

# **RELATÓRIO DE PESQUISA**

## **Classificação temática de Projetos de Lei por meio de inteligência artificial**

Fábio de Barros Correia Gomes

**Brasília**

**MAIO, 2024**

## Sumário

1 Introdução .....	3
2 Contexto .....	4
3 Metodologia .....	5
4 Atividades e Resultados .....	11
4.a. Ampliação da revisão da literatura .....	12
4.b. Preparação do banco de dados .....	14
4.c. Experimentos com modelos de ML e ANN .....	16
4.d. Análise da estrutura lógica da tipologia de classificação .....	31
5 Conclusões e Sugestões .....	32

## 1 Introdução

Este relatório detalha as atividades realizadas, os resultados alcançados e as conclusões extraídas de pesquisa sobre a classificação temática de projetos de lei (PL) por meio de métodos de aprendizado de máquina e de inteligência artificial; particularmente os relacionados à saúde (PLRS) no contexto da Câmara dos Deputados do Brasil. A execução deste estudo ocorreu entre 18 de março e 16 de maio de 2024, estando em conformidade com o projeto previamente aprovado. A supervisão esteve a cargo de integrante do Grupo de Pesquisa e Extensão do CEFOR da Câmara dos Deputados, focado em "Legislativo e Política de Saúde" (GPE-Saúde).<sup>1</sup>

Este esforço se insere num conjunto de investigações que buscam otimizar o processo de análise da produção legislativa, associada às diretrizes constitucionais de saúde, estabelecidas desde a criação do Sistema Único de Saúde (SUS) pela Constituição Federal de 1988.

A relevância da pesquisa é demonstrada pelo volume expressivo de legislação produzida anualmente, onde, por exemplo, entre 1999 e 2006, mais de 20.000 PL foram apresentados, com cerca de 5.000 desses relacionados à saúde; evidenciando a necessidade de metodologias automáticas para facilitar a classificação e análise de tais textos legislativos.

Antes da exposição das seções inicialmente mencionadas, será apresentado um resumo sobre o contexto e a metodologia empregada na pesquisa.

---

<sup>1</sup> O supervisor é doutor em ciências da saúde, coordenador da área de Saúde da Consultoria Legislativa da Câmara e membro do GPE-Saúde.

## 2 Contexto

Os antecedentes dessa pesquisa incluem uma série de experimentos conduzidos com o intuito de aplicar e avaliar técnicas de aprendizado de máquina para a classificação de PLRS. A primeira rodada de experimentos, realizada em 2019, utilizou o ambiente de programação R para analisar um conjunto de 8.327 PLs, sendo 2.328 identificados como PLRS (28%). A metodologia empregada incluiu a extração de texto de arquivos no formato “pdf” e a aplicação do modelo *Random Forest* (com 100 árvores), atingindo uma acurácia de 89,7%, especificidade de 97,4% e *recall* de 69,8%. O *Kernel SVM* alcançou um *recall* de 73,6%.<sup>2</sup>

Uma segunda rodada, em 2020, replicou e expandiu os experimentos anteriores no ambiente *Python*, corrigindo erros de classificação anteriores e incorporando modelos adicionais como *Stochastic Gradient Descent (SGD)* e *XGBoost*, mantendo a alta precisão e acurácia (em torno de 90%). O modelo *Random Forest* apresentou melhor desempenho de *recall* (79,8%) do que na primeira rodada.

Alguns experimentos foram realizados visando os grupos temáticos de saúde usando o modelo *Random Forest*, mas, enquanto a precisão foi em geral próxima a 90% e a acurácia em torno de 97%, a sensibilidade foi muito inferior à alcançada pelo grupo geral da saúde. Por exemplo, para o grupo de prevenção, o valor foi de 57%; para o grupo de direitos e responsabilidades, 45%; para o grupo

---

<sup>2</sup> Uma explicação sobre o significado desses indicadores é apresentada na seção sobre metodologia.

de gestão, recursos e pesquisa, 37%; e para o grupo de ações e serviços de saúde, 30%.

Em 2022, uma terceira rodada de experimentos foi realizada após uma revisão detalhada da classificação de PLs, resultando em melhorias nos resultados do modelo *Random Forest*. A acurácia foi de 91%; o *recall*, 79,4%; a precisão, 88%; e a medida F1, 83,6%. Também foi elaborada proposta de revisão na tipologia usada para a classificação de PLRS, de modo a otimizar o desempenho dos modelos de aprendizado de máquina através do refinamento da lógica de classificação.

A investigação também evidenciou a relevância dos dados disponíveis, com mais de 50 mil PLs classificados manualmente, representando um recurso valioso para futuras pesquisas.

Os resultados obtidos nas rodadas de pesquisa anteriores forneceram uma base sólida para a exploração de novas abordagens de inteligência artificial, além do aprendizado de máquina, na classificação de PLs, de modo a contribuir com a automação e eficiência na gestão da vasta quantidade de proposições produzida anualmente.

### **3 Metodologia**

Este estudo adotou uma metodologia centrada no processamento de linguagem natural (PLN) para a classificação temática de milhares de PLs em língua portuguesa, focando na sua relação com a saúde. A investigação se apoiou em métodos de pesquisa documental, incluindo levantamento bibliográfico e revisão da literatura nacional e internacional, além da elaboração e aplicação de modelos de

aprendizado de máquina e de outras formas de inteligência artificial para a tarefa de classificação.

A pesquisa continuou a empregar modelos supervisionados. Esses modelos foram treinados com um conjunto de dados já classificados por humanos, permitindo que o algoritmo aprendesse as relações matemáticas entre o texto de cada PL e as categorias da tipologia estabelecida, para posteriormente aplicar esse conhecimento na classificação de novos conjuntos de dados.

Os dados utilizados neste estudo incluíram os PLs apresentados na Câmara dos Deputados entre 2011 e 2014, além de mais 2.492 PLs do ano de 2022, num total de 10.067 PLs; todos codificados por classificadores humanos como relacionados ou não à saúde. Estes PLs estavam distribuídos em grupos temáticos como "ações e serviços de saúde", "prevenção", "direitos e responsabilidades", "gestão, recursos e pesquisa" e outros temas relacionados à saúde, abrangendo cerca de 60 categorias distintas.

A preparação do banco de dados foi realizada por meio do ambiente virtual do *Google Colab* para implementação de códigos em *Python*. Foi realizada a extração de URLs de *hiperlinks* fornecidos pelo sítio da Internet dos Dados Abertos da Câmara dos Deputados,<sup>3</sup> o que permitiu o *download* de arquivos em formato “pdf”, dos anos de 2015 a 2023 (para facilitar pesquisas futuras, pois já estavam incorporados ao banco de dados os anos de 2011 a 2014). Esses arquivos foram armazenados no *Google Drive*.

---

<sup>3</sup> BRASIL. Câmara dos Deputados. Dados abertos [Sítio da Internet]. Brasília: Câmara dos Deputados, 2024. Disponível em: <https://dadosabertos.camara.leg.br/swagger/api.html#staticfile>.

A extração automática de conteúdo dos arquivos “pdf”, era um gargalo presente nos experimentos anteriores <sup>4</sup> e foi superada com a adoção do pacote para *Python* denominado *Tesseract*.

Assim, os arquivos em “pdf” foram transformados em arquivos “txt” e o conteúdo desses arquivos foram inseridos no banco de dados no formato “csv”.

A preparação dos dados, em geral, envolveu processos de limpeza de texto, como a remoção de caracteres especiais, visando reduzir a complexidade dos dados e melhorar a performance dos modelos de PLN e inteligência artificial na classificação dos textos.

Para verificação da consistência dos dados foi utilizada a ferramenta *OpenRefine*, que permite a manipulação de grandes volumes de dados.

Outro avanço metodológico desse experimento foi a manipulação de variáveis de texto, resultando na divisão do texto completo dos PLs em variáveis contendo os artigos e a justificção em separado, permitindo maior variação nos experimentos.

Além da melhoria na qualidade do conjunto de treinamento e do aumento do banco de dados através da inclusão de novos PLs (de 2022), esta etapa de pesquisa explorou a tipologia usada na classificação (focando no nível de grupo temático), para auxiliar na resolução de inconsistências lógicas na classificação.

Para a análise, foram selecionados os modelos *SGDClassifier* e de *Artificial Neural Network (ANN)*. O primeiro modelo foi selecionado, pois foi realizada

---

<sup>4</sup> A extração dos conteúdos dos PLs é uma etapa crítica, pois nem todos os arquivos em “pdf” disponibilizados pela CD estão num formato que permite uma extração rápida e simplificada do texto por pacotes de programas de computador, uma vez que alguns PLs estão armazenados como imagens.

uma revisão dos experimentos de 2020 e 2022 e observou-se que tal modelo apresentou bom desempenho nas quatro métricas de avaliação, logo foi utilizado como parâmetro para esta etapa. O segundo modelo foi selecionado, pois representa avanço tecnológico em relação aos modelos prévios de ML.

Vale destacar que todos os modelos foram salvos para facilitar uso posterior.

O design experimental previu a utilização de 80% do conjunto de dados para o treinamento dos modelos, enquanto os 20% restantes foram reservados para a fase de teste, a fim de avaliar o desempenho da classificação.

Foram desenvolvidas várias versões de modelo de ANN, inclusive um com múltiplas entradas de variáveis de texto, na tentativa de melhorar a classificação temática de proposições legislativas.

Também foram testados diferentes algoritmos para classificar os grupos temáticos das proposições, com ênfase em técnicas para lidar com bancos de dados desbalanceados.

Foi utilizada a técnica *SMOTE* para lidar com desbalanceamento nos dados e realizada experimentação para redução de dimensionalidade.

Foi dada atenção à busca de alternativas para otimizar a classificação, incluindo a atribuição de pesos diferenciados às categorias e a análise da eficácia de *pipelines* e *ensembles* de modelos.

Para uma melhor compreensão dos resultados, também foi elaborado código que permite listar e observar o conteúdo dos PLs em que a predição do modelo diferiu da classificação por humano.



As mencionadas estratégias metodológicas são especificadas na seção de resultados.

Os modelos foram avaliados com base em métricas de desempenho estabelecidas, por meio de indicadores, que incluem: acurácia, precisão (ou valor preditivo positivo), sensibilidade (também conhecida como *recall*), especificidade e a medida F1 (*F1 score*).

Conforme detalhado por Igual e Seguí (2017, p. 72),<sup>5</sup> essas métricas utilizam em seus cálculos os casos de: verdadeiros positivos (VP) - quando o classificador prevê corretamente uma amostra como positiva -; falsos positivos (FP) - quando o classificador incorretamente prevê uma amostra como positiva; verdadeiros negativos (VN) - quando o classificador corretamente prevê uma amostra como negativa -; e falsos negativos (FN) - quando o classificador falha ao prever uma amostra positiva.

Por exemplo, a acurácia é definida pela proporção de previsões corretas do total de previsões feitas, calculada como a soma dos VP e VN, dividida pela soma de VP, VN FP e FN (IGUAL E SEGUÍ, 2017, p. 73) <sup>5</sup>. Este indicador reflete a capacidade geral do modelo de classificar corretamente as instâncias tanto positivas quanto negativas.

A precisão, ou valor preditivo positivo, foca na proporção de instâncias positivas corretamente identificadas do total de instâncias classificadas como positivas pelo modelo. É calculada dividindo-se os VP pela soma dos VP e FP.

---

<sup>5</sup> IGUAL, L.; SEGUÍ, S. Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications. Cham, Switzerland: Springer International Publishing AG, 2017. Disponível em: <https://github.com/DataScienceUB/introduction-datascience-python-book>.

Koehrsen (2018) <sup>6</sup> descreve a precisão como a capacidade do modelo de classificação de retornar apenas instâncias relevantes, enfatizando sua importância na qualidade das classificações positivas.

A sensibilidade, ou *recall*, mede a capacidade do modelo de identificar todas as instâncias relevantes, calculada pela divisão dos VP pela soma dos VP e FN. Este indicador é particularmente importante para avaliar se o modelo é capaz de captar a maioria das instâncias positivas, conforme destacado por Koehrsen (2018).

A especificidade, por outro lado, é a capacidade do modelo de classificação para identificar corretamente os VN entre todos os casos negativos, calculada pela divisão dos VN pela soma dos VN e FP, conforme explicado por Igual e Seguí (2017, p. 73). <sup>5</sup>

Por fim, a Medida F1 é uma métrica que combina precisão e sensibilidade em uma média harmônica, calculada como o dobro do produto da precisão e do *recall* dividido pela soma da precisão e do *recall*. Koehrsen (2018) <sup>6</sup> e Tharwat (2020, p. 174) <sup>7</sup> destacam que a média harmônica é utilizada em vez de uma média simples para penalizar valores extremos, tornando a Medida F1 um indicador robusto do desempenho equilibrado de classificação, variando de zero a um, onde valores mais altos indicam um desempenho superior.

---

<sup>6</sup> KOEHRSEN, W. Beyond Accuracy: Precision and Recall. Towards Data Science. Publicado Online: 03 de março de 2018. Disponível em: <https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c>.

<sup>7</sup> THARWAT, A. Classification assessment methods. Applied Computing and Informatics, vol 17, 1, 2020. Disponível em: <https://www.emerald.com/insight/content/doi/10.1016/j.aci.2018.08.003/full/pdf?title=classification-assessment-methods>.

Finalmente, a pesquisa conclui com uma análise da estrutura lógica da tipologia de classificação para PLRS. Para tanto, foi analisada a estrutura proposta na rodada anterior de experimentos e verificado o desempenho para o nível de grupos temáticos. Também foram utilizados subsídios decorrentes da verificação das discordâncias entre classificações de humanos e de modelos.

Em síntese, as atividades de pesquisa foram planejadas e executadas ao longo de um período de 60 dias, dividido em fases, começando com a revisão da literatura e preparação dos dados, seguido pelo treinamento e avaliação dos modelos de ML e de redes neurais, e na preparação do relatório sobre os resultados obtidos.

A viabilidade da execução desta pesquisa foi assegurada por diversos fatores, incluindo a disponibilidade de um robusto conjunto de dados previamente classificado e os avanços tecnológicos em PLN e inteligência artificial.

Na seção seguinte, são apresentadas tanto as atividades que contribuíram para o aperfeiçoamento metodológico quanto os resultados alcançados por este estudo.

#### **4 Atividades e Resultados**

As seguintes atividades foram realizadas no âmbito dessa pesquisa:

- a) Ampliação da revisão da literatura;
- b) Preparação do banco de dados;
- c) Experimentos com modelos de ML e ANN;
- d) Análise da estrutura lógica da tipologia de classificação.

Seguem as descrições de cada atividade e dos resultados obtidos.

#### 4.a. Ampliação da revisão da literatura

Revisões anteriores já haviam identificado 70 estudos de interesse, que abordam temas de “*machine learning*”, “metodologias de classificação” e “agenda da saúde em tramitação na Câmara dos Deputados”. Nessa etapa foi priorizada a busca por aplicações específicas para o caso de classificação de proposições do Legislativo, tendo sido encontradas duas experiências de elevado interesse. São elas:

- 1) **Título:** *A classification approach for estimating subjects of bills in the Brazilian Chamber of Deputies.*

**Data:** 2023.

**Autor:** Nunes, Rafael Oleques.

**Disponível em:** <http://hdl.handle.net/10183/267612>.

**Natureza:** Trabalho de Conclusão de Curso (TCC) do Curso de Ciência de Computação Universidade Federal do Rio Grande do Sul (UFRGS).

**Síntese:** O estudo identifica como problema de pesquisa a ausência de classificação de temas em uma grande parcela das proposições legislativas na Câmara dos Deputados do Brasil. Entre 1991 e 2022, aproximadamente 75% dessas proposições não possuíam classificação temática nos metadados associados, dificultando a acessibilidade e compreensão por parte dos cidadãos e aumentando o distanciamento entre sociedade e política.

Para enfrentar essa questão, Nunes propõe o uso de técnicas de aprendizado de máquina e PLN para estimar os temas das proposições legislativas. Seu estudo destaca a aplicação de dois modelos baseados em BERT (uma técnica

de PLN de última geração), adaptados para a língua portuguesa, e utiliza a ementa das proposições - um resumo dos principais pontos do documento - como base para a análise.

Os resultados obtidos por Nunes indicam que o modelo BERTimbau (uma variação do BERT para o português) alcançou uma pontuação F1 ponderada de 78,94% e uma pontuação F1 macro de 72,78%.

2) **Título:** *Legal Text Classification based on Greek Legislation.*

**Data:** 2020.

**Autor:** Papaloukas, Christos N.

**Disponível em:** <https://pergamos.lib.uoa.gr/uoa/dl/object/2931361/file.pdf>.

**Natureza:** Tese de mestrado na *National and Kapodistrian University of Athens School of Sciences. Department of Informatics and Telecommunications Program of Postgraduate Studies. Information and Data Management.*

**Síntese:** O processamento de textos jurídicos emergiu como um campo significativo no âmbito do PLN, apresentando aplicações valiosas não apenas para profissionais do direito, mas também para a sociedade em geral. Nesse contexto, foi conduzido um estudo focado na classificação de textos jurídicos baseados na legislação grega. Introduziu-se e disponibilizou-se publicamente um novo conjunto de dados baseado na legislação grega, compreendendo mais de 47 mil documentos da legislação grega oficial e classificada.

Foram experimentados e avaliados diversos métodos e classificadores avançados, abrangendo desde modelos de aprendizado de máquina tradicionais e modelos recorrentes até modelos de transferência de aprendizado de última geração. Os resultados evidenciaram as limitações dos classificadores de

aprendizado de máquina tradicionais frente a métodos mais sofisticados, apesar de estabelecerem linhas de base adequadas para a maioria das tarefas consideradas. Arquiteturas recorrentes, utilizando incorporações de palavras específicas ao domínio jurídico, ofereceram um desempenho geral melhorado, sendo competitivas até mesmo com modelos baseados em *Transformers*.

No entanto, modelos *Transformers* multilíngues e monolíngues de última geração disputaram a liderança no ranking de classificadores, suscitando questionamentos sobre a necessidade de treinar modelos de transferência de aprendizado monolíngues como uma prática padrão.

Este estudo representou a primeira ocasião em que a classificação de textos jurídicos gregos foi investigada com tal profundidade. Antecipou-se que os resultados serviriam como uma base sólida, contribuindo significativamente para as futuras pesquisas da comunidade de PLN grega.

#### 4.b. Preparação do banco de dados

Foi realizado o download de 33.282 arquivos em pdf a partir do site da Câmara dos Deputados, correspondentes aos PLs de 2015 a 2023 (ver Quadro 1), pois os de 2011 a 2014 já estavam incorporados ao banco de dados da pesquisa.

**Quadro 1. Quantitativo de arquivos de PLs em “pdf”, por ano.**

Ano	Arquivos
2015	4260
2016	2532
2017	2665
2018	1809
2019	5466
2020	4716
2021	3965

2022	2492
2023	5377
<b>Total</b>	<b>33282</b>

Cada arquivo em pdf gerou um arquivo em “txt”, de modo que foram gerados cerca de 66.500 arquivos, dispostos em 21 pastas e ocupando quase 8 gigabytes em espaço de disco.

Nesta etapa da pesquisa foram adicionados os 2.492 PLs de 2022, classificados por humanos, de modo que o banco de dados foi composto por 10.067 PLs.

Os PLs dos demais anos poderão ser incluídos gradativamente por meio de atividades regulares do GPE, uma vez que é preciso uma revisão nas classificações que são incorporadas ao banco, para garantir a sua qualidade.

No caso de 2022, foram adicionadas 37 classificações de PL que estavam faltando, e revisados os códigos de uma lista de 30 em que havia dúvida de classificação.

As variáveis do banco de dados são: ano, número, ementa, indexação, texto completo do PL (incluindo artigos e justificção), artigos do PL, justificção do PL, classificação de relação com saúde, grupo temático de assistência, grupo temático de prevenção, grupo temático de gestão e recursos, grupo temático de direitos e responsabilidades, e de outros temas da saúde (sem ocorrência no banco de dados), além de cinco variáveis para registro de até cinco subtemas por PL.

#### 4.c. Experimentos com modelos de ML e ANN

Os experimentos exploraram o desempenho de classificação de versões de modelos de ML e de redes neurais a partir de numerosas variações em sua modelagem e da fonte de dados, incluindo texto completo dos PLs, ementas, indexações e combinações destas variáveis. O Quadro 2 sintetiza os 50 experimentos realizados e seus resultados, medidos por métricas de acurácia, *recall*, precisão e *F1 score*.



## Quadro 2. Resultados de experimentos de ML e redes neurais.

#	01	#	02
Nome	1. SGD X - text - y - saúde - 11 a 14	Nome	2. SGD X - text - y - saúde - 2022
Modelo	SGD	Modelo	SGD
Período	2011 a 2014	Período	2022
Objetivo	Classificar PLs da saúde	Objetivo	Classificar PLs da saúde
X	PL completo	X	PL completo
Y	Saúde	Y	Saúde
Parâmetros	TF-IDF; pipeline	Parâmetros	TF-IDF; pipeline
Acurácia	91	Acurácia	88
Recall	76 (saúde)	Recall	68 (saúde)
Precisão	88 (saúde)	Precisão	88 (saúde)
F1 score	82 (saúde)	F1 score	76 (saúde)
TN/FP/FN/TP	1049 / 43 / 100 / 323	TN/FP/FN/TP	337 / 14 / 48 / 100
Observações		Observações	
#	03	#	04
Nome	3. SGD TFIDF texto e saúde 11 a 14 e 22	Nome	4. SGD - texto - y -saúde -11 a 14 e 22-Countvectorizer
Modelo	SGD	Modelo	SGD
Período	2011 a 2014 e 2022	Período	2011 a 2014 e 2022
Objetivo	Classificar PLs da saúde	Objetivo	Classificar PLs da saúde
X	PL completo	X	PL completo
Y	Saúde	Y	Saúde
Parâmetros	TF-IDF; Countvectorizer; gridsearch	Parâmetros	Countvectorizer sem TF-IDF
Acurácia	91,60	Acurácia	88,18
Recall	-	Recall	77,62
Precisão	-	Precisão	80,14
F1 score	91,40	F1 score	78,86
TN/FP/FN/TP	1398 / 44 / 125 / 447	TN/FP/FN/TP	1332 / 110 / 128 / 444
Observações	Código não calculou outras métricas.	Observações	

#	05	#	06
Nome	5. SGD - texto - y - saúde - 11 a 14 e 22 - TF-IDF	Nome	6. SGD Cópia 01 - texto - y - saúde - 11 a 14 e 22 - TF-IDF
Modelo	SGD	Modelo	SGD
Período	2011 a 2014 e 2022	Período	2011 a 2014 e 2022
Objetivo	Classificar PLs da saúde	Objetivo	Classificar PLs da saúde
X	PL completo	X	PL completo
Y	Saúde	Y	Saúde
Parâmetros	TF-IDF	Parâmetros	TF-IDF
Acurácia	90,96	Acurácia	90,61
Recall	90,96	Recall	90,61
Precisão	90,84	Precisão	90,49
F1 score	90,81	F1 score	90,44
TN/FP/FN/TP	1380 / 62 / 120 / 452	TN/FP/FN/TP	1379 / 63 / 126 / 446
Observações	Listou FP e FN	Observações	Listou FP e FN (modelo ... cópia.joblib)
#	07	#	08
Nome	7. SGD Cópia 02- texto - y - saúde - 11 a 14 e 22 - TF-IDF	Nome	8. SGD Cópia 03 - texto - y - saúde - 11 a 14 e 22 - TF-IDF
Modelo	SGD	Modelo	SGD
Período	2011 a 2014 e 2022	Período	2011 a 2014 e 2022
Objetivo	Classificar PLs da saúde	Objetivo	Classificar PLs da saúde
X	PL completo	X	PL completo
Y	Saúde	y	Saúde
Parâmetros	TF-IDF	Parâmetros	TF-IDF
Acurácia	-	Acurácia	90,96
Recall	-	Recall	90,96
Precisão	-	Precisão	90,84
F1 score	-	F1 score	90,81
TN/FP/FN/TP	-	TN/FP/FN/TP	1380 / 62 / 120 / 452
Observações	Métricas não calculadas. Importado modelo e salvos os discordantes.	Observações	Salvo modelo (usado em 7).

#	09	#	10
Nome	9. ANN X - texto - y - saúde - 11 a 14	Nome	10. ANN X - ementa - y - saúde - 11 a 14
Modelo	ANN	Modelo	ANN
Período	2011 a 2014	Período	2011 a 2014
Objetivo	Classificar PLs da saúde	Objetivo	Classificar PLs da saúde
X	PL completo	X	Ementa do PL
y	Saúde	y	Saúde
Parâmetros	TF-IDF; 10 épocas; batch=32	Parâmetros	TF-IDF; 10 épocas; batch=32
Acurácia	91,35	Acurácia	83,63
Recall	79,55	Recall	72,58
Precisão	89,84	Precisão	71,93
F1 score	84,38	F1 score	72,25
TN/FP/FN/TP	1030 / 40 / 91 / 354	TN/FP/FN/TP	944 / 126 / 122 / 323
Observações		Observações	
#	11.	#	12
Nome	11. ANN X - indexacao - y - saúde - 11 a 2014	Nome	12. ANN X - ementa e indexacao - y - saúde - 11 a 14
Modelo	ANN	Modelo	ANN
Período	2011 a 2014	Período	2011 a 2014
Objetivo	Classificar PLs da saúde	Objetivo	Classificar PLs da saúde
X	Indexação do PL	X	Ementa e Indexação do PL (concatenados)
y	Saúde	y	Saúde
Parâmetros	TF-IDF; 10 épocas; batch=32	Parâmetros	TF-IDF; 10 épocas; batch=32
Acurácia	86,60	Acurácia	86,86
Recall	72,58	Recall	72,35
Precisão	79,95	Precisão	80,90
F1 score	76,08	F1 score	76,36
TN/FP/FN/TP	989 / 81 / 122 / 323	TN/FP/FN/TP	994 / 76 / 123 / 322
Observações		Observações	

#	13	#	14
Nome	13. ANN X - ementa e indexação múltipla entrada- y - saúde - 11 a 14	Nome	14. ANN X - artigos - y - saúde - 11 a 14
Modelo	ANN	Modelo	ANN
Período	2011 a 2014	Período	2011 a 2014
Objetivo	Classificar PLs da saúde	Objetivo	Classificar PLs da saúde
X	Ementa e Indexação do PL (múltiplas entradas)	X	Artigos do PL
y	Saúde	y	Saúde
Parâmetros	TF-IDF; 10 épocas; batch=32	Parâmetros	TF-IDF; 10 épocas; batch=32
Acurácia	86	Acurácia	87,32
Recall	79 (saúde)	Recall	71,68
Precisão	83 (saúde)	Precisão	82,85
F1 score	79 (saúde)	F1 score	76,86
TN/FP/FN/TP	1003 / 67 / 114 / 331	TN/FP/FN/TP	1004 / 66 / 126 / 319
Observações	Código para dividir a variável “texto”.	Observações	
#	15	#	16
Nome	15. ANN X - artigos e indexação - y - saúde - 11 a 14	Nome	16. ANN X - múltiplos - y - saúde - 11 a 14 - versão 01
Modelo	ANN	Modelo	ANN
Período	2011 a 2014	Período	2011 a 2014
Objetivo	Classificar PLs da saúde	Objetivo	Classificar PLs da saúde
X	Artigos e Indexação do PL concatenados)	X	Ementa, Indexação, Artigos e Justificação do PL (múltiplas entradas)
y	Saúde	y	Saúde
Parâmetros	TF-IDF; 10 épocas; batch=32	Parâmetros	TF-IDF; 10 épocas; batch=32
Acurácia	88,11	Acurácia	91
Recall	74,15	Recall	77 (saúde)
Precisão	83,54	Precisão	92 (saúde)
F1 score	78,57	F1 score	84 (saúde)
TN/FP/FN/TP	1005 / 65 / 115 / 330	TN/FP/FN/TP	1039 / 31 / 101 / 344
Observações		Observações	Código para dividir a variável “texto”.

#	17	#	18
Nome	17. ANN X - múltiplos - y - saúde - 11 a 14 - versão 02	Nome	18. ANN oversampling adaptativo - text - y - saúde - 11 a 14
Modelo	ANN	Modelo	ANN
Período	2011 a 2014	Período	2011 a 2014
Objetivo	Classificar PLs da saúde	Objetivo	Classificar PLs da saúde
X	Ementa, Indexação, Artigos e Justificação do PL (múltiplas entradas)	X	PL completo
y	Saúde	y	Saúde
Parâmetros	TF-IDF; 10 épocas; batch=32	Parâmetros	Tokenizer e não TF-IDF; 10 épocas; batch=32, oversampling
Acurácia	91	Acurácia	-
Recall	82 (saúde)	Recall	-
Precisão	88 (saúde)	Precisão	-
F1 score	85 (saúde)	F1 score	-
TN/FP/FN/TP	1021 / 49 / 80 / 365	TN/FP/FN/TP	-
Observações	Código não difere do 8. Aleatória – na inicialização dos pesos da rede. Modelo salvo.	Observações	Experimento apresentou erro no código. Não concluído.
#	19	#	20
Nome	19. ANN SMOTE - text - y - saúde - 11 a 14	Nome	20. ANN weights - text - y - saúde - 11 a 14
Modelo	ANN	Modelo	ANN
Período	2011 a 2014	Período	2011 a 2014
Objetivo	Classificar PLs da saúde	Objetivo	Classificar PLs da saúde
X	PL completo	X	PL completo
y	Saúde	y	Saúde
Parâmetros	Stratify, TF-IDF, SMOTE, 10 épocas; batch=32	Parâmetros	Stratify, TF-IDF, class weights, 10 épocas; batch=32
Acurácia	72,07	Acurácia	89,70
Recall	0	Recall	76,12
Precisão	0	Precisão	85,41
F1 score	0	F1 score	80,49
TN/FP/FN/TP	1092 / 0 / 423 / 0	TN/FP/FN/TP	1037 / 55 / 101 / 322
Observações	Deve haver erro no código	Observações	

#	21	#	22
Nome	21. ANN X - múltiplos - y - grupos - 11 a 14	Nome	22. ANN X - múltiplos - y - grupos / saúde - 11 a 14
Modelo	ANN	Modelo	ANN
Período	2011 a 2014	Período	2011 a 2014
Objetivo	Classificar PLs de cada grupo temático e depois agregar como saúde	Objetivo	Classificar PLs de cada grupo temático e depois agregar como saúde
X	Ementa, Index, Artigos e Justifi...(múltiplas entradas)	X	Ementa, Index..., Artigos e Justifi...(múltiplas entradas)
y	Grupo de Direitos e Responsabilidades	y	Grupo de Direitos e Responsabilidades
Parâmetros	TF-IDF; 10 épocas; batch=32, random_state	Parâmetros	TF-IDF; 10 épocas; batch=32, random_state,
Acurácia	-	Acurácia	asa 93,00 / ppc 75,04 / grp 93,86 / dir 99,00 /
Recall	-	Recall	asa 00,00 / ppc 18,53 / grp 01,96 / dir 86,53
Precisão	-	Precisão	asa 00,00 / ppc 18,53 / grp 02,27 / dir 98,90
F1 score	-	F1 score	asa 00,00 / ppc 18,53 / grp 02,10 / dir 92,30
TN/FP/FN/TP	-	TN/FP/FN/TP	asa 1409 / 50 / 56 / 0; ppc 1094 / 189 / 189 / 43; grp 1421 / 43 / 50 / 1; dir 1410 / 1 / 14 / 90 .
Observações	Funcionou até certo ponto, mas falhou - falta de RAM	Observações	Usei modelos salvos para economizar RAM. Agregação para saúde deu erro. Desempenho inferior ao original. Calculado ROC para grupos.
#	23	#	24
Nome	23. ANN X - text - y - asa - 11 a 14	Nome	24. ANN X - múltiplos - y - asa - 11 a 14
Modelo	ANN	Modelo	ANN
Período	2011 a 2014	Período	2011 a 2014
Objetivo	Classificar PLs do grupo de Assistência à saúde	Objetivo	Classificar PLs do grupo de Assistência à saúde
X	PL completo	X	Ementa, Indexação, Artigos e Justificação do PL (múltiplas entradas)
y	Grupo de Assistência à saúde	y	Grupo de Assistência à saúde
Parâmetros	TF-IDF; 10 épocas; batch=32, random_state	Parâmetros	TF-IDF; 10 épocas; batch=32, random_state
Acurácia	97,09	Acurácia	98
Recall	31,57	Recall	44 (saúde)
Precisão	78,26	Precisão	86 (saúde)
F1 score	45,00	F1 score	58 (saúde)
TN/FP/FN/TP	1453 / 5 / 39 / 18	TN/FP/FN/TP	1454 / 4 / 32 / 25

Observações		Observações	
#	25	#	26
Nome	25. ANN X - ementa - y - asa - 11 a 14	Nome	26. ANN X - indexacao - y - asa 11 a 14
Modelo	ANN	Modelo	ANN
Período	2011 a 2014	Período	2011 a 2014
Objetivo	Classificar PLs do grupo de Assistência à saúde	Objetivo	Classificar PLs do grupo de Assistência à saúde
X	Ementa	X	Indexação
y	Grupo de Assistência à saúde	y	Grupo de Assistência à saúde
Parâmetros	TF-IDF; 10 épocas; batch=32, random_state	Parâmetros	TF-IDF; 10 épocas; batch=32, random_state
Acurácia	97,29	Acurácia	97,02
Recall	36,84	Recall	38,59
Precisão	70,00	Precisão	68,75
F1 score	48,27	F1 score	49,43
TN/FP/FN/TP	1449 / 9 / 36 / 21	TN/FP/FN/TP	1448 / 10 / 35 / 22
Observações		Observações	
#	27	#	28
Nome	27. ANN X - ementa e indexacao - y - asa - 11 a 14	Nome	28. ANN X - ementa e indexação múltipla entrada - y - asa - 11 a 14
Modelo	ANN	Modelo	ANN
Período	2011 a 2014	Período	2011 a 2014
Objetivo	Classificar PLs do grupo de Assistência à saúde	Objetivo	Classificar PLs do grupo de Assistência à saúde
X	Ementa e Indexação do PL (concatenados)	X	Ementa e Indexação do PL (múltipla entrada)
y	Grupo de Assistência à saúde	y	Grupo de Assistência à saúde
Parâmetros	TF-IDF; 10 épocas; batch=32, random_state	Parâmetros	TF-IDF; 10 épocas; batch=32, random_state
Acurácia	97,27	Acurácia	97
Recall	36,84	Recall	39 (saúde)
Precisão	70,27	Precisão	76 (saúde)
F1 score	55,31	F1 score	51 (saúde)
TN/FP/FN/TP	1447 / 11 / 31 / 26	TN/FP/FN/TP	1451 / 7 / 35 / 22
Observações		Observações	

#	29	#	30
Nome	29. ANN X - múltiplos - y - dir - 11 a 14	Nome	30. ANN X - text - y - dir - 11 a 14
Modelo	ANN	Modelo	ANN
Período	2011 a 2014	Período	2011 a 2014
Objetivo	Classificar PLs do grupo de Direitos e Responsabilidades	Objetivo	Classificar PLs do grupo de Direitos e Responsabilidades
X	Ementa, Indexação, Artigos e Justificação do PL (múltiplas entradas)	X	PL completo
y	Grupo de Direitos e Responsabilidades	y	Grupo de Direitos e Responsabilidades
Parâmetros	TF-IDF; 10 épocas; batch=32, random_state	Parâmetros	TF-IDF; 10 épocas; batch=32, random_state
Acurácia	95	Acurácia	94,25
Recall	45 (saúde)	Recall	40,47
Precisão	89 (saúde)	Precisão	80,95
F1 score	60 (saúde)	F1 score	53,96
TN/FP/FN/TP	1382 / 7 / 69 / 57	TN/FP/FN/TP	1377 / 12 / 75 / 51
Observações	Também usou F1 como métrica do history.	Observações	
#	31	#	32
Nome	31. ANN X - ementa - y - dir - 11 a 14	Nome	32. ANN X - indexacao - y - dir - 11 a 14
Modelo	ANN	Modelo	ANN
Período	2011 a 2014	Período	2011 a 2014
Objetivo	Classificar PLs do grupo de Direitos e Responsabilidades	Objetivo	Classificar PLs do grupo de Direitos e Responsabilidades
X	Ementa do PL	X	Indexação do PL
y	Grupo de Direitos e Responsabilidades	y	Grupo de Direitos e Responsabilidades
Parâmetros	TF-IDF; 10 épocas; batch=32, random_state	Parâmetros	TF-IDF; 10 épocas; batch=32, random_state
Acurácia	92,34	Acurácia	93,39
Recall	43,65	Recall	53,17
Precisão	55,00	Precisão	62,03
F1 score	48,67	F1 score	57,26
TN/FP/FN/TP	1344 / 45 / 71 / 55	TN/FP/FN/TP	1348 / 41 / 59 / 67
Observações		Observações	

#	33	#	34
---	----	---	----



<b>Nome</b>	33. ANN X - ementa e indexacao - y - dir - 11 a 14	<b>Nome</b>	34. ANN X - ementa e indexação múltipla entrada- y - dir - 11 a 14
<b>Modelo</b>	ANN	<b>Modelo</b>	ANN
<b>Período</b>	2011 a 2014	<b>Período</b>	2011 a 2014
<b>Objetivo</b>	Classificar PLs do grupo de Direitos e Responsabilidades	<b>Objetivo</b>	Classificar PLs do grupo de Direitos e Responsabilidades
<b>X</b>	Ementa e Indexação do PL (concatenados)	<b>X</b>	Ementa e Indexação do PL (múltipla entrada)
<b>y</b>	Grupo de Direitos e Responsabilidades	<b>y</b>	Grupo de Direitos e Responsabilidades
<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32, random_state	<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32, random_state
<b>Acurácia</b>	93,92	<b>Acurácia</b>	94
<b>Recall</b>	46,82	<b>Recall</b>	45 (saúde)
<b>Precisão</b>	70,23	<b>Precisão</b>	70 (saúde)
<b>F1 score</b>	56,19	<b>F1 score</b>	55 (saúde)
<b>TN/FP/FN/TP</b>	1364 / 25 / 67 / 59	<b>TN/FP/FN/TP</b>	1364 / 25 / 69 / 57
<b>Observações</b>		<b>Observações</b>	
<b>#</b>	35	<b>#</b>	36
<b>Nome</b>	35. ANN X - text - y - grp - 11 a 14	<b>Nome</b>	36. ANN X - ementa - y - grp - 11 a 14
<b>Modelo</b>	ANN	<b>Modelo</b>	ANN
<b>Período</b>	2011 a 2014	<b>Período</b>	2011 a 2014
<b>Objetivo</b>	Classificar PLs do grupo de Gestão e Recursos	<b>Objetivo</b>	Classificar PLs do grupo de Gestão e Recursos
<b>X</b>	PL completo	<b>X</b>	Ementa do PL
<b>y</b>	Grupo de Gestão e Recursos	<b>y</b>	Grupo de Gestão e Recursos
<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32, random_state	<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32, random_state
<b>Acurácia</b>	96,36	<b>Acurácia</b>	96,17
<b>Recall</b>	19,04	<b>Recall</b>	20,63
<b>Precisão</b>	75,00	<b>Precisão</b>	61,90
<b>F1 score</b>	30,37	<b>F1 score</b>	30,95
<b>TN/FP/FN/TP</b>	1448 / 4 / 51 / 12	<b>TN/FP/FN/TP</b>	1444 / 8 / 50 / 13
<b>Observações</b>		<b>Observações</b>	
<b>#</b>	37	<b>#</b>	38
<b>Nome</b>	37. ANN X - indexacao - y - grp - 11 a 14	<b>Nome</b>	38. ANN X - ementa e indexacao - y - grp - 11 a 14

<b>Modelo</b>	ANN	<b>Modelo</b>	ANN
<b>Período</b>	2011 a 2014	<b>Período</b>	2011 a 2014
<b>Objetivo</b>	Classificar PLs do grupo de Gestão e Recursos	<b>Objetivo</b>	Classificar PLs do grupo de Gestão e Recursos
<b>X</b>	Indexação do PL	<b>X</b>	Ementa e Indexação do PL (concatenados)
<b>y</b>	Grupo de Gestão e Recursos	<b>y</b>	Grupo de Gestão e Recursos
<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32, random_state	<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32, random_state
<b>Acurácia</b>	95,64	<b>Acurácia</b>	95,97
<b>Recall</b>	20,63	<b>Recall</b>	17,46
<b>Precisão</b>	44,82	<b>Precisão</b>	55,00
<b>F1 score</b>	28,26	<b>F1 score</b>	26,50
<b>TN/FP/FN/TP</b>	1436 / 16 / 50 / 13	<b>TN/FP/FN/TP</b>	1443 / 9 / 52 / 11
<b>Observações</b>		<b>Observações</b>	
<b>#</b>	39	<b>#</b>	40
<b>Nome</b>	39. ANN X - ementa e indexação múltipla entrada- y - grp - 11 a 14	<b>Nome</b>	40. ANN X - múltiplos - y - grp - 11 a 14
<b>Modelo</b>	ANN	<b>Modelo</b>	ANN
<b>Período</b>	2011 a 2014	<b>Período</b>	2011 a 2014
<b>Objetivo</b>	Classificar PLs do grupo de Gestão e Recursos	<b>Objetivo</b>	Classificar PLs do grupo de Gestão e Recursos
<b>X</b>	Ementa e Indexação do PL (múltipla entrada)	<b>X</b>	Ementa, Indexação, Artigos e Justificação do PL (múltiplas entradas)
<b>y</b>	Grupo de Gestão e Recursos	<b>y</b>	Grupo de Gestão e Recursos
<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32, random_state	<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32, random_state
<b>Acurácia</b>	96	<b>Acurácia</b>	96
<b>Recall</b>	16 (saúde)	<b>Recall</b>	16 (saúde)
<b>Precisão</b>	67 (saúde)	<b>Precisão</b>	77 (saúde)
<b>F1 score</b>	26 (saúde)	<b>F1 score</b>	26 (saúde)
<b>TN/FP/FN/TP</b>	1447 / 5 / 53 / 10	<b>TN/FP/FN/TP</b>	1449 / 3 / 53 / 10
<b>Observações</b>		<b>Observações</b>	
<b>#</b>	41	<b>#</b>	42
<b>Nome</b>	41. ANN X - text - y - ppc - 11 a 14	<b>Nome</b>	42. ANN X - ementa - y - ppc - 11 a 14

<b>Modelo</b>	ANN	<b>Modelo</b>	ANN
<b>Período</b>	2011 a 2014	<b>Período</b>	2011 a 2014
<b>Objetivo</b>	Classificar PLs do grupo de Prevenção	<b>Objetivo</b>	Classificar PLs do grupo de Prevenção
<b>X</b>	PL completo	<b>X</b>	Ementa
<b>y</b>	Grupo de Prevenção	<b>y</b>	Grupo de Prevenção
<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32, random_state	<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32, random_state
<b>Acurácia</b>	92,87	<b>Acurácia</b>	89,70
<b>Recall</b>	66,07	<b>Recall</b>	51,54
<b>Precisão</b>	82,87	<b>Precisão</b>	71,77
<b>F1 score</b>	73,52	<b>F1 score</b>	60,00
<b>TN/FP/FN/TP</b>	1257 / 31 / 77 / 150	<b>TN/FP/FN/TP</b>	1242 / 46 / 110 / 117
<b>Observações</b>		<b>Observações</b>	
<b>#</b>	43	<b>#</b>	44
<b>Nome</b>	43. ANN X - múltiplos - y - ppc - 11 a 14	<b>Nome</b>	44. ANN X - indexacao - y - ppc - 11 a 14
<b>Modelo</b>	ANN	<b>Modelo</b>	ANN
<b>Período</b>	2011 a 2014	<b>Período</b>	2011 a 2014
<b>Objetivo</b>	Classificar PLs do grupo de Prevenção	<b>Objetivo</b>	Classificar PLs do grupo de Prevenção
<b>X</b>	Ementa, Indexação, Artigos e Justificação do PL (múltiplas entradas)	<b>X</b>	Indexação
<b>y</b>	Grupo de Prevenção	<b>y</b>	Grupo de Prevenção
<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32, random_state	<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32, random_state
<b>Acurácia</b>	93	<b>Acurácia</b>	88,91
<b>Recall</b>	70 (saúde)	<b>Recall</b>	61,67
<b>Precisão</b>	83 (saúde)	<b>Precisão</b>	63,34
<b>F1 score</b>	76 (saúde)	<b>F1 score</b>	62,50
<b>TN/FP/FN/TP</b>	1255 / 33 / 68 / 159	<b>TN/FP/FN/TP</b>	1207 / 81 / 87 / 140
<b>Observações</b>		<b>Observações</b>	

<b>#</b>	45	<b>#</b>	46
----------	----	----------	----

<b>Nome</b>	45. ANN X - ementa e indexacao - y - ppc - 11 a 14	<b>Nome</b>	46. ANN X - ementa e indexação múltipla entrada- y - ppc - 11 a 14
<b>Modelo</b>	ANN	<b>Modelo</b>	ANN
<b>Período</b>	2011 a 2014	<b>Período</b>	2011 a 2014
<b>Objetivo</b>	Classificar PLs do grupo de Prevenção	<b>Objetivo</b>	Classificar PLs do grupo de Prevenção
<b>X</b>	Ementa e Indexação do PL (concatenados)	<b>X</b>	Ementa e Indexação do PL (múltipla entrada)
<b>y</b>	Grupo de Prevenção	<b>y</b>	Grupo de Prevenção
<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32, random_state	<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32, random_state
<b>Acurácia</b>	90,36	<b>Acurácia</b>	91
<b>Recall</b>	66,51	<b>Recall</b>	65 (saúde)
<b>Precisão</b>	68,32	<b>Precisão</b>	75 (saúde)
<b>F1 score</b>	67,41	<b>F1 score</b>	69 (saúde)
<b>TN/FP/FN/TP</b>	1218 / 70 / 76 / 151	<b>TN/FP/FN/TP</b>	1238 / 50 / 80 / 147
<b>Observações</b>		<b>Observações</b>	
<b>#</b>	47	<b>#</b>	48
<b>Nome</b>	47. ANN X - texto - y - saúde – 2022 – limpo	<b>Nome</b>	48. ANN X - texto - y - saúde – 2022 – com acentos
<b>Modelo</b>	ANN	<b>Modelo</b>	ANN
<b>Período</b>	2022	<b>Período</b>	2022
<b>Objetivo</b>	Classificar PLs da saúde	<b>Objetivo</b>	Classificar PLs da saúde
<b>X</b>	PL completo	<b>X</b>	PL completo
<b>Y</b>	Saúde	<b>y</b>	Saúde
<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32	<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32
<b>Acurácia</b>	86,37	<b>Acurácia</b>	87,17
<b>Recall</b>	64,77	<b>Recall</b>	67,29
<b>Precisão</b>	89,56	<b>Precisão</b>	89,91
<b>F1 score</b>	75,82	<b>F1 score</b>	76,97
<b>TN/FP/FN/TP</b>	328 / 12 / 56 / 103	<b>TN/FP/FN/TP</b>	328 / 12 / 52 / 107
<b>Observações</b>	Sem sinais ortográficos nos PLs	<b>Observações</b>	Com sinais ortográficos nos PLs

<b>#</b>	49	<b>#</b>	50
----------	----	----------	----

<b>Nome</b>	49. ANN X - texto - y - saúde – 11 a 14 e 22 – com acentos	<b>Nome</b>	50. ANN X - text - y - saúde - 11 a 14 e 22 - limpo
<b>Modelo</b>	ANN	<b>Modelo</b>	ANN
<b>Período</b>	2011 a 2014 e 2022	<b>Período</b>	2011 a 2014 e 2022
<b>Objetivo</b>	Classificar PLs da saúde	<b>Objetivo</b>	Classificar PLs da saúde
<b>X</b>	PL completo	<b>X</b>	PL completo
<b>Y</b>	Saúde	<b>Y</b>	Saúde
<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32	<b>Parâmetros</b>	TF-IDF; 10 épocas; batch=32
<b>Acurácia</b>	90,41	<b>Acurácia</b>	90,11
<b>Recall</b>	78,26	<b>Recall</b>	76,39
<b>Precisão</b>	86,35	<b>Precisão</b>	87,22
<b>F1 score</b>	82,11	<b>F1 score</b>	81,45
<b>TN/FP/FN/TP</b>	1378 / 70 / 123 / 443	<b>TN/FP/FN/TP</b>	1378 / 64 / 135 / 437
<b>Observações</b>	Faltou RAM, mesmo com ambiente T4! Funcionou com <b>TPU</b>	<b>Observações</b>	

Nos experimentos de 01 a 08 foi utilizado o modelo SGD no conjunto de dados de 2011 a 2014, de 2022 e na combinação de dados de 2011 a 2014 e 2022; utilizando o texto completo dos PLs como variável independente (X) e a variável sobre saúde como a variável dependente (y).

Observa-se que a acurácia variou em torno de 90% e o *F1 Score* chegou a 91,4% no período combinado.

Nos experimentos de 09 a 50 foi utilizado o modelo ANN nos períodos já mencionados, com muitas variações nas variáveis X e y. Em geral, obtiveram elevados desempenhos nas métricas e comparáveis ao do SGD.

Por exemplo, quando se utilizou o texto completo dos PLs como variável X e a variável sobre saúde como a variável y, o ANN obteve resultado um pouco superior em relação ao SGD no período de 2011 a 2014 (experimentos 1 e 9). O ANN também foi um pouco superior nesses parâmetros para o ano de 2022 (experimentos 2 e 48). Já para o período combinado de 2011 a 2014 e 2022, o SGD teve um desempenho um pouco superior ao do ANN (experimentos 3 a 8, 49 e 50).

Para o período de 2011 a 2014 foram realizados mais experimentos, considerando que esse banco foi objeto de uma revisão mais detalhada de sua qualidade.

Destaca-se que o desempenho utilizando o texto completo dos PLs como variável X (experimento 9) foi superior a utilização isolada ou de pares específicos das variáveis ementa, indexação, artigos e justificação (experimentos 10 a 15); contudo, o uso do texto completo foi um pouco inferior à utilização de todas essas variáveis como múltiplas entradas na rede neural (experimentos 16 e 17).

Os experimentos 18 e 19 tentaram abordar técnicas para lidar com o grande desbalanceamento do banco de dados, entretanto não se obteve êxito, carecendo de aperfeiçoamentos futuros na codificação. O uso de pesos (experimento 20) não ampliou o desempenho.

Os experimentos de 21 a 46 utilizaram grupos temáticos da saúde como variável  $y$ . O desempenho foi inferior ao obtido quando  $y$  foi a variável sobre saúde, sendo o *recall* a métrica com os desempenhos mais baixos.

Os experimentos 47 e 48 verificaram o efeito do uso de variáveis de texto com ou sem acentuação gráfica em português e o desempenho do modelo que utilizou a acentuação foi um pouco superior.

#### 4.d. Análise da estrutura lógica da tipologia de classificação

Estudo realizado em 2022 sugeriu revisão da estrutura lógica da tipologia de classificação de projetos de lei relacionados à saúde, em geral, mantendo os mesmos grupos temáticos e subtemas e excluindo o próximo nível hierárquico, o de destaques temáticos. Desse modo, seriam eliminados 21 destaques, reduzindo o número de categorias da tipologia.

Na pesquisa atual, os experimentos sobre grupos temáticos (21 a 46) permitiram observar que para este nível hierárquico, há necessidade de modificações na tipologia, em virtude do baixo *recall* evidenciado, principalmente para o grupo de gestão e recursos.

A verificação das discordâncias entre classificações de humanos e de modelos sugere que o modelo SGD capta bem expressões de saúde, mas quando tais expressões estão na justificativa apenas como argumento comparativo,

terminam por provocar erros de classificação. Por outro lado, o modelo ajuda a captar erros humanos, particularmente na área de riscos à saúde (ex.: ambiental, trânsito) e de lógica da tipologia (que diferencia benefícios por deficiências e por doença).

## 5 Conclusões e Sugestões

Essa pesquisa focou na aplicação de técnicas de aprendizado de máquina e redes neurais artificiais para a classificação temática de PLs, especialmente na área da saúde. A análise dos resultados revelou uma performance consistentemente elevada em termos de precisão e acurácia, aproximando-se de 90% e 97% respectivamente, nos modelos de melhor desempenho testados ao longo dos experimentos.

No entanto, o desempenho em termos de *recall* foi variável, destacando-se como uma área de potencial melhoria, principalmente nos grupos temáticos mais específicos, como gestão, recursos e pesquisa, que apresentaram *recall* significativamente menor.

A complexidade e a variabilidade dos textos legislativos, combinadas com a precisão requerida na classificação temática, ressaltam a importância de continuar refinando os modelos e as abordagens de treinamento.

Provavelmente, o uso futuro de métodos mais avançados, como transformadores (a exemplo do BERT) possibilitará redução de problemas associados à interpretação da justificação dos PLs.



Além disso, a análise da estrutura lógica da tipologia utilizada para a classificação indicou que ajustes podem ser benéficos, como simplificar a estrutura ao eliminar categorias de menor relevância, para aprimorar tanto a precisão quanto a usabilidade dos modelos.

Por fim, destacam-se avanços obtidos com as atividades de ampliação da revisão da literatura e de preparação do banco de dados.

No primeiro caso, foram identificados estudos com aplicação direta no Legislativo, e no segundo, foram superados gargalos críticos, como a conversão de arquivos em “pdf” para “txt”, e desenvolvidos métodos que garantem a expansão de novas etapas desta agenda de pesquisa.