

**INSTITUTO SERZEDELLO CORRÊA**  
**PÓS-GRADUAÇÃO EM ANÁLISE DE DADOS PARA O CONTROLE**

**RELATÓRIO FINAL DE PROJETO**

**CLASSIFICAÇÃO AUTOMÁTICA DE ACÓRDÃOS**  
**A PARTIR DE ENUNCIADOS**  
**DA JURISPRUDÊNCIA SELECIONADA**

TRABALHO FINAL DA DISCIPLINA: **TÉCNICAS DE MINERAÇÃO DE DADOS**

PROFESSOR: **WAGNER MEIRA JUNIOR**

ALUNO: **LEONARDO AUGUSTO DA SILVA PACHECO**

# 1. Motivação e Contexto

## 1.1. Descrição do problema

O Tribunal de Contas da União (TCU) disponibiliza suas decisões para o público por meio do Sistema de Pesquisa de Jurisprudência, localizada no endereço <https://contas.tcu.gov.br/pesquisaJurisprudencia/#/>. O sistema é composto por 4 bases principais de pesquisa:

- **Acórdãos:** pesquisa no inteiro teor (relatório, voto e acórdão) de acórdãos e decisões;
- **Jurisprudência selecionada:** pesquisa no enunciado elaborado a partir de deliberações selecionadas sob o critério de relevância jurisprudencial;
- **Publicações:** pesquisa em boletins elaborados (Jurisprudência, Pessoal, Licitações e contratos);
- **Súmulas:** pesquisa nos enunciados de súmulas.

A única pesquisa completa, que abrange todas as decisões colegiadas, é a pesquisa do inteiro teor. Porém, como o texto do inteiro teor, em geral é um texto muito grande, sua pesquisa se torna um tanto imprecisa. Isso cria dificuldade para que o usuário consiga encontrar facilmente o que procura ao realizar uma pesquisa, a não ser que providencie um termo de busca extremamente preciso e elaborado, o que não está ao alcance do usuário comum. Há exemplos de inteiros teores com mais de trezentas páginas (<https://contas.tcu.gov.br/sagas/SvlVisualizarRelVotoAcRtf?codFiltro=SAGAS-SESSAO-ENCERRADA&seOcultaPagina=S&item0=603589>).

Em outros tribunais superiores, para cada acórdão, uma equipe de jurisprudência elabora uma ementa, que resume as informações consideradas mais relevantes em cada decisão, o tema, um resumo da decisão tomada, legislação empregada, termos para indexação outras decisões citadas, entre outras. Uma pesquisa baseada apenas nos textos das ementas, indexadores e outras informações preparadas traria provavelmente maior precisão nas pesquisas realizadas. Podemos visualizar o exemplo de um acórdão do STF na figura 1, a seguir.

A elaboração de ementa, por outro lado, exigiria um trabalho de jurisprudência muito maior do que temos hoje, exigindo maior equipe, maior suporte de ferramentas e maior tempo de elaboração. Outra dificuldade é que a elaboração da ementa poderia atrasar a produção dos relatórios, acórdãos e publicação. Hoje, tão logo assinado, um acórdão pode ser publicado imediatamente, facilitando o acesso às partes e à sociedade. A elaboração da ementa traria o risco de que esse prazo seja ampliado. Conforme podemos visualizar no gráfico (figura 2), a quantidade de decisões do Tribunal tem crescido ano a ano (a linha de tendência traçada é uma parábola), o que parece tornar a elaboração de ementa uma possibilidade cada vez mais remota.

Figura 1 - Exemplo de acórdão publicado no STF.

## Pesquisa de Jurisprudência



### Acórdãos

Documentos encontrados: 684

(1 / 69) pági

<< | < | > | >> | Nova Pesquisa

1 | 2 | 3 | 4 | 5 | 6 | 7 | Próximo

Expressão de busca: (PRISAO EM SEGUNDA INSTANCIA)

Acompanhamento Processual | Inteiro Teor | DJ/DJe | Ementa sem Formatação



**HC 163718 AgR / SP - SÃO PAULO**

**AG.REG. NO HABEAS CORPUS**

**Relator(a): Min. LUIZ FUX**

**Julgamento: 30/11/2018**

**Órgão Julgador: Primeira Turma**

#### Publicação

#### PROCESSO ELETRÔNICO

DJe-264 DIVULG 07-12-2018 PUBLIC 10-12-2018

#### Parte(s)

AGTE.(S) : ANGELO APARECIDO SIVIERO

ADV.(A/S) : HELIO DA SILVA SANCHES

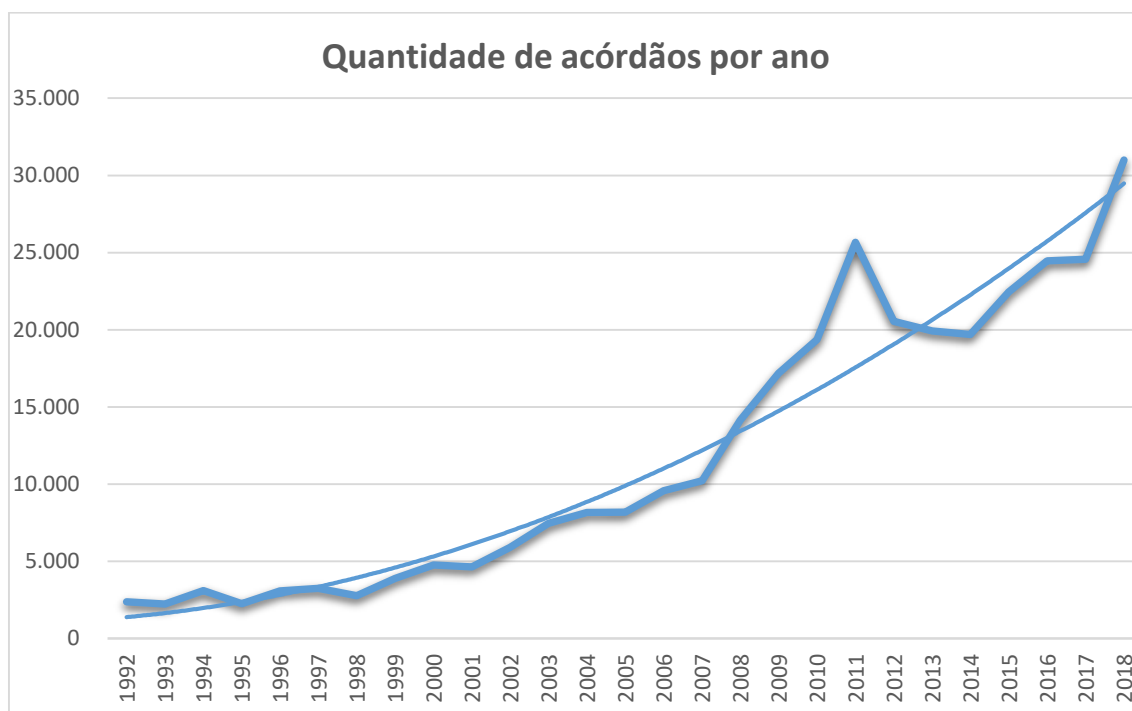
AGDO.(A/S) : RELATOR DO HC Nº 473.576 DO SUPERIOR TRIBUNAL DE JUSTIÇA

#### Ementa

**Ementa:** AGRAVO REGIMENTAL NO HABEAS CORPUS. PENAL E PROCESSO PENAL. CRIMES DE TRÁFICO ILÍCITO DE ENTORPECENTES, ASSOCIAÇÃO PARA O TRÁFICO E PORTE ILEGAL DE ARMA DE FOGO DE USO RESTRITO. ARTIGOS 33 E 35 DA LEI 11.343/06 E ARTIGO 16 DA LEI 10.826/03. ALEGADA NULIDADE PROCESSUAL. REDISCUSSÃO DOS CRITÉRIOS DE DOSIMETRIA DA PENA. PLEITO DE REVOGAÇÃO DA CUSTÓDIA CAUTELAR. ALEGADA NULIDADE PROCESSUAL. TEMAS NÃO DEBATIDOS PELAS **INSTÂNCIAS** PRECEDENTES. SUPRESSÃO DE **INSTÂNCIAS**. IMPOSSIBILIDADE. INEXISTÊNCIA DE ABUSO DE PODER OU FLAGRANTE ILEGALIDADE. REVOLVIMENTO DO CONJUNTO FÁTICO-PROBATÓRIO. INADMISSIBILIDADE NA VIA ELEITA. AGRAVO REGIMENTAL DESPROVIDO. 1. Inexiste excepcionalidade que permita a concessão da ordem de ofício ante a ausência de flagrante ilegalidade ou abuso de poder na decisão da Corte Superior que assentou ser inviável a análise do mérito da questão recorrida, sob pena de supressão de **instâncias**. 2. In casu, o recorrente foi condenado, pelo juízo natural, à pena de 15 (quinze) anos e 04 (quatro) meses de reclusão, **em** regime inicial fechado, pela prática dos crimes previstos nos artigos 33 e 35 da Lei 11.343/06 e 16 da Lei 10.826/03. 3. A dosimetria da pena, bem como os

Fonte: <http://stf.jus.br/portal/jurisprudencia/pesquisarJurisprudencia.asp>

Figura 2 - Evolução da quantidade de acórdãos nos últimos anos



Fonte: <https://pesquisa.apps.tcu.gov.br/#/pesquisa/acordao-completo>

O Tribunal vem adotando uma solução intermediária, que é a jurisprudência selecionada. A Secretaria das Sessões seleciona não todos, mas alguns acórdãos, após a decisão, a partir do qual geram um texto de enunciado, classificação em área, tema e subtema, outros termos indexadores e outros metadados. Com isso, possibilita uma pesquisa temática mais precisa, sobre a uma base de acórdãos restrita, que permite um entendimento aproximado jurisprudencial do Tribunal acerca de um determinado tema, e indica alguns acórdãos que demonstram aquela linha de entendimento.

Não sendo uma pesquisa completa, reduz os custos e riscos mencionados, mas possui limitações. Não conseguimos, para um acórdão qualquer, achar um resumo, entender quais são os assuntos dos quais ele trata. Além disso, ao realizar determinadas pesquisas, o usuário comumente encontra dificuldades em encontrar esse acórdão que seria relevante, conforme é demonstrado pelo registro dos logs do sistema de pesquisa.

## 1.2. Processo de geração do acórdão

A origem do acórdão é uma auditoria, monitoramento ou outro trabalho realizado pelo corpo técnico (Secretaria do Tribunal), cujo resultado é um relatório, que explica a situação analisada e traz uma proposta de encaminhamento da decisão do Tribunal àquele respeito. É autuado em um processo pelo sistema **e-TCU**, e geralmente assinado pelo auditor ou equipe de auditoria, diretor e secretário.

É designado um ministro relator, que produz um relatório final a partir do relatório técnico (comumente o teor é o mesmo), um voto e um rascunho do acórdão (decisão). O ministro revisor poderá seguir o voto do relator, emitir voto contrário ou complementar. Os ministros, reunidos em uma sessão do Tribunal, discutem, votam e produzem a decisão (o acórdão). O sistema **Sagas** faz todo o gerenciamento desde o momento de designação do relator do processo, passando pela sessão até a assinatura do acórdão.

Uma vez assinado e disponibilizado o acórdão no Sagas, o inteiro teor, composto pelo relatório, voto e acórdão, é publicado pelo **Sistema de Pesquisa de Jurisprudência**.

Ao final, a ata da sessão também é publicada, e ocorre o monitoramento das decisões pelo sistema **Radar**. Além disso, o acórdão pode ser objeto de recurso.

### **1.3. Processo de geração da jurisprudência selecionada**

Dentro da Secretaria das sessões, existe a Diretoria de Jurisprudência, responsável pela elaboração da jurisprudência selecionada. O trabalho ocorre utilizando o sistema **e-Juris**.

Primeiramente, são selecionados acórdãos que denotam linhas de entendimento do Tribunal acerca de diversos assuntos. Por exemplo, em 2017, foram proferidos 24.558 acórdãos, dos quais menos de mil foram selecionados. Do inteiro teor são extraídos excertos, ou seja, parágrafos e trechos do texto dos quais melhor se pode depreender o entendimento. Em seguida, produz-se um enunciado, que é o resumo do entendimento realizado. Foram 993 enunciados produzidos a partir de acórdãos de 2017. É possível, de um único acórdão, extrair mais de um enunciado.

O enunciado de jurisprudência é acompanhado de uma série de termos de indexação, organizados em Área, Tema e subtema. Tais termos pertencem a um vocabulário controlado, o VCE (Vocabulário de Controle Externo), produzido pela área de gestão documental do TCU. Além disso, são registradas a legislação empregada e outras informações relevantes.

### **1.4. Motivação**

Trazendo o conhecimento gerado pelo trabalho da jurisprudência selecionada para a pesquisa de acórdãos, poderemos acrescentar informações relevantes, facilitar o processo de busca e a associação de acórdãos afins, provendo, enfim, uma pesquisa mais útil para o usuário.

O próprio processo de produção da jurisprudência selecionada poderia ser facilitado a partir do processamento de decisões pretéritas, com o preenchimento automatizado de informações ou sugestões.

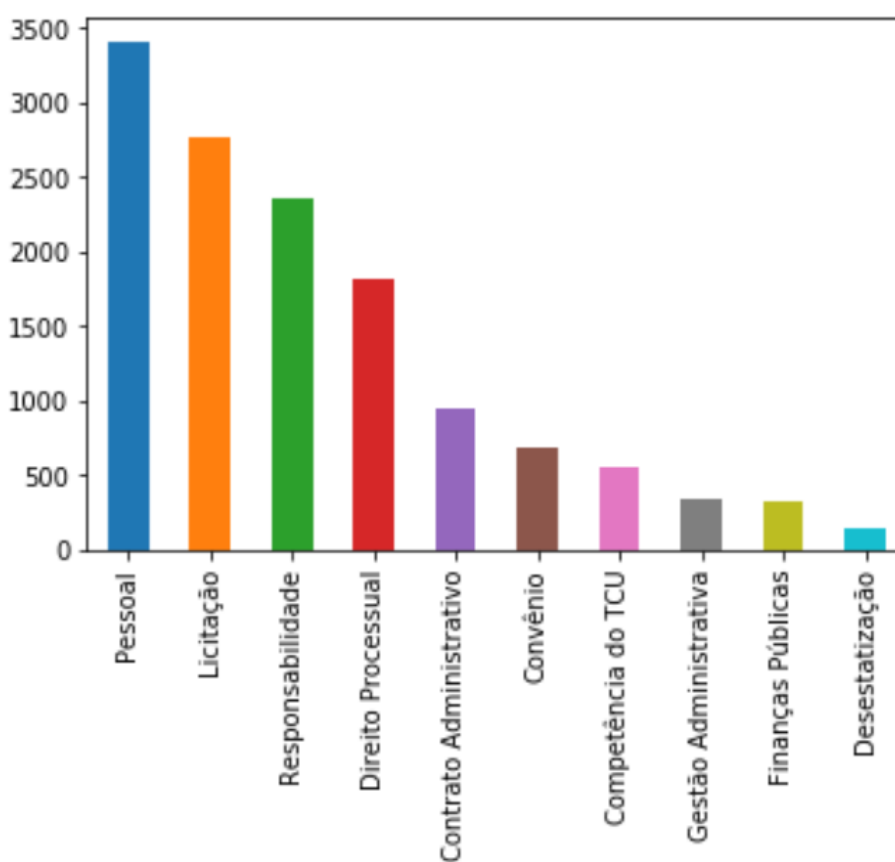
Finalmente, a utilização desses termos na jurisprudência pode trazer informações práticas de relacionamento de termos que podem retroalimentar o Thesaurus do TCU, o VCE.

## 2. Descrição

O presente projeto visa verificar a viabilidade de indexação e classificação automática de enunciados de jurisprudência e acórdãos, a partir de textos produzidos ou extraídos e metadados registrados nos sistemas de pesquisa de Jurisprudência, de processos (e-TCU) e do sistema de Jurisprudência (e-Juris).

Especificamente, queremos produzir a classificação de área. Atualmente, os enunciados são distribuídos em 10 áreas, conforme o gráfico da figura 3.

*Figura 3 - Quantidade de enunciados por área de jurisprudência*



A opção mais simples é extrair a partir do enunciado, que já contém a síntese do que deseja exprimir como entendimento jurisprudencial. Desejamos verificar a possibilidade de obtenção da área a partir do excerto do inteiro teor, que é uma informação menos tratada, ou diretamente do inteiro teor. Uma dificuldade enfrentada é a cardinalidade entre acórdãos e enunciados de jurisprudência, demonstrada na tabela abaixo: como de um acórdão pode ser extraído mais de um enunciado, cada qual com sua área, um mesmo acórdão pode ter mais de uma área de jurisprudência associada.

*Tabela 1 - Enunciados de jurisprudência por acórdão*

<b>Enunciados</b>	<b>Quantidade</b>	<b>Parte</b>	<b>Acumulado</b>
1	7.236	82,8%	82,8%
2	1.113	12,7%	95,5%
3	230	2,63%	98,2%
4	89	1,02%	99,2%
5	39	0,45%	99,6%
6	12	0,14%	99,8%
7	7	0,08%	99,9%
8	4	0,05%	99,9%
9	2	0,02%	99,9%
10	1	0,01%	99,9%
11	2	0,02%	100,0%
12	1	0,01%	100,0%
13	2	0,02%	100,0%

Outras variações são verificadas, como a possibilidade de utilização de informação do tipo de processo (assunto), a possibilidade de utilização apenas do excerto do voto, e a consideração apenas dos acórdãos com um enunciado associado.

Além da área, são testadas as classificações em temas e subtemas, e a sugestão de outros termos indexadores.

As informações do acórdão são fornecidas pelos sistemas Sagas e e-TCU. O sistema e-Juris fornece as informações dos enunciados de jurisprudência e sua associação com os acórdãos. O sistema de pesquisa de Jurisprudência também trata essas informações, sendo uma possível fonte secundária de informações.

As informações produzidas e manipuladas por meio de tais sistemas estão disponíveis na base de dados corporativa de produção (Oracle), e os textos estão contidos em documentos, armazenados em um sistema de Gestão Eletrônica de Documentos (GED), denominado Sisdoc.

## 3. Metodologia

### 3.1. Entrevistas

O levantamento de informações de negócio permite entender como funcionam os processos, quais os sistemas empregados, quais as informações mais importantes, como são produzidas, e quais atividades de mineração poderiam trazer mais retorno para a atividade do Tribunal. Para tanto, foram entrevistados dois diretores da Secretaria das Sessões: Marisa, da diretoria de sistemas, e Sérgio, da diretoria de jurisprudência.

Os detalhes sobre a obtenção de dados em banco de dados e no GED foram fornecidos pela equipe de pesquisa textual, na Secretaria de Soluções de Tecnologia da Informação, em especial o Leandro Ribeiro e a Alessandra Requena.

### 3.2. Preparação e pré-processamento de dados

Ao invés de buscarmos dados primários obtidos diretamente dos sistemas, damos preferência para dados secundários manipulados pelo sistema de pesquisa de jurisprudência, por já apresentarem algum tratamento de qualidade e cruzamento de informações, facilitando assim o presente projeto.

Os metadados principais da jurisprudência são obtidos da visão JURISPRUDENCIA.VW\_JURISPRUDENCIA\_SELECIONADA, entre eles, os códigos de documento do enunciado e do excerto, os códigos e descrições de área, tema e subtema, a chave do acórdão (número, ano e colegiado) e o tipo do processo (processo de instrução da auditoria).

Empregando uma servlet do Sisdoc (<https://contas.tcu.gov.br/egestao/ObterDocumentoSisdoc>), e fornecendo o código, baixamos cada documento utilizado (enunciados e excertos), em formato Word. Extraímos o conteúdo textual do documento por meio da ferramenta Tika (<https://tika.apache.org>).

Dos conteúdos textuais são extraídos os textos de enunciados e dos excertos. Os excertos são fracionados em relatório, voto, acórdão e, em poucos casos, sumário.

A pesquisa de Jurisprudência mantém uma tabela de *staging*, BUSCA\_TEXTUAL.ACORDAO, que contém o inteiro teor do acórdão já extraído, tratado, e armazenado como campo CLOB.

A análise de termos (temas de jurisprudência) foi obtida por meio de query sobre as tabelas da jurisprudência.

Os metadados e textos são objeto de análise (profiling). Os textos de enunciado, excerto e inteiro teor são pré-processados, com os seguintes passos:



- Normalização: remoção de pontuações
- Remoção de stopwords: conforme lista do português da biblioteca nltk
- Stemização: remoção da terminação das palavras conforme PorterStemmer

As informações intermediárias são armazenadas em arquivos separados por vírgula (CSV), e em memória, mantidas em dataframe Pandas (<https://pandas.pydata.org>) e matrizes numpy (<https://docs.scipy.org/doc/numpy/index.html>).

### 3.3. Processamento de dados e avaliação

Os textos lematizados são transformados em vetores esparsos por meio de técnica de *bag of words* (BOW) com TFIDF, com vocábulos de diferentes tamanhos.

Os algoritmos de mineração utilizam com dados de saída as classificações em áreas, e como dados de entradas as BOW de enunciados, excertos e acórdãos. Eventualmente, são utilizados também os tipos de processo (transformados em variáveis *dummy*), BOW de excertos de votos e temas e subtemas. Apesar de serem informações hierárquicas, as minerações de áreas, temas e subtemas foram efetuadas de forma independente. Foram empregadas técnicas multi-classes. Para as redes neurais, as áreas são pré-processadas pelo LabelBinarizer (*sklearn.preprocessing.LabelBinarizer*).

O principal algoritmo empregado é a **logistic regression**, mas foram também empregadas técnicas de **naive bayes**, **random forests**, **linear discriminant analysis** e redes neurais densas.

Como de um acórdão pode-se extrair mais de um enunciado com sua área respectiva, trata-se de um problema multi-classes e multi-labels. Verificamos a possível vantagem de se empregar apenas os acórdãos que produziram um enunciado apenas, e no emprego de todos os acórdãos com enunciados, empregamos a saída *predict\_proba* para extrair múltiplas possibilidades de classificação.

Para os estudos de viabilidade, os dados foram separados em dados de treinamento e teste (em geral, 80/20%). Em geral, empregado como avaliação a medida de acurácia, e produzidas matrizes de confusão.

Ao final, foi efetuada uma avaliação comparativa de classificação de áreas a partir do enunciado, do excerto e do inteiro teor do acórdão, empregando diversos algoritmos de mineração. O estudo comparativo empregou *cross-validation* sobre todos os dados, com 10 *folds*. A avaliação foi realizada com todas as métricas de F1: macro, micro, *weighted* e classe a classe, extraída a média e desvio padrão entre todas as 10 medidas realizadas.

Adicionalmente, realiza-se uma análise de padrões frequentes sobre os termos de indexação dos enunciados, para verificar a possibilidade de sugestão de termos adicionais a partir de termos já escolhidos na classificação de enunciados, empregando algoritmo *apriori*.

## 4. Implementação

A implementação foi efetuada em Python 3 e registradas em Jupiter Notebooks (<https://jupyter.org>). As principais tarefas de mineração empregaram algoritmos disponibilizados no scikit-learn (<https://scikit-learn.org/stable/>). Para redes neurais, empregamos a biblioteca Keras (<https://keras.io>) sobre a plataforma Tensorflow (<https://www.tensorflow.org>). A análise de padrões frequentes se ampara sobre a biblioteca MLXTEND (<http://rasbt.github.io/mlxtend/>).

### 4.1. Análise de padrões frequentes de temas

A primeira análise realizada foi dos padrões frequentes de temas de jurisprudência. Utilizando a função *apriori* do MLXTEND e a função *association\_rules*, empregando métrica lift como *threshold*, e limitando também pelo valor de confiança, obtemos em um dataframe a lista de regras de associação entre grupos de temas.

Essas regras foram armazenadas em arquivo CSV e também em planilha Excel.

### 4.2. Classificação a partir de enunciados

A partir dos textos de enunciados pré-processados e transformados em bag of words de 8000 palavras (*sklearn.feature\_extraction.text.TfidfVectorizer*), empregamos regressão logística (*sklearn.linear\_model.LogisticRegression*), com algoritmo solver Limited-memory BFGS, com regularização L2, multinomial e até 200 iterações, para produzir a classificação das áreas, dos temas e dos subtemas.

Alternativamente, incluímos teste com a inclusão de tipos de processos, transformados em variáveis *dummy* e testamos classificação por meio de rede neuronal densa (*keras.models.Sequential*).

### 4.3. Classificação a partir de excertos

A partir dos textos de excertos pré-processados e transformados em bag of words de 8000 palavras, empregamos regressão logística para produzir a classificação das áreas, dos temas e dos subtemas.

Adicionalmente, produzimos uma classificação de área, tema e subtema a partir dos excertos de votos, e utilização de rede neuronal.

## 4.4. Classificação a partir de acórdãos

A partir dos textos de inteiros teores de acórdãos pré-processados e transformados em bag of words de 8000 palavras, empregamos regressão logística para produzir a classificação das áreas.

Adicionalmente, produzimos uma classificação alternativa de áreas a partir apenas dos acórdãos relacionados com apenas um enunciado de jurisprudência, ou seja, pouco mais de 7 mil acórdãos conforme tabela 1.

## 4.5. Análise comparativa

Finalmente, fizemos uma análise comparativa de classificação de áreas de jurisprudência a partir de textos pré-processados de enunciados, excertos e acórdãos. Para tanto, inicialmente, fizemos o cruzamento (*merge*) de vários dataframes, cada um contendo um tipo de textos pré-processados, e armazenamos em um único arquivo CSV.

A primeira comparação realizada empregou **Logistic Regression** sobre vocabulários de 8000 palavras para cada tipo de texto, em uma cross-validation de 10 folds, sem dados de teste, e extraíndo ao final uma série de medidas de F1 exportadas para planilha Excel.

Na sequência, repetimos o mesmo procedimento, mas variando os tamanhos de vocabulários conforme a característica de variabilidade presumida do texto. Enunciados, que são textos mais curtos e contidos, foram limitados a um vocabulário de 8.000 palavras; excertos, a 12.000 palavras; e acórdãos, a 20.000 palavras.

Apesar de serem execuções separadas, em função de utilizarem a mesma semente de aleatoriedade, acreditamos que são comparáveis também entre si execuções diferentes, empregando algoritmos distintos.

Na sequência, repetimos o primeiro procedimento, empregando **Naive Bayes** (`sklearn.naive_bayes.BernoulliNB`) padrão. Em seguida, **Random Forest** (`sklearn.ensemble.RandomForestClassifier`), com 100 estimadores (`n_estimators=100`) e profundidade máxima de 20 níveis (`max_depth=20`). Por último, empregamos **Linear Discriminant Analysis** (LDA) (`sklearn.discriminant_analysis.LinearDiscriminantAnalysis`), com otimização por mínimos quadrados (`solver = 'lsqr'`), com encolhimento automático (`shrinkage = 'auto'`).

## 5. Avaliação

### 5.1. Análise de padrões frequentes de temas

Empregando como suporte mínimo de 0,002, threshold mínimo do lift de 1728 regras, com diferentes graus de confiança. Filtrando confiança acima de 90% e lift acima de 16, chegamos a 200 regras.

### 5.2. Classificação a partir de enunciados

Obtemos, na regressão logística, acurácia nos dados de testes de:

- 87% para classificação de áreas
- 62% para classificação de temas
- 46% para classificação de subtemas

Lembrando que, apesar de se tratarem de dados hierárquicos, optamos por fazer classificações independentes de área, tema e subtema.

Na classificação de áreas empregando também o tipo de processo, a acurácia caiu para 79%, indicando que o tipo de processo não é um bom preditor para temas de jurisprudência. Na classificação de áreas empregando redes neurais densas, chegamos a 93% de acurácia.

Pela matriz de confusão (tabela abaixo), conseguimos observar que em áreas com menos exemplos, temos uma acurácia de classificação bem menor. Esse fenômeno será abordado com maior profundidade na seção de avaliação comparativa entre classes (5.6).

*Tabela 2 - Matriz de confusão de Logistic Regression a partir de enunciados para predição de áreas*

	Competência do TCU	Contrato Administrativo	Convênio	Desestatização	Direito Processual	Finanças Públicas	Gestão Administrativa	Licitação	Pessoal	Responsabilidade
Competência do TCU	75	0	2	0	13	0	0	1	4	11
Contrato Administrativo	0	143	1	0	0	0	0	34	5	8
Convênio	1	2	94	0	3	0	0	11	6	15
Desestatização	0	2	0	6	0	1	0	9	3	2
Direito Processual	3	1	5	0	315	0	0	3	6	28
Finanças Públicas	0	0	5	0	1	33	1	10	7	7
Gestão Administrativa	3	1	1	0	3	0	29	9	8	7
Licitação	0	5	1	1	2	0	1	560	6	4
Pessoal	2	0	2	0	4	1	0	8	662	5
Responsabilidade	0	4	17	0	15	0	0	25	5	395

### 5.3. Classificação a partir de excertos

Obtemos, na regressão logística, acurácia nos dados de testes de:

- 84% para classificação de áreas
- 57% para classificação de temas
- 40% para classificação de subtemas

Na classificação de áreas empregando apenas os excertos relativos a votos, obtivemos uma acurácia de 83%, 55% e 37%, respectivamente. Isso indica que, apesar do voto, no excerto, conter informação relativamente mais valiosa para formação do enunciado e do tema de jurisprudência, vale mais a pena empregar todo o excerto.

A matriz de confusão indicou um desbalanceamento menor entre as classes, conforme será abordado com maior profundidade na seção de avaliação comparativa entre classes (5.6).

### 5.4. Classificação a partir de acórdãos

Obtemos, na regressão logística, acurácia nos dados de testes de 67% para classificação de áreas. Restringindo aos acórdãos com apenas um enunciado, obtivemos 64% de acurácia, ou seja, houve piora.

Trocando a forma de predição do modelo do `predict`, que retorna a classe escolhida, para o `predict_proba`, que retorna as probabilidades de cada classe, e selecionando não só a classe com maior probabilidade mas as três classes mais prováveis, conseguimos minorar o efeito da inacurácia do modelo, e faz sentido pois boa parte dos acórdãos tem realmente duas ou três classificações. Analisando a coluna de acumulado da tabela 1, verificamos que mais de 98% dos acórdãos selecionados possuem até 3 enunciados. Segue alguns exemplo de saída do algoritmo (ao lado da chave do acórdão, a classificação real, abaixo, as predições):

```
Acórdão 1645/2009-1  Pessoal
  classe 1 Pessoal (0.8275881328078026)          ok! (1)
  classe 2 Direito Processual (0.05804026687076472)
  classe 3 Responsabilidade (0.03712286098791601)
Acórdão 199/2016-P  Licitação
  classe 1 Licitação (0.9572207907147968)        ok! (1)
  classe 2 Contrato Administrativo (0.0113074320574)
  classe 3 Direito Processual (0.009423356058260652)
Acórdão 93/2009-P  Contrato Administrativo
  classe 1 Licitação (0.6019899117621723)
  classe 2 Contrato Administrativo (0.269293688075)  ok! (2)
  classe 3 Responsabilidade (0.03161306609552059)
Acórdão 1345/2012-P  Direito Processual          (errou todas)
  classe 1 Licitação (0.45065045571959594)
  classe 2 Contrato Administrativo (0.34046246474580655)
  classe 3 Responsabilidade (0.06772145259031845)
Acórdão 1457/2011-P  Contrato Administrativo
  classe 1 Licitação (0.2472683382652277)
  classe 2 Contrato Administrativo (0.200306811959)  ok! (2)
  classe 3 Responsabilidade (0.16951411307050376)
```

## 5.5. Análise comparativa global

Ao contrário das análises anteriores, que visavam primordialmente verificar a validade da classificação automática, esta análise pretende comparar de forma mais assertiva o desempenho da classificação a partir de enunciados, excertos e acórdãos, entre diversos algoritmos empregados.

As medidas de F1 foram obtidas sobre cada fração de validação (fold). Além da F1 micro, que é a acurácia, foram registradas também as de F1 macro, que são uma média aritmética da F1 de cada classe, e a F1 weighted, que é uma média ponderada pela representatividade de cada classe. Das 10 validações, tiramos a média e o desvio padrão. Foram registradas e organizadas todas as medidas realizadas em uma planilha Excel. A tabela a seguir mostra um resumo.

*Tabela 3 - Medidas F1 comparativas para classificação por áreas*

		Logistic Regression			Naive Bayes		
métrica		enunciados	excertos	acordaos	enunciados	excertos	acordaos
f1_macro	média	0,743	0,746	0,558	0,652	0,588	0,525
	desvio	0,022	0,016	0,026	0,027	0,017	0,028
f1_micro	média	0,867	0,832	0,672	0,826	0,724	0,598
	desvio	0,008	0,006	0,009	0,010	0,014	0,015
f1_weighted	média	0,860	0,827	0,654	0,814	0,732	0,609
	desvio	0,008	0,007	0,010	0,012	0,013	0,015

		Random Forest			Linear Discriminant Analysis		
métrica		enunciados	excertos	acordaos	enunciados	excertos	acordaos
f1_macro	média	0,474	0,513	0,480	0,592	0,735	0,562
	desvio	0,013	0,025	0,026	0,027	0,018	0,026
f1_micro	média	0,743	0,729	0,638	0,702	0,808	0,645
	desvio	0,013	0,009	0,010	0,012	0,006	0,009
f1_weighted	média	0,695	0,694	0,605	0,707	0,807	0,647
	desvio	0,016	0,013	0,011	0,012	0,007	0,011

As medidas tiradas da Logistic Regression, em geral, foram as melhores obtidas. As medidas sobre enunciados e excertos foram muito próximas. No caso da F1 macro temos um empate considerando a distribuição de t-sudent com nível de confiança de 95%. No caso dos acórdãos, temos medidas cerca de 20% menores. Além disso, podemos observar que as classificações sobre excertos apresentaram os menores desvios.

As medidas tiradas de outras técnicas de mineração foram piores, especialmente no caso da random forest. Vale um destaque para o LDA, que apesar de muito ruim na predição a partir dos enunciados, teve um desempenho relativamente bom para excertos e acórdãos, em comparação com a regressão logística.

## 5.6. Análise comparativa por classes (áreas)

De forma similar às medidas de F1 globais, foram obtidas medidas F1 para cada classes. Só que, ao invés de termos uma medição para cada fração de validação (fold), temos 10, uma para cada classe, o que aumenta a quantidade de informações obtidas. A tabela a seguir mostra um resumo apenas das medições sobre a regressão logística. Na última linha, a média das médias (F1 macro) e o desvio das médias de cada classe.

*Tabela 4 - Medidas F1 de classes, para regressão logística*

Classe ou área	medida	enunciados	excertos	acordãos
Competência do TCU	média	0,798	0,631	0,348
	desvio	0,056	0,066	0,049
Contrato Administrativo	média	0,768	0,726	0,382
	desvio	0,032	0,011	0,046
Convênio	média	0,730	0,644	0,365
	desvio	0,081	0,071	0,055
Desestatização	média	0,244	0,622	0,635
	desvio	0,110	0,128	0,156
Direito Processual	média	0,881	0,816	0,548
	desvio	0,021	0,017	0,036
Finanças Públicas	média	0,639	0,701	0,438
	desvio	0,080	0,085	0,090
Gestão Administrativa	média	0,666	0,692	0,604
	desvio	0,089	0,059	0,073
Licitação	média	0,897	0,875	0,746
	desvio	0,015	0,017	0,034
Pessoal	média	0,946	0,959	0,905
	desvio	0,012	0,010	0,010
Responsabilidade	média	0,861	0,794	0,608
	desvio	0,014	0,018	0,020
(entre as classes)	média	0,743	0,746	0,558
	desvio	0,202	0,112	0,181

Nas medidas extraídas da classificação a partir dos enunciados, verificamos que as classes mais representativas segundo a figura 3 apresentam F1 acima da média, e classes menos representativas apresentam F1 abaixo da média, com destaque negativo para “Desestatização”.

As medidas extraídas da classificação a partir dos excertos estão mais equilibradas entre as classes, o que se reflete no desvio entre as classes, bem mais baixo que os demais. As medidas são relativamente boas para todas as classes. Mesmo a média de “Desestatização”, a pior delas, está em 62%, contra 24% do grupo anterior. As medidas da classificação a partir dos acórdãos apresentaram um comportamento mais anômalo.

## 6. Conclusão

A mineração de padrões frequentes de temas gerou regras que podem ser úteis para auxiliar o cadastramento desses termos para indexação da jurisprudência selecionada. Essas informações de relacionamento entre os termos podem, ainda, ser uma retroalimentação interessante para o sistema de Thesaurus do TCU, o VCE. Isso tudo exige uma análise mais aprofundada, conversas com as áreas de negócio afetadas, e possível emprego de algoritmos de mineração de sequências de termos.

A classificação de área a partir dos enunciados e excertos apresentou acurácia muito boa e próxima entre elas, no caso da regressão logística, o que indica uma boa viabilidade como mecanismo de classificação automática. O fato da regressão logística sobre excertos apresentar desvio bem menor entre as medidas médias das classes, indica essa classificação com um excelente insumo para um *stacking* de modelos de classificação de área, talvez com um balanceamento entre as classes dos dados de treinamento.

A possível utilidade dessa classificação automática para o cadastro da jurisprudência deve ser avaliada. Uma possível utilização é justamente a discordância da classificação automática a partir do enunciado e do excerto, bem como a efetiva classificação efetuada pelo analista de jurisprudência. Tal discordância pode ser um indicativo de possível revisão do cadastro.

A acurácia das classificações de tema e subtema foram menores, o que é esperado pelo número de classes e, especialmente, pela falta de representatividade de certas classes. Ainda assim, são viáveis, dadas as quantidades de classes possíveis. Um possível teste futuro é a classificação dos temas utilizando as classificações das áreas, e dos subtemas a partir das áreas e temas, quebrando assim o paradigma de classificação independente que foi escolhido no presente trabalho.

A utilização de metadados adicionais aos textos pode ser melhor explorado. Tentamos, neste trabalho, a classificação a partir de textos de enunciados associados a tipos de processos, o que não se mostrou viável. Contudo, a utilização do tipo de processo em adição ao texto do acórdão pode se mostrar útil, e outros metadados podem ser buscados.

Os modelos de redes neurais profundas revelaram uma melhoria na acurácia na base de teste. Isso pode ser aprofundado em trabalhos futuros. As medidas tiradas de outras técnicas de mineração foram piores. Devemos considerar, porém, que ajustes nos hiperparâmetros e métodos de otimização podem trazer uma melhora sensível nesses números. Especialmente no caso da random forest, que apresentou o pior desempenho. Vale um destaque para o LDA, que apesar de muito ruim na predição a partir dos enunciados, teve um desempenho relativamente bom para excertos e acórdãos, em comparação com a regressão logística. Por isso, um aprofundamento na utilização desses modelos pode ser promissor para a melhoria da acurácia.

A classificação a partir do inteiro teor do acórdão apresentou a pior acurácia, na comparação, com cerca de 20% abaixo dos demais, no caso da regressão logística. Apesar disto,



parece ser viável a classificação, merecendo maior investimento em técnicas de mineração, ajustes e uso de técnicas de composição de modelos (ensembles e stacking), de modo a alcançarmos medidas melhores. A classificação a partir dos acórdãos é a mais promissora do ponto de vista da utilidade, por poder ser realizada antes do trabalho de jurisprudência, e poder ser aplicada à pesquisa de inteiro teor.

Outros trabalhos futuros possíveis envolvem a produção automática de textos, ou seja, a produção de excertos a partir dos acórdãos, e enunciados a partir dos acórdãos e excertos. Além disso, outras técnicas de extração de semântica dos textos podem ser úteis, a partir de modelos de linguagem e técnicas recentemente desenvolvidas de processamento de linguagem natural.