

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

ALINE MACOHIN

**MODELO DE TOMADA DE DECISÃO APLICADO A CÂMARAS
COLEGIADAS**

DISSERTAÇÃO

CURITIBA

2013

ALINE MACOHIN

**MODELO DE TOMADA DE DECISÃO APLICADO A CÂMARAS
COLEGIADAS**

Dissertação apresentada ao Programa de Pós-graduação em Computação Aplicada da Universidade Tecnológica Federal do Paraná como requisito parcial para obtenção do grau de “Mestre em Computação” – Área de Concentração: Engenharia de Sistemas Computacionais.

Orientador: Gustavo Alberto Giménez Lugo

Co-orientador: César Augusto Tacla

CURITIBA

2013

Dados Internacionais de Catalogação na Publicação

M171 Macohin, Aline

Modelo de Tomada de Decisão Aplicado a Câmaras Colegiadas/ Aline Macohin. – 2013.
112 f. : il. ; 30 cm

Orientador: Gustavo Alberto Giménez Lugo.

Coorientador: César Augusto Tacla.

Dissertação (Mestrado) – Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Computação Aplicada, Curitiba, 2013.

Bibliografia: f. 104-107.

1. Órgãos colegiados – Processo decisório. 2. Votação. 3. Análise de regressão. 4. Árvores de decisão. 5. Teoria da sensibilidade (Matemática). 6. Métodos de simulação. 7. Brasil. Superior Tribunal de Justiça – Jurisprudência – Estudo de casos. 8. Computação – Dissertações. I. Lugo, Gustavo Alberto Giménez, orient. II. Tacla, César Augusto, coorient. III. Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Computação Aplicada. IV. Título.

CDD (22. ed.) 004

Biblioteca Central da UTFPR, Campus Curitiba

ATA DA DEFESA DE DISSERTAÇÃO DE MESTRADO 10

DISSERTAÇÃO PARA OBTENÇÃO DO GRAU DE MESTRE EM COMPUTAÇÃO APLICADA

No dia 26 de agosto de 2013, às 14:00 horas, reuniu-se na Sala C-301 - bloco C - 3º andar do Câmpus Curitiba, a banca examinadora composta pelos professores doutores:

Prof. Gustavo Alberto Giménez Lugo, Dr. (Presidente)	UTFPR - CT
Prof. Andre Luís Vignatti, Dr.	UFPR
Prof. Celso Antônio Alves Kaestner, Dr.	UTFPR - CT
Prof. Etéocles da Silva Cavalcanti	UTFPR - CT

sob Presidência de **Gustavo Alberto Giménez Lugo** para examinar a dissertação do candidato **ALINE MACOHIN**, intitulada: “Modelo Computacional de Tomada de Decisão Aplicado ao Domínio Legal”. Após a apresentação, o candidato foi arguido pelos examinadores e foi dada a palavra aos presentes para formularem perguntas ao candidato. Os examinadores reunidos deliberaram pela _____ da dissertação.

O candidato foi informado que a concessão do referido grau, na área de concentração Computação Aplicada, está condicionada à (i) satisfação dos requisitos solicitados pela Banca Examinadora e lavrados na documentação entregue ao Candidato; (ii) entrega da dissertação em conformidade com as normas exigidas pela UTFPR; e (iii) entrega da documentação necessária para elaboração do Diploma. A Banca Examinadora determina um **prazo de _____ dias** para o cumprimento dos requisitos (desconsiderar esse parágrafo caso a dissertação seja reprovada).

Nada mais havendo a tratar, a sessão foi encerrada às _____, dela sendo lavrada a presente ata que segue assinada pela Banca Examinadora e pelo Candidato.

Prof. **Gustavo Alberto Giménez Lugo, Dr.**
presidente - (UTFPR - CT)

Prof. **Andre Luís Vignatti, Dr.**
(UFPR)

Prof. **Celso Antônio Alves Kaestner, Dr.**
(UTFPR - CT)

Prof. **Etéocles da Silva Cavalcanti**
(UTFPR - CT)

Candidato: _____

DECLARAÇÃO PARA A OBTENÇÃO DO GRAU DE MESTRE

A coordenação do Programa declara que foram cumpridos todos os requisitos exigidos pelo Programa de Pós-Graduação para a obtenção do grau de mestre.

Curitiba, ____ de _____ de 20 ____.

RESUMO

MACOHIN, Aline. MODELO DE TOMADA DE DECISÃO APLICADO A CÂMARAS COLEGIADAS. 112 f. Dissertação – Programa de Pós-graduação em Computação Aplicada, Universidade Tecnológica Federal do Paraná. Curitiba, 2013.

Este trabalho se refere ao desenvolvimento de um modelo de tomada de decisão de câmaras colegiadas aplicando os métodos de regressão logística binária e o de árvore de decisão, quanto ao julgamento de assuntos em colegiado. O modelo é aplicável a nível de voto individual, gerando informações sobre como diferentes influências (características dos fatos, colegas de julgamento e do próprio julgador), podem afetar os votos no âmbito das câmaras colegiadas. Como estudo de caso do modelo foram utilizados dados de jurisprudências de um tribunal brasileiro, entre o período de 29/09/1997 e 17/11/2011 e se obteve que as variáveis Colegiado, Presidente e Tipo de Réu apresentaram uma maior influência nos dados utilizados, em relação às outras variáveis. O modelo foi aplicado em análises de sensibilidade e simulações utilizando o ambiente Netlogo para o teste de diversos cenários. A porcentagem de acertos que se obteve do modelo de árvore de decisão na amostra de teste foi de 81,33%, e para o modelo de regressão a porcentagem de acertos foi de 65,4%.

Palavras-chave: Modelo de Tomada de Decisão, Câmaras Colegiadas, Regressão Logística, Árvore de Decisão, Análise de Sensibilidade

ABSTRACT

MACOHIN, Aline. DECISION MAKING MODEL APPLIED TO THE CHAMBERS COLLEGIATE. 112 f. Dissertação – Programa de Pós-graduação em Computação Aplicada, Universidade Tecnológica Federal do Paraná. Curitiba, 2013.

This work is about the development of a computational model for the decision making of collegiate chambers using the methods of binary logistical regression and decision tree, as to the trial of a fact in collegiate. The model is applicable at the individual vote level, creating data about how different influences (facts characteristics, trial colleagues and from the judge himself) may affect voting in collegiate chambers. Jurisprudence data from a brazilian tribunal, between 29/09/1997 and 17/11/2011, was used as a case study and the test resulted in the variables Collegiate, President and Type of Defendant showing a bigger influence on the data used for the experiment. The model was applied in sensitivity analysis and simulations using the Netlogo environment to test several scenarios. The percentage of correct guesses that was obtained using the decision tree model in the sample was of 81,33% and for the regression model of 65,4%.

Keywords: Decision Making Model, Chambers Collegiate, Logistic Regression, Decision Tree, Sensitivity Analysis

LISTA DE FIGURAS

FIGURA 1	– Classificação das variáveis.	15
FIGURA 2	– Tipos de correlações.	17
FIGURA 3	– Árvore representada por regras de classificação e fluxograma.	21
FIGURA 4	– Procedimento para classificação dos dados.	22
FIGURA 5	– Mapa espacial dos senadores no 90º Congresso Estadunidense.	26
FIGURA 6	– Passos para o cálculo da ideologia quando se possui somente uma dimensão.	27
FIGURA 7	– Função de utilidade quadrática e normal. As funções são unimodais e simétricas, e estão centradas no ponto ideal do legislador. A utilidade diminui com o aumento da distância dos pontos do espaço político comparado ao ponto ideal.	30
FIGURA 8	– Passos para a criação do modelo e simulação.	40
FIGURA 9	– Fases do experimento da simulação.	47
FIGURA 10	– Fluxograma para elaboração da análise de sensibilidade dos dados.	49
FIGURA 11	– Tela do NetLogo com o modelo “ <i>Preferential Attachment</i> ”.	50
FIGURA 12	– Estrutura do Poder Judiciário Brasileiro.	54
FIGURA 13	– Espelho de um Acórdão do STJ.	55
FIGURA 14	– Número de jurisprudências extraídas por data de julgamento sobre Direito Tributário comparado ao total de jurisprudências publicadas pelo STJ. ...	58
FIGURA 15	– Currículo de uma ministra disponível no <i>site</i> do tribunal.	61
FIGURA 16	– Diagrama entidade relacionamento que representa o contexto do trabalho.	71
FIGURA 17	– Diagrama entidade relacionamento que representa as possíveis influências entre as variáveis.	73
FIGURA 18	– Comparação da porcentagem de erros para cada tipo de voto dos modelos utilizados.	79
FIGURA 19	– Comparação da porcentagem de erros no modelo de árvore de decisão utilizando as respostas da amostra de teste.	82
FIGURA 20	– Comparação entre a porcentagem de acertos dos votos positivos e negativos de cada modelo.	82
FIGURA 21	– Fluxograma do comportamento da simulação.	84
FIGURA 22	– Simulação desenvolvida para testar os modelos sugeridos.	85
FIGURA 23	– Gráficos dos experimentos realizados.	92
FIGURA 24	– Teste da variável Colegiado.	94
FIGURA 25	– Teste da variável Tipo Réu.	95
FIGURA 26	– Teste da variável Tipo Autor.	96
FIGURA 27	– Teste da variável Estado do Processo.	96
FIGURA 28	– Teste da variável Faculdade.	97
FIGURA 29	– Teste da variável Ideologia.	98
FIGURA 30	– Teste da variável Indicador Ideologia.	98
FIGURA 31	– Teste da variável Presidente.	99
FIGURA 32	– Teste da variável Voto do Relator.	100
FIGURA 33	– Teste da variável Presidente x Tipo Autor.	100

FIGURA 34	–	Teste da variável Presidente x Tipo Réu.	101
FIGURA 35	–	Variáveis categóricas do modelo de regressão.	108
FIGURA 36	–	Variáveis categóricas do modelo de regressão.	109
FIGURA 37	–	Variáveis do modelo de regressão e seus respectivos coeficientes (B).	110
FIGURA 38	–	Variáveis do modelo de regressão e seus respectivos coeficientes (B).	111
FIGURA 39	–	Variáveis do modelo de regressão e seus respectivos coeficientes (B).	112

LISTA DE TABELAS

TABELA 1	– Valores de referência para o <i>Kappa statistic</i>	45
TABELA 2	– Número de ocorrências para cada assunto envolvendo direito tributário. . .	58
TABELA 3	– Número de ocorrências para cada tipo de resultado encontrado nas jurisprudências extraídas.	59
TABELA 4	– Número de ocorrências para cada tipo de decisão final encontrado nas jurisprudências sobre direito tributário. A decisão se refere ao recurso/processo enviado para o Superior Tribunal de Justiça, em que o colegiado decide se vai aceitá-lo e dar provimento. Foram separadas as quantidades de decisões por processos em que ocorreu unanimidade e processos em que ocorreu votação por maioria.	60
TABELA 5	– Significados dos valores utilizados no algoritmo do <i>W-NOMINATE</i> . Exemplo utilizando dados de senadores dos Estados Unidos da América. .	62
TABELA 6	– Significados dos valores utilizados no algoritmo do <i>W-NOMINATE</i> . Exemplo utilizando dados de juízes.	63
TABELA 7	– Valor da ideologia dos membros de um colegiado, obtido através de casos de direito tributário com votos por maioria.	64
TABELA 8	– Variáveis do modelo. As palavras “membro” representam os membros que podem fazer parte de um colegiado.	72
TABELA 9	– Tabela comparativa das variáveis utilizadas no modelo de Cameron e Cummings e no modelo sugerido neste trabalho.	74
TABELA 10	– Valores da codificação da variável categórica “Presidente”.	75
TABELA 11	– Resultados obtidos na aplicação dos dados de treinamento no modelo de regressão, e na aplicação do modelo obtido na amostra de teste. Neste teste foram utilizadas todas as variáveis.	76
TABELA 12	– Resultados obtidos na aplicação dos dados de treinamento no modelo de regressão, e na aplicação do modelo obtido na amostra de teste. Neste teste foram utilizadas as variáveis do modelo de Cameron e Cummings.	77
TABELA 13	– Resultados obtidos com os dados de treinamento no modelo de regressão, e na utilização do modelo na amostra de teste. Foram utilizadas as variáveis Colegiado, Tipo de Réu e Presidente.	78
TABELA 14	– Resultados obtidos na aplicação dos dados de treinamento no modelo de árvore, e na aplicação das regras da árvore de decisão obtidas na amostra de teste. Neste teste foram utilizadas todas as variáveis.	80
TABELA 15	– Resultados obtidos com os dados de treinamento no modelo de árvore, e na utilização do modelo na amostra de teste. Foram utilizadas as variáveis Colegiado, Tipo de Réu e Presidente.	81
TABELA 16	– Resultados obtidos com os dados de treinamento no modelo de árvore, e na utilização do modelo na amostra de teste. Foram utilizadas as variáveis do modelo de Cameron e Cummings.	81
TABELA 17	– Parâmetros da simulação.	87
TABELA 18	– Definição das variáveis independentes e de controle da simulação.	89

TABELA 19 – Definição das variáveis de controle e dependentes da simulação.	90
TABELA 20 – Valores dos parâmetros de entrada da simulação no ambiente <i>NetLogo</i> . Para as simulações foi utilizado o modelo de regressão logística binária. ..	91
TABELA 21 – Resultados obtidos com os experimentos realizados no ambiente <i>NetLogo</i> por tipo de coalizão de voto.	91

SUMÁRIO

1 INTRODUÇÃO	11
1.1 PROBLEMA	12
1.2 HIPÓTESE	12
1.3 OBJETIVOS	12
1.3.1 Objetivo Geral	12
1.3.2 Objetivos Específicos	12
1.4 JUSTIFICATIVA	13
1.5 ESTRUTURA DO TRABALHO	13
2 TOMADA DE DECISÃO EM CÂMARAS COLEGIADAS	14
2.1 ANÁLISE ESTATÍSTICA	14
2.1.1 População e Amostra	14
2.1.2 Tipos de Variáveis	15
2.1.3 Correlação	16
2.1.4 Regressão	17
2.1.5 Modelos de Classificação	20
2.2 CÁLCULO DO VALOR DA IDEOLOGIA	25
2.3 MODELOS	32
2.3.1 Judiciais	32
2.3.2 Estatísticos	35
2.4 SIMULAÇÃO EM CÂMARAS COLEGIADAS	37
3 MÉTODO DE PESQUISA	39
3.1 AQUISIÇÃO DO CONJUNTO DE DADOS	39
3.2 PRÉ-PROCESSAMENTO DOS DADOS	41
3.3 CRIAÇÃO DOS MODELOS	42
3.3.1 Seleção da Amostra	42
3.3.2 Seleção das Variáveis	43
3.3.3 Avaliação dos Modelos	43
3.3.3.1 Regressão Logística	43
3.3.3.2 Árvore de Decisão	45
3.4 PROJEÇÃO DOS DADOS	46
3.5 CONSTRUÇÃO DA SIMULAÇÃO	46
3.5.1 Análise de Sensibilidade	48
3.5.2 Simulação utilizando o NetLogo	49
4 DADOS COLETADOS	51
4.1 ESTUDO DE CASO	51
4.1.1 Superior Tribunal de Justiça	53
4.1.2 Documentos produzidos	53
4.1.3 Membros do Colegiado	55
4.2 EXTRAÇÃO DE DOCUMENTOS	56
4.3 CLASSIFICAÇÃO DAS INFORMAÇÕES	58
4.3.1 Jurisprudências	59

4.3.2 Juízes	61
4.4 REPRESENTAÇÃO FORMAL DOS DADOS	63
4.4.1 Dados dos membros do colegiado	64
4.4.2 Características dos membros do colegiado	65
4.4.3 Características dos colegas de julgamento	67
4.4.4 Dados dos membros nos processos	68
4.4.5 Características do processo	69
5 MODELO SUGERIDO	71
5.1 ATRIBUTOS	71
5.2 REGRESSÃO LOGÍSTICA	75
5.3 ÁRVORE DE DECISÃO	79
5.4 SIMULAÇÃO	82
6 EXPERIMENTOS E ANÁLISE DOS RESULTADOS	86
6.1 VARIÁVEIS	86
6.2 RESULTADOS E TENDÊNCIAS	88
6.2.1 Simulação com o NetLogo	88
6.2.2 Análise de Sensibilidade	93
6.2.2.1 Colegiado	94
6.2.2.2 Tipo Réu	94
6.2.2.3 Tipo Autor	95
6.2.2.4 Estado do Processo	95
6.2.2.5 Faculdade	96
6.2.2.6 Ideologia	97
6.2.2.7 Indicador Ideologia	97
6.2.2.8 Presidente	98
6.2.2.9 Voto do Relator	99
6.2.2.10 Presidente x Tipo Autor	99
6.2.2.11 Presidente x Tipo Réu	101
7 CONCLUSÃO E TRABALHOS FUTUROS	102
REFERÊNCIAS	104
Apêndice A – CONVERSÃO DAS VARIÁVEIS CATEGÓRICAS	108
Apêndice B – COEFICIENTES DO MODELO DE REGRESSÃO LOGÍSTICA	110

1 INTRODUÇÃO

A problemática de avaliar quais fatores influenciam na tomada de decisão de membros de um colegiado é foco do estudo de diversos especialistas de algumas áreas, como a judicial (HWONG, 2006) e a política (MATTILA, 2003). Na área judicial, Schubert (1965) levanta a seguinte pergunta “Até que ponto os atos públicos dos juízes são influenciados por suas crenças pessoais?”.

As câmaras colegiadas, também denominadas de órgãos colegiados, conselhos, comitês, entre outros, dizem respeito a uma forma de gestão a qual a direção é compartilhada por um conjunto de pessoas com igual autoridade, que reunidas, decidem sobre um tema em pauta.

As câmaras colegiadas são compostas por grupos representantes de diversos segmentos, como por exemplo no âmbito escolar, em que um colegiado é composto pela comunidade escolar, pais ou responsável legal, professores, funcionários, coordenadores pedagógicos, alunos, dirigentes, entre outros. Já no âmbito judicial, um colegiado pode ser composto por juízes com diferentes áreas de conhecimento.

Dentre as diversas câmaras colegiadas existentes, há as presentes no Congresso Nacional, conforme definição retirada do *site* da Câmara Legislativa:

“O Congresso Nacional é composto de duas Casas: Câmara dos Deputados e Senado Federal. Cada uma dessas Casas possui Comissões Parlamentares, Permanentes ou Temporárias, com funções legislativas e fiscalizadoras, na forma definida na Constituição Federal e nos seus Regimentos Internos. No cumprimento dessas duas funções básicas, de elaboração das leis e de acompanhamento das ações administrativas, no âmbito do Poder Executivo, as Comissões promovem, também, debates e discussões com a participação da sociedade em geral, sobre todos os temas ou assuntos de seu interesse.” (Disponível em: <http://www2.camara.leg.br/atividade-legislativa/comissoes>. Acesso em 15/12/2013.)

Outros exemplos de câmaras colegiadas incluem o Conselho Nacional de Turismo, Conselho Tutelar do Ministério Público, Conselhos de Administração, e, Turmas, Seções e Cortes dos tribunais brasileiros que julgam processos a partir de um grupo de juízes.

1.1 PROBLEMA

O problema abordado neste trabalho é a complexidade inerente à projeção de votos de membros de um colegiado, decorrente do grande número de variáveis existentes e que podem influenciar os votos. Há também dificuldade de se encontrar um modelo genérico adaptável aos vários assuntos tratados pelos colegiados, pelo fato de cada país e tema da pauta do julgamento terem suas respectivas peculiaridades. Além disso, a utilização de modelos estatísticos para projeção de votos, pode não representar adequadamente a dinâmica presente nos processos.

1.2 HIPÓTESE

Com base na questão em estudo dessa dissertação, foi formulada a seguinte hipótese: A inserção de variáveis de interações, entre outras não utilizadas nos modelos estudados, permite projetar com maior precisão os votos ?

1.3 OBJETIVOS

A seguir, serão detalhados o objetivo geral e os objetivos específicos.

1.3.1 OBJETIVO GERAL

Este trabalho tem por objetivo contribuir para a construção de modelos de tomada de decisão com a adição de novas variáveis, como por exemplo variáveis de interação, a fim de verificar se melhoram a projeção dos votos.

1.3.2 OBJETIVOS ESPECÍFICOS

O objetivo geral pode ser decomposto nos seguintes itens:

- Avaliar se as novas variáveis contribuem para o modelo de tomada de decisão;
- Avaliar o desempenho dos modelos propostos;
- Avaliar o comportamento do modelo adaptado em cenários específicos.

1.4 JUSTIFICATIVA

Uma das vantagens ao se criar modelos é a possibilidade de descrever uma grande quantidade de relacionamentos com um certo grau de precisão, além de ser relevante em pesquisas multidisciplinares (LEEuw, 2004).

Dentre os modelos de tomada de decisão já existentes, é necessária a inclusão de novas características para verificar se estas podem contribuir para a melhoria na projeção dos votos e/ou ainda apresentar uma maior influência que outras variáveis na projeção. Ainda, com a adição de variáveis como as de interações, não permitidas por modelos estatísticos, permite trabalhar de forma dinâmica.

Com a aplicação de métodos computacionais para simular características de diferentes membros do colegiado que podem participar dos julgamentos, permite obter tendências de votos a partir de um conjunto de características.

1.5 ESTRUTURA DO TRABALHO

No capítulo 2 é apresentada a revisão bibliográfica, abordando os modelos legais, computacionais e estatísticos já existentes, além dos métodos para o cálculo do valor da ideologia e trabalhos envolvendo simulações em câmaras colegiadas.

No capítulo 3 é introduzida a metodologia aplicada no desenvolvimento da extração e pré-processamento dos dados de documentos, criação dos modelos, projeção dos dados e construção da simulação.

No capítulo 4 são detalhados os dados obtidos nos documentos de julgamentos, para melhor compreensão do modelo. Além disso, são citadas também algumas estatísticas dos dados.

No capítulo 5 são apresentados os modelos sugeridos neste trabalho. São descritos também os atributos utilizados e seus respectivos valores de influência sobre a variável a ser projetada. Além disso, é abordado como foi construída a simulação que aplica os modelos.

No capítulo 6 são apresentados os experimentos e a análise dos resultados obtidos, o que inclui as simulações realizadas e a análise de sensibilidade.

Por fim, o capítulo 7 apresenta a conclusão sobre o trabalho realizado e as perspectivas de trabalhos futuros.

2 TOMADA DE DECISÃO EM CÂMARAS COLEGIADAS

O uso de modelos teóricos, estatísticos ou computacionais permite uma melhor compreensão sobre como as decisões dos membros de um colegiado são tomadas. Com base nisso, foram levantados diferentes estudos sobre a tomada de decisão, a fim de verificar os tipos de variáveis que podem ser utilizadas e os diferentes métodos disponíveis.

Na seção 2.1 será dada uma breve introdução a alguns conceitos e modelos estatísticos utilizados no decorrer do trabalho. Já na seção 2.2 será abordado o cálculo do valor da ideologia, pelo fato de alguns modelos utilizarem este valor em seus cálculos. Na seção 2.3 serão abordados os modelos de tomada de decisão. Por fim, na seção 2.4 é abordado o uso de simulação em tomadas de decisões de câmaras colegiadas.

2.1 ANÁLISE ESTATÍSTICA

A seguir, serão abordados alguns temas associados à estatística, com o intuito de esclarecer alguns termos citados na análise dos dados.

2.1.1 POPULAÇÃO E AMOSTRA

A população representa o universo estatístico, e, segundo Morettin e Bussab (2004), é o conjunto de todos os elementos ou resultados sob investigação. Já a amostra é qualquer subconjunto da população.

Para calcular a amostra é necessário ter conhecimento sobre a população, uma vez que as observações contidas em uma amostra são tanto mais informativas sobre a população quanto mais conhecimento explícito ou implícito tivermos dessa mesma população. Um exemplo citado por Morettin e Bussab (2004) é o da análise da quantidade de glóbulos brancos, obtida de algumas gotas de sangue da ponta do dedo de um paciente, que dará uma ideia geral da quantidade dos glóbulos brancos no corpo todo, pois sabe-se que a distribuição dos glóbulos brancos é homogênea, e de qualquer lugar que se tivesse retirado a amostra ela seria

representativa.

A amostra possui diferentes tipos de técnicas, tais como a probabilística e a não probabilística. A probabilística ocorre quando todos os elementos de uma população possuem uma probabilidade conhecida, e maior que zero, de pertencer a uma amostra, caso contrário será considerada não probabilística.

As técnicas do tipo probabilística incluem 1) amostragem aleatória simples, 2) amostragem sistemática, 3) amostragem estratificada e 4) amostragem por conglomerados. A amostragem aleatória simples é aquela em que todos os elementos possuem probabilidades iguais de pertencer à amostra. A amostragem sistemática consiste em retirar os elementos a cada intervalo de tempo/evento, por exemplo. A amostragem estratificada é mais indicada quando a população divide-se em subpopulações ou estratos, sendo recomendado que cada grupo tenha um número de indivíduos presentes na amostra. Por outro lado, a amostragem por conglomerados define que em vez de se analisar os elementos individualmente, devem ser analisados os conjuntos destes elementos.

2.1.2 TIPOS DE VARIÁVEIS

Segundo Morettin e Bussab (2004), as variáveis podem ser classificadas em qualitativas e quantitativas, que se subdividem em mais dois tipos cada uma, conforme figura 1.

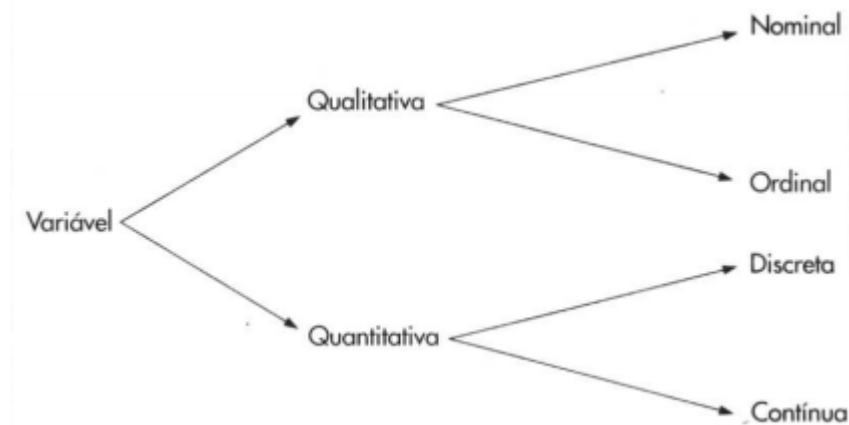


Figura 1: Classificação das variáveis.

Fonte: Morettin e Bussab (2004).

Os conceitos destes tipos de variáveis são definidos por Morettin e Bussab (2004) como:

“Dentre as variáveis qualitativas, ainda podemos fazer uma distinção entre dois tipos: variável qualitativa nominal, para a qual não existe nenhuma ordenação nas possíveis realizações, e variável qualitativa ordinal, para a qual existe uma ordem nos seus resultados.

[...]

De modo análogo, as variáveis quantitativas podem sofrer uma classificação dicotômica: (a) variáveis quantitativas discretas, cujos possíveis valores formam um conjunto finito ou enumerável de números, e que resultam, frequentemente, de uma contagem, como por exemplo número de filhos (0,1,2,...); (b) variáveis quantitativas contínuas, cujos possíveis valores pertencem a um intervalo de números reais de uma mensuração, como por exemplo estatura e peso de um indivíduo.

[...]

Existe um tipo de variável qualitativa para a qual essa quantificação é muito útil: a chamada variável dicotômica. Para essa variável só podem ocorrer duas realizações, usualmente chamadas sucesso e fracasso.” (MORETTIN; BUSSAB, 2004)

2.1.3 CORRELAÇÃO

Segundo Larson e Farber (2010), uma correlação é uma relação entre duas variáveis, em que os dados são representados por pares ordenados (x, y) , onde x é a variável independente (ou explanatória) e y é a variável dependente (ou resposta).

O diagrama mais indicado para verificar se duas variáveis possuem uma correlação linear é o diagrama de dispersão. A variável independente x é medida pelo eixo horizontal, e a variável dependente y é medida pelo eixo vertical, conforme figura 2. A correlação linear ainda pode ser negativa ou positiva, e, através do gráfico é possível identificar se não há correlação ou se esta é não linear. Na correlação linear negativa conforme x aumenta, y tende a decrescer. Já na correlação linear positiva, conforme x aumenta, y tende a aumentar.

Para calcular o coeficiente de correlação, que é uma medida da força e direção de uma relação linear entre duas variáveis, utiliza-se a fórmula abaixo. O símbolo r representa o coeficiente de correlação amostral, e, n representa o número de pares de dados. Já x e y representam as variáveis anteriormente citadas, cuja correlação se deseja obter.

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}} \quad (1)$$

Segundo Larson e Farber (2010), a amplitude do coeficiente de correlação é -1 para 1. Se x e y têm uma correlação linear positiva forte, r está próximo de 1. Se não há correlação linear ou uma correlação linear fraca, r está próximo a 0. É importante ressaltar que, se r está próximo de 0, não significa que não há relação entre x e y , significa somente que não há relação linear.

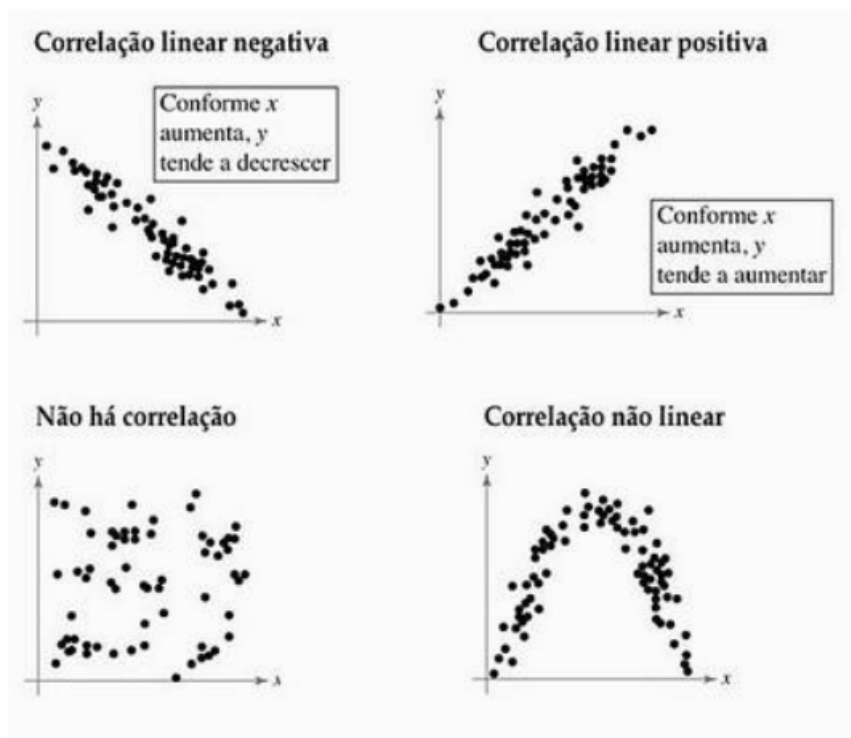


Figura 2: Tipos de correlações.

Fonte: Larson e Farber (2010).

2.1.4 REGRESSÃO

A regressão é um método estatístico utilizado para explorar e inferir relações entre variáveis. Segundo Levine et al. (2008), na regressão a variável que se deseja prever denomina-se variável dependente, já as variáveis utilizadas para realizar a previsão são chamadas de variáveis independentes. Levine et al. (2008) ainda citam:

“Além de prever valores para a variável dependente, a análise de regressão permite também que você identifique o tipo de relação matemática que existe entre uma variável dependente e uma variável independente, quantifique o efeito que mudanças na variável independente exercem sobre a variável dependente e identifique observações incomuns.” (LEVINE et al., 2008)

Dentre as diversas regressões existentes, estão a linear simples, a linear múltipla e as não lineares. Neste trabalho será abordado o modelo de regressão logística binária, um modelo não linear que foi projetado especificamente para variáveis dependentes binárias, como abordado neste trabalho pela variável voto.

Stock e Watson (2004) afirmam que em uma regressão com uma variável dependente binária, Y modela a probabilidade de que um determinado valor da variável dependente ocorra,

por isso, se adota uma formulação não linear que obrigue os valores previstos a se situarem entre zero e um, sendo utilizadas nas regressões *probit* e *logit*. A regressão *probit* utiliza a função de distribuição acumulada (f.d.a.) normal padrão. Já a regressão *logit* utiliza a f.d.a. logística.

O modelo *logit* da população, da variável dependente binária Y , com múltiplos regressores (X_1, \dots, X_k) é calculado da seguinte maneira:

$$P(Y = 1|X_1, X_2, \dots, X_k) = F(\beta_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k) = \frac{1}{1 + e^{-(\beta_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k)}} \quad (2)$$

Sendo:

B_0 = Constante do modelo;

B_1, \dots, B_k = Coeficientes de cada variável preditora.

Os coeficientes do modelo *logit* podem ser estimados por máxima verossimilhança. Stock e Watson (2004) ainda citam que o estimador é consistente e normalmente distribuído em amostras grandes, de modo que as estatísticas t e os intervalos de confiança para os coeficientes possam ser construídos de forma usual.

A definição do método de verossimilhança, é concebida por Stock e Watson (2004) conforme trecho abaixo:

“A função de verossimilhança é a distribuição de probabilidade conjunta dos dados, tratada como uma função dos coeficientes desconhecidos. O estimador de verossimilhança (E.M.V.) dos coeficientes desconhecidos consiste nos valores dos coeficientes que maximizam a função de verossimilhança. Como o E.M.V. escolhe os coeficientes desconhecidos para maximizar a função de verossimilhança, que por sua vez é a distribuição de probabilidade conjunta, temos que ele escolhe os valores dos parâmetros para maximizar a probabilidade de selecionar os dados efetivamente observados. Nesse sentido, os E.M.V.s são os valores dos parâmetros “mais prováveis” de terem produzidos os dados.

Para ilustrar a estimação de máxima verossimilhança, considere duas observações independentes e identicamente distribuídas (i.i.d.), Y_1 e Y_2 , de uma variável dependente binária sem nenhum regressor. Assim, Y é uma variável aleatória de Bernoulli e o único parâmetro desconhecido a ser estimado é a probabilidade p de que $Y = 1$, que é também a média de Y .

Para obter o estimador de máxima verossimilhança (E.M.V.), precisamos de uma expressão para a função de verossimilhança, que por sua vez requer uma expressão para a distribuição de probabilidade conjunta dos dados. A distribuição de probabilidade conjunta das duas observações, Y_1 e Y_2 é $P(Y_1 = y_1, Y_2 = y_2)$. Como Y_1 e Y_2 são independentemente distribuídas¹, a distribuição conjunta é o produto

¹ Duas variáveis aleatórias X e Y são independentemente distribuídas, ou independentes, se o conhecimento do valor de uma das variáveis não fornece nenhuma informação sobre a outra. Especificamente, X e Y são

das distribuições individuais, logo $P(Y_1 = y_1, Y_2 = y_2) = P(Y_1 = y_1)P(Y_2 = y_2)$. A distribuição de Bernoulli pode ser resumida na fórmula $P(Y = y) = p^y(1-p)^{1-y}$: quando $y = 1$, $P(Y = 1) = p^1(1-p)^0 = p$ e quando $y = 0$, $P(Y = 0) = p^0(1-p)^1 = 1-p$. Portanto, a distribuição de probabilidade conjunta de Y_1 e Y_2 é $P(Y_1 = y_1, Y_2 = y_2) = [p^{y_1}(1-p)^{1-y_1}] \times [p^{y_2}(1-p)^{1-y_2}] = p^{(y_1+y_2)}(1-p)^{2-(y_1+y_2)}$.” (STOCK; WATSON, 2004)

Logo, para $n = 2$ observações i.i.d. de variáveis aleatórias de Bernoulli, a função de verossimilhança é:

$$f(p; Y_1; Y_2) = p^{(Y_1+Y_2)}(1-p)^{2-(Y_1+Y_2)} \quad (3)$$

Sendo a variação do valor de p , utilizado para a maximização da função de verossimilhança até que obtenha um valor satisfatório de maximização da função. Stock e Watson (2004) ainda ressaltam que é conveniente maximizar não a verossimilhança, mas seu logaritmo, conforme equação a seguir:

$$S \ln(p) + (n - S) \ln(1 - p) = \frac{d}{dp} \ln[f_{bernoulli}(p; Y_1, \dots, Y_n)] = \frac{S}{p} - \frac{n - S}{1 - p} \quad (4)$$

Sendo:

$$S = \sum_{i=1}^n Y_i;$$

n = Número de observações.

No exemplo anterior foi citado o E.M.V. para a distribuição de Bernoulli, utilizado para verificar o valor da verossimilhança quando não há os preditores, entretanto, para os modelos *logit* incluindo os previsores, é necessário aplicar a seguinte fórmula:

$$\sum_{i=1}^n Y_i \ln \left[\frac{1}{1 + e^{-(\beta_0 + B_1 X_1 + B_2 X_2 + \dots + B_k X_k)}} \right] + \sum_{i=1}^n (1 - Y_i) \ln \left[1 - \left(\frac{1}{1 + e^{-(\beta_0 + B_1 X_1 + B_2 X_2 + \dots + B_k X_k)}} \right) \right] \quad (5)$$

Há também as medidas de ajustes, que são utilizadas na regressão logística, denominadas fração corretamente prevista e pseudo- R^2 . Segundo Stock e Watson (2004), a definição para ambas é:

independentes se a distribuição condicional de Y dado X é igual à distribuição marginal de Y . Isto é, X e Y são independentemente distribuídas se, para todos os valores de x e y $P(Y = y | X = x) = P(Y = y)$ (STOCK; WATSON, 2004).

“A fração corretamente prevista utiliza a seguinte regra: se $Y_i = 1$ e a probabilidade excede 50 por cento, ou se $Y_i = 0$ e a probabilidade prevista é menor do que 50 por cento, diz-se que Y_i é corretamente previsto. Caso contrário, diz-se que Y_i é incorretamente previsto.

[...]

O pseudo- R^2 mede o ajuste do modelo utilizando a função de verossimilhança. Como o E.M.V. maximiza essa função, a inclusão de outro regressor em um modelo *probit* ou *logit* aumenta o valor da verossimilhança maximizada, assim como a inclusão de um regressor necessariamente reduz a soma dos quadrados dos resíduos na regressão linear por M.Q.O. (Mínimos Quadrados Ordinários). Isso sugere a medição da qualidade do ajuste de um modelo *probit* pela comparação do valor da função de verossimilhança maximizada, incluindo todos os regressores com o valor da verossimilhança sem nenhum regressor.” (STOCK; WATSON, 2004)

Os regressores citados por Stock e Watson (2004) são as variáveis utilizadas para estimar a variável resposta, também denominada variável de saída. A fórmula para calcular o R^2 é representada na fórmula abaixo. Vale ressaltar que alguns *softwares* estatísticos utilizam variações do cálculo do valor do R^2 .

$$pseudo - R^2 = 1 - \frac{\ln(f_{logit}max)}{\ln(f_{bernoulli}max)} \quad (6)$$

Sendo:

$f_{bernoulli}max$ = Valor da verossimilhança que inclui os preditores no cálculo;

$f_{logit}max$ = Valor da verossimilhança que não inclui os preditores no cálculo.

2.1.5 MODELOS DE CLASSIFICAÇÃO

Segundo Han (2005), a classificação é o processo utilizado para encontrar um modelo ou função que descreva e distinga classes de dados, a fim de possibilitar a utilização do modelo para prever a classe de objetos cujo valor seja desconhecido². Os métodos de classificação são geralmente utilizados para prever variáveis categóricas.

Han (2005) cita também que os modelos são derivados da análise de um conjunto de dados cuja classe é conhecida, e este pode ser representado por diversas formas, como regras de classificação (SE - ENTÃO), árvores de decisão, fórmulas matemáticas ou ainda redes neurais. Pode-se citar ainda outros modelos de construção de modelos de classificação, como a classificação Bayesiana, máquina de vetores de suporte e o KNN (*K-Nearest Neighbor*).

Dentre os vários métodos de classificações existentes, foi escolhida a árvore de decisão para detalhamento, em vista de ser adequada para a proposta do trabalho, ser de fácil utilização

²Neste caso, a classe desconhecida é equivalente à variável dependente dos modelos de regressão.

e interpretação, além de permitir a descoberta exploratória de conhecimento a partir dos dados (HAN, 2005).

Uma árvore de decisão pode ser definida como um fluxograma com estrutura de árvore, em que cada nó representa testes sobre um valor de atributo ou valores de atributos, e cada ramo representa um resultado do teste. Já as folhas da árvore representam classes ou distribuições da classe, conforme figura 3.

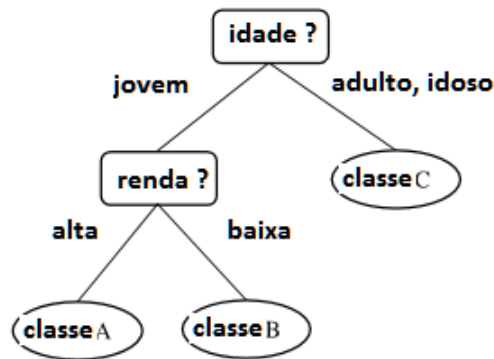
$$\begin{aligned} & idade(X, "jovem") \text{ E } renda(X, "alta") \rightarrow classe(X, "A") \\ & idade(X, "jovem") \text{ E } renda(X, "baixa") \rightarrow classe(X, "B") \\ & idade(X, "adulto") \rightarrow classe(X, "C") \\ & idade(X, "idoso") \rightarrow classe(X, "C") \end{aligned}$$


Figura 3: Árvore representada por regras de classificação e fluxograma.

Fonte: Han (2005). Tradução nossa.

Para chegar ao modelo de árvore de decisão é necessário realizar alguns procedimentos, conforme figura 4. O primeiro passo é selecionar uma amostra dos dados e a separar em tuplas, cuja classe seja conhecida, para aplicar um algoritmo de classificação pré-definido e obter um modelo que represente os dados da amostra.

A tupla pode ser representada por $X = (x_1, x_2, \dots, x_n)$, que representa as n características de um objeto (um item da amostra) que pertence a uma determinada classe já conhecida. Esta técnica é conhecida como aprendizado supervisionado, por se conhecer a classe a qual a tupla pertence.

Já o segundo passo é composto da obtenção da acurácia do modelo e a classificação dos novos dados. É necessário definir uma nova amostra, a amostra de teste, para testar a acurácia, devido ao modelo possuir tendências de estar altamente ajustado³ aos dados de treinamento.

³Durante o treinamento, o modelo pode incorporar algumas anomalias dos dados que não estão presentes nos

Após a aplicação das regras de classificação nos dados de teste (cuja classe também é conhecida), obtém-se a acurácia do modelo, que é a porcentagem das tuplas (dados) que foram classificadas corretamente. Caso o valor da acurácia seja satisfatório no estudo do caso proposto, pode-se aplicar as regras de classificação em dados cuja classe seja desconhecida e estabelecer assim valores de projeção.

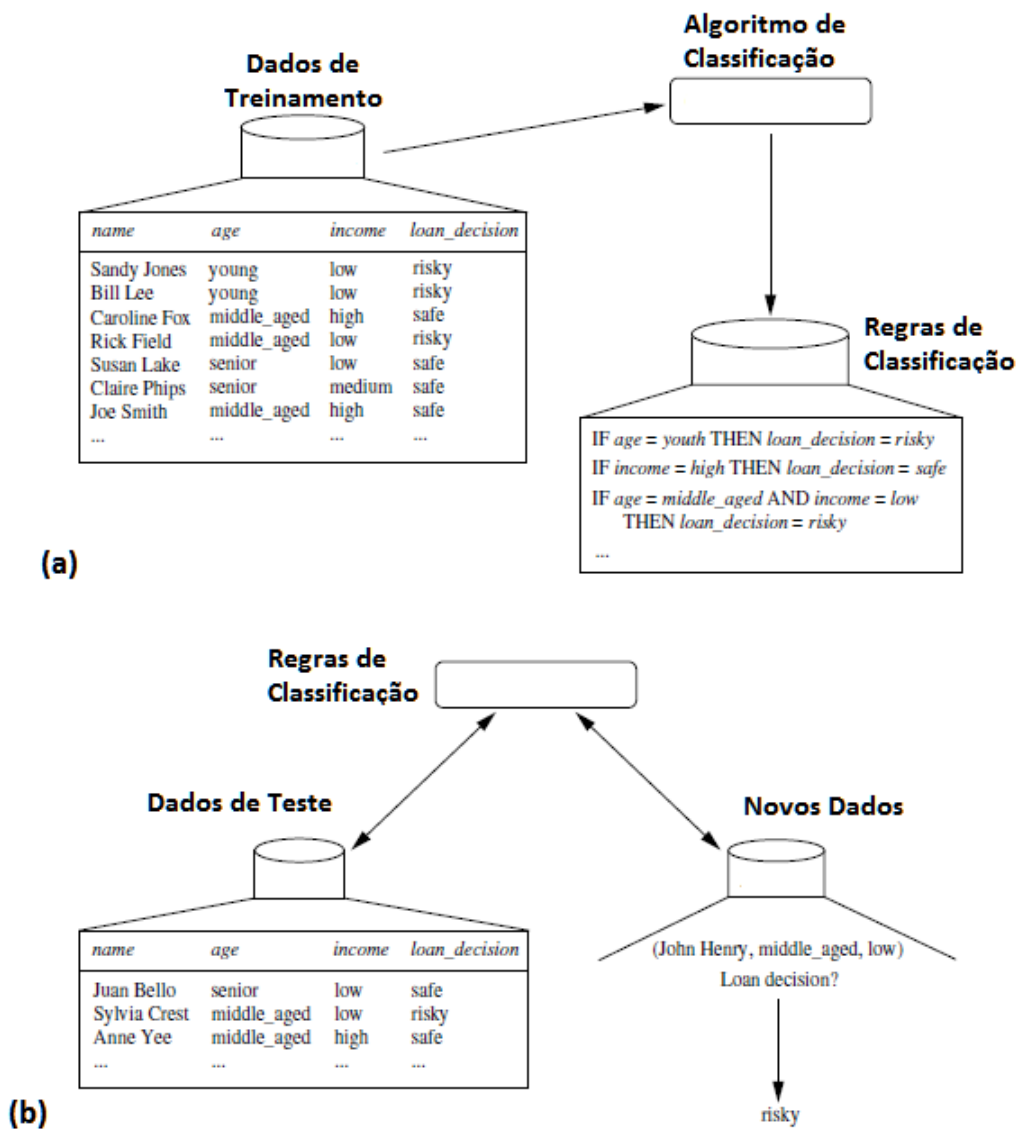


Figura 4: Procedimento para classificação dos dados.

Fonte: Han (2005). Tradução nossa.

Conforme figura 4, o item (a) representa o aprendizado, em que se aplica o algoritmo de classificação nos dados de treinamento e se obtêm as regras de classificação. Já o item (b) representa a classificação, que aplica os dados de teste nas regras de classificação para verificar dados em geral (HAN, 2005).

a acurácia do modelo. Caso a acurácia seja satisfatória, aplica-se nos novos dados que se deseja classificar.

Durante o final da década de 70 e início da década de 80, Quinlan (1986) desenvolveu o algoritmo de árvore de decisão denominado ID3 (*Iterative Dichotomizer 3*), que serviu de base para o desenvolvimento de novos algoritmos como o C4.5 de Quinlan (1993).

O algoritmo do ID3 pode ser verificado no pseudocódigo apresentado no algoritmo 1, conforme citado por Mitchell (1997), que consiste num processo de indução de árvore. A construção da árvore se dá de cima para baixo (*top-down*), com o objetivo de escolher o melhor atributo para cada nó da árvore de decisão através da utilização do método de ganho de informação⁴. A árvore utiliza métodos recursivos após a escolha do atributo para cada nó, começando pela raiz, e aplica o mesmo algoritmo aos descendentes do nó, até que se atenda aos critérios de interrupção do algoritmo.

Vale destacar ainda, a fórmula do cálculo do ganho de informação, que é utilizado no algoritmo ID3, e definida por:

$$\text{Ganho}(A) = \text{Entropia}(D) - \text{Entropia}_A(D) \quad (7)$$

Sendo:

D = Conjunto dos dados de treinamento;

$$\text{Entropia}(D) = -\sum_{i=1}^m p_i \log_2(p_i);$$

m = Número de classes existentes;

p_i = Probabilidade da classe C_i ocorrer em D ;

$$\text{Entropia}_A(D) = \sum_{j=1}^v \frac{D_j}{D} \times \text{Entropia}(D_j);$$

v = Conjunto de todos os valores possíveis para o atributo A ;

D_j = Subconjunto de D para o qual o atributo A tem o valor v_j .

Já o algoritmo C4.5 (QUINLAN, 1993) possui algumas melhorias em relação ao ID3, como a utilização da poda da árvore para eliminar o problema de *overfitting* (sobreajustamento dos dados), trabalhar com valores indisponíveis, com valores contínuos e derivar regras.

⁴É uma medida estatística utilizada para medir a eficácia de um atributo em classificar os dados do treino. A escolha do atributo mais eficaz, ou seja, aquele que reduz mais a entropia, faz com que as árvores tenham tendência de serem menos profundas e com mais ramificações. A entropia neste caso é o cálculo utilizado para obter o grau de desorganização das informações.

Algoritmo 1: Pseudocódigo do Algoritmo do ID3.

Entrada: *Exemplos* : Dados de treinamento;

AtributoAlvo : Atributo cujo valor a árvore irá predizer ;

Atributos : Atributos utilizados para predizer o *AtributoAlvo*.

Saída: Árvore de decisão que classifica da melhor maneira os dados de treinamento.

início

 Crie *Raiz*, que será o nó raiz da árvore;

if todos os dados de *Exemplos* forem positivos **then**

 Faça *Raiz.classe* = "+";

 Retorne *Raiz*;

end

if todos os dados de *Exemplos* forem negativos **then**

 Faça *Raiz.classe* = "-";

 Retorne *Raiz*;

end

if *Atributos* for uma lista vazia **then**

 Faça *Raiz.classe* = valor de *AtributoAlvo* mais frequente entre *Exemplos*;

 Retorne *Raiz*;

else

 Faça *A* ser o atributo de *Atributos* que melhor classifica os dados de *Exemplos* (atributo com o maior ganho de informação);

 Faça *Raiz.atributo* = *A*;

repita

 Adicione um novo ramo ao nó *Raiz*, correspondente ao teste

$A = v_i$;

 Faça *Exemplos_{v_i}* ser o subconjunto de *Exemplos* formado pelos dados que têm o atributo *A* com valor v_i ;

if conjunto *Exemplos_{v_i}* for vazio **then**

 Crie uma nova folha *F*;

 Faça *F.classe* = valor de *AtributoAlvo* mais frequente entre *Exemplos*;

 Adicione *F* ao ramo;

else

 Crie uma subárvore *SubArv*;

 Faça *SubArv* = ID3(*Exemplos_{v_i}*, *AtributoAlvo*, *Atrib* = {*A*});

 Adicione *SubArv* ao ramo;

end

até Para cada possível valor v_i do atributo *A*;

end

 Retorne *Raiz*, que é o nó raiz da árvore de decisão.

fim

2.2 CÁLCULO DO VALOR DA IDEOLOGIA

Há também o valor da ideologia, utilizado em alguns modelos de tomada de decisão, que pode assumir diversas interpretações como “Liberal” e “Conservador” (POOLE, 2005), “Governistas” e “Oposicionistas”, “Contra” e “A Favor”, e, “Social” e “Democratas” . No exemplo de “Liberal” e “Conservador”, com relação ao valor da ideologia, quando é superior a 0 indica ser conservador, e inferior a 0 indica ser liberal (POOLE, 2005), podendo ainda se adotar outras interpretações. Para tanto, vários autores criaram métodos diferentes para calcular esse valor, conforme pode ser visto abaixo.

Poole (2005) apresenta em seu livro como utilizar um simples modelo espacial de votação para analisar dados de votações de legisladores. Poole (2005) utiliza mapas espaciais por acreditar que é importante visualizar uma grande quantidade de valores em representações gráficas neste formato, além de permitir visualizar legisladores que votem de maneira similar, quais partidos estão envolvidos a cada período de tempo, entre outras vantagens.

Cada legislador é representado por um ponto, e cada votação é representada por dois pontos, “Sim” e “Não”. Em cada votação, cada legislador vota próximo do resultado final (“Sim” ou “Não”), probabilisticamente. Com o acúmulo de todos os votos, forma-se um mapa espacial que resume os dados das votações, conforme exemplo da figura 5, que mostra a posição dos senadores no 90º Congresso Estadunidense.

Conforme figura 5, a dimensão vertical representa os senadores que foram identificados como *Northerners* (próximo ao topo) e os candidatos identificados como *Southerners* (próximo à base) em relação aos Direito Civis. A dimensão horizontal representa os candidatos identificados como conservadores (próximos à extrema direita) e os candidatos identificados como liberais (próximos à extrema esquerda). As letras “S” identificam os *Southern Democrats* (Democratas do Sul) , as letras “D” identificam os *Northern Democrat* (Democratas do Norte), e, as letras “R” os Republicanos

A partir dos dados, pode-se verificar os pontos que estão mais próximos, ou seja, legisladores que votam de forma parecida. A distribuição dos legisladores pelo “mapa” indica quantas dimensões o mapa espacial apresenta. Vale ressaltar que os mapas espaciais são baseados na geometria euclidiana dos pontos. Para uma dimensão são definidos p legisladores e q sessões de votação e se assume que cada legislador possui um ponto ideal, associado a uma função de utilidade.

Poole (2005) apresenta uma teoria simples de mapas espaciais que denomina de Teoria da Ideologia do Espaço Físico (*Basic-Space Theory of Ideology*) e demonstra como construir e

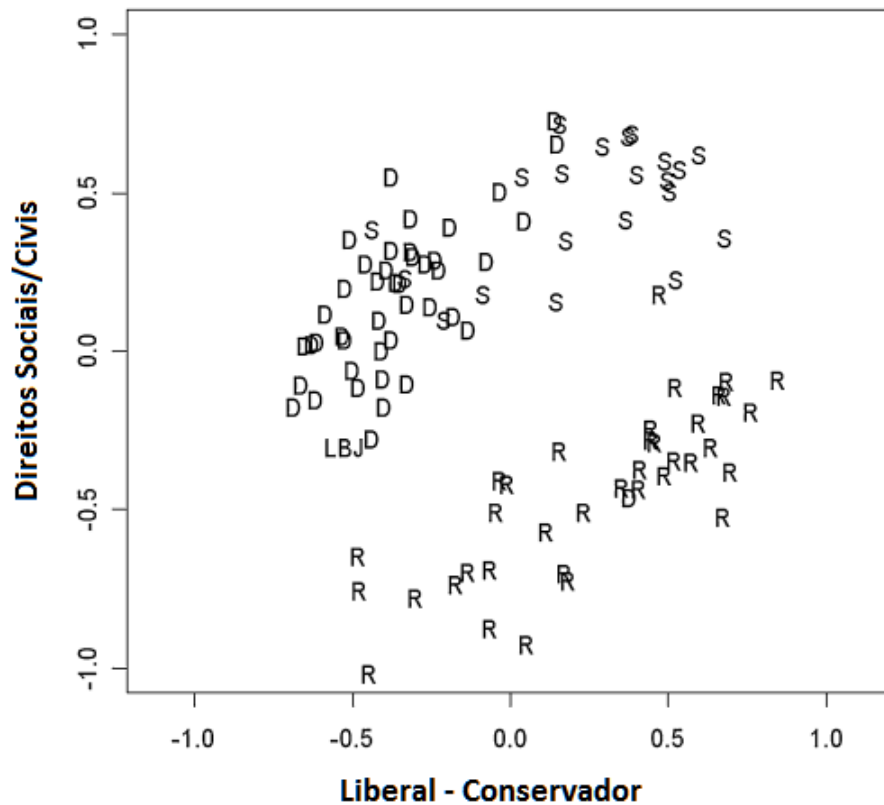


Figura 5: Mapa espacial dos senadores no 90º Congresso Estadunidense.

Fonte: Poole (2005). Tradução nossa.

interpretar os mapas espaciais. A palavra “teoria” é usada por Poole de forma ampla e inclui: (1) o aparato técnico do modelo espacial, (2) a teoria sobre como os legisladores tomam decisões, (3) a teoria de um sistema de crenças (ideologia) que está ligada à teoria do modelo espacial e à teoria de tomada de decisão, (4) o *software* que incorpora os itens 1, 2 e 3 e gera os mapas espaciais, (5) a compreensão do cenário político vigente na época dos dados. Todos esses dados são necessários para construir e compreender o mapa espacial.

Para encontrar o ponto ideal de cada legislador, deve-se seguir os 4 passos citados por Poole (2005): 1) calcular uma matriz de concordância legislador legislador; 2) converter a matriz nas distâncias ao quadrado; 3) calcular a matriz duplamente centrada; 4) calcular os autovalores, conforme figura 6 e explicação abaixo.

1) Calcular uma matriz de concordância legislador legislador

Para calcular a matriz de concordância (*Agreement Score*) para cada par de legisladores, que é a proporção de vezes que votaram da mesma maneira (neste caso legisladores *a* e *b*), em relação ao total de votações, deve-se executar a seguinte fórmula:

		Votações				
Legisladores	1	3	5	4	2	
Cinco	N	Y	N	N	N	
Seis	N	Y	Y	N	N	
Quatro	N	Y	N	Y	N	
Um	Y	N	N	Y	Y	
Dois	N	N	N	Y	Y	
Três	N	N	N	Y	N	

Valores de concordância					Distâncias Quadráticas				
1.0					.00				
.8	1.0				.04	.00			
.8	.6	1.0			.04	.16	.00		
.2	.0	.4	1.0		.64	1.00	.36	.00	
.4	.2	.6	.8	1.0	.36	.64	.16	.04	.00
.6	.4	.8	.6	.8	1.0	.16	.36	.04	.16

Matriz Duplamente Centrada					Pontos dos Legisladores				
.09					$x_5 = .3$				
.15	.25				$x_6 = .5$				
.03	.05	.01			$x_4 = .1$				
-.15	-.25	-.05	.25		$x_1 = -.5$				
-.09	-.15	-.03	.15	.09	$x_2 = -.3$				
-.03	-.05	-.01	.05	.03	.01	$x_3 = -.1$			

Figura 6: Passos para o cálculo da ideologia quando se possui somente uma dimensão.

Fonte: Poole (2005). Tradução nossa.

$$A_{ab} = \frac{\sum_{j=1}^q C_{abj}}{\sum_{j=1}^q \Delta_{abj}} \quad (8)$$

Sendo:

a = Legislador A;

b = Legislador B;

q = Número total de votações;

$C_{abj} = 1$ se os legisladores a e b votaram de maneira igual na votação j , caso contrário será igual a 0;

$\Delta_{abj} = 1$ se os legisladores a e b participaram da votação j , caso contrário será 0.

2) Converter a matriz nas distâncias ao quadrado

O segundo passo é converter a matriz de concordância em uma matriz de distâncias quadráticas, similar à distância Euclidiana. Para isto, deve ser subtraída a matriz de concordância de 1 e a elevar ao quadrado, através da seguinte fórmula:

$$d_{ab}^2 = (1 - A_{ab})^2 \quad (9)$$

Sendo:

A_{ab} = Matriz de concordância legislador legislador;

a = Legislador A;

b = Legislador B.

Vale ressaltar que quando não existe um índice de concordância entre os legisladores, Poole (2005) atribui um valor mediano às distâncias, ou seja, 0,5, sendo o seu quadrado igual a 0,25.

3) Calcular a matriz duplamente centrada

O terceiro passo é calcular a matriz duplamente centrada a partir da matriz de distâncias quadráticas. Esta matriz produzida é definida como uma matriz de produto cruzado (*cross product matrix*) das coordenadas do legislador (YOUNG; HOUSEHOLDER, 1938; ROSS; CLIFF, 1964). Para cada elemento da matriz de distâncias quadráticas, subtrai-se a média da linha, a média da coluna, adiciona-se a média da matriz e se divide por -2, conforme fórmula a seguir:

$$y_{ab} = \frac{d_{ab}^2 - d_b^2 - d_a^2 + d^2}{-2} \quad (10)$$

Sendo:

$d_b^2 = \frac{\sum_{a=1}^p d_{ab}^2}{p}$, que representa a média da b-ésima coluna;

$d_a^2 = \frac{\sum_{b=1}^p d_{ab}^2}{p}$, que representa a média da a-ésima linha;

$d^2 = \frac{\sum_{a=1}^p \sum_{b=1}^p d_{ab}^2}{p^2}$, que representa a média da matriz;

p = Número de legisladores;

a = Legislador A;

$b = \text{Legislador B.}$

4) Calcular os autovalores

Por fim, o último passo é obter os autovalores da matriz duplamente centrada, que podem ser considerados como o valor da ideologia, cuja fórmula é:

$$\det(A - \lambda I) = 0 \quad (11)$$

Sendo:

$A = \text{Matriz duplamente centrada;}$

$\lambda = \text{Autovalor;}$

$I = \text{Matriz identidade } I \text{ de } A.$

Apesar de se possuir o valor da ideologia, é necessária a utilização de métodos que projetem votos de acordo com a “utilidade” obtida em cada voto. Para calcular esta utilidade, pode-se utilizar o *Optimal Classification (O.C.)* (POOLE, 2005), que considera os erros que possam vir a existir durante a classificação. Apesar do O.C. maximizar a classificação correta, Poole (2005) cita que os principais métodos probabilísticos de projeção de votos do parlamento são o *NOMINATE*, *Quadratic-Normal (QN) Scaling* e o *Bayesian Markov Chain Monte Carlo (M.C.M.C.)* (METROPOLIS; ULAM, 1949).

No *Nominal Three-Step Estimation*, é utilizada a distribuição normal para definir o valor da utilidade, em que, conforme o valor se distancia do ponto ideal, menor é a sua utilidade⁵. Além disso, Poole (2005) considera que a distribuição normal é um modelo mais realista para comportamentos de tomada de decisões, pelo fato de que quando mais distante a posição ideológica do ponto ideal do legislador, menor será a sua utilidade, além de se basear em estudos de psicologia que utilizam métodos escalares multidimensionais.

A distribuição normal para o cálculo da utilidade é definida pela equação abaixo e pela figura 7, em que é comparada a distribuição normal com a distribuição quadrática. A equação abaixo define o valor da utilidade do voto “Sim” na votação j .

$$U_{ijy} = u_{ijy} + \varepsilon_{ijy} \quad (12)$$

⁵Neste caso, Poole (2005) cita o exemplo de um eleitor que deve votar em um candidato que possua interesses em comum, ou seja, que possua o valor de utilidade maior, portanto, este candidato deve ter sua posição política próxima ao ponto ideal do eleitor.

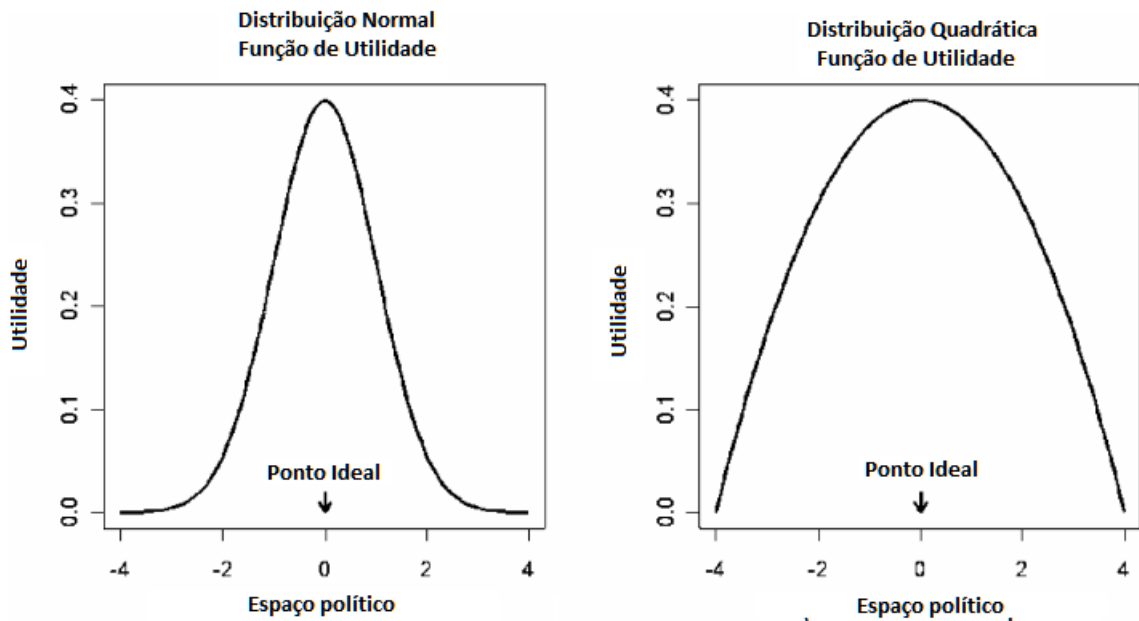


Figura 7: Função de utilidade quadrática e normal. As funções são unimodais e simétricas, e estão centradas no ponto ideal do legislador. A utilidade diminui com o aumento da distância dos pontos do espaço político comparado ao ponto ideal.

Fonte: Poole (2005). Tradução nossa.

Sendo:

j = Sessão de votação;

s = Número de dimensões;

i = Legislador i ;

ε_{ijy} = Termo estocástico, também denominado termo de erro;

$u_{ijy} = \beta \exp(-\frac{1}{2} \sum_{k=1}^s w_k d_{ijk}^2)$ = Parte determinística da função de utilidade;

w_k = “Peso” com $w_k > 0$;

β = Ajuste sobre o erro existente, e é proporcional à variância da distribuição de erros;

$f d_{ijk}^2 = (X_{ik} - O_{jky})^2$ = Matriz de distâncias quadráticas do legislador i para o resultado “Sim”, na dimensão k ;

X_{ik} = Coordenada do legislador i na dimensão k ;

$$O_{jky} = \frac{O_{jky} + O_{jkn}}{2} - \frac{O_{jky} - O_{jkn}}{2}.$$

Utilizando o valor da utilidade, Poole (2005) define a probabilidade de um legislador votar positivo ou negativo, conforme trecho abaixo.

$$\begin{aligned}
& “P(\text{Legislador } i \text{ votar Sim}) = P(U_{ijy} - U_{ijn} > 0) \\
& P(\text{Legislador } i \text{ votar Sim}) = P(\varepsilon_{ijn} - \varepsilon_{ijy} < u_{ijy} - u_{ijn}) \\
& P(\text{Legislador } i \text{ votar Não}) = P(U_{ijy} - U_{ijn} < 0) \\
& P(\text{Legislador } i \text{ votar Não}) = P(\varepsilon_{ijn} - \varepsilon_{ijy} > u_{ijy} - u_{ijn})” \text{ (POOLE, 2005)}
\end{aligned}$$

A partir destes cálculos, Keith Poole e Howard Rosenthal, desenvolveram o *NOMINATE* em 1982, um método de escalonamento multidimensional⁶ para analisar dados de votações e definir preferências políticas. Posteriormente, foi utilizado como base para a criação do *D-NOMINATE* e do *W-NOMINATE* que utilizam o modelo *Logit* (CRAMER, 2003) na distribuição do erro. O algoritmo *W-NOMINATE* é estático em relação ao seu ponto ideal e sua função de utilidade é ligeiramente diferente do *D-NOMINATE*, além de utilizar o modelo de distâncias com peso (*weighted euclidean metric*). Posteriormente, Carroll et al. (2009) criaram o *DW-NOMINATE*, que se diferencia do *W-NOMINATE* por ser dinâmico e utilizar a distribuição dos erros como uma distribuição normal.

Além do método de Keith Poole citado anteriormente, há outros métodos que calculam a ideologia, como o *Segal-Cover Score* e o *Martin-Quinn Score*.

O *Segal-Cover Score*, abordado no artigo *Ideological Values and the Votes of U.S. Supreme Court Justices* (SEGAL et al., 1995), define as preferências ideológicas através de valores entre -1 (extremamente conservador) e 1 (extremamente liberal). Esse valor é calculado através de notícias, em vista de nos Estados Unidos da América ser comum a imprensa dar ampla cobertura ao processo de indicação dos novos Ministros pelo Presidente da República, debatendo acerca da posição ideológica dos possíveis futuros Ministros. Alguns assuntos são colocados em pauta também para definir a posição ideológica, ou seja, como ser a favor (liberal) ou não (conservador) ao aborto. Entretanto, nem todos os assuntos são abordados pelas notícias, logo não é possível definir a posição ideológica dos juízes em todos os temas (RIBEIRO, 2011).

Por fim, no trabalho de Martin e Quinn (2002), é abordado o *Martin-Quinn Score*, que utilizam cadeias de *Markov* e modelos Bayesianos para estimar a posição ideológica dos juízes. Um diferencial encontrado no algoritmo de *Martin-Quinn* é considerar que a ideologia do juiz pode mudar, portanto, calcula-se dinamicamente esse valor, verificando-se a sua tendência ao longo do tempo. Martin e Quinn (2002) verificaram com esse algoritmo, que os juízes não mantêm uma posição ideológica constante ao longo do tempo e o algoritmo apresentou bom

⁶O escalonamento multidimensional é projetado para construir um diagrama mostrando os relacionamentos entre um certo número de objetos, sendo dada somente uma tabela de distâncias entre os objetos. O diagrama é um tipo de mapa que pode ser em uma dimensão (se os objetos caem em uma reta), em duas dimensões (se os objetos caem em um plano), em três dimensões (se os objetos podem ser representados por pontos no espaço), ou em um número mais alto de dimensões (caso em que uma simples representação geométrica não é possível) (MANLY, 2008).

desempenho, principalmente em questões de direitos civis. Outra vantagem do método é a possibilidade de utilizá-lo quando há poucos dados ou dados faltantes sobre determinado juiz.

2.3 MODELOS

A seguir, serão abordados os diferentes modelos existentes, tanto teóricos, quanto exemplos de aplicações com dados de tribunais de outros países.

2.3.1 JUDICIAIS

Segundo Hwong (2006), há cinco maneiras de explicar uma tomada de decisão judicial, através do modelo legal, atitudinal, atributos pessoais, estratégico e institucional, conforme descritos abaixo. Vale ressaltar que estes modelos podem ser utilizados em conjunto, e não servem apenas para identificar um comportamento de um tribunal.

Modelo Legal

Segundo Hwong (2006), o modelo legal é aquele que define que os juízes se baseiam somente na lei ao tomar suas decisões, ou seja, o modelo legal é a premissa que se adota, de que os membros poder judiciário são imparciais. Neste modelo, pode-se considerar também as jurisprudências como embasamento para as decisões.

Modelo Atitudinal

Já o modelo atitudinal é aquele que define que os juízes se baseiam em preferências políticas pessoais, mais que na lei, ao tomar suas decisões. Hwong (2006) cita o exemplo de Pritchett (1948), que afirmou que os juízes tendem a apoiar decisões de juízes mais próximos do seu círculo de amizade, desde que sejam iguais às suas preferências políticas ou próximas delas. No Canadá, Sidney Peck foi o primeiro a aplicar técnicas quantitativas para mapear padrões de votos dos juízes de acordo com esse modelo. Para analisar os dados da Suprema Corte do Canadá entre 1958 e 1966, utilizou-se da técnica denominada *scalogram analysis* (PECK, 1967), que é geralmente aplicada em casos não unânimes.

Um ponto interessante levantado por Peck, foi o de analisar o quanto os juízes são favoráveis às empresas em processos nos quais o governo é parte também. Nesse caso, Peck definiu uma faixa de porcentagem relativa ao número de casos não unânimes julgados em que empresa e governo eram partes, por exemplo, em 80% ou mais o juiz pode ser considerado fortemente a favor das empresas, entre 60% e 79% é definido como a favor das empresas, entre

41% e 59% é definido como neutro, entre 21% e 40% é definido como contra e 20% ou menos é classificado como fortemente contra empresas.

Peck também analisou casos sobre negligência, tributação e direito penal, e dividiu os votos em dois grupos, a favor ou contra a parte (por exemplo, voto a favor do governo ou contra o governo). Já Fouts (1967), fez um trabalho similar ao de Peck, entretanto extraiu dos casos da Suprema Corte do Canadá, entre os anos 1950 e 1960, os votos não unânimes e os classificou em dois tipos, os que tratam de liberdade dos civis e os que tratam do liberalismo econômico. Os casos classificados como liberdade dos civis, são aqueles em que há conflitos entre direitos pessoais e reivindicações de liberdade e a autoridade do governo. Já, os casos classificados como liberalismo econômico são aqueles em que há conflitos entre os mais carentes financeiramente e os que possuem muitas riquezas e detêm monopólios, como por exemplo, as empresas.

Modelo de Atributos Pessoais

Outro modelo é o de atributos pessoais, originado do modelo atitudinal, que define que os juízes se baseiam em preferências pessoais que podem vir a influenciar suas preferências políticas pessoais. Nas preferências pessoais incluem-se o perfil socioeconômico dos juízes, características regionais e afiliações políticas em cada caso. Hwong (2006) cita exemplos de vários pesquisadores, como o de George (2001), que inferiu que ex-professores de direito que foram nomeados para o Tribunal de Apelação dos EUA, apresentaram uma maior tendência, comparados aos outros juízes, a dar opiniões e reverter opiniões de tribunais inferiores, a fim de modificar o voto anterior, além de agir conforme sua posição ideológica acadêmica.

Outro exemplo citado é o de Brudney et al. (1999), que inferiram que os juízes democráticos tendem a votar favoravelmente aos sindicatos em um maior número de vezes, comparados aos juízes republicanos, e, mesmo no caso em que os juízes republicanos votam a favor dos sindicatos, as juízas tendem a ser mais favoráveis do que os juízes.

Modelo Estratégico

O modelo estratégico é aquele que define que os juízes modificam seus votos de acordo com as interações feitas com outros juízes. Hwong (2006) cita os exemplos de Murphy (1964) e Epstein e Knight (1997), que analisaram dois conjuntos de dados, um do mandato de 1983 e outro do mandato de Warren Burger entre 1969 e 1985, em que foi identificado que o juiz William Brennan, no mandato de 1983, votou estrategicamente para garantir que a revisão das decisões fossem concedidas.

Modelo Institucional

O modelo institucional é aquele que define que os juízes votam de acordo com a influência exercida pelo órgão em que trabalham, como fatores políticos ou procedimentais. Esse modelo é fortemente ligado ao modelo estratégico. Os fatores políticos e procedimentais incluem quais casos serão julgados.

Para exemplificar esse modelo, pode-se citar o trabalho de Russell (1969), que, através de uma visão institucional e estratégica, após analisar 1031 casos entre 1950 e 1964, concluiu que a Suprema Corte possui uma forte tendência a reverter as decisões do Tribunal de Apelação de Quebec.

Modelo Contextual

Ainda, o modelo contextual, segundo Cameron e Cummings (2003), explora as características de outros juízes que podem influenciar no voto de um juiz. Cameron e Cummings (2003) identificaram que pelo menos na Corte de Apelação dos Estados Unidos da América, houve interações sociais substancialmente significativas entre os juízes, entretanto a interpretação de seus efeitos não é trivial.

Modelo de Economia Social

O modelo de economia social, criado por Cameron e Cummings (2003), possui como objetivo estudar a influência de fatores na tomada de decisões judiciais em colegiados, e utiliza como estudo de caso a Corte Federal de Apelação. Este modelo é baseado no estudo de interações sociais de grupos, e ele pode ser utilizado juntamente com os modelos atitudinal e o contextual, conforme os trabalhos de Manski (2000) e Becker e Murphy (2000). Além de considerar as características de outros juízes que podem influenciar no voto de um juiz, considera também as ações tomadas pelos outros.

Para que as características de um juiz possam influenciar um grupo, considera-se que este grupo deve conter pessoas com características diversas nos aspectos raciais, de gênero sexual e ideologia, por exemplo. Esse aspecto está contido nos três que devem ser considerados separadamente no modelo da economia social: 1) o impacto das características de um juiz (raça, gênero sexual e ideologia) nele mesmo; 2) o impacto das características dos outros juízes em um juiz; 3) o impacto de cada ação dos outros juízes em cada juiz.

Há ainda outro fato que pode ser levado em conta nesse modelo. Segundo Cameron e Cummings (2003), pode haver uma norma de consenso, ou seja, votação unânime para garantir a legitimidade do tribunal que está proferindo a sentença e evitar uma possível reversão da

decisão por um tribunal superior.

2.3.2 ESTATÍSTICOS

No trabalho de Hwong (2006), é explorada a influência das características sociodemográficas (laços políticos, laços regionais, gênero sexual, anos de experiência como advogado, entre outros) dos juízes em suas decisões relativas ao Direito Tributário. Para Hwong (2006), seu trabalho contribui para a compreensão das decisões judiciais de Direito Tributário, com três ideias. A primeira é que as características sociodemográficas são variáveis influenciáveis, ou seja, juízes com características parecidas tendem a ter o mesmo comportamento durante os julgamentos. A segunda ideia é a de que cada tribunal possui seu conjunto de variáveis sociodemográficas e seu valor de influência. Já a terceira, é a de que uma análise quantitativa contribui para se conhecer melhor como são tomadas as decisões judiciais, agregada de uma análise qualitativa posterior.

A análise, apesar de explorar o comportamento dos juízes, não explora o seu comportamento individual e sim, quando estão inseridos em grupos, como nas decisões em colegiado. Hwong (2006) utilizou dois modelos como base, o modelo de atributos pessoais⁷ e o modelo de Schneider, além de análises bivariadas e multivariadas⁸.

Já o trabalho de Martin et al. (2004), na área de modelos estatísticos de tomada de decisão, tenta prever resultados de processos ainda não julgados na Suprema Corte dos EUA. Após a decisão real dos casos, há uma comparação com os resultados previstos. Para isso, Martin et al. (2004) selecionaram casos que estavam pendentes na Suprema Corte dos EUA e aplicaram dois métodos diferentes para compararem os resultados.

O primeiro método consiste na aplicação de um modelo estatístico com base em informações derivadas de decisões passadas da Suprema Corte, anteriores ao caso a ser previsto, em que foram construídas árvores de classificação com base nas características de cada caso. Já o segundo método, uma análise qualitativa, baseia-se em julgamentos de especialistas e pesquisadores na área jurídica que podem possuir opiniões baseadas nas leis, livros, opiniões de tribunais, preferências políticas e ideológicas.

Após analisarem 628 casos anteriormente decididos, Martin et al. (2004) criaram árvores de classificação, que são citadas como modelo estatístico pelos mesmos. Para cada caso, foram obtidas seis características que foram utilizadas como variáveis explanatórias, sendo

⁷Este modelo demonstra que os juízes tomam decisões influenciados por suas características pessoais.

⁸Segundo Manly (2008), o ponto principal de uma análise multivariada é considerar variáveis relacionadas simultaneamente, sendo todas consideradas igualmente importantes, pelo menos inicialmente.

elas: (a) Tribunal de origem do caso, (b) Tema do caso, (c) Tipo do autor do processo (por exemplo, EUA, pessoa lesada, empresa, entre outros), (d) Tipo do réu, (e) Ideologia da decisão de primeira instância, e, (f) Autor do processo argumentou sobre a constitucionalidade de uma lei.

Outro trabalho que pode ser citado é o de Boutilier e Lu (2011), que abordam sobre modelos de escolha social, utilizados na definição de como são formadas as decisões dos juízes e até na projeção de votos. No trabalho, Boutilier e Lu (2011) definem que há vários fatores que devem ser levados em conta na escolha social como: (a) é impraticável a obtenção do valor de preferências completas, tipicamente as preferências são estimadas observando-se o comportamento do indivíduo quando realiza escolhas, (b) quanto mais dados para treinar os modelos probabilísticos de escolha social, mais realista será o modelo, (c) a aproximação dos dados é necessária por várias razões, como por exemplo, a falta de dados. Outro ponto interessante levantado por Boutilier e Lu (2011) é o dos agentes que podem manipular as votações, ou seja, mudar o resultado de uma eleição, por exemplo, declarando informações incorretas.

Já Cameron e Cummings (2003), criaram um modelo utilizando conceitos de Economia Social e efeitos dos vizinhos ou pares (*neighborhood/peer effect*), como por exemplo, a prática de crimes, ingestão de bebidas alcoólicas, gravidez na adolescência, tabagismo entre adolescentes, abandono escolar, entre outros.

No trabalho de Cameron e Cummings (2003), são citados alguns valores de grande importância do modelo, como a utilidade para o juiz de cada tipo de voto proferido v_i , sendo $u_i(v_i) = h_i(v_i) + \theta_1 S_i(v_i, \bar{v}_i) + \varepsilon_i(v_i)$, que utiliza como base o impacto das características de um juiz i em si mesmo, o impacto das características dos outros juízes votantes no juiz i , e o impacto dos votos dos outros juízes no voto do juiz i . O valor $h_i(v_i)$ representa a utilidade do voto do juiz i para si mesmo; θ_1 representa um coeficiente de ajuste para definir a influência do voto dos demais juízes no juiz votante; $S_i(v_i, \bar{v}_i)$, a utilidade do voto do juiz i para o grupo; \bar{v}_i é o valor médio dos votos dos outros juízes que estão julgando e $\varepsilon_i(v_i)$ representa o comportamento peculiar do juiz i .

A utilidade privada reflete o valor do voto juiz i no que ele crê que seja correto, baseando-se na informação que possui do caso, compreensão da lei, jurisprudências anteriores e convicções pessoais, além da possível influência dos outros juízes do colegiado. Já a função da utilidade social ($S_i(v_i, \bar{v}_i)$) é obtida pela seguinte fórmula $S_i(v_i, \bar{v}_i) = 1 - |v_i - \bar{v}_i|$, em que no caso de três juízes, se o juízes 2 e 3 votam do mesmo modo, o juiz 1, recebe o valor 1 para a utilidade social, se ele votar da mesma forma, caso contrário receberá 0. Mas no caso do juiz 2

votar de um jeito e o juiz 3 de outro, o juiz 1 receberá 0,5, independentemente do seu voto.

2.4 SIMULAÇÃO EM CÂMARAS COLEGIADAS

Dentre diversos trabalhos sobre simulação em câmaras colegiadas existentes, pode-se citar os trabalhos de Gaines et al. (1996) e Tanford e Penrod (1983), que abordam sobre a tomada de decisões em júris.

Gaines et al. (1996) desenvolveram um modelo computacional dos processos cognitivos envolvidos na tomada de decisão de um júri e utilizaram como base o modelo *Story Model* de Pennington e Hastie (1986) e o modelo de Schank (1995). Os trabalhos utilizados como base por Gaines et al. (1996) contribuem para determinar que um júri ao votar, deve compreender bem todos os fatos e ser capaz de explicá-los, e, para isto deve seguir alguns passos como 1) selecionar a narrativa dos fatos que mais se aproxima da evidência de inocência ou culpa do réu, 2) aprender com o juiz os tipos de votos que podem proferir (culpado ou não culpado, por exemplo), e, 3) atribuir aos fatos da narrativa escolhida as sentenças definidas.

A seleção da narrativa dos fatos deve se dar através de três princípios, a cobertura, a coerência e a singularidade. A cobertura indica que a maioria das evidências deve ser contemplada pela narrativa escolhida. Já a coerência indica que não deve haver contradição nos argumentos e os eventos devem corresponder ao que os jurados acreditam como os fatos ocorram, além de não haverem lacunas na narrativa dos fatos. Já o princípio da singularidade indica que quanto mais chances de uma narrativa ser escolhida, em detrimento das outras, maior será a sua singularidade, entretanto esta deve corresponder também às evidências.

Diante destas premissas e de outras citadas no trabalho de Gaines et al. (1996), foi utilizado o modelo de simulação denominado *WAPNER* (PENNINGTON; HASTIE, 1988), que simula a tomada de decisão de um júri. Em primeiro lugar, o simulador indica a narrativa que melhor explica as evidências, e, posteriormente fornece o veredito e o nível de confiança do mesmo. Gaines et al. (1996) utilizaram o *WAPNER* em três experimentos distintos e chegaram a conclusão de que o simulador se mostrou eficiente ao que propõem.

Outro trabalho que aborda a simulação em câmaras colegiadas é o de Tanford e Penrod (1983), que busca investigar a influência de uma opinião intransigente em um júri utilizando simulações computacionais de tomada de decisão. Os autores citam que em alguns casos uma minoria pode influenciar a maioria, como verificado no estudo de Latané (1981) que calcula os efeitos da influência destes dois grupos. Latane e Wolf (1981), ainda trabalham a teoria do impacto social que está relacionada ao número de membros em cada grupo, e pode contribuir

para a compreensão da influência dos jurados.

Tanford e Penrod (1983), ressaltam a importância da simulação computacional para determinar o impacto das características de cada júri nos vereditos, além de permitir realizar combinações de diferentes tamanhos de júris, regras de decisão (quantidade de votantes da maioria), criação de jurados intransigentes, e, acompanhar os resultados obtidos.

Já Penrod et al. (1979), apresentam um modelo de tomada de decisão em júri que simula o poder de persuasão durante o processo de deliberação de um júri. Alguns dos parâmetros deste modelo incluem o tamanho do júri, regras de tomada de decisão, e, uma probabilidade inicial de um membro do júri tentar convencer os demais. O modelo de Penrod et al. (1979), denominado *DICE*, utiliza cadeias de *Markov* ao representar as deliberações dos júris, e, a probabilidade de mudança de opinião de um jurado pode ser obtido através de uma matriz de probabilidades de transição, causado pelo poder de persuasão da maioria ou ainda da minoria.

Um dos resultados obtidos por Tanford e Penrod (1983) indicou que quando os jurados com opiniões consistentes constituíam a minoria, a influência era muito mais forte em júris com seis membros, comparado aos júris com doze membros.

3 MÉTODO DE PESQUISA

Este capítulo apresenta as etapas do método de pesquisa para o desenvolvimento do modelo proposto neste trabalho. De modo geral a metodologia deste trabalho pode ser representada pela figura 8. Na seção 3.1 é abordado o protótipo de extração de dados desenvolvido no estudo de caso abordado neste trabalho. Na seção 3.2 é detalhado como foram selecionados e tratados os dados de um estudo de caso de câmaras colegiadas. Na seção 3.3 são abordados como foram selecionados os modelos e as variáveis que serão utilizadas. Na seção 3.4 são citadas as técnicas para projetar votos, a partir dos modelos obtidos. Por fim, na seção 3.5 é abordado como foram realizadas as simulações.

3.1 AQUISIÇÃO DO CONJUNTO DE DADOS

A primeira tarefa é o desenvolvimento do protótipo de extração de dados de documentos e a definição da fonte de dados que disponibilize documentos publicamente em seus *sites*.

É necessário escolher um tipo de assunto em pauta e uma câmara colegiada para analisar os dados e acompanhar o comportamento dos membros do colegiado de acordo com este assunto, uma vez que o voto pode variar de acordo com o assunto publicado. O critério para escolha do tribunal foi disponibilizar os documentos em formato *HTML (HyperText Markup Language)* para permitir de forma viável a sua extração. Diante disso, o escolhido foi o Superior Tribunal de Justiça, um órgão do Poder Judiciário Brasileiro que julga parte dos processos em colegiado.

Já os critérios para escolha do assunto em pauta, levaram em conta a relevância, encontrando-se uma quantidade considerável disponível no órgão escolhido e ser um assunto contínuo no tempo, ou seja, sempre haverá processos para acompanhar a evolução do comportamento dos membros do colegiado. Assim, foi escolhido o assunto “Tributário”.

Já a segunda tarefa consiste em verificar quais dados estão disponíveis nesses

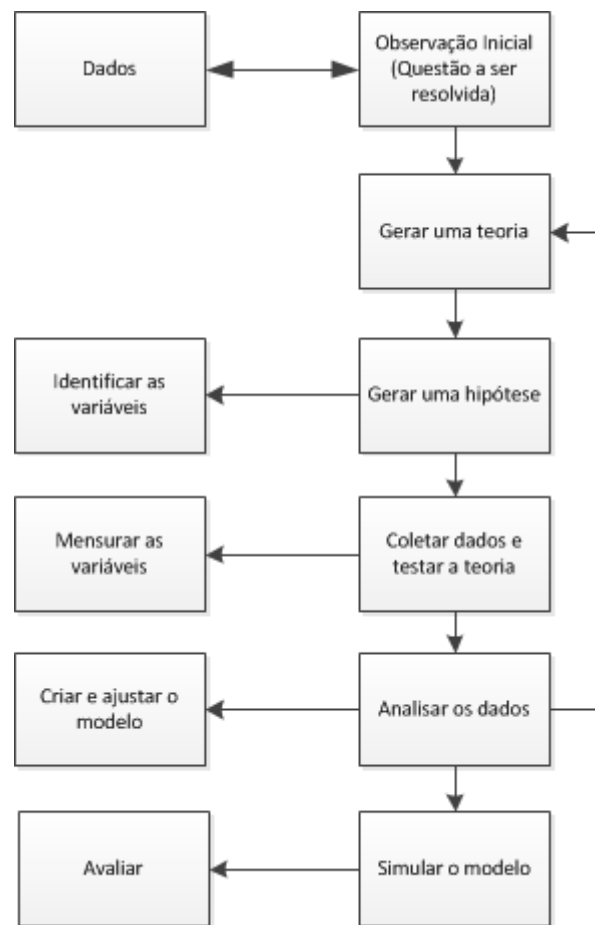


Figura 8: Passos para a criação do modelo e simulação.

Fonte: Própria.

documentos para realizar uma recuperação de informação e inseri-los em um banco de dados.

A terceira tarefa inclui a verificação da integridade das informações extraídas, para isso deve ser selecionada uma amostra dos dados utilizando-se a técnica de amostragem simples para a verificação manual dos documentos, e averiguar se os dados condizem com o que foi inserido no banco de dados. Para calcular o tamanho desta amostra, deve ser aplicar a seguinte fórmula:

$$n = (N * n_0) / (N + n_0) \quad (13)$$

Sendo:

$$n_0 = 1/E^2;$$

N = Tamanho da população;

E = Erro amostral, por convenção adota-se 0,05 ou 0,1 que representam

respectivamente 95% e 90% de confiança;

n = Tamanho da amostra, ou seja, quantidade de dados que devem ser testados.

Logo, através do cálculo do tamanho da amostra, em que $N = 650000$ documentos de acórdãos, e, adotando $E = 0,01$, deverão ser analisados manualmente 100 acórdãos para verificar se as informações contidas no banco de dados estão de acordo com o que foi disponibilizado *online*.

Os dados foram filtrados também em processos com decisão por maioria para identificar as coalizões de votos, e com somente um réu e um autor, para eliminar possíveis erros da análise automatizada dos tipos de decisões (positivas, negativas ou parciais).

3.2 PRÉ-PROCESSAMENTO DOS DADOS

Nesta seção são abordados os passos do pré-processamento dos dados para melhorar a precisão, eficiência e escalabilidade do processo de classificação e previsão, conforme citado por Han (2005).

De modo geral, primeiramente executa-se uma limpeza dos dados, a fim de eliminar ou reduzir o ruído, além da substituição de valores faltantes por valores de ocorrência mais frequente para o atributo, ou ainda, o valor mais provável.

Em seguida, realiza-se a análise de relevância dos atributos, eliminando os atributos redundantes, através da análise por correlação entre atributos fortemente correlacionados, ou ainda, a transformação de intervalos de valores em categorias, como por exemplo converter nomes de cidades em estados ou regiões. Pode-se também selecionar um conjunto menor de atributos a partir do conjunto completo, utilizando de técnicas heurísticas, como por exemplo, a indução por árvore de decisão.

Posteriormente, deve ser executada a transformação dos dados através da normalização, para reduzir atributos com valores muito grandes, como por exemplo a renda, ou ainda ano de nascimento.

É indicado ainda obter uma visão global dos dados, através de técnicas de sumarização para identificar propriedades típicas dos dados, e verificar valores que devem ser tratados como ruído ou discrepantes. Dentre as propriedades típicas, incluem-se as medidas de tendência central e a dispersão dos dados.

Dentro do pré-processamento dos dados inclui-se o cálculo do valor da ideologia. O valor da ideologia pode ser calculado de acordo com os votos dos membros do colegiado

extraídos dos documentos. Para calcular este valor foi utilizado o algoritmo *W-NOMINATE* de Keith Poole e Howard Rosenthal (POOLE et al., 2011).

Como dados de entrada do algoritmo, foram utilizados os dados dos votos dos membros em cada decisão que participaram, similar ao propósito do algoritmo original que utilizava dados de votos de senadores em sessões de votações de leis. No estudo de caso abordado por Poole et al. (2011), cada senador/deputado é representado por uma linha, já nas votações em que participam são representados por colunas, formando assim uma matriz das votações que cada senador participou e o voto que proferiu.

Após a execução do algoritmo, obtêm-se os valores da ideologia para cada membro do colegiado estudado, desde que hajam votos suficientes para o algoritmo calcular o valor.

3.3 CRIAÇÃO DOS MODELOS

Para a criação dos modelos, foram selecionados dois modelos distintos, o de regressão logística e o de árvore de decisão. O modelo de regressão logística binária foi escolhido pelo fato de não requerer que as variáveis independentes possuam distribuição normal ou tenham variância similar, além disso, este modelo não assume que as variáveis independentes devam ter relação linear com a variável dependente, e, geralmente é utilizado quando há muitos efeitos não lineares entre as variáveis, ou seja, viável para dados de ciências sociais que apresentam este tipo de comportamento. Já o modelo de árvore de decisão foi escolhido pelo fato de ser um método simples, de fácil aplicação e capaz de identificar as variáveis mais importantes para projetar a variável dependente.

Para o desenvolvimento dos modelos, buscou-se utilizar um modelo de tomada de decisão já existente para a escolha dos atributos, neste caso o de Cameron e Cummings (2003), conforme detalhado anteriormente.

A seguir, será exposto como foram escolhidas as amostras de treinamento e teste dos modelos, a seleção das variáveis disponíveis, e a avaliação dos modelos obtidos.

3.3.1 SELEÇÃO DA AMOSTRA

As amostras dividem-se em amostra de treinamento e teste. A amostra de treinamento é utilizada na criação de um modelo a partir destes dados. Já a amostra de teste, serve para testar o modelo obtido a partir da amostra de treinamento e verificar se este apresenta uma porcentagem de acertos adequada.

Para a escolha da amostra de treinamento, aplicou-se o método denominado amostra estratificada, com aproximadamente $2/3$ do total de registros em cada grupo da variável dependente, ou seja, a variável voto. A escolha pela amostra estratificada, se deu por possibilitar a obtenção da mesma quantidade de registros dos dois valores da variável voto, o voto positivo e o voto negativo.

3.3.2 SELEÇÃO DAS VARIÁVEIS

Para a escolha das variáveis que irão compor os modelos, foram identificadas todas as variáveis disponíveis nos currículos dos membros do colegiado e nos documentos. Em seguida, realizou-se uma correlação bivariada, de todas as variáveis para verificar se essas atendiam à regra de não terem uma alta correlação que prejudicasse a regressão.

Num primeiro momento, são utilizadas todas as variáveis, a fim de identificar quais exercem mais influência que as outras. Porém, num modelo mais refinado para o estudo de caso, é possível reduzir o número de variáveis.

3.3.3 AVALIAÇÃO DOS MODELOS

A avaliação do modelo de regressão logística é diferente do modelo de árvore de decisão, ou seja, cada modelo possui parâmetros específicos que indicam a qualidade do modelo.

3.3.3.1 Regressão Logística

No modelo de regressão logística binária, o modelo pode ser avaliado de modo geral e a nível de variáveis. Para avaliação do modelo no geral, deve-se verificar os seguintes valores¹:

- Log verossimilhança inicial: Valor de aderência do modelo quando nenhuma variável preditora está inclusa;
- Log verossimilhança final: Valor de aderência do modelo quando todas as variáveis preditores estão inclusas;
- Porcentagem de acerto dos votos positivos: Porcentagem de votos positivos que são projetados pelo modelo na amostra de teste de modo correto;

¹Estes valores podem ser obtidos ao executar o método de regressão logística binária no *software* SPSS v.21. da empresa IBM.

- Porcentagem de acerto dos votos negativos: Porcentagem de votos negativos que são previstos pelo modelo na amostra de teste de modo correto;
- R^2 de Cox & Snell (COX; SNELL, 1968): Métrica para avaliar a qualidade do modelo. O valor varia entre 0 e 1;
- R^2 de Nagelkerke (NAGELKERKE, 1991): Métrica para avaliar a qualidade do modelo. O valor varia entre 0 e 1.

Os valores de R^2 podem variar de 0 (previsores são inúteis na previsão da variável de saída) a 1 (o modelo prevê perfeitamente a variável de saída). Por exemplo, se o R^2 de um modelo é 0,8234, isto significa que 82,34% da variável dependente consegue ser explicada pelos regressores presentes no modelo. Entretanto, é comum encontrar o valor de R^2 baixo em fenômenos das ciências sociais, conforme citado por Hosmer e Lemeshow (2000).

Já para a avaliação das variáveis presentes no modelo, deve-se atentar para os seguintes valores obtidos:

- *B*: Valor da variável que é utilizado na equação da regressão logística para projetar a variável dependente;
- *Standard Error*: Erro padrão associado ao valor da constante;
- *Wald*: Utilizado para avaliar se o parâmetro é estatisticamente significativo. Quanto maior que zero for para o previsor, mais este contribui para a previsão da variável de saída.
- *Sig.* : Convencionalmente se define que caso o valor seja $< 0,05$, indica que a variável é significante no modelo. Deve ser analisado com o valor de *Wald*;
- *Exp(B)*: Exponenciação do coeficiente *B*. Caso o valor seja maior que 1, a medida que este valor aumenta, aumentam as chances de ocorrer uma determinada saída. Caso seja menor que 1, a medida que o previsor aumenta, diminuem as chances da saída determinada ocorrer (FIELD, 2005).

Vale ressaltar que a escolha das variáveis que melhor representam o modelo, impacta nos valores que indicam a qualidade do modelo no modo geral.

3.3.3.2 Árvore de Decisão

No modelo de árvore de decisão, o modelo pode ser avaliado de modo geral através dos seguintes valores²:

- *Correctly Classified Instances*: Representa o número de instâncias classificadas corretamente;
- *Incorrectly Classified Instances*: Representa o número de instâncias classificadas incorretamente;
- *Kappa statistic*: Medida de concordância usada em escalas nominais. Quanto menor o valor de Kappa, menor a confiança de observação, o valor 1 implica a correlação perfeita;
- *Mean absolute error*: Média da diferença entre os valores atuais e os projetados em todos os casos, é a média do erro da projeção;
- *Root mean squared error*: Utilizado para medir o sucesso de uma predição numérica;
- *Relative absolute error*: É o erro total absoluto. Em todas as mensurações de erro, valores mais baixos significam maior precisão do modelo, com o valor próximo de zero se atinge o modelo estatisticamente perfeito.

Com relação ao valor de *Kappa statistic*, deve-se considerar os valores da tabela 1 para verificar se o valor obtido através dos dados é adequado.

Tabela 1: Valores de referência para o *Kappa statistic*.

Fonte: Landis e Koch (1977). Tradução nossa.

Intervalo de Valores	Concordância
<0	Sem concordância
0 - 0,19	Pobre
0,20 - 0,39	Razoável
0,40 - 0,59	Moderada
0,60 - 0,79	Substancial
0,80 - 1	Excelente

²Estes valores podem ser obtidos ao executar o algoritmo J48 no *software* Weka 3.6.10 desenvolvido pela Universidade de Waikato .

3.4 PROJEÇÃO DOS DADOS

Para a projeção dos dados, deve-se utilizar a amostra de teste e verificar se o modelo possui porcentagens de acertos similares ao da amostra de treinamento, em relação à projeção da variável voto. Caso não, indica que modelo provavelmente necessite ser modificado, retirando variáveis conforme a avaliação apresentada na subseção 3.3.3, ou ainda inserindo novas variáveis que possam contribuir para o modelo.

Pode haver alguns problemas quanto a questão da amostra selecionada, sendo necessário nestes casos, a adoção de outros métodos que melhor representem a população dos dados.

3.5 CONSTRUÇÃO DA SIMULAÇÃO

Para a simulação, foi considerada a transparência do comportamento, além da eficácia na apresentação dos resultados. Segundo Lorscheid (2011), esse processo pode ser dividido em várias etapas: 1) concepção do modelo, 2) construção do modelo, 3) verificação, 4) validação e 5) análise de sensibilidade. Uma das principais questões durante a análise é como produzir e analisar dados de simulações, para responder às questões levantadas na pesquisa.

Neste mesmo sentido, Railsback e Grimm (2011) ressaltam que o pesquisador deve verificar se os resultados são realmente novos e importantes ou apenas consequência de uma decisão durante a modelagem. Faz-se necessário também encontrar discrepâncias entre a previsão e a saída dos resultados do modelo.

O processo de análise é dividido em duas partes: 1) análise pré-experimental, e, 2) análise experimental (MONTGOMERY, 2006), conforme mostra a figura 9.

Cada uma das fases é composta de algumas atividades. O primeiro passo consiste em definir o objetivo do experimento, a fim de garantir uma análise eficaz e eficiente dos resultados da simulação, para a questão de pesquisa poder ser respondida. Os principais objetivos desse passo são identificar os fatores importantes e seus efeitos sobre a resposta, e verificar se existem diferentes conjuntos de variáveis que permitem obter um desempenho aceitável (LAW; KELTON, 2000). Em algumas hipóteses da pesquisa, pode haver alguns conjuntos diferentes de variáveis para que as hipóteses sejam testadas.

A segunda atividade é a classificação das variáveis, a fim de distribuí-las nas categorias de independente, dependente e variáveis de controle (FIELD; HOLE, 2003). As variáveis de controle são aquelas fixadas por um valor ou distribuídas aleatoriamente por uma distribuição

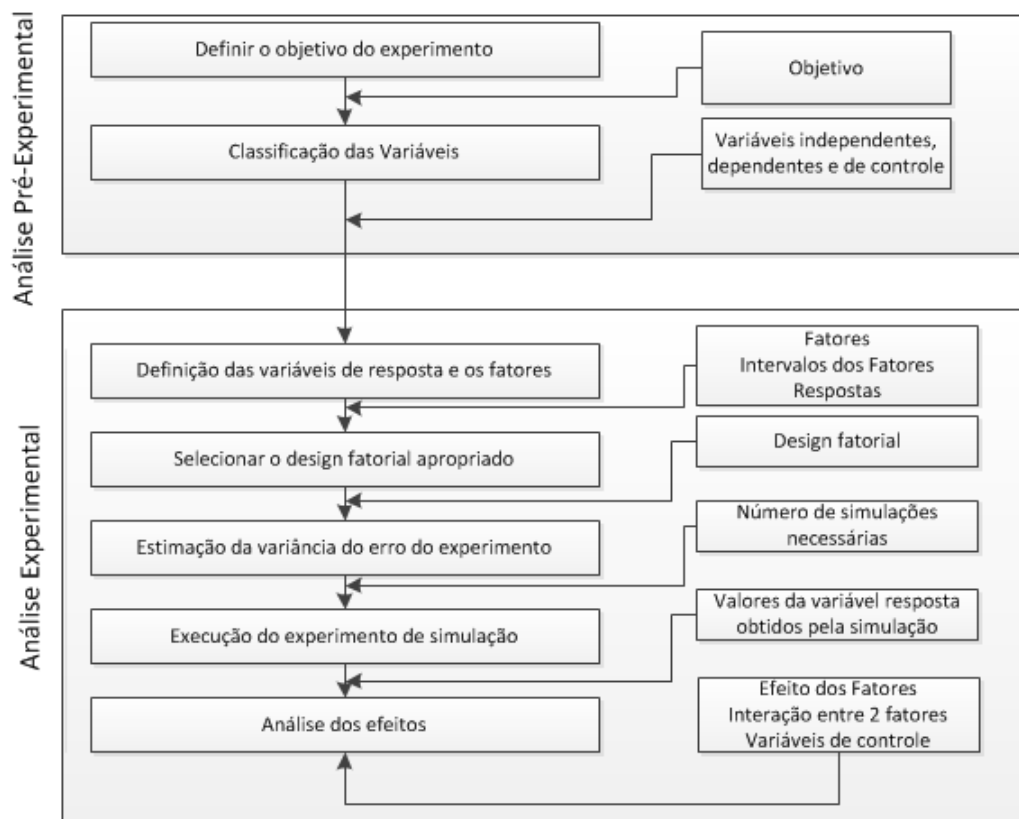


Figura 9: Fases do experimento da simulação.

Fonte: Montgomery (2006). Tradução nossa.

de probabilidade³. Todas as variáveis funcionam como parâmetros de entrada da simulação e através delas será possível projetar o valor da variável resposta, portanto, é fundamental que se defina todas as variáveis e os possíveis valores de entrada, para permitir sua fácil aplicação.

A terceira atividade é a definição das variáveis de resposta e os fatores, também conhecidas como variáveis independentes. As variáveis de resposta são utilizadas para medir o resultado experimental (LAW; KELTON, 2000). Com relação às variáveis independentes, requer-se a definição dos *factor levels*, ou seja, a faixa de valores (WU et al., 2000). Entretanto, deve-se tomar cautela quanto à definição das variáveis independentes como variáveis de controle, pois podem gerar efeitos indesejados no modelo.

A quarta atividade é selecionar o design fatorial apropriado. O design fatorial é uma técnica que analisa as faixas de valores das variáveis independentes de forma sistemática, possibilitando obter resultados válidos e objetivos, além de identificar as interações entre fatores. Os fatores interagem quando o efeito de um fator em relação à resposta é enfraquecido ou reforçado com o valor de um outro fator (FIELD; HOLE, 2003). Em modelos com mais de

³Um exemplo de variável de controle é o número de iterações por simulação.

um fator, o conhecimento sobre as interações existentes é importante para a interpretação dos resultados.

A quinta atividade é a estimação da variância do erro do experimento, que ocorre quando o modelo possui fatores de perturbação incontroláveis ou desconhecidos, ou seja, quando se executa a mesma simulação, com o mesmo conjunto de variáveis, obtêm-se variabilidades nas respostas, o que pode prejudicar a análise de diferentes conjuntos de variáveis. Portanto, Lorscheid (2011) cita que é necessário definir um número de experimentos/execuções necessários para cada simulação. Algumas medidas importantes para analisar a precisão da média e a variabilidade dos dados, são a soma dos quadrados dos erros, variância e o desvio padrão (FIELD; HOLE, 2003).

Já a sexta atividade, é a execução do experimento de simulação, de acordo com os passos definidos anteriormente. Os valores das respostas obtidas são armazenados em uma matriz de valores médios de cada conjunto de N simulações de combinação de fatores, para posterior análise.

E, por fim, o sétimo passo é a análise dos efeitos, que representa o quanto as variáveis dependentes escolhidas em cada simulação explicam a variável resposta, também conhecida pelo valor de R^2 .

3.5.1 ANÁLISE DE SENSIBILIDADE

A análise de sensibilidade consiste em criar um algoritmo para cruzamento automático dos mais diversos atributos e valores, e verificação dos votos projetados pelo modelos abordados. A abordagem utilizada neste trabalho, foi a de obter as características de processos reais e mudar um ou mais valores de variáveis por vez, a fim de verificar a mudança do voto individual projetado.

A partir disto, visualiza-se através de gráficos o impacto de cada variável, e permite verificar qual valor e grupos de valores podem resultar na mudança do voto projetado anteriormente, ou seja, visualizar se houveram mais votos positivos, negativos, ou ainda se a alteração do valor não resultou em grandes mudanças.

O algoritmo é representado pelo fluxograma da figura 10. Neste exemplo, foi utilizado apenas um atributo para análise de sensibilidade, mas podem ser utilizados dois ou mais atributos. Pode-se definir também os valores das variáveis que serão testados.

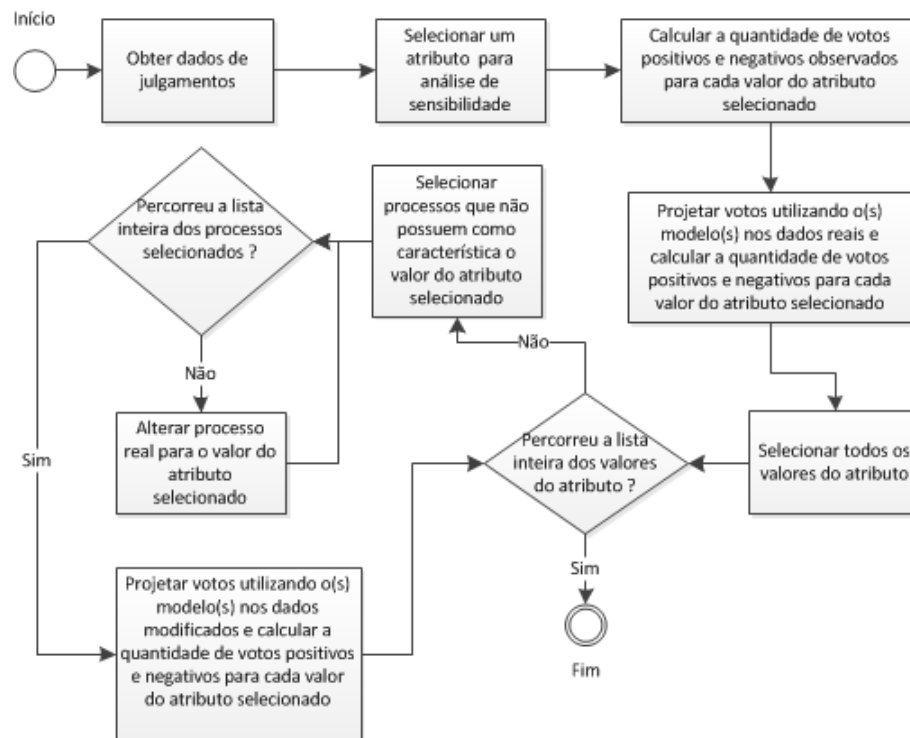


Figura 10: Fluxograma para elaboração da análise de sensibilidade dos dados.

Fonte: Própria.

3.5.2 SIMULAÇÃO UTILIZANDO O NETLOGO

Nesta simulação foi utilizado o ambiente do *NetLogo*. Este ambiente, disponível para *download* em <http://ccl.northwestern.edu/netlogo/>, é um ambiente de modelagem programável multiagente utilizado para simular fenômenos naturais e sociais.

De acordo com a documentação do programa disponível em seu *site*, os usuários do sistema ao criarem um modelo, podem dar instruções a centenas ou milhares de agentes que operam de modo independente. Portanto, torna possível explorar a conexão entre o comportamento a nível de indivíduos e a nível macro, que decorre das interações.

O *NetLogo* ainda permite que os usuários explorem o comportamento dos agentes sob diferentes condições e a partir dos mais diferentes modelos, permitindo também a criação de novas modelagens.

Um exemplo de modelo que se pode citar é o de *Preferential Attachment* disponível na biblioteca de modelos do *NetLogo*. Neste modelo, os agentes são representados por nós, e a cada iteração é criado um novo nó que é conectado a um nó já existente. Quanto mais conexões o nó já existente possui, maior a probabilidade de um novo nó se conectar a ele. O modelo pode

ser visualizado na figura 11

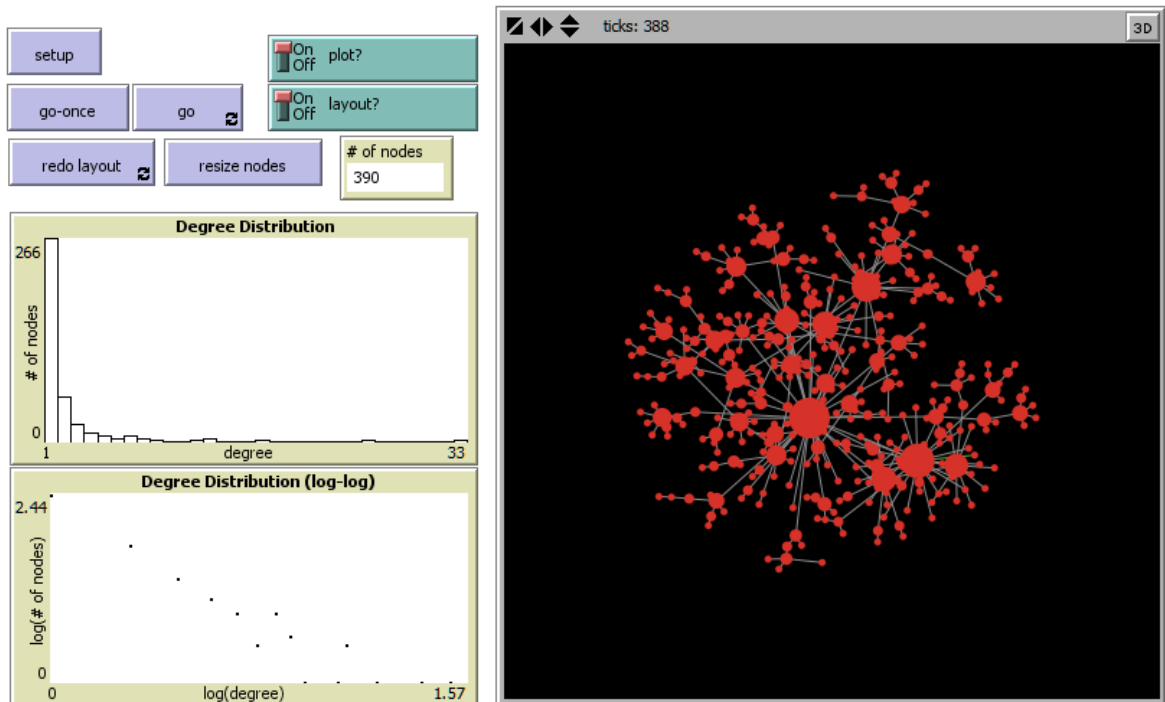


Figura 11: Tela do NetLogo com o modelo “*Preferential Attachment*”.

Fonte: Própria

A aplicação do *NetLogo* neste trabalho, é a possibilidade da criação de agentes representando os membros de um colegiado, e o seus comportamentos guiados de acordo com os modelos criados (árvore de decisão e regressão logística binária), além da inserção de novas variáveis de interação. A partir da aplicação dos modelos, viabiliza-se o teste de inúmeros cenários e os comportamentos dos agentes, de acordo com as características definidas.

4 DADOS COLETADOS

A seguir, serão apresentados os dados coletados que foram utilizados nos modelos sugeridos neste trabalho. Na seção 4.1 é citada a fonte de dados utilizada para coleta das informações e o motivo de sua escolha. Na seção 4.2 são abordados os documentos extraídos, que incluem a fonte, período de extração, e o contexto ao qual os dados pertencem. Na seção 4.3 são apresentadas as informações extraídas dos documentos, que incluem o voto proferido (positivo ou negativo), tipo de resultado (unanimidade ou maioria), informações dos currículos dos membros do colegiado, entre outros. Por fim, na seção 4.4 é apresentada a representação formal das variáveis identificadas a partir dos dados extraídos.

4.1 ESTUDO DE CASO

Como estudo de caso para aplicar o modelo, foram selecionados os colegiados presentes no Superior Tribunal de Justiça. Os colegiados representam grupos de juízes que julgam processos e formam jurisprudências, documentos que consolidam as decisões.

A análise deste estudo de caso é importante pois estudantes, profissionais da área jurídica e cidadãos, independente da metodologia adotada em qualquer pesquisa na área judicial, dependem de informações contidas em textos legais¹, entretanto, a interpretação das opiniões judiciais é bastante trabalhosa (EVANS et al., 2006), ou seja, a possibilidade de identificação de tendências das decisões se faz importante.

Há ainda, a problemática de avaliar se as decisões são tomadas baseadas somente na lei ou se há influência de fatores externos, ou ainda conforme citado por Schubert (1965) “Até que ponto os atos públicos dos juízes são influenciados por suas crenças pessoais?”. Este tema é abordado em trabalhos como o de Hwong (2006), que cita que para um grupo definido pelo autor como *American Legal Realists*, as decisões legais sofrem influências de fatores extralegis, como sociais, políticos e econômicos, e isso pode ser verificado através de análises quantitativas.

¹Os textos legais podem incluir leis, jurisprudências, súmulas, entre outros documentos emitidos pelo poder judiciário.

Entretanto, vale ressaltar, que não há como definir as características e influências que sejam aplicáveis à todas as situações.

Os documentos, disponibilizados *online* pelo poder judiciário, apresentam múltiplas possibilidades de exploração, incluindo a descrição das decisões proferidas pelos magistrados. À navegação temática (assunto do acórdão), agrega-se a navegação social (por exemplo, quais membros do colegiado estão julgando o processo) e temporal (qual a data do julgamento). Esse cenário requer o desenvolvimento de novas metodologias e técnicas para extrair, classificar e referenciar conteúdos.

A modelagem da tomada de decisão entre os membros do poder judiciário não pode ser dissociada do substrato social que a define: as interações que estabelecem preferências, ideias e influências em momentos particulares. Algumas das informações que podem ser retiradas das redes sociais associadas aos magistrados e às decisões² podem ser:

- Assuntos mais julgados no tempo³;
- Membros e decisões mais influentes no tempo⁴;
- Decisões e membros que desempenham o papel de intermediários;
- Relatores e membros do colegiado que participam juntos com mais frequência em acórdãos;
- Filiação e trajetória (acadêmica, social, política, entre outros) de grupos de influência;
- Verificação de quem originou as opiniões/decisões mais propagadas ao longo do tempo⁵.

O conteúdo presente nas decisões não fornece apenas uma riqueza pela sua variedade na definição de conceitos e pessoas associadas a cada decisão, essa base serve de infraestrutura para a descrição de alguns comportamentos como:

- Prováveis decisões de acordo com os assuntos, membros do colegiado, advogados e partes envolvidas nos processos;

²Neste contexto, o conceito de decisões referencia os documentos denominados “Acórdãos”.

³Tendo como base o que foi publicado nos documentos. No caso do Superior Tribunal de Justiça, alguns documentos não possuem o assunto definido.

⁴Como exemplo de influência de um membro de um colegiado, poder-se-ia ter a quantidade de citações nos processos em que foi relator, ou ainda, a quantidade de votos que acompanharam os seus processos. Já um exemplo de decisão influente, seria o número de vezes que um processo foi citado.

⁵O sentido de propagação aqui mencionado, é o de que se uma decisão que é muito citada, foi embasada em outra, logo a decisão original foi fortemente propagada.

- Prováveis ideologias dos membros de um colegiado;
- Quais jurisprudências e bibliografias podem ser citadas em cada julgamento, baseadas em jurisprudências de assuntos similares;
- Descrição dos prováveis motivos que possibilitam a difusão de opiniões.

4.1.1 SUPERIOR TRIBUNAL DE JUSTIÇA

O Superior Tribunal de Justiça do Brasil, segundo o seu próprio *site* (<http://www.stj.jus.br>), pode ser definido como:

“[...] A corte responsável por uniformizar a interpretação da lei federal em todo o Brasil, seguindo os princípios constitucionais e a garantia e defesa do Estado de Direito.

O STJ é a última instância da Justiça brasileira para as causas infraconstitucionais, não relacionadas diretamente à Constituição. Como órgão de convergência da Justiça comum, aprecia causas oriundas de todo o território nacional, em todas as vertentes jurisdicionais não-especializadas [...]” (Disponível em: <http://www.stj.jus.br>. Acesso em 15/12/2013.)

Dentro da hierarquia do poder judiciário brasileiro, o Superior Tribunal de Justiça localiza-se abaixo do Supremo Tribunal Federal, e no mesmo nível de outros tribunais superiores. A figura 12 representa a estrutura do poder judiciário brasileiro, a qual possui três graus de instâncias, ou seja, um mesmo caso pode ser julgado e passar por três tribunais do poder judiciário até obter uma decisão em que não cabe mais recurso. Em alguns casos, pode haver ainda a quarta instância, quando há lesão ou ameaça à Constituição Federal.

4.1.2 DOCUMENTOS PRODUZIDOS

O motivo pela escolha do STJ como estudo de caso, é devido pelo fato do órgão disponibilizar em seu *site* (www.stj.jus.br) as jurisprudências, ou seja, documentos que identificam os membros do colegiado e a decisão proferida, entre outras características.

De acordo com o *site* do Superior Tribunal de Justiça, a jurisprudência é a orientação que resulta de um conjunto de decisões judiciais proferidas num mesmo sentido sobre determinada matéria. No STJ as decisões são disponibilizadas somente após publicação no Diário da Justiça Eletrônico e inclusão na base de dados pela Secretaria de Jurisprudência⁶.

⁶Disponível em: <http://www.stj.jus.br>. Acesso em 15/12/2013.

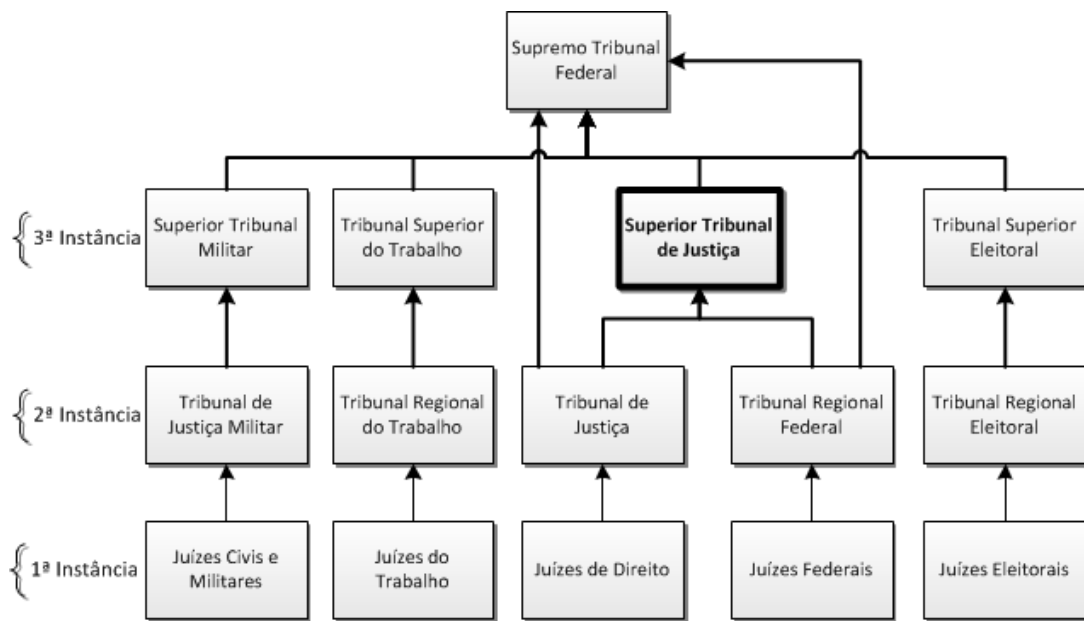


Figura 12: Estrutura do Poder Judiciário Brasileiro.

Fonte: Própria.

As jurisprudências podem ser denominadas também de acórdãos e, no caso do STJ, são disponibilizados em seu *site* os espelhos destes documentos, que representam um resumo do acórdão, conforme figura 13. Abaixo serão apresentados alguns campos presentes neste espelho, segundo o *site* do tribunal.

- **Processo:** Este campo apresenta a sigla da classe processual, o número do processo com a sigla da Unidade da Federação, o nome por extenso da classe processual e o número de registro;
- **Relator:** Neste campo é apresentado o nome do relator e, quando houver, do relator para acórdão ou do revisor;
- **Órgão Julgador:** Este campo indica o órgão julgador que proferiu a decisão;
- **Data do Julgamento:** Este campo indica a data em que ocorreu a sessão de julgamento;
- **Data da Publicação/Fonte:** Este campo indica a data em que ocorreu a publicação do acórdão e também a sua fonte de publicação (Diário da Justiça, Diário da Justiça Eletrônico, Revista do Superior Tribunal de Justiça, entre outros);
- **Ementa:** Este campo apresenta o resumo do conteúdo do acórdão. Sua elaboração é atribuição do Ministro relator ou, quando vencido, do Ministro designado relator para o acórdão;

Documento 4			
Íntegra do Acórdão	Acompanhamento Processual	Resultado sem Formatação	Imprimir/Salvar
Processo			
AgRg no REsp 1135432 / PR AGRAVO REGIMENTAL NO RECURSO ESPECIAL 2009/0069413-8			
Relator(a)			
Ministro CASTRO MEIRA (1125)			
Órgão Julgador			
T2 - SEGUNDA TURMA			
Data do Julgamento			
27/11/2012			
Data da Publicação/Fonte			
DJe 06/12/2012			
Ementa			
<p>TRIBUTÁRIO. COMPLEMENTAÇÃO DE APOSENTADORIA. ENTIDADE FECHADA DE PREVIDÊNCIA PRIVADA. CONTRIBUIÇÕES EFETUADAS SOB A ÉGIDE DA LEI 7.713/88. RESGATE. IMPOSTO DE RENDA. NÃO INCIDÊNCIA. LIQUIDAÇÃO DE SENTENÇA. VIOLAÇÃO A COISA JULGADA. NÃO CONFIGURADA. DIVERGÊNCIA JURISPRUDENCIAL NÃO DEMONSTRADA.</p> <p>1. A jurisprudência desta Corte orienta-se no sentido de se evitar a ocorrência da tributação em duplicidade quando do resgate das contribuições vertidas aos fundos de previdência privada ocorridas durante a vigência da Lei 7.713/88, cujo ônus tenha sido exclusivo do participante. Precedentes.</p> <p>2. Não se verifica a violação à coisa julgada, uma vez que o acórdão recorrido tão somente respaldou a decisão de piso, a qual fixou os parâmetros para a liquidação do julgado em consonância com o entendimento predominante do STJ.</p> <p>3. O Tribunal de origem não exarou cognição a respeito das disposições estabelecidas pelo art. 368 do CPC, configurando, pois, a ausência de prequestionamento. Incidência da Súmula 282/STF.</p> <p>4. Agravo regimental não provido.</p>			
Acórdão			
<p>Vistos, relatados e discutidos os autos em que são partes as acima indicadas, acordam os Ministros da Segunda Turma do Superior Tribunal de Justiça, por unanimidade, negar provimento ao agravo regimental, nos termos do voto do(a) Sr(a). Ministro(a) Relator(a). Os Srs. Ministros Humberto Martins, Herman Benjamin (Presidente), Mauro Campbell Marques e Diva Malerbi (Desembargadora convocada TRF 3a. Região) votaram com o Sr. Ministro Relator.</p>			
Informações Complementares			
Aguardando análise.			

Figura 13: Espelho de um Acórdão do STJ.

Fonte: Site do Superior Tribunal de Justiça

- **Acórdão:** Este campo apresenta o resultado final do julgamento, votação, presença ou ausência de ministro e sustentação oral, se houver;
- **Palavras de Resgate:** Este campo é complementar aos campos Ementa e Outras Informações, destinando-se à inclusão de palavras que possam auxiliar a pesquisa e facilitar o resgate dos acórdãos.

4.1.3 MEMBROS DO COLEGIADO

Conforme contido no *site* do tribunal:

“O Superior Tribunal de Justiça (STJ) é composto por, no mínimo, 33 ministros nomeados pelo presidente da República, após aprovação do Senado Federal. Essa composição é estabelecida pelo artigo 104 da Constituição Federal, segundo o qual o cargo deve ser preenchido por brasileiros com mais de 35 e menos de 65 anos de idade, de notável saber jurídico e reputação ilibada [...]” (Disponível em: <http://www.stj.jus.br>. Acesso em 15/12/2013.)

O Superior Tribunal de Justiça é organizado pelo critério da especialização, em vista disso, existem três seções de julgamento, cada uma delas composta por duas turmas, que analisam e julgam matérias de acordo com a natureza da causa submetida a apreciação. Além das seções e turmas, existem também a Corte Especial, órgão de maior hierarquia no tribunal.

A Corte Especial é dirigida pelo presidente do Tribunal e formada por quinze ministros, escolhidos pelo critério de antiguidade no STJ. Além de exercer funções administrativas, esse órgão julga os processos criminais de competência originária, aqueles que têm início no próprio Tribunal.

Em relação às seções, existem três seções especializadas de julgamento no Superior Tribunal de Justiça (STJ) e cada uma é formada por duas turmas especializadas, e cada turma é integrada por cinco ministros.

A Primeira Seção, composta por ministros da Primeira Turma e da Segunda Turma, aprecia matérias de Direito Público, com destaque para questões administrativas e tributárias, mandados de segurança contra ministros de Estado, entre outros temas.

A Segunda Seção, composta por ministros da Terceira Turma e da Quarta Turma, decide sobre matérias de Direito Privado, examinando questões de Direito Civil e Comercial.

Já a Terceira Seção, composta por ministros da Quinta Turma e da Sexta Turma, julga causas que envolvam matérias de Direito Penal, como habeas-corpus, bem como questões previdenciárias, mandados de segurança contra ministros de Estado e matérias de Direito Público e Privado não cobertas pela Primeira e Segunda seções.

4.2 EXTRAÇÃO DE DOCUMENTOS

Após a escolha do órgão do qual devem ser extraídas as informações e o assunto, foram extraídas de forma automática todas as decisões proferidas em colegiado do *site* do Superior Tribunal de Justiça (STJ), publicadas no período entre 29/09/1997 e 17/11/2011, cujo assunto envolvesse Direito Tributário. Após a extração, foram realizados os devidos tratamentos dos dados conforme seção 3.2, que aborda sobre o pré-processamento dos dados.

A análise textual do documento consistiu em separá-lo nos itens Data de Julgamento,

Tipo de Processo, Identificação do Processo, Relator, Assunto, Certidão de Julgamento, entre outros, para facilitar uma análise mais profunda posteriormente. Os itens foram separados dessa forma, pelo fato do STJ fornecer esses tipos de informações em um documento único. Além das informações contidas explicitamente nas jurisprudências, há outras informações que podem ser filtradas e calculadas, como o tipo de voto proferido (positivo, parcial ou negativo) e a ideologia dos juízes respectivamente.

Ao todo foram extraídas 686.450 jurisprudências, compreendendo o período entre 29/09/1997 e 17/11/2011, com 5 jurisprudências apresentando datas inválidas⁷ por provável erro de digitação do tribunal, e 52 jurisprudências não apresentaram data de julgamento. O intervalo de tempo entre 1997 e 2011 foi escolhido por ser o intervalo em que foram encontradas jurisprudências para *download* através do *site* do tribunal.

Deste total de jurisprudências, foram selecionados os acórdãos referentes ao assunto tributário, que totalizaram 184.704 ocorrências, correspondente a 26,91% do total. Vale ressaltar que 265 documentos não apresentaram nenhum assunto definido.

O critério utilizado para separação das jurisprudências que possuem o conteúdo sobre direito tributário é conter a palavra “tributário” na lista de palavras chaves identificadoras do processo⁸. Outra informação obtida, além dos assuntos dos processos, foi o tipo de resultado das votações, ou seja, maioria ou unanimidade.

Os dez assuntos mais citados nos documentos de direito tributário podem ser verificados na tabela 2. O que se pôde observar foi que existem 789 assuntos diferentes que foram associados à palavra “tributário”, e, o número total de ocorrências de palavras associadas à Direito Tributário foi de 779.467. Neste caso, os assuntos “FGTS” e “Fundo de Garantia por Tempo de Serviço” têm o mesmo significado, entretanto foram citadas em ambas as formas nos processos.

Na figura 14, pode ser verificada a distribuição de jurisprudências do assunto tributário por mês de julgamento, comparada ao número total de jurisprudências de cada mês. Nesta figura, verifica-se que se mantém constante o número de processos de Direito Tributário no tempo, logo, viabilizam-se testes dos modelos com dados futuros. As jurisprudências foram agrupadas por mês de julgamento e foram ignorados os dados anteriores a 2001 por conter poucas jurisprudências.

⁷Na jurisprudência <https://ww2.stj.jus.br/websecstj/cgi/revista/REJ.cgi/ITA?seq=55560&nreg=200801138376&dt=20090401&formato=HTML> é citada a data 30/12/1899.

⁸Em média, cada jurisprudência pôde ser identificada por 5 palavras.

Tabela 2: Número de ocorrências para cada assunto envolvendo direito tributário.

Fonte: Própria.

Assunto	Nº de Ocorrências
Direito Tributário	187.485
Contribuição	85.579
Social	85.179
FGTS	29.089
Fundo De Garantia por Tempo De Serviço	29.014
Correção Monetária Dos Depósitos	24.548
Índice Aplicável	23.527
Previdenciária	19.782
Imposto De Renda	16.148
Pis	13.939

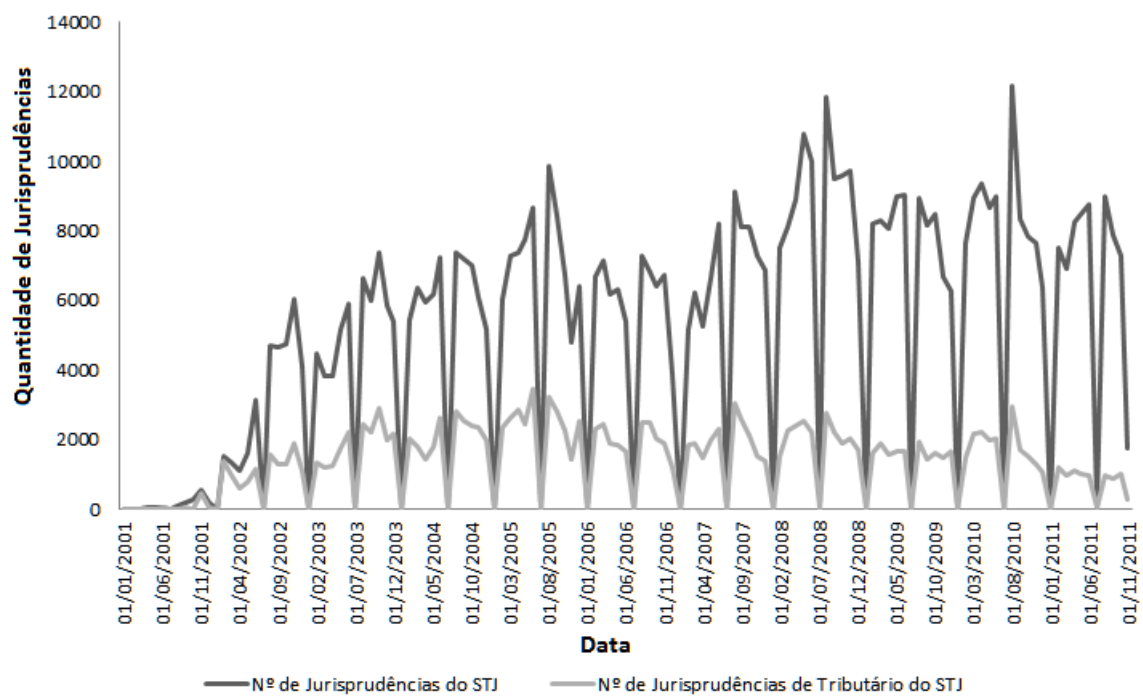


Figura 14: Número de jurisprudências extraídas por data de julgamento sobre Direito Tributário comparado ao total de jurisprudências publicadas pelo STJ.

Fonte: Própria

4.3 CLASSIFICAÇÃO DAS INFORMAÇÕES

A seguir, são tratados dados específicos das jurisprudências sobre Direito Tributário e dados dos juízes que puderam ser extraídos e classificados do Superior Tribunal de Justiça,

entre 29/09/1997 e 17/11/2011.

4.3.1 JURISPRUDÊNCIAS

A quantidade de votos de unanimidade e maioria pode ser verificada na tabela 3. O valor “outros” representa os processos em que os juízes demonstraram serem favoráveis/desfavoráveis por unanimidade a um quesito, e favoráveis/desfavoráveis por maioria em outro, ou ainda, o processo foi retirado de pauta.

Tabela 3: Número de ocorrências para cada tipo de resultado encontrado nas jurisprudências extraídas.

Fonte: Própria.

Tipo de Resultado	Nº de Ocorrências
Unanimidade	179.443
Outros	2.934
Maioria	2.327

Um exemplo de tipo de resultado de unanimidade é:

“A Turma, por unanimidade, negou provimento ao agravo regimental, nos termos do voto do Sr. Ministro Relator. Os Srs. Ministros Castro Filho e Eliana Calmon votaram com o Sr. Ministro Relator. Ausente, justificadamente, o Sr. Ministro Francisco Peçanha Martins.”

Um exemplo de tipo de resultado de maioria é:

“Prosseguindo no julgamento, após o voto vista da Sra. Ministra Laurita Vaz, a Turma, por maioria, deu provimento ao recurso, nos termos do voto da Sra. Ministra Eliana Calmon que lavrará o acórdão, vencido o Sr. Ministro Relator. Votou vencido o Sr. Ministro Francisco Peçanha Martins. Votaram com a Sra. Ministra Eliana Calmon os Srs. Ministros Franciulli Netto e Laurita Vaz. Não participou do julgamento o Sr. Ministro Paulo Medina.”

Dois exemplos de tipo de resultado de outros são:

“Retirado de pauta por indicação do Sr. Ministro Relator. O referido é verdade. Dou fé.”

“Prosseguindo-se no julgamento, após o voto vista do Sr. Ministro Relator, mantendo seu voto, a Turma, por unanimidade, não conheceu do recurso da empresa e, por maioria, vencido o Sr. Ministro Relator, deu provimento ao recurso do Estado do Rio de Janeiro, nos termos do voto da Sra. Ministra Eliana Calmon, que lavrará o acórdão. Votaram com a Sra. Ministra Eliana Calmon os Srs. Ministros Paulo Gallotti e Francisco Falcão. Não participaram do julgamento os Srs. Ministros Franciulli Netto, Laurita Vaz e Paulo Medina.”

Quanto ao resultado final da decisão dos processos envolvendo Direito Tributário, foram classificados em negativo, positivo, parcial e não encontrado. Nesse caso, foram ignorados 70 processos do tipo resultado maioria e 3.304 do tipo resultado unanimidade, por apresentarem mais de uma parte autora, o que dificultava a extração, uma vez que era proferido um voto para cada pedido de cada autor, que pode ser escrito de inúmeras maneiras. Os resultados obtidos podem ser verificados na tabela 4.

Vale ressaltar que os dados da regressão, tanto a amostra de teste, quanto a amostra de treinamento, se limitaram a utilizar apenas os processos que tiveram sua decisão final com voto positivo ou negativo, que totalizaram 1939 processos.

Tabela 4: Número de ocorrências para cada tipo de decisão final encontrado nas jurisprudências sobre direito tributário. A decisão se refere ao recurso/processo enviado para o Superior Tribunal de Justiça, em que o colegiado decide se vai aceitá-lo e dar provimento. Foram separadas as quantidades de decisões por processos em que ocorreu unanimidade e processos em que ocorreu votação por maioria.

Fonte: Própria.

Decisão	Nº de Ocorrências - Maioria	Nº de Ocorrências - Unanimidade
Negou provimento	1.381	132.054
Deu provimento	558	23.236
Deu parcial provimento	313	20.386
Erro	5	463

Um exemplo de decisão negativa é:

“A Turma, por unanimidade, negou provimento ao agravo regimental, nos termos do voto do Sr. Ministro Relator. Os Srs. Ministros Castro Meira, Francisco Peçanha Martins, Eliana Calmon e Franciulli Netto votaram com o Sr. Ministro Relator.”

Um exemplo de decisão positiva é :

“Prosseguindo no julgamento, após o voto vista da Sra. Ministra Laurita Vaz, a Turma, por maioria, deu provimento ao recurso, nos termos do voto da Sra. Ministra Eliana Calmon que lavrará o acórdão, vencido o Sr. Ministro Relator. Votou vencido o Sr. Ministro Francisco Peçanha Martins. Votaram com a Sra. Ministra Eliana Calmon os Srs. Ministros Franciulli Netto e Laurita Vaz. Não participou do julgamento o Sr. Ministro Paulo Medina.”

4.3.2 JUÍZES

A seguir, serão detalhados alguns tipos de dados extraídos sobre juízes de diferentes fontes de dados.

No *site* do tribunal está disponibilizado o currículo resumido de cada juiz que já trabalhou no órgão, conforme figura 15. Dentre as informações disponibilizadas, foram obtidos o grau de escolaridade, faculdade cursada na graduação, ano de nascimento, estado que nasceu e o período de exercício no tribunal.

Dados Pessoais

Nascimento: 14 de outubro de 1956 - São Paulo/SP

Formação Acadêmica

- Bacharel em Direito pela Faculdade de Direito da Universidade de São Paulo.
- Mestre em Direito Processual pela Faculdade de Direito da Universidade de São Paulo.
- Doutora em Direito Processual pela Faculdade de Direito da Universidade de São Paulo.
- Especialização em Direito Processual Penal da Faculdade de Direito da Pontifícia Universidade Católica de São Paulo.
- Especialização em Direito Penal Econômico e Europeu pela Faculdade de Coimbra, Instituto de Direito Penal Econômico Europeu e IBCCRIM.

Funções Atuais

- Ministra do Superior Tribunal de Justiça desde 9/8/2006;
- Presidente da Terceira Seção do STJ, biênio: 8/2011 a 7/2013;
- Membro da Comissão de Jurisprudência do STJ;
- Membro da Corte Especial do STJ;

Figura 15: Currículo de uma ministra disponível no *site* do tribunal.

Fonte: *Site* do STJ.

Diante dos diversos atributos existentes nos documentos, para o cálculo da ideologia⁹ foram utilizados os trechos do processo que identificam se o voto de cada membro do colegiado foi vencido ou venceu, o que permite obter o valor da ideologia de cada um, após aplicação do algoritmo *W-NOMINATE* de Keith Poole e Howard Rosenthal.

Para calcular o valor da ideologia foi realizada uma adaptação do exemplo de Poole

⁹Neste trabalho o termo citado é ideologia, por ser o mesmo citado por Poole (2005), entretanto, este termo deve ser considerado apenas no sentido de possuir opinião favorável ou contrária aos processos.

e Rosenthal na página [http : //www.voteview.com/icpsr.htm](http://www.voteview.com/icpsr.htm). A interpretação dos valores utilizados por Poole e Rosenthal, conforme exemplo abaixo que demonstra uma linha de dados que é lida pelo algoritmo, pode ser verificada na tabela 5.

1079991099 0USA 200 BUSH 111111111999119

Tabela 5: Significados dos valores utilizados no algoritmo do W-NOMINATE. Exemplo utilizando dados de senadores dos Estados Unidos da América.

Fonte: Própria.

Valor	Descrição
107	Número do Congresso
99910	Código de identificação do deputado/senador
99	Código de identificação do estado
0	0 caso Senado ou Número do Distrito do Congresso
USA	Nome do Estado do deputado/senador.
200	Código do Partido. 100 = Democrata; 200= Republicano
Bush	Nome do deputado/senador
11...99:	Dados das votações. Votos positivos (1,2,3); votos negativos (4,5,6); votos faltantes (7,8,9); não presente na legislatura (0)

Entretanto, o algoritmo é voltado para analisar a ideologia dos participantes de votações do Senado e Congresso dos Estados Unidos da América. Como nas votações do Senado e Congresso todos podem participar, e nas votações em colegiado do Superior Tribunal de Justiça somente um grupo de juízes pode participar por processo, foi realizada uma adaptação para calcular o valor da ideologia.

Adaptando os dados de entrada para os dados do STJ, defini-se que os dados com “-”, representam dados que não são possíveis de determinar por fazerem parte de outro contexto, portanto para todos os juízes foi considerado o mesmo valor. Cada membro do colegiado terá uma linha, e cada coluna dos dados de votações corresponde a um processo, ou seja, se foram analisados 600 processos, terão 600 colunas, e como cada processo é julgado por em média 5 juízes, logo os demais juízes que não participaram receberão o valor “0” nesta coluna. A interpretação dos valores contidos abaixo pode ser verificada na tabela 6.

0009991099 0BRA 200 JOSE 111111111999119

Tabela 6: Significados dos valores utilizados no algoritmo do *W-NOMINATE*. Exemplo utilizando dados de juízes.

Fonte: Própria.

Valor	Descrição
000	-
99910	Código de identificação do membro do colegiado (gerado aleatoriamente)
99	-
0	-
BRA	-
200	-
Bush	Nome do membro do colegiado
11...99:	Dados das votações. Votos positivos (1,2,3); votos negativos (4,5,6); faltou/impedido na votação (7,8,9); não compunha o colegiado julgador (0)

Ao ser calculada a ideologia, de todos os membros do colegiado nos processos de Direito Tributário em casos em que houve votação por maioria, o algoritmo *W-NOMINATE* conseguiu prever 78.6% dos votos positivos e 94% dos votos negativos. Para o cálculo, foram ignorados dados de 23 membros e 14 votos por não serem muito representativos para o cálculo¹⁰. Dos dados que foram aproveitados, estão 27 membros e 1.927 votos. Os valores da ideologia podem ser verificados na tabela 7.

4.4 REPRESENTAÇÃO FORMAL DOS DADOS

A representação formal dos conjuntos de dados depende, em primeiro lugar, do número de variáveis ou características de interesse. Quando se dispõem de n valores de cada uma de p variáveis, os dados podem ser organizar em uma matriz $n \times p$, em que se convencionou representar as variáveis nas colunas e os indivíduos ou unidades estatísticas nas linhas. Esta matriz é denominada matriz dos dados e é, em geral, representada por X :

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (14)$$

¹⁰Membros que possuíam pouca quantidade de processos julgados na amostra não puderam ter sua ideologia calculada.

Tabela 7: Valor da ideologia dos membros de um colegiado, obtido através de casos de direito tributário com votos por maioria.

Fonte: Própria.

Membro	Ideologia	Membro	Ideologia	Membro	Ideologia
A1	0.109	J1	-0.490	S1	-0.558
B1	0.154	K1	0.305	T1	0.976
C1	-0.427	L1	0.119	U1	-0.087
D1	0.334	M1	0.431	V1	0.948
E1	0.695	N1	0.866	W1	-0.141
F1	0.584	O1	-0.894	X1	-0.130
G1	0.597	P1	-0.951	Y1	-1.000
H1	-0.281	Q1	0.378	Z1	0.293
I1	0.973	R2	0.936	A2	-0.232

Adaptando esta matriz ao contexto do trabalho, o valor de n representa o número de processos julgados por membro, e p o número de características extraídas tanto do processo quanto de seus julgadores. Apesar do estudo de caso deste trabalho estar relacionado aos juízes, estes serão tratados como membros de um colegiado.

4.4.1 DADOS DOS MEMBROS DO COLEGIADO

Antes de representar as características dos membros do colegiado, é necessário citar o conjunto ao qual eles pertencem, no caso um órgão e os processos vinculados ao mesmo, que apesar de pertencerem ao órgão, são julgados pelos membros.

Conjunto de membros do colegiado pertencente a um determinado órgão

$$J = \{j_1, j_2, j_3, \dots, j_{nj}\}$$

$$\{j_i \in J | 0 < i < nj\}$$

$nj = n^o$ total de membros do colegiado

Conjunto de processos pertencentes a um determinado órgão

$$P = \{p_1, p_2, p_3, \dots, p_{np}\}$$

$$\{p_j \in P | 0 < j < np\}$$

$np=n^o$ total de processos

4.4.2 CARACTERÍSTICAS DOS MEMBROS DO COLEGIADO

A seguir, serão representadas as características dos membros do colegiado que foram extraídas e puderam ser quantificadas. Algumas das características tiveram seus intervalos de valores reduzidos em poucas categorias, para facilitar a aplicação nos modelos.

Os atributos raça, gênero sexual, valor da ideologia e ano de nascimento foram escolhidos por serem utilizados no modelo de Cameron e Cummings (2003). Já os atributos faculdade cursada e estado que o membro nasceu foram utilizados por serem citados no trabalho de Hwong (2006). Por fim, os demais atributos que incluem o grau de escolaridade, tipo de faculdade e membro como relator, foram utilizados para testar se estes possuem influência na projeção dos votos. Além de todos os atributos citados estarem disponíveis para extração.

Raça do membro

Classificada em duas categorias, branco e negro, a mesma abordagem adotada por Cameron e Cummings (2003).

$$R = \{0, 1\};$$

$$\{r_i \in R | 0 < i < nj\};$$

$$\begin{cases} if(r_i = 0)\{is-negro(r_i)\}; \\ if(r_i = 1)\{\sim is-negro(r_i)\}; \end{cases}$$

Gênero sexual do membro

Classificado em duas categorias, feminino e masculino.

$$S = \{0, 1\}$$

$$\{s_i \in S | 0 < i < nj\}$$

$$\begin{cases} if(s_i = 0)\{is-mulher(s_i)\}; \\ if(s_i = 1)\{\sim is-mulher(s_i)\}; \end{cases}$$

Valor da ideologia do membro

O valor da ideologia do membro é calculado através do algoritmo *W-NOMINATE* e de acordo com os processos que julgou. O valor para definição de conservador e liberal, se dá conforme citado por Cameron e Cummings (2003).

$$I = \{-1, \dots, 1\}$$

$$\{i_i \in I \mid 0 < i < nj\}$$

$$\begin{cases} if(i_i < -0.39)\{is-liberal(i_i)\}; \\ if(i_i > 0.34)\{is-conservador(i_i)\}; \end{cases}$$

Ano de nascimento do membro

Cada elemento corresponde a um ano.

$$A = \{1899, \dots, 1949\}$$

$$\{a_i \in A \mid 0 < i < nj\}$$

Faculdade cursada pelo membro na graduação

Cada elemento corresponde a uma universidade.

$$F = \{UERJ, UNB, UFRGS, UFMG, USP, \dots, UFPR\}$$

$$\{f_i \in F \mid 0 < i < nj\}$$

Grau de escolaridade máximo do membro

Classificado em quatro categorias. Considerado uma variável ordinal que varia entre 0 e 3. Graduação (0), Especialização (1), Mestrado (2) e Doutorado (3).

$$G = \{0, 1, 2, 3\}$$

$$\{g_i \in G \mid 0 < i < nj\}$$

Estado que o membro nasceu

Cada elemento corresponde a um estado do Brasil. Entretanto, pode ocorrer que um membro do colegiado tenha nascido fora do Brasil, representado portanto pelo valor “EST”, que indica estrangeiro.

$$EJ = \{AC, AL, AP, AM, BA, CE, DF, \dots, TO, EST\}$$

$$\{e_{ji} \in EJ \mid 0 < i < nj\}$$

Tipo de faculdade de graduação cursada pelo membro

Classificado em duas categorias, pública e privada.

$$TF = \{0, 1\}$$

$$\{tf_i \in TF | 0 < i < nj\}$$

$$\begin{cases} if(tf_i = 0)\{\sim \text{is-publica}(tf_i)\}; \\ if(tf_i = 1)\{\text{is-publica}(tf_i)\}; \end{cases}$$

Membro como relator do processo

Define se o membro foi o relator do processo.

$$RE = \{0, 1\}$$

$$\{re_{ij} \in RE | 0 < i < nj; 0 < j < np\}$$

$$\begin{cases} if(re_{ij} = 0)\{\sim \text{is-relator}(re_{ij})\}; \\ if(re_{ij} = 1)\{\text{is-relator}(re_{ij})\}; \end{cases}$$

4.4.3 CARACTERÍSTICAS DOS COLEGAS DE JULGAMENTO

A seguir, serão apresentados os diferentes tipos de dados vinculados aos colegas votantes de cada membro, e, em cada processo.

Os atributos indicador raça, indicador gênero sexual, indicador ideologia e indicador ano de nascimento foram escolhidos por serem utilizados no modelo de Cameron e Cummings (2003). Por fim, os demais atributos, que incluem o indicador grau de escolaridade e o indicador tipo de faculdade, foram utilizados para testar se estes possuem influência na projeção dos votos. Além de todos os atributos citados estarem disponíveis para extração.

Indicador da raça

Média da raça dos colegas do membro que estão votando no processo.

$$MR = \{0, \dots, 1\}$$

$$\{mr_{ij} \in MR | 0 < i < nj; 0 < j < np\}$$

Indicador do gênero sexual

Média do gênero sexual dos colegas do membro que estão votando no processo.

$$MS = \{0, \dots, 1\}$$

$$\{ms_{ij} \in MS | 0 < i < nj; 0 < j < np\}$$

Indicador da ideologia

Média da ideologia dos colegas do membro que estão votando no processo.

$$MI = \{-1, \dots, 1\}$$

$$\{mi_{ij} \in MI | 0 < i < nj; 0 < j < np\}$$

Indicador do ano de nascimento

Média do ano de nascimento dos colegas do membro que estão votando no processo.

$$MA = \{1899, \dots, 1949\}$$

$$\{ma_{ij} \in MA | 0 < i < nj; 0 < j < np\}$$

Indicador do tipo de faculdade

Média do tipo de faculdade dos colegas do membro que estão votando no processo.

$$MTF = \{0, \dots, 1\}$$

$$\{mtf_{ij} \in MTF | 0 < i < nj; 0 < j < np\}$$

Indicador do grau de escolaridade

Média do grau de escolaridade dos colegas do membro que estão votando no processo.

$$MGE = \{0, \dots, 3\}$$

$$\{mge_{ij} \in MGE | 0 < i < nj; 0 < j < np\}$$

4.4.4 DADOS DOS MEMBROS NOS PROCESSOS

A seguir, será apresentado o tipo de dado vinculado aos membros votantes, no caso o voto que o membro votante i proferiu no processo j .

$$V = \{0, 1\}$$

$$\{v_{ij} \in V | 0 < i < nj; 0 < j < np\}$$

$$\begin{cases} if(v_{ij} = 0) \{\sim \text{is-voto-positivo}(v_{ij})\}; \\ if(v_{ij} = 1) \{\text{is-voto-positivo}(v_{ij})\}; \end{cases}$$

4.4.5 CARACTERÍSTICAS DO PROCESSO

A seguir, serão apresentados os diferentes tipos de dados vinculados aos processos. Os atributos estado de origem do processo, tipo de réu e tipo de autor foram escolhidos por serem utilizados no modelo de Martin et al. (2004). Por fim, os demais atributos, voto do relator, presidente e colegiado, foram utilizados para testar se estes possuem influência na projeção dos votos. Além de todos os atributos citados estarem disponíveis para extração.

Estado de origem do processo

Cada elemento corresponde a um estado do Brasil.

$$EP = \{AC, AL, AP, AM, BA, CE, DF, \dots, TO\}$$

$$\{ep_j \in EP | 0 < j < np\}$$

Tipo de réu do processo

Classificado em três categorias, pessoa física, pessoa jurídica e Estado.

$$TR = \{Pessoa\ física, Pessoa\ jurídica, Estado\}$$

$$\{tr_j \in TR | 0 < j < np\}$$

Tipo de autor do processo

Classificado em três categorias, pessoa física, pessoa jurídica e Estado.

$$TA = \{Pessoa\ física, Pessoa\ jurídica, Estado\}$$

$$\{ta_j \in TA | 0 < j < np\}$$

Voto do relator

Voto que o relator do processo proferiu.

$$VR = \{0, 1\}$$

$$\{vr_j \in VR | 0 < j < np\}$$

$$\begin{cases} if(vr_j = 0) \{ \sim is-voto-relator-positivo(vr_j); \} \\ if(vr_j = 1) \{ is-voto-relator-positivo(vr_j); \} \end{cases}$$

Presidente

Presidente do Brasil na época que o processo foi julgado. Diante a amostra dos dados

analisada, obtiveram-se somente três presidentes, Fernando Henrique Cardoso, Luiz Inácio Lula da Silva e Dilma Rousseff.

$$P = \{FernandoHenrique, Lula, Dilma\}$$

$$\{p_j \in P | 0 < j < np\}$$

Colegiado

Colegiado que o membro está julgando um processo. Um membro pode fazer partes de vários colegiados, mas um processo é julgado por somente um colegiado.

$$C = \{PrimeiraTurma, SegundaTurma, PrimeiraSeção, CorteEspecial\}$$

$$\{c_{ij} \in C | 0 < i < nj; 0 < j < np\}$$

$$\{c_j \in C | 0 < j < np\}$$

5 MODELO SUGERIDO

Após a coleta dos dados e a devida classificação dos atributos, os dados foram aplicados aos modelos de regressão e à árvore de decisão, a fim de obter os coeficientes para a equação e as regras da árvore de decisão, respectivamente.

Na seção 5.1 são apresentados os atributos utilizados nos modelos e as possíveis influências entre as variáveis. Na seção 5.2 é apresentado o modelo de regressão logística com seus coeficientes e resultados. Na seção 5.3 é apresentado o modelo de árvore de decisão e os resultados obtidos para avaliação. Por fim, na seção 5.4 é apresentado o modo como a simulação foi estruturada.

5.1 ATRIBUTOS

A primeira tarefa para o desenvolvimento do modelo de tomada de decisão aplicado ao domínio legal, é a análise dos atributos que os acórdãos/membros do colegiado possuem e o que os modelos estudados apresentam. Com os atributos definidos, foram mapeadas as relações existentes entre as entidades possuidoras dos atributos, de acordo com a figura 16.

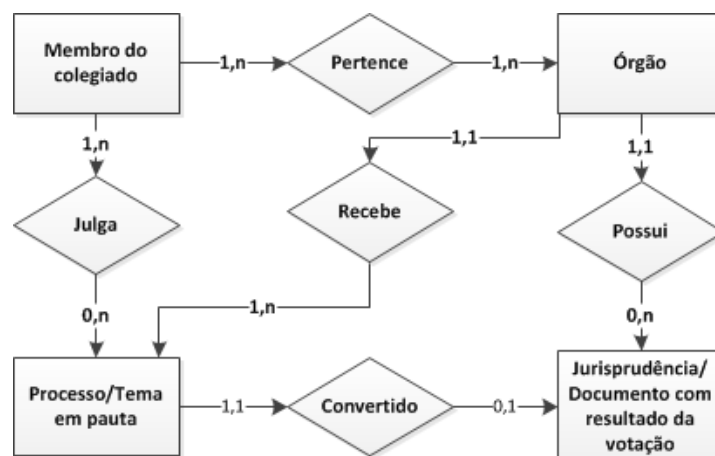


Figura 16: Diagrama entidade relacionamento que representa o contexto do trabalho.

Fonte: Própria.

A lista das entidades e suas respectivas variáveis podem ser verificadas na tabela 8. Esta análise é importante para compreender o contexto no qual as variáveis estão inseridas.

Tabela 8: Variáveis do modelo. As palavras “membro” representam os membros que podem fazer parte de um colegiado.

Fonte: Própria.

Entidade	Variável	Tipo Variável
Membro	Ano de nascimento	Discreta
	Relator do processo	Dicotômica (Sim, Não)
	Estado do membro	Nominal
	Gênero sexual	Dicotômica (Homem, Mulher)
	Grau de escolaridade	Ordinal
	Ideologia	Contínua
	Raça	Dicotômica (Branco, Negro)
	Tipo de universidade cursada	Dicotômica (Pública, Particular)
	Universidade cursada	Nominal
	Processo	Estado do processo
Tipo de réu		Nominal
Tipo de autor		Nominal
Voto do relator		Dicotômica (Positivo, Negativo)
Colegiado		Nominal
Presidente do Brasil		Nominal
Colegas	Indicador ano de nascimento	Contínua
	Indicador grau de escolaridade	Contínua
	Indicador raça	Contínua
	Indicador tipo de universidade cursada	Contínua
	Indicador gênero sexual	Contínua
	Indicador ideologia	Contínua

Para o cálculo das variáveis cujo nome contém a palavra “Indicador” foi utilizada a média simples dos atributos dos colegas, entretanto, para trabalhos futuros pode-se utilizar outro tipo de cálculo.

Na tabela 9 constam as comparações dos atributos utilizados no modelo de Cameron e Cummings (2003) e no modelo sugerido. As variáveis não utilizadas no modelo sugerido são a “Era doutrinária” e o “Indicador Voto”. A variável “Era doutrinária” é caracterizada no

modelo de Cameron e Cummings (2003) pelos períodos que houveram decisões significantes com relação ao tema “Ações Afirmativas”, sendo elas *Pre-Bakke* (até 1978), *Post-Bakke* (1978-1984), *Post-Stotts* (1984-1989), e, *Post-Croson* (1989-1995). Já a variável “Indicador Voto” não é utilizada pois o modelo sugerido tem como objetivo projetar votos, logo, se já for de conhecimento os votos dos membros não seria necessário projetá-los.

Quanto ao assunto, no modelo sugerido é utilizado o assunto tributário, já no modelo de Cameron e Cummings (2003), os assuntos utilizados são as palavras-chaves ligadas ao assunto emprego, dentre elas incluem-se contratações, promoções de cargos e sindicato.

Dentre os atributos apresentados, vale ressaltar os atributos utilizados no modelo de Cameron e Cummings e que puderam ser identificados nas jurisprudências brasileiras, que incluem a Ideologia, Gênero Sexual, Ano de Nascimento, Indicador Gênero Sexual, Indicador Ideologia, Indicador Ano de Nascimento, Raça, Indicador Raça e Estado Processo, em que este último equivale à corte de origem, citada no trabalho de Cameron e Cummings. Estes atributos são importantes, pois são utilizados na comparação do modelo sugerido, para verificar se apresentou um ganho na projeção dos votos corretamente.

Com a definição das entidades que pertencem ao processo de formação de uma jurisprudência, foram mapeadas as relações existentes entre as mesmas, e as informações que podem possuir um grau de influência na sentença, conforme figura 17.

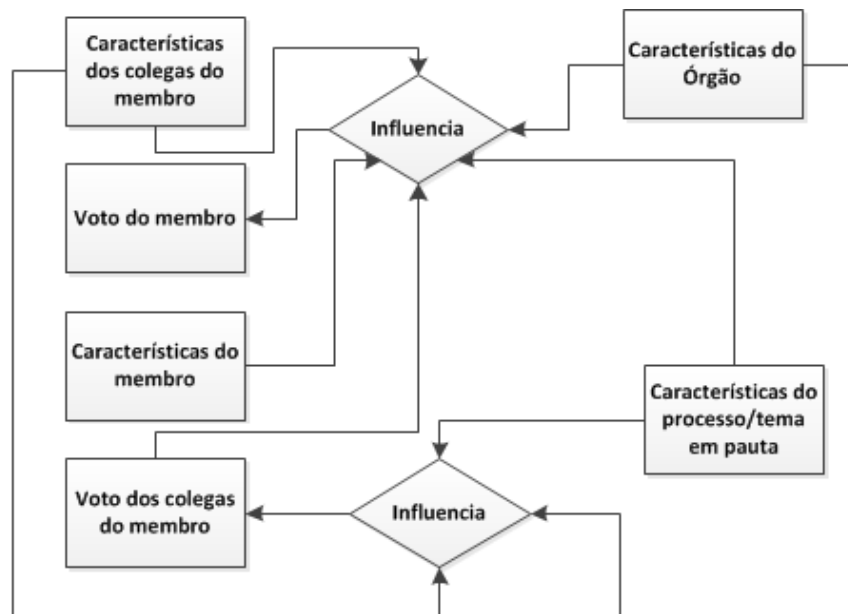


Figura 17: Diagrama entidade relacionamento que representa as possíveis influências entre as variáveis.

Fonte: Própria.

Tabela 9: Tabela comparativa das variáveis utilizadas no modelo de Cameron e Cummings e no modelo sugerido neste trabalho.

Fonte: Própria.

Variável	Cameron e Cummings	Sugerido
Ano de nascimento	X	X
Relator do processo		X
Estado do membro do colegiado		X
Gênero sexual	X	X
Grau de escolaridade		X
Ideologia	X	X
Raça	X	X
Tipo de universidade cursada		X
Universidade cursada		X
Estado do processo	X	X
Tipo de réu		X
Tipo de autor		X
Voto do relator		X
Colegiado		X
Presidente do Brasil		X
Indicador ano de nascimento	X	X
Indicador grau de escolaridade		X
Indicador raça	X	X
Indicador tipo de universidade cursada		X
Indicador gênero sexual	X	X
Indicador ideologia	X	X
Indicador voto	X	
Era doutrinária	X	
Assunto	X	X

5.2 REGRESSÃO LOGÍSTICA

Para projetar votos (positivos ou negativos), foi utilizada a regressão logística binária. Como variável dependente, foi utilizada a variável voto e como variáveis explicativas/independentes, foram utilizados todos os outros atributos citados na seção 5.1.

Antes de executar a regressão logística e obter o modelo, deve-se dar um tratamento especial às variáveis categóricas, o que é denominado codificação das categorias da variável qualitativa. Quando as variáveis são dicotômicas, ou seja, possuem somente dois valores (raça branca e negra, como utilizado no trabalho de Cameron e Cummings (2003)) para estimar os parâmetros do modelo, essas variáveis devem ser substituídas por valores numéricos, por exemplo, 0 para raça branca e 1 para raça negra. Entretanto, quando a variável categórica possui como valores mais de duas categorias, é necessário codificá-las conforme mostrado no Apêndice A, para permitir a inserção dos valores por categoria na equação da regressão. Um exemplo de codificação para a variável “Presidente” pode ser verificado na tabela 10.

Tabela 10: Valores da codificação da variável categórica “Presidente”.

Fonte: Própria.

Presidente	Categoria
Dilma Rousseff	1
Fernando Henrique Cardoso	2
Luiz Inácio Lula da Silva	3

Após converter as variáveis categóricas, foram calculados os coeficientes que forneceram ao modelo a máxima verossimilhança, obtendo-se a equação que permite determinar a probabilidade de um voto positivo ocorrer. Os coeficientes demonstram o efeito de cada uma das variáveis independentes na variável dependente.

A equação abaixo utilizou o método de regressão logística com os respectivos coeficientes das variáveis para estimar a probabilidade de um voto positivo ocorrer, com valor de corte 0,5 (>0,5 voto positivo e <0,5 voto negativo). Optou-se por utilizar somente duas variáveis para exemplificar, mas o modelo completo deve contemplar todos os coeficientes das variáveis contidas no Apêndice B, representados pelos valores B , conforme citado na subseção 2.1.4, fórmula 2.

$$P(1) = \frac{1}{1 - e^{-(-0,403 + Relator * 0,182 + Presidente(1) * 0,352 + Presidente(2) * -0,703)}} \quad (15)$$

Se, em um caso hipotético, o indivíduo votante for Relator e a presidente vigente for a Dilma Rousseff, a probabilidade do voto positivo ocorrer será:

$$P(1) = \frac{1}{1 - e^{-(0,403 + 1 \cdot 0,182 + 1 \cdot 0,352 + 0 \cdot -0,703)}} = 0,5327$$

Logo, se $P(1) > 0,5$ a probabilidade do voto positivo ocorrer é maior que a probabilidade do voto negativo, e infere-se que o voto seja positivo neste caso.

Para avaliar o modelo, devem ser verificados os dados apresentados na tabela 11. A amostra de treinamento representa os dados selecionados através da técnica de amostragem estratificada para treinamento do modelo, que é composto de 2/3 do total de registros de votos positivos e 2/3 do total de registros de votos negativos. Já a amostra de teste representa a amostra que foi utilizada para validar o modelo (1/3 do total de dados de cada tipo de voto).

Tabela 11: Resultados obtidos na aplicação dos dados de treinamento no modelo de regressão, e na aplicação do modelo obtido na amostra de teste. Neste teste foram utilizadas todas as variáveis.

Fonte: Própria.

	Amostra de Treinamento	Amostra de Teste
Instâncias classificadas corretamente	65,7%	65,4%
Votos negativos classificados como positivos	14,68%	15,22%
Votos positivos classificados como negativos	68,07%	68,06%

De acordo com a tabela 11, foi obtida uma porcentagem similar de acertos para a amostra de treinamento e teste, e, uma baixa previsão para os votos positivos. Ainda foram obtidos outros dados como os abaixo¹, que estatisticamente não são considerados satisfatórios, entretanto são aceitáveis, devido ao contexto abordado.

- Valor de log verossimilhança inicial: 10459,574;
- Valor de log verossimilhança final: 9540,126;
- R^2 de Cox & Snell: 0,109;
- R^2 de Nagelkerke: 0,149.

Foi realizado o mesmo teste utilizando somente as variáveis do modelo de Cameron e Cummings (2003), para verificar se as variáveis sugeridas por este trabalho contribuem

¹O significado de cada valor pode ser verificado na subseção 3.3.3 deste trabalho.

de maneira mais significativa que as variáveis citadas por Cameron e Cummings (2003). O resultado pode ser verificado na tabela 12.

Tabela 12: Resultados obtidos na aplicação dos dados de treinamento no modelo de regressão, e na aplicação do modelo obtido na amostra de teste. Neste teste foram utilizadas as variáveis do modelo de Cameron e Cummings.

Fonte: Própria.

	Amostra de Treinamento	Amostra de Teste
Instâncias classificadas corretamente	62%	61,1%
Votos negativos classificados como positivos	13,06%	14,21%
Votos positivos classificados como negativos	81,02%	81,57%

De acordo com a tabela 12, foi obtida uma porcentagem similar de acertos para a amostra de treinamento e teste, e, uma baixa previsão para os votos positivos. Além do desempenho ter sido menor, comparado ao modelo utilizando todas as variáveis deste trabalho, foram obtidos outros dados como:

- Valor de log verossimilhança inicial: 10459,574;
- Valor de log verossimilhança final: 9950,858;
- R^2 de Cox & Snell: 0,062;
- R^2 de Nagelkerke: 0,085.

Após analisar os dados do Apêndice B, verifica-se que é possível ajustar o modelo, ao eliminar algumas variáveis que não contribuem para a projeção. Portanto, devem ser consideradas algumas condições citadas por Field (2005) abaixo.

“1.Verificar a tabela denominada “Variáveis na equação” a fim de obter quais variáveis são significativas para prever a saída.

2.Para cada variável no modelo, verificar a estatística de *Wald* e sua significância (que deve ser menor que 0,05). Ainda deve-se utilizar o valor do $Exp(B)$ para a interpretação. Caso o valor seja maior do que 1, então quando o previsor aumentar, as chances de uma saída ocorrer aumentam. Já quando o valor é menor do que 1, indica que quando o previsor aumenta, as chances de uma saída ocorrer diminuem. Para a interpretação acima mencionada ser confiável, o intervalo de confiança de $Exp(B)$ não deve ultrapassar 1.

3.Conferir a tabela denominada “Variáveis que não estão na equação” para obter as variáveis que não preveem de forma significativa a saída.” (FIELD, 2005)

Conforme regra 1, quando são analisado os valores das variáveis presentes na equação, verifica-se que poucas variáveis apresentam um valor de *Sig.* inferior a 0,05, que incluem o Voto do Relator, Faculdade, Estado do Processo, Colegiado, Tipo de Réu e Presidente.

Conforme regra 2, as variáveis que atendem este quesito são o Relator, Indicador Ano de Nascimento, Voto do Relator, Indicador Grau de Escolaridade, Colegiado, Tipo de Réu, Tipo de Autor, Presidente. Entretanto, dentre as variáveis que atendem as regras 1 e 2, inclui-se apenas o Colegiado, Tipo de Réu e Presidente.

Já a regra 3, neste teste não é aplicada devido à todas as variáveis terem sido utilizadas no modelo.

As variáveis que atendem as três regras são o Colegiado, o Tipo de Réu e o Presidente. E, ao executar a regressão logística com estas variáveis, se obtêm os resultados contidos na tabela 13.

Tabela 13: Resultados obtidos com os dados de treinamento no modelo de regressão, e na utilização do modelo na amostra de teste. Foram utilizadas as variáveis Colegiado, Tipo de Réu e Presidente.

Fonte: Própria.

	Amostra de Treinamento	Amostra de Teste
Instâncias classificadas corretamente	63,9%	62,8%
Votos negativos classificados como positivos	6,25%	6,97%
Votos positivos classificados como negativos	87,63%	89,28%

De acordo com a tabela 13, verifica-se que foi obtida uma porcentagem similar de acertos para a amostra de treinamento e teste, e, uma baixa previsão para os votos positivos, menor ainda que a do teste anterior, devido à retirada da maior parte das variáveis. Ainda foram obtidos outros dados como:

- Valor de log verossimilhança inicial: 10459,574
- Valor de log verossimilhança final: 10085,413
- R^2 de Cox & Snell: 0,046
- R^2 de Nagelkerke: 0,063

Comparando os valores do modelo original, do modelo ajustado, e ainda o do modelo de Cameron e Cummings (2003), verifica-se que o modelo ajustado, apesar de possuir uma

melhor classificação dos votos negativos, apresentou uma diminuição na quantidade de acertos dos votos positivos, ou seja, conclui-se que algumas das variáveis retiradas eram de grande importância para a previsão dos votos positivos.

Já o modelo de Cameron e Cummings, além de obter uma grande porcentagem de erros dos votos positivos, teve uma porcentagem maior do total de erros dos votos, comparado aos modelos completo e ajustado. Os valores podem ser verificados na figura 18.

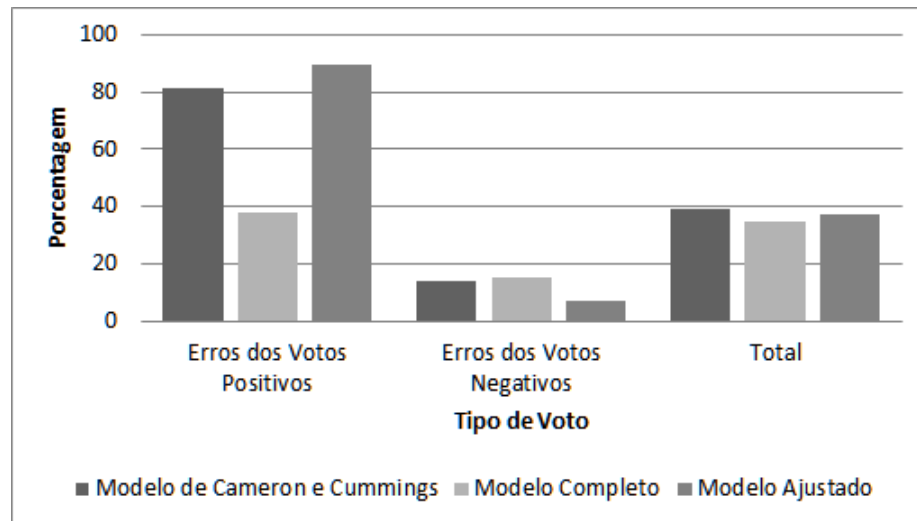


Figura 18: Comparação da porcentagem de erros para cada tipo de voto dos modelos utilizados.

Fonte: Própria.

5.3 ÁRVORE DE DECISÃO

Para obter o modelo de árvore de decisão, foi aplicado o algoritmo J48, que utilizou como variáveis todas as citadas anteriormente. Posteriormente, foi aplicado o algoritmo utilizando apenas as variáveis identificadas como relevantes no modelo de regressão, para verificar se houve um ganho na projeção dos votos.

Para o treinamento do modelo, foi utilizada a mesma amostra de treinamento da regressão, assim como a mesma amostra para teste do modelo.

Antes da aplicação do algoritmo J48, foram definidos alguns parâmetros para a execução do algoritmo, como o fator de confiança e o número mínimo de instâncias por folhas. O fator de confiança é utilizado para realizar a poda, em que menores valores implicam em maiores podas, neste caso utilizou-se o valor de 0,25, valor padrão do programa *Weka*, e também valor em que se obteve uma maior porcentagem de acertos. Já o número mínimo de instâncias

por folha, foi definido como 2, valor padrão do programa *Weka*, entretanto, caso os dados possuam muitos ruídos é necessário aumentar este valor.

Como resultado da aplicação do J48 utilizando todas as variáveis, foi obtido o resultado contido na tabela 14 para a amostra de treinamento e teste².

Tabela 14: Resultados obtidos na aplicação dos dados de treinamento no modelo de árvore, e na aplicação das regras da árvore de decisão obtidas na amostra de teste. Neste teste foram utilizadas todas as variáveis.

Fonte: Própria.

	Amostra de Treinamento	Amostra de Teste
Instâncias classificadas corretamente	86,12%	81,33%
Estatística Kappa	0,697	0,5921
Média de erro absoluto	0,2078	0,2448
Raiz quadrada do erro quadrático médio	0,3223	0,3794
Votos negativos classificados como positivos	9,01%	12,71%
Votos positivos classificados como negativos	22,26%	28,94%

Ao aplicar as variáveis Colegiado, Tipo de Réu e Presidente, interpretadas como relevantes a partir do modelo de regressão, e, ainda utilizando a mesma amostra de treinamento e teste, foram obtidos os resultados demonstrados na tabela 15. Pode-se verificar que diminuiu o número de classificações erradas para votos negativos, entretanto, houve um crescimento no número de votos positivos classificados de forma errônea.

Outra comparação realizada, foi a de utilizar somente os atributos citados por Cameron e Cummings (2003) e que puderam ser extraídos dos dados. De acordo com estes atributos, foi obtido o resultado disponível na tabela 16, após a aplicação do algoritmo J48.

Na figura 19 são apresentadas as porcentagens de erros do modelo sugerido na sua forma completa, do modelo sugerido ajustado de acordo com as variáveis interpretadas como influentes pelo modelo de regressão, e do modelo que utiliza as variáveis do trabalho de Cameron e Cummings. Verifica-se a menor porcentagem de erros em votos positivos, apresentado pelo modelo completo sugerido neste trabalho, em comparação ao conjunto de variáveis utilizados por Cameron e Cummings (2003) na sua previsão. Já o modelo ajustado, obteve uma melhora na projeção de votos negativos, e teve uma queda no desempenho dos votos positivos, que ocorreu devido à diminuição da quantidade de previsores do modelo.

²A interpretação dos valores obtidos deve se dar conforme subseção 3.3.3 deste trabalho.

Tabela 15: Resultados obtidos com os dados de treinamento no modelo de árvore, e na utilização do modelo na amostra de teste. Foram utilizadas as variáveis Colegiado, Tipo de Réu e Presidente.

Fonte: Própria.

	Amostra de Treinamento	Amostra de Teste
Instâncias classificadas corretamente	64,44%	63,49%
Estatística Kappa	0,0784	0,0506
Média de erro absoluto	0,4411	0,4429
Raiz quadrada do erro quadrático médio	0,4697	0,4719
Votos negativos classificados como positivos	4,72%	5,12%
Votos positivos classificados como negativos	88,76%	90,68%

Tabela 16: Resultados obtidos com os dados de treinamento no modelo de árvore, e na utilização do modelo na amostra de teste. Foram utilizadas as variáveis do modelo de Cameron e Cummings.

Fonte: Própria.

	Amostra de Treinamento	Amostra de Teste
Instâncias classificadas corretamente	76,7123%	71,4739%
Estatística Kappa	0,4773	0,3572
Média de erro absoluto	0,3228	0,3572
Raiz quadrada do erro quadrático médio	0,4017	0,443
Votos negativos classificados como positivos	12,76%	16,37%
Votos positivos classificados como negativos	41,45%	49,50%

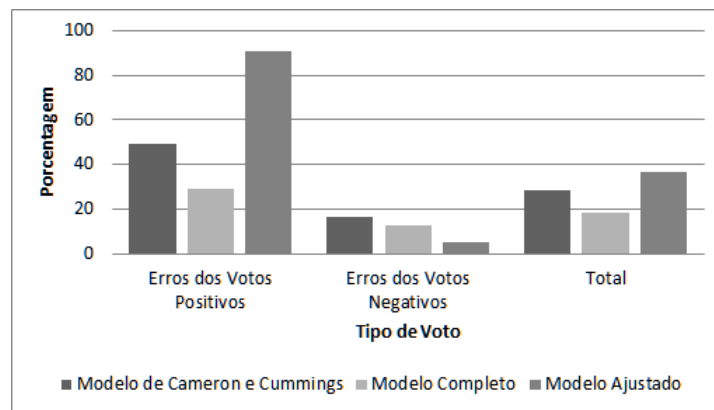


Figura 19: Comparação da porcentagem de erros no modelo de árvore de decisão utilizando as respostas da amostra de teste.

Fonte: Própria.

Comparando os resultados obtidos de modo geral nos modelos abordados, a figura 20 ilustra a porcentagem de votos positivos e negativos projetados de forma correta nos modelos completos, que foram os modelos que tiveram uma maior porcentagem de acertos. Pode-se verificar que o modelo de árvore de decisão se destaca ao projetar os votos positivos, que possui baixa projeção no modelo de regressão.

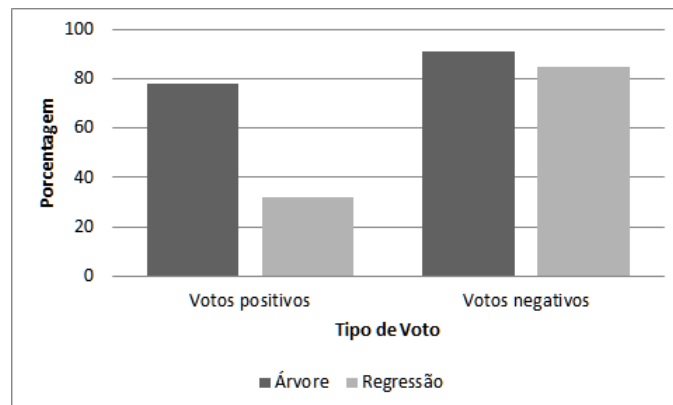


Figura 20: Comparação entre a porcentagem de acertos dos votos positivos e negativos de cada modelo.

Fonte: Própria.

5.4 SIMULAÇÃO

O comportamento desenvolvido neste trabalho para os agentes se dá conforme figura 21. A simulação neste ambiente é similar à análise de sensibilidade (ver subseção 3.5.1), entretanto, dá autonomia para o usuário testar diversos cenários. Entretanto, na simulação há

ainda a utilização de variáveis de interação para verificar o quanto influenciam no resultado final de cada votação.

A primeira variável de interação é a probabilidade de um membro do colegiado mudar seu voto para seguir o de um colega, ou seja, após calcular o seu voto de acordo com um modelo (árvore ou regressão), este pode alterar o voto. A segunda variável de interação está relacionada à primeira, pelo fato de que ao existir a probabilidade de mudar o voto, este pode seguir o voto de um colega aleatório, ou ainda, seguir o voto de um membro que possua mais votos que foram seguidos pelos outros ao longo de sua carreira, podendo-se inferir que possui uma maior influência.

Vale ressaltar, que nem sempre a mudança do voto implica em alterá-lo de positivo para negativo por exemplo, a mudança pode significar a mudança de um argumento, caso o voto copiado seja igual ao do membro que está alterando o voto. A mudança se dará conforme a probabilidade que se estabelece, em seguir o voto de um membro aleatório ou membro influente.

Na figura 22 pode-se verificar os controles de variáveis (1), os gráficos obtidos de acordo com os valores da simulação (2), além da tela que permite visualizar os agentes que julgam cada processo (3). Os botões de controle representam as variáveis contidas nas tabelas 18 e 19, e, a probabilidade de estes ocorrerem na criação dos agentes (processos/membros), são representados pelos botões com sufixo “%”. Um exemplo para melhor entendimento pode ser verificado abaixo.

“Suponha que todos os processos devam ter como presidente a senhora Dilma Rousseff para testar o impacto desta variável nas demais. Para isto, deve-se selecionar no botão de controle “presidente” o valor 1 e no botão de controle “%presidente” o valor 100%.”

Quanto a tela que possibilita visualizar a quantidade de agentes da simulação e os processos que são criados ao longo da execução, estão inclusos dois tipos de agentes, os processos que ainda não foram julgados representados pelos círculos em branco, e, os membros do colegiado representados pelos bonecos vermelhos (quando votam negativo) e azuis (quando votam positivo). As arestas que interligam os membros do colegiado com os processos representam os processos que ainda não foram julgados, ou seja, ainda há um vínculo do membro com o processo.

Conforme figura 22, o número 1 representa os botões de controle de variáveis em verde, o número 2 representa os gráficos para acompanhar a quantidade de valores de cada variável ao longo da simulação, e, o número 3 ilustra os agentes e processos que são criados ao longo da simulação.

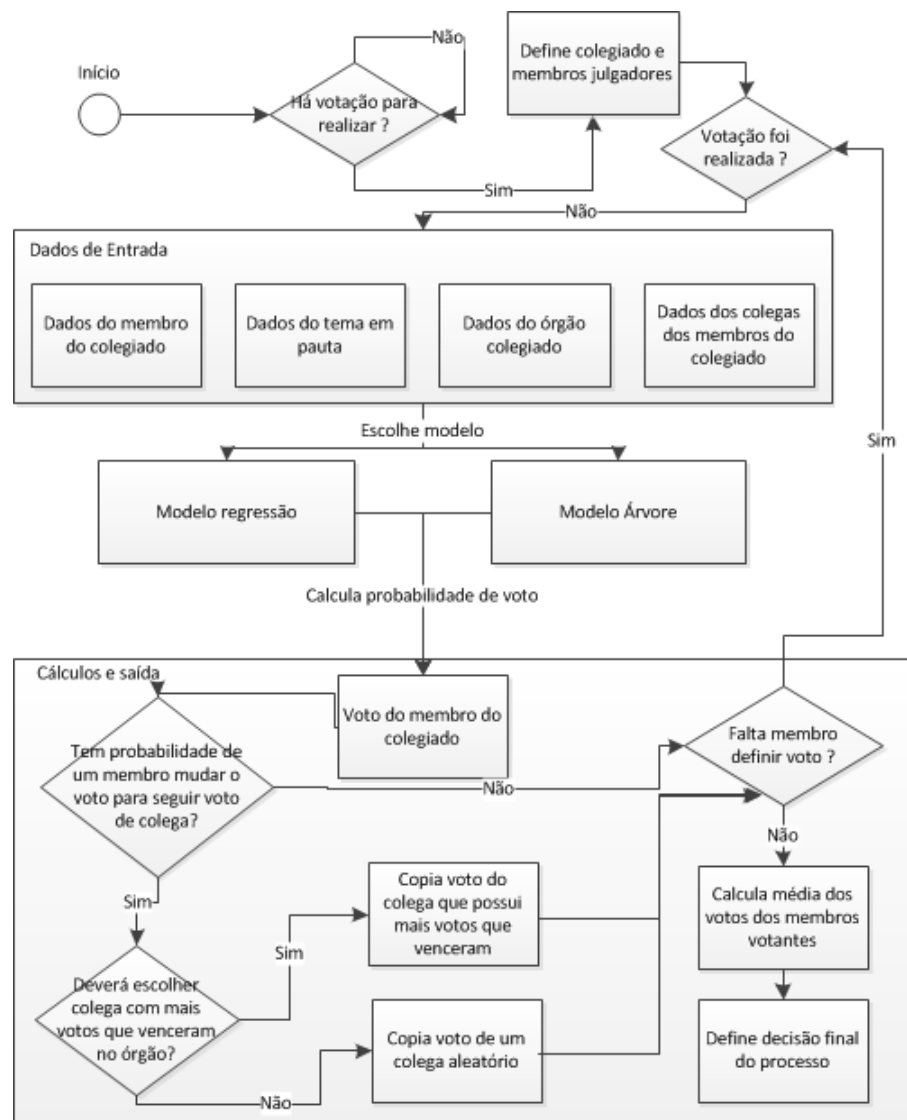


Figura 21: Fluxograma do comportamento da simulação.

Fonte: Própria.

A simulação pode ser executada de modo contínuo, ao se clicar no botão “rodando”, ou se executar uma iteração por vez, ao se clicar no botão “rodar uma vez”.

O primeiro passo executado é a definição da porcentagem de características que os agentes terão, como por exemplo a escolaridade, raça, ideologia, gênero sexual, entre outros. Escolheu-se o valor de porcentagem, pelo fato das características serem geradas de modo aleatório, mas caso o usuário deseje que todos os agentes tenham uma característica específica, este deve colocar o atributo como 100% nos botões de controle.

Após a escolha das características e clicado em “rodando” ou “rodar uma vez” serão criados os processos. As características dos processos podem ser mudadas no decorrer da

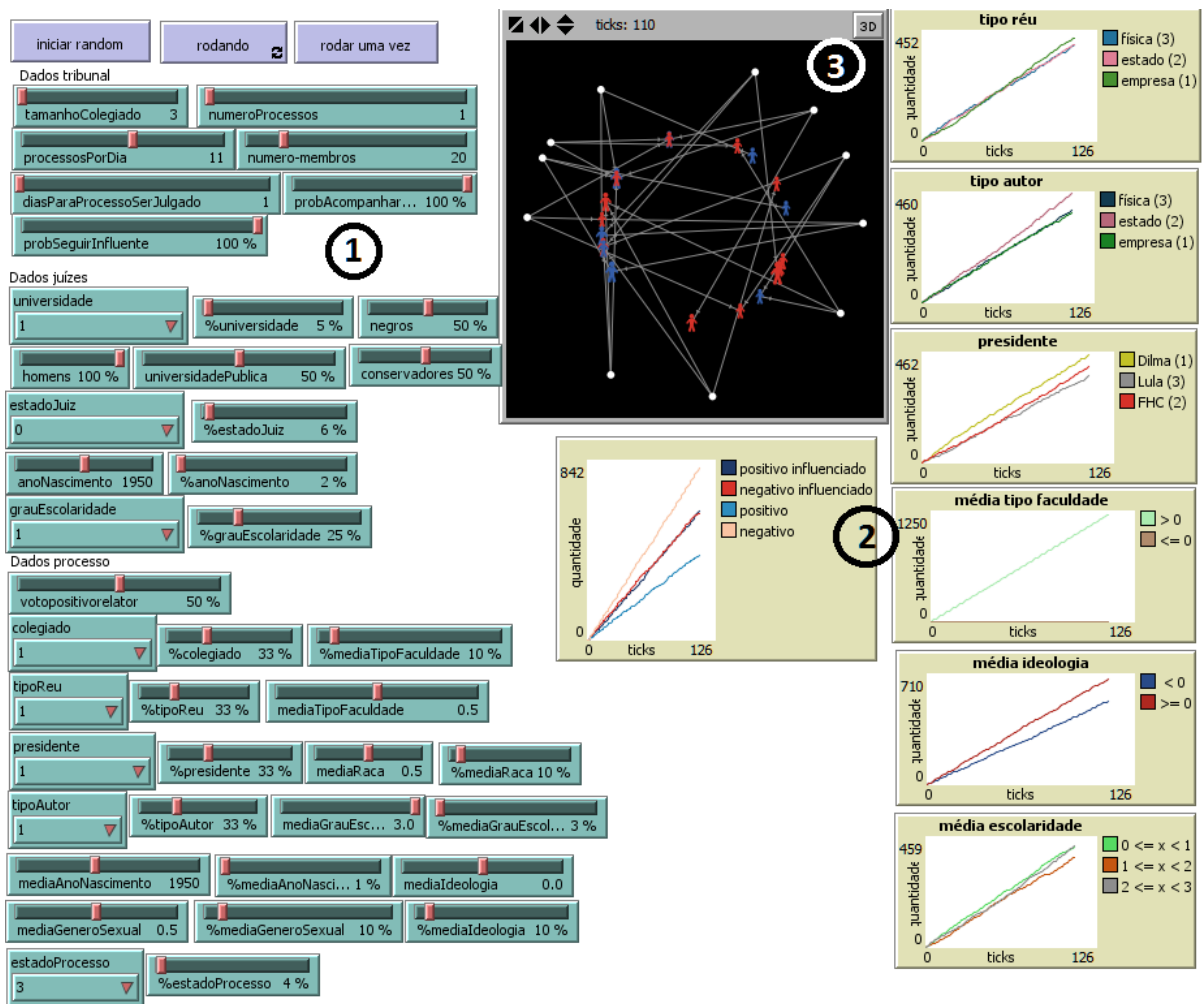


Figura 22: Simulação desenvolvida para testar os modelos sugeridos.

Fonte: Própria.

execução da simulação, que afetará somente os novos processos criados. Quanto a definição dos agentes membros do colegiado, estes são definidos quando clicado no botão “iniciar”, e, de acordo com as probabilidades e valores definidos pelo usuário.

Quando um processo passar pelo número de iterações (variável “dias para processo ser julgado”) necessários, este é julgado. A função que define se o processo foi julgado como positivo ou negativo, executa dois passos, 1) calcula através dos modelos abordados neste trabalho, o voto individual de cada membro do colegiado, e, 2) faz a média dos votos obtidos e define o processo como julgado e o seu voto final. Quando um processo é julgado, ele some da tela, para facilitar a visualização dos novos processos que estão sendo criados.

6 EXPERIMENTOS E ANÁLISE DOS RESULTADOS

A simulação visa examinar o fenômeno da influência dos mais diversos atributos nos votos dos membros de um colegiado. A análise das características com maior influência é importante para testar cenários, e, possíveis resultados da simulação criada a partir do ambiente do *NetLogo* e da análise de sensibilidade.

Na seção 6.1 são abordadas as variáveis utilizadas na simulação, com suas respectivas classificações. Na seção 6.2 é apresentada a simulação criada no *NetLogo*, que permite testar diversos cenários com os modelos criados, além de verificar o impacto das variáveis de interação. Além disso, na seção 6.2 também são apresentados os resultados obtidos através da análise de sensibilidade.

6.1 VARIÁVEIS

Para apresentar a simulação, os parâmetros são classificados em dois tipos, as variáveis e os parâmetros de suporte. Os parâmetros que influenciam o comportamento do modelo e os parâmetros que servem para avaliar o modelo, estão inclusos como variáveis. Já os “Parâmetros de suporte”, são utilizados para calcular os resultados da simulação, tanto votos individuais como o voto final de cada processo (média dos votos).

Para obter uma visão geral dos componentes do modelo, as variáveis são classificadas em variáveis independentes, dependentes e de controle. As variáveis dependentes são o voto e a probabilidade associada a ele. A variável independente é o método pelo qual se estimam os votos, e, as variáveis de controle são aquelas possíveis de serem manipuladas para verificar os votos projetados. As variáveis disponibilizadas como parâmetros da simulação estão presentes na tabela 17.

O próximo passo é transformar as variáveis em níveis de fatores, conforme as tabelas 18 e 19, em que as variáveis de resposta correspondem às variáveis dependentes. Dentre as variáveis, vale destacar a variável *num_iteracoes*, que é utilizada para testar o efeito das

Tabela 17: Parâmetros da simulação.**Fonte: Própria.**

Parâmetro	Variável	Outros parâmetros
int ano_de_nascimento	Ano de nascimento	Controle
boolean relator	Relator do processo	Controle
String estado_membro	Estado do membro do colegiado	Controle
int genero_sexual	Gênero sexual	Controle
int grau_escolaridade	Grau de escolaridade	Controle
double ideologia	Ideologia	Controle
int raca	Raça	Controle
int tipo_universidade	Tipo de universidade	Controle
String universidade	Universidade cursada	Controle
String estado_processo	Estado do processo	Controle
String tipo_reu	Tipo de réu	Controle
String tipo_autor	Tipo de autor	Controle
int voto_relator	Voto do relator	Controle
String colegiado	Colegiado	Controle
String presidente	Presidente do Brasil	Controle
double media_ano	Indicador ano de nascimento	Controle
double media_grau	Indicador grau de escolaridade	Controle
double media_raca	Indicador raça	Controle
double media_tipo_universidade	Indicador tipo de universidade	Controle
double media_genero	Indicador gênero sexual	Controle
double media_ideologia	Indicador ideologia	Controle
double prob_acompanhar_voto	Probabilidade de seguir voto	Controle
double prob_seguir_influente	Probabilidade de seguir influente	Controle
int num_iteracoes	Ticks/Número de iterações	Controle
int num_processos_dia	Número de processos por dia	Controle
int total_membros_disponiveis	Total de membros do órgão	Controle
int voto_projetado	-	Saída da simulação
int probabilidade_voto_1_ocorrer	-	Saída da simulação
int total_membros_processo	Total de membros votantes	Parâmetro de suporte
String metodo_projecao	Método para projeção de votos	Parâmetro de suporte
int num_dias_julgar	Nº de dias para julgar um tema	Parâmetro de suporte

variáveis de controle na simulação, pois define o número de iterações a serem realizadas.

Parte-se do pressuposto que, os colegiados devem ter no mínimo três membros votantes e que cada processo deve demorar no mínimo um dia para ser julgado, além de haver um ou mais processos a serem julgados por dia.

6.2 RESULTADOS E TENDÊNCIAS

Após a definição das variáveis, foram criadas duas simulações, 1) a primeira consistiu na criação de processos com características de acordo com os valores definidos nas variáveis de controle da tela de simulação, e, 2) a segunda, denominada análise de sensibilidade, que se baseou nos processos já existentes e alterou uma ou mais variáveis por vez, a fim de comparar se houve mudança de voto.

Vale ressaltar que na primeira simulação, foram inseridas duas variáveis de interação, 1) a probabilidade de um membro do colegiado seguir voto de outro membro votante, e, caso siga, será aplicada a 2) probabilidade de seguir o membro mais influente votante. A influência neste caso, é calculada pelo número de vezes que um membro teve o seu voto seguido, cujo valor pode aumentar no decorrer das iterações.

6.2.1 SIMULAÇÃO COM O NETLOGO

Com o objetivo de verificar o impacto das variáveis de interação na quantidade de votos positivos e negativos obtidas durante a simulação, foi desenvolvido um algoritmo no ambiente do *NetLogo*, conforme detalhado na seção 5.4. Dentre os diversos parâmetros existentes da simulação (tabela 17), foram realizados oito experimentos de acordo com os valores contidos na tabela 20.

Foram escolhidos os valores de *total_membros_processo* por serem os mesmos utilizados no trabalho de Tanford e Penrod (1983). Os valores de *num_dias_julgar* foram escolhidos aleatoriamente e podem ter um impacto grande no resultado, uma vez que a cada *dia/tick* que passa, o voto pode ser alterado, caso o valor de probabilidade de mudança de voto seja maior que 0.

A variável *num_processos_dia* não tem impacto diretamente nos votos, apenas implica na quantidade de processos que serão criados por *dia/tick*. A variável *total_membros_disponiveis* representa o número de membros que um órgão tem e estão disponíveis para participar das votações. Além disso, a cada processo os membros são escolhidos aleatoriamente.

Tabela 18: Definição das variáveis independentes e de controle da simulação.

Fonte: Própria.

Classificação das variáveis	Intervalo de níveis de fatores
Variáveis independentes	
metodo_projecao	{regressão logística, árvore de decisão}
num_iteracoes	N
num_processos_dia	> 0
total_membros_disponiveis	≥ 3
num_dias_julgar	> 0
total_membros_processo	≥ 3
Variáveis de controle	
raça	{0,1}
genero_sexual	{0,1}
relator	{0,1}
estado_membro	{AC,AL,AP,AM,BA,CE,DF,ES,GO,MA,MT,MS,MG,PA,PB,PR,PE,PI,RJ,RN,RS,RO,RR,SC,SP,SE,TO,ESTRANGEIRO}
grau_escolaridade	{0,1,2,3}
ano_nascimento	{1899 $\leq x \leq$ 1949}
ideologia	{-1 $\leq x \leq$ 1}
tipo_universidade	{0,1}
universidade	{FDSM,PUCMINAS,PUCSP,UCG,UERJ,UFAL,UFBA,UFC,UFJF,UFMG,UFPE,UFPR,UFRGS,UFRJ,UFRN,UFSC,UGF,UNIBENNET}
estado_processo	{AC,AL,AP,AM,BA,CE,DF,ES,GO,MA,MT,MS,MG,PA,PB,PR,PE,PI,RJ,RN,RS,RO,RR,SC,SP,SE,TO}
tipo_reu	{física,jurídica,estado}
tipo_autor	{física,jurídica,estado}
voto_relator	{0,1}
prob_membro_acompanhar_voto	{0 $\leq x \leq$ 1}
prob_membro_seguir_influente	{0 $\leq x \leq$ 1}
colegiado	{Corte Especial, Primeira Seção, Primeira Turma, Segunda Turma}

Tabela 19: Definição das variáveis de controle e dependentes da simulação.

Fonte: Própria.

Classificação das variáveis	Intervalo de níveis de fatores
Variáveis de controle	
presidente	{Fernando Henrique, Lula, Dilma}
media_ano	{1899 <= x <= 1949}
media_grau	{0 <= x <= 3}
media_raca	{0 <= x <= 1}
media_tipo_universidade	{0 <= x <= 1}
media_genero	{0 <= x <= 1}
media_ideologia	{-1 <= x <= 1}
Variáveis dependentes	
voto	
probabilidade_associada	

Já a variável *num_iteracoes* representa a quantidade de dias/ticks executados, e, a variável *prob_acompanhar_voto*, a primeira variável de interação do sistema, representa a probabilidade do membro do colegiado alterar o seu voto.

A segunda variável de interação, a *prob_seguir_influente*, depende da variável de probabilidade de acompanhar voto, pois, caso *prob_acompanhar_voto* seja maior que 0, será aplicada a probabilidade da *prob_seguir_influente* ocorrer, e, caso esta probabilidade não ocorra, será copiado o voto de um membro aleatório, ainda, caso contrário, quando ocorre a probabilidade de seguir o membro mais influente, o voto deste membro será copiado. Demonstrando desse modo, a influência das variáveis de interação em relação às variáveis de preferências pessoais.

A escolha pelo modelo de regressão se deu por conter os coeficientes que indicam o nível de influência de cada variável na projeção do voto. Quanto às variáveis de controle, foram definidos todos os valores com probabilidades iguais de ocorrerem, exceto o controle das variáveis que constam na tabela 20.

Após as simulações, obteve-se o resultado da tabela 21, que representa a quantidade de votos positivos por coalizões, como por exemplo o experimento 1, quando ocorreu 4 votos positivos e 2 votos negativos, totalizando 36 ocorrências. Foi possível verificar também que, quando o colegiado é maior, diminuem as ocorrências de coalizações cujo voto positivo é a

Tabela 20: Valores dos parâmetros de entrada da simulação no ambiente *NetLogo*. Para as simulações foi utilizado o modelo de regressão logística binária.

Fonte: Própria.

Experimento	1	2	3	4	5	6	7	8
num_dias_julgar	1	1	1	1	1	1	1	1
num_processos_dia	1	1	1	1	1	1	1	1
total_membros_processo	6	6	6	6	12	12	12	12
total_membros_disponiveis	25	25	25	25	25	25	25	25
num_iteracoes	500	500	500	500	500	500	500	500
prob_acompanhar_voto	0	0.5	0.5	1	0	0.5	0.5	1
prob_seguir_influente	0	0	0.5	1	0	0	0.5	1

maioria.

Tabela 21: Resultados obtidos com os experimentos realizados no ambiente *NetLogo* por tipo de coalizão de voto.

Fonte: Própria.

Coalizões de votos	1	2	3	4
6/6	0	10	9	149
5/6	5	26	22	0
4/6	36	50	47	0
Coalizões de votos	5	6	7	8
12/12	0	0	0	167
10/12	0	7	10	0
8/12	9	10	20	0

Já o impacto das alterações dos votos no resultado final dos processos, pode ser visualizado na figura 23. Apesar da quantidade de votos positivos e negativos, e seus valores influenciados respectivamente, forem um pouco diferentes, demonstra o impacto das variáveis de interações no resultado final dos processos, ao se comparar a quantidade de votos positivos quando não houve a mudança de votos, e quando houve.

Conforme figura 23, pode-se verificar a quantidade de processos com resultado final positivo e negativo. As legendas “positivo influenciado” e “negativo influenciado” indicam o resultado final quando há mudanças dos votos dos membros. As linhas dos votos positivos e

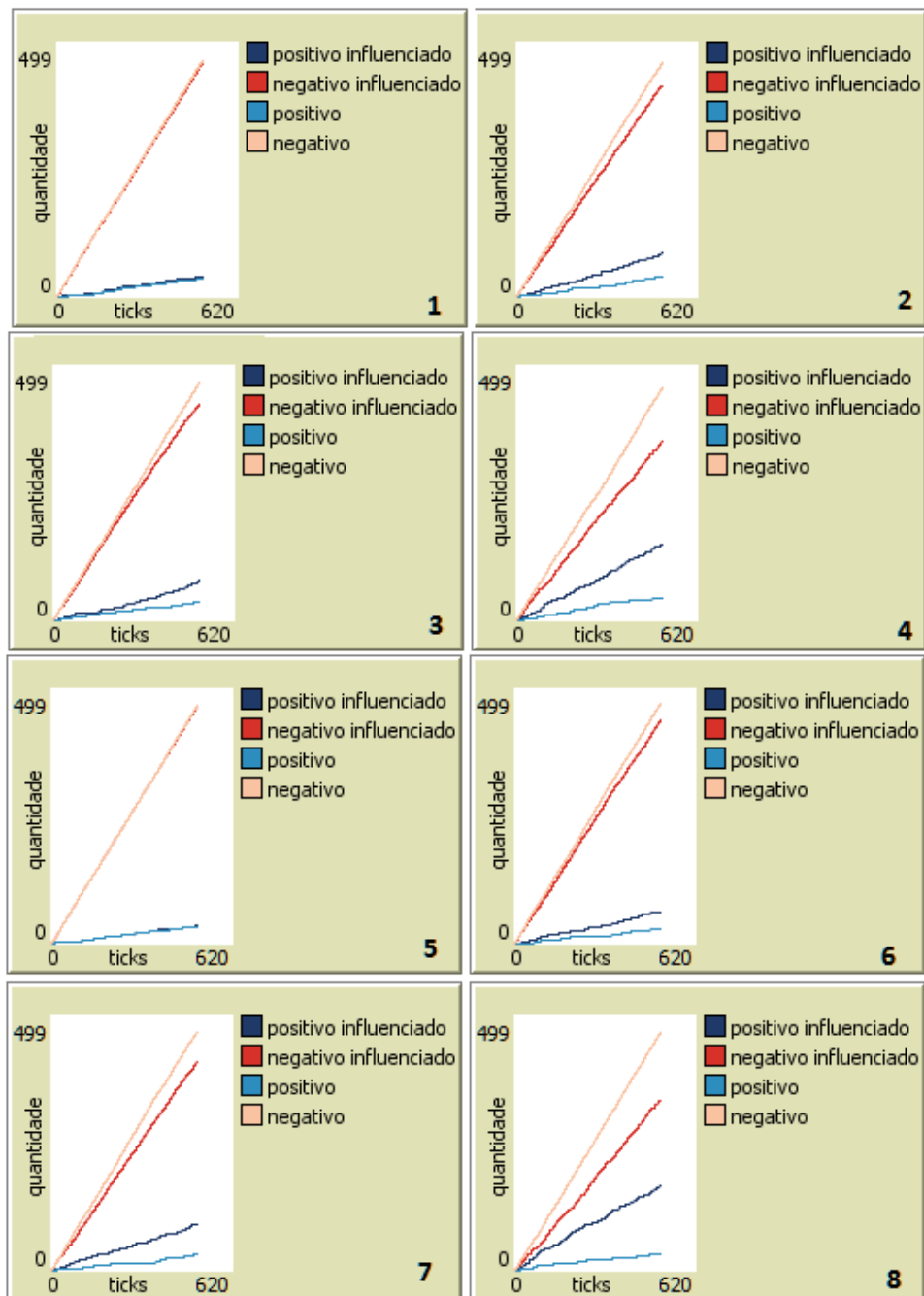


Figura 23: Gráficos dos experimentos realizados.

Fonte: Própria.

negativos, representam o voto original, ou seja, mesmo que alterado por influência de outros membros, ele foi mantido em uma variável auxiliar para viabilizar a comparação.

Na figura 23, visualiza-se também que ao um membro de um colegiado alterar seu voto, mas não necessariamente seguir o mais influente, causa mudanças também na quantidade de votos positivos e negativos. Mas pode ocorrer também que estas mudanças não sejam drásticas,

devido ao voto que é copiado ter chances de ser igual ao dele mesmo, apesar do argumento poder ser diferente, como ocorre nos julgamentos de processos reais.

6.2.2 ANÁLISE DE SENSIBILIDADE

Com o intuito de verificar a influência de cada variável no voto projetado, serão descritos abaixo os resultados obtidos com a análise.

Para a escolha dos modelos que serão utilizados, foi calculada a porcentagem de acertos dos votos positivos e negativos do modelo de regressão completo, ajustado e o modelo de árvore de decisão completo. Como resultado, pode-se verificar que o modelo de regressão e o de árvore completos foram os que se mantiveram na média, ou seja, acertaram tanto votos positivos, quanto negativos de forma similar.

A partir disto, foram utilizados os processos existentes e alterado uma ou mais variáveis para verificar o impacto destas no voto projetado. De modo geral, o comportamento da análise de sensibilidade se deu conforme a figura 10.

A análise das variáveis abaixo auxilia na compreensão de quais variáveis surtem mais efeito ao utilizar a simulação desenvolvida no *software Netlogo* para testar cenários. O total de jurisprudências analisadas foi 1.939 e o total de votos foi 12.055. As legendas presentes nas figuras a seguir, possuem as seguintes definições:

- Observado: Porcentagem de votos positivos em processos reais, cujo valor do atributo escolhido não tenha ocorrido;
- Projetado com dados originais - Árvore: Porcentagem de votos positivos, calculada através do modelo de árvore, em que os dados são de processos reais cujo valor do atributo escolhido não tenha ocorrido;
- Projetado com dados originais - Regressão: Porcentagem de votos positivos, calculada através do modelo de regressão, em que os dados são de processos reais cujo valor do atributo escolhido não tenha ocorrido;
- Projetado com dados alterados - Árvore: Porcentagem de votos positivos, calculada através do modelo de árvore. A amostra utilizada compreende as ocorrências diversas do valor da variável testada. Por exemplo, se a variável a ser testada é o colegiado e o valor é a Primeira Seção, a amostra compreende todas as jurisprudências que não tiveram como colegiado a Primeira Seção, para verificar o voto que o método projetou para elas.

Além de utilizar esta amostra, foram alterados todos os valores de colegiado para esta amostra, como sendo “Primeira Seção” por exemplo, para verificar se ao alterar o valor, houve uma modificação nos votos projetados;

- Projetado com dados alterados - Regressão: Porcentagem de votos positivos calculada através do modelo de regressão. A amostra compreende as ocorrências diversas do valor da variável testada, com a posterior modificação para verificação do voto projetado, conforme descrito no item anterior.

6.2.2.1 Colegiado

Ao se comparar os três valores desta variável, verificou-se que houve uma alteração nos votos projetados ao se modificar o valor da variável a testada. O modelo de regressão demonstrou este efeito, ao contrário do modelo de árvore que manteve a mesma projeção. É possível inferir que os membros ao participar da “Primeira Turma”, tendem a votar negativo, já os membros da “Corte Especial”, “Primeira Seção” e “Segunda Turma”, tendem a votar positivo mais vezes quando são destes colegiados, conforme figura 24.

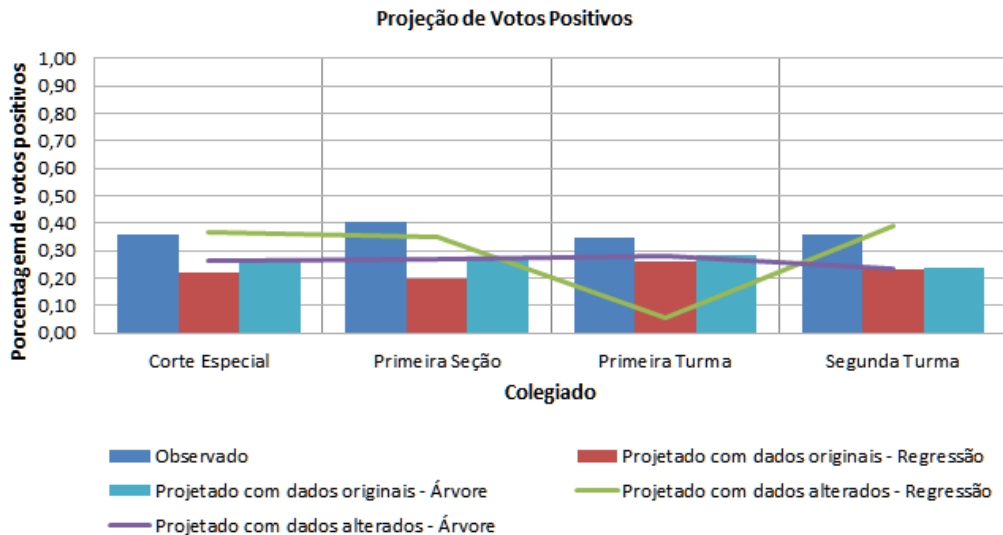


Figura 24: Teste da variável Colegiado.

Fonte: Própria.

6.2.2.2 Tipo Réu

Ao se comparar os três valores desta variável, verificou-se que houve uma alteração nos votos projetados ao se alterar o valor da variável a ser testada. O modelo de regressão

demonstrou este efeito melhor que o modelo de árvore. É possível inferir que os membros, quando o processo tem como réu o “Estado” e “Pessoa Física”, tendem a votar negativo à positivo, comparado quando o réu é o “Empresa”, conforme figura 25.

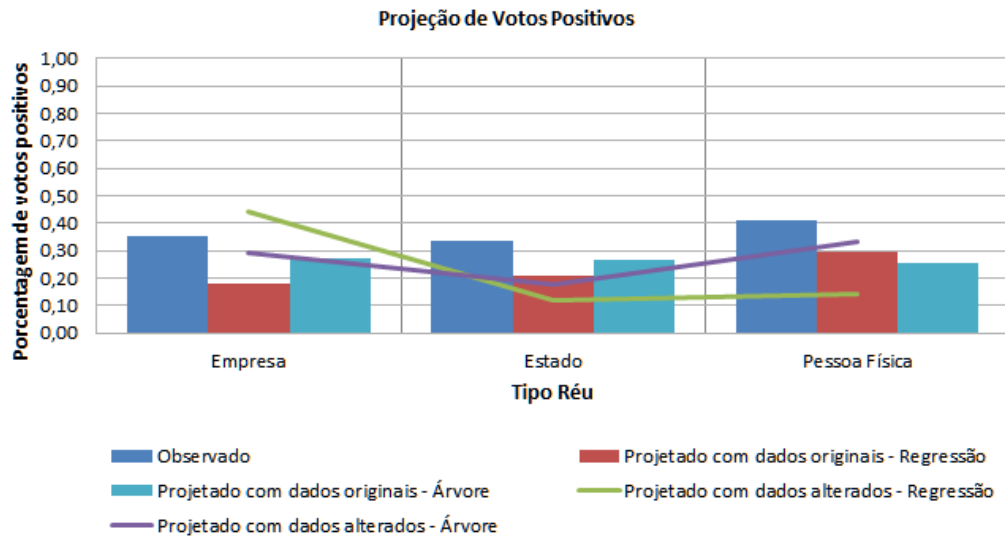


Figura 25: Teste da variável Tipo Réu.

Fonte: Própria.

6.2.2.3 Tipo Autor

Ao se comparar os três valores desta variável, verificou-se que houve uma alteração nos votos projetados ao se alterar o valor da variável a ser testada. O modelo de regressão ajustado demonstrou este efeito, apesar de pouca variação, ao contrário do modelo de árvore que manteve a mesma projeção. É possível inferir que os membros, quando o processo tem como autor o “Estado”, tendem a votar negativo à positivo, já quando o réu é “Pessoa Jurídica” ou “Pessoa Física” tendem a votar positivo mais vezes, conforme figura 26.

6.2.2.4 Estado do Processo

Ao se comparar os valores desta variável, verificou-se que houve uma alteração nos votos projetados ao se alterar o valor da variável a ser testada. Tanto o modelo de regressão, quanto o modelo de árvore demonstraram este efeito, apesar de não representarem as mesmas variâncias. É possível inferir, de acordo com o modelo de regressão, que os membros quando o processo tem como estado de origem os estados “AM”, “BA”, “CE”, “MT”, “PB”, “PI”, “RO”, “RR” e “TO”, tendem a votar negativo à positivo, já quando o processo tem como estado de origem os demais estados, tendem a votar positivo mais vezes, conforme figura 27.

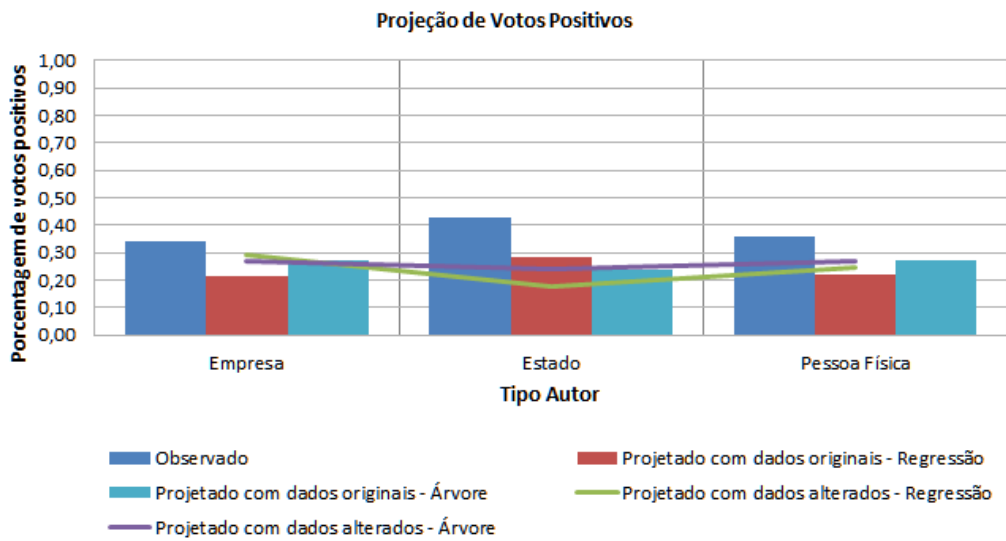


Figura 26: Teste da variável Tipo Autor.

Fonte: Própria.

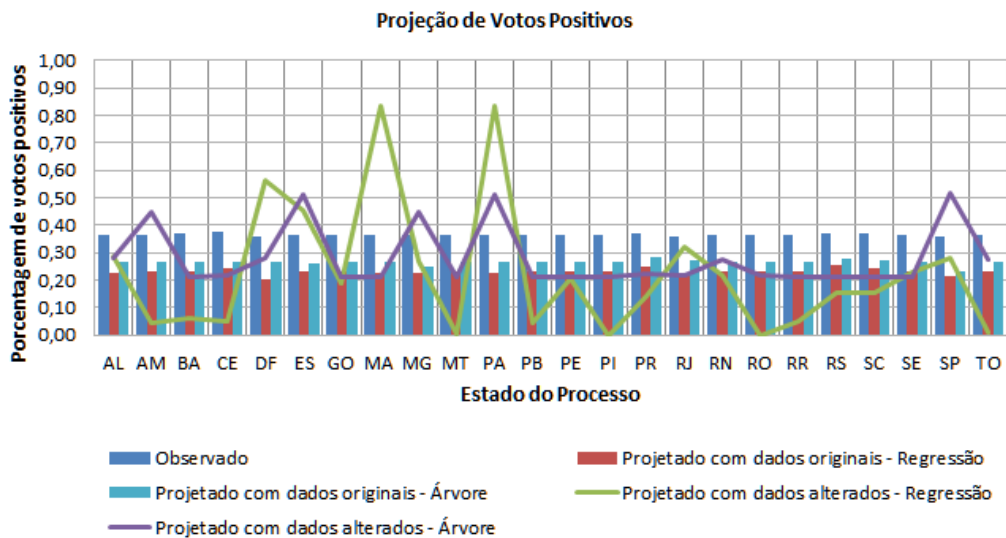


Figura 27: Teste da variável Estado do Processo.

Fonte: Própria.

6.2.2.5 Faculdade

Ao se comparar os valores desta variável, verificou-se que houve uma alteração nos votos projetados ao se alterar o valor da variável a ser testada. Tanto o modelo de regressão, quanto o modelo de árvore demonstraram este efeito. É possível inferir, de acordo com o modelo de regressão, que os membros do colegiado quando cursaram as faculdades “UERJ”, “UFAL”,

“UFBA”, “UFC”, “UFRGS” e “UFRJ”, tendem a votar positivo mais vezes, já quando cursaram as demais faculdades, tendem a votar negativo à positivo, conforme figura 28.

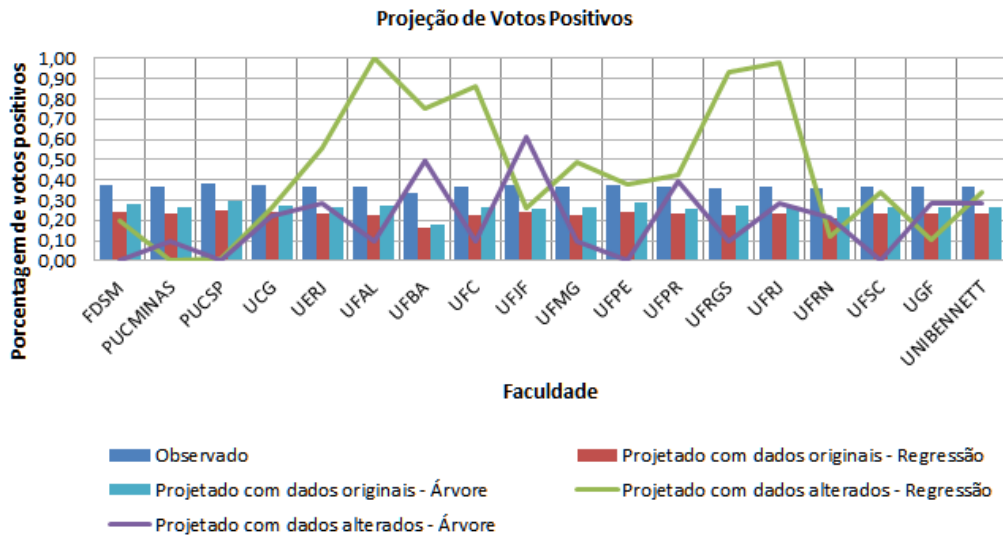


Figura 28: Teste da variável Faculdade.

Fonte: Própria.

6.2.2.6 Ideologia

Ao se comparar os valores desta variável, verificou-se que não houve uma alteração, por parte do modelo de árvore, nos votos projetados ao se alterar o valor da variável a ser testada, conforme figura 29. Já conforme modelo de regressão, houve uma alteração significativa, em que se demonstrou haver mais votos positivos quando o valor da ideologia é superior a 0.

6.2.2.7 Indicador Ideologia

Ao se comparar os valores desta variável, verificou-se que houve uma alteração nos votos projetados ao se alterar o valor da variável a ser testada. O modelo de regressão ajustado demonstrou este efeito, ao contrário do modelo de árvore que manteve a mesma projeção. É possível inferir que quando os colegas de julgamento do membro do colegiado possuem uma média simples de ideologia abaixo de 0,3, o membro tende a votar negativo à positivo, já quando a média está acima de 0,3, o membro do colegiado tende a votar positivo mais vezes, conforme figura 30.

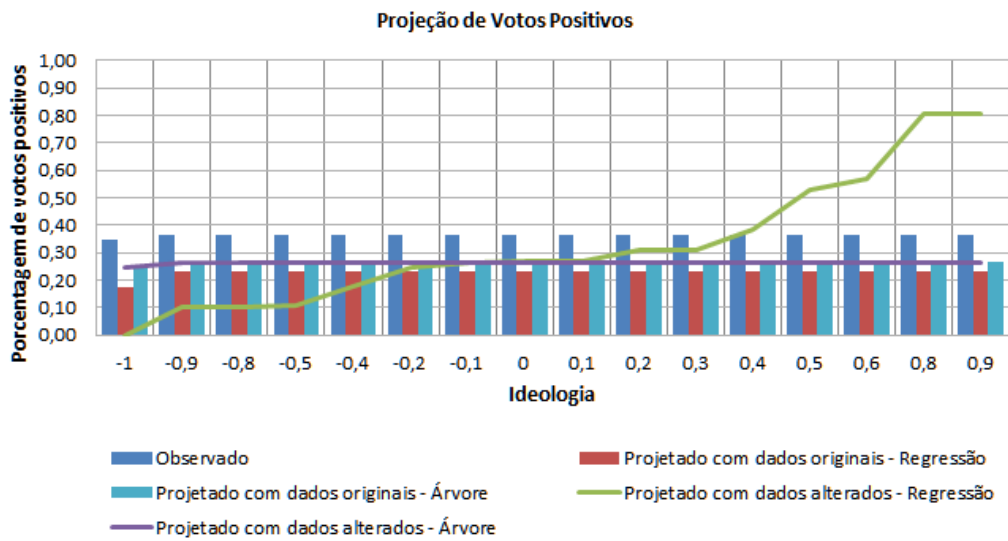


Figura 29: Teste da variável Ideologia.

Fonte: Própria.

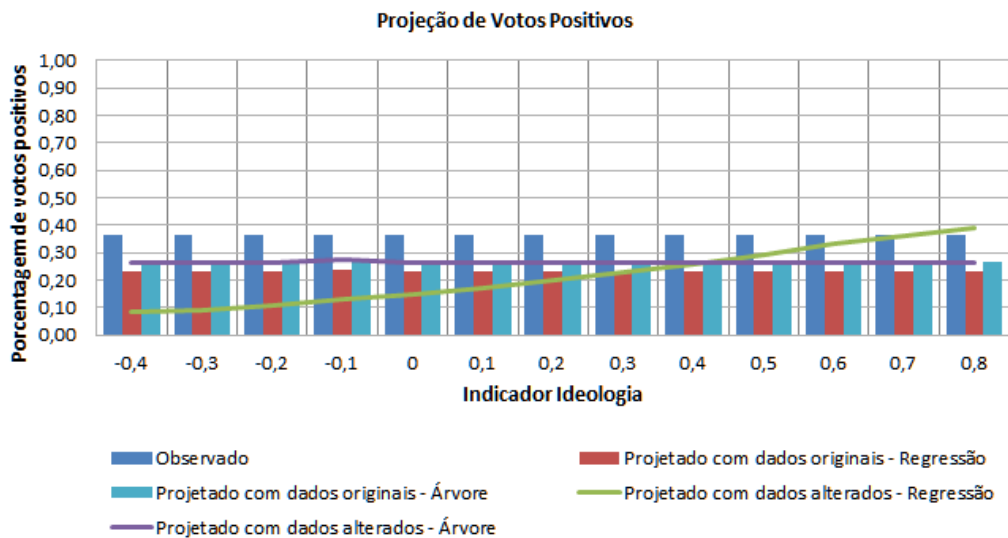


Figura 30: Teste da variável Indicador Ideologia.

Fonte: Própria.

6.2.2.8 Presidente

Ao se comparar os valores desta variável, verificou-se que houve uma alteração nos votos projetados ao se alterar o valor da variável a ser testada. O modelo de regressão ajustado demonstrou este efeito, ao contrário do modelo de árvore que manteve a mesma projeção. É possível inferir que quando o presidente do Brasil é o “Fernando Henrique Cardoso”, os

membros do colegiado tendem a votar negativo à positivo, já quando o presidente do Brasil é o “Luiz Inácio Lula da Silva” e a “Dilma Rousseff”, os membros tendem a votar positivo mais vezes, conforme figura 31.

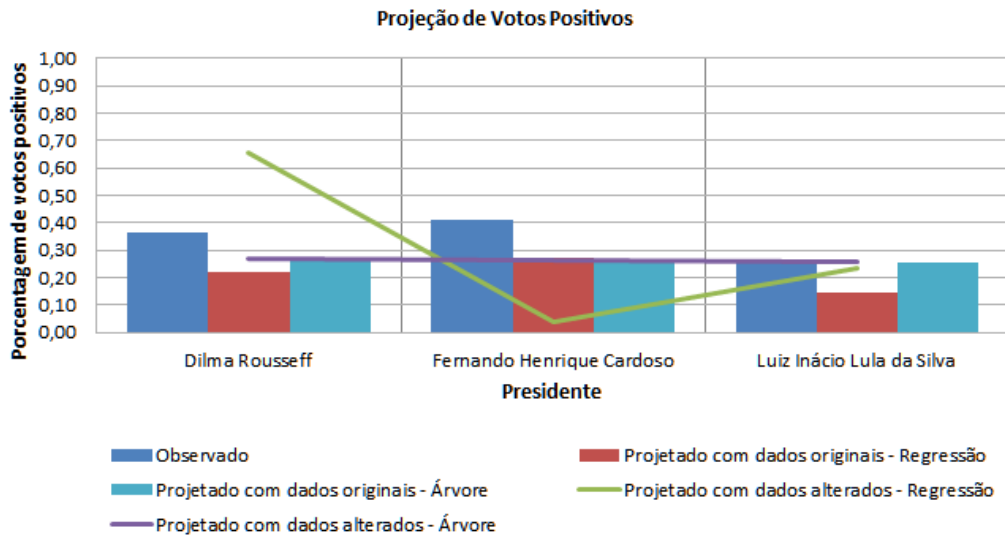


Figura 31: Teste da variável Presidente.

Fonte: Própria.

6.2.2.9 Voto do Relator

Ao se comparar os valores desta variável, verificou-se que houve uma alteração nos votos projetados ao se alterar o valor da variável a ser testada. Tanto o modelo de regressão, quanto o modelo de árvore demonstraram este efeito, apesar de indicarem variâncias diferentes. É possível inferir que os membros do colegiado quando o voto do relator é negativo, tendem a votar negativo à positivo, já quando o voto do relator é positivo tendem a votar positivo mais vezes, conforme figura 32.

6.2.2.10 Presidente x Tipo Autor

Ao se comparar os valores destas variáveis, verificou-se que houve uma alteração nos votos projetados ao se alterar o valor da variável a ser testada. O modelo de regressão ajustado demonstrou este efeito, ao contrário do modelo de árvore que manteve a mesma projeção. É possível inferir que os membros do colegiado no governo do “Fernando Henrique Cardoso” e o autor do processo é “Pessoa Jurídica”, tendem a votar negativo.

Já quando o governo é da “Dilma Rousseff” e o autor do processo é “Pessoa Jurídica”,

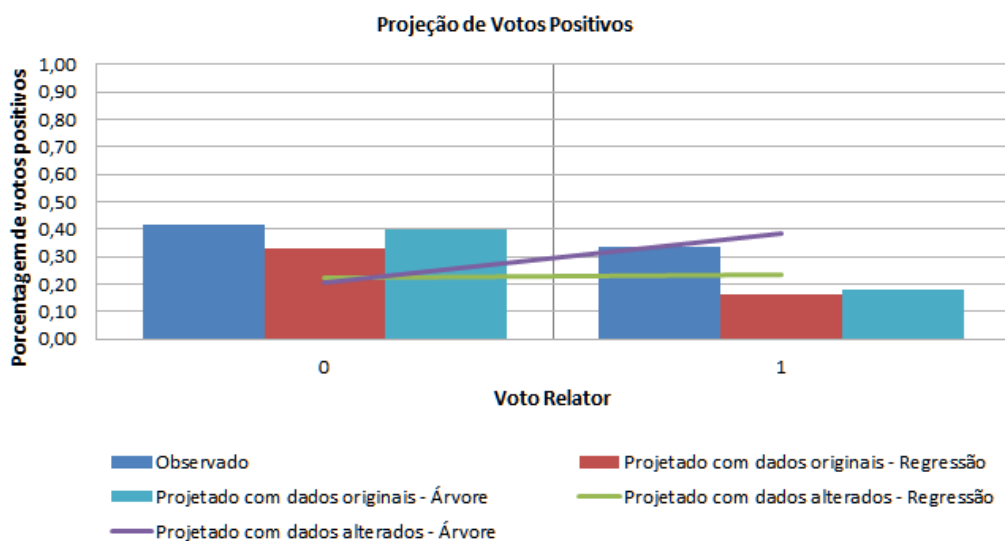


Figura 32: Teste da variável Voto do Relator.

Fonte: Própria.

“Estado” ou “Pessoa Física”, ou governo do “Luiz Inácio Lula da Silva” e o autor do processo é “Pessoa Jurídica” ou “Pessoa Física”, ou ainda, quando o governo é do “Fernando Henrique Cardoso” e o autor do processo é “Estado” ou “Pessoa Física”, os membros tendem a votar positivo mais vezes, conforme figura 33.

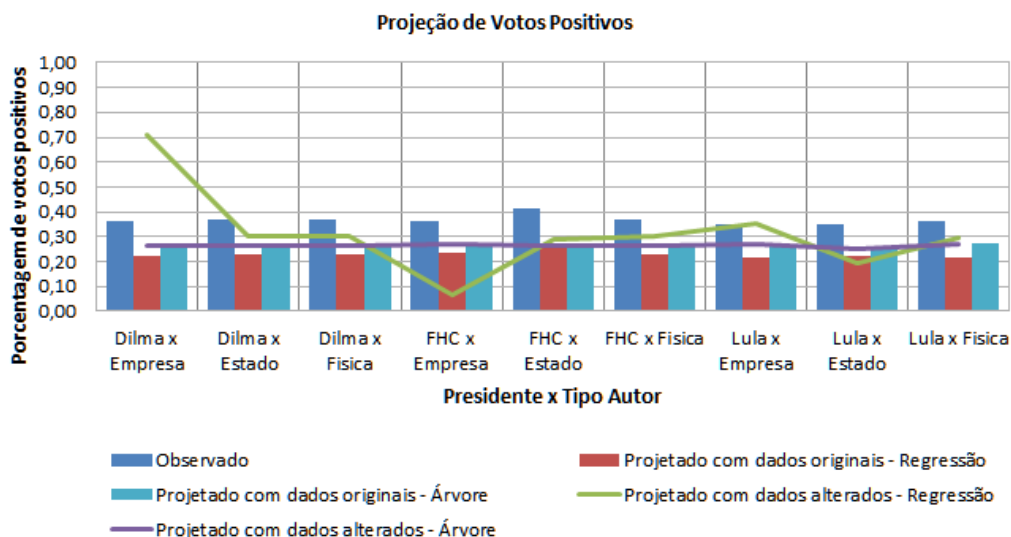


Figura 33: Teste da variável Presidente x Tipo Autor.

Fonte: Própria.

6.2.2.11 Presidente x Tipo Réu

Ao se comparar os valores destas variáveis, verificou-se que houve uma alteração nos votos projetados ao se alterar o valor da variável a ser testada. Tanto o modelo de regressão, quanto o modelo de árvore demonstraram este efeito. É possível inferir que os membros do colegiado no governo do “Fernando Henrique Cardoso” e o réu é “Pessoa Jurídica”, tendem a votar negativo. Já quando o governo é da “Dilma Rousseff” ou do “Luiz Inácio Lula da Silva” e o réu é “Pessoa Jurídica”, os membros tendem a votar positivo mais vezes, conforme figura 34.

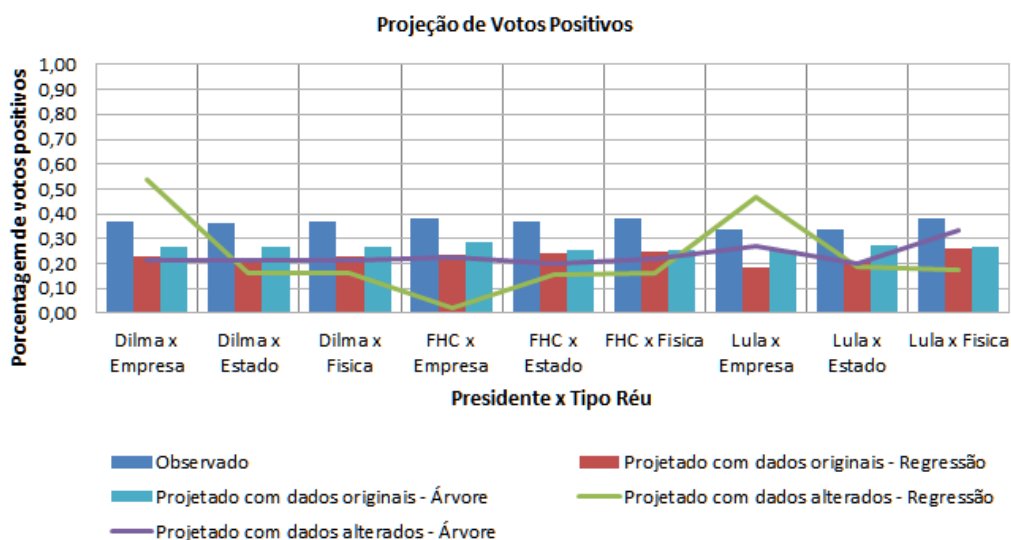


Figura 34: Teste da variável Presidente x Tipo Réu.

Fonte: Própria.

7 CONCLUSÃO E TRABALHOS FUTUROS

Diante do trabalho exposto, conclui-se que a simulação e os modelos sugeridos, que incluem a regressão logística e a árvore, permitem projetar votos individuais de membros de um colegiado durante um julgamento. Os dados utilizados para projeção apenas identificam probabilidades de um voto ocorrer, de acordo com os dados utilizados para o treinamento do modelo.

A principal contribuição deste trabalho é a modelagem do cenário de um colegiado e como os membros se comportam de acordo com as variáveis de influência. A partir disto, foram sugeridas novas variáveis que podem influenciar na projeção de votos em grupos.

Com os modelos sugeridos, foi possível comparar o resultado da projeção dos votos, e, com uma análise estatística mais refinada, viabilizou a diminuição da quantidade de variáveis utilizadas, por obter um resultado similar ao do modelo completo, que utiliza todas as variáveis preditoras. No modelo de regressão completo foram obtidos 65,4% de acertos no geral, com 21 variáveis preditoras, e no modelo de regressão ajustado obtiveram-se 61,1%, com apenas três variáveis.

De acordo com a realidade brasileira, os modelos sugeridos destacam-se frente aos modelos estudados, pois, ao se utilizar as variáveis citadas no modelo de Cameron e Cummings (2003) por exemplo, obteve-se uma baixa previsibilidade de votos positivos. Há ainda, o modelo de árvore de decisão que se destacou frente ao modelo de regressão, pois, ao aplicar o algoritmo de árvore C4.5, na amostra de teste e utilizando todas as variáveis, foram obtidos 81,33% de acertos na projeção dos votos.

Foi fundamental a aplicação dos dois métodos, a árvore de decisão e a regressão, pois permitiu verificar o grau de acerto de ambos, além de analisar se a projeção pode se diferenciar de acordo com os métodos aplicados. A vantagem do modelo de regressão em relação ao modelo de árvore, é a de fornecer os coeficientes que podem quantificar o valor de influência de cada variável sobre a variável resposta (voto), as analisando portanto, de modo independente.

Além do valor obtido pela regressão, através do cálculo da ideologia realizado com o

algoritmo de Poole et al. (2011) e com dados das votações, foi possível obter uma quantidade maior de acertos¹, entretanto, o algoritmo não considera as características do processo e partes durante as votações, as quais este trabalho teve enfoque.

Apesar de ser possível simular julgamentos, a partir da seleção de alguns atributos disponíveis *online* sobre os membros de um colegiado e documentos, o modelo não atendeu a alguns parâmetros de avaliação estatísticos, como o R^2 elevado. Entretanto, segundo Wooldridge (2008), nas ciências sociais é comum encontrar valores baixos para o R^2 , como por exemplo entre 0,1 e 0,3, mas, não implicam que modelos com estes valores devam ser inutilizados. Esse valor é esperado pela dificuldade em se prever o comportamento individual estatisticamente, ou seja, identificar todas as variáveis que possam influenciar e quantificá-las. Apesar disso, a simulação possibilitou verificar a mudança de votos projetados de acordo com os valores das variáveis de interação, já na análise de sensibilidade pôde-se verificar a mudança dos votos projetados com a alteração de alguns valores de variáveis preditoras.

Com a realização da simulação, foi possível fornecer um ambiente para testes de cenários e dos modelos propostos, além de facilitar a visualização do impacto das variáveis de interação sobre a decisão final de cada julgamento. Já com a refinação do modelo de regressão logística, pôde-se verificar que entre as variáveis que mais exercem influência, estão inclusas o Colegiado, Tipo de Réu e Presidente. Entretanto, a partir da análise de sensibilidade, pôde-se verificar que outros valores de variáveis têm forte influência, como identificado através dos gráficos das variáveis de Ideologia, Estado do processo, Faculdade, entre outras.

A obtenção das variáveis de maior influência, contrasta com as variáveis utilizadas em modelos de outros países, como por exemplo o de Cameron e Cummings (2003), uma vez que o resultado deste trabalho indicou que as variáveis de processo (características do processo/ambiente), exercem mais influência que as próprias características dos membros, em quantidade de variáveis deste tipo.

Como trabalhos futuros, indica-se a busca de novas variáveis para adicionar aos modelos e verificar se possuem um efeito considerável, além da análise individualizada e em grupos das variáveis existentes. A busca de novas variáveis pode se tornar um trabalho difícil, uma vez que não existem informações disponíveis prontas para o uso, como as atividades exercidas pelos membros ao longo de sua carreira profissional. Outro item, é a extração de jurisprudências não abordadas neste caso, para verificar a aderência do modelo proposto em novos julgamentos.

¹O algoritmo *W-NOMINATE* acertou cerca de 87.74% com apenas uma dimensão ao utilizar os dados de votos dos membros e os participantes da votação.

REFERÊNCIAS

- BECKER, G.; MURPHY, K. **Social Economics: Market Behavior in a Social Environment**. Belknap Press of Harvard University Press, 2000. (Belknap Series). ISBN 9780674011212. Disponível em: <<http://books.google.com.br/books?id=cUMgPIPp0-IC>>.
- BOUTILIER, C.; LU, T. Probabilistic and utility-theoretic models in social choice: Challenges for learning, elicitation, and manipulation. **IJCAI**, 2011.
- BRUDNEY, J. J.; SCHIAVONI, S.; MERRITT, D. J. **Judicial Hostility Toward Labor Unions? Applying the Social Background Model to a Celebrated Concern**. [S.l.]: 60 Ohio St. L.J. 1675., 1999.
- CAMERON, C. M.; CUMMINGS, C. P. **Diversity and Judicial Decision-Making: Evidence from Affirmative Action Cases in the Federal Courts of Appeals, 1971-1999**. [S.l.: s.n.], 2003.
- CARROLL, R. et al. Measuring Bias and Uncertainty in DW-NOMINATE Ideal Point Estimates via the Parametric Bootstrap. **Political Analysis**, v. 17, n. 3, p. 261–275, jun. 2009. Disponível em: <<http://dx.doi.org/10.1093/pan/mpp005>>.
- COX, D. R.; SNELL, E. J. A General Definition of Residuals. **Journal of the Royal Statistical Society. Series B (Methodological)**, Wiley, v. 30, n. 2, p. 248–275, 1968.
- CRAMER, J. S. Cambridge University Press, 2003. Disponível em: <<http://dx.doi.org/10.1017/CBO9780511615412.010>>.
- EPSTEIN, L.; KNIGHT, J. **The Choices Justices Make**. Washington, D.C.: CQ Press, 1997.
- EVANS, M. C. et al. Recounting the Courts? Applying Automated Content Analysis to Enhance Empirical Legal Research. **SSRN eLibrary**, SSRN, 2006.
- FIELD, A. **Discovering Statistics Using SPSS**. [S.l.]: SAGE Publications, 2005. ISBN 0761944524.
- FIELD, A.; HOLE, G. **How to Design and Report Experiments**. [S.l.]: Sage Publications Limited, 2003. ISBN 9780761973836.
- FOUTS, D. **The Canadian Supreme Court, 1950-1960: A Study of Judicial Policy Making**. Thesis (Ph.D.)–University of Minnesota, 1967. Disponível em: <<http://books.google.com.br/books?id=HHWKMwEACAAJ>>.
- GAINES, D. M.; BROWN, D. C.; DOYLE, T. K. A computer simulation model of juror decision making. **Expert Systems with Applications**, v. 11, n. 1, p. 13 – 28, 1996. ISSN 0957-4174. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0957417496000024>>.
- GEORGE, T. E. **Court Fixing**. [S.l.]: 43 Ariz. L.Rev. 9, 2001.

HAN, J. **Data Mining: Concepts and Techniques**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005. ISBN 1558609016.

HOSMER, D. W.; LEMESHOW, S. **Applied logistic regression (Wiley Series in probability and statistics)**. 2. ed. [S.l.]: Wiley-Interscience Publication, 2000. ISBN 0471356328.

HWONG, T. A **Quantitative Exploration of Judicial Decision Making in Canadian income Tax Cases**. 234 p. Tese (Doctor of philosophy in Law) — Faculty of Graduate Studies and Osgoode Hall Law School, Toronto, Ontario, 2006.

LANDIS, J. R.; KOCH, G. G. The Measurement of Observer Agreement for Categorical Data. **Biometrics**, v. 33, n. 1, p. 159–174, mar. 1977.

LARSON, R.; FARBER, B. **Elementary Statistics**. 4. ed. [S.l.]: Pearson Prentice Hall, 2010.

LATANÉ, B. The psychology of social impact. **American Psychologist**, v. 36, n. 4, p. 343–356, 1981.

LATANE, B.; WOLF, S. The Social Impact of Majorities and Minorities. **Psychological Review**, v. 88, n. 5, p. 438–53, set. 1981. Disponível em: <<http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ254474>>.

LAW, A. M.; KELTON, D. W. **Simulation Modeling and Analysis**. [S.l.]: McGraw-Hill Education.Europe, 2000. Paperback. ISBN 0071165371.

LEEUEW, S. E. van der. Why model? **Cybernetics and Systems**, v. 35, n. 2-3, p. 117–128, 2004.

LEVINE, D. M. et al. **Estatística. Teoria e Aplicações**. 5. ed. [S.l.]: Grupo Gen and LTC, 2008.

LORSCHIED, I. H. B.-O. M. M. **Opening the ‘Black Box’ of Simulations: Increased Transparency of Simulation Models and Effective Results Reports through the Systematic Design of Experiments**. Hamburg (Germany), 2011. v. 001.

MANLY, B. **Métodos estatísticos multivariados: uma introdução**. Bookman, 2008. ISBN 9788577801855. Disponível em: <<http://books.google.com.br/books?id=t6x2PgAACAAJ>>.

MANSKI, C. F. **Economic Analysis of Social Interactions**. [S.l.], March 2000. (Working Paper Series, 7580). Disponível em: <<http://www.nber.org/papers/w7580>>.

MARTIN, A. D.; QUINN, K. M. Dynamic ideal point estimation via markov chain monte carlo for the u.s. supreme court, 1953–1999. **Political Analysis**, v. 10, n. 2, p. 134–153, 2002.

MARTIN, A. D. et al. Competing approaches to predicting supreme court decision making. **Perspectives on Politics**, v. 2, n. 04, p. 761–767, 2004. Disponível em: <<http://dx.doi.org/10.1017/S1537592704040502>>.

MATTILA, M. Why bother? determinants of turnout in the european elections. **Electoral Studies**, v. 22, n. 3, p. 449 – 468, 2003. ISSN 0261-3794. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0261379401000543>>.

METROPOLIS, N.; ULAM, S. M. The Monte Carlo Method. **Journal of the American Statistical Association**, American Statistical Association, v. 44, n. 247, p. 335–341, set. 1949. ISSN 01621459.

MITCHELL, T. M. **Machine Learning**. 1. ed. New York, NY, USA: McGraw-Hill, Inc., 1997. ISBN 0070428077, 9780070428072.

MONTGOMERY, D. C. **Design and Analysis of Experiments**. [S.l.]: John Wiley Sons, 2006. ISBN 0470088109.

MORETTIN, P. A.; BUSSAB, W. d. O. **Estatística Básica**. 5. ed. [S.l.]: Saraiva, 2004.

MURPHY, W. F. Book. **Elements of judicial strategy / Walter F. Murphy**. [S.l.]: University of Chicago Press, Chicago :, 1964. xiii, 249 p. ; p. ISBN 0226553701 0226553698.

NAGELKERKE, N. J. D. A note on a general definition of the coefficient of determination. **Biometrika**, Oxford University Press, v. 78, n. 3, p. 691–692, set. 1991. ISSN 1464-3510. Disponível em: <<http://dx.doi.org/10.1093/biomet/78.3.691>>.

PECK, S. **A Behavioral Approach to the Judicial Process: Scalogram Analysis**. Osgoode Hall Law School of York University, 1967. Disponível em: <<http://books.google.com.br/books?id=NOUZHQAACAAJ>>.

PENNINGTON, N.; HASTIE, R. Evidence evaluation in complex decision making. **Journal of Personality and Social Psychology**, v. 51, n. 2, p. 242–258, 1986.

PENNINGTON, N.; HASTIE, R. Explanation-Based Decision-Making - effects of memory structure on judgment. **J Exp Psychol Learn J Exp Psychol Learn**, v. 14, n. 3, p. 521–533, 1988. Disponível em: <://A1988P126100015>.

PENROD, S.; HASTIE, R.; ASSOCIATION, A. P. **Models of jury decision making: a critical review**. American Psychological Association, 1979. Disponível em: <<http://books.google.com.br/books?id=ACxGAQAIAAJ>>.

POOLE, K. **Spatial Models of Parliamentary Voting**. [S.l.]: Cambridge University Press, 2005. (Analytical Methods for Social Research). ISBN 9780521851947.

POOLE, K. et al. Scaling roll call votes with wnominate in R. **Journal of Statistical Software**, v. 42, n. 14, p. 1–21, 2011. Disponível em: <<http://www.jstatsoft.org/v42/i14/>>.

PRITCHETT, C. **The Roosevelt Court: a study in judicial politics and values, 1937-1947**. Macmillan Co., 1948. (Quadrangle paperbacks). Disponível em: <<http://books.google.com.br/books?id=0PU-AAAAIAAJ>>.

QUINLAN, J. Induction of decision trees. **Machine Learning**, Kluwer Academic Publishers, v. 1, n. 1, p. 81–106, 1986. ISSN 0885-6125. Disponível em: <<http://dx.doi.org/10.1007/BF00116251>>.

QUINLAN, J. R. **C4.5: programs for machine learning**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993. ISBN 1-55860-238-0.

RAILSBACK, S. F.; GRIMM, V. **Agent-Based and Individual-Based Modeling: A Practical Introduction**. [S.l.]: Princeton Univ Press, 2011.

RIBEIRO, R. Preferências, custos da decisão e normas jurídicas no processo decisório das cortes: o modelo de múltiplos comportamentos. **Economic Analysis of Law Review**, v. 2, n. 2, 2011. ISSN 1980-8860.

ROSS, J.; CLIFF, N. A generalization of the interpoint distance model. **Psychometrika**, v. 29, n. 2, p. 167–176, 1964. Disponível em: <<http://EconPapers.repec.org/RePEc:spr:psycho:v:29:y:1964:i:2:p:167-176>>.

RUSSELL, P. H. Book. **The Supreme Court of Canada as a bilingual and bicultural institution [by] Peter H. Russell**. [S.l.]: Queen's Printer [Ottawa, 1969. xix, 282 p. p.

SCHANK, R. C. **Tell Me a Story: Narrative and Intelligence**. [S.l.]: Northwestern University Press, 1995. ISBN 0810113139.

SCHUBERT, G. A. Book. **The judicial mind : the attitudes and ideologies of Supreme Court justices, 1946-1963 / by Glendon Schubert**. [S.l.]: Northwestern U.P., Evanston, Ill. :, 1965. xiii, 295 p. : p.

SEGAL, J. A. et al. Ideological values and the votes of u.s. supreme court justices revisited. **The Journal of Politics**, v. 57, p. 812–823, 1995. ISSN 1468-2508. Disponível em: <<http://dx.doi.org/10.2307/2960194>>.

STOCK, J. H.; WATSON, M. W. **Econometria**. [S.l.]: Pearson, 2004.

TANFORD, S.; PENROD, S. Computer modeling of influence in the jury: The role of the consistent juror. **Social Psychology Quarterly.American Sociological Association.**, v. 46, n. 3, p. 200–212, 1983.

WOOLDRIDGE, J. **Introductory Econometrics: A Modern Approach (with Economic Applications, Data Sets, Student Solutions Manual Printed Access Card)**. 4. ed. [S.l.]: South-Western College Pub, 2008. ISBN 0324581629.

WU, C.; HAMADA, M.; WU, C. **Experiments: planning, analysis, and parameter design optimization**. [S.l.]: Wiley New York, 2000.

YOUNG, G.; HOUSEHOLDER, A. Discussion of a set of points in terms of their mutual distances. **Psychometrika**, v. 3, n. 1, p. 19–22, 1938. Disponível em: <<http://EconPapers.repec.org/RePEc:spr:psycho:v:3:y:1938:i:1:p:19-22>>.

APÊNDICE A – CONVERSÃO DAS VARIÁVEIS CATEGÓRICAS

É necessário o conhecimento da ordem das variáveis e como estas foram codificadas para a localização dos coeficientes no Apêndice B deste trabalho. A categorização das variáveis pode ser verificada nas figuras 35 e 36. Vale ressaltar que a sigla “EST”, presente na figura 36, significa que o membro do colegiado é estrangeiro e não nasceu em nenhum estado do Brasil.

Variável	Valor	Nº	Variável	Valor	Nº
EstadoProcesso	AL	1	EstadoProcesso	AM	2
EstadoProcesso	BA	3	EstadoProcesso	CE	4
EstadoProcesso	DF	5	EstadoProcesso	ES	6
EstadoProcesso	GO	7	EstadoProcesso	MA	8
EstadoProcesso	MG	9	EstadoProcesso	MT	10
EstadoProcesso	PA	11	EstadoProcesso	PB	12
EstadoProcesso	PE	13	EstadoProcesso	PI	14
EstadoProcesso	PR	15	EstadoProcesso	RJ	16
EstadoProcesso	RN	17	EstadoProcesso	RO	18
EstadoProcesso	RR	19	EstadoProcesso	RS	20
EstadoProcesso	SC	21	EstadoProcesso	SE	22
EstadoProcesso	SP	23	EstadoProcesso	TO	24
Presidente	Dilma	1	Presidente	Fernando Henrique	2
Presidente	Lula	3	TipoReu	Pessoa Jurídica	1
TipoReu	Estado	2	TipoReu	Pessoa Física	3
TipoAutor	Pessoa Jurídica	1	TipoReu	Estado	2
TipoAutor	Pessoa Física	3	Colegiado	Corte Especial	1
Colegiado	Primeira Turma	2	Colegiado	Segunda Turma	3

Figura 35: Variáveis categóricas do modelo de regressão.

Fonte: Própria.

Variável	Valor	Nº	Variável	Valor	Nº
Faculdade	FDSM	1	Faculdade	PUCMINAS	2
Faculdade	PUCSP	3	Faculdade	UCG	4
Faculdade	UERJ	5	Faculdade	UFAL	6
Faculdade	UFBA	7	Faculdade	UFC	8
Faculdade	UFJF	9	Faculdade	UFMG	10
Faculdade	UFPE	11	Faculdade	UFPR	12
Faculdade	UFRGS	13	Faculdade	UFRJ	14
Faculdade	UFRN	15	Faculdade	UFSC	16
Faculdade	UGF	17	Faculdade	UNIBENNET	18
EstadoMembro	AL	1	EstadoMembro	AM	2
EstadoMembro	BA	3	EstadoMembro	CE	4
EstadoMembro	EST	5	EstadoMembro	GO	6
EstadoMembro	MG	7	EstadoMembro	PB	8
EstadoMembro	PE	9	EstadoMembro	PR	10
EstadoMembro	RJ	11	EstadoMembro	RN	12
EstadoMembro	RS	13	EstadoMembro	SC	14
EstadoMembro	SP	15			

Figura 36: Variáveis categóricas do modelo de regressão.

Fonte: Própria.

APÊNDICE B – COEFICIENTES DO MODELO DE REGRESSÃO LOGÍSTICA

Os coeficientes da regressão logística podem ser verificados na coluna *B* das figuras 37, 38 e 39.

	B	S.E.	Wald	df	Sig.	Exp(B)
Ano de Nascimento	-2,019	1,602	1,588	1	,208	,133
Colegiado			38,312	2	,000	
Colegiado(1) - Corte Especial	,066	,179	,137	1	,712	1,068
Colegiado(2) - 1ª Turma	-,604	,109	30,918	1	,000	,547
Estado do Membro			3,251	4	,517	
Estado do Membro(1) - AL	-64,962	52,097	1,555	1	,212	,000
Estado do Membro(5) - EST	-8,364	6,801	1,512	1	,219	,000
Estado do Membro(10) - PR	13,828	11,492	1,448	1	,229	1013064,113
Estado do Membro(13) - RJ	-33,377	26,228	1,619	1	,203	,000
Estado do Processo			99,649	23	,000	
Estado do Processo(1) - AL	,910	,908	1,003	1	,317	2,484
Estado do Processo(2) - AM	,283	1,006	,079	1	,778	1,327
Estado do Processo(3) - BA	,408	,911	,200	1	,655	1,503
Estado do Processo(4) - CE	,295	,906	,106	1	,745	1,342
Estado do Processo(5) - DF	1,366	,900	2,304	1	,129	3,921
Estado do Processo(6) - ES	1,196	,967	1,531	1	,216	3,307
Estado do Processo(7) - GO	,726	,947	,588	1	,443	2,067
Estado do Processo(8) - MA	2,018	,989	4,169	1	,041	7,526
Estado do Processo(9) - MG	,871	,898	,940	1	,332	2,388
Estado do Processo(10) - MT	-,148	1,034	,020	1	,886	,862
Estado do Processo(11) - PA	2,000	,975	4,207	1	,040	7,386
Estado do Processo(12) - PB	,275	,929	0,088	1	,767	1,317
Estado do Processo(13) - PE	,755	,906	0,694	1	,405	2,127
Estado do Processo(14) - PI	-,416	1,100	,143	1	,705	,660
Estado do Processo(15) - PR	,637	,896	,505	1	,477	1,890
Estado do Processo(16) - RJ	,995	,898	1,228	1	,268	2,706

Figura 37: Variáveis do modelo de regressão e seus respectivos coeficientes (B).

Fonte: Própria.

	B	S.E.	Wald	df	Sig.	Exp(B)
Estado do Processo(17) - RN	,774	,941	,676	1	,411	2,169
Estado do Processo(18) - RO	-20,266	23167,308	,000	1	,999	,000
Estado do Processo(19) - RR	,328	1,535	,046	1	,831	1,388
Estado do Processo(20) - RS	,663	,895	,549	1	,459	1,941
Estado do Processo(21) - SC	,674	,896	,566	1	,452	1,963
Estado do Processo(22) - SE	,798	1,000	,636	1	,425	2,220
Estado do Processo(23) - SP	,911	,896	1,033	1	,309	2,487
Faculdade			115,945	16	,000	
Faculdade(1) - FDSM	-20,763	16,902	1,509	1	,219	,000
Faculdade(2) - PUCMINAS	-91,535	73,115	1,567	1	,211	,000
Faculdade(3) - PUCSP	-69,853	55,661	1,575	1	,209	,000
Faculdade(4) - UCG	-19,009	14,628	1,689	1	,194	,000
Faculdade(5) - UERJ	24,962	19,201	1,690	1	,194	69307568929
Faculdade(6) - UFAL	92,083	73,434	1,572	1	,210	9,801
Faculdade(7) - UFBA	25,562	20,483	1,557	1	,212	1,263
Faculdade(8) - UFC	32,227	25,453	1,603	1	,205	9,906
Faculdade(9) - UFJF	-18,099	14,285	1,605	1	,205	,000
Faculdade(10) - UFMG	21,125	16,824	1,577	1	,209	1494975685
Faculdade(11) - UFPE	6,685	4,793	1,945	1	,163	800,480
Faculdade(12) - UFPR	7,490	5,690	1,732	1	,188	1789,717
Faculdade(13) - UFRGS	36,067	28,535	1,598	1	,206	4,608
Faculdade(14) - UFRJ	38,752	30,562	1,608	1	,205	6,759
Faculdade(15) - UFRN	-24,710	20,949	1,391	1	,238	,000
Faculdade(17) - UGF	-34,793	27,605	1,589	1	,208	,000
Gênero Sexual	32,415	26,425	1,505	1	,220	1,195

Figura 38: Variáveis do modelo de regressão e seus respectivos coeficientes (B).

Fonte: Própria.

	B	S.E.	Wald	df	Sig.	Exp(B)
Grau	-5,290	4,190	1,594	1	,207	,005
Ideologia	25,826	21,254	1,476	1	,224	1,645
Indicador Ano de Nascimento	-,005	,014	,114	1	,736	,995
Indicador Gênero Sexual	,237	,348	,464	1	,496	1,267
Indicador Grau	-,091	,207	,194	1	,660	,913
Indicador Ideologia	,591	,234	6,394	1	,011	1,806
Indicador Raça	-,214	,987	,047	1	,828	,807
Indicador Tipo Faculdade	-,353	,390	,819	1	,365	,703
Presidente			28,373	2	,000	
Presidente(1) - Dilma Rousseff	,678	,233	8,439	1	,004	1,970
Presidente(2) - Fernando Henrique	-,728	,144	25,662	1	,000	,483
Raça	10,068	8,021	1,575	1	,209	23578,378
Relator	,077	,069	1,248	1	,264	1,080
Tipo de Autor			5,328	2	,070	
Tipo de Autor(1) - Pessoa Jurídica	,065	,101	,408	1	,523	1,067
Tipo de Autor(2) - Estado	-,174	,125	1,936	1	,164	,840
Tipo de Réu			67,162	2	,000	
Tipo de Réu(1) - Pessoa Jurídica	,564	,072	62,161	1	,000	1,758
Tipo de Réu(2) - Estado	,157	,109	2,082	1	,149	1,170
Tipo Faculdade	-47,197	37,842	1,556	1	,212	,000
Voto do Relator	,174	,053	10,969	1	,001	1,191
Constante	3932,497	3114,553	1,594	1	,207	

Figura 39: Variáveis do modelo de regressão e seus respectivos coeficientes (B).

Fonte: Própria.