

CLASSIFICAÇÃO DE SENTENÇAS DE JUIZADO ESPECIAL CÍVEL UTILIZANDO APRENDIZADO DE MÁQUINA

CLASSIFICATION OF SPECIAL CIVIL COURT JUDGMENTS USING MACHINE LEARNING

Isabela Cristina Sabo¹, Thiago Raulino Dal Pont², Aires José Rover³, Jomi Fred Hübner⁴

Artigo aceito como convidado.

Resumo

O artigo expõe um estudo de caso realizado no Juizado Especial Cível da Universidade Federal de Santa Catarina (JEC/UFSC), tendo como objetivo empregar diferentes técnicas de Aprendizado de Máquina para classificar sentenças sobre falhas no serviço de transporte aéreo (Direito do Consumidor) em quatro classes: “procedente”, “parcialmente procedente”, “improcedente” e “extinção sem resolução do mérito”. Foram realizados dois experimentos, sendo um com o texto integral da sentença e outro com a retirada da parte dispositiva, isto é, do texto que representa o resultado processual (a classe). A acurácia obtida pelos classificadores no segundo experimento sofreu uma mínima redução. De maneira geral, os modelos obtidos pela Regressão Logística, pela RNA e pela Floresta Aleatória alcançaram maior desempenho para as classes “procedente”, “parcialmente procedente” e “improcedente”, enquanto que para classe “extinção”, cuja amostra é menor, o maior desempenho foi obtido pelo SVM com Kernel RBF. Utilizou-se a ferramenta Orange, versão 3.22, software de código aberto executado em Python.

Palavras-chave

Direito do Consumidor; Transporte Aéreo; Mineração de Texto; Classificação de Texto; Orange.

Abstract

The paper exposes a case study conducted at the Special Civil Court of the Federal University of Santa Catarina, aiming to employ different Machine Learning techniques to classify judgments on failures in the air transport service (Consumer Law) in four labels: “well

¹ Doutoranda em Direito pela Universidade Federal de Santa Catarina. Bolsista do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). E-mail: isabelasabo@gmail.com.

² Mestrando em Engenharia de Automação e Sistemas pela Universidade Federal de Santa Catarina. Bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). E-mail: thiagordalpont@gmail.com.

³ Doutor em Direito pela Universidade Federal de Santa Catarina. Professor Associado da mesma universidade, lotado no Departamento de Direito. E-mail: aires.rover@gmail.com.

⁴ Doutor em Engenharia Elétrica pela Universidade de São Paulo. Professor Associado da Universidade Federal de Santa Catarina, lotado no Departamento de Engenharia de Automação e Sistemas. E-mail: jomifred@gmail.com.

founded”, “partly founded”, “not well founded” and “without prejudice”. Two experiments were performed, one with the full text of the judgments and other removing the dispositive part, the text that represents the procedural result (the label). The accuracy obtained by the classifiers in the second experiment was minimally reduced. In general, the models obtained by Logistic Regression, ANN and Random Forest achieved higher performance for the “well founded”, “partly founded” and “not well founded” labels, while for the “without prejudice” label, whose sample is smaller, the highest performance was achieved by SVM with RBF Kernel. Orange, version 3.22, an open source software running on Python, was used as tool.

Keywords

Consumer Law; Air Transport; Text Mining; Text Classification; Orange.

1 Introdução

De acordo com o último relatório estatístico “Justiça em Números”, publicado anualmente pelo Conselho Nacional de Justiça (CNJ), o Poder Judiciário brasileiro finalizou o ano de 2018 com 78,7 milhões de processos em tramitação, isto é, aguardando solução definitiva. Em todo o Poder Judiciário, no ano de 2018, ingressaram cerca de 28,1 milhões de processos, dos quais aproximadamente 4,8 milhões foram nos Juizados Especiais Cíveis (JECs) (BRASIL, 2019).

Os JECs são órgãos integrantes do Poder Judiciário, regulamentados pela Lei n. 9.099/1995, aos quais buscam facilitar o acesso do cidadão à Justiça por meio de procedimentos mais simples e ausência de custos. Com isso, os JECs tendem a se aproximar dos problemas jurídicos do cidadão comum que se vê envolvido em conflitos diários de pequena expressão econômica, seja nas compras que efetua, nos serviços que contrata ou nos acidentes que sofre (WATANABE, 1985).

Diante disso, o artigo expõe um estudo de caso realizado no Juizado Especial Cível da Universidade Federal de Santa Catarina (JEC/UFSC), tendo como objetivo principal empregar técnicas de Aprendizado de Máquina para classificar sentenças sobre falhas no serviço de transporte aéreo (Direito do Consumidor) em quatro possíveis resultados: a) procedente; b) parcialmente procedente; c) improcedente e d) extinção sem resolução do mérito.

Para tanto, o artigo encontra-se estruturado da seguinte forma:

Seção 2: contém um breve referencial teórico, contextualizando o Aprendizado de Máquina e sua relação com a Mineração de Dados e de Texto, bem como expando trabalhos em que técnicas destes domínios foram aplicadas em ambientes de textos jurídicos.

Seção 3: contém metodologia utilizada, envolvendo desde a coleta das sentenças, a ferramenta utilizada, as técnicas de pré-processamento, de classificação e de validação, e ainda, a estratégia experimental adotada.

Seção 4: contém a apresentação e análise dos resultados obtidos.

Justifica-se a pesquisa pela atualidade do tema e pela menor incidência de trabalhos empíricos envolvendo aplicações de Aprendizado de Máquina no Poder Judiciário brasileiro.

2 Referencial teórico

Uma máquina aprende quando sua performance é aprimorada com a experiência. Especificamente, diz-se que um programa de computador aprende com a experiência E em relação a alguma classe de tarefas T e medida de desempenho P , se seu desempenho nas tarefas em T , medido por P , melhorar com a experiência E (MITCHELL, 1997).

Aprendizado de Máquina consiste em programar computadores para otimizar um critério de desempenho usando dados de exemplo ou de experiências anteriores. O aprendizado se faz necessário em casos nos quais não seja possível escrever diretamente um programa de computador para solucionar um determinado problema, porém, necessita de dados de exemplo ou de experiência. Um caso em que o aprendizado é necessário, por exemplo, é quando a perícia humana é inexistente ou quando os humanos sejam incapazes de explicar suas perícias (ALPAYDIN, 2009).

O aprendizado pode ocorrer de modo supervisionado ou não supervisionado. No aprendizado supervisionado ou direcionado busca-se inferir uma função ou relacionamento com base nos dados de treinamento rotulados, utilizando-se essa função para mapear novos dados não rotulados, o que demanda um número suficiente de exemplos para aprender. Já no aprendizado não supervisionado, busca-se descobrir padrões ocultos em dados não rotulados, com base no relacionamento entre pontos dos próprios dados (DESHPANDE; KOTU, 2019).

Uma das formas de aprendizado supervisionado se dá por meio da classificação dos dados, que consiste na tarefa de aprender uma função alvo f que mapeie cada conjunto de atributos x para um dos rótulos de classes y pré-determinados. Essa função alvo também é conhecida como modelo de classificação, o qual pode ser descritivo (para distinguir quais atributos definem o rótulo de uma classe) ou preditivo (para prever o rótulo não conhecido de uma classe). Exemplos de técnicas de classificação incluem Árvores de Decisão, Redes Neurais Artificiais, *Support Vector Machine*, *Naive Bayes*, entre outros (TAN; STEINBACH; KUMAR, 2009).

Na classificação de texto, o corpus é dividido em classes e são fornecidos exemplos de treinamento que associam pontos de dados a rótulos de classe. Por exemplo, textos de treinamento extraídos de um portal de notícias sobre assuntos políticos podem pertencer a um desses três rótulos: “senado”, “congresso” e “legislação”. Depois, para um determinado conjunto de testes, o objetivo será colocá-los em uma dessas categorias com o uso de um modelo supervisionado que foi construído usando os exemplos de treinamento (AGGARWAL, 2018).

Sobre a classificação de textos jurídicos, alguns trabalhos devem ser mencionados.

No contexto internacional, destaca-se o experimento realizado no Ministério da Justiça Italiano - o Justice Relationship Management (eJRM), que fornece apoio a cidadãos leigos, estudantes e mediadores para a obtenção de soluções jurídicas sobre um determinado caso.

SABO, Isabela Cristina; DAL PONT, Thiago Raulino; ROVER, Aires José; HÜBNER, Jomi Fred. Classificação de sentenças de Juizado Especial Cível utilizando aprendizado de máquina. **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 94-106, 2019.

O sistema foi alimentado por decisões judiciais que são recuperadas no momento em que o caso em língua natural é apresentado. Sua arquitetura é composta por 4 etapas: 1) Indexação: destina-se a armazenar as decisões judiciais em um banco de dados; 2) Mineração de núcleo: treina um modelo de classificação para prever o campo jurídico a que determinada descrição do caso apresentado pertence; 3) Processamento de consulta: extrai os termos relevantes da consulta que será usada pela mineração do núcleo para prever o campo jurídico do texto em disputa; 4) Ranking: recupera e classifica decisões judiciais relevantes (pertencentes ao campo jurídico previsto) a serem apresentadas aos disputantes e mediadores. Foram utilizados os seguintes algoritmos de classificação: Naive Bayes, Árvore de decisão, Support Vector Machine (SVM) Linear e com Kernel Gaussiano, todos com e sem PCA (Principal Component Analysis) para a redução de dimensionalidade do texto. A acurácia máxima obtida foi de 91,3%, por meio do SVM Linear com PCA (EL JELALI; FERSINI; MESSINA, 2015).

Outro experimento ocorreu na Corte Europeia de Direitos Humanos (CEDH), com o intuito de prever decisões. Os pedidos que nela tramitam versam sobre violações, por algum Estado-membro, dos direitos civis e políticos estabelecidos na Convenção Europeia de Direitos Humanos. O objetivo foi prever se um artigo específico da Convenção foi violado devido à evidência textual extraída de um caso, que compreende partes específicas referentes a “fatos”, a “lei aplicável” e os “argumentos” apresentados pelas partes envolvidas. Trata-se de uma classificação binária tendo como entrada os dados textuais extraídos dos casos e como saída o julgamento real deles sobre a violação ou não a determinado artigo da Convenção. Um banco de dados foi formado a partir de decisões relacionadas a 3 artigos da Convenção. Os resultados indicaram que a seção “fatos” de um caso é o fator preditivo mais importante, isto é, a tomada de decisão judicial é significativamente afetada pelo estímulo dos fatos narrados. Foi utilizado o algoritmo de classificação SVM Linear, obtendo-se a acurácia máxima de 79% (ALETRAS et al., 2016).

Finalmente, no contexto nacional, destaca-se o experimento realizado no Supremo Tribunal Federal (STF), conduzido pelo projeto VICTOR, da Universidade de Brasília (UnB), atualmente em andamento, que tem como objetivo resolver um problema de reconhecimento de padrões em textos dos processos judiciais que lá ingressam. Especificamente, o problema a ser resolvido é a classificação (vinculação) de processos em temas de repercussão geral (RG). A RG é um instrumento processual que atua como “filtro recursal”, permitindo ao STF selecionar os recursos que irá analisar de acordo com critérios de relevância jurídica, política, social ou econômica. O projeto foi composto por 2 etapas: 1) Classificação de 5 tipos de peças dentro de um processo (meta secundária): acórdão, recurso extraordinário (RE), agravo de recurso extraordinário (ARE), despacho e sentença; 2) Classificação de temas de repercussão geral (meta principal): após a classificação de peças, inicia-se a fase de classificação de temas contando com a automação da segmentação dos 5 tipos de peças importantes para a identificação dos temas de RG. A meta principal é relacionar um processo inteiro a um ou mais temas de RG. Obteve-se, na meta secundária, a acurácia máxima de 90,35%, utilizando Redes Neurais Convolucionais como algoritmo de classificação (SILVA, 2018; SILVA et al., 2018).

3 Metodologia de pesquisa

A metodologia de pesquisa pode ser visualizada a partir de três frentes: método de abordagem, procedimento e técnicas.

O método de abordagem consistiu no indutivo, uma vez que foi possível inferir conhecimento sobre os textos coletados. O procedimento adotado foi o estudo de caso que, como já exposto, diz respeito às sentenças sobre o consumo do serviço de transporte aéreo no âmbito do JEC/UFSC. E as técnicas, por sua vez, integraram desde a coleta, a preparação (pré-processamento) e a classificação por meio de diferentes algoritmos.

3.1 Coleta de dados

Foram coletadas 665 sentenças que versam especificamente sobre o assunto e que foram publicadas no período de 01/01/2014 a 03/06/2019. Os documentos foram obtidos por meio do sistema de processo eletrônico utilizado pelo JEC/UFSC, o “SAJ” (Sistema de Automação da Justiça).

Esleveu-se como análise os processos judiciais sobre falha na prestação do serviço de transporte aéreo pois, em contato com os servidores do JEC/UFSC, a equipe de pesquisa foi informada de que é a matéria atualmente mais discutida nos casos que tramitam no local.

As sentenças coletadas foram importadas em texto puro (“txt”) com codificação UTF-8.

3.2 Ferramenta

O experimento foi realizado por intermédio da ferramenta Orange, versão 3.22, software de código aberto desenvolvido por pesquisadores da Universidade de Liubliana, Eslovênia, que dispõe de componentes de Aprendizado de Máquina, Estatística, Mineração de Dados (incluindo Mineração de Texto) e Visualização. O seu uso se dá pela criação de fluxos de trabalho ou através da linguagem Python como biblioteca (DEMŠAR et al., 2013).

3.3 Pré-processamento

No caso em estudo fez-se necessário a preparação dos dados coletados, que é, via de regra, uma fase obrigatória e consiste em um conjunto de técnicas que inicializam os dados adequadamente para servir como entrada para um determinado algoritmo (GARCÍA; LUENGO; HERRERA, 2015). As sentenças foram submetidas às seguintes etapas de preparação:

- Transformação: Conversão ou consolidação dos dados para que o resultado da mineração seja mais eficiente (GARCÍA; LUENGO; HERRERA, 2015). No caso em estudo, todo o texto foi convertido em caixa baixa (*lowercase*).
- Tokenização: As sequências de palavras (*string*) são divididas em pedaços, utilizando um conjunto de delimitadores como espaço e pontuações (vírgulas, pontos etc.) (LEE et al., 1999). No caso em estudo, o critério utilizado para essa divisão foi o de expressões regulares (*regular expression* - “*regexp*”), isto é, padrões que associam sequências de caracteres no texto.

- Stemização (*stemming*): Redução das palavras variantes ao seu radical (*stem* - tronco). Isso é feito considerando que cada radical está semanticamente relacionado, de modo que ele não precisa ser uma palavra existente no dicionário (JIVANI, 2011). A técnica mais comum em mineração de texto é o método Porter, cujo objetivo básico é remover e/ou substituir o sufixo das palavras. Embora seja considerado extremamente eficiente pela literatura, é possível que o método cometa erros (DESHPANDE; KOTU, 2019), como no caso em estudo, a palavra “morais” foi reduzida a “moraí”, não existente no dicionário da língua portuguesa.
- Filtro: Remoção das palavras sem valor (*stopwords*), isto é, palavras que não contribuem para o contexto ou conteúdo do documento, como artigos, preposições, pronomes etc. Sua presença na mineração poderá apresentar obstáculos na compreensão do conteúdo do documento (KANNAN; GURUSAMY, 2014).
- *N-grams*: Agrupamento de palavras que normalmente surgem juntas. Trata-se, normalmente, da etapa final de pré-processamento (DESHPANDE; KOTU, 2019). No caso, o “n” foi definido em 2 (bigrams). Por exemplo, “dano moral”, “transporte aéreo”, “processo civil” etc.

Nas etapas de stemização e filtro foi necessário a especificação de um dicionário da língua portuguesa.

Essas etapas de pré-processamento têm como resultado a geração de uma *bag of words* (bolsa de palavras), na qual cada documento é representado como um vetor de palavras que ocorrem no documento. Nela é efetuada a contagem dos termos e o cálculo da frequência em que cada termo aparece no documento (MATSUBARA; MARTINS; MONARD, 2003).

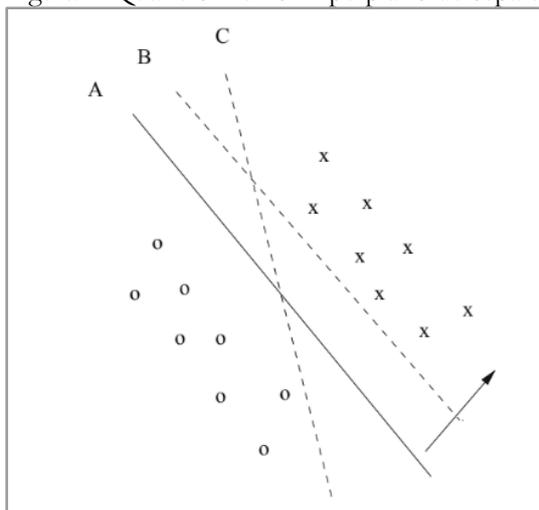
3.4 Classificação

Diferentes algoritmos de classificação foram utilizados neste trabalho, sendo eles: k Vizinhos Mais Próximos (k *Nearest Neighbors* - kNN), Máquina de Vetores de Suporte (*Support Vector Machine* - SVM) com função de base radial de Kernel (*Radial Basis Function* - Kernel RBF), Floresta Aleatória (*Random Forest*), Rede Neural Artificial (RNA), *Naive Bayes* e Regressão Logística.

O classificador kNN parte do princípio de que instâncias semelhantes possuem rótulos semelhantes. A ideia básica é identificar os vizinhos mais próximos de um ponto de teste e calcular o número de pontos que pertencem a cada classe. A classe com o maior número de pontos é indicada como relevante (AGGARWAL, 2018). Neste trabalho, o número de vizinhos foi definido como quatro, sendo a distância euclidiana (em linha reta) a métrica de dissimilaridade.

O objetivo do classificador SVM, por sua vez, é determinar o que pode melhor separar as diferentes classes. Por exemplo, na Figura 1, há duas classes indicadas por 'x' e 'o' e três hiperplanos de separação indicados por A, B e C. O hiperplano A fornece a melhor separação entre as diferentes classes, porque a distância normal de qualquer um dos pontos de dados é a maior (AGGARWAL, ZHAI, 2012).

Figura 1. Qual é o melhor hiperplano de separação?



Fonte: (AGGARWAL, ZHAI, 2012).

Quando as classes não podem ser separadas por uma reta como na Figura 1, o SVM não é suficiente para gerar bons resultados. Nesse caso, como solução é incluída ao SVM a função *Kernel*, que consiste em mapear os vetores de entrada em outras dimensões nas quais há separação linear entre os dados. A partir disso, novas formas de hiperplanos podem ser construídas (RUSSELL; NORVIG, 2009). Dentre as diversas funções existentes, utilizou-se neste trabalho a RBF, a qual abstrai o comportamento de uma função circular (CHAPELLE; HAFFNER; VAPNIK, 1999). O SVM normalmente apresenta bom desempenho em relação a dados textuais devido à natureza esparsa de alta dimensão do texto, na qual poucas características são irrelevantes, mas tendem a estar correlacionadas entre si (AGGARWAL, ZHAI, 2012).

O classificador *Naive Bayes* utiliza um modelo probabilístico que assume o corpus como uma mistura de diferentes classes. O processo, que é aplicado uma vez para cada documento observado, é o seguinte: a) selecione a r classe (componente da mistura) C_r com probabilidade anterior $\alpha_r = P(C_r)$; b) gere a classe do próximo documento a partir da distribuição de probabilidade para C_r (AGGARWAL, 2018). *Naive Bayes* talvez seja o classificador mais simples e também o mais usado, pois ele modela a distribuição dos documentos em cada classe com independência sobre a distribuição dos diferentes termos. Duas classes de modelos são comumente usadas para esse classificador (AGGARWAL, ZHAI, 2012).

As RNAs são modelos inspirados no sistema nervoso dos seres vivos. Elas têm a capacidade de adquirir e manter conhecimento (baseado em informações) e podem ser definidas como um conjunto de unidades de processamento, representadas por neurônios artificiais, interligadas por várias interconexões (sinapses artificiais) e implementadas por vetores e matrizes de pesos sinápticos (SILVA et al., 2017). Assim, as RNAs tentam simular o comportamento das redes de neurônios, como a capacidade de aprendizado e abstração, o reconhecimento de padrões, entre outros (TAN; STEINBACH; KUMAR, 2009).

Neste trabalho utilizou-se uma RNA do tipo *Multi Layer Perceptron*, que é composta por uma camada de neurônios de entrada, outra de saída, e uma ou mais camadas intermediárias. O

número de neurônios das camadas de entrada e de saída são iguais às entradas e saídas. No caso, foram utilizadas 114.310 neurônios na camada de entrada e apenas um na camada de saída. O alto número de neurônios de entrada decorre do uso de *n-grams*. Assim, além de todas as palavras contidas nas sentenças, inclui-se as combinações em pares destas palavras na ordem em que ocorrem nos textos. Por fim, a RNA contou com duas camadas intermediárias com 100 e 50 neurônios, respectivamente, ambas com função de ativação tangente hiperbólica (TAN; STEINBACH; KUMAR, 2009).

Florestas Aleatórias são implementações de conjuntos de árvores de decisão, conhecidas por serem altamente robustas e precisas. Uma árvore de decisão, por sua vez, é uma divisão hierárquica do espaço de dados. A ideia é dividi-lo em regiões de atributos fortemente influenciadas por uma classe específica durante a fase de treinamento. Portanto, as divisões estão associadas aos seus principais rótulos de classe (AGGARWAL, 2018).

Por fim, Regressão Logística é uma forma especializada de regressão, que consiste em uma técnica estatística, utilizada para prever e explicar uma variável categórica binária, isto é, duas classes, podendo ser adaptado para problemas com múltiplas classes. O objetivo principal é identificar a classe a qual pertence um objeto através da investigação de uma relação de dependência (HAIR JR. et al., 2009).

3.5 Treinamento e teste

O objetivo deste trabalho consiste em classificar as sentenças por meio das técnicas expostas em quatro classes, de acordo com os resultados processuais possíveis: a) procedente (acolhimento total do pedido do consumidor); b) parcialmente procedente (acolhimento parcial do pedido do consumidor); c) improcedente (rejeição do pedido do consumidor) e d) extinção sem resolução do mérito (acolhimento das preliminares indicadas no art. 337 do Código de Processo Civil⁵).

A base de dados de treinamento contou com um total de 665 documentos de sentenças, abrangendo as quatro classes. Entretanto, algumas destas sentenças apresentaram mais de um resultado, ou seja, contêm mais de uma classe. Nesses casos, as sentenças foram replicadas para cada uma das classes indicadas em seu dispositivo, resultando em um total de 673 sentenças para o treinamento. Assim, para a classe “procedente”, foram incluídas 214

⁵ Art. 337. Incumbe ao réu, antes de discutir o mérito, alegar: I - inexistência ou nulidade da citação; II - incompetência absoluta e relativa; III - incorreção do valor da causa; IV - inépcia da petição inicial; V - perempção; VI - litispendência; VII - coisa julgada; VIII - conexão; IX - incapacidade da parte, defeito de representação ou falta de autorização; X - convenção de arbitragem; XI - ausência de legitimidade ou de interesse processual; XII - falta de caução ou de outra prestação que a lei exige como preliminar; XIII - indevida concessão do benefício de gratuidade de justiça.

§ 1º Verifica-se a litispendência ou a coisa julgada quando se reproduz ação anteriormente ajuizada.

§ 2º Uma ação é idêntica a outra quando possui as mesmas partes, a mesma causa de pedir e o mesmo pedido.

§ 3º Há litispendência quando se repete ação que está em curso.

§ 4º Há coisa julgada quando se repete ação que já foi decidida por decisão transitada em julgado.

§ 5º Excetuadas a convenção de arbitragem e a incompetência relativa, o juiz conhecerá de ofício das matérias enumeradas neste artigo.

§ 6º A ausência de alegação da existência de convenção de arbitragem, na forma prevista neste Capítulo, implica aceitação da jurisdição estatal e renúncia ao juízo arbitral.

sentenças; para classe “improcedente”, 70 sentenças; para classe “parcialmente procedente”, 379 sentenças; e para a classe “extinção”, 10 sentenças.

Foram realizados dois experimentos, sendo um com o texto integral da sentença e outro com a retirada da parte dispositiva⁶, isto é, do texto que representa o resultado processual. Essa estratégia foi adotada para possibilitar uma comparação do desempenho das técnicas onde a classe é indicada no corpus textual e onde a classe não é indicada no corpus textual.

Para cada uma das quatro classes foram utilizados os seis classificadores indicados. O método de amostragem para criação dos conjuntos de treino e teste foi a validação cruzada (*cross validation*). Tal técnica divide a base de dados em k conjuntos disjuntos com tamanhos aproximadamente iguais. A partir daí, a cada etapa do treinamento, um conjunto é utilizado para teste e os demais como treino, de tal maneira que, cada conjunto criado será utilizado tanto para treino quanto para teste (WONG, 2015). Neste trabalho foi utilizado um total de 10 divisões. Como resultado de cada rodada na validação cruzada, será verificada a medida de acurácia para as quatro classes. Tal métrica indica a proporção de acertos em relação ao total de classificações feitas. Ao final, a avaliação se dará através da média das k acurácias calculadas para cada classe, respectivamente (WONG, 2015). Os resultados serão discutidos adiante.

4 Resultados e discussão

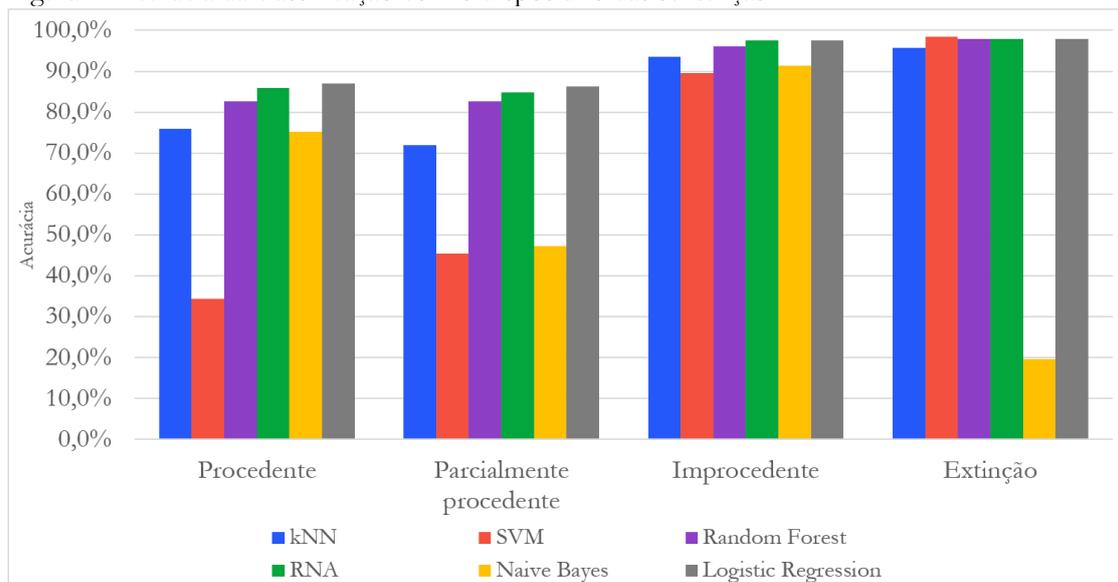
O Quadro 1 contém a acurácia obtida por cada algoritmo em cada uma das quatro classes utilizando o texto integral das sentenças. De um lado, observa-se o maior desempenho da Regressão Logística, seguido da RNA, para as classes “procedente”, “parcialmente procedente” e “improcedente”. De outro, observa-se um desempenho significativamente menor do SVM com *Kernel RBF* para as classes “procedente” e “parcialmente procedente”, e do *Naive Bayes* para a classe “parcialmente procedente”. Todavia, para a classe “extinção”, cuja amostra é menor, o SVM com *Kernel RBF* atingiu o maior desempenho, ao passo que o *Naive Bayes* atingiu significativamente o menor. A acurácia alcançada por cada classificador pode ser melhor comparada a partir da Figura 2.

Quadro 1. Acurácia da classificação com o dispositivo das sentenças

Classe	kNN	SVM <i>Kernel RBF</i>	<i>Random Forest</i>	RNA	<i>Naive Bayes</i>	<i>Logistic Regression</i>
Procedente	75,9%	34,3%	82,6%	85,9%	75,2%	87,1%
Parcialmente procedente	71,9%	45,5%	82,6%	84,9%	47,3%	86,2%
Improcedente	93,5%	89,6%	96,1%	97,5%	91,4%	97,6%
Extinção	95,7%	98,5%	97,9%	97,9%	19,6%	97,9%

⁶ Art. 489. São elementos essenciais da sentença: I - o relatório, que conterá os nomes das partes, a identificação do caso, com a suma do pedido e da contestação, e o registro das principais ocorrências havidas no andamento do processo; II - os fundamentos, em que o juiz analisará as questões de fato e de direito; III - o dispositivo, em que o juiz resolverá as questões principais que as partes lhe submeterem.

Figura 2. Acurácia da classificação com o dispositivo das sentenças

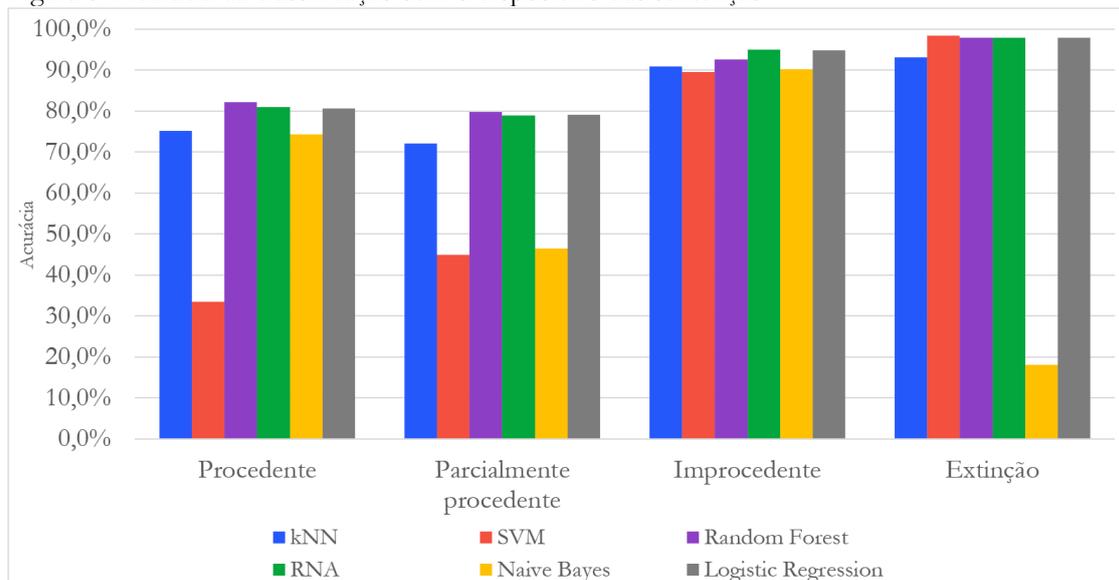


O Quadro 2 contém a acurácia obtida por cada algoritmo em cada uma das quatro classes utilizando o texto das sentenças com a retirada do dispositivo (resultado do processo). De um lado, observa-se o maior desempenho da Floresta Aleatória, seguido da RNA e da Regressão Logística, para as classes “procedente”, “parcialmente procedente” e “improcedente”. De outro, observa-se um desempenho significativamente menor do SVM com *Kernel RBF* para as classes “procedente” e “parcialmente procedente”, e do *Naive Bayes* para a classe “parcialmente procedente”. Todavia, para a classe “extinção”, cuja amostra é menor, novamente o SVM com *Kernel RBF* atingiu o maior desempenho, ao passo que o *Naive Bayes* obteve resultado significativamente menor. A acurácia alcançada por cada classificador pode ser melhor comparada a partir da Figura 3.

Quadro 2. Acurácia da classificação sem o dispositivo das sentenças

Classe	kNN	SVM <i>Kernel RBF</i>	<i>Random Forest</i>	RNA	<i>Naive Bayes</i>	<i>Logistic Regression</i>
Procedente	75,2%	33,4%	82,2%	81,0%	74,3%	80,7%
Parcialmente procedente	72,1%	45,0%	79,9%	79,0%	46,5%	79,2%
Improcedente	90,9%	89,6%	92,7%	95,1%	90,2%	94,9%
Extinção	93,2%	98,5%	97,9%	97,9%	18,1%	97,9%

Figura 3. Acurácia da classificação sem o dispositivo das sentenças



O Quadro 3 apresenta a diferença da acurácia entre os experimentos, indicando uma média das acurácias obtidas por cada classificador nas quatro classes. De maneira geral, infere-se pelas médias que a acurácia no experimento com a retirada do resultado processual (parte do texto em que se indica a classe a qual ele pertence) sofreu uma redução mínima, o que demonstra que os classificadores mantiveram seu desempenho apenas com o texto no qual se relata os fatos narrados pelas partes no processo e os fundamentos jurídicos aplicáveis ao caso. Isso se torna pressuposto para a realização de experimentos com textos de processos judiciais nos quais ainda não haja sentença, ou seja, nos quais o juiz ainda não tenha decidido o resultado.

Observa-se, ainda, que a maior diferença se obteve com a Regressão Logística para a classe “parcialmente procedente”, cuja amostra é maior, enquanto que a diferença obtida pelo kNN para a mesma classe foi negativa, o que significa que esse classificador se desempenhou melhor no texto das sentenças sem o dispositivo.

Quadro 3. Diferença da acurácia entre os Quadros 1 e 2

Classe	kNN	<i>SVM</i> <i>Kernel RBF</i>	<i>Random</i> <i>Forest</i>	RNA	<i>Naive</i> <i>Bayes</i>	<i>Logistic</i> <i>Regression</i>
Procedente	0,7%	0,9%	0,4%	4,9%	0,9%	6,4%
Parcialmente procedente	-0,2%	0,5%	2,7%	5,9%	0,8%	7,0%
Improcedente	2,6%	0,0%	3,4%	2,4%	1,2%	2,7%
Extinção	2,5%	0,0%	0,0%	0,0%	1,5%	0,0%
Média	1,4%	0,4%	1,6%	3,3%	1,1%	4,0%

5 Conclusões

O artigo em questão tratou de experimentos iniciais realizados nas sentenças do JEC/UFSC, indicando-se apenas quatro possíveis classes a que pertencem os textos. De modo geral, foi possível obter um panorama de como as diversas técnicas de Aprendizado de Máquina se comportam diante dos textos jurídicos (específicos sobre Direito do Consumidor e falha no serviço de transporte aéreo), avaliando-se quais modelos de classificação alcançaram maior e menor desempenho.

Como trabalhos futuros almeja-se aprimorar a representação do texto da *bag of words* para uma representação mais robusta, como a *Word Embeddings*, capaz de associar, numericamente, contextos e semântica às palavras. Além disso, pretende-se construir uma nova estrutura para predição de sentenças, composta de classificadores em cascata, utilizando atributos extraídos dos textos por meio de clusterização. Assim, permite-se fazer classificações intermediárias que servirão de base para uma classificação final.

Agradecimentos

À CAPES e ao CNPq por financiarem parte desta pesquisa, e à Dra. Vânia Petermann, juíza responsável pelo Juizado Especial Cível da Universidade Federal de Santa Catarina, por facilitar o acesso os dados utilizados.

Referências

- AGGARWAL, Charu C. **Machine learning for text**. Cham, Switzerland: Springer International Publishing, 2018.
- AGGARWAL, Charu C.; ZHAI, ChengXiang. A survey of text classification algorithms. In: AGGARWAL, Charu C.; ZHAI, ChengXiang (Eds). **Mining text data**. New York: Springer Science, 2012. p. 163-222.
- ALETRAS, Nikolaos; TSARAPATSANIS, Dimitrios; PREOȚIUC-PIETRO, Daniel; LAMPOS, Vasileios. Predicting judicial decisions of the European Court of Human Rights: a natural language processing perspective. **PeerJ Computer Science**, v. 2, n. 93, p. 1-19, 2016.
- ALPAYDIN, Ethem. **Introduction to machine learning**. 2. ed. Cambridge, MA: MIT Press, 2009.
- BRASIL. Conselho Nacional de Justiça (CNJ). **Justiça em Números 2019**: ano-base 2019. Brasília: CNJ, 2019a.
- CHAPELLE, Olivier; HAFFNER, Patrick.; VAPNIK, Vladimir N.. Support vector machines for histogram-based image classification. **Institute Of Electrical And Electronics Engineers (IEEE)**. [s.l.], p. 1055-1064. set. 1999.
- DEMŠAR, Janez; CURK, Tomaž; ERJAVEC, Aleš; GORUP, Črt; HOČEVAR, Tomaž; MILUTINVIČ, Mitar; MOŽINA, Martin; POLAJNAR, Matija; TOPLAK, Marko; STARIČ, Anže; ŠTAJDOHAR, Miha; UMEK, Lan; ŽAGAR, Lan; ŽBONTAR, Jure; ŽITNIK, Marinka; ZUPAN, Blaž. Orange: Data Mining Toolbox in Python. **Journal of Machine Learning Research**, n. 14, p. 2349-2353, aug. 2013.
- DESHPANDE, Bala; KOTU, Vijay. **Data science: concepts and practice**. 2. ed. Cambridge: Morgan Kaufmann (Elsevier), 2019.
- EL JELALI, Soufiane; FERSINI, Elisabetta; MESSINA, Enza. **Legal retrieval as support to eMediation: matching disputant's case and court decisions**. Artificial Intelligence and Law, v. 23, n. 1, p. 1-22, mar. 2015.
- GARCÍA, Salvador; LUENGO, Julián; HERRERA, Francisco. **Data preprocessing in data mining**. Springer International Publishing, 2015.

SABO, Isabela Cristina; DAL PONT, Thiago Raulino; ROVER, Aires José; HÜBNER, Jomi Fred. Classificação de sentenças de Juizado Especial Cível utilizando aprendizado de máquina. **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 94-106, 2019.

- HAIR JR., Joseph F.; BLACK, William C.; BABIN, Barry J.; ANDERSON, Rolph E.; TATHAM, Ronald L. **Análise multivariada de dados**. Tradução de Adonai Schlup Sant'Anna. 6. ed. Porto Alegre: Bookman, 2009.
- JIVANI, Anjali Ganesh. A comparative study of stemming algorithms. **Int. J. Comp. Tech. Appl.**, v. 2, n. 6, p. 1930-1938, 2011.
- KANNAN, Subbu; GURUSAMY, Vairaprakash. Preprocessing techniques for text mining. In: **Proceedings of Recent Trends and Research Issues in Computer Science (RTRICS)**, oct. 2014.
- LEE, Mong-Li; LING, Tok Wangng; LU, Hongjun; KO, Yee Teng. Cleansing data for mining and warehousing. In: **Proceedings of the 10th International Conference on Database and Expert Systems Applications (DEXA)**, Florence (Italy), p. 751-760, aug./set. 1999.
- MATSUBARA, Edson Takashi; MARTINS, Claudia Aparecida; MONARD, Maria Carolina. PreTexT: uma ferramenta para pré-processamento de textos utilizando a abordagem bag-of-words. **Relatórios Técnicos do ICMC**, São Carlos, n. 209, ago. 2003.
- MITCHELL, Tom M. **Machine Learning**. Ithaca, NY: McGraw-Hill, 1997.
- RUSSELL, Stuart J.; NORVIG, Peter. **Artificial Intelligence: A Modern Approach**. 3. ed. Upper Saddle River, NJ, USA: Prentice Hall Press, 2009.
- SILVA, Ivan Nunes da; SPATTI, Danilo Hernane; FLAUZINO, Rogerio Andrade; LIBONI, Luisa Helena Bartocci; ALVES, Silas Franco dos Reis. **Artificial neural networks: a practical course**. Switzerland: Springer International Publishing, 2017.
- SILVA, Nilton Correia da. Notas iniciais sobre a evolução dos algoritmos do VICTOR: o primeiro projeto de inteligência artificial em supremas cortes do mundo. In: FERNANDES, Ricardo Vieira de Carvalho; CARVALHO, Angelo Gamba Prata de Carvalho (Coord.). **Tecnologia jurídica & direito digital: II Congresso Internacional de Direito, Governo e Tecnologia – 2018**. Belo Horizonte: Fórum, 2018. p. 89-94.
- SILVA, Nilton Correia da; BRAZ, Fabricio Ataides; CAMPOS, Teófilo Emídio de; GUEDES, André Bernardes Soares; MENDES, Danilo Barros; BEZERRA, Davi Alves; GUSMÃO, Davi Benevides; CHAVES, Felipe Borges de Souza; ZIEGLER, Gabriel Gomes; HORINOUCI, Lucas Hiroshi; FERREIRA, Marcelo Hertton Pereira; INAZAWA, Pedro Henrique Gonçalves; COELHO, Victor Hugo Dias; FERNANDES, Ricardo Vieira de Carvalho; PEIXOTO, Fabiano Hartmann; MAIA FILHO, Mamede Said; SUKIENNIK, Bernardo Pablo; ROSA, Lahis da Silva; SILVA, Roberta Zumblick Martins da; JUNQUILHO, Tainá Aguiar; CARVALHO, Gustavo H. T. A. Document type classification for Brazil's supreme court using a Convolutional Neural Network. In: **Proceedings of 10th International Conference on Forensic Computer Science and Cyber Law (ICoFCS)**, São Paulo, Brazil, oct. 2018.
- TAN, Pang-ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao data mining: mineração de dados**. Rio de Janeiro: Ciência Moderna Ltda., 2009.
- WATANABE, Kazuo. Filosofia e características básicas do Juizado Especial de Pequenas Causas. In: WATANABE, Kazuo (Coord.). **Juizado Especial de Pequenas Causas: lei n. 7.244/1984**. São Paulo: Revista dos Tribunais, 1985.
- WONG, Tzu-tsung. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. **Pattern Recognition**. [s.l.], p. 2839-2846. set. 2015.