

USO DE *MACHINE LEARNING* PARA A ANÁLISE DE PROJETOS LEGISLATIVOS DE DESENVOLVIMENTO REGIONAL: O CASO DA ZONA FRANCA DE MANAUS

Use of machine learning for the analysis of regional development legislative projects: the case of the Manaus Free Zone

DOI: [10.48075/igepec.v26i2.28611](https://doi.org/10.48075/igepec.v26i2.28611)

Waldecy Rodrigues
Victor Cannavale
Daniela Mascarenhas de Queiroz Trevisan
David Prata

USO DE *MACHINE LEARNING* PARA A ANÁLISE DE PROJETOS LEGISLATIVOS DE DESENVOLVIMENTO REGIONAL: O CASO DA ZONA FRANCA DE MANAUS

Use of machine learning for the analysis of regional development legislative projects: the case of the Manaus Free Zone

Waldecy Rodrigues
Victor Cannavale
Daniela Mascarenhas de Queiroz Trevisan
David Prata

Resumo: Este artigo tem como objetivo principal analisar o risco de projetos de lei para o modelo Zona Franca de Manaus (ZFM) serem transformados em norma jurídica, com o uso de *machine learning*. Trata-se de uma pesquisa com análise estruturada sobre os dados abertos da Câmara dos Deputados Federais, no período de 2011-2018, com aplicação de modelo de regressão logística. Quanto à análise dos projetos com temas pertinentes ao desenvolvimento regional da ZFM, ficou demonstrado que, apesar de serem temas importantes, as atividades do polo e, na maioria, estarem na agenda do executivo para o ano, eles não recebem a atenção de grupos e atores que podem representar influência para a sua efetivação até o momento. A única exceção é o projeto de isenção de Imposto de Produtos Industrializados para bicicletas comuns e elétricas que pelo modelo apresentou uma maior chance estatística de ser aprovado, o que pode levar uma grande perda de competitividade da Zona Franca de Manaus.

Palavras-chave: Amazônia. Zona Franca de Manaus. Processo Legislativo. Regressão Logística.

Abstract: *This paper has as main objective to analyze the risk of bills for the Free Trade Zone of Manaus model (FTZM) being transformed into a legal norm, with the use of machine learning. This is a research with exploratory and structured analysis on open data from the Chamber of Deputies, in the period 2011-2018, with the application of a logistic regression model. As for the analysis of projects with themes relevant to regional development from FTZM, it was shown that, despite being important themes, the activities of the pole and, for the most part, being on the executive's agenda for the year, they do not receive the attention of groups and actors who may represent an influence for its effectiveness so far. The only exception is the Tax on Industrialized Products for common and electric bicycles, which, according to the model, had a highest statistical chance of being approved, which could lead to a great loss of competitiveness in the Manaus Free Trade Zone.*

Keywords: Amazon. Manaus Free Trade Zone. Legislative Process. Logistic Regression.

Resumen: *Este artículo tiene como principal objetivo analizar el riesgo de que las facturas del modelo de la Zona Franca de Manaus (ZFM) se transforme en una norma legal, con el uso de machine learning. Se trata de una investigación con análisis estructurado sobre datos abiertos de la Cámara de Diputados de la Federación, en el período 2011-2018, con la aplicación de un modelo de regresión logística. En cuanto al análisis de proyectos con temas relevantes para el desarrollo regional de la ZFM, se demostró que, apesar de ser temas importantes, las actividades del polo y, en su mayor parte, estar en la agenda del ejecutivo para el año, no reciben la atención de grupos y actores que pueden representar influencia para su realización hasta el momento. La única excepción es el proyecto de exención del Imposto de Productos Industrializados para bicicletas comunes y eléctricas, que según el modelo presenta una mayor probabilidad estadística de ser aprobado, lo que podría conllevar una gran pérdida de competitividad en la Zona Franca de Manaus.*

Palabras clave: Amazonia. Zona Franca de Manaus. Proceso Legislativo. Regresión Logística.

INTRODUÇÃO

O modelo Zona Franca de Manaus (ZFM) foi criado no Brasil em 1967 e tem a abrangência em toda a Amazônia Ocidental - composta pelos Estados do Acre, Amazonas, Rondônia e Roraima, além da Área de Livre Comércio de Macapá/Santana, no Amapá. A ZFM tem como principal objetivo o desenvolvimento da região Norte e, somente em 2015, movimentou R\$25 bilhões de incentivos fiscais trazendo impactos efetivos para a região na área socioeconômica e ambiental (HOLLAND et al., 2019).

Apesar de prorrogado, em 2014, por mais 50 anos, vêm tendo de lidar com oportunidades de atuação esquecidas nas políticas de governo. Dessa forma, novos caminhos dentro das estruturas administrativas precisam ser desenvolvidos, capazes de oferecer maior previsibilidade sobre ações de acompanhamento legislativo.

Este modelo alicerçado em vantagens tributárias, tem alta relevância para a região, gerando ações de desenvolvimento, oportunidades de negócios, atraindo investimentos, realizando integração econômica com parcerias entre governos e instituições de ensino, pesquisa e desenvolvimento, além de investimentos em projetos de estímulo ao empreendedorismo (SUFRAMA, 2019). Contudo, parte dessas ações tem sido mitigada por crises financeiras e fiscais, dinamizadas por diferentes atores, na chamada guerra fiscal.

Devido ao aumento dos gastos públicos, da crise da dívida externa e dos efeitos da nova constituição que passa a impactar os benefícios previdenciários e as transferências constitucionais da União, nos anos 80, o estado brasileiro enfrentou um período de crise em suas finanças públicas (GIAMBIAGI, 1997). Esse movimento teve reflexo no comportamento dos estados que, com maior autonomia dada pela constituição, passaram a utilizar a renúncia fiscal para atrair investimentos e promover o seu próprio desenvolvimento regional.

No contexto da relação horizontal, a Zona Franca de Manaus tem sofrido questionamentos na esfera administrativa, judicial e principalmente legislativa sobre sua política de incentivos fiscais e de seu modelo de desenvolvimento regional. Diante destes desafios econômicos e políticos a Superintendência da Zona Franca de Manaus - SUFRAMA tem encontrado na sua base política de apoio, a forma para reagir a estas novas proposições. Este processo de decisão política da Zona Franca de Manaus e intervenção institucional exigem esforço societário expresso nas políticas públicas das jurisdições subnacionais, nacionais e supranacionais e na relação com seus atores (MENDONÇA, 2013).

São cada vez mais recorrentes as propostas legislativas que isentam ou reduzem os impostos federais no modelo de Zona Franca. Desta forma, para a redução de impactos e abertura de diálogos na Câmara de Deputados é preciso conhecer as condicionantes políticas dos projetos de lei propostos.

A partir da dinâmica complexa destas relações, o objetivo deste trabalho é estruturar um modelo de predição baseado em *machine learning* para verificar a chance estatística da transformação de um projeto de lei em norma jurídica de impacto para a Zona Franca de Manaus. Uma contribuição relevante deste artigo do ponto de vista metodológico é que suas predições no modelo logístico, ocorre mediante o uso de técnicas de aprendizado de máquina, com a biblioteca scikit-learn. O treinamento foi realizado sobre a amostra de treino e a predição na amostra de teste, para ter simultaneamente melhores níveis de acurácia e predição. Isto trás resultados muito superiores aos modelos estáticos principalmente na prevenção de preditores falsos positivos ou falsos negativos.

2 – MATERIAIS E MÉTODOS

Para a análise foram utilizados os dados abertos da Câmara dos Deputados¹. Os dados do órgão são divididos por ano e por assuntos e foram coletados considerando cinco estruturas para concatenar as informações e criar as variáveis independentes em estudo, essas estruturas são: *Proposições*; *Proposições de Temas*; *Orientação de Votação*; *Órgãos Deputados*; e *Proposições Autores*.

Segundo Meneguín et al. (2017), as condicionantes podem ser determinadas pelo *tempo*; pelo conjunto de *discussões políticas*; pelo conjunto de *atores* capazes de influenciar o processo de inclusão de determinados temas na agenda; pelo de *regime de tramitação*; e pelas próprias *votações*. Para o autor, dentre estas condicionantes, o *tempo médio de tramitação* das proposições tem forte influência sobre a transformação em norma jurídica. Esta média é influenciada por uma quantidade pequena de projetos cuja tramitação recomeça em momento político oportuno.

Momento este definido por uma *agenda*, ou seja, um conjunto de discussões políticas, entendidas como questões legítimas e que chamam a atenção do sistema político (CAPELLA, 2018) e que tem um papel desafiador na definição de temas que deverão emergir no cenário político. Segundo Santos e Borges (2018), a agenda ocorre por um processo de definição de alternativas, onde se reduz o foco das atividades decisórias a determinados temas.

Alguns atores envolvidos são capazes de influenciar o processo de inclusão de determinados temas (*agenda holders*). Segundo Silva e Araújo (2010), a forma como estes atores se movimenta e exerce pode influenciar no potencial de transformação jurídica de projetos de lei. Quanto à condicionante *regime de tramitação*, afirma Santos (2014), que a sua urgência implica, a impossibilidade de parlamentares e grupos de interesse interferirem no processo decisório. Ou seja, o processo mais rápido de tramitação impede os possíveis debates sobre as matérias legislativas.

Por fim, outro fator condicionante, a *votação* nas comissões tem papel fundamental para aprovação de matérias, permitem ao legislativo, ação independente em relação ao Executivo. Segundo Junqueira (2016), ao subdividir os parlamentares, possibilita-se que, antes minorias em plenário, possam se transformar em maiorias no seu interior. Pereira e Mueller (2000), ainda ressaltam, que os líderes partidários dentro destas comissões temáticas têm capacidade de influenciar a sua composição, como também é comum, afirma Nicolau (1999), os líderes orientarem suas bancadas como votar.

Os dados coletados foram divididos por ano e, para esta pesquisa, foram utilizados os registros que vão de 2011 a 2018, período correspondente a dois mandatos legislativos. As amostras referentes a estes anos têm no total 219.647 registros e, para o trabalho, utilizou-se uma amostra com 19.652 registros, referentes a projetos de leis em tramitação no período citado. Este recorte apresenta 19.106 projetos de lei que não foram transformados em norma jurídica e 546 que passaram pela transformação. Esta amostra foi balanceada, conforme seção 3.2.

O trabalho foi dividido nas etapas de manipulação de dados e análise. Para as duas etapas foi utilizada a linguagem de programação *Python*.

¹<https://dadosabertos.camara.leg.br/swagger/api.html>

2.1 – PROCESSAMENTO DOS DADOS

Após a coleta dos dados, foram definidas cinco condicionantes, utilizando-se de uma ou mais tabelas contidas nos dados abertos da Câmara dos Deputados. *Regime de tramitação*; *Tempo de tramitação*; *Convergência de Agendas* (convergência de agenda entre o executivo e legislativo); *Orientação de Votação* pelos Partidos; e a *Agenda Holder* (capacidade de agenda parlamentar pelo deputado).

Para a verificação do *regime de tramitação* do projeto de lei, foi criada a variável X_2 , utilizando dados referentes ao seu *status* legislativo. Atualmente, estão disponíveis quatro categorias de *status*, sendo dois com subdivisões: tramitação em prioridade; tramitação ordinária; tramitação especial; tramitação em urgência.

Na transformação dos dados e sua normalização, utilizou-se a forma binomial. Sendo 1 para os projetos de lei com regime de urgência e prioridade e 0 para os demais. O objetivo é verificar o grau de importância que o *status* da tramitação tem sobre a variável dependente.

No processamento da variável *tempo de tramitação* do projeto de lei, variável X_3 , utilizou-se como referência a diferença entre o tempo final e inicial da tramitação, em dias. A partir da análise dos dados 2011 e 2018, foi possível estipular o tempo médio de tramitação de projetos de lei que se transformaram em norma jurídica.

A partir da média de dias de tramitação para projetos transformados em lei, foi possível dividir os projetos em grupos, entre os que tramitam com o prazo inferior à média (978 dias), ou seja, têm possibilidade de se transformar em norma jurídica, em relação àqueles que, por estarem em um prazo superior, vão perdendo a capacidade de se transformar em norma no decurso do tempo. Assim, para normalização dos dados, foi utilizado o número 1 para os projetos de lei com tempo de tramitação inferior à média e, para os que têm tempo de tramitação superior, foi utilizado o número 0.

Para a análise da *convergência de agendas* entre o executivo e a Câmara dos Deputados, criou-se a variável X_4 . Essa variável surge por meio da análise dos temas envolvidos em proposições de iniciativa de cada esfera. Para mapear os temas envolvidos na agenda do Executivo, foram utilizadas como referência as medidas provisórias. Assim, por meio da análise dos cinco principais temas envolvidos nestas propostas no ano, verificou-se a agenda do executivo e se esses temas estão presentes nos projetos de lei que foram iniciados na Câmara dos Deputados naquele ano.

Como a proposta se baseia na análise da convergência entre as agendas, foi criado uma sistematização em que, para cada projeto de lei que estivesse presente pelo menos um tema convergente com as medidas provisórias, se utiliza então o valor 1 e, então, 0 quando o projeto de lei não tem nenhuma categoria de tema convergente. Um projeto de lei pode ter um ou mais temas, por isso foi preciso agrupar os dados categóricos de modo que ele responda de forma numérica. Desse modo, os dados foram formatados por meio da técnica de pivoteamento, ou seja, as linhas com as temáticas foram transformadas em colunas numéricas.

De modo a analisar o impacto do direcionamento dos partidos em relação à *orientação de votação* a respeito do projeto de lei, foi proposta a criação da variável X_5 . Em suma, a cada projeto de lei é possível estabelecer três *status*, são eles: sim; não; liberado.

Os dados foram dispostos de forma que, para cada projeto com orientação “Sim”, recebesse o valor 1. E, para os com *status* “Não” ou “Liberado”, o valor 0. Ressalta-se que cada projeto de lei pode ter mais de uma orientação por partido. Nesse caso, o agrupamento de dados foi feito com a possibilidade de que, se, em algum momento, houver a orientação “Sim”, automaticamente inclui-se a possibilidade 1.

Para a análise da *agenda holder* foram verificadas se as posições (cargos) na Câmara dos Deputados conferem, ao agente político, capacidade de articular melhor os projetos de sua autoria, implicando, assim, êxito nas suas propostas. Assim, foram estabelecidas categorias de cargo que mais se adequam a essa capacidade de agenda, são elas: presidente e relator.

Assim, a variável *X6*, *agenda holder*, com objetivo de identificar se a ocupação de cargos relevantes na Câmara dos Deputados representa também o êxito de projetos de lei desses autores. Caso esse autor ocupasse o papel de presidente ou relator, foi atribuído o valor 1; caso não ocupasse nenhuma posição entre estas, foi atribuído com o valor 0. Ressalta-se que um projeto de lei pode ter um ou mais autores, portanto, para criar a variável *X6*, foi utilizado como critério o fato de, se a proposição tem apenas um autor com o papel de agenda, foi atribuído o número 1; caso contrário, 0.

Após a criação das variáveis, os dados passaram pelo processo de limpeza para tratar dados ausentes. Nesta etapa foi adotado o padrão de preenchimento automático, sugerido por Silva, Peres e Boscaroli (2017), preenchendo essas lacunas com o valor mais frequente para aquele dado. A Tabela 1 apresenta a estruturação do pré-processamento dos dados, conforme exposto nesta seção.

Tabela 1 – Método de pré-processamento dos dados

Variáveis	Descrição
Regime de Tramitação (X2)	Dummy: Urgência e Prioridade (1); Demais (0)
Tempo de Tramitação (X3)	Dummy: > 978 dias (1); < 978 dias (0)
Convergência de Agenda (X4)	Dummy: Convergência (1); Não tem convergência (0)
Orientação de Voto (X5)	Dummy: Orientação (1); Sem orientação (0)
Agenda Holder (X6)	Dummy: Ocupou cargo representativo (1); Não ocupou (0)
Risco de Conversão (X1)	Dummy: Sim (1); Não (0)

Fonte: Elaboração própria com dados da Câmara de Deputados (2021).

2.2 – ANÁLISE DE DADOS

A primeira seleção da amostra resultou em um conjunto de 19.106 projetos de lei que não foram transformados em norma jurídica e 546 que passaram pela transformação. Diante da diferença de dados entre as amostras que passaram por transformações jurídicas e as que não passaram, foi criada outra amostra aleatória com balanceamento dos dados, a fim de comparação e calibração do melhor modelo.

Como se trata de aprendizado de máquina, a aplicação de técnicas de balanceamento permite comparar os modelos e verificar se há distorção na predição para a classe mais popular, ou seja, a classe 0. A técnica utilizada para o balanceamento dos dados é chamada *oversampling*. Essa técnica consiste em replicar dados aleatórios das classes minoritárias para que a amostra tenha a mesma proporção de 0 e 1.

O modelo de regressão logística é um método amplamente utilizado para variáveis independentes de dois grupos binários. O modelo se caracteriza por sua robustez, facilidade de interpretação e diagnóstico. A função *logit* como a função que mapeia a probabilidade de pertencimento a uma classe como amplitude de 0 ou 1. O cálculo da regressão logística baseia-se no conceito de *chance* de determinado evento ocorrer (LEVIN, 1986; HAIR, 2005; RODRIGUES E PARREIRA, 2013; BRUCE e BRUCE, 2019).

O cálculo da regressão logística baseia-se no conceito de chance de determinado evento ocorrer, segundo Levin (1986), as chances de um evento ocorrer podem ser

definidas como a probabilidade deste evento ocorrer dividido pela probabilidade de ele não ocorrer:

$$\text{Chances} = P / 1 - P \quad (1)$$

Conforme Levin (1986), como se trata de variáveis dependentes dicotômicas, estas devem ser transformadas em probabilidade logarítmica, ou logit, com o objetivo de modelar o logaritmo natural das chances:

$$L = \log(P / 1 - P) \quad (2)$$

Explica o autor, que a probabilidade logarítmica de uma variável dependente, é baseada na combinação linear de variáveis independentes, desta forma podendo ser assim ser matematicamente representada, com duas variáveis independentes, X_1 e X_2 , da seguinte forma:

$$L = b_0 + b_1X_1 + b_2X_2 + e \quad (3)$$

Neste sentido, L representa a função logit, b_0 a constante (ou intercepto Y), b_1 e b_2 os coeficientes para as variáveis prognosticadoras X_1 e X_2 e o termo de erro ou resíduo.

Deste modo, considerando que probabilidade prevista é dada pela função de resposta logística, temos a seguinte equação de resultado:

$$\hat{P} = e^{b_0 + b_1X_1 + b_2X_2} / 1 + e^{b_0 + b_1X_1 + b_2X_2} \quad (4)$$

Finalmente, o modelo de regressão binomial escolhido para representar o logaritmo de chance de um projeto de lei virar norma jurídica com os fatores determinantes seria dado por:

$$\hat{P} = e^{b_0 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6} / 1 + e^{b_0 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6} \quad (5)$$

Seguindo a metodologia de construção e avaliação de modelos preditivos propostos por Silva, Peres e Boscardioli (2017), usou-se inicialmente a estratégia *holdout* que pressupõe a criação de dois subconjuntos de dados disjuntos, a partir do conjunto de dados disponível para o uso na indução do modelo. Um conjunto de dados será usado para treinamento e o outro para teste após a construção do modelo.

Desta forma, a amostra foi dividida em 67% dos dados para a amostra de treino e 33% para as amostras de teste, porcentagem estipulados na metodologia proposta por Silva, Peres e Boscardioli (2017), ficando dispostas conforme a Tabela 2.

Tabela 2 – Disposição dos dados de treinamento e teste para os dados balanceados e não balanceados

Amostra de Treino Não Balanceado	Amostra de Teste Não Balanceado	Amostra de Treino Balanceado	Amostra de Teste Balanceado
13166	6486	25598	12608

Fonte: Elaboração própria com dados da Câmara de Deputados (2021).

Ainda em relação à preparação do modelo foi necessário verificar o coeficiente de correlação das variáveis independentes - X_2 , X_3 , X_4 , X_5 , X_6 - com a variável dependente X_1 e o grau de significância representativo de cada uma, de forma a melhor determinar quais variáveis poderiam agregar ao modelo proposto. Esta seleção, chamada *stepwise* segundo Hair et al. (2008), tem objetivo de determinar, entre as variáveis independentes, aquelas que têm poder explicativo experimental e podem acrescentar ao modelo.

Após a preparação dos dados, foi realizado o treinamento do modelo. Nesta etapa do aprendizado de máquina, foi utilizada a biblioteca *scikit-learn*, que oferece ferramentas importantes para as várias fases dos projetos de *machine learning*. O treinamento foi realizado sobre a amostra de treino e a predição na amostra de teste, conforme citam Silva, Peres e Boscaroli (2017). Os autores afirmam que a amostra de teste é utilizada após o término do treinamento e, conseqüentemente, para aplicação de medidas de avaliação do modelo.

Quanto às medidas de avaliação propostas, foram analisados a acurácia, *f1-score* e a matriz de confusão, seguindo as medidas mais comumente usadas para avaliação de classificadores, conforme metodologia descrita por Silva, Peres e Boscaroli (2017). Para os autores a acurácia pode ser descrita como erro de generalização e uma função de perda binária, sendo interpretada como a probabilidade de ocorrer uma classificação correta.

3 – RESULTADOS E DISCUSSÕES

Após o pré-processamento dos dados, iniciou-se o processo de execução da análise. Na amostra com dados completos balanceados, usando *oversampling*, todas as variáveis explicativas são estatisticamente significativas 0.0001%. Dessa forma, foi mantida para o treinamento e teste do modelo, como apresentado na Tabela 3.

Tabela 3 – Modelo de determinantes do risco legislativo com dados balanceados (*oversampling* e sem *stepwise*) – Câmara Federal - 2021

Generalized Linear Model Regression Results						
Dep. Variable:	X1	No. Observations:	38206			
Model:	GLM	Df Residuals:	38200			
Model Family:	Binomial	Df Model:	5			
Link Function:	logit	Scale:	1.0000			
Method:	IRLS	Log-Likelihood:	-23281.			
Date:	Thu, 01 Apr 2021	Deviance:	46563.			
Time:	13:58:10	Pearson chi2:	3.84e+04			
No. Iterations:	7					
Covariance Type:	nonrobust					
	coef	std err	z	P> z	[0.025	0.975]
Intercept	-1.3611	0.026	-51.953	0.000	-1.412	-1.310
X2	1.0683	0.023	45.748	0.000	1.023	1.114
X3	0.1424	0.023	6.183	0.000	0.097	0.188
X4	-0.3678	0.024	-15.369	0.000	-0.415	-0.321
X5	3.0654	0.229	13.413	0.000	2.618	3.513
X6	1.1551	0.027	43.421	0.000	1.103	1.207

Fonte: Elaboração própria com dados da Câmara de Deputados (2021).

Após a análise discriminante de dados, foi realizada a etapa de treinamento do modelo de regressão logística, com os dados não balanceados e balanceados. Com o uso do *oversampling*, o modelo apresentou indicadores de acurácia (70%) e *f1-score* ($Z_0 = 72\%$ e $Z_1 = 67\%$), reduzindo o risco de preditores de falsos negativos e falsos positivos. Ao final do processo de medição, verificou-se a matriz de confusão de cada amostra de teste sobre os dados da amostra de teste, Tabela 4.

Tabela 4 – Matriz de confusão - amostra de aprendizagem do modelo logístico de risco para projetos de desenvolvimento regional na Zona Franca de Manaus – 2021.

Matriz 1 – dados não balanceados			Matriz 2 - dados balanceados		
X1	0	1	X1	0	1
0	6311	3	0	4944	1353
1	166	6	1	2477	3834

Fonte: Elaboração própria com dados da Câmara de Deputados (2021).

No modelo final de regressão logística, pode-se perceber que as variáveis X_2 , X_3 , X_5 , X_6 afetaram, significativamente, os modelos, enquanto a variável X_4 mostrou-se significativa para o modelo balanceado e sem significância para o não balanceado. Em geral, os dois modelos propostos apresentaram, aproximadamente, o mesmo coeficiente de regressão. Destacam-se os coeficientes de regressão altos das variáveis X_2 , X_5 e X_6 em relação à variável independente X_1 e negativo para a variável X_4 .

Como a variável X_2 trata do processo de tramitação especial de projetos de lei, era esperado, como indicado na literatura, que ele tivesse maior probabilidade de transformação em norma jurídica. Porém, além dela, a variável X_6 que representa os *agenda holders* também obteve um coeficiente de regressão promissor e próximo ao percentual obtido com a variável X_2 . Dessa forma, a variável X_6 , apesar de processos de transformações de dados mais complexos do que a variável X_2 , conseguiu, em termos percentuais, traduzir inicialmente o fenômeno social proposto.

Quanto à variável X_3 , que representa o tempo de tramitação, apesar de ter apresentado significância, não obteve desempenho similar às demais variáveis. Dessa forma, pelo fato de a base de construção estar relacionada à média de tempo, trabalhos futuros podem utilizar outras métricas que maximizem o resultado na separação dos grupos.

Paralelamente, a variável X_5 apresentou o índice de correlação mais relevante entre as variáveis, nos dois modelos. Dessa forma, podemos determinar que a orientação de voto de bancada tem grande relevância no processo, apesar de rara na quantidade de processos que tramitam na Câmara dos Deputados. Isto indica que a quantidade de influenciadores na tramitação de projetos de lei tem peso sobre o processo legislativo, assim como retrata a importância daquele tema para a agenda do parlamento, e não somente para determinados agentes.

Em suma, os dois modelos tiveram diferenças relevantes, o que determinou a escolha de um deles como o melhor modelo, de acordo com a metodologia utilizada. Assim, o modelo sem balanceamento, apesar de ter apresentado uma acurácia de 97%, ao mesmo tempo, apresentou um *f1-score* baixo para a predição “1”, de apenas 7%. Ao contrário do modelo balanceado que apresentou acurácia menor, porém com *f1-score* mais equilibrado para predizer “0” e “1”, respectivamente 72% e 67%.

Além disso, a matriz de confusão no modelo balanceado apresentou resultados satisfatórios para a predição de valores verdadeiros positivos e falsos negativos, enquanto o modelo não balanceado não soube predizer com eficiência os valores

verdadeiros positivos, apresentando maior quantidade de falsos positivos. Em vista desses resultados, optou-se pela utilização do modelo balanceado para análise dos dados, pois este se encontra com padrões mais homogêneos em relação às medidas de avaliação utilizadas.

Os projetos de lei analisados nesta pesquisa se baseiam em 4 grupos temáticos principais que afetam as atividades desenvolvidas pelo Polo Industrial de Manaus, são eles: criação de zonas francas; ampliação de área de abrangências; processo produtivo básico; e incentivos fiscais. Foram analisados somente os projetos de lei com tramitação iniciada em 2019. Ressalta-se, ainda, que a tramitação negativa e positiva foi estabelecida de acordo com a análise do tema e as atribuições do modelo Zona Franca de Manaus. Dessa forma, segue a relação de assunto, tema e tramitações para cada projeto a ser analisado no modelo proposto, conforme exposto na Tabela 5.

Quadro 2 — Projetos em tramitação 2019 e 2020 correlatos à SUFRAMA para análise preditiva e explicativa

Projeto de Lei	Assunto	Tema	Tramitação	Chance de Aprovação
1077/2019	Altera o Decreto-Lei n.º 288, de 28 de fevereiro de 1967, para estabelecer prazo máximo para análise de proposta de um Processo Produtivo Básico - PPB.	Processo Produtivo Básico	Positiva	22%
4247/2019	Cria a Zona Franca do Distrito Federal e Entorno	Criação de zonas francas	Negativa	17%
801/2019	Altera a Lei n.º 7.965, de 22 de dezembro de 1989, que “Cria Área de Livre Comércio no Município de Tabatinga, no Estado do Amazonas, e dá outras providências”.	Ampliação Área de Abrangência	Positiva	17%
3107/2019	Isenta de IPI bicicletas comuns e elétricas	Incentivos Fiscais	Negativa	39%

Fonte: Elaboração própria com dados da Câmara de Deputados (2021).

Para análise explicativa do tema processo produtivo básico, temos a PL 1077/2019. Esse projeto de lei trata do Decreto-Lei n.º 288, de fevereiro de 1967, que estabelece o prazo máximo para análise de proposta de um processo produtivo básico. Atualmente, há demora na aprovação de processos produtivos básicos, o que prejudica as empresas a se instalarem no Polo Industrial de Manaus, pois é condicionante para produção de novos produtos tecnológicos. O tempo é crucial principalmente quando se trata de um mercado competitivo.

Pelo fato de o projeto facilitar o tempo de tramitação de processos produtivos básicos, este foi avaliado como positivo para modelo. Trata-se de um tema sensível, com repercussão para melhorar as atividades do polo. O projeto em questão apresentou no modelo, uma chance de 22% para transformação em norma jurídica. Como está na fase inicial de tramitação, ou seja, somente passou por uma comissão até o momento, algumas variáveis podem sofrer alteração, de modo a tornar o projeto mais consistente. Dessa forma, destaca-se a capacidade adaptativa dos dados e do modelo, ao longo do tempo.

Para a análise do tema referente à criação de novas zonas francas, usou-se como base ilustrativa o PL 4247/2019. Esse projeto de lei trata especificamente da criação Zona Franca do Distrito Federal e do entorno e visa avançar com novas políticas de desenvolvimento regional com o estímulo à atividade econômica. Atualmente, tramitam na Câmara de Deputados outros projetos com a mesma temática, como o projeto de lei 1525/2019 que cria a Zona Franca Da Baixada Fluminense, no Estado do Rio de Janeiro e o projeto de lei 1858/2019 que cria a Zona Franca da Indústria Calçadista. No modelo estatístico proposto este projeto apresentou uma chance estatística de 17% para a transformação em norma jurídica. Como está na fase inicial de tramitação, ou seja, somente passou por uma comissão até o momento, algumas variáveis podem sofrer alteração, de modo a tornar o projeto mais consistente. Dessa forma, destaca-se a capacidade adaptativa dos dados e do modelo, ao longo do tempo.

A ampliação da área de abrangência do modelo Zona Franca de Manaus é importante, pois impulsiona o desenvolvimento em toda a Amazônia Ocidental. Atualmente, além do Polo Industrial de Manaus, o modelo conta com as áreas de livre comércio em região de fronteira, como a Área de Livre Comércio de Tabatinga. Esta área está localizada entre os municípios de Tabatinga e Letícia e tem como ponto forte o comércio da cidade, além da pesca e agricultura. Como se trata de um ponto estratégico de desenvolvimento dessa região de fronteira, esta conta benefícios fiscais e administração pela SUFRAMA.

Especificamente no caso das áreas de livre comércio tramitam, na Câmara dos Deputados, projetos que visam ampliar a sua área geográfica de modo a abarcar maior quantidade de empresas e ampliar aquelas já existentes. A ampliação de área ocorre pela delimitação de zona de comprimento e faixa institucionalizada pelo decreto que criou a Zona Franca de Manaus. Um desses projetos de lei, refere-se ao PL 801/2019, que altera a Lei n.º 7.965, de 22 de dezembro de 1989, que cria Área de Livre Comércio no Município de Tabatinga, no Estado do Amazonas, e dá outras providências.

Esse projeto, com tema positivo para o modelo, proporciona maior desenvolvimento para região incentivada e proporciona a atração de novos investimentos, além da área delimitada. No modelo, o projeto de lei 801/2019 apresentou uma chance de 17% de se transformar em norma jurídica. Como está na fase inicial de tramitação, ou seja, somente passou por uma comissão até o momento, algumas variáveis podem sofrer alteração, de modo a tornar o projeto mais consistente. Dessa forma, destaca-se a capacidade adaptativa dos dados e do modelo, ao longo de tempo.

Os incentivos fiscais são o grande atrativo de investimentos no modelo Zona Franca de Manaus. Entre os impostos concedidos estão o Imposto de Importação (II), Imposto de Produtos Industrializados (IPI), Programa de Integração Social e Financiamento da Seguridade Social (PIS/COFINS) e Redução do Imposto de Renda (IR). É constante a atuação legislativa de modo a levar esses benefícios fiscais para determinadas categorias, beneficiando, assim, estruturas fora da Amazônia Ocidental. Um dos polos que enfrenta esta situação, é também um dos mais desenvolvidos no modelo, o polo de duas rodas, representado por motocicletas, ciclomotores, motonetas, bicicletas e similares.

O polo de duas rodas, no ano de 2013, foi responsável por 16,71% do total do faturamento do modelo, movimentando cerca de 14 bilhões de reais, e foi responsável por bons números na geração de emprego. Apesar importante para a diminuição do número de veículos motorizados nas ruas, essa mesma medida proposta pela PL 3107/2019, que isenta de IPI bicicletas comuns e elétricas, pode levar a fuga de capitais da zona franca para outros estados mais desenvolvidos, aumentando, assim, a desigualdade intrarregional.

Logo, no modelo, o PL 3107/2019 apresentou uma predição de chances de aprovação de 39%. Como está na fase inicial de tramitação, ou seja, somente passou por uma comissão até o momento, algumas variáveis podem sofrer alteração, de modo a tornar o projeto mais consistente. Dessa forma, destaca-se a capacidade adaptativa dos dados e do modelo, ao longo de tempo.

Quanto à análise dos projetos com temas pertinentes à Zona Franca de Manaus, ficou demonstrado que, apesar de serem temas importantes, as atividades do polo e, na maioria, estarem na agenda do executivo para o ano, eles não recebem a atenção de grupos e atores que podem representar influência para a sua efetivação até o momento. Por isso, destaca-se a importância e a necessidade do acompanhamento diário das matérias, assim como o estabelecimento de padrões que tornem a atualização dos dados uma realidade, pois a dinâmica do processo legislativo exige respostas rápidas.

Por fim, foi possível perceber que o avanço dos dados, além de enriquecedor para as políticas públicas, agregando informações aos tomadores de decisão, também permitem inovações tanto na área de transparência, como na visualização de dados. Abre-se caminho com este trabalho, para a melhoria das representações institucionais, não somente da SUFRAMA, mas dos órgãos que atuam juntamente ao Congresso Nacional e que podem valer-se de outras informações estratégicas, além daquelas típicas da atividade.

CONSIDERAÇÕES FINAIS

Para esta pesquisa utilizado o modelo de regressão logística, com *machine learning*, de modo a prever a capacidade de um projeto de lei se transformar em norma jurídica. Para a realização deste trabalho, utilizou-se dos dados abertos legislativos, metodologia clara e aprendizagem de máquina com mais de 19 mil projetos de lei obtidos dos dados abertos da Câmara de Deputados Federais. De forma que pode ser utilizado para qualquer predição de matérias que tramitam no Congresso Nacional.

Pelo processo de aprendizagem de máquinas com método de balanceamento oversampling foi possível encontrar um modelo estatístico logístico com boa acurácia (70%) e com boa precisão (72%; 67%), com bom nível de prevenção de predições de falsos positivos e falsos negativos. Também se destaca que o nível de significância das variáveis X₂, X₃, X₄, X₅ e X₆, respectivamente relacionadas ao regime de tramitação, tempo de tramitação, convergência de agendas, orientação de votação e agenda holder foi de 0.0001%, o que demonstra grande robustez das predições, muito acima de quando apenas se roda modelos estatísticos estáticos, no caso foi utilizado base de treinamento, com sucessivas fases de processamento, onde os resultados foram aprimorados a cada fase.

Quanto à análise dos projetos com temas pertinentes à Zona Franca de Manaus, ficou demonstrado que, apesar de serem temas importantes, as atividades do polo e, na maioria, estarem na agenda do executivo para o ano, eles não recebem a atenção de grupos e atores que podem representar influência para a sua efetivação até o momento. A única exceção é o projeto de isenção de IPI bicicletas comuns e elétricas que pelo modelo apresentou uma chance estatística de 39% de ser aprovado, o que pode levar uma grande perda de competitividade da Zona Franca de Manaus.

Por fim, foi possível perceber que o avanço dos dados, além de enriquecedor para as políticas públicas, agregando informações aos tomadores de decisão, também permitem inovações tanto na área de transparência, como na visualização de dados. Abre-se caminho com este trabalho, para a melhoria das representações institucionais, não somente da SUFRAMA, mas dos órgãos que atuam juntamente ao Congresso

Nacional e que podem valer-se de outras informações estratégicas, além daquelas típicas da atividade.

REFERÊNCIAS

- BRUCE, A.; BRUCE, P. **Estatística Prática para Cientistas de Dados**. Alta Books, v. 3, f. 196, 2019. 392 p.
- CAPELLA, A. C. **Formulação de Políticas Públicas**. Brasília: Enap, 2018. 152 p.
- GIAMBIAGI, F. Crise fiscal da União: O que aconteceu recentemente? **Revista de Economia Política**, v. 17, n. 1, p. 65, 1997.
- HAIR, J. F., BLACK, W. C., BABIN, B. J., ANDERSON, R. E., & TATHAM, R. L. **Análise multivariada de dados - 6ed**. Bookman Editora, f. 344, 2008. 688 p.
- HOLLAND, M., GURGEL, A., CERQUEIRA, C., SERIGATTI, F., GELCER, D., ANDRADE, J., & PIERI, R. Zona Franca de Manaus: Impacto, Efetividade e Oportunidades. **FGV EESP**, São Paulo, p. 102, 2019.
- JUNQUEIRA, A. Desvendando o Papel da Estrutura de Comissões para a atuação da Oposição. **VI Seminário Discente do Programa de Pós-Graduação em Ciência Política**, p. 26, 2016.
- LEVIN, J. **Estatística aplicada às ciências humanas**, f. 199. 1986. 397 p.
- MENDONÇA, M. B. **O processo de decisão política e a Zona Franca de Manaus**. São Paulo, 2013. 291 p. Tese (Administração Públicas e Governo) - Fundação Getúlio Vargas, São Paulo, 2013.
- MENEGUIN, F. B., & SILVEIRA, R. **Avaliação de impacto legislativo: cenários e perspectivas para sua aplicação**. Brasília: Senado Federal, 2017. 127 p.
- NASCIMENTO, S. P. Guerra Fiscal: uma avaliação comparativa entre alguns estados participantes. **Economia Aplicada**, v. 12, n. 4, p. 667-706, 2008.
- NICOLAU, J. Disciplina partidária e base parlamentar na Câmara dos Deputados no primeiro governo Fernando Henrique Cardoso (1995-1998). **Dados**, v. 43, n. 4, p. 00-00, 1999.
- PEREIRA, C.; MUELLER, B. Uma Teoria da Preponderância do Poder Executivo: O sistema de Comissões no Legislativo Brasileiro. **Revista Brasileira de Ciências Sociais**, v. 15, n. 43, p. 23, 2000.
- RODRIGUES, W.; PARREIRA, L. A. Análise do risco ao desemprego entre grupos demográficos no município de Palmas-TO no ano de 2008: uma aplicação do modelo de regressão logístico binomial. **Informe GEPEC**, v. 17, n. 1, p. 23-33, 2013.
- SANTOS, F.; BORGES, M. **Poder de agenda**. Brasília: ENAP, 2018. 95 p.

SANTOS, M. Representação de Interesses na Câmara dos Deputados: o Lobby e o Sucesso da Agenda Legislativa da Indústria. **Revista Ibero-Americana de Estudos Legislativos**, Rio de Janeiro, v. 3, n. 1, p. 52-70, 2014.

SILVA, L. A.; PERES, S. M.; BOSCARIOLI, C. **Introdução à Mineração de Dados**: Com Aplicações em R. Elsevier Brasil, v. 2, f. 148, 2017. 296 p.

SILVA, R. S. e; ARAÚJO, S. M. V. G. Apropriação da agenda do legislativo: como aferir esse fenômeno. **exto para Discussão. Centro de Estudos da Consultoria**, Senado, v. 76, p. 40, 2010.

SUPERINTENDÊNCIA DA ZONA FRANCA DE MANAUS (SUFRAMA). Relatório de Gestão 2018. **Governo Federal**, p. 146, 2019.

VARSANO, R. A guerra fiscal do ICMS: Quem ganha e quem perde. **Planejamento e Políticas Públicas**, Rio de Janeiro, n. 15, p. 18, 1997.

Submetido em 15/12/2021.
Aprovado em 20/05/2022.