

APLICAÇÃO DE TÉCNICAS DE PROCESSAMENTO DE LINGUAGEM NATURAL PARA ANÁLISE DA OBSERVÂNCIA DE ARTIGOS SELECIONADOS DA REFORMA TRABALHISTA (LEI N.º 13.467/2017)

APPLICATION OF NATURAL LANGUAGE PROCESSING TECHNIQUES FOR THE ANALYSIS OF THE OBSERVANCE OF SELECTED ARTICLES OF THE LABOR REFORM (LAW 13.467/2017)

Tiago Ribeiro Alves de Melo¹, Shaiala Ribeiro de Castro Araujo Marques², Francine Athaide Cadore³

Recebido em: 18/04/2019. Aprovado em: 22/04/2019.

Resumo

O presente artigo apresenta o resultado de uma análise de observância da Reforma Trabalhista em processos judiciais de primeira instância dos TRTs da 2ª, 3ª e 15ª Regiões, que tiveram decisão de 11/11/2017 a 30/05/2018. Sua aplicabilidade foi analisada nos temas: (a) Horas *in itinere* – Art. 58; (b) Terceirização – Art 4º-A; Correção monetária pela TR – Art. 879; (d) Contribuição sindical – Art. 579; (e) intervalo intrajornada – Art. 71; e (f) Negociado sobre o legislado – Art. 611-A. O estudo utilizou técnicas de processamento de linguagem natural para a classificação das decisões que se referenciam à Reforma Trabalhista e o enquadramento dos méritos da decisão em “Não Aplica” e “Admite aplicação”. As classificações foram construídas seguindo uma abordagem de construção de dicionários semânticos para cada tema, com a determinação de classes, orientada pela técnica de *feature engineering*, e a construção de matrizes termo-documento a partir de técnicas de TFDCF e TFIDF.

Palavras-chave

Inteligencia Artificial; jurimetria; processamento de linguagem natural.

¹ Doutor em Economia de la Empresa, pela Universidade de Salamanca, Mestre em Business Management pela Liverpool University. Atualmente é Executivo de Produto na Softplan. E-mail: melotiago@hotmail.com.

² Mestre em Direito pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Especialista em Direito dos Contratos e Responsabilidade Civil pela Universidade do Vale do Rio dos Sinos (Unisinos), Bacharel em Ciências Jurídicas e Sociais pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Atualmente é Data Science Legal Advisor na Softplan. E-mail: shaiala.marques@gmail.com.

³ Especialista em Marketing Estratégico com ênfase em Marketing Digital pela Escola Superior de Propaganda e Marketing (ESPM), Especialista em Gestão da Comunicação Pública e Empresarial pela Universidade Tuiuti do Paraná, Bacharel em Comunicação Social – Habilitação em Jornalismo pela Universidade Federal de Santa Maria (UFSM). Atualmente é coordenadora de Assessoria de Comunicação na Softplan. E-mail: cadorefrancine@gmail.com.

Abstract

This article presents the technical approach based on artificial intelligence, used in a project to analyse the observance of the Labor Reform in judicial proceedings of the first instance of the Labor Tribunals of the 2nd, 3rd and 15th Regions, with decisions from 11/11/2017 to 30/05/2018. Its applicability was examined within the following themes : (a) Hours in itinere – Art. 58; (b) Outsourcing – Art 4º-A; Monetary correction – Art. 879; (d) Union contribution – Art. 579; (e) Intra-journey interval – Art. 71; and (f) Negotiated over the legislated – Art. 611-A. The study successfully employed natural language processing techniques to classify the decisions that refer to the labor reform and the framing of the merits of the decision in "do not apply" and "admits application". The classifications were constructed following an approach based on semantic dictionaries for each theme, with the determination of classes, guided by the technique of Feature Engineering; and the construction of thermo-document matrices from TFDCF and TFIDF techniques.

Keywords

Artificial Intelligence; jurimetrics; natural language processing.

1 Introdução

No dia 11 de novembro de 2017, entrou em vigor a Reforma Trabalhista – Lei Federal nº 13.467/2017 – que alterou significativamente diversos pontos da Consolidação das Leis do Trabalho (CLT).

A Softplan foi contratada no início de 2018 por uma grande entidade patronal interessada em analisar a observância de alguns dos artigos alterados pela Reforma Trabalhista em decisões de primeira instância nos Tribunais Regionais de Trabalho da 2ª, 15ª e 3ª Regiões.

Em respeito a acordos de confidencialidade e atenção às regras de proteção de propriedade intelectual da Softplan, estamos impedidos de revelar detalhes e resultados específicos desse estudo técnico-jurídico.

O objetivo do presente artigo é descrever o processo de concepção e abordagens técnicas utilizadas no projeto. Ou seja, demonstrar como a inteligência artificial, em especial sua subárea de processamento de linguagem natural, pode ser usada para resolver problemas reais. Apresentamos o artigo em forma de relatório técnico de execução do projeto, sem pretensões de realizar uma revisão ou aprofundamento do referencial teórico que fundamenta a aplicação do ferramental técnico-computacional utilizado no projeto.

O objetivo do projeto foi entregar à entidade patronal informações atualizadas e com alto nível de acurácia sobre as decisões proferidas no período de 11/11/2017 a 30/05/2018, dos tribunais selecionados, classificando-as, quando possível, em “Não aplica” e “Admite aplicação”. Esse trabalho será doravante chamado de Observatório da Reforma Trabalhista.

O Quadro 1 relaciona os temas objetos deste estudo.

Quadro 1 - Relação de artigos objetos de estudo

Tema	Fundamentação Legal
Horas <i>in itinere</i> . Horas em trânsito, Transporte, Horas no trajeto não será jornada de trabalho.	Art. 58 - A duração normal do trabalho, para os empregados em qualquer atividade privada, não excederá de 8 (oito) horas diárias, desde que não seja fixado expressamente outro limite. § 2º O tempo despendido pelo empregado desde a sua residência até a efetiva ocupação do posto de trabalho e para o seu retorno, caminhando ou por qualquer meio de transporte, inclusive o fornecido pelo empregador, não será computado na jornada de trabalho, por não ser tempo à disposição do empregador.
Terceirização: atividade principal; atividade-fim, meio, prestação de serviços; Atividade de apoio	Art. 4º-A. Considera-se prestação de serviços a terceiros a transferência feita pela contratante da execução de quaisquer de suas atividades, inclusive sua atividade principal, à pessoa jurídica de direito privado prestadora de serviços que possua capacidade econômica compatível com a sua execução.
Correção monetária pela TR. Atualização dos créditos	Art. 879 - Sendo ilíquida a sentença exequenda, ordenar-se-á, previamente, a sua liquidação, que poderá ser feita por cálculo, por arbitramento ou por artigos § 7º A atualização dos créditos decorrentes de condenação judicial será feita pela Taxa Referencial (TR), divulgada pelo Banco Central do Brasil, conforme a Lei nº8.177, de 1º de março de 1991.
Contribuição sindical	Art. 579 do Decreto-Lei nº 5.452, atualizado pela Lei Federal nº 13.467/2017. Art. 579. O desconto da contribuição sindical está condicionado à autorização prévia e expressa dos que participarem de uma determinada categoria econômica ou profissional, ou de uma profissão liberal, em favor do sindicato representativo da mesma categoria ou profissão ou, inexistindo este, na conformidade do disposto no art. 591 desta Consolidação. Art. 8, IV da Constituição Federal Art. 8º É livre a associação profissional ou sindical, observado o seguinte: IV - a assembleia geral fixará a contribuição que, em se tratando de categoria profissional, será descontada em folha, para custeio do sistema confederativo da representação sindical respectiva, independentemente da contribuição prevista em lei.
Intervalo intrajornada	Art. 71 - Em qualquer trabalho contínuo, cuja duração exceda de 6 (seis) horas, é obrigatória a concessão de um intervalo para repouso ou alimentação, o qual será, no mínimo, de 1 (uma) hora e, salvo acordo escrito ou contrato coletivo em contrário, não poderá exceder de 2 (duas) horas. § 4º A não concessão ou a concessão parcial do intervalo intrajornada mínimo, para repouso e alimentação, a empregados urbanos e rurais, implica o pagamento, de natureza indenizatória, apenas do período suprimido, com acréscimo de 50% (cinquenta por cento) sobre o valor da remuneração da hora normal de trabalho.
Negociado sobre o legislado: Negociação sindical prevalece sobre o que está na legislação	Art. 611-A. A convenção coletiva e o acordo coletivo de trabalho têm prevalência sobre a lei quando, entre outros, dispuserem sobre: I - pacto quanto à jornada de trabalho, observados os limites constitucionais; II - banco de horas anual; III - intervalo intrajornada, respeitado o limite mínimo de trinta minutos para jornadas superiores a seis horas; IV - adesão ao Programa Seguro-Emprego (PSE), de que trata a Lei nº 13.189, de 19 de novembro de 2015; V - plano de cargos, salários e funções compatíveis com a condição pessoal do empregado, bem como identificação dos cargos que se enquadram como funções de confiança; VI - regulamento empresarial; VII - representante dos trabalhadores no local de trabalho; VIII - teletrabalho, regime de sobreaviso, e trabalho intermitente; IX - remuneração por produtividade, incluídas as gorjetas percebidas pelo empregado, e remuneração por desempenho individual; X - modalidade de registro de jornada de trabalho;

MELO, Tiago Ribeiro Alves de; MARQUES, Shaiala Ribeiro de Castro Araujo; CADORE, Francine Athaide. Aplicação de técnicas de processamento de linguagem natural para análise da observância de artigos selecionados da Reforma Trabalhista (Lei n.º 13.467/2017). **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 58-69, 2019.

	<p>XI - troca do dia de feriado;</p> <p>XII - enquadramento do grau de insalubridade;</p> <p>XIII - prorrogação de jornada em ambientes insalubres, sem licença prévia das autoridades competentes do Ministério do Trabalho;</p> <p>XIV - prêmios de incentivo em bens ou serviços, eventualmente concedidos em programas de incentivo;</p> <p>XV - participação nos lucros ou resultados da empresa.</p> <p>§ 1º No exame da convenção coletiva ou do acordo coletivo de trabalho, a Justiça do Trabalho observará o disposto no § 3º do art. 8º desta Consolidação.</p> <p>§ 2º A inexistência de expressa indicação de contrapartidas recíprocas em convenção coletiva ou acordo coletivo de trabalho não ensejará sua nulidade por não caracterizar um vício do negócio jurídico.</p> <p>§ 3º Se for pactuada cláusula que reduza o salário ou a jornada, a convenção coletiva ou o acordo coletivo de trabalho deverão prever a proteção dos empregados contra dispensa imotivada durante o prazo de vigência do instrumento coletivo.</p> <p>§ 4º Na hipótese de procedência de ação anulatória de cláusula de convenção coletiva ou de acordo coletivo de trabalho, quando houver a cláusula compensatória, esta deverá ser igualmente anulada, sem repetição do indébito.</p> <p>§ 5º Os sindicatos subscritores de convenção coletiva ou de acordo coletivo de trabalho deverão participar, como litisconsortes necessários, em ação individual ou coletiva, que tenha como objeto a anulação de cláusulas desses instrumentos.</p>
--	---

Fonte: Autores.

2 Quem é a Softplan

A Softplan é uma empresa de tecnologia da informação e desenvolvimento de sistemas sediada na cidade de Florianópolis, Brasil, com 28 anos de atuação no mercado da justiça. Entre outras soluções, ela é a desenvolvedora do SAJ (Sistema de Automação da Justiça), sistema de gestão de processos judiciais privado com maior abrangência no Brasil, em uso atualmente em Ministérios Públicos (Mato Grosso do Sul, Santa Catarina, Acre, entre outros), Procuradorias Jurídicas Municipais e Estaduais (São Paulo, Pernambuco, Distrito Federal, Santa Catarina, entre outros), e Tribunais de Justiça (São Paulo, Santa Catarina, Amazonas, Ceará, entre outros). Além do SAJ e desses segmentos, a Softplan desenvolve soluções, produtos e análises para a advocacia privada. Aproximadamente 50% dos processos na justiça estadual do Brasil tramitam no SAJ.

3 Abordagem técnica

A abordagem técnica deste estudo está dividida em questões relacionadas a:

- (a) Engenharia de dados: envolvendo a identificação de processos candidatos à população da análise, a estratégia de captura para garantia da maior abrangência possível e a execução da captura de dados e documentos processuais;
- (b) Ciência de dados: abrangendo as técnicas e ferramentas empregadas para leitura de peças processuais, extração e identificação de processos que compõem o objeto de estudo e a classificação objetiva da observância ou não da aplicação da Reforma Trabalhista.

MELO, Tiago Ribeiro Alves de; MARQUES, Shaiala Ribeiro de Castro Araujo; CADORE, Francine Athaide. Aplicação de técnicas de processamento de linguagem natural para análise da observância de artigos selecionados da Reforma Trabalhista (Lei n.º 13.467/2017). **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 58-69, 2019.

3.1 *Engenharia de dados*

O estudo apresentou uma série de desafios técnicos derivados das características intrínsecas do sistema judiciário brasileiro:

- (a) Cada tribunal tem um sistema único e próprio de acesso aos dados públicos;
- (b) Esses sistemas apresentam requisitos distintos de segurança, *layout* e abrangência dos dados;
- (c) Eles também têm tabelas próprias de categorização das decisões (ou se não são diferentes, não têm homogeneidade de práticas), e os teores das decisões não estão disponíveis em forma de texto, estão dentro de documentos.

Essas características são obstáculos estruturais que dificultam uma busca em larga escala para a elaboração de um diagnóstico completo e confiável, como é o caso da construção do Observatório.

3.1.1 *Estratégia de identificação de processos para compor a população do estudo*

Dado que o objetivo do Observatório é capturar as decisões que versam sobre artigos específicos da Reforma Trabalhista e considerando que é imprescindível analisar todas as decisões desses tribunais de 11/11/2017 a 30/05/2018, optou-se pela seguinte estratégia para garantir a máxima abrangência possível:

A) Diários de Justiça Eletrônicos:

- Desenvolvimento de robôs conectados aos Diários de Justiça Eletrônicos dos respectivos tribunais;
- Os robôs capturaram todas as movimentações publicáveis dos diários;
- Em seguida, uma inteligência de negócio classificou todos os tipos de movimentação que tinham em sua denominação caráter de sentença, decisão, julgamento ou qualquer outro teor terminativo de processo;
- Por fim, os números de processos que continham esse tipo de movimentação foram guardados para uma busca ativa no site de acesso público de processos de cada tribunal.

B) Buscas ativas nos sites de acesso dos tribunais

Para garantir a abrangência máxima, além dos processos selecionados a partir da leitura dos Diários de Justiça Eletrônicos, foi executada uma consulta ativa de processos nos sites dos tribunais. Por meio de uma técnica estatística de predição de números sequenciais de processos – a partir do sequenciamento do número do CNJ e seu comportamento histórico – foram encontrados todos os processos distribuídos nos tribunais a partir de 01/01/2015, que também foram adicionados à população do estudo para verificação de movimentações de decisões a partir de 11/11/2017.

3.1.2 *Coleta e captura de dados e documentos*

A coleta dos dados processuais deu-se a partir da lista de processos gerada por capturas ativas e oriundas de Diários de Justiça – por meio da técnica de *crawling* dos dados processuais disponibilizados nos sites públicos de acesso dos tribunais trabalhistas.

A técnica de *crawling* consiste no desenvolvimento de rotinas automatizadas que executam buscas em determinados sites ou sistemas simulando uma interação humana. Por ser automatizada e iterativa, essa técnica garante que as buscas tenham alta precisão e sejam executadas em alta escala.

O processo automatizado de coleta de dados atende a requisitos específicos para que não afete o ambiente onde estão hospedados os sistemas de atuação, tais como:

- Utilização da automação em horários de baixa atividade;
- Restrição de qualquer métrica/configuração na automação que não influencie no comportamento dos sistemas em que atua, garantindo que não ocorra degradação de performance e qualquer outro tipo de sobrecarga decorrente do processo.

Foram coletados os seguintes dados processuais: números de processo, assuntos primários, assuntos secundários, partes e advogados, unidades judiciais, comarcas e tribunais, data e tipo de todas as movimentações (essas são posteriormente normalizadas para classificação de status de processo, data de distribuição, julgamento e baixa).

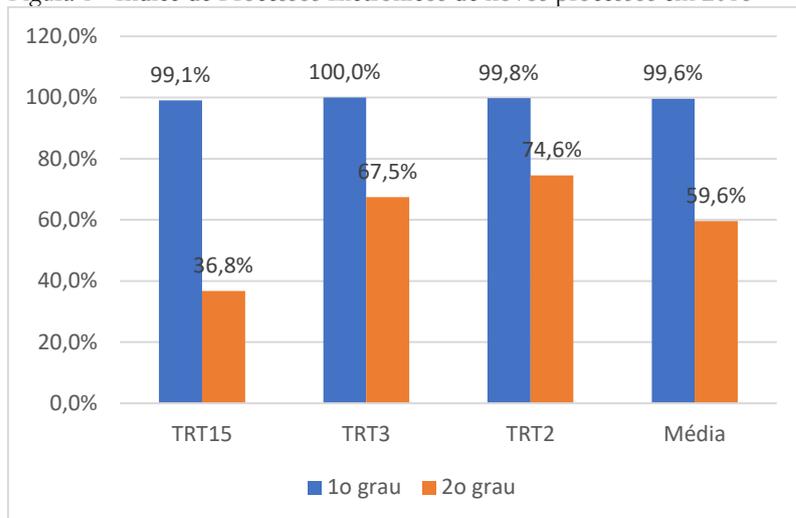
Além dos dados processuais, foram coletados todos os documentos categorizados em movimentações terminativas de processos, que foram em seguida objeto de processamento para as classificações deste estudo, a partir de técnicas de processamento de linguagem natural.

3.2 *Ciência de dados*

A disciplina de ciência de dados está em bastante evidência em praticamente todos os segmentos de mercado nos últimos cinco anos. A oportunidade na área de ciência de dados tem se expandido para diversos campos e as aplicações podem ser concebidas nas mais diversas áreas do conhecimento, como aponta Fawcett e Provost. (2013). No caso da Justiça, a partir de 2016, com a consolidação do processo digital, sua aplicação passou de experimental para efetiva.

De acordo com a Figura 1, no ano de 2016, 99,6% dos novos processos que entraram nos tribunais objetos de análise, fizeram-no em sua forma digital. Isso significa que praticamente todos os processos que compõem a população deste estudo estão disponíveis e acessíveis, aqui contemplando os metadados (que não são os dados dispostos em forma de texto estruturado nos sites dos tribunais) e os documentos (também disponíveis nos sites dos tribunais, porém de forma não estruturada, dentro dos documentos que compõem as peças processuais).

Figura 1 - Índice de Processos Eletrônicos de novos processos em 2016



Fonte: Justiça em Números 2017.

A subárea da ciência de dados que estuda a extração e classificação de elementos a partir de texto é o processamento de linguagem natural. Dentro dessa subárea, existem diversas técnicas e ferramentas que podem ser aplicadas de forma individual e coletivamente para a classificação de elementos a partir de uma extensiva massa de dados (grande número de documentos).

A literatura especializada é bastante rica na argumentação sobre as limitações cognitivas humanas para o processamento e entendimento de quantidades massivas de dados e documentos. Em um artigo seminal de 1972, intitulado “*Subjective Probability: A Judgment of Representativeness*”, os professores Daniel Kahneman e Amos Tversky listaram algumas dessas limitações: “ancoragem, ilusão da frequência, viés da confirmação, ilusão da validação, viés do otimismo”.

O objetivo do processamento de linguagem natural é desenvolver técnicas computacionais e matemáticas que, a partir de uma orientação/curadoria especializada, executem as tarefas de entendimento, extração, classificação e interpretação de texto de forma automatizada, reduzindo o efeito das limitações cognitivas humanas.

O caso específico do projeto do Observatório sobre a Reforma Trabalhista apresenta uma série de características únicas que o qualificam como um projeto de altíssima complexidade técnica, demandando o emprego de múltiplas abordagens, inclusive uma intensiva curadoria especialista.

A academia classifica o trabalho de busca e de estruturação de texto em dados não estruturados (como peças processuais) como a disciplina de *Information Retrieval*. Essa disciplina está composta por um conjunto de técnicas e artifícios matemáticos e computacionais aplicáveis de acordo com a problemática de cada situação.

Para o presente estudo, a atividade de classificação consistiu nas tarefas de agrupamento dos processos entre:

MELO, Tiago Ribeiro Alves de; MARQUES, Shaiala Ribeiro de Castro Araujo; CADORE, Francine Athaide. Aplicação de técnicas de processamento de linguagem natural para análise da observância de artigos selecionados da Reforma Trabalhista (Lei n.º 13.467/2017). **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 58-69, 2019.

- (a) Os que referenciam a Reforma Trabalhista;
- (b) Os que referenciam a um dos temas objetos deste estudo;
- (c) Os que utilizam o tema específico para balizar sua decisão;
- (d) Os que balizaram sua decisão e cujo mérito foi objetivamente definido;
- (e) O mérito da decisão entre os grupos de “Não Aplica”, “Aplica” e “Admite aplicação”:
 - Não Aplica: Por declaração de Inconstitucionalidade e Por outros motivos
 - Aplica: Declara a aplicação da lei no respectivo artigo, de forma líquida e textual
 - Admite aplicação: Infere-se a aplicação em decorrência da argumentação da decisão em relação à temporalidade do fato. Quando houver casos julgados baseados em fatos após a promulgação da lei, a mesma será aplicada

A) Filtro: Faz referência à Reforma Trabalhista e a um dos temas objeto deste estudo?

A Figura 2 detalha a cascata das classificações desde a coleta inicial de processos que tiveram decisão no período especificado e nos três tribunais pesquisados. Foram identificados 672.000 processos candidatos a participar do estudo. Destes, 101.880 fizeram algum tipo de referência à Reforma Trabalhista. O segundo filtro considerou se houve uma citação do tema específico deste estudo. Foram identificados 44.579 processos que se referiam aos artigos objetos deste estudo.

Nos dois filtros iniciais, não se aplicaram técnicas semânticas de *Information Retrieval*. Foi realizada basicamente uma análise de similaridade e morfológica com um conjunto de *corpus* que se relacionam à Reforma Trabalhista (tais como: reforma trabalhista, Reforma Trabalhista e as variações do número da nova lei). Levou-se em consideração os aspectos da matéria e a forma com que seria possível relacionar os termos à necessidade final, entendendo que não bastaria um treinamento do *corpus* para a obtenção de resultado, como algumas técnicas comuns sugerem (AERY, RAMAMURTHY, ASLANDOGAN, 2003).

B) Filtro: Os temas objeto deste estudo são usados como balizadores da decisão?

As técnicas mais refinadas de processamento de linguagem natural foram empregadas a partir do terceiro filtro de classificação. Nesse filtro, o contexto da referência dos artigos objetos de estudo passa a ser crítico, visto que, dado à imensa publicidade dada à Reforma Trabalhista, o termo pode ser sido usado de forma genérica, ilustrativa ou sem força argumentativa decisória.

A literatura cita uma série de vícios semânticos inerentes à linguagem natural que se apresentaram de forma acentuada no presente estudo. Um texto jurídico tradicional, apesar de compreender um vocabulário reduzido, comparado com o espectro total do idioma português, é intrinsecamente mais rebuscado, formal e subjetivo que a linguagem coloquial. O contexto semântico nesse cenário torna-se fundamental para a compreensão do real significado de palavras, expressões e sentenças.

A presença de hiperônimos e hipônimos, por exemplo, são bastante elementares em textos jurídicos, utilizados para fins de argumentação técnica. A título de exemplo, “Taxa Legal”, tema referenciado no artigo de Correção Monetária pela TR, configura-se como um hipônimo, do hiperônimo “Taxa”; enquanto que “Código de Processo Civil” e “Código Civil” são hipônimos do hiperônimo “Código”.

Uma forma de controlar o efeito dessas propriedades semânticas é aplicar uma abordagem técnica de Raciocínio Baseado em Casos (RBC), em que se estudam casos reais de aplicação de determinado evento e, a partir do aprendizado histórico, replica-se o entendimento no presente texto. Como a Reforma Trabalhista é um assunto novo e singular, descartou-se a aplicação dessa abordagem.

Alternativamente, podem-se aplicar técnicas relacionadas à criação de *Thesaurus* (dicionários temáticos), em que, por indução, determinado evento pode ser descoberto. No entanto, como o presente estudo tem o caráter vanguardista, não há referencial teórico ou histórico que pudesse balizar a construção e aplicação de *thesaurus* como método único de abordagem técnica.

A estratégia adotada, portanto, focou na construção de um *Thesaurus* fortemente conduzido por análises massivas dos documentos. Ou seja, optou-se por um método bastante conservador de se apoiar no conhecimento técnico aplicado de especialistas jurídicos internos, que efetivamente leram e anotaram manualmente um conjunto significativo de características que conduziram à criação das classes.

Para o entendimento se o tema referenciado foi fundamentador de uma decisão, foram lidos por um especialista interno e classificados em torno de 1.000 (um mil) processos. O termo técnico dado a esse trabalho realizado por um especialista de negócio é o *Feature Engineering*. Essa técnica, aplicada com rigor metodológico, conduz à construção de classes (*features*) e matrizes de termos-documento de alto refinamento técnico e escalabilidade no campo específico de estudo. Ela não é replicável em cenários mais abrangentes.

Isso foi comprovado quando as matrizes de classificações elaboradas inicialmente para um dos tribunais não obtiveram o mesmo desempenho em outros tribunais. O contexto da argumentação jurídica deste estudo provou-se extremamente individualizado.

Nessa etapa, foram descartados um conjunto bastante representativo de decisões, em que o mérito não foi objeto do julgamento, como nos casos de: extinção por abandono, desistência, homologação de acordo, correções de fluxo, inépcia da inicial, entre outros.

Figura 2 - Cascata de classificações primárias dos processos (em mil processos)



Fonte: Softplan.

MELO, Tiago Ribeiro Alves de; MARQUES, Shaiala Ribeiro de Castro Araujo; CADORE, Francine Athaide. Aplicação de técnicas de processamento de linguagem natural para análise da observância de artigos selecionados da Reforma Trabalhista (Lei n.º 13.467/2017). **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 58-69, 2019.

Aliado à construção técnico-jurídica das classes e matrizes, foi aplicado um conjunto de técnicas computacionais que maximizaram o potencial das classificações. Dependendo da característica da matriz termo-documento de cada classe, empregou-se a técnica de TFDCF (*term frequency, disjoint corpora frequency*) ou de TFIDF (*term frequency - inverse document frequency*). A primeira controla matematicamente o efeito de repetições de texto – ou extratos de texto – semanticamente dispensáveis, evitando assim um peso relativo desproporcional a palavras ou expressões. A segunda foca na associação de palavras ou expressões em grupos (bigramas, trigramas, etc.) que possam exprimir sentido de forma coletiva.

Tais técnicas, além de compatíveis com o objeto de estudo, mostram-se aplicáveis à ideia geral de resultado em que se busca obter (MORENO, ROMÁ-FERRI, MOREDA; 2017).

Foram encontrados 5.772 processos que se enquadram como fundamentadores da decisão, 12,9% do total de referências a um dos seis artigos objetos de estudo.

C) Filtro: O mérito da decisão está definido?

Esse penúltimo filtro, para capturar a definição do mérito da decisão, contou com participação direta e indispensável da equipe jurídica do cliente como curadoria especializada. Adicionou-se, portanto, mais uma camada de *Feature Engineering*, trazendo maior acurácia às classificações.

Buscou-se aqui segregar as decisões tácitas e sem uma objetividade líquida do mérito, para não enviesar a extração do mérito (Não Aplica/Aplica/Admite Aplicação) na fase seguinte.

Nessa etapa, pôde-se observar claramente a tendência à referência à Reforma Trabalhista nas decisões, porém sem que essa citação tenha qualquer balizamento de seu mérito. Nesses casos, para evitar falsos positivos e uma classificação de mérito equivocada, preferiu-se criar uma classe “Indefinida”. Nela foram enquadrados os processos em que o nível médio de certeza na classificação de “Não aplica”, “Aplica” ou “Admite Aplicação” não atingiu 80% de acurácia.

D) Filtro: O mérito da decisão é classificado em “Não Aplica”, “Aplica” e “Admite Aplicação”

Uma vez segregados os processos em que o mérito não pôde ser objetivamente determinado, permaneceram aqueles em que o mérito foi classificado como definido.

Estes passaram pela classificação mais importante do trabalho e objeto da contratação do cliente. Para essas classificações, foram aplicadas as mesmas técnicas e abordagens dos filtros que empregam o conjunto de técnicas de processamento de linguagem natural orientadas pelo *feature engineering* interno da Softplan e do cliente.

Além da importância do contexto, como nos agrupamentos anteriores, o mérito da Aplicabilidade e o motivo da Não Aplicação buscou entendimentos com alto grau de liquidez, em que as declarações foram próximas do literal (considerando a aplicação das técnicas já abordadas de TFDCF e TFIDF).

Os processos classificados como passíveis de aplicação foram divididos nos grupos de:

- Aplica: quando há aplicação direta e expressa das condições oriundas da Reforma Trabalhista;
- Admite Aplicação: quando há alegação de não aplicabilidade em razão da ocorrência do feito ser anterior à promulgação da Reforma Trabalhista. Infere-se, portanto, que haverá a Aplicação da Reforma Trabalhista para os novos processos, distribuídos a partir de 11/11/2017.

Os processos classificados como “Não Aplica” foram divididos nos grupos de:

- Inconstitucionalidade: quando há uma discordância e declaração formal de que o tema objeto deste estudo fere algum preceito ou ordenação constitucional;
- Outros motivos: quando não houve enquadramento nos grupos anteriores.

4 Conclusões

O presente artigo apresentou a dinâmica adotada para a resolução de um problema legal real com uso intensivo de técnicas de inteligência artificial.

Apesar do impedimento em divulgar os resultados específicos do projeto em termos do quantitativo de unidades judiciais que aplicaram ou não aplicaram os artigos da Reforma Trabalhista, esse artigo demonstrou com sucesso a aplicação de diversas técnicas derivadas da inteligência artificial, incluindo as constantes interações da computação com a matemática.

Ficou claro na abordagem técnica do projeto discorrido no artigo que a disponibilidade dos dados (neste caso, as decisões judiciais) é uma etapa crítica e obrigatória para o sucesso do projeto. Essa condição é acentuada no contexto de múltiplos sistemas de origem de dados públicos e altíssimo volume de documentos – e sua representativa massa de dados – a utilização do ferramental do *Big data* (serviços de *crawling*, armazenamento em nuvem, conversão de documentos em texto e bases de dados não relacionais).

Não existe inteligência artificial sem dados.

Outro aspecto interessante que ficou demonstrado por esse artigo é que, ainda que sejam usadas as melhores técnicas, ferramentas e profissionais envolvidos na disciplina de ciência de dados, o papel do especialista de negócio é um requisito imprescindível para o sucesso desse tipo de projeto. É esse profissional que, baseado em seu conhecimento aplicado e experiência, orienta o desenvolvimento (mais ainda no caso de análises inéditas como foi o caso da Reforma Trabalhista), valida os resultados e orienta a calibragem dos algoritmos.

Não existe inteligência artificial aplicada sem especialistas de negócios.

5 Referências

- AERY, Manu; RAMAMURTHY, Naveen; ASLANDOGAN, Y. Alp. Topic identification of textual data. **The University of Texas at Arlington**, Arlington, Texas, 2003.
- FRIEDMAN, Thomas L. **O mundo é plano: O mundo globalizado no século XXI**. 3. ed. atual. E amp. Rio de Janeiro: Objetiva, 2009.

MELO, Tiago Ribeiro Alves de; MARQUES, Shaiala Ribeiro de Castro Araujo; CADORE, Francine Athaide. Aplicação de técnicas de processamento de linguagem natural para análise da observância de artigos selecionados da Reforma Trabalhista (Lei n.º 13.467/2017). **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 58-69, 2019.

- Justiça em Números 2017. **Conselho Nacional de Justiça**, Brasília, 2017. Disponível em: www.cnj.jus.br/files/conteudo//2017/12/b60a659e5d5cb79337945c1dd137496c.pdf. Acesso em: 15 de abril de 2019.
- KAHNEMAN, D.; TVERSKY, A. **Subjective Probability**: A Judgment of Representativeness, *Cognitive Psychology*, no 3, 1972.
- MORENO, Isabel; ROMÁ-FERRI, María Teresa; MOREDA, Paloma. Combining Profiles and Local Information for Named Entity Classification: Adjustment of a Domain and Language Independent Approach. In: **Proceedings of the Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, LTC**. 2017. p. 17-19.
- PROVOST, Foster; FAWCETT, Tom. **Data Science for Business: What you need to know about data mining and data-analytic thinking**. " O'Reilly Media, Inc.", 2013.

MELO, Tiago Ribeiro Alves de; MARQUES, Shaiala Ribeiro de Castro Araujo; CADORE, Francine Athaide. Aplicação de técnicas de processamento de linguagem natural para análise da observância de artigos selecionados da Reforma Trabalhista (Lei n.º 13.467/2017). **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 58-69, 2019.