

# Mecanismo de Previsão de Abstenção em Eleições por Município

Pedro Ferreira  
Dept. of Electrical Engineering  
University of Brasilia  
Brasilia, Brazil  
pedrohgf@globomail.com

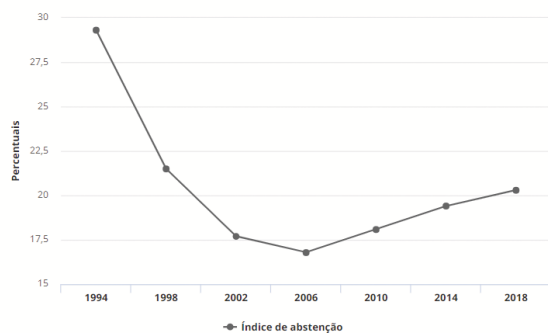
Thiago Dantas  
Dept. of Electrical Engineering  
University of Brasilia  
Brasilia, Brazil  
thiagodma52@gmail.com

**Resumo**—Este trabalho consiste no desenvolvimento de um modelo de previsão de nível de abstenção por município brasileiro por meio de características socioeconômicas do município por meio da aplicação de técnicas de aprendizado de máquina. Para isso, múltiplos algoritmos foram empregados e comparados quanto à performance, com os melhores resultados sendo obtidos pela técnica de regressão por Florestas Randômicas, produzindo um modelo capaz de explicar em torno de 50% da variância da amostra.

## I. INTRODUÇÃO

Um dos principais desafios de regimes democráticos é a elaboração de mecanismos para universalizar o pleno exercício da cidadania, garantindo, independente de obstáculos de cunho socioeconômico ou geográfico, o acesso dos cidadãos a seus direitos, bem como promovendo a participação popular nos processos da sociedade.

Atualmente, no Brasil, embora o sufrágio universal seja um direito garantido constitucionalmente, sendo um dos elementos fundamentais da representatividade e da democracia, apenas 80% das pessoas aptas a votar exercem seu direito.



Fonte: TSE

Figura 1: Nível de Abstenção Média por Eleição no Brasil

Uma vez que a decisão de se abster é tomada individualmente por cada eleitor, considerando diversos fatores socioeconômicos, políticos e culturais, como a facilidade de acesso ao local de votação, o valor atribuído ao voto pelo eleitor, a afinidade com os candidatos e a competitividade

da eleição, a utilização de um modelo em escala estadual ou nacional, ainda que esses sejam capazes de promover um bom ajuste, não permite a adoção de forma orientada de políticas públicas eficazes para o estimular o comparecimento do eleitorado.

Dessa forma, a elaboração de um modelo de previsão de nível de abstenção em escala municipal, levando em conta fatores socioeconômicos e políticos, surge como uma alternativa adequada para a avaliação do comparecimento do eleitorado, auxiliando na aferição do impacto de fatores estruturais sobre as taxas de abstenção, bem como permitindo a identificação de municípios com abstenções excessivamente elevadas para sua configuração socioeconômica, facilitando, assim, o planejamento e a adoção de políticas públicas orientadas.

Embora estudos sobre abstenção em processos eleitorais sejam um tópico de interesse na ciência política, estando bem documentados na literatura, a maior parte dos trabalhos realizados no contexto nacional consistem na elaboração de modelos em escala nacional ou estadual [4], utilizando, para a estimação, técnicas de regressão linear ou logística baseadas sobretudo em características políticas subjetivas do pleito [4]- [6].

Este trabalho se propõe a desenvolver um modelo de previsão de níveis de abstenção, em escala de município, baseado, principalmente, nas características socioeconômicas do eleitorado, permitindo uma análise mais detalhada sobre os fatores que influenciam na abstenção.

Para assegurar um bom ajuste do modelo, foram testadas diferentes técnicas de aprendizado de máquina, como regressão linear multivariada, máquina de suporte de vetores para regressão, regressão de k-vizinhos mais próximos e florestas randômicas, sendo comparadas quanto ao seu desempenho principalmente em termos de seu coeficiente de determinação,  $R^2$ . Foi testada, também, uma abordagem baseada em uma rede neural MLP para regressão.

## II. MODELO

### A. Descrição dos Dados

Buscando agregar dados regionais e socioeconômicos de cada município em quantidade suficiente para a realização de uma boa regressão, foram utilizados os dados do Censo

de 2010 do IBGE [1], compostos pelo perfil detalhado dos 5564 municípios quanto à 114 características de sua população.

Por sua vez, os dados de abstenção nas eleições subsequentes ao censo (2012, 2014, 2016 e 2018) foram obtidas por meio dos boletins de urna de cada estado, disponíveis no *site* do TSE [2]. Os resultados de cada seção foram agregados no resultado de sua zona eleitoral correspondente e os resultados de todas as zonas de cada municípios foram agregados, gerando os resultados da eleição por município.

Para cada município, 42 características socioeconômicas foram extraídas diretamente do Censo ou geradas por meio de combinações entre características disponíveis. Por sua vez, foram obtidos, também, 4 parâmetros eleitorais a partir dos boletins de urna.

Assim, cada município teve o comportamento avaliado de acordo com as 46 características dispostas na tabela (I). As primeiras *features* da primeira coluna foram escolhidas visando esclarecer a contribuição de dificuldades geográficas nas abstenções - em municípios com maior área, o acesso às urnas poderia ser mais difícil, o que poderia se traduzir em aumento nas abstenções -, sendo complementado por uma caracterização do município em termos estruturais, considerando aspectos como a facilidade de acesso à educação, saúde, o poder de compra da população e seu IDH.

Em sequência, ainda na primeira coluna, estão listadas as características relacionadas à discriminação da população do município em termos de gênero, preferência religiosa, poder de compra, finalidade social do espaço (rural ou urbano) ocupado, concentração de renda e grau de escolaridade. A segunda coluna mantém o padrão iniciado pela primeira, sendo composta por características de discriminação da população quanto à etnia, renda familiar e dependência de programas sociais. Há, também, uma caracterização cultural e social dos núcleos familiares, sendo representada pelos atributos de tipo de unidade domiciliar. Por fim, são representados atributos eleitorais específicos de cada eleição, como a votação do primeiro colocado, a votação do segundo colocado, a apatia eleitoral - medida pela proporção de votos inválidos - e a entropia da eleição, definida como a entropia no sentido de Shannon do processo eleitoral, sendo uma medida do quão equilibrada está a eleição.

### B. Determinação do Problema e Abordagem

Esse trabalho se desenvolve em torno da determinação de um modelo de previsão de nível de abstenção, no Brasil, à nível de município, capaz de auxiliar na elaboração de políticas públicas para aumentar os níveis de participação do eleitorado.

Nesse cenário, uma importante restrição para a elaboração do modelo é a questão da precedência temporal. Como a adoção de políticas públicas para estimular a participação do eleitorado deve ser feita antes da eleição, é necessário que as características utilizadas para prever

o nível de abstenção estejam disponíveis ou possam ser estimados de alguma outra forma antes da realização do pleito. É importante perceber que, no cenário onde algumas das características assumem valores estimados, os valores estimados devem ser substituídos pelos valores reais aferidos no processo de votação antes de realimentarem o modelo.

Considerando que a estimação da abstenção será realizada em nível municipal, foi feita a opção de utilizar o Censo do IBGE para a caracterização socioeconômica da população, uma vez que este é o único levantamento estruturado com abrangência suficiente para permitir a determinação das características. Por sua vez, optou-se por utilizar as informações disponibilizadas pelo TSE para a determinação do perfil político de cada pleito.

Uma vez que os Censos são realizados apenas a cada dez anos, enquanto são realizadas eleições a cada dois anos, é necessário realizarmos uma nova etapa de processamento nos dados. Uma possível estratégia seria buscar atualizar, para cada eleição, os dados disponíveis no Censo utilizando índices gerados por outros órgãos. No entanto, como as variações ocorrem de forma não-uniformes entre os municípios e entre o eleitorado, essa estratégia acabaria por introduzir ruído nos dados.

Dessa forma, partindo do pressuposto de que o nível de abstenção médio em um município é, na realidade, produto da característica socioeconômica e de fatores políticos comuns em todas as eleições, foi feita a opção de se desmembrar a geração do modelo em duas partes: uma composta pela abstenção média de cada município e outra composta pela abstenção específica de cada eleição, isto é, pela diferença entre a abstenção da eleição e a abstenção média.

Dessa forma, é adequado utilizar, para a determinação da primeira parcela do modelo, apenas as características socioeconômicas dos municípios, enquanto é possível estimar a segunda parcela utilizando apenas as características próprias de cada eleição, isto é, suas características políticas particulares.

Uma vez que a primeira parcela da abstenção representa a abstenção média, por analogia com a teoria de sinais, essa parcela será doravante identificada por "abstenção DC", enquanto a segunda parcela, correspondente à uma abstenção de uma eleição específica e, portanto, variável no tempo, será denominada "abstenção AC".

## III. SOLUÇÃO E ANÁLISE

Nesta seção, todos os testes foram feitos usando uma separação dos dados em dados de teste (30%) e dados de treinamento (70%).

### A. Aquisição dos Dados

Os dados utilizados neste projeto foram nos *sites* do IBGE (censo) e do TSE (abstenções). Como já era esperado, diversos desafios relacionados à situação dos dados surgiram.

Área do Município	População com Ensino Superior Completo
Número de Estabelecimentos do SUS	Unidades Domésticas do Tipo Homens com Filhos
Alunos Matriculados no Ensino Médio	Unidades Domésticas do Tipo Mulheres com Filhos
Alunos Matriculados no Ensino Fundamental	Percentual de Amarelos na População
Alunos Matriculados em Creche ou Escola Primária	Percentual de Brancos na População
PIB per capita	Percentual de Indígenas na População
IDH	Percentual de Pardos na População
População - Homens	Percentual de Negros na População
População - Mulheres	Percentual de Pessoas de Cor Não-Declarada na População
População residente de Religião Espírita	Percentual de Beneficiários de Programas Assistenciais
População residente de Religião Evangélica	Percentual de Famílias com Renda Familiar de até 1/8 de Salário Mínimo
População residente de Religião Católica	Percentual de Famílias com Renda Familiar de até 1/4 de Salário Mínimo
Rendimento Médio em Domicílios Rurais	Percentual de Famílias com Renda Familiar de até 1/2 de Salário Mínimo
Rendimento Mediano em Domicílios Rurais	Percentual de Famílias com Renda Familiar de até 1 Salário Mínimo
Rendimento Médio em Domicílios Rurais	Percentual de Famílias com Renda Familiar de até 2 Salários Mínimos
Rendimento Mediano em Domicílios Urbanos	Percentual de Famílias com Renda Familiar de até 3 Salários Mínimos
População Ocupada no Município de Residência	Percentual de Famílias com Renda Familiar de até 5 Salários Mínimos
População Ocupada em Outros Municípios	Percentual de Famílias com Renda Familiar de até 10 Salários Mínimos
Taxa de Analfabetismo da População no Ano de 2000	Percentual de Famílias com Renda Familiar superior à 10 Salários Mínimos
Taxa de Analfabetismo da População no Ano de 2010	Percentual de Votos do 1º Colocado da Eleição
População Sem Instrução ou com Ensino Fundamental Incompleto	Percentual de Votos do 2º Colocado da Eleição
População com Ensino Fundamental Completo e Médio Incompleto	Percentual de Votos Inválidos na Eleição
População com Ensino Médio Completo e Superior Incompleto	Entropia da Eleição

Tabela I: Atributos utilizados para a previsão da abstenção

Primeiramente, foi necessário escolher quais seriam as *features* utilizadas no modelo DC. Esses atributos foram escolhidos a partir dos dados do censo de 2010. O procedimento utilizado para seleção das *features* foi, basicamente, olhar uma a uma das planilhas dos dados do censo de 2010 e escolher as que se considerou mais relevantes para estimar a abstenção a nível municipal.

Com as *features* determinadas, começou a etapa do tratamento dos dados. Para isso foram utilizadas as bibliotecas do *Python*: *pandas* e *numpy*. Algumas ações tomadas nesta etapa foram: correção de dados incorretos, conversão dos tipos de variáveis, padronização dos nomes dos municípios (para facilitar a combinação dos bancos de dados) entre outras.

Para obtenção das *features* do modelo AC foram utilizados os dados do TSE. Esses dados foram passados em um *parser* feito em MATLAB que gerou as *features* do modelo AC e a abstenção por município, turno e ano de eleição.

#### B. Manipulação dos Dados e Determinação dos Alvos dos Regressores

Com todos os dados em mão, partiu-se então para a determinação do alvo do regressor no modelo DC. Para isso, basicamente foi feita a média aritmética das abstenções por município, ou seja, foram somadas as taxas de abstenção de todos os anos em todos os turnos e o resultado dividido pelo número de parcelas da soma.

De forma similar, o alvo do regressor no modelo AC foi calculado como a subtração dois indicadores. O primeiro deles sendo a taxa de abstenção em um determinado ano, turno e município e o segundo deles sendo a abstenção média do município (esse é o alvo do modelo DC).

Por fim, com as *features* para os modelos DC e AC determinadas e tratadas, elas foram normalizadas para

evitar que diferenças na ordem de grandeza dos atributos afetem na estimação feita pelo regressor. Este é um ponto importante para o bom funcionamento de diversos algoritmos de aprendizado de máquina.

#### C. Determinação do Regressor Utilizado para o Modelo DC

Para determinação do regressor que será utilizado para fazer a estimação do nível de abstenção foram elencados alguns regressores que acredita-se que são válidos para o problema em questão:

- 1) **Florestas Randômicas para Regressão:** esse modelo foi escolhido como candidato pois é um regressor capaz de aprender funções não-lineares complexas. Outra razão para a escolha desse modelo está relacionada ao fato de que após o treinamento é possível compreender quais atributos foram mais relevantes para a determinação da previsão. Essa característica desse modelo é de grande valia para o problema em questão.
- 2) **K-Vizinhos mais Próximos para Regressão:** esse modelo foi escolhido como candidato pois é um regressor capaz de aprender funções não-lineares complexas.
- 3) **Regressão de Menor Ângulo:** esse modelo foi escolhido como candidato pois trata-se de um modelo linear que utiliza como métrica de distância a distância de Mahalanobis.
- 4) **Perceptron Multicamadas para Regressão:** esse modelo foi escolhido como candidato pois é um regressor capaz de aprender funções não-lineares complexas.
- 5) **Máquina de Vetor de Suporte para Regressão Linear:** esse modelo foi escolhido como candidato

para avaliar se a função que deseja-se aproximar é linear ou aproximável por uma reta.

- 6) **Máquina de Vetor de Suporte para Regressão Não-Linear**: esse modelo foi escolhido como candidato pois é um regressor capaz de aprender funções não-lineares complexas.

Para escolher qual dos regressores aproxima melhor a função objetivo, foram feitos diversos testes e utilizada como métrica de qualidade do regressor obtido o coeficiente de determinação  $R^2$ . Os resultados obtidos estão dispostos na tabela II.

Tabela II: Comparação entre os Regressores

Algoritmo	$R^2$
Floresta Randômica	0.5050
K-Vizinhos	0.3821
Regressão de Menor Ângulo	-22453589152591.74
Perceptron Multicamadas	0.1060
Máquina de Vetor de Suporte Linear	-0.04697
Máquina de Vetor de Suporte Não-Linear	-0.04697

O coeficiente de determinação é uma métrica utilizada para avaliar a qualidade de um regressor, de forma que quanto maior é o  $R^2$ , maior a capacidade do modelo regressivo explicar a variância observada na amostra. Valores negativos de  $R^2$  indicam que o regressor se ajusta aos dados pior do que um regressor trivial (aquele que aproxima qualquer valor pela média da amostra).

Avaliando-se os resultados obtidos conclui-se que o modelo de regressor que melhor se ajustou aos dados foi o modelo de Floresta Randômica. Esse resultado somado com o fato de que o algoritmo de florestas randômicas não é uma caixa preta, em que o processo de regressão é algo obscuro, mas que permite avaliar quais atributos são os mais importantes para a regressão, fortalece a hipótese que, entre os algoritmos testados e dado o objetivo do projeto, o de Floresta Randômica será o mais apropriado.

#### D. Determinação dos Parâmetros da Floresta Randômica do Modelo DC

Os resultado obtido na seção anterior corresponde a uma Floresta Randômica definida com os seguinte parâmetros:

- $n\_estimators$ : 500;
- $max\_depth$ : 100;

Os outros parâmetros do regressor foram mantidos em seus valores *default*. A documentação da classe utilizada com os valores padrão de cada parâmetro da Floresta Randômica pode ser encontrada em [7].

Para avaliar se os parâmetros escolhidos são razoáveis, serão feitos alguns testes seguindo a seguinte metodologia:

- 1) Varia-se o parâmetro  $n\_estimators$  com o  $max\_depth$  fixo em 100 e identifica-se para qual valor de  $n\_estimators$  obtém-se o maior  $R^2$ .
- 2) Fixa-se o parâmetro  $n\_estimators$  para o valor ótimo na etapa anterior e varia-se o parâmetro  $max\_depth$  para determinar seu valor ótimo.

Os resultados obtidos por estes testes estão dispostos nas figuras 2 e 3.

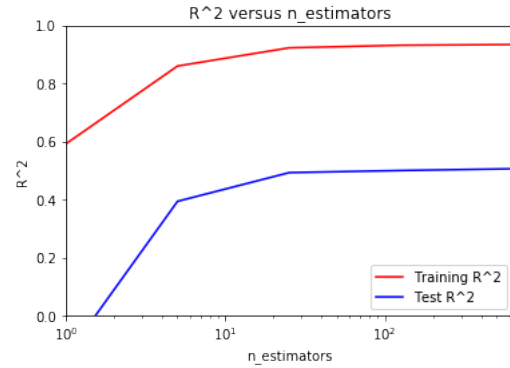


Figura 2

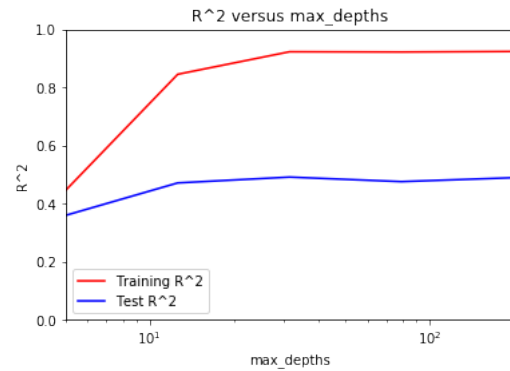


Figura 3

Avaliando os gráficos na figura 2, conclui-se que a região em que o número de estimadores é menor que 30 consistem em uma região de *underfitting*, em que ambos  $R^2$  de treinamento e teste são baixos. Percebe-se que o  $R^2$  para de aumentar a partir do número de estimadores igual a 30. Dessa forma, o valor de  $n\_Estimators$  que estabelece um compromisso entre complexidade computacional e performance em termos de  $R^2$  é  $n\_Estimators = 30$ .

Fazendo uma análise semelhante para os gráficos da figura 3, conclui-se que um valor razoável para o parâmetro  $max\_depth$  é 30.

#### E. Determinação do Regressor Utilizado para o Modelo AC

Para determinação do regressor que será utilizado para fazer a estimação do nível de abstenção AC foram elencados os mesmos regressores propostos para o modelo DC:

- 1) **Florestas Randômicas para Regressão**
- 2) **K-Vizinhos mais Próximos para Regressão**
- 3) **Regressão de Menor Ângulo**
- 4) **Perceptron Multicamadas para Regressão**
- 5) **Máquina de Vetor de Suporte para Regressão Linear.**
- 6) **Máquina de Vetor de Suporte para Regressão Não-Linear**

Para escolher qual dos regressores aproxima melhor a função objetivo, foram feitos diversos testes e utilizada como métrica de qualidade do regressor obtido o coeficiente de determinação  $R^2$ . Os resultados obtidos estão dispostos na tabela III.

Tabela III: Comparação entre os regressores

Algoritmo	$R^2$
Floresta Randômica	-0.041
K-Vizinhos	0.0251
Regressão de Menor Ângulo	0.0390
Perceptron Multicamadas	0.004
Máquina de Vetor de Suporte Linear	-0.001
Máquina de Vetor de Suporte Não-Linear	-0.001

Analisando-se os resultados obtidos, percebe-se que nenhum dos regressores obteve performance aceitável em termos de  $R^2$ . Uma provável causa para isso é que as *features* escolhidas não têm peso determinante para a estimação da abstenção AC. Esse resultado reforça a hipótese de que o nível de abstenção é determinado primordialmente por fatores socioeconômicos. Como as *features* escolhidas para esse caso são de cunho político, os regressores tiveram performance ruim. Por esses motivos, o regressor AC não foi mais levado em consideração.

#### IV. RESULTADOS

Utilizando-se as Florestas Randômicas com os parâmetros determinados na seção III-D, atingiu-se o valor de  $R^2 = 0.47$  para a abstinência DC.

Uma vez que, embora o valor de  $R^2$  encontrado supere consideravelmente os valores de  $R^2$  obtidos em problemas similares na literatura [4], ainda há uma grande parcela da variância da amostra não explicada, é prudente estabelecer uma métrica auxiliar antes de validar o modelo gerado. Dessa forma, além do coeficiente de determinação, foram avaliados, também, o erro absoluto e o erro quadrático médio do modelo, apresentados na tabela IV.

Ao incluirmos as métricas auxiliares na análise, podemos verificar que, embora o valor do coeficiente  $R^2$  esteja consideravelmente distante de 1, tanto o erro médio em valores absolutos quanto o erro quadrático médio são pequenos, indicando que o modelo aproxima muito bem a abstenção real.

Considerando que a variação de  $R^2$  no conjunto de testes e no conjunto de treinamento estão representadas em um mesmo gráfico (figuras 2 e 3), é possível verificar, por inspeção visual, que o modelo não está em uma região

Tabela IV: Diferentes Métricas de Performance para o Modelo gerado por Floresta Randômica

Métrica	Valor
$R^2$	0.4783
MAE	0.0132
MSE	0.0002

de *overfitting*, dispensando, assim, o cálculo dos critérios de informação de Akaike e Schwartz.

Dessa forma, podemos perceber que o modelo escolhido é válido e tem bom poder de explicação sobre os dados, embora boa parte de sua variância não possa ser explicada pelas *features* escolhidas. Dentre as hipóteses possíveis para explicar esse fenômeno, convém destacar a possibilidade dos dados levantados estarem contaminados por alguma forma de ruído, de que características de outras naturezas (não-socioeconômicas) não avaliadas também sejam preponderantes na decisão do eleitor de se abster e na possibilidade de que a escala municipal ainda esteja consideravelmente distante da escala onde a decisão de abstenção realmente é identificável (escala de zona eleitoral).

Para avaliar a qualidade do regressor, foram comparadas a abstenção estimada e real para alguns municípios. Estes dados estão dispostos na tabela V.

Avaliando-se os dados na tabela V, percebe-se um ponto interessante: o regressor, em geral, tem mais facilidade estimar corretamente o nível de abstenção para municípios menores. Esse é um resultado que faz sentido, pois, era esperado que em regiões mais pobres e com menores graus de educação, os atributos escolhidos (que levam em conta aspectos socioeconômicos) seriam mais relevantes para a estimação. Outro fator que justifica esse resultado é que os dados obtidos em municípios grandes são mais propícios a ruídos, devido à quantidade de pessoas, e são naturalmente mais complexos de avaliar que um município pequeno.

Analisando a árvore obtida no final do processo de treinamento, os 5 atributos mais relevantes foram:

- 1) Percentual de Negros;
- 2) Unidades domésticas por tipo - nuclear - homem com filho(s);
- 3) Percentual de pessoas que recebem entre meio e um salário mínimo;
- 4) Percentual Beneficiários;
- 5) Estabelecimentos de Saúde SUS.

A identificação dos 5 atributos mais relevantes para estimar o nível de abstenção por município traz informações muito importantes acerca deste tema. O fato de o atributo relacionado ao percentual de negros no município ser o mais relevante causa um impacto social muito forte. Foi observado, em geral, que municípios com maior percentual de negros tendem a ter um número maior de abstenções. Um motivo que justifica esse resultado é que a grande maioria da população negra no Brasil tem condições financeiras e educacionais precárias. Dessa forma, como em geral pessoas com menor nível de instrução tendem a dar menos valor ao voto, o fato de (infelizmente) a majoritariedade da população negra não ter um bom nível de instrução, justifica o resultado obtido. Outra justificativa para esse resultado pode ser o fato de que a grande maioria dos candidatos a cargos políticos não são negros. Dessa forma, a população negra pode-se sentir não representada e, dessa forma, desmotivada a votar.

Tabela V: Comparação entre abstenção real e predita para alguns municípios.

Município	Estado	Abstenção Real (%)	Abstenção Predita (%)
Cabixi	RO	21.06	21.24
Boa Vista	RO	14.91	20.70
Boa Vista dos Ramos	AM	19.39	20.30
Manaus	AM	15.72	18.26
Acrelândia	AC	20.19	20.23
Rio Branco	AC	16.32	21.36
Barra de Santana	PB	16.55	18.62
João Pessoa	PB	12.65	20.78
Aroazes	PI	16.74	16.66
Teresina	PI	13.99	17.62

O segundo atributo mais relevante, unidades domésticas por tipo - nuclear - homem com filho(s), reflete o perfil e a estrutura familiar dos municípios. O terceiro atributo mais relevante, percentual de pessoas que recebem entre meio e um salário mínimo, confirma mais uma vez a hipótese de que municípios mais pobres tendem a ter níveis de abstenção maiores. Por outro lado, o terceiro atributo mais relevante, percentual beneficiários, mostra que a quantidade de beneficiários em um município tem impacto forte no nível de abstenção. Isso faz sentido, pois pessoas que recebem benefícios que foram propostos por um determinado partido ou político tendem a ser mais fiéis em termos de votação e sempre votar naquele que lhe traz o benefício. Por exemplo, beneficiários do Bolsa Família tendem a votar em candidatos do PT e não se abster.

Por fim, o número de estabelecimentos de saúde SUS foi o quinto atributo mais relevante. Os resultados indicaram que municípios com maior número de estabelecimentos de saúde SUS tendem a ter menores níveis de abstenção. Provavelmente, esse atributo reflete de alguma forma a esperança do eleitor em ver mudanças a partir do seu voto. Como vivem em condições precárias de saúde, não acreditam que novos políticos sejam capazes de mudar a situação e sem absterem de votar.

## V. CONCLUSÕES E RECOMENDAÇÕES

Analisando os resultados obtidos foi possível concluir que o nível de abstenção pode ser predito, com uma determinada tolerância, a nível municipal, utilizando dados com características socioeconômicas do censo. Vale ressaltar, que o modelo DC utilizado tenta prever a abstenção média no período de validade do censo, proporcionando informações úteis a respeito dos principais fatores responsáveis pelo nível de abstenção nos municípios.

A utilização de Florestas Randômicas, a possibilidade de obter uma árvore final e analisar os atributos mais significativos escolhidos pelo regressor traz informações de extrema relevância para a determinação de políticas públicas direcionadas por município para combater a abstenção de votos e incentivar de formas mais direcionadas, objetivas e efetivas os eleitores a exercerem seu direito de votar.

O fato de se obter, em geral, predições mais acuradas para municípios menores e menos acuradas para mu-

nicipios maiores sugere que as *features* escolhidas não são suficientes para determinar o nível de abstenção de municípios grandes de forma mais confiável. Dessa forma, conclui-se duas coisas: a primeira é que os resultados obtidos neste trabalho são mais confiáveis para municípios pequenos; a segunda é que a elaboração de *features* mais complexas, que envolvam elementos temporais, econômicos, elementos de contexto mundial, tendências do momento deve acarretar em estimativas mais precisas para os municípios grandes.

## REFERÊNCIAS

- [1] Base de Dados do IPEAGeo, disponível em <http://www.ipea.gov.br/ipeageo/bases.html>
- [2] Estatísticas Eleitorais do TSE, disponível em <http://www.tse.jus.br/eleicoes/estatisticas/eleicoes/eleicoes-antecedentes/estatisticas-eleitorais-anos-antecedentes>
- [3] Alexei V. Zakharov, *A model of electoral competition with abstaining voters*, Mathematical and Computer Modelling, Volume 48, Issues 9–10, 2008, Pages 1527–1553
- [4] Wagner Antonio Kamakura, *Using Voter-choice Modeling to Plan Final Campaigns in Runoff Elections*, Revista de Administração Contemporânea
- [5] Gabriel Katz, Ines Levin *A General Model of Abstention Under Compulsory Voting*, Political Science Research and Methods Vol 6, No. 3, 489–508
- [6] Timothy J. Power, J. Timmons Robert, *Compulsory Voting, Invalid Ballots, and Abstention in Brazil*, Political Research Quarterly, Vol. 48, No. 3 (December 1995), pp. 795–82
- [7] <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.Random>