

um produto desenvolvido  
a partir da pesquisa  
“Identificação de alunos com  
risco de evasão escolar  
no ensino médio integrado  
por meio da mineração de  
dados com aplicação das  
Regras de Associação”

**MINERAÇÃO DE  
DADOS**

**EDUCACIONAIS:  
UM GUIA METODOLÓGICO  
DE IDENTIFICAÇÃO  
DE ALUNOS COM  
RISCO PARA**

**EVASÃO  
ESCOLAR**



**PRODUTO  
EDUCACIONAL**

## Dados internacionais de catalogação na fonte

T231m Tavares, Nicolle Oliveira  
Mineração de dados educacionais : um guia metodológico de identificação de alunos com risco para evasão escolar / Nicolle Oliveira Tavares – 6 – MT, 2022.  
72 f. : il. color.

Orientador(a) Ed Wilson Tavares Ferreira  
Dissertação. (CBA 0 Mestrado Profissional em Educação Profissional e Tecnológica) – Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso, Campus Cuiabá, 2022.  
Bibliografia incluída

1. Predição. 2. Dados. 3. Padrões. 4. IFMT. 5. Weka. I. Título.

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).  
Bibliotecário(as): Jorge Nazareno Martins Costa (CRB1-3205)

**Autora**  
**Orientador**

**Nicolle Oliveira Tavares**  
**Dr. Ed Wilson Tavares Ferreira**

**A pesquisa**

**“Identificação de alunos com risco de evasão escolar no ensino médio integrado por meio da mineração de dados com aplicação das Regras de Associação”**

## CRÉDITOS DAS IMAGENS |

As imagens, elementos e ícones utilizados neste Produto Educacional foram extraídas dos sítios *on-line* FREEPIK e GRATISPNG e podem ser utilizadas e distribuídas desde que sem fins comerciais.

## **SOBRE A AUTORA**

É aluna do Programa de Mestrado Profissional em Educação Profissional e Tecnológica (ProfEPT), no Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso (IFMT). Especialista em Gestão Pública pela Faculdade Educacional da Lapa (2018) e graduada em Administração pela Universidade Federal de Mato Grosso (2015). Atualmente é Administradora da Universidade Federal de Mato Grosso (UFMT).

Link para o currículo Lattes:  
<http://lattes.cnpq.br/5071489886462431>



## PALAVRAS DA AUTORA

*Este guia é resultado de uma pesquisa de mestrado intitulada “Identificação de alunos com risco de evasão escolar no ensino médio integrado por meio da mineração de dados com aplicação das Regras de Associação”. Ele é destinado aos gestores, docentes, coordenadores de cursos e demais interessados no tema.*

*Este material apresenta como contexto a evasão escolar no Ensino Médio Integrado ao ensino técnico. Pesquisar sobre a evasão escolar é necessário, haja vista não ser todo brasileiro que possui acesso à Educação e, ainda existem fatores que implicam na dificuldade tanto ao acesso quanto à permanência no sistema educacional do país.*

*Trata-se de um evento que é associado a situações como a retenção e repetência de um aluno, saída do aluno de uma instituição ou do sistema educacional, a não conclusão de um nível de ensino ou ainda abandono da escola e seu retorno após determinado período (DORE e LUCHER, 2011).*

*O Ensino Médio Integrado ao ensino técnico enfrenta uma série de obstáculos que vão desde a tentativa de implantação de uma base unitária para*

*todos até a evasão escolar e embora presente as menores taxas de evasão escolar se comparada aos demais níveis e modalidades de ensino, algumas exceções ocorrem e devem ser investigadas. Segundo Ciavatta (2014), Ensino Médio Integrado ou formação integrada é um conceito que vai além da articulação entre o ensino médio e a educação profissional, pois engloba a concepção politécnica, omnilateral e a escola unitária no atual contexto histórico e sob uma específica correlação de forças entre as classes.*

*De acordo com os dados da Plataforma Nilo Peçanha de 2019, ano base 2018 (BRASIL, 2019), os cursos de Ensino Médio Integrado ao ensino técnico do Campus Cuiabá – Cel. Octayde Jorge da Silva (CBA-OJS), do IFMT, apresentaram uma taxa de evasão de 5,9%. O curso que apresenta a maior taxa de evasão é o curso EMI Técnico em Eletrônica com 26,8%, seguido pelo curso de EMI Técnico em Eletrotécnica, com 16,7% e em terceiro o curso de EMI Técnico em Telecomunicações com 10,9%. Os referidos cursos fazem parte do Departamento de área de Eletroeletrônica do CBA-OJS/IFMT, juntamente com o curso de Técnico em Eletroeletrônica, que apresentou taxa de evasão de 2%.*

*Posto isso, o critério utilizado para inclusão dos cursos a serem exemplificados neste guia foram os cursos de Ensino Médio Integrado ao ensino técnico, com maiores taxas de evasão escolar no ano de 2018 no CBA-OJS/IFMT.*

*Ressalta-se que o aluno evadido é aquele que apresenta a situação de matrícula como abandono, cancelada, desligada, transferência externa ou transferência interna. Sendo assim, este foi o conceito de aluno evadido adotado neste guia que engloba os cursos de Ensino Médio Integrado ao ensino técnico do DAEE, por apresentarem as maiores taxas de evasão no CBA-OJS/IFMT, no ano de 2018 (BRASIL, 2019).*

*Dada essa problemática, a grande quantidade de dados armazenados em instituições escolares carrega um conjunto de valiosas informações, que quando transformadas por meio da Mineração de Dados permitem a identificação de elementos determinantes para a predição da evasão escolar.*

*Por meio da utilização da técnica de Mineração de Dados é possível investigar o fenômeno da evasão escolar, visto que, a sua aplicação permite a criação de elementos para a identificação precoce de alunos com risco de evasão escolar.*

*Para tanto, o objetivo deste guia é servir como recurso de apoio no processo de mitigação da problemática da evasão escolar. Dessa forma, foi apresentado um passo a passo de como realizar a detecção e categorização dos elementos que contribuem para a identificação de padrões de alunos com risco de evasão escolar, mediante a aplicação*

*da técnica de Mineração de Dados Educacionais.*

*Portanto, este guia está estruturado em 3 seções, que apresentam informações sobre os conceitos relacionados à Mineração de Dados Educacionais, um diagnóstico da situação do conjunto de dados dos alunos do DAEE/CBA-OJS/IFMT, durante o período de 2009 a 2019 e o passo a passo para a aplicação da Mineração de Regras de Associação.*

*Desejo que este material seja útil e aplicado no cotidiano das atividades do lócus desta pesquisa, bem como possa ser aplicado em outras instituições de ensino.*

*Boa leitura!*

*Nicolle*

# SUMÁRIO

1

A Mineração de Dados Educacionais:  
uma aliada aos trabalhos de mitigação  
da evasão escolar

1.1 Como aplicar a Mineração de Dados no  
âmbito escolar?

07

05

SAIBA MAIS

09

2

Diagnóstico da situação do conjunto de  
dados dos alunos do DAEE/CBA-OJS/IFMT,  
durante o período de 2009 a 2018

2.1 Conhecendo o perfil dos evadidos

2.1.1 Perfil pessoal

2.1.2 Perfil socioeconômico

2.1.3 Perfil acadêmico

2.1.3.1 Análise das disciplinas por curso

12

12

13

17

19

10

SAIBA MAIS

22

3

Passo a Passo da aplicação da Mineração  
de Regras de Associação

23

3.1 Passo a passo

24

3.1.1 Obtenção e seleção dos dados

24

3.1.2 Pré-processamento e tratamento  
dos dados

27

3.1.3 Transformação dos dados

29

3.1.4 Mineração dos dados

31

3.1.5 Interpretação e avaliação dos  
resultados

54

PALAVRAS  
FINAIS

66

REFERÊNCIAS

67

SAIBA MAIS

65

1

# A MINERAÇÃO DE DADOS EDUCACIONAIS: UMA ALIADA AOS TRABALHOS DE MITIGAÇÃO DA EVASÃO ESCOLAR

Padrões

KDD

Etapas

DATA MINING

Em um ambiente de negócios cada vez mais inovador e inconstante, a tecnologia se consolida como um elemento indispensável ao proporcionar qualidade, celeridade, maior produtividade, eficiência, entre outros benefícios. Os sistemas de informação constituem-se como importantes aliados nos processos de tomada de decisão ao permitir que um grande volume de dados gerados diariamente seja processado e armazenado.

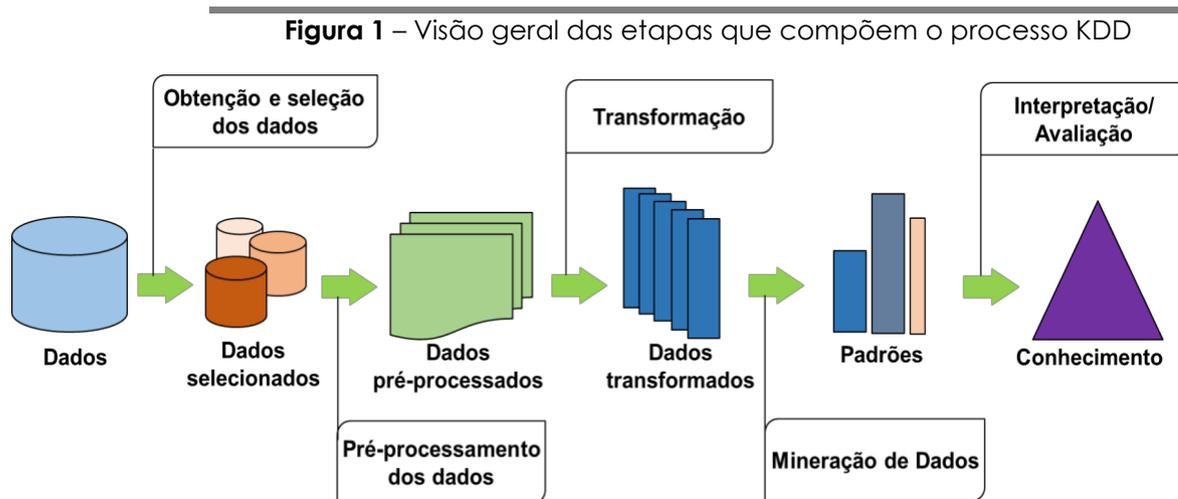
Neste contexto, uma relevante área disponível para a geração de conhecimento é a Mineração de Dados, do Inglês Data Mining. Esta área proporciona várias técnicas para a produção de um novo conhecimento ao gestor, por meio da transformação de uma ampla quantidade de dados armazenados em padrões válidos. Além disso, é possível promover uma externalização do conhecimento, do implícito para o explícito.

De acordo com Cardoso e Machado (2008, p. 497),

Data mining, ou mineração de dados, é uma técnica que faz parte de uma das etapas da descoberta de conhecimento em banco de dados. Ela é capaz de revelar, automaticamente, o conhecimento que está implícito em grandes quantidades de informações armazenadas nos bancos de dados de uma organização.

Um termo muito utilizado no contexto da Mineração de Dados é o KDD (*Knowledge Discovery in Databases* ou Descoberta de Conhecimento nas Bases de Dados, cujo conceito mais difundido e relevante de KDD foi desenvolvido por Fayyad, Piatetsky-Shapiro e Smyth em 1996, que o definiram como um processo de identificação de padrões válidos, novos úteis e compreensíveis a partir de dados.

Fayyad, Piatetsky-Shapiro e Smyth (1996) desenvolveram os passos do processo de KDD que são amplamente utilizados até os dias atuais. Os passos foram agrupados em cinco fases, conforme a Figura 1:



Fonte: Adaptado pela pesquisadora de Fayyad, Piatetsky-Shapiro e Smyth (1996)

## 1.1 COMO APLICAR A MINERAÇÃO DE DADOS NO ÂMBITO ESCOLAR?

No âmbito escolar, a quantidade de dados armazenados vem aumentando cada vez mais. São dados de registro escolar, dados socioeconômicos e acadêmicos de alunos, dados de Ambientes Virtuais de Aprendizagem (AVA), dados do corpo docente, da área administrativa, dados de produção acadêmica, entre outros. Muitos destes ficam armazenados durante anos na instituição apenas para registro e controle, mas quando são transformados podem propiciar aos gestores um conjunto de informações extremamente valiosas.

Neste contexto, embora seja uma área recente, a mineração de dados vem sendo utilizada com frequência no

contexto educacional, em especial para detectar elementos/fatores podem contribuir para a identificação de padrões de alunos com risco potencial de evasão escolar. O termo destinado para essa aplicação é Mineração de Dados Educacionais (MDE). Ao realizar pesquisas em bases científicas é possível encontrar diversos trabalhos com utilização da mineração de dados voltadas para a evasão escolar.

Neste guia será abordado o modelo de descrição. O modelo de descrição apresenta sua ênfase na busca de padrões de ocorrências frequentes, tendências e generalizações sobre os dados sem que haja a interferência do usuário. Dentre as atividades descritivas as regras de associação são consideradas uma das tarefas de mineração mais importantes e atrai muita atenção da academia e indústria.

De forma geral, os trabalhos que utilizam a mineração de dados educacionais adotam algoritmos ou técnicas utilizados tradicionalmente na mineração de dados e seguem as etapas de KDD, no qual os dados são obtidos, em alguns casos é realizada uma seleção dos atributos considerados pelo pesquisador como mais relevantes para a ocorrência da evasão escolar. Na etapa posterior, ocorre o tratamento dos dados. Na etapa da mineração de dados, geralmente são realizados vários testes utilizando algoritmos para que sejam feitas análises entre os resultados obtidos. Por fim, ocorre a interpretação e avaliação dos resultados, sendo possível identificar padrões de alunos para evasão escolar.

De acordo com Han, Pei e Kamber (2011), a mineração de dados possui dois grandes grupos de atividades: As atividades preditivas, que buscam um modelo de conhecimento previsão de valores de certos atributos em novas situações por meio de um histórico de dados e as atividades descritivas, que descrevem de forma compreensível um conhecimento existente em um conjunto de dados.

## SAIBA MAIS

*Algumas sugestões de leitura para aprofundamento das temáticas tratadas neste guia:*



*ARAÚJO, Cristiane F; SANTOS, Roseli. A. A educação profissional de nível médio e os fatores internos/externos às instituições que causam a evasão escolar. In: INTERNATIONAL CONGRESS UNIVERSITY - INDUSTRY COOPERATION, 4., 2012, Taubaté. Anais: [...]. São Paulo: Universidade de Taubaté, 2012. Disponível em: <http://www.unitau.br/unindu/artigos/pdf525.pdf>. Acesso em: 30 ago. 2019.*



*BÜRKLE, Paula Yamada. Um método de pós-processamento de regras de associação com base nas relações de dependência entre os atributos. Dissertação de Mestrado em Computação, Otimização combinatória e inteligência artificial. Universidade Federal Fluminense. Rio de Janeiro, RJ, 107p, 2006.*



*PASTA, Arquelau. Aplicação da técnica de Data Mining na base de dados do ambiente de gestão educacional: um estudo de caso de uma instituição de ensino superior de Blumenau, SC. 153 f. Dissertação (Mestrado em Computação Aplicada) - Universidade do Vale do Itajaí, São José, 2011.*

2

Diagnóstico da situação do conjunto de dados dos alunos do DAEE - IFMT, Campus Cuiabá - Cel. Octayde Jorge da Silva, durante o período de 2009 a 2019

POWERBI

Análise  
descritiva

Perfil  
dos evadidos

Nesta seção será apresentado um diagnóstico da situação do conjunto de dados dos alunos do DAEE/CBA-OJS/IFMT. Estas informações são fundamentais para que o pesquisador tenha conhecimento do comportamento do conjunto de dados anteriormente à etapa de Mineração de Dados, além de auxiliar posteriormente na interpretação e análise das regras de associação geradas.

Trata-se de um estudo de caso, e a metodologia aplicada pode ser empregada em demais cursos de diferentes instituições. Neste estudo foi utilizado um conjunto de dados fornecido pelo CBA-OJS/IFMT, entretanto dados de outras instituições também podem ser utilizados. Portanto, os nomes das colunas podem variar conforme o sistema acadêmico utilizado pela instituição.

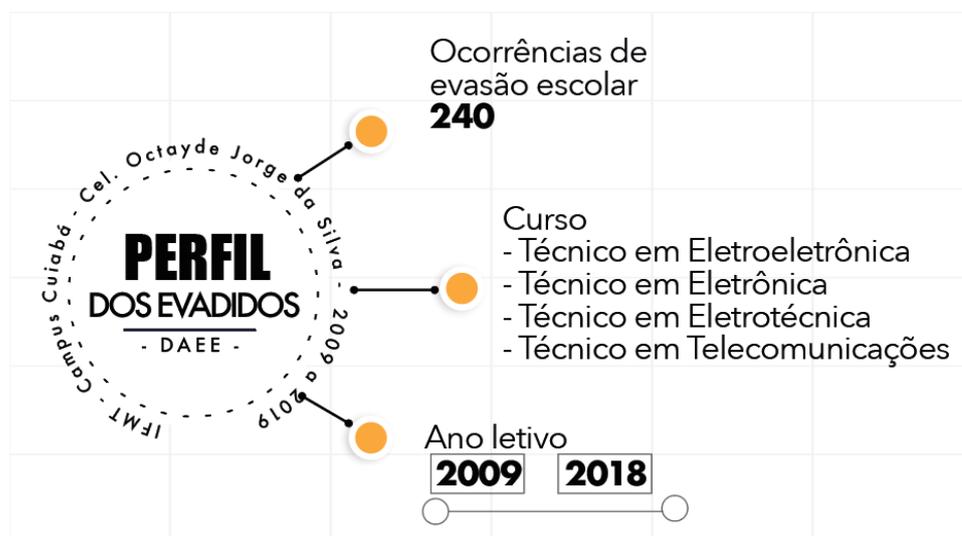
Foi solicitado o conjunto de dados dos cursos de Ensino Médio Integrado ao ensino técnico do DAEE/CBA-OJS/IFMT, referente ao período de 2009 a 2019, via Sistema Eletrônico do Serviço de Informações ao Cidadão (e-SIC). O sistema permite que qualquer pessoa, física ou jurídica, encaminhe pedidos de acesso à informação. Com o intuito de conhecer o perfil dos alunos evadidos, foram selecionados 240 do total de 721 alunos que apresentavam a "Situação\_matrícula" como "evasão", "cancelado", "transferido externo" e "transferido interno", segundo o conceito de aluno evadido da Plataforma Nilo Peçanha. Destaca-se que para a análise dos alunos evadidos foi considerado o período de 2009 a 2018, pois em 2019 não foi identificada nenhuma ocorrência de evasão.

Por meio de um Dashboard elaborado foram realizadas segmentações de dados que incluem o Curso e o Ano letivo. Desta forma, o leitor pode selecionar os dados de acordo com os cursos e anos específicos ou um período temporal entre 2009 a 2018, por exemplo, Curso de EMI Técnico em Eletrotécnica no período de 2011 a 2017.

Para este diagnóstico foi escolhido o *software* PowerBI Desktop, desenvolvido pela *Microsoft*, por ser um *software* gratuito e de fácil manuseio. Entretanto, outros *softwares*

podem ser adotados para esta atividade, como o *Google Data Studio*, *Tableau*, entre outros.

**Figura 2** – Segmentações de dados do perfil dos alunos evadidos dos cursos de EMI Técnico do DAEE/CBA-OJS/IFMT no *Dashboard* do *PowerBI Desktop*



Fonte: Adaptada pela pesquisadora, de *PowerBI Desktop* (2021)

## 2.1 CONHECENDO O PERFIL DOS EVADIDOS

### 2.1.1 Perfil Pessoal

No perfil pessoal foram analisados os atributos de Sexo, Idade, Raça/Cor, Estado Civil, Deficiência e Naturalidade UF.

Para análise de Sexo e Idade, foi elaborado um gráfico cruzando as variáveis Sexo por Idade. A maior parte dos alunos é do sexo masculino (81,25%) e possuía 14 anos de idade no momento da inscrição (46,66%).

No tocante à Raça/Cor dos alunos, 39,58% dos alunos são brancos, 34,58% dos alunos são mestiços, 17,92% são negros, 14,17% se autodeclararam outra raça, 1,67% são indígenas e 0,42% são asiáticos. Em relação ao Estado Civil, 99,17% dos



**Gráfico 1** – Renda familiar dos alunos evadidos dos cursos de EMI Técnico do DAEE/CBA-OJS/IFMT, no período de 2009 a 2018



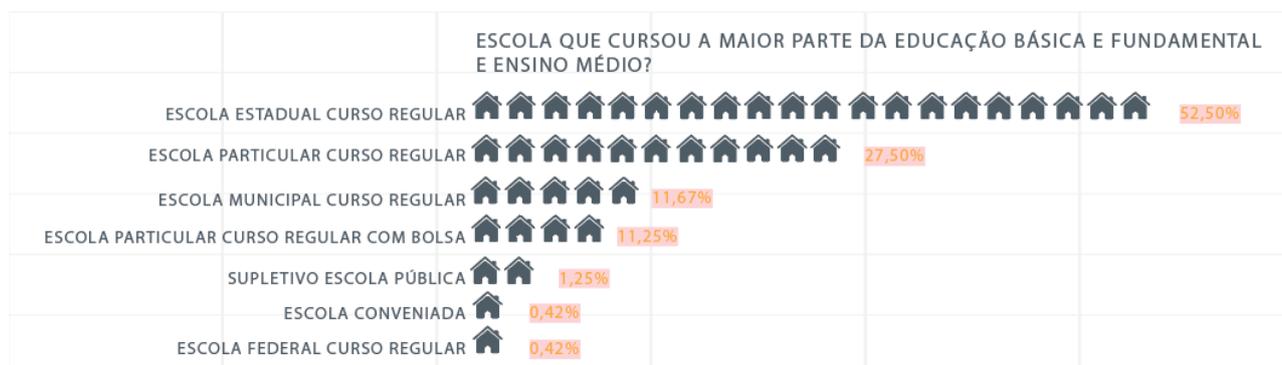
Fonte: Elaborado pela pesquisadora a partir de dados obtidos por meio dos sistemas Qacadêmico e Qseleção do IFMT (2021)

A maior parcela dos alunos (41,25%) possui renda familiar acima de 3 salários-mínimos. A quantidade média de pessoas por casa é de 4,15.

Quanto ao uso de computador em casa, verificou-se que 85,42% possuem computador e 19,17% não possuem computador em casa. Em relação ao tipo de serviço médico utilizado, 51,79% dos alunos utilizam Clínicas e hospitais públicos; 31,87% utilizam clínicas e hospitais particulares ou públicos contanto que possuam plano de saúde; 12,35% utilizam clínicas e hospitais particulares e 3,98% utilizam farmácias e remédios caseiros. No tocante à origem dos alunos, 98,35% são oriundos de zonas urbanas e 1,65% são de zonas rurais.

O tipo de escola que os alunos cursaram a maior parte da educação básica, fundamental e ensino médio é apresentada no gráfico 2.

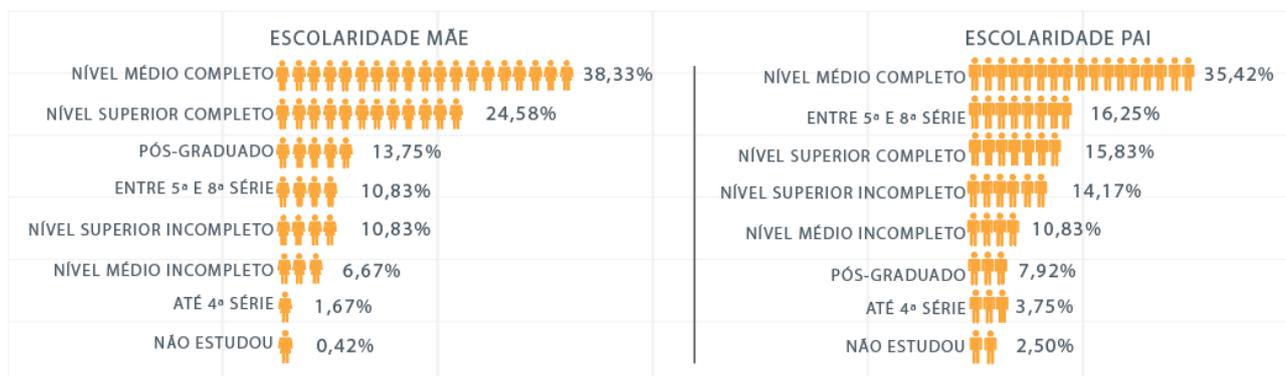
**Gráfico 2** – Escola que os alunos evadidos dos cursos de EMI Técnico do DAEE/CBA-OJS/IFMT, no período de 2009 a 2018, cursaram a maior parte da educação



Fonte: Elaborado pela pesquisadora a partir de dados obtidos por meio dos sistemas Qacadêmico e Qseleção do IFMT (2021)

É importante ressaltar que o maior percentual de alunos é proveniente de escolas estaduais e municipais públicas. Em seguida foi verificada a escolaridade dos pais dos alunos.

**Gráfico 3** – Escolaridade da mãe e do pai dos alunos evadidos dos cursos de EMI Técnico do DAEE/CBA-OJS/IFMT, no período de 2009 a 2018



Fonte: Elaborado pela pesquisadora a partir de dados obtidos por meio dos sistemas Qacadêmico e Qseleção do IFMT (2021)

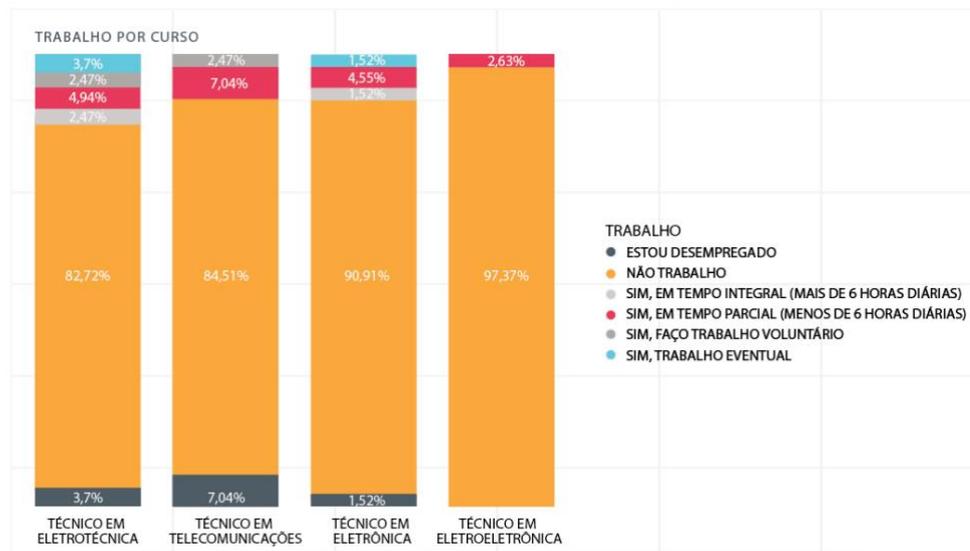
Constatou-se que a maior parcela de mães e pais apresentam apenas o nível médio completo. Verificou-se ainda que o nível de escolaridade das mães é superior à escolaridade dos pais. Vale destacar que a percentual de mães que possuem pós-graduação é aproximadamente o dobro do percentual de pais que possuem pós-graduação, bem como,

o percentual de mães que possuem nível superior completo é 8,75% acima do percentual de pais com a mesma escolaridade.

Em relação aos meios de transporte mais utilizados pelos alunos, o ônibus é utilizado por 64,17% dos alunos, em segundo lugar 20,42% dos alunos se locomovem a pé, em terceiro o carro próprio com 12,08% dos alunos, em quarto a bicicleta com 5,83%, em quinto a carona com 2,92%, em sexto o transporte escolar com 2,92% e em último lugar a motocicleta com 0,83%.

Para a investigação do atributo “Trabalho” foi realizada uma análise mais detalhada e segmentada, por meio do cruzamento com o atributo “Curso”, conforme o gráfico abaixo.

**Gráfico 4** – Situação de “Trabalho por curso” dos alunos evadidos dos cursos de EMI Técnico do DAEE/CBA-OJS/IFMT, no período de 2009 a 2018



Fonte: Elaborado pela pesquisadora a partir de dados obtidos por meio dos sistemas Qacadêmico e Qseleção do IFMT (2021)

Verificou-se que a maior parte dos alunos também não trabalhava ou estava desempregada. Os resultados, em ordem decrescente, foram os seguintes: no curso de EMI Técnico em Eletrotécnica, houve a menor quantidade de alunos que não

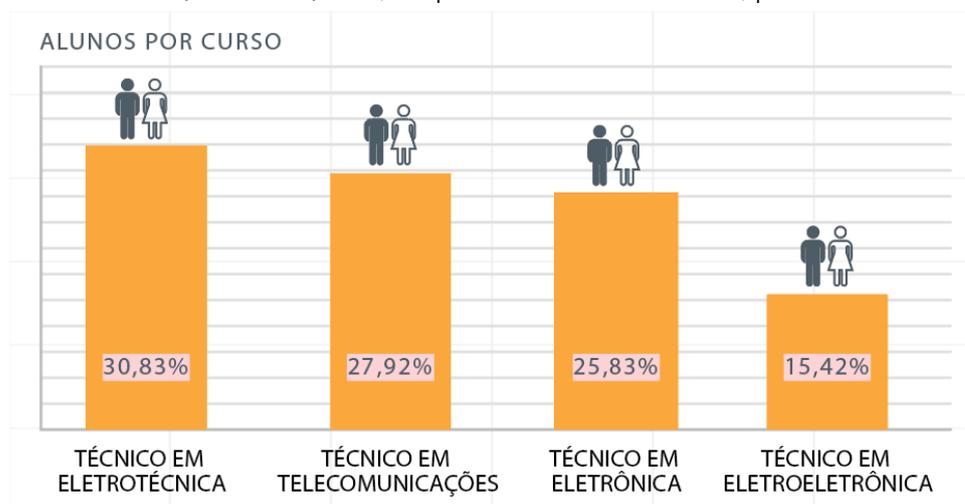
trabalhavam (82,72%) e a maior quantidade de alunos que exerciam algum tipo de trabalho (17,28%); o próximo curso foi o de EMI Técnico em Telecomunicações, em que 84,51% dos alunos não trabalhavam ou estavam desempregados e 15,49% dos alunos exerciam algum tipo de trabalho. Em seguida, no curso de EMI Técnico em Eletrônica, 93,43% dos alunos não trabalhavam ou estavam desempregados e 7,59% dos alunos exerciam algum tipo de trabalho. Por último, no curso de EMI Técnico em Eletroeletrônica, apresentou-se a maior quantidade de alunos que não trabalhavam ou estavam desempregados (97,37%) e, conseqüentemente, a menor quantidade de alunos que exerciam algum tipo de trabalho (2,63%).

É importante destacar que o curso de EMI Técnico em Eletrotécnica apresentou a maior quantidade de alunos que exerciam algum tipo de trabalho, além de ser aquele com a menor quantidade de alunos nos dados completos e a maior quantidade de alunos evadidos. Vale ressaltar ainda que o curso de EMI Técnico em Eletroeletrônica teve a maior quantidade de alunos que não trabalhavam, apresentou a maior quantidade de alunos nos dados completos e a menor quantidade de alunos evadidos.

### **2.1.3 Perfil Acadêmico**

No perfil acadêmico foi explorada a Quantidade de alunos por curso e as disciplinas por curso. O gráfico abaixo apresenta a situação geral dos cursos de Ensino Médio integrado ao técnico do DAEE/CBA-OJS/IFMT, por meio do cruzamento dos atributos "Matrícula" (contagem distinta) e o "Curso".

**Gráfico 5** – Situação dos alunos evadidos dos cursos de EMI Técnico do DAEE/CBA-OJS/IFMT, no período de 2009 a 2018, por curso



Fonte: Elaborado pela pesquisadora a partir de dados obtidos por meio dos sistemas Qacadêmico e Qseleção do IFMT (2021)

O curso de EMI Técnico em Eletroeletrônica possui 247 do total de 721 alunos, sendo o curso com maior percentual de alunos (34,26%) e também é o curso com o menor percentual de alunos evadidos (15,42%), conforme o gráfico acima.

O curso de EMI Técnico em Eletrotécnica possui a maior quantidade de alunos evadidos, 74 dos 240 alunos, em seguida o curso de EMI Técnico em Telecomunicações com 67 alunos evadidos, seguido de EMI Técnico em Eletrônica com 62 alunos evadidos e por último o curso de EMI Técnico em Eletroeletrônica com 37 alunos evadidos.

Dessa forma, é possível afirmar que, no período entre 2009 e 2018, dos 143 alunos do curso de EMI Técnico em Eletrotécnica, aproximadamente 51,74% dos alunos evadiram; no curso de EMI Técnico em Eletrônica, cerca de 35,63% evadiram; no curso de EMI Técnico em Telecomunicações, 42,67% aproximadamente evadiram e, por fim, no curso de EMI

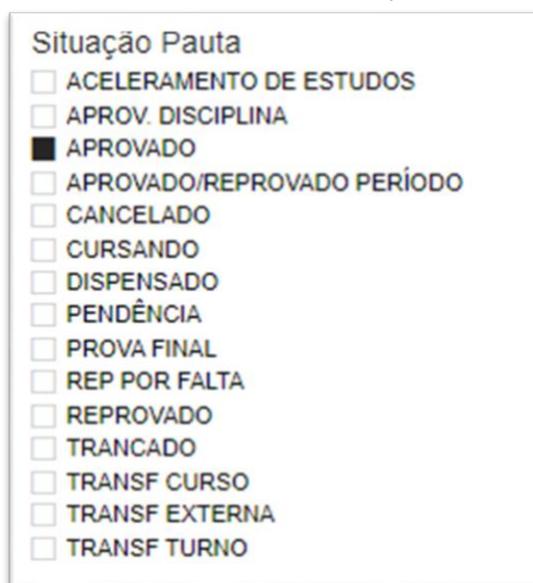
Técnico em Eletroeletrônica, apenas 14,97% dos alunos evadiram.

### 2.1.3.1 Análise das Disciplinas por curso

O *Dashboard* permite ao leitor realizar diferentes análises, conforme o curso desejado, o período, a disciplina e a situação de pauta. Por meio da segmentação de dados do atributo "Situação\_Pauta" é possível filtrar todas as ocorrências de matrículas nas disciplinas nos respectivos cursos.

---

**Figura 3** – Segmentações da "Situação de Pauta" do perfil dos alunos evadidos dos cursos de EMI Técnico do DAEE/CBA-OJS/IFMT, no *Dashboard* do *PowerBI Desktop*

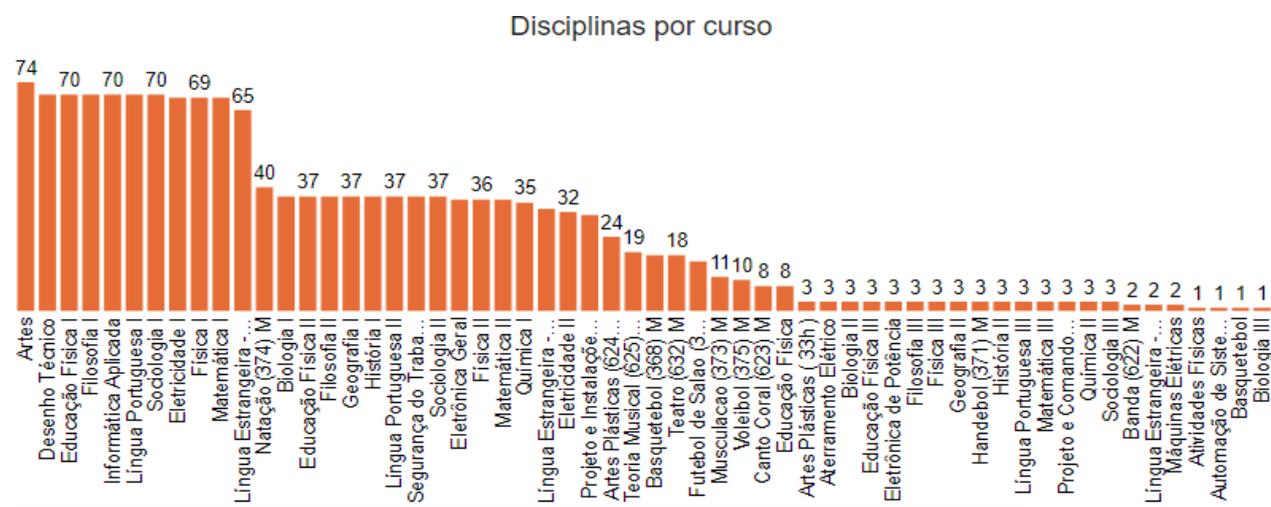


Fonte: PowerBI Desktop

---

A seguir será detalhada a análise de disciplinas do curso de EMI Técnico em Eletrotécnica por apresentar o maior quantitativo de ocorrências em disciplinas, durante o período temporal de 2009 a 2018. O gráfico abaixo demonstra a situação de pauta "Aprovado".

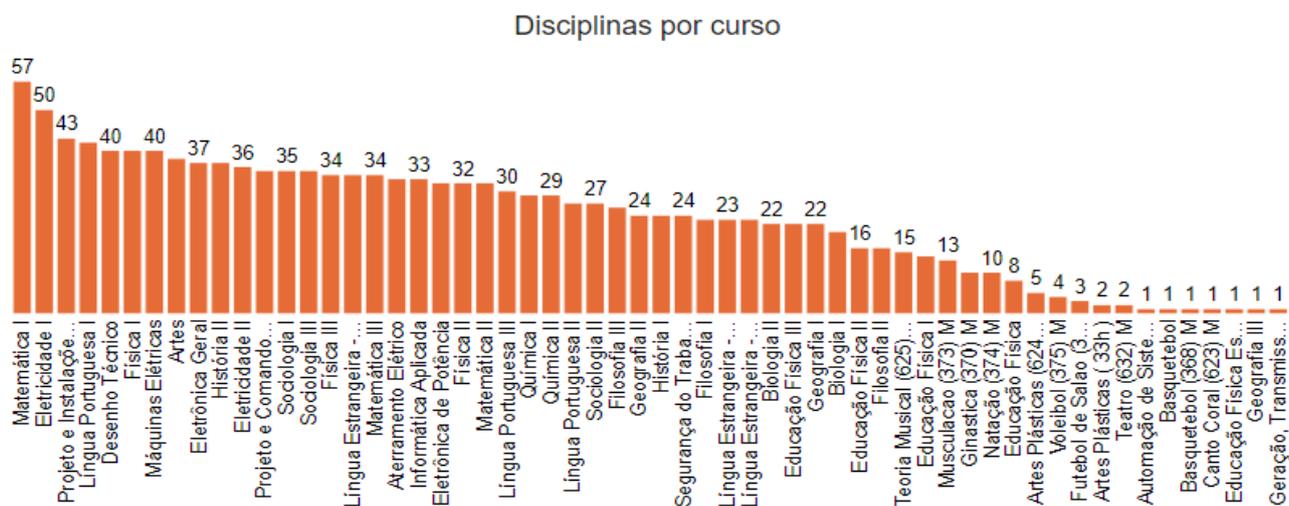
**Gráfico 6** – Ocorrências em disciplinas no curso EMI Técnico em Eletrotécnica do DAEE/CBA-OJS/IFMT, no período de 2009 a 2018, Situação de Pauta: Aprovado



Fonte: Elaborado pela pesquisadora a partir de dados obtidos por meio dos sistemas Qacadêmico e Qseleção do IFMT (2021)

No curso de EMI Técnico em Eletrotécnica, as 5 disciplinas com maiores índices de aprovação são Artes, Desenho Técnico, Educação Física, Filosofia e Informática. Dentre elas, apenas a disciplina de Desenho Técnico não pertence ao núcleo de “Disciplinas da Base comum”. No próximo gráfico foi aplicado o filtro de situação de pauta “Reprovado”.

**Gráfico 7** – Ocorrências em disciplinas no curso EMI Técnico em Eletrotécnica do DAEE/CBA-OJS/IFMT, no período de 2009 a 2018, Situação de Pauta: Reprovado



Fonte: Elaborado pela pesquisadora a partir de dados obtidos por meio dos sistemas Qacadêmico e Qseleção do IFMT (2021)

As 5 disciplinas com maiores índices de reprovação são Matemática I, Eletricidade I, Projetos e Instalações Elétricas em Alta tensão, Língua Portuguesa I e Desenho Técnico. Destaca-se que Eletricidade I, Projetos e Instalações Elétricas em Alta tensão e Desenho Técnico pertencem ao núcleo “Disciplinas da Base Profissional”.

## SAIBA MAIS

A seção 2 apresentou um diagnóstico da situação do conjunto de dados dos alunos dos cursos de EMI Técnico do DAEE/CBA-OJS/IFMT e aqui são apontados os QR Codes para acesso aos Dashboards elaborados no PowerBI. Para acesso, aponte a câmera do seu celular ou acesse o link ao lado de cada item.



*Dashboard dos dados dos alunos evadidos.*



*Link: <https://bit.ly/3C1mV5T>*



*Dashboard dos dados completos.*



*Link: <https://bit.ly/3heLfYk>*

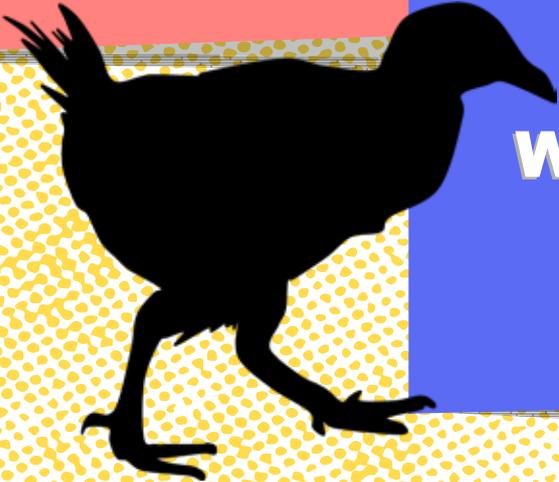
3

## Passo a Passo da aplicação da Mineração de Regras de Associação

PASSO A PASSO

Conjunto de dados  
Regras de Associação

WEKA

A black silhouette of a penguin is positioned to the left of a blue rectangular box with a white border and a white arrow pointing downwards. The word "WEKA" is written in white capital letters on the blue box.

Por meio da aplicação de Mineração de Regras de Associação é possível detectar e categorizar os elementos que contribuem para a identificação de padrões de alunos com risco de evasão escolar.

A seguir será apresentado o passo a passo de como executar a aplicação desta técnica no *software* Weka (*Waikato Environment for Knowledge Analysis*) 3.8.6, implementado em linguagem Java e desenvolvido pela Universidade de Waikato, na Nova Zelândia, em 1999.

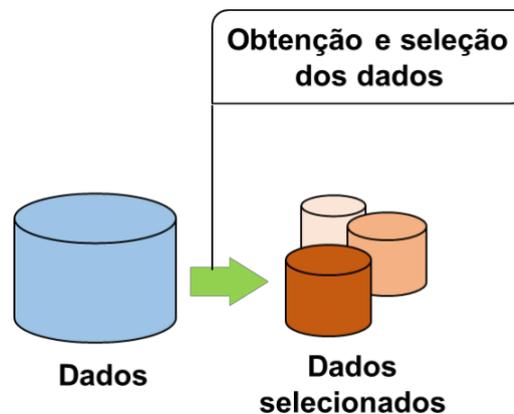
O Weka é um *software* livre de código aberto para a execução das atividades de mineração de dados. É composto por ferramentas de preparação de dados, classificação, regressão, agrupamento, mineração de regras de associação e visualização (Weka, 2019). O *software* é desenvolvido em Java, seu uso é gratuito e seu idioma é o Inglês. O download do Weka é realizado diretamente no site do *software*, que contém informações sobre a instalação, disponível para os sistemas operacionais Windows, MacOS, Linux, entre outros.

Os procedimentos executados seguem os passos do processo de KDD, tendo como exemplo a aplicação realizada no conjunto de dados dos alunos dos cursos de Ensino Médio Integrado ao ensino técnico do DAEE/CBA-OJS/IFMT, no período de 2009 a 2019, a fim de auxiliar o entendimento do leitor.

### **3.1 PASSO A PASSO**

#### **3.1.1 Obtenção e Seleção dos dados**

**Figura 4** – Etapa de “Obtenção e seleção dos dados”



Fonte: Elaborada pela pesquisadora, com o uso do software Power Point (2022)

A primeira etapa para realização da Mineração de Regras de Associação consiste na obtenção e seleção de um conjunto de dados.

Para a identificação de padrões de alunos com risco de evasão escolar é importante considerar dados de diferentes tipos como pessoais, sociais, econômicos e acadêmicos, visto que a evasão é um fenômeno multifatorial.

No conjunto de dados dos alunos dos cursos de Ensino Médio Integrado ao ensino técnico do DAEE/CBA-OJS/IFMT, os dados foram recebidos em três arquivos separados, em formato Excel, oriundos de duas bases de dados distintas: Os dados pessoais e acadêmicos foram extraídos do banco do sistema acadêmico denominado “Qacadêmico” e os dados socioeconômicos foram extraídos do sistema de inscrições “Qseleção”.

#### **De olho na dica**

Solicite os dados com um “identificador comum”, por exemplo o número de matrícula do aluno. Este identificador funciona como um “atributo-chave” caso seja necessária a unificação do conjunto de dados.

O agrupamento dos conjuntos de dados pessoais e socioeconômicos dos cursos do DAEE/CBA-OJS/IFMT foi

realizado por um script desenvolvido especificamente para este fim, na linguagem Python.

É importante destacar que o conjunto de dados exemplificado neste guia apresentou um cenário ruim, tendo em vista a sua divisão em três arquivos. Contudo, caso o leitor tenha acesso a um conjunto de dados num único arquivo, a etapa de unificação dos dados pode ser suprimida.

**Figura 5** – Conjunto de dados pessoais e socioeconômicos dos alunos evadidos dos cursos de EMI Técnico do DAEE/CBA-OJS/IFMT, no período 2009 a 2019

A	B	C	D	E	F	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	
	COD_M	NOME	ANO LI	PERIOD	MATRI	DESC C	COD_C	NR INS	SIT_CA	COD_C	DESC I	COD_C	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
1	0	9acc420e3eb4c97394	2009	1	c7b2a8826	Edital 20019b1fafd14b749fa825d	INSCRIÇA	143	Eletrônica	7159	Feminino	14 ANOS	SOLTEIRC	BRANCA	OUTRA	NÃO TRAE	4	D		
2	1	84da14d2fe22cf00c1	2009	1	30104ae7b	Edital 2001b4b4a12ec491173596	INSCRIÇA	143	Eletrônica	3123	Feminino	14 ANOS	SOLTEIRC	BRANCA	CATOLICANÃO TRAE	5	A			
3	2	522de1831a9201221e	2009	1	3ff750414f	Edital 20016676319ccf09093216	INSCRIÇA	143	Eletrônica	2309	Feminino	Até 13 AN	SOLTEIRC	MESTIÇA	CATOLICANÃO TRAE	4	A			
4	3	ca4b8961c2914edec	2009	1	e694ade0c	Edital 2001dd792a02e45155d043	INSCRIÇA	143	Eletrônica	2734	Masculino	14 ANOS	SOLTEIRC	OUTRA	CATOLICANÃO TRAE	4	A			
5	4	7767383a15b0ded8ec	2009	1	621010c5e	Edital 20011bc1bcb8e2ee1aa43a	INSCRIÇA	143	Eletrônica	8610	Feminino	14 ANOS	SOLTEIRC	MESTIÇA	CATOLICANÃO TRAE	4	A			
6	5	77503f5a2ec2e2d21a	2009	1	63fd1feb7c	Edital 200196f15ef3afca9e263d	INSCRIÇA	143	Eletrônica	7780	Masculino	Até 13 AN	SOLTEIRC	BRANCA	CATOLICANÃO TRAE	7	A			
7	6	de9af6ee812bc1a68fe	2009	1	3b9177daf	Edital 200117708cc6e5b72c4ca7	INSCRIÇA	143	Eletrônica	8646	Feminino	14 ANOS	SOLTEIRC	ASIÁTICO	NENHUM/NÃO TRAE	7	D			
8	7	0d90da36e3e6d921a5	2009	1	f600917d8f	Edital 2001d6148c1519741f79a	INSCRIÇA	143	Eletrônica	2798	Feminino	14 ANOS	SOLTEIRC	BRANCA	NENHUM/NÃO TRAE	4	A			
9	8	e06766d7abddeeabb	2009	1	c6639ad85	Edital 200158ad24b97a52b27e7	INSCRIÇA	143	Eletrônica	3109	Feminino	14 ANOS	SOLTEIRC	BRANCA	CATOLICANÃO TRAE	5	A			
10	9	8fe4c8daefc3aa9d931	2009	1	c591d2213	Edital 20012644500db3ad32a3e	INSCRIÇA	143	Eletrônica	6633	Feminino	Até 13 AN	SOLTEIRC	NEGRA	EVANGÉL/NÃO TRAE	7	A			
11	10	bf3a92366:5443a35df	2009	1	e5e2810512	Edital 20012a10acc7f4f2706558	INSCRIÇA	143	Eletrônica	3279	Masculino	14 ANOS	SOLTEIRC	MESTIÇA	CATOLICANÃO TRAE	4	A			
12	11	eb304693417a7e59bf	2009	1	75510387e	Edital 2001139171fad_cfa1150f17	INSCRIÇA	143	Eletrônica	2838	Masculino	14 ANOS	SOLTEIRC	OUTRA	EVANGÉL/NÃO TRAE	4	D			
13	12	b72d2136e3d39e4d57	2009	1	5888c01a3f	Edital 2001abc55662decc38f8	INSCRIÇA	143	Eletrônica	7858	Masculino	14 ANOS	SOLTEIRC	BRANCA	CATOLICANÃO TRAE	3	D			
14	13	b72d2136e3d39e4d57	2009	1	5888c01a3f	Edital 200156268a12ceef19c5431	CADASTR	2648	Eletrônica	203694	Masculino	15 ANOS	SOLTEIRC	BRANCA	CATOLICANÃO TRAE	3	D			
15	14	b72d2136e3d39e4d57	2009	1	5888c01a3f	Edital 200156268a12ceef19c5431	CADASTR	2648	Eletrônica	203694	Masculino	15 ANOS	SOLTEIRC	BRANCA	CATOLICANÃO TRAE	3	D			
16	15	49b697e9747dbd5eaa	2009	1	64a81c145	Edital 20012650a12dc4c629c4c3	INSCRIÇA	143	Eletrônica	2374	Masculino	14 ANOS	SOLTEIRC	OUTRA	NÃO TRAE	5	A			
17	16	be27a9d8a1537cec3e	2009	1	29d9df361f	Edital 2001f15bb57fdb8157eff838	INSCRIÇA	143	Eletrônica	3071	Masculino	15 ANOS	SOLTEIRC	MESTIÇA	ESPIRITA NÃO TRAE	5	A			
18	17	906b5c568be1cf86168	2009	1	418c7de23	Edital 20012af731898_7a7ea1512	INSCRIÇA	143	Eletrônica	7945	Feminino	14 ANOS	SOLTEIRC	MESTIÇA	EVANGÉL/NÃO TRAE	4	D			
19	18	b7e628c84679c24511	2009	1	867794815	Edital 2001db1908671afc3b9d34	INSCRIÇA	143	Eletrônica	3104	Masculino	14 ANOS	SOLTEIRC	BRANCA	CATOLICANÃO TRAE	5	A			
20	19	433a39e053ab760cdf	2009	1	64615d336	Edital 2001f5733b67215a929e48f	INSCRIÇA	143	Eletrônica	2525	Masculino	14 ANOS	SOLTEIRC	BRANCA	CATOLICANÃO TRAE	6	A			
21	20	b1e87953fe16c413ce	2009	1	5be086d5c	Edital 2001bf1d3bbcc71dd4b9d0e	INSCRIÇA	143	Eletrônica	3591	Masculino	14 ANOS	SOLTEIRC	ASIÁTICO	NENHUM/NÃO TRAE	8	A			
22	21	c3b85c79c1660f1d0d	2009	1	019186995	Edital 2001cb8618ba34adbea83	INSCRIÇA	143	Eletrônica	8182	Masculino	14 ANOS	SOLTEIRC	BRANCA	CATOLICANÃO TRAE	5	A			
23	22	c664073c:7372800ca	2009	1	f1f6a896ff5	Edital 2001f66c45317:178116e63	INSCRIÇA	143	Eletrônica	7817	Masculino	14 ANOS	SOLTEIRC	MESTIÇA	CATOLICANÃO TRAE	3	D			
24	23	1ae0f0e5c146ef0defa5	2009	1	b060ec81e	Edital 20019c117dc239170981d5	INSCRIÇA	143	Eletrônica	3333	Feminino	Até 13 AN	SOLTEIRC	MESTIÇA	CATOLICANÃO TRAE	4	A			
25	24	89f0da5e7:1984e9a5c	2009	1	245e6b0e0	Edital 2001488264a3e77f8d53c1	INSCRIÇA	143	Eletrônica	2890	Masculino	14 ANOS	SOLTEIRC	NEGRA	CATOLICANÃO TRAE	4	A			
26	25	9633c992e9c0dec63e	2009	1	5dac81133	Edital 200140b3c4d649a3d61273	INSCRIÇA	143	Eletrônica	2190	Feminino	14 ANOS	SOLTEIRC	MESTIÇA	CATOLICANÃO TRAE	5	A			
27	26	c06b7bd94b3018a3c4	2009	1	e03996974	Edital 20016ca1e9a4e3219b5be7	INSCRIÇA	143	Eletrônica	2801	Feminino	15 ANOS	SOLTEIRC	ASIÁTICO	CATOLICANÃO TRAE	5	A			
28	27	31327845a8c595e0cfe	2009	1	244faea1a	Edital 2001a79141fe11581a8e263	INSCRIÇA	143	Eletrônica	3558	Masculino	15 ANOS	SOLTEIRC	NEGRA	CATOLICANÃO TRAE	4	D			
29	28	337dedaf981cc2564bfb	2009	1	6245e783c	Edital 200108ab947873d4bce24c	INSCRIÇA	143	Eletrônica	7903	Masculino	Até 13 AN	SOLTEIRC	BRANCA	OUTRA NÃO TRAE	5	D			
30	29	1ae0f0e5c146ef0defa5	2009	1	b060ec81e	Edital 20019c117dc239170981d5	INSCRIÇA	143	Eletrônica	3333	Feminino	Até 13 AN	SOLTEIRC	MESTIÇA	CATOLICANÃO TRAE	4	A			
31	30	be27a9d8a1537cec3e	2009	1	29d9df361f	Edital 2001f15bb57fdb8157eff838	INSCRIÇA	143	Eletrônica	3071	Masculino	15 ANOS	SOLTEIRC	MESTIÇA	ESPIRITA NÃO TRAE	5	A			
32	31	c06b7bd94b3018a3c4	2009	1	e03996974	Edital 20016ca1e9a4e3219b5be7	INSCRIÇA	143	Eletrônica	2801	Feminino	15 ANOS	SOLTEIRC	ASIÁTICO	CATOLICANÃO TRAE	5	A			
33	32	4ca7e1e7a5e4996b1	2009	1	1c933d6dc	Edital 2001cb0dbd12431e32558f	INSCRIÇA	143	Eletrônica	5662	Masculino	15 ANOS	SOLTEIRC	MESTIÇA	EVANGÉL/NÃO TRAE	4	A			
34	33	433a39e053ab760cdf	2009	1	64615d336	Edital 2001f5733b67215a929e48f	INSCRIÇA	143	Eletrônica	2525	Masculino	14 ANOS	SOLTEIRC	BRANCA	CATOLICANÃO TRAE	6	A			
35	34	c3b85c79c1660f1d0d	2009	1	019186995	Edital 2001cb8618ba34adbea83	INSCRIÇA	143	Eletrônica	8182	Masculino	14 ANOS	SOLTEIRC	BRANCA	CATOLICANÃO TRAE	5	A			
36	35	c664073c:7372800ca	2009	1	f1f6a896ff5	Edital 2001f66c45317:178116e63	INSCRIÇA	143	Eletrônica	7817	Masculino	14 ANOS	SOLTEIRC	MESTIÇA	CATOLICANÃO TRAE	3	D			

Fonte: Elaborada pela pesquisadora a partir de dados obtidos por meio dos sistemas Qacadêmico e Qseleção do IFMT (2021)

Lembre-se de que dados sensíveis devem ser codificados ou suprimidos, conforme as orientações da Lei Geral de Proteção de Dados Pessoais (LGPD).

Os dados acadêmicos permaneceram em um arquivo separado, visto que as regras de associação destes foram geradas de forma separada.

Depois da unificação dos dados, é importante selecionar apenas os dados que serão utilizados para a geração das regras de associação. Caso você queira analisar um departamento ou curso específico, aplique um filtro!

No conjunto de dados pessoais e socioeconômicos foram selecionadas 735 ocorrências de um total de 613 alunos evadidos e concluintes dos cursos de EMI Técnico do DAEE/CBA-OJS/IFMT, no período de 2009 a 2019. No conjunto de dados acadêmicos foram selecionadas 621 ocorrências de um total de 613 alunos evadidos e concluintes dos cursos de EMI Técnico do DAEE/CBA-OJS/IFMT, no período de 2009 a 2019.

### Atenção!

É importante identificar no seu conjunto de dados a coluna que contém dados acerca da situação de matrícula do aluno. No caso do conjunto de dados dos cursos do DAEE/CBA-OJS/IFMT são empregados os termos apresentados na Plataforma Nilo Peçanha.

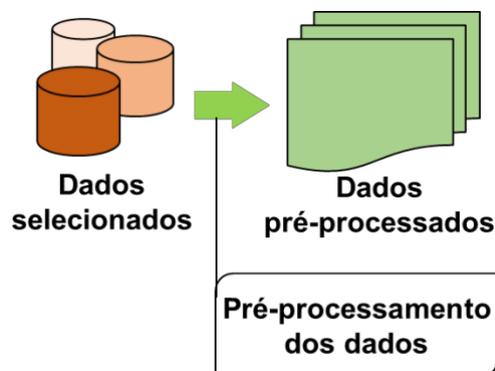
Os alunos considerados como evadidos apresentam a "Situação\_matrícula" como "evasão", "cancelado", "transferido externo" e "transferido interno", segundo o conceito de aluno evadido da Plataforma Nilo Peçanha. Em relação aos alunos concluídos foram considerados aqueles que apresentam a "Situação\_matrícula" como "concluído", "concludente" e "estagiário (concludente)".

Para a execução da técnica de Mineração de regras de associação não foram considerados os alunos com a "Situação\_matrícula" como "matriculado".

Importante lembrar que alunos que apresentam a "Situação\_matrícula" como "matriculado", ainda não concluíram sua vida acadêmica. Para identificação de padrões de alunos com risco de evasão escolar, é inviável realizar a comparação destes alunos com os alunos evadidos e concluintes.

### 3.1.2 Pré-processamento e tratamento dos dados

**Figura 6** – Etapa de “Pré-processamento e tratamento de dados”



Fonte: Elaborada pela pesquisadora, com o uso do software Power Point (2022)

Esse é o momento de verificar possíveis inconsistências, dados faltantes e realizar as adequações no seu conjunto de dados!

A etapa de pré-processamento dos dados é responsável pela compreensão e preparação do conjunto de dados para a etapa de “Data Mining ou Mineração de Dados” no Weka.

No conjunto de dados dos alunos nos cursos de Ensino Médio Integrado ao ensino técnico do DAEE/CBA-OJS/IFMT, foram selecionados o total de 54.349 dados dos alunos.

Durante o pré-processamento foram detectados 3397 dados socioeconômicos inconsistentes e 8067 dados socioeconômicos faltantes, que foram excluídos, **resultando no total de 42.885 dados.**

Devido à disposição das disciplinas no conjunto de dados acadêmicos foi necessária a realização de um tratamento de dados. As disciplinas encontravam-se dispostas em linhas e dessa forma cada aluno constava em mais de uma ocorrência, o que impossibilitava a geração das regras de associação no Weka.

O tratamento dos dados foi realizado por meio da ferramenta de gerenciamento de dados “Microsoft SQL Server Management Studio 2017”. As disciplinas, que anteriormente encontravam-se dispostas em linhas, foram reorganizadas em

colunas. Cada linha passou a representar um aluno e os valores apresentaram as seguintes situações de pauta do aluno em determinada disciplina: “Aprovado” e “Reprovado”. Dessa forma, o aluno constava em apenas uma linha, sem duplicações, conforme figura 7:

**Figura 7** – Conjunto de dados acadêmicos dos alunos evadidos dos cursos de EMI Técnico do DAEE/CBA-OJS/IFMT, no período de 2009 a 2019, após o tratamento

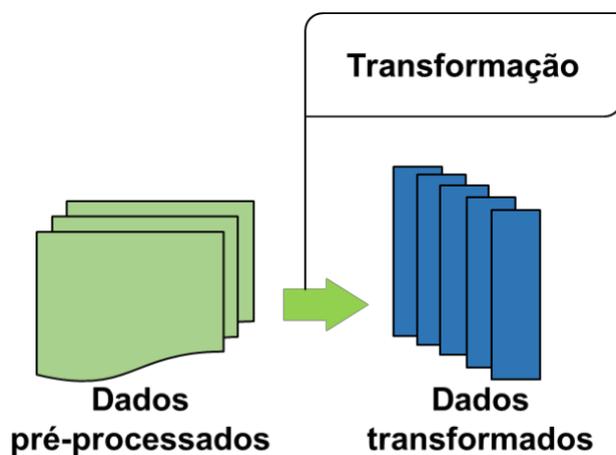
A	B	C	E	F	G	H	
1	COD_MATRICULA	ANO LETIVO	PERIODO LETIVO	Análise de Circuitos	Análise de Redes Telefônicas	Arte Contemporânea	Artes
2	000290e4dd7ca176ed6564ceefc4c06965def23a	2016	1				APROVADO
3	002489bacde0758c0b221fc313b3c3e6400b2e75	2011	1				APROVADO
4	01973b5d0ae4eac9e2368e2fc8979dc44803f21	2015	1				CANCELADO
5	01e3ff1043cf649ef7884821446e639faa568481	2011	1				APROVADO
6	01f89517c81b199723ddf535d28fc0ec0cc262e3	2017	1			APROVADO	APROVADO
7	0290b06e02a7197bfb88a4cfa8e0db35cbb1f8d3	2011	1		APROVADO		APROVADO
8	03f49c64d6c4bacdbb263e752f40ff1de3d12aed	2013	1		APROVADO		APROVADO
9	04b42f134040e1da0dbc3a0f0e538e77d1cdd376	2012	1				APROVADO
10	04e0e135ba348cd085d9b850a25c930f58c8c01	2011	1				APROVADO
11	05734b0e8009ebfca5540d74cee7dd879fb1485c	2012	1				REPROVADO
12	06042c23bfa9ebf00078b0a649004685ad35d1c5	2017	1				APROVADO
13	063254e2f54a372656811ad5623fcdca4f760726b	2017	1				TRANSF EXTERNA
14	065a133a8dc72044bcc91c3c4430b63f351764e8	2013	1	APROVADO			APROVADO
15	06a9089d83b4a2d1982e6892f9109ee8153cf834	2016	1				APROVADO/REPROVADO PERÍODO
16	06a9aadbb0baafe0fd5418a15f2811854152df63	2014	1				APROVADO/REPROVADO PERÍODO
17	06d0acfc3d7d9336f3311505c068e154e8447166	2012	1				APROVADO
18	07040d70beeb0112e7885d6716ccf795e5e19541	2017	1				APROVADO
19	072e55d9e8886374b6c258b1c689992d4e5c4995	2017	1				APROVADO
20	075f693933d5599464ed440176ae552762f3658	2010	1				APROVADO/REPROVADO PERÍODO
21	07a2f824529d9f6a45f78a493ec1637465323ec7	2014	1	APROVADO			APROVADO
22	086879e94c385d9eb3601c5d0fd29a20b109593	2016	1				REPROVADO
23	08d2603475e036cd40cda84323d59194d051d6fb	2012	1				APROVADO
24	0a8009a155e87330dc8c29cb7b86c7eb15f35a89	2011	1				REPROVADO
25	0a917da58713975539bceebcf62e82728f8cfbf	2010	1				APROVADO
26	0ac9ba0227821cedcbb0317e60480a424e87ad3	2013	1				APROVADO
27	0b42678ea55b1dee994d605be032a27dc3413156	2016	1				APROVADO
28	0b81975b4d2a8817f0ed376de3c5ef2a34b1eeed	2014	1				APROVADO
29	0bff3ad123a9538309e5a928717ae3fa9b6997ce	2017	1				APROVADO
30	0c9ef5e6102ffb008f809005648496b22c95e4cb	2012	1		APROVADO		APROVADO

Fonte: Elaborada pela pesquisadora, a partir de dados obtidos por meio dos sistemas Qacadêmico e Qseleção do IFMT (2022)

### 3.1.3 Transformação dos dados

---

**Figura 8** – Etapa de “Transformação dos dados”



---

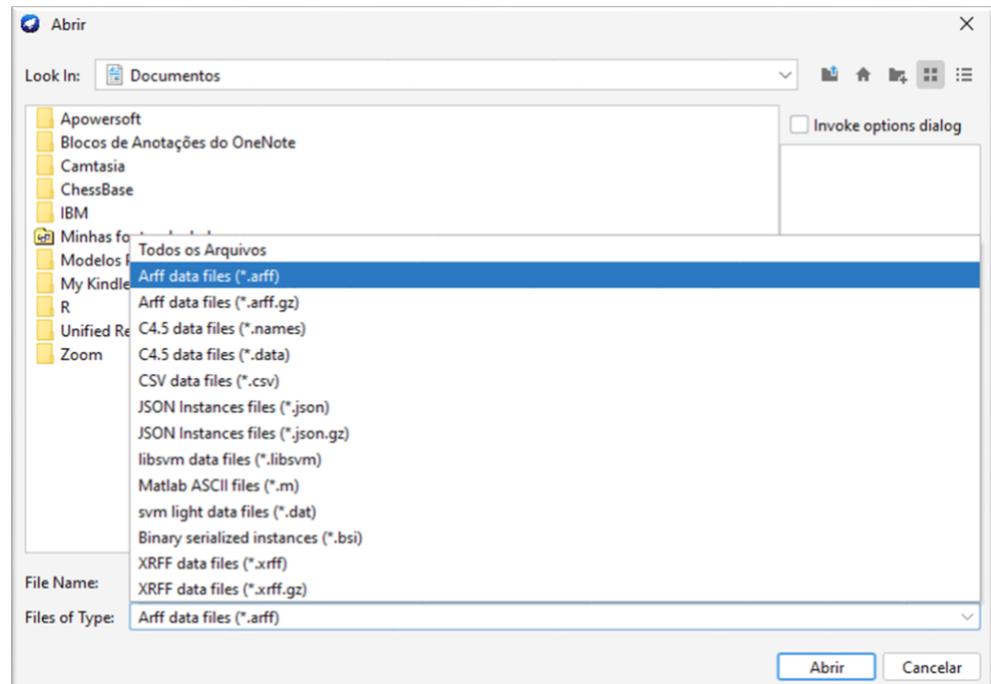
Fonte: Elaborada pela pesquisadora, a partir de dados obtidos por meio dos sistemas Qacadêmico e Qseleção do IFMT (2022)

---

Para a inserção do conjunto de dados do *software* Weka é necessário fazer a conversão do arquivo contendo o banco para a extensão adequada.

Embora o Weka seja compatível com diversas extensões, o formato de arquivo padrão utilizado pelo *software* é a extensão ARFF, que apresenta uma seção de cabeçalho composta por sequências numéricas, nominais ou data e uma seção de dados com os valores.

**Figura 09** – Tela “Open File” do Weka 3.8.6



Fonte: Weka 3.8.6

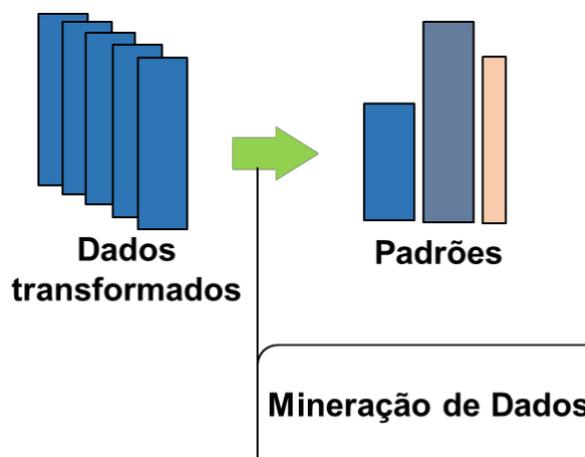
Para transformar o seu conjunto de dados em ARFF, o primeiro passo é salvá-los em CSV. Alguns ajustes podem ser necessários como substituição de ponto e vírgula por somente vírgula.

#### **De olho na dica**

O próprio Weka permite que o usuário salve um arquivo no formato ARFF, visto que ao abrir um arquivo CSV no Weka é possível salvá-lo em ARFF em seu computador.

### 3.1.4 Mineração dos Dados

**Figura 10** – Etapa de “Mineração de dados”



Fonte: Elaborada pela pesquisadora, com o uso do software Power Point (2022)

A etapa de Mineração dos Dados no Weka é responsável pela detecção e categorização dos elementos que contribuem para a identificação de um aluno com risco de evasão escolar.

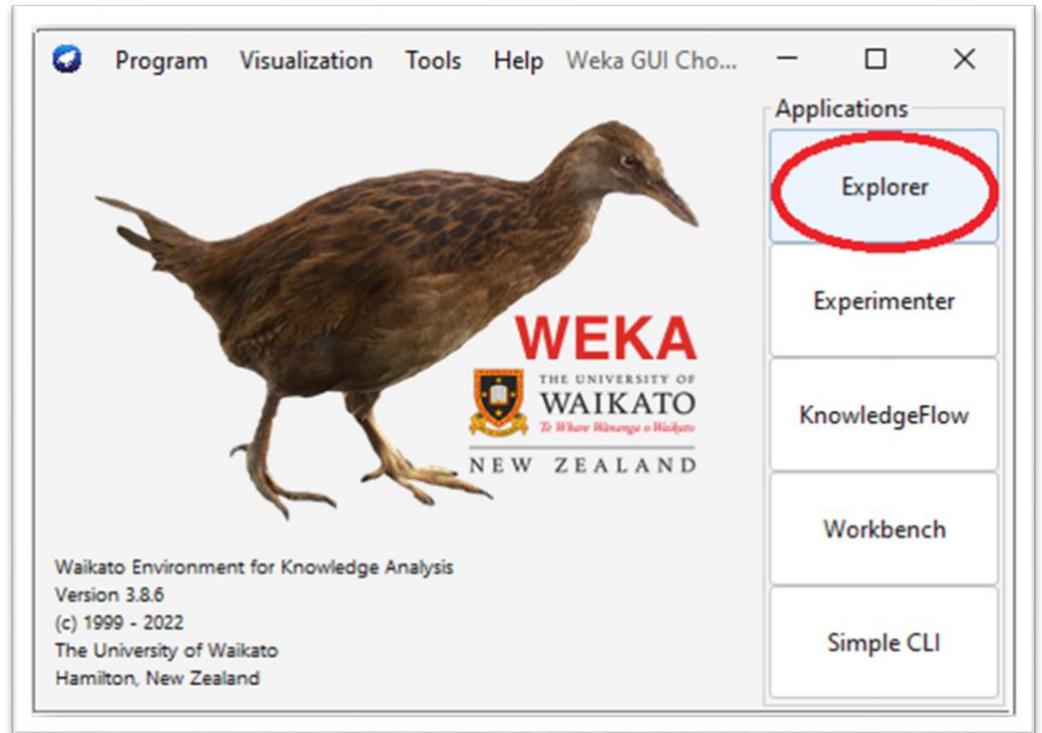
Devido à grande quantidade de atributos no conjunto de dados apresentado neste guia e com o intuito de realizar uma análise de resultados mais detalhada, optou-se por gerar as regras de associação dos dados pessoais e socioeconômicos separada dos dados acadêmicos.

Os passos detalhados a seguir referem-se à geração de regras dos dados acadêmicos, entretanto destaca-se que os procedimentos executados são os mesmos para os dados pessoais e socioeconômicos, bem como para outros conjuntos de dados.

Na etapa de Mineração de Dados todos os passos são executados no Weka 3.8.6. Ao clicar no botão “Explorer”, abre-se a tela principal do Weka.

Weka advém do pássaro Weka! Ele é um pássaro que não voa e é encontrado apenas na Nova Zelândia

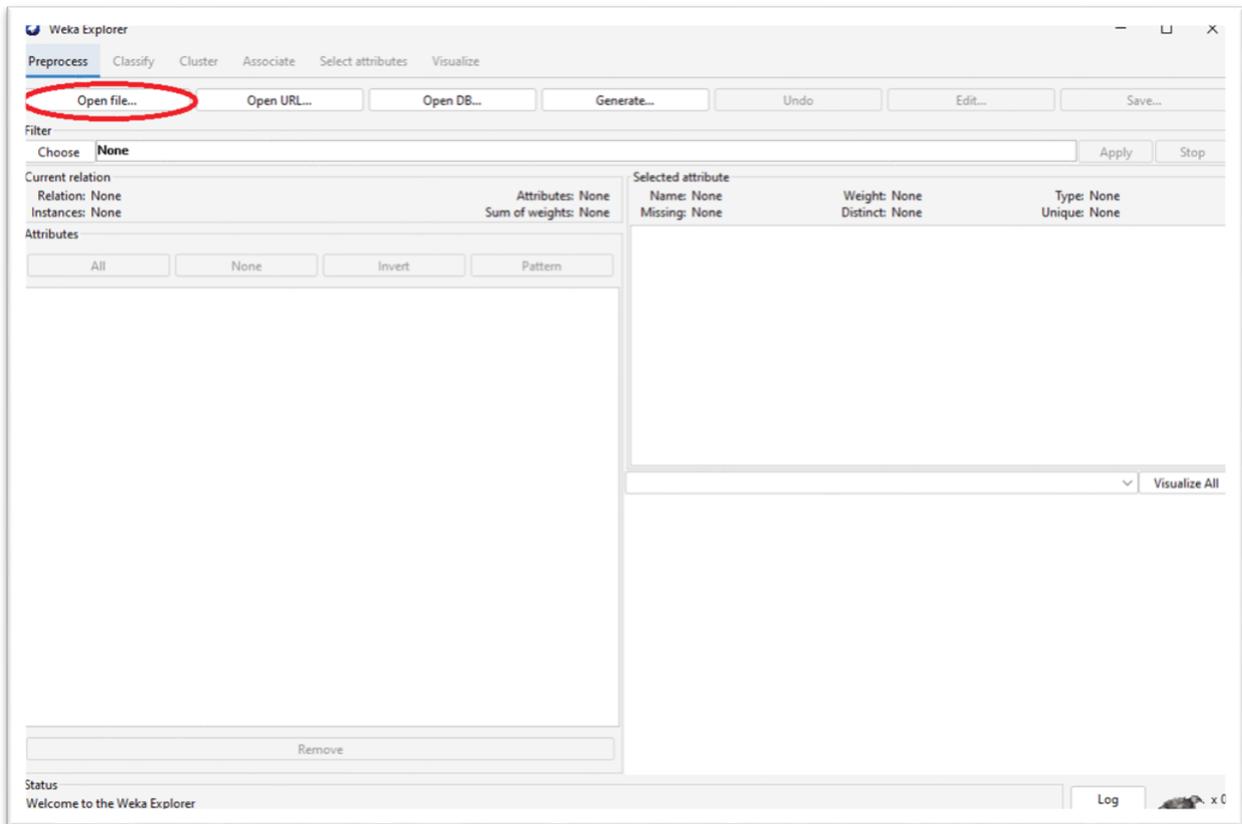
Figura 11 – Tela inicial do Weka 3.8.6 – Seleção do “Explorer”



Fonte: Weka 3.8.6

Ao selecionar o botão “Open File” é possível realizar a busca por arquivos compatíveis com o Weka no computador. Conforme mencionado na etapa anterior, o arquivo padrão do Weka é a extensão Arff.

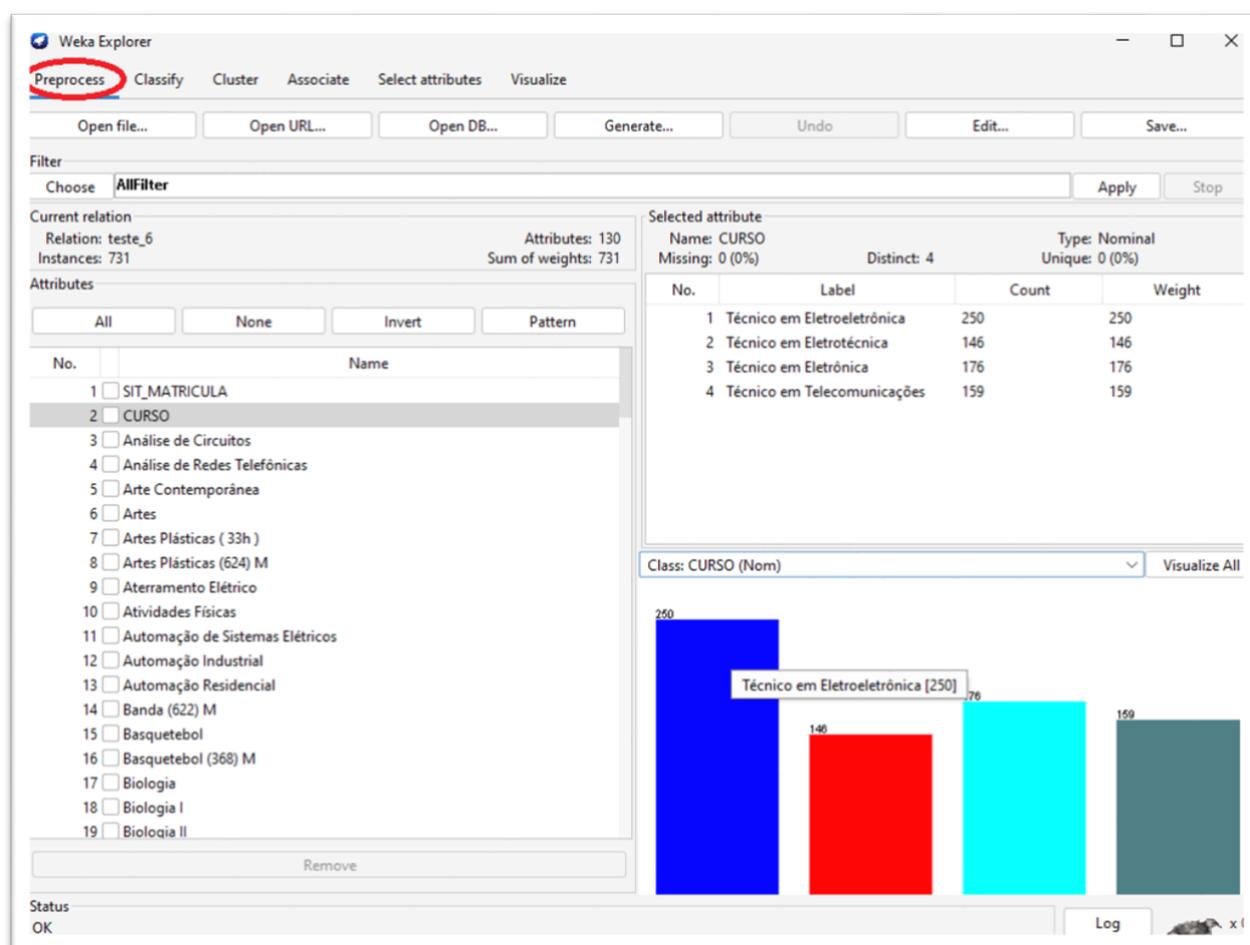
**Figura 12** – Tela principal do Weka 3.8.6 – Seleção do “Open File”



Fonte: Weka 3.8.6

Após a escolha do arquivo a ser utilizado, o usuário deve selecionar a aba “Preprocess”, na qual são descritas as informações do conjunto de dados em tabelas e gráficos. A tabela apresenta o nome do dado, tipo de dado, quantidade de valores distintos e percentual de valores ausentes. No gráfico existe a possibilidade de cruzamento de informações de diferentes variáveis. Essas variáveis são chamadas de “*attributes*” ou atributos pelo Weka.

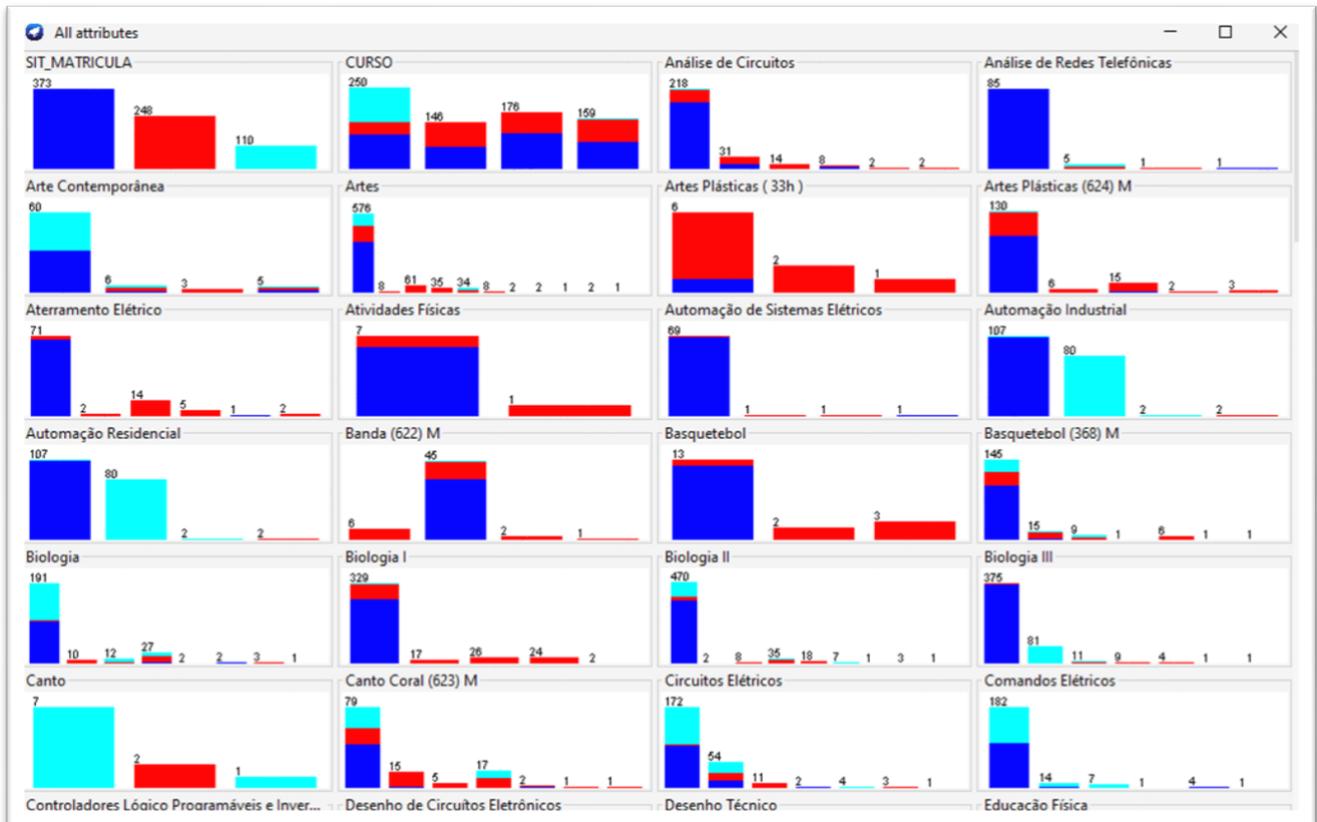
**Figura 13** – Tela principal do Weka 3.8.6 – Seleção do “Preprocess”



Fonte: Weka 3.8.6

A figura 14 demonstra um exemplo do painel de gráficos com todos os atributos.

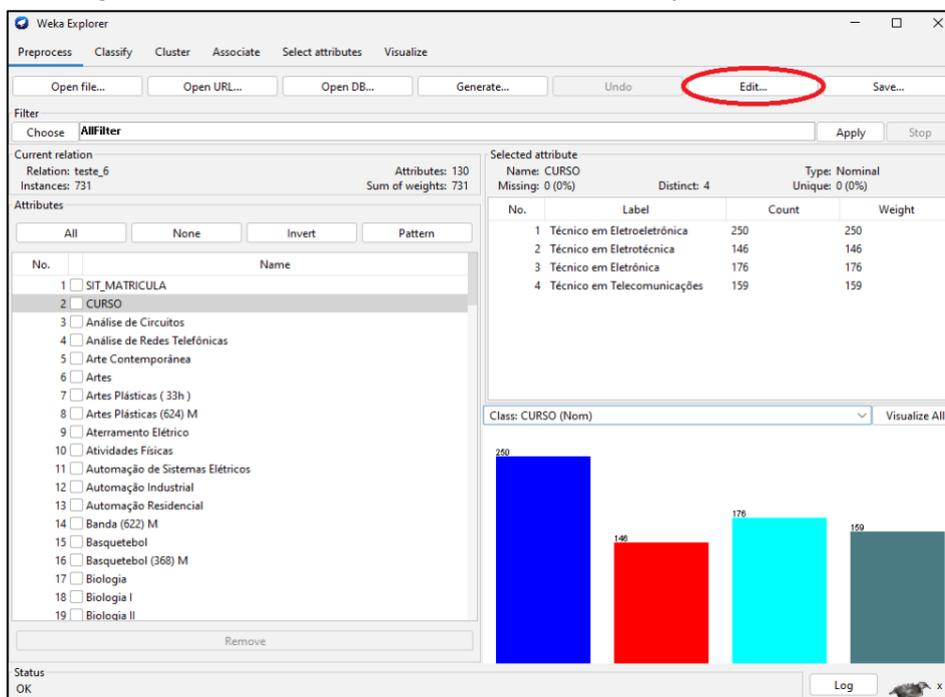
**Figura 14** – Painel de gráficos com todos os atributos do Weka 3.8.6



Fonte: Weka 3.8.6

Ao clicar no botão “Edit” o usuário visualiza o conjunto de dados.

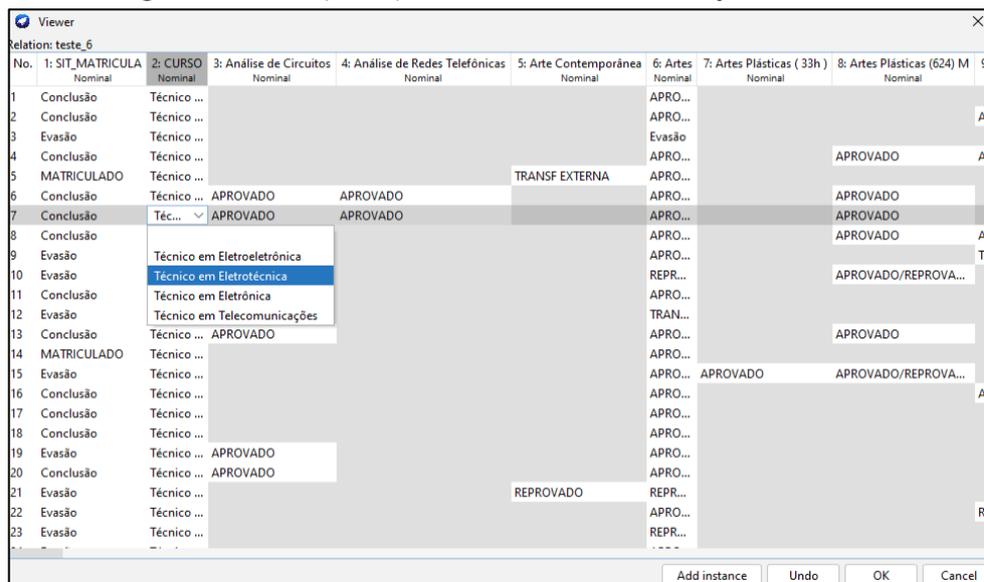
**Figura 15** – Tela principal do Weka 3.8.6 – Seleção do “Edit”



Fonte: Weka 3.8.6

A tela “Edit” apresenta o conjunto de dados completo e possibilita a seleção e edição, conforme a figura 16.

**Figura 16** – Tela principal do Weka 3.8.6 – Seleção do “Edit”



Fonte: Weka 3.8.6

A utilização das Regras de Associação possui como objetivo encontrar relacionamentos ou padrões frequentes em um conjunto de dados. De acordo com Pasta (2011), a tarefa de associação identifica e descreve as associações entre as variáveis no mesmo item ou associações entre os itens diferentes que ocorrem concomitantemente, de uma forma frequente em bases de dados.

Uma limitação encontrada em regras de associação é a quantidade de regras de associação. Considerando que ao utilizar muitas informações de todos os atributos, o *software* pode gerar muitas regras de associação desconectadas com a realidade e, portanto, irrelevantes para a pesquisa. Bürkle (2006) esclarece que a extração de regras de associação pode produzir centenas e até milhares de regras dificultando a sua avaliação pelo usuário. Além disso, grande parte dos resultados produzidos contém regras óbvias e redundantes.

Desta forma, opta-se por restringir a quantidade de atributos ou de informações destes, deixando os mais relevantes de acordo com o elencado pela literatura científica. Schonhorst (2010) também alertou para o fato de que o número de regras geradas depende da quantidade de atributos considerados e do suporte e confiança mínimos especificados e, na maior parte dos casos, este número torna inviável a observação de todas as regras geradas para a obtenção de conhecimento que auxilie no processo de tomada de decisão.

O processo de restrição das informações dos atributos é realizado por meio do botão “Choose” e, também é utilizado para segmentar a geração de regras de associação.

**Figura 17** – Tela principal do Weka 3.8.6 – Seleção do “Choose”

The screenshot shows the Weka Explorer interface with the 'Filter' dialog box open. The 'Choose' button is highlighted with a red circle. The dialog displays the 'Selected attribute' 'CURSO' with a table of its values and a bar chart below it.

No.	Label	Count	Weight
1	Técnico em Eletroeletrônica	250	250
2	Técnico em Eletrotécnica	146	146
3	Técnico em Eletrônica	176	176
4	Técnico em Telecomunicações	159	159

Class: CURSO (Nom) Visualize All

Bar chart showing the distribution of the 'CURSO' attribute:

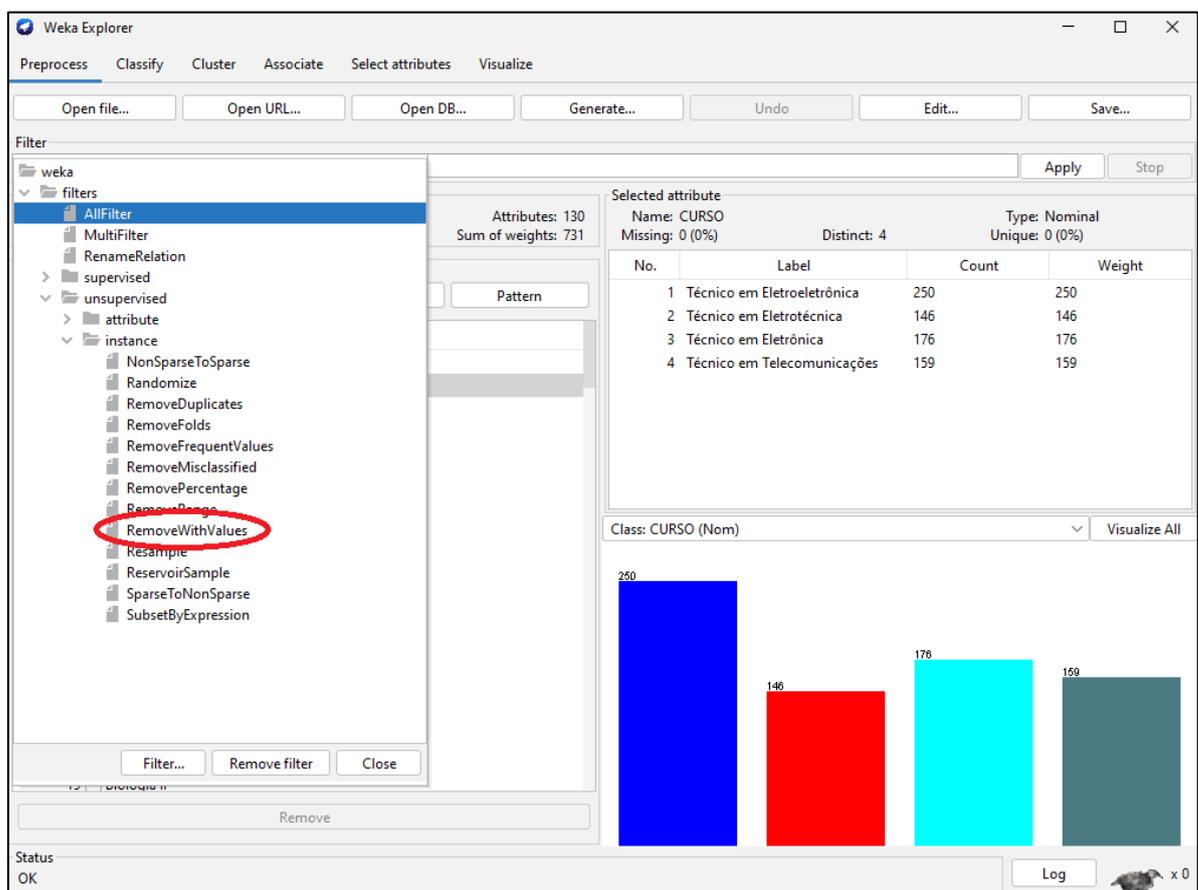
- Label 1: 250
- Label 2: 146
- Label 3: 176
- Label 4: 159

Fonte: Weka 3.8.6

O filtro “*RemoveWithValues*” é utilizado para suprimir do conjunto de dados as informações dos atributos menos importantes para a pesquisa.

Durante a etapa de mineração dos dados pessoais e socioeconômicos no Weka foram descartados os atributos que não contribuíam para gerar as regras de associação, por exemplo, número de matrícula; nome do aluno codificado; tipo sanguíneo, entre outros, tendo em vista que esses não interferem no resultado pretendido.

**Figura 18** – Tela principal do Weka 3.8.6 – Seleção do filtro “*RemoveWithValues*”



Fonte: Weka 3.8.6

Neste exemplo, estão sendo suprimidas as informações 2,3 e 4 (EMI Técnico em Eletrotécnica, EMI Técnico em

Eletrônica e EMI Técnico em Telecomunicações) do atributo 2 (CURSO).

Figura 19 – Tela principal do Weka 3.8.6 – Informações suprimidas

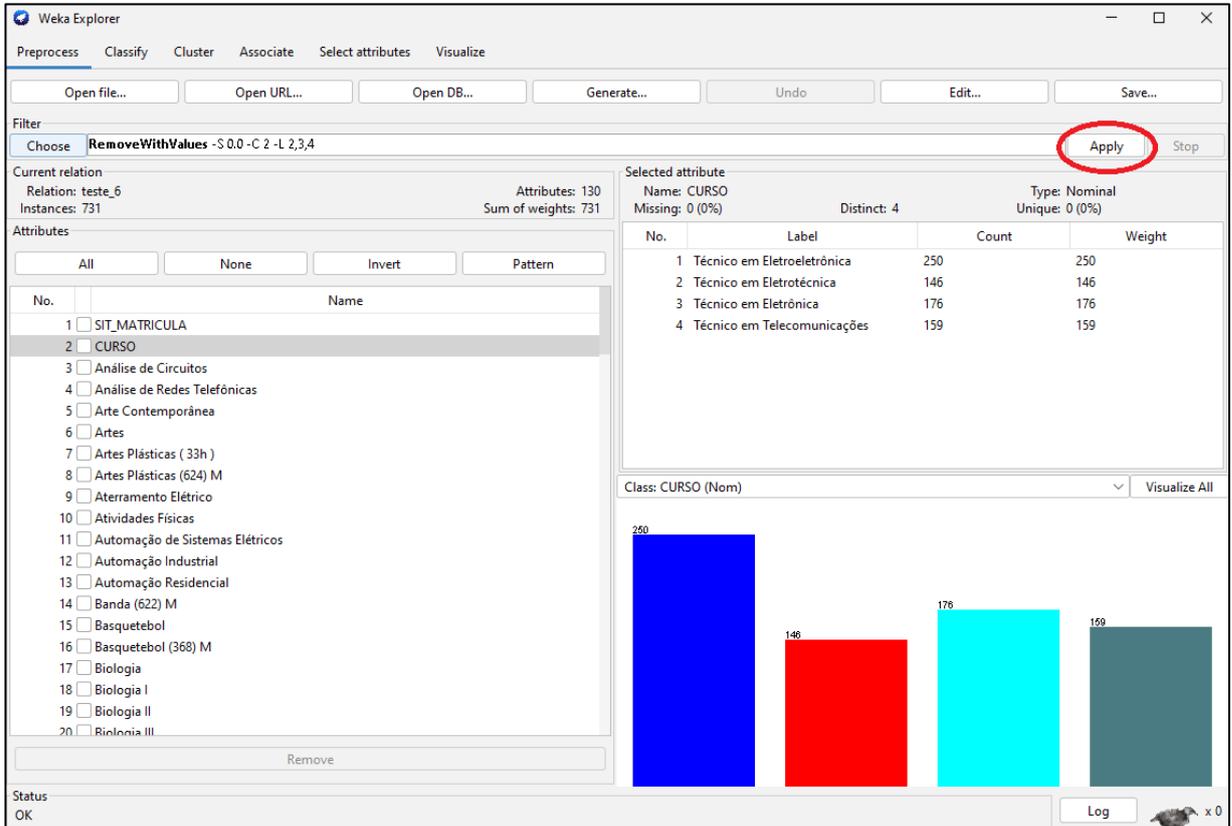
The screenshot shows the Weka Explorer interface. A dialog box titled 'weka.gui.GenericObjectEditor' is open, showing the configuration for the 'RemoveWithValues' filter. The 'attributeIndex' is set to 2 and 'nominalIndices' is set to 2,3,4. In the background, a bar chart displays the distribution of the 'CURSO' attribute, with four bars representing the categories: Técnico em Eletroeletrônica (250), Técnico em Eletrotécnica (146), Técnico em Eletrônica (176), and Técnico em Telecomunicações (159).

No.	Label	Count	Weight
1	Técnico em Eletroeletrônica	250	250
2	Técnico em Eletrotécnica	146	146
3	Técnico em Eletrônica	176	176
4	Técnico em Telecomunicações	159	159

Fonte: Weka 3.8.6

Após selecionados os atributos, é necessário aplicar a formatação por meio do botão “Apply”.

**Figura 20** – Tela principal do Weka 3.8.6 – Seleção do “Apply”



Fonte: Weka 3.8.6

Na figura 21, verifica-se a nova apresentação dos dados após supressão das informações dos atributos.

**Figura 21** – Tela principal do Weka 3.8.6 – Nova apresentação dos dados após supressão

The screenshot shows the Weka Explorer interface with the 'RemoveWithValues' filter applied. The 'Current relation' is 'teste\_6-weka.filters.unsupervised.instance.RemoveWithValues-S0...'. The 'Attributes' list on the left includes 'CURSO' (selected). The 'Selected attribute' table shows the distribution of the 'CURSO' attribute values.

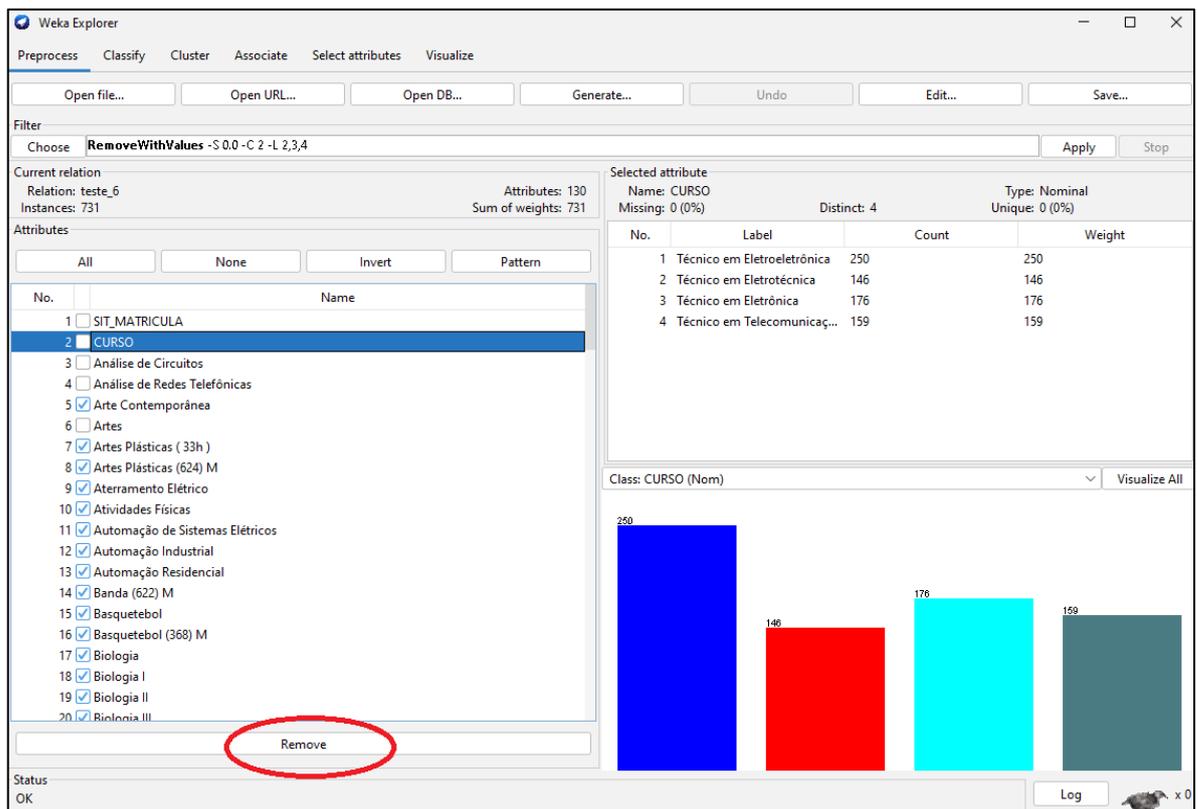
No.	Label	Count	Weight
1	Técnico em Eletroeletrônica	250	250
2	Técnico em Eletrotécnica	0	0
3	Técnico em Eletrônica	0	0
4	Técnico em Telecomunicaç...	0	0

The bar chart below the table shows a single blue bar representing the count of 250 for the 'CURSO' attribute. The status bar at the bottom indicates 'OK' and 'Log'.

Fonte: Weka 3.8.6

Outra forma de restringir as informações é remover o atributo por completo. Para isso, seleciona-se os atributos a serem removidos e clica-se no botão “Remove”.

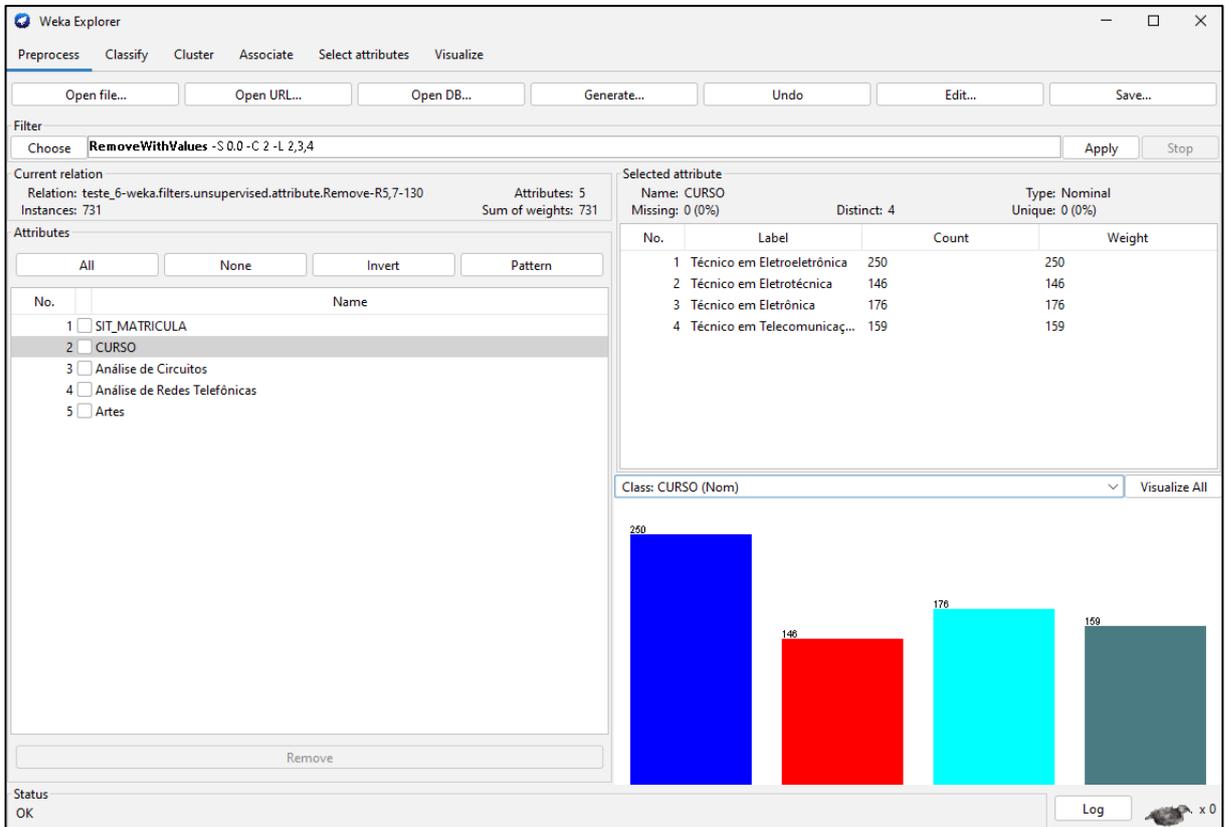
**Figura 22** – Tela principal do Weka 3.8.6 – Seleção do “Remove”



Fonte: Weka 3.8.6

Tela de apresentação dos dados após remoção de atributos.

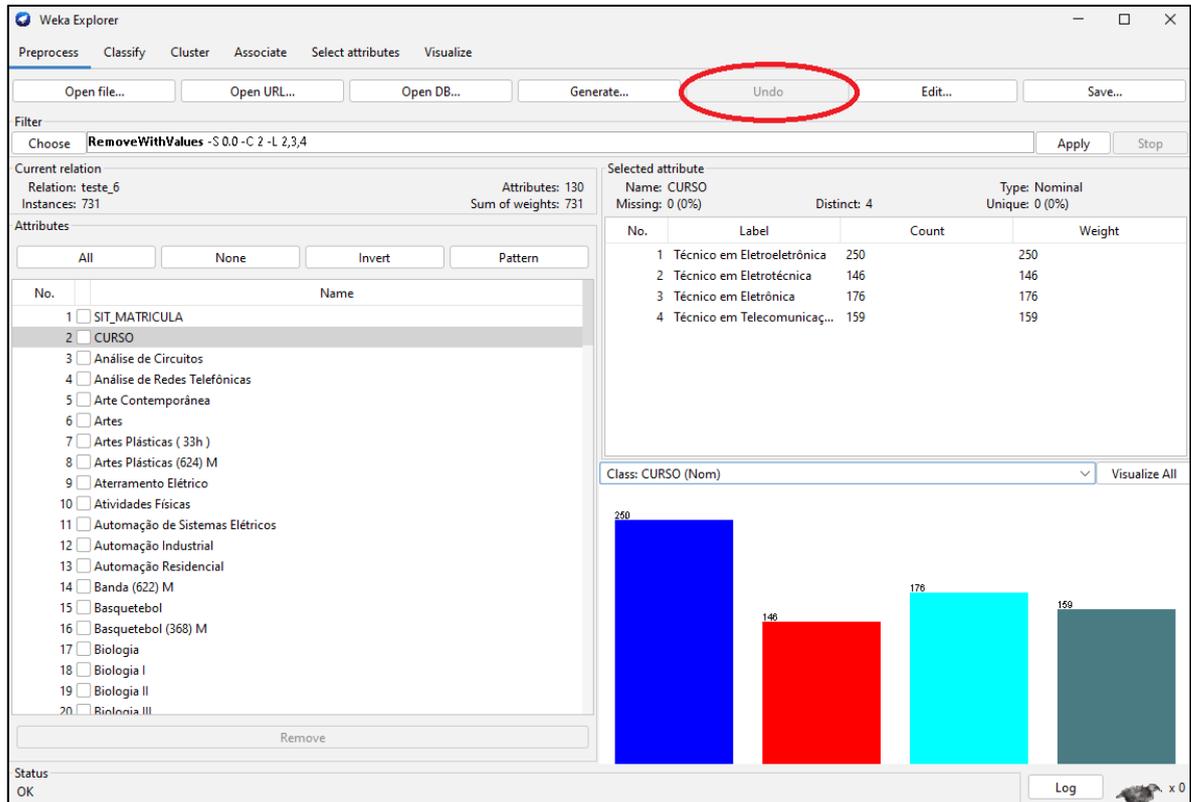
**Figura 23** – Tela principal do Weka 3.8.6 – Nova apresentação dos dados após supressão (2ª opção)



Fonte: Weka 3.8.6

O botão “Undo” é utilizado para desfazer qualquer ação realizada nessa etapa de processamento dos dados.

**Figura 24** – Tela principal do Weka 3.8.6 – Seleção do “Undo”



Fonte: Weka 3.8.6

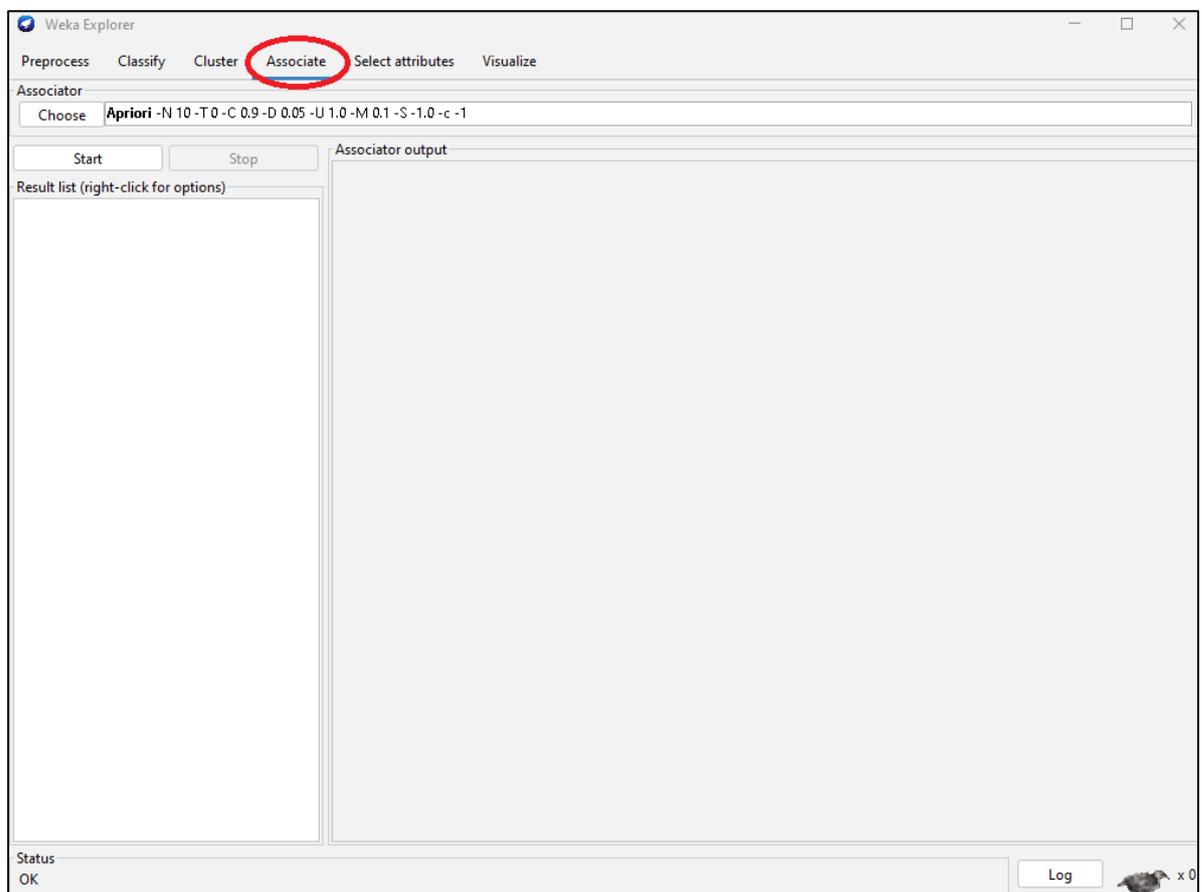
Na aba “Associate” é onde serão executados os comandos dos dados previamente processados. Para o objetivo desta pesquisa foi escolhido o algoritmo Apriori.

O algoritmo Apriori foi desenvolvido por Agrawal et al (1993) e é considerado o algoritmo mais popular e clássico para gerar as regras de associação. Segundo Silva et al. (2010), trata-se de um algoritmo que realiza buscas recursivas no conjunto de dados à procura de conjuntos frequentes e que atendem a um suporte mínimo especificado.

O passo inicial para aplicação do algoritmo Apriori consiste em encontrar conjunto de itens frequentes. Desta forma, o usuário especifica um suporte mínimo, que

corresponde com o conjunto de itens que ocorrem em pelo menos uma dada porcentagem e o algoritmo seleciona as transações acima desse suporte pré-estabelecido. Em seguida, são construídas as regras dos itens selecionados anteriormente, e o Apriori calcula a confiabilidade de cada regra e mantém apenas aquelas em que a confiança é superior ao limite definido pelo usuário (PASTA, 2011).

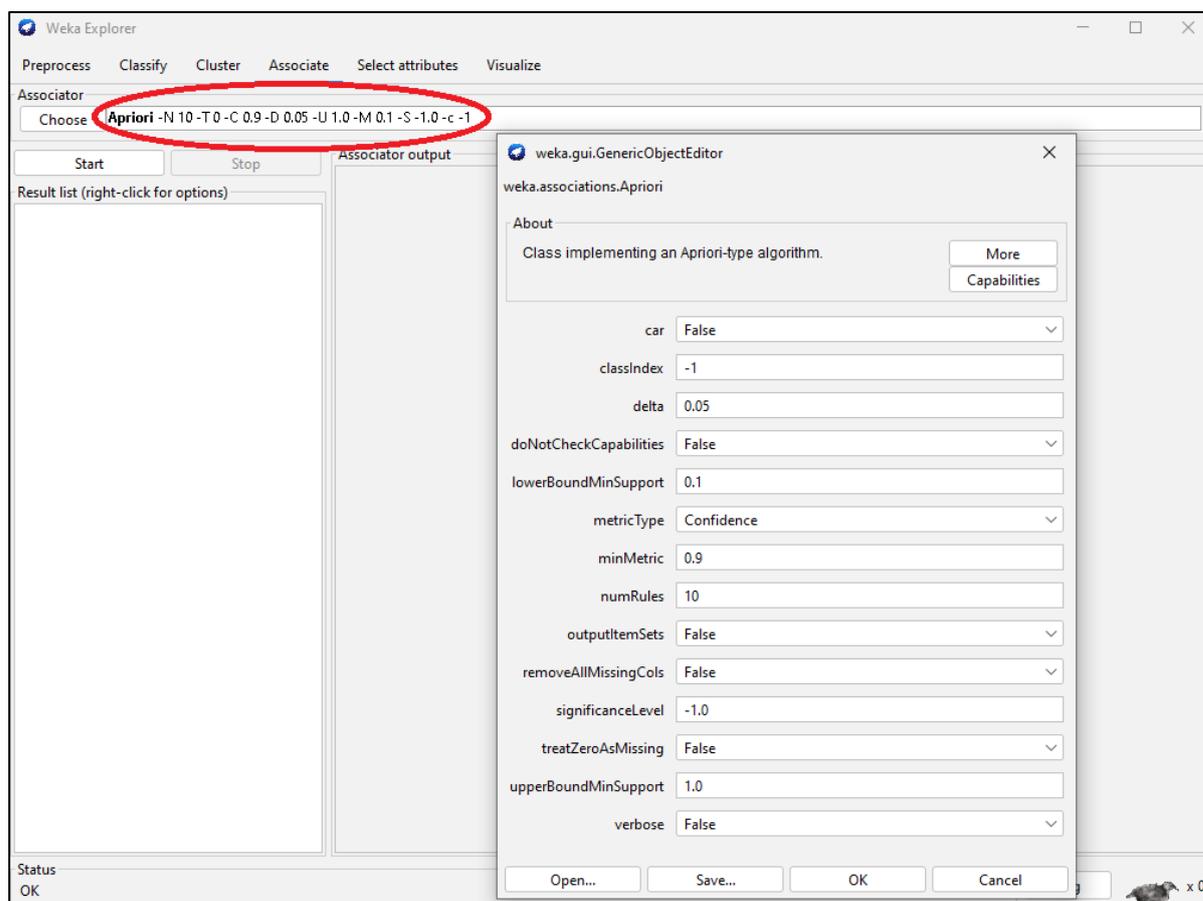
**Figura 25** – Tela principal do Weka 3.8.6 – Seleção do “Associate”



Fonte: Weka 3.8.6

Clicando na janela de formatação abre-se uma caixa na qual é possível informar alguns parâmetros para elaboração das regras.

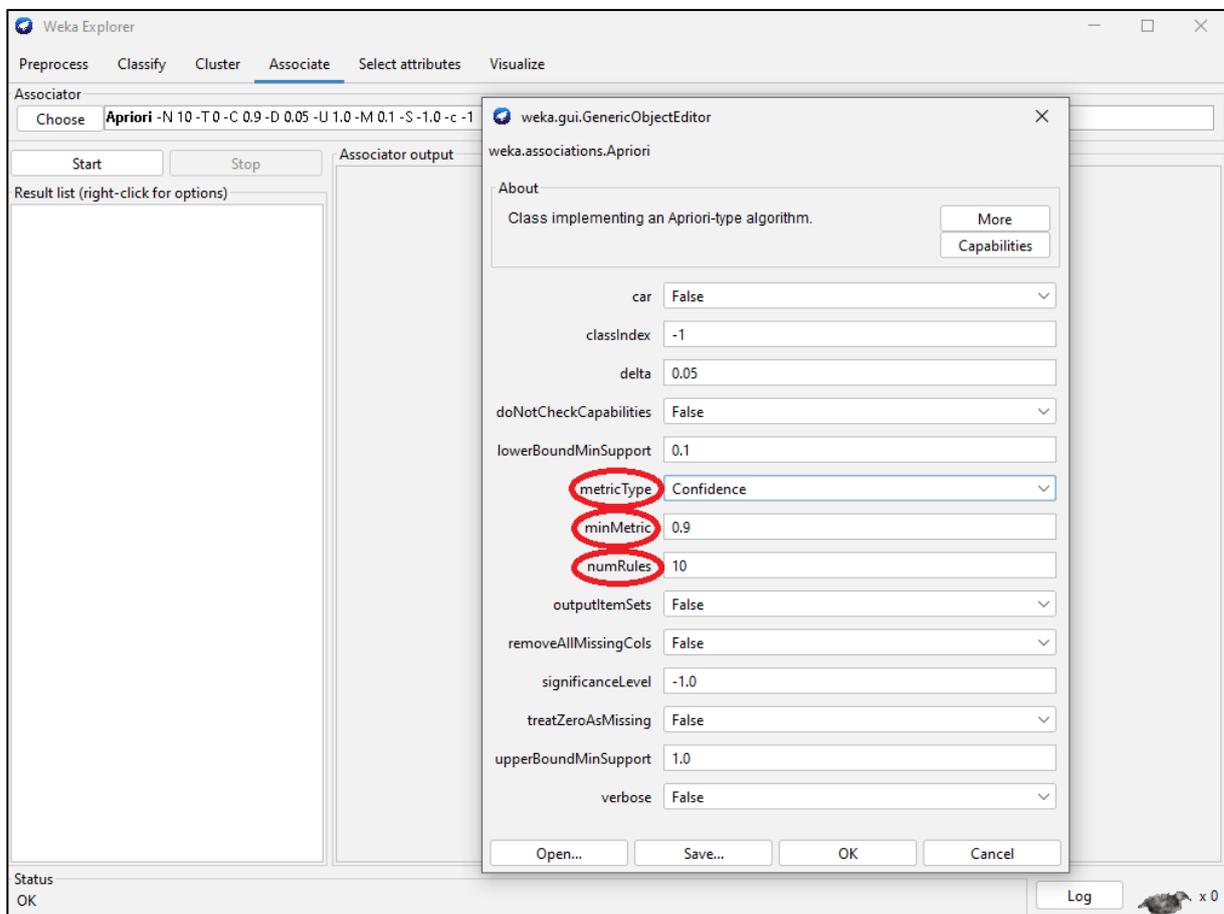
**Figura 26** – Tela principal do Weka 3.8.6 – Seleção do “Apriori”



Fonte: Weka 3.8.6

No algoritmo Apriori, as regras de associação geradas devem atender a um suporte e confiança mínimos especificados pelo decisor. Alguns parâmetros significativos são apresentados: O Tipo de Métrica (*MetricType*); Número de regras (*numRules*), indica a quantidade de regras desejada (default = 12); Confiança (*minMetric*), corresponde a precisão mínima exigida (default = 0,9 = 90%); Suporte Mínimo (*lowerboundminsupport*), especifica o suporte mínimo desejado (default = 0,1 = 10%). Então, observa-se que quanto mais próximo de 1,0, ou 100%, maior será a confiabilidade dos dados, isto é, maior será a confiança e o suporte da regra.

**Figura 27** – Tela principal do Weka 3.8.6 – Seleção das métricas



Fonte: Weka 3.8.6

Os critérios mais utilizados na definição do *metricType* são os níveis de confiança (*Confidence*), *lift* e suporte.

Romão *et al.* (1999) destacam que o suporte corresponde à frequência com que ocorrem os padrões em toda a base, isto é, a porcentagem dos pesquisadores da base que possuem X e Y, já a confiança é uma medida da força das regras, a porcentagem de pesquisadores de X que possuem Y. Corroborando com Zhang e Zhang (2002), o suporte representa a frequência dos padrões e a confiança a força da implicação, por exemplo, em pelo menos 50% das vezes que o antecedente ocorrer nas transações, o consequente também ocorrerá.

Bürkle (2006) explica que valores de confiança altos são importantes para garantir uma elevada coesão entre os itens

analisados, pois uma confiança baixa não apresenta qualquer padrão de comportamento e valores de confiança acima de 50% também refletem no número de regras produzidas. O suporte mínimo é exigido de forma a garantir a significância estatística dos padrões gerados, além disso restringe a quantidade de regras na saída do algoritmo de mineração.

A medida de interesse *lift*, também denominada de *interest* é uma das mais utilizadas para avaliar dependências. Em uma regra de associação  $A \rightarrow B$ , o *lift* indicará o quão frequente torna-se B quando o A ocorre. Por exemplo, se  $lift(A \rightarrow B) = 4$ , significa que B tem 4 vezes mais chances de ocorrer quando A ocorre. Esta medida varia entre 0 e  $\infty$  e pode-se dizer que quanto maior o valor do lift, mais interessante é a regra (SCHONHORST, 2010).

O valor do lift é calculado pela seguinte fórmula:

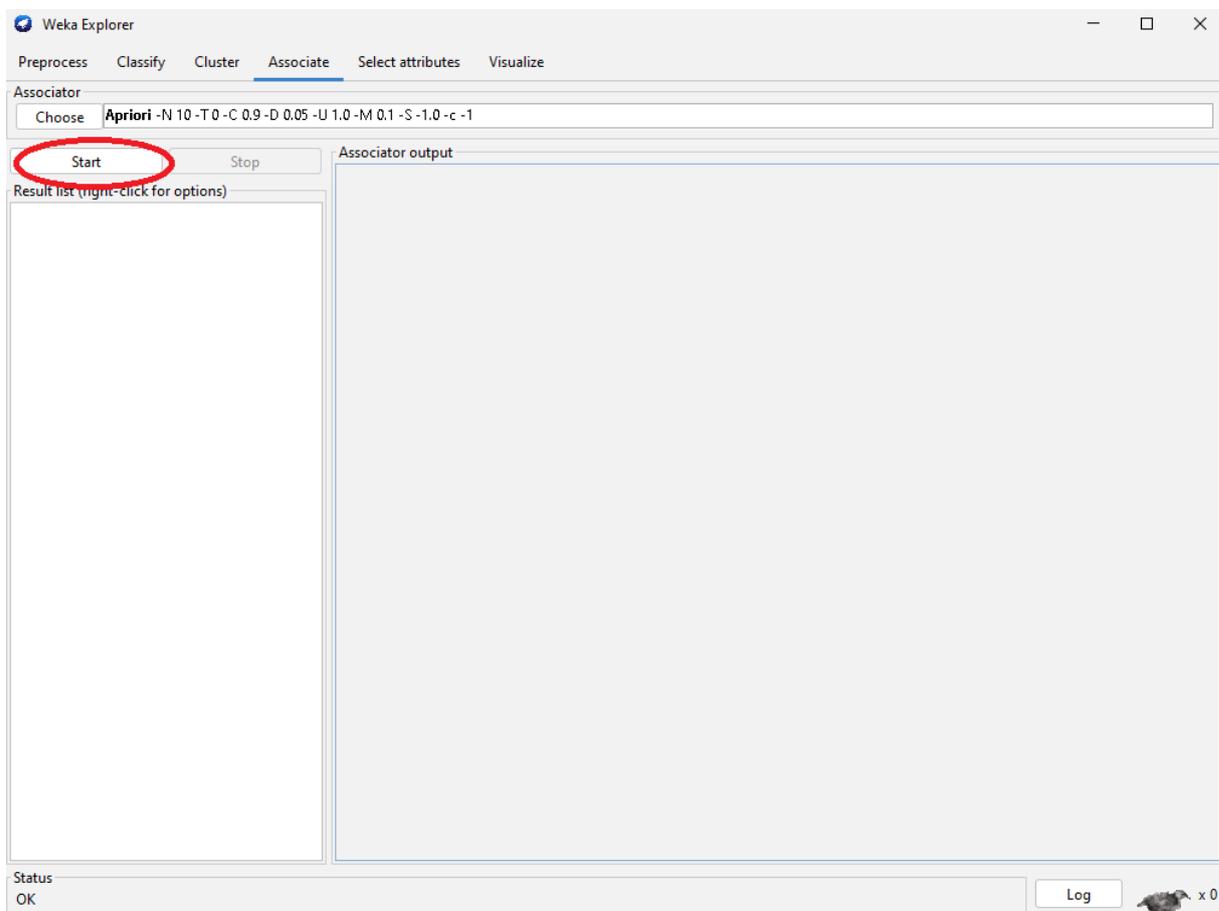
$$lift(A \rightarrow B) = \frac{\text{confiança}(A \rightarrow B)}{\text{suporte}(B)}$$

Se  $lift(A \rightarrow B) = 1$ , então A e B são independentes, porém se  $lift(A \rightarrow B) > 1$ , então A e B são positivamente dependentes e caso  $lift(A \rightarrow B) < 1$ , A e B são negativamente dependentes.

Depois de definida a métrica a ser utilizada deve-se ajustar o valor mínimo que o software deverá retornar com as regras desse critério, e definir quantas regras serão necessárias.

Depois de definidos os parâmetros, clicando no botão "Start" o software irá processar os dados e retornar o número de regras solicitadas, desde que seja possível existir essa quantidade, com base no valor mínimo do parâmetro que foi escolhido.

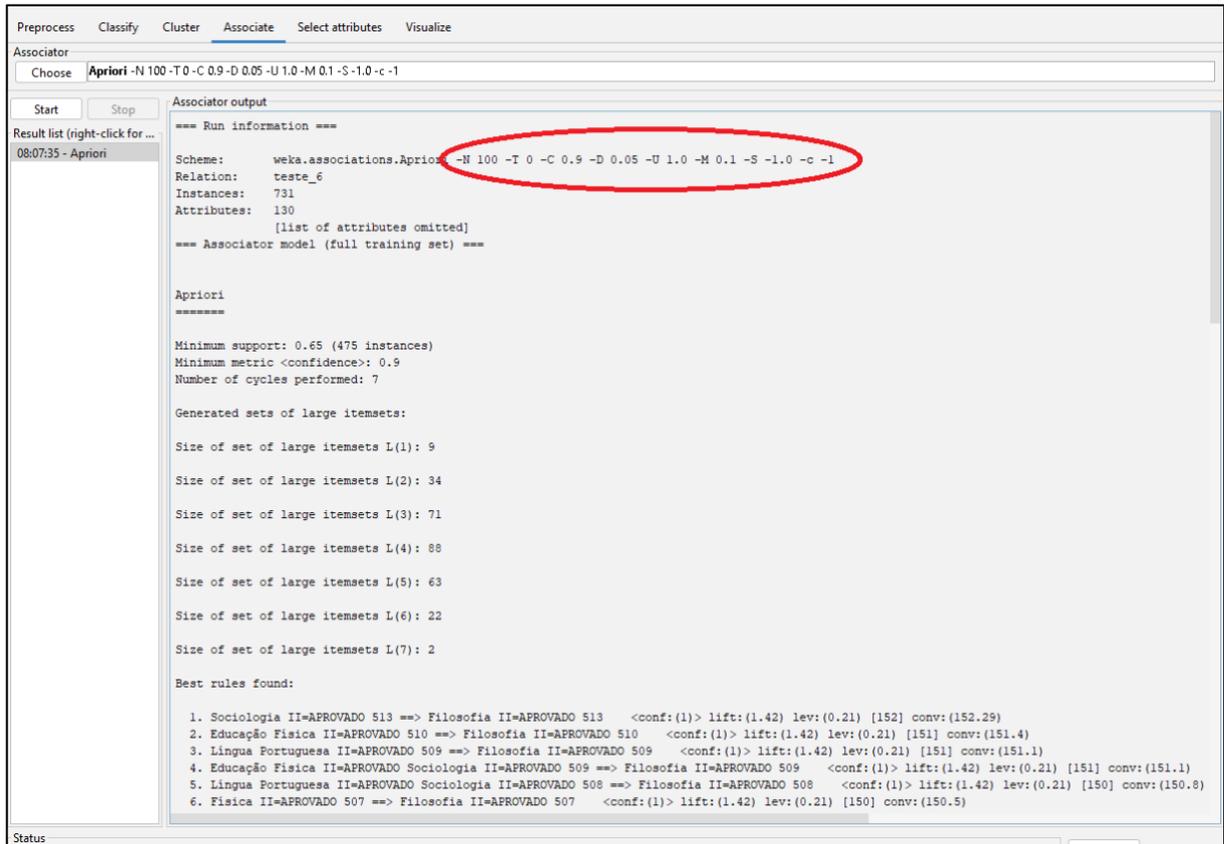
**Figura 28** – Tela principal do Weka 3.8.6 – Seleção do “Start”



Fonte: Weka 3.8.6

Na janela de saída do Weka são apresentadas algumas informações relevantes. Em “Scheme” são evidenciados os parâmetros utilizados como número de regras, nível de confiança ou de *lift*.

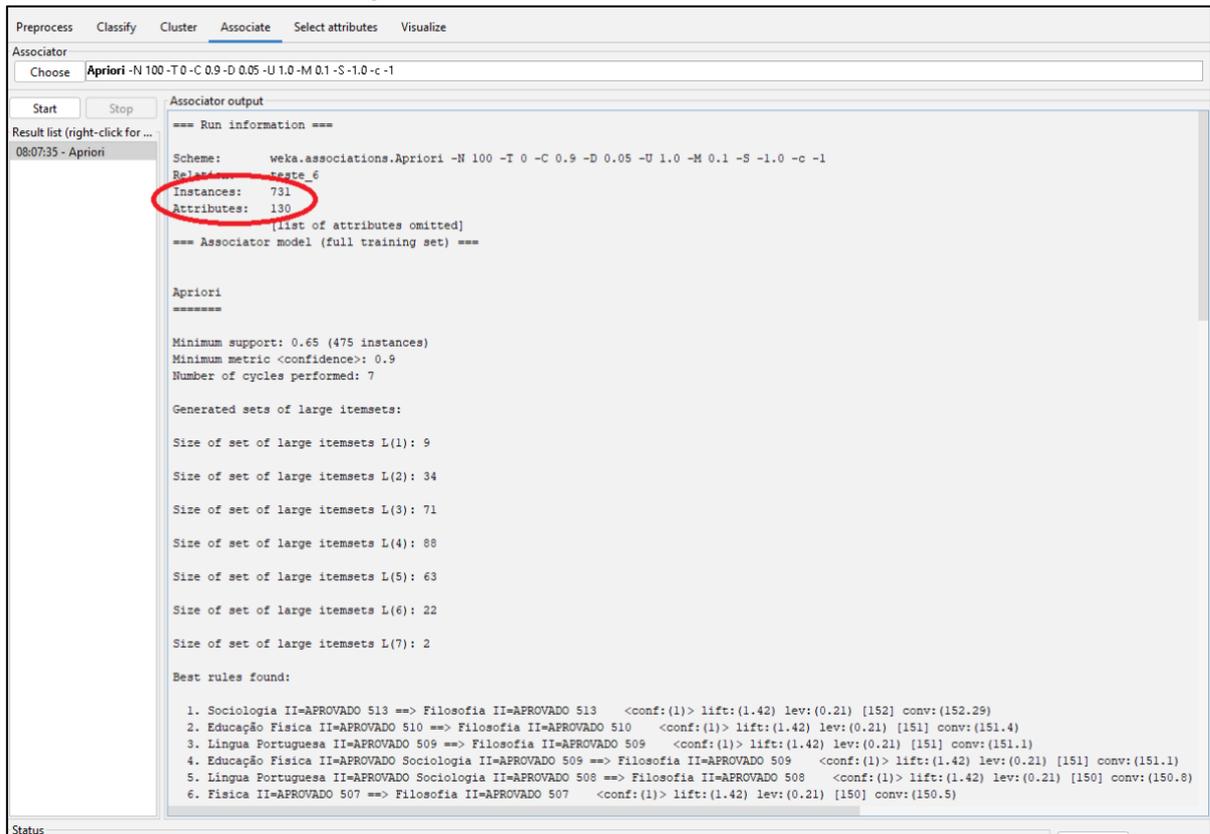
**Figura 29** – Tela do Weka 3.8.6 – Informação sobre o “Scheme”



Fonte: Weka 3.8.6

“Instances” são o número de ocorrências ou dos sujeitos da pesquisa analisados (linhas) e “Attributes” a quantidade de variáveis ou atributos executados no processo (colunas).

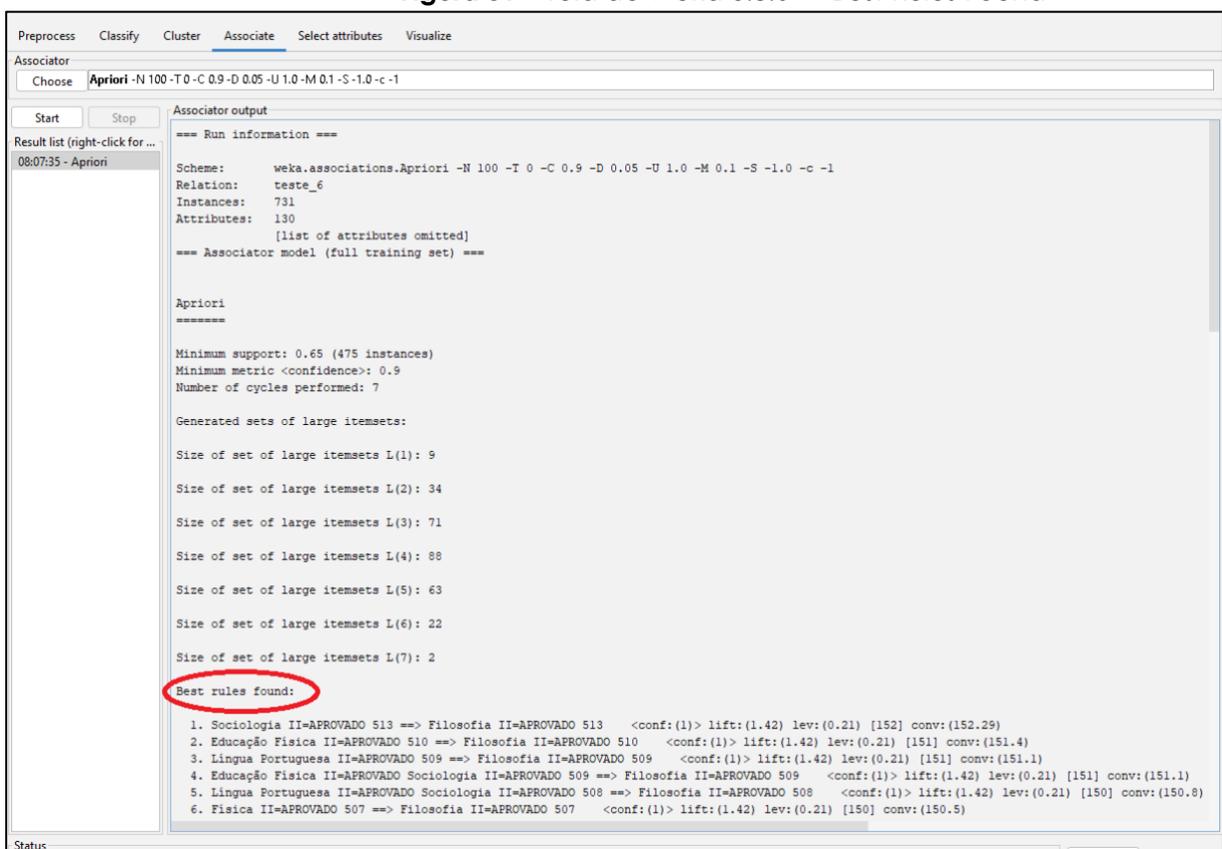
Figura 30 – Tela do Weka 3.8.6 – “Instances” e “Attributes”



Fonte: Weka 3.8.6

Em “Best Rules Found” são elencadas as melhores regras encontradas com base nos parâmetros que foram definidos. O Weka apresenta todas as regras organizadas de forma decrescente em relação ao parâmetro determinado, cabendo ao pesquisador identificar e extrair dessa lista de regras aquelas que melhor atendem seu objetivo de pesquisa.

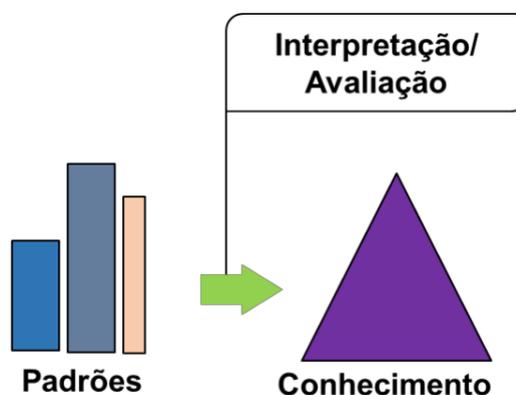
Figura 31 – Tela do Weka 3.8.6 – “Best Rules Found”



Fonte: Weka 3.8.6

### 3.1.5 Interpretação e avaliação dos resultados

**Figura 32** – Etapa de “Interpretação e avaliação dos resultados”



Fonte: Elaborada pela pesquisadora, com o uso do software Power Point (2022)

Para melhor entendimento, a interpretação das regras de associação deve ser realizada da seguinte forma:

**Pão→Leite; suporte = 20%; confiança = 50%. Isto significa que, em 20% de todas as compras realizadas, pão e leite foram comprados juntos; Ao considerar apenas as compras do item pão, em 50% destas também foram comprados leite.**

Na tabela a seguir serão apresentadas as principais regras obtidas como resultado da Mineração dos dados pessoais e socioeconômicos dos cursos de Ensino Médio Integrado ao ensino técnico do DAEE/CBA-OJS/IFMT, durante o período de 2009 a 2019.

Tabela 1 – Principais regras de associação do conjunto de dados pessoais e socioeconômicos dos alunos evadidos e concluintes dos cursos de EMI Técnico do DAEE/CBA-OJS/IFMT, no período de 2009 a 2019

REGRA (ÍNDICE)	CATEGORIA 1	→	CATEGORIA 2	Suporte %	Confiança %	Lift
1	Situação de matrícula = Evasão 299	→	Tipo de escola cursada anteriormente = Pública 198	0.27	0.66	1.09
2	Tipo de escola cursada anteriormente = Particular 289	→	Situação de matrícula = Concluído 188	0.26	0.65	1.1
3	Situação de matrícula = Evasão 299	→	Meio de transporte = Ônibus 182	0.25	0.61	1.03

Fonte: Elaborada pela pesquisadora a partir de dados obtidos por meio dos sistemas Qacadêmico e Qseleção do IFMT (2021)

A seguir será apresentada a interpretação das regras geradas de acordo com o índice exposto na tabela 1.

Em 27% de todas as ocorrências, aparecem em conjunto as seguintes situações: Evasão do curso e aluno que estudou anteriormente em escola pública. Considerando apenas os alunos que evadiram do curso, 66% desses estudaram anteriormente em escolas públicas.

Em 26% de todas as ocorrências, aparecem em conjunto as seguintes situações: Aluno que estudou anteriormente em escola particular e conclusão do curso. Considerando apenas os que estudaram anteriormente em escola particular, 65% concluíram o curso.

Em 25% de todas as ocorrências, aparecem em conjunto as seguintes situações: Evasão do curso e ônibus como meio de transporte. Considerando apenas os alunos que evadiram do curso, 61% deles utilizam o ônibus como meio de transporte.

Na análise dos dados pessoais e socioeconômicos foram identificados os elementos que contribuem para a detecção de padrões de alunos com risco de evasão escolar como os resultados referentes a categoria “Tipo de escola cursada

anteriormente", na qual verificou-se uma associação entre alunos que estudaram anteriormente em escola pública e evasão do curso e os alunos que estudaram anteriormente em escola particular e conclusão do curso. A categoria "Meio de Transporte" também apresentou associação com a evasão do curso visto que em 25% de todas as ocorrências, os alunos que utilizam o ônibus como meio de transporte evadem do curso e ao analisar apenas os alunos que evadiram do curso, 61% deles utilizam o ônibus como meio de transporte.

Verificou-se uma correlação dos resultados desta pesquisa com a literatura. Em seus estudos, Resende (2012) aponta quase todos os alunos evadidos em sua amostra eram originários de escolas públicas estaduais e afirmaram que o Ensino em um Instituto Federal era muito difícil e não estavam acostumados a estudar. Sendo assim, devido a um excesso de disciplinas em cursos de Ensino Médio Integrado ao ensino técnico foi relatado uma frequente sensação de fracasso e desânimo nos estudantes que preferem voltar para a escola de origem por terem maior chance de recuperação e aprovação.

A Mineração de Regras de Associação dos dados acadêmicos foi realizada de forma separada por curso e por ano. No curso de EMI Técnico em Eletrônica foram analisadas 176 ocorrências de alunos evadidos e concluintes durante o período de 2009 a 2019, entretanto o *software* gerou regras referentes apenas ao primeiro ano de curso, conforme tabela 2.

Tabela 2 – Principais regras de associação do conjunto de dados acadêmicos dos alunos evadidos e concluintes do curso de EMI Técnico em Eletrônica do DAEE/CBA-OJS/IFMT, no período de 2009 a 2019

REGRA (ÍNDICE)	CATEGORIA 1	→	CATEGORIA 2	Suporte %	Confiança %	Lift
1 – 1º ano	Artes = Reprovado; Eletricidade Básica = Reprovado; Filosofia I = Reprovado 18	→	Situação de matrícula = Evasão; Física I = Reprovado; Matemática I = Reprovado 18	0.10	1	9.26
2 – 1º ano	Situação de matrícula = Evasão; Artes = Reprovado; Eletricidade Básica = Reprovado; Filosofia I = Reprovado 18	→	Física I = Reprovado; Matemática I = Reprovado 18	0.10	1	9.26
3 – 1º ano	Artes = Reprovado; Física I = Reprovado 18	→	Situação de Matrícula = Evasão; Eletricidade Básica = Reprovado; Filosofia I = Reprovado; Matemática I = Reprovado 18	0.10	1	9.26

Fonte: Elaborada pela pesquisadora a partir de dados obtidos por meio dos sistemas Qacadêmico e Qseleção do IFMT (2021)

A seguir será apresentada a interpretação das regras geradas de acordo com o índice exposto na tabela 2.

1º ano: Em 10% de todas as ocorrências, aparecem em conjunto as seguintes situações: Aluno reprovado em Artes, Eletricidade Básica, Filosofia I, Física I e Matemática I e Evasão de curso. Um aluno que reprova em Artes, Eletricidade Básica e Filosofia I possui 9,26 chances a mais de reprovar em Física I, Matemática I e evadir do curso.

1º ano: Em 10% de todas as ocorrências, aparecem em conjunto as seguintes situações: Reprovação em Artes, Eletricidade Básica, Filosofia I, Física I e Matemática I e Evasão do curso. Um aluno que evade do curso e reprova em Artes, Eletricidade Básica, Filosofia I possui 9,26 chances a mais de reprovar em Física I e Matemática I.

1º ano: Em 10% de todas as ocorrências, aparecem em conjunto as seguintes situações: Reprovação em Artes, Física I, Eletricidade Básica, Filosofia I, Matemática I e Evasão do curso. Um aluno que reprova em Artes e Física I possui 9,26 chances a mais de reprovar em Eletricidade Básica, Filosofia I, Matemática I e evadir do curso.

As regras geradas no curso de Técnico em Eletrônica apresentaram uma associação entre a reprovação em disciplinas iniciais e a evasão do curso. A partir da predominância de regras relacionadas à evasão do curso no primeiro ano, foi possível concluir que a reprovação em disciplinas iniciais impacta na evasão do aluno.

Garcia, Azevedo e Sobrinho (2019) analisaram o fracasso escolar a partir das taxas de reprovação, de distorção idade-série e de abandono escolar dos alunos do Ensino Médio das escolas estaduais dos 10 estados brasileiros durante o período de 2010 a 2016. Os resultados quantitativos apresentaram uma ampliação na reprovação, principalmente no primeiro ano, de 2010 para 2016, o que aumentou o índice de fracasso escolar. Já no terceiro ano verificou-se uma redução nas taxas de reprovação e abandono.

A seguir serão apresentadas as principais regras de associação do curso de EMI Técnico em Eletrotécnica, de acordo com a tabela abaixo. No curso de EMI Técnico em Eletrotécnica foram analisadas 146 ocorrências de alunos evadidos e concluintes durante o período de 2009 a 2019 e foram geradas regras referentes aos quatro anos de curso.

Tabela 3 – Principais regras de associação do conjunto de dados acadêmicos dos alunos evadidos e concluintes do curso de EMI Técnico em Eletrônica do DAEE/CBA-OJS/IFMT, no período de 2009 a 2019

REGRA (ÍNDICE)	CATEGORIA 1	→	CATEGORIA 2	Suporte %	Confiança %	Lift
1 – 1º ano	Desenho Técnico = Reprovado 16	→	Situação de matrícula = Evasão; Eletricidade I = Reprovado; Matemática I = Reprovado 15	0.10	0.94	7.2
2 – 2º ano	Física II = Reprovado 15	→	Situação de matrícula = Evasão; Matemática II = Reprovado 15	0.10	1	9.13
3 – 3º ano	Biologia III = Aprovado; Física III = Aprovado 68	→	Situação de matrícula = Conclusão; Geografia II = Aprovado; Química III = Aprovado 67	0.45	0.99	2.15
4 – 4º ano	Automação de Sistemas Elétricos = Aprovado 69	→	Situação de matrícula = Conclusão; Geografia III = Aprovado 68	0.46	0.99	2.12

Fonte: Elaborada pela pesquisadora a partir de dados obtidos por meio dos sistemas Qacadêmico e Qseleção do IFMT (2021)

A seguir será apresentada a interpretação das regras geradas de acordo com o índice exposto na tabela 3.

1º ano: Em 10% de todas as ocorrências, aparecem em conjunto as seguintes situações: Reprovação em Desenho Técnico, Eletricidade I e Matemática I e Evasão do Curso. Um aluno que reprova em Desenho Técnico possui 7,2 chances a mais de evadir do curso e reprovar em Eletricidade I e Matemática I.

2º ano: Em 10% de todas as ocorrências aparecem em conjunto as seguintes situações: Reprovação em Física II, Matemática II e evasão do curso. Um aluno que reprova em Física II apresenta 9,13 chances a mais de evadir do curso e reprovar em Matemática II.

3º ano: Em 45% de todas as ocorrências aparecem em conjunto as seguintes situações: Aprovação em Biologia III, Física III, Geografia II, Química III e conclusão do curso. Um aluno que é aprovado em Biologia III e Física III possui 2,15

chances a mais de concluir o curso e ser aprovado em Geografia II e Química III.

4º ano: Em 46% de todas as ocorrências aparecem em conjunto as seguintes situações: Aprovação em Automação de Sistemas Elétricos, Geografia III e conclusão do curso. Um aluno que é aprovado em Automação de Sistemas Elétricos possui 2,12 chances a mais de concluir o curso e ser aprovado em Geografia III.

A partir dos resultados do curso de EMI Técnico em Eletrotécnica foi verificada uma relação entre reprovações em disciplinas e a evasão do curso nos dois primeiros anos de curso. Já nos 3º e 4º anos de curso as regras apresentam uma maior associação entre aprovações em disciplinas e a conclusão do curso pelo aluno. Detectou-se, portanto uma tendência de evasão escolar maior nos primeiros anos do curso.

Diversos estudos apontam uma tendência de evasão escolar maior nos primeiros anos do curso, tendo em vista as frequentes reprovações no início do processo educacional. Peixoto, Braga e Bogutchi (2003) afirmam que o rendimento escolar nos primeiros anos de curso possui forte influência nas taxas de abandono, uma vez que tal resultado condiciona o aluno à um sentimento de vergonha devido ao fracasso acadêmico.

A seguir serão apresentadas as principais regras de associação do curso de EMI Técnico em Eletroeletrônica, conforme tabela abaixo. Para este curso foram analisadas 144 ocorrências de alunos evadidos e concluintes durante o período de 2009 a 2019. Destaca-se que diferentemente dos demais cursos do DAEE/CBA-OJS/IFMT, o curso de EMI Técnico em Eletroeletrônica possui 3 anos de duração. As regras geradas no primeiro e terceiro ano de curso não apresentavam

relação com a situação de matrícula do aluno, portanto foram analisadas as regras referentes ao segundo ano do curso.

Tabela 4 – Regras de associação do conjunto de dados acadêmicos dos alunos evadidos e concluintes do curso EMI Técnico de Eletroeletrônica do DAEE/CBA-OJS/IFMT, no período de 2009 a 2019

REGRA (ÍNDICE)	CATEGORIA 1	→	CATEGORIA 2	Suporte %	Confiança %	Lift
1 – 2º Ano	Biologia II = Aprovado 104	→	Situação de matrícula = Conclusão; Língua Portuguesa II = Aprovado 101	0.70	0.97	1.38
2 – 2º ano	Geografia II = Aprovado 104	→	Situação de matrícula = Conclusão; Língua Portuguesa II = Aprovado 101	0.70	0.97	1.38
3 – 2º ano	História II = Aprovado 104	→	Situação de matrícula = Conclusão; Língua Portuguesa II = Aprovado 101	0.70	0.97	1.38

Fonte: Elaborada pela pesquisadora a partir de dados obtidos por meio dos sistemas Qacadêmico e Qseleção do IFMT (2021)

A seguir será apresentada a interpretação das regras geradas de acordo com o índice exposto na tabela 4.

2º ano: Em 70% de todas as ocorrências aparecem em conjunto as seguintes situações: Aprovação em Biologia II, Língua Portuguesa II e conclusão do curso. Um aluno que é aprovado em Biologia II possui 1,38 chances a mais de concluir o curso e ser aprovado em Língua Portuguesa II.

2º ano: Em 70% de todas as ocorrências aparecem em conjunto as seguintes situações: Aprovação em Geografia II, Língua Portuguesa II e conclusão do curso. Um aluno que é

aprovado em Geografia II possui 1,38 chances a mais de concluir o curso e ser aprovado em Língua Portuguesa II.

2º ano: Em 70% de todas as ocorrências aparecem em conjunto as seguintes situações: Aprovação em História II, Língua Portuguesa II e conclusão de curso. Um aluno que é aprovado em História II possui 1,38 chances a mais de concluir o curso e ser aprovado em Língua Portuguesa II.

As regras geradas referentes ao segundo ano do curso de EMI Técnico em Eletretrônica apresentam como relação comum a conclusão no curso e a aprovação em “Disciplinas da Base comum” do curso. Corroborando com a análise descritiva realizada anteriormente, as maiores ocorrências de aprovação são de disciplinas pertencentes às “Disciplinas da Base comum” e as disciplinas com maiores ocorrências de reprovação são pertencentes às “Disciplinas da Base Profissional”. Narciso (2015) apontou como um fator decisivo para a evasão escolar no Ensino Médio Integrado o baixo desempenho em disciplinas de exatas e as disciplinas técnicas ou profissionais.

A seguir serão apresentadas as principais regras de associação do curso de EMI Técnico em Telecomunicações, conforme tabela abaixo. Para este curso foram analisadas 155 ocorrências de alunos evadidos e concluintes durante o período de 2009 a 2019. As regras geradas no primeiro e terceiro ano de curso não apresentavam relação com a situação de matrícula do aluno e o segundo ano não gerou nenhuma regra. Portanto foram analisadas as regras referentes ao quarto ano do curso.

Tabela 5 – Regras de associação do conjunto de dados acadêmicos dos alunos evadidos e concluintes do curso EMI Técnico de Telecomunicações do DAEE/CBA-OJS/IFMT, no período de 2009 a 2019

REGRA (ÍNDICE)	CATEGORIA 1	→	CATEGORIA 2	Suporte	Confiança	Lift
1 – 4º ano	Análise de Redes Telefônicas = Aprovado 85	→	Situação de matrícula = Conclusão; Biologia III = Aprovado 85	0.54	1	1.82
2 – 4º ano	Análise de Redes Telefônicas = Aprovado 85	→	Situação de matrícula = Conclusão; Geografia III = Aprovado 85	0.54	1	1.82
3 – 4º ano	Análise de Redes Telefônicas = Aprovado 85	→	Situação de matrícula = Conclusão; História III = Aprovado 85	0.54	1	1.82

Fonte: Elaborada pela pesquisadora a partir de dados obtidos por meio dos sistemas Qacadêmico e Qseleção do IFMT (2021)

A seguir será apresentada a interpretação das regras geradas de acordo com o índice exposto na tabela 5.

4º ano: Em 54% de todas as ocorrências aparecem em conjunto as seguintes situações: Aprovação em Análise de Redes Telefônicas, Biologia III e conclusão de curso. Um aluno que é aprovado em Análise de Redes Telefônicas possui 1,82 chances a mais de concluir o curso e ser aprovado em Biologia III.

4º ano: Em 54% de todas as ocorrências aparecem em conjunto as seguintes situações: Aprovação em Análise de Redes Telefônicas, Geografia III e conclusão de curso. Um aluno que é aprovado em Análise de Redes Telefônicas possui 1,82 chances a mais de concluir o curso e ser aprovado em Geografia III.

4º ano: Em 54% de todas as ocorrências aparecem em conjunto as seguintes situações: Aprovação em Análise de Redes Telefônicas, História III e conclusão de curso. Um aluno que é aprovado em Análise de Redes Telefônicas possui 1,82 chances a mais de concluir o curso e ser aprovado em História III.

A partir da interpretação das regras geradas no curso de EMI Técnico em Telecomunicações foi verificado que o aluno aprovado em uma disciplina de base profissional possui mais chances de ser aprovado em uma disciplina de base comum. Ressaltou-se ainda a constatação realizada nos cursos anteriores em que as maiores ocorrências de evasão ocorrem nos primeiros anos de curso, o que permite concluir que o foco da predição à evasão escolar deve ocorrer nas séries iniciais do Ensino Médio Integrado ao ensino técnico.

Resende (2012) aponta que as taxas de evasão são mais elevadas nas primeiras séries e na 3ª série do Ensino Médio apresenta-se bastante reduzida. A autora destaca ainda que o combate à evasão deve ser priorizado na série inicial do ensino médio.

Dados do Censo Escolar apontam uma redução nas taxas de abandono escolar ao longo das séries do Ensino Médio, conforme tabela abaixo.

**Tabela 6 – Taxa de abandono total do Ensino Médio no Brasil, por ano, em 2019**

Taxa de abandono total - 2019	
Ensino Médio	Taxa de abandono total
1ª ano	6,1
2ª ano	4,6
3ª ano	3,0

Fonte: Taxa de rendimento escolar (INEP, 2019)

De acordo com o INEP (2019), taxa de abandono indica a porcentagem de alunos que deixou de frequentar a escola após a data de referência do Censo. Ao analisar os dados do Censo Escolar referentes ao ano de 2019 publicados pelo INEP nota-se que as taxas de abandono escolar na primeira série do ensino médio são mais elevadas que nas demais séries, visto que na 3ª série a taxa de abandono reduziu significativamente.

## SAIBA MAIS

A seção 3 apresentou um passo a passo de como executar a aplicação da técnica de Mineração de Regras de Associação no software Weka. A seguir são demonstrados alguns elementos para auxiliá-lo no trabalho com os dados. Para acesso às sugestões, aponte a câmera do seu celular para o QR Code ou acesse o link abaixo de cada item.



Site da Plataforma Nilo Peçanha, desenvolvida pela Rede Federal de Educação Profissional, Científica e Tecnológica SETEC/MEC



<http://plataformanilopecanha.mec.gov.br/>



Para download do software Weka



link: [https://waikato.github.io/weka-wiki/downloading\\_weka/](https://waikato.github.io/weka-wiki/downloading_weka/)



Conjunto de dados acadêmicos dos alunos dos cursos de Ensino Médio Integrado ao ensino técnico do DAEE/CBA-OJS/IFMT utilizados como exemplo neste guia



link: [encurtador.com.br/tvBH6](http://encurtador.com.br/tvBH6)



Conjunto de dados acadêmicos dos alunos dos cursos de Ensino Médio Integrado ao ensino técnico do DAEE/CBA-OJS/IFMT, utilizados como exemplo neste guia



link: [encurtador.com.br/kzBC9](http://encurtador.com.br/kzBC9)



Arquivo com a Legenda das colunas (Atributos) dos dados socioeconômicos



link: [encurtador.com.br/gmwA2](http://encurtador.com.br/gmwA2)

## **PALAVRAS FINAIS**

*A evasão escolar é um fenômeno de alta complexidade que atinge a todas as Instituições e modalidades de Ensino. Garantir a permanência escolar vai além do êxito do aluno, constitui-se da materialização de um direito constitucional que estabelece como dever do Estado e da família a oferta de um ensino pautado no princípio de igualdade.*

*Diante disso, este guia apresentou um passo a passo de como detectar e categorizar os elementos que contribuem para a identificação de padrões de alunos com risco de evasão escolar mediante a utilização da técnica de Mineração de Dados Educacionais com aplicação das Regras de Associação. O exemplo do conjunto de dados dos cursos de Ensino Médio Integrado ao ensino técnico do DAEE/CBA-OJS/IFMT, foi utilizado para auxiliar o entendimento do leitor.*

*Destaca-se que a metodologia apresentada neste guia pode ser executada em outros conjuntos de dados, seja de outros cursos do IFMT ou ainda de outras instituições. É importante ressaltar que o processo de Mineração de Dados com utilização do KDD também pode ser desenvolvido em outros softwares de mineração, bem como outros métodos de pré-processamento dos dados podem ser adotados para a geração das regras de associação.*

*O processo de descoberta de conhecimento por meio da Mineração de Dados é fantástico! Transformar dados armazenados em conhecimento é de extrema relevância pois permite a detecção de padrões válidos e úteis, além de embasar as tomadas de decisões nas instituições de forma a prever fenômenos tão complexos, como a evasão escolar.*

*Portanto, espera-se que, por meio do passo a passo apresentado nesse guia, seja possível a aplicação da técnica de Mineração de Regras de Associação em outros conjuntos de dados, além de contribuir para a mitigação da evasão escolar nos cursos de Ensino Médio Integrado ao ensino técnico do DAEE/CBA-OJS/IFMT, conjunto de dados exemplificado neste guia e apoiar os trabalhos dos gestores, docentes, coordenadores de cursos e demais interessados no tema.*

Obrigada!

# REFERÊNCIAS

AGRAWAL, R.; SRIKANT, R. *Fast Algorithms for Mining Association Rules*. In: ACM VLDB INT. CONFERENCE ON VERY LARGE DATABASES, 20., 1994. Proceedings... Hove: Morgan Kaufmann, 1994. p.487-499.

BRASIL. *Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD)*. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/l13709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm). Acesso em: 15 fev. 2022.

BRASIL. *Ministério da Educação. Secretaria de Educação Superior. Plataforma Nilo Peçanha - PNP 2019 (Ano Base 2018)*. Brasília: MEC, 2019. Disponível em: <http://plataformanilopecanha.mec.gov.br/>. Acesso em: 29 out. 2019

BÜRKLE, Paula Yamada. *Um método de pós-processamento de regras de associação com base nas relações de dependência entre os atributos*. Dissertação de Mestrado em Computação, Otimização combinatória e inteligência artificial. Universidade Federal Fluminense. Rio de Janeiro, RJ, 107p, 2006.

CARDOSO, Olinda Nogueira Paes; MACHADO, Rosa Teresa Moreira. *Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras*. *RAP. Revista Brasileira de Administração Pública*, ISSN 0034-7612. 42(3): 495-528, Rio de Janeiro, v. 42, n. 3, p. 495-528, maio/jun. 2008. Disponível em: <http://repositorio.ufla.br/jspui/handle/1/184>. Acesso em: 31 out. 2019.

CIAVATTA, Maria. *Ensino Integrado, a Politecnia e a Educação Omnilateral: por que lutamos?* *Revista Trabalho & Educação*, v. 23, n. 1, p. 187-205, 2014. Disponível em: <https://seer.ufmg.br/index.php/trabedu/article/view/9303>. Acesso em: 15 ago. 2019.

DORE, Rosemary; LÜSCHER, Ana Zuleima. *Permanência e evasão na educação técnica de nível médio em Minas Gerais*. *Cadernos de pesquisa*, v. 41, n. 144, set./dez. 2011.

FAYYAD, Usama M.; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic, 1996. *From data mining to knowledge discovery in Databases*. *AI magazine*. v. 17, n. 3, 1996. Disponível em: <https://www.aaai.org/ojs/index.php/aimagazine/article/viewFile/1230/1131>. Acesso em: 03 dez. 2019

GARCIA, Paulo Sergio; AZEVEDO, Giovanna Rocha; SOBRINHO, Alex Moura. Um estudo sobre o fracasso escolar no Ensino Médio entre os anos de 2010, 2013 e 2016. *Revista Internacional d'Humanitats*, v. 45, p. 103-122, 2019.

HAN, Jiawei; PEI, Jian; KAMBER, Micheline. *Data mining: concepts and techniques*. Terceira edição. Elsevier, 2011.

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Taxas de rendimento escolar, 2020*. Disponível em: [https://download.inep.gov.br/educacao\\_basica/educacenso/situacao\\_aluno/documentos/2020/taxas\\_de\\_rendimento\\_escolar.pdf](https://download.inep.gov.br/educacao_basica/educacenso/situacao_aluno/documentos/2020/taxas_de_rendimento_escolar.pdf). Acesso em: 08 fev. de 2022.

NARCISO, Luciana Gusmão de Souza. *Análise da Evasão nos Cursos Técnicos do Instituto Federal do Norte de Minas Gerais – Câmpus Arinos: Exclusão da Escola ou Exclusão na Escola?*/ Luciana Gusmão de Souza Narciso; orientador, Erni José Seibel-Florianópolis, SC, 2015.

PASTA, Arquelau. *Aplicação da técnica de Data Mining na base de dados do ambiente de gestão educacional: um estudo de caso de uma instituição de ensino superior de Blumenau, SC*. 153 f. Dissertação (Mestrado em Computação Aplicada) - Universidade do Vale do Itajaí, São José, 2011.

PEIXOTO, M. C. L.; BRAGA, M. M.; BOGUTCHI, T. F. A evasão no ensino superior brasileiro: o caso da UFMG. *Revista da Rede de Avaliação Institucional da Educação Superior*. Campinas, v. 8, n.1, pp.161-189, mar, 2003.

RESENDE, Maria Liliana do Amaral. *Evasão escolar no primeiro ano do Ensino Médio Integrado do IFSULDEMINAS - Campus Machado*. 35º Encontro Anual da Anpocs GT08 – Educação e Sociedade, 2012.

ROMÃO, Wesley et al. *Extração de regras de associação em C&T: O algoritmo Apriori*. XIX Encontro Nacional em Engenharia de Produção, v. 34, p. 37-39, 1999.

SCHONHORST, Gustavo Bonnard. *Mineração de regras de associação aplicada à modelagem dos dados transacionais de um supermercado* / Gustavo Bonnard Schonhorst. -- Itajubá, (MG) : [s.n.], 2010. Disponível em: <https://repositorio.unifei.edu.br/jspui/handle/123456789/1495>. Acesso em: 24 set. 2021.

SILVA, Cassiana Fagundes; RODRIGUES, Cleuton. T.; MONTEIRO, Marcus Vinícius. B. *Uso de regras de associação para descoberta de conhecimento na produtividade de açaí no Estado do Amapá, 2010* In: *Anais do V Congresso Sul Brasileiro de Computação*. Disponível em: <http://periodicos.unesc.net/sulcomp/issue/view/25>. Acesso em:

SILVA, Cassiana Fagundes; RODRIGUES, Cleuton. T.; MONTEIRO, Marcus Vinícius. B. *Uso de regras de associação para descoberta de conhecimento na produtividade de açaí no Estado do Amapá, 2010* In: *Anais do V Congresso Sul Brasileiro de Computação*. Disponível em: <http://periodicos.unesc.net/sulcomp/issue/view/25>. Acesso em: 30 set. 2021

WEKA. *Machine Learning Software in Java. Versão 3.8: University of Waikato, 2019*. Disponível em: <https://www.cs.waikato.ac.nz/ml/weka/downloading.html>. Acesso em: 03 dez. 2019.

ZHANG, C.; ZHANG, S. *Association Rules Mining: Models and Algorithms. Lecture Notes in Artificial Intelligence*, v. 2307, Springer, 2002.