



Ranniery Dias de Brito

ANÁLISE E PREDIÇÃO NAS VOTAÇÕES DE LEIS FEDERAIS NA CÂMARA DOS DEPUTADOS

Recife

2022

Ranniery Dias de Brito

ANÁLISE E PREDIÇÃO NAS VOTAÇÕES DE LEIS FEDERAIS NA CÂMARA DOS DEPUTADOS

Monografia apresentada ao Curso de Bacharelado em Ciência da Computação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

Universidade Federal Rural de Pernambuco – UFRPE

Departamento de Computação

Curso de Bacharelado em Ciência da Computação

Orientador: Kellyton dos Santos Brito

Recife

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

B862a BRITO, RANNIERY
ANÁLISE E PREDIÇÃO NAS VOTAÇÕES DE LEIS FEDERAIS NA CÂMARA DOS DEPUTADOS /
RANNIERY BRITO. - 2022.
37 f. : il.

Orientador: Kellyton dos Santos Brito.
Inclui referências.

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Rural de Pernambuco,
Bacharelado em Ciência da Computação, Recife, 2022.

1. Processo legislativo. 2. Predição. 3. Aprendizagem de máquina. 4. Deep Learning. 5. Análise
descritiva. I. Brito, Kellyton dos Santos, orient. II. Título

CDD 004



**MINISTÉRIO DA EDUCAÇÃO E DO ESPORTO
UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO (UFRPE)
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

<http://www.bcc.ufrpe.br>

FICHA DE APROVAÇÃO DO TRABALHO DE CONCLUSÃO DE CURSO

Trabalho defendido por Ranniery Dias de Brito às 10 horas do dia 27 de maio de 2022, no link <https://meet.google.com/zbf-iyat-jhk>, como requisito para conclusão do curso de Bacharelado em Ciência da Computação da Universidade Federal Rural de Pernambuco, intitulado “Análise e Predição das Votações de Leis Federais na Câmara dos Deputados”, orientado por Kellyton dos Santos Brito e aprovado pela seguinte banca examinadora:

Kellyton dos Santos Brito
DC/UFRPE

George Cabral
DC/UFRPE



**MINISTÉRIO DA EDUCAÇÃO E DO ESPORTO
UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO (UFRPE)
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

<http://www.bcc.ufrpe.br>

FICHA DE APROVAÇÃO DO TRABALHO DE CONCLUSÃO DE CURSO

Trabalho defendido por Ranniery Dias de Brito às 10 horas do dia 27 de maio de 2022, no link <https://meet.google.com/zbf-iyat-jhk>, como requisito para conclusão do curso de Bacharelado em Ciência da Computação da Universidade Federal Rural de Pernambuco, intitulado “Análise e Predição das Votações de Leis Federais na Câmara dos Deputados”, orientado por Kellyton dos Santos Brito e aprovado pela seguinte banca examinadora:

Kellyton dos Santos Brito
DC/UFRPE

George Cabral
DC/UFRPE

Aos meus pais, que desde o princípio ouviram todas minhas sonhos e inquietações.

Agradecimentos

Agradeço à todos os que até aqui me acompanharam e aos que foram perdidos no meio do caminho, Em especial, agradeço imensamente aos meus pais, Ivonete Dias e Roberto Brito que desde o começo da minha vida acreditaram em mim e me educado com amor e carinho, e me prepararam para o mundo com seus ensinamentos e conselhos.

Obrigado ao meu irmão Washington Dias, que sempre foi uma inspiração para mim, despertou meu espírito científico, ainda que de maneira amadora e empírica, e sempre me incentivou desde nossa infância. Sou grato a todos os membros da minha família que me disseram palavras de força e perseverança em minha jornada acadêmica.

Agradeço ao professor Drº Kellyton dos Santos Brito, pelo esforço, dedicação, paciência, compreensão, ímpeto e orientação excepcional. Muito obrigado por me acompanhar nos mais difíceis momentos desta pesquisa, e também nos mais empolgantes. Você foi parte fundamento para conclusão desse trabalho.

Gostaria de agradecer também aos professores que fizeram parte da minha vida acadêmica, e que me fizeram aprender a ser um cientista. Hoje me torno a soma de cada um de vocês. Agradeço a todos os que fazem parte do corpo docente do curso de Bacharelado em Ciência da Computação, aos funcionários administrativos, ao Departamento de computação, a toda a UFRPE por ser uma das minhas casas durante todos esses anos. Agradeço a Sandra Xavier, que sempre foi muito prestativa, competente e atenciosa.

Quero também agradecer a minha companheira, Maria Luiza, que foi meu porto seguro durante minha vida acadêmica, me incentivando, me ouvindo, não me deixando desistir, corrigindo meus textos e me ajudando de toda forma possível. Eu não teria chegado até aqui sem o seu apoio e atenção.

A todos os meus amigos, e colegas, da universidade, trabalho e da vida. Muito obrigado, vocês também me ajudaram mais do que imaginam, sou muito agradecido a todo o apoio, risadas e incentivos que vocês me proporcionaram.

A todos, sou muito grato.

“Se eu vi mais longe, foi por estar sobre ombros de gigantes.”
(Isaac Newton)

Resumo

Este estudo tem por objetivo analisar algoritmos de aprendizagem de máquina e *deep learning* para a tarefa de previsibilidade de aprovação de PLs. É seguida uma abordagem pós-positivista, adotando o paradigma quali-quantitativo como metodologia. Na busca por resultados foram feitos experimentos utilizando os dados disponíveis no Portal da Câmara dos Deputados, seguindo as etapas de revisão bibliográfica, definição de ambiente de experimentação, análise descritiva e predição. Buscou-se ainda realizar uma análise descritiva e prever possíveis resultados no processo de votação de proposições legislativas tendo como foco projetos de lei que tenham sido votados.

Palavras-chave: Processo legislativo, Predição, Aprendizagem de máquina, deep learning, Análise descritiva.

Abstract

This study aims to analyze machine learning algorithms and deep learning for the task of predictability of approval of bills. It follows a post-positivist approach, adopting the quali-quantitative paradigm as a methodology. In the search for results, experiments were carried out using the available data on the Portal of the Chamber of Deputies, following the steps of bibliographic review, definition of experimentation environment, descriptive analysis and prediction. It was also sought to do a descriptive analysis and to predict possible outcomes in the voting process of legislative proposals focusing on bills that have been voted.

Keywords: Legislative process, Prediction, Machine Learning, Deep Learning, Descriptive analysis.

Lista de ilustrações

Figura 1 – Infográfico do Processo Legislativo	15
Figura 2 – Fonte: (OLIVEIRA, 2018)	19
Figura 3 – Proporção entre projetos os projetos apresentados e votados	24
Figura 4 – Temas mais apresentados	25
Figura 5 – Partidos que mais apresentaram projetos	25
Figura 6 – População de deputados(as) por sexo	26
Figura 7 – Proporção dos projetos apresentados por sexo	26

Lista de tabelas

Tabela 1 – Características selecionadas para o dataset	22
Tabela 2 – Configurações de multilayer perceptron	23
Tabela 3 – Configurações de Random Forests	23
Tabela 4 – Configuração de Regressão Logística	23
Tabela 5 – Proporção entre os tipos de propostas analisadas	25
Tabela 6 – Tipos de proponente e projetos apresentados	26
Tabela 7 – Projetos apresentados por faixa etária	27
Tabela 8 – Projetos apresentados por escolaridade do proponente	27
Tabela 9 – Proporção de projetos apresentados por estados	27
Tabela 10 – Proporção de projetos apresentados por proponentes do mesmo partido do presidente da república	28
Tabela 11 – Proporção de projetos apresentados por proponentes do mesmo partido do presidente da república	28
Tabela 12 – Desempenho das configurações dos modelos de MLP	29
Tabela 13 – Desempenho das configurações dos modelos de Florestas Aleatórias	29
Tabela 14 – Resultado da avaliação dos modelos	30
Tabela 15 – Desempenho dos atributos	30

Lista de abreviaturas e siglas

PL	Projeto de Lei
PLO	Projeto de Lei Ordinária
PEC	Projeto de Emenda a Constituição
PLP	Projeto de Lei complementar
LR	Regressão Logística
RF	Florestas Aleatórias
MLP	Multilayer Perceptron

Sumário

	Lista de ilustrações	7
1	INTRODUÇÃO	11
2	FUNDAMENTAÇÃO TEÓRICA	13
2.1	O processo legislativo	13
2.2	Dados abertos do legislativo	14
2.3	Aprendizado de máquina	16
3	TRABALHOS RELACIONADOS	18
4	METODOLOGIA	21
4.1	Posicionamento filosófico	21
4.2	Paradigma de pesquisa	21
4.3	Etapas	21
5	RESULTADOS E DISCUSSÃO	24
5.1	Análise descritiva	24
5.2	Aprendizagem de máquina para detecção de aprovações	29
6	CONCLUSÃO E TRABALHOS FUTUROS	32
6.1	Trabalhos Futuros	32
	REFERÊNCIAS	34

1 INTRODUÇÃO

Análises computacionais vêm sendo realizadas nos mais diversos setores da sociedade, como predição de corrupção(COLLIRI; ZHAO, 2019), identificação de mudanças políticas (CHAN; ZHONG, 2019) e classificação de atividades políticas (PURPURA; HILLARD, 2006), para citar alguns exemplos. Além disso, entender como o poder legislativo funciona é um objeto de estudo de diversas áreas do conhecimento, incluindo a Ciência da Computação.

O aumento da transparência pública e a divulgação de dados abertos governamentais vem se expandindo com o passar dos anos, tendo como destaque a formação da Parceria de Governo Aberto em 2011 do qual o Brasil fez parte deste o início (OPEN...,). Essa mudança alterou a forma de interação entre os governos e a sociedade, levando ao aumento da responsabilização pública, da participação e colaboração cidadã. Nesse contexto, a Web 2.0 se tornou o principal meio de publicação e consumo de dados, possibilitando que iniciativas de publicação de portais de dados abertos fossem lançadas (ZUIDERWIJK; JANSSEN; DAVIS, 2014). Assim, novos benefícios como a prestação de melhores serviços públicos e o aumento da eficiência e eficácia do governo puderam ser alcançados (WAHID, 2012), principalmente devido à possibilidade de a sociedade analisar os dados divulgados, permitindo a geração de valores voltados para transparência e responsabilização(JANSSEN; CHARALABIDIS; ZUIDERWIJK, 2012).

O Processo legislativo além de ser consideravelmente complexo, há um grande volume de projetos propostos. Somente em 2019 foram feitas 5331 propostas na câmara dos deputados(DADOS...,), o que dá uma média de aproximadamente 444 proposições por mês, um número considerável e que torna o acompanhamento pela população em geral algo difícil de ser realizado. Ainda que grande parte destas proposições não sejam aprovadas e que exista um prazo para que entrem em vigor, por vezes ocorre uma urgência em se adequar a uma lei e dependendo do caso pode ocorrer escassez de alguns recursos necessários para conformidade.

Um dos grandes desafios históricos da análise de dados é a análise e entendimento do processo legislativo. A proposição e o voto dos parlamentares nos projetos de lei, e sua posterior aprovação e transformação em lei, é um dos principais pilares da democracia. Porém, o entendimento e eventual predição dos votos nos projetos não é uma tarefa simples, pois o processo legislativo é complexo e influenciado por diversos fatores, que vão de ideológicos e sociais até pressões de grupos de interesse, sindicatos, ONGs e grandes corporações. Apesar de apresentar alguns resultados pro-

missores, esta área de pesquisa ainda está em estágios iniciais, e os resultados são restritos em grande parte à análise do contexto dos Estados Unidos.

Analisar dados referentes ao cenário político do Brasil naturalmente levanta certas indagações, que foram constantes no processo de pesquisa para este trabalho. É possível prever que um projeto vai ser aprovado dado que ele seja votado? Existem padrões identificáveis presentes nesses dados? Fatores demográficos influenciam no processo legislativo? Estes questionamentos serviram não apenas para estimular a busca por respostas, mas também para compreender com mais profundidade este cenário.

Considerando os aspectos acima, este trabalho tem como objetivo realizar experimentos com algoritmos de aprendizagem de máquina e *deep learning* para a tarefa de previsibilidade de aprovação de PLs. Isso se dará através do emprego de diferentes algoritmos baseados em AP e AM que sejam aplicáveis ao tema proposto. Foram definidas métricas para fins de avaliação e abordagens utilizadas em outros países e possíveis aplicações no contexto do processo legislativo brasileiro foram levadas em consideração.

Além do cumprimento destes objetivos, este trabalho através da utilização de ferramentas técnicas e modelos computacionais, busca fornecer uma melhor compreensão do processo legislativo para população ou possíveis interessados, Através da extração de informação contidas na base de dados na plataforma da câmara dos deputados. Com a criação deste precedente, estima-se que mais pesquisas, inclusive aquelas em que o(a) pesquisador(a) por ventura não possua conhecimento computacional, possam vir a utilizar as referências aqui presentes, além de terem este trabalho como inspiração para estudos futuros nesta área.

No presente capítulo encontra-se a contextualização a respeito deste trabalho. O capítulo 2 apresenta a fundamentação teórica contendo os principais conceitos utilizados no decorrer do experimento. O capítulo 3 discute os trabalhos relacionados a este, realizando uma comparação com o que já existe na literatura e o que foi desenvolvido, enquanto que no capítulo 4 é descrito todo o processo metodológico utilizado, desde a concepção até a realização dos experimentos. No capítulo 5 são apresentados os resultados da pesquisa e uma discussão acerca deles e, por fim, no capítulo 6, encontra-se a conclusão, que demonstra as contribuições deste trabalho e possibilidades futuras.

2 FUNDAMENTAÇÃO TEÓRICA

A discussão das proposições legislativas no Congresso Nacional é uma etapa extremamente importante da instituição das políticas públicas e dos programas governamentais, que na maior parte das vezes só podem ser implementados após a aprovação e o início da vigência de lei federal. Essas iniciativas afetam a vida de todos os residentes no país em áreas como educação, saúde, segurança pública, meio ambiente e mercado financeiro(MENEGUIN, 2010).

2.1 O processo legislativo

Um projeto de lei ou proposição é toda matéria sujeita à deliberação da Câmara. Apesar dessa ampla definição, os tipos de proposição considerados principais são:

- Propostas de Emenda à Constituição (PEC) - Proposição legislativa destinada a alterar a Constituição Federal.
- Projetos de Lei Complementar (PLP) - Proposição destinada a elaboração de Lei Complementar.
- Projetos de Lei Ordinária (PL) - Proposição destinada a dispor sobre matéria de competência normativa da União e pertinente às atribuições do Congresso Nacional. Sujeita-se, após aprovado, à sanção ou ao veto presidencial.
- Projetos de Decreto Legislativo (PDC) - Proposição que visa a regular as matérias de competência exclusiva do Poder Legislativo, sem a sanção do Presidente da República.
- Projetos de Resolução (PRC) - Proposição destinada a elaboração de Resolução da Câmara dos Deputados.
- Medidas Provisórias (MPV) - Norma Jurídica de iniciativa exclusiva do Presidente da República, com força de lei ordinária, adotada em caso de urgência e relevância, com produção de efeitos desde sua edição. A conversão em lei depende de apreciação pelo Congresso Nacional. No caso de rejeição ou não apreciação pelo Congresso Nacional no prazo determinado, a medida provisória perde seus efeitos, e as relações jurídicas constituídas na sua vigência serão disciplinadas em até sessenta dias por decreto legislativo ou, na ausência deste, continuarão regidas pela medida provisória.

(ACESSO...,). Este trabalho é focado exclusivamente em **PECs, PLs e PLPs**.

Os projetos podem ser propostos por qualquer deputado ou senador, qualquer comissão, pelo Presidente da República, os Tribunais Superiores e Supremo Federal, o Procurador-geral da República e pela iniciativa popular. Todos os projetos de lei começam a tramitar na Câmara dos Deputados, exceto quando são apresentados por senador ou comissão do Senado, em que começam pelo próprio Senado(ENTENDA...,).

Como podemos observar na Figura1, o processo legislativo brasileiro é consideravelmente complexo, de acordo com as informações contidas no site oficial da Câmara legislativa (ENTENDA...,). O projeto de lei é apresentado pelo proponente e em seguida é realizada a análise de conteúdo por comissões. A maioria dos projetos em tramitação na Câmara só precisa passar pelas comissões, ou seja, têm tramitação conclusiva nas comissões. Se forem aprovados por todas elas, vão direto para o Senado. A partir daqui existem 2 caminhos para o PL seguir:

- Ser dirigido para o plenário caso seja necessário. É realizada então a votação e caso já tenha passado anteriormente pelo senado ele segue para o presidente da república, caso contrário o projeto vai ser votado pelo senado.
- Caso não seja necessária a votação pelo plenário, a PL é recebida no senado, que pode aprovar com mudanças. Neste caso, o projeto volta para a câmara dos deputados. Já no caso de ser aprovada sem quaisquer alterações o PL é recebido pelo presidente da república.

Caso o projeto chegue na presidência da república, ele pode:

- Ser aprovado sem veto, tornando-se lei.
- Ser aprovada com vetos, onde a parte sancionada se torna lei, e os vetos são dirigidos ao congresso.
- Ser vetado integralmente, e então o projeto volta integralmente para o congresso.

2.2 Dados abertos do legislativo

Desde o século XVIII os governos mundiais vêm se preocupando com o 'direito de saber' e a transparência e, iniciando com o governo dos Estados Unidos da América na década de 50, essas preocupações vêm sido formalmente tratadas (IVESTER, 1977) (LITTLE; TOMPKINS, 1974) (PARKS, 1957). Os governos concordam desde

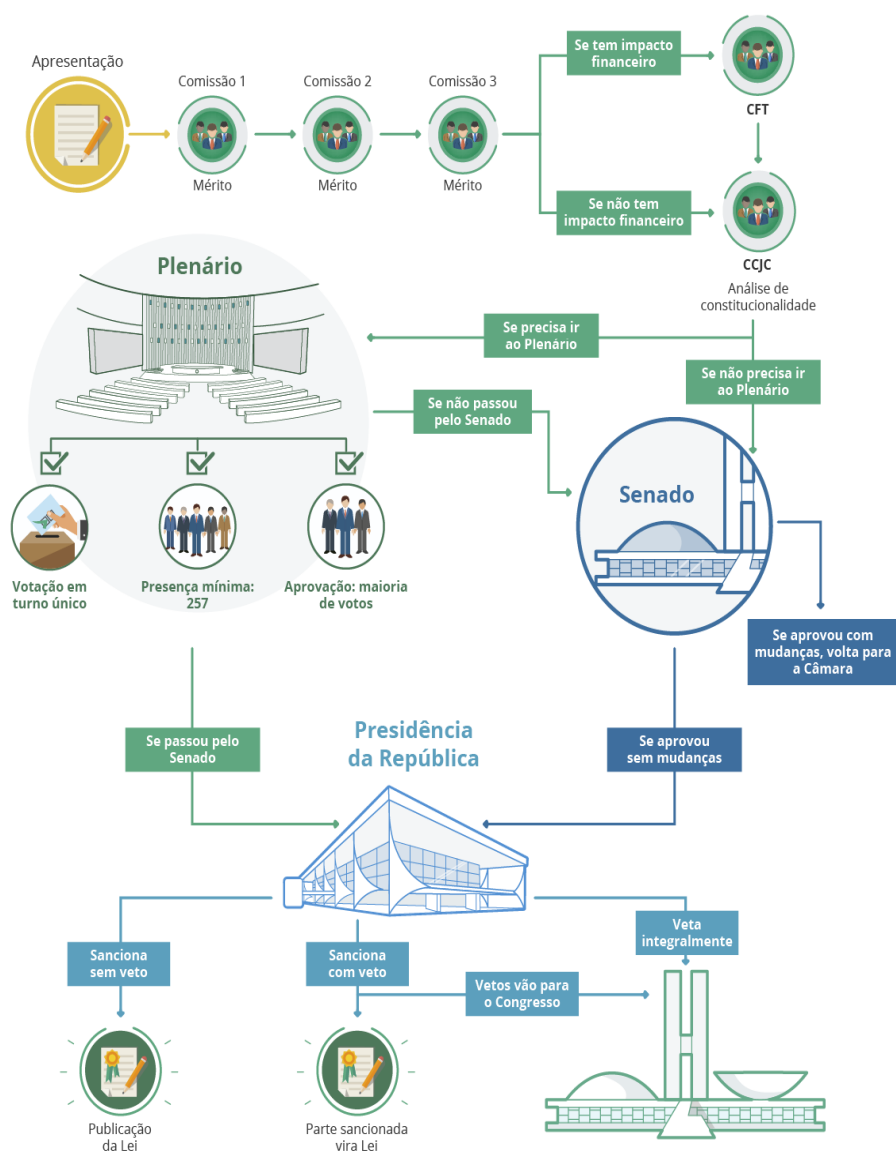


Figura 1 – Infográfico do Processo Legislativo (ENTENDA...,)

aquela época que publicar as informações governamentais tem o potencial de aumentar a responsabilização pública, a participação e a colaboração cidadã, além de proporcionar uma melhoria dos serviços públicos, aumentando assim a eficiência e efetividade do governo, bem como a diminuição da corrupção (ANDERSEN, 2009)(BERTOT; JAEGER; GRIMES, 2010)(FRANCOLI, 2011)(WAHID, 2012)(WONG; WELCH, 2004).

Para este trabalho foi utilizado o Sistema de Dados Abertos da Câmara Legislativa (Disponível em: <https://dadosabertos.camara.leg.br>). Esta plataforma contém diversas informações, como Partidos, Deputados e Proposições, para citar algumas.

O Sistema dispõe de uma série de opções para obtenção dos dados desejados, como XML, JSON, CSV, XLSX e ODS para arquivos, mas além desses há também

disponível uma API Restfull. Esta API permite que sejam realizadas consultas mais específicas, como pedir para listar os autores de um determinado projeto de lei, ou listar os membros de um determinado partido. Ela contém ainda diversos filtros, como por exemplo a listagem de todos os projetos que foram apresentados em determinado ano.

2.3 Aprendizado de máquina

Neste trabalho serão utilizados alguns conceitos de aprendizagem de máquina, setor da Computação que se tornou popular pela sua capacidade de simular a forma do ser humano de aprender. Considerando que nesta pesquisa busca-se prever a aprovação ou não de projetos de lei, utilizar uma ferramenta como esta foi algo essencial para que os resultados fossem obtidos.

”...O Aprendizado de Máquina é uma área de IA cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática. Um sistema de aprendizado é um programa de computador que toma decisões baseado em experiências acumuladas através da solução bem sucedida de problemas anteriores...”(MONARD; BARANAUSKAS, 2003)

Um dos tipos de aprendizagem de máquina aqui utilizados é o **aprendizado supervisionado**, onde é fornecido ao algoritmo de aprendizado, ou indutor, um conjunto de exemplos de treinamento para os quais o rótulo da classe associada é conhecido. Em geral, cada exemplo é descrito por um vetor de valores de características, ou atributos, e o rótulo da classe associada. O objetivo do algoritmo de indução é construir um classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados, ou seja, exemplos que não tenham o rótulo da classe. Para rótulos de classe discretos, esse problema é conhecido como classificação, e para valores contínuos como regressão. É importante não confundir com o **aprendizado não supervisionado**, onde o indutor analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira, formando agrupamentos ou *clusters*(MONARD; BARANAUSKAS, 2003). Neste trabalho não será utilizado o aprendizado não supervisionado.

Outro conceito importante é a **deep learning**, uma subárea de Aprendizagem de Máquina que é uma evolução de redes neurais que tem a capacidade de tratar dados sem pre-processamento em tarefas como reconhecimento visual, reconhecimento de fala e processamento de linguagem natural. Algoritmos de *deep learning* objetivam produzir representações hierárquicas de alto nível dos dados de entrada, por meio

de camadas de processamento sequencial em uma rede neural artificial(BEZERRA, 2016).

Além disso, em estatística, o modelo logístico é um modelo estatístico que modela a probabilidade de um evento de duas alternativas(binário) ocorrer fazendo com que as probabilidades logarítmicas para o evento sejam uma combinação linear de um ou mais variáveis independentes (TOLLES; MEURER, 2016). Uma das principais vantagens da Regressão Logísticas é que ela pode naturalmente fornecer probabilidades e se estender a problemas de classificação multi-classe. (MAALOUF, 2011)

3 Trabalhos relacionados

A utilização de aprendizagem de máquina vem sendo realizada na maioria das áreas de conhecimento para as mais diversas finalidades. Isso reverbera diretamente na produção científica, fato que é demonstrado através do crescimento do número de trabalhos que abordam essa ferramenta.

Alguns trabalhos recentes analisam diversos aspectos desta problemática, como em (COLLIRI; ZHAO, 2019), onde é utilizado Redes Temporais, que é uma rede em que os *links* estão ativos durante uma fração de tempo como uma chamada telefônica ou e-mail (KARSAI; PERRA; VESPIGNANI, 2014), para prever a dinâmica de votação de projetos de lei e prever possíveis casos de corrupção e/ou condenações entre deputados brasileiros através de redes temporais,. Já em (CHAN; ZHONG, 2019) é utilizado o jornal Chinês com publicações entre os anos 1951 e 2019 para prever e identificar mudanças na política da China. (PURPURA; HILLARD, 2006) utiliza classificação automatizada para classificar as atividades legislativas dos EUA. Em (CASAS; DENNY; WILKERSON, 2020) é mostrado que projetos "caroneiros" têm mais facilidade em virar leis como apêndices de outros projetos.

Em relação ao mesmo foco deste trabalho, que é análise e predição, destaca-se a tese (OLIVEIRA, 2018), onde foi criado um artefato capaz de explicar e prever se um Projeto de Lei Ordinária (PLO) será aprovado ou não na Câmara dos Deputados. Para isto o autor coletou, através do portal de Dados Abertos da Câmara dos Deputados (DADOS...,), dados das PLOs aprovadas e arquivadas no intervalo de 2003 à 2016. Os dados então foram limpos e organizados em biênios, já que nesse intervalo de tempo ocorrem eleições. O autor então extraiu inicialmente 27 características pertencentes ao projeto de lei que foram:

Características	Descrição
autores_grupo_mesclado	Se há mais de um autor, se todos os autores pertencerem à oposição/situação ou se possui autores de ambos os grupos
estado	Indica o total de cadeiras ocupadas na Câmara pelos parlamentares do mesmo estado do autor principal do projeto de lei.
governo	Indica se o autor principal do projeto é membro de algum partido da coalizão do governo.
lideranca	Indica se o autor era líder de governo e de partido no momento da apresentação do projeto
liderancas_total	Indica quantas vezes o autor foi líder de partido ou de governo até o momento da apresentação do projeto.
mandatos	Indica o total de mandatos que o autor possuía até o momento da apresentação do projeto.
membro_comissao	Indica se o autor principal do projeto é membro, vice-presidente ou presidente de uma das comissão que analisará o projeto no momento de sua apresentação.
partido_proporcao	Indica a proporção de cadeiras ocupadas pelo partido do autor na Câmara dos Deputados.
presidente_flag	Indica se o autor é o do mesmo partido do presidente da Câmara dos Deputados no momento da apresentação do projeto de lei.
tema_projeto_score	Indica o total de projetos de lei do mesmo tema do que o autor havia apresentado no momento da apresentação do projeto de lei.
tema_relatorias_score	Indica quantas relatorias sobre o mesmo tema da proposição o autor principal havia apresentado no momento da apresentação do projeto de lei
tipo_autor	Indica se o autor principal pertence ao Poder Legislativo ou ao Poder Executivo ou ao Poder Judiciário ou ao Ministério Público
vice_presidente_partido_flag	Indica se o autor é o do mesmo partido do vice-presidente da Câmara dos Deputados no momento da apresentação do projeto de lei
apensado_flag	Indica se o projeto foi apensado à outro durante sua tramitação.
assunto	Indica o tema do projeto de lei.
casa_revisora_flag	Indica se o projeto é proveniente do Senado
emendas	Total de propostas de emendas que o projeto recebeu.
total_pareceres	Indica o total de pareceres que o um determinado projeto de
processo	Indica se o projeto é conclusivo pelas comissões ou se precisará passar pelo plenário.
regime	Indica o regime de tramitação do projeto.
sessao_legislativa	Indica em qual ano da legislatura o projeto foi apresentado
total_apensados	Indica o total de projetos de lei foi apensado à um determi
total_autores	Indica o total de autores do projeto de lei.
total_dias	Total de dias que o projeto ficou em tramitação.
total_orgaos	Total de comissões da Câmara dos Deputados que foram designadas para analisar o projeto.
total_tramitacoes	Total de tramitações que o projeto recebeu.
velocidade	Total de tramitações dividido por total de dias no últimos 6 meses de tramitação do projeto antes dele ser arquivado ou aprovado.

Figura 2 – Fonte: (OLIVEIRA, 2018)

Obtendo como resultado acurácia de 90,1% para processos arquivados e 58% para predições de processos aprovados

No artigo (NAY, 2017) o autor, através de experimentos, buscou encontrar um modelo para buscar prever se um projeto de lei seria aprovado ou não no congresso americano. A base de dados utilizada nos experimentos foram os projetos de leis propostos no intervalo dos anos 1993 até o ano de 2015. Seguidamente, foi realizada a extração de 12 características: região correspondente ao estado que o patrocinador representa, proporção de câmara no partido do proponente, número de mandatos que o proponente atuou no Congresso, tempo médio de permanência do proponente nas comissões em que o projeto de lei está atribuído, posição de liderança do proponente em qualquer comitê, se o proponente é ou não do partido majoritário, número de co-proponentes, sessão do Congresso que corresponde à data do texto completo, se é um projeto de lei da Câmara, mês em que a proposta é apresentada, tema principal da proposição e por último, número de caracteres do texto completo. Com posse dessas características foi realização da análise dos dados. Em seguida foi treinado um modelo utilizando versões do *Wod2Vec* e modelos baseado em árvore além do uso da técnica de *ensemble stacking* obtendo um sucesso de 96% de sucesso em prever se um projeto vai ser reprovado.

Já em (CHENG et al., 2017), também focado no congresso americano, além do texto do projeto também são utilizados dados de perfil dos legisladores. Para realizar a análise do texto dos projetos, o autor utilizou *Bag of Words*, e também a forma como os dados de perfil ideológico de legisladores foram utilizados em um modelo espacial euclidiano chamado *policy location*. A grosso modo, é possível afirmar que neste modelo

os legisladores são representados por sua política ideológica mais favorável em um ponto, chamado de "ponto ideal", e a estimativa é feita medindo a distância entre a política ideológica do projeto e o ponto ideal do legislador, resultando em um paradigma semi-supervisionado.

([KARIMI et al., 2019](#)), também propõe um processo de predição baseado no congresso americano, como os partidos com maioria de participação são apenas 2, e os campos ideológicos desses partidos vêm se distanciando um do outro. O processo foi relativamente semelhante ao produzido por ([CHENG et al., 2017](#)). Este processo foi dividido em 2 etapas: na primeira etapa, 2 conjuntos de dados foram analisados, o primeiro conjunto são os fatores ideológicos, que são bem reconhecidos por desempenhar um papel importante no congresso dos EUA. O segundo conjunto de fatores são fatores sociais e estão relacionados a 1) a filiação partidária dos representantes e 2) como seu registro de votação anterior se entrelaça com outros representantes. E a segunda etapa foi a utilização dos textos dos projetos para aprender diretamente uma incorporação que codifica suas informações ideológicas semânticas.

4 METODOLOGIA

Este capítulo descreverá a metodologia utilizada neste trabalho. Aqui se encontra o posicionamento filosófico utilizado, seguido pelo Paradigma de pesquisa escolhido para a realização do trabalho e, por fim, as fases realizadas da pesquisa.

4.1 Posicionamento filosófico

Este trabalho segue uma abordagem pós-positivista (pós-empiricista) que teve o Sir Karl Popper como um dos fundadores (ROBSON, 2002). A análise se utiliza do reducionismo, quebrando as atividades complexas em partes mais simples e verificando através de experimentos se existe uma relação de causa e efeito no objeto de estudo.

4.2 Paradigma de pesquisa

O paradigma adotado nesta pesquisa foi o quali-quantitativo, sendo este uma fusão dos paradigmas qualitativo e quantitativo. A escolha foi feita tendo em vista a necessidade de analisar as relações entre as variáveis estudadas, com base em dados qualitativos, e também a realização de análises descritivas e representações numéricas dos dados.

4.3 Etapas

Para este trabalho, foram realizadas 4 etapas para a realização do experimento: revisão bibliográfica, definição de ambiente de experimentação, análise descritiva e predição.

Na revisão bibliográfica foi realizado o estudo de livros, artigos, periódicos, documentos oficiais e demais materiais relacionados à mineração de dados, dados abertos governamentais e ao ciclo de vida das propostas de leis, levantando métodos e técnicas para abordagem.

Na fase de ambiente de experimentação foram definidos que, para a coleta dos dados na API do portal da câmara dos deputados, o desenvolvimento da aplicação é composto pela *stack* de desenvolvimento .NET Core da Microsoft, além do uso de banco de dados MySQL. Inicialmente, a aquisição da base de dados foi realizada utilizando a API Restful na plataforma de dados abertos da câmara dos deputados, onde foram obtidas diversas informações como: projetos de leis ordinárias, tramitações dos

projetos de lei, autores dos projetos, temas das proposições, deputados, mesas diretora, partidos, legislaturas, contendo como recorte os projetos dos tipos: PEC, PL e PLP, sendo do intervalo de anos: 2000 até 2020, inseridos em uma base de dados relacionais. Porém a API, até o momento da escrita deste trabalho, não dispõe de filtros e algumas informações que foram necessárias para o experimento, como por exemplo a classificação se um projeto está aprovado ou não. Como a pesquisa é focada em projetos que foram a plenário, ou seja, que foram para votação na câmara dos deputados, foram utilizados os códigos de tramitação das proposições para realizar as classificações. Os códigos em questão são:

- Aprovação em plenário: 244 e 1235
- Reprovação em plenário: 197 e 1236

Com base nos dados analisados, todos os projetos foram classificados entre: **aprovados** - projetos que foram aprovados tendo sido votados em plenário, e **reprovados** - projetos que foram reprovados tendo sido votados em plenário. Nem todas as informações necessárias para o experimento estavam disponíveis até o momento da escrita deste trabalho. Sendo assim, foi necessário realizar um cruzamento de informações para obtenção de dados como faixa etária dos proponentes e se o projeto proposto é do mesmo partido do presidente da república.

Em seguida foi realizada uma análise descritiva nos dados obtidos para extração de informação. O capítulo 5 apresenta todas as informações obtidas através desta análise.

Tabela 1 – Características selecionadas para o dataset

Característica	Descrição
IdAutor	Id do autor do projeto
tipoProponente	Tipo do proponente
nomecivil	Nome civil do proponente
sexo	Sexo do proponente
ufnascimento	Estado natal do proponente
escolaridade	Escolaridade do proponente
siglaPartido	Sigla do partido do proponente
siglaUf	Sigla do estado
idLegislatura	Id da legislatura
ordemassinatura	Ordem de assinatura do proponente
assinaturas	Número de assinaturas do projeto
idProjeto	Id do projeto
numero	Número da proposição
siglatipo	Sigla tipo da proposição
ano	Ano em que proposição foi realizada

Um dataset contendo todos os projetos da base de dados que foram a plenário foi selecionado contendo as características demonstradas na tabela 1.

Por fim, para realização da etapa de aprendizado de máquina e predições, foram experimentados 3 modelos: MLPs, Random Forests e Regressão Logística. Foi realizada a seleção de todos os projetos que foram a plenário e através de *bootstrapping*, amostragem replicável (determinística) e amostragem estratificada dividindo em 2 grupos da e assim utilizados para o treinamento dos modelos de aprendizagem de máquina e para realizar a avaliação dos modelos, foram utilizados critérios mensurar o desempenho como Area Under the "ROC"Curve (AUC), Classification Accuracy (CA), F1-measure, Precisão e Recall. A aplicação Orange, que é um conjunto de ferramentas para visualização de dados, aprendizado de máquina e mineração de dados de código aberto, foi utilizada para realização dessa etapa.

Tabela 2 – Configurações de multilayer perceptron

Model	Neurons	Activation	Solver	Alpha	Maximal Iteration
1	3	Logistic	L-BFGS-B	0,05	100
2	3	ReLu	L-BFGS-B	0,05	100
3	3	Tanh	L-BFGS-B	0,05	100

Na tabela 2 podemos notar as configurações utilizadas em 3 experimentos utilizando o *dataset* nos modelos MLP.

Tabela 3 – Configurações de Random Forests

Model	Nºof Trees	Nºof atributes	Replicable Training	Balance class distribution	Limitdepth of individual trees	Don'tsplit subsets smaller than
1	3	Logistic	L-BFGS-B	0,05	100	5
2	3	ReLu	L-BFGS-B	0,05	100	5
3	3	Tanh	L-BFGS-B	0,05	100	7

Na tabela 3 podemos notar as configurações utilizadas em 3 experimentos utilizando o *dataset* nos modelos de Random Forest.

Tabela 4 – Configuração de Regressão Logística

Regularization Type	Stregth	Balance Class Distribution
LASSO L1	C = 1	No

Para o modelo de regressão logística todos os testes utilizando o *dataset* não demonstraram muita diferença um do outro ao alterar suas configurações, a tabela 3 contem a configuração utilizada na avaliação

5 RESULTADOS E DISCUSSÃO

Neste capítulo estão dispostos os resultados obtidos através do processo de experimentos apresentados anteriormente. Na seção 5.1 é descrita a análise estatística realizada utilizando a base de dados. Na seção seguinte estão contidas informações a respeito dos algoritmos de predição e resultado de suas operações.

5.1 Análise descritiva

Ao todo, o número dos projetos apresentados que foram votados no plenário, representam apenas 1% de todas as proposições apresentadas desde o ano 2000 até o ano de 2020.

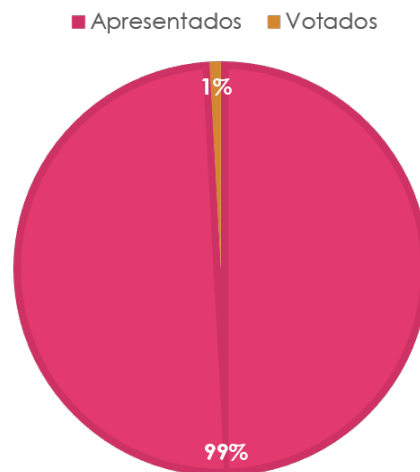


Figura 3 – Proporção entre projetos os projetos apresentados e votados

Como podemos observar na Figura 3, de todos os projetos apresentados até o momento apenas 1% são votados no plenário. Grande parte dessas propostas não é votada, mas sim pensada, caducada ou retirada.

De todos os projetos que foram apresentados e votados, conseguimos constatar na tabela abaixo como é disposta a distribuição dos tipos de PLOs. Como pode-se observar, as PLs compõem a maioria dos projetos apresentados, mais precisamente 90%, e desses, apenas 0,74% foram votados.

Ao todo foram obtidos 32 Temas, como por exemplo Previdência e Assistência Social, Economia, Energia, Recursos Hídricos e Minerais.

A título de demonstração, foi selecionada uma amostragem com as 10 proposições com as temáticas mais apresentadas na Figura 4. Pode-se observar que o Tema

Tabela 5 – Proporção entre os tipos de propostas analisadas

Tipo	Apresentados	Votados
PL	4.7490 (90%)	388 (0,74%)
PLP	2.541 (5%)	34 (0,06%)
PEC	2.424 (5%)	29 (0,06%)

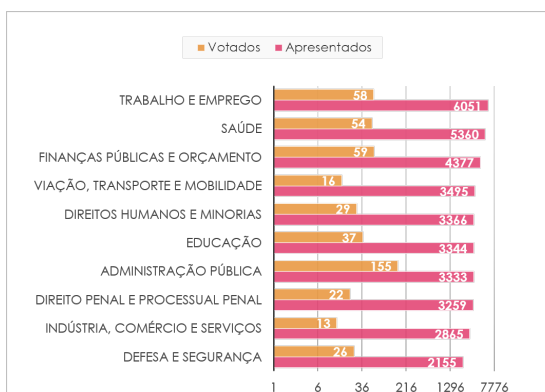


Figura 4 – Temas mais apresentados

mais apresentado foi o de Trabalho e Emprego, representando 10,93% de todos os projetos apresentados, seguido por Saúde e Finanças Públicas e Orçamento. Os projetos com tema Trabalho e Emprego representaram apenas 9,39% de todos os projetos votados, enquanto que o tema com mais votação foi Administração Pública, com 25,08%

Ao todo, foram obtidos 34 partidos na API, mas assim como os temas, a título de demonstração foi selecionada uma amostra com os 10 partidos que mais apresentaram projetos, como demonstrado na Figura 5. Nota-se que o Partido dos Trabalhadores - PT, é o partido que mais apresentou projetos, seguido do Partido da Social Democracia Brasileira - PSDB, e os demais partidos obtiveram porcentagens bem próximas umas das outras.

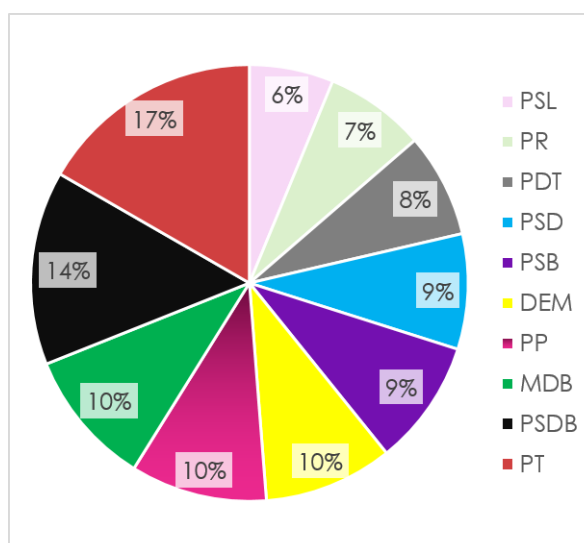


Figura 5 – Partidos que mais apresentaram projetos

Tabela 6 – Tipos de proponente e projetos apresentados

Tipo	Apresentados
Deputado	51348 (97,89%)
Órgão do Poder Executivo	862 (1,64%)
Órgão COMISSÃO PERMANENTE	46 (0,09%)
COMISSÃO PARLAMENTAR DE INQUÉRITO	16 (0,03%)
COMISSÃO MISTA PERMANENTE	2 (0,0%)

Quanto ao tipo proponente, são 6 ao total: Órgão do Poder Judiciário, Órgão do Poder Executivo, Comissão Permanente, Deputado, Comissão Parlamentar de Inquérito, Comissão Mista Permanente, ao olhar a Tabela 6, nota-se que quase 98% das propostas advém de deputados.

Já em relação ao sexo dos(as) parlamentares, vale ressaltar que todos os dados aqui mostrados são o histórico de toda população de deputados que possuem cadastro na plataforma de dados abertos da câmara. Ou seja, a soma de todos(as) os(as) parlamentares que já passaram pela câmara.

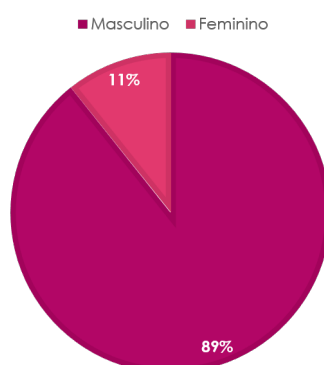


Figura 6 – População de deputados(as) por sexo

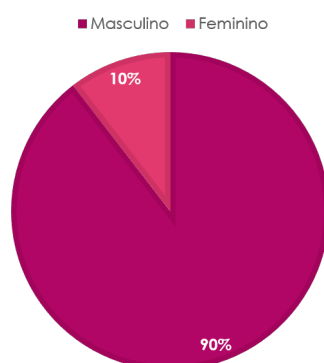


Figura 7 – Proporção dos projetos apresentados por sexo

É notado que na Figura 6, 89% de todos o corpo legislativo é composto por homens e que a parcela de projetos de leis segue essa proporção como demonstrado na Figura 7

Além disso, foram consideradas as faixas etárias dos proponentes, a obtenção desses dados foi feita através de cruzamento de informações e consultas de banco de dados, já que esta informação não está disponível na API. Quanto a disposição da faixa etária do corpo legislativo, esta ficou dividido em 5 grupos.

Tabela 7 – Projetos apresentados por faixa etária

Faixa Etária	Apresentados	Votados
21 – 33	5255	23
34 – 45	20771	67
46 – 57	19872	81
58 – 69	4626	49
70 – 86	484	7

Podemos observar na Tabela 7 que grande parte dos projetos apresentados e votados estão concentrados em 2 grupos etários: **34 - 45** e **46 - 57** anos.

Tabela 8 – Projetos apresentados por escolaridade do proponente

Escolaridade	Apresentados	Votados
Superior	33881	137
Superior Incompleto	4205	15
Ensino Médio	1762	2
PósGraduação	5616	38
Secundário	990	2
	⋮	

Ao todo 16 tipos de escolaridade estão registradas e também foram analisadas, sendo o ensino superior a modalidade com maior número de projetos apresentados e votados conforme na Tabela 8.

Tabela 9 – Proporção de projetos apresentados por estados

Estado	Apresentados	Votados
SP	8628	40
RJ	6432	26
MG	5847	29
RS	4005	19
PR	2629	11
	⋮	

Já em relação aos estados de nascimento dos proponentes, São Paulo é o estado natal da maioria dos proponentes com 16,45% de todos os projetos propostos, como mostrado na Tabela 9.

Tabela 10 – Proporção de projetos apresentados por proponentes do mesmo partido do presidente da república

Presidente	Apresentados	Votados
Luiz Inácio Lula da Silva	1744	7
Dilma Rousseff	1205	5
Jair Bolsonaro	736	4
Fernando Henrique Cardoso	346	3

Assim como as faixas etárias, até o momento da escrita deste trabalho, não existe a informação de que os projetos apresentados são do mesmo partido que o atual presidente no momento de sua apresentação, esta informação foi obtida através de cruzamento de informações. Na Tabela 10, vemos que os presidentes Luiz Inácio Lula da Silva foi o que mais teve correligionários apresentando projetos no seu mandato, bem como teve mais proposições votadas.

Tabela 11 – Proporção de projetos apresentados por proponentes do mesmo partido do presidente da república

Vice-Presidente	Apresentados	Votados
Michel Temer	1172	7
José Alencar	189	0
Marco Maciel	139	0

Quanto aos vice-presidentes, semelhantemente as informações de presidente, estes dados foram obtidos através de cruzamento de informações. É notado na Tabela 11 que o Michel Temer é o vice-presidente com mais correligionários apresentando projetos durante seu mandato, e também o único que teve proposições aprovadas durante seu período como vice-presidente.

Como principais resultados encontrados, foi ressaltado os resultados de duas análises distintas: uma sobre os projetos apresentados, e outra sobre os projetos aprovados.

Descobriu-se que os projetos mais apresentados tinham as seguintes características:

- **trabalho e emprego** como tema;
- **partido** que mais apresentou projetos foi o **PT**;
- o tipo de proponente foi **deputados**;
- do **sexo masculino**;
- do **estado de São Paulo**
- com **ensino superior completo**

Por outro lado, como dito anteriormente, o número de projetos que vão para votação no plenário corresponde a apenas 1% dentro do período avaliado. Dentre esses, destaca-se como características dos projetos com maior percentual de aprovação, os projetos com:

- **temática de direito e justiça**
- apresentados pelo **partido DEM** ou **apresentados pelo poder judiciário**
- com **mestrado completo**.

5.2 Aprendizagem de máquina para detecção de aprovações

Tendo todas essas informações em mãos, deu-se início à etapa de predição. Para isto, foi feito um recorte contendo todas as informações referentes aos projetos que foram a plenário, contendo todos os atributos dos projetos, autores do projeto, dados do autor, partidos, mesas, etc., separados em 2 grupos: O primeiro grupo contendo 2/3 do recorte inicial e o segundo grupo contendo o 1/3 restante. Após essa segmentação, foram treinados 3 tipos de modelos de AM: Regressão Logística (LR), Florestas Aleatórias (RF) e Rede Neural Artificial do tipo *multilayer Perceptron* (MLP). Várias configurações foram avaliadas para cada um dos 3 modelos.

Tabela 12 – Desempenho das configurações dos modelos de MLP

Model	AUC	ACC	F-1	Precision	Recall
1	28,20%	96,30%	94,50%	92,70%	96,30%
2	50,00%	3,70%	0,30%	0,10%	3,70%
3	50,00%	96,30%	94,50%	92,70%	96,30%

Tabela 13 – Desempenho das configurações dos modelos de Florestas Aleatórias

Model	AUC	ACC	F-1	Precision	Recall
1	73,80%	96,70%	95,30%	96,80%	96,70%
2	84,40%	97,40%	96,70%	97,50%	97,40%
3	86,30%	95,90%	95,20%	94,90%	95,90%

Os modelos foram avaliadas pelas métricas Area under the ROC curve (AUC), Precisão de Classificação, F-1 Score, Precisão e Recall

A AUC é uma medida de desempenho de classificação bidimensional do desempenho de classificação. Pode ser entendido como um gráfico da probabilidade de classificar os exemplos positivos em relação à taxa de classificar incorretamente os exemplos negativos verdadeiros. Nesse sentido, pode-se interpretar curva como uma comparação do desempenho do classificador em toda a gama de distribuições

de classe e custos de erro (RAKOTOMAMONJY, 2004). De acordo com (LÜDEMANN et al., 2006), resultados que variam entre 0.9-1 são considerados excelentes, entre 0.8-0.9 são bons resultados, valores razoáveis ficam entre 0.7-0.8, já os resultados que atingem o intervalo de 0.6-0.7 são considerados como resultados pobres, e por fim abaixo de 0.6 é um modelo de desempenho de falha.

A precisão de classificação ou Classification Accuracy, é uma métrica que resume o desempenho de um modelo de classificação como o número de predições corretas dividido pelo total de números de predições, um modelo que é bom tem tipicamente acima de 90% de acurácia (BRANCO; TORGO; RIBEIRO, 2016).

Precisão, dentre todas as classificações de classe Positivo que o modelo fez, quantas estão corretas, (BEGER, 2016).

Recall, taxa que representa os verdadeiros positivos de um modelo, é estimado dividindo os classificados corretamente pelo total de positivos (OLSON; DELEN, 2008).

F-1 Score, é uma métrica que consiste na combinação de precisão e recall, ou seja, uma média harmônica entre precisão e recall.

Nas tabelas 4, 3 e 2 são apresentadas respectivamente as configurações que obtiveram as melhores desempenhos.

Tabela 14 – Resultado da avaliação dos modelos

Model	AUC	CA	F-1	Precision	Recall
Random Forest	84,40%	97,40%	96,70%	97,50%	97,40%
Neural Net	50,00%	96,30%	94,50%	92,70%	96,30%
Logistic Regression	72,50%	96,30%	95,10%	94,90%	96,30%

As melhores combinações de configurações com os melhores resultados são demonstradas na tabela 14. Como pode-se observar, o modelo que obteve o melhor desempenho foi o Random Forest, com 5 árvores, com treino replicável, sem distribuição de balanceamento de classe, sem limite de profundidade individual de árvores, e sem dividir subgrupos menores que 5. É notado que a AUC (area under the ROC curve) nas configurações de MLPs não ultrapassam 50%, o resultado não é esperado e estudos adicionais são necessários para identificar o porquê deste resultado

Tabela 15 – Desempenho dos atributos

Feature	Ganho de informação
Estado natal	0.174
Tema	0.167
Partido	0.145
Legislatura	0.036
Numero de assinaturas	0.034

Após a predição foi realizada uma nova análise, dessa vez utilizando o modelo RF, que foi o que obteve melhor performance. Descobriu-se que, como demonstrado na tabela 15, as 5 características com maior ganho de informação foram: estado natal, tema, partido, id de legislatura e o número de assinaturas. Destaca-se como características dos projetos com maior percentual de aprovação, os projetos com:

- **estado natal São Paulo;**
- **temática administração pública;**
- **legislatura de número 2;**
- **número de assinaturas é 1**

Os trabalhos (NAY, 2017), (CHENG et al., 2017) e (KARIMI et al., 2019) que têm o mesmo foco, apresentam bons resultados, porém estes se restringem exclusivamente ao cenário do americano. Já em (OLIVEIRA, 2018) o cenário é mais semelhante a este trabalho, porém com um recorte do intervalo de 2003 até 2016 um pouco menor, em comparação ao aqui demonstrado foi durante o período do ano 2000 ao ano 2020. Além disto, as características de maior impacto reveladas no texto foram o total de tramitações e órgãos analisados, porém para um projeto ser votado/aprovado ele tem que passar por todos os estados de tramitação e ser aprovado pelos órgãos de análise de projetos. Outro ponto é que o autor não leva em consideração características relacionadas aos proponentes, como estado natal por exemplo, onde neste trabalho foi a característica com maior ganho de informação.

6 CONCLUSÃO E TRABALHOS FUTUROS

Com base nos resultados obtidos através de todo o processo desde a pesquisa bibliográfica até a fase de predição, o entendimento do processo legislativo e a eventual predição do resultado da votação dos projetos não é uma tarefa simples, pois o processo legislativo é complexo e influenciado por diversos fatores como ideológicos, sociais, e inclusive pressões de grupos de interesse, sindicatos, ONGs, empresários e grandes corporações e que demanda conhecimento computacional.

Foi desenvolvido um processo para a realização do experimento onde foram coletados os dados relativos a projetos, tramitações, deputados dentre outras informações referentes ao intervalo entre os anos 2000 e 2020. Em seguida foram realizadas análises descritivas para extração de informações, predições utilizando algoritmos de aprendizagem de máquina, e a avaliação de desempenho desses modelos utilizando métricas de performance. Descobriu-se pelos resultados que fatores demográficos como estado de nascimento do proponente podem influenciar na aprovação.

Posteriormente foi realizada a descoberta das principais características utilizando análise descritiva e através de experimentação com modelos de aprendizagem de máquina, relacionadas aos projetos de lei propostos, à sua votação no plenário da Câmara dos Deputados, e à sua aprovação. Dadas as condições específicas neste documento, e dentro do cenário proposto, é possível prever quando um projeto, dado que ele está sendo votado, se ele vai ser aprovado ou não, utilizando modelos com precisão superior a 92%.

Por fim, este trabalho se propõe a ser uma ferramenta para ter um melhor entendimento do processo legislativo, contribuindo para área de pesquisa análise e predição em área pouco explorada pela computação. Além de ser um complemento as necessidades dos trabalhos que foram expostas no capítulo anterior, expandindo o tema de tendo contribuído gerando resultados que podem ser utilizados em ciências sociais para análises do processo legislativo. Outro contribuição é a possibilidade de avaliar o desempenho dos projetos de lei.

6.1 Trabalhos Futuros

Dadas as restrições de tempo e de escopo de um Trabalho de Conclusão de Curso, é sabido que há pontos de melhoria, podendo citar por exemplo:

Realização de teste estatístico, é preciso também testar a normalidade da distribuição de erros e caso necessário, implementar algum tipo de transformação.

Estudos adicionais são necessários, tendo em vista os resultados referentes a AUC (area under the ROC curve) das redes neurais multilayer perceptron não ultrapassarem 50%, que é um comportamento não esperado, se fazendo necessário uma investigação deste comportamento.

Utilização da ementa e do texto da proposição Uma hipótese levantada inicialmente era a de que talvez a ementa detalhada do projeto de lei possa influenciar de alguma forma na votação,

Avaliar situação proporcional dos partidos e frentes parlamentares, sendo um fator que dispõe de grande impacto nas votações, tendo em vista que quanto maior é o grupo da base aliada do proponente, maior será o número de votos para a proposta.

Análise de sentimento da população, tendo em vista a natureza de alguns temas e propostas, e o envolvimento de redes sociais como plataforma de comunicação entre proponentes e a população, é possível que exista um impacto da opinião expressadas em redes sociais pelos cidadãos na votação.

Referências

- ACESSO À INFORMAÇÃO. CAMARA DOS DEPUTADOS. <https://www2.camara.leg.br/transparencia/aceso-a-informacao/copy_of_perguntas-frequentes/processo-legislativo>. Acessado em: 09-07-2021. Citado na página 14.
- ANDERSEN, T. B. E-government as an anti-corruption strategy. *Information Economics and Policy*, Elsevier, v. 21, n. 3, p. 201–210, 2009. Citado na página 15.
- BEGER, A. Precision-recall curves. 2016. Citado na página 30.
- BERTOT, J. C.; JAEGER, P. T.; GRIMES, J. M. Crowd-sourcing transparency: Icts, social media, and government transparency initiatives. In: *Proceedings of the 11th Annual International Digital Government Research Conference on Public Administration Online: Challenges and Opportunities*. [S.l.: s.n.], 2010. p. 51–58. Citado na página 15.
- BEZERRA, E. Introdução à aprendizagem profunda. *Artigo–31º Simpósio Brasileiro de Banco de Dados–SBB2016–Salvador*, 2016. Citado na página 17.
- BRANCO, P.; TORGO, L.; RIBEIRO, R. P. A survey of predictive modeling on imbalanced domains. *ACM Computing Surveys (CSUR)*, ACM New York, NY, USA, v. 49, n. 2, p. 1–50, 2016. Citado na página 30.
- CASAS, A.; DENNY, M. J.; WILKERSON, J. More effective than we thought: Accounting for legislative hitchhikers reveals a more inclusive and productive lawmaking process. *American Journal of Political Science*, Wiley Online Library, v. 64, n. 1, p. 5–18, 2020. Citado na página 18.
- CHAN, J. T.; ZHONG, W. Reading china: Predicting policy change with machine learning. AEI Economics Working Paper Series, 2019. Citado 2 vezes nas páginas 11 e 18.
- CHENG, Y. et al. Legislative prediction with dual uncertainty minimization from heterogeneous information. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, Wiley Online Library, v. 10, n. 2, p. 107–120, 2017. Citado 3 vezes nas páginas 19, 20 e 31.
- COLLIRI, T.; ZHAO, L. Analyzing the bills-voting dynamics and predicting corruption-convictions among brazilian congressmen through temporal networks. *Scientific Reports*, Nature Publishing Group, v. 9, n. 1, p. 1–11, 2019. Citado 2 vezes nas páginas 11 e 18.
- DADOS Abertos. CAMARA DOS DEPUTADOS. <<https://dadosabertos.camara.leg.br/swagger/api.html>>. Acessado em: 15-07-2021. Citado 2 vezes nas páginas 11 e 18.
- ENTENDA O PROCESSO LEGISLATIVO. CAMARA DOS DEPUTADOS. <<https://www.camara.leg.br/entenda-o-processo-legislativo/>>. Acessado em: 09-07-2021. Citado 2 vezes nas páginas 14 e 15.

- FRANCOLI, M. What makes governments' open? ejournal of edemocracy & open government. 2011. Citado na página 15.
- IVESTER, D. M. The constitutional right to know. *Hastings Const. LQ*, HeinOnline, v. 4, p. 109, 1977. Citado na página 14.
- JANSSEN, M.; CHARALABIDIS, Y.; ZUIDERWIJK, A. Benefits, adoption barriers and myths of open data and open government. *Information systems management*, Taylor & Francis, v. 29, n. 4, p. 258–268, 2012. Citado na página 11.
- KARIMI, H. et al. Multi-factor congressional vote prediction. In: *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. [S.l.: s.n.], 2019. p. 266–273. Citado 2 vezes nas páginas 20 e 31.
- KARSAI, M.; PERRA, N.; VESPIGNANI, A. Time varying networks and the weakness of strong ties. *Scientific reports*, Nature Publishing Group, v. 4, n. 1, p. 1–7, 2014. Citado na página 18.
- LITTLE, J. W.; TOMPKINS, T. Open government laws: An insider's view. *NCL rev.*, HeinOnline, v. 53, p. 451, 1974. Citado na página 14.
- LÜDEMANN, L. et al. Glioma assessment using quantitative blood volume maps generated by t1-weighted dynamic contrast-enhanced magnetic resonance imaging: a receiver operating characteristic study. *Acta Radiologica*, Taylor & Francis, v. 47, n. 3, p. 303–310, 2006. Citado na página 30.
- MAALOUF, M. Logistic regression in data analysis: an overview. *International Journal of Data Analysis Techniques and Strategies*, Inderscience Publishers Ltd, v. 3, n. 3, p. 281–299, 2011. Citado na página 17.
- MENEGUIN, F. B. Avaliação de impacto legislativo no brasil. 2010. Citado na página 13.
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, Manole Ltda, v. 1, n. 1, p. 32, 2003. Citado na página 16.
- NAY, J. J. Predicting and understanding law-making with word vectors and an ensemble model. *PloS one*, Public Library of Science San Francisco, CA USA, v. 12, n. 5, p. e0176999, 2017. Citado 2 vezes nas páginas 19 e 31.
- OLIVEIRA, D. A. *Compreendendo e Prevendo o Processo Legislativo na Câmara dos Deputados do Brasil*. Dissertação (Mestrado), 2018. Citado 4 vezes nas páginas 7, 18, 19 e 31.
- OLSON, D. L.; DELEN, D. *Advanced data mining techniques*. [S.l.]: Springer Science & Business Media, 2008. Citado na página 30.
- OPEN Government Partnership, "Open Government Partnership," Official Website,. <www.opengovpartnership.org>. Acessado em: 15-05-2022. Citado na página 11.
- PARKS, W. Open government principle: Applying the right to know under the constitution. *Geo. Wash. L. Rev.*, HeinOnline, v. 26, p. 1, 1957. Citado na página 14.

PURPURA, S.; HILLARD, D. Automated classification of congressional legislation. In: *Proceedings of the 2006 international conference on Digital government research*. [S.l.: s.n.], 2006. p. 219–225. Citado 2 vezes nas páginas 11 e 18.

RAKOTOMAMONJY, A. Optimizing area under roc curve with svms. In: *ROCAI*. [S.l.: s.n.], 2004. p. 71–80. Citado na página 30.

ROBSON, C. *Real world research: A resource for social scientists and practitioner-researchers*. [S.l.]: Wiley-Blackwell, 2002. Citado na página 21.

TOLLES, J.; MEURER, W. J. Logistic regression: relating patient characteristics to outcomes. *Jama*, American Medical Association, v. 316, n. 5, p. 533–534, 2016. Citado na página 17.

WAHID, F. The current state of research on e-government in developing countries: A literature review. In: SPRINGER. *International Conference on Electronic Government*. [S.l.], 2012. p. 1–12. Citado 2 vezes nas páginas 11 e 15.

WONG, W.; WELCH, E. Does e-government promote accountability? a comparative analysis of website openness and government accountability. *Governance*, Wiley Online Library, v. 17, n. 2, p. 275–297, 2004. Citado na página 15.

ZUIDERWIJK, A.; JANSSEN, M.; DAVIS, C. Innovation with open data: Essential elements of open data ecosystems. *Information polity*, IOS Press, v. 19, n. 1-2, p. 17–33, 2014. Citado na página 11.