

**INSTITUTO FEDERAL GOIANO – CAMPUS CERES
BACHARELADO EM SISTEMAS DE INFORMAÇÃO
ANA PAULA TEODORO DE MATOS**

**APLICAÇÃO DE TÉCNICA DE MINERAÇÃO DE DADOS COMO PERSPECTIVA
DE ANÁLISE DO CENÁRIO PARTIDÁRIO DAS ELEIÇÕES MUNICIPAIS DO
INTERIOR GOIANO**

**CERES – GO
2021**

ANA PAULA TEODORO DE MATOS

**APLICAÇÃO DE TÉCNICA DE MINERAÇÃO DE DADOS COMO PERSPECTIVA
DE ANÁLISE DO CENÁRIO PARTIDÁRIO DAS ELEIÇÕES MUNICIPAIS DO
INTERIOR GOIANO**

Trabalho de curso apresentado ao curso de Sistemas de Informação do Instituto Federal Goiano – Campus Ceres, como requisito parcial para a obtenção do título de Bacharel em Sistemas de Informação, sob orientação da Prof^a. Ramayane Bonacin Braga.

CERES – GO

2021

Sistema desenvolvido pelo ICMC/USP
Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas - Instituto Federal Goiano

Ma Matos, Ana Paula
 Aplicação de Técnica de Mineração de Dados como
 Perspectiva de Análise do Cenário Partidário das
 Eleições Municipais do Interior Goiano / Ana Paula
 Matos; orientadora Ramayane Braga. -- Ceres, 2021.
 53 p.

Monografia (Graduação em Sistemas de Informação) --
Instituto Federal Goiano, Campus Ceres, 2021.

1. Mineração de dados. 2. Dados eleitorais. 3.
Classificação. 4. Associação. 5. Eleições Municipais.
I. Braga, Ramayane, orient. II. Título.



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO



Repositório Institucional do IF Goiano - RIIF Goiano
Sistema Integrado de Bibliotecas

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR PRODUÇÕES TÉCNICO-CIENTÍFICAS NO REPOSITÓRIO INSTITUCIONAL DO IF GOIANO

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia Goiano, a disponibilizar gratuitamente o documento no Repositório Institucional do IF Goiano (RIIF Goiano), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IF Goiano.

Identificação da Produção Técnico-Científica

- | | |
|--|---|
| <input type="checkbox"/> Tese | <input type="checkbox"/> Artigo Científico |
| <input type="checkbox"/> Dissertação | <input type="checkbox"/> Capítulo de Livro |
| <input type="checkbox"/> Monografia - Especialização | <input type="checkbox"/> Livro |
| <input checked="" type="checkbox"/> TCC - Graduação | <input type="checkbox"/> Trabalho Apresentado em Evento |
| <input type="checkbox"/> Produto Técnico e Educacional - Tipo: _____ | |

Nome Completo do Autor: Ana Paula Teodoro de Matos

Matrícula: 2017103202030384

Título do Trabalho: APLICAÇÃO DE TÉCNICA DE MINERAÇÃO DE DADOS COMO PERSPECTIVA DE ANÁLISE DO CENÁRIO PARTIDÁRIO DAS ELEIÇÕES MUNICIPAIS DO INTERIOR GOIANO

Restrições de Acesso ao Documento

Documento confidencial: Não Sim, justifique: _____

Informe a data que poderá ser disponibilizado no RIIF Goiano: 19/03/2021

O documento está sujeito a registro de patente? Sim Não

O documento pode vir a ser publicado como livro? Sim Não

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O/A referido/a autor/a declara que:

1. o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
2. obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia Goiano os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
3. cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia Goiano.

Ceres, 19 de março de 2021.

(Assinado Eletronicamente)

Assinatura do Autor e/ou Detentor dos Direitos Autorais

Ciente e de acordo:

(Assinado Eletronicamente)

Assinatura da orientadora: Ramayane Bonacin Braga (1189232)

Ceres, 19 de março de 2021.

Documento assinado eletronicamente por:

- Ana Paula Teodoro de Matos, 2017103202030384 - Discente, em 19/03/2021 09:13:07.
- Ramayane Bonacin Braga, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 19/03/2021 09:08:48.

Este documento foi emitido pelo SUAP em 19/03/2021. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifgoiano.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 250802

Código de Autenticação: f9edbf9fe5



INSTITUTO FEDERAL GOIANO

Campus Ceres

Rodovia GO-154, Km.03, Zona Rural, None, CERES / GO, CEP 76300-000

(62) 3307-7100



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO

ATA DE DEFESA DE TRABALHO DE CURSO

Aos oito dias do mês de março do ano de dois mil e vinte e um realizou-se a defesa de Trabalho de Curso da acadêmica **Ana Paula Teodoro de Matos**, do Curso de Bacharelado em Sistemas de Informação, matrícula 2017103202030384, cujo título é **“Aplicação de Técnica de Mineração de Dados como Perspectiva de Análise do Cenário Partidário das Eleições Municipais do Interior Goiano”**. A defesa iniciou-se às 19 horas e 01 minuto, finalizando-se às 19 horas e 31 minutos. A banca examinadora considerou o trabalho APROVADO com média 7,7 no trabalho escrito, média 9,1 no trabalho oral, apresentando assim média aritmética final de 8,4 pontos, estando a estudante APTA para fins de conclusão do Trabalho de Curso.

Após atender às considerações da banca e respeitando o prazo disposto em calendário acadêmico, a estudante deverá fazer a submissão da versão corrigida em formato digital (.pdf) no Repositório Institucional do IF Goiano – RIIF, acompanhado do Termo Ciência e Autorização Eletrônico (TCAE), devidamente assinado pela autora e orientadora.

Os integrantes da banca examinadora assinam a presente.

(Assinado Eletronicamente)

Ramayane Bonacin Braga
Presidenta da Banca Examinadora

(Assinado Eletronicamente)

Indiana Esteva Gontijo
Membro da Banca Examinadora

(Assinado Eletronicamente)

Fabiola Gonçalves Coelho Ribeiro
Membro 2 Banca Examinadora

Documento assinado eletronicamente por:

- Fabiola Gonçalves Coelho Ribeiro, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 10/03/2021 15:52:04.
- Indiana Esteva Gontijo, PROF ENS BAS TEC TECNOLOGICO-SUBSTITUTO, em 08/03/2021 21:40:21.
- Ramayane Bonacin Braga, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 08/03/2021 21:24:29.

Este documento foi emitido pelo SUAP em 08/03/2021. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifgoiano.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 246888
Código de Autenticação: 7861a30030



Dedico este trabalho a todos que contribuíram para a sua realização e pelo apoio incondicional.

AGRADECIMENTOS

Agradeço, primeiramente, à minha mãe, Maria Aparecida de Matos, por todo amor e suporte que me deu desde meu nascimento e a quem devo muitas de minhas conquistas. Foi imprescindível para que eu chegasse até aqui. À minha orientadora, Prof^a. Ramayane Bonacin Braga, pelo voto de confiança, pela paciência e pelas orientações ao longo deste trabalho e da graduação como um todo. Aos meus colegas discentes que tive ao longo da carreira acadêmica, pela inspiração contínua. Muito obrigada!

“Aprendemos quando resolvemos nossas dúvidas, superamos
nossas incertezas, satisfazemos nossa curiosidade”.

Maria Teresa Mantoan

RESUMO

Atualmente, o Brasil é um país de referência mundial nas eleições devido seu processo de modernização por urnas eletrônicas totalmente informatizadas desde o ano de 2000, garantindo assim, a segurança dos votos e a liberdade democrática. Em cada eleição, uma grande quantidade de dados é gerada e, com base em números estatísticos, são efetuados estudos para construir aplicações ou descobrir novas informações. Este trabalho tem como objetivo efetuar uma pesquisa para aplicação de técnicas de descoberta de conhecimento em mineração de dados, visando encontrar uma possível tendência a ser observada no perfil do eleitorado das zonas eleitorais 039 e 072 no Estado de Goiás. Com o auxílio da ferramenta WEKA para realizar essa tarefa, é analisado um conjunto de dados de votantes e partidos eleitorais em âmbito municipal, ocorrido entre os anos de 2000 a 2016, obtidos a partir do repositório de dados disponibilizado pelo TSE (Tribunal Superior Eleitoral) e, posteriormente, é aplicada a técnica de Árvore de Decisão utilizando o algoritmo de classificação de dados J48, para que possam ser geradas regras de associações. Por fim, são apresentados os resultados encontrados através de quatro experimentos efetuados, explicitando se determinados padrões no perfil do eleitorado possuem influência na vitória de determinado pleito eleitoral partidário das eleições municipais das zonas aludidas.

Palavras-chave: Mineração de dados. Dados eleitorais. Classificação. Associação. Eleições Municipais.

ABSTRACT

Currently, Brazil is a country of world reference in elections due to its modernization process by fully computerized electronic ballot boxes since 2000, thus guaranteeing the security of votes and democratic freedom. In each election, a large amount of data is generated and, based on statistical numbers, studies are carried out to build applications or discover new information. This work aims to conduct a research to apply knowledge discovery techniques in data mining, aiming to find a possible trend to be observed in the electorate profile of electoral zones 039 and 072 in the State of Goiás. With the help of the WEKA tool to perform this task, a set of data from voters and electoral parties at the municipal level, which took place between 2000 and 2016, obtained from the data repository made available by the TSE (Superior Electoral Court), was subsequently applied to Decision Tree technique using the J48 data classification algorithm, so that association rules can be generated. Finally, the results found through four experiments carried out are presented, explaining whether certain patterns in the electorate profile have an influence on the victory of a given party electoral election in the municipal elections in the areas mentioned.

Keywords: Data mining. Electoral data. Classification. Association. Municipal Elections.

LISTA DE ABREVIATURAS E SIGLAS

KDD – (Knowledge Discovery in Databases) Descoberta de Conhecimento em Base de Dados

PFL - Partido da Frente Liberal

PMDB - Partido do Movimento Democrático Brasileiro

PP - Partido Progressista

PSD - Partido Social Democrático

PSDB - Partido da Social Democracia Brasileira

PSDC - Partido Social Democrata Cristão

PT - Partido dos Trabalhadores

PTB - Partido Trabalhista Brasileiro

PTC - Partido Trabalhista Cristão

PT do B - Partido Trabalhista do Brasil

PTN - Partido Trabalhista Nacional

PR - Partido Republicano

PV - Partido Verde

TER - Tribunal Regional Eleitoral

TSE - Tribunal Superior Eleitoral

UF - Unidade Federativa

XSLX – Extensão de arquivo em formato do Microsoft Excel

WEKA - Ambiente Waikato para análise de conhecimento

LISTA DE ILUSTRAÇÕES

Figura 1 - Etapas do processo de KDD.....	20
Figura 2 - Registros de atributos por linha.....	25
Figura 3 - Tela principal do software WEKA.....	32
Figura 4 - Árvore de decisão utilizando o algoritmo J48.....	33
Figura 5 - Árvore de agrupamento do conjunto (atributo “sexo”).....	35
Figura 6 - Árvore de decisão (Parte 1).....	36
Figura 7 - Árvore de decisão (Parte 2).....	37
Figura 8 - Árvore de decisão (Parte 3).....	37
Figura 9 - Árvore de agrupamento do conjunto (atributo “grau de escolaridade”)....	38
Figura 10 - Árvore de decisão (Parte 4).....	39
Figura 11 - Árvore de decisão (Parte 5).....	39
Figura 12 - Árvore de decisão (Parte 6).....	40
Figura 13 - Árvore de agrupamento do conjunto (atributo “município”).....	41
Figura 14 - Árvore de decisão (Parte 7).....	42
Figura 15 - Árvore de decisão (Parte 8).....	43
Figura 16 - Árvore de decisão (Parte 9).....	44

Figura 17 - Árvore de agrupamento do conjunto (atributo “faixa etária”).....	45
Figura 18 - Árvore de decisão (Parte 10).....	45
Figura 19 - Árvore de decisão (Parte 11).....	46
Figura 20 - Árvore de decisão (Parte 12).....	47

LISTA DE TABELAS

Tabela 1- Partidos que tiveram maior depósito de votos em cada eleição municipal para o cargo de chefe do poder executivo.....	26
--	----

SUMÁRIO

1 INTRODUÇÃO.....	18
2 REVISÃO DA LITERATURA.....	19
2.1 KDD.....	20
2.2 MINERAÇÃO DE DADOS.....	21
2.3 ÁRVORE DE DECISÃO.....	21
2.4 REGRAS DE ASSOCIAÇÃO.....	22
2.4.1 Algoritmo J48.....	24
3 MATERIAIS E MÉTODOS.....	25
3.1 METODOLOGIA UTILIZADA.....	25
3.2 FERRAMENTAS UTILIZADAS.....	30
3.3 PRÉ-PROCESSAMENTO E LIMPEZA.....	30
3.4 TRANSFORMAÇÃO DOS DADOS.....	31
3.5 MINERAÇÃO DOS DADOS.....	32
4 RESULTADOS E DISCUSSÃO.....	34
4.1 EXPERIMENTO 1.....	34
4.1.1 Resultados Obtidos.....	34
4.1.2 Análise do Resultado.....	35

4.2 EXPERIMENTO 2.....	38
4.2.1 Resultados Obtidos.....	38
4.2.2 Análise do Resultado.....	39
4.3 EXPERIMENTO 3.....	40
4.3.1 Resultados Obtidos.....	40
4.3.2 Análise do Resultado.....	41
4.4 EXPERIMENTO 4.....	44
4.4.1 Resultados Obtidos.....	44
4.4.2 Análise do Resultado.....	45
4.5 INTERPRETAÇÃO QUALITATIVA DOS RESULTADOS.....	48
5 CONSIDERAÇÕES FINAIS.....	50
REFERÊNCIAS.....	51

1. INTRODUÇÃO

Atualmente, alguns conceitos como KDD (do inglês, *Knowledge Discovery in Databases*) e as tarefas de Mineração de Dados são imprescindíveis para a compreensão de tendências, e diversas áreas fazem proveito dessas técnicas, como setores de vendas, administração, investimentos, saúde, educação, política, entre outros. Em tese, setores que disponham de um banco com diversos dados e uma capacidade razoável de processamento, pode ser uma área aplicável de KDD.

Um processo não trivial de identificação de padrões, que tem por objetivo a descoberta de conhecimento em banco de dados, a fim de extrair informações implícitas e potencialmente úteis (FAYYAD et. al., 1996). Este perfil de aplicação de descoberta de conhecimento se torna viável, pois é o mesmo perfil desejado para esse estudo.

Este trabalho se deu com intuito de realizar uma análise de elementos de utilidade pública para cidadãos de uma determinada região de municípios da unidade federativa do Estado de Goiás, selecionando na base todos os dados de eleições municipais referentes às Zonas 039 e 072. Através das experimentações utilizando o algoritmo J48, essa proposta anseia por destacar e reforçar um conhecimento implícito que condiz com a realidade vivida pelos eleitores estudados ao decorrer dos anos objetivando agregar informações satisfatórias e certas.

Após o acesso aos dados das eleições municipais de 2000 a 2016, foi necessário avaliar o nível de relevância de cada variável para a construção do resultado final, que é analisar se há relação do perfil do eleitorado com os partidos que obtiveram vitórias nas eleições. Para isso foi utilizada a ferramenta WEKA (do inglês *Waikato Environment for Knowledge Analysis*), uma suíte totalmente gratuita, desenvolvida pela Universidade de Waikato (Nova Zelândia).

Além disso, apropriada para o objeto de estudo por possuir diversos algoritmos nativos em sua biblioteca, os dados foram obtidos em formato CSV (do inglês, *Comma-separated values*) e estes foram importados ao WEKA, dando início, portanto, às análises descritas nos próximos capítulos almejando estudar eventuais tendências relevantes.

2. REVISÃO DE LITERATURA

Nos dias atuais, o Brasil vive um período democrático, marcado principalmente pelo amplo direito de voto por parte do cidadão e, a cada mês, o TSE - Tribunal Superior Eleitoral consolida dados estatísticos sobre o eleitorado brasileiro, deixando as informações de forma pública para serem extraídas do Cadastro Nacional de Eleitores. De acordo com as estatísticas da Justiça Eleitoral, o eleitorado de votantes teve um crescimento de 3,14% em 2018 se comparado aos números de 2014, constatando uma evolução para os brasileiros que estavam em condições de votar (TSE, 2018).

A teoria da escolha racional, idealizada por Anthony Downs, na segunda metade da década de 50, afirma que o comportamento eleitoral acontece conforme um auto interesse, em que o indivíduo direciona seu voto para maximizar seus próprios ganhos e reduzir seus custos (LAVAREDA, 2011). Sendo assim, o eleitorado se posiciona como um juiz do governo: se a economia flui de maneira positiva, os governantes ganham mais votos; se não, a oposição se beneficia. As ideologias, a identificação partidária e os valores dão espaço a um sistema de interesses.

Na bibliografia especializada, utilizando *data mining* com bases de dados eleitorais, tem-se três trabalhos recentes que se encaixam nesse contexto. No trabalho descrito em Kira (2014), foi aplicado várias técnicas e ferramentas de Mineração de Dados a fim de estudar a presença de padrões indicando a condição de eleito ou não eleito de um candidato. Enquanto em Pereira (2013) temos o estudo de KDD para encontrar possíveis protótipos em dados eleitorais de candidatos a prefeito das eleições do ano de 2012, utilizando a ferramenta WEKA e diversas técnicas. Guedes (2018) aplica técnicas de mineração de dados com o objetivo de prever campanhas eleitorais.

A diferença de proposta para esse estudo está em o foco central ser voltado ao eleitorado de apenas duas Zonas Eleitorais do Estado de Goiás (039 e 072) e na utilização de uma técnica e um algoritmo de aprendizado de máquina específicos, sucessivamente: Árvore de decisão e J48.

2.1. KDD

“A descoberta de conhecimento em bases de dados tem como objetivo encontrar padrões intrínsecos aos dados nela contidos, apresentando-os de forma a facilitar sua assimilação como conhecimento”. (SILVA; PERES; BOSCARIOLI, 2016, p.11). Encontrar o que até então estava desconhecido remete-se ao ato de uma descoberta, considerando que geralmente as bases de dados são extensas e que o conhecimento pode estar implícito, é necessário um trabalho de busca detalhado, o que é denominado de mineração de dados explorando processos analíticos, sistemáticos e se possível, automatizados (SILVA; PERES; BOSCARIOLI, 2016).

Embora os termos KDD e *Data Mining* sejam frequentemente entendidos como sinônimos, é importante frisar que o KDD é um processo de várias etapas e maior que *Data Mining*, compreendendo a identificação de padrões a partir da existência de dados, já a Mineração de Dados é uma das etapas do processo (FAYYAD et. al., 1996).

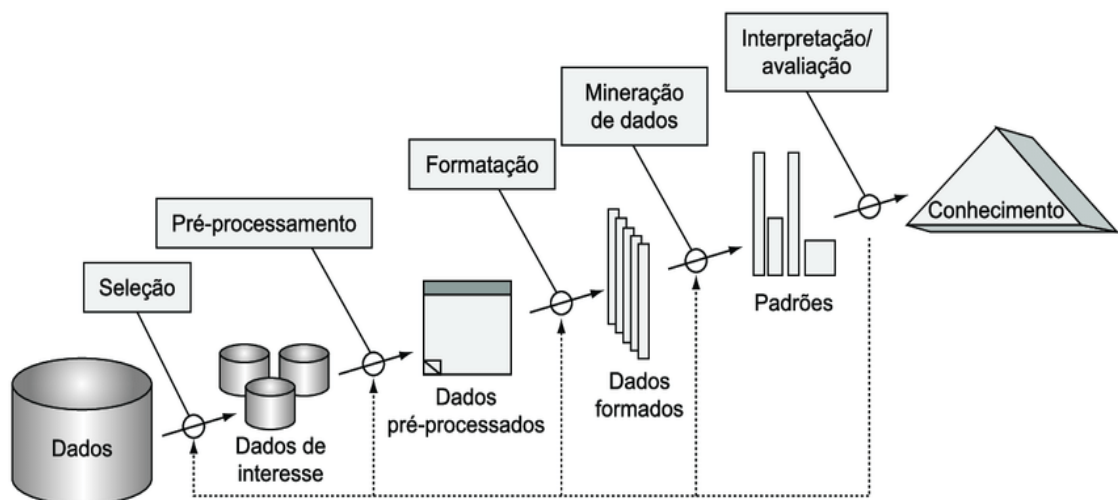


Figura 1 - Etapas do processo de KDD.

Fonte: (FAYYAD et. al., 1996).

Na Figura 1, pode-se visualizar as fases do processo de KDD, que se inicia com a organização da base de dados que contém os dados da área de interesse, a fim de que seja descoberto algum tipo de conhecimento útil. Esses dados passam por uma série de procedimentos de seleção aplicados ao pré-processamento como, por exemplo, a organização de um único repositório, eliminadas as instâncias repetidas e selecionados os valores dos atributos que sejam considerados relevantes para o algoritmo de análise. Com os dados transformados, inicia-se a tarefa de mineração de

dados, envolvendo a resolução, interpretação e avaliação de possíveis padrões de conhecimento que podem ser formatados em gráficos, tabelas e relatórios estruturados.

2.2. MINERAÇÃO DE DADOS

Sobre Mineração de Dados (do inglês Data Mining) podemos afirmar que

é definida em termos de esforços para descoberta de padrões em bases de dados. A partir dos padrões descobertos, têm-se condições de gerar conhecimento útil para um processo de tomada de decisão. Trata-se, portanto, da aplicação de técnicas, implementadas por meio de algoritmos computacionais, capazes de receber, como entrada, um conjunto de fatos ocorridos no mundo real e devolver, como saída, um padrão de comportamento, o qual pode ser expresso, por exemplo, como uma regra de associação, uma função de mapeamento ou a modelagem de um perfil. (SILVA; PERES; BOSCAROLI, 2016, p.11).

Essa etapa é a principal do processo de KDD e compreende a busca efetiva por conhecimentos implícitos e úteis. Para os autores, mineração de dados é a combinação entre inteligência artificial, análise estatística, aprendizado de máquina e tecnologia de banco de dados, com a finalidade de descobrir padrões relevantes que ocorrem nos dados e que sejam importantes para embasar a assimilação de informação importante, suportando a geração de conhecimento (SILVA; PERES; BOSCAROLI, 2016).

2.3. ÁRVORES DE DECISÃO

As árvores de decisão visam a classificação de exemplos utilizando uma estrutura de árvore (composta de raiz e nós ou folhas) para representar um número de possíveis caminhos de decisão e um resultado para cada caminho, também podem ser consideradas um modelo que possui como principal característica a fácil, compreensível e intuitiva interpretação (GRUS, 2019).

De acordo com Carvalho (2001, p.70), uma árvore é um grafo, estrutura matemática composta por arestas e vértices, não tem ciclos dentro de sua estrutura, e as arestas só podem ser percorridas em uma direção. Os vértices, também conhecidos como nós, guardam informações e as arestas ligam os nós, podendo ter

várias arestas saindo de um nó, porém, apenas uma chegando. Dentre estes nós, alguns são considerados peculiares, quais sejam, as raízes e as folhas.

A raiz é o nó de onde a árvore se origina, por definição, nenhuma aresta chega na raiz. Em compensação, as folhas são o fim da árvore, delas não saem nenhuma aresta. Além disso, podemos definir um caminho pela árvore a partir da escolha de arestas sucessivas começando da raiz e terminando em alguma das folhas, já um nó é considerado filho se existe uma aresta partindo do originário (pai) e chegando nele (CARVALHO, 2001).

Sobre as árvores de decisão podemos afirmar que

podem lidar naturalmente com uma mistura de atributos numéricos e determinantes; podem até classificar os dados para os atributos que estão faltando. Ao mesmo tempo, elas possuem muitas recomendações, sendo assim, encontrar a árvore de decisão perfeita para um conjunto de dados em treinamento é computacionalmente um problema muito complexo. (GRUS, 2019, p. 202).

Sua constituição é viabilizada por algoritmos, dentre os quais, o J48, este algoritmo é dividido em duas partes: expansão e poda. A expansão consiste na geração da árvore a partir de sua raiz, de modo que o conjunto de padrões é herdado do pai. Caso o nó seja a raiz, todos os exemplos podem ser classificados; o atributo e a condição são escolhidos através de um critério de escolha de atributos, normalmente, o ganho de informação, de forma a elevar ao máximo a divisão entre as classes. A poda, que é um método bastante utilizado e eficaz, consiste em melhorar a taxa de acerto do modelo para novos exemplos e facilitar a interpretação por parte do usuário, os quais foram utilizados no treinamento.

2.4. REGRAS DE ASSOCIAÇÃO

Para Silva, Peres e Boscaroli (2016, p.199), as regras de associação são definidas como “o processo de analisar os relacionamentos existentes entre atributos de uma base de dados transacional, com o objetivo de encontrar associações ou correlações”.

Segundo os autores

as regras de associação, independentemente de expressarem associações simples, ordenadas ou correlações, são comumente representadas por meio de afirmações do tipo SE ENTÃO, sendo também interpretadas como implicações do antecedente da regra para o seu conseqüente. Assim, algoritmos usados para descobrir tais regras são baseados no pressuposto de que a presença de um atributo em um evento implica a presença de outro atributo no mesmo evento. As áreas de aplicações para a descoberta de regras de associação são diversas. (Silva, Peres e Boscaroli, 2016, p.199).

As aplicações mais comuns são as relacionadas com análise de compras, entretanto, outros domínios podem se beneficiar da resolução da tarefa de mineração, como a análise de comportamento eleitoral, que é o foco deste trabalho (SILVA, PERES E BOSCARIOLI, 2016). Geralmente, as regras de associação são apresentadas da seguinte forma: SE o eleitor tiver idade superior a 70 anos, ENTÃO possui ensino fundamental incompleto.

A forma de averiguar se uma regra é válida ou não em um domínio sob análise é por meio das medidas de suporte e confiança. A existência de associações ou a correlação entre os atributos implica no suporte, que determina o quão comum os fatos podem ocorrer simultaneamente em uma massa de dados, e na confiança, que determina a frequência de observação um atributo num conjunto de alguns itens aparece em transações contendo um segundo atributo nesse mesmo conjunto de transações (CARVALHO, 2001).

Segundo Silva, Peres e Boscaroli (2016, p.205)

espera-se que os valores obtidos para essas medidas na análise de cada regra de associação estejam acima de limiares mínimos estabelecidos pelo usuário (um limiar mínimo para o suporte e um limiar mínimo para a confiança). Diz-se que uma regra que possui suporte e confiança acima dos respectivos limiares mínimos é uma regra forte e, portanto, considerada válida.

Uma regra de associação com um suporte muito baixo, pode ser um indício de que a associação ocorre apenas ao acaso. Em uma outra perspectiva, o suporte baixo pode não ser interessante do ponto de vista do modelo de negócios, pois o custo em promover algum tipo de política para uma regra com suporte baixo pode não ser muito diferente do custo de promover uma regra com suporte alto. Dadas estas razões, as regras de associação com alta porcentagem de incorrelação, as informações tornam-se irrelevantes para o processo de descoberta de padrões nesse trabalho, que visa analisar os perfis do eleitorado.

Na sequência será introduzido o algoritmo J48, que será utilizado para a execução desse trabalho.

2.4.1. Algoritmo J48

O J48 possui código aberto, foi desenvolvido por J. Ross Quinlan e implementado no software WEKA. É um algoritmo que pode lidar tanto com atributos contínuos e discretos, quanto com valores categóricos e ausentes. O tratamento de atributos contínuos envolve a consideração de todos os valores presentes no conjunto de treinamento, fazendo com que tais dados sejam ordenados de forma crescente através dos valores presentes nos dados de treinamento e, após esta ordenação, seja selecionado o valor que favorecerá a redução da entropia. Já os atributos discretos, possuem um conjunto de valores finito ou contavelmente infinito, frequentemente representados como variáveis inteiras.

Tal algoritmo utiliza dois métodos de poda, o primeiro é conhecido como substituição de subárvore, ou seja, os nós de uma árvore de decisão podem ser substituídos por uma folha, basicamente, diminuindo o número de testes ao longo de determinado percurso. O segundo tipo é denominado de capturação de subárvore que, por sua vez, possibilita que um nó possa ser movido para cima em direção à raiz da árvore, substituindo os outros nós pelo caminho.

No método de construção da árvore de decisão J48, sempre que um conjunto de itens é encontrado, o procedimento identifica o atributo que melhor discrimina os diversos modelos, ou seja, entre os valores possíveis, caso haja qualquer valor que não encontre ambiguidade, esse ramo é encerrado e o resultado obtido é imposto a ele.

Fazendo uma breve explanação, a construção da árvore de decisão pelo algoritmo J48 se dá por meio da abordagem top-down, em que o atributo mais significativo, quando comparado aos outros atributos do conjunto, é considerado a raiz da árvore. Na sequência da construção, o próximo nó da árvore será o segundo atributo mais significativo e assim, sucessivamente, até gerar o nó da folha, que representa o atributo alvo da instância.

3. MATERIAIS E MÉTODOS

Através dos atributos selecionados para análise, disponibilizados em bases de dados pelo TSE, o presente capítulo descreve os materiais e os métodos desenvolvidos para o aprimoramento de padrões do eleitorado das Zonas 039 (Guarinos, Hidrolina, Itapaci, Pilar de Goiás e São Luiz do Norte) e 072 (Ceres, Nova Glória, Rialma, Rianópolis e Santa Isabel) no Estado de Goiás, utilizando uma das principais ferramentas de mineração de dados, a WEKA, responsável por extrair o conhecimento pretendido.

Além disso, foram determinados os métodos descritos ao decorrer dos itens 3.1 a 3.5 e algoritmo J48 que serão utilizados nesse procedimento. Seguindo a técnica de mineração dos dados por Árvore de Decisão, os resultados foram obtidos, a fim de analisar e descobrir prováveis padrões e tendências.

3.1 METODOLOGIA UTILIZADA

A extração dos dados foi feita a partir do Portal do Tribunal Superior Eleitoral, de onde foram obtidas 5 pastas de arquivos na opção “Eleitorado” através de download, referentes às eleições dos anos de 2000, 2004, 2008, 2012 e 2016. Os arquivos estão disponíveis em formato .txt e organizados de forma que, cada um contém um registro por linha com as seguintes informações: município, estado civil, faixa etária, escolaridade, sexo e a quantidade total de eleitores pertencentes ao perfil, deixando tais atributos separados pela simbologia (;) - ponto e vírgula.

Após essa realização de filtragem dos dados disponíveis, para um processo mais acurado, foram selecionadas apenas as zonas 039 e 072, consideradas úteis para o trabalho proposto.

```
"29/09/2016";"23:24:16";"2016";"GO";"94137";"ITAPACI";"39";"12";"3";"CASADO";"7";  
|"45 A 59 ANOS";"8";"SUPERIOR COMPLETO";"2";"MASCULINO";"1"
```

Figura 2 - Registros de atributos por linha.

Fonte: Tribunal Superior Eleitoral (2016).

Na Figura 2 pode-se observar o valor: “29/09/2016” descreve a data de geração do arquivo, “23:24:16” indica a hora da extração. O valor “2016” representa o ano da eleição em questão, “GO”; “94137”; “ITAPACI”; “39”; retrata a sigla da unidade da federação, composta pelo código TSE do município, nome do município e número da zona eleitoral. “12” informa a seção eleitoral, “MASCULINO” representa o sexo, “45 A 59 ANOS” o perfil etário, “SUPERIOR COMPLETO” indica o grau de escolaridade declarada pelo eleitor, “1” a quantidade de eleitores que compõem o perfil, “3”, “7”, “8”, “2” são os dados referentes ao código dos registros seguintes a sua identificação.

Com relação à proposta de mineração de dados para esse trabalho, foram selecionados os partidos que obtiveram êxito em cada eleição municipal estudada para o cargo de chefe do poder executivo, podendo avaliar se sua vitória eleitoral tem relação com algumas características dos cidadãos do município: social, cultural e financeira.

Tabela 1 - Partidos que tiveram maior depósito de votos em cada eleição municipal para o cargo de chefe do poder executivo.

Zona	Ano	Município	Partido	Quant. votos
	2000	Guarinos	PSDC	1058 votos
			PMDB	915 votos
		Hidrolina	PFL	1621 votos
			PMDB	1566 votos
		Itapaci	PREJUDICADO	PREJUDICADO
		Pilar de Goiás	PSDB	1230 votos
	PMDB		1150 votos	
	2004	São Luiz do Norte	PMDB	1826 votos
			PFL	1188 votos
		Guarinos	PSDB	908 votos
			PP	837 votos
		Hidrolina	PMDB	1669 votos
PT			1219 votos	
Itapaci	PT	5669 votos		
	PFL	2416 votos		

039

Pilar de Goiás	PMDB	842 votos
	PSDB	781 votos
São Luiz do Norte	PMDB	1323 votos
	PFL	1215 votos
Guarinos	PTN	1102 votos
	PP	666 votos
Hidrolina	PP	1722 votos
	PT	1540 votos
Itapaci	PMDB	4651 votos
	PSDB	3729 votos
Pilar de Goiás	PSDB	1161 votos
	PMDB	1117 votos
São Luiz do Norte	PP	2134 votos
	PREJUDICADO	PREJUDICADO
Guarinos	PSD	1107 votos
	PT	564 votos
Hidrolina	PMDB	2490 votos
	PSDB	954 votos
Itapaci	PMDB	4981 votos
	PTB	4794 votos
Pilar de Goiás	PMDBP	1034 votos
	PSDB	814 votos
São Luiz do Norte	PMDB	1856 votos
	PP	1601 votos
Guarinos	PSD	1109 votos
	PMDB	927 votos
Hidrolina	PMDB	2149 votos
	PP	1483 votos
Itapaci	PSDB	5852 votos
	PTC	5803 votos
Pilar de Goiás	PMDB	1682 votos
	PV	1162 votos
São Luiz do Norte	PMDB	2155 votos

			DEM	1634 votos
072	2000	Ceres	PP	5908 votos
			PSD	5892 votos
		Nova Glória	PTB	2758 votos
			PFL	2663 votos
		Rialma	PMDB	3082 votos
			PSDB	2274 votos
		Rianópolis	PP	1564 votos
			PSDB	1523 votos
		Santa Isabel	PFL	1464 votos
			PP	654 votos
	2004	Ceres	PSDB	7723 votos
			PP	5025 votos
		Nova Glória	PSDB	3490 votos
			PMDB	2235 votos
		Rialma	PSDB	3409 votos
			PMDB	1290 votos
		Rianópolis	PSDB	1936 votos
			PSC	950 votos
	Santa Isabel	PFL	1505 votos	
		PP	1133 votos	
2008	Ceres	PMDB	6948 votos	
		PP	5067 votos	
	Nova Glória	PP	3356 votos	
		PSDB	2881 votos	
	Rialma	PSDB	3495 votos	
		PP	3311 votos	
	Rianópolis	PSDB	1575 votos	
		PMDB	1544 votos	
	Santa Isabel	DEM	1300 votos	
		PR	1250 votos	
	Ceres	PT	6233 votos	
		PT do B	5580 votos	

2012	Nova Glória	PSDB	3826 votos	
		PTB	2225 votos	
	Rialma	PSDB	3822 votos	
		PSD	3237 votos	
	Rianópolis	PSDB	1935 votos	
		PMDB	1500 votos	
	Santa Isabel	DEM	2389 votos	
		PSDB	843 votos	
	2016	Ceres	PSDB	10228 votos
			PMDB	2596 votos
Nova Glória		PP	3721 votos	
		PMDB	1983 votos	
Rialma		PTB	4168 votos	
		PSDB	2485 votos	
Rianópolis		PSDB	1772 votos	
		PMDB	1386 votos	
Santa Isabel		PSDB	1631 votos	
		PP	1459 votos	

Fonte: ELEIÇÕES ANTERIORES. Tribunal Superior Eleitoral (2020).

A partir das informações expostas na Tabela 1, nota-se os partidos mais bem votados na 1ª e 2ª posição das zonas 039 e 072. Observa-se que em alguns municípios houve maior diferença de votos de um partido para o outro e que em outros municípios a diferença de votos foi inferior, sendo que no ano de eleição municipal seguinte, o partido classificado em 2º lugar na eleição anterior, ganhou as eleições, que é o caso dos municípios de Hidrolina 2000-2004; Pilar de Goiás 2000-2004/ 2008-2012; Nova Glória 2008-2012; Rialma 2004-2008; Rianópolis 2000-2004 e Santa Isabel 2012-2016.

Em determinados municípios, alguns partidos tiveram vitórias recorrentes: Guarinos 2012-2016; Hidrolina 2012-2016; Itapaci 2008-2012; Pilar de Goiás 2012-2016; São Luiz do Norte 2000-2004/2012-2016; Rialma 2004-2008-2012; Rianópolis 2004-2008-2012-2016 e Santa Isabel 2000-2004/2008-2012.

Foi observado também que na zona 039 o partido que obteve maior quantidade de vitórias foi o PMDB e na zona 072 foi o partido PSDB. Será analisado através do

perfil do eleitorado se a missão de cada partido tem influência no resultado das eleições.

3.2 FERRAMENTA UTILIZADA

O trabalho foi desenvolvido utilizando a ferramenta WEKA (do inglês *Waikato Environment for Knowledge Analysis*); uma suíte totalmente gratuita, desenvolvida pela Universidade de Waikato (Nova Zelândia). O software disponibiliza diferentes algoritmos de classificação de aprendizado de máquina necessários para fazer pré-processamento dos dados, classificação, regressão, regras de associação e seleção de atributos, além de ter um rápido desempenho (MATSUMOTA, 2018). Devido a essas funcionalidades, WEKA foi a ferramenta escolhida para executar os algoritmos de mineração de dados.

3.3 PRÉ-PROCESSAMENTO E LIMPEZA

A partir da base de dados disponibilizada pelo TSE, são aplicadas as técnicas de pré-processamento, que são de suma importância para o resultado final deste trabalho. Para facilitar a implementação, optou-se por fazer a limpeza dos dados, como primeira parte do pré-processamento. Para isso, foram feitas, nesta ordem, as seguintes operações: exclusão das variáveis data de geração do arquivo, hora de geração do arquivo, UF, código TSE do município, quantidade de eleitores no perfil, estado civil do candidato e remoção de acentuação das palavras.

Entre os atributos disponíveis, esses itens foram excluídos para que se possa ter uma quantidade de porcentagem maior de instâncias classificadas corretamente e foram selecionados aqueles considerados mais relevantes para obtenção do conhecimento pretendido, ou seja, aqueles que retornaram um resultado eficaz, que foram os seguintes: ano, município, zona, sexo, faixa etária e grau de escolaridade. Utilizando essas variáveis espera-se que seja extraída uma satisfatória correlação dos dados explícitos na aplicação do algoritmo J48.

De início, inseriu-se a base de dados na planilha eletrônica do Excel 2010, software da Microsoft, para verificar cada um dos atributos, por meio da opção de importação de dados. Durante este procedimento foram encontrados alguns problemas que deveriam ser solucionados na fase da limpeza dos dados.

Verificou-se que antes das eleições municipais do ano de 2016, os municípios de Rialma e Santa Isabel compunham a zona eleitoral de número 112 do TRE- GO e que somente a partir do ano de 2018 passaram a fazer parte da zona 72. Deste modo, optou-se por alterar as células para a zona que se encontram atualmente, a fim de não interferir no conhecimento gerado.

Após, constatou-se que as nomenclaturas da descrição do grau de escolaridade dos anos de 2000 e 2004 eram diferentes dos anos de 2008, 2012 e 2016. Naqueles constam as seguintes frases: “PRIMEIRO GRAU INCOMPLETO”, “PRIMEIRO GRAU COMPLETO”, “SEGUNDO GRAU INCOMPLETO” e “SEGUNDO GRAU COMPLETO”. Nestes constam as seguintes frases: “ENSINO FUNDAMENTAL INCOMPLETO”, “ENSINO FUNDAMENTAL COMPLETO”, “ENSINO MÉDIO INCOMPLETO” e “ENSINO MÉDIO COMPLETO”. Esse problema foi resolvido através da atualização da nomenclatura dos anos de 2000 e 2004 à equivalente aos anos consecutivos trabalhados.

Por fim, verificou-se que existiam inúmeros perfis do eleitorado municipal que eram idênticos, os quais estavam indicados em apenas uma linha. Foi preciso fazer um trabalho de distribuição de todos esses eleitores em linhas unitárias, mesmo que sejam dados iguais, para obtenção de um resultado eficaz ao final da pesquisa. Ademais, o arquivo em formato .CSV foi gerado para ser utilizado na mineração.

3.4 TRANSFORMAÇÃO DOS DADOS

Para maior aproveitamento da ferramenta WEKA, alguns atributos foram alterados, com o objetivo de reduzir os valores, por exemplo, o atributo “MASCULINO”, deixando apenas como atributo “M”.

Realizadas estas etapas, foi criado um arquivo XLSX (arquivo no formato da planilha Excel 2010) apenas com os dados que serão utilizados para realizar o processo de mineração de dados (ano, município, zona, sexo, faixa etária e grau de escolaridade). Após a criação deste arquivo foi realizada a conversão do mesmo para o formato CSV, utilizado pela ferramenta WEKA. Realizada a conversão, foi preciso processar o arquivo criado na ferramenta WEKA, a fim de obter os resultados pretendidos.

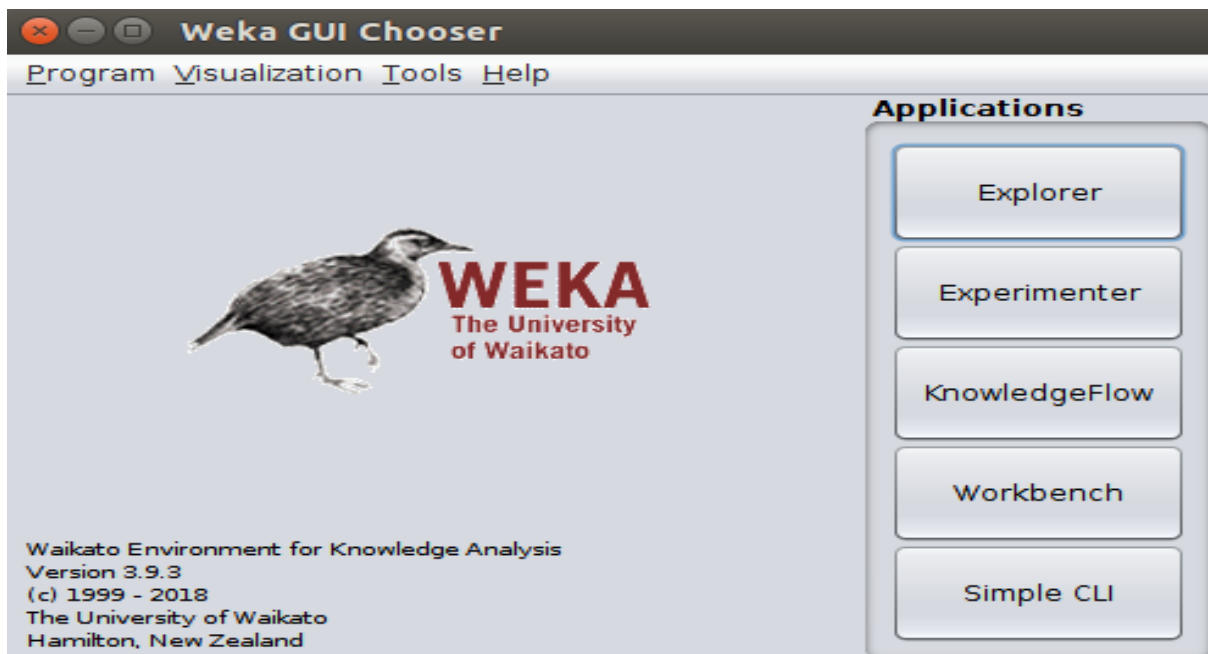


Figura 3 - Tela principal do software WEKA.

Fonte: WEKA. The University of Waikato (1999-2018).

A Figura 3 refere-se a tela principal do software WEKA, onde foi carregado o arquivo .CSV e exibido as informações preliminares.

3.5 MINERAÇÃO DOS DADOS

Após filtragem e seleção, os dados foram carregados para a ferramenta WEKA, a fim de serem submetidos ao algoritmo de mineração de dados. Entre os diversos tipos de algoritmos disponíveis na ferramenta, foi escolhido o J48, do tipo classificação, pois as árvores de decisão, geradas a partir desse algoritmo, são de fácil entendimento.

Os resultados da mineração dos dados referentes às eleições municipais da zona 039 serão comparados aos resultados obtidos a partir dos dados das eleições municipais da zona 072, identificando possíveis tendências ou diferenças.

A árvore de decisão foi o tipo de modelo definido pois pode ser visualizado com facilidade, o que ajudaria na análise das regras geradas. Após carregar os dados e selecionar o algoritmo J48, foi realizada sua execução, utilizando o seguinte esquema de execução (scheme): `-weka.classifiers.trees.J48 -S -C 0.25 -M 2`

A Figura 4 a seguir, refere-se a tela onde é definido o algoritmo e suas devidas configurações, bem como, onde a mineração de dados é executada e o resultado é obtido.

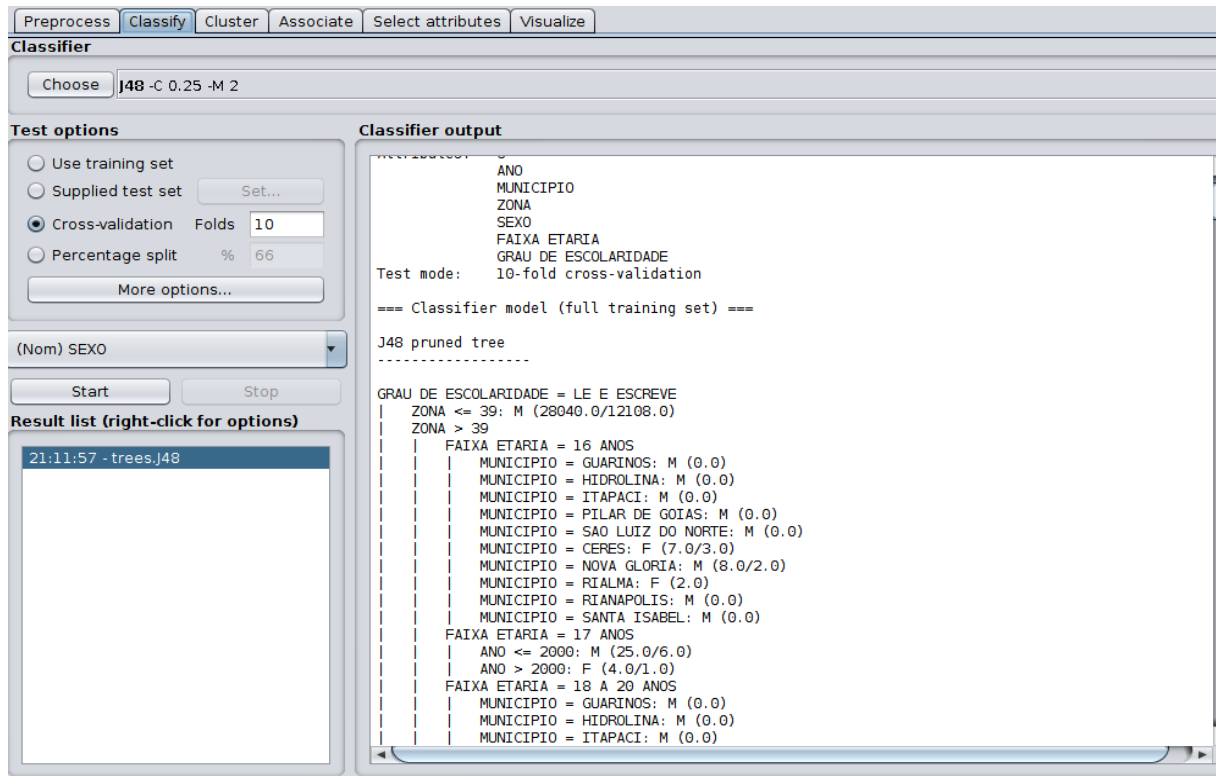


Figura 4 - Árvore de decisão utilizando o algoritmo J48.

Fonte: Arquivo pessoal (2020).

4. RESULTADOS E DISCUSSÃO

A fim de cumprir o objetivo central do trabalho, a mineração de dados proposta necessita ter uma grande abrangência, podendo assim coletar possíveis tendências de diferentes municípios do Estado de Goiás, com suas características culturais, sociais e políticas. Para isso, foi realizado processo de mineração de dados selecionando apenas as zonas 039 e 072, tornando o processo mais reduzido e facilitado. O critério selecionado para seleção destas zonas foi a quantidade de habitantes e por serem municípios mais próximos da autora do trabalho.

4.1 EXPERIMENTO 1

A configuração do experimento 1 foi realizada com a utilização do conjunto dos atributos: ano, município, zona, sexo, faixa etária e grau de escolaridade.

4.1.1 Resultados Obtidos

Com os referidos atributos foram feitos testes prévios e para efetiva seleção, o atributo escolhido foi o valor “SEXO”, considerado o método com melhor estimativa de acurácia de 54.642% e conseqüentemente, uma incorrelação de 45.358%. Nesse sentido, esperava-se uma correspondência de maior porcentagem para uma melhor determinância na análise do resultado. Nota-se, que na Figura 5, foi gerada pelo algoritmo J48, uma grande quantidade de dados na árvore.

Níveis da árvore gerada:

- Quantidade de folhas: 352

- Tamanho da árvore: 440

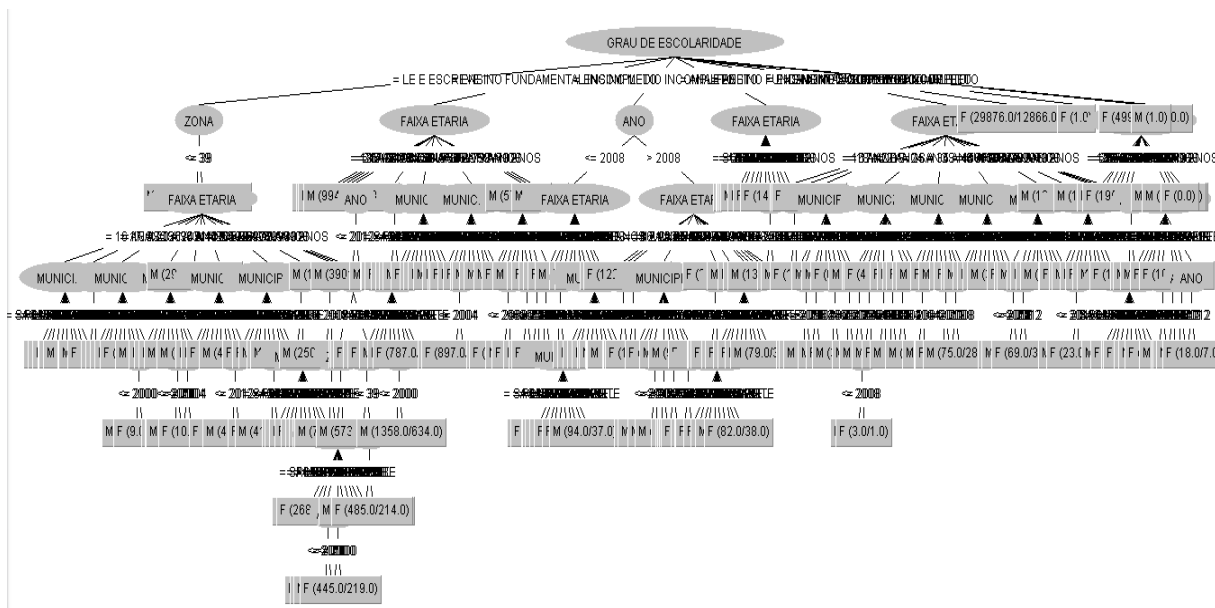


Figura 5- Árvore de agrupamento do conjunto (atributo “sexo”).

Fonte: Arquivo pessoal (2020).

4.1.2 Análise do Resultado

A árvore gerada apresentou-se robusta devido ao elevado número de folhas, mesmo expandindo a ilustração, a ferramenta não traz uma forma que permita visualizar todos os itens na tela ao mesmo tempo, por isso, serão analisados alguns dos resultados por partes.

Na Figura 6 a seguir observa-se os eleitores que possuíam ensino médio incompleto nos anos de 2012 e 2016:

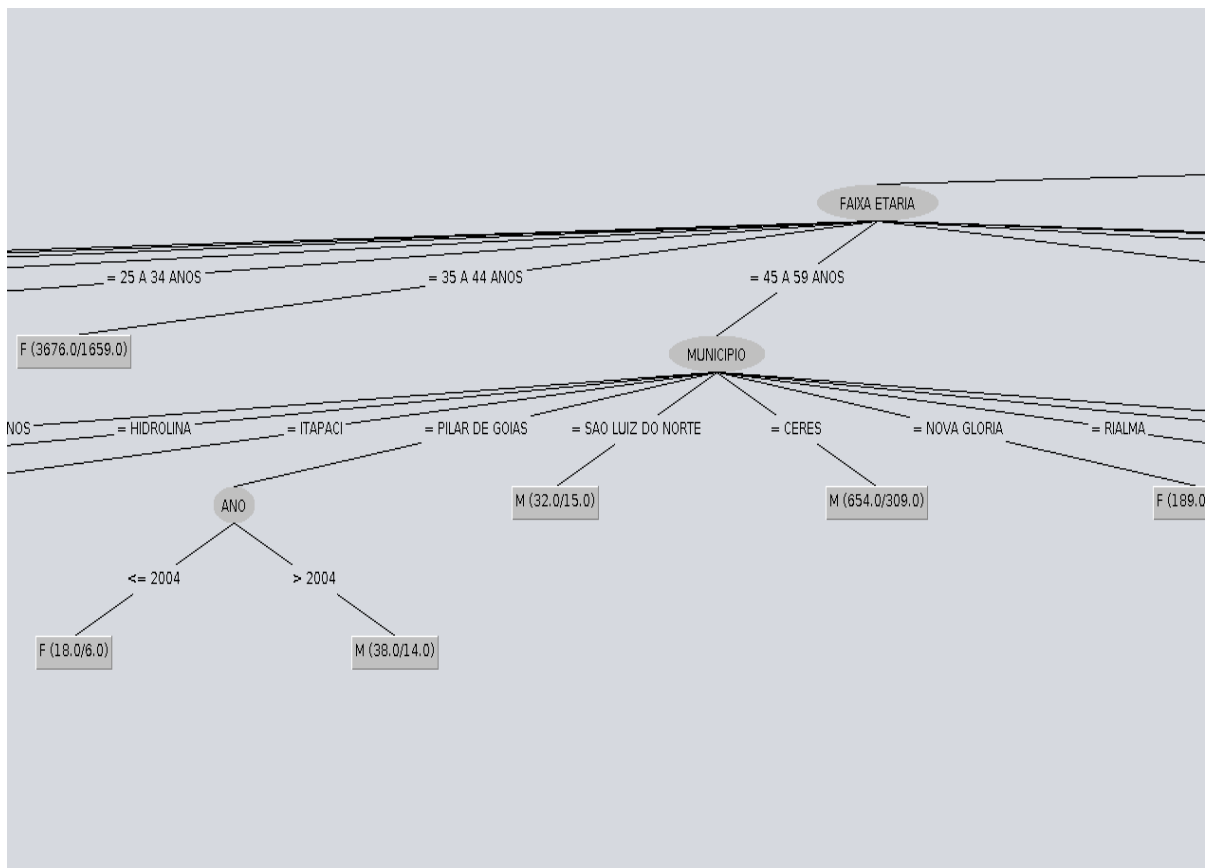


Figura 6 - Árvore de decisão (Parte 1).

Fonte: Arquivo pessoal (2020).

A partir da imagem, no município de Pilar de Goiás, os votantes do sexo masculino com idade entre 45 a 59 anos, após o ano de 2008, possuíam ensino médio incompleto. Doutro lado, o sexo feminino prevaleceu nos anos de 2000 e 2004.

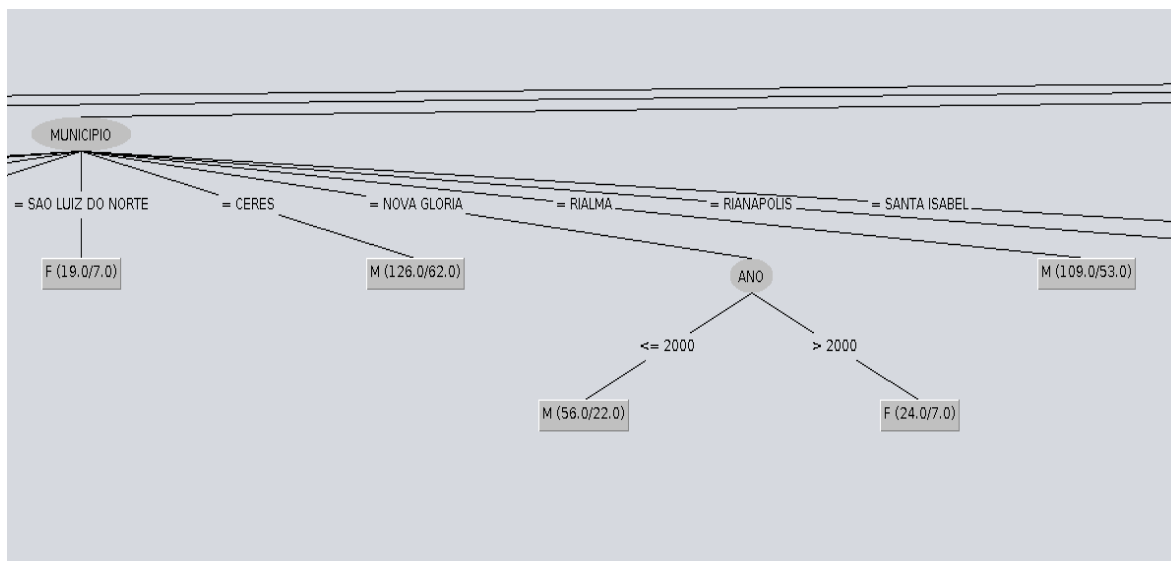


Figura 7 - Árvore de decisão (Parte 2)

Fonte: Arquivo pessoal (2020).

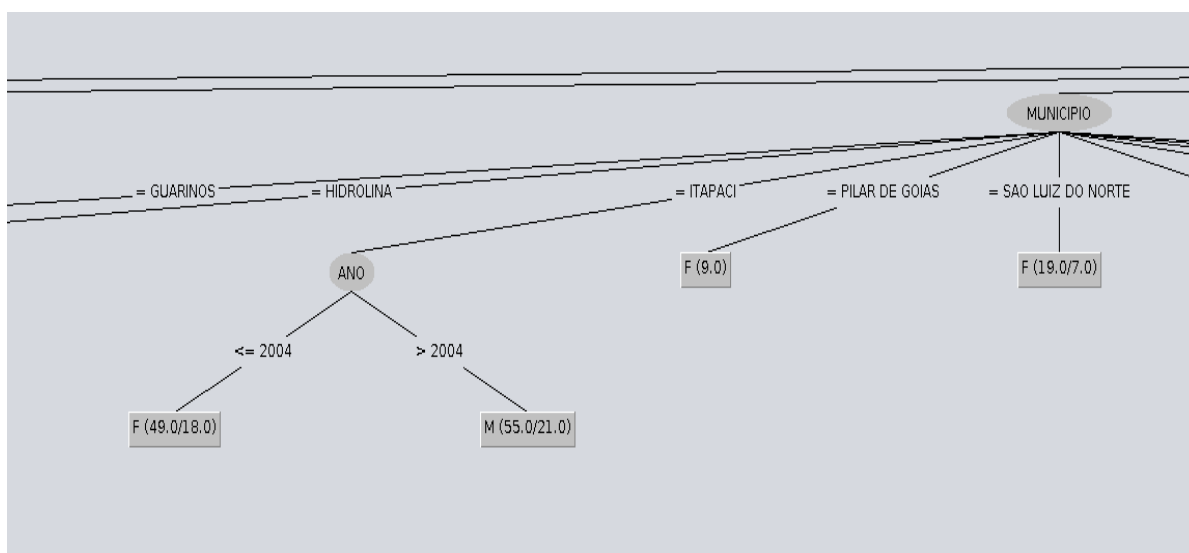


Figura 8 - Árvore de decisão (Parte 3)

Fonte: Arquivo pessoal (2020).

As Figuras 7 e 8 expressam os municípios com eleitores de faixa etária entre 18 a 20 anos. Conforme resultado emitido pelo algoritmo, o grau de escolaridade da maior parte dos votantes nessas municipalidades é de ensino médio incompleto. Como exemplo, no município de Nova Glória, a maioria dos votantes após o ano de 2004 são do sexo feminino, por sua vez, em Itapaci, a partir de 2008, sobrepôs-se o sexo masculino.

4.2 EXPERIMENTO 2

Configuração com utilização dos atributos: ano, município, zona, sexo, faixa etária e grau de escolaridade.

4.2.1 Resultados Obtidos

A correlação das instâncias se deu em 46.3288% e automaticamente, uma incorrelação de 53.6712% com o atributo de valor “GRAU DE ESCOLARIDADE”. Dessa forma, esperava-se uma similitude ainda maior que o experimento anterior para uma melhor determinância na análise do resultado. Infere-se da imagem abaixo que a árvore gerada pelo algoritmo J48 possui uma alta quantidade de informações.

Níveis da árvore gerada:

- Quantidade de folhas: 148
- Tamanho da árvore: 214

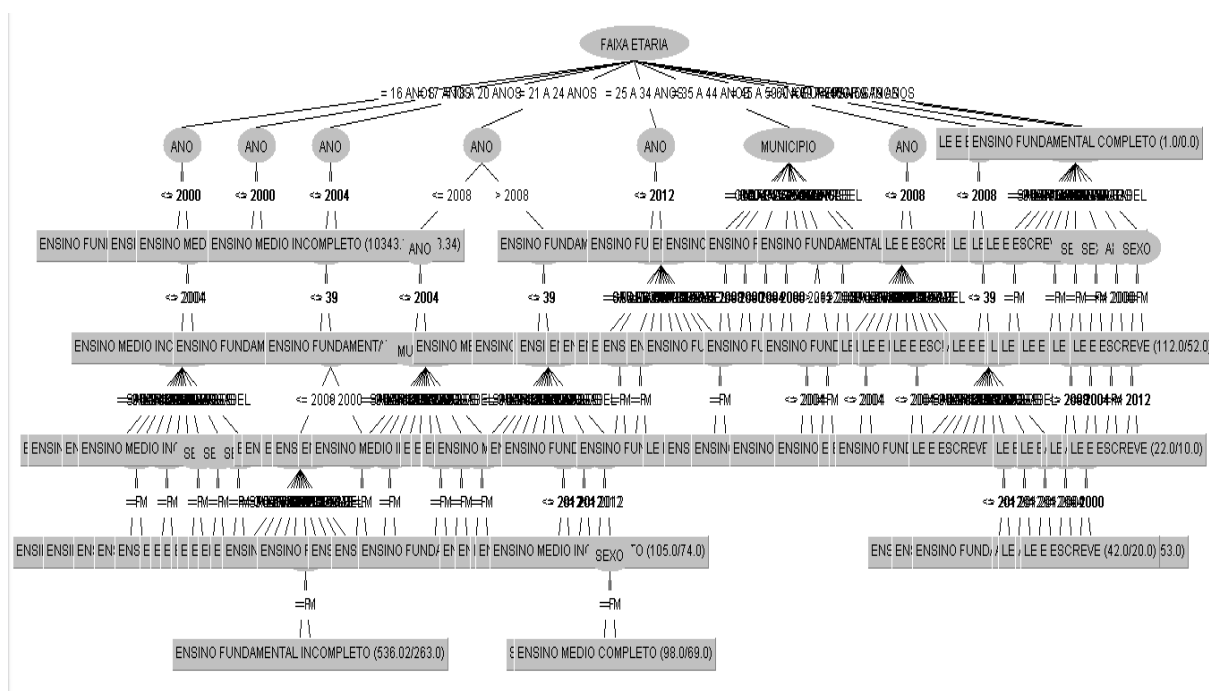


Figura 9 - Árvore de agrupamento do conjunto (atributo “grau de escolaridade”).

Fonte: Arquivo pessoal (2020).

4.2.2 Análise do Resultado

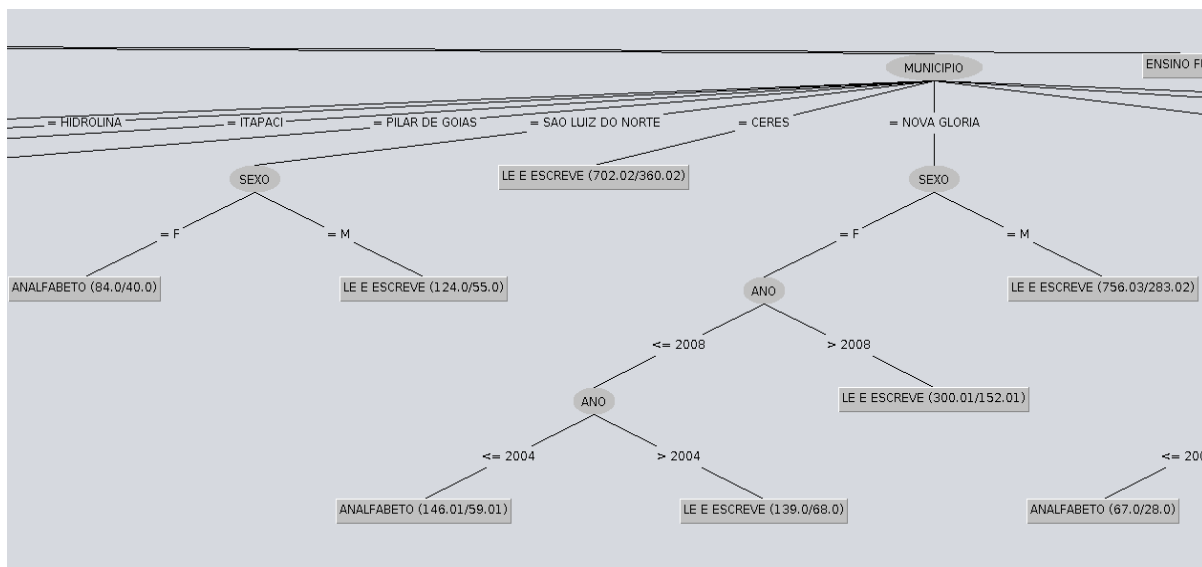


Figura 10 - Árvore de decisão (Parte 4).

Fonte: Arquivo pessoal (2020).

Na Figura 10 podemos observar um comportamento relevante nos anos de 2000 e 2004 com eleitores votantes de idade superior a 79 anos. Em Nova Glória, a maioria dos eleitores do sexo feminino são analfabetos. O interessante é que nos anos seguintes houve uma evolução para esse gênero, sendo esta enquadrada com grau de escolaridade de quem lê e escreve. Já o sexo masculino, não obteve progresso em seu grau de escolaridade no decorrer dos derradeiros anos.

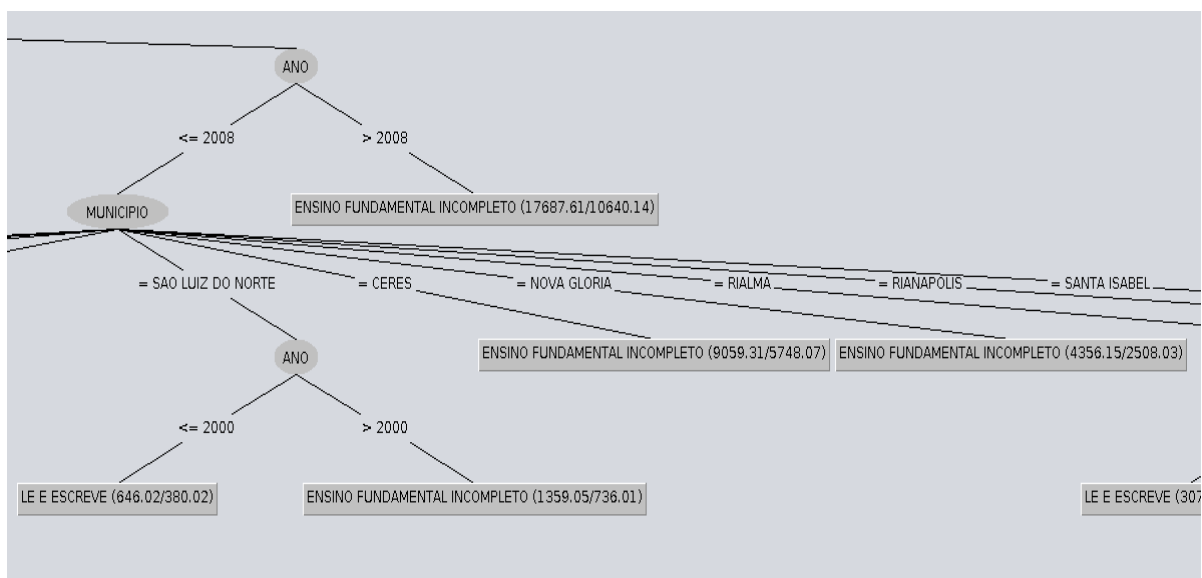


Figura 11 - Árvore de decisão (Parte 5).

Fonte: Arquivo pessoal (2020).

Exsurge da Figura 11 que a faixa etária dos votantes entre 35 a 44 anos dos municípios, possuem, majoritariamente, ensino fundamental incompleto. Nos municípios de São Luiz do Norte (PMDB: 2000-2004), Rialma (PSDB: 2004-2008) e Rianópolis (PSDB: 2004-2008) antes das eleições 2008, os representantes dos partidos reeleitos considerando a faixa etária acima, tiveram eleitores votantes com grau de escolaridade de quem lê e escreve.

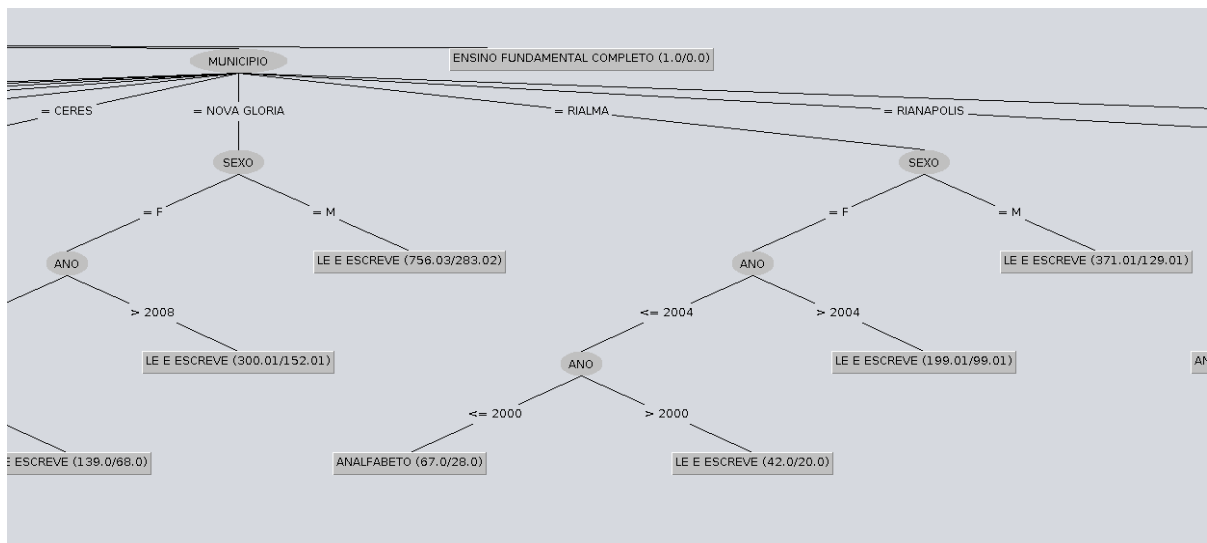


Figura 12 - Árvore de decisão (Parte 6).

Fonte: Arquivo pessoal (2020).

Na Figura 12, verifica-se que o grau de escolaridade “lê e escreve” entre os eleitores votantes do sexo masculino, com idade superior a 79 anos, é predominante nos municípios. Já para o gênero feminino, o analfabetismo é dominante. Com isso, vê-se que as oportunidades de aprendizagem para as mulheres em anos anteriores não surgiram como nos tempos atuais.

4.3 EXPERIMENTO 3

Configuração com utilização do conjunto dos atributos: ano, município, zona, sexo, faixa etária e grau de escolaridade.

4.3.1 Resultados Obtidos

Neste experimento o poder de acurácia foi 46.3567% e consequentemente, incorrelação de 53.6433% com o atributo de valor “MUNICÍPIO”. Verifica-se na imagem abaixo que a árvore gerada pelo algoritmo J48, mesmo com baixa correlação, possui uma grande quantidade de dados extraídos.

Níveis da árvore gerada:

- Quantidade de folhas: 91
- Tamanho da árvore: 128

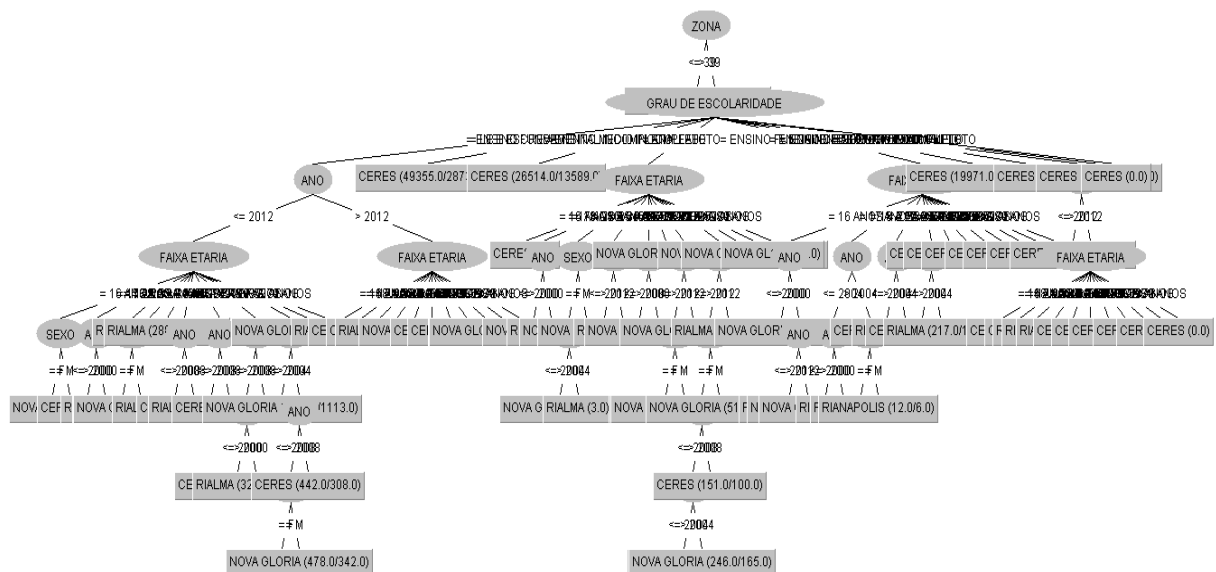


Figura 13 - Árvore de agrupamento do conjunto (atributo “município”).

Fonte: Arquivo pessoal (2021).

4.3.2 Análise do Resultado

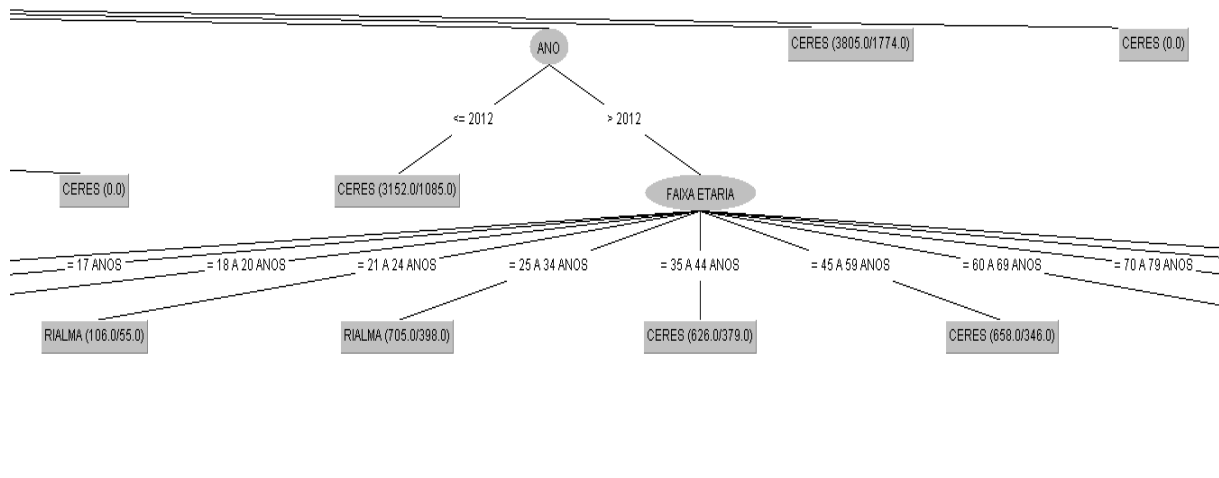


Figura 14 - Árvore de decisão (Parte 7).

Fonte: Arquivo pessoal (2021).

A Figura 14 traz a informação de dados dos eleitores votantes com grau de escolaridade de ensino superior incompleto da Zona Eleitoral 039. Porém, nota-se que, mesmo assim, a ferramenta trouxe também resposta dos municípios da Zona Eleitoral 072. Isso se dá devido à alta porcentagem no grau de incorrelação.

Nas eleições dos anos de 2000 até 2012, a maioria dos eleitores fazem parte do município de Ceres, isso se dá em razão da população desse município ser a mais populosa das cidades trabalhadas neste estudo. No ano de 2016, os eleitores com idade igual ou maior a 35 anos continuaram sendo da cidade de Ceres, contudo os que possuem idade abaixo de 34 anos, maioria de Rialma. Essa alteração pode ter sido ocasionada em razão do decurso de tempo, no qual o município vizinho vem tomando maior dimensão populacional e, ainda, provendo uma maior valorização municipal superior do que alguns dos outros estudados.

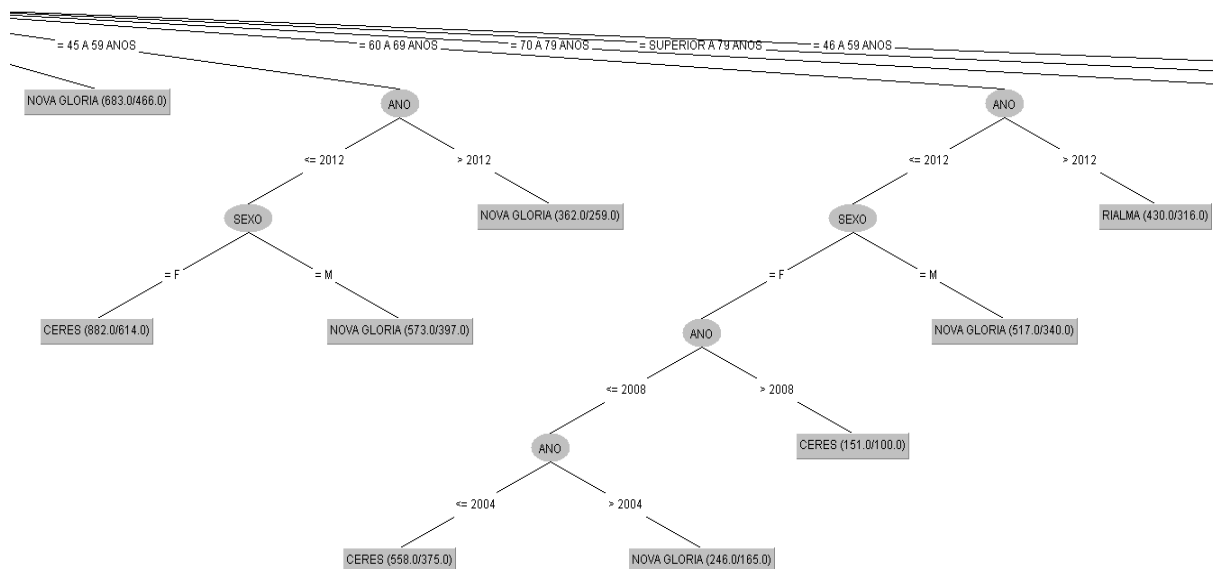


Figura 15- Árvore de decisão (Parte 8).

Fonte: Arquivo pessoal (2021).

A Figura 15 também traz a informação de dados dos eleitores da Zona Eleitoral 039, incluindo a Zona 072 erroneamente. Nos anos de 2004 a 2012, os eleitores votantes analfabetos com idade entre 21 até superior a 79 anos, obtiveram maior concentração no município de Nova Glória e foram do sexo masculino.

Doutra banda, houve diversificação do sexo feminino.

-De 18 a 20 anos: Ano de 2000, Nova Glória. A partir de 2004, Rialma.

-De 45 a 59 anos: Até 2012, Ceres. Ano de 2016, Nova Glória.

-De 60 a 69 anos: Ano de 2000, Ceres. Ano de 2004, Nova Glória. Ano de 2008, Ceres. A partir de 2012, Rialma.

Verifica-se dos dados que o gênero feminino com analfabetismo se destaca como maioria em diversos municípios, anos e faixas etárias. Mesmo com a transformação social e educacional das mulheres no decorrer dos anos, é notável a desigualdade no que se refere ao grau de instrução entre os gêneros no contexto histórico no comparativo da quantidade populacional do município de Nova Glória.

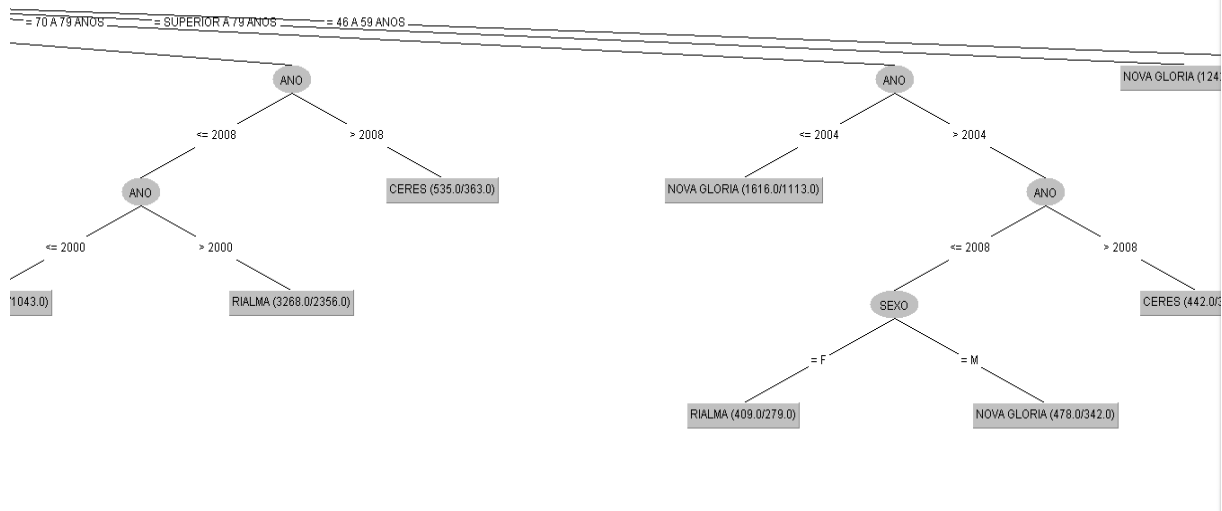


Figura 16 - Árvore de decisão (Parte 9).

Fonte: Arquivo pessoal (2021).

Na apresentação, os eleitores que possuem faixa etária entre 60 a 69 anos no ano de 2000 residentes na cidade de Ceres, possuem grau de escolaridade de quem lê e escreve. Nas eleições dos anos 2004 a 2012, os eleitores de Rialma, do sexo feminino, possuíam instrução idêntica a supracitada. Do mesmo modo, eleitores com idade entre 70 a 79 anos, nas eleições de 2000 e 2004 pertencem ao município de Nova Glória e do sexo masculino. Por fim, no ano de 2016, a maioria do município de Ceres. Vê-se que há informações mais detalhadas para as Eleições 2000 e 2004.

4.4 EXPERIMENTO 4

Configuração com utilização do conjunto dos atributos: ano, município, zona, sexo, faixa etária e grau de escolaridade.

4.4.1 Resultados Obtidos

Por último, foi efetuado teste com o atributo de valor “FAIXA ETÁRIA”. A correlação das instâncias se deu em 30.6951%, sendo a mais baixa porcentagem dos experimentos, e uma incorrelação de 69.3049%.

Níveis da árvore gerada:

- Quantidade de folhas: 229

- Tamanho da árvore: 321

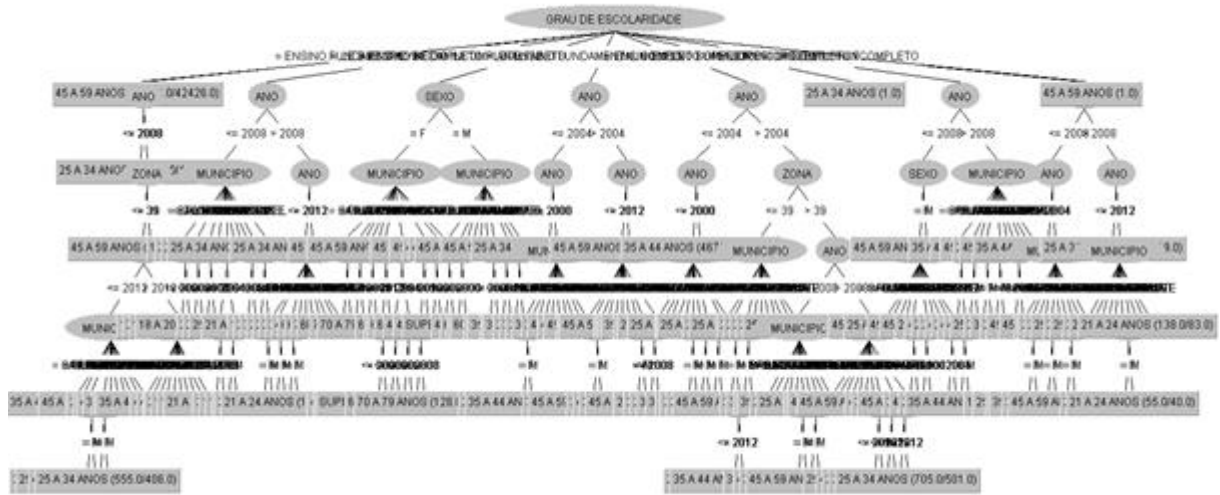


Figura 17- Árvore de agrupamento do conjunto (atributo “faixa etária”).

Fonte: Arquivo pessoal (2021).

4.4.2 Análise do Resultado

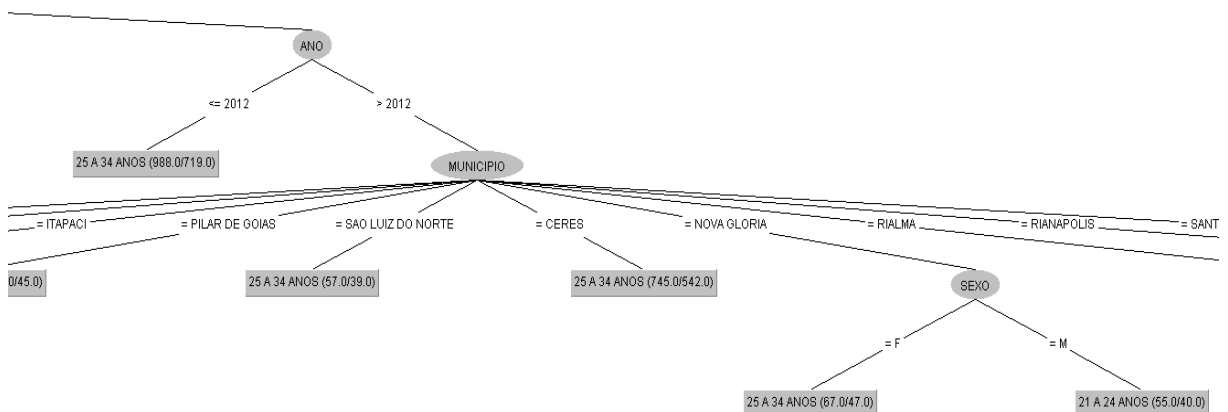


Figura 18- Árvore de decisão (Parte 10).

Fonte: Arquivo pessoal (2021).

Nas eleições dos anos de 2012 e 2016, a maior parte dos eleitores que possuíam grau de escolaridade superior incompleto estavam com idade entre 25 a 34

anos, exceto no município de Nova Glória, onde o sexo masculino se destaca majoritariamente com idade entre 21 a 24 anos. Vê-se que esse gênero nesta municipalidade alcançou uma escolaridade mais jovem que o sexo feminino nesse tipo de ensino.

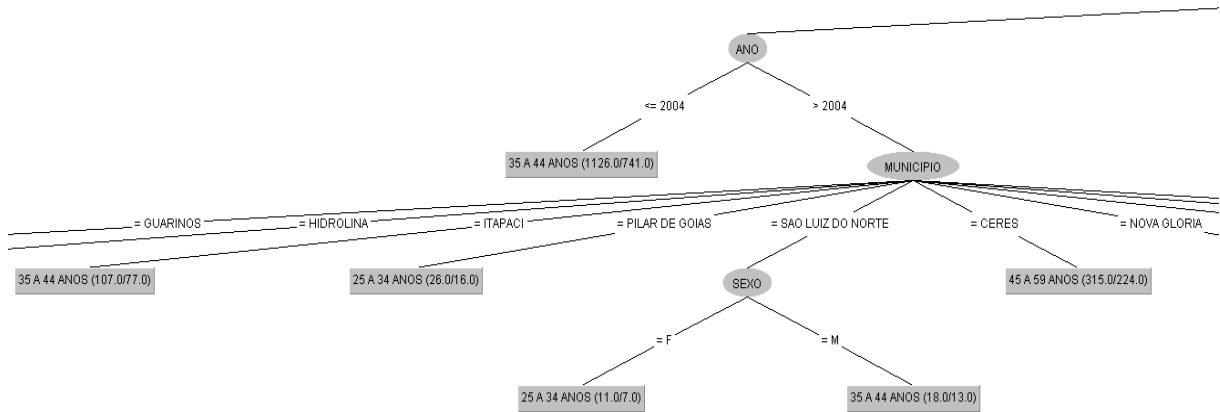


Figura 19- Árvore de decisão (Parte 11).

Fonte: Arquivo pessoal (2021).

Já na figura acima, são trazidas também informações dos votantes com ensino superior incompleto dos anos anteriores a 2008, e sua maior concentração oscila entre as idades de 25 a 49 anos. Os municípios de São Luiz do Norte (F= 25 a 34 anos/ M= 35 a 44 anos), Rialma (F= 35 a 44 anos/ M= 45 a 59 anos) e Hidrolina (F= 18 a 20 anos/ M= 25 a 34 anos) trazem dados significativos: mulheres com idade mais jovial já alcançaram esse nível de instrução, enquanto homens obtiveram a mesma formação em idade relativamente superior.

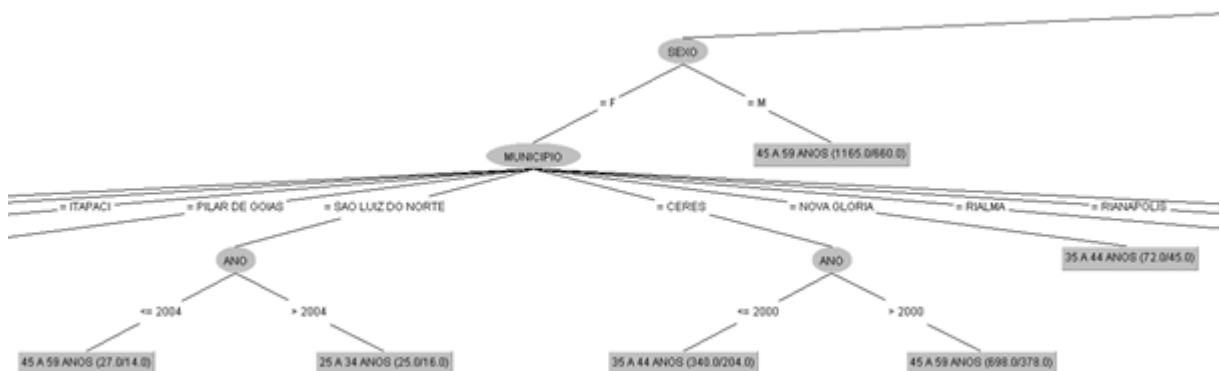


Figura 20- Árvore de decisão (Parte 12).

Fonte: Arquivo pessoal (2021).

Por fim, através do algoritmo, os dados mostram eleitores votantes nas eleições dos anos de 2000 a 2008 com educação superior completa, que pertencentes ao sexo masculino, sendo a maioria entre 45 a 59 anos, enquanto o sexo feminino figura entre 25 a 59 anos, gerando tendência de desproporcionalidade entre os pares. Nota-se que, com o passar dos anos, as mulheres vêm empreendendo esforços a fim de obter formação em nível superior, o que tem gerado reflexos positivos de disparidades para o gênero.

4.5 INTERPRETAÇÃO QUALITATIVA DOS RESULTADOS

Com base nas análises individuais de cada experimento, pode ser feita uma análise geral entre os diferentes perfis dos eleitores votantes nessas duas zonas trabalhadas.

É importante salientar que, conforme resultado gerado pelo algoritmo J48, o partido que obteve mais reeleições entre os municípios (PSDB- Rialma (2004-2008-2012) e Rianópolis (2004-2008-2012-2016)) possui em sua grande maioria eleitores com grau de escolaridade “lê e escreve” e faixa etária entre 21 a 24 anos de idade. Através dessa constatação, percebe-se que as informações retornadas podem não estar totalmente condizentes com a realidade devido aos altos valores de incorrelação.

Doutra banda, a maioria dos eleitores votantes com idade superior a 79 anos do sexo feminino possuem analfabetismo, enquanto os do sexo masculino, possuem leitura e escrita. No entanto, os eleitores do sexo feminino na faixa etária de 16 a 79 anos possuem ensino médio incompleto, contudo, os do sexo masculino possuem ensino fundamental incompleto. Sendo assim, constata-se a partir dos resultados, que com o passar dos últimos anos, as mulheres tiveram maiores oportunidades na obtenção de conhecimento e aprendizagem gerando, conseqüentemente, um maior grau de instrução do que os homens.

O algoritmo J48 não pôde categorizar, através da aplicação de árvore de decisão, com boa acurácia os dados pré-processados e não permitiu que a autora desse trabalho fizesse as devidas análises com total firmeza no reconhecimento de padrões devido à alta porcentagem de incorrelação. Tais medidas requerem uma análise detalhada no intuito de que atenda às expectativas e se adeque à realidade e complexidade do problema.

Ao avaliar o trabalho como uma proposta de previsibilidade partidária eleitoral, os resultados obtidos mostram-se não conclusivos devido o algoritmo ter tido mais treinamento, porém não conseguiu ter maior poder de acurácia. Com os resultados deste trabalho, não foi possível afirmar, por exemplo, que o partido do PSDB venceria novamente pela 5ª vez consecutiva as eleições 2020 no município de Rianópolis, bem

como, não é possível concluir, a partir do estudado até o momento, se a missão de cada partido tem influência no resultado das eleições.

Contudo, há indícios de que, em trabalhos futuros, este projeto possa ser referência para pesquisas nessa temática.

5. CONSIDERAÇÕES FINAIS

Durante o período eleitoral, as informações disponibilizadas ao eleitorado podem ser transmitidas através de diferentes maneiras, seja na comunidade em que o cidadão vive (propaganda eleitoral), por meio de diálogos com as pessoas que fazem parte do seu cotidiano ou de forma direta, que são plataformas de comunicação social (redes sociais, televisão, rádio, internet).

O objetivo deste trabalho foi usar a técnica de mineração de dados do tipo árvore de decisão para identificar padrões no repositório aberto de dados eleitorais do TSE, nas eleições municipais dos anos de 2000 a 2016. No entanto, essa pesquisa não possibilitou mensurar o quanto as características do eleitorado nas Zonas 039 e 072 tiveram impacto determinante no resultado das eleições municipais relatadas.

Foram empreendidos diversos esforços para desenvolver a presente versão. Objetivando obter um melhor aproveitamento da ferramenta WEKA, foram realizadas diligências a fim de encontrar a adequada preparação dos dados, efetuando a tarefa de seleção e filtragem desses dados, principalmente pela grande quantidade de informações e atributos disponíveis e, com isso, foi uma das etapas de organização dos dados que mais consumiu tempo e trabalho.

Através da experiência obtida até o momento, como prováveis desdobramentos e expansão de possível versão futura, torna-se imprescindível avaliar os padrões encontrados nos experimentos usando o algoritmo J48, podendo comparar resultados e informações das duas zonas eleitorais trabalhadas, extraindo um conhecimento detalhado em determinados aspectos e buscando encontrar novos meta-atributos que possam ser significativos para a classificação.

REFERÊNCIAS

ABERNETHY, Michael. **Mineração de dados com o Weka**. IBM, 2010. Disponível em: <https://www.ibm.com/developerworks/br/opensource/library/os-weka1/index.html>. Acesso em: 24 de mar. de 2020.

APURAÇÃO. G1, 2012. Disponível em: <http://g1.globo.com/go/goias/apuracao/itapaci.html>. Acesso em: 28 de abr. de 2020.

BRASIL tem 147,3 milhões de eleitores aptos a votar nas eleições 2018. TSE, 2018. Disponível em: <http://www.tse.jus.br/imprensa/noticias-tse/2018/Agosto/brasil-tem-147-3-milhoes-de-eleitores-aptos-a-votar-nas-eleicoes-2018/>. Acesso em: 26 de mar. de 2020.

CARVALHO, L. A. V. **Dataminig: a mineração de dados no marketing, medicina, economia, engenharia e administração**: 1. ed. São Paulo: Érica, 2001.

CIÊNCIA de dados Weka. Minerando Dados, 2019. Disponível em: <https://minerandodados.com.br/ciencia-dados-weka/weka-tela-parametros-algoritmo-j48-classificacao/>. Acesso em: 27 de mar. de 2020.

FÁVERO, Luiz Paulo. **KDD e Data Mining: mais do que apenas conceitos**. It Forum 365, 2019. Disponível em: <https://www.itforum365.com.br/colunas/kdd-e-data-mining-mais-do-que-apenas-conceitos/>. Acesso em: 27 de mar. de 2020.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory e SMYTH, Padhraic. **Knowledge Discovery and Data Mining: Towards a Unifying Framework**. AAAI, 1996. Disponível em: <https://www.aaai.org/Papers/KDD/1996/KDD96-014.pdf>. Acesso em: 27 de mar. de 2020.

GONÇALVES, Eduardo Corrêa. **Mineração de regras de associação com a ferramenta de Data Mining Weka**. DevMedia, 2011. Disponível em: <https://www.devmedia.com.br/mineracao-de-regras-de-associacao-com-a-ferramenta-de-data-mining-weka/20478>. Acesso em: 27 de mar. de 2020.

GRUS, Joel. **Data Science do Zero**. 1. ed. Rio de Janeiro: Alta Books, 2016.

GUEDES, Pedro Casali. **Aplicação de técnicas de data mining para previsibilidade eleitoral**. 120f. Trabalho de Conclusão de Curso (Bacharel) Universidade Federal de Santa Catarina, Florianópolis, 2018.

KIRA, Gustavo. **Aplicação de técnicas de mineração de dados à um recorte do repositório de dados históricos sobre eleições brasileiras do Tribunal Superior**

Eleitoral brasileiro. 116f. Trabalho de Conclusão de Curso (Bacharel) Universidade Tecnológica Federal do Paraná, Curitiba, 2014.

LAS CASAS, Pedro Henrique Bragioni. **Mineração de dados aplicada.** Homepages, 2014. Disponível em: https://homepages.dcc.ufmg.br/~pedro.lascasas/aula_2_minera%C3%A7%C3%A3o_de_dados_aplicada_weka.pdf. Acesso em: 27 de mar. de 2020.

LAVAREDA, Antônio. (2011). **Neuropolítica: o papel das emoções e do inconsciente.** Revista USP, (90) 120-147. Disponível em: <https://doi.org/10.11606/issn.2316-9036.v0i90p120-147/>. Acesso em: 27 de mar. de 2020.

MATSUMOTA, Leonardo. **Data Mining (Mineração de Dados) utilizando Weka.** Management and It Innovation, 2018. Disponível em: <https://leonardo-matsumota.com/2018/03/08/data-mining-mineracao-de-dados-utilizando-weka/>. Acesso em: 27 de mar. de 2020.

PEREIRA, Francisco Carvalho. **Descoberta de conhecimento a partir de dados das eleições municipais das regiões do Brasil.** 59f. Trabalho de Conclusão de Curso (Especialização) Universidade Federal do Pampa, Bagé, 2013.

PROCESSO eleitoral no Brasil. TSE, 2015. Disponível em: <http://www.tse.jus.br/eleicoes/processo-eleitoral-brasileiro/funcionamento-do-processo-eleitoral-no-brasil/>. Acesso em: 24 de mar. de 2020.

REPOSITÓRIO de dados eleitorais. TSE, 2016. Disponível em: <http://www.tse.jus.br/eleicoes/estatisticas/repositorio-de-dados-eleitorais-1/repositorio-de-dados-eleitorais/>. Acesso em: 24 de mar. de 2020.

RESULTADO da eleição. TSE, 2008. Disponível em: http://www.tse.jus.br/hotsites/estatistica2008/est_result/resultadoEleicao.htm/. Acesso em: 02 de mai. de 2020.

RESULTADO da eleição 2000. TSE, 2016. Disponível em: <http://www.tse.jus.br/eleicoes/eleicoes-antiores/eleicoes-2000/resultado-da-eleicao-2000/>. Acesso em: 02 de mai. de 2020.

RESULTADO da eleição 2004. TSE, 2016. Disponível em: <http://www.tse.jus.br/eleicoes/eleicoes-antiores/eleicoes-2004/candidaturas-votacao-e-resultados/resultado-da-eleicao-2004/>. Acesso em: 02 de mai. de 2020.

RESULTADO das eleições 2016 por região. Gazeta do povo, 2016. Disponível em: <https://www.gazetadopovo.com.br/apuracao/resultados-eleicoes-2016-primeiroturno/>. Acesso em: 28 de abr. de 2020.

SILVA, L. A; PERES, S. M; BOSCARIOLI, C. **Introdução à mineração de dados: com aplicações em R**: 1. ed. Rio de Janeiro: Elsevier, 2016.