

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI  
FACULDADE DE CIÊNCIAS EXATAS  
CURSO DE SISTEMAS DE INFORMAÇÃO

**APLICAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS PARA AUXÍLIO À  
GESTÃO ACADÊMICA DA UFVJM**

**Marcus Vinícius Barbosa Costa**

Diamantina  
2012

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI  
FACULDADE DE CIÊNCIAS EXATAS

**APLICAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS PARA AUXÍLIO À  
GESTÃO ACADÊMICA DA UFVJM**

**Marcus Vinícius Barbosa Costa**

Orientador:

Prof<sup>o</sup>. Mestre Áthila Rocha Trindade

Trabalho de Conclusão de Curso apresentado ao curso de Sistemas de Informação como parte dos requisitos exigidos para a obtenção do título de Bacharel em Sistemas de Informação da Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM).

Diamantina

2012

**APLICAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS PARA AUXÍLIO À  
GESTÃO ACADÊMICA DA UFVJM**

**Marcus Vinícius Barbosa Costa**

Orientador:

Prof<sup>o</sup>. Mestre Áthila Rocha Trindade

Trabalho de Conclusão de Curso apresentado  
como parte dos requisitos exigidos para a  
obtenção do título de Bacharel em Sistemas de  
Informação e aprovado em \_\_\_\_/\_\_\_\_/\_\_\_\_  
pela banca examinadora constituída pelos pro-  
fessores:

---

Prof<sup>o</sup>. Mestre Áthila Rocha Trindade  
Orientador

---

Prof<sup>o</sup>. Doutor Alexandre Ramos Fonseca  
Universidade Federal dos Vales do  
Jequitinhonha e Mucuri

---

Prof<sup>a</sup>. Mestre Caroline Queiroz Santos  
Universidade Federal dos Vales do  
Jequitinhonha e Mucuri

*Dedico este trabalho aos meus pais  
por todo apoio e carinho.*

## AGRADECIMENTOS

Ao fim deste trabalho, deixo aqui meus agradecimentos:

- a Deus por tudo;
- ao prof<sup>o</sup>. Áthila Rocha Trindade por toda paciência e dedicação;
- aos professores da banca Alexandre Ramos Fonseca e Caroline Queiroz Santos pelos valiosos comentários;
- a todos os outros professores do curso de Sistemas de Informação da Universidade Federal dos Vales do Jequitinhonha e Mucuri;
- a minha família por todo o incentivo;
- a todos os amigos do curso de Sistemas de Informação pela agradável convivência ao longo desses anos.

## RESUMO

Os avanços na área de Tecnologia da Informação e a constante redução dos custos de armazenamento fizeram com que o volume de dados produzidos e armazenados por diversas organizações crescessem em uma velocidade extraordinária. Porém, tamanha quantidade de dados torna a análise dos mesmos somente através de métodos tradicionais muito complexa e dispendiosa, gerando a necessidade de técnicas e ferramentas computacionais específicas que permitam a obtenção de conhecimento relevante. A Mineração de Dados surgiu exatamente com esse propósito. Ela une conhecimentos das áreas de Estatística, Banco de Dados e Aprendizado de Máquina e conta com diversas técnicas (algoritmos) que, quando aplicadas adequadamente, podem revelar conhecimentos de grande valia para a gestão do negócio. Diante disso, neste trabalho foram analisadas as bases de dados referentes aos anos de 2002 a 2009 do processo seletivo unificado da Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), com o objetivo de obtenção de conhecimento acerca do perfil sócio-econômico dos candidatos. Foi utilizado o algoritmo de Mineração de Dados *Apriori* que implementa a extração de conhecimento através da obtenção de regras de associação, e o algoritmo *K-means* que a implementa através da formação de grupos com características em comum, o que é denominado clusterização. Os resultados obtidos foram satisfatórios, comprovando a eficácia destas técnicas para a descoberta de conhecimento em grandes bases de dados.

**Palavras-chave:** Mineração de Dados, Regras de Associação, *Apriori*, *K-means*, Clusterização, Descoberta de Conhecimento.

## **ABSTRACT**

*The advances in the area of Information Technology and the constant reduction of the storage costs made the volume of data produced and stored by several organizations grow in an extraordinary speed. However, so much amount of data makes the analysis by itself through of traditional methods very complex and expensive, generating the need of the computational specific techniques and tools that allow obtaining relevant knowledge. The Data Mining emerged exactly with this purpose. It combines knowledge of the areas of Statistics, Database and Machine Learning and it has several techniques (or algorithms) that, when applied properly, can reveal valuable knowledge to business management. Therefore, in this work were analyzed the databases regarding to the years 2002 to 2009 of the UFVJM University entrance process, for the purpose in the extraction of knowledge about the socio-economic profile of the candidates. It were used the algorithm of the Data Mining called Apriori, which implements the extraction of knowledge by obtaining association rules, and the algorithm called K-means which implements it by obtaining groups with similar characteristics, called clustering. The results obtained were satisfactory, proving the effectiveness this techniques for the knowledge discovery in large databases.*

**Keywords:** *Data Mining, Association Rules, Apriori, K-means, Clustering, Knowledge Discovery.*

## LISTA DE FIGURAS

<b>Figura 1:</b> Organização hierárquica do dado, informação, conhecimento e decisão . . .	2
<b>Figura 2:</b> Tela inicial do WEKA . . . . .	10
<b>Figura 3:</b> Aba <i>Preprocess</i> do WEKA . . . . .	10
<b>Figura 4:</b> Esquema da relação entre as tabelas utilizadas . . . . .	14
<b>Figura 5:</b> Etapas do KDD . . . . .	17
<b>Figura 6:</b> Áreas envolvidas na Mineração de Dados . . . . .	20
<b>Figura 7:</b> Pseudocódigo do algoritmo <i>Apriori</i> . . . . .	34
<b>Figura 8:</b> <i>Apriori</i> : poda de conjuntos candidatos usando a propriedade <i>apriori</i> . . . . .	38
<b>Figura 9:</b> <i>Apriori</i> : poda de regras de associação usando a medida de confiança . . . . .	40
<b>Figura 10:</b> <i>K-means</i> : escolha dos centros iniciais . . . . .	43
<b>Figura 11:</b> <i>K-means</i> : associação de um objeto ao centro de um <i>cluster</i> . . . . .	44
<b>Figura 12:</b> <i>K-means</i> : <i>clusters</i> obtidos após o cálculo de similaridade . . . . .	44
<b>Figura 13:</b> <i>K-means</i> : movimentação dos centros dos <i>clusters</i> . . . . .	45
<b>Figura 14:</b> Pseudocódigo do algoritmo <i>K-means</i> . . . . .	46
<b>Figura 15:</b> Aba <i>Preprocess</i> do WEKA . . . . .	50
<b>Figura 16:</b> Tela de configuração dos parâmetros do algoritmo <i>Apriori</i> no WEKA . . . . .	51
<b>Figura 17:</b> Regras extraídas pelo algoritmo <i>Apriori</i> no WEKA . . . . .	52
<b>Figura 18:</b> Tela de configuração dos parâmetros do algoritmo <i>K-means</i> no WEKA . . . . .	52
<b>Figura 19:</b> Resultado da execução do algoritmo <i>K-means</i> no WEKA . . . . .	53
<b>Figura 20:</b> Gráfico dos atributos sexo e local de residência (segundo semestre de 2009) . . . . .	67
<b>Figura 21:</b> Gráfico dos atributos aprovado e local de residência (segundo semestre de 2009) . . . . .	68
<b>Figura 22:</b> Gráfico dos atributos como conheceu o vestibular e local de residência (segundo semestre de 2009) . . . . .	69
<b>Figura 23:</b> Gráfico dos atributos aprovado e tipo da escola (segundo semestre de 2009) . . . . .	70
<b>Figura 24:</b> Gráfico dos atributos por que escolheu o curso e sexo (segundo semestre de 2009) . . . . .	71



**Figura 25:** Gráfico dos atributos aprovado e atividade remunerada (segundo semestre de 2009) . . . . . 72

**Figura 26:** Gráfico dos atributos aprovado e frequência pré-vestibular (segundo semestre de 2009) . . . . . 73

## LISTA DE TABELAS

<b>Tabela 1:</b>	Associação de uma pergunta do questionário às suas possíveis respostas . . .	13
<b>Tabela 2:</b>	Base de dados transacional de um supermercado . . . . .	30
<b>Tabela 3:</b>	Geração de combinações através do Modelo Suporte-Confiança . . . . .	33
<b>Tabela 4:</b>	Base de dados transacional a ser processada pelo algoritmo <i>Apriori</i> . . . . .	34
<b>Tabela 5:</b>	<i>Apriori</i> : conjuntos frequentes de tamanho 1 . . . . .	35
<b>Tabela 6:</b>	<i>Apriori</i> : conjuntos candidatos de tamanho 2 . . . . .	35
<b>Tabela 7:</b>	<i>Apriori</i> : conjuntos frequentes de tamanho 2 . . . . .	36
<b>Tabela 8:</b>	<i>Apriori</i> : conjuntos candidatos de tamanho 3 . . . . .	36
<b>Tabela 9:</b>	<i>Apriori</i> : conjuntos candidatos de tamanho 3 após aplicação de poda . . . . .	37
<b>Tabela 10:</b>	<i>Apriori</i> : conjuntos frequentes de tamanho 3 . . . . .	37
<b>Tabela 11:</b>	<i>Apriori</i> : conjuntos candidatos de tamanho 4 . . . . .	37
<b>Tabela 12:</b>	<i>Apriori</i> : conjuntos candidatos de tamanho 4 após aplicação de poda . . . . .	37
<b>Tabela 13:</b>	<i>Apriori</i> : regras de associação candidatas para um conjunto de 4 itens . . . . .	39
<b>Tabela 14:</b>	Base de dados a ser processada pelo algoritmo <i>K-means</i> . . . . .	42
<b>Tabela 15:</b>	Associação de uma pergunta do questionário às suas possíveis respostas . . .	49
<b>Tabela 16:</b>	Regras de associação envolvendo os atributos local de residência e tipo da escola . . . . .	57
<b>Tabela 17:</b>	Regras de associação envolvendo os atributos local de residência e curso . . .	83
<b>Tabela 18:</b>	Regras de associação envolvendo os atributos tipo da escola e curso . . . . .	85
<b>Tabela 19:</b>	Regras de associação envolvendo os atributos atividade remunerada e curso . .	88
<b>Tabela 20:</b>	Regras de associação envolvendo os atributos frequência pré-vestibular e notas	90
<b>Tabela 21:</b>	Regras de Associação envolvendo os atributos tipo da escola e notas . . . . .	94
<b>Tabela 22:</b>	Regras de associação envolvendo os atributos local de residência e notas . . .	100
<b>Tabela 23:</b>	Regras de Associação envolvendo os atributos curso e notas . . . . .	102
<b>Tabela 24:</b>	Regras de Associação envolvendo os atributos aprovado, tipo da escola e curso	104
<b>Tabela 25:</b>	Regras de Associação envolvendo os atributos aprovado, local de residência e curso . . . . .	110

<b>Tabela 26:</b> Regras de Associação envolvendo os atributos aprovado, atividade remunerada e curso . . . . .	115
<b>Tabela 27:</b> Regras de Associação envolvendo os atributos aprovado, frequência pré-vestibular e curso . . . . .	120

## LISTA DE ABREVIATURAS E SIGLAS

<b>ARFF</b>	<i>Attribute-Relation File Format</i>
<b>Conf</b>	Confiança
<b>Cont</b>	Continuação
<b>COPESE</b>	Comissão Permanente de Processos Seletivos
<b>GPL</b>	<i>General Public License</i>
<b>JDBC</b>	<i>Java Database Connectivity</i>
<b>KDD</b>	<i>Knowledge Discovery in Databases</i>
<b>Sem</b>	Semestre
<b>SGBD</b>	Sistema de Gerenciamento de Banco de Dados
<b>SQL</b>	<i>Structured Query Language</i>
<b>Sup</b>	Suporte
<b>UFVJM</b>	Universidade Federal dos Vales do Jequitinhonha e Mucuri
<b>WEKA</b>	<i>Waikato Environment for Knowledge Analysis</i>

**LISTA DE SÍMBOLOS**

$\approx$	Aproximadamente
$\emptyset$	Conjunto vazio
$\subseteq$	Contido
$\neq$	Diferente
$\div$	Divisão
$\wedge$	E
$\Rightarrow$	Implicação
$\cap$	Interseção
$\geq$	Maior ou igual a
$\leq$	Menor ou igual a
$\times$	Multiplicação
$\notin$	Não pertence
$\in$	Pertence
$\%$	Por cento
$\subset$	Propriamente contido
$ $	Tal que
$\cup$	União
$C_k$	Conjunto de itens candidatos de tamanho $k$
$F_k$	Conjunto de itens frequentes de tamanho $k$

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>1</b>
<b>2</b>	<b>OBJETIVO</b> .....	<b>4</b>
2.1	Objetivo Geral . . . . .	4
2.2	Objetivo Específico . . . . .	4
<b>3</b>	<b>JUSTIFICATIVA</b> .....	<b>5</b>
<b>4</b>	<b>METODOLOGIA</b> .....	<b>8</b>
4.1	Ferramentas Computacionais . . . . .	9
4.1.1	WEKA . . . . .	9
4.1.2	MySQL . . . . .	11
4.2	Descrição da Base de Dados . . . . .	12
<b>5</b>	<b>REVISÃO DE LITERATURA</b> .....	<b>15</b>
5.1	KDD ( <i>Knowledge Discovery in Databases</i> ) . . . . .	16
5.2	Mineração de Dados . . . . .	19
5.3	Regras de Associação . . . . .	23
5.4	Clusterização . . . . .	25
<b>6</b>	<b>REGRAS DE ASSOCIAÇÃO E CLUSTERIZAÇÃO: CONCEITOS BÁSICOS E ALGORITMOS</b> .....	<b>28</b>
6.1	Regras de Associação . . . . .	28
6.1.1	Descrição formal . . . . .	28
6.1.2	As medidas de Suporte, Confiança e <i>Lift</i> . . . . .	28
6.1.3	Modelo Suporte-Confiança . . . . .	32
6.1.4	O algoritmo <i>Apriori</i> . . . . .	33
6.2	Clusterização . . . . .	41
6.2.1	Descrição formal . . . . .	41
6.2.2	O algoritmo <i>K-means</i> . . . . .	41

<b>7 ESTUDO DE CASO DO VESTIBULAR DA UFVJM.....</b>	<b>47</b>
7.1 Seleção dos dados . . . . .	48
7.2 Limpeza dos dados . . . . .	48
7.3 Transformação dos dados . . . . .	48
7.4 Mineração de dados . . . . .	49
7.5 Análise e interpretação dos resultados . . . . .	54
<b>8 RESULTADOS COMPUTACIONAIS .....</b>	<b>55</b>
8.1 Regras de Associação . . . . .	55
8.1.1 Análise dos atributos local de residência e tipo da escola . . . . .	56
8.1.2 Análise dos atributos local de residência e curso . . . . .	59
8.1.3 Análise dos atributos tipo da escola e curso . . . . .	59
8.1.4 Análise dos atributos atividade remunerada e curso . . . . .	60
8.1.5 Análise dos atributos frequência pré-vestibular e notas . . . . .	60
8.1.6 Análise dos atributos tipo da escola e notas . . . . .	61
8.1.7 Análise dos atributos local de residência e notas . . . . .	61
8.1.8 Análise dos atributos curso e notas . . . . .	61
8.1.9 Análise dos atributos aprovado, tipo da escola e curso . . . . .	62
8.1.10 Análise dos atributos aprovado, local de residência e curso . . . . .	63
8.1.11 Análise dos atributos aprovado, atividade remunerada e curso . . . . .	64
8.1.12 Análise dos atributos aprovado, frequência pré-vestibular e curso . . . . .	65
8.2 Clusterização . . . . .	67
8.2.1 Análise dos atributos sexo e local de residência . . . . .	67
8.2.2 Análise dos atributos aprovado e local de residência . . . . .	68
8.2.3 Análise dos atributos como conheceu o vestibular e local de residência . . . . .	69
8.2.4 Análise dos atributos aprovado e tipo da escola . . . . .	70
8.2.5 Análise dos atributos por que escolheu o curso e sexo . . . . .	71
8.2.6 Análise dos atributos aprovado e atividade remunerada . . . . .	72
8.2.7 Análise dos atributos aprovado e frequência pré-vestibular . . . . .	73
<b>9 CONCLUSÕES E TRABALHOS FUTUROS .....</b>	<b>74</b>
<b>REFERÊNCIAS.....</b>	<b>78</b>
<b>APÊNDICE A – REGRAS GERADAS PELO ALGORITMO <i>APRIORI</i>.....</b>	<b>82</b>

## 1 INTRODUÇÃO

Nas últimas décadas, houve um aumento gigantesco no volume de dados produzidos pelas diversas organizações. A evolução das tecnologias de *software* e *hardware* aliada a uma redução constante nos custos de armazenamento contribuíram para que mais e mais empresas passassem a adquirir e utilizar computadores e dispositivos de armazenamento para a execução de suas atividades. A informatização possibilitou a coleta de quantidades cada vez maiores de dados sobre clientes, produtos, fornecedores, enfim, toda e qualquer atividade realizada no meio organizacional [18][27].

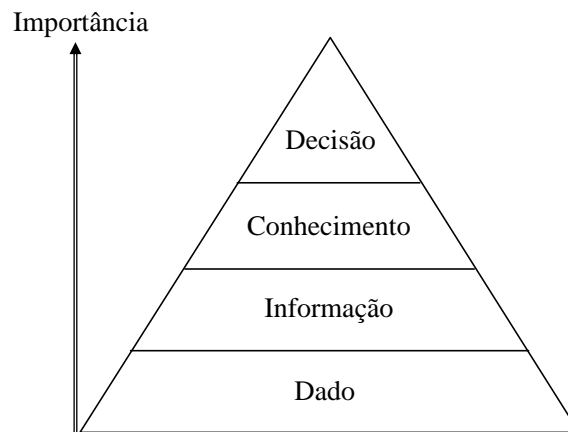
Além de gigantesca, essa quantidade de dados cresce continuamente. Em seu trabalho, Sferra e Corrêa [31] mencionam que há uma estimativa de que a quantidade de informação no mundo dobra a cada 20 meses, ao passo que o tamanho e o número de bases de dados crescem com velocidade ainda maior. Diante dessa enorme proliferação de dados e da crescente necessidade de obter conhecimento para a tomada de decisões, criou-se uma demanda por novas técnicas e ferramentas que pudessem auxiliar na obtenção do conhecimento útil e implícito nesses dados. Os esforços para satisfazer essa necessidade acabaram levando ao surgimento de uma nova área de pesquisa conhecida como Mineração de Dados [4].

Conforme mostra a pirâmide ilustrada na Figura 1, os dados são de fundamental importância para as organizações, pois representam a matéria-prima do processo decisório. Isso justifica os esforços na área de Mineração de Dados em desenvolver técnicas e ferramentas para estudá-los.

Como é possível perceber, os dados encontram-se na base da pirâmide, podendo ser representados por números, textos ou algum outro tipo de mídia como imagem, som ou vídeo. O dado é o elemento bruto e não transmite significado por si só. A menos que seja trabalhado, é pouco útil para se tomar decisões. Dados correspondem aos valores dos campos de um registro em uma base de dados [27].

A informação é o conjunto de dados organizados de forma a transmitir algum significado. Informações podem ser obtidas até mesmo através de consultas em um Sistema de Gerenciamento de Banco de Dados (SGBD) [29]. Uma consulta em um SGBD poderia forne-





**Figura 1:** Organização hierárquica do dado, informação, conhecimento e decisão (adaptado de [28])

cer uma informação do tipo: "15 celulares foram vendidos para o cliente X na data 10/01/2012" [28].

O conhecimento, por sua vez, é representado como um conjunto de dados e informações que agregam valor a um negócio, podendo até mesmo fornecer algum grau de certeza para ele [28]. Segundo Romão et al [29], a Mineração de Dados pode extrair um conhecimento do tipo "SE (idade = 30 a 40 anos) E (profissão = advogado) ENTÃO (produto = *notebook*) com um grau de certeza de 85%". Isso possibilitaria responder a uma pergunta de suma importância para a tomada de decisão: "Quais clientes tem maior probabilidade de comprar *notebooks*?", o que permitiria à organização promover ações para direcionar seus esforços para esse público.

Embora as organizações venham coletando e armazenando quantidades cada vez maiores de dados, a maior parte delas ainda não usa essa quantidade de dados para convertê-la em conhecimentos que possam ser utilizados para a gestão das suas próprias atividades. Nesse contexto se insere a UFVJM. Embora a instituição possua bases com quantidades significativas de dados acumulados no decorrer de vários anos, apenas uma vez a Mineração de Dados foi realizada. No trabalho de Pires [25], o processo foi aplicado na base de dados de um semestre do vestibular da instituição, utilizando uma das técnicas (algoritmos) de Mineração de Dados.

Considerando a importância do conhecimento para a gestão de qualquer negócio, este trabalho se propôs a realizar a Mineração de Dados nos outros semestres (dos anos de 2002 a 2009) em que o processo seletivo unificado foi promovido aplicando-se duas técnicas distintas. O objetivo foi obter conhecimento relevante sobre o perfil sócio-econômico dos candidatos ao vestibular, de forma a fornecer subsídios para o processo de tomada de decisão da instituição.

O órgão da universidade atualmente responsável pelo gerenciamento desses processos seletivos (vestibulares) é a Comissão Permanente de Processos Seletivos (COPESE). Portanto, para a realização deste trabalho, tal órgão disponibilizou as bases de dados dos anos mencionados e nelas foram feitas certas adaptações para que fossem submetidas à aplicação das técnicas de Mineração de Dados.

Este trabalho está estruturado em 9 capítulos. Neste foi apresentado o contexto em que o trabalho está inserido. O restante dele está organizado conforme os itens a seguir:

- No segundo Capítulo serão apresentados os objetivos geral e específico ao desenvolver este trabalho;
- No Capítulo 3 serão demonstrados os fatores que motivaram o desenvolvimento, destacando casos de aplicação da Mineração de Dados;
- O Capítulo 4 apresentará a metodologia empregada, onde serão descritos também as ferramentas computacionais utilizadas, a estrutura das bases de dados cedidas pela COPESE, bem como as alterações realizadas nessas bases para a aplicação da Mineração de Dados;
- Uma revisão da literatura incluindo todos os conceitos relacionados, bem como exemplos de aplicação das técnicas de Mineração de Dados utilizadas neste trabalho será abordada no Capítulo 5;
- O Capítulo 6 demonstrará aspectos relativos às tarefas de Regras de Associação e Clusterização, apresentando exemplos práticos que demonstram o mecanismo de funcionamento das suas principais técnicas ou algoritmos;
- No Capítulo 7 será feito um estudo do processo seletivo unificado da UFVJM, demonstrando passo-a-passo como foi realizada a Mineração de Dados e como foram analisados e interpretados, de uma forma geral, os resultados;
- No Capítulo 8 serão demonstrados os resultados obtidos com a aplicação das técnicas de Mineração de Dados para diferentes semestres de realização do vestibular;
- Por fim, no 9º Capítulo serão apresentadas as considerações finais e as propostas de trabalhos futuros.

## **2 OBJETIVO**

### **2.1 Objetivo Geral**

O objetivo deste trabalho foi aplicar técnicas de Mineração de Dados nas bases de dados dos processos seletivos unificados da UFVJM, no período de 2002 a 2009, com a finalidade de obter conhecimento útil sobre o perfil sócio-econômico dos candidatos, de forma a auxiliar na tomada de decisões da instituição em nível da gestão acadêmica. Este conhecimento foi baseado em informações acerca dos candidatos, a saber: sua origem, tipo da escola em que cursou o ensino médio, curso escolhido, se exerce atividade remunerada, sexo, frequência em pré-vestibular, como conheceu o vestibular da UFVJM, aprovação ou reprovação no vestibular e desempenho (notas).

### **2.2 Objetivo Específico**

O presente trabalho teve como objetivo específico cumprir os seguintes pontos:

- Entender o mecanismo de funcionamento do processo seletivo unificado da UFVJM;
- Estudar as tarefas de Regras de Associação e Clusterização, bem como os principais algoritmos utilizados por cada uma delas;
- Preparar as bases de dados dos vestibulares para serem submetidas à Mineração de Dados;
- Aplicar os algoritmos de extração de regras de associação e obtenção de grupos nessas bases;
- Analisar os resultados obtidos a fim de obter conhecimento útil para a tomada de decisão no que se refere ao perfil sócio-econômico dos candidatos ao processo seletivo unificado;
- Produzir o Trabalho de Conclusão do Curso de Sistemas de Informação;
- Publicar os resultados obtidos na forma de artigo científico.

### 3 JUSTIFICATIVA

A cada dia, as organizações vêm coletando e armazenando maiores quantidades de dados, resultantes de suas operações diárias. Entretanto, esse aumento contínuo no volume de dados produzidos e armazenados excede a capacidade e habilidade do ser humano de examiná-los e interpretá-los. Dessa forma, surge a necessidade de se utilizar técnicas e ferramentas computacionais que permitam realizar uma análise mais profunda e precisa desses dados, de forma a convertê-los em conhecimentos que possam ser utilizados para a gestão do negócio.

A COPESE, órgão da UFVJM responsável pelo gerenciamento dos processos seletivos, insere-se nesta situação. Embora o órgão possua bases com grande quantidade de dados sobre os candidatos ao vestibular acumulados no decorrer de vários anos, apenas em um semestre do processo seletivo esses dados foram submetidos a aplicação da Mineração de Dados, como pode ser visto em [25]. Porém, conforme demonstrado no Capítulo 1, o conhecimento é o suporte para a tomada de decisão e, portanto, sua obtenção contínua é de extrema importância para a organização, o que levou a realização da Mineração nas bases de dados dos demais semestres do vestibular da instituição.

Em Pires [25], o processo foi aplicado sobre a base de dados do segundo semestre de 2009 do vestibular unificado da UFVJM, utilizando a técnica *Apriori* para geração de regras de associação. Como descobertas deste trabalho, pode-se citar o fato de que a maior parte dos candidatos aos cursos de Educação Física e Bacharelado em Humanidades são de Diamantina e, portanto, a maioria dos aprovados são dessa cidade. Essa mesma situação ocorre com Ciências Contábeis, Ciências Econômicas, Administração e Serviço Social cuja maioria de candidatos são de Teófilo Otoni, e o curso de Odontologia que é essencialmente formado por candidatos de outras regiões de Minas Gerais. No geral, a maior parte dos candidatos aprovados não exercem atividade remunerada, com exceção de Ciências Contábeis. A maioria dos candidatos de Odontologia e Farmácia cursaram o ensino médio em escola particular e, portanto, a maior parte dos aprovados são desse tipo de escola. Essa mesma situação ocorre com Serviço Social, Bacharelado em Humanidades, Administração, Ciências Econômicas e Ciências Contábeis, que são essencialmente formados por candidatos de escola estadual. Por fim, a maioria dos candidatos

de Diamantina, Vale do Jequitinhonha e Vale do Mucuri estudaram em escola estadual, ao passo que a maior parte dos candidatos de Belo Horizonte e outras regiões de Minas Gerais estudaram em escola particular. Neste trabalho, não foi analisado o desempenho (notas) no vestibular.

Um outro projeto similar a esse foi realizado em 2004, na Universidade de Formiga (UNIFOR) [24]. Através da utilização do algoritmo *Apriori*, pertencente à tarefa de Regras de Associação, e da ferramenta de Mineração de Dados WEKA, foi feita uma análise sobre os dados relacionados ao questionário sócio-econômico-cultural, aplicado durante o processo seletivo da instituição. Os conhecimentos obtidos com a realização do trabalho foram satisfatórios, uma vez que foram utilizados pelos responsáveis da instituição como apoio às tomadas de decisão e na formação de conhecimento sobre os alunos. O projeto permitiu descobrir, entre outras coisas, que a maioria dos candidatos haviam cursado o ensino médio em escola pública no período diurno, estavam concorrendo ao processo seletivo pela primeira vez, residiam com os pais e localizavam-se na faixa etária de 17 a 20 anos. Além disso, uma boa parcela dos candidatos que residiam em Formiga escolheram a instituição por estar mais perto de casa e os que se localizavam a uma distância máxima de 100 quilômetros da cidade ficaram sabendo da Universidade por panfleto e a escolheram em razão do conceito. Isso tornou possível melhorar a qualidade das informações divulgadas, atingindo um público ainda maior.

A Mineração de Dados também tem sido muito utilizada na área da saúde. Segundo Freitas [16], a medicina, por exemplo, tem gerado grandes quantidades de dados que não são devidamente exploradas. Um exemplo de aplicação nessa área ocorreu em 2010, no Paraná [41]. O trabalho buscou aplicar técnicas de Mineração de Dados para identificar características relacionadas à mortalidade infantil. Para a realização do projeto, foram integradas as bases de dados de 2000 a 2004 de três órgãos responsáveis pelo controle, monitoramento e avaliação da mortalidade infantil: SIM (Sistema de Informações sobre Mortalidade), SINASC (Sistema de Informações sobre Nascidos Vivos) e SIMI (Sistema de Investigação da Mortalidade Infantil). Utilizando a técnica de árvores de decisão da tarefa de classificação e a ferramenta WEKA, procurou-se analisar aspectos relacionados à causa do óbito, de forma a auxiliar na sua redução. O trabalho permitiu obter o perfil da mortalidade infantil no estado em função de aspectos como sexo, idade e escolaridade da mãe, estado civil, peso e raça da criança, tipo de parto, número de consultas pré-natais, entre outros. O trabalho mostrou que o sexo masculino era predominante para óbitos; as mães adolescentes (menos de 16 anos) com escolaridade razoável (de 4 a 7 anos), sem filho morto, mas já com outro filho, apresenta risco de desconforto respiratório ao recém-

nascido. Além disso, o baixo peso ao nascer (menos de 2,5 quilogramas) com pós-datismo (gestação entre 40 e 42 semanas) estava relacionado ao óbito em 30 dias de vida.

Outro trabalho nessa área consistiu em analisar as bases de dados do CadSUS (Cartão Nacional de Saúde) e do CLEITOS (Sistema Nacional de Leitos), ambas cedidas pela Secretaria Municipal de Saúde de Londrina - PR [34]. Essas bases armazenam, respectivamente, dados sócio-econômicos da população e dados referentes à internação do paciente. A partir da utilização do algoritmo *Apriori*, o trabalho permitiu conhecer melhor as condições de vida da população e descobrir alguns aspectos importantes. Um deles é o fato de que a incidência de procedimentos de parto e pediátricos de urgência ou emergência é alta em áreas menos favorecidas e em pessoas de nível social mais baixo. Além disso, a grande maioria das pessoas que moram em domicílios de madeira são internadas com urgência ou emergência e a maior parte das cirurgias múltiplas e internações em psiquiatria são realizadas em pessoas do sexo masculino.

Um outro projeto envolvendo Mineração de Dados buscou analisar uma base de dados do Banco do Brasil de Guarapuava (Paraná) que continha registros de diversas empresas atendidas pela agência, sendo 266 adimplentes e 73 inadimplentes [20]. Essa agência detém uma grande fatia do mercado de pessoas jurídicas (micro, pequenos e médios empresários) da cidade, no que diz respeito ao crédito bancário. Aplicando as técnicas de Redes Neurais e Árvores de Decisão, ambas pertencentes à tarefa de Classificação, o objetivo foi obter conhecimentos que pudessem ser úteis na classificação de novas empresas como prováveis adimplentes ou inadimplentes, como forma de facilitar a análise de crédito. Para os autores, análises desse tipo podem trazer vantagens como necessidade de menos pessoas envolvidas com a análise de crédito, as quais podem ser aproveitadas em outras atividades; maior rapidez no processamento dos pedidos de crédito e direcionamento mais eficaz do mesmo. Ainda de acordo com os autores, as técnicas se mostraram eficientes na classificação dessas empresas e, de fato, foram de grande valia para os analistas de crédito.

Assim como nos casos citados anteriormente, a prática da Mineração de Dados nas bases de dados dos vestibulares da UFVJM também acarretará grandes benefícios. A aplicação do processo permitirá à instituição conhecer melhor o perfil sócio-econômico dos candidatos ao vestibular em função de diferentes aspectos como o local e escola de origem, tipo de atividade remunerada, frequência em pré-vestibular, entre outros, permitindo à Universidade uma tomada de decisão mais sólida e uma maior reflexão sobre as melhores ações a serem realizadas.

## 4 METODOLOGIA

Este trabalho foi marcado inicialmente por um estudo das tarefas de Mineração de Dados, mais especificamente das Regras de Associação e da Clusterização, bem como, respectivamente, dos algoritmos *Apriori* e *K-means*. Esse estudo se baseou em discussões entre orientado e orientador, na leitura de artigos científicos, capítulos de teses e dissertações, livros e apostilas a respeito do tema.

Em um segundo momento, foi realizada uma análise das bases de dados dos vestibulares unificados de 2002 a 2009 da UFVJM, com a finalidade de compreender a estrutura das tabelas e os relacionamentos existentes.

Para dar continuidade ao trabalho, foi efetuada a preparação dos dados para que eles pudessem ser submetidos à aplicação da Mineração. Isso envolveu três etapas fundamentais: a seleção, a limpeza e a transformação dos dados. Basicamente, essas etapas consistiram em, respectivamente, coletar os dados mais relevantes para o objetivo em questão; eliminar inconsistência, duplicidade e outros problemas existentes; bem como realizar algumas outras adaptações, como criação e alteração de tabelas e construção de visões<sup>1</sup> (*views*) e funções, de forma a facilitar a leitura e análise dos resultados. A preparação dos dados será melhor detalhada na Seção 4.2 e no Capítulo 7.

Uma vez completada essa etapa, a ferramenta de Mineração de Dados escolhida foi instalada e estudada, através da leitura de manuais e documentação disponíveis no site do desenvolvedor.

Concluída a preparação, essas bases de dados foram submetidas à Mineração de Dados, dando entrada na ferramenta escolhida. Nesse momento, os algoritmos mencionados anteriormente foram aplicados à essas bases e os resultados obtidos foram analisados e estudados, a fim de obter conhecimento útil.

Paralelamente a todas as atividades citadas anteriormente e com a coordenação do orientador, o Trabalho de Conclusão de Curso foi escrito conforme o padrão exigido.

Ao final, os resultados obtidos com a utilização das técnicas de Mineração de Dados

---

<sup>1</sup>Uma visão ou *view* é uma tabela virtual formada pela junção (união) de várias outras tabelas.

serão publicados na forma de artigo científico.

Com o uso crescente da Mineração de Dados, diversas ferramentas vem sendo propostas e desenvolvidas. Na Seção seguinte será apresentada uma delas, conjuntamente com a ferramenta selecionada para apoiar a análise e preparação das bases de dados.

## 4.1 Ferramentas Computacionais

### 4.1.1 WEKA

O WEKA (*Waikato Environment for Knowledge Analysis*) é um *software* para Mineração de Dados livre e gratuito. Essa ferramenta foi desenvolvida pela Universidade de Waikato na Nova Zelândia, sendo implementado pela primeira vez em 1997. Inicialmente, o desenvolvimento do *software* era voltado para a investigação de técnicas de aprendizagem de máquina, enquanto sua aplicação inicial foi direcionada para a agricultura, uma área de destaque na economia do país [36].

O WEKA conta com uma série de algoritmos para Mineração de Dados e utiliza a licença GPL (*General Public License*), o que permite estudar e alterar o seu código-fonte. É escrito em linguagem Java, podendo ser executado e transportado para diferentes plataformas [36]. O *software* possui uma interface gráfica que facilita a interação dos usuários e disponibiliza relatórios contendo gráficos e dados estatísticos de cada tarefa executada. O WEKA se encontra disponível em <http://www.cs.waikato.ac.nz/ml/weka>.

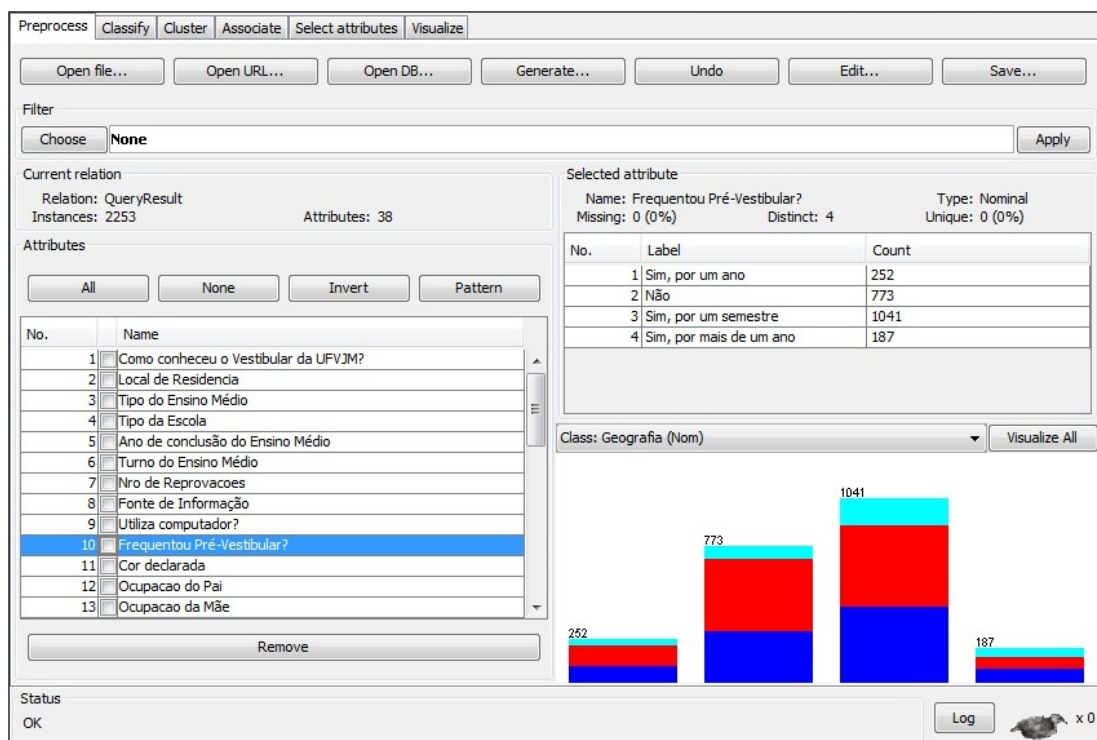
A interface inicial do WEKA, ilustrada na Figura 2, possibilita escolher entre os quatro modos de trabalho com a ferramenta: *Explorer*, *Experimenter*, *KnowledgeFlow* e *SimpleCLI*. Para este trabalho, será considerado o modo de operação WEKA *Explorer*.





**Figura 2:** Tela inicial do WEKA

O WEKA *Explorer* (Figura 3) é o ambiente principal do programa. Através dele, é possível executar todos os algoritmos de mineração de dados disponibilizados pela ferramenta. Essa janela contempla as seguintes abas: *Preprocess*, *Classify*, *Cluster*, *Associate*, *Select attributes* e *Visualize*. Neste trabalho, serão utilizadas as abas *Preprocess*, *Cluster* e *Associate*.



**Figura 3:** Aba *Preprocess* do WEKA

A aba *Preprocess* conta com alguns botões, dentre eles *Open File*, *Open URL* e *Open DB* que permitem, abrir, respectivamente, bases de dados a partir de arquivos ARFF, que é o

formato padrão do WEKA, bases que estejam disponíveis na web e diferentes bancos de dados via *driver* JDBC<sup>2</sup> (*Java Database Connectivity*).

O grupo *Attributes* exibe todos os atributos de uma base de dados, permitindo ao usuário selecionar aqueles que serão submetidos à Mineração de Dados. Para um atributo específico desse grupo, suas informações quantitativas podem ser visualizadas no grupo *Selected attribute*. Logo abaixo desse grupo, é possível também visualizar esses valores graficamente.

A aba *Cluster* do WEKA *Explorer* oferece opções relacionadas à tarefa de Clusterização em Mineração de Dados. Nela, é possível selecionar o algoritmo de Clusterização desejado, bem como configurar os seus parâmetros.

A aba *Associate* permite trabalhar com a tarefa de Regras de Associação. Assim como na aba anterior, na aba *Associate* pode-se definir o algoritmo a ser utilizado e configurar os respectivos parâmetros.

As interfaces das abas *Associate* e *Cluster* serão melhor apresentadas na Seção 7.4.

#### 4.1.2 MySQL

O MySQL é um SGBD<sup>3</sup> baseado no uso de comandos SQL (*Structured Query Language*). Ele foi criado na Suécia por David Axmark, Allan Larsson e Michael Monty Widenius, os quais têm trabalhado juntos no sistema desde 1980 [22].

O MySQL caracteriza-se por ser um *software* livre baseado no uso da licença GPL, ter bom desempenho, estabilidade, portabilidade, compatibilidade com outras linguagens, suportar tabelas de grande tamanho, ser rápido, confiável, multiplataforma e pouco exigente quanto a consumo de *hardware*, além de possuir um site com completo material de referência [38]. Por todos esses fatores, ele vem ganhando grande popularidade ao redor do mundo, sendo utilizado por grandes empresas. Seu desenvolvimento e manutenção contam com a participação de diversos profissionais e com a contribuição de várias pessoas pelo mundo, o que faz desse sistema um dos gerenciadores de bancos de dados mais utilizados [38].

Atualmente, o MySQL pode ser obtido em <http://www.mysql.com>, onde também é disponibilizada uma ampla documentação, bem como atualizações do sistema.

<sup>2</sup>Um *driver* JDBC permite que uma aplicação desenvolvida em linguagem Java possa se comunicar com um Sistema de Gerenciamento de Banco de Dados.

<sup>3</sup>Um SGBD é um *software* desenvolvido para manipular bases de dados, permitindo o armazenamento e recuperação dos dados [18].

O uso do MySQL auxiliou na análise das tabelas e dos relacionamentos existentes, além de facilitar o processo de preparação das bases de dados, através da adição de tabelas, criação de funções e *views* e correção de erros.

## 4.2 Descrição da Base de Dados

As bases de dados a serem utilizadas foram cedidas pelo órgão gerenciador dos processos seletivos da UFVJM, a COPESE. Elas armazenam diversos dados a respeito dos candidatos aos vestibulares nos anos de 2002 a 2009. Como cada base de dados contempla um grande conjunto de tabelas, optou-se por utilizar aquelas que armazenam dados pessoais e sócio-econômicos dos candidatos, bem como dados relacionados à aprovação ou reprovação desses candidatos.

A seguir, é apresentada uma breve descrição de cada uma das tabelas utilizadas:

- A primeira tabela armazena os códigos referentes às respostas fornecidas no questionário sócio-econômico que é preenchido no momento da inscrição no vestibular através da internet;
- A segunda tabela guarda, além do número de inscrição do candidato e do curso escolhido (que são comuns à todas as tabelas), os dados pessoais como nome, endereço, sexo, data de nascimento, telefones de contato, CPF, identidade, dados sobre ENEM (inscrição e ano), se é deficiente, se recebeu isenção, entre outros, que também são importantes;
- A terceira armazena as notas somente dos candidatos que foram aprovados no processo seletivo unificado, incluindo aquelas referentes ao ENEM, além da classificação de cada candidato;
- A quarta e última tabela é similar à anterior, porém armazena as notas de todos os candidatos ao vestibular sendo eles aprovados ou não.

Para a realização desse trabalho, cada base de dados foi adaptada para que pudesse ser submetida à Mineração de Dados. A seguir, é descrito brevemente como isso foi realizado.

Para cada pergunta do questionário sócio-econômico foi criada uma tabela para armazenar a resposta de forma descritiva. O número de questões existentes em cada questionário sofreu variação de acordo com o ano e semestre de realização do vestibular. Na tabela que registra os

códigos das respostas, havia uma coluna para cada pergunta do questionário sócio-econômico que armazenava a respectiva resposta de forma numérica, o que dificultava a leitura e análise. Para modificar isso e facilitar a análise dos dados, cada coluna foi associada à sua respectiva tabela que contém as possíveis respostas de forma descritiva. A Tabela a seguir exemplifica como foi associada a pergunta "Em que local você reside?" aos seus respectivos dados:

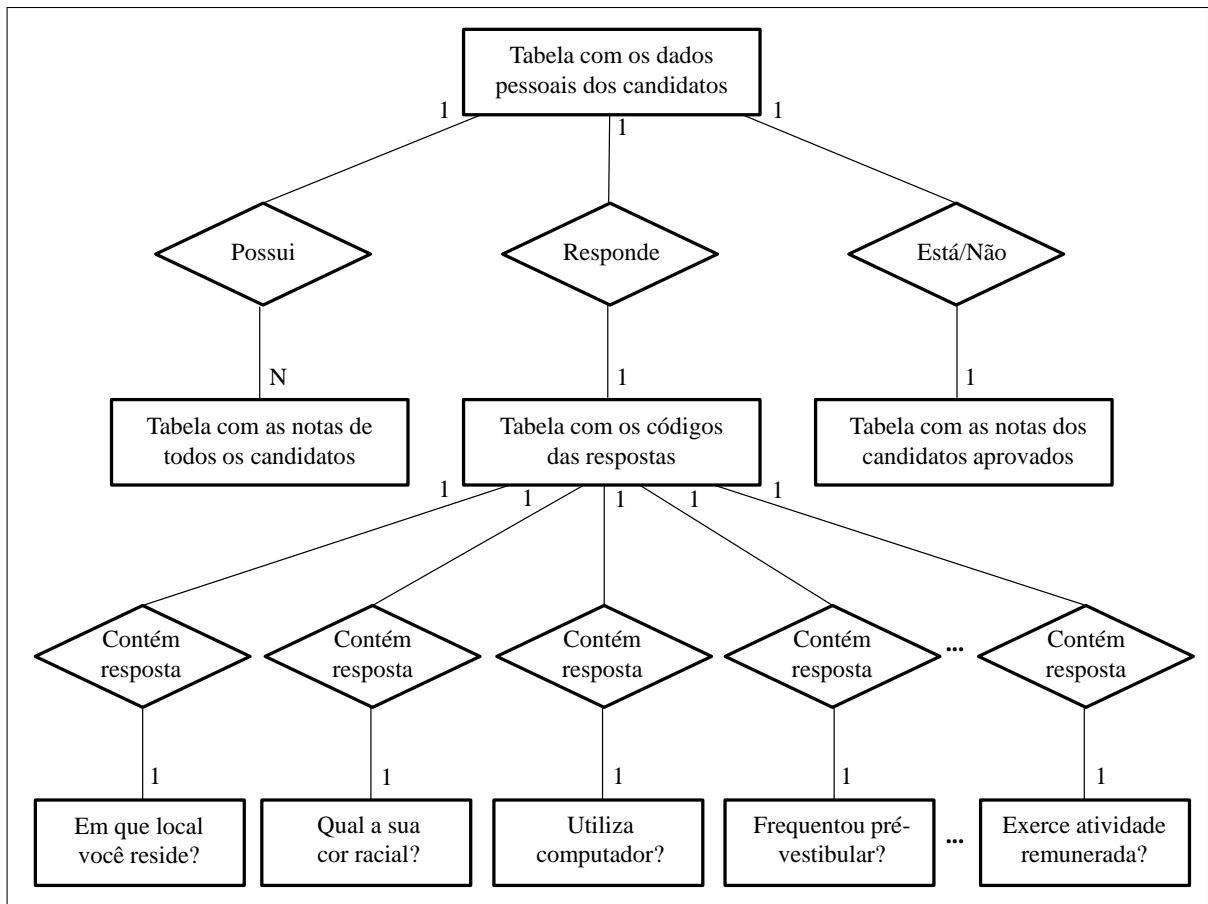
**Tabela 1:** Exemplo de associação das respostas do questionário socioeconômico para a pergunta: "Em que local você reside?"

<b>Residencia</b>	
<b>Código</b>	<b>Resposta</b>
1	Belo Horizonte
2	Diamantina
3	Teófilo Otoni
4	Na região do Vale do Jequitinhonha
5	Na região do Vale do Mucuri
6	Outra região de Minas Gerais
7	Outros estados

Um procedimento similar a este foi realizado para a tabela de notas com a adoção de conceitos descritivos (péssimo, regular, bom, muito bom e ótimo) ao invés de valores numéricos de acordo com o aproveitamento obtido em cada disciplina.

Uma vez concluído, foi criada uma visão para relacionar os dados de todas as tabelas. Com essa visão foi possível analisar o perfil pessoal e sócio-econômico dos candidatos aprovados ou não levando em consideração as notas obtidas por cada um deles em cada disciplina.

O diagrama a seguir demonstra como as tabelas das bases de dados foram interligadas para serem submetidas à Mineração de Dados.



**Figura 4:** Esquema da relação entre as tabelas utilizadas para Mineração de Dados

No esquema apresentado, os quatro primeiros retângulos representam as tabelas que foram utilizadas em cada base de dados. Conforme mostra a Figura 4, todos os candidatos possuem um conjunto de notas referentes a cada uma das matérias exigidas no vestibular, as quais são armazenadas na tabela de notas. De acordo com tais conceitos, estes candidatos podem ou não serem aprovados no processo seletivo. Aqueles que obtiveram o aproveitamento mínimo são registrados na tabela de candidatos aprovados. Ao se inscrever no vestibular, cada candidato deve ainda responder o questionário sócio-econômico. Tal questionário é formado por um determinado número de questões. Nesse caso, "Em que local você reside?", "Qual a sua cor racial?", "Utiliza computador?", "Frequentou pré-vestibular?" e "Exerce atividade remunerada?" correspondem a algumas das tabelas que armazenam as respostas para cada pergunta do questionário de forma descritiva. Portanto, todas são ligadas à tabela encarregada de armazenar o código referente à resposta de cada uma dessas perguntas.

## 5 REVISÃO DE LITERATURA

Nas últimas décadas, houve um contínuo e acelerado crescimento no volume de dados espalhado pelo mundo. A informatização das empresas, a queda no custo dos dispositivos de armazenamento e a evolução das tecnologias de *software* e *hardware* foram as grandes molas propulsoras desse crescimento [18]. Diante disso, a capacidade do ser humano de coletar e armazenar estes dados tornou-se muito superior à sua capacidade de analisá-los e compreendê-los. Dessa forma, as decisões importantes eram tomadas não com base nos dados ricos em conhecimento armazenados nas bases de dados, mas, simplesmente, na intuição do profissional já que o mesmo não dispunha de ferramentas eficientes para lidar com tamanho volume de dados [19]. De acordo com Han e Kamber [19], as bases de dados transformavam-se em verdadeiros "túmulos" de dados, ou seja, raramente eram visitadas.

Tal situação gerou a necessidade de técnicas e ferramentas computacionais que permitissem a extração de conhecimento a partir desses grandes volumes de dados de forma precisa e ágil, possibilitando seu posterior uso para a tomada de decisão [14]. Isso levou ao surgimento de uma nova área de pesquisa em banco de dados conhecida como Mineração de Dados. Segundo Fayyad, Piatetsky-Shapiro e Smyth [14], a Mineração de Dados é a principal etapa do processo de Descoberta de Conhecimento em Bases de Dados ou KDD (*Knowledge Discovery in Databases*), embora muitos vejam os termos como sinônimos. Porém, de acordo com Han e Kamber [19], ainda não há um consenso sobre isso e a Mineração de Dados vem adquirindo maior popularidade.

Muitas pessoas tratam de Mineração de Dados como um sinônimo de outro termo popularmente usado, *Knowledge Discovery in Databases* ou KDD. Alternativamente, os outros vêem Mineração de Dados como simplesmente um passo essencial no processo de Descoberta de Conhecimento em Bases de Dados. No entanto, na indústria, na mídia e no meio de pesquisa de banco de dados, a Mineração de Dados está se tornando, a longo prazo, mais popular do que a Descoberta de Conhecimento em Bases de Dados [19].

Em razão da falta de consenso entre os termos, nas Seções seguintes serão descritos, com mais detalhes, cada um deles.

## 5.1 KDD (*Knowledge Discovery in Databases*)

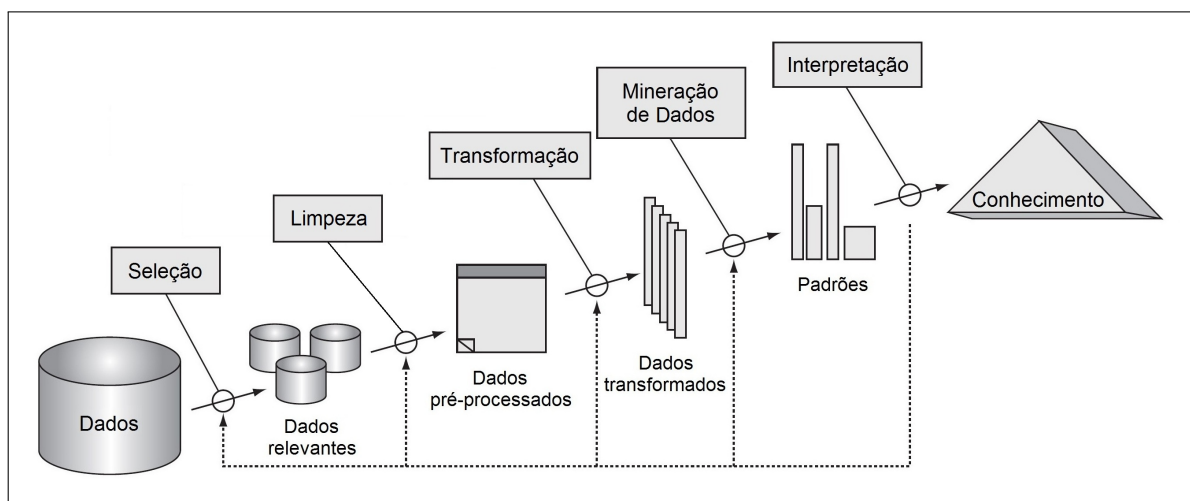
O termo KDD surgiu em 1989 para designar o processo não-trivial de identificar nos dados padrões novos, válidos, úteis e compreensíveis [14]. Segundo Silva [36] e Romão [28], cada um desses termos presentes no conceito de KDD podem ser definidos da seguinte forma:

- **processo:** é uma seqüência de passos que envolve preparação de dados, busca de padrões, avaliação de conhecimento, iteração e modificação;
- **não-trivial:** esse termo indica que o KDD envolve também busca e avaliação de padrões e não simplesmente o processamento dos dados e geração dos resultados;
- **padrão:** um padrão é uma expressão em alguma linguagem que descreva um conjunto de dados. Uma regra de associação representa um exemplo de padrão;
- **novos:** os padrões descobertos devem revelar um fato desconhecido e inesperado;
- **válido:** um padrão deve possuir algum grau de certeza, ou seja, deve satisfazer limites que garantam que os casos cobertos pelo padrão sejam aceitáveis. Por exemplo, um usuário pode decidir que somente os padrões em que um desses limites seja superior a 50% sejam considerados, já que valores inferiores a este podem representar uma minoria ou exceção;
- **útil:** os padrões precisam fornecer algum suporte para a tomada de decisão e levar a alguma ação que resulte em benefícios para o negócio. Por exemplo, os padrões obtidos podem ser utilizados para que se realizem ações que elevem o retorno financeiro de uma organização;
- **compreensível:** por fim, os padrões devem estar escritos em uma linguagem que possa ser facilmente compreendida pelos usuários. Simplicidade de representação constitui um bom fator de compreensão de acordo com os autores.

Sferra e Corrêa [31] descrevem sucintamente cada uma das etapas do KDD que incluem a seleção dos dados, a limpeza dos dados, a transformação dos dados, a Mineração de Dados e a avaliação e interpretação dos resultados, ressaltando que a limpeza dos dados é a que demanda maior tempo devido à dificuldade de integração de diferentes bases de dados. Um exemplo seria a existência de dados divergentes entre os setores de Vendas e Recursos Humanos, os quais possuem bases de dados independentes.

O KDD é um processo cíclico, uma vez que pode ser necessário voltar a alguma etapa para que, por exemplo, uma nova análise dos dados seja realizada ou caso novas necessidades venham a surgir [19]. Por exemplo, se em algum momento for detectada qualquer inconsistência ou identificada a necessidade de inserir novos dados para análise, faz-se necessário voltar à etapa de limpeza ou seleção, respectivamente.

O esquema a seguir apresenta cada uma dessas etapas. Antes de aplicá-las, é fundamental ter em mente os objetivos, identificar a fonte de dados e compreender a sua estrutura [7].



**Figura 5:** Etapas do KDD (adaptado de [13])

- **Seleção:** esta etapa consiste em escolher e reunir os dados mais relevantes para o objetivo em questão [18]. Por exemplo, o usuário pode concluir que telefone e endereço não são importantes para analisar se um cliente é bom pagador ou não. A etapa de seleção dos dados pode ser simples ou não em virtude de aspectos como a quantidade de bases de dados a serem utilizadas e a estrutura de cada uma [15].
- **Limpeza:** também conhecida como pré-processamento, nessa etapa os dados obtidos são analisados a fim de eliminar problemas existentes como inconsistência, duplicidade, contradição, campos em branco [18] [7]. A inconsistência surge quando, para um mesmo atributo, existem valores diferentes que representam a mesma coisa [10]. Um exemplo seria o atributo Estado que pode assumir os valores MG, Minas Gerais, Mg ou minas gerais. A duplicidade ocorre quando um mesmo registro é armazenado mais de uma vez na base de dados [18]. Um exemplo prático seria uma base de dados possuir um mesmo re-



gistro de um cliente armazenado duas vezes, porém com identificações diferentes. Como exemplo de contradição, tem-se uma base de dados contendo um registro com idade 19 anos e escolaridade Mestrado.

- **Transformação:** essa etapa consiste em converter os dados pré-processados para um formato que possa ser utilizado pelos algoritmos de Mineração de Dados [18]. Barroso e Neto [7] citam como exemplos a transformação de atributos qualitativos em quantitativos. Um exemplo seria atribuir 1 e 2 para os valores qualitativos sim e não. De acordo com Freitas [16], o processo inverso também pode ser necessário, já que muitos algoritmos de aprendizagem só trabalham com atributos qualitativos. Gonçalves [18] aponta a possibilidade de se agrupar valores, já que podem existir diversos valores diferentes para um mesmo atributo. Como exemplo, a partir do atributo CEP poderia ser formado o atributo Região, o qual seria resultante do agrupamento de diversos CEP's e, portanto, assumiria uma quantidade bem menor de valores possíveis. Filho [15] afirma que, nessa etapa, pode-se realizar junções entre as tabelas da base de dados de forma a gerar uma tabela única contendo todos os dados com os quais deseja-se trabalhar. Isso permite armazenar os dados de uma forma mais organizada, pois inclui-se apenas aqueles que serão utilizados no processo.
- **Mineração de Dados:** é a principal etapa do processo e, portanto, merece uma seção dedicada a ela. A Mineração de Dados será descrita com maiores detalhes na Seção seguinte. Basicamente, essa etapa consiste em aplicar os algoritmos de Mineração de Dados sobre os dados transformados, a fim de encontrar padrões interessantes [18]. Para isso, é necessário identificar quais tarefas serão utilizadas e escolher o algoritmo a ser aplicado antes de dar início a essa etapa [7].
- **Interpretação:** por fim, os padrões obtidos são analisados com o propósito de obter algum conhecimento útil para que, posteriormente, este possa se converter em ações a serem aplicadas em benefício do negócio [18].

## 5.2 Mineração de Dados

Na literatura são encontradas definições distintas para Mineração de Dados, sendo algumas delas:

*Data Mining* é o processo de descoberta de conhecimento interessante a partir de grandes quantidades de dados armazenados em bases de dados, *Data Warehouses*<sup>1</sup> e outros repositórios [19].

A Mineração de Dados faz parte de uma das etapas da Descoberta de Conhecimento em Bases de Dados. Ela é capaz de revelar o conhecimento que está implícito em grandes quantidades de informações armazenadas nas bases de dados de uma organização [8].

A Mineração de Dados é a extração de informação implícita, previamente desconhecida e potencialmente útil a partir dos dados [42].

Mineração de dados ou *Data Mining* é termo utilizado para nomear o processo de análise de conjuntos de dados com o objetivo de encontrar padrões que representem informações úteis e não triviais. Para tanto, utiliza-se de métodos matemáticos, heurísticas e algoritmos [26].

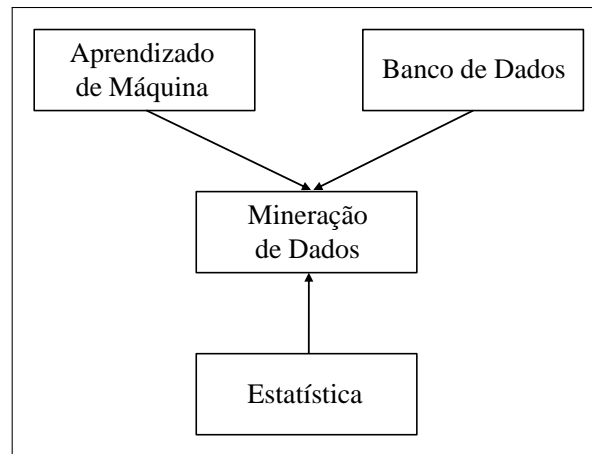
Conforme mostra a Figura 6, a Mineração de Dados une conhecimentos das áreas de Estatística, Banco de Dados e Aprendizado de Máquina.

A Estatística envolve conceitos que são muito importantes no estudo dos dados como variância, distribuição normal, desvio simples, intervalo de confiança, frequência, análise de conjuntos, entre outros [6]. Alguns desses conceitos são muito utilizados pela Mineração de Dados, conforme será possível perceber no decorrer do Capítulo 6. Como exemplo, tem-se as Regras de Associação em Mineração de Dados que são amplamente baseadas em análises estatísticas.

O Aprendizado de Máquina objetiva encontrar uma forma de fazer com que os programas aprendam com os dados estudados e tomem decisões tendo em vista as características

---

<sup>1</sup>*Data Warehouse* é um depósito de dados utilizado para armazenar um conjunto de dados referentes às atividades de uma organização, os quais oferecem suporte à tomada de decisão [26].



**Figura 6:** Áreas envolvidas na Mineração de Dados (extraído de [18])

destes dados [6]. Uma das aplicações são métodos de Classificação que utilizam técnicas estatísticas e de aprendizado de máquina para, a partir de um estudo prévio dos dados, realizar análises antecipadas com o intuito de prever comportamentos futuros [4]. Isto será mais detalhado na Página 21 deste Capítulo.

A Mineração de Dados resultou de uma evolução na área de Banco de Dados [18]. Atualmente, as diversas bases de dados são analisadas visando obter conhecimento novo, ao contrário de antigamente em que elas eram usadas apenas para processamento de transações de rotina das organizações como folha de pagamento, controle de estoque, entre outros [18]. Essa mudança ocorreu a partir do momento em que o acúmulo excessivo de dados resultantes das transações diárias das organizações começou a despertar a preocupação dos profissionais, que precisavam encontrar uma forma de obter conhecimento útil a partir desses dados e utilizá-lo em benefício da organização [4].

A Mineração de Dados combina todo um conjunto de tarefas e técnicas. De acordo com Gonçalves [18] e Amo [4], uma tarefa específica o tipo de conhecimento a ser minerado da base de dados, podendo este ser representado na forma de regras, modelos, grupos, entre outros. Cada tarefa possui um conjunto de técnicas, que correspondem aos métodos (algoritmos) que serão empregados.

Quanto às tarefas de Mineração de Dados, de acordo com Amo [4] e Han e Kamber [19], as principais são:

- **Regras de Associação:** Uma regra de associação é uma expressão na forma  $X \Rightarrow Y$  (lê-se "X implica em Y" ou "se X então Y"), onde X e Y representam conjuntos de itens. O

objetivo dessa tarefa é identificar itens que ocorram simultaneamente em transações de uma base de dados. Um caso clássico aplicável é o de análise de cestas de compras, que consiste em identificar quais produtos são mais vendidos em conjunto e, dessa forma, colocá-los próximos uns dos outros nas prateleiras para induzir os clientes a comprarem esses itens. Um exemplo de regra de associação seria  $\{\text{café}\} \Rightarrow \{\text{pão}\}$ , que indica que uma porcentagem dos clientes adquirem café e pão na mesma compra. Essa tarefa será estudada com mais detalhes nas Seções 5.3 e 6.1 em virtude de ser um dos focos deste trabalho.

- **Padrões Sequenciais:** Essa tarefa tem como objetivo descobrir conjuntos de itens que ocorram frequentemente em uma base de dados obedecendo a uma determinada sequência de aparição. É semelhante à tarefa de Regras de Associação porém, nesse caso, importa a ordem em que os itens ocorrem. Um exemplo seria a sequência  $\{\text{maçã}\} \Rightarrow \{\text{banana}\}$  que indica que os clientes compram maçã e, algum tempo depois, compram banana.
- **Classificação e Predição:** A Classificação consiste em encontrar modelos que descrevem classes de itens. A Predição consiste em utilizar tais modelos para classificar futuros itens. Em outras palavras, a partir de itens já classificados busca-se criar um modelo que permita classificar futuros itens de acordo com suas características a fim de prever algum comportamento futuro. Um exemplo prático seria descobrir qual o risco de fornecer empréstimo a um cliente. A empresa poderia identificar três modelos de classificação: "clientes da faixa econômica C com idade superior a 45 anos representam alto risco", "clientes da faixa econômica B com idade entre 45 e 55 anos representam um risco médio" e "clientes da faixa econômica B com idade entre 30 e 40 anos representam um risco baixo". A partir desses modelos, será possível prever se um futuro cliente oferecerá um alto, baixo ou médio risco para a empresa tendo em vista suas características e, com isso, decidir pela concessão ou não do empréstimo.
- **Clusterização:** Esse tipo de tarefa faz com que os itens sejam agrupados de acordo com o grau de semelhança entre eles. Dessa forma, aqueles que possuírem alto grau de similaridade permanecerão no mesmo grupo ou *cluster* e os que possuírem baixo grau de semelhança permanecerão em *clusters* distintos. A Clusterização poderia ser usada para identificar diferentes grupos de clientes. Uma empresa poderia realizar a Clusterização para obter grupos de clientes que possuem o mesmo perfil de aquisição de serviços, por

exemplo. Em razão dessa tarefa também ser um dos focos deste trabalho, ela será detalhada nas Seções 5.4 e 6.2.

- **Análise de *Outliers*:** *Outliers* são dados que não condizem com o comportamento da maioria dos outros. Esses valores são tratados como exceção e, por isso, são muitas vezes excluídos ou simplesmente desconsiderados. Porém, em muitos casos, eles podem ser importantes. A Análise de *Outliers* poderia ser útil, por exemplo, para identificar fraudes em cartão de crédito que ocorre quando há um gasto muito alto, ou seja, fora do padrão de compra daquele cliente. Nesse caso, esses valores fora do padrão poderiam ser a chave para identificar essas fraudes.

Um aspecto de destaque da Mineração de Dados é a possibilidade dela ser utilizada nas mais diversas áreas. Finanças, ramo alimentício, vendas, medicina, indústria e detecção de fraudes constituem algumas das áreas em que a Mineração de Dados tem sido empregada [9] [7] [20] [30]. A seguir são apresentados alguns exemplos práticos de acordo com Lemos, Steiner e Nievola [20].

- **Vendas:** detectar perfis de compra de clientes, os quais podem ser usados em futuras promoções; identificar quais produtos são mais adquiridos conjuntamente e obter o perfil do cliente de acordo com a região demográfica;
- **Medicina:** prever diagnósticos, antecipar tratamentos e prever quais pacientes têm maior chance de adquirir uma doença com base nos registros históricos das doenças e dos pacientes;
- **Detecção de fraudes:** identificar fraudes em cartões de crédito e telecomunicação em razão da existência de alto número de compras e chamadas ou mesmo em ligações de água e energia elétrica em função de consumo fora do habitual.

Devido a popularidade assumida pela Mineração de Dados e a importância de se obter o conhecimento implícito em bases de dados, conforme afirmam Han e Kamber [19], um número amplo de publicações vem se dedicando ao tema. Nas seções 5.3 e 5.4 serão apresentados alguns trabalhos já realizados envolvendo a Mineração de Dados, com foco nas tarefas e técnicas utilizadas neste trabalho.

### 5.3 Regras de Associação

A tarefa de Regras de Associação foi introduzida por Agrawal, Imielinski e Swami [2] e tem como objetivo encontrar conjuntos de itens que ocorram simultaneamente em transações de uma base de dados e formar regras a partir desses conjuntos. A seguir são apresentadas algumas definições para Regras de Associação:

As Regras de Associação abrangem a busca por itens que frequentemente ocorram de forma simultânea em transações da base de dados [37].

A tarefa de Regras de Associação se propõe a encontrar todas as associações relevantes entre um conjunto de itens aplicados a outros itens, e utiliza alguns algoritmos para realizar seu objetivo [34].

Uma Regra de Associação é representada como uma implicação na forma  $LHS \Rightarrow RHS$ , em que LHS e RHS são respectivamente o antecedente (*Left Hand Side*) e o conseqüente (*Right Hand Side*) da regra [30].

As Regras de Associação identificam itens que ocorrem com frequência, por exemplo, itens que costumam ser comprados pelo mesmo cliente [32].

De acordo com Gonçalves [17], as regras de associação são avaliadas com base em certas medidas que mensuram o quanto elas são relevantes para o usuário. Tais medidas podem ser classificadas como objetivas ou subjetivas. As medidas de interesse objetivas envolvem uma análise estatística e incluem o suporte, a confiança e o *lift*. Por serem medidas de grande importância no estudo das regras de associação, elas serão detalhadas no Capítulo seguinte. Quanto maiores forem os valores dessas medidas, mais interessante se torna a regra. As medidas de interesse subjetivas, por sua vez, focam mais o interesse de quem analisa as regras, uma vez que estas podem ser interessantes para uma pessoa mas não para outra. O que determina se uma regra é interessante ou não do ponto de vista subjetivo é a sua capacidade de revelar fatos úteis e inesperados para quem dela faz uso. É fundamental ainda atentar para o fato de que, apesar das regras de associação proporcionarem conhecimentos valiosos, elas podem, muitas vezes, também revelar relacionamentos óbvios entre os itens. É como dizer, por exemplo, que os clientes que compram arroz também compram feijão.

A intuitividade e a simplicidade de representação são as características marcantes das regras de associação e o que as tornam amplamente utilizadas. Tal tarefa conta com alguns algoritmos para extração de conhecimento, sendo o *Apriori* o mais conhecido e utilizado. Proposto por Agrawal e Srikant [3], esse algoritmo ganhou destaque por ter resolvido os problemas existentes no modelo anterior de geração de regras de associação, o modelo Suporte-Confiança. Isso foi possível graças a utilização de uma propriedade conhecida como *apriori* e de técnicas de junção e poda [18], que serão detalhadas no Capítulo seguinte.

Em 1999, Romão et al [29] buscou verificar a viabilidade do *Apriori* como um algoritmo para extração de regras de associação. Para isso, ele realizou a Mineração de Dados envolvendo a tarefa de Regras de Associação na base de dados do diretório dos grupos de pesquisa no Brasil, cedido pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico). Esse diretório contava com 33.765 registros de pesquisadores, onde cada registro era formado por diversos campos. O foco do trabalho foi analisar o perfil dos pesquisadores e, segundo o autor, o algoritmo gerou diversas regras interessantes confirmando o seu potencial para extração de regras de associação.

Silva, Boscaroli e Peres [35] realizaram um trabalho em que fizeram uso da tarefa e do algoritmo *Apriori* para investigar o perfil dos usuários que acessam o site de um provedor de serviços de internet. Para isso, foram analisados os arquivos de *log* de um servidor Web a fim de obter as páginas, serviços e horários mais acessados para oferecer serviços, produtos e atendimento personalizado. O estudo levou à conclusão de que nos períodos da manhã e da tarde havia um fluxo maior de visitantes do que de clientes em oposição ao que ocorria no período da noite.

Barroso e Neto [7] também fizeram uso da tarefa e do algoritmo para analisar a base de dados de uma locadora de filmes com a finalidade de obter conhecimento para a tomada de decisão. O trabalho revelou, entre outros pontos, que homens, no geral, tem uma tendência maior a locar filmes de ação e na faixa de 31 a 50 anos a locar filmes de desenho animado pois, provavelmente, tem filhos.

Por fim, Silva, Rodrigues e Monteiro [33] apresentaram um trabalho em que foi analisada a base de dados do Instituto de Pesquisas Científicas e Tecnológicas do Estado do Amapá (IEPA). Utilizando a ferramenta WEKA e o algoritmo *Apriori*, o objetivo foi identificar quais os fatores que influenciavam na produtividade do Açáí no estado. O trabalho permitiu concluir que a grande maioria dos produtores possuíam outra atividade e empregavam mão-de-obra fa-

miliar, praticamente não tinham acesso a financiamento e assistência técnica e utilizavam áreas pequenas a médias para cultivo, o que fazia a produção ser pouco expressiva.

A aplicação mais clássica de Regras de Associação é a análise de cestas de compras. Neste tipo de problema, procura-se identificar quais produtos são mais adquiridos em conjunto [18], conforme já citado na Página 21. Schonhorst [30] demonstra isso ao utilizar a tarefa de Regras de Associação e o algoritmo *Apriori* para fazer a análise de cesta de compras de um supermercado do Sul de Minas Gerais com cerca de 15.000 itens. Além de identificar quais itens eram comprados em conjunto, o autor também analisou as vendas em diferentes dias da semana e períodos do dia. Um exemplo prático de regra de associação desse tipo é:  $\{\text{café}\} \wedge \{\text{pão}\} \Rightarrow \{\text{biscoito}\}$ . Esta regra demonstra que as pessoas que compram café e pão tendem a comprar biscoito. Mais exemplos práticos disso serão demonstrados no Capítulo seguinte.

Juntamente com as Regras de Associação será utilizada a tarefa de Clusterização que será descrita na Seção seguinte.

#### 5.4 Clusterização

Clusterização, Análise de Agrupamentos ou Análise de Grupos é o termo usado para referenciar a tarefa de agrupar itens com base em algum grau de semelhança entre eles. Cada grupo é também denominado *cluster*. O agrupamento ideal é aquele em que os itens com alto grau de similaridade se localizam no mesmo *cluster* e os itens com baixo grau de semelhança se encontrem em *clusters* distintos [39]. As duas definições a seguir descrevem bem o propósito da Clusterização em Mineração de Dados:

A Análise de Agrupamentos associa um item a uma ou várias classes categóricas (ou *clusters*). Os *clusters* são definidos por meio do agrupamento de dados baseados em medidas de similaridade ou modelos probabilísticos, visando detectar a existência de diferentes grupos dentro de um determinado conjunto de dados e, em caso de sua existência, determinar quais são eles [12].

A Análise de Agrupamentos visa segmentar um conjunto de dados em um número de subgrupos homogêneos ou *clustering*. Seu objetivo é formar grupos baseados no princípio de que esses grupos devem ser o mais homogêneos em si e mais heterogêneos entre si [11].



De acordo com Ochi [23], o número de grupos a serem encontrados pode ser conhecido ou não. Caso esse número seja conhecido, o problema é denominado problema de  $k$ -clusterização, em que  $k$  corresponde ao número de grupos a serem obtidos. Caso não seja conhecido, é chamado problema de clusterização automática. Para as situações em que  $k$  é desconhecido, o número de soluções possíveis é ampliado. Por exemplo, para um conjunto de 10 elementos há 115.975 possibilidades, uma vez que os elementos deverão ser combinados em *clusters* de tamanho 1 a 10. Mas, se o número de *clusters* for fornecido, a quantidade de soluções decresce. Por exemplo, para um mesmo conjunto de 10 itens, dado a necessidade de se formar 2 *clusters* distintos, há 511 soluções possíveis. De acordo com o autor, o número de soluções possíveis varia em função das seguintes fórmulas:

- número de soluções em que a quantidade de *clusters* é conhecida:

$$N(n, k) = \frac{1}{k!} \sum_{i=0}^k (-1)^i C_{(k,i)} (k-i)^n$$

- número de soluções em que a quantidade de *clusters* é desconhecida:

$$N(n) = \sum_{k=1}^n \frac{1}{k!} \sum_{i=0}^k (-1)^i C_{(k,i)} (k-i)^n$$

em que  $n$  e  $k$  representam, respectivamente, o número de elementos a serem agrupados e a quantidade de *clusters* a serem formados. Embora estas fórmulas forneçam todas as possibilidades de agrupamento possíveis, isso não é o suficiente em uma tarefa de Clusterização. É preciso, além disso, adotar algum critério que permita inserir em um mesmo grupo apenas os itens que possuam algum grau de semelhança. Portanto, outro aspecto fundamental a ser considerado em relação a essa tarefa é como medir o quanto um elemento é semelhante a outro para que seja possível verificar se ambos devem fazer parte do mesmo grupo ou não. Isso é feito através da utilização de alguma medida de similaridade [23]. Na Subseção 6.2.2 será apresentado com mais detalhes o uso de uma dessas medidas no processo de execução de um algoritmo de Clusterização.

Assim como as Regras de Associação, a Clusterização conta com diversos algoritmos, sendo o *K-means* o mais utilizado. Proposto por J. B. MacQueen, o *K-means* caracteriza-se pela facilidade de entendimento, por possuir versões implementadas em diversas ferramentas de Mineração de Dados e por ter passado por várias melhorias ao longo do tempo [18].

Ochi [23] apresenta mais detalhes sobre o uso da tarefa de Clusterização expondo con-

ceitos, algoritmos e algumas das possibilidades de aplicações que incluem células de manufatura, roteamento de veículos, clusterização de grafos, computação médica e biologia computacional.

Um trabalho interessante envolvendo essa tarefa e o algoritmo *K-means* foi realizado pela Universidade do Extremo Sul Catarinense (UNESC), que procurou analisar a incidência de cárie dental na Região Sul do Brasil em crianças de 6 a 12 anos de idade [21]. Para isso, foram considerados os dados das escolas públicas das capitais dos estados do Sul: Curitiba, Florianópolis e Porto Alegre. A análise possibilitou concluir que a maior incidência de cáries em dentes permanentes ocorria nas idades de 8 a 12 anos e em dentes temporários de 6 a 7 anos. Além disso, a maior parte dos casos ocorria no estado de Santa Catarina e em crianças de 6 anos de idade do sexo feminino, levando a necessidade de uma divulgação maior nesse estado para esse público.

Os casos apresentados no decorrer deste Capítulo levam à conclusão de que a Mineração de Dados é uma alternativa eficaz para extração de conhecimento a partir de grandes bases de dados. A aplicação da Mineração de Dados pode ajudar diversas organizações a analisar riscos (durante concessão de crédito, por exemplo), identificar perfis de compra de clientes, detectar fraudes, entre inúmeras outras vantagens. Assim como citado por Han e Kamber [19] e confirmado por Cardoso e Machado [8] em seu trabalho, o conhecimento obtido através da Mineração de Dados é a chave para se tomar decisões baseadas realmente em fatos e não em apenas suposições. Portanto, em um ambiente tão competitivo como o atual, a Mineração de Dados pode ser a solução para a empresa estar a um passo a frente da concorrência.

## 6 REGRAS DE ASSOCIAÇÃO E CLUSTERIZAÇÃO: CONCEITOS BÁSICOS E ALGORITMOS

### 6.1 Regras de Associação

#### 6.1.1 Descrição formal

Formalmente, as regras de associação são descritas da seguinte maneira:  $D$  é uma base de dados constituída por um conjunto de itens  $I = \{i_1, i_2, \dots, i_n\}$  e por um conjunto de transações  $T = \{t_1, t_2, \dots, t_n\}$ . Cada transação  $t_i$  é composta por um conjunto de itens (*itemset*). A regra de associação, portanto, é uma expressão do tipo  $X \Rightarrow Y$  tal que  $X \subset I$ ,  $Y \subset I$ ,  $X \neq \emptyset$ ,  $Y \neq \emptyset$  e  $X \cap Y = \emptyset$ . Nesse caso,  $X$  é denominado antecedente e  $Y$  é denominado conseqüente da regra e ambos representam conjuntos de itens (*itemset*) distintos com pelo menos um item [2] [17].  $X$  ou  $Y$  não podem conter todos os itens de  $I$ , pois, dessa forma, não haveria como formar dois conjuntos de itens distintos não-vazios.

Para cada conjunto de itens, denomina-se comprimento desse conjunto a quantidade de itens pertencentes a ele. Ou seja, se um conjunto possui 20 itens, diz-se que seu comprimento é *20-itemset* e se possui  $k$  itens o comprimento é *k-itemset* [18].

#### 6.1.2 As medidas de Suporte, Confiança e Lift

Conforme mencionando anteriormente, as regras de associação são analisadas com base em medidas de interesse, que podem ser objetivas ou subjetivas. Para este trabalho, serão consideradas as medidas de interesse objetivas de Suporte, Confiança e *Lift*. A seguir, são apresentados os conceitos de cada termo considerando  $X$  o antecedente e  $Y$  o conseqüente da regra.

O suporte de uma regra de associação,  $\text{Sup}(X \Rightarrow Y)$ , representa a porcentagem de transações que contém os itens  $X$  e  $Y$  simultaneamente em relação ao número total de transações [17].

O suporte é dado por:

$$\text{Sup}(X \Rightarrow Y) = \frac{\text{Transações em que X e Y ocorrem simultaneamente}}{\text{Total de transações}}$$

Um exemplo disso seria: "35% do total de transações contém refrigerante e biscoito", onde 35% representa o suporte da regra de associação. Quanto maior o suporte, maior o número de transações em que X e Y estão presentes simultaneamente [39].

A confiança de uma regra de associação, representada por  $\text{Conf}(X \Rightarrow Y)$ , indica a porcentagem de transações que contêm X e Y dentre as que contêm X [17], ou seja:

$$\text{Conf}(X \Rightarrow Y) = \frac{\text{Transações em que X e Y ocorrem simultaneamente}}{\text{Transações em que X ocorre}}$$

Um exemplo prático seria: "Entre os clientes que possuem emprego e imóvel quitado, 75% tem empréstimo aprovado", em que 75% representa a confiança da regra. Quanto maior a confiança, maior a chance de que Y esteja presente em transações em que X está presente [39].

O *lift* indica o quanto mais frequente torna-se Y quando X ocorre [17] e é dado por:

$$\text{Lift}(A \Rightarrow B) = \frac{\text{Conf}(X \Rightarrow Y)}{\text{Sup}(Y)}$$

Como exemplo tem-se: "As pessoas que compram leite tem uma chance 3 vezes maior de também comprar pão". O *lift* avalia o nível de dependência entre os itens. Dessa forma, um resultado de *lift* igual a 1 revela uma independência entre os itens, ou seja, a compra de um item não tem qualquer influência sobre a compra do outro. Caso o valor seja menor que 1, diz-se que há uma dependência negativa entre eles, o que implica que a compra de um item diminui a chance de aquisição do outro. Mas, se por outro lado, o valor for superior a 1, a dependência torna-se positiva, o que significa que a compra de um item eleva a chance de compra do outro. Quanto maiores forem os valores do suporte, da confiança e do *lift*, mais interessante se torna a regra de associação [17].

Para exemplificar esses conceitos e facilitar o entendimento, serão efetuados alguns cálculos e análises de suporte, confiança e *lift* utilizando, para isso, a seguinte base de dados:

**Tabela 2:** Base de dados transacional de um supermercado (extraída de [4])

TID	Lista de Itens
1	pão, açúcar, manteiga
2	pão, leite, açúcar, manteiga, cerveja
3	pão, leite, suco, iogurte
4	pão, leite, açúcar, manteiga, iogurte
5	pão, açúcar, suco, manteiga, fralda, refrigerante
6	leite, refrigerante, iogurte

Como pode-se perceber, cada linha da base de dados representa uma transação de compra e venda que é composta por um conjunto de itens. Esse tipo de base de dados é conhecida como Base de Dados Transacional. Cada uma dessas transações possui um código exclusivo, denominado identificador da transação (TID).

Abaixo seguem alguns exemplos de cálculo e análise de suporte, confiança e *lift*.

**Exemplo 1:**

Dada a regra de associação  $\{\text{pão}\} \Rightarrow \{\text{leite}\}$ , tem-se que:

- O suporte de pão, dado por  $\text{Sup}(\text{pão})$ , é  $\frac{5}{6} \approx 0,83$  ou 83%, pois pão se encontra em 5 das 6 transações.
- O suporte de leite,  $\text{Sup}(\text{leite})$ , é  $\frac{4}{6} \approx 0,66$  ou 66%, uma vez que leite está presente em 4 das 6 transações.
- O suporte da regra,  $\text{Sup}(\text{pão} \Rightarrow \text{leite})$ , é  $\frac{3}{6} = 0,5$  ou 50% pois pão e leite aparecem juntos em 3 das 6 transações.
- A confiança da regra,  $\text{Conf}(\text{pão} \Rightarrow \text{leite})$ , equivale a  $\frac{3}{5} = 0,6$  ou 60%.
- O *lift* da regra,  $\text{Lift}(\text{pão} \Rightarrow \text{leite})$ , corresponde a  $\frac{18}{20} = 0,9$ .

Como pode-se perceber, a confiança é igual a 60%, o que indica que esta é a probabilidade do cliente comprar leite dado que comprou pão. Porém, a chance de qualquer cliente adquirir leite é de 66%, o que indica que a compra de pão diminuiu a probabilidade de um cliente comprar leite, gerando uma dependência negativa entre os itens, conforme mostra o valor do *lift*.

**Exemplo 2:**

A partir da regra de associação  $\{\text{pão}\} \Rightarrow \{\text{manteiga}\}$ , tem-se que:

- O suporte de pão,  $\text{Sup}(\text{pão})$ , é  $\frac{5}{6} \approx 0,83$  ou 83%.
- O suporte de manteiga,  $\text{Sup}(\text{manteiga})$ , é  $\frac{4}{6} \approx 0,66$  ou 66%.
- O suporte da regra,  $\text{Sup}(\text{pão} \Rightarrow \text{manteiga})$ , é  $\frac{4}{6} \approx 0,66$  ou 66%;
- A confiança da regra,  $\text{Conf}(\text{pão} \Rightarrow \text{manteiga})$ , é  $\frac{4}{5} = 0,8$  ou 80%.
- O *lift* da regra,  $\text{Lift}(\text{pão} \Rightarrow \text{manteiga})$ , equivale a  $\frac{24}{20} = 1,2$ .

Nesse caso, a confiança equivale a 80%, o que demonstra que esta é a probabilidade de um cliente comprar manteiga dado que comprou pão. Tendo em vista que a chance de qualquer cliente comprar manteiga é 66%, conclui-se que a compra de pão aumentou a compra de manteiga.

O valor obtido para o *lift* demonstra esse aumento, o que revela a existência de uma dependência positiva entre os itens.

**Exemplo 3:**

Para a regra de associação  $\{\text{refrigerante}\} \Rightarrow \{\text{iogurte}\}$ , conclui-se que:

- O suporte de refrigerante,  $\text{Sup}(\text{refrigerante})$ , é  $\frac{2}{6} \approx 0,33$  ou 33%;
- O suporte de iogurte,  $\text{Sup}(\text{iogurte})$ , é  $\frac{3}{6} = 0,5$  ou 50%;
- O suporte da regra,  $\text{Sup}(\text{refrigerante} \Rightarrow \text{iogurte})$ , corresponde a  $\frac{1}{6} \approx 0,16$  ou 16%;
- A confiança da regra,  $\text{Conf}(\text{refrigerante} \Rightarrow \text{iogurte})$ , equivale a  $\frac{1}{2} = 0,5$  ou 50%.
- O *lift* da regra,  $\text{Lift}(\text{refrigerante} \Rightarrow \text{iogurte})$ , é  $\frac{6}{6} = 1$ .

A confiança diz que em 50% das compras de refrigerante também houve compra de iogurte, mas em 50% das transações em geral o iogurte foi comprado, ou seja, a compra de um produto não influenciou na compra do outro. O valor do *lift* confirma essa relação de independência existente entre os itens.

### 6.1.3 Modelo Suporte-Confiança

O primeiro modelo usado para mineração de regras de associação ficou conhecido como Modelo Suporte-Confiança. Proposto por Agrawal, Imielinski e Swami [2], ele consiste em obter todas as regras que possuam valores de suporte e confiança maiores ou iguais a um suporte e confiança mínimos. Tais valores mínimos de suporte e confiança são especificados pelo próprio usuário.

No Modelo Suporte-Confiança, o processo de mineração de regras de associação é dividido em duas fases [17]:

- A primeira delas consiste em encontrar todos os conjuntos de itens cujo valor de suporte seja maior ou igual ao suporte mínimo. Esses conjuntos, denominados frequentes, possuem um tamanho que varia de 1 a  $m$ , em que  $m$  representa o total de itens da base de dados;
- A segunda etapa visa gerar, para cada conjunto identificado na etapa anterior, as regras de associação cujo valor de confiança seja maior ou igual à confiança mínima.

Esse modelo foi criticado em virtude do alto número de combinações que podiam ser geradas na etapa 1. Após a geração de todas essas combinações, o suporte de cada conjunto ainda deve ser calculado de forma a atender o suporte mínimo especificado. Em seguida, para cada conjunto frequente identificado, as regras de associação ainda devem ser geradas de forma a também atender a confiança mínima estipulada.

Para uma dada quantidade de itens, o número de combinações é determinado pela expressão  $2^m - 1$ , em que  $m$  corresponde ao número de itens da base de dados. Gonçalves [18] confirma e exemplifica isso ao mostrar que para um conjunto de quatro itens {café, leite, manteiga, pão} podem ser geradas 15 combinações diferentes:

**Tabela 3:** Combinações geradas no Modelo Suporte-Confiança para um conjunto de 4 itens

1 item	2 itens	3 itens	4 itens
{café}	{café, leite}	{café, leite, manteiga}	{café, leite, manteiga, pão}
{leite}	{café, manteiga}	{café, leite, pão}	
{manteiga}	{café, pão}	{café, manteiga, pão}	
{pão}	{leite, manteiga}	{leite, manteiga, pão}	
	{leite, pão}		
	{manteiga, pão}		

Considerando o fato de que as empresas trabalham com grande quantidade de itens, a prática de um modelo desse gênero torna-se extremamente inviável. Para 100 itens, por exemplo, poderiam ser geradas aproximadamente  $1,26 \times 10^{30}$  combinações diferentes, o que torna as instruções SQL altamente complexas em virtude do grande número de combinações de diferentes tamanhos que podem ser formadas (de tamanho 1 a 100). Para resolver esse problema, um algoritmo foi proposto e será apresentado a seguir.

#### 6.1.4 O algoritmo *Apriori*

Como afirmado na Seção 5.3, o *Apriori* é o algoritmo mais utilizado para minerar regras de associação. Seu surgimento se deu no intuito de resolver problemas decorrentes do Modelo Suporte-Confiança.

O uso do *Apriori* permitiu reduzir significativamente o número de combinações a serem realizadas bem como a contagem de suporte utilizando, para isso, uma propriedade denominada *apriori*. Essa propriedade afirma que um conjunto que possui um subconjunto infrequente, também é infrequente. Denomina-se conjunto infrequente aquele que possui suporte menor que o suporte mínimo estipulado pelo usuário [18]. Por exemplo, se o subconjunto {café, pão} é infrequente, então os conjuntos {café, pão, açúcar} e {café, pão, leite, biscoito} também são. Com isso, tem-se uma diminuição no tempo de execução do algoritmo, pois se, por exemplo, o subconjunto {café, pão} não é frequente, não há necessidade de percorrer a base de dados novamente para calcular o suporte dos conjuntos {café, pão, açúcar} e {café, pão, leite, biscoito}.

Para exemplificar o funcionamento do *Apriori*, será utilizada a base de dados a seguir:



**Tabela 4:** Base de dados transacional que será processada pelo algoritmo *Apriori* (adaptada de [5])

TID	Lista de Itens
1	a, b, e
2	b, d
3	b, c
4	a, b, d
5	a, c
6	b, c
7	a, c
8	a, b, c, e
9	a, b, c

A Figura abaixo exibe o pseudocódigo do *Apriori*. Como é possível perceber, o algoritmo recebe uma base de dados ( $D$ ) e um suporte mínimo ( $SupMin$ ) como parâmetros de entrada.

```

Entrada:
1   $D$  – base de dados transacional
2   $SupMin$  – suporte mínimo

Algoritmo:
1   $F_1 = \{ \text{Conjunto de itemsets frequentes de tamanho 1} \}$ 
2  para ( $k=2; F_{k-1} \neq \emptyset; k++$ )
3       $C_k = \text{apriori-gen}(F_{k-1})$ 
4      para todas as transações  $t$  faça
5           $C_t = \text{subset}(C_k, t)$ 
6          para todos os candidatos  $c$  em  $C_t$  faça
7               $c.\text{contagem}++$ 
8          fim para todos
9      fim para todas
10      $F_k = \{ c \text{ em } C_k \mid c.\text{contagem} \geq SupMin \}$ 
11 fim para
12 Saída: união de todos os conjuntos  $F_k$ 

```

**Figura 7:** Pseudocódigo do algoritmo *Apriori* (adaptado de [3] e [18])

O primeiro passo consiste em encontrar todos os conjuntos frequentes com 1 item ( $F_1$ ), que são aqueles que atendem ao valor mínimo de suporte. Nesse exemplo, serão determinados os conjuntos frequentes que ocorram em pelo menos 2 das 9 transações (suporte mínimo igual a 2 ou aproximadamente 22%).

A Tabela abaixo mostra os conjuntos obtidos após a pesquisa realizada na base de dados:

**Tabela 5:** Conjuntos frequentes de tamanho 1

$F_1$				
{a}	{b}	{c}	{d}	{e}
6	7	6	2	2

Como todos os conjuntos satisfazem o valor mínimo de suporte, o algoritmo passa para a etapa seguinte que consiste em determinar os conjuntos candidatos de tamanho 2 ( $k=2$ ). Desse momento em diante, o *Apriori* utiliza os conjuntos frequentes da etapa anterior ( $F_{k-1}$ ) para gerar os conjuntos candidatos da etapa atual [5], ou seja, os conjuntos frequentes de tamanho 1 serão usados para produzir os conjuntos candidatos de tamanho 2. Esse passo é executado até que o conjunto anterior seja um conjunto vazio (linha 2).

A geração dos conjuntos candidatos de tamanho  $k$  é feita através de uma operação de junção entre os conjuntos frequentes da etapa anterior [18]. Esse processo é executado pela função *apriori-gen* (linha 3), que recebe como parâmetro o conjunto frequente da etapa anterior ( $F_{k-1}$ ) e retorna o conjunto candidato da etapa atual ( $C_k$ ) [5]. Dessa forma,  $C_2$ , será produzido pela junção  $F_1 \times F_1$ , conforme mostra a Tabela a seguir:

**Tabela 6:** Conjuntos candidatos de tamanho 2

$C_2(F_1 \times F_1)$									
{a, b}	{a, c}	{a, d}	{a, e}	{b, c}	{b, d}	{b, e}	{c, d}	{c, e}	{d, e}

Posteriormente, a base de dados é novamente percorrida a fim de obter o suporte de cada um dos conjuntos candidatos formados anteriormente. A função encarregada dessa tarefa é a *subset* (linha 5). Ela recebe como parâmetro o conjunto  $C_k$  e as transações da base de dados. Para cada subconjunto de  $C_k$ , a função verifica se ele está presente em uma transação e, caso

isso ocorra, incrementa o contador de suporte (linha 7) [5]. A Tabela 7 mostra o resultado da contagem de suporte:

**Tabela 7:** Conjuntos frequentes de tamanho 2

$F_2$									
{a, b}	{a, c}	{a, d}	{a, e}	{b, c}	{b, d}	{b, e}	{c, d}	{c, e}	{d, e}
4	4	1	2	4	2	2	0	1	0

Como foram encontrados conjuntos com suporte menor que 2, estes são removidos de  $F_2$ .

Na etapa seguinte, o *Apriori* determina os conjuntos candidatos de tamanho 3 ( $k=3$ ), efetuando a junção  $F_2 \times F_2$ . Porém, nesse momento a junção ocorre de uma forma diferente. Ela somente será possível entre conjuntos candidatos que compartilhem o mesmo prefixo de tamanho  $k-2$  [18]. Para simplificar, serão considerados dois exemplos:

- Considerando que  $k=3$ , a junção é possível entre os conjuntos candidatos {a, b} e {a, c}, uma vez que estes compartilham o mesmo prefixo de tamanho  $k-2$ , que é o prefixo {a}.
- A junção não é possível entre os conjuntos candidatos {a, b} e {b, c}, já que não compartilham o mesmo prefixo.

A Tabela a seguir exibe os conjuntos candidatos obtidos após a aplicação da operação de junção:

**Tabela 8:** Conjuntos candidatos de tamanho 3

$C_3(F_2 \times F_2)$					
{a, b, c}	{a, b, e}	{a, c, e}	{b, c, d}	{b, c, e}	{b, d, e}

Observando a Tabela 7, pode-se perceber que os subconjuntos {a, d}, {c, d}, {c, e} e {d, e} são infreqüentes. Utilizando o conceito da propriedade *apriori*, conclui-se que o mesmo é válido para os conjuntos {a, c, e}, {b, c, d}, {b, c, e} e {b, d, e} da Tabela 8. Diante disso, a operação de poda, também realizada pela função *apriori-gen* (linha 3), é aplicada sobre eles:

**Tabela 9:** Conjuntos candidatos de tamanho 3 após aplicação da operação de poda

$C_3(F_2 \times F_2)$					
{a, b, c}	{a, b, e}	{a, c, e}	{b, c, d}	{b, c, e}	{b, d, e}
		Podado	Podado	Podado	Podado

Para cada um dos conjuntos candidatos restantes, o algoritmo realiza a verificação do suporte, conforme mostra a Tabela 10:

**Tabela 10:** Conjuntos frequentes de tamanho 3

$F_3$	
{a, b, c}	{a, b, e}
2	2

Como todos os conjuntos atendem ao valor de suporte mínimo, o algoritmo percorre novamente a base de dados a fim de gerar os conjuntos candidatos de tamanho 4 ( $k=4$ ). A junção  $F_3 \times F_3$  é realizada utilizando a mesma linha de raciocínio anterior:

**Tabela 11:** Conjuntos candidatos de tamanho 4

$C_4(F_3 \times F_3)$
{a, b, c, e}

Nesse momento a propriedade *apriori* e a técnica de poda são novamente aplicadas, eliminando o conjunto {a, b, c, e} por conter subconjuntos infrequentes como mostram as Tabelas 7 e 9. O resultado após a aplicação da operação de poda é exibido na Tabela seguinte:

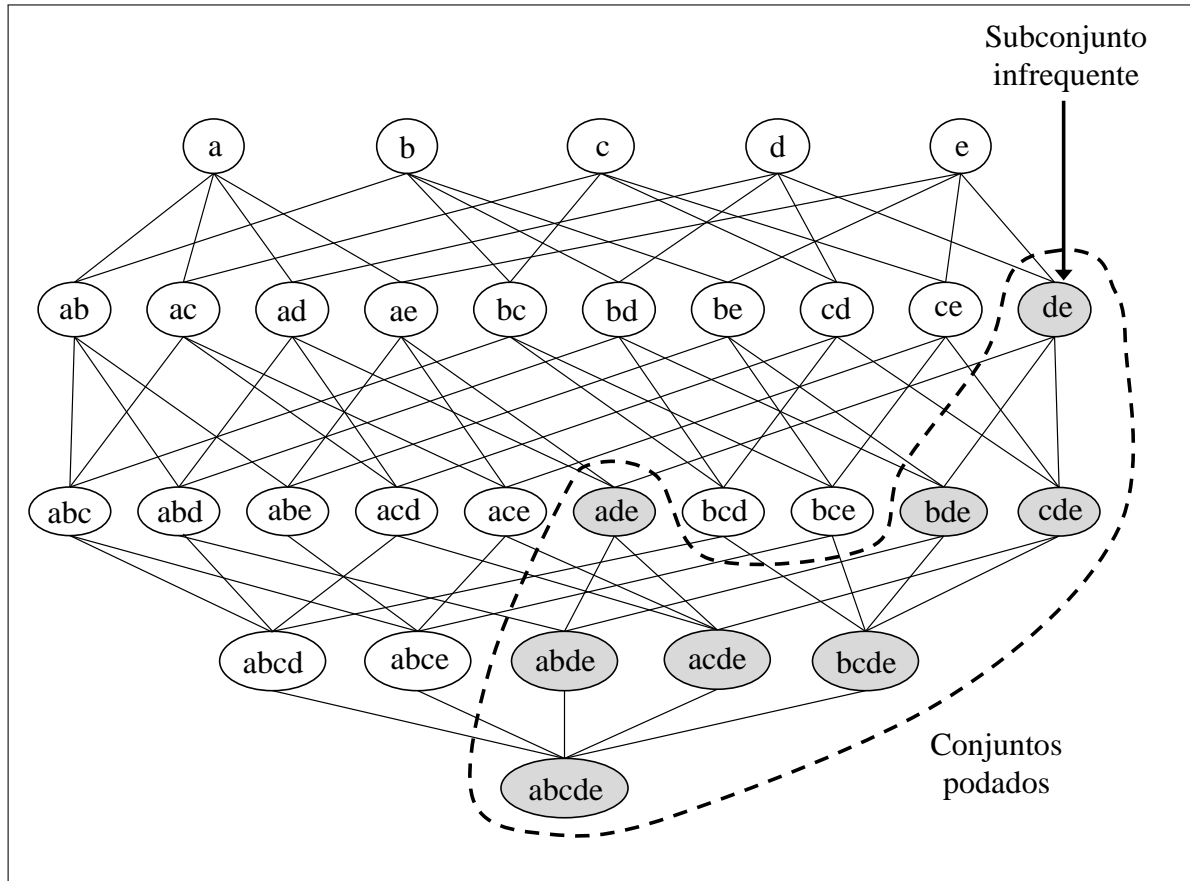
**Tabela 12:** Conjuntos candidatos de tamanho 4 após aplicação da operação de poda

$C_4(F_3 \times F_3)$
{a, b, c, e}
Podado

Como a lista de conjuntos candidatos de tamanho 4 torna-se vazia ( $C_4=\emptyset$ ), logo  $F_4=\emptyset$  e, dessa forma, ao tentar produzir  $C_5$ , o *Apriori* encerra o processo de geração de conjuntos. Logo,

$F_1$ ,  $F_2$  e  $F_3$  constituem os conjuntos frequentes obtidos com a aplicação do algoritmo (saída do algoritmo).

A Figura a seguir demonstra como ocorre o processo de geração de conjuntos candidatos e aplicação da operação de poda sobre eles.



**Figura 8:** Poda de conjuntos candidatos usando a propriedade *apriori* (adaptado de [39])

Como o processo de geração de conjuntos no algoritmo *Apriori* ocorre em níveis, um subconjunto infrequente de um nível elimina os conjuntos do nível abaixo que contenham esse subconjunto [39]. Portanto, se o subconjunto  $\{d, e\}$  é infrequente e aplicando o conceito da propriedade *apriori*, os conjuntos  $\{a, d, e\}$ ,  $\{b, d, e\}$  e  $\{c, d, e\}$  também são. Consequentemente, este subconjunto leva à poda dos conjuntos do nível logo abaixo dele até que o último nível seja atingido.

Uma vez encerrada a fase de determinação dos conjuntos frequentes, o próximo passo é gerar as regras de associação. Cada conjunto de  $k$  itens frequentes, pode produzir até  $2^k - 2$  regras de associação [39]. Por exemplo, para o conjunto  $\{a, b, c, d\}$ , em que  $k=4$ , podem ser geradas 14 regras de associação:

**Tabela 13:** Regras de associação candidatas para um conjunto de 4 itens

Nº de itens no conseqüente		
1 item	2 itens	3 itens
$abc \Rightarrow d$	$ab \Rightarrow cd$	$a \Rightarrow bcd$
$abd \Rightarrow c$	$ac \Rightarrow bd$	$b \Rightarrow abd$
$acd \Rightarrow b$	$ad \Rightarrow bc$	$c \Rightarrow abd$
$bcd \Rightarrow a$	$bc \Rightarrow ad$	$d \Rightarrow abc$
	$bd \Rightarrow ac$	
	$cd \Rightarrow ab$	

De forma sucinta, a geração de regras no algoritmo *Apriori* ocorre da seguinte maneira: para cada conjunto frequente  $f \subseteq I$ , o algoritmo localiza todos os subconjuntos  $f'$  de  $f$ . Então, para cada subconjunto  $f'$  de  $f$  encontrado, é produzida uma regra de associação na forma  $f' \Rightarrow (f - f')$ , desde que  $\text{Sup}(f) \div \text{Sup}(f')$  seja maior ou igual à confiança mínima estabelecida [5] [30]. Como exemplo, pode-se citar o conjunto frequente  $f = \{a, b\}$  e o subconjunto  $f' = \{a\}$ . Nesse caso, pode ser produzida a regra de associação  $a \Rightarrow b$ , onde  $\{a\} \Rightarrow \{a, b\} - \{a\}$ , contanto que a confiança mínima seja atingida, ou seja,  $\text{Conf}(a \Rightarrow b) \geq \text{Sup}(a, b) \div \text{Sup}(a)$ .

No processo de geração de regras de associação, o *Apriori* também utiliza uma abordagem em níveis, em que cada nível possui as regras de acordo com o número de itens do conseqüente [39]. Ou seja, no primeiro nível estão todas as regras de associação que possuem apenas um item no conseqüente. A partir dessas regras, novas regras são geradas contendo dois itens no conseqüente, posteriormente três itens e assim por diante.

Assim como no processo de geração de conjuntos candidatos e frequentes, a etapa de produção de regras de associação também está sujeita a aplicação da operação de poda. Para aplicar a técnica de poda com base na confiança das regras, o *Apriori* baseia-se no seguinte teorema:

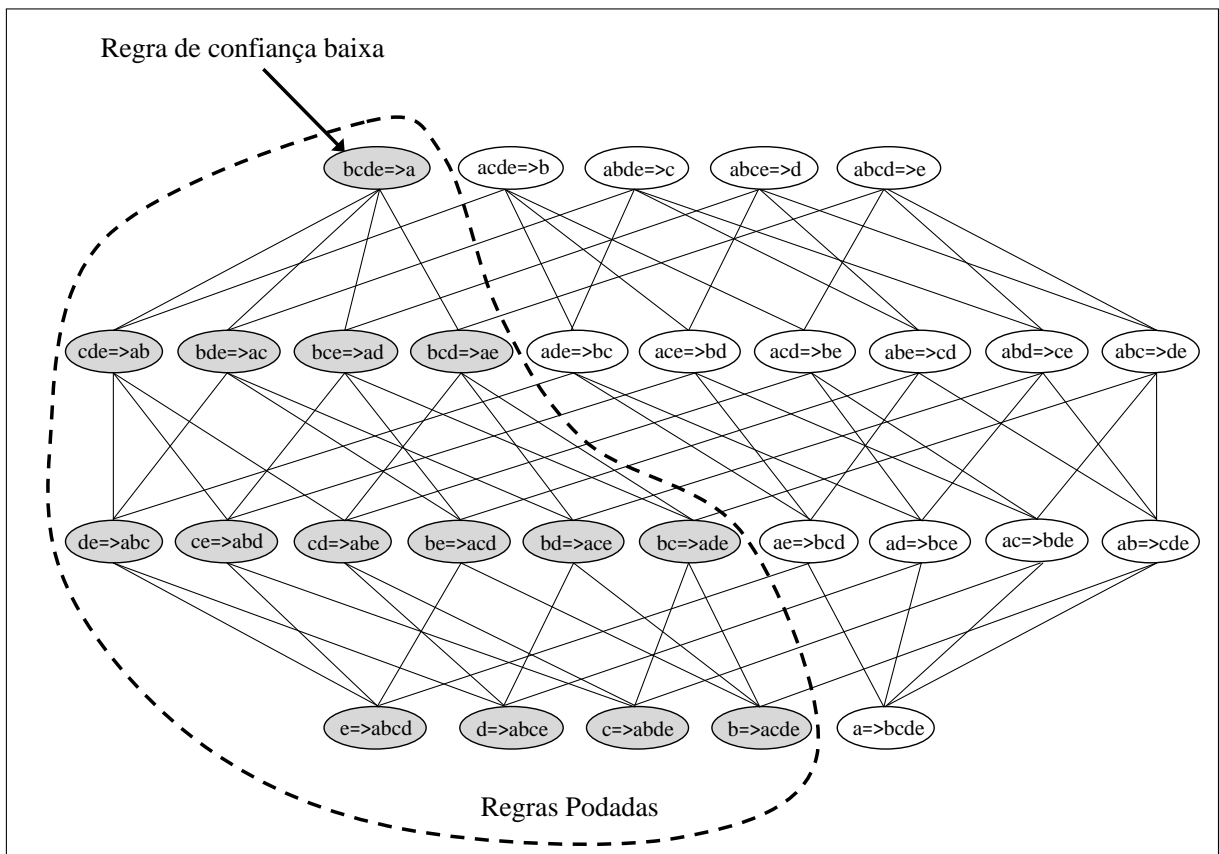
**Teorema 1.** Se uma regra  $X \Rightarrow Y - X$  não atender o limite de confiança, então qualquer regra  $X' \Rightarrow Y - X'$ , em que  $Y$  representa todos os itens da base de dados e  $X'$  é um subconjunto de  $X$ , não deve atender o limite de confiança também [39].

Para simplificar esse teorema, basta considerar as regras  $X' \Rightarrow Y - X'$  e  $X \Rightarrow Y - X$ . Considerando  $Y$  como sendo o conjunto de todos os itens da base de dados, a confiança dessas regras é dada, respectivamente, por  $\text{Sup}(Y) \div \text{Sup}(X')$  e  $\text{Sup}(Y) \div \text{Sup}(X)$ . Tendo em vista que  $X'$  é um subconjunto de  $X$ , pode-se concluir que  $\text{Sup}(X') \geq \text{Sup}(X)$  e, portanto, a confiança da regra

$X' \Rightarrow Y - X'$  não pode ser maior que a da regra  $X \Rightarrow Y - X$  [39].

Um exemplo prático seria considerar  $Y = \{a, b, c, d, e\}$ ,  $X = \{a, b\}$  e  $X' = \{a\}$ . Nesse caso, a regra  $X \Rightarrow Y - X$  seria dada por  $ab \Rightarrow cde$  e a regra  $X' \Rightarrow Y - X'$  poderia ser representada como  $a \Rightarrow bcde$ , uma vez que  $X'$  é um subconjunto de  $X$ . Se a primeira regra não atender o limite de confiança, então a segunda também não atenderá. Isso ocorre porque o suporte de  $X'$  (denominador da segunda regra) é sempre maior ou igual ao suporte de  $X$  (denominador da primeira regra), o que implica que a divisão da segunda regra (dada por  $\text{Sup}(Y) \div \text{Sup}(X')$ ) terá um resultado (valor de confiança) menor ou igual ao da divisão da primeira regra (que é  $\text{Sup}(Y) \div \text{Sup}(X)$ ).

A Figura a seguir demonstra como esse processo ocorre:



**Figura 9:** Poda de regras de associação usando a medida de confiança (adaptado de [39])

Para simplificar, considera-se que a regra  $\{bcde\} \Rightarrow \{a\}$  não satisfaz o valor mínimo de confiança. Pela aplicação do teorema, conclui-se que todas as demais regras que possuam algum subconjunto do antecedente dessa regra podem ser podadas, uma vez que também não atenderão ao valor mínimo de confiança. Isso ocorre pois, conforme já demonstrado anteriormente, o suporte de cada subconjunto de  $bcde$  ( $cde$ ,  $bde$ ,  $bce$ ,  $bcd$  e demais conforme mostra a Figura 9)

é sempre maior ou igual ao do próprio conjunto. Dessa forma, as regras que possuem algum subconjunto de bcde no antecedente terão sempre valor de confiança menor ou igual ao da regra cujo antecedente é o conjunto bcde.

## 6.2 Clusterização

### 6.2.1 Descrição formal

De acordo com Ochi [23], a Clusterização pode ser descrita formalmente da seguinte maneira: dado um conjunto  $Y = \{Y_1, Y_2, Y_3, \dots, Y_n\}$  composto  $n$  elementos, a Clusterização consiste em obter um conjunto  $C = \{C_1, C_2, C_3, \dots, C_k\}$  formado por  $k$  *clusters*, de forma que os elementos de um *cluster*  $C_i$  tenham mais semelhança entre si do que com os elementos dos outros *clusters* em  $C$ . Somente há uma Clusterização se forem atendidas as seguintes condições:

$$C_i \neq \emptyset, \text{ para } 1 \leq i \leq k$$

$$C_i \cap C_j = \emptyset, \text{ para } 1 \leq i, j \leq k \text{ e } i \neq j$$

$$\bigcup_{i=1}^k C_i = Y$$

De acordo com as propriedades acima, uma clusterização somente será válida desde que, dado um conjunto inicial, cada *cluster* contenha pelo menos um elemento deste conjunto (primeira propriedade), de forma que um mesmo elemento jamais esteja presente em dois ou mais *clusters* distintos (segunda propriedade). Considerando que a clusterização deve abranger todos os elementos de um conjunto sem que haja repetição dos mesmos, a união de todos os grupos formados deve ser igual ao conjunto inicial (terceira propriedade).

### 6.2.2 O algoritmo *K-means*

Como mencionado na Seção 5.4, o algoritmo de Clusterização mais utilizado para encontrar grupos com elementos similares é o *K-means*. Diante disso, nessa Subseção será simulado o seu processo de execução utilizando a seguinte base de dados bidimensional<sup>1</sup>:

---

<sup>1</sup>Considera-se dimensão cada atributo em uma base de dados.



**Tabela 14:** Base de dados que será processada pelo algoritmo *K-means* (extraída de [18])

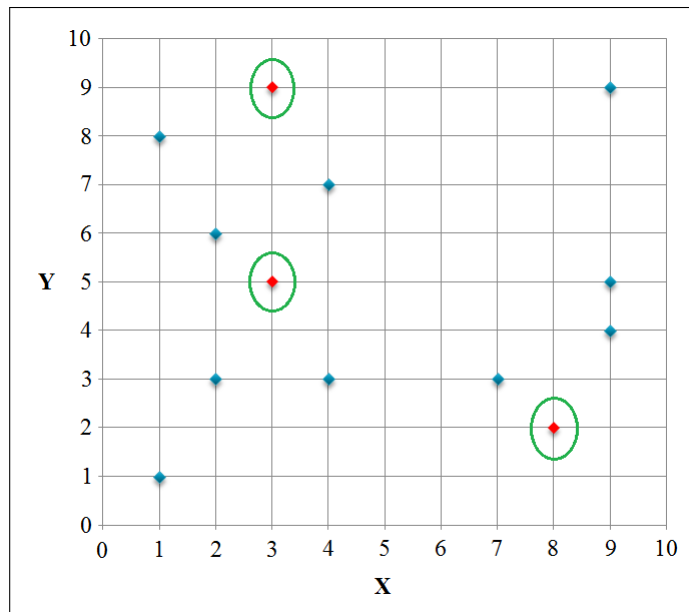
X	Y
3	9
3	5
8	2
1	1
9	4
9	5
1	8
2	3
4	3
7	3
2	6
9	9
4	7

A base de dados a ser utilizada deverá ser passada como parâmetro de entrada, assim como o número de grupos que o usuário deseja formar e a quantidade máxima de iterações do algoritmo. Para essa simulação, será adotado o número de *clusters* igual a 3 ( $k=3$ ).

Existem algumas medidas úteis para determinar a similaridade entre os elementos. Nesse caso, será utilizada a distância euclidiana. Esta medida considera a distância entre dois elementos  $X_i$  e  $X_j$  em um espaço  $n$ -dimensional. Dessa forma, quanto menor a distância encontrada entre dois elementos  $X_i$  e  $X_j$ , maior será o nível de semelhança entre eles [23]. A distância euclidiana é dada pela seguinte fórmula:

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{in} - x_{jn})^2}$$

O processo de execução do *K-means* é basicamente dividido em 3 etapas. A primeira delas consiste em escolher o centro inicial de cada *cluster*, também denominado centróide ou semente. Como  $k=3$ , serão escolhidos 3 centróides iniciais. No algoritmo original, estes centros eram sempre os três primeiros registros da base de dados. Mas hoje existem versões do *K-means* que trabalham de forma diferente selecionando os centróides até mesmo aleatoriamente [18]. Na Figura 10, os pontos vermelhos representam os centros iniciais dos *clusters*. Para este exemplo, serão adotados como centros iniciais os três primeiros registros da base de dados.



**Figura 10:** Escolha dos centros iniciais (adaptado de [18])

A segunda etapa consiste em associar cada um dos demais registros a um centro inicial. Isso é feito medindo a distância de cada um deles aos centróides. Dessa forma, cada registro será associado ao *cluster* cujo centróide estiver mais próximo dele [18]. Para exemplificar, será considerado o registro (9, 4). Calculando a distância entre ele e cada centro inicial, tem-se:

- Do registro (9,4) ( $x_i$ ) ao centróide (3,9) ( $x_j$ ):

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2} = \sqrt{(9 - 3)^2 + (4 - 9)^2} = \sqrt{36 + 25} = \sqrt{61} \approx 7,81$$

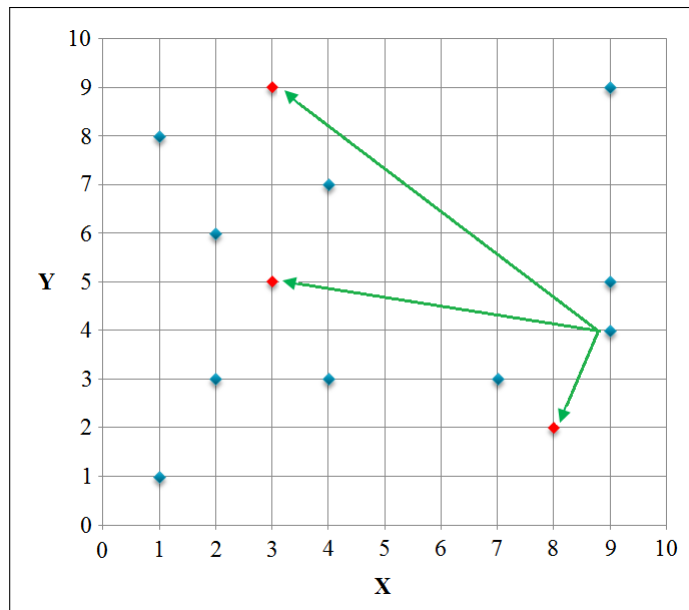
- Do registro (9,4) ( $x_i$ ) ao centróide (3,5) ( $x_j$ ):

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2} = \sqrt{(9 - 3)^2 + (4 - 5)^2} = \sqrt{36 + 1} = \sqrt{37} \approx 6,08$$

- Do registro (9,4) ( $x_i$ ) ao centróide (8,2) ( $x_j$ ):

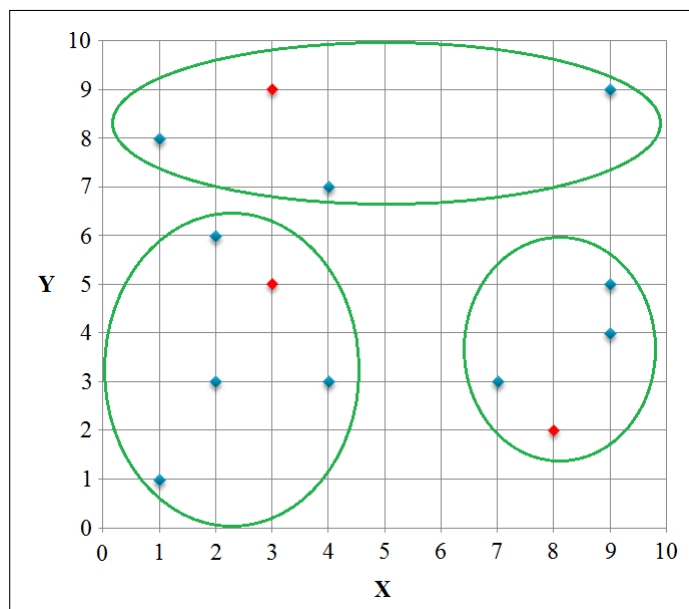
$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2} = \sqrt{(9 - 8)^2 + (4 - 2)^2} = \sqrt{1 + 4} = \sqrt{5} \approx 2,23$$

Como pode-se perceber, o registro (9,4) está mais próximo do *cluster* cujo centro é (8,2), conforme mostra a Figura 11:



**Figura 11:** O objeto (9,4) é associado ao *cluster* cujo centro é (8,2) (adaptado de [18])

Esse mesmo cálculo é efetuado para todos os demais registros, sendo obtido o seguinte resultado ao final dessa segunda etapa:



**Figura 12:** *Clusters* obtidos após o cálculo de similaridade (adaptado de [18])

Observando a Figura 12, é possível perceber que foram formados três *clusters* distintos, os quais são constituídos pelos seguintes pontos:

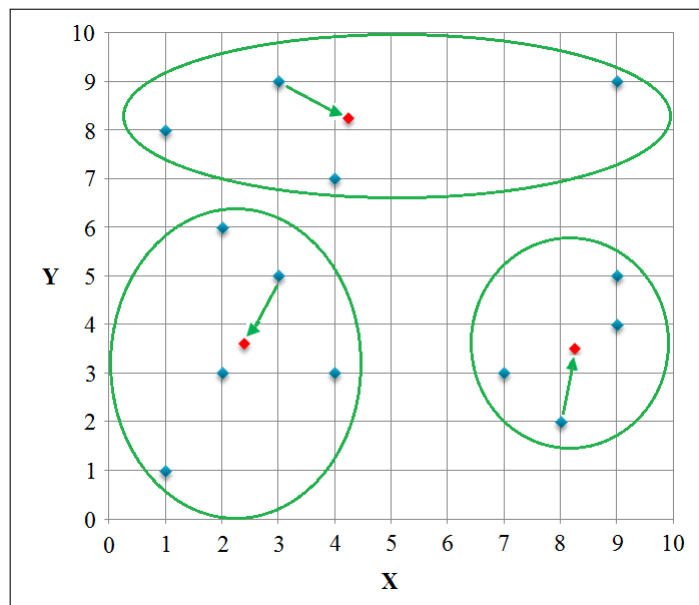
- $C_1 = (1,1), (2,3), (2,6), (3,5)$  e  $(4,3)$ ;

- $C_2 = (1,8), (3,9), (4,7)$  e  $(9,9)$ ;
- $C_3 = (7,3), (8,2), (9,4)$  e  $(9,5)$ .

Um vez associados os registros aos seus centróides, o algoritmo passa para a terceira etapa que consiste em recalculer os centros dos *clusters*, a fim de verificar se houve movimentação dos mesmos. Isso é feito de forma simples, utilizando a média aritmética dos valores das coordenadas dos pontos [18].

- $C_1 = \frac{1}{5} \times (1+2+2+3+4, 1+3+6+5+3) = \frac{1}{5} \times (12,18) = (\frac{12}{5}, \frac{18}{5}) = (2.4, 3.6)$
- $C_2 = \frac{1}{4} \times (1+3+4+9, 8+9+7+9) = \frac{1}{4} \times (17,33) = (\frac{17}{4}, \frac{33}{4}) = (4.25, 8.25)$
- $C_3 = \frac{1}{4} \times (7+8+9+9, 3+2+4+5) = \frac{1}{4} \times (33,14) = (\frac{33}{4}, \frac{14}{4}) = (8.25, 3.5)$

Os resultados comprovam que os centros dos *clusters* se movimentaram ao final do processo. A Figura a seguir demonstra o ocorrido, em que os pontos vermelhos representam a posição dos centróides após se movimentarem:



**Figura 13:** Os centros dos *clusters* se modificaram após a primeira iteração do algoritmo (adaptado de [18])

Diante disso, o algoritmo retorna para a segunda etapa e reassocia cada registro a um centróide usando os novos valores obtidos. Caso os centros dos *clusters* não tivessem se movimentado, o *K-means* encerraria sua execução. Mas, se esses centros nunca parassem de se mover, a execução do algoritmo seria encerrada após atingir o número máximo de iterações fornecido como parâmetro pelo usuário no início do processo [18].

A seguir, é exibido o pseudo-código do algoritmo:

<b>Entrada:</b>	
1	$D$ – base de dados n-dimensional
2	$K$ – número de grupos a serem formados
3	$MAX$ – número máximo de iterações do algoritmo
<b>Algoritmo:</b>	
1	Escolher $K$ centros de <i>clusters</i> iniciais $C_1, C_2, \dots, C_n$ entre os $n$ registros de $D$ ;
2	Alocar cada ponto $x_i$ ( $i = 1, 2, \dots, n$ ) ao <i>cluster</i> cujo centro estiver mais próximo, utilizando alguma medida de distância;
3	Computar os novos centros de <i>clusters</i> iniciais $C_1, C_2, \dots, C_n$ utilizando a média aritmética;
4	<b>se</b> ( <i>os novos centros de clusters são iguais aos anteriores</i> ) <b>ou</b> ( <i>já ocorreram mais do que</i>
	$MAX$ iterações) <b>então</b>
5	terminar execução
6	<b>senão</b>
7	retornar ao passo 2
8	<b>fim se</b>

**Figura 14:** Pseudocódigo do algoritmo *K-means* (extraído de [18])

## 7 ESTUDO DE CASO DO VESTIBULAR DA UFVJM

A UFVJM realizava, a cada seis meses, o processo seletivo unificado (vestibular tradicional) para provimento de vagas nos seus cursos de graduação. Esta era, até então, a única forma de ingresso na instituição. O órgão da universidade responsável pelo gerenciamento desses processos seletivos é a COPESE. No momento da inscrição, cada candidato devia, além de informar seus dados pessoais, responder um questionário sócio-econômico, criando-se, assim, bases de dados sobre esses candidatos. A resposta a essas questões poderia permitir conhecer melhor o perfil dos candidatos ao vestibular em função de diversos aspectos.

Com o acúmulo cada vez maior dos dados ao longo do tempo, tornou-se extremamente difícil obter conhecimentos que permitissem identificar o perfil dos ingressantes sem o uso de técnicas e ferramentas computacionais apropriadas. Tal análise é de fundamental importância, uma vez que pode permitir a instituição obter conhecimentos úteis para auxiliar na tomada de decisões.

Diante dessa situação, percebeu-se a necessidade de se utilizar ferramentas e técnicas de Mineração de Dados com o propósito de obter o conhecimento implícito nessa grande quantidade de dados.

Após definidos o objetivo e a fonte de dados a ser analisada, foram selecionados o *hardware* e o *software* a serem utilizados. Dessa forma, todo o trabalho foi realizado em um *notebook* com processador Intel *Core i7* de 2.0 *gigahertz*, 6 *gigabytes* de memória RAM e disco rígido de 640 *gigabytes*. O sistema operacional utilizado foi o *Windows 7 Home Premium*. Em razão dos fatores já mencionados na Seção 4.1, as aplicações específicas selecionadas incluíram o Sistema de Gerenciamento de Banco de Dados MySQL 5.2 e a ferramenta de Mineração de Dados WEKA 3.6.

Ao longo dos tópicos seguintes será apresentado, detalhadamente, como foi realizada cada etapa do processo.

## **7.1 Seleção dos dados**

Primeiramente, foram coletados os dados julgados mais relevantes. Como cada base de dados possuía muitas tabelas e atributos que não teriam grande importância para o objetivo em questão, optou-se por utilizar somente aquelas que guardam os dados pessoais e sócio-econômicos dos candidatos, bem como aspectos relacionados à aprovação ou reprovação destes candidatos. A descrição de cada uma das tabelas que foram selecionadas nessa etapa pode ser visualizada na Seção 4.2.

## **7.2 Limpeza dos dados**

Nesta etapa, alguns problemas identificados como valores ausentes ou duplicados foram corrigidos. Para os campos em branco foi atribuído um valor padrão, sendo este tratado como desconhecido e, portanto, desconsiderado na análise. Já os registros duplicados foram removidos.

## **7.3 Transformação dos dados**

Para que os dados fossem submetidos à Mineração de Dados, algumas adaptações foram realizadas. Conforme foi mencionado na Seção 4.2, a tabela que armazena as respostas do questionário possui várias colunas, onde cada uma corresponde a uma questão e armazena a respectiva resposta em formato numérico. Para facilitar o processo de análise dos resultados, foi criada uma tabela para cada pergunta do questionário que passou a armazenar a resposta de forma descritiva.

A Tabela a seguir exemplifica como foi relacionada a pergunta "Você frequentou ou frequenta curso pré-vestibular?" aos respectivos dados:

**Tabela 15:** Exemplo de associação das respostas do questionário socioeconômico para a pergunta: "Você frequentou ou frequenta curso pré-vestibular?"

Frequencia PreVestibular	
id	resposta
1	Não
2	Sim, por menos de um semestre
3	Sim, por um semestre
4	Sim, por um ano
5	Sim, por mais de um ano
6	Sim, integrado ao Ensino Médio (2º grau)

Um procedimento similar a este foi realizado para a tabela de notas com a adoção de conceitos descritivos (péssimo, regular, bom, muito bom, ótimo) ao invés de valores numéricos de acordo com o aproveitamento obtido em cada disciplina.

Uma vez concluído, foi criada uma visão para relacionar os dados das tabelas selecionadas. É essa visão que deu entrada na ferramenta de Mineração de Dados escolhida.

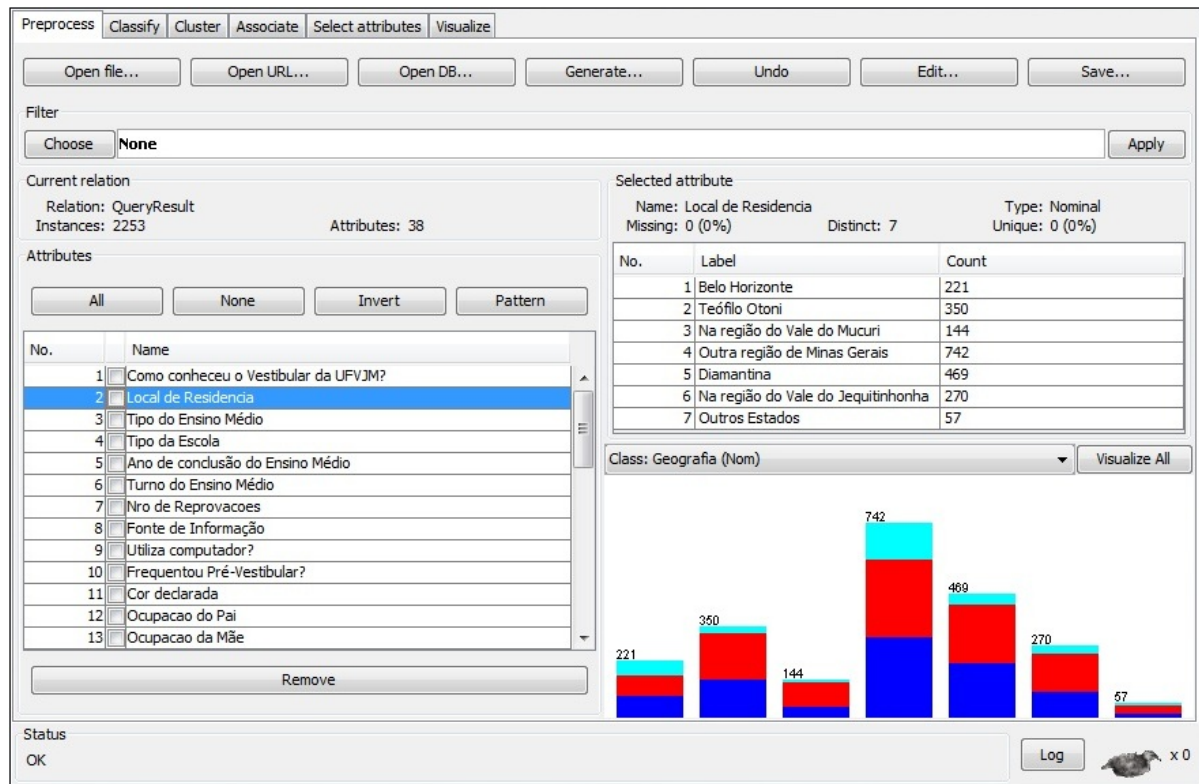
#### 7.4 Mineração de dados

Antes de dar início à Mineração de Dados, foram definidas as tarefas e algoritmos que seriam utilizados. Por razões já mencionadas nas seções 5.3 e 5.4, foram utilizadas as tarefas de Regras de Associação e Clusterização, bem como, respectivamente, os algoritmos *Apriori* e *K-means*.

Como citado anteriormente, a ferramenta de Mineração de Dados escolhida foi o WEKA. Para estabelecer uma conexão entre ela e o MySQL, foi utilizado um driver JDBC, uma vez que o WEKA foi totalmente desenvolvido em linguagem Java. A tarefa de configuração da conexão foi realizada através da opção *Open DB* localizada na aba *Preprocess*.

Uma vez estabelecida a comunicação entre os dois sistemas, a consulta inserida no WEKA foi realizada no SGBD e os resultados foram exibidos no programa como mostra a Figura 15. Através do botão *Save*, localizado nessa mesma aba, é possível salvar todos os dados da consulta realizada no SGBD no formato ARFF (formato padrão do WEKA), agilizando o processo de carregamento dos dados da consulta no WEKA futuramente.





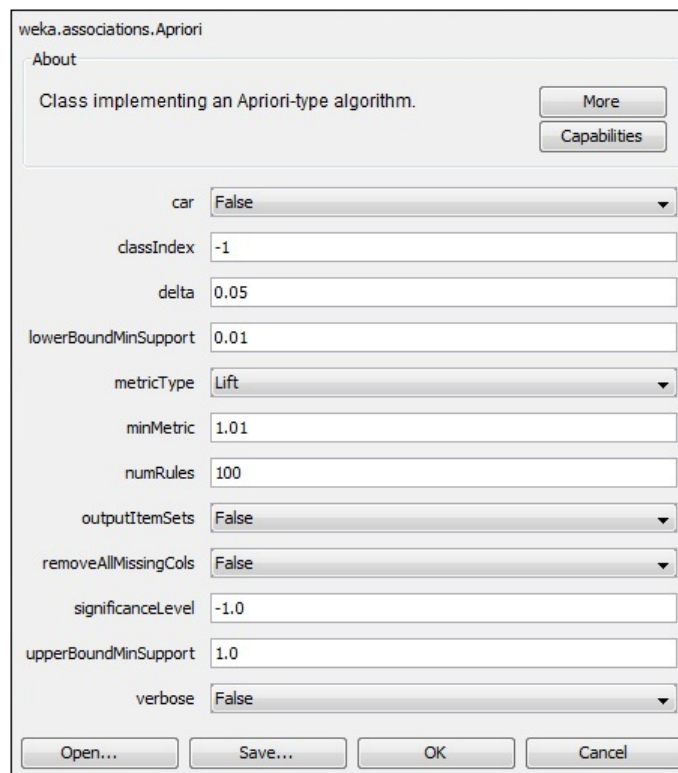
**Figura 15:** Aba *Preprocess* do WEKA

O próximo passo consistiu em acessar a aba correspondente às tarefas de Mineração de Dados escolhidas. Neste trabalho será utilizada a aba *Associate* que permite trabalhar com Regras de Associação e a aba *Cluster* para acessar as opções referentes à tarefa de Clusterização. Em cada uma dessas abas, os parâmetros dos respectivos algoritmos foram configurados. A seguir, é demonstrado como esse passo foi realizado para cada tarefa.

A Figura 16 representa a tela de configuração dos parâmetros do algoritmo *Apriori*, usado para extração de regras de associação. Esta tela contém diversos campos, sendo os mais utilizados:

- ***delta***: o valor especificado nesse campo reduz o suporte a cada iteração, partindo do limite máximo até que o valor mínimo seja alcançado ou até que o número máximo de regras estipulado seja atingido;
- ***lowerBoundMinSupport***: é o valor de suporte mínimo.
- ***metricType***: tipo de métrica (medida) que será considerada para avaliar as regras. Neste trabalho serão utilizadas as medidas de *confidence* (confiança) e *lift*.

- ***minMetric***: este campo é utilizado para definir o valor mínimo da medida selecionada, ou seja, os valores mínimos de confiança e *lift*.
- ***numRules***: número máximo de regras a serem geradas.
- ***outputItemSets***: caso essa opção seja habilitada, além das regras geradas, será exibido também os *itemsets* frequentes de tamanho 1 a n, em que n representa o número de atributos selecionados;
- ***removeAllMissingCols***: se essa opção for habilitada, os atributos que possuem valores ausentes serão desconsiderados.
- ***upperBoundMinSupport***: esse campo é utilizado para especificar o valor máximo que o suporte pode assumir.
- ***Verbose***: caso seja habilitado, exibe mensagens na execução do *Apriori*.



The image shows a dialog box titled "weka.associations.Apriori". It has an "About" section at the top with the text "Class implementing an Apriori-type algorithm." and two buttons: "More" and "Capabilities". Below this are several configuration fields:

- car**: A dropdown menu set to "False".
- classIndex**: A text input field containing "-1".
- delta**: A text input field containing "0.05".
- lowerBoundMinSupport**: A text input field containing "0.01".
- metricType**: A dropdown menu set to "Lift".
- minMetric**: A text input field containing "1.01".
- numRules**: A text input field containing "100".
- outputItemSets**: A dropdown menu set to "False".
- removeAllMissingCols**: A dropdown menu set to "False".
- significanceLevel**: A text input field containing "-1.0".
- upperBoundMinSupport**: A text input field containing "1.0".
- verbose**: A dropdown menu set to "False".

At the bottom of the dialog box are four buttons: "Open...", "Save...", "OK", and "Cancel".

**Figura 16:** Tela de configuração dos parâmetros do algoritmo *Apriori*

Uma vez configurados os parâmetros, foi acionado o botão *OK* e, na janela seguinte, o botão *Start*. Após isso, o algoritmo iniciou o processo de extração de regras de associação e exibiu os resultados conforme mostra a Figura 17:

1. Local de Residência=Belo Horizonte 596 ==> Tipo da Escola=Em Escola Municipal 27	conf:(0.05) < lift:(2.18)>
2. Tipo da Escola=Em Escola Municipal 56 ==> Local de Residência=Belo Horizonte 27	conf:(0.48) < lift:(2.18)>
3. Local de Residência=Diamantina 392 ==> Tipo da Escola=Em Escola Estadual Curso=NUTRICA0 54	conf:(0.14) < lift:(2.08)>
4. Tipo da Escola=Em Escola Estadual Curso=NUTRICA0 179 ==> Local de Residência=Diamantina 54	conf:(0.3) < lift:(2.08)>
5. Local de Residência=Diamantina 392 ==> Tipo da Escola=Em Escola Estadual Curso=FISIOTERAPIA 58	conf:(0.15) < lift:(2.02)>
6. Tipo da Escola=Em Escola Estadual Curso=FISIOTERAPIA 198 ==> Local de Residência=Diamantina 58	conf:(0.29) < lift:(2.02)>
7. Local de Residência=Diamantina 392 ==> Tipo da Escola=Em Escola Estadual Curso=ENGENHARIA FLORESTAL 33	conf:(0.08) < lift:(1.93)>
8. Tipo da Escola=Em Escola Estadual Curso=ENGENHARIA FLORESTAL 118 ==> Local de Residência=Diamantina 33	conf:(0.28) < lift:(1.93)>
9. Tipo da Escola=Em Escola Estadual 1290 ==> Local de Residência=Diamantina Curso=ENGENHARIA FLORESTAL 33	conf:(0.03) < lift:(1.87)>
10. Local de Residência=Diamantina Curso=ENGENHARIA FLORESTAL 37 ==> Tipo da Escola=Em Escola Estadual 33	conf:(0.89) < lift:(1.87)>
11. Local de Residência=Vale do Jequitinhonha 201 ==> Tipo da Escola=Em Escola Estadual Curso=ENFERMAGEM 39	conf:(0.19) < lift:(1.86)>
12. Tipo da Escola=Em Escola Estadual Curso=ENFERMAGEM 282 ==> Local de Residência=Vale do Jequitinhonha 39	conf:(0.14) < lift:(1.86)>
13. Local de Residência=Belo Horizonte Tipo da Escola=Em Escola Particular 339 ==> Curso=ODONTOLOGIA 93	conf:(0.27) < lift:(1.77)>
14. Curso=ODONTOLOGIA 418 ==> Local de Residência=Belo Horizonte Tipo da Escola=Em Escola Particular 93	conf:(0.22) < lift:(1.77)>
15. Tipo da Escola=Em Escola Estadual 1290 ==> Local de Residência=Diamantina Curso=NUTRICA0 54	conf:(0.04) < lift:(1.77)>
16. Local de Residência=Diamantina Curso=NUTRICA0 64 ==> Tipo da Escola=Em Escola Estadual 54	conf:(0.84) < lift:(1.77)>
17. Local de Residência=Diamantina Tipo da Escola=Em Escola Estadual 302 ==> Curso=NUTRICA0 54	conf:(0.18) < lift:(1.69)>
18. Curso=NUTRICA0 286 ==> Local de Residência=Diamantina Tipo da Escola=Em Escola Estadual 54	conf:(0.19) < lift:(1.69)>
19. Tipo da Escola=Em Escola Estadual 1290 ==> Local de Residência=Vale do Jequitinhonha Curso=ENFERMAGEM 39	conf:(0.03) < lift:(1.67)>
20. Local de Residência=Vale do Jequitinhonha Curso=ENFERMAGEM 49 ==> Tipo da Escola=Em Escola Estadual 39	conf:(0.8) < lift:(1.67)>
21. Local de Residência=Diamantina 392 ==> Tipo da Escola=Em Escola Estadual Curso=ENFERMAGEM 67	conf:(0.17) < lift:(1.64)>
22. Tipo da Escola=Em Escola Estadual Curso=ENFERMAGEM 282 ==> Local de Residência=Diamantina 67	conf:(0.24) < lift:(1.64)>
23. Local de Residência=Belo Horizonte 596 ==> Tipo da Escola=Em Escola Particular Curso=ODONTOLOGIA 93	conf:(0.16) < lift:(1.62)>
24. Tipo da Escola=Em Escola Particular Curso=ODONTOLOGIA 260 ==> Local de Residência=Belo Horizonte 93	conf:(0.36) < lift:(1.62)>
25. Tipo da Escola=Em Escola Estadual 1290 ==> Local de Residência=Diamantina Curso=FISIOTERAPIA 58	conf:(0.04) < lift:(1.62)>
26. Local de Residência=Diamantina Curso=FISIOTERAPIA 75 ==> Tipo da Escola=Em Escola Estadual 58	conf:(0.77) < lift:(1.62)>
27. Local de Residência=Diamantina 392 ==> Tipo da Escola=Em Escola Estadual 302	conf:(0.77) < lift:(1.61)>
28. Tipo da Escola=Em Escola Estadual 1290 ==> Local de Residência=Diamantina 302	conf:(0.23) < lift:(1.61)>
29. Tipo da Escola=Em Escola Estadual 1290 ==> Local de Residência=Diamantina Curso=ENFERMAGEM 67	conf:(0.05) < lift:(1.61)>
30. Local de Residência=Diamantina Curso=ENFERMAGEM 87 ==> Tipo da Escola=Em Escola Estadual 67	conf:(0.77) < lift:(1.61)>

**Figura 17:** Regras extraídas pelo algoritmo *Apriori*

Uma vez concluída a tarefa de extração de regras de associação, o próximo passo é analisar os possíveis grupos a serem formados.

A Figura 18 demonstra a tela de configuração dos parâmetros do algoritmo *K-means*, usado para obtenção de grupos com características similares.

weka.clusterers.SimpleKMeans

About

Cluster data using the k means algorithm. More  
Capabilities

displayStdDevs  False

distanceFunction Choose

dontReplaceMissingValues  False

maxIterations

numClusters

preserveInstancesOrder  False

seed

Open... Save... OK Cancel

**Figura 18:** Tela de configuração dos parâmetros do algoritmo *K-means*

Como pode-se perceber a tela contém alguns campos, sendo utilizados neste trabalho os seguintes:

- ***distanceFunction***: permite seleccionar a medida de distância a ser utilizada, sendo, por padrão, a distância euclidiana, já demonstrada na Subsecção 6.2.2;
- ***maxIterations***: número máximo de iteração do algoritmo. Ao ser atingido, o algoritmo encerra a execução, conforme já mencionado Subsecção 6.2.2;
- ***numClusters***: número de grupos (*clusters*) a serem formados.

Uma vez configurados os parâmetros, foi acionado o botão *OK* e, na janela seguinte, o botão *Start*. Após isso, o algoritmo iniciou o processo de formação de grupos e exibiu os resultados conforme mostra a Figura 19.

```
Cluster centroids:
Attribute          Full Data          Cluster#
                   (3504)             0           1           2           3
-----
Tipo da Escola     Escola Estadual Escola Particular Escola Estadual Escola Municipal Outra

Time taken to build model (full training data) : 0.02 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      958 ( 27%)
1     2370 ( 68%)
2       86 (  2%)
3       90 (  3%)

Class attribute: Física
Classes to Clusters:
  0   1   2   3 <-- assigned to cluster
713 2153 77 82 | PESSIMO
152 160  5  6 | REGULAR
 69  40  2  2 | BOM
 22  17  2  0 | MUITO BOM
  2  0  0  0 | OTIMO

Cluster 0 <-- REGULAR
Cluster 1 <-- PESSIMO
Cluster 2 <-- MUITO BOM
Cluster 3 <-- BOM
```

**Figura 19:** Resultado da execução do algoritmo *K-means*

## 7.5 Análise e interpretação dos resultados

Nessa seção será demonstrado um exemplo de como os resultados exibidos no WEKA foram analisados para cada uma das técnicas utilizadas.

A Figura 17 apresenta algumas regras mineradas através do algoritmo *Apriori*. Analisando a regra número 10 tem-se que:

Local de Residência=Diamantina Curso=Engenharia Florestal 37  $\Rightarrow$  Tipo da Escola=Em Escola Estadual 33 conf:(0.89) <lift:(1.87)>

Essa regra demonstra que dos 37 candidatos residentes em Diamantina que escolheram o curso de Engenharia Florestal, 33 cursaram o ensino médio em escola estadual com uma confiança de 89%. De acordo com o valor do *lift*, a chance de um candidato de Diamantina que escolheu esse curso ter estudado em escola estadual é quase 2 vezes maior.

Na Figura 19, são mostrados os resultados da aplicação do algoritmo *K-means* envolvendo os atributos tipo da escola e notas (Física). Nesta Figura, são apresentados os grupos formados (de 0 a 3). No segundo retângulo desta Figura, cada coluna corresponde a um tipo de escola, sendo 0 para escola particular, 1 para escola estadual, 2 para escola municipal e 3 para outras escolas, conforme é descrito no primeiro retângulo. Por exemplo, para escola particular (0), há 713 conceitos péssimo, 152 regular, 69 bom, 22 muito bom e 2 conceitos ótimo. Observando a Figura 19, é possível constatar que a maior parte dos candidatos de todas as escolas obtiveram conceito péssimo na disciplina de Física, seguido dos conceitos regular, bom, muito bom e ótimo. Para os três últimos conceitos, tem-se maior predominância de candidatos de escola particular.

Os demais resultados obtidos para regras de associação foram analisados e comparados para cada semestre de realização do vestibular e estão listados no Capítulo seguinte. A Clusterização, por sua vez, foi realizada apenas no segundo semestre do vestibular de 2009, que é a base de dados mais recente disponibilizada pelo órgão regulador dos processos seletivos unificados da UFVJM.

## 8 RESULTADOS COMPUTACIONAIS

### 8.1 Regras de Associação

Nesta Seção serão apresentados os resultados obtidos com a aplicação do algoritmo *Apriori* através da ferramenta de Mineração de Dados WEKA. Os atributos escolhidos foram aqueles que possibilitaram a realização de análises do perfil sócio-econômico dos candidatos conforme descrito na Seção 2.1. Em razão do alto número de regras obtidas, somente aquelas julgadas mais relevantes foram selecionadas. Como critério de seleção dessas regras, somente foram coletadas para posterior análise as que obtiveram:

- suporte mínimo de 1%;
- confiança mínima de 50%;
- *lift* mínimo de 1.01.

Em razão de não ter sido disponibilizado o questionário sócio-econômico referente ao primeiro e segundo semestre de 2006, não foi possível efetuar a análise de alguns atributos para estes vestibulares.

Todas as análises que envolviam o atributo local de residência não foram realizadas nos anos de 2002 a 2004 em virtude da forma como o questionário sócio-econômico foi formulado, em que ele considerava interior de Minas Gerais como todas as cidades do estado exceto a capital (Belo Horizonte), o que dificultava uma análise mais específica.

Por fim, algumas análises também não puderam ser realizadas nas bases de dados do ano de 2002 em virtude de um problema existente nelas, o qual impedia a interligação das tabelas que foram utilizadas nesse trabalho.

### **8.1.1 Análise dos atributos local de residência e tipo da escola**

A partir dessa primeira análise foi possível perceber, conforme mostra a Tabela 16, que a maior parte dos candidatos residentes em Diamantina, Teófilo Otoni e demais cidades do Vale do Jequitinhonha e do Vale do Mucuri cursaram o Ensino Médio em escola da rede estadual, ao passo que os candidatos de Belo Horizonte e outras regiões de Minas Gerais o fizeram, essencialmente, na rede particular de ensino.

Apenas nesta análise será demonstrada a tabela com as regras geradas, que permitiram chegar à conclusão do parágrafo anterior. Em virtude da extensão das outras tabelas, nas demais análises serão apresentados apenas os comentários relacionados. As tabelas correspondentes a cada uma das demais análises podem ser visualizadas no Apêndice A.

**Tabela 16:** Regras de associação envolvendo os atributos local de residência e tipo da escola

Ano/Sem	Consequente da Regra (tipo da escola)							
	Escola Estadual				Escola Particular			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2005/1	Diamantina 392 ⇒ Escola Estadual 302	0.1	0.77	1.61	Belo Horizonte 596 ⇒ Escola Particular 339	0.12	0.57	1.21
	Na Região do Vale do Jequitinhonha 201 ⇒ Escola Estadual 146	0.05	0.73	1.52	Outra Região de Minas Gerais 1496 ⇒ Escola Particular 800	0.27	0.53	1.14
2005/2	Diamantina 334 ⇒ Escola Estadual 264	0.11	0.79	1.73	Outra Região de Minas Gerais 1241 ⇒ Escola Particular 702	0.29	0.57	1.18
	Na Região do Vale do Jequitinhonha 131 ⇒ Escola Estadual 92	0.04	0.7	1.53	Belo Horizonte 430 ⇒ Escola Particular 241	0.1	0.56	1.17
2007/1	Diamantina 559 ⇒ Escola Estadual 448	0.14	0.8	1.53	Outra Região de Minas Gerais 1405 ⇒ Escola Particular 844	0.26	0.6	1.44
	Na Região do Vale do Mucuri 123 ⇒ Escola Estadual 95	0.03	0.77	1.47	Belo Horizonte 482 ⇒ Escola Particular 273	0.08	0.57	1.35
	Na Região do Vale do Jequitinhonha 364 ⇒ Escola Estadual 266	0.08	0.73	1.39				
	Teófilo Otoni 345 ⇒ Escola Estadual 243	0.07	0.7	1.34				
2007/2	Na Região do Vale do Mucuri 129 ⇒ Escola Estadual 111	0.04	0.86	1.46	Outra Região de Minas Gerais 1114 ⇒ Escola Particular 612	0.23	0.55	1.53
	Diamantina 443 ⇒ Escola Estadual 376	0.14	0.85	1.45	Belo Horizonte 374 ⇒ Escola Particular 200	0.07	0.53	1.49
	Teófilo Otoni 379 ⇒ Escola Estadual 309	0.11	0.82	1.39				
	Na Região do Vale do Jequitinhonha 279 ⇒ Escola Estadual 210	0.08	0.75	1.28				
2008/1	Diamantina 598 ⇒ Escola Estadual 494	0.14	0.83	1.35	Belo Horizonte 431 ⇒ Escola Particular 237	0.07	0.55	1.33
	Na Região do Vale do Jequitinhonha 428 ⇒ Escola Estadual 356	0.1	0.83	1.35	Outra Região de Minas Gerais 1218 ⇒ Escola Particular 665	0.19	0.55	1.61
	Na Região do Vale do Mucuri 194 ⇒ Escola Estadual 159	0.05	0.82	1.33				
	Teófilo Otoni 591 ⇒ Escola Estadual 463	0.13	0.78	1.28				
2008/2	Na Região do Vale do Mucuri 113 ⇒ Escola Estadual 100	0.05	0.88	1.44	Belo Horizonte 266 ⇒ Escola Particular 146	0.07	0.55	1.63
	Na Região do Vale do Jequitinhonha 213 ⇒ Escola Estadual 183	0.09	0.86	1.4	Outra Região de Minas Gerais 643 ⇒ Escola Particular 348	0.17	0.54	1.61
	Diamantina 436 ⇒ Escola Estadual 360	0.18	0.83	1.34				
	Teófilo Otoni 264 ⇒ Escola Estadual 216	0.11	0.82	1.33				
2009/1	Diamantina 647 ⇒ Escola Estadual 552	0.16	0.85	1.26	Belo Horizonte 340 ⇒ Escola Particular 188	0.05	0.55	2.02
	Na Região do Vale do Mucuri 262 ⇒ Escola Estadual 220	0.06	0.84	1.24				
	Na Região do Vale do Jequitinhonha 493 ⇒ Escola Estadual 410	0.12	0.83	1.23				
	Teófilo Otoni 794 ⇒ Escola Estadual 619	0.18	0.78	1.15				

Continua na página seguinte...



**Tabela 16:** Regras de associação envolvendo os atributos local de residência e tipo da escola (cont)

		Consequente da Regra (tipo da escola)						
		Escola Estadual			Escola Particular			
Ano/Sem	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2009/2	Diamantina 469 ⇒ Escola Estadual 404	0.18	0.86	1.37	Belo Horizonte 221 ⇒ Escola Particular 122	0.05	0.55	1.74
	Na Região do Vale do Mucuri 144 ⇒ Escola Estadual 116	0.05	0.81	1.28	Outra Região de Minas Gerais 742 ⇒ Escola Particular 389	0.17	0.52	1.65
	Na Região do Vale do Jequitinhonha 270 ⇒ Escola Estadual 214	0.09	0.79	1.26				
	Teófilo Otoni 350 ⇒ Escola Estadual 273	0.12	0.78	1.24				

### **8.1.2 Análise dos atributos local de residência e curso**

A partir desta análise, foi possível constatar que a maior parte dos candidatos aos cursos de Educação Física e Bacharelado em Humanidades (BHU) são de Diamantina. Já a maior parte dos candidatos aos cursos de Administração, Ciências Econômicas, Ciências Contábeis e Serviço Social residem em Teófilo Otoni. Em contrapartida, a maioria dos candidatos aos cursos de agrárias (Agronomia, Engenharia Florestal e Zootecnia), bem como Odontologia e Farmácia, ambos da área de saúde, são de outra região do estado de Minas Gerais, embora Farmácia seja menos frequente (apenas ocorre em 2005/1, 2005/2, 2007/1 e 2007/2). Isso quer dizer que a maior parte desses candidatos não residem em Diamantina, Teófilo Otoni, Belo Horizonte, bem como nenhuma outra cidade localizada no Vale do Jequitinhonha ou no Vale do Mucuri.

Para o curso de Bacharelado em Ciência e Tecnologia de Teófilo Otoni, no primeiro semestre de 2009 (ano em que o curso começou a ser oferecido no campus da cidade) a maior parte dos candidatos eram residentes na própria cidade. No semestre seguinte não houveram regras com a confiança mínima especificada envolvendo este curso que pudessem fortalecer essa afirmação. Quanto ao Bacharelado em Ciência e Tecnologia de Diamantina, no segundo semestre de 2009 a maior parte dos candidatos eram de outra região de Minas Gerais.

As regras correspondentes a esta análise podem ser visualizadas na Tabela 17 do Apêndice A.

### **8.1.3 Análise dos atributos tipo da escola e curso**

Conforme mostra a Tabela 18 do Apêndice A, a maior parte dos candidatos aos cursos de Odontologia e Farmácia cursaram o ensino médio em escola particular. Já a maioria dos candidatos aos cursos de Nutrição (embora não esteja presente em 2007/2, 2008/1 e 2009/1), Fisioterapia (embora apareça apenas em semestres intercalados), Serviço Social, Ciências Econômicas, Ciências Contábeis, Educação Física, Administração, Sistemas de Informação, Ciências Biológicas e Bacharelado em Humanidades cursaram o ensino médio em escola estadual.

No que se refere aos cursos da área de agrárias, percebe-se que Zootecnia apresentou inicialmente (2002/2, 2003/1 e 2003/2) maioria de candidatos oriundos de escola estadual. Em 2007/1, 2008/1 e 2008/2 houve uma mudança onde a maioria dos candidatos passaram a ser de

escola particular. Engenharia Florestal (presente em 2002/1, 2002/2, 2003/1, 2003/2 e 2005/1) e Agronomia (presente em 2002/1, 2002/2, 2005/1 e 2009/2) são pouco frequentes, mas sempre associados a maioria de alunos de escola estadual.

Conforme pode ser visto na Tabela 18 do Apêndice A, essa análise não foi demonstrada no ano de 2004 pois não foi gerada nenhuma regra no formato "curso  $\Rightarrow$  tipo da escola" com a confiança mínima estabelecida.

#### **8.1.4 Análise dos atributos atividade remunerada e curso**

De acordo com essa análise, os cursos de Agronomia, Engenharia Florestal e Zootecnia (embora só ocorram a partir do primeiro semestre de 2007) conjuntamente com Farmácia, Odontologia, Fisioterapia, Nutrição e Enfermagem, estes cinco últimos da área de saúde, assim como Bacharelado em Ciência e Tecnologia de Diamantina são, essencialmente, formados por candidatos que não exercem atividade remunerada.

O curso de Enfermagem só ocorre no primeiro semestre de cada ano pois só havia realização de vestibular para este curso ao final de cada ano.

A Tabela 19 do Apêndice A demonstra as regras correspondentes a esta análise.

#### **8.1.5 Análise dos atributos frequência pré-vestibular e notas**

A análise desses atributos permite concluir que a grande maioria dos candidatos obtiveram um conceito péssimo, independentemente de frequência em pré-vestibular. Os candidatos que obtiveram conceitos maiores ou iguais a regular, embora estes conceitos sejam em menor quantidade, frequentaram pré-vestibular por um semestre. Há um destaque para História e Português e Literatura. Embora isso ocorra em poucos semestres, pode-se perceber que foram as únicas matérias em que pouco mais da metade dos candidatos que frequentaram pré-vestibular por pelo menos um ano obtiveram um conceito regular.

A frequência pré-vestibular com duração de menos de um semestre somente foi demonstrada até o segundo semestre de 2004, uma vez que a partir do primeiro semestre de 2005 essa opção de frequência foi removida do questionário, o qual passou a considerar um semestre como frequência mínima.

A Tabela 20 do Apêndice A apresenta as regras obtidas para esses atributos.

### **8.1.6 Análise dos atributos tipo da escola e notas**

Conforme Tabela 21 do Apêndice A, a maior parte dos candidatos vindos de escola municipal e estadual obtiveram conceito péssimo nas disciplinas. Para as escolas federais, a ocorrência foi muito pequena (apenas 9 regras divididas entre os anos de 2003 e 2004), mas sempre com maioria de candidatos com conceito péssimo. Por representar um número muito pequeno, as regras correspondentes a esse tipo de escola não foram incluídas na tabela referente a esta análise.

O conceito regular (em menor ocorrência do que todos os demais) está associado tanto a escolas estaduais quanto particulares. Já os conceitos maiores que regular (bom, muito bom e ótimo) estão sempre associados com alunos provenientes de escola da rede particular de ensino.

### **8.1.7 Análise dos atributos local de residência e notas**

A análise desses atributos permite perceber que a maior parte dos candidatos de Diamantina, Teófilo Otoni e demais cidades do Vale do Jequitinhonha e do Vale do Mucuri obtiveram um conceito péssimo. Os conceitos maiores ou iguais a regular, embora estes ocorram em pequeno número, estão sempre associados às demais cidades do estado (exceto Belo Horizonte).

É possível perceber, conforme Tabela 22 do Apêndice A, que nesse período houve uma variação no número de regras presentes envolvendo os melhores conceitos. Como a situação nos anos posteriores a 2007 se mostrou similar para Diamantina, Teófilo Otoni, Vale do Jequitinhonha e Vale do Mucuri, optou-se por apresentar as regras somente até este ano. Para as demais cidades do estado, a situação foi semelhante ao primeiro semestre de 2007, ou seja, houve redução significativa no número de regras com conceitos maiores ou iguais a regular associados a outra região de Minas Gerais.

As regras de associação podem ser vistas na Tabela 22 do Apêndice A.

### **8.1.8 Análise dos atributos curso e notas**

Com a análise desses atributos foi possível perceber que a grande maioria dos candidatos aos cursos, ao longo dos semestres, obtiveram um conceito péssimo para a maioria das disciplinas. Além desse conceito, foi verificado somente a presença do conceito regular o qual

se encontrava em menor quantidade em relação ao anterior.

Sendo assim, conforme pode ser visto na Tabela 23 do Apêndice A, foram listados somente os cursos que obtiveram conceito regular para disciplinas. Para os demais casos, houve presença somente do conceito péssimo.

Um fato interessante é que, no geral, quando o conceito regular ocorre, apenas está relacionado às disciplinas de Geografia, História ou Português e Literatura, com maior destaque para os cursos de Agronomia, Zootecnia e Odontologia. Esta situação é perceptível apenas a partir do segundo semestre de 2004, o que justifica o fato dos semestres anteriores a este não terem sido mostrados na Tabela.

Para os semestres anteriores a 2004/2, houve ocorrência apenas do conceito péssimo.

### **8.1.9 Análise dos atributos aprovado, tipo da escola e curso**

Analisando a Tabela 24 do Apêndice A, pode-se perceber que, ao longo dos anos (exceto para o primeiro semestre do ano de 2003), os cursos de Odontologia e Farmácia apresentaram geralmente alto índice de reprovação entre alunos de escola particular, embora a maior parte dos aprovados nesses cursos ainda venham desse tipo de escola. Para as escolas estaduais, a maioria dos candidatos foram reprovados em ambos os cursos ao longo dos anos, embora no primeiro semestre de 2003 apenas seja possível perceber um bom índice de aprovação em Farmácia.

Um fato interessante é em relação ao curso de Enfermagem. Embora a maior parte dos aprovados sejam de escola particular a partir de 2005/1, em 2003/1 pode-se perceber que a maior parte dos aprovados eram, essencialmente, de escola estadual. Mas de 2004/1 em diante as escolas estaduais passaram a apresentar alto índice de reprovação neste curso, ficando a escola particular responsável pela maior parte das aprovações.

Além disso, as escolas estaduais também respondem por um alto índice de reprovação em engenharia florestal (exceto em 2003/2), sendo mais frequente a partir do segundo semestre de 2005, conforme mostra a Tabela. Enquanto as escolas estaduais apresentam alto índice de reprovação em Engenharia Florestal, as escolas particulares respondem pela maior parte das aprovações no mesmo.

Para os cursos de Ciências Econômicas, Ciências Contábeis, Serviço Social, Sistemas de Informação, Administração, Educação Física e Bacharelado em Humanidades, a maior parte dos aprovados são oriundos de escolas estaduais. Embora as escolas estaduais como um todo

apresentem bom índice de aprovação para a maioria desses cursos, o mesmo não ocorre com Administração e Serviço Social, em que a maior parte dos alunos de escolas estaduais foram reprovados.

Quanto ao curso de Bacharelado em Ciência e Tecnologia, tanto de Diamantina quanto de Teófilo Otoni, tem-se um bom índice de aprovação tanto para escolas estaduais quanto particulares. Observando o ano de 2009, é possível verificar que, embora ambos tenham bom índice de aprovação, o curso de Bacharelado em Ciência e Tecnologia de Diamantina tem um índice de aprovação um pouco maior do que o de Teófilo Otoni.

Por fim, pode-se perceber algo interessante também sobre os cursos de Fisioterapia e Ciências Biológicas. De 2003/2 a 2007/2, as escolas estaduais, como um todo, apresentaram grande número de reprovações em Fisioterapia e, a partir de 2008/2, a situação se inverte e estende até 2009/2. Embora as escolas particulares respondessem pela maior parte das aprovações em ambos os cursos, pode-se perceber uma outra mudança a partir do segundo semestre de 2008, em que as escolas estaduais passaram a responder pela maior parte das aprovações nesses cursos.

Por fim, para o curso de Nutrição ocorre uma modificação semelhante. Do segundo semestre de 2003 ao segundo semestre de 2008, a maioria dos candidatos de escolas estaduais foram reprovados, mas em 2009/1 a maior parte dos alunos desse tipo de escola que escolheram o curso de Nutrição foram aprovados.

#### **8.1.10 Análise dos atributos aprovado, local de residência e curso**

Ao longo dos anos, Diamantina apresentou alto percentual de reprovação nos cursos de Enfermagem, Farmácia e Nutrição (este apenas presente nos anos de 2005, 2007 e 2008). Há um destaque para o curso de Fisioterapia que, embora tenha apresentado inicialmente alto índice de reprovação nos anos de 2005 a 2007, foi possível perceber uma mudança no segundo semestre dos anos de 2008 e 2009, em que Diamantina passou a apresentar um bom índice de aprovação nesse curso. O mesmo ocorre para Ciências Biológicas que, ao contrário do primeiro semestre dos anos de 2007 e 2008, apresentou um bom índice de aprovação entre os candidatos de Diamantina no último semestre dos anos de 2008 e 2009. Além disso, a cidade apresenta bom índice de aprovação nos cursos de Educação Física, Bacharelado em Humanidades e Bacharelado em Ciência e Tecnologia de Diamantina. A maior parte dos aprovados em Educação

Física e Bacharelado em Humanidades são dessa cidade.

Já o Vale do Jequitinhonha se destacou mais pelo alto índice de reprovação no curso de Enfermagem ao longo dos anos. Porém, no ano de 2009 a região passou a apresentar um índice de aprovação significativo no curso de Bacharelado em Ciência e Tecnologia de Diamantina.

Para Teófilo Otoni, é possível perceber que há um forte índice de aprovação nos cursos de Ciências Contábeis (principalmente) e Ciências Econômicas. Em contrapartida, Teófilo Otoni não possui um índice bom de aprovação em Serviço Social e Administração, embora todos os cursos citados sejam oferecidos na própria cidade. Essa ocorrência de Serviço Social e Administração é pouco frequente, mas é possível perceber que sempre que os mesmos surgem, estão relacionados a algum índice de reprovação. Apesar disso, a maior parte dos candidatos aprovados em nível geral nesses cursos são dessa cidade. Além disso, a cidade, como um todo, apresenta também um bom índice de aprovação em Bacharelado em Ciência e Tecnologia, exclusivamente de Teófilo Otoni.

Assim como o Vale do Jequitinhonha, o Vale do Mucuri apresenta bom índice de aprovação no Bacharelado em Ciência e Tecnologia, porém de Teófilo Otoni, o que se justifica pela cidade estar localizada nessa região. Mas, assim como Teófilo Otoni, a região possui um índice maior de reprovação em Serviço Social.

Para Belo Horizonte, tem-se alto índice de reprovação em Farmácia e Odontologia.

Para outras regiões de Minas Gerais, as regras apontam para um alto índice de reprovação nos cursos de Enfermagem, Farmácia, Odontologia e Engenharia Florestal, mas também aponta para um índice maior de aprovação em Zootecnia, Agronomia e Bacharelado em Ciência e Tecnologia de Diamantina. Embora as demais regiões apresentem um índice alto de reprovação em Odontologia e Engenharia Florestal, a maioria dos aprovados nesses cursos, assim como Zootecnia e Agronomia, são provenientes dessas regiões.

As regras de associação referentes a esses resultados se encontram na Tabela 25 do Apêndice A.

### **8.1.11 Análise dos atributos aprovado, atividade remunerada e curso**

Conforme mostra a Tabela 26 do Apêndice A, os cursos de Odontologia, Enfermagem e Farmácia possuem, em geral, alto índice de reprovação mesmo entre os candidatos que não exercem algum tipo de atividade remunerada, exceto para o primeiro semestre de 2003. Já os cursos

de Ciências Biológicas, Educação Física e Fisioterapia (este mais recentemente) possuem um maior índice de aprovação entre os candidatos que não exercem atividade econômica. Há um destaque para Fisioterapia que, embora tenha apresentado um alto número de reprovações do segundo semestre de 2003 ao primeiro semestre de 2007, foi possível perceber uma mudança nos três últimos semestres em que o vestibular foi realizado, onde o curso passou a apresentar um índice considerável de aprovações entre os candidatos que não exercem algum tipo de atividade remunerada.

Quanto aos cursos da área de Agrárias, a análise permitiu perceber que Agronomia e Zootecnia são os únicos cursos da área que possuem um bom índice de aprovação entre os candidatos que não exercem atividade econômica, ao contrário do que ocorre com Engenharia Florestal. Apesar da menor ocorrência, tal curso, quando presente, sempre está associado a um maior índice de reprovação.

Ciências Contábeis, Bacharelado em Humanidades e Bacharelado em Ciência e Tecnologia de Teófilo Otoni são os únicos cursos que apresentaram bom índice de aprovação tanto entre os candidatos que não exercem atividade remunerada quanto aqueles que a praticam em tempo integral.

Para os cursos de Ciências Contábeis e Ciências Econômicas, tem-se um bom índice de aprovação para os candidatos que não exercem atividade econômica. Quanto ao curso de Serviço Social, embora inicialmente a maior parte dos candidatos tenham sido aprovados, pode-se perceber uma mudança a partir do primeiro semestre de 2008 que se estende até o primeiro semestre de 2009, em que a maior parte dos candidatos que não exercem atividade remunerada foram reprovados.

Para os cursos de Ciências Biológicas, Educação Física, Enfermagem, Farmácia, Fisioterapia, Odontologia e Nutrição (todos da área de saúde), Agronomia, Engenharia Florestal e Zootecnia (da área de agrárias) e Bacharelado em Ciência e Tecnologia de Diamantina, a maioria dos aprovados não exercem atividade remunerada.

#### **8.1.12 Análise dos atributos aprovado, frequência pré-vestibular e curso**

De acordo com Tabela 27 do Apêndice A, a maioria dos candidatos que frequentaram pré-vestibular, independentemente do tempo de duração, para os cursos de Enfermagem, Farmácia e Odontologia foram reprovados, exceto no primeiro semestre de 2003.



Os cursos de Fisioterapia e Nutrição compartilham uma mesma situação. No primeiro semestre de 2003, a maior parte dos candidatos que frequentaram pré-vestibular foram aprovados. Dos semestres posteriores até o primeiro semestre de 2008 há uma inversão, ou seja, a maioria foi reprovada. No segundo semestre dos anos de 2008 e 2009, tem-se uma situação similar ao primeiro semestre de 2003. Para todos os cursos mencionados até então, a maioria dos candidatos sem frequência em pré-vestibular foram reprovados.

Para os cursos de Agrárias, o número de regras geradas foi pequeno. Mas o que se pode perceber é que quando as regras envolvendo Agronomia e Zootecnia estiveram presentes, a maioria dos candidatos com frequência em pré-vestibular foram aprovados. Para Engenharia Florestal, essa situação se manteve mais dividida, ou seja, ora a maior parte foi aprovada, ora reprovada.

Os cursos de Ciências Contábeis, Ciências Econômicas, Bacharelado em Ciência e Tecnologia de Teófilo Otoni, Bacharelado em Ciência e Tecnologia de Diamantina e Bacharelado em Humanidades apresentaram um índice de aprovação significativo mesmo entre candidatos que não frequentaram pré-vestibular.

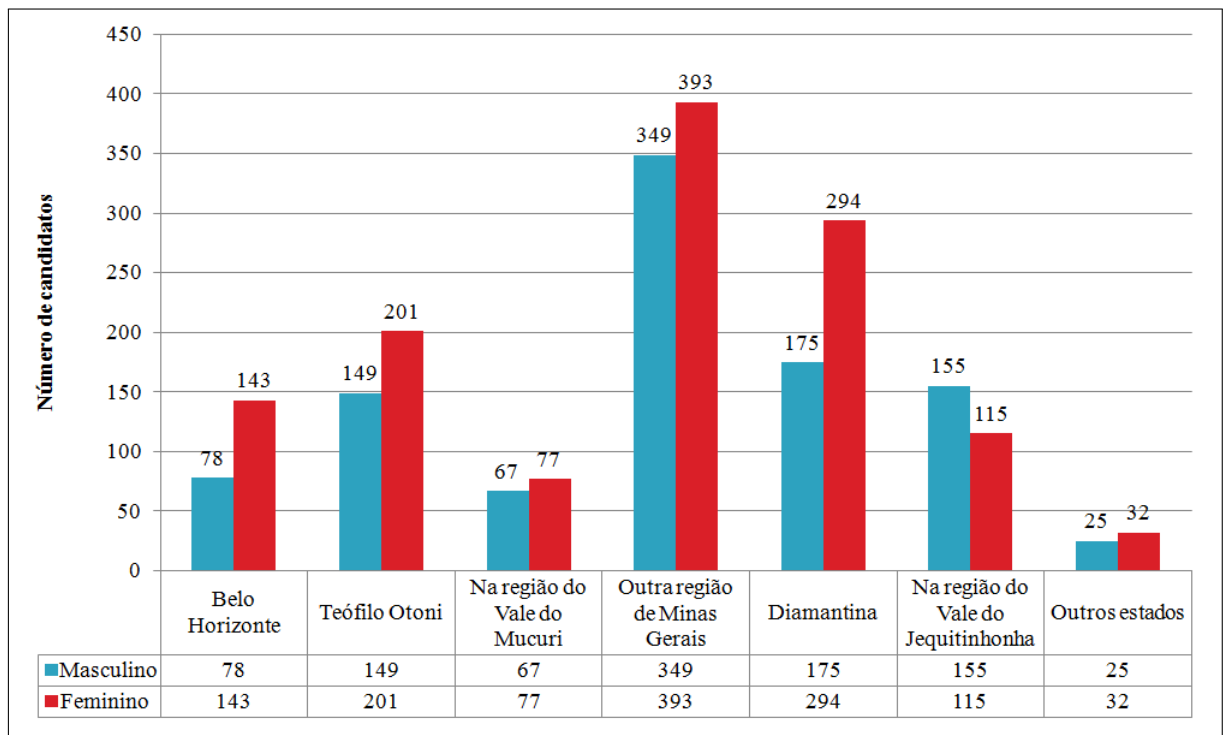
Os cursos de Administração, Ciências Contábeis, Ciências Econômicas, Serviço Social, Bacharelado em Ciência e Tecnologia de Teófilo Otoni e Bacharelado em Humanidades são formados, essencialmente, por candidatos que não frequentam pré-vestibular.

A maior parte dos aprovados em Administração, Serviço Social, Ciências Contábeis, Bacharelado em Ciência e Tecnologia de Teófilo Otoni e Bacharelado em Humanidades não frequentaram pré-vestibular.

## 8.2 Clusterização

### 8.2.1 Análise dos atributos sexo e local de residência

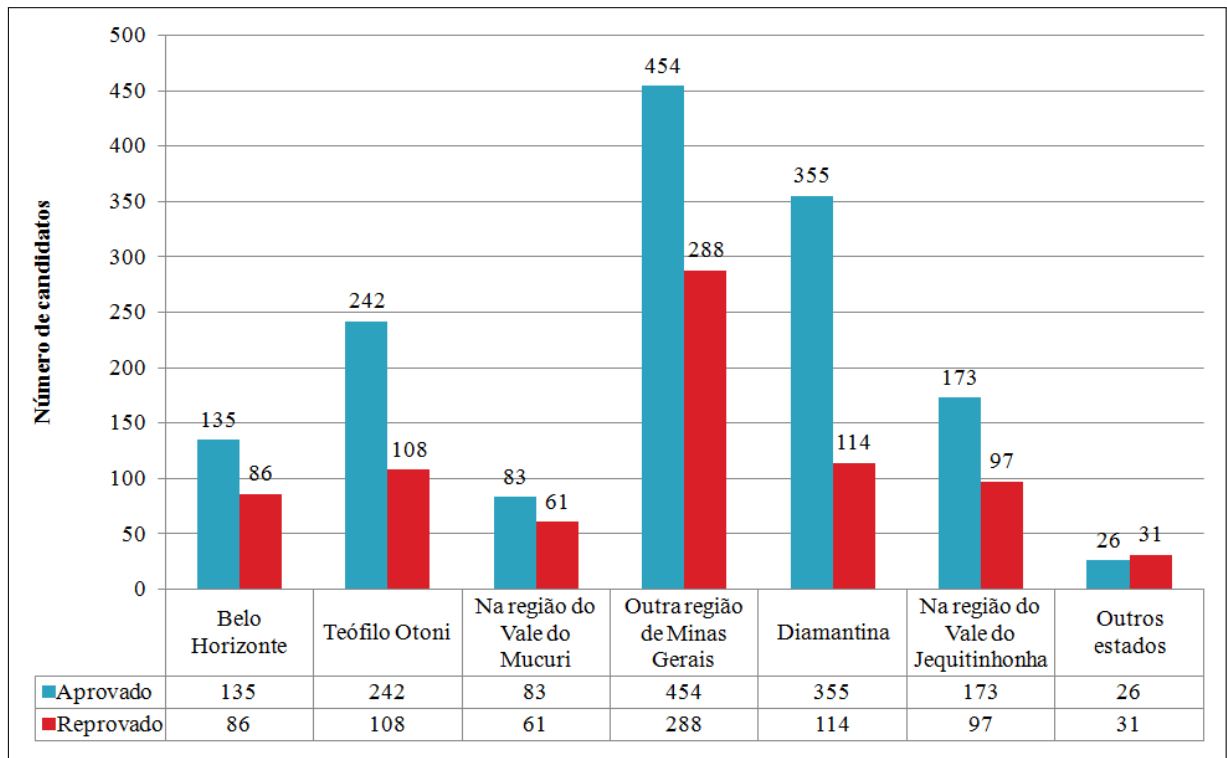
Conforme mostra essa análise, em todos os locais de residência, exceto Vale do Jequitinhonha, há mais candidatas do sexo feminino do que do sexo masculino.



**Figura 20:** Gráfico dos atributos sexo e local de residência (segundo semestre de 2009)

## 8.2.2 Análise dos atributos aprovado e local de residência

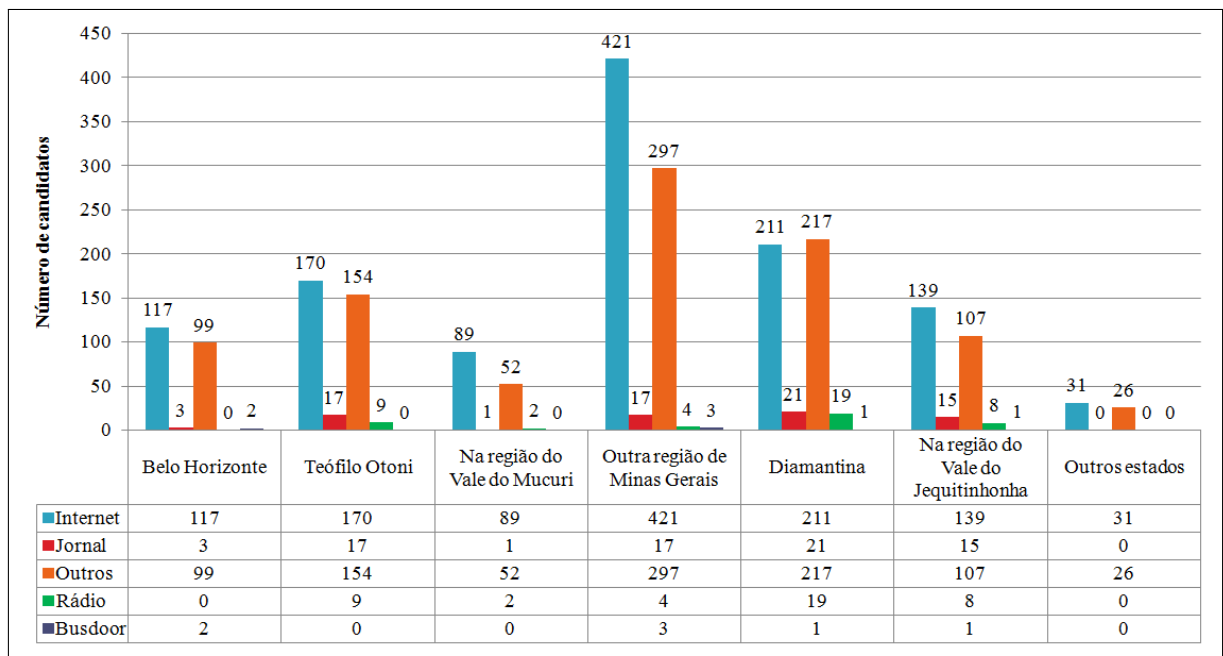
De acordo com os dados da análise, o número de aprovações foi superior ao de reprovações em todos os locais de residência, exceto em outros estados onde a quantidade de reprovações foi maior.



**Figura 21:** Gráfico dos atributos aprovado e local de residência (segundo semestre de 2009)

### 8.2.3 Análise dos atributos como conheceu o vestibular e local de residência

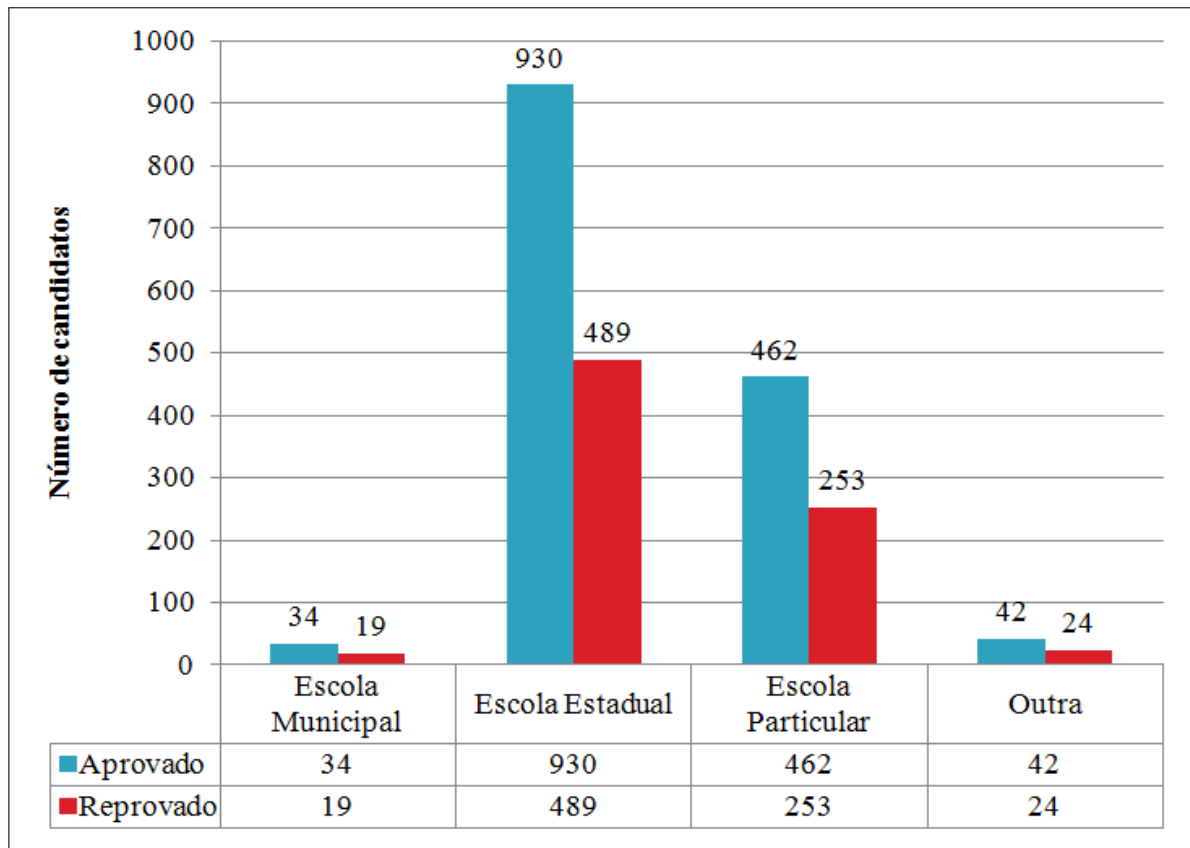
Observando o gráfico da Figura 22, é possível perceber que em todos os locais a maioria dos candidatos tomaram conhecimento do vestibular através da internet e de outros meios, sendo a internet o meio de divulgação que atingiu o maior número de candidatos. Apenas em Diamantina prevaleceram outros meios. Em terceiro lugar, encontra-se o jornal para todos os locais, exceto o Vale do Mucuri, onde o rádio foi o terceiro meio de divulgação que mais atingiu a população, embora a diferença entre este e o jornal tenha sido extremamente pequena nesta região.



**Figura 22:** Gráfico dos atributos como conheceu o vestibular e local de residência (segundo semestre de 2009)

### 8.2.4 Análise dos atributos aprovado e tipo da escola

Essa análise mostra que, para todos os tipos de escola, a maior parte dos candidatos ao vestibular foram aprovados.

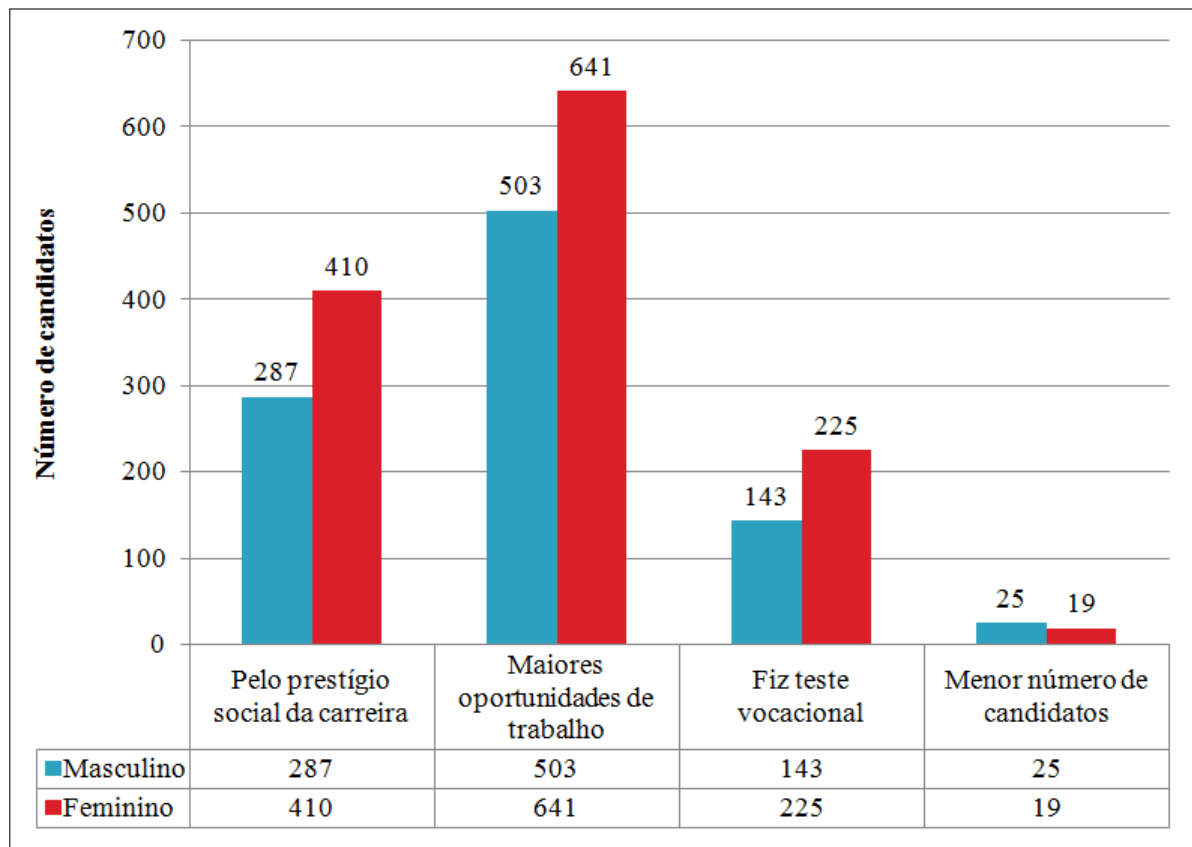


**Figura 23:** Gráfico dos atributos aprovado e tipo da escola (segundo semestre de 2009)

### 8.2.5 Análise dos atributos por que escolheu o curso e sexo

Entre os candidatos que escolheram o curso em razão do prestígio social da carreira, por maiores oportunidades de trabalho ou através da realização de teste vocacional, a maioria é do sexo feminino. Já entre os que escolheram em virtude de possuir menor número de candidatos, a maior parte é do sexo masculino.

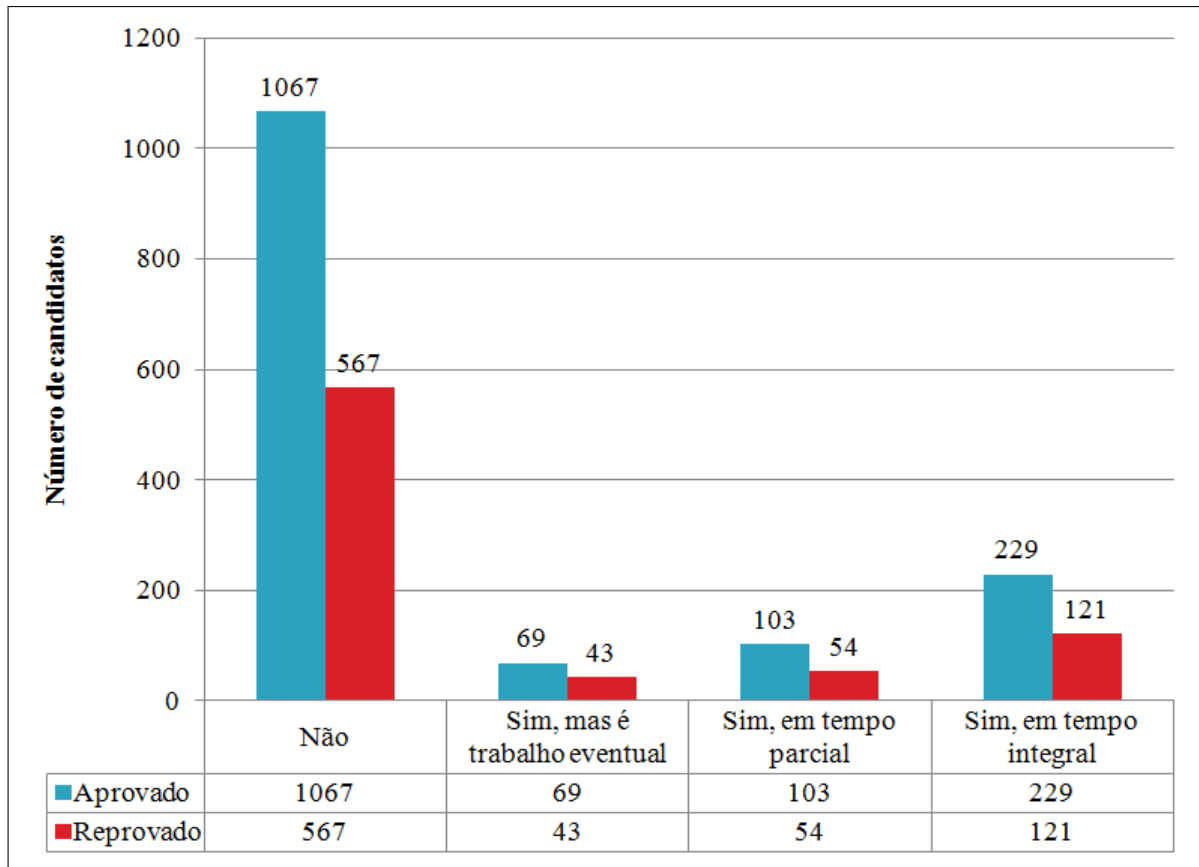
De forma geral, a maioria dos candidatos de ambos os sexos escolheram o curso em razão dele oferecer maiores oportunidades de trabalho.



**Figura 24:** Gráfico dos atributos por que escolheu o curso e sexo (segundo semestre de 2009)

### 8.2.6 Análise dos atributos aprovado e atividade remunerada

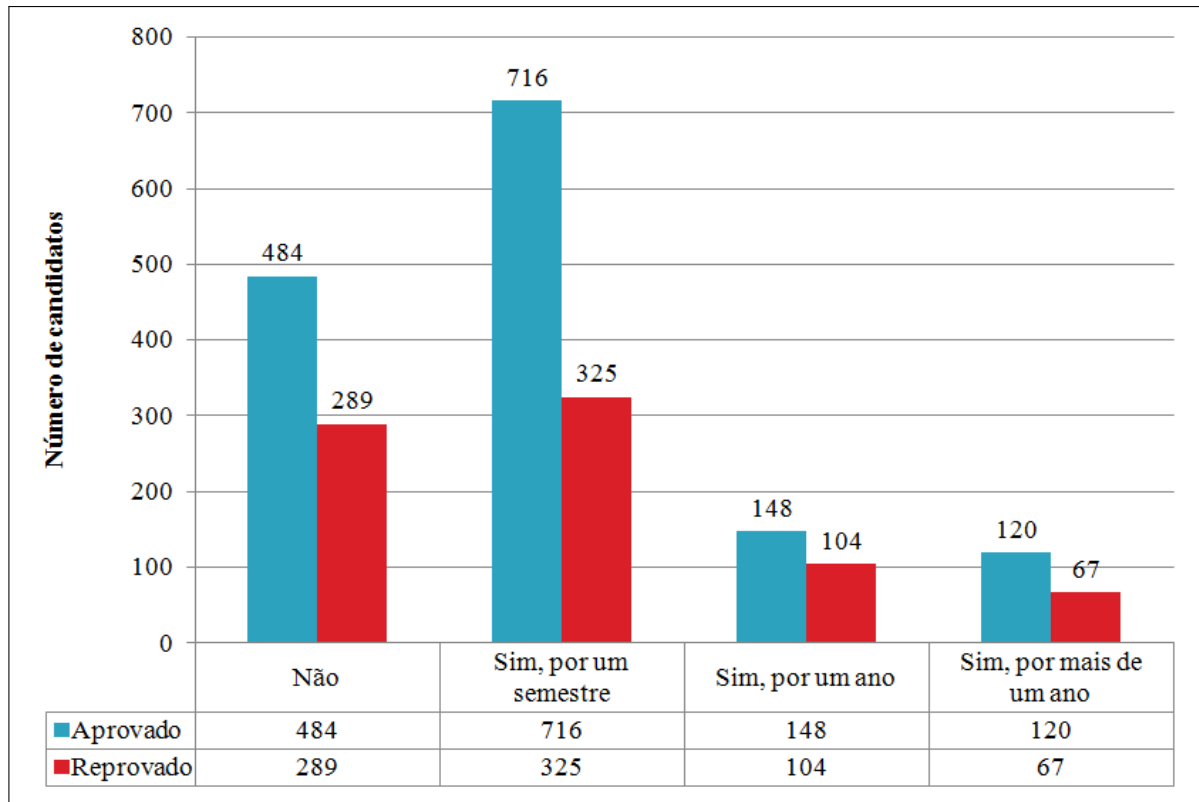
Conforme análise mostrada a seguir, entre os candidatos que não exercem ou mesmo exercem algum tipo de atividade remunerada seja em caráter eventual, parcial ou integral, o número de aprovações foi superior ao de reprovações.



**Figura 25:** Gráfico dos atributos aprovado e atividade remunerada (segundo semestre de 2009)

### 8.2.7 Análise dos atributos aprovado e frequência pré-vestibular

A partir dessa análise, verificou-se que embora, no geral, o índice de aprovações tenha sido superior ao de reprovações, o maior número de aprovações ocorreu entre aqueles que frequentaram pré-vestibular por um semestre.



**Figura 26:** Gráfico dos atributos aprovado e frequência pré-vestibular (segundo semestre de 2009)



## 9 CONCLUSÕES E TRABALHOS FUTUROS

Conforme foi demonstrado, este trabalho utilizou duas técnicas ou algoritmos de Mineração de Dados, *Apriori* e *K-means*, para extração de regras de associação e obtenção de grupos com características semelhantes, respectivamente. Ambos os algoritmos são os mais utilizados para esses propósitos por razões já mencionadas nas seções 5.3 e 5.4, razões essas que motivaram a escolha de ambos para uso neste trabalho.

Após a análise das regras de associação e dos grupos obtidos com a aplicação desses algoritmos, foi possível concluir que:

- Há uma diferença no perfil dos candidatos quanto à localização e o tipo da escola onde estudaram: candidatos de Diamantina, Teófilo Otoni, Vale do Jequitinhonha e Vale do Mucuri, em sua maioria, cursaram o ensino médio em escola estadual. Em contrapartida, a maior parte dos candidatos de Belo Horizonte e demais regiões do estado de Minas Gerais cursaram o ensino médio em escola particular.
- Há uma diferença no perfil dos candidatos quanto à localização e alguns cursos escolhidos: a maioria dos candidatos aos cursos de Educação Física e Bacharelado em Humanidades residem em Diamantina. Para os cursos de Administração, Ciências Econômicas, Ciências Contábeis e Serviço Social, a maioria reside em Teófilo Otoni. Agronomia, Engenharia Florestal, Zootecnia, Odontologia e Farmácia apresentam maioria de candidatos residentes em outras regiões de Minas Gerais (excluindo-se Diamantina, Teófilo Otoni, Belo Horizonte e demais cidades do Vale do Jequitinhonha e do Vale do Mucuri).
- Conseqüentemente, há uma diferença no perfil dos candidatos quanto ao tipo da escola e os cursos escolhidos: nos cursos de Odontologia e Farmácia, a maior parte dos candidatos realizaram o ensino médio em escola particular, em oposição ao que ocorre com a maioria dos candidatos aos cursos de Nutrição, Fisioterapia, Serviço Social, Ciências Econômicas, Ciências Contábeis, Educação Física, Administração, Sistemas de Informação, Ciências Biológicas e Bacharelado em Humanidades.
- O desempenho (notas), em geral, é ruim. Os poucos bons desempenhos estão associados,

em sua maioria, a candidatos oriundos de outra região de Minas Gerais, que estudaram em escola particular e frequentaram pré-vestibular.

- Não há nenhum curso cujos candidatos se destaquem nas notas de uma forma geral. Os poucos bons resultados ocorrem com maior frequência nos cursos de Agronomia, Zootecnia e Odontologia.
- Dentre os candidatos de Diamantina, pode-se notar um bom índice de aprovação nos cursos de Educação Física, Bacharelado em Ciência e Tecnologia, Bacharelado em Humanidades, Fisioterapia (mais recentemente) e Ciências Biológicas. Dentre os candidatos de Teófilo Otoni, pode-se notar um bom índice de aprovação nos cursos de Ciências Contábeis, Ciências Econômicas e Bacharelado em Ciência e Tecnologia. Para o Vale do Jequitinhonha, tem-se maior índice de aprovação em Bacharelado em Ciência e Tecnologia de Diamantina e para o Vale do Mucuri em Bacharelado em Ciência e Tecnologia de Teófilo Otoni. Para outras regiões de Minas Gerais, foi possível perceber que o maior índice de aprovação se deu nos cursos de Zootecnia, Agronomia e Bacharelado em Ciência e Tecnologia de Diamantina. Para os cursos de Odontologia, Enfermagem e Farmácia, a maioria dos aprovados são de outras regiões de Minas Gerais.
- De forma geral, a maioria dos aprovados nos cursos são os que não exercem atividade remunerada. Isto pode ser justificado pelo fato de que a maioria dos candidatos não exerce nenhuma atividade. Apenas os cursos de Ciências Contábeis, Bacharelado em Humanidades e Bacharelado em Ciência e Tecnologia de Teófilo Otoni apresentaram bom índice de aprovação entre candidatos que exercem algum tipo de atividade remunerada.
- A internet foi o meio de divulgação que atingiu o maior número de candidatos, salvo Diamantina onde predominaram outros meios (excluindo-se internet, rádio, jornal e *bus-door*).
- Para todos os locais, exceto Vale do Jequitinhonha, há mais candidatos do sexo feminino do que do sexo masculino.
- Entre todos os locais, Diamantina foi a cidade com a maior porcentagem de candidatos aprovados (cerca de 75% dos candidatos dessa cidade foram aprovados).

- Não há escola que se destaque percentualmente em aprovação em relação às demais. Para todas elas, embora o número de aprovações tenha sido maior que o de reprovações, a diferença entre a porcentagem de aprovados é pequena entre as escolas.
- No geral, a maioria dos candidatos escolheram o curso em razão de maiores oportunidades de trabalho.

Os resultados obtidos podem auxiliar setores da Universidade. Por exemplo, identificadas as deficiências dos alunos com relação a desempenho na maioria das disciplinas, ações poderiam ser tomadas em conjunto com a comunidade regional para amenizar tal problema, na forma de cursos preparatórios para o ENEM (atualmente única forma de ingresso na universidade) ou mesmo programas de nivelamento para os calouros.

Além disso, os resultados envolvendo os cursos das áreas de saúde (Odontologia e Farmácia) e agrárias (Agronomia, Engenharia Florestal e Zootecnia) atestam que é preciso que as autoridades educacionais dos Vales do Jequitinhonha e Mucuri se reúnam e verifiquem a possibilidade de se realizar ações que aumentem as chances de acesso dos estudantes dos Vales aos cursos das áreas de saúde e agrárias, uma vez que eles são atualmente compostos, em sua essência, por candidatos provenientes de outras regiões do estado de Minas Gerais.

Outro exemplo de ação que poderia ser tomada seria o investimento na divulgação da UFVJM e seus cursos, pesquisas, programas de extensão, entre outros, pela internet e em outros meios de comunicação nas diversas regiões.

Os cursos de Ciências Contábeis, Bacharelado em Humanidades e Bacharelado em Ciência e Tecnologia possuem uma boa porcentagem de aprovados que exercem atividade remunerada. Dessa forma, outra possível ação seria uma maior implementação de programas de maior apoio extra-classe para esses alunos como monitoria, por exemplo, já que esses candidatos estudam e trabalham e, portanto, fornecer um maior apoio a eles seria de grande valia.

Como foi possível perceber, a maioria dos candidatos aprovados no vestibular não exercem atividade remunerada. Sendo assim, constitui-se como outra possível ação a ser realizada pela instituição uma maior implementação em programas extra-curriculares como monitoria, bolsas de pesquisa, entre outros, uma vez que tais candidatos, por não exercerem atividade, dispõem de maior tempo livre para se dedicarem a esses programas.

Conforme foi possível perceber ao longo da Seção 8.1, um dos grandes problemas do uso do algoritmo *Apriori* é que este pode gerar um número muito grande de regras a serem

analisadas, o que levou a combinação de no máximo três atributos de cada base de dados. Dessa forma, como um dos trabalhos futuros, propõe-se encontrar maneiras de simplificar o processo de análise por parte do usuário, reduzindo o número de regras apresentadas a ele. Para exemplificar isso, considera-se os cursos da área de agrárias. Se Agronomia, Engenharia Florestal e Zootecnia são essencialmente compostos por candidatos que não exercem atividade remunerada, logo os cursos de agrárias, como um todo, são formados por maioria de candidatos que não exercem atividade remunerada. Isso facilitaria a análise reduzindo o número de regras apresentadas de três para uma.

Quando se utiliza bases com grande quantidade de dados, a tarefa de analisar regras torna-se ainda mais difícil, pois além do número ainda maior de regras que são produzidas, há uma segunda preocupação que é como conseguir representar os resultados de uma maneira simples, fácil de ser entendida pelos leitores. Neste trabalho foi utilizado o método mais comum de exibição de resultados em tabelas. Como segunda proposta, sugere-se buscar por novas formas de representar esses resultados para, assim, facilitar a visualização e compreensão por parte dos leitores.

Por fim, como foi possível perceber, em algumas análises houveram situações em que os resultados se modificaram de um semestre para outro, principalmente no primeiro e segundo semestre de 2009. Sendo assim, a Mineração em bases de dados posteriores poderiam confirmar se estas modificações se mantêm nos semestres seguintes ou se constituíram apenas uma exceção.

## REFERÊNCIAS

- [1] ABERNETHY, M. *Mineração de Dados com Weka*. 2010. Disponível em: <<http://www.ibm.com/developerworks/br/opensource/library/os-weka1/index.html>>. Acesso em: 13 out. 2011.
- [2] AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. *Mining Association Rules between Sets of Items in Large Databases*. Washington, 1993. Disponível em: <<http://rakesh.agrawal-family.com/papers/sigmod93assoc.pdf>>. Acesso em: 17 dez. 2011.
- [3] AGRAWAL, R.; SRIKANT, R. *Fast Algorithms for Mining Association Rules*. Santiago, 1994. Disponível em: <<http://rakesh.agrawal-family.com/papers/vldb94apriori.pdf>>. Acesso em: 19 dez. 2011.
- [4] AMO, S. de. *Técnicas de Mineração de Dados*. Uberlândia, 2004. Disponível em: <<http://www.deamo.prof.ufu.br/arquivos/JAI-cap5.pdf>>. Acesso em: 26 out. 2011.
- [5] ARBEX, E. C.; SABOREDO, A. de P.; MIRANDA, D. *Implementação e estudo de caso do algoritmo Apriori para Mineração de Dados*. Resende, 2006. Disponível em: <[http://www.aedb.br/seget/artigos04/193\\_artigo\\_SEGET.doc](http://www.aedb.br/seget/artigos04/193_artigo_SEGET.doc)>. Acesso em: 30 nov. 2011.
- [6] BARBOSA, D. C. C. *Mineração de Dados usando o Software WizRule em Base de Dados de Compras de Itens de TI*. Dissertação (Mestrado Profissionalizante em Administração) — Faculdade de Economia e Finanças IBMEC, Rio de Janeiro, 2007.
- [7] BARROSO, B. da C.; NETO, P. N. F. *Descoberta de conhecimento na base de dados de uma locadora de filmes*. Monografia (Graduação em Ciência da Computação) — Universidade Federal do Pará, Belém, 2006.
- [8] CARDOSO, O. N. P.; MACHADO, R. T. M. *Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras*. Lavras, 2008. Disponível em: <<http://scielo.br/pdf/rap/v42n3/a04v42n3.pdf>>. Acesso em: 17 out. 2011.
- [9] CARVALHO, L. A. V. de. *Data Mining: A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração*. Rio de Janeiro: Editora Ciência Moderna Ltda, 2005.
- [10] CASTANHEIRA, L. G. *Aplicação de Técnicas de Mineração de Dados em Problemas de Classificação de Padrões*. Dissertação (Pós-Graduação em Engenharia Elétrica) — Universidade Paulista, Belo Horizonte, 2008.
- [11] CÔRTEZ, S. da C.; PORCARO, R. M.; LIFSCHITZ, S. *Mineração de Dados - Funcionalidades, Técnicas e Abordagens*. Rio de Janeiro, 2002. Disponível em: <[ftp://139.82.16.194/pub/docs/techreports/02\\_10\\_cortes.pdf](ftp://139.82.16.194/pub/docs/techreports/02_10_cortes.pdf)>. Acesso em: 25 nov. 2011.
- [12] DANTAS, E. R. G. et al. *O Uso da Descoberta de Conhecimento em Base de Dados para Apoiar a Tomada de Decisões*. João Pessoa, 2008. Disponível em: <[http://www.aedb.br/seget/artigos08/331\\_331\\_Artigo\\_SEGET\\_EJDR\\_Versao\\_Final\\_010808.pdf](http://www.aedb.br/seget/artigos08/331_331_Artigo_SEGET_EJDR_Versao_Final_010808.pdf)>. Acesso em: 19 out. 2011.

- [13] FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. *From Data Mining to Knowledge Discovery in Databases*. AI Magazine, Providence, jul. 1997. Disponível em: <<http://www.aaai.org/aitopics/assets/PDF/AIMag17-03-2-article.pdf>>. Acesso em: 26 out. 2011.
- [14] FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. *The KDD Process for Extracting Useful Knowledge from Volumes of Data*. Communications of the ACM, v. 39, n.11, nov. 1996. Disponível em: <[http://shawndra.pbworks.com/f/The KDD process for extracting useful knowledge from volumes of data.pdf](http://shawndra.pbworks.com/f/The+KDD+process+for+extracting+useful+knowledge+from+volumes+of+data.pdf)>. Acesso em: 26 out. 2011.
- [15] FILHO, P. D. *Análise e Critérios de Tomada de Decisões Baseados na Técnica de Data Mining*. Dissertação (Pós-Graduação em Engenharia de Produção) — Universidade Paulista, São Paulo, 2011.
- [16] FREITAS, J. A. da S. *Uso de Técnicas de Data Mining para Análise de Bases de Dados Hospitalares com Finalidades de Gestão*. Tese (Doutorado em Ciências Empresariais) — Universidade do Porto, Porto, 2006.
- [17] GONÇALVES, E. C. *Regras de Associação e suas Medidas de Interesse Objetivas e Subjetivas*. Niterói, 2005. Disponível em: <<http://www.dcc.ufla.br/infocomp/artigos/v4.1/art04.pdf>>. Acesso em: 18 out. 2011.
- [18] GONÇALVES, E. C. *Mineração de Dados (Data Mining)*. Escola Nacional de Ciências Estatísticas. Rio de Janeiro, 2010.
- [19] HAN, J.; KAMBER, M. *Data Mining: Concepts and Techniques*. 2. ed. San Francisco: Morgan Kaufmann, 2005.
- [20] LEMOS, E. P.; STEINER, M. T. A.; NIEVOLA, J. C. *Análise de crédito bancário por meio de redes neurais e árvores de decisão: uma aplicação simples de data mining*. São Paulo, 2005. Disponível em: <<http://www.rausp.usp.br/download.asp?file=V4003225.pdf>>. Acesso em: 11 set. 2011.
- [21] MATTOS, M. C. de; SELINGER, T. C.; SIMÕES, P. W. T. de. *Data Mining em Saúde Bucal por meio da Técnica de Clusterização e do Algoritmo K-Means*. Santa Catarina, 2004. Disponível em: <<http://www.sbis.org.br/cbis9/arquivos/204.pdf>>. Acesso em: 31 ago. 2011.
- [22] MYSQL. *MySQL - The world's most popular open source database*. 2011. Web site da ferramenta. Disponível em: <<http://www.mysql.com>>. Acesso em: 14 out. 2011.
- [23] OCHI, L. S.; DIAS, C. R.; SOARES, S. S. F. *Clusterização em Mineração de Dados*. Niterói, 2004. Disponível em: <<http://www.ic.uff.br/~satoru/conteudo/artigos/ERI-Minicurso-SATORU.pdf>>. Acesso em: 19 out. 2011.
- [24] OLIVEIRA, A. G. de; GARCIA, D. F. *Mineração da Base de Dados de um Processo Seletivo Universitário*. Formiga, 2004. Disponível em: <<http://www.dcc.ufla.br/infocomp/artigos/v3.2/art07.pdf>>. Acesso em: 27 ago. 2011.
- [25] PIRES, M. *Mineração de Dados no Estudo de Caso do Vestibular da UFVJM*. Monografia (Graduação em Sistemas de Informação) — Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina, 2010.

- [26] PITONI, R. M. *Mineração de Regras de Associação nos Canais de Informação do Direto*. Dissertação (Trabalho de Diplomação em Ciência da Computação) — Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.
- [27] REZENDE, S. O. *Mineração de Dados*. São Leopoldo, 2005. Disponível em: <[http://www.addlabs.uff.br/enia\\_site/dw/mineracaodedados.pdf](http://www.addlabs.uff.br/enia_site/dw/mineracaodedados.pdf)>. Acesso em: 28 dez. 2011.
- [28] ROMÃO, W. *Descoberta de Conhecimento Relevante em Banco de Dados sobre Ciência e Tecnologia*. Tese (Doutorado em Engenharia de Produção) — Universidade Federal de Santa Catarina, Florianópolis, 2002.
- [29] ROMÃO, W. et al. *Extração de Regras de Associação em C&T: o Algoritmo Apriori*. Florianópolis, 1999. Disponível em: <[http://www.abepro.org.br/biblioteca/ENEGEP1999\\_A0901.pdf](http://www.abepro.org.br/biblioteca/ENEGEP1999_A0901.pdf)>. Acesso em: 19 out. 2011.
- [30] SCHONHORST, G. B. *Mineração de Regras de Associação Aplicada à Modelagem dos Dados Transacionais de um Supermercado*. Dissertação (Pós-Graduação em Engenharia de Produção) — Universidade Federal de Itajubá, Itajubá, 2010.
- [31] SFERRA, H. H.; CORRÊA, A. M. C. J. *Conceitos e Aplicações de Data Mining*. Piracicaba, 2003. Disponível em: <<http://www.unimep.br/phpg/editora/revistaspdf/rct22art02.pdf>>. Acesso em: 15 set. 2011.
- [32] SILBERCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. *Sistema de Banco de Dados*. 5. ed. Rio de Janeiro: Editora Campus, 2006.
- [33] SILVA, C. F.; RODRIGUES, C. T.; MONTEIRO, M. V. B. *Uso de Regras de Associação para Descoberta de Conhecimento na Produtividade de Açaí no Estado do Amapá*. Macapá, 2010. Disponível em: <<http://periodicos.unesc.net/index.php/sulcomp/article/view/297/304>>. Acesso em: 3 jan. 2012.
- [34] SILVA, G. C. *Mineração de Regras de Associação Aplicada a Dados da Secretaria Municipal de Saúde de Londrina - PR*. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.
- [35] SILVA, M. P. da; BOSCARIOLI, C.; PERES, S. M. *Análise de logs da Web por meio de técnicas de Data Mining*. Cascavel, 2003. Disponível em: <[http://conged.deinfo.uepg.br/ iconged/Artigos/Artigo\\_03.pdf](http://conged.deinfo.uepg.br/ iconged/Artigos/Artigo_03.pdf)>. Acesso em: 5 jan. 2012.
- [36] SILVA, M. P. dos S. *Conceitos, Aplicações e Experimentos com Weka*. São José dos Campos, 2004. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/erirjes/2004/004.pdf>>. Acesso em: 13 out. 2011.
- [37] SOUZA, H. B. de; NAGLIS, D. L. *Mineração de Dados: uma aplicação na base de dados de artigos periódicos científicos das áreas de informação (ABCDM)*. Monografia (Especialização em Engenharia Elétrica) — Universidade de Brasília, Brasília, 2008.
- [38] SUEHRING, S. *MySQL a Bíblia*. 7. ed. São Paulo: Elsevier, 2002.

- [39] TAN, P.-N.; STEINBACH, M.; KUMAR, V. *Introdução ao Data Mining Mineração de Dados*. Rio de Janeiro: Editora Ciência Moderna Ltda, 2009.
- [40] UNIVERSITY OF WAIKATO. *Weka 3 – Machine Learning Software in Java*. 2011. Web site da ferramenta. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka>>. Acesso em: 14 out. 2011.
- [41] VIANNA, R. C. X. F. et al. *Mineração de dados e características da mortalidade infantil*. Curitiba, 2010. Disponível em: <<http://www.scielo.org/pdf/csp/v26n3/11.pdf>>. Acesso em: 12 set. 2011.
- [42] WITTEN, I. H.; FRANK, E. *Data Mining: Pratical Machine Learning Tools and Techniques*. 2. ed. San Francisco: Elsevier, 2005.



## **APÊNDICE A – REGRAS GERADAS PELO ALGORITMO *APRIORI***

Nas tabelas seguintes, estão listadas as regras de associação geradas pelo algoritmo *Apriori* com seus respectivos valores de suporte, confiança e *lift* para cada combinação de atributos.

**Tabela 17:** Regras de associação envolvendo os atributos local de residência e curso

Ano/Sem	Consequente da Regra (local de residência)											
	Diamantina (D)				Teófilo Otoni (TO)				Outra Região de Minas Gerais (OR)			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2005/1									Engenharia Florestal 212 ⇒ OR 139	0.05	0.66	1.18
									Farmácia-Bioq./A. Clínica <sup>1</sup> 254 ⇒ OR 159	0.05	0.63	1.13
									Agronomia 199 ⇒ OR 121	0.04	0.61	1.1
									Odontologia 418 ⇒ OR 241	0.08	0.58	1.04
									Zootecnia 196 ⇒ OR 111	0.04	0.57	1.02
									Farmácia-Bioq./Industrial <sup>2</sup> 192 ⇒ OR 107	0.04	0.56	1.01
2005/2									Agronomia 204 ⇒ OR 138	0.06	0.68	1.18
									Engenharia Florestal 206 ⇒ OR 134	0.06	0.65	1.13
									Odontologia 417 ⇒ OR 268	0.11	0.64	1.12
									Zootecnia 156 ⇒ OR 100	0.04	0.64	1.12
									Farmácia-Bioq./A. Clínica 282 ⇒ OR 163	0.07	0.58	1.01
2007/1					Ciências Contábeis 82 ⇒ TO 51	0.02	0.62	5.95	Agronomia 142 ⇒ OR 90	0.03	0.63	1.49
					Administração 166 ⇒ TO 103	0.03	0.62	5.94	Odontologia 465 ⇒ OR 293	0.09	0.63	1.48
					Serviço Social 163 ⇒ TO 83	0.03	0.51	4.87	Zootecnia 118 ⇒ OR 73	0.02	0.62	1.45
									Engenharia Florestal 172 ⇒ OR 98	0.03	0.57	1.34
									Farmácia 349 ⇒ OR 197	0.06	0.56	1.33
2007/2					Administração 177 ⇒ TO 112	0.04	0.63	4.54	Zootecnia 85 ⇒ OR 60	0.02	0.71	1.72
					Ciências Econômicas 72 ⇒ TO 41	0.02	0.57	4.09	Agronomia 129 ⇒ OR 79	0.03	0.61	1.49
					Ciências Contábeis 104 ⇒ TO 58	0.02	0.56	4	Odontologia 389 ⇒ OR 232	0.09	0.6	1.46
					Serviço Social 190 ⇒ TO 100	0.04	0.53	3.78	Engenharia Florestal 189 ⇒ OR 108	0.04	0.57	1.39
								Farmácia 378 ⇒ OR 191	0.07	0.51	1.23	

Continua na página seguinte...

<sup>1</sup>Farmácia-Bioquímica/Análise Clínica

<sup>2</sup>Farmácia-Bioquímica/Industrial

**Tabela 17:** Regras de associação envolvendo os atributos local de residência e curso (cont)

Ano/Sem	Consequente da Regra (local de residência)											
	Diamantina (D)				Teófilo Otoni (TO)				Outra Região de Minas Gerais (OR)			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2008/1	Educação Física 124 ⇒ D 67	0.02	0.54	3.13	Administração 281 ⇒ TO 176	0.05	0.63	3.67	Zootecnia 103 ⇒ OR 66	0.02	0.64	1.82
					Ciências Contábeis 138 ⇒ TO 82	0.02	0.59	3.48	Agronomia 122 ⇒ OR 72	0.02	0.59	1.68
					Serviço Social 288 ⇒ TO 167	0.05	0.58	3.4	Odontologia 474 ⇒ OR 275	0.08	0.58	1.65
					Ciências Econômicas 114 ⇒ TO 64	0.02	0.56	3.29	Engenharia Florestal 220 ⇒ OR 117	0.03	0.53	1.51
2008/2	Educação Física 87 ⇒ D 52	0.03	0.6	2.77	Administração 129 ⇒ TO 77	0.04	0.6	4.57	Zootecnia 48 ⇒ OR 26	0.01	0.54	1.7
					Ciências Contábeis 84 ⇒ TO 46	0.02	0.55	4.19	Agronomia 109 ⇒ OR 54	0.03	0.5	1.56
					Ciências Econômicas 61 ⇒ TO 32	0.02	0.52	4.02				
					Serviço Social 134 ⇒ TO 70	0.03	0.52	4				
2009/1	BHU <sup>3</sup> 185 ⇒ D 120	0.03	0.65	3.51	Administração 286 ⇒ TO 184	0.05	0.64	2.84	Odontologia 391 ⇒ OR 202	0.06	0.52	2.08
					Ciências Contábeis 133 ⇒ TO 82	0.02	0.62	2.72				
					BCT - TO <sup>4</sup> 447 ⇒ TO 281	0.08	0.63	2.77				
					Serviço Social 253 ⇒ TO 143	0.04	0.57	2.49				
2009/2	BHU 186 ⇒ D 121	0.05	0.65	3.13	Administração 140 ⇒ TO 87	0.04	0.62	4	Zootecnia 61 ⇒ OR 36	0.02	0.59	1.79
	Educação Física 59 ⇒ D 37	0.02	0.63	3.01	Serviço Social 105 ⇒ TO 60	0.03	0.57	3.68	Odontologia 283 ⇒ OR 157	0.07	0.55	1.68
					Ciências Contábeis 78 ⇒ TO 43	0.02	0.55	3.55	Engenharia Florestal 127 ⇒ OR 67	0.03	0.53	1.6
					Ciência Econômicas 56 ⇒ TO 29	0.01	0.52	3.33	BCT - D <sup>5</sup> 286 ⇒ OR 154	0.07	0.54	1.63

<sup>3</sup>Bacharelado em Humanidades

<sup>4</sup>Bacharelado em Ciência e Tecnologia - Teófilo Otoni

<sup>5</sup>Bacharelado em Ciência e Tecnologia - Diamantina

**Tabela 18:** Regras de associação envolvendo os atributos tipo da escola e curso

Ano/Sem	Consequente da regra (tipo da escola)							
	Escola Estadual				Escola Particular			
	Descrição da regra	Sup	Conf	Lift	Descrição da regra	Sup	Conf	Lift
2002/1	Engenharia Florestal 107 ⇒ Escola Estadual 64	0.02	0.6	1.3	Odontologia 628 ⇒ Escola Particular 404	0.13	0.64	1.38
	Agronomia 90 ⇒ Escola Estadual 46	0.01	0.51	1.11	Farmácia-Bioquímica/Industrial 241 ⇒ Escola Particular 138	0.04	0.57	1.23
	Nutrição 338 ⇒ Escola Estadual 194	0.06	0.57	1.25				
2002/2	Engenharia Florestal 171 ⇒ Escola Estadual 95	0.04	0.56	1.18	Odontologia 442 ⇒ Escola Particular 264	0.12	0.6	1.37
	Zootecnia 174 ⇒ Escola Estadual 93	0.04	0.53	1.14	Farmácia-Bioquímica/Industrial 192 ⇒ Escola Particular 103	0.05	0.54	1.23
	Nutrição 370 ⇒ Escola Estadual 194	0.09	0.52	1.12				
	Agronomia 137 ⇒ Escola Estadual 70	0.03	0.51	1.09				
	Fisioterapia 468 ⇒ Escola Estadual 238	0.11	0.51	1.08				
2003/1	Zootecnia 232 ⇒ Escola Estadual 118	0.04	0.51	1.11	Farmácia-Bioquímica/Industrial 224 ⇒ Escola Particular 122	0.04	0.54	1.28
	Engenharia Florestal 191 ⇒ Escola Estadual 97	0.03	0.51	1.11	Odontologia 475 ⇒ Escola Particular 251	0.08	0.53	1.24
	Nutrição 364 ⇒ Escola Estadual 184	0.06	0.51	1.1				
2003/2	Nutrição 316 ⇒ Escola Estadual 184	0.09	0.58	1.22	Odontologia 405 ⇒ Escola Particular 222	0.11	0.55	1.28
	Zootecnia 127 ⇒ Escola Estadual 69	0.03	0.54	1.14	Farmácia-Bioquímica/Industrial 182 ⇒ Escola Particular 96	0.05	0.53	1.23
	Fisioterapia 479 ⇒ Escola Estadual 252	0.12	0.53	1.1				
	Engenharia Florestal 152 ⇒ Escola Estadual 76	0.04	0.5	1.05				
2005/1	Nutrição 286 ⇒ Escola Estadual 179	0.06	0.63	1.31	Odontologia 419 ⇒ Escola Particular 260	0.09	0.62	1.33
	Engenharia Florestal 212 ⇒ Escola Estadual 118	0.04	0.56	1.16	Farmácia-Bioquímica/Industrial 192 ⇒ Escola Particular 105	0.04	0.55	1.17
	Agronomia 199 ⇒ Escola Estadual 104	0.04	0.52	1.09				
2005/2	Nutrição 323 ⇒ Escola Estadual 184	0.08	0.57	1.24	Odontologia 417 ⇒ Escola Particular 244	0.1	0.59	1.22
	Fisioterapia 348 ⇒ Escola Estadual 181	0.08	0.52	1.14	Farmácia-Bioquímica/Industrial 225 ⇒ Escola Particular 120	0.05	0.53	1.11
2007/1	Serviço Social 163 ⇒ Escola Estadual 135	0.04	0.83	1.58	Odontologia 465 ⇒ Escola Particular 310	0.09	0.67	1.59
	Ciências Contábeis 82 ⇒ Escola Estadual 62	0.02	0.76	1.44	Zootecnia 118 ⇒ Escola Particular 67	0.02	0.57	1.36
	Ciências Econômicas 73 ⇒ Escola Estadual 53	0.02	0.73	1.39	Farmácia 349 ⇒ Escola Particular 181	0.05	0.52	1.24
	Educação Física 145 ⇒ Escola Estadual 100	0.03	0.69	1.32				
	Administração 166 ⇒ Escola Estadual 112	0.03	0.67	1.29				

Continua na página seguinte...

**Tabela 18:** Regras de associação envolvendo os atributos tipo da escola e curso (cont)

Ano/Sem	Consequente da regra (tipo da escola)											
	Escola Estadual				Escola Particular							
	Descrição da regra	Sup	Conf	Lift	Descrição da regra	Sup	Conf	Lift				
<b>2007/1</b> (cont)	Sistemas de Informação 178 ⇒ Escola Estadual 116	0.04	0.65	1.24								
	Ciências Biológicas 199 ⇒ Escola Estadual 108	0.03	0.54	1.04								
	Nutrição 174 ⇒ Escola Estadual 92	0.03	0.53	1.01								
<b>2007/2</b>	Ciências Contábeis 104 ⇒ Escola Estadual 88	0.03	0.85	1.44	Odontologia 389 ⇒ Escola Particular 228	0.08	0.59	1.63				
	Serviço Social 190 ⇒ Escola Estadual 160	0.06	0.84	1.43								
	Administração 177 ⇒ Escola Estadual 143	0.05	0.81	1.38								
	Ciências Econômicas 72 ⇒ Escola Estadual 55	0.02	0.76	1.3								
	Educação Física 124 ⇒ Escola Estadual 89	0.03	0.72	1.22								
	Fisioterapia 203 ⇒ Escola Estadual 132	0.05	0.65	1.11								
	Sistemas de Informação 131 ⇒ Escola Estadual 82	0.03	0.63	1.07								
<b>2008/1</b>	Serviço Social 288 ⇒ Escola Estadual 243	0.07	0.84	1.37	Odontologia 474 ⇒ Escola Particular 279	0.08	0.59	1.74				
	Ciências Contábeis 138 ⇒ Escola Estadual 113	0.03	0.82	1.33					Farmácia 347 ⇒ Escola Particular 182	0.05	0.52	1.55
	Educação Física 124 ⇒ Escola Estadual 100	0.03	0.81	1.31					Zootecnia 103 ⇒ Escola Particular 52	0.02	0.5	1.49
	Administração 281 ⇒ Escola Estadual 221	0.06	0.79	1.28								
	Sistemas de Informação 166 ⇒ Escola Estadual 123	0.04	0.74	1.21								
	Ciências Econômicas 14 ⇒ Escola Estadual 82	0.02	0.72	1.17								
	Ciências Biológicas 185 ⇒ Escola Estadual 122	0.04	0.66	1.07								
<b>2008/2</b>	Serviço Social 134 ⇒ Escola Estadual 122	0.06	0.91	1.48	Odontologia 309 ⇒ Escola Particular 167	0.08	0.54	1.61				
	Educação Física 87 ⇒ Escola Estadual 75	0.04	0.86	1.4					Zootecnia 48 ⇒ Escola Particular 25	0.01	0.52	1.55
	Ciências Contábeis 84 ⇒ Escola Estadual 71	0.04	0.85	1.38					Farmácia 253 ⇒ Escola Particular 128	0.06	0.51	1.51
	Administração 129 ⇒ Escola Estadual 107	0.05	0.83	1.35								
	Sistemas de Informação 108 Escola Estadual 85	0.04	0.79	1.28								
	Ciências Econômicas 61 ⇒ Escola Estadual 44	0.02	0.72	1.17								
	Ciências Biológicas 129 ⇒ Escola Estadual 92	0.05	0.71	1.16								
	Nutrição 138 ⇒ Escola Estadual 88	0.04	0.64	1.04								
	Fisioterapia 109 ⇒ Escola Estadual 68	0.03	0.62	1.02								

Continua na página seguinte...

**Tabela 18:** Regras de associação envolvendo os atributos tipo da escola e curso (cont)

Consequente da regra (tipo da escola)								
Escola Estadual					Escola Particular			
Ano/Sem	Descrição da regra	Sup	Conf	Lift	Descrição da regra	Sup	Conf	Lift
2009/1	Ciências Contábeis 133 ⇒ Escola Estadual 119	0.03	0.89	1.32	Odontologia 391 ⇒ Escola Particular 221	0.06	0.57	2.07
	Serviço Social 253 ⇒ Escola Estadual 222	0.06	0.88	1.3				
	Educação Física 88 ⇒ Escola Estadual 76	0.02	0.86	1.28				
	Ciências Econômicas 81 ⇒ Escola Estadual 66	0.02	0.81	1.2				
	Bacharelado em Humanidades 185 ⇒ Escola Estadual 130	0.04	0.81	1.2				
	Administração 286 ⇒ Escola Estadual 231	0.07	0.81	1.19				
	Sistemas de Informação 128 ⇒ Escola Estadual 101	0.03	0.79	1.17				
	Ciências Biológicas 128 ⇒ Escola Estadual 95	0.03	0.74	1.1				
2009/2	Serviço Social 105 ⇒ Escola Estadual 92	0.04	0.88	1.39	Odontologia 283 ⇒ Escola Particular 167	0.07	0.59	1.86
	Educação Física 59 ⇒ Escola Estadual 49	0.02	0.83	1.32				
	Ciências Contábeis 78 ⇒ Escola Estadual 64	0.03	0.82	1.3				
	Administração 140 ⇒ Escola Estadual 113	0.05	0.81	1.28				
	Bacharelado em Humanidades 186 ⇒ Escola Estadual 150	0.07	0.81	1.28				
	Ciências Econômicas 56 ⇒ Escola Estadual 45	0.02	0.8	1.28				
	Ciências Biológicas 112 ⇒ Escola Estadual 81	0.04	0.72	1.15				
	Nutrição 83 ⇒ Escola Estadual 60	0.03	0.72	1.15				
	Fisioterapia 87 ⇒ Escola Estadual 61	0.03	0.7	1.11				
	Agronomia 75 ⇒ Escola Estadual 48	0.02	0.64	1.02				

**Tabela 19:** Regras de associação envolvendo os atributos atividade remunerada e curso

Ano/Sem	Antecedente da regra (curso)			
	Descrição da regra	Sup	Conf	Lift
2002/1	Odontologia 628 ⇒ Atividade remunerada=Não 559	0.18	0.89	1.1
	Farmácia-Bioquímica/Análise Clínica 400 ⇒ Atividade remunerada=Não 335	0.11	0.84	1.03
	Farmácia-Bioquímica/Industrial 241 ⇒ Atividade remunerada=Não 199	0.06	0.83	1.02
	Fisioterapia 616 ⇒ Atividade remunerada=Não 503	0.16	0.82	1.01
2002/2	Odontologia 442 ⇒ Atividade remunerada=Não 387	0.18	0.88	1.11
	Fisioterapia 468 ⇒ Atividade remunerada=Não 384	0.17	0.82	1.04
2003/1	Farmácia-Bioquímica/Industrial 224 ⇒ Atividade remunerada=Não 191	0.06	0.85	1.08
	Fisioterapia 531 ⇒ Atividade remunerada=Não 435	0.14	0.82	1.04
	Farmácia-Bioquímica/Análise Clínica 229 ⇒ Atividade remunerada=Não 186	0.06	0.81	1.03
	Odontologia 475 ⇒ Atividade remunerada 384	0.13	0.81	1.03
2003/2	Nutrição 364 ⇒ Atividade remunerada=Não 291	0.1	0.8	1.02
	Odontologia 405 ⇒ Atividade remunerada=Não 358	0.17	0.88	1.09
	Farmácia-Bioquímica/Análise Clínica 294 ⇒ Atividade remunerada=Não 254	0.12	0.86	1.07
2004/1	Fisioterapia 479 ⇒ Atividade remunerada=Não 393	0.19	0.82	1.01
	Nutrição 367 ⇒ Atividade remunerada=Não 230	0.07	0.63	1.15
2004/2	Farmácia-Bioquímica/Industrial 241 ⇒ Atividade remunerada=Não 147	0.05	0.61	1.12
	Enfermagem 691 ⇒ Atividade remunerada=Não 411	0.13	0.59	1.09
2005/1	Farmácia-Bioquímica/Industrial 209 ⇒ Atividade remunerada=Não 109	0.06	0.52	1.21
	Farmácia-Bioquímica/Análise Clínica 254 ⇒ Atividade remunerada=Não 233	0.08	0.92	1.08
	Farmácia-Bioquímica/Industrial 192 ⇒ Atividade remunerada=Não 176	0.06	0.92	1.07
	Odontologia 418 ⇒ Atividade remunerada=Não 374	0.13	0.89	1.05
2005/2	Fisioterapia 411 ⇒ Atividade remunerada=Não 361	0.12	0.88	1.03
	Odontologia 41 ⇒ Atividade remunerada=Não 378	0.16	0.91	1.05
	Farmácia-Bioquímica/Industrial 225 ⇒ Atividade remunerada=Não 202	0.08	0.9	1.04
2007/1	Nutrição 323 ⇒ Atividade remunerada=Não 284	0.12	0.88	1.01
	Farmácia-Bioquímica/Análise Clínica 282 ⇒ Atividade remunerada=Não 246	0.1	0.87	1.01
	Farmácia 349 ⇒ Atividade remunerada=Não 324	0.1	0.93	1.15
	Odontologia 465 ⇒ Atividade remunerada=Não 431	0.13	0.93	1.15
	Zootecnia 118 ⇒ Atividade remunerada=Não 106	0.03	0.9	1.11
	Nutrição 174 ⇒ Atividade remunerada=Não 154	0.05	0.89	1.1
	Agronomia 142 ⇒ Atividade remunerada=Não 123	0.04	0.87	1.07
	Fisioterapia 228 ⇒ Atividade remunerada=Não 195	0.06	0.86	1.06
2007/2	Enfermagem 411 ⇒ Atividade remunerada=Não 366	0.11	0.89	1.1
	Engenharia Florestal 172 ⇒ Atividade remunerada=Não 141	0.04	0.82	1.02
	Odontologia 389 ⇒ Atividade remunerada=Não 353	0.13	0.91	1.15
	Farmácia 378 ⇒ Atividade remunerada=Não 340	0.13	0.9	1.14
	Fisioterapia 203 ⇒ Atividade remunerada=Não 181	0.07	0.89	1.13
	Engenharia Florestal 189 ⇒ Atividade remunerada=Não 163	0.06	0.86	1.1
	Agronomia 129 ⇒ Atividade remunerada=Não 111	0.04	0.86	1.09
Nutrição 186 ⇒ Atividade remunerada=Não 158	0.06	0.85	1.08	
Zootecnia 85 ⇒ Atividade remunerada=Não 72	0.03	0.85	1.08	

Continua na página seguinte...

**Tabela 19:** Regras de associação envolvendo os atributos atividade remunerada e curso (cont)

Ano/Sem	Antecedente da regra (curso)			
	Descrição da regra	Sup	Conf	Lift
2008/1	Farmácia 347 ⇒ Atividade remunerada=Não 319	0.09	0.92	1.21
	Odontologia 474 ⇒ Atividade remunerada=Não 427	0.12	0.9	1.19
	Nutrição 219 ⇒ Atividade remunerada=Não 197	0.06	0.9	1.19
	Agronomia 122 ⇒ Atividade remunerada=Não 107	0.03	0.88	1.16
	Fisioterapia 161 ⇒ Atividade remunerada=Não 141	0.04	0.88	1.16
	Enfermagem 279 ⇒ Atividade remunerada=Não 233	0.07	0.84	1.1
	Engenharia Florestal 220 ⇒ Atividade remunerada=Não 183	0.05	0.83	1.1
	Zootecnia 103 ⇒ Atividade remunerada=Não 81	0.02	0.79	1.04
2008/2	Odontologia 309 ⇒ Atividade remunerada=Não 282	0.14	0.91	1.2
	Farmácia 253 ⇒ Atividade remunerada=Não 229	0.11	0.91	1.19
	Nutrição 138 ⇒ Atividade remunerada=Não 120	0.06	0.87	1.14
	Zootecnia 48 ⇒ Atividade remunerada=Não 41	0.02	0.85	1.12
	Fisioterapia 109 ⇒ Atividade remunerada=Não 88	0.04	0.81	1.06
2009/1	Odontologia 391 ⇒ Atividade remunerada=Não 353	0.1	0.9	1.27
	Nutrição 112 ⇒ Atividade remunerada=Não 99	0.03	0.88	1.24
	Zootecnia 73 ⇒ Atividade remunerada=Não 63	0.02	0.86	1.21
	Farmácia 216 ⇒ Atividade remunerada=Não 183	0.05	0.85	1.19
	Engenharia Florestal 204 ⇒ Atividade remunerada=Não 169	0.05	0.83	1.16
	Agronomia 120 ⇒ Atividade remunerada=Não 99	0.03	0.83	1.16
	Enfermagem 185 ⇒ Atividade remunerada=Não 152	0.04	0.82	1.15
	Fisioterapia 111 ⇒ Atividade remunerada=Não 88	0.03	0.79	1.11
	Bacharelado em Ciência e Tecnologia - Diamantina 273 ⇒ Atividade remunerada=Não 213	0.06	0.78	1.09
2009/2	Odontologia 283 ⇒ Atividade remunerada=Não 263	0.12	0.93	1.28
	Fisioterapia 87 ⇒ Atividade remunerada=Não 77	0.03	0.89	1.22
	Farmácia 173 ⇒ Atividade remunerada=Não 149	0.07	0.86	1.19
	Nutrição 83 ⇒ Atividade remunerada=Não 71	0.03	0.86	1.18
	Bacharelado em Ciência e Tecnologia - Diamantina 286 ⇒ Atividade remunerada=Não 244	0.11	0.85	1.18
	Agronomia 75 ⇒ Atividade remunerada=Não 63	0.03	0.84	1.16
	Zootecnia 61 ⇒ Atividade remunerada=Não 48	0.02	0.79	1.08



**Tabela 20:** Regras de associação envolvendo os atributos frequência pré-vestibular e notas

Ano/Sem	Regras em que há alguma frequência em pré-vestibular				Regras em que não há frequência em pré-vestibular			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2003/1	Sim, por menos de um semestre 353 ⇒ Biologia=Péssimo 210	0.07	0.59	1.1	Não 685 ⇒ Biologia=Péssimo 434	0.14	0.63	1.17
	Sim, por um semestre 563 ⇒ Biologia=Péssimo 324	0.11	0.58	1.07	Não 685 ⇒ Matemática=Péssimo 362	0.12	0,53	1.07
	Sim, por menos de um semestre 353 ⇒ Matemática=Péssimo 190	0.06	0.54	1.09	Não 685 ⇒ Física=Péssimo 544	0.18	0.79	1.05
	Sim, por um semestre 563 ⇒ Matemática=Péssimo 295	0.1	0.52	1.06	Não 685 ⇒ História=Péssimo 393	0.13	0.57	1.04
	Sim, por um semestre 563 ⇒ Física=Péssimo 448	0.15	0.8	1.05	Não 685 ⇒ Língua Estrangeira=Péssimo 364	0.12	0.53	1.16
	Sim, por menos de um semestre 353 ⇒ Física=Péssimo 273	0.09	0.77	1.02				
	Sim, por menos de um semestre 353 ⇒ História=Péssimo 217	0.07	0.61	1.11				
	Sim, por um semestre 563 ⇒ História=Péssimo 327	0.11	0.58	1.05				
	Sim, por um semestre 563 ⇒ Língua Estrangeira=Péssimo 287	0.09	0.51	1.11				
2003/2	Sim, por um semestre 587 ⇒ Matemática=Péssimo 388	0.19	0.66	1.01	Não 220 ⇒ Biologia=Péssimo 137	0.07	0.62	1.41
	Sim, por um semestre 587 ⇒ Física=Péssimo 359	0.17	0.61	1.01	Não 220 ⇒ Matemática=Péssimo 155	0.07	0.7	1.08
	Sim, por menos de um semestre 537 ⇒ Língua Estrangeira=Péssimo 378	0.18	0.7	1.01	Não 220 ⇒ Química=Péssimo 112	0.05	0.51	1.39
	Sim, por um semestre 587 ⇒ Língua Estrangeira=Péssimo 412	0.2	0.7	1.01	Não 220 ⇒ Geografia=Péssimo 133	0.06	0.6	1.31
					Não 220 ⇒ Física=Péssimo 163	0.08	0.74	1.23
					Não 220 ⇒ História=Péssimo 117	0.06	0.53	1.12
					Não 220 ⇒ Língua Estrangeira=Péssimo 176	0.08	0.8	1.15
2004/1	Sim, por menos de um semestre 257 ⇒ Biologia=Péssimo 129	0.04	0.5	1.17	Não 456 ⇒ Biologia=Péssimo 240	0.08	0.53	1.23
	Sim, por menos de um semestre 257 ⇒ Matemática=Péssimo 210	0.07	0.82	1.07	Não 456 ⇒ Matemática=Péssimo 357	0.11	0.78	1.02
	Sim, por menos de um semestre 257 ⇒ Física=Péssimo 232	0.07	0.9	1.06	Não 456 ⇒ Física=Péssimo 394	0.12	0.86	1.02
	Sim, por um semestre 422 ⇒ Física=Péssimo 366	0.12	0.87	1.02	Não 456 ⇒ Geografia=Péssimo 290	0.09	0.64	1.02
	Sim, por menos de um semestre 257 ⇒ Geografia=Péssimo 178	0.06	0.69	1.11	Não 456 ⇒ História=Péssimo 239	0.08	0.52	1.01
	Sim, por um semestre 422 ⇒ Geografia=Péssimo 274	0.09	0.65	1.04	Não 456 ⇒ Língua Estrangeira=Péssimo 298	0.09	0.65	1.07
	Sim, por um semestre 422 ⇒ História=Péssimo 239	0.08	0.57	1.09				
	Sim, por menos de um semestre 257 ⇒ História=Péssimo 145	0.05	0.56	1.09				
	Sim, por menos de um semestre 257 ⇒ Língua Estrangeira=Péssimo 179	0.06	0.7	1.14				
	Sim, por um semestre 422 ⇒ Língua Estrangeira=Péssimo 274	0.09	0.65	1.06				

Continua na página seguinte...

**Tabela 20:** Regras de associação envolvendo os atributos frequência pré-vestibular e notas (cont)

Ano/Sem	Regras em que há alguma frequência em pré-vestibular				Regras em que não há frequência em pré-vestibular			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2004/2	Sim, por um semestre 269 ⇒ Matemática=Péssimo 206	0.11	0.77	1.14	Não 101 ⇒ Matemática=Péssimo 70	0.04	0.69	1.03
	Sim, por mais de um ano 189 ⇒ Matemática=Péssimo 130	0.07	0.69	1.03	Não 101 ⇒ Química=Péssimo 74	0.04	0.73	1.18
	Sim, por menos de um semestre 231 ⇒ Matemática=Péssimo 158	0.09	0.68	1.02	Não 101 ⇒ Física=Péssimo 59	0.03	0.58	1.07
	Sim, por um semestre 269 ⇒ Química=Péssimo 184	0.1	0.68	1.1				
	Sim, por menos de um semestre 231 ⇒ Química=Péssimo 147	0.08	0.64	1.02				
	Sim, por mais de um ano 189 ⇒ Física=Péssimo 115	0.06	0.61	1.12				
	Sim, por um ano 133 ⇒ Física=Péssimo 77	0.04	0.58	1.06				
	Sim, por menos de um semestre 231 ⇒ Física=Péssimo 133	0.07	0.58	1.06				
	Sim, por um ano 133 ⇒ História=Regular 76	0.04	0.57	1.09				
2005/1	Sim, por um semestre 787 ⇒ Matemática=Péssimo 573	0.2	0.73	1.06	Não 748 ⇒ Matemática=Péssimo 548	0.19	0.73	1.07
	Sim, por um semestre 787 ⇒ Química=Péssimo 448	0.15	0.57	1.11	Não 748 ⇒ Química=Péssimo 454	0.16	0.61	1.19
	Sim, por um semestre 787 ⇒ Física=Péssimo 477	0.16	0.61	1.06	Não 748 ⇒ Física=Péssimo 455	0.16	0.61	1.07
	Sim, por um semestre 787 ⇒ História=Péssimo 401	0.14	0.51	1.12	Não 748 ⇒ História=Péssimo 374	0.13	0.5	1.1
	Sim, por mais de um ano 438 ⇒ História=Regular 218	0.07	0.5	1.09				
	Sim, por um ano 720 ⇒ História=Regular 358	0.12	0.5	1.08				
2005/2	Biologia=Regular 371 ⇒ Sim, por um semestre 195	0.08	0.53	1.04	Não 265 ⇒ Biologia=Péssimo 143	0.04	0.54	1.43
	Matemática=Regular 340 ⇒ Sim, por um semestre 188	0.08	0.55	1.09	Não 265 ⇒ Matemática=Péssimo 204	0.09	0.77	1.12
	Sim, por um semestre 1093 ⇒ Química=Péssimo 577	0.24	0.53	1.02	Não 265 ⇒ Química=Péssimo 173	0.07	0.65	1.26
	Química=Regular 498 ⇒ Sim, por um semestre 255	0.11	0.51	1.01	Não 265 ⇒ Física=Péssimo 246	0.1	0.93	1.04
	Sim, por um semestre 1093 ⇒ Física=Péssimo 977	0.41	0.89	1.01	Não 265 ⇒ Geografia=Péssimo 148	0.06	0.56	1.17
	Língua Estrangeira=Ótimo 86 ⇒ Sim, por um semestre 49	0.02	0.57	1.13	Não 265 ⇒ História=Péssimo 138	0.06	0.52	1.35
	Português e Literatura=Bom ⇒ 284 Sim, por um semestre 149	0.06	0.52	1.04				
2007/1	Sim, por um semestre 949 ⇒ Biologia=Péssimo 475	0.14	0.5	1.08	Não 1301 ⇒ Biologia=Péssimo 736	0.22	0.57	1.22
	Sim, por um semestre 949 ⇒ Matemática=Péssimo 574	0.17	0.6	1.03	Não 1301 ⇒ Matemática=Péssimo 792	0.24	0.61	1.04
	Sim, por um semestre 949 ⇒ Química=Péssimo 667	0.2	0.7	1.08	Não 1301 ⇒ Química=Péssimo 937	0.28	0.72	1.11
	Sim, por um semestre 949 ⇒ Física=Péssimo 739	0.22	0.78	1.04	Não 1301 ⇒ Física=Péssimo 1022	0.31	0.78	1.05

Continua na página seguinte...

**Tabela 20:** Regras de associação envolvendo os atributos frequência pré-vestibular e notas (cont)

Ano/Sem	Regras em que há alguma frequência em pré-vestibular				Regras em que não há frequência em pré-vestibular			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
<b>2007/1</b> (cont)	Sim, por mais de um ano 339 ⇒ História=Regular 176	0.05	0.52	1.12				
	Sim, por um ano 702 ⇒ História=Regular 354	0.11	0.5	1.09				
	Sim, por mais de um ano 339 ⇒ Português e Literatura=Regular 183	0.06	0.54	1.03				
	Sim, por um ano 702 ⇒ Português e Literatura=Regular 372	0.11	0.53	1.01				
<b>2007/2</b>	Sim, por um semestre 1224 ⇒ Matemática=Péssimo 675	0.25	0.55	1.01	Não 760 ⇒ Biologia=Péssimo 437	0.16	0.57	1.37
	Sim, por um semestre 1224 ⇒ Física=Péssimo 869	0.32	0.71	1.01	Não 760 ⇒ Matemática=Péssimo 442	0.16	0.58	1.06
	Sim, por mais de um ano 325 ⇒ História=Regular 166	0.06	0.51	1.18	Não 760 ⇒ Química=Péssimo 558	0.21	0.73	1.23
					Não 760 ⇒ Física=Péssimo 586	0.22	0.77	1.09
					Não 760 ⇒ Geografia=Péssimo 456	0.17	0.6	1.14
					Não 760 ⇒ Língua Estrangeira=Péssimo 498	0.18	0.66	1.16
<b>2008/1</b>	Sim, por um semestre 938 ⇒ Biologia=Péssimo 746	0.22	0.8	1.04	Não 1605 ⇒ Biologia=Péssimo 1306	0.38	0.81	1.07
	Sim, por um semestre 938 ⇒ Química=Péssimo 591	0.17	0.63	1.02	Não 1605 ⇒ Matemática=Péssimo 1293	0.37	0.81	1.05
	Sim, por um semestre 938 ⇒ Física=Péssimo 831	0.24	0.89	1.01	Não 1605 ⇒ Química=Péssimo 1137	0.33	0.71	1.15
					Não 1605 ⇒ Física=Péssimo 1455	0.42	0.91	1.03
					Geografia=Péssimo 1510 ⇒ Não 756	0.22	0.5	1.08
					História=Péssimo 1148 ⇒ Não 599	0.17	0.52	1.13
					Não 1605 ⇒ Língua Estrangeira=Péssimo 907	0.26	0.57	1.17
					Não 1605 ⇒ Português e Literatura=Péssimo 857	0.25	0.53	1.17
<b>2008/2</b>	Matemática=Bom 142 ⇒ Sim, por um semestre 80	0.04	0.56	1.35	Não 628 ⇒ Biologia=Péssimo 510	0.25	0.81	1.11
	Sim, por um ano 295 ⇒ História=Regular 157	0.08	0.53	1.12	Não 628 ⇒ Matemática=Péssimo 515	0.25	0.82	1.12
	Sim, por um ano 295 ⇒ Português e Literatura=Regular 147	0.07	0.5	1.17	Não 628 ⇒ Química=Péssimo 540	0.27	0.86	1.1
					Não 628 ⇒ Física=Péssimo 521	0.26	0.83	1.09
					Não 628 ⇒ Geografia=Péssimo 376	0.19	0.6	1.33
					Não 628 ⇒ Língua Estrangeira=Péssimo 428	0.21	0.68	1.17
					Não 628 ⇒ Português e Literatura=Péssimo 358	0.18	0.57	1.36

Continua na página seguinte...

**Tabela 20:** Regras de associação envolvendo os atributos frequência pré-vestibular e notas (cont)

Ano/Sem	Regras em que há alguma frequência em pré-vestibular				Regras em que não há frequência em pré-vestibular			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2009/1	Sim, por um semestre 947 ⇒ Biologia=Péssimo 677	0.19	0.71	1.01	Não 1855 ⇒ Biologia=Péssimo 1438	0.41	0.78	1.1
	Sim, por um semestre 947 ⇒ Matemática=Péssimo 691	0.2	0.73	1.02	Não 1855 ⇒ Matemática=Péssimo 1380	0.39	0.74	1.04
	Sim, por um ano 511 ⇒ Português e Literatura=Regular 270	0.08	0.53	1.11	Não 1855 ⇒ Química=Péssimo 1343	0.38	0.72	1.12
	Sim, por mais de um ano 191 ⇒ Português e Literatura=Regular 97	0.03	0.51	1.07	Não 1855 ⇒ Física=Péssimo 1663	0.47	0.9	1.04
					Geografia=Péssimo 987 ⇒ Não 608	0.17	0.62	1.16
					Não 1855 ⇒ História=Péssimo 1006	0.29	0.54	1.1
					Não 1855 ⇒ Língua Estrangeira=Péssimo 1162	0.33	0.63	1.08
					Português e Literatura=Péssimo 1487 ⇒ Não 888	0.25	0.6	1.13
2009/2	Biologia=Muito bom 246 ⇒ Sim, por um semestre 122	0.05	0.5	1.07	Não 773 ⇒ Biologia=Péssimo 455	0.2	0.59	1.13
	Matemática=Muito bom 111 ⇒ Sim, por um semestre 61	0.03	0.55	1.19	Não 773 ⇒ Matemática=Péssimo 541	0.24	0.7	1.05
	Matemática=Regular 640 ⇒ Sim, por um semestre 317	0.14	0.5	1.07	Não 773 ⇒ Química=Péssimo 427	0.19	0.55	1.18
	Química=Muito bom 388 ⇒ Sim, por um semestre 201	0.09	0.52	1.12	Não 773 ⇒ Física=Péssimo 526	0.23	0.68	1.12
	Física=Regular 760 ⇒ Sim, por um semestre 394	0.17	0.52	1.12	Não 773 ⇒ Geografia=Péssimo 407	0.18	0.53	1.14
	Língua Estrangeira=Muito bom 616 ⇒ Sim, por um semestre 306	0.14	0.5	1.08	Não 773 ⇒ História=Péssimo 457	0.2	0.59	1.11
	Português e Literatura=Muito bom 57 ⇒ Sim, por um semestre 32	0.01	0.56	1.22				
	Português e Literatura=Bom 380 ⇒ Sim, por um semestre 198	0.09	0.52	1.13				

**Tabela 21:** Regras de Associação envolvendo os atributos tipo da escola e notas

Ano/Sem	Escola Municipal (EM)				Escola Estadual (EE)				Escola Particular (EP)			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2003/1	EM 76 ⇒ Mat <sup>6</sup> =Péssimo 41	0.01	0.54	1.09	EE 1393 ⇒ Bio <sup>7</sup> =Péssimo 823	0.27	0.59	1.09	Bio=Muito bom 272 ⇒ EP 141	0.05	0.52	1.22
	EM 76 ⇒ Fís <sup>8</sup> =Péssimo 59	0.02	0.78	1.02	EE 1393 ⇒ Mat=Péssimo 769	0.25	0.55	1.12	Bio=Ótimo 164 ⇒ EP 83	0.03	0.51	1.19
					Qui <sup>9</sup> =Péssimo 560 ⇒ EE 338	0.11	0.6	1.32	Mat=Ótimo 148 ⇒ EP 78	0.03	0.53	1.24
					Qui=Regular 1253 ⇒ EE 627	0.21	0.5	1.09	Qui=Ótimo 245 ⇒ EP 135	0.04	0.55	1.3
					EE 1393 ⇒ Fís=Péssimo 1130	0.37	0.81	1.07	Qui=Muito bom 982 ⇒ EP 520	0.17	0.53	1.25
					Geo <sup>10</sup> =Péssimo 1271 ⇒ EE 666	0.22	0.52	1.14	Fís=Muito bom 85 ⇒ EP 52	0.02	0.61	1.44
					EE 1393 ⇒ His <sup>11</sup> Péssimo 848	0.28	0.61	1.1	Fís=Bom 176 ⇒ EP 106	0.03	0.6	1.42
					EE 1393 ⇒ Lín Est <sup>12</sup> =Péssimo 759	0.25	0.54	1.19	Geo=Ótimo 153 ⇒ EP 82	0.03	0.54	1.26
					Por e Lit <sup>13</sup> =Péssimo 378 ⇒ EE 220	0.07	0.58	1.27	His=Ótimo 156 ⇒ EP 82	0.03	0.53	1.24
					Por e Lit=Regular 1065 ⇒ EE 528	0.17	0.5	1.08	His=Muito bom 262 ⇒ EP 135	0.04	0.52	1.21
									Lín Est=Muito bom 526 ⇒ EP 301	0.1	0.57	1.35
									Lín Est=Ótimo 190 ⇒ EP 97	0.03	0.51	1.2
									Por e Lit=Ótimo 239 ⇒ EP 125	0.04	0.52	1.23
								Por e Lit=Muito bom 751 ⇒ EP 384	0.13	0.51	1.2	
2003/2	EM 57 ⇒ Mat=Péssimo 42	0.02	0.74	1.13	EE 994 ⇒ Bio=Péssimo 502	0.24	0.51	1.14	Bio=Ótimo 55 ⇒ EP 39	0.02	0.71	1.66
					EE 994 ⇒ Mat=Péssimo 718	0.34	0.72	1.1	Bio=Muito bom 394 ⇒ EP 202	0.1	0.51	1.2
					Qui=Péssimo 765 ⇒ EE 415	0.2	0.54	1.14	Mat=Muito bom 86 ⇒ EP 50	0.02	0.58	1.36
					EE 994 ⇒ Fís=Péssimo 642	0.31	0.65	1.07	Mat=Regular 634 ⇒ EP 323	0.15	0.51	1.19
					EE 994 ⇒ Geo=Péssimo 540	0.26	0.54	1.18	Qui=Muito bom 370 ⇒ EP 205	0.1	0.55	1.3

Continua na página seguinte...

<sup>6</sup>Matemática

<sup>7</sup>Biologia

<sup>8</sup>Física

<sup>9</sup>Química

<sup>10</sup>Geografia

<sup>11</sup>História

<sup>12</sup>Língua Estrangeira

<sup>13</sup>Português e Literatura

**Tabela 21:** Regras de Associação envolvendo os atributos tipo da escola e notas (cont)

Ano/Sem	Escola Municipal (EM)				Escola Estadual (EE)				Escola Particular (EP)			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2003/2 (cont)					EE 994 ⇒ His=Péssimo 511	0.25	0.51	1.08	Fís=Muito bom 177 ⇒ EP 106	0.05	0.6	1.4
					EE 994 ⇒ Lín Est=Péssimo 775	0.37	0.78	1.12	Fís=Bom 230 ⇒ EP 120	0.06	0.52	1.22
					Por e Lit=Péssimo 615 ⇒ EE 336	0.16	0.55	1.15	Geo=Muito bom 289 ⇒ EP 175	0.08	0.61	1.42
									Lín Est=Ótimo 80 ⇒ EP 62	0.03	0.78	1.81
									Lín Est=Muito bom 164 ⇒ EP 117	0.06	0.71	1.67
									Por e Lit=Muito bom 314 ⇒ EP 166	0.08	0.53	1.24
2004/1	EM 51 ⇒ Mat=Péssimo 43	0.01	0.84	1.1	EE 1030 ⇒ Mat=Péssimo 840	0.26	0.82	1.07				
	EM 51 ⇒ Lín Est=Péssimo 33	0.01	0.65	1.06	EE 1030 ⇒ Fís=Péssimo 923	0.29	0.9	1.05				
					EE 1030 ⇒ His=Péssimo 569	0.18	0.55	1.06				
					EE 1030 ⇒ Lín Est=Péssimo 713	0.22	0.69	1.13				
2004/2					EE 546 ⇒ Mat=Péssimo 423	0.23	0.77	1.16				
					EE 546 ⇒ Quí=Péssimo 359	0.2	0.66	1.05				
					EE 546 ⇒ Fís=Péssimo 348	0.19	0.64	1.17				
					EE 546 ⇒ Geo=Péssimo 278	0.15	0.51	1.16				
2005/1	EM 56 ⇒ Mat=Péssimo 43	0.01	0.77	1.12	Bio=Péssimo 920 ⇒ EE 510	0.17	0.55	1.16	Bio=Muito bom 417 ⇒ EP 234	0.08	0.56	1.2
	EM 56 ⇒ Fís=Péssimo 43	0.01	0.77	1.35	EE 1290 ⇒ Mat=Péssimo 1003	0.34	0.78	1.14	Bio=Ótimo 266 ⇒ EP 149	0.05	0.56	1.19
	EM 56 ⇒ His=Péssimo 34	0.01	0.61	1.34	EE 1290 ⇒ Quí=Péssimo 738	0.25	0.57	1.12	Mat=Bom 107 ⇒ EP 73	0.02	0.68	1.46
					EE 1290 ⇒ Fís=Péssimo 822	0.28	0.64	1.12	Mat=Muito Bom 76 ⇒ EP 49	0.02	0.64	1.38
					Geo=Péssimo 867 ⇒ EE 492	0.17	0.57	1.19	Mat=Regular 658 ⇒ EP 382	0.13	0.58	1.24
					His=Péssimo 1224 ⇒ EE 645	0.22	0.53	1.1	Quí=Muito bom 99 ⇒ EP 63	0.02	0.64	1.36
					Lín Est=Péssimo 759 ⇒ EE 432	0.15	0.57	1.19	Quí=Bom 137 ⇒ EP 80	0.03	0.58	1.25
					Lín Est=Regular 875 ⇒ EE 476	0.16	0.54	1.14	Quí=Regular 1067 ⇒ EP 533	0.18	0.5	1.07
					Por e Lit=Péssimo 923 ⇒ EE 501	0.17	0.54	1.14	Fís=Muito bom 71 ⇒ EP 51	0.02	0.72	1.53
									Fís=Bom 93 ⇒ EP 59	0.02	0.63	1.35
									Fís=Regular 986 ⇒ EP 526	0.18	0.53	1.14
								Geo=Muito bom 256 ⇒ EP 151	0.05	0.59	1.26	

Continua na página seguinte...

**Tabela 21:** Regras de Associação envolvendo os atributos tipo da escola e notas (cont)

Ano/Sem	Escola Municipal (EM)				Escola Estadual (EE)				Escola Particular (EP)			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2005/1 (cont)									His=Muito bom 93 ⇒ EP 57	0.02	0.61	1.31
									His=Regular 1238 ⇒ EP 622	0.21	0.5	1.07
									Lín Est=Ótimo 280 ⇒ EP 208	0.07	0.74	1.58
									Lín Est=Muito bom 517 ⇒ EP 288	0.1	0.56	1.19
									Por e Lit=Muito bom 140 ⇒ EP 86	0.03	0.61	1.31
									Por e Lit=Bom 505 ⇒ EP 289	0.1	0.57	1.22
2005/2	EM 46 ⇒ Mat=Péssimo 35	0.01	0.76	1.11	Bio=Péssimo 814 ⇒ EE 457	0.19	0.56	1.23	Bio=Ótimo 107 ⇒ EP 72	0.03	0.67	1.4
	EM 46 ⇒ Quí=Péssimo 29	0.01	0.63	1.22	EE 990 ⇒ Mat=Péssimo 761	0.32	0.77	1.12	Bio=Muito bom 480 ⇒ EP 288	0.12	0.6	1.25
	EM 46 ⇒ Fís=Péssimo 43	0.02	0.93	1.05	EE 990 ⇒ Quí=Péssimo 597	0.25	0.6	1.17	Mat=Muito bom 136 ⇒ EP 101	0.04	0.74	1.55
					EE 990 ⇒ Fís=Péssimo 898	0.38	0.91	1.02	Mat=Bom 194 ⇒ EP 125	0.05	0.64	1.34
					EE 990 ⇒ Geo=Péssimo 540	0.23	0.53	1.15	Mat=Regular 340 ⇒ EP 177	0.07	0.52	1.09
					His=Péssimo 834 ⇒ EE 449	0.19	0.54	1.18	Quí=Muito bom 213 ⇒ EP 140	0.06	0.66	1.37
					Lín Est=Péssimo 762 ⇒ EE 416	0.17	0.55	1.19	Quí=Bom 325 ⇒ EP 189	0.08	0.58	1.21
					Por e Lit=Péssimo 577 ⇒ EE 328	0.14	0.57	1.24	Quí=Regular 498 ⇒ EP 257	0.11	0.52	1.08
									Fís=Bom 60 ⇒ EP 38	0.02	0.63	1.32
									Fís=Regular 161 ⇒ EP 88	0.04	0.55	1.14
									Geo=Muito bom 265 ⇒ EP 162	0.07	0.61	1.28
									His=Muito bom 373 ⇒ EP 225	0.09	0.6	1.26
									His=Regular 926 ⇒ EP 469	0.2	0.51	1.06
								Lín Est=Ótimo 86 ⇒ EP 68	0.03	0.79	1.65	
								Lín Est=Muito bom 447 ⇒ EP 269	0.11	0.6	1.26	
								Por e Lit=Ótimo 60 ⇒ EP 42	0.02	0.7	1.46	
								Por e Lit=Muito bom 324 ⇒ EP 207	0.09	0.64	1.33	
2007/1	EM 74 ⇒ Bio=Péssimo 38	0.01	0.51	1.11	EE 1730 ⇒ Bio=Péssimo 957	0.29	0.55	1.19	Bio=Ótimo 160 ⇒ EP 100	0.03	0.63	1.49
	EM 74 ⇒ Quí=Péssimo 50	0.02	0.68	1.04	EE 1730 ⇒ Mat=Péssimo 1120	0.34	0.65	1.1	Bio=Muito bom 664 ⇒ EP 378	0.11	0.57	1.36
	EM 74 ⇒ Fís=Péssimo 59	0.02	0.8	1.06	EE 1730 ⇒ Quí=Péssimo 1246	0.38	0.72	1.11	Mat=Muito bom 242 ⇒ EP 145	0.04	0.6	1.43

Continua na página seguinte...

**Tabela 21:** Regras de Associação envolvendo os atributos tipo da escola e notas (cont)

Ano/Sem	Escola Municipal (EM)				Escola Estadual (EE)				Escola Particular (EP)			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2007/1 (cont)	EM 74 ⇒ His=Péssimo 38	0.01	0.51	1.22	Geo=Péssimo 1114 ⇒ EE 688	0.21	0.62	1.18	Mat=Bom 422 ⇒ EP 219	0.07	0.52	1.24
					EE 1730 ⇒ Fís=Péssimo 1374	0.42	0.79	1.06	Quí=Muito bom 243 ⇒ EP 143	0.04	0.59	1.41
					Geo=Regular 1315 ⇒ EE 709	0.21	0.54	1.03	Quí=Bom 347 ⇒ EP 182	0.06	0.52	1.25
					His=Péssimo 1387 ⇒ EE 795	0.24	0.57	1.09	Fís=Muito bom 130 ⇒ EP 82	0.02	0.63	1.51
					Lín Est=Péssimo 1275 ⇒ EE 747	0.23	0.59	1.12	Fís=Bom 230 ⇒ EP 116	0.04	0.5	1.21
					Lín=Regular 1258 ⇒ EE 684	0.21	0.54	1.04	Geo=Muito bom 789 ⇒ EP 448	0.14	0.57	1.36
					Por e Lit=Péssimo 970 ⇒ EE 554	0.17	0.57	1.09	Lín=Ótimo 98 ⇒ EP 69	0.02	0.7	1.68
					Por e Lit=Regular 1732 ⇒ EE 919	0.28	0.53	1.01	Lín=Muito bom 670 ⇒ EP 360	0.11	0.54	1.28
								Por e Lit=Muito bom 222 ⇒ EP 119	0.04	0.54	1.28	
2007/2	EM 62 ⇒ Bio=Péssimo 35	0.01	0.56	1.34	Bio=Péssimo 1145 ⇒ EE 787	0.29	0.69	1.17	Bio=Muito bom 453 ⇒ EP 253	0.09	0.56	1.56
	EM 62 ⇒ Mat=Péssimo 41	0.02	0.66	1.21	EE 1597 ⇒ Mat=Péssimo 926	0.34	0.58	1.06	Mat=Muito bom 147 ⇒ EP 77	0.03	0.52	1.46
	EM 62 ⇒ Quí=Péssimo 45	0.02	0.73	1.22	EE 1597 ⇒ Quí=Péssimo 1078	0.4	0.68	1.13	Quí=Muito bom 256 ⇒ EP 146	0.05	0.57	1.59
	EM 62 ⇒ Fís=Péssimo 54	0.02	0.87	1.24	EE 1597 ⇒ Fís=Péssimo 1209	0.44	0.76	1.07	Quí=Bom 351 ⇒ EP 180	0.07	0.51	1.43
	EM 62 ⇒ Geo=Péssimo 41	0.02	0.66	1.26	EE 1597 ⇒ Geo=Péssimo 934	0.34	0.58	1.11	Fís=Muito bom 90 ⇒ EP 56	0.02	0.62	1.74
	EM 62 ⇒ His=Péssimo 31	0.01	0.5	1.17	His=Péssimo 1159 ⇒ EE 751	0.28	0.65	1.1	Fís=Bom 223 ⇒ EP 111	0.04	0.5	1.39
	EM 62 ⇒ Lín Est=Péssimo 36	0.01	0.58	1.03	EE 1597 ⇒ Lín Est=Péssimo 984	0.36	0.62	1.09	Geo=Muito bom 308 ⇒ EP 162	0.06	0.53	1.47
	EM 62 ⇒ Por e Lit=Péssimo 39	0.01	0.63	1.26	EE 1597 ⇒ Por e Lit=Péssimo 869	0.32	0.54	1.09	Lín Est=Muito bom 254 ⇒ EP 147	0.05	0.58	1.61
								Por e Lit=Bom 134 ⇒ EP 72	0.03	0.54	1.5	
								Por e Lit=Muito bom 64 ⇒ EP 34	0.01	0.53	1.48	
2008/1	EM 73 ⇒ Bio=Péssimo 60	0.02	0.82	1.08	EE 2128 ⇒ Bio=Péssimo 1711	0.49	0.8	1.05	Bio=Muito bom 126 ⇒ EP 89	0.03	0.71	2.09
	EM 73 ⇒ Quí=Péssimo 53	0.02	0.73	1.18	EE 2128 ⇒ Mat=Péssimo 1752	0.51	0.82	1.08	Mat=Muito bom 131 ⇒ EP 87	0.03	0.66	1.96
	EM 73 ⇒ Fís=Péssimo 67	0.02	0.92	1.04	EE 2128 ⇒ Quí=Péssimo 1483	0.43	0.7	1.13	Mat=Bom 225 ⇒ EP 113	0.03	0.5	1.49
					EE 2128 ⇒ Fís=Péssimo 1953	0.56	0.92	1.04	Quí=Ótimo 58 ⇒ EP 43	0.01	0.74	2.19
					Geo=Péssimo 1510 ⇒ EE 992	0.29	0.66	1.07	Quí=Muito bom 423 ⇒ EP 241	0.07	0.57	1.68
					Geo=Regular 1418 ⇒ EE 881	0.25	0.62	1.01	Fís=Bom 98 ⇒ EP 62	0.02	0.63	1.87
					His=Péssimo 1148 ⇒ EE 770	0.22	0.67	1.09	Por e Lit=Muito bom 275 ⇒ EP 158	0.05	0.57	1.7

Continua na página seguinte...



**Tabela 21:** Regras de Associação envolvendo os atributos tipo da escola e notas (cont)

Ano/Sem	Escola Municipal (EM)				Escola Estadual (EE)				Escola Particular (EP)			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
<b>2008/1</b> (cont)					EE 2128 ⇒ Lín Est=Péssimo 1126	0.32	0.53	1.09				
					EE 2128 ⇒ Por e Lit=Péssimo 1117	0.32	0.52	1.15				
<b>2008/2</b>	EM 39 ⇒ Mat=Péssimo 32	0.02	0.82	1.12	EE 1242 ⇒ Bio=Péssimo 962	0.48	0.77	1.06	Bio=Muito bom 49 ⇒ EP 28	0.01	0.57	1.7
	EM 39 ⇒ Quí=Péssimo 35	0.02	0.9	1.15	EE 1242 ⇒ Mat=Péssimo 1002	0.5	0.81	1.11	Mat=Bom 142 ⇒ EP 81	0.04	0.57	1.7
	EM 39 ⇒ Fís=Péssimo 31	0.02	0.79	1.04	EE 1242 ⇒ Quí=Péssimo 1041	0.51	0.84	1.07	Mat=Muito bom 69 ⇒ EP 39	0.02	0.57	1.68
	EM 39 ⇒ Geo=Péssimo 22	0.01	0.56	1.26	EE 1242 ⇒ Fís=Péssimo 1038	0.51	0.84	1.1	Quí=Muito bom 47 ⇒ EP 32	0.02	0.68	2.03
	EM 39 ⇒ Por e Lit=Péssimo 21	0.01	0.54	1.29	EE 1242 ⇒ Geo=Péssimo 666	0.33	0.54	1.19	Quí=Bom 115 ⇒ EP 62	0.03	0.54	1.61
					His=Péssimo 634 ⇒ EE 470	0.23	0.74	1.21	Fís=Muito bom 67 ⇒ EP 40	0.02	0.6	1.78
					EE 1242 ⇒ Lín Est=Péssimo 821	0.41	0.66	1.13	Fís=Regular 280 ⇒ EP 142	0.07	0.51	1.51
					Por e Lit=Péssimo 845 ⇒ EE 612	0.3	0.72	1.18	Fís=Bom 136 ⇒ EP 68	0.03	0.5	1.49
								Geo=Muito bom 324 ⇒ EP 170	0.08	0.52	1.56	
								Lín Est=Muito bom 172 ⇒ EP 100	0.05	0.58	1.73	
								Por e Lit=Muito bom 125 ⇒ EP 69	0.03	0.55	1.64	
<b>2009/1</b>	EM 86 ⇒ Bio=Péssimo 67	0.02	0.78	1.1	EE 2370 ⇒ Bio=Péssimo 1814	0.52	0.77	1.08	Bio=Muito bom 225 ⇒ EP 130	0.04	0.58	2.11
	EM 86 ⇒ Quí=Péssimo 63	0.02	0.73	1.13	EE 2370 ⇒ Mat=Péssimo 1853	0.53	0.78	1.1	Mat=Muito bom 192 ⇒ EP 106	0.03	0.55	2.02
	EM 86 ⇒ Fís=Péssimo 77	0.02	0.9	1.04	EE 2370 ⇒ Quí=Péssimo 1710	0.49	0.72	1.11	Quí=Muito bom 361 ⇒ EP 187	0.05	0.52	1.89
	EM 86 ⇒ His=Péssimo 55	0.02	0.64	1.3	EE 2370 ⇒ Fís=Péssimo 2153	0.61	0.91	1.05	Fís=Bom 113 ⇒ EP 69	0.02	0.61	2.23
	EM 86 ⇒ Lín Est=Péssimo 59	0.02	0.69	1.19	Geo=Regular 1522 ⇒ EE 1122	0.32	0.74	1.09	Lín Est=Ótimo 50 ⇒ EP 36	0.01	0.72	2.63
	EM 86 ⇒ Por e Lit=Péssimo 43	0.01	0.5	1.18	Geo=Péssimo 987 ⇒ EE 718	0.2	0.73	1.08	Lín Est=Muito bom 301 ⇒ EP 161	0.05	0.53	1.96
					EE 2370 ⇒ His=Péssimo 1258	0.36	0.53	1.08				
					EE 2370 ⇒ Lín Est=Péssimo 1499	0.43	0.63	1.09				
				Por e Lit=Péssimo 1487 ⇒ EE 1068	0.3	0.72	1.06					
<b>2009/2</b>	EM 53 ⇒ Mat=Péssimo 39	0.02	0.74	1.1	EE 1419 ⇒ Bio=Péssimo 821	0.36	0.58	1.11	Bio=Muito bom 246 ⇒ EP 130	0.06	0.53	1.67
	EM 53 ⇒ Quí=Péssimo 30	0.01	0.57	1.21	EE 1419 ⇒ Mat=Péssimo 1016	0.45	0.72	1.07	Mat=Muito bom 111 ⇒ EP 72	0.03	0.65	2.04
	EM 53 ⇒ Fís=Péssimo 34	0.02	0.64	1.06	Quí=Regular 811 ⇒ EE 520	0.23	0.64	1.02	Fís=Muito bom 124 ⇒ EP 68	0.03	0.55	1.73
	EM 53 ⇒ His=Péssimo 30	0.01	0.57	1.06	EE 1419 ⇒ Quí=Péssimo 719	0.32	0.51	1.08	Geo=Muito bom 313 ⇒ EP 159	0.07	0.51	1.6

Continua na página seguinte...

**Tabela 21:** Regras de Associação envolvendo os atributos tipo da escola e notas (cont)

Ano/Sem	Escola Municipal (EM)				Escola Estadual (EE)				Escola Particular (EP)			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2009/2 (cont)					EE 1419 ⇒ Fis=Péssimo 927	0.41	0.65	1.08	Por e Lit=Muito bom 57 ⇒ EP 31	0.01	0.54	1.71
					EE 1419 ⇒ Geo=Péssimo 741	0.33	0.52	1.13				
					EE 1419 ⇒ His=Péssimo 783	0.35	0.55	1.04				
					Lín Est=Péssimo 852 ⇒ EE 594	0.26	0.7	1.11				
					Lín Est=Regular 785 ⇒ EE 504	0.22	0.64	1.02				
					Por e Lit=Regular 1041 ⇒ EE 709	0.31	0.68	1.08				
					Por e Lit=Péssimo 771 ⇒ EE 504	0.22	0.65	1.04				

**Tabela 22:** Regras de associação envolvendo os atributos local de residência e notas

Ano/Sem	Local de residência	Descrição da Regra	Sup	Conf	Lift
2005/1	Diamantina	Diamantina 392 ⇒ Matemática=Péssimo 299	0.1	0.76	1.12
		Diamantina 392 ⇒ Física=Péssimo 263	0.09	0.67	1.18
		Diamantina 392 ⇒ Química=Péssimo 209	0.07	0.53	1.04
		Diamantina 392 ⇒ História=Péssimo 214	0.07	0.55	1.2
	Vale do Jequitinhonha	Vale do Jequitinhonha 201 ⇒ Matemática=Péssimo 158	0.05	0.79	1.15
		Vale do Jequitinhonha 201 ⇒ Física=Péssimo 143	0.05	0.71	1.25
		Vale do Jequitinhonha 201 ⇒ Química=Péssimo 128	0.04	0.64	1.25
		Vale do Jequitinhonha 201 ⇒ História=Péssimo 111	0.04	0.55	1.22
		Vale do Jequitinhonha 201 ⇒ Biologia=Péssimo 105	0.04	0.52	1.53
	Outra Região de MG	Biologia=Bom 266 ⇒ Outra Região de Minas Gerais 154	0.05	0.58	1.04
		Biologia=Regular 831 ⇒ Outra Região de Minas Gerais 481	0.16	0.58	1.04
		Matemática=Muito bom 76 ⇒ Outra Região de Minas Gerais 46	0.02	0.61	1.09
		Química=Bom 137 ⇒ Outra Região de Minas Gerais 78	0.03	0.57	1.03
		Física=Bom 93 ⇒ Outra Região de Minas Gerais 62	0.02	0.67	1.2
		Geografia=Regular 1234 ⇒ Outra Região de Minas Gerais 690	0.24	0.56	1.01
Língua Estrangeira=Regular 875 ⇒ Outra Região de Minas Gerais 500		0.17	0.57	1.03	
Língua Estrangeira=Bom 269 ⇒ Outra Região de Minas Gerais 151		0.05	0.56	1.01	
Português e Literatura=Bom 505 ⇒ Outra Região de Minas Gerais 283		0.1	0.56	1.01	
Português e Literatura=Regular 1131 ⇒ Outra Região de Minas Gerais 630	0.22	0.56	1.01		
2005/2	Diamantina	Diamantina 334 ⇒ Matemática=Péssimo 260	0.11	0.78	1.13
		Diamantina 334 ⇒ Geografia=Péssimo 189	0.08	0.57	1.19
		Diamantina 334 ⇒ Química=Péssimo 184	0.08	0.55	1.06
	Vale do Jequitinhonha	Vale do Jequitinhonha 131 ⇒ Física=Péssimo 119	0.05	0.91	1.02
		Vale do Jequitinhonha 131 ⇒ Matemática=Péssimo 106	0.04	0.81	1.18
		Vale do Jequitinhonha 131 ⇒ Química=Péssimo 90	0.04	0.69	1.33
		Vale do Jequitinhonha 131 ⇒ Biologia=Péssimo 74	0.03	0.56	1.5
		Vale do Jequitinhonha 131 ⇒ Geografia=Péssimo 82	0.03	0.63	1.32
		Vale do Jequitinhonha 131 ⇒ História=Péssimo 67	0.03	0.51	1.33
	Outra Região de MG	Biologia=Muito bom 480 ⇒ Outra Região de Minas Gerais 297	0.12	0.62	1.08
		Matemática=Muito bom 136 ⇒ Outra Região de Minas Gerais 84	0.04	0.62	1.08
		Física=Bom 60 ⇒ Outra Região de Minas Gerais 38	0.02	0.63	1.1
		Geografia=Muito bom 265 ⇒ Outra Região de Minas Gerais 173	0.07	0.65	1.14
		História=Muito bom 373 ⇒ Outra Região de Minas Gerais 236	0.1	0.63	1.1
		Língua Estrangeira=Ótimo 86 ⇒ Outra Região de Minas Gerais 55	0.02	0.64	1.1
Português e Literatura=Bom 284 ⇒ Outra Região de Minas Gerais 177		0.07	0.62	1.09	
2007/1	Diamantina	Diamantina 559 ⇒ Física=Péssimo 429	0.13	0.77	1.02
		Diamantina 559 ⇒ Química=Péssimo 371	0.11	0.66	1.02
		Diamantina 559 ⇒ Matemática=Péssimo 342	0.1	0.61	1.04
	Teófilo Otoni	Teófilo Otoni 345 ⇒ Física=Péssimo 289	0.09	0.84	1.12
		Teófilo Otoni 345 ⇒ Química=Péssimo 271	0.08	0.79	1.21
		Teófilo Otoni 345 ⇒ Matemática=Péssimo 240	0.07	0.7	1.18
		Teófilo Otoni 345 ⇒ Biologia=Péssimo 225	0.07	0.65	1.41

Continua na página seguinte...

**Tabela 22:** Regras de associação envolvendo os atributos local de residência e notas (cont)

Ano/Sem	Local de residência	Descrição da Regra	Sup	Conf	Lift
2007/1 (cont)	Vale do Jequitinhonha	Na região do Vale do Jequitinhonha 364 ⇒ Física=Péssimo 285	0.09	0.78	1.04
		Na região do Vale do Jequitinhonha 364 ⇒ Química=Péssimo 267	0.08	0.73	1.13
		Na região do Vale do Jequitinhonha 364 ⇒ Matemática=Péssimo 230	0.07	0.63	1.08
		Na região do Vale do Jequitinhonha 364 ⇒ Biologia=Péssimo 201	0.06	0.55	1.19
	Vale do Mucuri	Na região do Vale do Mucuri 123 ⇒ Física=Péssimo 101	0.03	0.82	1.09
		Na região do Vale do Mucuri 123 ⇒ Química=Péssimo 102	0.03	0.83	1.28
		Na região do Vale do Mucuri 123 ⇒ Biologia=Péssimo 91	0.03	0.74	1.59
		Na região do Vale do Mucuri 123 ⇒ Matemática=Péssimo 81	0.02	0.66	1.12
		Na região do Vale do Mucuri 123 ⇒ Língua Estrangeira=Péssimo 78	0.02	0.63	1.64
		Na região do Vale do Mucuri 123 ⇒ Geografia=Péssimo 69	0.02	0.56	1.66
		Na região do Vale do Mucuri 123 ⇒ História=Péssimo 68	0.02	0.55	1.32
	Outra Região de MG	Biologia=Bom 441 ⇒ Local de Residência=Outra região de Minas Gerais 220	0.07	0.5	1.17
	2007/2	Diamantina	Diamantina 443 ⇒ Matemática=Péssimo 278	0.1	0.63
Diamantina 443 ⇒ Química=Péssimo 270			0.1	0.61	1.02
Diamantina 443 ⇒ Física=Péssimo 332			0.12	0.75	1.06
Diamantina 443 ⇒ Geografia=Péssimo 257			0.09	0.58	1.1
Diamantina 443 ⇒ Português e Literatura=Péssimo 225			0.08	0.51	1.02
Teófilo Otoni		Teófilo Otoni 379 ⇒ Biologia=Péssimo 243	0.09	0.64	1.52
		Teófilo Otoni 379 ⇒ Matemática=Péssimo 232	0.09	0.61	1.12
		Teófilo Otoni 379 ⇒ Química=Péssimo 300	0.11	0.79	1.33
		Teófilo Otoni 379 ⇒ Física=Péssimo 309	0.11	0.82	1.16
		Teófilo Otoni 379 ⇒ Geografia=Péssimo 246	0.09	0.65	1.24
		Teófilo Otoni 379 ⇒ História=Péssimo 205	0.08	0.54	1.27
		Teófilo Otoni 379 ⇒ Língua Estrangeira=Péssimo 258	0.09	0.68	1.21
		Teófilo Otoni 379 ⇒ Português e Literatura=Péssimo 239	0.09	0.63	1.26
Vale do Jequitinhonha		Vale do Jequitinhonha 279 ⇒ Biologia=Péssimo 145	0.05	0.52	1.23
		Vale do Jequitinhonha 279 ⇒ Matemática=Péssimo 163	0.06	0.58	1.07
		Vale do Jequitinhonha 279 ⇒ Química=Péssimo 199	0.07	0.71	1.19
		Vale do Jequitinhonha 279 ⇒ Física=Péssimo 206	0.08	0.74	1.05
		Vale do Jequitinhonha 279 ⇒ Geografia=Péssimo 167	0.06	0.6	1.14
		Vale do Jequitinhonha 279 ⇒ Língua Estrangeira=Péssimo 182	0.07	0.65	1.16
		Vale do Jequitinhonha 279 ⇒ Português e Literatura=Péssimo 150	0.06	0.54	1.07
Vale do Mucuri		Vale do Mucuri 129 ⇒ Biologia=Péssimo 75	0.03	0.58	1.38
		Vale do Mucuri 129 ⇒ Matemática=Péssimo 80	0.03	0.62	1.13
		Vale do Mucuri 129 ⇒ Química=Péssimo 96	0.04	0.74	1.25
		Vale do Mucuri 129 ⇒ Física=Péssimo 98	0.04	0.76	1.08
		Vale do Mucuri 129 ⇒ Geografia=Péssimo 86	0.03	0.67	1.27
		Vale do Mucuri 129 ⇒ Língua Estrangeira=Péssimo 97	0.04	0.75	1.34
		Vale do Mucuri 129 ⇒ Português e Literatura=Péssimo 81	0.03	0.63	1.26
Outra Região de MG	Biologia=Muito bom 453 ⇒ Outra Região de Minas Gerais 239	0.09	0.53	1.29	
	Matemática=Muito bom 147 ⇒ Outra Região de Minas Gerais 83	0.03	0.56	1.38	
	Química=Muito bom 351 ⇒ Outra Região de Minas Gerais 193	0.07	0.55	1.34	
	Física=Muito bom 90 ⇒ Outra Região de Minas Gerais 48	0.02	0.53	1.3	

**Tabela 23:** Regras de Associação envolvendo os atributos curso e notas

Ano/Sem	Antecedente da Regra (Curso)			
	Descrição da Regra	Sup	Conf	Lift
2004/2	Farmácia-Bioquímica/Análise Clínica 170 ⇒ Geografia=Regular 87	0.05	0.51	1.14
	Engenharia Florestal 172 ⇒ História=Regular 189	0.1	0.54	1.03
	Odontologia 352 ⇒ História=Regular 189	0.1	0.54	1.03
	Nutrição 274 ⇒ História=Regular 145	0.08	0.53	1.01
2005/1	Agronomia 199 ⇒ Geografia=Regular 112	0.04	0.56	1.23
	Agronomia 199 ⇒ História=Regular 100	0.03	0.5	1.09
	Odontologia 418 ⇒ História-Regular 209	0.07	0.5	1.09
2005/2	Zootecnia 156 ⇒ Português e Literatura=Regular 78	0.03	0.5	1.18
2006/1	Farmácia-Bioquímica/Industrial 284 ⇒ Geografia=Regular 147	0.06	0.52	1.21
	Farmácia-Bioquímica/Industrial 284 ⇒ História=Regular 147	0.06	0.52	1.14
2006/2	Fisioterapia 213 ⇒ Geografia=Regular 113	0.04	0.53	1.14
	Farmácia 377 ⇒ Geografia=Regular 196	0.07	0.52	1.12
	Ciências Biológicas 189 ⇒ Geografia=Regular 97	0.03	0.51	1.1
	Engenharia Florestal 166 ⇒ Geografia=Regular 84	0.03	0.51	1.09
	Agronomia 138 ⇒ Geografia=Regular 69	0.02	0.5	1.07
	Zootecnia 103 ⇒ Geografia=Regular 51	0.02	0.5	1.06
	Odontologia 355 ⇒ História=Regular 188	0.06	0.53	1.17
	Engenharia Florestal 166 ⇒ História=Regular 86	0.03	0.52	1.15
	Ciências Biológicas 189 ⇒ História=Regular 95	0.03	0.5	1.11
	Agronomia 138 ⇒ História=Regular 69	0.02	0.5	1.11
	Sistemas de Informação 124 ⇒ Português e Literatura=Regular 68	0.02	0.55	1.21
	Ciências Econômicas 101 ⇒ Português e Literatura=Regular 55	0.02	0.54	1.2
	Agronomia 138 ⇒ Português e Literatura=Regular 75	0.03	0.54	1.2
	Ciências Biológicas 189 ⇒ Português e Literatura=Regular 99	0.03	0.52	1.16
2007/1	Farmácia 349 ⇒ História=Regular 182	0.06	0.52	1.12
	Agronomia 142 ⇒ História=Regular 72	0.02	0.51	1.09
	Odontologia 465 ⇒ História=Regular 233	0.07	0.5	1.08
	Ciências Biológicas 200 ⇒ História=Regular 99	0.03	0.5	1.07
	Ciências Econômicas 73 ⇒ Português e Literatura=Regular 47	0.01	0.64	1.23
	Química 71 ⇒ Português e Literatura=Regular 42	0.01	0.59	1.13
	Ciências Biológicas 200 ⇒ Português e Literatura=Regular 114	0.03	0.57	1.09
	Zootecnia 118 ⇒ Português e Literatura=Regular 66	0.02	0.56	1.07
	Engenharia Florestal 172 ⇒ Português e Literatura=Regular 96	0.03	0.56	1.06
	Sistemas de Informação 178 ⇒ Português e Literatura=Regular 98	0.03	0.55	1.05
	Agronomia 142 ⇒ Português e Literatura=Regular 77	0.02	0.54	1.03
	Turismo 133 ⇒ Português e Literatura=Regular 71	0.02	0.53	1.02
	Farmácia 349 ⇒ Português e Literatura=Regular 186	0.06	0.53	1.02
Fisioterapia 228 ⇒ Português e Literatura=Regular 121	0.04	0.53	1.01	
2007/2	Turismo 89 ⇒ História=Regular 49	0.02	0.55	1.28
	Ciências Econômicas 72 ⇒ História=Regular 36	0.01	0.5	1.16
	Química 56 ⇒ Português e Literatura=Regular 33	0.01	0.59	1.38

Continua na página seguinte...

**Tabela 23:** Regras de Associação envolvendo os atributos curso e notas (cont)

Ano/Sem	Antecedente da Regra (Curso)			
	Descrição da Regra	Sup	Conf	Lift
2008/1	Agronomia 122 ⇒ História=Regular 66	0.02	0.54	1.23
	Ciências Econômicas 114 ⇒ História=Regular 60	0.02	0.53	1.2
	Ciências Biológicas 185 ⇒ História=Regular 96	0.03	0.52	1.18
	Ciências Contábeis 138 ⇒ História=Regular 70	0.02	0.51	1.15
2008/2	Fisioterapia 109 ⇒ História=Regular 69	0.03	0.63	1.33
	Odontologia 309 ⇒ História=Regular 168	0.08	0.54	1.14
	Zootecnia 48 ⇒ História=Regular 25	0.01	0.52	1.1
	Educação Física 87 ⇒ História=Regular 45	0.02	0.52	1.09
	Nutrição 138 ⇒ História=Regular 69	0.03	0.5	1.05
	Sistemas de Informação 108 ⇒ Português e Literatura=Regular 60	0.03	0.56	1.31
	Ciências Biológicas 129 ⇒ Português e Literatura=Regular 71	0.04	0.55	1.29
2009/1	Educação Física 88 ⇒ Geografia=Regular 48	0.01	0.55	1.26
	Zootecnia 73 ⇒ Geografia=Regular 37	0.01	0.51	1.17
	Ciências Contábeis 133 ⇒ Geografia=Regular 67	0.02	0.5	1.16
	Bacharelado em Humanidades 185 ⇒ Geografia=Regular 93	0.03	0.5	1.16
	Farmácia 216 ⇒ Português e Literatura=Regular 128	0.04	0.59	1.24
	Ciências Biológicas 128 ⇒ Português e Literatura=Regular 75	0.02	0.59	1.23
	Nutrição 112 ⇒ Português e Literatura=Regular 63	0.02	0.56	1.18
	Fisioterapia 111 ⇒ Português e Literatura=Regular 60	0.02	0.54	1.13
	Agronomia 120 ⇒ Português e Literatura=Regular 64	0.02	0.53	1.12
	Ciências Econômicas 81 ⇒ Português e Literatura=Regular 41	0.01	0.51	1.06
	Engenharia Florestal 204 ⇒ Português e Literatura=Regular 101	0.03	0.5	1.04
2009/2	Agronomia 75 ⇒ Geografia=Regular 40	0.02	0.53	1.34
	Agronomia 75 ⇒ Português e Literatura=Regular 43	0.02	0.57	1.24
	Fisioterapia 87 ⇒ Português e Literatura=Regular 49	0.02	0.56	1.22
	Zootecnia 61 ⇒ Português e Literatura=Regular 33	0.01	0.54	1.17
	Educação Física 59 ⇒ Português e Literatura=Regular 31	0.01	0.53	1.14

**Tabela 24:** Regras de Associação envolvendo os atributos aprovado, tipo da escola e curso

Ano/Sem	Regras envolvendo escola estadual (EE)				Regras envolvendo escola particular (EP)			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2003/1	EE $\wedge$ Farmácia-Bioq./Industrial <sup>14</sup> 78 $\Rightarrow$ Aprovado=S 66	0.02	0.85	1.09	Aprovado=S $\wedge$ Farmácia-Bioq./Industrial 191 $\Rightarrow$ EP 103	0.03	0.54	1.27
	Aprovado=S $\wedge$ Enfermagem 491 $\Rightarrow$ EE 247	0.08	0.5	1.1	Aprovado=S $\wedge$ Odontologia 370 $\Rightarrow$ EP 197	0.06	0.53	1.25
					Aprovado=S $\wedge$ Fisioterapia 422 $\Rightarrow$ EP 213	0.07	0.5	1.19
					EP $\wedge$ Farmácia-Bioq./Industrial 122 $\Rightarrow$ Aprovado=S 103	0.03	0.84	1.09
					EP $\wedge$ Farmácia-Bioq./A. Clínica <sup>15</sup> 103 $\Rightarrow$ Aprovado=S 86	0.03	0.83	1.08
2003/2	EE $\wedge$ Farmácia-Bioq./A. Clínica 132 $\Rightarrow$ Aprovado=N 123	0.06	0.93	1.31	EP $\wedge$ Odontologia 251 $\Rightarrow$ Aprovado=S 197	0.06	0.78	1.01
	EE $\wedge$ Fisioterapia 252 $\Rightarrow$ Aprovado=N 219	0.11	0.87	1.22	Aprovado=S $\wedge$ Farmácia-Bioq./A. Clínica 53 $\Rightarrow$ EP 40	0.02	0.75	1.76
	EE $\wedge$ Odontologia 149 $\Rightarrow$ Aprovado=N 128	0.06	0.86	1.21	Aprovado=S $\wedge$ Odontologia 93 $\Rightarrow$ EP 65	0.03	0.7	1.63
	EE $\wedge$ Farmácia-Bioq./Industrial 72 $\Rightarrow$ Aprovado=N 61	0.03	0.85	1.19	Aprovado=S $\wedge$ Farmácia-Bioq./Industrial 46 $\Rightarrow$ EP 32	0.02	0.7	1.63
	EE $\wedge$ Nutrição 184 $\Rightarrow$ Aprovado=N 151	0.07	0.82	1.15	Aprovado=S $\wedge$ Fisioterapia 95 $\Rightarrow$ EP 52	0.02	0.55	1.28
	EE $\wedge$ Engenharia Florestal 76 $\Rightarrow$ Aprovado=S 40	0.02	0.53	1.82	EP $\wedge$ Farmácia-Bioq./A. Clínica 141 $\Rightarrow$ Aprovado=N 101	0.05	0.72	1.01
2004/1	EE $\wedge$ Fisioterapia 190 $\Rightarrow$ Aprovado=N 175	0.06	0.92	1.12	EP $\wedge$ Odontologia 163 $\Rightarrow$ Aprovado=N 138	0.04	0.85	1.03
	EE $\wedge$ Farmácia-Bioq./A. Clínica 61 $\Rightarrow$ Aprovado=N 56	0.02	0.92	1.11	EP $\wedge$ Farmácia-Bioq./Industrial 90 $\Rightarrow$ Aprovado=N 76	0.02	0.84	1.02
	EE $\wedge$ Enfermagem 246 $\Rightarrow$ Aprovado=N 224	0.07	0.91	1.1				
	EE $\wedge$ Odontologia 105 $\Rightarrow$ Aprovado=N 94	0.03	0.9	1.09				
	EE $\wedge$ Farmácia-Bioq./Industrial 69 $\Rightarrow$ Aprovado=N 59	0.02	0.86	1.04				
	EE $\wedge$ Nutrição 158 $\Rightarrow$ Aprovado=N 135	0.04	0.85	1.04				
2004/2	EE $\wedge$ Farmácia Bioq./Industrial 57 $\Rightarrow$ Aprovado=N 49	0.03	0.86	1.26	EP $\wedge$ Farmácia-Bioq./Industrial 64 $\Rightarrow$ Aprovado=N 49	0.03	0.77	1.12
	EE $\wedge$ Odontologia 74 $\Rightarrow$ Aprovado=N 59	0.03	0.8	1.17				
	EE $\wedge$ Fisioterapia 128 $\Rightarrow$ Aprovado=N 101	0.06	0.79	1.16				
	EE $\wedge$ Farmácia-Bioq./A. Clínica 41 $\Rightarrow$ Aprovado=N 32	0.02	0.78	1.15				
	EE $\wedge$ Nutrição 98 $\Rightarrow$ Aprovado=N 74	0.04	0.76	1.11				
2005/1	EE $\wedge$ Farmácia Bioq./A. Clínica 106 $\Rightarrow$ Aprovado=N 100	0.03	0.94	1.15	Fisioterapia $\wedge$ Aprovado=S 65 $\Rightarrow$ EP 46	0.02	0.71	1.51
	EE $\wedge$ Farmácia Bioq./Industrial 81 $\Rightarrow$ Aprovado=N 76	0.03	0.94	1.15	Odontologia $\wedge$ Aprovado=S 61 $\Rightarrow$ EP 39	0.01	0.64	1.36

Continua na página seguinte...

<sup>14</sup>Farmácia-Bioquímica/Industrial

<sup>15</sup>Farmácia-Bioquímica/Análise Clínica

**Tabela 24:** Regras de Associação envolvendo os atributos aprovado, tipo da escola e curso (cont)

Ano/Sem	Regras envolvendo escola estadual (EE)				Regras envolvendo escola particular (EP)			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2005/1 (cont)	EE $\wedge$ Fisioterapia 198 $\Rightarrow$ Aprovado=N 183	0.06	0.92	1.13	Enfermagem $\wedge$ Aprovado=S 72 $\Rightarrow$ EP 45	0.02	0.63	1.33
	EE $\wedge$ Enfermagem 282 $\Rightarrow$ Aprovado=N 257	0.09	0.91	1.11	Engenharia Florestal $\wedge$ Aprovado=S 52 $\Rightarrow$ EP 27	0.01	0.52	1.11
	EE $\wedge$ Odontologia 131 $\Rightarrow$ Aprovado=N 118	0.04	0.9	1.1	EP $\wedge$ Odontologia 260 $\Rightarrow$ Aprovado=N 221	0.08	0.85	1.04
2005/2	EE $\wedge$ Farmácia Bioq./A. Clínica 124 $\Rightarrow$ Aprovado=N 113	0.05	0.91	1.21	Farmácia Bioq./Industrial $\wedge$ Aprovado=S 43 $\Rightarrow$ EP 33	0.01	0.77	1.6
	EE $\wedge$ Farmácia Bioq./Industrial 95 $\Rightarrow$ Aprovado=N 86	0.04	0.91	1.2	Odontologia $\wedge$ Aprovado=S 81 $\Rightarrow$ EP 61	0.03	0.75	1.57
	EE $\wedge$ Odontologia 147 $\Rightarrow$ Aprovado=N 132	0.06	0.9	1.19	Farmácia Bioq./A. Clínica $\wedge$ Aprovado=S 42 $\Rightarrow$ EP 30	0.01	0.71	1.49
	EE $\wedge$ Fisioterapia 181 $\Rightarrow$ Aprovado=N 157	0.07	0.87	1.15	Fisioterapia $\wedge$ Aprovado=S 82 $\Rightarrow$ EP 55	0.02	0.67	1.4
	EE $\wedge$ Nutrição 184 $\Rightarrow$ Aprovado=N 159	0.07	0.86	1.14	Engenharia Florestal $\wedge$ Aprovado=S 68 $\Rightarrow$ EP 41	0.02	0.6	1.26
	EE $\wedge$ Engenharia Florestal 94 $\Rightarrow$ Aprovado=N 74	0.03	0.79	1.04	EP $\wedge$ Farmácia Bioq./A. Clínica 145 $\Rightarrow$ Aprovado=N 115	0.05	0.79	1.05
2007/1	Aprovado=S $\wedge$ Serviço Social 74 $\Rightarrow$ EE 64	0.02	0.86	1.65	Aprovado=S $\wedge$ Odontologia 73 $\Rightarrow$ EP 55	0.02	0.75	1.8
	Aprovado=S $\wedge$ Ciências Econômicas 58 $\Rightarrow$ EE 42	0.01	0.72	1.38	Aprovado=S $\wedge$ Enfermagem 76 $\Rightarrow$ EP 51	0.02	0.67	1.6
	Aprovado=S $\wedge$ Ciências Contábeis 60 $\Rightarrow$ EE 42	0.01	0.7	1.34	Aprovado=S $\wedge$ Farmácia 72 $\Rightarrow$ EP 47	0.01	0.65	1.56
	Aprovado=S $\wedge$ Educação Física 75 $\Rightarrow$ EE 50	0.02	0.67	1.27	Aprovado=S $\wedge$ Engenharia Florestal 60 $\Rightarrow$ EP 36	0.01	0.6	1.43
	Aprovado=S $\wedge$ Administração 75 $\Rightarrow$ EE 41	0.01	0.55	1.04	Aprovado=S $\wedge$ Ciências Biológicas 73 $\Rightarrow$ EP 38	0.01	0.52	1.24
	Aprovado=S $\wedge$ Sistemas de Informação 73 $\Rightarrow$ EE 39	0.01	0.53	1.02	Aprovado=S $\wedge$ Fisioterapia 73 $\Rightarrow$ EP 37	0.01	0.51	1.21
	EE $\wedge$ Enfermagem 203 $\Rightarrow$ Aprovado=N 183	0.06	0.9	1.4	EP $\wedge$ Odontologia 310 $\Rightarrow$ Aprovado=N 255	0.08	0.82	1.28
	EE $\wedge$ Odontologia 135 $\Rightarrow$ Aprovado=N 121	0.04	0.9	1.4	EP $\wedge$ Farmácia 181 $\Rightarrow$ Aprovado=N 134	0.04	0.74	1.15
	EE $\wedge$ Farmácia 149 $\Rightarrow$ Aprovado=N 131	0.04	0.88	1.37	EP $\wedge$ Ciências Biológicas 74 $\Rightarrow$ Aprovado=S 38	0.01	0.51	1.43
	EE $\wedge$ Ciências Econômicas 53 $\Rightarrow$ Aprovado=S 42	0.01	0.79	2.21				
	EE $\wedge$ Nutrição 92 $\Rightarrow$ Aprovado=N 69	0.02	0.75	1.17				
	EE $\wedge$ Fisioterapia 116 $\Rightarrow$ Aprovado=N 85	0.03	0.73	1.14				
	EE $\wedge$ Ciências Biológicas 108 $\Rightarrow$ Aprovado=N 79	0.02	0.73	1.14				
	EE $\wedge$ Engenharia Florestal 81 $\Rightarrow$ Aprovado=N 58	0.02	0.72	1.12				
	EE $\wedge$ Ciências Contábeis 62 $\Rightarrow$ Aprovado=S 42	0.01	0.68	1.89				
	EE $\wedge$ Sistemas de Informação 116 $\Rightarrow$ Aprovado=N 77	0.02	0.66	1.03				
EE $\wedge$ Educação Física 100 $\Rightarrow$ Aprovado=S 50	0.02	0.5	1.4					

Continua na página seguinte...



**Tabela 24:** Regras de Associação envolvendo os atributos aprovado, tipo da escola e curso (cont)

Ano/Sem	Regras envolvendo escola estadual (EE)				Regras envolvendo escola particular (EP)			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2007/2	Aprovado=S $\wedge$ Ciências Contábeis 74 $\Rightarrow$ EE 62	0.02	0.84	1.43	Aprovado=S $\wedge$ Odontologia 94 $\Rightarrow$ EP 73	0.03	0.78	2.17
	Aprovado=S $\wedge$ Serviço Social 90 $\Rightarrow$ EE 75	0.03	0.83	1.42	Aprovado=S $\wedge$ Farmácia 90 $\Rightarrow$ EP 60	0.02	0.67	1.86
	Aprovado=S $\wedge$ Administração 90 $\Rightarrow$ EE 66	0.02	0.73	1.25	Aprovado=S $\wedge$ Engenharia Florestal 75 $\Rightarrow$ EP 45	0.02	0.6	1.67
	Aprovado=S $\wedge$ Ciências Econômicas 51 $\Rightarrow$ EE 37	0.01	0.73	1.24	EP $\wedge$ Ciências Biológicas 62 $\Rightarrow$ Aprovado=S 43	0.02	0.69	1.45
	Aprovado=S $\wedge$ Educação Física 90 $\Rightarrow$ EE 63	0.02	0.7	1.19	EP $\wedge$ Odontologia 228 $\Rightarrow$ Aprovado=N 155	0.06	0.68	1.3
	Aprovado=S $\wedge$ Sistemas de Informação 92 $\Rightarrow$ EE 56	0.02	0.61	1.04	EP $\wedge$ Farmácia 181 $\Rightarrow$ Aprovado=N 121	0.04	0.67	1.28
	EE $\wedge$ Odontologia 141 $\Rightarrow$ Aprovado=N 125	0.05	0.89	1.7				
	EE $\wedge$ Farmácia 182 $\Rightarrow$ Aprovado=N 153	0.06	0.84	1.61				
	EE $\wedge$ Engenharia Florestal 92 $\Rightarrow$ Aprovado=N 67	0.02	0.73	1.39				
	EE $\wedge$ Educação Física 89 $\Rightarrow$ Aprovado=S 63	0.02	0.7	1.19				
	EE $\wedge$ Ciências Contábeis 88 $\Rightarrow$ Aprovado=S 62	0.02	0.7	1.48				
	EE $\wedge$ Sistemas de Informação 82 $\Rightarrow$ Aprovado=S 56	0.02	0.68	1.43				
	EE $\wedge$ Ciências Econômicas 55 $\Rightarrow$ Aprovado=S 37	0.01	0.67	1.41				
	EE $\wedge$ Nutrição 108 $\Rightarrow$ Aprovado=N 69	0.03	0.64	1.22				
	EE $\wedge$ Fisioterapia 132 $\Rightarrow$ Aprovado=N 78	0.03	0.59	1.13				
	EE $\wedge$ Administração 143 $\Rightarrow$ Aprovado=N 77	0.03	0.54	1.03				
EE $\wedge$ Serviço Social 160 $\Rightarrow$ Aprovado=N 85	0.03	0.53	1.02					
2008/1	Aprovado=S $\wedge$ Ciências Contábeis 73 $\Rightarrow$ EE 58	0.02	0.79	1.29	Aprovado=S $\wedge$ Odontologia 72 $\Rightarrow$ EP 59	0.02	0.82	2.42
	Aprovado=S $\wedge$ Educação Física 72 $\Rightarrow$ EE 56	0.02	0.78	1.27	Aprovado=S $\wedge$ Farmácia 74 $\Rightarrow$ EP 54	0.02	0.73	2.16
	Aprovado=S $\wedge$ Serviço Social 72 $\Rightarrow$ EE 51	0.01	0.71	1.15	Aprovado=S $\wedge$ Enfermagem 72 $\Rightarrow$ EP 39	0.01	0.54	1.6
	Aprovado=S $\wedge$ Ciências Econômicas 73 $\Rightarrow$ EE 51	0.01	0.7	1.14	Aprovado=S $\wedge$ Ciências Biológicas 72 $\Rightarrow$ EP 36	0.01	0.5	1.48
	Aprovado=S $\wedge$ Administração 73 $\Rightarrow$ EE 48	0.01	0.66	1.07	EP $\wedge$ Odontologia 279 $\Rightarrow$ Aprovado=N 220	0.06	0.79	1.2
	EE $\wedge$ Odontologia 171 $\Rightarrow$ Aprovado=N 160	0.05	0.94	1.42	EP $\wedge$ Farmácia 182 $\Rightarrow$ Aprovado=N 128	0.04	0.7	1.07
	EE $\wedge$ Farmácia 153 $\Rightarrow$ Aprovado=N 137	0.04	0.9	1.36	EP $\wedge$ Ciências Biológicas 56 $\Rightarrow$ Aprovado=S 36	0.01	0.64	1.88
	EE $\wedge$ Nutrição 126 $\Rightarrow$ Aprovado=N 106	0.03	0.84	1.28				
	EE $\wedge$ Enfermagem 170 $\Rightarrow$ Aprovado=N 139	0.04	0.82	1.24				
	EE $\wedge$ Engenharia Florestal 131 $\Rightarrow$ Aprovado=N 106	0.03	0.81	1.23				

Continua na página seguinte...

**Tabela 24:** Regras de Associação envolvendo os atributos aprovado, tipo da escola e curso (cont)

Ano/Sem	Regras envolvendo escola estadual (EE)				Regras envolvendo escola particular (EP)			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2008/1 (cont)	EE $\wedge$ Serviço Social 243 $\Rightarrow$ Aprovado=N 192	0.06	0.79	1.2				
	EE $\wedge$ Administração 221 $\Rightarrow$ Aprovado=N 173	0.05	0.78	1.19				
	EE $\wedge$ Ciências Biológicas 122 $\Rightarrow$ Aprovado=N 88	0.03	0.72	1.1				
	EE $\wedge$ Ciências Econômicas 82 $\Rightarrow$ Aprovado=S 51	0.01	0.62	1.82				
	EE $\wedge$ Educação Física 100 $\Rightarrow$ Aprovado=S 56	0.02	0.56	1.64				
	EE $\wedge$ Ciências Contábeis 113 $\Rightarrow$ Aprovado=S 58	0.02	0.51	1.5				
2008/2	Aprovado=S $\wedge$ Serviço Social 65 $\Rightarrow$ EE 59	0.03	0.91	1.48	Aprovado=S $\wedge$ Odontologia 91 $\Rightarrow$ EP 62	0.03	0.68	2.03
	Aprovado=S $\wedge$ Educação Física 72 $\Rightarrow$ EE 62	0.03	0.86	1.4	Aprovado=S $\wedge$ Farmácia 90 $\Rightarrow$ EP 57	0.03	0.63	1.89
	Aprovado=S $\wedge$ Ciências Contábeis 67 $\Rightarrow$ EE 56	0.03	0.84	1.36	Aprovado=S $\wedge$ Engenharia Florestal 75 $\Rightarrow$ EP 39	0.02	0.52	1.55
	Aprovado=S $\wedge$ Administração 76 $\Rightarrow$ EE 62	0.03	0.82	1.33	EP $\wedge$ Ciências Biológicas 34 $\Rightarrow$ Aprovado=S 29	0.01	0.85	1.53
	Aprovado=S $\wedge$ Sistemas de Informação 77 $\Rightarrow$ EE 60	0.03	0.78	1.27	EP $\wedge$ Odontologia 167 $\Rightarrow$ Aprovado=N 105	0.05	0.63	1.42
	Aprovado=S $\wedge$ Ciências Econômicas 46 $\Rightarrow$ EE 33	0.02	0.72	1.17	EP $\wedge$ Farmácia 128 $\Rightarrow$ Aprovado=N 71	0.04	0.55	1.26
	Aprovado=S $\wedge$ Ciências Biológicas 90 $\Rightarrow$ EE 60	0.03	0.67	1.09				
	Aprovado=S $\wedge$ Fisioterapia 75 $\Rightarrow$ EE 47	0.02	0.63	1.02				
	EE $\wedge$ Educação Física 75 $\Rightarrow$ Aprovado=S 62	0.03	0.83	1.48				
	EE $\wedge$ Odontologia 118 $\Rightarrow$ Aprovado=N 94	0.05	0.8	1.8				
	EE $\wedge$ Ciências Contábeis 71 $\Rightarrow$ Aprovado=S 56	0.03	0.79	1.41				
	EE $\wedge$ Farmácia 114 $\Rightarrow$ Aprovado=N 86	0.04	0.75	1.71				
	EE $\wedge$ Ciências Econômicas 44 $\Rightarrow$ Aprovado=S 33	0.02	0.75	1.34				
	EE $\wedge$ Sistemas de Informação 85 $\Rightarrow$ Aprovado=S 60	0.03	0.71	1.26				
	EE $\wedge$ Fisioterapia 68 $\Rightarrow$ Aprovado=S 47	0.02	0.69	1.24				
	EE $\wedge$ Engenharia Florestal 89 $\Rightarrow$ Aprovado=N 58	0.03	0.65	1.48				
	EE $\wedge$ Ciências Biológicas 92 $\Rightarrow$ Aprovado=S 60	0.03	0.65	1.17				
	EE $\wedge$ Administração 107 $\Rightarrow$ Aprovado=S 62	0.03	0.58	1.04				
	EE $\wedge$ Nutrição 88 $\Rightarrow$ Aprovado=N 47	0.02	0.53	1.21				
	EE $\wedge$ Serviço Social 122 $\Rightarrow$ Aprovado=N 63	0.03	0.52	1.17				

Continua na página seguinte...

**Tabela 24:** Regras de Associação envolvendo os atributos aprovado, tipo da escola e curso (cont)

Ano/Sem	Regras envolvendo escola estadual (EE)			Regras envolvendo escola particular (EP)				
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2009/1	Aprovado=S $\wedge$ Ciências Contábeis 72 $\Rightarrow$ EE 63	0.02	0.88	1.29	Aprovado=S $\wedge$ Odontologia 73 $\Rightarrow$ EP 59	0.02	0.81	2.96
	Aprovado=S $\wedge$ Educação Física 57 $\Rightarrow$ EE 48	0.01	0.84	1.25	Aprovado=S $\wedge$ Farmácia 72 $\Rightarrow$ EP 45	0.01	0.63	2.29
	Aprovado=S $\wedge$ Ciências Econômicas 51 $\Rightarrow$ EE 42	0.01	0.82	1.22	EP $\wedge$ BCT - D <sup>16</sup> 88 $\Rightarrow$ Aprovado=S 74	0.02	0.84	1.7
	Aprovado=S $\wedge$ Bacharelado em Humanidades 160 $\Rightarrow$ EE 130	0.04	0.81	1.2	EP $\wedge$ BCT - TO <sup>17</sup> 122 $\Rightarrow$ Aprovado=S 89	0.03	0.73	1.48
	Aprovado=S $\wedge$ Serviço Social 73 $\Rightarrow$ EE 57	0.02	0.78	1.15	EP $\wedge$ Odontologia 221 $\Rightarrow$ Aprovado=N 162	0.05	0.73	1.45
	Aprovado=S $\wedge$ Sistemas de Informação 72 $\Rightarrow$ EE 55	0.02	0.76	1.13				
	Aprovado=S $\wedge$ Administração 81 $\Rightarrow$ EE 58	0.02	0.72	1.06				
	Aprovado=S $\wedge$ Ciências Biológicas 72 $\Rightarrow$ EE 50	0.01	0.69	1.03				
	EE $\wedge$ Odontologia 155 $\Rightarrow$ Aprovado=N 142	0.04	0.92	1.81				
	EE $\wedge$ Bacharelado em Humanidades 150 $\Rightarrow$ Aprovado=S 130	0.04	0.87	1.75				
	EE $\wedge$ BCT - D 171 $\Rightarrow$ Aprovado=S 142	0.04	0.83	1.68				
	EE $\wedge$ Farmácia 121 $\Rightarrow$ Aprovado=N 95	0.03	0.79	1.55				
	EE $\wedge$ Engenharia Florestal 116 $\Rightarrow$ Aprovado=N 89	0.03	0.77	1.52				
	EE $\wedge$ Administração 231 $\Rightarrow$ Aprovado=N 173	0.05	0.75	1.48				
	EE $\wedge$ Serviço Social 222 $\Rightarrow$ Aprovado=N 165	0.05	0.74	1.47				
	EE $\wedge$ Enfermagem 121 $\Rightarrow$ Aprovado=N 84	0.02	0.69	1.37				
	EE $\wedge$ Ciências Econômicas 66 $\Rightarrow$ Aprovado=S 42	0.01	0.64	1.29				
	EE $\wedge$ Educação Física 76 $\Rightarrow$ Aprovado=S 48	0.01	0.63	1.28				
	EE $\wedge$ BCT - TO 296 $\Rightarrow$ Aprovado=S 182	0.05	0.61	1.24				
	EE $\wedge$ Fisioterapia 72 $\Rightarrow$ Aprovado=S 42	0.01	0.58	1.18				
EE $\wedge$ Sistemas de Informação 101 $\Rightarrow$ Aprovado=S 55	0.02	0.54	1.1					
EE $\wedge$ Nutrição 67 $\Rightarrow$ Aprovado=S 36	0.01	0.54	1.09					
EE $\wedge$ Ciências Contábeis 119 $\Rightarrow$ Aprovado=S 63	0.02	0.53	1.07					
EE $\wedge$ Ciências Biológicas 95 $\Rightarrow$ Aprovado=S 50	0.01	0.53	1.07					

Continua na página seguinte...

<sup>16</sup>Bacharelado em Ciência e Tecnologia - Diamantina

<sup>17</sup>Bacharelado em Ciência e Tecnologia - Teófilo Otoni

**Tabela 24:** Regras de Associação envolvendo os atributos aprovado, tipo da escola e curso (cont)

Ano/Sem	Regras envolvendo escola estadual (EE)				Regras envolvendo escola particular (EP)			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2009/2	Aprovado=S $\wedge$ Ciências Contábeis 58 $\Rightarrow$ EE 49	0.02	0.84	1.34	Aprovado=S $\wedge$ Odontologia 90 $\Rightarrow$ EP 66	0.03	0.73	2.31
	Aprovado=S $\wedge$ Educação Física 45 $\Rightarrow$ EE 38	0.02	0.84	1.34	Aprovado=S $\wedge$ Farmácia 92 $\Rightarrow$ EP 46	0.02	0.5	1.58
	Aprovado=S $\wedge$ Serviço Social 49 $\Rightarrow$ EE 41	0.02	0.84	1.33	EP $\wedge$ BCT - D 123 $\Rightarrow$ Aprovado=S 99	0.04	0.8	1.24
	Aprovado=S $\wedge$ Bacharelado em Humanidades 149 $\Rightarrow$ EE 122	0.05	0.82	1.3	EP $\wedge$ BCT - TO 77 $\Rightarrow$ Aprovado=S 60	0.03	0.78	1.2
	Aprovado=S $\wedge$ Ciências Econômicas 48 $\Rightarrow$ EE 39	0.02	0.81	1.29	EP $\wedge$ Odontologia 167 $\Rightarrow$ Aprovado=N 101	0.04	0.6	1.74
	Aprovado=S $\wedge$ Ciências Biológicas 81 $\Rightarrow$ EE 60	0.03	0.74	1.18				
	Aprovado=S $\wedge$ Administração 75 $\Rightarrow$ EE 55	0.02	0.73	1.16				
	Aprovado=S $\wedge$ Fisioterapia 66 $\Rightarrow$ EE 48	0.02	0.73	1.15				
	EE $\wedge$ Ciências Econômicas 45 $\Rightarrow$ Aprovado=S 39	0.02	0.87	1.33				
	EE $\wedge$ Sistemas de Informação 35 $\Rightarrow$ Aprovado=S 29	0.01	0.83	1.27				
	EE $\wedge$ Odontologia 106 $\Rightarrow$ Aprovado=N 86	0.04	0.81	2.33				
	EE $\wedge$ Bacharelado em Humanidades 150 $\Rightarrow$ Aprovado=S 122	0.05	0.81	1.25				
	EE $\wedge$ BCT - D 145 $\Rightarrow$ Aprovado=S 117	0.05	0.81	1.24				
	EE $\wedge$ Fisioterapia 61 $\Rightarrow$ Aprovado=S 48	0.02	0.79	1.21				
	EE $\wedge$ Educação Física 49 $\Rightarrow$ Aprovado=S 38	0.02	0.78	1.19				
	EE $\wedge$ Ciências Contábeis 64 $\Rightarrow$ Aprovado=S 49	0.02	0.77	1.18				
	EE $\wedge$ BCT - TO 148 $\Rightarrow$ Aprovado=S 111	0.05	0.75	1.15				
	EE $\wedge$ Ciências Biológicas 81 $\Rightarrow$ Aprovado=S 60	0.03	0.74	1.14				
	EE $\wedge$ Farmácia 90 $\Rightarrow$ Aprovado=N 51	0.02	0.57	1.63				
	EE $\wedge$ Serviço Social 92 $\Rightarrow$ Aprovado=N 51	0.02	0.55	1.59				
EE $\wedge$ Administração 113 $\Rightarrow$ Aprovado=N 58	0.03	0.51	1.47					

**Tabela 25:** Regras de Associação envolvendo os atributos aprovado, local de residência e curso

Ano/Sem	Local de residência	Antecedente da regra							
		Local de residência e curso				Aprovado e curso			
		Descrição da regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2005/1	Diamantina (D)	D $\wedge$ Enfermagem 87 $\Rightarrow$ Aprovado=N 78	0.03	0.9	1.1				
		D $\wedge$ Fisioterapia 75 $\Rightarrow$ Aprovado=N 67	0.02	0.89	1.09				
		D $\wedge$ Nutrição 64 $\Rightarrow$ Aprovado=N 56	0.02	0.88	1.07				
2005/1	Vale do Jequitinhonha (VJ)	VJ $\wedge$ Enfermagem 49 $\Rightarrow$ Aprovado=N 49	0.02	1	1.22				
		OR $\wedge$ Enfermagem 267 $\Rightarrow$ Aprovado=N 242	0.08	0.91	1.11	Engenharia Florestal $\wedge$ Aprovado=S 52 $\Rightarrow$ OR 35	0.01	0.67	1.21
		OR $\wedge$ Farmácia-Bioq./A. Clínica <sup>18</sup> 159 $\Rightarrow$ Aprovado=N 144	0.05	0.91	1.11	Zootecnia $\wedge$ Aprovado=S 59 $\Rightarrow$ OR 37	0.01	0.63	1.13
2005/1	Outra Região de MG (OR)	OR $\wedge$ Odontologia 241 $\Rightarrow$ Aprovado=N 209	0.07	0.87	1.06	Agronomia $\wedge$ Aprovado=S 53 $\Rightarrow$ OR 31	0.01	0.58	1.06
		D $\wedge$ Farmácia-Bioq./A. Clínica 32 $\Rightarrow$ Aprovado=N 31	0.01	0.97	1.28				
		D $\wedge$ Farmácia-Bioq./Industrial <sup>19</sup> 35 $\Rightarrow$ Aprovado=N 32	0.01	0.91	1.21				
2005/2	Diamantina (D)	D $\wedge$ Fisioterapia 80 $\Rightarrow$ Aprovado=N 68	0.03	0.85	1.13				
		D $\wedge$ Nutrição 90 $\Rightarrow$ Aprovado=N 75	0.03	0.83	1.1				
		BH $\wedge$ Farmácia-Bioq./A. Clínica 63 $\Rightarrow$ Aprovado=N 56	0.02	0.89	1.18				
2005/2	Belo Horizonte (BH)	OR $\wedge$ Odontologia 268 $\Rightarrow$ Aprovado=N 218	0.09	0.81	1.08	Agronomia $\wedge$ Aprovado=S 67 $\Rightarrow$ OR 48	0.02	0.72	1.25
		OR $\wedge$ Farmácia-Bioq./A. Clínica 163 $\Rightarrow$ Aprovado=N 131	0.05	0.8	1.06	Zootecnia $\wedge$ Aprovado=S 74 $\Rightarrow$ OR 49	0.02	0.66	1.15
		OR $\wedge$ Farmácia-Bioq./Industrial 123 $\Rightarrow$ Aprovado=N 98	0.04	0.8	1.06	Engenharia Florestal $\wedge$ Aprovado=S 68 $\Rightarrow$ OR 44	0.02	0.65	1.13
2007/1	Diamantina (D)	Odontologia $\wedge$ Aprovado=S 81 $\Rightarrow$ OR 50	0.02	0.62	1.07				
		D $\wedge$ Farmácia 53 $\Rightarrow$ Aprovado=N 48	0.01	0.91	1.41				
		D $\wedge$ Enfermagem 76 $\Rightarrow$ Aprovado=N 67	0.02	0.88	1.37				
2007/1	Diamantina (D)	D $\wedge$ Ciências Biológicas 66 $\Rightarrow$ Aprovado=N 45	0.01	0.68	1.06				

Continua na página seguinte...

<sup>18</sup>Farmácia-Bioquímica/Análise Clínica

<sup>19</sup>Farmácia-Bioquímica/Industrial

**Tabela 25:** Regras de Associação envolvendo os atributos aprovado, local de residência e curso (cont)

Ano/Sem	Local de residência	Antecedente da regra								
		Local de residência e curso				Aprovado e curso				
		Descrição da regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift	
2007/1 (cont)	<b>Teófilo Otoni (TO)</b>	TO $\wedge$ Ciências Contábeis 51 $\Rightarrow$ Aprovado=S 40	0.01	0.78	2.19	Aprovado=S $\wedge$ Ciências Contábeis 60 $\Rightarrow$ TO 40	0.01	0.67	6.38	
					Aprovado=S $\wedge$ Administração 75 $\Rightarrow$ TO 46	0.01	0.61	5.87		
					Aprovado=S $\wedge$ Serviço Social 74 $\Rightarrow$ TO 38	0.01	0.51	4.91		
	<b>Vale do Jequitinhonha (VJ)</b>	VJ $\wedge$ Enfermagem 60 $\Rightarrow$ Aprovado=N 57	0.02	0.95	1.48					
	<b>Belo Horizonte (BH)</b>	BH $\wedge$ Odontologia 111 $\Rightarrow$ Aprovado=N 76	0.02	0.68	1.07					
	<b>Outra Região de MG (OR)</b>	OR $\wedge$ Odontologia 293 $\Rightarrow$ Aprovado=N 258	0.08	0.88	1.37	Aprovado=S $\wedge$ Agronomia 60 $\Rightarrow$ OR 38	0.01	0.63	1.49	
		OR $\wedge$ Enfermagem 158 $\Rightarrow$ Aprovado=N 133	0.04	0.84	1.31	Aprovado=S $\wedge$ Engenharia Florestal 60 $\Rightarrow$ OR 38	0.01	0.63	1.49	
		OR Farmácia 197 $\Rightarrow$ Aprovado=N 154	0.05	0.78	1.22	Aprovado=S $\wedge$ Zootecnia 62 $\Rightarrow$ OR 35	0.01	0.56	1.33	
2007/2	<b>Diamantina (D)</b>	D $\wedge$ Farmácia 56 $\Rightarrow$ Aprovado=N 48	0.02	0.86	1.64	Aprovado=S $\wedge$ Educação Física 90 $\Rightarrow$ D 49	0.02	0.54	3.34	
		D $\wedge$ Educação Física 61 $\Rightarrow$ Aprovado=S 49	0.02	0.8	1.68					
		D $\wedge$ Nutrição 48 $\Rightarrow$ Aprovado=N 30	0.01	0.63	1.2					
		D $\wedge$ Fisioterapia 56 $\Rightarrow$ Aprovado=N 33	0.01	0.59	1.13					
		<b>Teófilo Otoni (TO)</b>	TO $\wedge$ Ciências Contábeis 58 $\Rightarrow$ Aprovado=S 44	0.02	0.76	1.59	Aprovado=S $\wedge$ Administração 90 $\Rightarrow$ TO 54	0.02	0.6	4.3
						Aprovado=S $\wedge$ Ciências Contábeis 74 $\Rightarrow$ TO 44	0.02	0.59	4.27	
						Aprovado=S $\wedge$ Serviço Social 90 $\Rightarrow$ TO 48	0.02	0.53	3.83	
		<b>Belo Horizonte (BH)</b>	BH $\wedge$ Odontologia 84 $\Rightarrow$ Aprovado=N 60	0.02	0.71	1.37				
			BH $\wedge$ Farmácia 76 $\Rightarrow$ Aprovado=N 54	0.02	0.71	1.36				
		<b>Outra Região de MG (MG)</b>	OR $\wedge$ Zootecnia 60 $\Rightarrow$ Aprovado=S 46	0.02	0.77	1.61	Aprovado=S $\wedge$ Zootecnia 62 $\Rightarrow$ OR 46	0.02	0.74	1.81
		OR $\wedge$ Odontologia 232 $\Rightarrow$ Aprovado=N 176	0.06	0.76	1.45	Aprovado=S $\wedge$ Engenharia Florestal 75 $\Rightarrow$ OR 47	0.02	0.63	1.53	
		OR $\wedge$ Farmácia 191 $\Rightarrow$ Aprovado=N 137	0.05	0.72	1.37	Aprovado=S $\wedge$ Agronomia 78 $\Rightarrow$ OR 48	0.02	0.62	1.5	
		OR $\wedge$ Agronomia 79 $\Rightarrow$ Aprovado=S 48	0.02	0.61	1.27	Aprovado=S $\wedge$ Odontologia 94 $\Rightarrow$ OR 56	0.02	0.6	1.45	
		OR $\wedge$ Engenharia Florestal 108 $\Rightarrow$ Aprovado=N 61	0.02	0.56	1.08					

Continua na página seguinte...

**Tabela 25:** Regras de Associação envolvendo os atributos aprovado, local de residência e curso (cont)

Ano/Sem	Local de residência	Antecedente da regra							
		Local de residência e curso				Aprovado e curso			
		Descrição da regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2008/1	Diamantina (D)	D $\wedge$ Farmácia 60 $\Rightarrow$ Aprovado=N 51	0.01	0.85	1.29	Aprovado=S $\wedge$ Educação Física 72 $\Rightarrow$ D 42	0.01	0.58	3.38
		D $\wedge$ Nutrição 47 $\Rightarrow$ Aprovado=N 37	0.01	0.79	1.2				
		D $\wedge$ Ciências Biológicas 73 $\Rightarrow$ Aprovado=N 54	0.02	0.74	1.13				
		D $\wedge$ Enfermagem 75 $\Rightarrow$ Aprovado=N 54	0.02	0.72	1.1				
		D $\wedge$ Educação Física 67 $\Rightarrow$ Aprovado=S 42	0.01	0.63	1.83				
	Teófilo Otoni (TO)	TO $\wedge$ Ciências Econômicas 64 $\Rightarrow$ Aprovado=S 49	0.01	0.77	2.24	Aprovado=S $\wedge$ Ciências Econômicas 73 $\Rightarrow$ TO 49	0.01	0.67	3.94
TO $\wedge$ Administração 176 $\Rightarrow$ Aprovado=N 130		0.04	0.74	1.12	Aprovado=S $\wedge$ Administração 73 $\Rightarrow$ TO 46	0.01	0.63	3.7	
TO $\wedge$ Serviço Social 167 $\Rightarrow$ Aprovado=N 123		0.04	0.74	1.12	Aprovado=S $\wedge$ Serviço Social 72 $\Rightarrow$ TO 44	0.01	0.61	3.58	
Vale do Jequitinhonha (VJ)	VJ $\wedge$ Enfermagem 40 $\Rightarrow$ Aprovado=N 36	0.01	0.9	1.37	Aprovado=S $\wedge$ Ciências Contábeis 73 $\Rightarrow$ TO 40	0.01	0.55	3.21	
Vale do Mucuri (VM)	VM $\wedge$ Serviço Social 46 $\Rightarrow$ Aprovado=N 41	0.01	0.89	1.36					
Belo Horizonte (BH)	BH $\wedge$ Odontologia 113 $\Rightarrow$ Aprovado=N 81	0.02	0.72	1.09					
Outra Região de MG (OR)	OR $\wedge$ Odontologia 275 $\Rightarrow$ Aprovado=N 243	0.07	0.88	1.34	Aprovado=S $\wedge$ Engenharia Florestal 60 $\Rightarrow$ OR 41	0.01	0.68	1.94	
	OR $\wedge$ Farmácia 164 $\Rightarrow$ Aprovado=N 129	0.04	0.79	1.2	Aprovado=S $\wedge$ Zootecnia 63 $\Rightarrow$ OR 43	0.01	0.68	1.94	
	OR $\wedge$ Enfermagem 113 $\Rightarrow$ Aprovado=N 88	0.03	0.78	1.18					
	OR $\wedge$ Zootecnia 66 $\Rightarrow$ Aprovado=S 43	0.01	0.65	1.9					
2008/2	Diamantina (D)	D $\wedge$ Educação Física 52 $\Rightarrow$ Aprovado=S 43	0.02	0.83	1.48	Aprovado=S $\wedge$ Educação Física 72 $\Rightarrow$ D 43	0.02	0.6	2.77
D $\wedge$ Fisioterapia 34 $\Rightarrow$ Aprovado=S 24		0.01	0.71	1.26					
D $\wedge$ Farmácia 54 $\Rightarrow$ Aprovado=N 38		0.02	0.7	1.59					
D $\wedge$ Ciências Biológicas 54 $\Rightarrow$ Aprovado=S 34		0.02	0.63	1.13					
D $\wedge$ Nutrição 54 $\Rightarrow$ Aprovado=N 30		0.01	0.56	1.26					

Continua na página seguinte...

**Tabela 25:** Regras de Associação envolvendo os atributos aprovado, local de residência e curso (cont)

Ano/Sem	Local de residência	Antecedente da regra							
		Local de residência e curso				Aprovado e curso			
		Descrição da regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2008/2 (cont)	Teófilo Otoni (TO)	TO $\wedge$ Ciências Contábeis 46 $\Rightarrow$ Aprovado=S 41	0.02	0.89	1.6	Aprovado=S $\wedge$ Ciências Contábeis 67 $\Rightarrow$ TO 41	0.02	0.61	4.69
		TO $\wedge$ Ciências Econômicas 32 $\Rightarrow$ Aprovado=S 25	0.01	0.78	1.4	Aprovado=S $\wedge$ Serviço Social 65 $\Rightarrow$ TO 37	0.02	0.57	4.36
	Belo Horizonte (BH)	BH $\wedge$ Odontologia 69 $\Rightarrow$ Aprovado=N 48	0.02	0.7	1.58	Aprovado=S $\wedge$ Administração 76 $\Rightarrow$ TO 43	0.02	0.57	4.33
2009/1	Outra Região de MG (OR)	OR $\wedge$ Farmácia 51 $\Rightarrow$ Aprovado=N 27	0.01	0.53	1.2	Aprovado=S $\wedge$ Ciências Econômicas 46 $\Rightarrow$ TO 25	0.01	0.54	4.16
		OR $\wedge$ Agronomia 54 $\Rightarrow$ Aprovado=S 39	0.02	0.72	1.29				
	OR $\wedge$ Odontologia 152 $\Rightarrow$ Aprovado=N 106	0.05	0.7	1.58	Aprovado=S Zootecnia 40 $\Rightarrow$ OR 21	0.01	0.53	1.65	
	OR $\wedge$ Zootecnia 26 $\Rightarrow$ Aprovado=S 21	0.01	0.81	1.45	Aprovado=S Agronomia 75 $\Rightarrow$ OR 39	0.02	0.52	1.64	
	OR $\wedge$ Farmácia 100 $\Rightarrow$ Aprovado=N 60	0.03	0.6	1.36	Aprovado=S $\wedge$ Odontologia 91 $\Rightarrow$ OR 46	0.02	0.51	1.59	
		OR $\wedge$ Engenharia Florestal 76 $\Rightarrow$ Aprovado=N 40	0.02	0.53	1.19				
2009/1	Diamantina (D)	D $\wedge$ BCT - D <sup>20</sup> 111 $\Rightarrow$ Aprovado=S 96	0.03	0.86	1.75	Aprovado=S $\wedge$ BHU <sup>21</sup> 160 $\Rightarrow$ D 103	0.03	0.64	3.49
		D $\wedge$ BHU 120 $\Rightarrow$ Aprovado=S 103	0.03	0.86	1.74				
		D $\wedge$ Enfermagem 74 $\Rightarrow$ Aprovado=N 51	0.01	0.69	1.36				
	Teófilo Otoni (TO)	TO $\wedge$ Administração 184 $\Rightarrow$ Aprovado=N 134	0.04	0.73	1.44	Aprovado=S $\wedge$ Ciências Contábeis 72 $\Rightarrow$ TO 48	0.01	0.67	2.94
TO $\wedge$ Serviço Social 143 $\Rightarrow$ Aprovado=N 104		0.03	0.73	1.44	Aprovado=S $\wedge$ Administração 81 $\Rightarrow$ TO 50	0.01	0.62	2.72	
		TO $\wedge$ BCT - TO <sup>22</sup> 281 $\Rightarrow$ Aprovado=S 186	0.05	0.66	1.34	Aprovado=S $\wedge$ Serviço Social 73 $\Rightarrow$ TO 39	0.01	0.53	2.36
		TO $\wedge$ Ciências Contábeis 82 $\Rightarrow$ Aprovado=S 48	0.01	0.59	1.18				
	Vale do Jequitinhonha (VJ)	VJ $\wedge$ BCT - D 43 $\Rightarrow$ Aprovado=S 37	0.01	0.86	1.74				
	Vale do Mucuri (VM)	VM $\wedge$ Serviço Social 52 $\Rightarrow$ Aprovado=N 39	0.01	0.75	1.48				
		VM $\wedge$ BCT - TO 69 $\Rightarrow$ Aprovado=S 39	0.01	0.57	1.14				

Continua na página seguinte...

<sup>20</sup>Bacharelado em Ciência e Tecnologia - Diamantina

<sup>21</sup>Bacharelado em Humanidades

<sup>22</sup>Bacharelado em Ciência e Tecnologia - Teófilo Otoni



**Tabela 25:** Regras de Associação envolvendo os atributos aprovado, local de residência e curso (cont)

Ano/Sem	Local de residência	Antecedente da regra							
		Local de residência e curso				Aprovado e curso			
		Descrição da regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2009/1 (cont)	Belo Horizonte (BH)	BH $\wedge$ Odontologia 89 $\Rightarrow$ Aprovado=N 64	0.02	0.72	1.42				
	Outra Região de MG (OR)	OR $\wedge$ Odontologia 202 $\Rightarrow$ Aprovado=N 167	0.05	0.83	1.63				
		OR $\wedge$ BCT - D 73 $\Rightarrow$ Aprovado=S 57	0.02	0.78	1.58				
		OR $\wedge$ Engenharia Florestal 91 $\Rightarrow$ Aprovado=N 66	0.02	0.73	1.43				
		OR $\wedge$ Farmácia 87 $\Rightarrow$ Aprovado=N 53	0.02	0.61	1.2				
2009/2	Diamantina (D)	D $\wedge$ Fisioterapia 27 $\Rightarrow$ Aprovado=S 23	0.01	0.85	1.31	Aprovado=S $\wedge$ BHU 149 $\Rightarrow$ D 102	0.05	0.68	3.29
		D $\wedge$ BCT - D 66 $\Rightarrow$ Aprovado=S 56	0.02	0.85	1.3	Aprovado=S $\wedge$ Educação Física 45 $\Rightarrow$ D 30	0.01	0.67	3.2
		D $\wedge$ BHU 121 $\Rightarrow$ Aprovado=S 102	0.05	0.84	1.29				
		D $\wedge$ Educação Física 37 $\Rightarrow$ Aprovado=S 30	0.01	0.81	1.24				
		D $\wedge$ Ciências Biológicas 58 $\Rightarrow$ Aprovado=S 45	0.02	0.78	1.19				
		D $\wedge$ Farmácia 39 $\Rightarrow$ Aprovado=N 23	0.01	0.59	1.69				
	Teófilo Otoni (TO)	TO $\wedge$ Ciências Econômicas 29 $\Rightarrow$ Aprovado=S 28	0.01	0.97	1.48	Aprovado=S $\wedge$ Administração 75 $\Rightarrow$ TO 50	0.02	0.67	4.29
		TO $\wedge$ BCT - TO 98 $\Rightarrow$ Aprovado=S 81	0.04	0.83	1.27	Aprovado=S $\wedge$ Serviço Social 49 $\Rightarrow$ TO 31	0.01	0.63	4.07
		TO $\wedge$ Ciências Contábeis 43 $\Rightarrow$ Aprovado=S 33	0.01	0.77	1.18	Aprovado=S $\wedge$ Ciências Econômicas 48 $\Rightarrow$ TO 28	0.01	0.58	3.76
			Aprovado=S $\wedge$ Ciências Contábeis 58 $\Rightarrow$ TO 33	0.01	0.57	3.66			
	Vale do Jequitinhonha (VJ)	VJ $\wedge$ BCT - D 38 $\Rightarrow$ Aprovado=S 30	0.01	0.79	1.21				
	Vale do Mucuri (VM)	VM $\wedge$ BCT - TO 41 $\Rightarrow$ Aprovado=S 28	0.01	0.68	1.05				
	Belo Horizonte (BH)	BH $\wedge$ BCT - TO 39 $\Rightarrow$ Aprovado=S 29	0.01	0.74	1.14				
BH $\wedge$ Farmácia 35 $\Rightarrow$ Aprovado=S 24		0.01	0.69	1.05					
BH $\wedge$ Odontologia 71 $\Rightarrow$ Aprovado=N 40		0.01	0.56	1.62					
Outra Região de MG (OR)	OR $\wedge$ BCT - TO 181 $\Rightarrow$ Aprovado=S 143	0.06	0.79	1.21	Aprovado=S Engenharia Florestal 75 $\Rightarrow$ OR 43	0.02	0.57	1.74	
	OR $\wedge$ Odontologia 157 $\Rightarrow$ Aprovado=N 109	0.05	0.69	1.99	Aprovado=S $\wedge$ Zootecnia 46 $\Rightarrow$ OR 25	0.01	0.54	1.65	
	OR $\wedge$ Zootecnia 36 $\Rightarrow$ Aprovado=S 25	0.01	0.69	1.07	Aprovado=S $\wedge$ Odontologia 90 $\Rightarrow$ OR 48	0.02	0.53	1.62	

**Tabela 26:** Regras de Associação envolvendo os atributos aprovado, atividade remunerada e curso

Ano/Sem	Antecedente da regra							
	Atividade remunerada (Ativ Rem) e curso				Aprovado e curso			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2003/1	Ativ Rem=Não $\wedge$ Farmácia-Bioq./Industrial <sup>23</sup> 191 $\Rightarrow$ Aprovado=S 161	0.05	0.84	1.09	Aprovado=S $\wedge$ Farmácia-Bioq./A. Clínica <sup>24</sup> 176 $\Rightarrow$ Ativ Rem=Não 150	0.05	0.85	1.08
	Ativ Rem=Não $\wedge$ Fisioterapia 435 $\Rightarrow$ Aprovado=S 352	0.12	0.81	1.04	Aprovado=S $\wedge$ Farmácia-Bioq./Industrial 191 $\Rightarrow$ Ativ Rem=Não 161	0.05	0.84	1.07
	Ativ Rem=Não $\wedge$ Farmácia-Bioq./A. Clínica 186 $\Rightarrow$ Aprovado=S 150	0.05	0.81	1.04	Aprovado=S $\wedge$ Fisioterapia 422 $\Rightarrow$ Ativ Rem=Não 352	0.12	0.83	1.06
	Ativ Rem=Não $\wedge$ Enfermagem 489 $\Rightarrow$ Aprovado=S 393	0.13	0.8	1.04	Aprovado=S $\wedge$ Odontologia 370 $\Rightarrow$ Ativ Rem=Não 307	0.1	0.83	1.05
	Ativ Rem=Não $\wedge$ Odontologia 384 $\Rightarrow$ Aprovado=S 307	0.1	0.8	1.03	Aprovado=S $\wedge$ Nutrição 284 $\Rightarrow$ Ativ Rem=Não 232	0.08	0.82	1.04
					Aprovado=S $\wedge$ Enfermagem 491 $\Rightarrow$ Ativ Rem=Não 393	0.13	0.8	1.02
2003/2	Ativ Rem=Não $\wedge$ Farmácia-Bioq./A. Clínica 254 $\Rightarrow$ Aprovado=N 203	0.1	0.8	1.12	Aprovado=S $\wedge$ Farmácia-Bioq./A. Clínica 53 $\Rightarrow$ Ativ Rem=Não 51	0.02	0.96	1.19
	Ativ Rem=Não $\wedge$ Fisioterapia 393 $\Rightarrow$ Aprovado=N 310	0.15	0.79	1.11	Aprovado=S $\wedge$ Odontologia 93 $\Rightarrow$ Ativ Rem=Não 86	0.04	0.92	1.14
	Ativ Rem=Não $\wedge$ Odontologia 358 $\Rightarrow$ Aprovado=N 272	0.13	0.76	1.07	Aprovado=S $\wedge$ Nutrição 77 $\Rightarrow$ Ativ Rem=Não 69	0.03	0.9	1.1
	Ativ Rem=Não $\wedge$ Farmácia-Bioq./Industrial 142 $\Rightarrow$ Aprovado=N 102	0.05	0.72	1.01	Aprovado=S $\wedge$ Fisioterapia 95 $\Rightarrow$ Ativ Rem=Não 83	0.04	0.87	1.08
	Ativ Rem=Não $\wedge$ Zootecnia 94 $\Rightarrow$ Aprovado=S 59	0.03	0.63	2.17	Aprovado=S $\wedge$ Farmácia-Bioq./Industrial 46 $\Rightarrow$ Ativ Rem=Não 40	0.02	0.87	1.07
	Ativ Rem=Não $\wedge$ Agronomia 101 $\Rightarrow$ Aprovado=S 59	0.03	0.58	2.02				
2004/1	Ativ Rem=Não $\wedge$ Enfermagem 411 $\Rightarrow$ Aprovado=N 361	0.11	0.88	1.07	Aprovado=S $\wedge$ Fisioterapia 74 $\Rightarrow$ Ativ Rem=Não 48	0.02	0.65	1.19
	Ativ Rem=Não $\wedge$ Odontologia 263 $\Rightarrow$ Aprovado=N 226	0.07	0.86	1.04	Aprovado=S $\wedge$ Nutrição 66 $\Rightarrow$ Ativ Rem=Não 42	0.01	0.64	1.17
	Ativ Rem=Não $\wedge$ Farmácia-Bioq./Industrial 147 $\Rightarrow$ Aprovado=N 123	0.04	0.84	1.01	Aprovado=S $\wedge$ Enfermagem 86 $\Rightarrow$ Ativ Rem=Não 50	0.02	0.58	1.07
	Ativ Rem=Não $\wedge$ Fisioterapia 288 $\Rightarrow$ Aprovado=N 240	0.08	0.83	1.01				
	Ativ Rem=Não $\wedge$ Farmácia-Bioq./A. Clínica 130 $\Rightarrow$ Aprovado=N 108	0.03	0.83	1.01				
2004/2	Ativ Rem=Não $\wedge$ Farmácia-Bioq./Industrial 109 $\Rightarrow$ Aprovado=N 89	0.05	0.82	1.2	Fisioterapia $\wedge$ Aprovado=S 94 $\Rightarrow$ Ativ Rem=Não 49	0.03	0.52	1.21
	Ativ Rem=Não $\wedge$ Odontologia 131 $\Rightarrow$ Aprovado=N 98	0.05	0.75	1.1				
	Ativ Rem=Não $\wedge$ Farmácia-Bioq./A. Clínica 74 $\Rightarrow$ Aprovado=N 53	0.03	0.72	1.05				
	Ativ Rem=Não $\wedge$ Fisioterapia 169 $\Rightarrow$ Aprovado=N 120	0.07	0.71	1.04				
	Ativ Rem=Não $\wedge$ Agronomia 51 $\Rightarrow$ Aprovado=S 28	0.02	0.55	1.72				
	Ativ Rem=Não $\wedge$ Zootecnia 59 $\Rightarrow$ Aprovado=S 31	0.02	0.53	1.65				

Continua na página seguinte...

<sup>23</sup>Farmácia-Bioquímica/Industrial

<sup>24</sup>Farmácia-Bioquímica/Análise Clínica

**Tabela 26:** Regras de Associação envolvendo os atributos aprovado, atividade remunerada e curso (cont)

Ano/Sem	Antecedente da regra							
	Atividade remunerada (Ativ Rem) e curso				Aprovado e curso			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2005/1	Ativ Rem=Não $\wedge$ Farmácia-Bioq./A. Clínica 233 $\Rightarrow$ Aprovado=N 202	0.07	0.87	1.06	Farmácia-Bioq./A. Clínica $\wedge$ Aprovado=S 32 $\Rightarrow$ Ativ Rem=Não 31	0.01	0.97	1.14
	Ativ Rem=Não $\wedge$ Enfermagem 448 $\Rightarrow$ Aprovado=N 382	0.13	0.85	1.04	Fisioterapia $\wedge$ Aprovado=S 65 $\Rightarrow$ Ativ Rem=Não 62	0.02	0.95	1.12
	Ativ Rem=Não $\wedge$ Odontologia 374 $\Rightarrow$ Aprovado=N 316	0.11	0.84	1.03	Odontologia $\wedge$ Aprovado=S 61 $\Rightarrow$ Ativ Rem=Não 58	0.02	0.95	1.11
	Ativ Rem=Não $\wedge$ Fisioterapia 361 $\Rightarrow$ Aprovado=N 299	0.1	0.83	1.01	Farmácia-Bioq./Industrial $\wedge$ Aprovado=S 35 $\Rightarrow$ Ativ Rem=Não 33	0.01	0.94	1.11
					Enfermagem $\wedge$ Aprovado=S 72 $\Rightarrow$ Ativ Rem=Não 66	0.02	0.92	1.07
					Nutrição $\wedge$ Aprovado=S 64 $\Rightarrow$ Ativ Rem=Não 57	0.02	0.89	1.04
					Zootecnia $\wedge$ Aprovado=S 59 $\Rightarrow$ Ativ Rem=Não 52	0.02	0.88	1.03
2005/2	Ativ Rem=Não $\wedge$ Farmácia-Bioq./A. Clínica 246 $\Rightarrow$ Aprovado=N 205	0.09	0.83	1.1	Farmácia-Bioq./A. Clínica $\wedge$ Aprovado=S 42 $\Rightarrow$ Ativ Rem=Não 41	0.02	0.98	1.13
	Ativ Rem=Não $\wedge$ Farmácia-Bioq./Industrial 202 $\Rightarrow$ Aprovado=N 161	0.07	0.8	1.06	Odontologia $\wedge$ Aprovado=S 81 $\Rightarrow$ Ativ Rem=Não 78	0.03	0.96	1.11
	Ativ Rem=Não $\wedge$ Odontologia 378 $\Rightarrow$ Aprovado=N 300	0.13	0.79	1.05	Nutrição $\wedge$ Aprovado=S 72 $\Rightarrow$ Ativ Rem=Não 69	0.03	0.96	1.11
					Farmácia-Bioq./Industrial $\wedge$ Aprovado=S 43 $\Rightarrow$ Ativ Rem=Não 41	0.02	0.95	1.1
					Engenharia Florestal $\wedge$ Aprovado=S 68 $\Rightarrow$ Ativ Rem=Não 64	0.03	0.94	1.09
					Fisioterapia $\wedge$ Aprovado=S 82 $\Rightarrow$ Ativ Rem=Não 75	0.03	0.91	1.05
					Agronomia $\wedge$ Aprovado=S 67 $\Rightarrow$ Ativ Rem=Não 60	0.03	0.9	1.03
				Zootecnia $\wedge$ Aprovado=S 74 $\Rightarrow$ Ativ Rem=Não 65	0.03	0.88	1.01	
2007/1	Ativ Rem=Não $\wedge$ Odontologia 431 $\Rightarrow$ Aprovado=N 363	0.11	0.84	1.31	Aprovado=S $\wedge$ Nutrição 66 $\Rightarrow$ Ativ Rem=Não 64	0.02	0.97	1.2
	Ativ Rem=Não $\wedge$ Enfermagem 366 $\Rightarrow$ Aprovado=N 299	0.09	0.82	1.27	Aprovado=S $\wedge$ Farmácia 72 $\Rightarrow$ Ativ Rem=Não 69	0.02	0.96	1.19
	Ativ Rem=Não $\wedge$ Ciências Econômicas 43 $\Rightarrow$ Aprovado=S 35	0.01	0.81	2.27	Aprovado=S $\wedge$ Odontologia 73 $\Rightarrow$ Ativ Rem=Não 68	0.02	0.93	1.15
	Ativ Rem=Não $\wedge$ Farmácia 324 $\Rightarrow$ Aprovado=N 255	0.08	0.79	1.23	Aprovado=S $\wedge$ Fisioterapia 73 $\Rightarrow$ Ativ Rem=Não 68	0.02	0.93	1.15
	Ativ Rem=Não $\wedge$ Ciências Contábeis 45 $\Rightarrow$ Aprovado=S 34	0.01	0.76	2.11	Aprovado=S $\wedge$ Engenharia Florestal 60 $\Rightarrow$ Ativ Rem=Não 55	0.02	0.92	1.14
	Ativ Rem=Não $\wedge$ Fisioterapia 195 $\Rightarrow$ Aprovado=N 127	0.04	0.65	1.02	Aprovado=S $\wedge$ Agronomia 60 $\Rightarrow$ Ativ Rem=Não 54	0.02	0.9	1.11
	Ativ Rem=Não $\wedge$ Educação Física 102 $\Rightarrow$ Aprovado=S 57	0.02	0.56	1.56	Aprovado=S $\wedge$ Enfermagem 76 $\Rightarrow$ Ativ Rem=Não 67	0.02	0.88	1.09
	Ativ Rem=Não $\wedge$ Serviço Social 100 $\Rightarrow$ Aprovado=S 52	0.02	0.52	1.45	Aprovado=S $\wedge$ Zootecnia 62 $\Rightarrow$ Ativ Rem=Não 54	0.02	0.87	1.08
	Ativ Rem=Não $\wedge$ Zootecnia 106 $\Rightarrow$ Aprovado=S 54	0.02	0.51	1.42	Aprovado=S $\wedge$ Ciências Biológicas 73 $\Rightarrow$ Ativ Rem=Não 61	0.02	0.84	1.04

Continua na página seguinte...

**Tabela 26:** Regras de Associação envolvendo os atributos aprovado, atividade remunerada e curso (cont)

Ano/Sem	Antecedente da regra							
	Atividade remunerada (Ativ Rem) e curso				Aprovado e curso			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2007/2	Ativ Rem=Não $\wedge$ Ciências Contábeis 54 $\Rightarrow$ Aprovado=S 41	0.02	0.76	1.59	Aprovado=S $\wedge$ Farmácia 90 $\Rightarrow$ Ativ Rem=Não 89	0.03	0.99	1.26
	Ativ Rem=Não $\wedge$ Educação Física 89 $\Rightarrow$ Aprovado=S 67	0.02	0.75	1.58	Aprovado=S $\wedge$ Odontologia 94 $\Rightarrow$ Ativ Rem=Não 92	0.03	0.98	1.25
	Ativ Rem=Não $\wedge$ Odontologia 353 $\Rightarrow$ Aprovado=N 261	0.1	0.74	1.41	Aprovado=S $\wedge$ Nutrição 76 $\Rightarrow$ Ativ Rem=Não 70	0.03	0.92	1.17
	Ativ Rem=Não $\wedge$ Farmácia 340 $\Rightarrow$ Aprovado=N 251	0.09	0.74	1.41	Aprovado=S $\wedge$ Fisioterapia 95 $\Rightarrow$ Ativ Rem=Não 87	0.03	0.92	1.17
	Ativ Rem=Não $\wedge$ Zootecnia 72 $\Rightarrow$ Aprovado=S 51	0.02	0.71	1.48	Aprovado=S $\wedge$ Engenharia Florestal 75 $\Rightarrow$ Ativ Rem=Não 68	0.03	0.91	1.15
	Ativ Rem=Não $\wedge$ Agronomia 111 $\Rightarrow$ Aprovado=S 68	0.03	0.61	1.28	Aprovado=S $\wedge$ Agronomia 78 $\Rightarrow$ Ativ Rem=Não 68	0.03	0.87	1.11
	Ativ Rem=Não $\wedge$ Ciências Biológicas 125 $\Rightarrow$ Aprovado=S 71	0.03	0.57	1.19	Aprovado=S $\wedge$ Zootecnia 62 $\Rightarrow$ Ativ Rem=Não 51	0.02	0.82	1.05
	Ativ Rem=Não $\wedge$ Engenharia Florestal 163 $\Rightarrow$ Aprovado=N 95	0.03	0.58	1.11				
	Ativ Rem=Não $\wedge$ Serviço Social 111 $\Rightarrow$ Aprovado=S 56	0.02	0.5	1.06				
2008/1	Ativ Rem=Não $\wedge$ Odontologia 427 $\Rightarrow$ Aprovado=N 357	0.1	0.84	1.27	Aprovado=S $\wedge$ Odontologia 72 $\Rightarrow$ Ativ Rem=Não 70	0.02	0.97	1.28
	Ativ Rem=Não $\wedge$ Farmácia 319 $\Rightarrow$ Aprovado=N 249	0.07	0.78	1.19	Aprovado=S $\wedge$ Farmácia 74 $\Rightarrow$ Ativ Rem=Não 70	0.02	0.95	1.25
	Ativ Rem=Não $\wedge$ Enfermagem 233 $\Rightarrow$ Aprovado=N 167	0.05	0.72	1.09	Aprovado=S $\wedge$ Nutrição 60 $\Rightarrow$ Ativ Rem=Não 56	0.02	0.93	1.23
	Ativ Rem=Não $\wedge$ Serviço Social 153 $\Rightarrow$ Aprovado=N 109	0.03	0.71	1.08	Aprovado=S $\wedge$ Enfermagem 72 $\Rightarrow$ Ativ Rem=Não 66	0.02	0.92	1.21
	Ativ Rem=Não $\wedge$ Engenharia Florestal 183 $\Rightarrow$ Aprovado=N 128	0.04	0.7	1.06	Aprovado=S $\wedge$ Fisioterapia 72 $\Rightarrow$ Ativ Rem=Não 66	0.02	0.92	1.21
	Ativ Rem=Não $\wedge$ Educação Física 92 $\Rightarrow$ Aprovado=S 58	0.02	0.63	1.84	Aprovado=S $\wedge$ Engenharia Florestal 60 $\Rightarrow$ Ativ Rem=Não 55	0.02	0.92	1.21
	Ativ Rem=Não $\wedge$ Ciências Econômicas 68 $\Rightarrow$ Aprovado=S 42	0.01	0.62	1.8	Aprovado=S $\wedge$ Agronomia 61 $\Rightarrow$ Ativ Rem=Não 55	0.02	0.9	1.19
	Ativ Rem=Não $\wedge$ Zootecnia 81 $\Rightarrow$ Aprovado=S 49	0.01	0.6	1.77	Aprovado=S $\wedge$ Ciências Biológicas 72 $\Rightarrow$ Ativ Rem=Não 60	0.02	0.83	1.1
	Ativ Rem=Não $\wedge$ Ciências Contábeis 70 $\Rightarrow$ Aprovado=S 39	0.01	0.56	1.63	Aprovado=S $\wedge$ Educação Física 72 $\Rightarrow$ Ativ Rem=Não 58	0.02	0.81	1.06
	Ativ Rem=Não $\wedge$ Agronomia 107 $\Rightarrow$ Aprovado=S 55	0.02	0.51	1.5	Aprovado=S $\wedge$ Zootecnia 63 $\Rightarrow$ Ativ Rem=Não 49	0.01	0.78	1.03
2008/2	Ativ Rem=Tempo integral $\wedge$ Ciências Contábeis 34 $\Rightarrow$ Aprovado=S 29	0.01	0.85	1.53	Aprovado=S $\wedge$ Odontologia 91 $\Rightarrow$ Ativ Rem=Não 85	0.04	0.93	1.23
	Ativ Rem=Não $\wedge$ Educação Física 66 $\Rightarrow$ Aprovado=S 56	0.03	0.85	1.52	Aprovado=S $\wedge$ Farmácia 90 $\Rightarrow$ Ativ Rem=Não 84	0.04	0.93	1.23
	Ativ Rem=Não $\wedge$ Zootecnia 41 $\Rightarrow$ Aprovado=S 33	0.02	0.8	1.44	Aprovado=S $\wedge$ Engenharia Florestal 75 $\Rightarrow$ Ativ Rem=Não 67	0.03	0.89	1.18
	Ativ Rem=Não $\wedge$ Ciências Econômicas 34 $\Rightarrow$ Aprovado=S 28	0.01	0.82	1.47	Aprovado=S $\wedge$ Nutrição 75 $\Rightarrow$ Ativ Rem=Não 66	0.03	0.88	1.16
	Ativ Rem=Não $\wedge$ Ciências Biológicas 106 $\Rightarrow$ Aprovado=S 77	0.04	0.73	1.3	Aprovado=S $\wedge$ Ciências Biológicas 90 $\Rightarrow$ Ativ Rem=Não 77	0.04	0.86	1.13
	Ativ Rem=Não $\wedge$ Agronomia 83 $\Rightarrow$ Aprovado=S 60	0.03	0.72	1.29	Aprovado=S $\wedge$ Zootecnia 40 $\Rightarrow$ Ativ Rem=Não 33	0.02	0.83	1.09

Continua na página seguinte...

**Tabela 26:** Regras de Associação envolvendo os atributos aprovado, atividade remunerada e curso (cont)

Ano/Sem	Antecedente da regra							
	Atividade remunerada (Ativ Rem) e curso				Aprovado e curso			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
(2008/2) (cont)	Ativ Rem=Não $\wedge$ Ciências Contábeis 41 $\Rightarrow$ Aprovado=S 29	0.01	0.71	1.27	Aprovado=S $\wedge$ Fisioterapia 75 $\Rightarrow$ Ativ Rem=Não 61	0.03	0.81	1.07
	Ativ Rem=Não $\wedge$ Odontologia 282 $\Rightarrow$ Aprovado=N 197	0.1	0.7	1.58	Aprovado=S $\wedge$ Agronomia 75 $\Rightarrow$ Ativ Rem=Não 60	0.03	0.8	1.05
	Ativ Rem=Não $\wedge$ Fisioterapia 88 $\Rightarrow$ Aprovado=S 61	0.03	0.69	1.24	Aprovado=S $\wedge$ Educação Física 72 $\Rightarrow$ Ativ Rem=Não 56	0.03	0.78	1.02
	Ativ Rem=Não $\wedge$ Farmácia 229 $\Rightarrow$ Aprovado=N 145	0.07	0.63	1.43				
	Ativ Rem=Não $\wedge$ Engenharia Florestal 153 $\Rightarrow$ Aprovado=N 86	0.04	0.56	1.27				
	Ativ Rem=Não $\wedge$ Serviço Social 64 $\Rightarrow$ Aprovado=N 35	0.02	0.55	1.24				
2009/1	Ativ Rem=Tempo integral $\wedge$ BHU <sup>25</sup> 51 $\Rightarrow$ Aprovado=S 44	0.01	0.86	1.75	Aprovado=S $\wedge$ Farmácia 72 $\Rightarrow$ Ativ Rem=Não 69	0.02	0.96	1.34
	Ativ Rem=Não $\wedge$ BHU 91 $\Rightarrow$ Aprovado=S 78	0.02	0.86	1.74	Aprovado=S $\wedge$ Odontologia 73 $\Rightarrow$ Ativ Rem=Não 68	0.02	0.93	1.31
	Ativ Rem=Não $\wedge$ BCT - D <sup>26</sup> 213 $\Rightarrow$ Aprovado=S 181	0.05	0.85	1.72	Aprovado=S $\wedge$ Enfermagem 72 $\Rightarrow$ Ativ Rem=Não 65	0.02	0.9	1.27
	Ativ Rem=Não $\wedge$ Odontologia 353 $\Rightarrow$ Aprovado=N 285	0.08	0.81	1.6	Aprovado=S $\wedge$ Nutrição 60 $\Rightarrow$ Ativ Rem=Não 54	0.02	0.9	1.26
	Ativ Rem=Não $\wedge$ Educação Física 58 $\Rightarrow$ Aprovado=S 40	0.01	0.69	1.4	Aprovado=S $\wedge$ Engenharia Florestal 60 $\Rightarrow$ Ativ Rem=Não 53	0.02	0.88	1.24
	Ativ Rem=Não $\wedge$ Engenharia Florestal 169 $\Rightarrow$ Aprovado=N 116	0.03	0.69	1.36	Aprovado=S $\wedge$ Fisioterapia 68 $\Rightarrow$ Ativ Rem=Não 58	0.02	0.85	1.2
	Ativ Rem=Não $\wedge$ BCT - TO <sup>27</sup> 284 $\Rightarrow$ Aprovado=S 189	0.05	0.67	1.35	Aprovado=S $\wedge$ Agronomia 61 $\Rightarrow$ Ativ Rem=Não 52	0.01	0.85	1.2
	Ativ Rem=Não $\wedge$ Fisioterapia 88 $\Rightarrow$ Aprovado=S 58	0.02	0.66	1.33	Aprovado=S $\wedge$ Zootecnia 44 $\Rightarrow$ Ativ Rem=Não 37	0.01	0.84	1.18
	Ativ Rem=Não $\wedge$ Serviço Social 137 $\Rightarrow$ Aprovado=N 91	0.03	0.66	1.31	Aprovado=S $\wedge$ BCT - D 228 $\Rightarrow$ Ativ Rem=Não 181	0.05	0.79	1.11
	Ativ Rem=Tempo integral $\wedge$ BCT - TO 77 $\Rightarrow$ Aprovado=S 49	0.01	0.64	1.29	Aprovado=S $\wedge$ Ciências Biológicas 72 $\Rightarrow$ Ativ Rem=Não 55	0.02	0.76	1.07
	Ativ Rem=Não $\wedge$ Ciências Biológicas 89 $\Rightarrow$ Aprovado=S 55	0.02	0.62	1.25				
	Ativ Rem=Não $\wedge$ Farmácia 183 $\Rightarrow$ Aprovado=N 114	0.03	0.62	1.23				
	Ativ Rem=Não $\wedge$ Zootecnia 63 $\Rightarrow$ Aprovado=S 37	0.01	0.59	1.19				
	Ativ Rem=Não $\wedge$ Ciências Contábeis 65 $\Rightarrow$ Aprovado=S 37	0.01	0.57	1.15				
	Ativ Rem=Não $\wedge$ Enfermagem 152 $\Rightarrow$ Aprovado=N 87	0.02	0.57	1.13				
Ativ Rem=Não $\wedge$ Agronomia 99 $\Rightarrow$ Aprovado=S 52	0.01	0.53	1.06					

Continua na página seguinte...

<sup>25</sup>Bacharelado em Humanidades

<sup>26</sup>Bacharelado em Ciência e Tecnologia - Diamantina

<sup>27</sup>Bacharelado em Ciência e Tecnologia - Teófilo Otoni

**Tabela 26:** Regras de Associação envolvendo os atributos aprovado, atividade remunerada e curso (cont)

Antecedente da regra								
Atividade remunerada (Ativ Rem) e curso					Aprovado e curso			
Ano/Sem	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2009/2	Ativ Rem=Não $\wedge$ Ciências Econômicas 31 $\Rightarrow$ Aprovado=S 26	0.01	0.84	1.29	Aprovado=S $\wedge$ Odontologia 90 $\Rightarrow$ Ativ Rem=Não 85	0.04	0.94	1.3
	Ativ Rem=Não $\wedge$ BHU 100 $\Rightarrow$ Aprovado=S 81	0.04	0.81	1.24	Aprovado=S $\wedge$ Farmácia 92 $\Rightarrow$ Ativ Rem=Não 86	0.04	0.93	1.29
	Ativ Rem=Tempo integral $\wedge$ Ciências Contábeis 36 $\Rightarrow$ Aprovado=S 29	0.01	0.81	1.24	Aprovado=S $\wedge$ Nutrição 50 $\Rightarrow$ Ativ Rem=Não 45	0.02	0.9	1.24
	Ativ Rem=Tempo integral $\wedge$ BHU 52 $\Rightarrow$ Aprovado=S 41	0.02	0.79	1.21	Aprovado=S $\wedge$ Agronomia 51 $\Rightarrow$ Ativ Rem=Não 45	0.02	0.88	1.22
	Ativ Rem=Não $\wedge$ BCT - TO 410 $\Rightarrow$ Aprovado=S 321	0.14	0.78	1.2	Aprovado=S $\wedge$ Fisioterapia 66 $\Rightarrow$ Ativ Rem=Não 58	0.03	0.88	1.21
	Ativ Rem=Tempo integral $\wedge$ BCT - TO 54 $\Rightarrow$ Aprovado=S 42	0.02	0.78	1.19	Aprovado=S $\wedge$ BCT - D 226 $\Rightarrow$ Ativ Rem=Não 192	0.09	0.85	1.17
	Ativ Rem=Não $\wedge$ Fisioterapia 77 $\Rightarrow$ Aprovado=S 58	0.03	0.75	1.16	Aprovado=S $\wedge$ Engenharia Florestal 75 $\Rightarrow$ Ativ Rem=Não 59	0.03	0.79	1.08
	Ativ Rem=Não $\wedge$ Ciências Biológicas 82 $\Rightarrow$ Aprovado=S 61	0.03	0.74	1.14	Aprovado=S $\wedge$ Zootecnia 46 $\Rightarrow$ Ativ Rem=Não 35	0.02	0.76	1.05
	Ativ Rem=Não $\wedge$ Educação Física 35 $\Rightarrow$ Aprovado=S 26	0.01	0.74	1.14	Aprovado=S $\wedge$ Ciências Biológicas 81 $\Rightarrow$ Ativ Rem=Não 61	0.03	0.75	1.04
	Ativ Rem=Não $\wedge$ Zootecnia 48 $\Rightarrow$ Aprovado=S 35	0.02	0.73	1.12				
	Ativ Rem=Não $\wedge$ Agronomia 63 $\Rightarrow$ Aprovado=S 45	0.02	0.71	1.1				
	Ativ Rem=Não $\wedge$ Odontologia 263 $\Rightarrow$ Aprovado=N 178	0.08	0.68	1.94				

**Tabela 27:** Regras de Associação envolvendo os atributos aprovado, frequência pré-vestibular e curso

Ano/Sem	Regras em que há alguma frequência em pré-vestibular				Regras em que não há frequência em pré-vestibular			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2003/1	Mais de 1 ano $\wedge$ Odontologia 73 $\Rightarrow$ Aprovado=S 61	0.02	0.84	1.08				
	1 ano $\wedge$ Odontologia 113 $\Rightarrow$ Aprovado=S 91	0.03	0.81	1.04				
	1 sem $\wedge$ Odontologia 64 $\Rightarrow$ Aprovado=S 51	0.02	0.8	1.03				
	Mais de 1 ano $\wedge$ Enfermagem 106 $\Rightarrow$ Aprovado=S 96	0.03	0.91	1.17				
	1 ano $\wedge$ Enfermagem 158 $\Rightarrow$ Aprovado=S 129	0.04	0.82	1.05				
	1 sem $\wedge$ Enfermagem 131 $\Rightarrow$ Aprovado=S 103	0.03	0.79	1.01				
	1 ano $\wedge$ Fisioterapia 137 $\Rightarrow$ Aprovado=S 111	0.04	0.81	1.04				
	Menos de 1 sem $\wedge$ Fisioterapia 51 $\Rightarrow$ Aprovado=S 41	0.01	0.8	1.04				
	1 sem $\wedge$ Farmácia-Bioq./Industrial <sup>28</sup> 37 $\Rightarrow$ Aprovado=S 33	0.01	0.89	1.15				
	1 ano $\wedge$ Farmácia-Bioq./Industrial 65 $\Rightarrow$ Aprovado=S 55	0.02	0.85	1.09				
	1 ano $\wedge$ Farmácia-Bioq./A. Clínica <sup>29</sup> 60 $\Rightarrow$ Aprovado=S 53	0.02	0.88	1.14				
1 sem $\wedge$ Nutrição 70 $\Rightarrow$ Aprovado=S 55	0.02	0.79	1.01					
2003/2	Menos de 1 sem $\wedge$ Odontologia 109 $\Rightarrow$ Aprovado=N 90	0.04	0.83	1.16	Não $\wedge$ Odontologia 32 $\Rightarrow$ Aprovado=N 29	0.01	0.91	1.27
	1 sem $\wedge$ Odontologia 107 $\Rightarrow$ Aprovado=N 88	0.04	0.82	1.16	Não $\wedge$ Fisioterapia 46 $\Rightarrow$ Aprovado=N 44	0.02	0.96	1.34
	Menos de 1 sem $\wedge$ Fisioterapia 115 $\Rightarrow$ Aprovado=N 99	0.05	0.86	1.21	Não $\wedge$ Farmácia-Bioq./A. Clínica 31 $\Rightarrow$ Aprovado=N 25	0.01	0.81	1.13
	1 sem $\wedge$ Fisioterapia 151 $\Rightarrow$ Aprovado=N 117	0.06	0.77	1.09	Não $\wedge$ Nutrição 35 $\Rightarrow$ Aprovado=N 33	0.02	0.94	1.33
	1 ano $\wedge$ Fisioterapia 73 $\Rightarrow$ Aprovado=N 55	0.03	0.75	1.06				
	1 ano $\wedge$ Farmácia-Bioq./Industrial 38 $\Rightarrow$ Aprovado=N 32	0.02	0.84	1.18				
	Menos de 1 sem $\wedge$ Farmácia-Bioq./Industrial 45 $\Rightarrow$ Aprovado=N 35	0.02	0.78	1.09				
	1 sem $\wedge$ Farmácia-Bioq./Industrial 44 $\Rightarrow$ Aprovado=N 32	0.02	0.73	1.02				
	1 sem $\wedge$ Farmácia-Bioq./A. Clínica 83 $\Rightarrow$ Aprovado=N 70	0.03	0.84	1.19				
	Menos de 1 sem $\wedge$ Farmácia-Bioq./A. Clínica 83 $\Rightarrow$ Aprovado=N 69	0.03	0.83	1.17				
	1 ano $\wedge$ Farmácia-Bioq./A. Clínica 38 $\Rightarrow$ Aprovado=N 31	0.01	0.82	1.15				
	Menos de 1 sem $\wedge$ Nutrição 73 $\Rightarrow$ Aprovado=N 55	0.03	0.75	1.06				
	1 sem $\wedge$ Nutrição 89 $\Rightarrow$ Aprovado=N 66	0.03	0.74	1.04				

Continua na página seguinte...

<sup>28</sup>Farmácia-Bioquímica/Industrial

<sup>29</sup>Farmácia-Bioquímica/Análise Clínica

**Tabela 27:** Regras de Associação envolvendo os atributos aprovado, frequência pré-vestibular e curso (cont)

Ano/Sem	Regras em que há alguma frequência em pré-vestibular				Regras em que não há frequência em pré-vestibular			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2003/2 (cont)	Menos de 1 sem $\wedge$ Agronomia 32 $\Rightarrow$ Aprovado=S 21	0.01	0.66	2.27				
	1 sem $\wedge$ Agronomia 40 $\Rightarrow$ Aprovado=S 21	0.01	0.53	1.82				
	Menos de 1 sem $\wedge$ Engenharia Florestal 43 $\Rightarrow$ Aprovado=S 27	0.01	0.63	2.17				
	1 sem $\wedge$ Engenharia Florestal 43 $\Rightarrow$ Aprovado=S 27	0.01	0.63	2.17				
	Menos de 1 sem $\wedge$ Zootecnia 37 $\Rightarrow$ Aprovado=S 24	0.01	0.65	2.25				
2004/1	1 sem $\wedge$ Odontologia 57 $\Rightarrow$ Aprovado=N 52	0.02	0.91	1.11	Não $\wedge$ Odontologia 59 $\Rightarrow$ Aprovado=N 53	0.02	0.9	1.09
	Mais de 1 ano $\wedge$ Odontologia 61 $\Rightarrow$ Aprovado=N 54	0.02	0.89	1.07	Não $\wedge$ Enfermagem 107 $\Rightarrow$ Aprovado=N 92	0.03	0.86	1.04
	1 ano $\wedge$ Odontologia 72 $\Rightarrow$ Aprovado=N 61	0.02	0.85	1.03	Não $\wedge$ Fisioterapia 76 $\Rightarrow$ Aprovado=N 65	0.02	0.86	1.04
	1 ano $\wedge$ Enfermagem 104 $\Rightarrow$ Aprovado=N 94	0.03	0.9	1.1	Não $\wedge$ Farmácia-Bioq./A. Clínica 39 $\Rightarrow$ Aprovado=N 33	0.01	0.85	1.03
	1 sem $\wedge$ Enfermagem 90 $\Rightarrow$ Aprovado=N 78	0.02	0.87	1.05	Não $\wedge$ Nutrição 61 $\Rightarrow$ Aprovado=N 52	0.02	0.85	1.03
	Mais de 1 ano $\wedge$ Enfermagem 98 $\Rightarrow$ Aprovado=N 84	0.03	0.86	1.04				
	Menos de 1 sem $\wedge$ Fisioterapia 45 $\Rightarrow$ Aprovado=N 41	0.01	0.91	1.1				
	1 ano $\wedge$ Fisioterapia 68 $\Rightarrow$ Aprovado=N 61	0.02	0.9	1.09				
	1 sem $\wedge$ Fisioterapia 80 $\Rightarrow$ Aprovado=N 67	0.02	0.84	1.02				
	1 ano $\wedge$ Farmácia-Bioq./Industrial 38 $\Rightarrow$ Aprovado=N 33	0.01	0.87	1.05				
	1 sem $\wedge$ Farmácia-Bioq./Industrial 38 $\Rightarrow$ Aprovado=N 32	0.01	0.84	1.02				
	1 sem $\wedge$ Nutrição 53 $\Rightarrow$ Aprovado=N 45	0.01	0.85	1.03				
2004/2	1 sem $\wedge$ Odontologia 43 $\Rightarrow$ Aprovado=N 37	0.02	0.86	1.26				
	1 ano $\wedge$ Odontologia 23 $\Rightarrow$ Aprovado=N 19	0.01	0.83	1.21				
	Mais de 1 ano $\wedge$ Odontologia 32 $\Rightarrow$ Aprovado=N 24	0.01	0.75	1.1				
	Menos de 1 sem $\wedge$ Odontologia 38 $\Rightarrow$ Aprovado=N 27	0.01	0.71	1.04				
	Menos de 1 sem $\wedge$ Fisioterapia 47 $\Rightarrow$ Aprovado=N 37	0.02	0.79	1.16				
	1 sem $\wedge$ Fisioterapia 52 $\Rightarrow$ Aprovado=N 38	0.02	0.73	1.07				
	1 sem $\wedge$ Farmácia-Bioq./Industrial 48 $\Rightarrow$ Aprovado=N 38	0.02	0.79	1.16				
	Menos de 1 sem $\wedge$ Farmácia-Bioq./A. Clínica 25 $\Rightarrow$ Aprovado=N 19	0.01	0.76	1.12				
	Menos de 1 sem $\wedge$ Nutrição 33 $\Rightarrow$ Aprovado=N 28	0.02	0.85	1.25				
	1 sem $\wedge$ Nutrição 43 $\Rightarrow$ Aprovado=N 32	0.02	0.74	1.09				

Continua na página seguinte...



**Tabela 27:** Regras de Associação envolvendo os atributos aprovado, frequência pré-vestibular e curso (cont)

Ano/Sem	Regras em que há alguma frequência em pré-vestibular				Regras em que não há frequência em pré-vestibular			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2005/1	1 sem $\wedge$ Odontologia 107 $\Rightarrow$ Aprovado=N 93	0.03	0.87	1.06	Não $\wedge$ Odontologia 119 $\Rightarrow$ Aprovado=N 111	0.04	0.93	1.14
	1 ano $\wedge$ Odontologia 108 $\Rightarrow$ Aprovado=N 90	0.03	0.83	1.02	Não $\wedge$ Enfermagem 133 $\Rightarrow$ Aprovado=N 126	0.04	0.95	1.16
	1 sem $\wedge$ Enfermagem 146 $\Rightarrow$ Aprovado=N 138	0.05	0.95	1.16	Não $\wedge$ Fisioterapia 110 $\Rightarrow$ Aprovado=N 98	0.03	0.89	1.09
	1 ano $\wedge$ Enfermagem 154 $\Rightarrow$ Aprovado=N 127	0.04	0.82	1.01	Não $\wedge$ Farmácia-Bioq./A. Clínica 80 $\Rightarrow$ Aprovado=N 70	0.02	0.88	1.07
	1 sem $\wedge$ Fisioterapia 113 $\Rightarrow$ Aprovado=N 97	0.03	0.86	1.05				
	1 sem $\wedge$ Farmácia-Bioq./Industrial 53 $\Rightarrow$ Aprovado=N 50	0.02	0.94	1.15				
	1 sem $\wedge$ Farmácia-Bioq./A. Clínica 74 $\Rightarrow$ Aprovado=N 67	0.02	0.91	1.11				
	1 ano $\wedge$ Farmácia-Bioq./A. Clínica 61 $\Rightarrow$ Aprovado=N 54	0.02	0.89	1.08				
1 sem $\wedge$ Nutrição 94 $\Rightarrow$ Aprovado=N 81	0.03	0.86	1.05					
2005/2	1 sem $\wedge$ Odontologia 187 $\Rightarrow$ Aprovado=N 154	0.06	0.82	1.09	Não $\wedge$ Odontologia 55 $\Rightarrow$ Aprovado=N 50	0.02	0.91	1.2
	1 ano $\wedge$ Odontologia 76 $\Rightarrow$ Aprovado=N 60	0.03	0.79	1.05	Não $\wedge$ Fisioterapia 41 $\Rightarrow$ Aprovado=N 36	0.02	0.88	1.16
	1 ano $\wedge$ Fisioterapia 67 $\Rightarrow$ Aprovado=N 52	0.02	0.78	1.03	Não $\wedge$ Farmácia-Bioq./A. Clínica 34 $\Rightarrow$ Aprovado=N 32	0.01	0.94	1.25
	1 ano $\wedge$ Farmácia-Bioq./Industrial 32 $\Rightarrow$ Aprovado=N 28	0.01	0.88	1.16	Não $\wedge$ Nutrição 30 $\Rightarrow$ Aprovado=N 27	0.01	0.9	1.19
	1 sem $\wedge$ Farmácia-Bioq./Industrial 125 $\Rightarrow$ Aprovado=N 100	0.04	0.8	1.06				
	1 sem $\wedge$ Farmácia-Bioq./A. Clínica 138 $\Rightarrow$ Aprovado=N 120	0.05	0.87	1.15				
	1 ano $\wedge$ Farmácia-Bioq./A. Clínica 54 $\Rightarrow$ Aprovado=N 46	0.02	0.85	1.13				
1 sem $\wedge$ Nutrição 169 $\Rightarrow$ Aprovado=N 140	0.06	0.83	1.1					
2007/1	1 sem $\wedge$ Odontologia 101 $\Rightarrow$ Aprovado=N 91	0.03	0.9	1.4	Não $\wedge$ Odontologia 167 $\Rightarrow$ Aprovado=N 149	0.05	0.89	1.39
	1 ano $\wedge$ Odontologia 134 $\Rightarrow$ Aprovado=N 107	0.03	0.8	1.24	Não $\wedge$ Enfermagem 147 $\Rightarrow$ Aprovado=N 136	0.04	0.93	1.44
	Mais de 1 ano $\wedge$ Odontologia 63 $\Rightarrow$ Aprovado=N 45	0.01	0.71	1.11	Não $\wedge$ Fisioterapia 75 $\Rightarrow$ Aprovado=N 56	0.02	0.75	1.16
	1 sem $\wedge$ Enfermagem 106 $\Rightarrow$ Aprovado=N 89	0.03	0.84	1.31	Não $\wedge$ Farmácia 131 $\Rightarrow$ Aprovado=N 120	0.04	0.92	1.43
	1 ano $\wedge$ Enfermagem 86 $\Rightarrow$ Aprovado=N 61	0.02	0.71	1.11	Não $\wedge$ Ciências Biológicas 51 $\Rightarrow$ Aprovado=N 36	0.01	0.71	1.1
	Mais de 1 ano $\wedge$ Enfermagem 71 $\Rightarrow$ Aprovado=N 48	0.01	0.68	1.05	Não $\wedge$ Ciências Econômicas 47 $\Rightarrow$ Aprovado=S 39	0.01	0.83	2.32
	1 sem $\wedge$ Fisioterapia 66 $\Rightarrow$ Aprovado=N 49	0.01	0.74	1.16	Aprovado=S $\wedge$ Administração 75 $\Rightarrow$ Não 49	0.01	0.65	1.66
	1 sem $\wedge$ Farmácia 90 $\Rightarrow$ Aprovado=N 71	0.02	0.79	1.23	Aprovado=S $\wedge$ Serviço Social 74 $\Rightarrow$ Não 41	0.01	0.55	1.41
	1 ano $\wedge$ Farmácia 86 $\Rightarrow$ Aprovado=N 58	0.02	0.67	1.05	Administração 166 $\Rightarrow$ Não 108	0.03	0.65	1.65
				Ciências Econômicas 73 $\Rightarrow$ Não 47	0.01	0.64	1.63	

Continua na página seguinte...

**Tabela 27:** Regras de Associação envolvendo os atributos aprovado, frequência pré-vestibular e curso (cont)

Ano/Sem	Regras em que há alguma frequência em pré-vestibular				Regras em que não há frequência em pré-vestibular			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
<b>2007/1</b> <b>(cont)</b>					Ciências Contábeis 82 ⇒ Não 49	0.01	0.6	1.52
					Serviço Social 163 ⇒ Não 91	0.03	0.56	1.42
<b>2007/2</b>	1 sem ∧ Odontologia 173 ⇒ Aprovado=N 136	0.05	0.79	1.5	Não ∧ Odontologia 50 ⇒ Aprovado=N 47	0.02	0.94	1.8
	1 ano ∧ Odontologia 88 ⇒ Aprovado=N 63	0.02	0.72	1.37	Não ∧ Farmácia 73 ⇒ Aprovado=N 63	0.02	0.86	1.65
	Mais de 1 ano ∧ Odontologia 78 ⇒ Aprovado=N 49	0.02	0.63	1.2	Não ∧ Ciências Contábeis 67 ⇒ Aprovado=S 45	0.02	0.67	1.41
	1 sem ∧ Fisioterapia 106 ⇒ Aprovado=N 60	0.02	0.57	1.08	Aprovado=S ∧ Ciências Contábeis 74 ⇒ Não 45	0.02	0.61	2.18
	1 sem ∧ Farmácia 167 ⇒ Aprovado=N 136	0.05	0.81	1.56	Aprovado=S ∧ Administração 90 ⇒ Não 53	0.02	0.59	2.11
	1 ano ∧ Farmácia 68 ⇒ Aprovado=N 50	0.02	0.74	1.41	Aprovado=S ∧ Serviço Social 90 ⇒ Não 52	0.02	0.58	2.07
	1 sem ∧ Nutrição 97 ⇒ Aprovado=N 56	0.02	0.58	1.1	Serviço Social 190 ⇒ Não 124	0.05	0.65	2.33
	1 sem ∧ Ciências Biológicas 77 ⇒ Aprovado=S 42	0.02	0.55	1.14	Ciências Contábeis 104 ⇒ Não 67	0.02	0.64	2.3
	1 sem ∧ Agronomia 73 ⇒ Aprovado=S 43	0.02	0.59	1.23	Administração 177 ⇒ Não 111	0.04	0.63	2.24
	1 sem ∧ Engenharia Florestal 95 ⇒ Aprovado=N 58	0.02	0.61	1.17				
1 sem ∧ Zootecnia 51 ⇒ Aprovado=S 36	0.01	0.71	1.48					
<b>2008/1</b>	1 sem ∧ Odontologia 113 ⇒ Aprovado=N 103	0.03	0.91	1.39	Não ∧ Odontologia 181 ⇒ Aprovado=N 167	0.05	0.92	1.4
	1 ano ∧ Odontologia 118 ⇒ Aprovado=N 89	0.03	0.75	1.15	Não ∧ Enfermagem 96 ⇒ Aprovado=N 79	0.02	0.82	1.25
	Mais de 1 ano ∧ Odontologia 62 ⇒ Aprovado=N 43	0.01	0.69	1.05	Não ∧ Farmácia 131 ⇒ Aprovado=N 112	0.03	0.85	1.3
	1 sem ∧ Enfermagem 76 ⇒ Aprovado=N 57	0.02	0.75	1.14	Não ∧ Nutrição 80 ⇒ Aprovado=N 64	0.02	0.8	1.22
	1 ano ∧ Enfermagem 71 ⇒ Aprovado=N 48	0.01	0.68	1.03	Não ∧ Ciências Biológicas 57 ⇒ Aprovado=N 40	0.01	0.7	1.07
	1 sem ∧ Farmácia 93 ⇒ Aprovado=N 81	0.02	0.87	1.32	Não ∧ Ciências Econômicas 75 ⇒ Aprovado=S 46	0.01	0.61	1.79
	1 sem ∧ Nutrição 64 ⇒ Aprovado=N 48	0.01	0.75	1.14	Não ∧ Ciências Contábeis 95 ⇒ Aprovado=S 48	0.01	0.51	1.48
	1 sem ∧ Engenharia Florestal 56 ⇒ Aprovado=N 43	0.01	0.77	1.17	Aprovado=S ∧ Ciências Contábeis 73 ⇒ Não 48	0.01	0.66	1.42
					Aprovado=S ∧ Serviço Social 72 ⇒ Não 42	0.01	0.58	1.26
					Aprovado=S ∧ Administração 73 ⇒ Não 38	0.01	0.52	1.12
				Ciências Contábeis 138 ⇒ Não 95	0.03	0.69	1.49	
				Administração 281 ⇒ Não 191	0.06	0.68	1.47	
				Serviço Social 288 ⇒ Não 195	0.06	0.68	1.46	
				Ciências Econômicas 114 ⇒ Não 75	0.02	0.66	1.42	

Continua na página seguinte...

**Tabela 27:** Regras de Associação envolvendo os atributos aprovado, frequência pré-vestibular e curso (cont)

Ano/Sem	Regras em que há alguma frequência em pré-vestibular				Regras em que não há frequência em pré-vestibular			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2008/2	1 sem $\wedge$ Odontologia 121 $\Rightarrow$ Aprovado=N 84	0.04	0.69	1.57	Não $\wedge$ Odontologia 55 $\Rightarrow$ Aprovado=N 47	0.02	0.85	1.93
	1 ano $\wedge$ Odontologia 56 $\Rightarrow$ Aprovado=N 38	0.02	0.68	1.54	Não $\wedge$ Farmácia 41 $\Rightarrow$ Aprovado=N 32	0.02	0.78	1.77
	Mais de 1 ano $\wedge$ Odontologia 75 $\Rightarrow$ Aprovado=N 48	0.02	0.64	1.45	Não $\wedge$ Ciências Econômicas 32 $\Rightarrow$ Aprovado=S 25	0.01	0.78	1.4
	1 sem $\wedge$ Fisioterapia 48 $\Rightarrow$ Aprovado=S 36	0.02	0.75	1.34	Não $\wedge$ Ciências Contábeis 48 $\Rightarrow$ Aprovado=S 35	0.02	0.73	1.31
	1 ano $\wedge$ Farmácia 46 $\Rightarrow$ Aprovado=N 32	0.02	0.7	1.58	Aprovado=S $\wedge$ Administração 76 $\Rightarrow$ Não 46	0.02	0.61	1.95
	1 sem $\wedge$ Farmácia 114 $\Rightarrow$ Aprovado=N 72	0.04	0.63	1.43	Aprovado=S $\wedge$ Serviço Social 65 $\Rightarrow$ Não 39	0.02	0.6	1.93
	1 sem $\wedge$ Nutrição 60 $\Rightarrow$ Aprovado=S 34	0.02	0.57	1.01	Aprovado=S $\wedge$ Ciências Contábeis 67 $\Rightarrow$ Não 35	0.02	0.52	1.68
	1 sem $\wedge$ Ciências Biológicas 42 $\Rightarrow$ Aprovado=S 30	0.01	0.71	1.28	Administração 129 $\Rightarrow$ Não 84	0.04	0.65	2.1
	1 sem $\wedge$ Agronomia 47 $\Rightarrow$ Aprovado=S 36	0.02	0.77	1.37	Serviço Social 134 $\Rightarrow$ Não 86	0.04	0.64	2.07
	1 sem $\wedge$ Engenharia Florestal 92 $\Rightarrow$ Aprovado=N 54	0.03	0.59	1.33	Ciências Contábeis 84 $\Rightarrow$ Não 48	0.02	0.57	1.84
1 sem $\wedge$ Zootecnia 32 $\Rightarrow$ Aprovado=S 28	0.01	0.88	1.57	Ciências Econômicas 61 $\Rightarrow$ Não 32	0.02	0.52	1.69	
2009/1	1 sem $\wedge$ Odontologia 92 $\Rightarrow$ Aprovado=N 75	0.02	0.82	1.61	Não $\wedge$ Odontologia 161 $\Rightarrow$ Aprovado=N 136	0.04	0.84	1.67
	1 ano $\wedge$ Odontologia 97 $\Rightarrow$ Aprovado=N 77	0.02	0.79	1.57	Não $\wedge$ Enfermagem 80 $\Rightarrow$ Aprovado=N 55	0.02	0.69	1.36
	Mais de 1 ano $\wedge$ Odontologia 51 $\Rightarrow$ Aprovado=N 28	0.01	0.55	1.58	Não $\wedge$ Farmácia 87 $\Rightarrow$ Aprovado=N 61	0.02	0.7	1.39
	1 sem $\wedge$ Enfermagem 47 $\Rightarrow$ Aprovado=N 37	0.01	0.79	1.56	Não $\wedge$ BHU <sup>30</sup> 114 $\Rightarrow$ Aprovado=S 96	0.03	0.84	1.7
	1 sem $\wedge$ BHU 44 $\Rightarrow$ Aprovado=S 40	0.01	0.91	1.84	Não $\wedge$ BCT - D <sup>31</sup> 122 $\Rightarrow$ Aprovado=S 99	0.03	0.81	1.64
	1 sem $\wedge$ BCT - D 85 $\Rightarrow$ Aprovado=S 74	0.02	0.87	1.76	Não $\wedge$ BCT - TO <sup>32</sup> 270 $\Rightarrow$ Aprovado=S 162	0.05	0.6	1.21
	1 sem $\wedge$ BCT - TO 126 $\Rightarrow$ Aprovado=S 88	0.03	0.7	1.41	Aprovado=S $\wedge$ Serviço Social 73 $\Rightarrow$ Não 46	0.01	0.63	1.19
	1 sem $\wedge$ Farmácia 56 $\Rightarrow$ Aprovado=N 45	0.01	0.8	1.59	Aprovado=S $\wedge$ Administração 81 $\Rightarrow$ Não 49	0.01	0.6	1.14
					Aprovado=S $\wedge$ BHU 160 $\Rightarrow$ Não 96	0.03	0.6	1.13
					Aprovado=S $\wedge$ Ciências Contábeis 72 $\Rightarrow$ Não 43	0.01	0.6	1.13
				Aprovado=S $\wedge$ BCT - TO 288 $\Rightarrow$ Não 162	0.05	0.56	1.06	
				Ciências Contábeis 133 $\Rightarrow$ Não 92	0.03	0.69	1.31	
				Administração 286 $\Rightarrow$ Não 192	0.05	0.67	1.27	

Continua na página seguinte...

<sup>30</sup>Bacharelado em Humanidades

<sup>31</sup>Bacharelado em Ciência e Tecnologia - Diamantina

<sup>32</sup>Bacharelado em Ciência e Tecnologia - Teófilo Otoni

**Tabela 27:** Regras de Associação envolvendo os atributos aprovado, frequência pré-vestibular e curso (cont)

Ano/Sem	Regras em que há alguma frequência em pré-vestibular				Regras em que não há frequência em pré-vestibular			
	Descrição da Regra	Sup	Conf	Lift	Descrição da Regra	Sup	Conf	Lift
2009/1 (cont)					Ciências Econômicas 81 ⇒ Não 54	0.02	0.67	1.26
					Serviço Social 253 ⇒ Não 168	0.05	0.66	1.25
					BHU 185 ⇒ Não 114	0.03	0.62	1.16
					BCT - TO 447 ⇒ Não 270	0.08	0.6	1.14
2009/2	1 ano ∧ Odontologia 50 ⇒ Aprovado=N 35	0.02	0.7	2.01	Não ∧ Odontologia 44 ⇒ Aprovado=N 34	0.02	0.77	2.22
	1 sem ∧ Odontologia 138 ⇒ Aprovado=N 96	0.04	0.7	2	Não ∧ Farmácia 39 ⇒ Aprovado=N 27	0.01	0.69	1.99
	1 sem ∧ Fisioterapia 48 ⇒ Aprovado=S 41	0.02	0.85	1.31	Não ∧ BHU 73 ⇒ Aprovado=S 55	0.02	0.75	1.16
	1 sem ∧ BHU 84 ⇒ Aprovado=S 70	0.03	0.83	1.28	Não ∧ BCT - D 68 ⇒ Aprovado=S 56	0.02	0.82	1.26
	1 sem ∧ BCT - D 156 ⇒ Aprovado=S 127	0.06	0.81	1.25	Não ∧ BCT - TO 124 ⇒ Aprovado=S 94	0.04	0.76	1.16
	1 sem ∧ BCT - TO 88 ⇒ Aprovado=S 68	0.03	0.77	1.19	Não ∧ Ciências Contábeis 45 ⇒ Aprovado=S 30	0.01	0.67	1.02
	1 sem ∧ Nutrição 41 ⇒ Aprovado=S 27	0.01	0.66	1.01	Aprovado=S ∧ Administração 75 ⇒ Não 42	0.02	0.56	1.63
	1 sem ∧ Ciências Biológicas 49 ⇒ Aprovado=S 39	0.02	0.8	1.22	Aprovado=S ∧ Ciências Contábeis 58 ⇒ Não 30	0.01	0.52	1.51
	1 sem ∧ Agronomia 44 ⇒ Aprovado=S 30	0.01	0.68	1.05	Aprovado=S ∧ BCT - TO 181 ⇒ Não 94	0.04	0.52	1.51
	1 sem ∧ Engenharia Florestal 70 ⇒ Aprovado=S 50	0.02	0.71	1.1	Administração 140 ⇒ Não 85	0.04	0.61	1.77
	1 sem ∧ Zootecnia 31 ⇒ Aprovado=S 23	0.01	0.74	1.14	Serviço Social 105 ⇒ Não 63	0.03	0.6	1.75
					Ciências Contábeis 78 ⇒ Não 45	0.02	0.58	1.68
				BCT - D 286 ⇒ 1 Sem 156	0.07	0.55	1.18	
				BCT - TO 237 ⇒ Não 124	0.06	0.52	1.52	