.", 2012.

RATNER, A. Accelerating Machine Learning With Training Data Management. PHD Thesis. n. August, 2019.

RATNER, A. et al. Weak supervision: the new programming paradigm for machine learning. **Hazy Research**. Available via <https://dawn.cs.stanford.edu/2017/07/16/weak-supervision/>. Accessed, p. 05–09, 2019.

RATNER, A. et al. Snorkel: Rapid training data creation with weak supervision. In: NIH PUBLIC ACCESS. **Proceedings of the VLDB Endowment. International Conference on Very Large Data Bases.** [S.l.], 2017. v. 11, n. 3, p. 269.

RATNER, A. et al. Training complex models with multi-task weak supervision. In: **Proceedings of the AAAI Conference on Artificial Intelligence.** [S.l.: s.n.], 2019. v. 33, n. 01, p. 4763–4771.

RAY, S. A quick review of machine learning algorithms. In: IEEE. **2019 International conference on machine learning, big data, cloud and parallel computing (COMIT-Con).** [S.l.], 2019. p. 35–39.

RUSSELL, S.; NORVIG, P. Artificial intelligence: a modern approach. 2002.

SALAMA, B.; CARLOTTI, D.; YEUNG, L. **As decisões da Justiça Trabalhista são imprevisíveis?** [S.l.]: Insper, 2018.

SALAMA, B.; CARLOTTI, D.; YEUNG, L. Quando litigar vale mais a pena do que fazer acordo: os grandes litigantes na justiça trabalhista. **Disponível: <<https://www.insper.edu.br/wp-content/uploads/2019/01/LitigarXFazer-Acordo-Justica-Trabalhista.pdf>>. Acesso em, v. 7, 2019.**

SALAMA, B. M. et al. Dano moral no brasil. **Série Pensando o Direito**, p. 53, 2011.

SCHÜTZE, H.; MANNING, C. D.; RAGHAVAN, P. **Introduction to information retrieval.** [S.l.]: Cambridge University Press Cambridge, 2008.

SHARF, Z.; RAZZAK, M. The informative vector selection in active learning using divisive analysis. **INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS, SCIENCE & INFORMATION SAI ORGANIZATION LTD 19 BOLLING RD, BRADFORD, WEST . . .**, v. 8, n. 10, p. 67–75, 2017.

SHI, C.; SOURDIN, T.; LI, B. The smart court-a new pathway to justice in china? **International Journal for Court Administration, Forthcoming**, 2021.

SNIJDERS, T. Hypothesis testing: Methodology and limitations. In: **International Encyclopedia of the Social and Behavioral Sciences (vol. 10).** [S.l.]: Elsevier-Pergamon, 2002. p. 7121–7127.

SONG, Y. et al. Employing auto-annotated data for government document classification. In: **Proceedings of the 2019 3rd International Conference on Innovation in Artificial Intelligence.** [S.l.: s.n.], 2019. p. 121–125.

STRICKSON, B.; IGLESIA, B. D. L. Legal judgement prediction for uk courts. In: **Proceedings of the 2020 The 3rd International Conference on Information Science and System.** [S.l.: s.n.], 2020. p. 204–209.

TRT4, S. **Sindicato dos Aeroviários e Gol firmam acordo de R\$ 10,3 milhões em processo que envolve 59 trabalhadores.** 2017. Available from Internet: <<https://www.trt4.jus.br/portais/trt4/modulos/noticias/149267>>.

TRT4, S.-G. J. **Conciliação na Justiça do Trabalho.** 2020. Available from Internet: <<https://www.trt4.jus.br/portais/trt4/conciliacao-trt4>>.

VARMA, P. et al. Learning dependency structures for weak supervision models. In: PMLR. **International Conference on Machine Learning**. [S.l.], 2019. p. 6418–6427.

VIEIRA, L. C. A discricionariedade do juiz e o principio da integridade proposto por ronald dworkin. **Legis Augustus**, v. 6, n. 2, p. 102–118, 2015.

WANG, A.; HOANG, C. D. V.; KAN, M.-Y. Perspectives on crowdsourcing annotations for natural language processing. **Language resources and evaluation**, Springer, v. 47, n. 1, p. 9–31, 2013.

WANG, Y. et al. Generalizing from a few examples: A survey on few-shot learning. **ACM Computing Surveys (CSUR)**, ACM New York, NY, USA, v. 53, n. 3, p. 1–34, 2020.

WAZLAWICK, R. **Metodologia de pesquisa para ciência da computação**. [S.l.]: Elsevier Brasil, 2017.

APÊNDICE A — APÊNDICE**A.1 Exemplo de acórdão do Tribunal Regional do Trabalho da 4ª Região**

Acórdão: 0021264-87.2017.5.04.0019 (ROT)

Redator: BEATRIZ RENCK

Órgão julgador: 6ª Turma

Data: 26/09/2019

PODER JUDICIÁRIO

JUSTIÇA DO TRABALHO

TRIBUNAL REGIONAL DO TRABALHO DA 4ª REGIÃO

Identificação

PROCESSO nº 0021264-87.2017.5.04.0019 (RO)

RECORRENTE: (Ocultado devido a questões de sigilo)

RECORRIDO: (Ocultado devido a questões de sigilo)

RELATOR: FERNANDO LUIZ DE MOURA CASSAL

EMENTA

FGTS. PRESCRIÇÃO. PARCELAS RECONHECIDAS EM DEMANDA PRETÉRITA. A prescrição incidente sobre diferenças salariais reconhecidas em demanda pretérita transitada em julgado é a trintenária, uma vez que a hipótese se assemelha àquela própria dos recolhimentos relativos aos valores pagos no curso do contrato de trabalho.

ACÓRDÃO

Vistos, relatados e discutidos os autos.

ACORDAM os Magistrados integrantes da 6ª Turma do Tribunal Regional do Trabalho da 4ª Região: por maioria, vencido o Exmo. Relator, DAR PROVIMENTO AO RECURSO ORDINÁRIO DO AUTOR, para condenar a reclamada a efetuar os depósitos do FGTS incidentes sobre as parcelas deferidas nos autos do processo nº 01674.004/93-1, em valores a serem apurados em liquidação de sentença, acrescidos de juros e correção monetária, na forma da lei. Valor da condenação provisoriamente arbitrado em R10.000,00, *comcustasdeR* 200,00, para os fins legais.

Intime-se.

Porto Alegre, 25 de setembro de 2019 (quarta-feira).

RELATÓRIO

O reclamante interpõe recurso ordinário (ID. 11322c1), inconformado com a sentença (ID. 62cea61) mediante a qual a ação foi extinta com resolução do mérito.

Inicialmente o autor requer seja afastada a prescrição total pronunciada. No mais reitera o pedido recolhimento do FGTS incidente sobre as parcelas e diferenças reconhecidas e apuradas nos autos do processo nº 01674.004/93-1, incluindo os expurgos inflacionários incidentes no período, com juros de mora e correção monetária.

Apresentadas contrarrazões (ID. 959d684), os autos são encaminhados a este Tribunal para julgamento.

É o relatório.

FUNDAMENTAÇÃO

RECURSO ORDINÁRIO DO AUTOR.

1. PRESCRIÇÃO. FGTS INCIDENTE SOBRE AS PARCELAS DEFERIDAS NO PROCESSO Nº 01674.004/93-1.

Requer o autor a reforma da sentença que extinguiu o feito, com resolução do mérito, por entender incidente a prescrição quinquenal quanto ao FGTS incidente sobre as parcelas deferidas no processo nº 01674.004/93-1. Aduz que o contrato de trabalho do autor encontra-se em vigor. Pondera que a pretensão de pagamento de FGTS sobre parcelas que compõe o contrato de trabalho em vigor, ainda que tenham sido deferidas judicialmente, possui como marco inicial para a contagem da prescrição total a data da rescisão contratual. Invoca a Súmula 362 do TST.

Ao exame.

A sentença extinguiu o feito, pelos seguintes fundamentos:

É incontroverso o fato de que a reclamada sucedeu a CEEE no contrato de emprego do autor, sendo assim responsável pelas parcelas postuladas pelo autor, inclusive aquelas anteriormente assumidas pela sucedida, nos termos dos artigos 10 e 448 da CLT, conforme exposto em preliminar.

Em relação à prescrição do Fundo de Garantia por Tempo de Serviço, como pedido principal, aplicar-se-ia o disposto na Súmula 362, II, do C. TST, o que prevê que "II - Para os casos em que o prazo prescricional já estava em curso em 13.11.2014, aplica-se o prazo prescricional que se consumir primeiro: trinta anos, contados do termo inicial, ou cinco anos, a partir de 13.11.2014."

Contudo, tal entendimento não se enquadra ao presente caso, uma vez que os valores pleiteados decorrem apenas da incidência do FGTS nas parcelas postuladas no processo nº 01674.004/93-1, tratando-se, portanto, de pedido de caráter acessório, que deve seguir a prescrição quinzenal aplicada ao pedido principal.

Dessa forma, considerando que a ação mencionada foi ajuizada em 16/12/1993, aplica-se à hipótese o entendimento consubstanciado na Súmula 206 do TST:

FGTS. INCIDÊNCIA SOBRE PARCELAS PRESCRITAS. A prescrição da pretensão relativa às parcelas remuneratórias alcança o respectivo recolhimento da contribuição para o FGTS.

Nesse mesmo sentido, transcreve-se recente decisão do TRT da 4ª Região:

Como observo, o pedido de condenação relativo aos depósitos de FGTS se refere a parcelas remuneratórias e consectários salariais, não se tratando de não-recolhimento de contribuição para o FGTS, como orienta a Súmula 362, do TST. Aplica-se, portanto, o entendimento da Súmula 206 do TST, como pretende a defesa. (TRT da 4ª Região, 11ª Turma, 0020903-61.2017.5.04.0022 RO, em 16/11/2018, Desembargador Roger Ballejo Villarinho)

Ante o exposto, julga-se extinto o pedido, com resolução do mérito, nos termos do art. 487, II, do CPC.

Entendo que a sentença não merece reparos.

No caso, é certo que o autor postula o recolhimento do FGTS incidente sobre as verbas deferidas nos autos do processo 01674.004 /93-1, que tramitou perante a 4ª Vara do Trabalho de Porto Alegre, em que obteve o reconhecimento do direito de correção do enquadramento no Quadro de Carreira implantado pela empregadora em 01/07/1991, em razão do desvio de função havido, no cargo de Mecânico de Equipamentos, no

Plano de Cargos de Natureza Operacional, Nível E, na Referência determinada de acordo com o art. 24, §1º, do Regulamento do Quadro, com o pagamento de todas as diferenças daí decorrentes, em parcelas vencidas e vincendas, com juros de mora e correção monetária.

Trata-se, a toda evidência, de pedido acessório, porquanto decorre diretamente do reconhecimento do direito ao pagamento de diferenças salariais por desvio de função, e por esse motivo deve ser observada a prescrição quinquenal, conforme preceitua o entendimento vertido na Súmula 206 do TST.

Logo, correta a sentença ao reconhecer a prescrição.

Nesta esteira, nego provimento ao recurso.

FERNANDO LUIZ DE MOURA CASSAL

Relator

VOTOS

DESEMBARGADORA BEATRIZ RENCK:

RECURSO ORDINÁRIO DO AUTOR.

1. PRESCRIÇÃO. FGTS INCIDENTE SOBRE AS PARCELAS DEFERIDAS NO PROCESSO Nº 01674.004/93-1.

Muito embora o entendimento do Exmo. Relator quanto à prescrição incidente na espécie, diverjo.

Busca o reclamante o depósito dos valores relativos ao FGTS incidente sobre as diferenças salariais por reenquadramento reconhecidas e pagas nos autos do processo nº 01674.004/93-1, cuja decisão transitou em julgado em 04.03.1999.

Diversamente do quanto concebido na origem, entendo incidente na espécie a prescrição trintenária. Isso por que, não se trata de recolhimento de FGTS sobre parcelas salariais sujeitas à prescrição quinquenal, o que seria inviável, uma vez que quando do perecimento do principal não há como se reconhecer o acessório. No caso destes autos as diferenças salariais por reenquadramento já foram reconhecidas e pagas nos autos de demanda pretérita, restando, apenas, o FGTS sobre elas incidente, o que se enquadra na regra prescricional aplicável aos depósitos sonegados no curso do contrato, sobre verbas já alcançadas ao trabalhador.

Neste sentido é a jurisprudência do C. TST, fartamente citada nas razões de recurso ordinário.

Diante desta realidade, dou provimento ao apelo para condenar a reclamada a efetuar os depósitos do FGTS incidentes sobre as parcelas deferidas nos autos do processo nº 01674.004/93-1.

Os valores devem ser apurados em liquidação de sentença, acrescidos de juros e correção monetária, na forma da lei.

Tendo em vista a natureza da parcela, não há incidência de descontos previdenciários e fiscais.

Custas revertidas à reclamada.

DESEMBARGADORA MARIA CRISTINA SCHAAN FERREIRA:

Acompanho a divergência.

PARTICIPARAM DO JULGAMENTO:

DESEMBARGADOR FERNANDO LUIZ DE MOURA CASSAL
(RELATOR)

DESEMBARGADORA BEATRIZ RENCK

DESEMBARGADORA MARIA CRISTINA SCHAAN FERREIRA

A.2 Aplicações práticas da técnica de Supervisão Fraca utilizando Snorkel Framework

A seguir são destacados trabalhos de pesquisa que se concentram na aplicação prática de técnica de Supervisão Fraca utilizando o Snorkel Framework.

A pesquisa de Dunnmon et al. (2020) focou na verificação prática da técnica de Supervisão Fraca na área da saúde. Nesse caso, as bases de dados previamente anotadas manualmente foram desenvolvidas com a utilização de profissionais da saúde especializados no assunto durante meses, o que acentuou ainda mais o custo de desenvolvimento desse ativo. Enfim, esses profissionais analisaram imagens de exames médicos e aplicaram um rótulo indicando o diagnóstico provável.

Diferentemente dos médicos que analisaram as imagens e aplicaram um rótulo manualmente, a tarefa de anotação automática fez uso de recurso não utilizado previamente. Além das imagens, é comum esses exames médicos serem acompanhados de relatórios

de texto livre contendo um breve estudo explicativo. Assim, médicos e engenheiros trabalharam em conjunto para desenvolver Funções de Rotulação de modo a processar esses documentos de texto. Tal tarefa levou aproximadamente 8h de trabalho clínico na qual foram desenvolvidas funções de extração de palavras-chave e processamento de heurísticas baseadas em ontologias médicas.

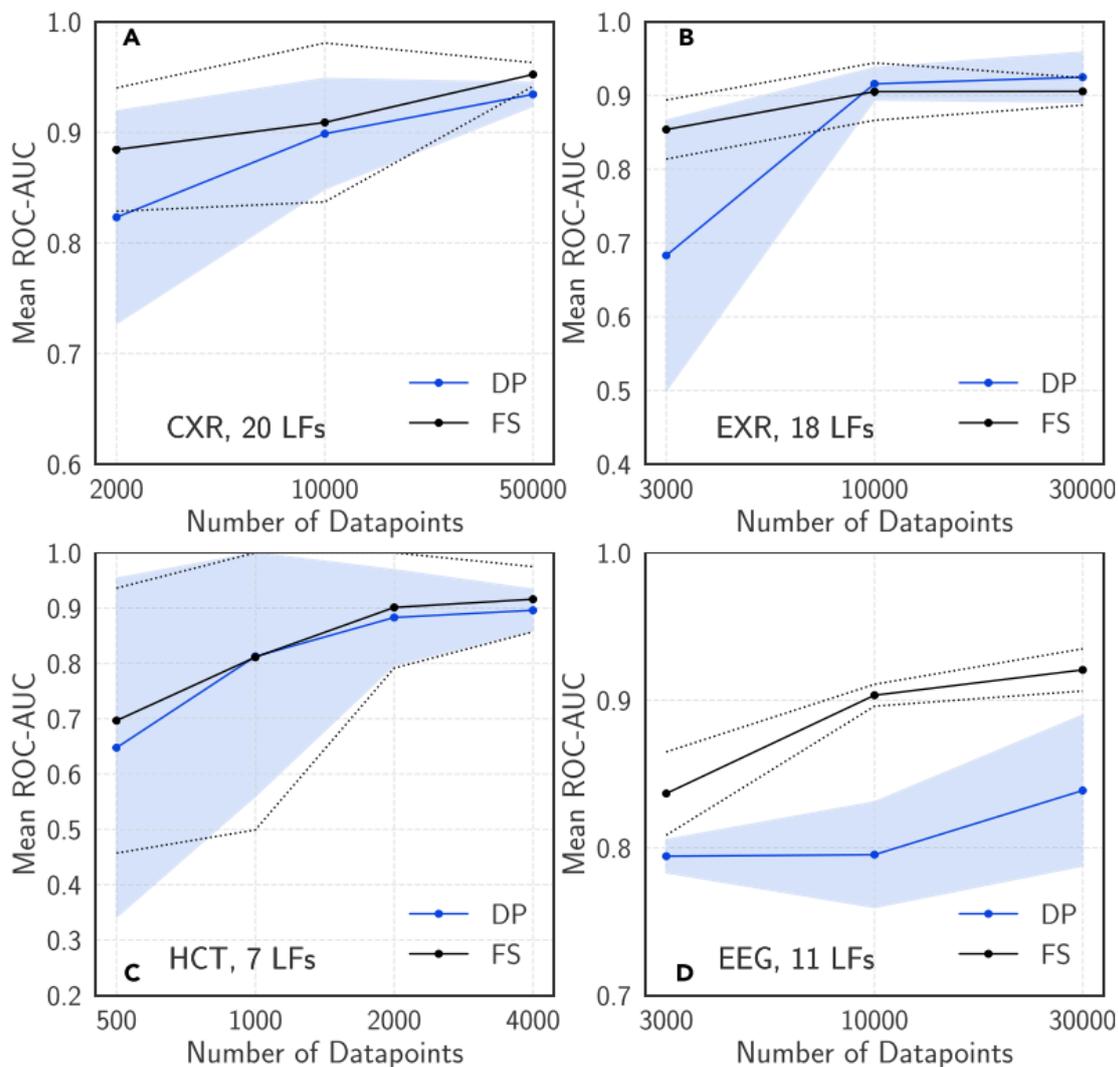
Enfim, ambas as bases de documentos anotadas manualmente e automaticamente foram utilizadas para o desenvolvimento de modelos de Aprendizado de Máquina Supervisionado utilizando redes neurais. A Figura A.1 apresenta pontuação média *ROC – AUC* dos modelos, sendo possível observar pontuações aproximadas em quantidades média de documentos utilizados. Além disso, conforme a quantidade de documentos utilizados aumenta, é possível verificar que a performance dos modelos treinados utilizando Supervisão Fraca tem a tendência de aumentar sem a necessidade de emprego de especialistas médicos na tarefa de anotação manual de documentos.

Também houve a realização de trabalho similar na Intel, é o caso de Bringer et al. (2019), que apresentam um estudo em que são analisados Tweets para o monitoramento de clientes em relação ao lançamento de novos produtos, realização de parcerias e também em relação a fusão e aquisição de outras empresas. Assim, a tarefa já era realizada na empresa por meio de um sistema de regras e heurísticas de modo a extrair todas as relações entre palavras, obtendo 0,85 pontos de precisão em média. Em contrapartida foi possível observar que apenas 0,05% dos dados extraídos do Twitter eram relevantes, o que se traduz em uma tarefa de Aprendizado de Máquina com grande desbalanceamento.

Apesar do alto desbalanceamento de classes, esse efeito foi mitigado em virtude da utilização das regras e heurísticas programadas utilizando a busca por palavras-chave, as quais apresentaram alto nível de precisão. Assim, no sistema desenvolvida na Intel, eram codificadas funções em busca de certas palavras que retornam um peso para a instância indicando a sua relevância, além de incluírem palavras modificadoras que poderiam aumentar ou diminuir o peso final. Essa abordagem apresenta como grande ponto negativo, por exemplo, a necessidade da aplicação manual de pesos as palavras-chave e heurísticas.

Assim, Bringer et al. (2019) realizaram a transformação das regras e heurística já desenvolvidas na empresa em Funções de Rotulação, entretanto, em funções de complexidade reduzida. Portanto, regras múltiplas que aumentavam ou diminuía o peso da instância foram divididas em funções de rotulação que simplesmente assinalavam um rótulo positivo ou negativo. Desse modo, o processamento de toda a base de documentos com essas funções gerou diversas instâncias com rótulos sobrepostos e conflitantes, en-

Figura A.1: O quadro A apresenta radiografia de tórax (*CXR*), o quadro B apresenta radiografia de extremidade (*EXR*), o quadro C, radiografia de cabeça (*HCT*) e o quadro D apresenta eletroencefalografia (*EEG*). *DP* significa *Data-Programming*, ou seja, técnica de *Weak Supervision*. *FS* significa *Full hand-labeled supervision*, ou seja, técnica de Aprendizado Supervisionado com base de dados anotada manualmente. A linha tracejada representa intervalo de confiança para *FS* e a área sombreada para *DP*. Ambas em relação a cinco treinamentos com sementes aleatórias.



Fonte: Dunmon et al. (2020)

tretanto a aplicação do modelo generativo realizou a pesagem de todas essas funções de acordo com os rótulos aplicados, tornando assim desnecessária a fase de pesagem manual que era realizada na técnica utilizada anteriormente na empresa.

Enfim, o estudo expõe que houve ganhos significativos na métrica de precisão em três tarefas de extração de relações aplicadas a projetos reais da empresa. Além disso, são enfatizados os benefícios da técnica em relação a técnica utilizada previamente na empresa, como, por exemplo, o custo muito menor para desenvolvimento do algoritmo.

Figura A.2: Distribuições de unigramas da base de dados do TRT da 3ª Região.

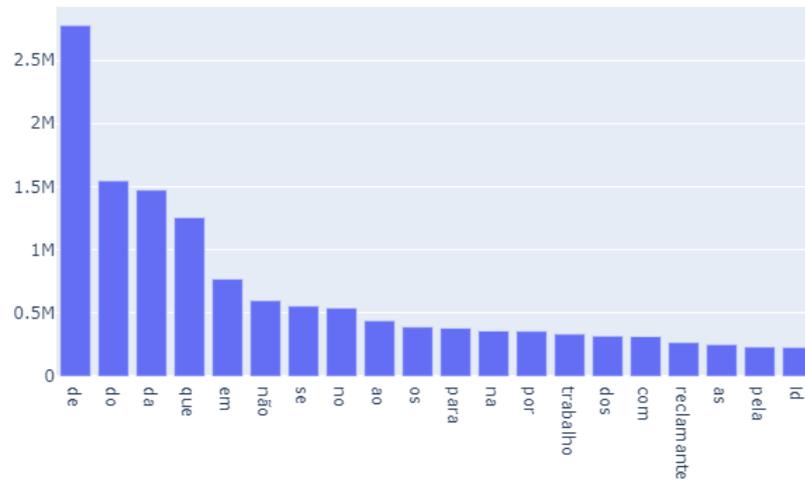
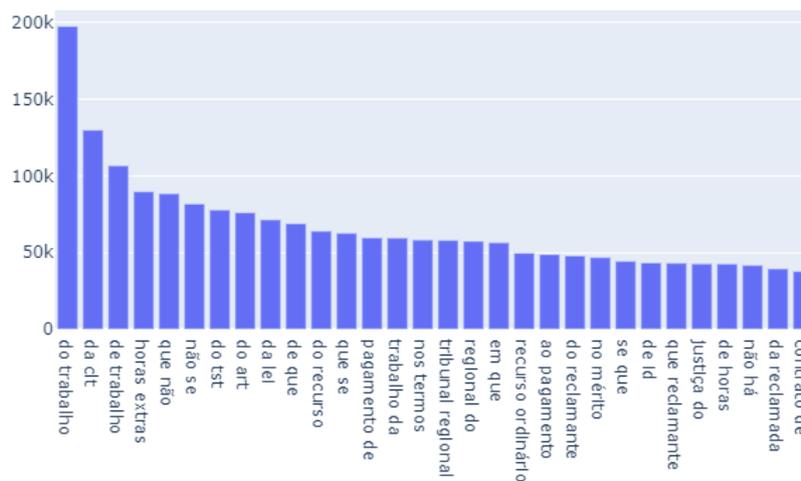


Figura A.3: Distribuições de bigramas da base de dados do TRT da 3ª Região.



Além disso, a facilidade de manutenção do modelo diante de alterações dos objetivos de negócio as quais podem ser rapidamente implementadas com alterações nas funções de rotulação.

A.3 Distribuições dos unigramas, bigramas e trigramas na fase de exploração inicial dos documentos extraídos da internet

Os Gráficos A.2, A.3, A.4, A.5, A.6 e A.7 apresentam as distribuições dos unigramas, bigramas e trigramas na fase de exploração inicial dos documentos extraídos da internet.

Figura A.4: Distribuições de trigramas da base de dados do TRT da 3ª Região.

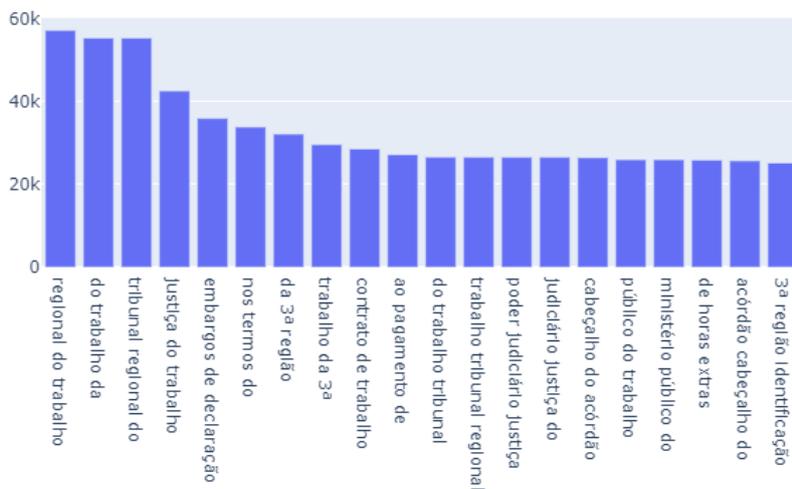


Figura A.5: Distribuições de unigramas da base de dados do TRT da 4ª Região.

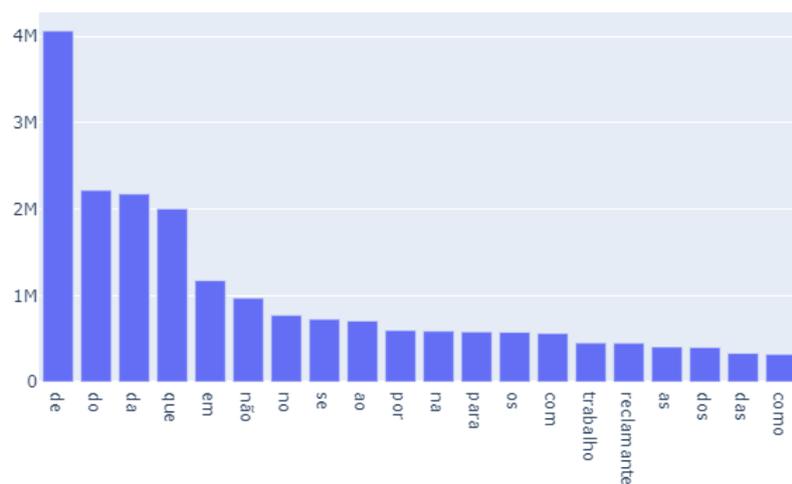


Figura A.6: Distribuições de bigramas da base de dados do TRT da 4ª Região.

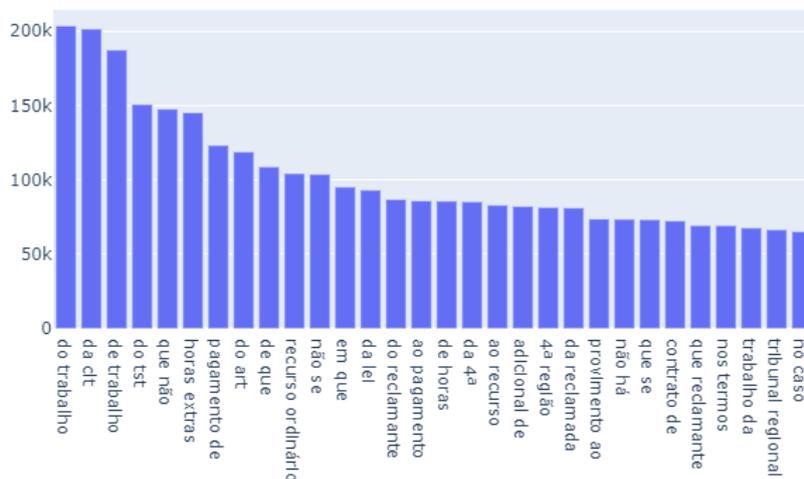
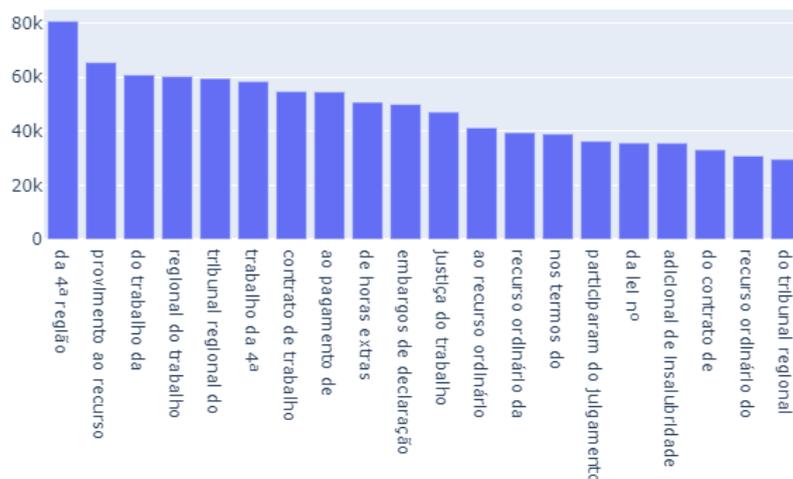


Figura A.7: Distribuições de trigramas da base de dados do TRT da 4ª Região.



A.4 Diretrizes para anotação manual de documentos jurídicos para pesquisa de mestrado de Rhuan Barros

Introdução

A presente pesquisa sendo realizada por Rhuan Barros para defesa de tese de mestrado na UFRGS tem por objetivo a análise jurimétrica de decisões judiciais por meio de algoritmos de Inteligência Artificial. Simplificadamente, o artefato de programação a ser desenvolvido é um algoritmo computacional que permita a “leitura” de uma decisão judicial e a aplicação de um rótulo indicando o teor de deferimento ou não do julgamento.

Desse modo, a técnica de Inteligência Artificial a ser empregada exige o processamento de uma base de dados de decisões judiciais que contenham um rótulo indicando o teor de deferimento ou não do julgamento. Entretanto, tal base de documentos com rótulos aplicados não existe atualmente. Assim, torna-se necessária a criação manual dessa base para a realização da pesquisa.

Desse modo, o trabalho de anotação consiste na leitura da decisão judicial e de interpretação quanto ao deferimento ou não. Entretanto, devido a análise pelo autor da pesquisa, verificou-se que para a interpretação quanto ao deferimento ou não do julgamento é necessário apenas a leitura do dispositivo da sentença. Ou seja, é necessário apenas a interpretação de um parágrafo de aproximadamente 3 linhas.

Portanto, foram selecionadas 500 decisões do Tribunal do Trabalho da 3ª Região e 500 decisões do Tribunal do Trabalho da 4ª Região. Após, foram extraídas automaticamente apenas o dispositivo de cada decisão e inseridas em um aplicativo de planilhas.

Assim, foi criada uma planilha com uma coluna com a lista de decisões judiciais e outra coluna em branco. A coluna em branco serve para ser inserida a informação quanto ao deferimento ou não do julgamento.

Além disso, de modo a aumentar o nível de qualidade da pesquisa, torna-se necessário que cada documento seja anotado por pelo menos duas pessoas. Assim, o autor da pesquisa realizou manualmente a anotação de 1000 decisões judiciais. Nesse ponto entra a sua tarefa como anotador. Após a realização da anotação das decisões por você, elas serão comparadas com as anotações realizadas pelo autor da pesquisa de modo a verificar a quantidade de concordância nos rótulos aplicados. Se houver discordância, o autor da pesquisa entrará em contato com o anotador e será decidido em conjunto qual será o rótulo final aplicado.

Após a compilação dos rótulos de 1000 decisões, o autor irá utilizar a base de dados para desenvolver o algoritmo computacional de Inteligência Artificial. Enfim, é de extrema importância que os documentos sejam rotulados corretamente, pois qualquer erro de rotulação acaba por gerar erros no algoritmo.

Como fazer a anotação manual das decisões judiciais

A anotação das decisões judiciais deve ser realizada em aplicativo de planilha do Google Suite disponibilizado por meio de link para acesso via computador. O preenchimento deve ser realizado online. Há uma coluna com as decisões e outra para inserir o rótulo em cada linha, respectivamente. A Figura A.8 apresenta foto da tela da planilha com algumas decisões rotuladas.

Rótulos

Há 3 possibilidades de rótulos a serem aplicados: DEFERIMENTO, INDEFERIMENTO, SEM ANÁLISE MÉRITO.

- **DEFERIMENTO:** aplicado quando houve análise dos pedidos contidos no recurso e eles foram concedidos total ou parcialmente.
- **INDEFERIMENTO:** aplicado quando houve análise dos pedidos contidos no recurso e eles foram negados totalmente.
- **SEM_ANALISE_MERITO:** aplicado quando não houve análise dos pedidos contidos no recurso. Isso pode ocorrer no caso, por exemplo, de anulação da sentença a que as partes fazem recurso devido a falha grave cometida a nível de primeiro grau ou de declaração de incompetência da Justiça do Trabalho em função de que o objeto da ação deve ser julgado por outro ramo da Justiça.

Casos de exemplo de anotação

Figura A.8: Planilha contendo decisões judiciais com alguns rótulos de exemplo.

INDICE	DISPOSITIVO	ROTULO
993	Cabeçalho do acórdão Acórdão FUNDAMENTOS PELOS QUAIS, A Segunda Turma, do Egrégio Tribunal Regional do Trabalho da Terceira Região, em sessão hoje realizada, à unanimidade, conheceu do agravo de petição; no mérito, sem divergência, negou provimento ao apelo; fixou custas no importe de R\$44,26, pela executada. Presidente: Exmo. Desembargador Sebastião Geraldo de Oliveira. Tomaram parte no julgamento: Exmo. Juiz Mauro César Silva (Relator, convocado para substituir o Exmo. Desembargador Lucas Vanucci Lins, em férias, nos termos do RI, art. 69, § 1º). Exmo. Desembargador Sebastião Geraldo de Oliveira e o Exmo. Desembargador Jales Valadão Cardoso. Procurador do Trabalho: Dr. Eduardo Maia Botelho. Secretária da sessão: Eleonora Leonel da Mata Silva. Belo Horizonte, 06 de agosto de 2019. Assinatura MAURO CÉSAR SILVA, Juiz Relator. c/ass VOTOS	INDEFERIDO
7092	Cabeçalho do acórdão Acórdão FUNDAMENTOS PELOS QUAIS, O Tribunal Regional do Trabalho da Terceira Região, em sessão ordinária da Décima Turma, hoje realizada, julgou o presente processo e, por unanimidade, conheceu os embargos de declaração opostos pelo reclamante, no mérito, sem divergência, deu-lhes provimento apenas para prestar esclarecimentos que passam a fazer parte integrante do julgado, sem, contudo, imprimir-lhe efeito modificativo. Tomaram parte no julgamento as(o) Exmas(o): Desembargadora Taisa Maria Macena de Lima (Relatora), Desembargadora Rosemary de Oliveira Pires (Presidente) e Desembargador Paulo Maurício Ribeiro Pires. Presente ao julgamento a il. representante do Ministério Público do Trabalho: Dra. Junia Castelar Savaget. Belo Horizonte, 10 de julho de 2019. Assinatura TAISSA MARIA MACENA DE LIMA, Relatora. VOTOS	DEFERIDO
335	Cabeçalho do acórdão SUMULA DO VOTO FUNDAMENTOS PELOS QUAIS, o Tribunal Regional do Trabalho da Terceira Região, em sessão ordinária da Nona Turma, hoje realizada, à unanimidade, conheceu do recurso, exceto quanto às horas extras decorrentes da redução da hora noturna, por constituir pretensão não deduzida na inicial, no mérito, por maioria de votos, deu-lhe provimento parcial para condenar a reclamada ao pagamento da PLR proporcional ao ano de 2014, como se apurou, vencida parcialmente a Exma. Desembargadora Maria Laura Franco Lima de Faria quanto à indenização substitutiva do cartão alimentação; inverteu o ônus da condenação, custas pela reclamada, no importe de R\$20,00, calculada sobre R\$1.000,00, valor fixado para a condenação; declarou que a parcela possui natureza salarial; incidem juros e correção monetária, na forma da Legislação aplicável. Tomaram parte no julgamento: Exmas. Desembargadora Mônica Sette Lopes (Presidente e Relatora), Desembargadora Maria Stela Álvares da Silva Campos e Desembargadora Maria Laura Franco Lima de Faria. Procurador do Trabalho: Dr. Sebastião Vieira Caiçeta. Belo Horizonte, 16 de maio de 2017. Assinatura MÔNICA SETTE LOPES, Desembargadora Relatora. MSL/1 VOTOS	DEFERIDO
307	Cabeçalho do acórdão Acórdão O Tribunal Regional do Trabalho da Terceira Região, em sessão ordinária da Primeira Turma, hoje realizada, julgou o presente processo e, preliminarmente, à unanimidade, conheceu dos embargos de declaração opostos pela reclamada no Id 1b54605, porque satisfeitos os pressupostos de admissibilidade, e, no mérito, não conheceu dos embargos de declaração de Id 9c0410a, é que em consonância com o princípio da unirecorribilidade ou unicidade recursal, contra ato decisório é cabível um único recurso; assim, na hipótese de a parte valer-se de dois recursos, o segundo não pode ser conhecido, porque alcançado pelo instituto da preclusão consumativa; assim, o segundo recurso interposto pela reclamada não pode ser conhecido (Id 9c0410a) por preclusão consumativa e afronta ao princípio da unirecorribilidade das decisões; no mérito, sem divergência, negou-lhes provimento. RAZÕES DE DECIDIR: Em síntese, a embargante alega que a decisão foi omissa quanto à análise dos cartões de ponto e distribuição do ônus da prova. Objetiva, ainda, prequestionar a matéria, nos termos da Súmula 297 do TST. Examinou. Nos termos do art. 1022 do NCPC, que reproduz as hipóteses de cabimento dos embargos de declaração, esses são cabíveis quando o julgado apresentar obscuridade ou contradição, ou quando for omitido ponto sobre o qual o juízo devia se pronunciar. Ainda, de acordo com o art. 897-A da CLT, os embargos também são cabíveis ante a presença de manifesto equívoco no exame dos pressupostos extrínsecos do recurso. Nessa esteira, em sede de embargos declaratórios, a contradição a ser suprida é a ausência de solução para uma questão controvertida. A contradição a ser sanada é aquela insita à própria decisão, ou seja, a existente dentro de seus	INDEFERIDO
4945	Cabeçalho do acórdão Acórdão Fundamentos pelos quais, O Tribunal Regional do Trabalho da Terceira Região, em sessão ordinária da sua Oitava Turma, hoje realizada, sob a Presidência do Exmo. Desembargador Márcio Ribeiro do Valle, presente a Exma. Procuradora Maria Amélia Bracks Duarte, representante do Ministério Público do Trabalho e, computados os votos dos Exmos. Desembargador Sérgio da Silva Paçanita e Juiz Conocada Luciana Alves Viotti (substituindo a Desembargadora Ana Maria Amorim Reboças). JULGOU o presente processo e, preliminarmente, à unanimidade, conheceu do recurso ordinário interposto pelo Banco Reclamado, rejeitando a preliminar de não conhecimento do recurso, por suposta ofensa ao princípio da dialeticidade, erigido em contrarrazões; no mérito, sem divergência, negou-lhe provimento. Belo Horizonte, 18 de abril de 2018. Assinatura MÁRCIO RIBEIRO DO VALLE, Desembargador Relator. MRV/1 VOTOS	INDEFERIDO

- Dispositivo: “Cabeçalho do acórdão Acórdão FUNDAMENTOS PELOS QUAIS, A Segunda Turma, do Egrégio Tribunal Regional do Trabalho da Terceira Região, em sessão hoje realizada, à unanimidade, conheceu dos embargos de declaração da reclamada e no mérito, sem divergência, negou-lhes provimento. Presidente: Exmo. Desembargador Jales Valadão Cardoso.”
 - Rótulo: INDEFERIDO
 - Explicação: nesse caso trata-se de um recurso de Embargos de Declaração. É possível observar que o magistrado escreve “...conheceu...”, mas isso não significa que houve deferimento de algum pedido. Após essa parte é possível ler “...negou-lhes provimento...”, o que significa que o pedido foi INDEFERIDO.
- Dispositivo: “Cabeçalho do acórdão Acórdão Fundamentos pelos quais O Tribunal Regional do Trabalho da Terceira Região, em sessão ordinária da sua Sétima Turma, hoje realizada, sob a presidência do Exmo. Desembargador Fernando Luiz Gonçalves Rios Neto, presente o Exmo. Procurador Genderson Silveira Lisboa, representante do Ministério Público do Trabalho, computados os votos da Exma. Des. Cristiana Maria Valadares Fenelon e do Exmo. Juiz Convocado Cléber Lúcio de Almeida (substituindo o Exmo. Des. Paulo Roberto de Castro), JULGOU o presente processo e, unanimemente, conheceu do recurso; no mérito, sem divergência, deu-lhe provimento parcial para estabelecer que as diferenças salariais por equipa-

ração e reflexos reconhecidos na sentença são devidos pelo período contratual não prescrito, ficando mantidos os demais termos e paramentos definidos na sentença. Mantém-se o valor da condenação, porque compatível. Belo Horizonte, 23 de fevereiro de 2017. Assinatura FERNANDO LUIZ GONÇALVES RIOS NETO Relator acp VOTOS”

- Rótulo: DEFERIDO
- Explicação: Novamente é possível observar que o recurso foi conhecido, mas isso não significa que foi deferido ou indeferido, apenas apreciado. Depois, é possível ler “...deu-lhe provimento parcial...” o que indica o DEFERIMENTO.

- Dispositivo: “Cabeçalho do acórdão Acórdão FUNDAMENTOS PELOS QUAIS, A Segunda Turma, do Egrégio Tribunal Regional do Trabalho da Terceira Região, em sessão hoje realizada, à unanimidade, conheceu do recurso da reclamada; no mérito, por maioria de votos, deu-lhe parcial provimento para afastar da condenação a "indenização correspondente a 1/12 do total das comissões auferidas durante o período de vigência do contrato de representação comercial, a teor do art. 27, alínea "j", da Lei n. 4.886/65" e a "indenização no valor correspondente a 1/3 das comissões auferidas nos três meses anteriores ao término do contrato (art. 34 da Lei n. 4.886/65)", vencido o Exmo. Juiz Convocado Relator, que desprovia integralmente o apelo; reduzido o valor da condenação para R\$10.000,00 e o das custas processuais para R\$200,00 encargo da reclamada, que poderá requerer a restituição do valor recolhido a maior. Presidente em exercício: Exmo. Desembargador Jales Valadão Cardoso Tomaram parte no julgamento: Exmo. Juiz Carlos Roberto Barbosa (Relator, convocado para substituir o Exmo. Desembargador Sebastião Geraldo de Oliveira, afastado nos termos da RA n. 25/2019), Exmo. Desembargador Jales Valadão Cardoso e a Exma. Desembargadora Maristela Íris da Silva Malheiros. Procurador do Trabalho: Dr. Eduardo Maia Botelho. Secretária da Sessão: Eleonora Leonel da Mata Silva. Belo Horizonte, 26 de março de 2019. Assinatura CARLOS ROBERTO BARBOSA Juiz Convocado Relator VOTOS”

- Rótulo: DEFERIDO
- Explicação: É possível ler “...no mérito, por maioria de votos, deu-lhe parcial provimento...” o que indica que o recurso foi DEFERIDO. Apesar de na sequência haver a palavra “...afastar...”, essa não indica indeferimento do recurso.

- Dispositivo: “Cabeçalho do acórdão CONCLUSÃO CONCLUSÃO O Tribunal Regional do Trabalho da Terceira Região, em sessão ordinária da sua Oitava Turma, hoje realizada, sob a Presidência da Exma. Desembargadora Ana Maria Amorim Rebouças, presente a Exma. Procuradora Maria Amélia Bracks Duarte, representante do Ministério Público do Trabalho e, computados os votos das Exmas. Desembargadora Ana Maria Amorim Rebouças e Juíza Convocada Ana Maria Espí Cavalcanti (substituindo o Desembargador José Marlon de Freitas); JULGOU o presente processo e, preliminarmente, à unanimidade, conheceu do Recurso Ordinário interposto pela Reclamante (Id. 92ff657), porquanto presentes os pressupostos objetivos e subjetivos de admissibilidade; no mérito, sem divergência, negou provimento ao apelo, adotando como razões de decidir os fundamentos da sentença, conforme autorização contida no artigo 895, §1º, inciso IV, da CLT. Belo Horizonte, 01 de fevereiro de 2017. Assinatura ANTÔNIO CARLOS RODRIGUES FILHO Juiz Convocado Relator VOTOS”
 - Rótulo: INDEFERIDO
 - Explicação: É possível ler “...no mérito, sem divergência, negou provimento ao apelo...” o que indica que o recurso foi INDEFERIDO.
- Dispositivo: “Cabeçalho do acórdão CERTIDÃO CERTIFICO que o Tribunal Regional do Trabalho da Terceira Região, em sessão ordinária da sua Sétima Turma, hoje realizada, sob a presidência do Exmo. Desembargador Fernando Luiz Gonçalves Rios Neto, presente o Exmo. Procurador Genderson Silveira Lisboa, representante do Ministério Público do Trabalho, computados os votos do Exmo. Des. Fernando Luiz Gonçalves Rios Neto e da Exma. Desa. Cristiana Maria Valadares Fenelon, JULGOU o presente processo e, unanimemente, conheceu do recurso ordinário interposto pelo reclamante e, no mérito, sem divergência, negou provimento ao apelo. FUNDAMENTOS : ADICIONAL DE INSALUBRIDADE. Em observância ao art. 195 da CLT, tornou-se necessária a realização da perícia para a caracterização e classificação da insalubridade, segundo as normas do Ministério do Trabalho. O i. perito oficial constatou que, "conforme PPRA da ré, anexo 1, no cargo de carpinteiro, função do reclamante, existia nível de ruído médio = 87,19dB(A) estando, pois, acima do limite de tolerância máxima permissível = 85dB(A)"(ID 85e05ed - pág. 03). Salientou o louvado que "após análise técnica da ficha de controle de entrega de EPI do reclamante, Id 67027e7, foi verificado fornecimento de um único protetor auricular tipo inserção pré-moldado, de silicone

- CA 5745 "(85e05ed - pág. 04). E, em resposta ao quesito "d" formulado pela reclamada, informou o expert que o fornecimento de EPI não atendeu ao prescrito pela legislação vigente, isso porque a reclamada não comprovou estabelecimento de normas para promover a conservação e higiene do EPI. Elucidou que o PPRA da reclamada limitou-se a recomendar o uso do protetor auricular, não estabelecendo, na totalidade, as normas de proteção, razão pela qual concluiu que não foi possível considerar neutralizado o agente insalubre. Em que pese a necessidade de fixação pela empresa de normas de proteção ao empregado, entendo que a ausência de tal circunstância, por si só, não enseja, automaticamente, a caracterização da insalubridade. In casu, conforme apurado pelo expert, foi fornecido durante todo o pacto laboral um protetor auricular de silicone tipo pré-moldado, o qual, segundo pesquisa ao endereço eletrônico <https://trt-3.jusbrasil.com.br/noticias/3141639/durabilidade-de-protetor-auditivo-varia-de-4-a-12-meses>, possui vida útil de 04 a 12 meses. Logo, considerando-se que o contrato de trabalho do recorrente vigorou de 01/09/2015 a 16/11/2015, ou seja, por pouco mais de 2 meses, não há irregularidade alguma no fornecimento de apenas um EPI por todo o lapso contratual, sendo, portanto, o necessário para neutralizar o ruído existente no local de trabalho do obreiro. Assim, considerando-se que o próprio autor não nega o uso efetivo do EPI fornecido pela ré, torna-se indevido o pagamento do adicional de insalubridade (inteligência da Súmula 289 do c. TST). Frise-se que, nos termos do artigo 479 do CPC/15, o Juízo não está vinculado às conclusões do perito, podendo firmar sua convicção em outros elementos e fatos existentes nos autos, conforme ocorreu no caso vertente. Pelo exposto, mantenho a r. sentença, no particular, por seus próprios e jurídicos fundamentos. Nada a reparar. MULTA DO ART. 477 DA CLT. Diante da decisão proferida por este Eg. Regional em 08/10/2015, julgando o Incidente de Uniformização de Jurisprudência de nº 01451-2013-005-03-00-2, foi editada a Súmula nº 48 com a seguinte redação: "MULTA DO §8º DO ART 477 DA CLT. FALTA DE PAGAMENTO DAS VERBAS RESCISÓRIAS NO PRAZO LEGAL. CABIMENTO. A aplicação da multa prevista no §8º do art. 477 da CLT está restrita à falta de pagamento das verbas rescisórias no prazo fixado pelo §6º". Sob este prisma, nos termos da Súmula acima transcrita, a ausência de entrega das guias TRCT e CD/SD não autoriza a aplicação da multa em comento. Nego provimento. Belo Horizonte, 23 de fevereiro de 2017. Assinatura SABRINA DE FARIA FRÓES LEÃO Juíza Relatora Convocada SFFL/NON VOTOS”

- Rótulo: INDEFERIDO
- Explicação: Apesar do tamanho do dispositivo da decisão, é possível verificar que foi escrito “...no mérito, sem divergência, negou provimento ao apelo...” o que indica claramente que o recurso foi INDEFERIDO. Nesse caso, o tamanho do dispositivo se deve ao fato que o magistrado inseriu as fundamentações no dispositivo o que não é habitual.
- Dispositivo: “ACORDAM os Magistrados integrantes da 1ª Turma do Tribunal Regional do Trabalho da 4ª Região: por unanimidade, NÃO ACOLHER OS EMBARGOS DE DECLARAÇÃO DA RECLAMANTE, Leticia Gonzales Ludwig.”
 - Rótulo: INDEFERIDO
 - Explicação: Nesse caso, foi escrito apenas “...NÃO ACOLHER...” e nada mais o que significa que o recurso foi INDEFERIDO.
- Dispositivo: “TRT4-443 ACORDAM os Magistrados integrantes da 7ª Turma do Tribunal Regional do Trabalho da 4ª Região: preliminarmente, por unanimidade, DECLARAR de ofício, a nulidade da sentença e atos posteriores, determinando o retorno dos autos à origem para que seja proferida nova decisão observando a forma legal com a identificação, no dispositivo, das parcelas objeto da condenação, restando prejudicada a análise do recurso ordinário da reclamada (Marisa Lojas S.A.) e do recurso adesivo da reclamante (Márcia Adolfo Câmara).”
 - Rótulo: SEM_ANALISE_MERITO
 - Explicação: Nesse caso, houve a “...nulidade da sentença...” o que indica que os pedidos do recurso não foram sequer analisados.
- Dispositivo: “Vistos, relatados e discutidos os autos. ACORDAM os Magistrados integrantes da 1ª Turma do Tribunal Regional do Trabalho da 4ª Região: por unanimidade, ACOLHER OS EMBARGOS DE DECLARAÇÃO DA RECLAMADA, Kepler Weber Industrial S/A, para, sanando a omissão apontada e sem efeito modificativo, fazer constar, no item 'b' do acórdão (Id 7092c85 - Pág. 1), a condenação da reclamada à satisfação de 10 minutos a título de horas in itinere, com adicional de 50%, e reflexos em 13º salário, férias acrescidas de 1/3, aviso-prévio e FGTS, quando o trabalho ocorreu no turno das 16h18min às 01h17min. Intime-se. Porto Alegre, 04 de abril de 2018 (quarta-feira).Cabeçalho do acórdão Acórdão”
 - Rótulo: DEFERIDO

- Explicação: Nesse caso de Embargos de Declaração, é comum os magistrados usarem a palavra “ACOLHER” para indicar o deferimento e alteração da sentença de primeiro grau.

Pontos importantes

- Ler atentamente cada uma das decisões antes de marcar o rótulo.
- Cuidar para não marcar o rótulo errado devido ao mau uso do aplicativo.
- Apenas a coluna do rótulo está liberada para edição, as outras colunas estão bloqueadas.
- Decisões em que houve deferimento parcial devem ser marcadas simplesmente como DEFERIMENTO independente se o deferimento foi total ou parcial.
- Por tratar-se de serviço de considerável esforço intelectualmente, recomenda-se a realização da atividade em 30 a 40 minutos diários.

A.5 Exemplos de instâncias em que houve divergência de anotação

- Dispositivo: “Cabeçalho do acórdão CONCLUSÃO Fundamentos pelos quais O Tribunal Regional do Trabalho da Terceira Região, em sessão ordinária da sua Sétima Turma, hoje realizada, sob a presidência do Exmo. Desembargador Marcelo Lamego Pertence, presente o Exmo. Procurador Arlélcio de Carvalho Lage, representante do Ministério Público do Trabalho, computados os votos da Exma. Des. Cristiana Maria Valadares Fenelon e do Exmo. Des. Paulo Roberto de Castro, JULGOU o presente processo e, unanimemente, não conheceu o Agravo de Petição, por ausência de ataque aos fundamentos da r. Decisão recorrida (Súmula 422 do c. TST). Custas processuais pela Executada, no importe de R\$44,26, nos termos do artigo 789-A, IV, da CLT. Belo Horizonte, 7 de junho de 2018. Assinatura FERNANDO ANTÔNIO VIÉGAS PEIXOTO Desembargador Relator VOTOS”
 - Rótulo aplicado pelo autor: SEM_ANALISE_MERITO
 - Rótulo aplicado pelo voluntário: INDEFERIMENTO
 - Rótulo adjudicado: SEM_ANALISE_MERITO
 - Explicação: O voluntário incorreu em erro ou por falta de atenção ou por não ter percebido que não há nenhuma informação indicando o vencedor da causa e que também há uma informação clara de que o recurso nem foi apreciado o que pode ser percebido pelas palavras “não conheceu”.

- Dispositivo: “Cabeçalho do acórdão CONCLUSÃO CONCLUSÃO O Tribunal Regional do Trabalho da Terceira Região, em sessão ordinária da sua Oitava Turma, hoje realizada, sob a Presidência do Exmo. Desembargador Sérgio da Silva Peçanha, presente a Exma. Procuradora Maria Amélia Bracks Duarte, representante do Ministério Público do Trabalho e, computados os votos dos Exmos. Desembargadores Ana Maria Amorim Rebouças e José Marlon de Freitas; JULGOU o presente processo e, preliminarmente, à unanimidade, conheceu do Recurso Ordinário interposto pelo Reclamante (fls. 103/112), porquanto presentes os pressupostos de cabimento e de admissibilidade; no mérito, sem divergência, deu-lhe parcial provimento para acrescer a condenação da Reclamada ao pagamento de: a) 10 horas extras semanais, acrescidas do adicional de 50%, no período contratual de fevereiro de 2013 até agosto de 2014, com reflexos em 13º salário, férias+1/3, FGTS (este ser depositado na conta vinculada do autor); b) dobra de domingos, bem como dos feriados previstos nas Leis federais de número 662/49 e 6.802/80, quais sejam: 1º de janeiro, 21 de abril, 1º de maio, 7 de setembro, 12 de outubro, 2 de novembro, 15 de novembro e 25 de dezembro, bem como aqueles religiosos amplamente consagrados, quais sejam, sexta-feira da Paixão e Corpus Christi, no período contratual de fevereiro de 2013 a agosto de 2014 e de dezembro de 2016 a janeiro de 2017, com reflexos em férias + 1/3, 13º salário e FGTS (este ser depositado na conta vinculada do autor), conforme se apurar em sede de liquidação; c) 1 (uma) hora extra diária, decorrente da supressão do intervalo intrajornada nos períodos em que não foram juntados os cartões de ponto (fevereiro de 2013 a agosto de 2014 e de dezembro 2016 a janeiro de 2017), com adicional legal de 50%, e reflexos em férias + 1/3, 13ºs. salários, FGTS; para fins previdenciários, declarou-se a natureza salarial das parcelas condenatórias ora deferidas, à exceção das seguintes parcelas que possuem natureza indenizatória: reflexos em férias indenizadas + 1/3 e FGTS (este ser depositado na conta vinculada do autor); acresceu à condenação o valor de R\$3.000,00(três mil reais), com custas igualmente acrescidas de R\$60,00(sessenta reais), a cargo da Reclamada, que, com a publicação deste acórdão, fica intimada ao seu pagamento, nos termos da Súmula nº 25 do TST; quanto aos demais tópicos e argumentos recursais, negou provimento ao Apelo, adotando, como razões de decidir, os”

– Rótulo aplicado pelo autor: DEFERIMENTO

– Rótulo aplicado pelo voluntário: INDEFERIMENTO

Figura A.9: Matriz de confusão construída utilizando a base de 30% reservada previamente para testes do modelo de classificação do deferimento ou não da decisão.

Rocchio classifier

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	135	12	17
DEFERIMENTO	13	110	10
SEM_ANALISE_MERITO	1	0	2

- Rótulo adjudicado: DEFERIMENTO
- Explicação: O voluntário incorreu em erro possivelmente em virtude de que o dispositivo apresenta bem mais palavras do que a média o que torna a leitura mais difícil. Por outro lado, é possível observar que há uma informação clara de que o recurso foi deferido o que pode ser percebido pelas palavras “deu-lhe parcial provimento”.

A.6 Figuras das matrizes de confusão do classificador quanto ao deferimento ou não da decisão treinado com base de dados criada manualmente

As Figuras A.9, A.10, A.11, A.12, A.13, A.14 e A.15 apresentam respectivamente as matrizes de confusão em relação aos algoritmos classificadores Rocchio, Gradient Boosting, Naive Bayes, K-nearest Neighbor, Support Vector Machine (SVM), Decision Tree e Random Forest desenvolvidos na Seção 6.4.2.1. Importante ressaltar que o eixo horizontal representa o rótulo verdadeiro e o eixo vertical representa o rótulo previsto.

Figura A.10: Matriz de confusão construída utilizando a base de 30% reservada previamente para testes do modelo de classificação do deferimento ou não da decisão.

Gradient Boosting Classifier

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	162	2	0
DEFERIMENTO	1	130	2
SEM_ANALISE_MERITO	0	0	3

Figura A.11: Matriz de confusão construída utilizando a base de 30% reservada previamente para testes do modelo de classificação do deferimento ou não da decisão.

Naive Bayes Classifier

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	160	4	0
DEFERIMENTO	57	76	0
SEM_ANALISE_MERITO	3	0	0

Figura A.12: Matriz de confusão construída utilizando a base de 30% reservada previamente para testes do modelo de classificação do deferimento ou não da decisão.

K-nearest Neighbor

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	158	6	0
DEFERIMENTO	67	66	0
SEM_ANALISE_MERITO	3	0	0

Figura A.13: Matriz de confusão construída utilizando a base de 30% reservada previamente para testes do modelo de classificação do deferimento ou não da decisão.

Support Vector Machine (SVM)

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	158	6	0
DEFERIMENTO	9	124	0
SEM_ANALISE_MERITO	3	0	0

Figura A.14: Matriz de confusão construída utilizando a base de 30% reservada previamente para testes do modelo de classificação do deferimento ou não da decisão.

Decision Tree

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	160	4	0
DEFERIMENTO	7	124	2
SEM_ANALISE_MERITO	2	0	1

Figura A.15: Matriz de confusão construída utilizando a base de 30% reservada previamente para testes do modelo de classificação do deferimento ou não da decisão.

Random Forest

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	156	8	0
DEFERIMENTO	5	128	0
SEM_ANALISE_MERITO	3	0	0

Figura A.16: Matriz de confusão construída utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	523	5	11
DEFERIMENTO	27	401	13
SEM_ANALISE_MERITO	0	0	20

A.7 Figuras das matrizes de confusão do classificador quanto ao deferimento ou não da decisão treinado com base de dados balanceada criada programaticamente

As Figuras A.21, A.17, A.19, A.18, A.22, A.16, A.20 apresentam respectivamente as matrizes de confusão em relação aos algoritmos classificadores Rocchio, Gradient Boosting, Naive Bayes, K-nearest Neighbor, Support Vector Machine (SVM), Decision Tree e Random Forest desenvolvidos na Seção 6.4.2.2. Importante ressaltar que o eixo horizontal representa o rótulo verdadeiro e o eixo vertical representa o rótulo previsto.

A.8 Figuras das matrizes de confusão do classificador quanto ao deferimento ou não da decisão treinado com base de dados completa criada programaticamente

As Figuras A.28, A.24, A.26, A.25, A.29, A.23, A.27 apresentam respectivamente as matrizes de confusão em relação aos algoritmos classificadores Rocchio, Gradient Boosting, Naive Bayes, K-nearest Neighbor, Support Vector Machine (SVM), Decision Tree e Random Forest desenvolvidos na Seção 6.4.2.3. Importante ressaltar que o eixo horizontal representa o rótulo verdadeiro e o eixo vertical representa o rótulo previsto.

Figura A.17: Matriz de confusão construída utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	529	1	9
DEFERIMENTO	31	398	12
SEM_ANALISE_MERITO	0	1	19

Figura A.18: Matriz de confusão construída utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	470	27	42
DEFERIMENTO	161	260	20
SEM_ANALISE_MERITO	8	1	11

Figura A.19: Matriz de confusão construída utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	408	28	103
DEFERIMENTO	37	322	82
SEM_ANALISE_MERITO	0	2	18

Figura A.20: Matriz de confusão construída utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	521	12	6
DEFERIMENTO	13	415	13
SEM_ANALISE_MERITO	0	0	20

Figura A.21: Matriz de confusão construída utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	339	37	163
DEFERIMENTO	19	332	90
SEM_ANALISE_MERITO	0	2	18

Figura A.22: Matriz de confusão construída utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	519	12	8
DEFERIMENTO	16	405	20
SEM_ANALISE_MERITO	0	2	18

Figura A.23: Matriz de confusão construída utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	529	4	6
DEFERIMENTO	38	403	0
SEM_ANALISE_MERITO	0	1	19

Figura A.24: Matriz de confusão construída utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	531	4	4
DEFERIMENTO	36	404	1
SEM_ANALISE_MERITO	0	1	19

Figura A.25: Matriz de confusão construída utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	532	6	1
DEFERIMENTO	168	273	0
SEM_ANALISE_MERITO	17	2	1

Figura A.26: Matriz de confusão construída utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	508	31	0
DEFERIMENTO	103	338	0
SEM_ANALISE_MERITO	16	4	0

Figura A.27: Matriz de confusão construída utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	531	4	4
DEFERIMENTO	29	412	0
SEM_ANALISE_MERITO	0	3	17

Figura A.28: Matriz de confusão construída utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	364	38	137
DEFERIMENTO	33	339	69
SEM_ANALISE_MERITO	0	2	18

Figura A.29: Matriz de confusão construída utilizando a base de dados padrão-ouro para testes do modelo de classificação do deferimento ou não da decisão.

	INDEFERIMENTO	DEFERIMENTO	SEM_ANALISE_MERITO
INDEFERIMENTO	529	7	3
DEFERIMENTO	25	416	0
SEM_ANALISE_MERITO	0	4	16