



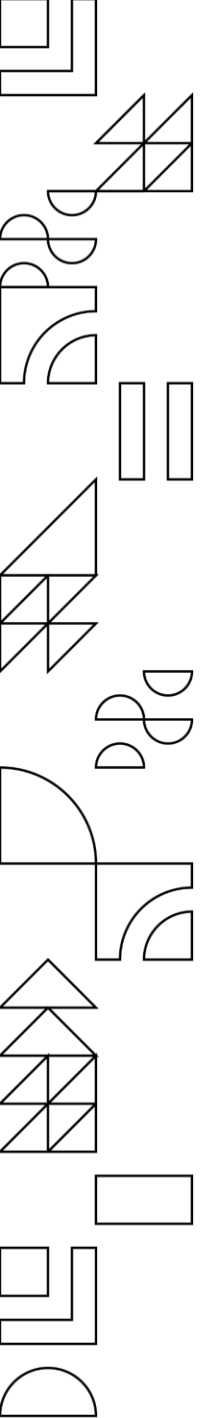
Escola Nacional de Administração Pública – ENAP
Bootcamp Machine Learning

Projeto Final

Classificação das jurisprudências do TCU

ALUNAS: LIVIA SANTOS E ISABELA CARDOSO

ORIENTADOR: FERNANDO SOLA



Problema

- Contexto

Necessidade classificar as jurisprudências do TCU em categorias

- Objetivo

Automatizar a classificação das jurisprudências do TCU a partir do texto do enunciado

- Resultado

O modelo acertou 89% das classificações

Fonte de dados e limitações



Fonte dos dados: Portal do TCU - Pesquisa Integrada do TCU - Jurisprudência Seleccionada – Pesquisa por árvore de classificação - Exportar resultado das pesquisas por área (total 10 áreas).

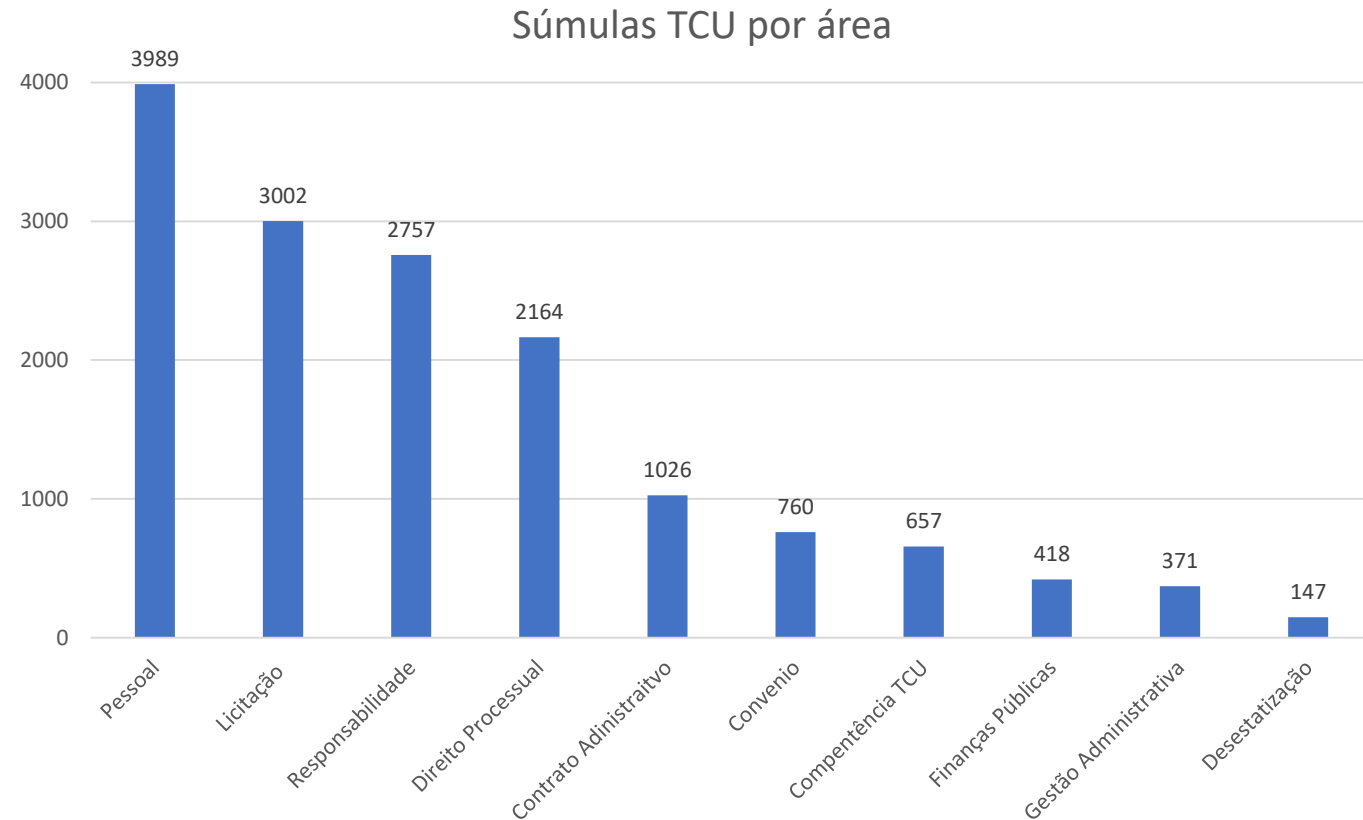


Limitação: única variável Texto da Súmula (enunciado).

Quantidade de amostras por categoria

Nome área	Código Área	Total amostras
Pessoal	8	3989
Licitação	7	3002
Responsabilidade	9	2757
Direito Processual	4	2164
Contrato Administrativo	1	1026
Convênio	2	760
Competência TCU	0	657
Finanças Públicas	5	418
Gestão Administrativa	6	371
Desestatização	3	147

Total 15291



*Classes desbalanceadas

Abordagem para solucionar o problema

Pré-processamento

- Criada coluna numérica (sequencial) para representar as áreas (categorias de classificação).

Separação dos dados em treino e teste

- Train test split – separação 80% treino e 20% para teste (sem balanceamento entre as classes)
- Random Over Sample (ROS) – replica amostras das classes minoritárias balanceando o dataset.

Term Frequency – Inverse Document Frequency (TF – IDF) – vocabulário e frequência

- Importância de uma palavra em um documento em relação a uma coleção de documentos ou corpus linguístico.
- **Estudo de caso:** corpus - enunciados das súmulas do TCU.
- Decisões selecionadas dos tribunais superiores e dos principais tribunais de contas estaduais, nas diversas áreas de atuação do controle externo.
- Parâmetros: exclusão de stop-words, limitação a 5000 palavras presentes em no máximo 70% dos documentos e no mínimo em 3 documentos

Modelos

Modelo 1 - TF-IDF sem stopwords + SGDClassifier (0,88 acurácia)

***Modelo 2** - TF-IDF sem stopwords + ROS (treino com classes balanceadas) + SGD Classifier (**0,89 acurácia**)

Modelo 3 - TF-IDF sem stopwords e números (regex)+ SGD Classifier (0,83 acurácia)

Modelo 4 - TF-IDF sem stopwords + Random Forest (0,84 acurácia)

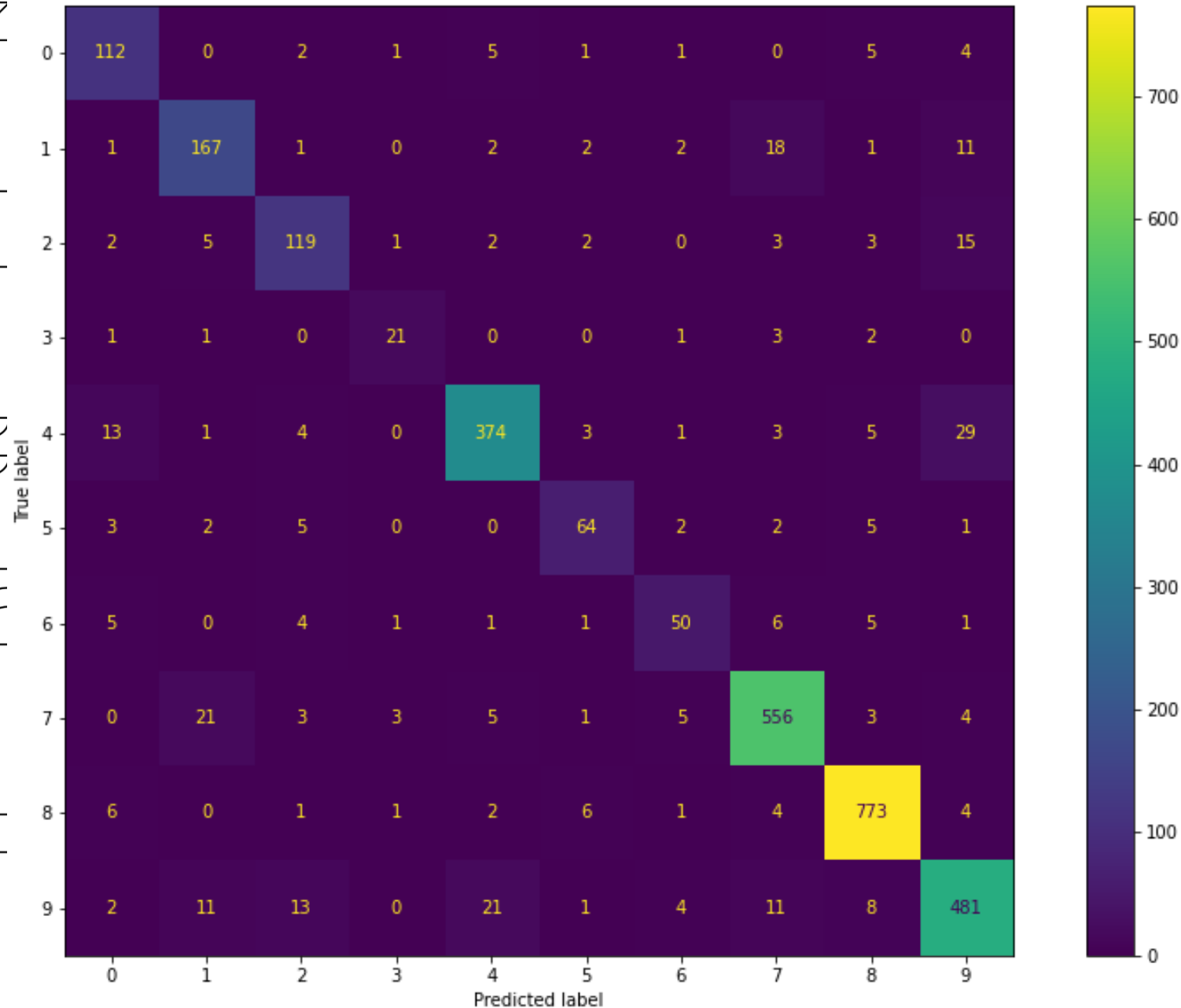
Modelo 5 - TF-IDF sem stopwords + SGDClassifier (0,87 acurácia) + GridSearchCV

Desempenho do modelo

	precision	recall	f1-score	support
0	0.77	0.85	0.80	131
1	0.81	0.81	0.81	205
2	0.74	0.78	0.76	152
3	0.75	0.72	0.74	29
4	0.90	0.88	0.89	433
5	0.78	0.77	0.78	84
6	0.74	0.69	0.71	74
7	0.92	0.93	0.92	601
8	0.95	0.97	0.96	798
9	0.89	0.86	0.87	552
accuracy			0.89	3059
macro avg	0.82	0.83	0.82	3059
weighted avg	0.89	0.89	0.89	3059

*Melhor desempenho com classes balanceadas (ROS)

Matriz de confusão



Nome categoria	Código categoria	Total amostras	Teste 20%
Pessoal	8	3989	798
Licitação	7	3002	601
Responsabilidade	9	2757	552
Direito Processual	4	2164	433
Contrato Administrativo	1	1026	205
Convênio	2	760	152
Competência TCU	0	657	131
Finanças Públicas	5	418	84
Gestão Administrativa	6	371	74
Desestatização	3	147	29



Predição na área 9 com mais de 0,95% prob

- A ocorrência de mudanças alinhadas com à concepção do objeto do convênio e de seus objetivos não caracteriza desvio de finalidade ou de objeto, mas falha de natureza formal. (real 2)
- A mudança de fundamentação de multa aplicada com base no art. 57 para o art. 58 da Lei 8.443/1992, a fim de apenar responsável que conseguiu, em grau de recurso, elidir o débito, mas não outras irregularidades analisadas pelo TCU, somente é possível caso o valor da pena pecuniária seja superior ao piso definido no regimento interno, levando-se em consideração a consequente redução da multa em função do afastamento do débito. (real 4)
- Sempre que juridicamente viável, deve ser promovida a citação solidária da sociedade empresária com seu dirigente, nos casos de tomada de contas especial que envolva renúncia fiscal, a exemplo das atividades desenvolvidas com recursos financeiros captados com amparo na Lei 8.313/1991 (Lei Rouanet) , mormente porque a omissão no dever de prestar contas configura infração que merece ser imputada à pessoa física do gestor, e não apenas à empresa. (real 4)
- O prazo decadencial previsto no art. 54 da Lei 9.784/1999 é aplicável ao TCU apenas como meio de autotutela no desempenho de sua função administrativa, não se aplicando ao exercício de sua competência constitucional de controle externo. (real 4)



Predição na área 4 com mais de 0,95% prob

- As peculiaridades e circunstâncias do caso concreto podem justificar a atuação do gestor em sentido diverso daquele indicado pelo TCU em caráter normativo (consulta) . (real 9)
- No juízo de responsabilidade acerca do descumprimento de diligência ou determinação do TCU, devem ser verificadas, objetivamente, as ações adotadas pelo agente público para a efetivação do comando do Tribunal, inexistindo espaço para se reabrir o debate acerca da questão de mérito, possível apenas mediante a via recursal adequada. (real 9)
- As declarações constantes no Sisac possuem presunção de veracidade. Por economia processual, eventual falta ou insuficiência de informação fica suprida pela declaração da autoridade responsável do órgão de origem, confirmada pelo órgão de controle interno. Os gestores têm que fornecer informações corretas e baseadas em provas documentais, sob pena de responsabilização nas esferas civil, penal e administrativa. (real 8)
- Eventual demora na instauração da tomada de contas especial constitui falha de procedimento atribuível à autoridade que tinha o dever de adotar essa medida, a qual deve ser apurada e, se for o caso, aplicada as sanções correspondentes em cada processo. (real 9)



Estudos futuros:

Retirar números não significativos e dar peso a Legislações

Utilizar parâmetros no modelo Random Forest

Criar expressões significativas para classes que se confundem

Usar expressões para classificação automática

Considerações finais:

O modelo SGD Classifier apresentou resultado satisfatório

Facilidade de replicação em atividades similares no setor público por tratar-se de modelo simples



Escola Nacional de Administração Pública – ENAP
Bootcamp Machine Learning

Obrigada

ALUNAS: LIVIA SANTOS E ISABELA CARDOSO

ORIENTADOR: FERNANDO SOLA