



Universidade Federal do Piauí
Centro de Ciências da Natureza
Programa de Pós-Graduação em Ciência da Computação

**Descoberta de Conhecimento através de
Métodos de Aprendizagem de Máquina
Simbólicos aplicados ao Ensino a Distância da
Universidade Federal do Piauí**

Aline Montenegro Leal Silva

Número de Ordem PPGCC: M001

Teresina-PI, 23 de março de 2018

Aline Montenegro Leal Silva

**Descoberta de Conhecimento através de Métodos de
Aprendizagem de Máquina Simbólicos aplicados ao
Ensino a Distância da Universidade Federal do Piauí**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFPI (área de concentração: Sistemas de Computação), como parte dos requisitos necessários para a obtenção do Título de Mestre em Ciência da Computação.

Universidade Federal do Piauí – UFPI

Centro de Ciências da Natureza

Programa de Pós-Graduação em Ciência da Computação

Orientador: Vinícius Ponte Machado

Teresina-PI

23 de março de 2018

Aline Montenegro Leal Silva

Descoberta de Conhecimento através de Métodos de Aprendizagem de Máquina Simbólicos aplicados ao Ensino a Distância da Universidade Federal do Piauí/
Aline Montenegro Leal Silva. – Teresina-PI, 23 de março de 2018-

90 p. : il. (algumas color.) ; 30 cm.

Orientador: Vinícius Ponte Machado

Dissertação (Mestrado) – Universidade Federal do Piauí – UFPI

Centro de Ciências da Natureza

Programa de Pós-Graduação em Ciência da Computação, 23 de março de 2018.

1. Descoberta de conhecimento. 2. Aprendizagem de máquina. 3. Ensino a distância. I. Vinícius Ponte Machado. II. Universidade Federal do Piauí. III. Descoberta de Conhecimento através de Métodos de Aprendizagem de Máquina Simbólicos aplicados ao Ensino a Distância da Universidade Federal do Piauí

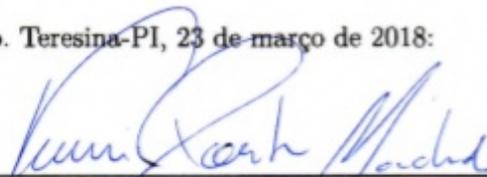
CDU 02:141:005.7

Aline Montenegro Leal Silva

**Descoberta de Conhecimento através de Métodos de
Aprendizagem de Máquina Simbólicos aplicados ao
Ensino a Distância da Universidade Federal do Piauí**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFPI (área de concentração: Sistemas de Computação), como parte dos requisitos necessários para a obtenção do Título de Mestre em Ciência da Computação.

Trabalho aprovado. Teresina-PI, 23 de março de 2018:



Vinícius Ponte Machado
Orientador



Francisco Chagas de Lima Júnior
Membro Externo (UERN)



Raimundo Santos Moura
Membro Interno



Rodrigo de Melo Souza Veras
Membro Interno

Teresina-PI
23 de março de 2018

*Ao meu esposo Jenner Vaz Silva e aos meus filhos, Levi Leal Silva e Luan Leal Silva,
por sempre estarem presentes na minha vida e me fazerem tão feliz.*

Agradecimentos

Agradeço primeiramente a Deus, pelo dom da vida, pela saúde diária, paz, família, amigos e pela oportunidade da realização satisfatória de mais um desafio.

Agradeço ao meu esposo Jenner Vaz Silva e aos meus filhos, Levi Leal Silva e Luan Leal Silva, pelo carinho e paciência nos momentos de ausência. A minha família é meu porto seguro, a força que preciso para seguir em frente.

Agradeço imensamente ao meu orientador, Vinícius Ponte Machado, por todos os ensinamentos, pela serenidade, apoio, amizade e motivação constante nesse período. Um verdadeiro exemplo de postura e competência intelectual. Seu posicionamento crítico me fez amadurecer profissionalmente e a cada dia de convívio aprendo coisas que guardarei para o resto da minha vida. Obrigada pela confiança depositada em mim desde o início.

Aos meus colegas de Mestrado, pelo companheirismo nos momentos de realizações e de dificuldades.

A todos os meus amigos, em especial aos estimados Gildásio Guedes Fernandes e Ricardo de Andrade Lira Rabêlo, pelo constante incentivo à realização deste projeto.

Aos nobres professores Laurindo de Sousa Britto Neto, Raimundo Santos Moura e Rodrigo de Melo Souza Veras, pelos conhecimentos transmitidos e pela incessante dedicação.

“Sempre busque ser o melhor! Mas não melhor que os outros, apenas o melhor de si!”
(Marcílio Flávio Rangel de Farias)

Resumo

Os cursos superiores na modalidade a distância tiveram um grande crescimento quantitativo na última década. Frente a esse crescimento, surge a preocupação com a qualidade do ensino e conseqüentemente com a performance acadêmica dos alunos. Com o intuito de identificar os perfis dos estudantes dessa modalidade educacional, especialmente dos cursos de Licenciatura em Computação, Sistemas de Informação e História, o presente trabalho propõe um processo de descoberta de conhecimento através de métodos de Aprendizagem de Máquina (AM) supervisionados aplicados ao ensino superior a distância, mais especificamente à base de dados do Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA) da Universidade Federal do Piauí (UFPI), cujos registros foram coletados nas duas últimas entradas do vestibular da Universidade Aberta do Brasil, as quais ocorreram em meados de 2014 e no início de 2017, respectivamente. Neste processo de descoberta de conhecimento, realizou-se a identificação de perfis a partir de uma correlação entre o Índice de Rendimento Acadêmico (IRA) e os aspectos sociais desses alunos. Foram utilizados três algoritmos de Aprendizagem de Máquina supervisionados com o paradigma simbólico: *J48*, *RandomTree* e *SimpleCart*. Observou-se que o *J48* obteve a melhor performance dentre os algoritmos aplicados, exibindo regras de produção bastante concisas que melhor representam a correlação do IRA com os demais atributos. Os perfis descobertos tendem a auxiliar os gestores do sistema de educação a distância na tomada de decisões em relação a melhorias no processo de ensino-aprendizagem já que, através da mineração de dados, teve-se uma ideia do desempenho do aluno, ao mostrar que a deficiência acadêmica possui correlações com aspectos sociais. A partir dessas informações é possível definir-se estratégias diferenciadas em relação a esses alunos.

Palavras-chaves: Descoberta de Conhecimento. Aprendizagem de Máquina. Educação a Distância.

Abstract

Distance learning higher education has experienced a great deal of quantitative growth in the last decade. Facing this growth, arises the concern with the quality of teaching and consequently with the academic performance of the students. In order to identify students' profiles of this educational modality, especially the courses of Degree in Computing, Information Systems and History, the present work proposes a process of knowledge discovery through supervised Machine Learning (ML) methods applied to distance higher education, more specifically to the SIGAA/UFPI database, whose records were collected in the last two entrance exams of the Open University of Brazil, which occurred in mid-2014 and early 2017, respectively. In this process of knowledge discovery, profiles were identified based on a correlation between the Academic Performance Index (API) and the social aspects of these students. Three supervised Machine Learning algorithms with symbolic paradigm were used: *J48*, *RandomTree* e *SimpleCart*. It was observed that *J48* obtained the best performance among the algorithms applied, showing very concise production rules that best represent the correlation between the API with the further attributes. The discovered profiles tend to assist managers of the distance education system in decisions making regarding improvements in the teaching learning process since, through the data mining, an idea of student performance was shown, showing that the deficiency has correlations with social aspects. Based on this information, it is possible to define differentiated strategies in relation to these students.

Keywords: Knowledge Discovery. Machine Learning. Distance Education.

Lista de ilustrações

Figura 1 – Número de matrículas em cursos superiores de graduação por modalidade de ensino no Brasil entre 2005 e 2015 - Fonte: (MEC/INEP, 2015)	3
Figura 2 – Etapas do processo de MSE - Fonte: Ferramenta The End (THEEND, 2015)	8
Figura 3 – Técnicas de IC encontradas nos EPs.	12
Figura 4 – Algoritmos de AM encontrados nos EPs.	12
Figura 5 – Distribuição dos estudos primários ao longo dos anos.	13
Figura 6 – Fases do processo de descoberta do conhecimento (adaptado de Fayyad et al. (1996)).	15
Figura 7 – Estratégias do Aprendizado de Máquina (Adaptado de Araujo (2014)).	20
Figura 8 – Hierarquia do Aprendizado (Adaptado de Araujo (2014)).	21
Figura 9 – Algoritmos de Aprendizado Simbólico (Adaptado de Araujo (2014)). .	22
Figura 10 – Árvore resultante do algoritmo <i>J48</i> para uma das bases de dados trabalhadas nesta pesquisa.	23
Figura 11 – Árvore resultante do algoritmo <i>RandomTree</i> para uma das bases de dados utilizadas nesta pesquisa.	24
Figura 12 – Funcionamento do SimpleCart através de regras geradas a partir das bases consideradas nesse trabalho.	25
Figura 13 – Resultado do Faixa_IRA em cada polo do Curso de Licenciatura em Computação da UAB4: Azul (Satisfatório) e Vermelho (Insatisfatório). .	36
Figura 14 – Acurácia resultante dos algoritmos <i>J48</i> , <i>RandomTree</i> e <i>SimpleCart</i> aplicados às sete bases de dados do SIGAA/UFPI apresentadas.	38
Figura 15 – Índice <i>Kappa</i> resultante dos algoritmos <i>J48</i> , <i>RandomTree</i> e <i>SimpleCart</i> aplicados às sete bases de dados do SIGAA/UFPI apresentadas.	39
Figura 16 – Parte das regras geradas pelo algoritmo <i>J48</i> (Base_LC_UAB4).	41
Figura 17 – Regras geradas pelo algoritmo <i>J48</i> (Base_LC_UAB4).	42
Figura 18 – Regras de produção geradas pelo algoritmo <i>J48</i> - Parte I (Base LC_UAB5).	43
Figura 19 – Regras de produção geradas pelo algoritmo <i>J48</i> - Parte II (Base LC_UAB5).	44
Figura 20 – Regras geradas pelo algoritmo <i>J48</i> (Base_SI_UAB4).	45
Figura 21 – Regras de produção geradas pelo algoritmo <i>J48</i> - Parte I (Base SI_UAB5).	47
Figura 22 – Regras de produção geradas pelo algoritmo <i>J48</i> - Parte II (Base SI_UAB5).	47
Figura 23 – Regras de produção geradas pelo algoritmo <i>J48</i> - Parte I (Base LC_SI_UAB4- _UAB5).	49

Figura 24 – Regras de produção geradas pelo algoritmo $J48$ - Parte II (Base LC_SI_UAB4-UAB5).	50
Figura 25 – Regras de produção geradas pelo algoritmo $J48$ - Parte III (Base LC_SI_UAB4-UAB5).	51
Figura 26 – Regras de produção geradas pelo algoritmo $J48$ - Parte IV (Base LC_SI_UAB4-UAB5).	51
Figura 27 – Regras geradas pelo algoritmo $J48$ (Base HI_UAB4).	53
Figura 28 – Regras de produção geradas pelo algoritmo $J48$ (Base HI_UAB5).	54

Lista de tabelas

Tabela 1 – Resultado das pesquisas nas bibliotecas digitais	9
Tabela 2 – Matriz de Confusão.	25
Tabela 3 – Interpretação dos valores do coeficiente <i>Kappa</i> (LANDIS; KOCH, 1977)	26
Tabela 4 – Discretização inicial do atributo Faixa_IRA (Teste T1).	34
Tabela 5 – Nova Discretização do atributo Faixa_IRA (Teste T2).	34
Tabela 6 – Discretização posterior do atributo Faixa_IRA (Teste T3).	34
Tabela 7 – Discretização do atributo Faixa_IRA com um conjunto de dados reduzido (Teste T4).	35
Tabela 8 – Discretização final do atributo Faixa_IRA (Teste T5)	35
Tabela 9 – Matriz de Confusão sem balanceamento de classe (<i>J48</i> - Base LC_UAB4).	35
Tabela 10 – Matriz de Confusão com balanceamento de classe (<i>J48</i> - Base LC_UAB4).	36
Tabela 11 – Matriz de Confusão (<i>J48</i> - Base SI_UAB4).	36
Tabela 12 – Matriz de Confusão (<i>J48</i> - Base HI_UAB4).	36
Tabela 13 – Matriz de Confusão (<i>J48</i> - Base LC_UAB5).	37
Tabela 14 – Matriz de Confusão (<i>J48</i> - Base SI_UAB5).	37
Tabela 15 – Matriz de Confusão (<i>J48</i> - Base HI_UAB5).	37
Tabela 16 – Matriz de Confusão (<i>J48</i> - Base LC_SI_UAB4_UAB5).	37
Tabela 17 – Resultado da acurácia e índice <i>Kappa</i> correspondentes aos algoritmos <i>J48</i> , <i>RandomTree</i> e <i>SimpleCart</i>	38
Tabela 18 – Perfis descobertos através das bases trabalhadas.	55

Lista de abreviaturas e siglas

AM	Aprendizado de Máquina
AVA	Ambiente Virtual de Aprendizagem
BD	Base de Dados
CR	Coefficiente de Rendimento
DCBD	Descoberta de Conhecimento em Base de Dados
EaD	Educação a Distância
Ea-SII	Engenharia de Software e Informática Industrial
EP	Estudos Primários
IC	Inteligência Computacional
IRA	Índice de Rendimento Acadêmico
LMS	Sistema de Gerenciamento de Aprendizagem
MSE	Mapeamento Sistemático de Estudo
QP1	Questão de Pesquisa 1
SIGAA	Sistema Integrado de Gestão de Atividades Acadêmicas
TICs	Tecnologias de Informação e Comunicação
UAB	Universidade Aberta do Brasil
UAB4	4ª Entrada do Vestibular da Universidade Aberta do Brasil
UAB5	5ª Entrada do Vestibular da Universidade Aberta do Brasil
UERJ	Universidade Estadual do Rio de Janeiro
UFBA	Universidade Federal da Bahia
UFJF	Universidade Federal de Juiz de Fora
UFPB	Universidade Federal da Paraíba
UFPI	Universidade Federal do Piauí
UFRN	Universidade Federal do Rio Grande do Norte

Sumário

1	INTRODUÇÃO	1
	INTRODUÇÃO	1
	Contextualização e Motivação	1
	Definição do Problema	2
	Justificativa	2
	Objetivos	3
	Contribuições	3
	Produções Bibliográficas	4
	Estrutura do Trabalho	4
2	TRABALHOS RELACIONADOS	5
2.1	Introdução	5
2.1.1	Estudos encontrados no Brasil	5
2.1.2	Estudos encontrados no mundo	7
2.1.2.1	Planejamento	7
2.1.2.1.1	Objetivos	8
2.1.2.1.2	Questões de Pesquisa	8
2.1.2.1.3	Base de Dados	9
2.1.2.1.4	Strings de Busca	9
2.1.2.1.5	Critérios de Inclusão	10
2.1.2.2	Processo de Seleção de Estudos	10
2.1.2.2.1	Processo de Extração de Dados	10
2.1.2.3	Resultados	10
2.2	Considerações Finais	13
3	DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS	15
3.1	Coleta	15
3.2	Pré-Processamento	16
3.2.1	Tratamento de valores desconhecidos	16
3.2.2	Tratamento de classes desbalanceadas	16
3.2.3	Seleção de Atributos	17
3.2.3.1	Ganho de Informação	17
3.2.3.2	Razão de Ganho	18
3.2.3.3	Gini	18
3.3	Transformação de Dados	19

3.4	Mineração de Dados	19
3.4.1	Aprendizagem de Máquina (AM)	19
3.4.1.1	Hierarquia do Aprendizado	20
3.4.1.2	O Paradigma Simbólico	21
3.4.1.3	Avaliação de Algoritmos	24
3.4.2	Ferramenta WEKA	26
3.5	Avaliação e Interpretação dos Resultados	26
3.6	Sistema Integrado de Gestão de Atividades Acadêmicas(SIGAA)	26
3.7	Considerações Finais	27
4	MÉTODO	29
4.1	Considerações Finais	31
5	AVALIAÇÃO DOS RESULTADOS	33
5.1	Algoritmo <i>J48</i>	33
5.1.1	Base_LC_UAB4	33
6	INTERPRETAÇÃO DOS RESULTADOS	41
6.1	Base LC_UAB4	41
6.2	Base LC_UAB5	43
6.3	Base SI_UAB4	44
6.4	Base SI_UAB5	46
6.5	Base Integrada LC_SI_UAB4_UAB5	48
6.6	Base HI_UAB4	52
6.7	Base HI_UAB5	54
6.8	Perfis Descobertos	55
7	CONCLUSÕES	57
7.1	Limitações e Continuidade da Pesquisa	58
	REFERÊNCIAS	61
	APÊNDICES	65
	APÊNDICE A – REGRAS DE PRODUÇÃO GERADAS ATRAVÉS DESTA PESQUISA	
	DESTA PESQUISA	67
A.1	Algoritmo <i>RandomTree</i>	67
A.2	Algoritmo <i>SimpleCart</i>	89

1 Introdução

Neste capítulo será discutido a importância do tema abordado, com a apresentação da contextualização e motivação, bem como a definição do problema, justificativa e os objetivos deste trabalho. Também serão apresentados as produções bibliográficas, contribuições e estrutura do trabalho.

Contextualização e Motivação

A Educação a Distância (EaD) se caracteriza por uma modalidade educacional na qual a mediação didático-pedagógica nos processos de ensino-aprendizagem ocorre com a utilização de tecnologias de informação e comunicação (TICs), com estudantes e professores desenvolvendo atividades educativas em lugares ou tempos diversos (BRASIL, 2005).

O Decreto nº 5.800 de 8 de junho de 2006 institui o Sistema Universidade Aberta do Brasil (UAB), voltado para o desenvolvimento da modalidade de Educação a distância, com a finalidade de expandir e interiorizar a oferta de cursos e programas de educação superior no País (BRASIL, 2006).

As transformações mais marcantes ocorridas na Educação a Distância refletem a mudança de um ambiente informacional para um ambiente de conhecimento, onde a figura do professor tutor é considerada como mediador da aprendizagem e facilitador do acesso ao conhecimento com base no diálogo e na interação. Dessa forma, essa modalidade de educação propicia a criação de um ambiente educacional em que o aluno precisa ser um indivíduo capaz de mostrar autonomia e compromisso com a aquisição de conhecimento, já que ele ocorre em lugares ou tempos diferentes (FARIAS, 2013). Fato esse que pode gerar uma dificuldade no acompanhamento do aluno e conseqüentemente um comprometimento do seu rendimento acadêmico, caso não haja boas estratégias de ensino-aprendizagem.

Considerando o cenário da educação superior nacional na última década, a EaD tem se caracterizado como a modalidade de ensino com maior crescimento, atingindo quase 1,4 milhão de alunos em 2015, o que já representa uma participação de 17,4% do total de matrículas da educação superior (MEC/INEP, 2015). Esse fato pode por em risco a qualidade desses cursos, desde que não haja melhorias no processo de ensino aprendizagem.

Na EaD, os ambientes de gestão da aprendizagem (LMS, do inglês *Learning Management System*) ou Ambientes Virtuais de Aprendizagem (AVAs), a exemplo do Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA), abrangem funcionalidades para armazenar, distribuir e gerenciar conteúdos de aprendizado, de forma interativa e gradativa.

Os AVAs são desenvolvidos para permitir abordagens didáticas que auxiliem a promoção do ensino e da aprendizagem em situações de mediação virtual ou semi presencial. Acumulam muitos dados também, já que todas as atividades do aluno encontram-se armazenadas em um banco de dados (BD), geridas pelos AVAs (CARVALHO et al., 2011).

Atualmente, é alarmante a distância crescente entre a geração de dados e a capacidade de analisá-los e compreendê-los. A medida que o volume de dados aumenta, a proporção dos dados que é analisada e entendida pelas pessoas diminui. Escondido entre todo este volume de dados está a informação potencialmente útil (BATISTA et al., 2003). Existe, portanto, a necessidade de uma nova geração de técnicas e ferramentas, algumas descritas nesse trabalho, que possibilite a identificação de padrões, os quais são objetos de estudo de uma área de pesquisa chamada Descoberta de Conhecimento em Base de Dados (DCBD). Para isso, usam-se ferramentas que são capazes de criar por si próprias, a partir da experiência passada, uma hipótese ou função, capaz de resolver o problema que se deseja tratar. A esse processo dá-se o nome de Aprendizado de Máquina (AM) (FACELI et al., 2011). Já a mineração de dados é uma das principais fases do processo de DCBD, em que algoritmos de AM são utilizados para a descoberta de conhecimento. Ela constitui uma ferramenta frequentemente utilizada para a análise de dados provenientes de AVA.

Definição do Problema

Na última década, a Educação a Distância tem sido a modalidade de ensino que tem apresentado maior crescimento no cenário da educação superior nacional. O número de alunos na modalidade a distância continua crescendo, atingindo quase 1,4 milhão em 2015, o que já representa uma participação de 17,4% do total de matrículas da educação superior, como mostra a Figura 1. Comparado com 2013, o número de ingressos nos cursos superiores a distância cresceu 41,2%, o que é uma evidência de que os cursos a distancia estão em clara expansão (MEC/INEP, 2015).

Frente ao crescimento quantitativo desta modalidade educacional, surge a preocupação com a qualidade do ensino e conseqüentemente com o desempenho acadêmico dos alunos. Portanto, devido à grande quantidade de dados armazenados nos AVAs, é de fundamental importância a identificação das informações relevantes presentes nesses ambientes virtuais, as quais podem ajudar na obtenção de êxito no desempenho dos alunos.

Justificativa

Para não perder a qualidade dos cursos de ensino superior a distância é importante haver tomada de decisões constantes em relação a melhorias no processo de ensino-aprendizagem. Em se tratando dos cursos da modalidade a distância da Universidade

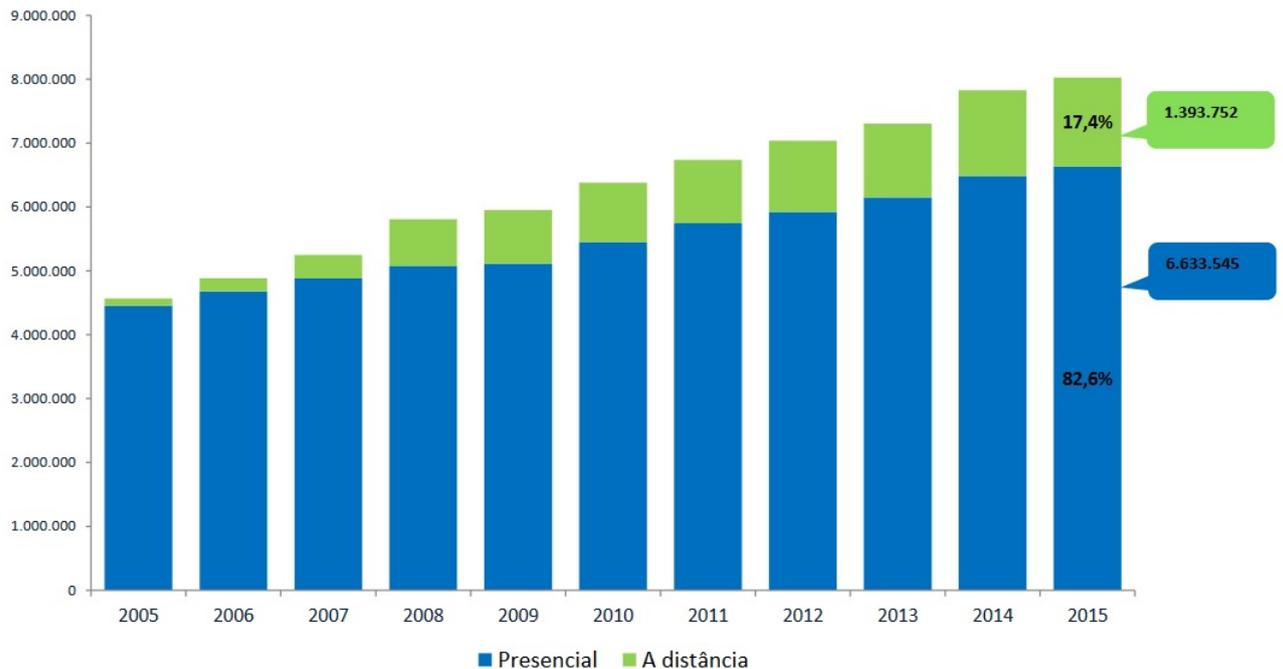


Figura 1 – Número de matrículas em cursos superiores de graduação por modalidade de ensino no Brasil entre 2005 e 2015 - Fonte: (MEC/INEP, 2015)

Federal do Piauí, é necessário estabelecer estratégias diferenciadas em relação aos alunos com baixo desempenho acadêmico. Tais estratégias podem ser definidas a partir da identificação dos perfis de alunos com base na correlação entre o IRA e os seus aspectos sociais, pois os alunos piores sucedidos para ingresso ao ensino superior são aqueles provenientes de classes sociais menos favorecidas, de acordo com Mendes Junior (2014) e Peixoto et al. (2016). Fato esse que levou, por exemplo, à implantação do sistema de cotas para acesso às Universidades Públicas Brasileiras.

Objetivos

Este trabalho tem como objetivo realizar um processo de descoberta de conhecimento através de métodos de Aprendizagem de Máquina (AM) supervisionados aplicados à base de dados do Ambiente Virtual de Aprendizagem SIGAA da Universidade Federal do Piauí a partir de uma identificação de perfis de alunos dos cursos da modalidade a distância, em especial os estudantes dos cursos de Licenciatura em Computação, Sistemas de Informação e História.

Contribuições

Como contribuições deste trabalho, destacam-se:

1. Lista com os perfis dos alunos dos cursos de Licenciatura em Computação, Sistemas de Informação e História da Universidade Federal do Piauí na modalidade a distância.
2. Guia para auxiliar gestores do sistema de educação a distância da UFPI na tomada de decisões em relação a melhorias no processo de ensino-aprendizagem, através de estratégias diferenciadas relativas aos alunos.

Produções Bibliográficas

As principais produções bibliográficas resultantes desse trabalho de pesquisa são:

- SILVA, A.M.L. et al. Descoberta de Conhecimento através de Métodos de Aprendizagem de Máquina Supervisionados aplicados ao SIGAA/UFPI. In: Revista de Sistemas e Computação - RSC, v.7, n.1, 2017. p.68-78. ISSN 2237-2903.
- SILVA, A.M.L. et al. Descoberta de Conhecimento através de Métodos de Aprendizado de Máquina Supervisionados aplicados ao Ambiente Virtual de Aprendizagem SIGAA/UFPI. Foz do Iguaçu: 23º CIAED - ABED, 2017. DOI 10.17143.
- SILVA, A.M.L. et al. Descoberta de Conhecimento através de Métodos de Aprendizado de Máquina em Sistemas de Gerenciamento de Aprendizagem: um mapeamento sistemático. Picos: ERIPI, v.1, n.1, 2017. p.182-188. ISBN: 978-85-7669-395-6.

Estrutura do Trabalho

O restante deste trabalho está organizado da seguinte maneira: no Capítulo 1 são discutidos trabalhos que inspiraram o desenvolvimento da pesquisa; no Capítulo 2 é apresentado todas as etapas do processo de DCBD e também a ferramenta usada para auxílio à descoberta do conhecimento e o ambiente virtual de aprendizagem utilizado; no Capítulo 3 é mostrado todo o método para a obtenção dos resultados; no Capítulo 4 são discutidos os resultados; a interpretação dos resultados é exibida no Capítulo 5 e por fim, no Capítulo 6 são apresentadas as conclusões, limitações e continuidade da pesquisa realizada.

2 Trabalhos Relacionados

2.1 Introdução

Para melhor apresentar os trabalhos que influenciaram essa pesquisa, o capítulo de trabalhos relacionados foi dividido em duas partes. A primeira mostra estudos encontrados que abordam técnicas de mineração de dados aplicadas à educação a distância no Brasil. Entretanto, outros estudos não utilizaram nenhuma técnica de inteligência computacional, mas foram importantes ao revelar que existe uma correlação entre o desempenho dos alunos no ensino superior e a classe social em que eles estão inseridos, também encontrados no país. A segunda parte exhibe estudos que abordam técnicas de Inteligência Computacional e algoritmos de Aprendizagem de Máquina para a determinação de padrões em base de dados relacionados a sistemas de gerenciamento de aprendizagem ou ambientes virtuais de aprendizagem encontrados no mundo.

2.1.1 Estudos encontrados no Brasil

Mendes Junior (2014) apresentou uma análise da progressão de desempenho dos alunos cotistas da Universidade Estadual do Rio de Janeiro (UERJ). Em relação aos coeficientes de rendimento (CR), foi possível perceber um pior desempenho entre os cotistas. Além disto, verificou-se que a diferença entre os não cotistas e cotistas é maior em cursos com alta dificuldade relativa e que os diferenciais de CR não tem diminuído significativamente ao longo do curso. Entre os que possuíam matrícula ativa em 2009, a diferença a favor dos não cotistas se mostrou positiva em 6,72%. Quando o universo dos alunos passou a ser a dos concluintes até o ano de 2012, a diferença aumentou para 8,50%. Percebeu-se também que a magnitude dos diferenciais de desempenho em termos de CR médio entre os alunos subiu com o aumento da dificuldade relativa do curso, atingindo os 16,35% entre os concluintes das carreiras mais difíceis.

Em Guércio et al. (2014), discutiu-se questões relevantes do desempenho acadêmico de estudantes do curso de Licenciatura em Computação da Universidade Federal de Juiz de Fora (UFJF), da modalidade a distância. Para isso, foi realizada uma descoberta de conhecimento na base de dados do ambiente virtual de aprendizagem *Moodle*, considerando três disciplinas do curso: Algoritmos, Inglês Instrumental e Língua Portuguesa. Como critérios para seleção de tais disciplinas foram analisados o número de estudantes que concluíram as mesmas, a quantidade de ofertas para as disciplinas e a disponibilidade dos recursos. Foram criados três conjuntos de treinamento, divididos por tempo de acordo com o período de duração das disciplinas. Os conjuntos foram divididos nos períodos

de 6, 12 e 18 semanas após o início do curso, para avaliar o comportamento dos alunos durante esses intervalos de tempo. Para cada um dos períodos de tempo descritos, foram analisados o total de acesso à plataforma e a interação ocorrida nos fóruns. Para a análise da interação entre alunos e tutores, foram selecionadas a quantidade de postagens criadas por cada aluno e o total dessas postagens que foram respondidas por alunos e tutores. Três algoritmos de classificação baseados em árvore de decisão foram utilizados: *J48*, *RandomTree* e *RandomForest*, sendo que o *J48* obteve a melhor performance dentre os três. Portanto, constatou-se que a partir da análise do desempenho de um aluno em um período de tempo pré-determinado, um professor é capaz de acompanhar seu rendimento, verificando se ele possui potencial para ser aprovado ou não.

Silva et al. (2015) investigou como os dados armazenados em um AVA podem ser transformados em informações potencialmente úteis para apoiar o acompanhamento de alunos em cursos EaD. Os experimentos foram realizados utilizando conjuntos de dados obtidos de um curso Técnico em Informática de uma instituição de ensino, que utiliza o AVA *Moodle*. O curso é composto de 25 disciplinas e com um total de 1171 alunos. Dentre essas disciplinas, foi selecionada a de Sistemas Operacionais, oferecida no segundo módulo. Ela disponibilizou três fóruns de discussão e do quantitativo de alunos do curso, considerou-se apenas 502 para a criação dos conjuntos de dados, porque eles participaram em algum momento da disciplina. Através dos experimentos realizados buscou-se obter um modelo preditivo que fosse capaz de prever quando um aluno da modalidade a distância apresenta características tendenciosas à evasão a partir de suas interações em fóruns de discussão. Neste trabalho, os modelos gerados a partir de árvores de decisão utilizaram os algoritmos *J48*, *BFTree* e *SimpleCart* e os baseados em estatística utilizaram os algoritmos *BayesNet* e *NaiveBayes*. Constatou-se que através das regras de classificação geradas pelo algoritmo *J48* foi possível apontar quais fatores são mais indicativos para diagnosticar alunos com tendência à evasão e reprovação. Concluiu-se que as técnicas baseadas em árvores de decisão são recomendadas nesse contexto educacional, uma vez que elas geram um resultado mais compreensível e de fácil interpretação pelo usuário.

Em Peixoto et al. (2016), comparou-se o desempenho de alunos cotistas e não cotistas da Universidade Federal da Bahia (UFBA), reunindo evidências que podiam embasar ações institucionais no sentido de assegurar sólida formação a todos os seus alunos. A amostra foi composta por todos os alunos ativos, um total de 26.175. As análises indicaram desempenho superior de não cotistas, sendo a magnitude da diferença 6,81% quando os dois grupos são diretamente comparados. O déficit na formação básica dos alunos cotistas, principalmente no domínio da matemática, aparece como um possível fator de influência nesta diferença de desempenho. Os resultados sugerem um fenômeno complexo e multifacetado. Concluiu-se que, para além das disputas ideológicas é preciso intervir nesta disparidade para uma efetiva inclusão dos cotistas na universidade.

Rabelo et al. (2017) descreveu a aplicação de técnicas de Mineração de Dados Educacionais por meio de árvores de decisão, utilizando os algoritmos *ID3* e *J48*, através de um ambiente virtual de aprendizagem educacional, a plataforma *Moodle*, durante a realização de cursos de graduação à distância pela UFRN. Com esse modelo, foi possível prever se o aluno teve sucesso ou insucesso acadêmico no decorrer do curso, mediante seu nível de participação, interação ou desempenho na plataforma. Dessa forma, não foi necessário esperar pelo final da disciplina para saber o desempenho final desses discentes, contribuindo para propor ações de ajuste de conduta durante o processo de ensino e aprendizagem e colaborando para diminuir os índices de evasão.

Para auxiliar o processo de descoberta de conhecimento proposto, foi realizado um mapeamento sistemático com o objetivo de identificar e sumarizar as técnicas de Inteligência Computacional e os algoritmos de Aprendizagem de Máquina existentes que oferecem suporte à descoberta de padrões na área de sistema de gerenciamento de aprendizagem ou ambientes virtuais de aprendizagem.

2.1.2 Estudos encontrados no mundo

Um Mapeamento Sistemático de Estudo (MSE) é um método de pesquisa que permite identificar de forma sistemática a literatura em uma área de conhecimento, apresentando na maioria das vezes um resumo visual (mapa) dos resultados encontrados (PETERSEN et al., 2008).

Para apoiar o desenvolvimento do trabalho em questão fez-se uso da ferramenta The End¹, desenvolvida pelo Laboratório de Engenharia de Software e Informática Industrial (Ea-SII), vinculado a Universidade Federal do Piauí. Ela oferece um serviço Web capaz de auxiliar o planejamento, execução e sumarização dos resultados de um MSE, tornando a pesquisa mais ágil e replicável (BRAGA et al., 2015). A Figura 2 exibe as etapas que compõem o processo de Mapeamento Sistemático implementado pela ferramenta The End.

Os resultados do MSE apresentados nesse trabalho seguem um conjunto de diretrizes propostas por Kitchenham, Pickard e Pfleeger (1995) com o objetivo de identificar e sumarizar os algoritmos de aprendizagem de máquina existentes que oferecem suporte à descoberta de padrões na área de sistema de gerenciamento de aprendizagem. A seguir apresenta-se o protocolo com as diretrizes que guiaram o estudo em questão.

2.1.2.1 Planejamento

Nessa etapa, os pesquisadores devem planejar a execução do estudo, definindo os objetivos, as questões de pesquisa, as bases de dados que serão utilizadas, a *string* de busca a ser usada nas bibliotecas digitais selecionadas, os critérios para inclusão e exclusão

¹ <http://easii.ufpi.br/theend/home>



Figura 2 – Etapas do processo de MSE - Fonte: Ferramenta The End ([THEEND, 2015](#))

dos trabalhos obtidos, formulários para extração de informações relevantes, dentre outros aspectos.

2.1.2.1.1 Objetivos

O objetivo deste trabalho é identificar as principais técnicas de Inteligência Computacional (IC) e os algoritmos de Aprendizagem de Máquina para a determinação de padrões em base de dados relacionados a sistemas de gerenciamento de aprendizagem ou ambientes virtuais de aprendizagem.

2.1.2.1.2 Questões de Pesquisa

As questões de pesquisa foram derivadas a partir do objetivo do mapeamento, apresentado anteriormente.

Estudos primários correspondem a investigações originais, que constituem as publicações encontradas nas bibliotecas digitais.

Visando estimular o levantamento e compreensão dos Estudos Primários (EP) sobre o assunto foco desse estudo, a seguinte questão de pesquisa direcionou esse trabalho.

Quais técnicas de Inteligência Computacional e algoritmos de Aprendizagem de Máquina apoiam pesquisadores e gestores na descoberta de padrões em relação aos sistemas de gerenciamento de aprendizagem? Essa questão orientou a elaboração das seguintes sub-questões de pesquisa:

1. Quais técnicas de Inteligência Computacional foram utilizadas?
2. Quais algoritmos de aprendizagem de máquina foram usados?

2.1.2.1.3 Base de Dados

No trabalho em questão, os artigos científicos foram pesquisados tomando como referência as bases de dados digitais: *Engineering Village*², *Scopus*³ e *Web of Science*⁴, pois essas bases publicam artigos das principais conferências e autores da área de Inteligência Artificial.

2.1.2.1.4 Strings de Busca

A pesquisa inicia-se pela formação das strings de busca, que são os termos utilizados nas bibliotecas digitais para a busca de estudos primários. Em todas as bases consideradas, utilizou-se a mesma string de busca: “machine learning” and “learning management system”.

A Tabela 1 apresenta a quantidade de resultados obtidos ao aplicar as strings de busca nas base de dados digitais.

Tabela 1 – Resultado das pesquisas nas bibliotecas digitais

Bases de Dados	Resultados
Engineering Village	19
Scopus	29
Web of Science	7
Total	55

Dos 55 artigos candidatos, 19 estavam em duplicidade, restando portanto, 36 artigos para serem analisados de acordo com os critérios de inclusão e exclusão.

² *Engineering Village*: <https://www.engineeringvillage.com/search/quick.url>

³ *Scopus*: <http://www.scopus.com/>

⁴ *Web of Science*: <https://webofknowledge.com/>

2.1.2.1.5 Critérios de Inclusão

A inclusão de um trabalho é determinada pela relevância em relação às questões levantadas. Baseado nisso, decidiu-se que seriam incluídos na pesquisa os estudos que:

- Devem ser escritos em Inglês;
- Devem estar publicados em workshop, conferência, revista ou jornal entre os anos de 2000 e 2016;
- Devem apresentar técnicas de Inteligência Computacional aplicadas à área de Educação a Distância (EaD);
- Devem utilizar algoritmos de aprendizagem de máquina supervisionados ou não supervisionados aplicados ao Sistema de Gerenciamento de Aprendizagem.

2.1.2.2 Processo de Seleção de Estudos

O processo de seleção de estudos foi realizado pelas fases a seguir:

1. Leitura de títulos e resumo: nessa etapa, aplicaram-se os critérios de inclusão no título e resumo de todos os 55 trabalhos candidatos à EP identificados durante a pesquisa. Desses 55 trabalhos, 25 foram aceitos, 19 trabalhos encontravam-se duplicados e 11 foram rejeitados, resultando na eliminação dos trabalhos rejeitados e duplicados.
2. Leitura completa dos trabalhos: nessa etapa, dos 25 trabalhos aceitos, 7 foram removidos por estarem disponíveis somente em forma de abstracts ou resumos expandidos e 7 foram removidos após a leitura das seções de introdução e conclusão, resultando em 11 trabalhos aceitos.

2.1.2.2.1 Processo de Extração de Dados

A extração das informações foi realizada com o auxílio da ferramenta The End e de planilhas eletrônicas contendo campos gerais (título, autores, ano etc.) e campos específicos para cada questão de pesquisa, tais como técnicas de IC e algoritmos de AM utilizados.

2.1.2.3 Resultados

Esta seção apresenta a sumarização dos resultados baseados nos 11 trabalhos incluídos como EP da pesquisa e que encontram-se listados abaixo, fornecendo uma identificação das técnicas de Inteligência Computacional (Figura 3) e algoritmos de Aprendizagem de Máquina (Figura 4) para a determinação de padrões em base de dados relacionada a

sistemas de gerenciamento de aprendizagem ou ambientes virtuais de aprendizagem, que é o objetivo deste trabalho.

- Analyzing learning concepts in intelligent tutoring systems (GÜNEL; POLAT; KURT, 2016).
- Educational Data Mining on Learning Management Systems using SCORM (CHANDRA; RAMAN, 2014).
- Using Feature Selection and Association Rule Mining to Evaluate Digital Courseware (SINGH; LAL, 2013b).
- Towards improvements on domain-independent measurements for collaborative assessment (ANAYA; BOTICARIO, 2010).
- Educational Courseware Evaluation Using Machine Learning Techniques (SINGH; LAL, 2013a).
- Application of knowledge based decision technique to predict student enrollment decision (BORAH et al., 2011).
- Multidimensional adaptations for open learning management systems (BALDIRIS et al., 2008).
- Modelling collaborative competence level using machine learning techniques (VALETTS; NAVARRO; GESA, 2008).
- Question recommender with ML business logic (BURDESCU; MIHAESCU; LOGOFATU, 2008).
- Creating glossaries using pattern-based and machine learning techniques (WESTERHOUT; MONACHESI, 2008).
- Enhancing the assessment environment within a learning management systems (BURDESCU; MIHAESCU, 2007).

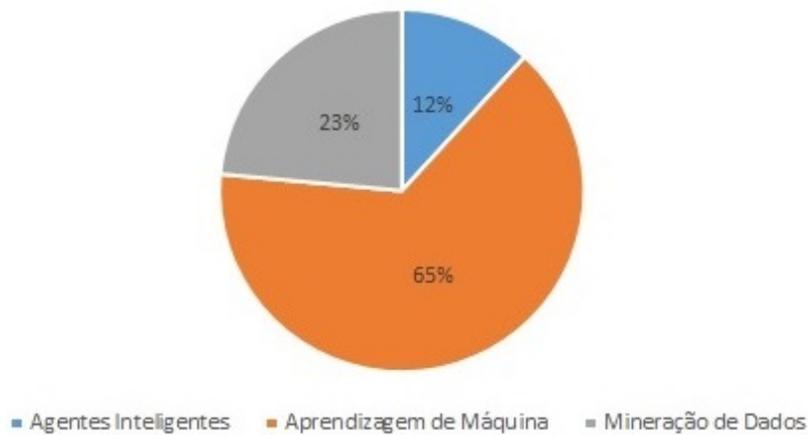


Figura 3 – Técnicas de IC encontradas nos EPs.

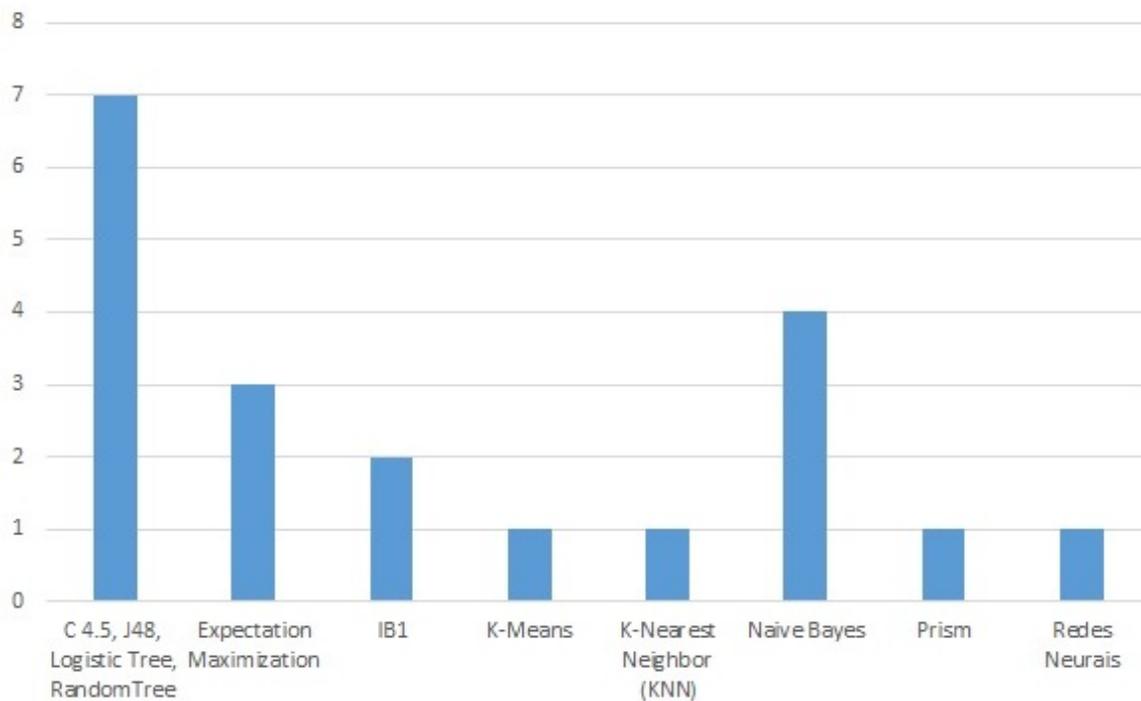


Figura 4 – Algoritmos de AM encontrados nos EPs.

Apesar da busca ter sido limitada a um período de tempo (2000 a 2016), observa-se que a maioria dos trabalhos concentraram-se nos últimos nove anos. Isto evidencia que pesquisas visando a identificação das técnicas de IC e dos algoritmos de Aprendizagem de Máquina utilizados para a determinação de padrões em base de dados relacionada a sistemas de gerenciamento de aprendizagem ou ambientes virtuais de aprendizagem têm aumentado na última década. O gráfico da Figura 5 ilustra a concentração dos estudos por ano.

De acordo com os EPs resultantes do MSE, conclui-se que a Aprendizagem de Máquina e a Mineração de Dados constituem as técnicas de inteligência computacional

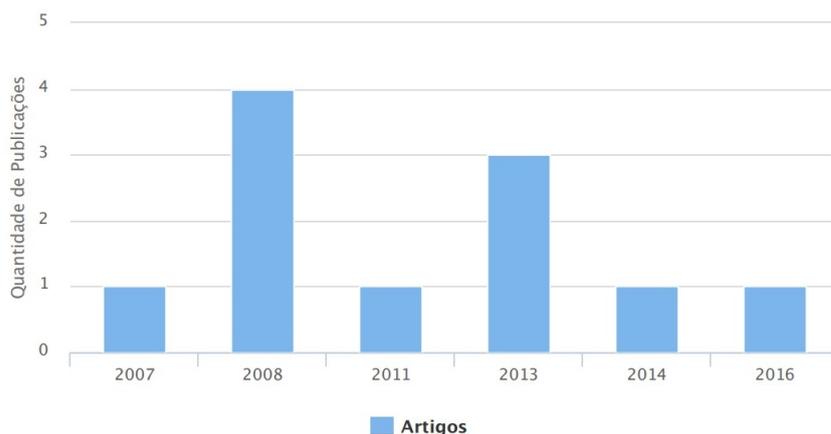


Figura 5 – Distribuição dos estudos primários ao longo dos anos.

mais utilizadas atualmente e que os algoritmos supervisionados têm se sobressaído quantitativamente em relação aos não-supervisionados durante a última década. Dentre os algoritmos supervisionados, aqueles com o paradigma simbólico baseados em árvores de decisão foram os mais utilizados, segundo os autores [Singh e Lal \(2013b\)](#), [Anaya e Boticario \(2010\)](#), [Singh e Lal \(2013a\)](#), [Borah et al. \(2011\)](#) e [Baldiris et al. \(2008\)](#), por possuírem uma classificação mais objetiva e de fácil interpretação, bem como serem computacionalmente eficientes para verificação do diagnóstico do aluno no Ambiente Virtual de Aprendizagem.

2.2 Considerações Finais

Este capítulo apresentou estudos relevantes para o desenvolvimento da pesquisa atual ao constatar que técnicas de mineração de dados através do paradigma simbólico são muito utilizadas para o auxílio na tomada de decisão em ambientes virtuais de aprendizagem. Tais técnicas também puderam ser observadas na maioria dos estudos encontrados no MSE realizado. Observou-se ainda por meio dos estudos apresentados que os alunos piores sucedidos para ingresso ao ensino superior são aqueles provenientes de classes sociais menos favorecidas, de acordo com [Mendes Junior \(2014\)](#) e [Peixoto et al. \(2016\)](#).

3 Descoberta de Conhecimento em Base de Dados

Existem diversas definições para Descoberta de Conhecimento em Base de Dados (DCBD), mas uma das mais utilizadas é a proposta por [Fayyad et al. \(1996\)](#) apud [HOLZINGER; DEHMER; JURISICA, 2014](#)), que define Descoberta do Conhecimento em Base de Dados como um processo não trivial para identificar padrões válidos, novos, potencialmente úteis e compreensíveis aos analistas humanos, se não imediatamente, ao menos após algum pós-processamento. Por não trivial é entendido que alguma busca ou inferência é utilizada. Essa definição para DCBD é também usada nesse trabalho.

O processo de DCBD considerado aqui encontra-se estruturado em 5 (cinco) fases: a) coleta: obtenção do conjunto de dados, b) pré-processamento: realiza a remoção de ruídos (dados inconsistentes) e balanceamento de classes, c) transformação: modificação da forma com que os dados estão representados para a aplicação dos algoritmos de Aprendizagem de Máquina, d) mineração de dados: aplicação dos algoritmos de AM para a obtenção de padrões, e) avaliação e interpretação dos resultados: corresponde à descoberta do conhecimento adquirido. Este processo é mostrado na Figura 6.

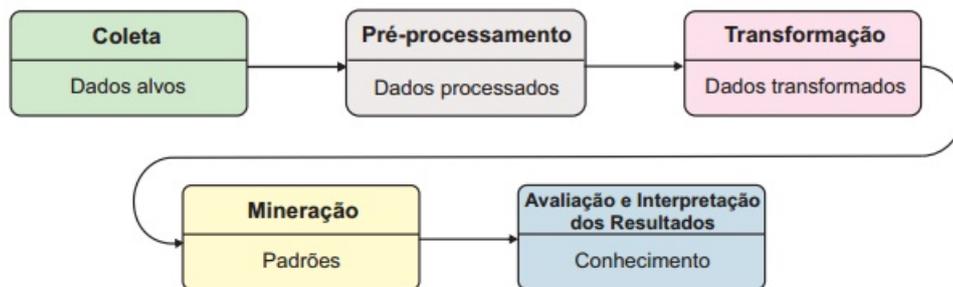


Figura 6 – Fases do processo de descoberta do conhecimento (adaptado de [Fayyad et al. \(1996\)](#)).

3.1 Coleta

Corresponde ao estágio inicial no qual os dados relativos à pesquisa são obtidos e posteriormente analisados.

Nesta pesquisa, os dados foram obtidos a partir do Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA) da Universidade Federal do Piauí.

3.2 Pré-Processamento

A literatura mostra que a etapa de pré-processamento de dados é um processo semiautomático. Por semiautomático entende-se que este processo depende da capacidade de identificar problemas (dados ausentes, fora de padrões, redundantes, inconsistentes) presentes nos dados e utilizar os métodos mais apropriados para solucionar cada um deles (BATISTA et al., 2003).

As principais tarefas encontradas nessa fase são: tratamento de valores desconhecidos, tratamento de classes desbalanceadas e seleção de atributos.

3.2.1 Tratamento de valores desconhecidos

Um problema relevante em qualidade de dados é a presença de valores desconhecidos ou valores ausentes. Esse problema surge, por exemplo, a partir do não registro dos valores de um atributo, por defeitos de equipamentos, recusa por parte dos entrevistados em responder determinadas perguntas, ausência de valores de preenchimento não obrigatório. O uso da imputação é uma das formas mais comuns de tratamento dos valores desconhecidos. Os métodos de imputação substituem valores desconhecidos por valores estimados. Esses valores estimados são calculados por meio de alguma informação extraída do conjunto de dados como a moda ou a média. Imputação pela moda e pela média são dois exemplos dessas técnicas, nelas os valores desconhecidos são substituídos pela moda ou pela média de todos os outros valores do conjunto (BATISTA et al., 2003).

3.2.2 Tratamento de classes desbalanceadas

O problema de classes desbalanceadas representa um domínio onde uma classe é representada por um grande número de exemplos (classe majoritária), enquanto que a outra é representada por poucos exemplos (classe minoritária). A maioria dos algoritmos de aprendizagem de máquina têm dificuldades em criar um modelo que classifique com precisão os exemplos da classe minoritária (BATISTA et al., 2003). Esse problema agrava-se ainda mais quando o custo da classificação incorreta da classe minoritária é muito maior do que o custo da classificação incorreta da classe majoritária.

Uma das formas mais diretas de lidar com classes desbalanceadas por meios de métodos de pré-processamento de dados é alterar a distribuição dessas classes de forma a tornar o conjunto de dados o mais balanceado possível. Existem dois métodos clássicos para balancear a distribuição das classes: i) remover exemplos da classe majoritária e ii) inserir exemplos na classe minoritária. Em sua versão mais simples, essa remoção/adição é feita de maneira aleatória. Nesses casos, os métodos são comumente chamados respectivamente de *random under-sampling* e *random over-sampling* (JAPKOWICZ; STEPHEN, 2002).

3.2.3 Seleção de Atributos

Embora considerando que todos os atributos da base de dados possam ser utilizados, é comum que existam atributos menos relevantes aos objetivos do processo pretendido, sendo importante selecionar um determinado subconjunto de atributos. A maioria dos autores defende a seleção de atributos relevantes na tentativa de diminuir a complexidade do problema e alcançar um bom desempenho no processo de classificação (NEVES, 2003). Vale destacar que os efeitos imediatos da seleção de atributos relevantes são a execução mais rápida e o aumento do desempenho do algoritmo. Algumas métricas utilizadas atualmente na seleção de atributos são descritas a seguir.

3.2.3.1 Ganho de Informação

É uma medida baseada no conceito de entropia e teve origem na área de Teoria da Informação. O algoritmo ID3 (QUINLAN, 1986), fundamentado em indução de árvores de decisão, utiliza essa medida. Para determinar o quão boa é uma condição de teste realizada é necessário comparar o grau de entropia do nó-pai (antes da divisão) com o grau de entropia dos nós-filhos (após a divisão). O atributo que gerar uma maior diferença é escolhido como condição de teste. O ganho é dado pela Equação 3.1, na forma:

$$\text{Ganho_de_informação} = \text{entropia}(\text{pai}) - \sum_{j=1}^n \left[\frac{N(v_j)}{N} \right] \text{entropia}(v_j) \quad (3.1)$$

onde n é o número de valores do atributo, ou seja, o número de nós-filhos, N é o número total de objetos do nó-pai e $N(v_j)$ é o número de exemplos associados ao nó filho v_j .

O grau de entropia é definido pela Equação 3.2 a seguir:

$$\text{entropia}(\text{nó}) = - \sum_{i=1}^c p\left(\frac{i}{\text{nó}}\right) \cdot \log_2 \left[p\left(\frac{i}{\text{nó}}\right) \right] \quad (3.2)$$

onde $p(i/\text{nó})$ é a fração dos registros pertencentes à classe i no nó e c é o número de classes.

O critério de ganho seleciona como atributo-teste aquele que maximiza o ganho de informação. O problema de se utilizar o ganho de informação é que ele dá preferência a atributos com muitos valores possíveis (número de arestas).

Um exemplo claro desse problema ocorreria ao utilizar um atributo totalmente irrelevante (por exemplo, um identificador único). Nesse caso, seria criado um nó para cada valor possível e o número de nós seria igual ao número de identificadores. Cada um desses nós teria apenas um exemplo, o qual pertence a uma única classe, ou seja, os exemplos

seriam totalmente discriminados. Assim, o valor da entropia seria mínimo porque, em cada nó, todos os exemplos (no caso um só) pertencem à mesma classe. Essa divisão geraria um ganho máximo, embora seja totalmente inútil.

3.2.3.2 Razão de Ganho

Para solucionar o problema do ganho de informação, foi proposto por [Quinlan \(1993\)](#) a Razão de Ganho (do inglês *Gain Ratio*), que nada mais é do que o ganho de informação relativo (ponderado) como critério de avaliação.

$$\text{Razão_de_ganho} = \frac{\text{ganho_de_informação}}{\text{entropia (nó)}} \quad (3.3)$$

Pela Equação 3.3, é possível perceber que a razão não é definida quando o denominador é igual a zero. Além disso, a razão de ganho favorece atributos cujo denominador, ou seja, a entropia, possui valor pequeno. Em [Quinlan \(1988\)](#), é sugerido que a razão de ganho seja realizada em duas etapas.

Na primeira etapa, é calculado o ganho de informação para todos os atributos. Após isso, considerar apenas aqueles atributos que obtiveram um ganho de informação acima da média e então escolher aquele que apresentar a melhor razão de ganho.

Dessa forma, Quinlan mostrou que a razão de ganho supera o ganho de informação tanto em termos de acurácia quanto em termos de complexidade das árvores de decisão geradas.

3.2.3.3 Gini

Outra medida bastante conhecida é o Gini, a qual emprega um índice de dispersão estatística proposto em 1912 pelo estatístico italiano Corrado Gini. Este índice é muito utilizado em análises econômicas e sociais, por exemplo, para quantificar a distribuição de renda em um certo país.

Ele é utilizado no algoritmo CART ([BREIMAN et al., 1984](#)). Para um problema de c classes, o `gini_index` é definido pela Equação 3.4, na forma:

$$\text{gini_index}(\text{nó}) = 1 - \sum_{i=1}^c p\left(\frac{i}{\text{nó}}\right) \quad (3.4)$$

Assim como no cálculo do ganho de informação, basta calcular a diferença entre o `gini_index` antes e após a divisão. Essa diferença Gini é representada pela Equação 3.5:

$$\text{Gini} = \text{gini_index}(\text{pai}) - \sum_{j=1}^n \left[\frac{N(v_j)}{N} \right] \text{gini_index}(v_j) \quad (3.5)$$

onde n é o número de valores do atributo, ou seja, o número de nós-filhos, N é o número total de objetos do nó-pai e $N(v_j)$ é o número de exemplos associados ao nó filho v_j . Assim, é selecionado o atributo que gerar um maior valor para Gini.

3.3 Transformação de Dados

O objetivo da etapa de transformação de dados é modificar a forma com que os dados estão representados. Desta forma, frequentemente tem-se atributos sendo transformados em outro com a mesma informação, mas com um formato que os algoritmos sejam melhores de tratar (BATISTA et al., 2003). Segundo esse autor, uma das transformações de dados mais comuns é a discretização, em que alguns algoritmos de mineração de dados, especialmente os algoritmos de classificação, requerem que os dados estejam na forma de atributos qualitativos. Assim, muitas vezes é necessário transformar atributos contínuos em faixa de valores nominais.

3.4 Mineração de Dados

O propósito da Mineração de Dados é detectar automaticamente padrões de associações úteis e não óbvios em grande quantidade de dados e para isso faz uso de algoritmos de Aprendizagem de Máquina (QUILICI-GONZALEZ; ZAMPIROLI, 2015).

3.4.1 Aprendizagem de Máquina (AM)

A Aprendizagem de Máquina, do inglês *machine learning*, é utilizada na fase de mineração de dados do processo de descoberta do conhecimento em base de dados e surgiu da percepção de criar programas computacionais que aprendam um determinado comportamento ou padrão automaticamente, a partir de exemplos ou observações. A ideia por trás da aprendizagem é que percepções devem ser usadas não apenas para agir, mas também para melhorar a habilidade do sistema para agir no futuro (RUSSELL; NORVIG; INTELLIGENCE, 1995).

Existem várias estratégias de aprendizagem que podem ser utilizadas por um sistema computacional como, por exemplo, o aprendizado por hábito, por instrução, por dedução, por analogia e por indução (COPPIN, 2015), como apresenta a Figura 7.

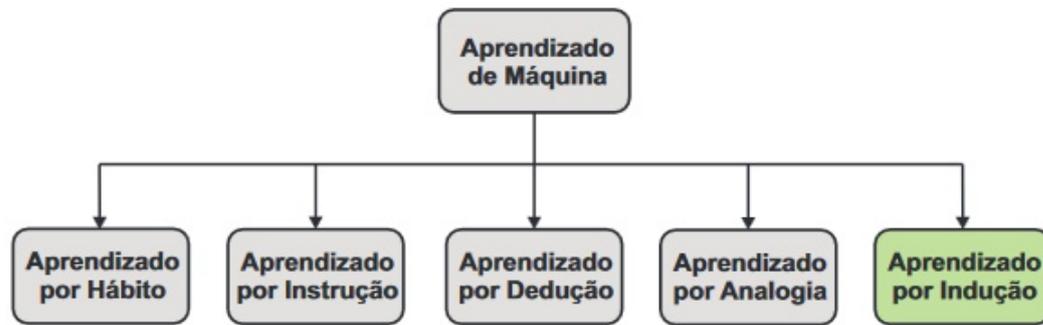


Figura 7 – Estratégias do Aprendizado de Máquina (Adaptado de Araujo (2014)).

3.4.1.1 Hierarquia do Aprendizado

O aprendizado por indução é caracterizado pelo raciocínio que parte do específico para o geral. É um modo de inferência lógica que permite obter generalizações a partir de exemplos para induzir um conceito que pode ou não preservar a verdade. Justamente por ser essa uma estratégia de aprendizado complexa, uma vez que o aprendizado desempenha a maior parte do esforço para a aquisição do conhecimento, esta permite que conceitos mais amplos possam ser aprendidos. Portanto, constitui uma das estratégias de aprendizado de interesse para as pesquisas relacionadas ao aprendizado de máquina.

O Aprendizado Indutivo pode ser dividido basicamente em Aprendizado Supervisionado e Não Supervisionado. No primeiro, busca-se a criação de um modelo preciso em relação à predição de valores para novos dados enquanto que no segundo o objetivo é encontrar características que podem resumir os dados. A diferença básica entre esses dois modos de aprendizagem é a presença ou não do atributo que rotula os exemplos do conjunto de dados. No Aprendizado Supervisionado esse rótulo é conhecido, ao passo que no aprendizado não supervisionado os exemplos não estão previamente rotulados. Adicionalmente, existe outro modo de aprendizagem, conhecido como aprendizado semi-supervisionado, no qual somente poucos exemplos encontram-se rotulados. Esse fato impossibilita o uso direto de algoritmos de Aprendizado Supervisionado, pois esse modo de aprendizagem requer um número razoável de exemplos rotulados (BLUM; MITCHELL, 1998).

Portanto, no aprendizado de máquina supervisionado, o objetivo é induzir um classificador (ou hipótese), por meio de um conjunto expressivo de dados previamente rotulados, para classificar novos exemplos ainda não rotulados. Se os rótulos das classes possuem valores discretos, o problema é conhecido como classificação. Caso as classes possuam valores contínuos, o problema é conhecido como regressão. A Figura 8 mostra a hierarquia do aprendizado descrita.

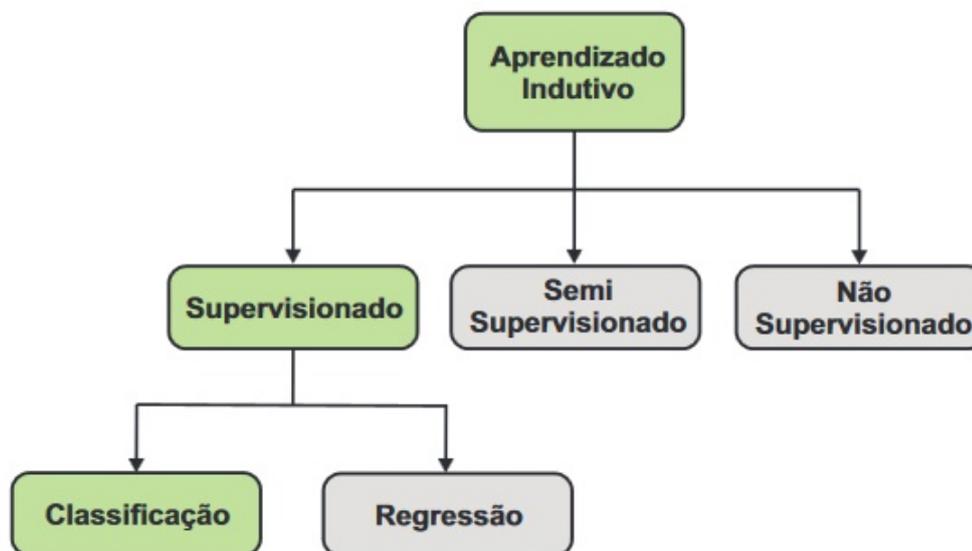


Figura 8 – Hierarquia do Aprendizado (Adaptado de Araujo (2014)).

Como o atributo classe considerado nesse trabalho possui valores discretos, optou-se por fazer uma descoberta de conhecimento a partir de uma classificação previamente conhecida, portanto escolheu-se os algoritmos supervisionados e consequentemente o paradigma simbólico, por apresentar regras de produção que podem ser facilmente interpretadas.

3.4.1.2 O Paradigma Simbólico

Os sistemas de aprendizagem simbólicos buscam aprender construindo representações simbólicas de um conceito através da análise de exemplos e contra-exemplos desse conceito. Segundo Gomes (2002), as representações simbólicas estão tipicamente na forma de alguma expressão lógica, árvore de decisão, regras de produção ou rede semântica. De acordo com Singh e Lal (2013b), Anaya e Boticario (2010), Singh e Lal (2013a), Borah et al. (2011), Baldiris et al. (2008), as árvores de decisão são bastante utilizadas atualmente no contexto educacional, gerando resultados mais compreensíveis e de fácil interpretação.

As árvores de decisão são constituídas por uma estrutura hierárquica que atua como meio de previsão/classificação (QUINLAN, 1999). Em todos os níveis da árvore são tomadas decisões acerca da estrutura do nível seguinte, até que sejam atingidos os nós terminais. Diversos autores Quinlan (1999), Banfield et al. (2007) referem-se à utilização destes modelos como sendo baseados no princípio de “dividir para conquistar”. Isto significa que em cada nível da árvore, um problema mais complexo de previsão/classificação é decomposto em problemas mais simples, traduzindo-se na geração de nós descendentes, onde a heterogeneidade da variável resposta é atenuada, originando previsões com menos riscos para cada um dos nós gerados. Este método é supervisionado, onde a variável resposta é explicada à custa das variáveis independentes medidas em qualquer escala. Quando a variável resposta é de natureza contínua, as árvores de decisão designam-se de

árvores de regressão (modelo de predição com regressão); quando é qualitativa, designam-se de árvores de classificação (modelo de predição com classificação). Neste último caso, o objetivo da previsão visa determinar a classe a que uma certa observação pertence.

A seguir são apresentados algoritmos supervisionados de classificação, baseados em árvores de decisão, de acordo com a Figura 9.

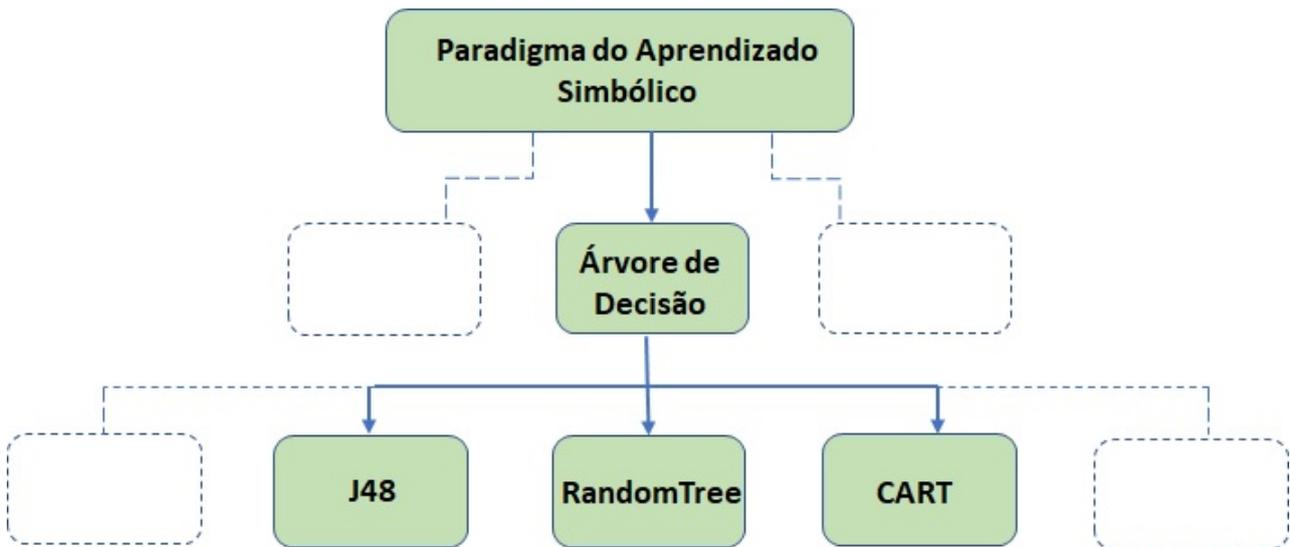


Figura 9 – Algoritmos de Aprendizado Simbólico (Adaptado de Araujo (2014)).

Algoritmo *J48*:

Algoritmo proposto por Salzberg (1994), surgiu da necessidade de recodificar o algoritmo C4.5, o qual constitui uma evolução do ID3, para a linguagem *Java*. O *J48* utiliza a abordagem de dividir-para-conquistar para a montagem da árvore, onde um problema complexo é decomposto em subproblemas mais simples, aplicando recursivamente a mesma estratégia a cada subproblema, dividindo o espaço definido pelos atributos em subespaços, associando-se a eles uma classe (WITTEN; FRANK, 2005). O algoritmo interrompe o processo de segmentação dos dados até que um nível de semelhança entre as classes finais, previamente determinado pelo usuário, seja atingido. Essas subdivisões dos dados são baseadas em procedimentos estatísticos, considerando-se os erros dos nós e seus descendentes. A identificação da raiz e dos seus descendentes é dada através dos cálculos de entropia e razão do ganho da informação. A entropia avalia a capacidade de previsão de um atributo, enquanto que a razão de ganho é usada em cada nó para escolha do atributo que mais eficientemente subdivide o conjunto das amostras em subconjuntos homogêneos e caracterizados por sua classe (ELMASRI; NAVATHE; MORAIS, 2005).

A Figura 10 ilustra uma árvore resultante do algoritmo *J48* para uma das bases de dados trabalhadas nesta pesquisa.

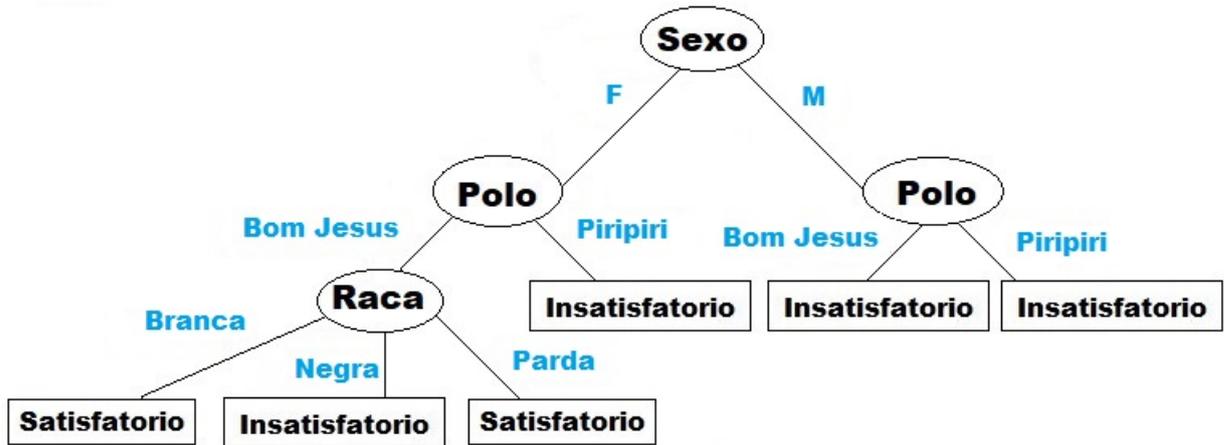


Figura 10 – Árvore resultante do algoritmo *J48* para uma das bases de dados trabalhadas nesta pesquisa.

Algoritmo *RandomTree*:

Algoritmo proposto por [Aldous \(1991\)](#), a *RandomTree* é uma árvore induzida aleatoriamente a partir de um conjunto de árvores possíveis, usando m atributos aleatórios em cada nó. O termo “aleatoriamente” significa que cada árvore tem uma chance igual de ser amostrada. Uma das formas mais utilizadas de se construir uma *RandomTree*, é através do processo de ramificação Galton-Watson ([RODRIGUEZ, 2012](#)).

De modo geral, podemos pensar um processo de ramificação Galton-Watson da seguinte maneira. Suponha que temos árvores que dão nascimento a novas árvores segundo a variável aleatória X com valores no conjunto $0,1,2,\dots$ e função de distribuição de probabilidade expressa por $P(X = k) = p_k$, para $k = 0, 1, 2, \dots, n$.

Inicialmente suponha que, temos uma única árvore na iteração $n = 0$. Essa árvore dá nascimento, na iteração $n = 1$, a novas árvores segundo a variável aleatória X . Dizemos que essas novas árvores são descendentes diretas da anterior. Isto é, ela dá nascimento a k descendentes diretos com probabilidade p_k . Em geral, se temos um certo número de árvores na iteração n , cada uma delas dá nascimento a novas árvores na iteração $n + 1$. Em cada caso, isto é feito de acordo uma variável aleatória independente e identicamente distribuída à variável aleatória X . Dizemos que a árvore da iteração $n = 0$ constitui a geração 0 e as árvores que nascem na iteração n constituem a n -ésima geração do processo, $n > 1$. Denotamos por Z_n a variável aleatória que conta o número de árvores da n -ésima geração ([RODRIGUEZ, 2012](#)).

A Figura 11 ilustra uma árvore resultante do algoritmo *RandomTree* para uma das bases de dados utilizadas nesta pesquisa.

Algoritmo *SimpleCart*:

O algoritmo *CART* (Classification and Regression Trees) foi proposto por [Breiman](#)

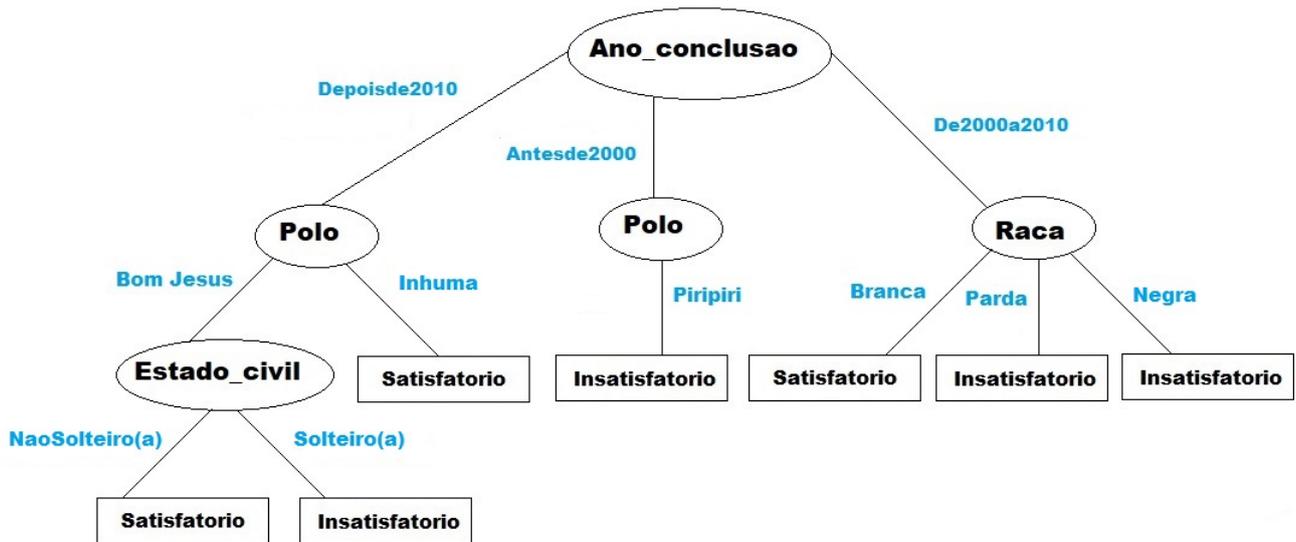


Figura 11 – Árvore resultante do algoritmo *RandomTree* para uma das bases de dados utilizadas nesta pesquisa.

et al. (1984) e consiste de uma técnica não-paramétrica que induz tanto árvores de classificação quanto árvores de regressão, dependendo se o atributo é nominal (classificação) ou contínuo (regressão).

As árvores geradas pelo algoritmo *CART* são sempre binárias, as quais podem ser percorridas da sua raiz até as folhas respondendo apenas a questões simples do tipo “sim” ou “não”.

O critério usado pelo algoritmo para a escolha dos nós é o índice de Gini, que mede a heterogeneidade dos dados. Quando este índice é igual a zero, o nó é puro. Nesse caso tem-se um nó terminal. Quando, o critério de Gini é utilizado tende-se a isolar em um ramo os registros que representam a classe mais frequente.

Diferente das abordagens adotadas por outros algoritmos, os quais utilizam pré-poda, o *CART* expande a árvore exaustivamente, realizando pós-poda por meio da redução do fator custo-complexidade (BREIMAN et al., 1984).

A Figura 12 ilustra um exemplo da execução de regras geradas pelo *SimpleCart* a partir das bases propostas nesse trabalho de pesquisa.

3.4.1.3 Avaliação de Algoritmos

A maioria dos critérios de análise dos resultados de uma classificação parte de uma matriz de confusão, que indica a quantidade de classificações corretas e incorretas para cada uma das classes. Uma matriz de confusão é criada baseada em quatro valores: verdadeiro positivo (número de exemplos corretamente identificados da classe de interesse), verdadeiro negativo (número de exemplos corretamente identificados que não pertencem a classe de interesse), falso positivo (número de exemplos que foram incorretamente atribuídos



Figura 12 – Funcionamento do SimpleCart através de regras geradas a partir das bases consideradas nesse trabalho.

a classe de interesse) e falso negativo (número de exemplos que não foram identificados como sendo da classe de interesse) e encontra-se ilustrada na Tabela 2 para uma situação de classificação binária (SOKOLOVA; LAPALME, 2009).

Tabela 2 – Matriz de Confusão.

		Predição	
		Positiva	Negativa
Exemplo	Positivo	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Negativo	Falso Positivo (FP)	Verdadeiro Negativo (VN)

A partir dessas quantidades algumas métricas podem ser calculadas para avaliar o desempenho dos classificadores, tais como: Acurácia e Índice *Kappa* (LANDIS; KOCH, 1977).

- Acurácia: mede a quantidade de instâncias corretamente classificadas (taxa de acerto) e com isenção de erros;
- Índice *Kappa*: expressa a medida da diferença entre a concordância dos dados de referência e a classificação automática e a probabilidade de concordância entre os dados de referência e a classificação aleatória. Conceitualmente, o índice *Kappa* pode ser definido de acordo com:

$$K = \frac{p_0 - p_e}{1 - p_e} = 1 - \frac{1 - p_0}{1 - p_e} \quad (3.6)$$

onde p_0 é a taxa de aceitação relativa e p_e é a taxa hipotética de aceitação.

A Tabela 3 mostra a interpretação dos valores do coeficiente *Kappa*.

Tabela 3 – Interpretação dos valores do coeficiente *Kappa* (LANDIS; KOCH, 1977)

Valores de Kappa	Interpretação
<0	Sem concordância
0 - 0,20	Pobre
0,20 - 0,39	Ligeira
0,40 - 0,59	Moderada
0,60 - 0,79	Substancial
0,80 - 1,0	Excelente

3.4.2 Ferramenta **WEKA**

Para apoiar o desenvolvimento deste trabalho, fez-se uso da ferramenta *Weka*¹, que contém uma coleção de algoritmos de Aprendizagem de Máquina e ferramentas de pré-processamento de dados projetadas para que possam experimentar os métodos existentes em novos conjuntos de dados. Ainda fornece suporte para todo o processo de mineração de dados experimental, incluindo a preparação dos dados de entrada, avaliação de sistemas de aprendizagem estatística, visualização dos dados de entrada e o resultado da aprendizagem (BOUCKAERT et al., 2016).

As bases de dados utilizadas nesse trabalho de pesquisa foram submetidas aos algoritmos de Aprendizagem de Máquina supervisionados da ferramenta WEKA (*J48*, *RandomTree* e *SimpleCart*) durante a fase de Mineração de Dados.

3.5 Avaliação e Interpretação dos Resultados

Após a fase de Mineração de Dados, o processo de DCBD entra na etapa de avaliação e interpretação dos resultados, onde se tenta descobrir se os algoritmos utilizados no processo atingiram as expectativas, avaliando os resultados de acordo com a matriz de confusão e com algumas métricas, como acurácia e índice *Kappa*, por exemplo (BATISTA et al., 2003).

3.6 Sistema Integrado de Gestão de Atividades Acadêmicas(SIGAA)

Desde 2003, a Universidade Federal do Rio Grande do Norte iniciou um projeto denominado Bases de Dados Integradas que tinha como propósito construir um único banco de dados que integrasse a área acadêmica, administrativa e de recursos humanos e que tal base de dados servisse como repositório de informações para qualquer atividade da sua comunidade. Mais tarde, esse sistema passou a ser chamado de SIGAA.

¹ <http://www.cs.waikato.ac.nz/ml/weka/>

Esse sistema informatiza os procedimentos da área acadêmica através dos módulos de: graduação, pós-graduação (stricto e lato sensu), ensino técnico, ensinos médio e infantil, submissão e controle dos projetos de ensino (monitoria e inovações), registro e relatórios da produção acadêmica dos docentes, atividades de ensino a distância e um ambiente virtual de aprendizagem denominado Turma Virtual. Atualmente o SIGAA está presente em mais de 29 universidades públicas do Brasil, incluindo a Universidade Federal do Piauí.

Através do portal do discente os alunos têm acesso às turmas virtuais do AVA, nas quais são armazenadas uma grande quantidade de informações, dentre elas fóruns de discussão, listas de exercícios e trabalhos produzidos nas disciplinas. Essas informações após mineradas podem ser bastantes úteis para a evolução do desempenho acadêmico dos alunos, já que a identificação de padrões auxiliam na tomada de decisões em relação a melhorias no processo de ensino-aprendizagem. Neste trabalho de pesquisa, todas as bases utilizadas são provenientes do SIGAA da Universidade Federal do Piauí.

3.7 Considerações Finais

Esse capítulo apresentou as etapas da Descoberta de Conhecimento em Base de Dados (coleta, pré-processamento, transformação, mineração e avaliação e interpretação dos resultados) propostas por [Fayyad et al. \(1996\)](#), bem como os critérios utilizados na avaliação de desempenho dos algoritmos, os quais envolviam as matrizes de confusão e as métricas acurácia e índice *Kappa*. Exibiu também a ferramenta *Weka*, utilizada para obtenção das regras de produção apresentadas através dos algoritmos de Aprendizagem de Máquina aplicados à base de dados do Ambiente Virtual de Aprendizagem SIGAA, também exibido nesse trabalho.

4 Método

Neste trabalho, o método empregado considerou o esquema de Descoberta do Conhecimento em Base de Dados proposto por [Fayyad et al. \(1996 apud HOLZINGER; DEHMER; JURISICA, 2014\)](#). Para a obtenção dos perfis dos alunos dos cursos na modalidade a distância, especialmente os alunos dos cursos de Licenciatura em Computação, Sistemas de Informação e História da UFPI, a partir de uma correlação entre o IRA e os aspectos sociais de tais estudantes, executou-se sequencialmente todas as fases do esquema de DCBD considerado (coleta, pré-processamento, transformação, mineração de dados, avaliação e interpretação dos resultados) em cada uma das sete bases utilizadas provenientes do SIGAA da Universidade Federal do Piauí. A correlação com aspectos sociais se deve ao fato de que os alunos piores sucedidos para ingresso ao ensino superior são aqueles provenientes de classes sociais menos favorecidas, de acordo com [Mendes Junior \(2014\)](#), [Peixoto et al. \(2016\)](#).

As bases de dados usadas foram as seguintes: Base LC_UAB4, Base LC_UAB5, Base SI_UAB4, Base SI_UAB5, Base integrada LC_SI_UAB4_UAB5, Base HI_UAB4 e Base HI_UAB5. Convém enfatizar que a nomenclatura UAB4 e UAB5 corresponde às entradas do vestibular da Universidade Aberta do Brasil, que ocorrem aproximadamente a cada dois anos. A UAB4 teve início no segundo semestre de 2014, enquanto que a UAB5 iniciou no primeiro semestre de 2017.

Considerou-se, a princípio, a base de dados do curso de Licenciatura em Computação proveniente da 4ª entrada do vestibular da EaD, intitulada de UAB4, por ser uma base menor e mais fácil de se trabalhar. Essa base foi chamada de Base LC_UAB4.

Inicialmente constatou-se que os dados dos alunos na Base LC_UAB4 encontravam-se disponibilizados em uma tabela com 21 atributos e 290 instâncias, cujos registros foram coletados a partir do segundo semestre de 2014.

Em seguida, observou-se que alguns atributos não tinham relevância para a pesquisa em questão, pois o objetivo era obter uma correlação entre o IRA dos alunos dos cursos de Licenciatura em Computação, Sistemas de Informação e História na modalidade a distância e os aspectos sociais de tais estudantes e também certos atributos, como Segundo_grau_ano_conclusao e IRA, não possuíam formato compatível com os algoritmos que foram utilizados na fase de mineração. Portanto, foi realizado um pré-processamento para diminuir a quantidade de informação irrelevante e melhorar a qualidade dos dados.

No primeiro passo do pré-processamento foi executada uma etapa de seleção de atributos. Esta seleção foi feita de duas formas: seleção manual e seleção automática. Na seleção manual, os atributos que não possuíam nenhum valor preenchido ou não tinham

correlação do IRA com os aspectos sociais dos alunos foram eliminados. Dessa forma, foram eliminados 12 atributos (Id_discente, Ano_Ingresso, Status_discente, Id_polo, Id_Curso, Curso_nome, Id_estado_civil, Id_raca, Id_tipo_necessidade_especial, Tipo_necessidade_especial, Id_cidade_endereco, Pais_nacionalidade), portanto somente 9 atributos continuaram na base de dados do curso de Licenciatura em Computação (Polo, Sexo, Conclusao_em_escola_publica, Desc_estado_civil, Raca, Segundo_grau_ano_conclusao, IRA, Cidade_endereco, Bairro).

Logo após, foi realizada a seleção automática de atributos em que calculou-se a razão de ganho para cada um dos 9 atributos restantes e aqueles com razão de ganho nulo foram eliminados. Destes, 2 obtiveram razão de ganho nulo e foram eliminados da base. Com isso, a base de dados resultante continha apenas 7 atributos (Polo, Sexo, Conclusao_em_escola_publica, Desc_estado_civil, Raca, Segundo_grau_ano_conclusao, IRA).

De posse dessas informações, foi necessário tratar o problema de valores desconhecidos, pois alguns dos atributos selecionados possuíam uns valores não preenchidos. Neste caso, optou-se por realizar a imputação pela média para preencher tais valores.

Para finalizar o pré-processamento, as técnicas de *Random Undersampling* e *Random Oversampling* foram utilizadas para criar subconjuntos de treinamento balanceados. Essas técnicas eliminam exemplos da classe majoritária e aumentam o número de exemplos da classe minoritária nos subconjuntos de treinamento, por meio da replicação aleatória dos exemplos existentes nessa classe para posterior utilização nos algoritmos de aprendizado, sendo, para isso, usado o *Resample* da ferramenta *Weka*.

A transformação nos dados foi necessária, já que existia atributos que precisavam estar discretizados. Portanto, foi feita a discretização do atributo IRA em Faixa_IRA, já que o IRA continha valores contínuos e os algoritmos de classificação trabalham com valores discretos, tornando assim Faixa_IRA atributo classe, pois o trabalho objetiva fazer uma correlação do IRA com os aspectos sociais dos alunos.

O atributo classe recebeu, após uma sequência de testes, as faixas de valores Insatisfatório (correspondente ao IRA de 0 a 6,9) e Satisfatório (IRA de 7,0 a 10,0), conforme será vista no capítulo 4.

O atributo Segundo_grau_ano_conclusao, no caso se refere ao ano de conclusão do Ensino Médio, também foi discretizado por possuir valores contínuos e consequentemente dividido na seguinte faixa de valores: Antesde2000, De2000a2010 e Depoisde2010.

A etapa seguinte consistiu na execução da fase de mineração de dados, em que algoritmos supervisionados simbólicos foram utilizados para a identificação dos perfis desejados. Visando a obtenção da extração do conhecimento para a identificação dos perfis, foram escolhidos os algoritmos *J48*, *RandomTree* e *SimpleCart* por gerarem regras

claras, principalmente o *J48*, além da facilidade de interpretação dos resultados. Eles foram submetidos ao método de teste *Cross Validation* (Validação Cruzada), que divide o conjunto de dados em K (por padrão $K=10$) partições (*Folds*) e depois separa aleatoriamente uma parte para teste e realiza o treinamento com as demais. Este procedimento é repetido para todas as partes e por último calcula a média entre elas.

Nesse trabalho utilizou-se a matriz de confusão, sendo formada pelos valores verdadeiros positivos e negativos e ainda pelo falso negativo e falso positivo. Em seguida, as métricas acurácia e índice *Kappa* foram usadas para avaliação da qualidade das regras obtidas, já que o intuito desse trabalho é descobrir perfis através de regras de produção claras e concisas provenientes de um classificador com resultado confiável.

Por último, na fase de avaliação dos resultados, foi apresentado o resultado obtido pelos algoritmos de classificação *J48*, *RandomTree* e *SimpleCart* utilizados na fase de Mineração de Dados.

4.1 Considerações Finais

Após realizar todas as fases de DCBD com a Base LC_UAB4, foram executadas as mesmas etapas com as seguintes bases:

- Base LC_UAB5: base de dados do curso de Licenciatura em Computação proveniente da 5ª e última entrada do vestibular da EaD, intitulada de UAB5. Esta base continha 21 atributos e 617 instâncias, cujos registros foram coletados a partir do primeiro semestre de 2017, provenientes da base real do SIGAA fornecida pela Universidade Federal do Piauí.
- Base SI_UAB4: base de dados do curso de Sistemas de Informação proveniente da 4ª entrada do vestibular da EaD, intitulada de UAB4. Esta base continha 21 atributos e 391 instâncias, cujos registros foram coletados a partir do segundo semestre de 2014, provenientes da base real do SIGAA fornecida pela Universidade Federal do Piauí.
- Base SI_UAB5: base de dados do curso de Sistemas de Informação proveniente da 5ª e última entrada do vestibular da EaD, intitulada de UAB5. Esta base continha 21 atributos e 813 instâncias, cujos registros foram coletados a partir do primeiro semestre de 2017, provenientes da base real do SIGAA fornecida pela Universidade Federal do Piauí.
- Base integrada LC_SI_UAB4_UAB5: corresponde às quatro bases anteriores integradas em um único arquivo. Esta base continha 22 atributos e 2018 instâncias, provenientes da base real do SIGAA fornecida pela Universidade Federal do Piauí.

- Base HI_UAB4: base de dados do curso de História proveniente da 4^a entrada do vestibular da EaD, intitulada de UAB4. Esta base continha 21 atributos e 224 instâncias, cujos registros foram coletados a partir do segundo semestre de 2014, provenientes da base real do SIGAA fornecida pela Universidade Federal do Piauí.
- Base HI_UAB5: base de dados do curso de História proveniente da 5^a e última entrada do vestibular da EaD, intitulada de UAB5. Esta base continha 21 atributos e 493 instâncias, cujos registros foram coletados a partir do primeiro semestre de 2017, provenientes da base real do SIGAA fornecida pela Universidade Federal do Piauí.

Destaca-se que todas essas sete bases (Base LC_UAB4, Base LC_UAB5, Base SI_UAB4, Base SI_UAB5, Base integrada LC_SI_UAB4_UAB5, Base HI_UAB4 e Base HI_UAB5) receberam o mesmo tratamento em relação ao pré-processamento, transformação e mineração de dados. Portanto, possuem os mesmos atributos (Polo, Sexo, Conclusao_em_escola_publica, Desc_estado_civil, Raca, Segundo_grau_ano_conclusao e Faixa_IRA), mas com quantidade de instâncias e resultados diferentes. A única exceção foi para a Base integrada LC_SI_UAB4_UAB5 que possuía um atributo a mais que as outras. Esse atributo era o curso (Licenciatura em Computação ou Sistemas de Informação), já que se trata de uma base integrada envolvendo dois cursos de graduação na modalidade a distância da Universidade Federal do Piauí.

5 Avaliação dos Resultados

Nesta seção, são apresentados os resultados dos três algoritmos com o paradigma simbólico baseados em árvores de decisão *J48*, *RandomTree* e *SimpleCart*, aplicados a cada uma das sete bases de dados provenientes do SIGAA/UFPI descritas anteriormente: Base_LC_UAB4, Base_LC_UAB5, Base_SI_UAB4, Base_SI_UAB5, Base integrada LC_SI_UAB4_UAB5, Base_HI_UAB4 e Base_HI_UAB5. A escolha por árvores de decisão é por sua classificação gerar regras claras e de fácil interpretação. Além disso, Chikalov (2011), discorre que as árvores oferecem meios que direcionam ao conhecimento proposicional, a fim de auxiliar no processo de tomada de decisão e classificação preditiva de objetos, tais como o desempenho do aluno no AVA.

Para chegarmos ao resultado obtido, foi feita uma sequência de cinco testes (T1, T2, T3, T4 e T5) na tentativa de encontrar os perfis dos alunos dos cursos de Licenciatura em Computação, Sistemas de Informação e História na modalidade a distância, a partir de uma correlação entre o IRA e os aspectos sociais de tais alunos. As diferenças entre os testes realizados se deu por causa da discretização do atributo Faixa_IRA. Em todos os três classificadores considerou-se os mesmos atributos (**Polo, Sexo, Conclusao_em_escola_publica, Desc_estado_civil, Raca, Segundo_grau_ano_conclusao, Faixa_IRA**), quantidade de instâncias e atributo classe (Faixa_IRA) e todos os testes foram realizados para cada uma das bases acima relacionadas. Como descrito anteriormente, a única exceção foi para a Base integrada LC_SI_UAB4_UAB5 que possuía um atributo a mais que as outras bases. Esse atributo era o curso (Licenciatura em Computação ou Sistemas de Informação), já que se trata de uma base integrada envolvendo dois cursos de graduação (Licenciatura em Computação e Sistemas de Informação) na modalidade a distância da Universidade Federal do Piauí.

5.1 Algoritmo *J48*

5.1.1 Base_LC_UAB4

Corresponde à base de dados do curso de Licenciatura em Computação proveniente da 4ª entrada do vestibular da EaD, intitulada de UAB4. Esta base continha a princípio 21 atributos e 290 instâncias, cujos registros foram coletados a partir do segundo semestre de 2014, provenientes da base real do SIGAA fornecida pela Universidade Federal do Piauí.

Inicialmente, os testes foram realizados com o algoritmo *J48* aplicado à base de dados LC_UAB4. No primeiro momento, o atributo classe foi discretizado, havendo uma transformação de valores contínuos em valores discretos, baseados em conceitos.

Essa discretização do atributo classe foi realizada com uma casa decimal e com um grande conjunto de dados objetivando uma redução na perda de informações gerada pela discretização, como mostra a Tabela 4 (Teste T1).

Tabela 4 – Discretização inicial do atributo Faixa_IRA (Teste T1).

IRA	Faixa_IRA
[0,0;0,9]	Um
[1,0;1,9]	Dois
[2,0;2,9]	Três
[3,0;3,9]	Quatro
[4,0;4,9]	Cinco
[5,0;5,9]	Seis
[6,0;6,9]	Sete
[7,0;7,9]	Oito
[8,0;8,9]	Nove
[9,0;10,0]	Dez

A acurácia alcançada através do teste T1 foi de 21,4521% e o índice *Kappa* de 0,0877. Pretendendo-se obter uma classificação mais confiável, realizou-se um novo teste de acordo com um conjunto menor de dados. Nesse segundo teste (T2), as 290 instâncias do atributo Faixa_IRA foram divididas em 5 grupos, resultando em uma acurácia de 32,0132% e coeficiente *Kappa* de 0,1485, conforme a tabela 5.

Tabela 5 – Nova Discretização do atributo Faixa_IRA (Teste T2).

IRA	Faixa_IRA
[0,0;1,1]	Insuficiente
[1,2;3,4]	Baixo
[3,5;6,2]	Regular
[6,3;7,8]	Bom
[7,9;10,0]	Ótimo

Um novo teste foi realizado com um conjunto ainda menor de dados com o intuito de melhorar a classificação (Teste T3). Dessa vez, a acurácia obtida foi de 53,4653% e o índice *Kappa* de 0,2157, o que gerou resultados não muito satisfatórios (Tabela 6).

Tabela 6 – Discretização posterior do atributo Faixa_IRA (Teste T3).

IRA	Faixa_IRA
[0,0;3,9]	Baixo
[4,0;6,9]	Regular
[7,0;8,9]	Bom
[9,0;10,0]	Ótimo

Um quarto teste foi realizado objetivando encontrar uma melhor classificação (Teste T4), dessa vez considerando um conjunto mais reduzido de dados para o Faixa_IRA, de acordo com a Tabela 7.

Tabela 7 – Discretização do atributo Faixa_IRA com um conjunto de dados reduzido (Teste T4).

IRA	Faixa_IRA
[0,0;3,9]	Baixo
[4,0;6,9]	Regular
[7,0;10,0]	Bom

Observou-se que a acurácia obtida por meio de T4 não foi muito satisfatória, sendo de 54,4554% e o índice *Kappa* de 0,2316. Então, finalmente, resolveu-se realizar a discretização com o menor conjunto de dados dentre todos já testados. A tabela 8 mostra a discretização final do atributo Faixa_IRA (Teste T5).

Tabela 8 – Discretização final do atributo Faixa_IRA (Teste T5)

IRA	Faixa_IRA
[0,0;6,9]	Insatisfatório
[7,0;10,0]	Satisfatório

Com esse novo teste (T5), conseguimos a maior acurácia alcançada até o momento para o atributo classe Faixa_IRA. A acurácia foi de 72,069% e o índice *Kappa* 0,3966. A Matriz de Confusão pode ser vista na Tabela 9.

Tabela 9 – Matriz de Confusão sem balanceamento de classe (*J48* - Base LC_UAB4).

	Satisfatório	Insatisfatório
Satisfatório	65	41
Insatisfatório	40	144

Constatou-se que o perfil Satisfatório ou Insatisfatório é suficiente para nosso processo, já que através dele é possível verificar se o aluno possui um IRA considerável ou não.

A partir do resultado obtido correspondente ao teste T5, permitiu-se fazer um comparativo desse resultado com aquele alcançado através de técnica de balanceamento de classe. Utilizou-se, portanto, o *Resample* da ferramenta Weka o qual gerou uma subamostragem estratificada do conjunto de dados, após a realização das técnicas de *random under-sampling* e *random over-sampling*, sendo realizadas uma sobre a outra, produzindo uma acurácia de 97,2414% e índice *Kappa* de 0,9408. A Matriz de Confusão após uso do *Resample* pode ser visualizada na Tabela 10.

Tabela 10 – Matriz de Confusão com balanceamento de classe ($J48$ - Base LC_UAB4).

	Satisfatório	Insatisfatório
Satisfatório	103	3
Insatisfatório	5	179

Verificou-se ainda que os resultados após o balanceamento tiveram uma melhora significativa, gerando regras claras e concisas provenientes de um classificador confiável. O balanceamento se fez necessário em virtude da presença de IRA elevado em determinados polos de apoio presencial do sistema de Educação a Distância, em contraposição a IRA baixo em outros polos, como pode ser visto na Figura 13.

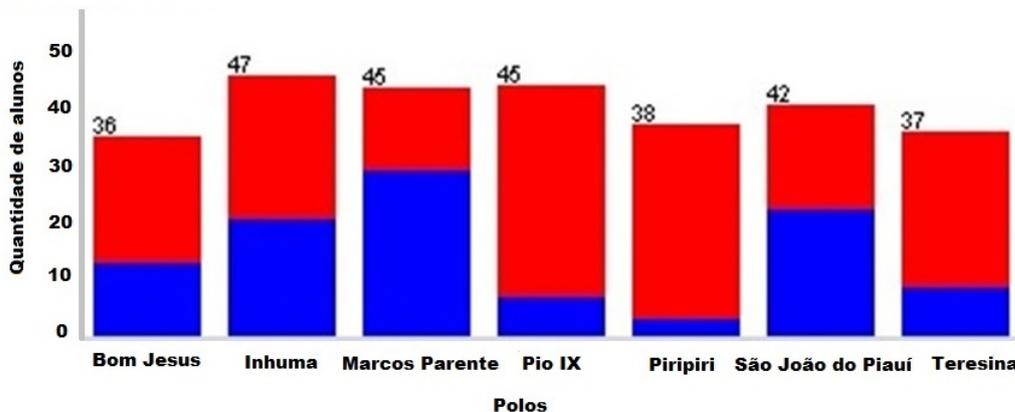


Figura 13 – Resultado do Faixa_IRA em cada polo do Curso de Licenciatura em Computação da UAB4: Azul (Satisfatório) e Vermelho (Insatisfatório).

A seguir encontram-se as Matrizes de Confusão referentes às demais bases submetidas ao algoritmo $J48$, após balanceamento de classe.

A Tabela 11 mostra a Matriz de Confusão correspondente à base SI_UAB4.

Tabela 11 – Matriz de Confusão ($J48$ - Base SI_UAB4).

	Satisfatório	Insatisfatório
Satisfatório	96	14
Insatisfatório	13	268

A Tabela 12 apresenta a Matriz de Confusão correspondente à base HI_UAB4.

Tabela 12 – Matriz de Confusão ($J48$ - Base HI_UAB4).

	Satisfatório	Insatisfatório
Satisfatório	132	5
Insatisfatório	2	85

A Tabela 13 exibe a Matriz de Confusão correspondente à base LC_UAB5.

Tabela 13 – Matriz de Confusão (*J48* - Base LC_UAB5).

	Satisfatório	Insatisfatório
Satisfatório	292	6
Insatisfatório	7	312

A Tabela 14 mostra a Matriz de Confusão correspondente à base SI_UAB5.

Tabela 14 – Matriz de Confusão (*J48* - Base SI_UAB5).

	Satisfatório	Insatisfatório
Satisfatório	285	4
Insatisfatório	16	508

A Tabela 15 apresenta a Matriz de Confusão correspondente à base HI_UAB5.

Tabela 15 – Matriz de Confusão (*J48* - Base HI_UAB5).

	Satisfatório	Insatisfatório
Satisfatório	328	16
Insatisfatório	7	142

A Tabela 16 exibe a Matriz de Confusão correspondente à base LC_SI_UAB4_UAB5.

Tabela 16 – Matriz de Confusão (*J48* - Base LC_SI_UAB4_UAB5).

	Satisfatório	Insatisfatório
Satisfatório	693	72
Insatisfatório	60	1193

Ao observar as matrizes de confusão submetidas ao algoritmo *J48* após o balanceamento de classe, conclui-se que os falso positivos e os falso negativos constituem uma quantidade menor quando comparado aos verdadeiro positivos e verdadeiro negativos, o que resulta em uma boa classificação.

Os demais algoritmos e bases de dados também fizeram uso do teste T5 descrito anteriormente, por ter sido o melhor entre todos os testes realizados.

Convém destacar que o objetivo deste trabalho não é encontrar o algoritmo com melhor desempenho, mas sim, descobrir perfis através de regras de produção concisas provenientes de um classificador com resultado confiável. Tais regras serão úteis na tomada de decisão em relação a melhorias no processo de ensino-aprendizagem, já que através da mineração de dados tem-se uma ideia do desempenho do aluno, ao mostrar que a deficiência acadêmica possui correlações com aspectos sociais.

A Tabela 17 apresenta a acurácia e índice *Kappa* resultante dos algoritmos *J48*, *RandomTree* e *SimpleCart* aplicados às sete bases de dados do SIGAA/UFPI apresentadas, após o balanceamento de classe.

Tabela 17 – Resultado da acurácia e índice *Kappa* correspondentes aos algoritmos *J48*, *RandomTree* e *SimpleCart*.

Métricas	Bases de Dados	Algoritmo		
		J48 (%)	RandomTree (%)	SimpleCart (%)
Acurácia	LC_UAB4	97,2414	98,9655	94,1379
	SI_UAB4	93,0946	94,8849	83,1202
	HI_UAB4	98,2143	95,0893	93,75
	LC_UAB5	97,893	92,3825	90,5997
	SI_UAB5	97,54	95,572	82,6568
	HI_UAB5	95,3347	91,8864	91,4807
	LC_SI_UAB4_UAB5	92,0714	90,8325	87,2646
Índice Kappa	LC_UAB4	0,9408	0,9777	0,8744
	SI_UAB4	0,8288	0,8714	0,5555
	HI_UAB4	0,9623	0,899	0,8706
	LC_UAB5	0,9578	0,8472	0,8114
	SI_UAB5	0,9468	0,9034	0,6236
	HI_UAB5	0,8913	0,8098	0,7917
	LC_SI_UAB4_UAB5	0,831	0,8079	0,7225

A Figura 14 apresenta o gráfico da acurácia resultante dos algoritmos *J48*, *RandomTree* e *SimpleCart* aplicados às sete bases de dados do SIGAA/UFPI apresentadas, após o balanceamento de classe.

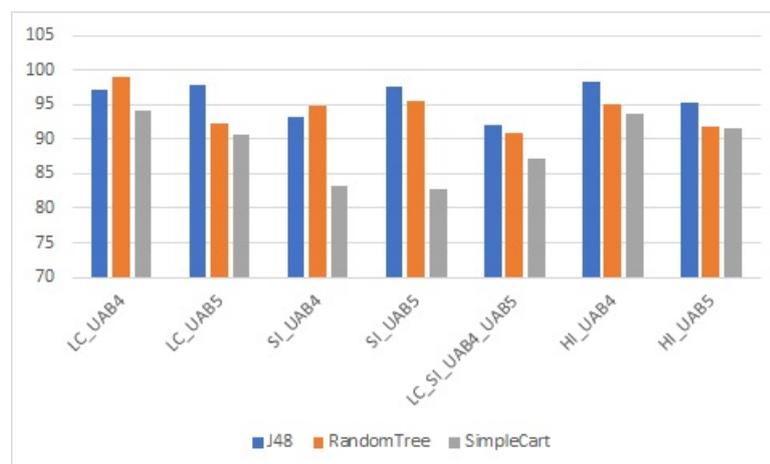


Figura 14 – Acurácia resultante dos algoritmos *J48*, *RandomTree* e *SimpleCart* aplicados às sete bases de dados do SIGAA/UFPI apresentadas.

A Figura 15 apresenta o gráfico do índice *Kappa* resultante dos algoritmos *J48*, *RandomTree* e *SimpleCart* aplicados às sete bases de dados do SIGAA/UFPI apresentadas, após o balanceamento de classe.

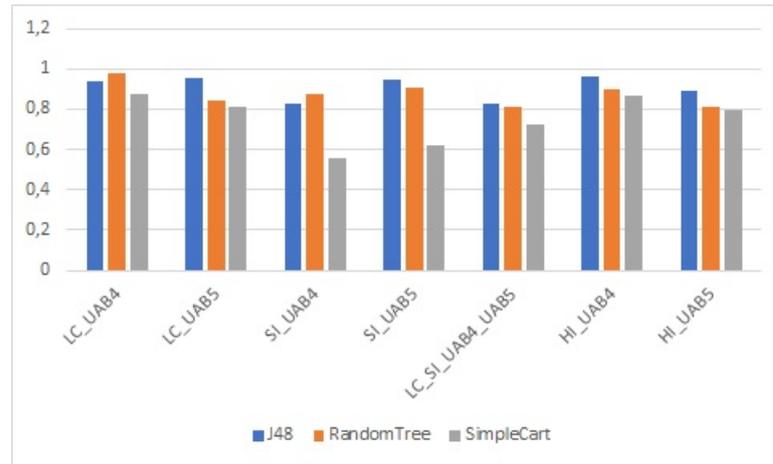


Figura 15 – Índice *Kappa* resultante dos algoritmos *J48*, *RandomTree* e *SimpleCart* aplicados às sete bases de dados do SIGAA/UFPI apresentadas.

6 Interpretação dos Resultados

Considerando que o *J48* exibiu os melhores resultados dentre os três classificadores apresentados para cinco das sete bases propostas (LC_UAB4, LC_UAB5, SI_UAB4, SI_UAB5, Base integrada LC_SI_UAB4_UAB5, HI_UAB4 e HI_UAB5), tendo parte desses resultados já sido publicados em [Silva et al. \(2017\)](#), optou-se por mostrar as regras de produção geradas por esse algoritmo para cada base de dados trabalhada, por serem concisas, de fácil interpretação e provenientes de um classificador confiável. As regras completas geradas pelos algoritmos *RandomTree* e *SimpleCart* para as bases acima citadas são exibidas no Apêndice A.

Com o intuito de entender melhor as regras apresentadas, a Figura 16 esclarece a estrutura dessas regras a partir de um fragmento gerado pela base LC_UAB4 quando submetida ao algoritmo *J48*, pois foi o que apresentou regras mais claras e facilmente interpretáveis.

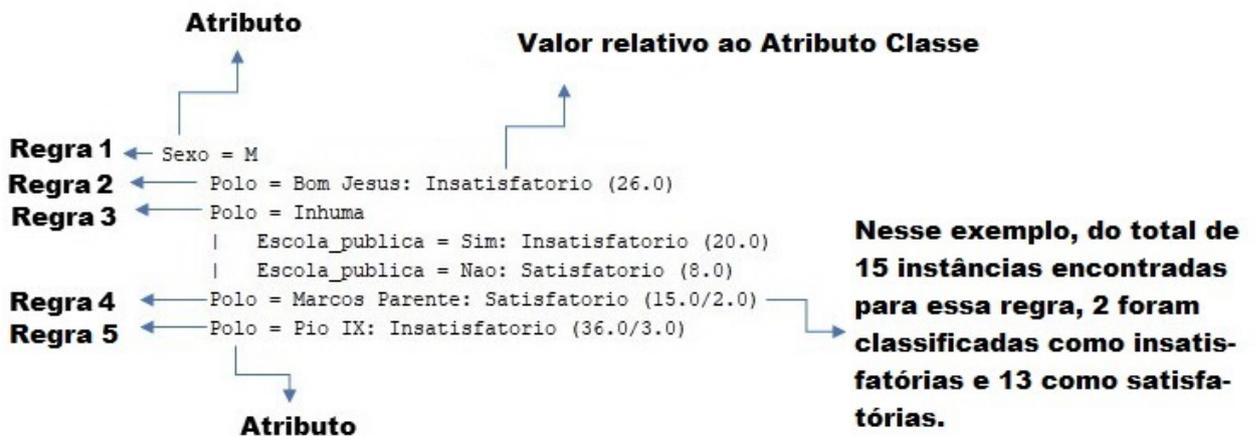


Figura 16 – Parte das regras geradas pelo algoritmo *J48* (Base_LC_UAB4).

6.1 Base LC_UAB4

Ao observar a Figura 17 que exhibe as regras de produção geradas pelo *J48* para a Base LC_UAB4, constata-se que:

- O polo da cidade de Marcos Parente foi o que obteve o melhor desempenho acadêmico, seguido pelo polo da cidade de São João do Piauí;
- O pior desempenho ficou para o polo da cidade de Piripiri, principalmente para os alunos do sexo masculino;

```

Sexo = F
| Polo = Bom Jesus
| | Raca = Branco: Satisfatorio (3.0)
| | Raca = Negro: Insatisfatorio (2.0)
| | Raca = Pardo: Satisfatorio (0.0)
| Polo = Inhuma
| | Ano_conclusao = Depoisde2010: Insatisfatorio (8.0)
| | Ano_conclusao = Antesde2000: Satisfatorio (7.0)
| | Ano_conclusao = De2000a2010: Satisfatorio (11.0)
| Polo = Marcos Parente: Satisfatorio (9.0)
| Polo = Pio IX
| | Estado_civil = NaoSolteiro(a): Satisfatorio (11.0)
| | Estado_civil = Solteiro(a): Insatisfatorio (4.0)
| Polo = Piripiri: Insatisfatorio (5.0)
| Polo = Sao Joao do Piaui
| | Ano_conclusao = Depoisde2010: Satisfatorio (6.0)
| | Ano_conclusao = Antesde2000: Insatisfatorio (3.0)
| | Ano_conclusao = De2000a2010: Satisfatorio (2.0)
| Polo = Teresina: Satisfatorio (23.0)
Sexo = M
| Polo = Bom Jesus: Insatisfatorio (26.0)
| Polo = Inhuma
| | Escola_publica = Sim: Insatisfatorio (20.0)
| | Escola_publica = Nao: Satisfatorio (8.0)
| Polo = Marcos Parente: Satisfatorio (15.0/2.0)
| Polo = Pio IX: Insatisfatorio (36.0/3.0)
| Polo = Piripiri: Insatisfatorio (41.0)
| Polo = Sao Joao do Piaui
| | Estado_civil = NaoSolteiro(a): Insatisfatorio (4.0)
| | Estado_civil = Solteiro(a): Satisfatorio (10.0)
| Polo = Teresina: Insatisfatorio (36.0)

```

Figura 17 – Regras geradas pelo algoritmo J_48 (Base_LC_UAB4).

- Os polos das cidades de Bom Jesus, Pio IX, Teresina e Inhuma obtiveram um desempenho intermediário entre os mais satisfatórios e os menos satisfatórios;
- Os alunos que concluíram o ensino médio depois de 2010 possuem um desempenho inferior àqueles cuja conclusão aconteceu anteriormente;
- As mulheres não solteiras se sobressaem academicamente melhor que os homens não solteiros;
- Os homens solteiros possuem um melhor desempenho do que as mulheres solteiras;
- Os alunos do sexo feminino possuem um desempenho superior aos do sexo masculino;
- Os alunos não solteiros têm um desempenho melhor que os alunos solteiros;
- Os estudantes da raça parda ou negra possuem um desempenho inferior aos da raça branca, sobretudo no polo de Bom Jesus.

- Os alunos que estudaram em escola pública têm uma performance acadêmica abaixo daqueles que não estudaram na mesma, principalmente no polo de Inhumas.

6.2 Base LC_UAB5

A Figura 18 exibe a primeira parte das regras de produção geradas pelo *J48* para a Base LC_UAB5.

```

Polo = Alegrete do Piaui
| Ano_conclusao = Depoisde2010: Satisfatorio (19.0)
| Ano_conclusao = Antesde2000: Insatisfatorio (0.0)
| Ano_conclusao = De2000a2010: Insatisfatorio (24.0)
Polo = Campo Alegre de Lourdes: Satisfatorio (43.0)
Polo = Campo Maior
| Ano_conclusao = Depoisde2010: Satisfatorio (23.0)
| Ano_conclusao = Antesde2000: Satisfatorio (0.0)
| Ano_conclusao = De2000a2010
| | Estado_civil = NaoSolteiro(a): Satisfatorio (4.0)
| | Estado_civil = Solteiro(a): Insatisfatorio (13.0)
Polo = Castelo do Piaui: Satisfatorio (26.0)
Polo = Esperantina
| Estado_civil = NaoSolteiro(a): Insatisfatorio (11.0)
| Estado_civil = Solteiro(a)
| | Escola_publica = Sim: Insatisfatorio (8.0)
| | Escola_publica = Nao: Satisfatorio (4.0)
Polo = Inhumas
| Estado_civil = NaoSolteiro(a): Satisfatorio (3.0)
| Estado_civil = Solteiro(a): Insatisfatorio (14.0/3.0)
Polo = Marcos Parente
| Raca = Branco: Insatisfatorio (6.0)
| Raca = Negro: Satisfatorio (0.0)
| Raca = Pardo: Satisfatorio (14.0/2.0)
Polo = Picos
| Estado_civil = NaoSolteiro(a): Satisfatorio (6.0)
| Estado_civil = Solteiro(a): Insatisfatorio (29.0)
Polo = Paes Landim
| Ano_conclusao = Depoisde2010
| | Raca = Branco: Satisfatorio (5.0)
| | Raca = Negro: Satisfatorio (6.0/1.0)
| | Raca = Pardo: Insatisfatorio (8.0)
| Ano_conclusao = Antesde2000: Insatisfatorio (3.0)
| Ano_conclusao = De2000a2010: Satisfatorio (12.0)
Polo = Pio IX
| Sexo = F: Insatisfatorio (34.0)
| Sexo = M: Satisfatorio (27.0/1.0)
Polo = Santa Cruz do Piaui
| Ano_conclusao = Depoisde2010: Satisfatorio (17.0/1.0)
| Ano_conclusao = Antesde2000: Satisfatorio (16.0)
| Ano_conclusao = De2000a2010: Insatisfatorio (17.0/1.0)

```

Figura 18 – Regras de produção geradas pelo algoritmo *J48* - Parte I (Base LC_UAB5).

A Figura 19 exibe o restante das regras de produção geradas pelo *J48* para a Base LC_UAB5.

```

Polo = SaoJoaodoPiaui
| Sexo = F: Insatisfatorio (123.0)
| Sexo = M
| | Raca = Branco: Satisfatorio (0.0)
| | Raca = Negro: Insatisfatorio (14.0)
| | Raca = Pardo: Insatisfatorio (6.0)
Polo = Uniao
| Sexo = F: Satisfatorio (5.0)
| Sexo = M: Insatisfatorio (14.0/2.0)

```

Figura 19 – Regras de produção geradas pelo algoritmo *J48* - Parte II (Base LC_UAB5).

Ao observar as regras de produção geradas pelo *J48* para a Base LC_UAB5, verifica-se que:

- Os polos cujos alunos possuem os melhores desempenhos acadêmicos são os das cidades de Campo Alegre de Lourdes (BA) e Castelo do Piauí;
- O polo que possui alunos com o pior índice acadêmico é o da cidade de Picos.
- Os polos das cidades de São João do Piauí, União, Inhumas, Pio IX, Esperantina, Paes Landim, Santa Cruz do Piauí e Campo Maior também apresentaram desempenho acadêmico considerado intermediário entre os mais satisfatórios e os menos satisfatórios;
- Os alunos que concluíram o ensino médio entre 2000 e 2010 possuem um desempenho inferior àqueles cuja conclusão aconteceu anteriormente ou posteriormente;
- Os homens possuem Índice de Rendimento Acadêmico melhor que o das mulheres;
- Os alunos não solteiros possuem um desempenho acadêmico superior ao dos solteiros;
- Os alunos cuja raça é branca se sobressaem academicamente melhor que os alunos de raça parda ou negra.
- Os estudantes que estudaram em escola pública têm uma performance acadêmica inferior àqueles que não estudaram na mesma.

6.3 Base SI_UAB4

A Figura 20 apresenta as regras de produção geradas pelo *J48* para a Base SI_UAB4.

```

Escola_publica = Sim
|  Raca = Branco
|  |  Estado_civil = NaoSolteiro(a): Insatisfatorio (18.0)
|  |  Estado_civil = Solteiro(a)
|  |  |  Polo = Barras: Insatisfatorio (6.0)
|  |  |  Polo = Campo Maior: Satisfatorio (0.0)
|  |  |  Polo = Esperantina: Satisfatorio (0.0)
|  |  |  Polo = Jaicos: Satisfatorio (27.0)
|  |  |  Polo = Oeiras: Insatisfatorio (4.0)
|  |  |  Polo = Paes Landim: Satisfatorio (0.0)
|  |  |  Polo = Regeneracao: Satisfatorio (2.0/1.0)
|  |  |  Polo = Valenca do Piaui: Satisfatorio (0.0)
|  |  |  Polo = Uniao: Satisfatorio (0.0)
|  Raca = Negro
|  |  Ano_conclusao = Depoisde2010: Satisfatorio (7.0)
|  |  Ano_conclusao = Antesde2000: Insatisfatorio (4.0)
|  |  Ano_conclusao = De2000a2010: Insatisfatorio (5.0)
|  Raca = Pardo
|  |  Polo = Barras: Insatisfatorio (29.0/2.0)
|  |  Polo = Campo Maior
|  |  |  Sexo = F: Satisfatorio (6.0)
|  |  |  Sexo = M
|  |  |  |  Ano_conclusao = Depoisde2010: Insatisfatorio (14.0)
|  |  |  |  Ano_conclusao = Antesde2000: Insatisfatorio (0.0)
|  |  |  |  Ano_conclusao = De2000a2010: Satisfatorio (3.0)
|  |  Polo = Esperantina: Insatisfatorio (11.0)
|  |  Polo = Jaicos
|  |  |  Estado_civil = NaoSolteiro(a)
|  |  |  |  Sexo = F: Insatisfatorio (3.0)
|  |  |  |  Sexo = M: Satisfatorio (13.0/4.0)
|  |  |  Estado_civil = Solteiro(a): Insatisfatorio (65.0/1.0)
|  |  Polo = Oeiras: Insatisfatorio (18.0/1.0)
|  |  Polo = Paes Landim: Insatisfatorio (57.0)
|  |  Polo = Regeneracao: Insatisfatorio (19.0/1.0)
|  |  Polo = Valenca do Piaui
|  |  |  Sexo = F: Satisfatorio (6.0)
|  |  |  Sexo = M
|  |  |  |  Estado_civil = NaoSolteiro(a): Insatisfatorio (5.0)
|  |  |  |  Estado_civil = Solteiro(a): Satisfatorio (20.0/9.0)
|  |  Polo = Uniao: Insatisfatorio (11.0)
Escola_publica = Nao
|  Estado_civil = NaoSolteiro(a): Satisfatorio (34.0)
|  Estado_civil = Solteiro(a): Insatisfatorio (4.0/1.0)

```

Figura 20 – Regras geradas pelo algoritmo J_48 (Base_SI_UAB4).

Ao analisar as regras de produção geradas pelo J_48 para a Base SI_UAB4, infere-se que:

- O polo da cidade de Valença do Piauí foi o que obteve o melhor desempenho acadêmico;

- Os piores desempenhos foram dos polos das cidades de Paes Landim, Esperantina e União, seguido dos polos das cidades de Oeiras, Barras e Regeneração, principalmente para os alunos da raça parda;
- Os alunos que concluíram o ensino antes de 2000 possuem um desempenho acadêmico inferior àqueles cuja conclusão aconteceu posteriormente;
- Os alunos não solteiros possuem um desempenho superior ao dos alunos solteiros;
- Os estudantes do sexo feminino se sobressaem academicamente melhor que os do sexo masculino;
- Os alunos da raça parda ou negra possuem um desempenho inferior aos da raça branca.
- Os alunos que estudaram em escola pública têm uma performance acadêmica abaixo daqueles que não estudaram na mesma.

6.4 Base SI_UAB5

A Figura 21 exibe a primeira parte das regras de produção geradas pelo $J48$ para a Base SI_UAB5.

```

Polo = Agua Branca
| Ano_conclusao = Depoisde2010: Satisfatorio (46.0)
| Ano_conclusao = Antesde2000: Insatisfatorio (5.0)
| Ano_conclusao = De2000a2010: Insatisfatorio (34.0)
Polo = Barras: Insatisfatorio (25.0)
Polo = Buriti dos Lopes: Insatisfatorio (38.0)
Polo = Cajazeiras do Piaui: Insatisfatorio (22.0)
Polo = Canto do Buriti
| Estado_civil = NaoSolteiro(a): Satisfatorio (72.0/6.0)
| Estado_civil = Solteiro(a)
| | Sexo = F
| | | Ano_conclusao = Depoisde2010: Satisfatorio (12.0)
| | | Ano_conclusao = Antesde2000: Insatisfatorio (0.0)
| | | Ano_conclusao = De2000a2010: Insatisfatorio (20.0)
| | Sexo = M: Insatisfatorio (60.0)
Polo = Castelo do Piaui: Insatisfatorio (28.0)
Polo = Campo Maior: Insatisfatorio (13.0)
Polo = Elesbao Veloso
| Sexo = F: Satisfatorio (17.0)
| Sexo = M
| | Estado_civil = NaoSolteiro(a): Satisfatorio (11.0)
| | Estado_civil = Solteiro(a): Insatisfatorio (27.0)
Polo = Esperantina
| Ano_conclusao = Depoisde2010: Satisfatorio (4.0)
| Ano_conclusao = Antesde2000: Insatisfatorio (2.0)
| Ano_conclusao = De2000a2010: Insatisfatorio (54.0)
Polo = Gilbues
| Estado_civil = NaoSolteiro(a): Insatisfatorio (21.0)
| Estado_civil = Solteiro(a): Satisfatorio (3.0)
Polo = Jaicos
| Estado_civil = NaoSolteiro(a): Insatisfatorio (11.0)
| Estado_civil = Solteiro(a)
| | Sexo = F: Insatisfatorio (9.0)
| | Sexo = M: Satisfatorio (44.0)
Polo = Oeiras
| Estado_civil = NaoSolteiro(a): Insatisfatorio (21.0)
| Estado_civil = Solteiro(a): Satisfatorio (64.0/7.0)
Polo = Regeneracao: Insatisfatorio (25.0)
Polo = Simoes
| Raca = Branco: Satisfatorio (2.0)
| Raca = Negro: Insatisfatorio (0.0)
| Raca = Pardo: Insatisfatorio (26.0)

```

Figura 21 – Regras de produção geradas pelo algoritmo $J48$ - Parte I (Base SI_UAB5).

A Figura 22 exibe o restante das regras de produção geradas pelo $J48$ para a Base SI_UAB5.

```

Polo = Simplicio Mendes: Insatisfatorio (42.0)
Polo = Valenca do Piaui
| Raca = Branco: Satisfatorio (14.0)
| Raca = Negro: Satisfatorio (0.0)
| Raca = Pardo: Insatisfatorio (9.0)
Polo = Uniao
| Escola_publica = Sim: Insatisfatorio (17.0)
| Escola_publica = Nao: Satisfatorio (15.0/2.0)

```

Figura 22 – Regras de produção geradas pelo algoritmo $J48$ - Parte II (Base SI_UAB5).

Ao analisar as regras de produção geradas pelo *J48* para a Base SI_UAB5, conclui-se que:

- Os polos das cidades de Jaicós, Oeiras e Valença do Piauí são os que possuem os alunos com os melhores desempenhos acadêmicos.
- Os polos das cidades de Simplício Mendes, Buriti dos Lopes, Castelo do Piauí, Regeneração, Barras, Cajazeiras, Gilbués, Campo Maior, Esperantina e Simões possuem os piores desempenhos acadêmicos;
- Os polos das cidades de União, Canto do Buriti, Água Branca e Elesbão Veloso possuem desempenho acadêmico considerado intermediário entre os mais satisfatórios e os menos satisfatórios;
- Os alunos que concluíram o ensino médio entre 2000 e 2010 possuem um desempenho acadêmico inferior àqueles cuja conclusão aconteceu anteriormente ou posteriormente;
- As mulheres se sobressaem academicamente melhor que os homens;
- Os alunos solteiros possuem um desempenho acadêmico inferior ao dos não solteiros;
- Os estudantes que possuem a raça parda ou negra têm um desempenho acadêmico abaixo dos alunos cuja raça é branca.
- Os alunos que estudaram em escola pública têm uma performance acadêmica inferior àqueles que não estudaram na mesma;

6.5 Base Integrada LC_SI_UAB4_UAB5

A Figura 23 exibe a primeira parte das regras de produção geradas pelo *J48* para a Base integrada LC_SI_UAB4_UAB5.

```

Polo = Agua Branca
| Raca = Branco: Insatisfatorio (2.0)
| Raca = Negro: Satisfatorio (3.0)
| Raca = Pardo: Insatisfatorio (34.0/6.0)
Polo = Alegrete do Piaui
| Sexo = F: Insatisfatorio (32.0/8.0)
| Sexo = M: Satisfatorio (22.0)
Polo = Barras: Insatisfatorio (100.0/18.0)
Polo = Bom Jesus
| Ano_conclusao_segundo_grau = Antesde2000
| | Raca = Branco: Satisfatorio (12.0)
| | Raca = Negro: Insatisfatorio (0.0)
| | Raca = Pardo: Satisfatorio (8.0)
| Ano_conclusao_segundo_grau = De2000a2010: Insatisfatorio (31.0)
| Ano_conclusao_segundo_grau = Depoisde2010: Insatisfatorio (2.0)
Polo = Buriti dos Lopes: Insatisfatorio (41.0/1.0)
Polo = Cajazeiras do Piaui: Insatisfatorio (57.0)
Polo = Campo Alegre de Lourdes
| Estado_civil = Solteiro(a): Satisfatorio (23.0/4.0)
| Estado_civil = NaoSolteiro(a): Insatisfatorio (11.0)
Polo = Campo Maior
| Ano_conclusao_segundo_grau = Antesde2000: Insatisfatorio (2.0)
| Ano_conclusao_segundo_grau = De2000a2010
| | Curso = Computacao
| | | Estado_civil = Solteiro(a): Satisfatorio (23.0)
| | | Estado_civil = NaoSolteiro(a): Insatisfatorio (5.0)
| | Curso = Sistemas de Informacao: Satisfatorio (31.0)
| Ano_conclusao_segundo_grau = Depoisde2010: Insatisfatorio (67.0/1.0)
Polo = Canto do Buriti: Satisfatorio (58.0)
Polo = Castelo do Piaui
| Raca = Branco: Satisfatorio (15.0)
| Raca = Negro: Satisfatorio (3.0)
| Raca = Pardo
| | Curso = Computacao: Satisfatorio (68.0/5.0)
| | Curso = Sistemas de Informacao
| | | Conclusao_escola_publica = Sim: Insatisfatorio (21.0/5.0)
| | | Conclusao_escola_publica = Nao: Satisfatorio (3.0)
Polo = Elesbao Veloso: Satisfatorio (49.0/4.0)
Polo = Esperantina: Insatisfatorio (61.0/18.0)
Polo = Gilbues: Insatisfatorio (9.0)
Polo = Inhumas
| Ano_conclusao_segundo_grau = Antesde2000: Insatisfatorio (2.0)
| Ano_conclusao_segundo_grau = De2000a2010
| | Sexo = F: Satisfatorio (12.0)
| | Sexo = M: Insatisfatorio (17.0/5.0)
| Ano_conclusao_segundo_grau = Depoisde2010: Satisfatorio (21.0)

```

Figura 23 – Regras de produção geradas pelo algoritmo $J48$ - Parte I (Base LC_SI_UAB4_UAB5).

A Figura 24 exibe a segunda parte das regras de produção geradas pelo $J48$ para a Base LC_SI_UAB4_UAB5.

```

Polo = Jaicos: Insatisfatorio (60.0)
Polo = Marcos Parente
|  Sexo = F
|  |  Estado_civil = Solteiro(a): Satisfatorio (15.0)
|  |  Estado_civil = NaoSolteiro(a): Insatisfatorio (2.0)
|  Sexo = M
|  |  Ano_conclusao_segundo_grau = Antesde2000: Insatisfatorio (0.0)
|  |  Ano_conclusao_segundo_grau = De2000a2010
|  |  |  Raca = Branco: Insatisfatorio (2.0)
|  |  |  Raca = Negro: Satisfatorio (4.0)
|  |  |  Raca = Pardo: Insatisfatorio (6.0)
|  |  Ano_conclusao_segundo_grau = Depoisde2010: Insatisfatorio (13.0/1.0)
Polo = Oeiras
|  Ano_conclusao_segundo_grau = Antesde2000: Insatisfatorio (9.0)
|  Ano_conclusao_segundo_grau = De2000a2010: Insatisfatorio (15.0)
|  Ano_conclusao_segundo_grau = Depoisde2010: Satisfatorio (34.0/1.0)
Polo = Paes Landim
|  Raca = Branco: Satisfatorio (10.0)
|  Raca = Negro
|  |  Estado_civil = Solteiro(a): Insatisfatorio (7.0)
|  |  Estado_civil = NaoSolteiro(a): Satisfatorio (5.0)
|  Raca = Pardo
|  |  Curso = Computacao: Satisfatorio (3.0/1.0)
|  |  Curso = Sistemas de Informacao: Insatisfatorio (21.0)
Polo = Picos
|  Ano_conclusao_segundo_grau = Antesde2000: Insatisfatorio (0.0)
|  Ano_conclusao_segundo_grau = De2000a2010
|  |  Conclusao_escola_publica = Sim: Insatisfatorio (4.0)
|  |  Conclusao_escola_publica = Nao: Satisfatorio (5.0)
|  Ano_conclusao_segundo_grau = Depoisde2010: Insatisfatorio (15.0)
Polo = Pio IX
|  Estado_civil = Solteiro(a): Insatisfatorio (64.0)
|  Estado_civil = NaoSolteiro(a)
|  |  Ano_conclusao_segundo_grau = Antesde2000: Satisfatorio (0.0)
|  |  Ano_conclusao_segundo_grau = De2000a2010: Satisfatorio (27.0/2.0)
|  |  Ano_conclusao_segundo_grau = Depoisde2010: Insatisfatorio (2.0)
Polo = Piripiri: Insatisfatorio (63.0)
Polo = Regeneracao
|  Raca = Branco: Satisfatorio (12.0)
|  Raca = Negro
|  |  Ano_conclusao_segundo_grau = Antesde2000: Insatisfatorio (0.0)
|  |  Ano_conclusao_segundo_grau = De2000a2010: Insatisfatorio (53.0)
|  |  Ano_conclusao_segundo_grau = Depoisde2010
|  |  |  Sexo = F: Insatisfatorio (5.0)
|  |  |  Sexo = M: Satisfatorio (26.0/3.0)
|  Raca = Pardo: Insatisfatorio (59.0)

```

Figura 24 – Regras de produção geradas pelo algoritmo *J48* - Parte II (Base LC_SI_UAB4-UAB5).

A Figura 25 exibe a terceira parte das regras de produção geradas pelo *J48* para a Base LC_SI_UAB4_UAB5.

```

Polo = Santa Cruz do Piaui
| Raca = Branco: Satisfatorio (12.0)
| Raca = Negro: Insatisfatorio (2.0)
| Raca = Pardo: Satisfatorio (29.0)
Polo = Simplicio Mendes
| Raca = Branco: Satisfatorio (6.0)
| Raca = Negro: Insatisfatorio (0.0)
| Raca = Pardo: Insatisfatorio (8.0)
Polo = Simoes
| Raca = Branco: Insatisfatorio (2.0)
| Raca = Negro: Satisfatorio (10.0)
| Raca = Pardo: Insatisfatorio (38.0)
Polo = Sao Joao do Piaui
| Raca = Branco: Insatisfatorio (0.0)
| Raca = Negro
| | Estado_civil = Solteiro(a): Satisfatorio (13.0)
| | Estado_civil = NaoSolteiro(a): Insatisfatorio (28.0)
| Raca = Pardo
| | Estado_civil = Solteiro(a)
| | | Sexo = F: Satisfatorio (56.0/22.0)
| | | Sexo = M: Insatisfatorio (48.0/9.0)
| | Estado_civil = NaoSolteiro(a)
| | | Sexo = F
| | | | Ano_conclusao_segundo_grau = Antesde2000: Insatisfatorio (11.0)
| | | | Ano_conclusao_segundo_grau = De2000a2010: Insatisfatorio (0.0)
| | | | Ano_conclusao_segundo_grau = Depoisde2010: Satisfatorio (3.0)
| | | | Sexo = M: Insatisfatorio (40.0)
Polo = Teresina
| Estado_civil = Solteiro(a): Insatisfatorio (40.0)
| Estado_civil = NaoSolteiro(a): Satisfatorio (11.0)

```

Figura 25 – Regras de produção geradas pelo algoritmo *J48* - Parte III (Base LC_SI_UAB4_UAB5).

A Figura 26 exibe o restante das regras de produção geradas pelo *J48* para a Base LC_SI_UAB4_UAB5.

```

Polo = Uniao
| Curso = Computacao
| | Raca = Branco: Insatisfatorio (3.0)
| | Raca = Negro: Insatisfatorio (1.0)
| | Raca = Pardo: Satisfatorio (68.0/6.0)
| Curso = Sistemas de Informacao
| | Raca = Branco: Satisfatorio (2.0/1.0)
| | Raca = Negro
| | | Estado_civil = Solteiro(a): Satisfatorio (18.0)
| | | Estado_civil = NaoSolteiro(a)
| | | | Conclusao_escola_publica = Sim: Insatisfatorio (9.0)
| | | | Conclusao_escola_publica = Nao: Satisfatorio (9.0)
| | Raca = Pardo
| | | Estado_civil = Solteiro(a): Insatisfatorio (21.0)
| | | Estado_civil = NaoSolteiro(a)
| | | | Conclusao_escola_publica = Sim
| | | | | Ano_conclusao_segundo_grau = Antesde2000: Satisfatorio (10.0)
| | | | | Ano_conclusao_segundo_grau = De2000a2010: Insatisfatorio (10.0/4.0)
| | | | | Ano_conclusao_segundo_grau = Depoisde2010: Satisfatorio (0.0)
| | | | | Conclusao_escola_publica = Nao: Insatisfatorio (10.0)
Polo = Valenca do Piaui: Insatisfatorio (109.0/10.0)

```

Figura 26 – Regras de produção geradas pelo algoritmo *J48* - Parte IV (Base LC_SI_UAB4_UAB5).

Ao observar as regras de produção geradas pelo *J48* para a Base LC_SI_UAB4_UAB5, infere-se que:

- Dos 29 polos relativos aos cursos de Licenciatura em Computação e Sistemas de Informação da UAB4 e UAB5, o polo em que os alunos possuem os melhores desempenhos acadêmicos é o da cidade de Canto do Buriti, seguido pelos polos das cidades de Santa Cruz do Piauí, Elesbão Veloso, Campo Alegre de Lourdes e Castelo do Piauí;
- Os polos das cidades de Cajazeiras do Piauí, Gilbués, Jaicós, Piripiri e Picos possuem os piores níveis acadêmicos, seguidos pelos polos das cidades de Buriti dos Lopes, Valença do Piauí, Barras, Simões, Teresina, Regeneração, Água Branca, Pio IX e Esperantina. Os alunos dos demais polos tiveram desempenho considerado intermediário entre os mais satisfatórios e os menos satisfatórios;
- Os alunos que concluíram o ensino médio entre 2000 e 2010 possuem um desempenho acadêmico inferior àqueles cuja conclusão aconteceu anteriormente ou posteriormente;
- As mulheres se sobressaem academicamente melhor que os homens;
- Os alunos solteiros possuem um desempenho acadêmico inferior ao dos não solteiros;
- Os alunos da raça parda ou negra têm um desempenho acadêmico abaixo dos alunos cuja raça é branca.
- Os estudantes que não estudaram em escola pública têm uma performance acadêmica superior àqueles que estudaram na mesma.

6.6 Base HI_UAB4

A Figura 27 apresenta as regras de produção geradas pelo *J48* para a Base HI_UAB4.

```

Sexo = F
| Polo = Castelo do Piauí: Satisfatorio (16.0)
| Polo = Esperantina
| | Estado_civil = NaoSolteiro(a): Insatisfatorio (8.0)
| | Estado_civil = Solteiro(a): Satisfatorio (5.0)
| Polo = Inhumá: Insatisfatorio (13.0)
| Polo = Jaicós
| | Escola_publica = Nao: Insatisfatorio (2.0)
| | Escola_publica = Sim: Satisfatorio (15.0/2.0)
| Polo = Urucui: Satisfatorio (2.0)
Sexo = M
| Raca = Branco: Satisfatorio (2.0)
| Raca = Negro
| | Polo = Castelo do Piauí: Insatisfatorio (0.0)
| | Polo = Esperantina: Satisfatorio (6.0)
| | Polo = Inhumá: Insatisfatorio (0.0)
| | Polo = Jaicós: Insatisfatorio (0.0)
| | Polo = Urucui: Insatisfatorio (8.0)
| Raca = Pardo: Insatisfatorio (56.0/1.0)

```

Figura 27 – Regras geradas pelo algoritmo $J48$ (Base HI_UAB4).

Ao verificar as regras de produção geradas pelo $J48$ para a Base HI_UAB4, observa-se que:

- O polo da cidade de Castelo do Piauí foi o que obteve o melhor desempenho acadêmico, seguido pelo polo da cidade de Jaicós;
- O pior desempenho foi do polo da cidade de Inhumá, principalmente para os alunos do sexo feminino;
- Os polos das cidades de Esperantina e Uruçuí possuem desempenho acadêmico considerado intermediário entre os mais satisfatórios e os menos satisfatórios;
- Os alunos solteiros do sexo feminino se sobressaem academicamente melhor que os alunos não solteiros do mesmo sexo;
- Os alunos do sexo feminino têm um desempenho acadêmico superior àqueles do sexo masculino;
- Os estudantes da raça parda ou negra possuem um desempenho inferior aos da raça branca.
- Os alunos que estudaram em escola pública têm uma performance acadêmica superior àqueles que não estudaram na mesma, principalmente no polo de Jaicós. Esse fato, foi justamente o inverso do que ocorreu nos cursos de Licenciatura em Computação e Sistemas de Informação, em que os alunos provenientes de escola pública tiveram um desempenho acadêmico inferior quando comparados àqueles que estudaram em escolas privadas.

6.7 Base HI_UAB5

A Figura 28 exibe as regras de produção geradas pelo *J48* para a Base HI_UAB5.

```

Raca = Branco
| Polo = Redencao do Gurgueia: Satisfatorio (19.0)
| Polo = Santa Cruz do Piaui: Satisfatorio (4.0)
| Polo = Sao Jose do Peixe: Satisfatorio (0.0)
| Polo = Castelo do Piaui: Satisfatorio (2.0)
| Polo = Corrente: Satisfatorio (9.0)
| Polo = Esperantina: Insatisfatorio (46.0)
| Polo = Floriano: Satisfatorio (0.0)
| Polo = Inhumas: Satisfatorio (55.0)
| Polo = Jaicos: Satisfatorio (9.0)
| Polo = Urucui: Satisfatorio (0.0)
Raca = Negro: Insatisfatorio (12.0)
Raca = Pardo
| Polo = Redencao do Gurgueia: Satisfatorio (45.0/4.0)
| Polo = Santa Cruz do Piaui
| | Segundo_grau_ano_conclusao = Antesde2000: Satisfatorio (4.0)
| | Segundo_grau_ano_conclusao = Depoisde2010: Insatisfatorio (18.0)
| | Segundo_grau_ano_conclusao = de2000a2010: Satisfatorio (9.0)
| Polo = Sao Jose do Peixe
| | Segundo_grau_ano_conclusao = Antesde2000: Satisfatorio (7.0)
| | Segundo_grau_ano_conclusao = Depoisde2010: Insatisfatorio (24.0/8.0)
| | Segundo_grau_ano_conclusao = de2000a2010: Satisfatorio (0.0)
| Polo = Castelo do Piaui: Satisfatorio (19.0)
| Polo = Corrente: Satisfatorio (2.0/1.0)
| Polo = Esperantina: Satisfatorio (25.0/1.0)
| Polo = Floriano: Satisfatorio (16.0)
| Polo = Inhumas
| | Escola_publica = Nao: Insatisfatorio (14.0)
| | Escola_publica = Sim: Satisfatorio (15.0/1.0)
| Polo = Jaicos
| | Segundo_grau_ano_conclusao = Antesde2000: Satisfatorio (0.0)
| | Segundo_grau_ano_conclusao = Depoisde2010: Satisfatorio (36.0)
| | Segundo_grau_ano_conclusao = de2000a2010
| | | Estado_civil = NaoSolteiro(a): Satisfatorio (6.0)
| | | Estado_civil = Solteiro(a): Insatisfatorio (12.0)
| Polo = Urucui
| | Estado_civil = NaoSolteiro(a)
| | | Segundo_grau_ano_conclusao = Antesde2000: Insatisfatorio (5.0)
| | | Segundo_grau_ano_conclusao = Depoisde2010: Insatisfatorio (18.0/1.0)
| | | Segundo_grau_ano_conclusao = de2000a2010
| | | | Sexo = F: Satisfatorio (5.0)
| | | | Sexo = M: Insatisfatorio (18.0/4.0)
| | Estado_civil = Solteiro(a): Satisfatorio (2.0)

```

Figura 28 – Regras de produção geradas pelo algoritmo *J48* (Base HI_UAB5).

Ao verificar as regras de produção geradas pelo *J48* para a Base HI_UAB5, constata-se que:

- Os polos das cidades de Castelo do Piauí, Floriano, Redenção do Gurgueia e Corrente são os que possuem os alunos com os melhores desempenhos acadêmicos;
- Os polos das cidades de Uruçuí, Esperantina, Santa Cruz do Piauí e São José do Peixe possuem os piores desempenhos acadêmicos;
- Os polos das cidades de Inhumas e Jaicós possuem desempenho considerado intermediário entre os mais satisfatórios e os menos satisfatórios;
- Os alunos que concluíram o ensino médio depois de 2010 possuem um desempenho acadêmico inferior àqueles cuja conclusão aconteceu anteriormente;
- As mulheres se sobressaem academicamente melhor que os homens;
- Os estudantes solteiros possuem um desempenho acadêmico inferior ao dos não solteiros;
- Os alunos que possuem a raça parda ou negra têm um desempenho acadêmico abaixo dos alunos cuja raça é branca.
- Os alunos que estudaram em escola pública têm uma performance acadêmica superior àqueles que não estudaram na mesma, principalmente no polo de Inhumas, sendo esse mesmo perfil confirmado no curso de História da UAB4. Esse padrão encontrado foi justamente o contrário do que ocorreu nos cursos de Licenciatura em Computação e Sistemas de Informação, em que os alunos provenientes de escola pública tiveram um desempenho acadêmico inferior quando comparados àqueles que estudaram em escolas privadas.

6.8 Perfis Descobertos

Ao analisar as sete bases trabalhadas (LC_UAB4, LC_UAB5, SI_UAB4, SI_UAB5, Base integrada LC_SI_UAB4_UAB5, HI_UAB4 e HI_UAB5), obteve-se os perfis apresentados na tabela 18 a seguir.

Tabela 18 – Perfis descobertos através das bases trabalhadas.

Bases	Polo(s)	Perfis de Melhor Desempenho				
		Conclusão 2º Grau	Sexo	Estado Civil	Raça	Escola Pública
LC_UAB4	Marcos Parente	Anterior a 2010	Feminino	Não Solteiro	Branca	Não
SI_UAB4	Valença do Piauí	A partir de 2000	Feminino	Não Solteiro	Branca	Não
HI_UAB4	Castelo do Piauí	-	Feminino	Solteiro	Branca	Sim
LC_UAB5	Campo Alegre de Lourdes (BA) e Castelo do Piauí	Anterior a 2000 e a partir de 2010	Masculino	Não Solteiro	Branca	Não
SI_UAB5	Jaicós, Oeiras e Valença do Piauí	Anterior a 2000 e a partir de 2010	Feminino	Não Solteiro	Branca	Não
HI_UAB5	Castelo do Piauí, Floriano, Redenção do Gurgueia e Corrente	Anterior a 2010	Feminino	Não Solteiro	Branca	Sim
LC_SI_UAB4_UAB5	Canto do Buriti, Santa Cruz do Piauí, Elesbão Veloso, Campo Alegre de Lourdes (BA) e Castelo do Piauí	Anterior a 2000 e a partir de 2010	Feminino	Não Solteiro	Branca	Não

Através dos perfis encontrados, observou-se que entre os alunos dos cursos de Licenciatura em Computação, Sistemas de Informação e História da UAB4 e UAB5, os melhores desempenhos acadêmicos são de alunos da raça branca e que não estudaram em escola pública. A exceção foi para o curso de História em que os alunos que estudaram em escola pública obtiveram um melhor desempenho. Em relação ao sexo e estado civil, constatou-se que os alunos do sexo feminino possuem uma performance mais satisfatória que os estudantes do sexo masculino e os não solteiros têm um desempenho acadêmico superior ao dos solteiros. Quanto ao ano de conclusão do 2º grau, três das sete bases (LC_UAB5, SI_UAB5 e LC_SI_UAB4_UAB5) mostraram um melhor desempenho para os anos anteriores a 2000 e posteriores a 2010.

Verificou-se também que o polo de Valença do Piauí obteve o melhor desempenho acadêmico do curso de Sistemas de Informação, tanto da UAB4, quanto da UAB5 e que o polo de Castelo do Piauí se destacou entre os melhores academicamente nos cursos de História (UAB4 e UAB5) e Licenciatura em Computação (UAB5). Já os dois polos com as melhores performances na base integrada LC_SI_UAB4_UAB5 (Campo Alegre de Lourdes (BA) e Castelo do Piauí) se confirmaram na base LC_UAB5.

A partir dos perfis descobertos pode-se propor estratégias diferenciadas aos gestores do sistema de educação a distância para auxílio na tomada de decisões em relação a melhorias no processo de ensino-aprendizagem. Algumas dessas estratégias encontram-se listadas a seguir.

- Realização de um acompanhamento presencial por parte dos tutores nos polos de apoio do sistema de educação a distância em que foi identificado perfis de alunos com deficiência no desempenho acadêmico;
- Realização de um nivelamento antes do início do curso na modalidade a distância, para os alunos que não se enquadram nos perfis de desempenho acadêmico satisfatório.

Outras estratégias mais específicas ficarão a critério dos gestores do sistema de educação a distância, os quais poderão tomar decisões a partir da lista de perfis descobertos nesse trabalho de pesquisa.

7 Conclusões

Com o intuito de identificar perfis dos alunos dos cursos de Licenciatura em Computação, Sistemas de Informação e História da Universidade Federal do Piauí na modalidade a distância, esse trabalho apresentou um processo de descoberta de conhecimento a partir de uma correlação entre o IRA e os aspectos sociais dos estudantes, cujos registros foram coletados inicialmente a partir do segundo semestre de 2014 correspondendo à quarta entrada do vestibular (UAB4) e posteriormente com início no primeiro semestre de 2017, relativo à quinta entrada do vestibular (UAB5). Para isso, utilizou sete base de dados (LC_UAB4, LC_UAB5, SI_UAB4, SI_UAB5, Base integrada LC_SI_UAB4_UAB5, HI_UAB4 e HI_UAB5) presentes no SIGAA/UFPI e fez uso de três algoritmos de Aprendizagem de Máquina (AM) supervisionados com o paradigma simbólico aplicados a essas bases: *J48*, *RandomTree* e *SimpleCart*.

Cinco dessas bases (LC_UAB4, LC_UAB5, SI_UAB4, SI_UAB5, Base integrada LC_SI_UAB4_UAB5) foram escolhidas com o propósito de verificar se os perfis descobertos se confirmam para o mesmo curso ou para área afim, como é o caso de Licenciatura em Computação e Sistemas de Informação. As outras duas bases (HI_UAB4 e HI_UAB5) foram selecionadas a partir da área de Ciências Humanas com o intuito de observar se os perfis descobertos se assemelham ou não aos de uma área distinta, como a de Ciências Exatas.

Convém destacar que o objetivo deste trabalho não era encontrar o classificador com melhor desempenho, mas sim, descobrir perfis através de regras de produção concisas e de fácil interpretação provenientes de um possível classificador confiável.

Observou-se que o *J48* exibiu em grande parte, um melhor desempenho em relação aos outros classificadores, mostrando regras de produção bastante concisas que melhor representam a correlação do IRA com os demais atributos considerados.

Conseguiu-se encontrar padrões entre as sete bases testadas. Percebeu-se, de modo geral, que os alunos que estudaram em escola pública tiveram um desempenho acadêmico inferior àqueles que não estudaram na mesma. A única exceção foi para o curso de História em que os alunos provenientes de escola pública tiveram um desempenho acadêmico superior àqueles que estudaram em instituições privadas. Observou-se também que os estudantes da raça branca se sobressaíram academicamente melhor em relação àqueles da raça parda ou negra. Em relação ao sexo, as mulheres possuem um perfil de desempenho acadêmico mais satisfatório que os homens. A única exceção para este fato foi no curso de Licenciatura em Computação referente à 5ª entrada do vestibular (Base_LC_UAB5), onde os homens tiveram um desempenho acadêmico superior ao das mulheres.

Outros padrões descobertos foram que os alunos que concluíram o ensino médio antes de 2000 e depois de 2010, por exemplo, possuem desempenho acadêmico superior àqueles cujo ano de conclusão do ensino médio foi entre essas datas. Descobriu-se também que os alunos não solteiros tiveram uma performance acadêmica superior quando comparados aos solteiros. A única exceção foi para a base HI_UAB4, onde os solteiros tiveram um melhor desempenho acadêmico. Esses padrões descobertos tornam evidentes que os alunos de classes sociais menos favorecidas possuem desempenho acadêmico inferior àqueles estudantes que não se enquadram nessa situação, principalmente quando o curso possui um grau de dificuldade maior.

Verificou-se que o polo de Valença do Piauí obteve o melhor desempenho acadêmico do curso de Sistemas de Informação, tanto da UAB4, quanto da UAB5 e que o polo de Castelo do Piauí se destacou entre os melhores academicamente nos cursos de História (UAB4 e UAB5) e Licenciatura em Computação (UAB5). Já os dois polos com as melhores performances na base integrada LC_SI_UAB4_UAB5 (Campo Alegre de Lourdes (BA) e Castelo do Piauí) se confirmaram na base LC_UAB5.

Observou-se ainda que os alunos provenientes de escola pública possuem mais dificuldade quando o curso tem um grau de exigência maior, como é o caso dos cursos da área de Exatas, pois de acordo com [Mendes Junior \(2014\)](#) e [Peixoto et al. \(2016\)](#), o déficit na formação básica dos alunos de classes menos favorecidas, principalmente no domínio da matemática, aparece como um possível fator de influência na diferença de desempenho desses alunos em relação àqueles oriundos de escolas privadas.

Considera-se através desse trabalho que esse processo de descoberta de conhecimento pode ser usado por outras bases de dados e também por outros Ambientes Virtuais de Aprendizagem e que os perfis descobertos podem auxiliar os gestores do sistema de educação a distância na tomada de decisões em relação a melhorias no processo de ensino-aprendizagem, já que através da mineração de dados tem-se uma ideia do desempenho do aluno, ao mostrar que a deficiência acadêmica possui correlações com aspectos sociais. Com base nessas informações encontradas, é possível definir-se estratégias diferenciadas em relação a esses alunos, como um acompanhamento presencial constante por parte dos tutores nos polos de apoio do sistema de educação a distância ou até mesmo submeter esses estudantes a um nivelamento antes do início do curso.

7.1 Limitações e Continuidade da Pesquisa

Durante o desenvolvimento do presente trabalho, alguns elementos foram considerados possíveis limitações para a abordagem. Dentre esses elementos, destacam-se:

- A abordagem contemplou somente os cursos de Licenciatura em Computação, Siste-

mas de Informação e História da modalidade a distância da UFPI. Portanto, não foi possível verificar se os perfis descobertos se confirmam para os demais cursos dessa modalidade de ensino;

- O trabalho levou em consideração apenas os exemplos classificados corretamente. Logo, nenhum estudo foi feito em relação aos exemplos incorretamente classificados.

Baseado nas limitações encontradas nessa abordagem, alguns pontos são primordiais para a continuidade da pesquisa. Como trabalho futuro, pretende-se executar as atividades descritas a seguir.

1. Realizar testes com os algoritmos *J48*, *RandomTree* e *SimpleCart* para bases de dados de outros cursos na modalidade a distância da UFPI com o propósito de verificar a compatibilidade dos perfis descobertos;
2. Realizar uma análise dos exemplos incorretamente classificados no trabalho atual para verificar até que ponto eles influenciam nas regras encontradas.
3. Executar testes com outros atributos que também correlacionem o IRA com os aspectos sociais dos alunos.
4. Identificar perfis de alunos da modalidade presencial, a partir de uma correlação entre o IRA e os aspectos sociais dos estudantes.

Referências

- ALDOUS, D. The continuum random tree. i. *The Annals of Probability*, JSTOR, p. 1–28, 1991. Citado na página 23.
- ANAYA, A. R.; BOTICARIO, J. G. Towards improvements on domain-independent measurements for collaborative assessment. In: *Educational Data Mining 2011*. [S.l.: s.n.], 2010. Citado 3 vezes nas páginas 11, 13 e 21.
- ARAUJO, F. *Descoberta de conhecimento em base de dados para o aprendizado da regulação médica/odontológica em operadora de plano de saúde*. Dissertação (Mestrado) — Universidade Federal do Piauí, 2014. Citado 4 vezes nas páginas 13, 20, 21 e 22.
- BALDIRIS, S. et al. Multidimensional adaptations for open learning management systems. In: IEEE. *Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on*. [S.l.], 2008. v. 3, p. 352–356. Citado 3 vezes nas páginas 11, 13 e 21.
- BANFIELD, R. E. et al. A comparison of decision tree ensemble creation techniques. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 29, n. 1, p. 173–180, 2007. Citado na página 21.
- BATISTA, G. E. d. A. P. et al. *Pré-processamento de dados em aprendizado de máquina supervisionado*. Tese (Doutorado) — Universidade de São Paulo, 2003. Citado 4 vezes nas páginas 2, 16, 19 e 26.
- BLUM, A.; MITCHELL, T. Combining labeled and unlabeled data with co-training. In: ACM. *Proceedings of the eleventh annual conference on Computational learning theory*. [S.l.], 1998. p. 92–100. Citado na página 20.
- BORAH, M. D. et al. Application of knowledge based decision technique to predict student enrollment decision. In: IEEE. *Recent Trends in Information Systems (ReTIS), 2011 International Conference on*. [S.l.], 2011. p. 180–184. Citado 3 vezes nas páginas 11, 13 e 21.
- BOUCKAERT, R. R. et al. Weka manual for version 3-6-14. 2016. Citado na página 26.
- BRAGA, R. et al. Ferramentas para desenvolvimento de sistemas baseados em inteligência computacional: Um mapeamento sistemático. *Anais do Simpósio Brasileiro de Automação Inteligente*, v. 12, 2015. Citado na página 7.
- BRASIL. *Decreto nº 5.622, DE 19 de dezembro de 2005, que Regulamenta o art. 80 da Lei nº 9.394, de 20 de dezembro de 1996, o qual estabelece as diretrizes e bases da educação nacional*. 2005. Disponível em: <http://portal.mec.gov.br/seed/arquivos/pdf/dec_5622.pdf>. Acesso em: 15 nov. 2017. Citado na página 1.
- BRASIL. *Decreto nº 5.800, DE 8 DE JUNHO DE 2006, que dispõe sobre o Sistema Universidade Aberta do Brasil*. 2006. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/decreto/d5800.htm>. Acesso em: 10 out. 2017. Citado na página 1.

- BREIMAN, L. et al. *Classification and regression trees*. [S.l.]: CRC press, 1984. Citado 2 vezes nas páginas 18 e 24.
- BURDESCU, D. D.; MIHAESCU, M. C. Enhancing the assessment environment within a learning management systems. In: IEEE. *EUROCON, 2007. The International Conference on "Computer as a Tool"*. [S.l.], 2007. p. 2438–2443. Citado na página 11.
- BURDESCU, D. D.; MIHAESCU, M. C.; LOGOFATU, B. Question recommender with ml business logic. In: IEEE. *Information Technology Interfaces, 2008. ITI 2008. 30th International Conference on*. [S.l.], 2008. p. 639–644. Citado na página 11.
- CARVALHO, R. S. et al. Integração entre o sistema de gestão acadêmica e o sistema de gestão da aprendizagem: identificando necessidades e prototipando requisitos favoráveis a prática docente. *Revista Brasileira de Computação Aplicada*, v. 4, n. 1, p. 81–91, 2011. Citado na página 2.
- CHANDRA, D. G.; RAMAN, A. C. Educational data mining on learning management systems using scorm. In: IEEE. *Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on*. [S.l.], 2014. p. 362–368. Citado na página 11.
- CHIKALOV, I. *Average Time Complexity of Decision Trees*. [S.l.]: Springer, 2011. Citado na página 33.
- COPPIN, B. *Inteligência artificial*. [S.l.]: Grupo Gen-LTC, 2015. Citado na página 19.
- ELMASRI, R.; NAVATHE, S. B.; MORAIS, R. de O. *Sistemas de banco de dados*. Pearson Addison Wesley, 2005. Citado na página 22.
- FACELI, K. et al. *Inteligência artificial: Uma abordagem de aprendizado de máquina*. *Rio de Janeiro: LTC*, v. 2, p. 192, 2011. Citado na página 2.
- FARIAS, S. C. Os benefícios das tecnologias de informação e comunicação (tic) no processo de educação a distância (ead). *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, v. 11, n. 3, p. 15–29, 2013. Citado na página 1.
- FAYYAD, U. M. et al. Knowledge discovery and data mining: Towards a unifying framework. In: *KDD*. [S.l.: s.n.], 1996. v. 96, p. 82–88. Citado 4 vezes nas páginas 13, 15, 27 e 29.
- GOMES, A. K. *Análise do conhecimento extraído de classificadores simbólicos utilizando medidas de avaliação e de interessabilidade*. Tese (Doutorado) — Universidade de São Paulo, 2002. Citado na página 21.
- GUÉRCIO, H. et al. Análise do desempenho estudantil na educação a distância aplicando técnicas de mineração de dados. In: *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*. [S.l.: s.n.], 2014. v. 3, n. 1, p. 641. Citado na página 5.
- GÜNEL, K.; POLAT, R.; KURT, M. Analyzing learning concepts in intelligent tutoring systems. *Int. Arab J. Inf. Technol.*, v. 13, n. 2, p. 281–286, 2016. Citado na página 11.
- HOLZINGER, A.; DEHMER, M.; JURISICA, I. Knowledge discovery and interactive data mining in bioinformatics-state-of-the-art, future challenges and research directions. *BMC bioinformatics*, BioMed Central, v. 15, n. 6, p. I1, 2014. Citado 2 vezes nas páginas 15 e 29.

- JAPKOWICZ, N.; STEPHEN, S. The class imbalance problem: A systematic study. *Intelligent data analysis*, IOS Press, v. 6, n. 5, p. 429–449, 2002. Citado na página 16.
- KITCHENHAM, B.; PICKARD, L.; PFLEEGER, S. L. Case studies for method and tool evaluation. *IEEE software*, IEEE, v. 12, n. 4, p. 52–62, 1995. Citado na página 7.
- LANDIS, J. R.; KOCH, G. G. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, JSTOR, p. 363–374, 1977. Citado 3 vezes nas páginas 15, 25 e 26.
- MAUDES, J. et al. Random feature weights for decision tree ensemble construction. *Information Fusion*, Elsevier, v. 13, n. 1, p. 20–30, 2012. Nenhuma citação no texto.
- MEC/INEP. *NOTAS ESTATÍSTICAS - Censo da Educação Superior 2015*. 2015. Disponível em: <http://download.inep.gov.br/educacao_superior/centso_superior/documentos/2015/Notas_Estatisticas_Censo_Superior_2015.pdf>. Acesso em: 8 nov. 2017. Citado 4 vezes nas páginas 13, 1, 2 e 3.
- MENDES JUNIOR, A. A. F. Uma análise da progressão dos alunos cotistas sob a primeira ação afirmativa brasileira no ensino superior: o caso da universidade do estado do rio de janeiro. *Ensaio: avaliação e políticas públicas em educação*, SciELO Brasil, v. 22, n. 82, p. 31–56, 2014. Citado 5 vezes nas páginas 3, 5, 13, 29 e 58.
- NEVES, R. d. C. D. d. *Pré-processamento no processo de descoberta de conhecimento em banco de dados*. Dissertação (Mestrado), 2003. Citado na página 17.
- PEIXOTO, A. d. L. A. et al. Cotas e desempenho acadêmico na ufba: um estudo a partir dos coeficientes de rendimento. *Avaliação: Revista da Avaliação da Educação Superior*, Universidade de Sorocaba, v. 21, n. 2, p. 569–591, 2016. Citado 5 vezes nas páginas 3, 6, 13, 29 e 58.
- PETERSEN, K. et al. Systematic mapping studies in software engineering. In: *EASE*. [S.l.: s.n.], 2008. v. 8, p. 68–77. Citado na página 7.
- QUILICI-GONZALEZ, J. A.; ZAMPIROLI, F. de A. *Sistemas Inteligentes e Mineração de Dados*. [S.l.: s.n.], 2015. Citado na página 19.
- QUINLAN, J. C4. 5: Programs for machine learning morgan kaufmann publishers san francisco. *CA Google Scholar*, 1993. Citado na página 18.
- QUINLAN, J. R. Induction of decision trees. *Machine learning*, Springer, v. 1, n. 1, p. 81–106, 1986. Citado na página 17.
- QUINLAN, J. R. Decision trees and multivalued attributes. *Machine Intelligence 11*, Oxford University Press, p. 305–318, 1988. Citado na página 18.
- QUINLAN, J. R. Simplifying decision trees. *International Journal of Human-Computer Studies*, Elsevier, v. 51, n. 2, p. 497–510, 1999. Citado na página 21.
- RABELO, H. et al. Utilização de técnicas de mineração de dados educacionais para predição de desempenho de alunos de ead em ambientes virtuais de aprendizagem. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2017. v. 28, n. 1, p. 1527. Citado na página 7.

- RODRIGUEZ, P. M. Processos de ramificação: teoria e aplicações. ICMC-USP, São Carlos, 2012. Citado na página 23.
- RUSSELL, S.; NORVIG, P.; INTELLIGENCE, A. A modern approach. *Artificial Intelligence. Prentice-Hall, Englewood Cliffs*, v. 25, p. 27, 1995. Citado na página 19.
- SALZBERG, S. L. C4. 5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993. *Machine Learning*, Springer, v. 16, n. 3, p. 235–240, 1994. Citado na página 22.
- SILVA, A. M. L. et al. Descoberta de conhecimento através de métodos de aprendizagem de máquina supervisionados aplicados ao sigaa/ufpi. *Revista de Sistemas e Computação-RSC*, v. 7, n. 1, 2017. Citado na página 41.
- SILVA, F. et al. Uma ferramenta para visualização das tendências de evasão geradas por mineração de dados a partir das interações em fóruns de discussão. *Anais temporários do LACLO 2015*, v. 10, n. 1, p. 237, 2015. Citado na página 6.
- SINGH, S.; LAL, S. P. Educational courseware evaluation using machine learning techniques. In: IEEE. *e-Learning, e-Management and e-Services (IC3e), 2013 IEEE Conference on*. [S.l.], 2013. p. 73–78. Citado 3 vezes nas páginas 11, 13 e 21.
- SINGH, S.; LAL, S. P. Using feature selection and association rule mining to evaluate digital courseware. In: IEEE. *ICT and Knowledge Engineering (ICT&KE), 2013 11th International Conference on*. [S.l.], 2013. p. 1–7. Citado 3 vezes nas páginas 11, 13 e 21.
- SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, Elsevier, v. 45, n. 4, p. 427–437, 2009. Citado na página 25.
- THEEND. *Desenvolvida pelo Laboratório de Engenharia de Software e Informática Industrial (Ea-SII), vinculado a Universidade Federal do Piauí*. 2015. Disponível em: <<http://easii.ufpi.br/theend/home>>. Acesso em: 10 mar. 2017. Citado 2 vezes nas páginas 13 e 8.
- VALETTIS, L. M.; NAVARRO, S. B.; GESA, R. F. Modelling collaborative competence level using machine learning techniques. In: *e-Learning*. [S.l.: s.n.], 2008. p. 56–60. Citado na página 11.
- WESTERHOUT, E.; MONACHESI, P. Creating glossaries using pattern-based and machine learning techniques. In: *LREC*. [S.l.: s.n.], 2008. Citado na página 11.
- WITTEN, I.; FRANK, E. Data mining: practical machine learning tools and techniques. *Morgan Kaufmann series in data management systems*, Elsevier, v. 7, p. 47, 2005. Citado na página 22.

Apêndices

APÊNDICE A – Regras de produção geradas através desta pesquisa

A.1 Algoritmo *RandomTree*

```

Escola_publica = Sim
|   Polo = Bom Jesus
|   |   Raca = Branco : Satisfatorio (4/0)
|   |   Raca = Negro : Insatisfatorio (27/0)
|   |   Raca = Pardo : Satisfatorio (0/0)
|   Polo = Inhuma
|   |   Sexo = F
|   |   |   Estado_civil = NaoSolteiro(a) : Satisfatorio (11/0)
|   |   |   Estado_civil = Solteiro(a)
|   |   |   |   Ano_conclusao = Depoisde2010 : Insatisfatorio (5/0)
|   |   |   |   Ano_conclusao = Antesde2000 : Satisfatorio (0/0)
|   |   |   |   Ano_conclusao = De2000a2010 : Satisfatorio (11/0)
|   |   |   Sexo = M : Insatisfatorio (21/0)
|   Polo = Marcos Parente
|   |   Estado_civil = NaoSolteiro(a) : Satisfatorio (4/1)
|   |   Estado_civil = Solteiro(a) : Satisfatorio (19/0)
|   Polo = Pio IX
|   |   Estado_civil = NaoSolteiro(a)
|   |   |   Sexo = F : Satisfatorio (7/0)
|   |   |   Sexo = M : Insatisfatorio (1/0)
|   |   Estado_civil = Solteiro(a)
|   |   |   Ano_conclusao = Depoisde2010 : Insatisfatorio (7/0)
|   |   |   Ano_conclusao = Antesde2000 : Satisfatorio (0/0)
|   |   |   Ano_conclusao = De2000a2010
|   |   |   |   Raca = Branco : Insatisfatorio (2/0)
|   |   |   |   Raca = Negro : Satisfatorio (0/0)
|   |   |   |   Raca = Pardo : Insatisfatorio (28/2)
|   Polo = Piripiri : Insatisfatorio (41/0)
|   Polo = Sao Joao do Piaui
|   |   Raca = Branco : Insatisfatorio (5/0)
|   |   Raca = Negro : Satisfatorio (11/0)
|   |   Raca = Pardo
|   |   |   Estado_civil = NaoSolteiro(a)
|   |   |   |   Ano_conclusao = Depoisde2010 : Satisfatorio (3/0)
|   |   |   |   Ano_conclusao = Antesde2000 : Insatisfatorio (4/0)
|   |   |   |   Ano_conclusao = De2000a2010 : Satisfatorio (0/0)
|   |   |   Estado_civil = Solteiro(a) : Satisfatorio (2/0)
|   Polo = Teresina : Insatisfatorio (37/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte I (Base LC_UAB4).

```

Escola_publica = Nao
| Raca = Branco : Satisfatorio (11/0)
| Raca = Negro : Satisfatorio (9/0)
| Raca = Pardo
| | Ano_conclusao = Depoisde2010 : Satisfatorio (0/0)
| | Ano_conclusao = Antesde2000 : Insatisfatorio (3/0)
| | Ano_conclusao = De2000a2010
| | | Polo = Bom Jesus : Satisfatorio (0/0)
| | | Polo = Inhuma : Satisfatorio (9/0)
| | | Polo = Marcos Parente : Satisfatorio (0/0)
| | | Polo = Pio IX : Satisfatorio (0/0)
| | | Polo = Piripiri : Satisfatorio (0/0)
| | | Polo = Sao Joao do Piaui : Satisfatorio (4/0)
| | | Polo = Teresina : Insatisfatorio (4/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte II (Base LC_UAB4).

```

Ano_conclusao = Depoisde2010
| Polo = Barras : Insatisfatorio (14/0)
| Polo = Campo Maior
| | Escola_publica = Sim
| | | Sexo = F : Satisfatorio (5/0)
| | | Sexo = M : Insatisfatorio (8/1)
| | Escola_publica = Nao : Satisfatorio (2/0)
| Polo = Esperantina : Satisfatorio (0/0)
| Polo = Jaicos
| | Raca = Branco : Satisfatorio (14/0)
| | Raca = Negro : Satisfatorio (0/0)
| | Raca = Pardo
| | | Sexo = F : Satisfatorio (2/0)
| | | Sexo = M : Insatisfatorio (42/0)
| Polo = Oeiras : Insatisfatorio (4/0)
| Polo = Paes Landim : Insatisfatorio (37/0)
| Polo = Regeneracao
| | Sexo = F : Satisfatorio (2/0)
| | Sexo = M
| | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (8/0)
| | | Estado_civil = Solteiro(a) : Satisfatorio (2/0)
| Polo = Valenca do Piaui
| | Estado_civil = NaoSolteiro(a) : Insatisfatorio (3/0)
| | Estado_civil = Solteiro(a) : Satisfatorio (1/0)
| Polo = Uniao : Satisfatorio (1/0)
Ano_conclusao = Antesde2000
| Sexo = F : Insatisfatorio (5/0)
| Sexo = M
| | Escola_publica = Sim
| | | Estado_civil = NaoSolteiro(a)
| | | | Polo = Barras : Satisfatorio (1/0)
| | | | Polo = Campo Maior : Satisfatorio (0/0)
| | | | Polo = Esperantina : Satisfatorio (0/0)
| | | | Polo = Jaicos : Insatisfatorio (9/0)
| | | | Polo = Oeiras : Insatisfatorio (7/0)
| | | | Polo = Paes Landim : Insatisfatorio (7/0)
| | | | Polo = Regeneracao : Satisfatorio (0/0)
| | | | Polo = Valenca do Piaui : Satisfatorio (0/0)
| | | | Polo = Uniao : Insatisfatorio (4/0)
| | | Estado_civil = Solteiro(a) : Insatisfatorio (2/0)
| | Escola_publica = Nao : Satisfatorio (4/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte I (Base SI_UAB4).

```

Ano_conclusao = De2000a2010
|   Raca = Branco
|   |   Sexo = F
|   |   |   Polo = Barras : Satisfatorio (0/0)
|   |   |   Polo = Campo Maior : Satisfatorio (0/0)
|   |   |   Polo = Esperantina : Satisfatorio (0/0)
|   |   |   Polo = Jaicos : Satisfatorio (0/0)
|   |   |   Polo = Oeiras : Satisfatorio (0/0)
|   |   |   Polo = Paes Landim : Satisfatorio (2/0)
|   |   |   Polo = Regeneracao : Insatisfatorio (7/0)
|   |   |   Polo = Valenca do Piaui : Satisfatorio (0/0)
|   |   |   Polo = Uniao : Satisfatorio (0/0)
|   |   Sexo = M
|   |   |   Polo = Barras : Insatisfatorio (10/0)
|   |   |   Polo = Campo Maior : Satisfatorio (0/0)
|   |   |   Polo = Esperantina : Satisfatorio (0/0)
|   |   |   Polo = Jaicos
|   |   |   |   Estado_civil = NaoSolteiro(a) : Insatisfatorio (12/0)
|   |   |   |   Estado_civil = Solteiro(a) : Satisfatorio (9/0)
|   |   |   |   Polo = Oeiras : Insatisfatorio (6/0)
|   |   |   |   Polo = Paes Landim : Satisfatorio (0/0)
|   |   |   |   Polo = Regeneracao : Satisfatorio (0/0)
|   |   |   |   Polo = Valenca do Piaui : Insatisfatorio (1/0)
|   |   |   |   Polo = Uniao : Insatisfatorio (1/0)
|   Raca = Negro : Insatisfatorio (4/0)
|   Raca = Pardo
|   |   Polo = Barras
|   |   |   Escola_publica = Sim
|   |   |   |   Sexo = F : Satisfatorio (2/0)
|   |   |   |   Sexo = M : Insatisfatorio (11/0)
|   |   |   |   Escola_publica = Nao : Satisfatorio (1/0)
|   |   |   Polo = Campo Maior : Satisfatorio (24/0)
|   |   |   Polo = Esperantina : Insatisfatorio (14/0)
|   |   |   Polo = Jaicos
|   |   |   |   Estado_civil = NaoSolteiro(a) : Satisfatorio (16/7)
|   |   |   |   Estado_civil = Solteiro(a) : Insatisfatorio (8/0)
|   |   |   |   Polo = Oeiras : Insatisfatorio (13/2)
|   |   |   |   Polo = Paes Landim : Insatisfatorio (21/0)
|   |   |   |   Polo = Regeneracao : Insatisfatorio (8/0)
|   |   |   |   Polo = Valenca do Piaui
|   |   |   |   |   Sexo = F : Satisfatorio (11/0)
|   |   |   |   |   Sexo = M
|   |   |   |   |   |   Estado_civil = NaoSolteiro(a) : Satisfatorio (5/0)
|   |   |   |   |   |   Estado_civil = Solteiro(a) : Satisfatorio (13/3)
|   |   |   |   |   Polo = Uniao : Insatisfatorio (8/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte II (Base SI_UAB4).

```

Polo = Castelo do Piaui
|   Sexo = F : Satisfatorio (17/0)
|   Sexo = M
|   |   Escola_publica = Nao : Insatisfatorio (11/0)
|   |   Escola_publica = Sim : Insatisfatorio (18/5)
Polo = Esperantina
|   Estado_civil = NaoSolteiro(a) : Insatisfatorio (8/0)
|   Estado_civil = Solteiro(a) : Satisfatorio (31/0)
Polo = Inhumá : Insatisfatorio (29/0)
Polo = Jaicos
|   Estado_civil = NaoSolteiro(a)
|   |   Escola_publica = Nao : Insatisfatorio (2/0)
|   |   Escola_publica = Sim : Satisfatorio (72/0)
|   Estado_civil = Solteiro(a)
|   |   Sexo = F : Insatisfatorio (1/0)
|   |   Sexo = M : Insatisfatorio (20/5)
Polo = Urucui
|   Raca = Branco : Satisfatorio (1/0)
|   Raca = Negro : Insatisfatorio (8/0)
|   Raca = Pardo : Satisfatorio (6/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* (Base HI_UAB4).

```

Polo = Alegrete do Piaui
| Estado_civil = NaoSolteiro(a)
| | Ano_conclusao = Depoisde2010 : Satisfatorio (13/0)
| | Ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | Ano_conclusao = De2000a2010 : Insatisfatorio (8/0)
| Estado_civil = Solteiro(a)
| | Sexo = F : Insatisfatorio (10/0)
| | Sexo = M : Satisfatorio (3/0)
Polo = Campo Alegre de Lourdes : Satisfatorio (47/0)
Polo = Campo Maior
| Ano_conclusao = Depoisde2010 : Satisfatorio (18/0)
| Ano_conclusao = Antesde2000 : Satisfatorio (4/0)
| Ano_conclusao = De2000a2010
| | Escola_publica = Sim : Insatisfatorio (10/0)
| | Escola_publica = Nao : Satisfatorio (5/0)
Polo = Castelo do Piaui : Satisfatorio (23/0)
Polo = Esperantina
| Escola_publica = Sim
| | Sexo = F : Insatisfatorio (1/0)
| | Sexo = M
| | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (11/0)
| | | Estado_civil = Solteiro(a)
| | | | Raca = Branco : Satisfatorio (2/0)
| | | | Raca = Negro : Satisfatorio (0/0)
| | | | Raca = Pardo
| | | | | Ano_conclusao = Depoisde2010 : Satisfatorio (19/1)
| | | | | Ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | | | | Ano_conclusao = De2000a2010 : Satisfatorio (5/0)
| | Escola_publica = Nao : Insatisfatorio (8/0)
Polo = Inhuma
| Ano_conclusao = Depoisde2010 : Insatisfatorio (4/0)
| Ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| Ano_conclusao = De2000a2010
| | Estado_civil = NaoSolteiro(a) : Satisfatorio (3/0)
| | Estado_civil = Solteiro(a) : Insatisfatorio (9/3)
Polo = Marcos Parente
| Estado_civil = NaoSolteiro(a) : Insatisfatorio (4/0)
| Estado_civil = Solteiro(a)
| | Ano_conclusao = Depoisde2010
| | | Sexo = F : Insatisfatorio (10/4)
| | | Sexo = M : Satisfatorio (10/0)
| | Ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | Ano_conclusao = De2000a2010 : Insatisfatorio (9/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte I (Base LC_UAB5).

```

Polo = Picos
|  Sexo = F : Satisfatorio (17/0)
|  Sexo = M
|  |  Ano_conclusao = Depoisde2010 : Insatisfatorio (5/0)
|  |  Ano_conclusao = Antesde2000 : Satisfatorio (2/0)
|  |  Ano_conclusao = De2000a2010 : Insatisfatorio (34/0)
Polo = Paes Landim
|  Sexo = F
|  |  Raca = Branco : Satisfatorio (3/0)
|  |  Raca = Negro : Insatisfatorio (9/0)
|  |  Raca = Pardo : Satisfatorio (0/0)
|  Sexo = M
|  |  Raca = Branco : Satisfatorio (0/0)
|  |  Raca = Negro : Satisfatorio (4/0)
|  |  Raca = Pardo
|  |  |  Ano_conclusao = Depoisde2010 : Satisfatorio (4/2)
|  |  |  Ano_conclusao = Antesde2000 : Satisfatorio (0/0)
|  |  |  Ano_conclusao = De2000a2010 : Satisfatorio (19/0)
Polo = Pio IX
|  Ano_conclusao = Depoisde2010
|  |  Escola_publica = Sim
|  |  |  Raca = Branco : Insatisfatorio (24/6)
|  |  |  Raca = Negro : Satisfatorio (0/0)
|  |  |  Raca = Pardo
|  |  |  |  Sexo = F : Insatisfatorio (19/0)
|  |  |  |  Sexo = M : Satisfatorio (3/0)
|  |  Escola_publica = Nao : Insatisfatorio (4/0)
|  Ano_conclusao = Antesde2000 : Satisfatorio (3/0)
|  Ano_conclusao = De2000a2010
|  |  Sexo = F : Insatisfatorio (2/0)
|  |  Sexo = M : Satisfatorio (16/0)
Polo = Santa Cruz do Piaui
|  Raca = Branco : Satisfatorio (15/4)
|  Raca = Negro : Satisfatorio (7/0)
|  Raca = Pardo
|  |  Estado_civil = NaoSolteiro(a) : Insatisfatorio (3/0)
|  |  Estado_civil = Solteiro(a)
|  |  |  Ano_conclusao = Depoisde2010
|  |  |  |  Sexo = F : Insatisfatorio (15/0)
|  |  |  |  Sexo = M : Satisfatorio (4/0)
|  |  |  Ano_conclusao = Antesde2000 : Satisfatorio (0/0)
|  |  |  Ano_conclusao = De2000a2010 : Insatisfatorio (8/1)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte II (Base LC_UAB5).

```

Polo = SaoJoaodoPiaui
| Ano_conclusao = Depoisde2010
| | Sexo = F : Insatisfatorio (60/0)
| | Sexo = M
| | | Estado_civil = NaoSolteiro(a) : Satisfatorio (11/0)
| | | Estado_civil = Solteiro(a)
| | | | Raca = Branco : Insatisfatorio (1/0)
| | | | Raca = Negro : Insatisfatorio (2/0)
| | | | Raca = Pardo : Insatisfatorio (9/4)
| Ano_conclusao = Antesde2000 : Satisfatorio (6/0)
| Ano_conclusao = De2000a2010
| | Raca = Branco
| | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (12/0)
| | | Estado_civil = Solteiro(a) : Satisfatorio (2/0)
| | Raca = Negro : Satisfatorio (0/0)
| | Raca = Pardo : Satisfatorio (3/0)
Polo = Uniao
| Sexo = F
| | Raca = Branco : Insatisfatorio (2/0)
| | Raca = Negro : Insatisfatorio (3/0)
| | Raca = Pardo
| | | Ano_conclusao = Depoisde2010 : Satisfatorio (4/0)
| | | Ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | | Ano_conclusao = De2000a2010 : Insatisfatorio (23/8)
| Sexo = M
| | Ano_conclusao = Depoisde2010 : Insatisfatorio (2/0)
| | Ano_conclusao = Antesde2000 : Insatisfatorio (1/0)
| | Ano_conclusao = De2000a2010 : Insatisfatorio (20/4)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte III (Base LC_UAB5).

```

Sexo = F
| Ano_conclusao = Depoisde2010
| | Raca = Branco : Insatisfatorio (9/0)
| | Raca = Negro : Insatisfatorio (1/0)
| | Raca = Pardo
| | | Polo = Agua Branca : Satisfatorio (3/0)
| | | Polo = Barras : Satisfatorio (0/0)
| | | Polo = Buriti dos Lopes : Satisfatorio (0/0)
| | | Polo = Cajazeiras do Piaui : Satisfatorio (0/0)
| | | Polo = Canto do Buriti : Satisfatorio (13/0)
| | | Polo = Castelo do Piaui : Insatisfatorio (3/0)
| | | Polo = Campo Maior : Satisfatorio (0/0)
| | | Polo = Elesbao Veloso : Satisfatorio (10/0)
| | | Polo = Esperantina : Satisfatorio (7/0)
| | | Polo = Gilbues : Satisfatorio (0/0)
| | | Polo = Jaicos
| | | | Estado_civil = NaoSolteiro(a) : Satisfatorio (4/0)
| | | | Estado_civil = Solteiro(a) : Insatisfatorio (8/1)
| | | Polo = Oeiras : Satisfatorio (25/5)
| | | Polo = Regeneracao : Satisfatorio (0/0)
| | | Polo = Simoes : Satisfatorio (0/0)
| | | Polo = Simplicio Mendes : Insatisfatorio (4/0)
| | | Polo = Valenca do Piaui : Satisfatorio (0/0)
| | | Polo = Uniao : Satisfatorio (0/0)
| Ano_conclusao = Antesde2000 : Insatisfatorio (11/0)
| Ano_conclusao = De2000a2010
| | Estado_civil = NaoSolteiro(a)
| | | Polo = Agua Branca : Insatisfatorio (2/0)
| | | Polo = Barras : Satisfatorio (0/0)
| | | Polo = Buriti dos Lopes : Insatisfatorio (1/0)
| | | Polo = Cajazeiras do Piaui : Satisfatorio (0/0)
| | | Polo = Canto do Buriti : Satisfatorio (0/0)
| | | Polo = Castelo do Piaui : Satisfatorio (0/0)
| | | Polo = Campo Maior : Satisfatorio (0/0)
| | | Polo = Elesbao Veloso : Satisfatorio (0/0)
| | | Polo = Esperantina : Satisfatorio (0/0)
| | | Polo = Gilbues : Satisfatorio (0/0)
| | | Polo = Jaicos : Satisfatorio (4/0)
| | | Polo = Oeiras : Satisfatorio (0/0)
| | | Polo = Regeneracao : Insatisfatorio (1/0)
| | | Polo = Simoes : Insatisfatorio (3/0)
| | | Polo = Simplicio Mendes : Satisfatorio (7/0)
| | | Polo = Valenca do Piaui : Satisfatorio (0/0)
| | | Polo = Uniao : Satisfatorio (0/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte I (Base SI_UAB5).

```

| | Estado_civil = Solteiro(a)
| | | Raca = Branco
| | | | Polo = Agua Branca : Satisfatorio (0/0)
| | | | Polo = Barras : Satisfatorio (0/0)
| | | | Polo = Buriti dos Lopes : Satisfatorio (0/0)
| | | | Polo = Cajazeiras do Piaui : Satisfatorio (0/0)
| | | | Polo = Canto do Buriti : Satisfatorio (0/0)
| | | | Polo = Castelo do Piaui : Satisfatorio (0/0)
| | | | Polo = Campo Maior : Satisfatorio (0/0)
| | | | Polo = Elesbao Veloso : Satisfatorio (0/0)
| | | | Polo = Esperantina : Insatisfatorio (4/1)
| | | | Polo = Gilbues : Satisfatorio (0/0)
| | | | Polo = Jaicos : Satisfatorio (0/0)
| | | | Polo = Oeiras : Satisfatorio (0/0)
| | | | Polo = Regeneracao : Satisfatorio (0/0)
| | | | Polo = Simoes : Satisfatorio (0/0)
| | | | Polo = Simplicio Mendes : Satisfatorio (0/0)
| | | | Polo = Valenca do Piaui : Satisfatorio (11/0)
| | | | Polo = Uniao : Satisfatorio (0/0)
| | | Raca = Negro : Insatisfatorio (5/0)
| | | Raca = Pardo
| | | | Polo = Agua Branca : Satisfatorio (0/0)
| | | | Polo = Barras : Satisfatorio (0/0)
| | | | Polo = Buriti dos Lopes : Satisfatorio (0/0)
| | | | Polo = Cajazeiras do Piaui : Satisfatorio (0/0)
| | | | Polo = Canto do Buriti : Insatisfatorio (2/0)
| | | | Polo = Castelo do Piaui : Insatisfatorio (6/0)
| | | | Polo = Campo Maior : Satisfatorio (3/0)
| | | | Polo = Elesbao Veloso : Satisfatorio (0/0)
| | | | Polo = Esperantina : Insatisfatorio (18/0)
| | | | Polo = Gilbues : Satisfatorio (0/0)
| | | | Polo = Jaicos : Satisfatorio (0/0)
| | | | Polo = Oeiras : Satisfatorio (0/0)
| | | | Polo = Regeneracao : Insatisfatorio (4/0)
| | | | Polo = Simoes : Satisfatorio (0/0)
| | | | Polo = Simplicio Mendes : Satisfatorio (0/0)
| | | | Polo = Valenca do Piaui : Satisfatorio (0/0)
| | | | Polo = Uniao : Satisfatorio (1/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte II (Base SI_UAB5).

```

Sexo = M
| Polo = Agua Branca
| | Raca = Branco : Insatisfatorio (3/0)
| | Raca = Negro
| | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (5/0)
| | | Estado_civil = Solteiro(a) : Satisfatorio (20/0)
| | Raca = Pardo : Insatisfatorio (28/0)
| Polo = Barras
| | Raca = Branco : Satisfatorio (1/0)
| | Raca = Negro : Insatisfatorio (4/0)
| | Raca = Pardo
| | | Ano_conclusao = Depoisde2010 : Insatisfatorio (29/0)
| | | Ano_conclusao = Antesde2000 : Satisfatorio (1/0)
| | | Ano_conclusao = De2000a2010
| | | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (2/0)
| | | | Estado_civil = Solteiro(a) : Satisfatorio (3/1)
| Polo = Buriti dos Lopes : Insatisfatorio (44/0)
| Polo = Cajazeiras do Piaui
| | Raca = Branco : Insatisfatorio (4/0)
| | Raca = Negro : Insatisfatorio (3/0)
| | Raca = Pardo
| | | Ano_conclusao = Depoisde2010 : Insatisfatorio (17/0)
| | | Ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | | Ano_conclusao = De2000a2010 : Insatisfatorio (12/2)
| Polo = Canto do Buriti
| | Estado_civil = NaoSolteiro(a)
| | | Raca = Branco : Satisfatorio (0/0)
| | | Raca = Negro : Satisfatorio (4/0)
| | | Raca = Pardo
| | | | Ano_conclusao = Depoisde2010 : Insatisfatorio (6/0)
| | | | Ano_conclusao = Antesde2000 : Satisfatorio (21/0)
| | | | Ano_conclusao = De2000a2010 : Satisfatorio (12/1)
| | Estado_civil = Solteiro(a)
| | | Ano_conclusao = Depoisde2010 : Insatisfatorio (29/0)
| | | Ano_conclusao = Antesde2000 : Satisfatorio (2/0)
| | | Ano_conclusao = De2000a2010 : Insatisfatorio (15/0)
| Polo = Castelo do Piaui
| | Raca = Branco : Insatisfatorio (1/0)
| | Raca = Negro : Satisfatorio (3/0)
| | Raca = Pardo : Insatisfatorio (17/0)
| Polo = Campo Maior
| | Ano_conclusao = Depoisde2010 : Insatisfatorio (2/0)
| | Ano_conclusao = Antesde2000
| | | Raca = Branco : Satisfatorio (0/0)
| | | Raca = Negro : Insatisfatorio (1/0)
| | | Raca = Pardo : Satisfatorio (1/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte III (Base SI_UAB5).

```

| | Ano_conclusao = De2000a2010
| | | Escola_publica = Sim : Insatisfatorio (8/0)
| | | Escola_publica = Nao
| | | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (1/0)
| | | | Estado_civil = Solteiro(a) : Satisfatorio (3/0)
| Polo = Elesbao Veloso
| | Ano_conclusao = Depoisde2010
| | | Estado_civil = NaoSolteiro(a) : Satisfatorio (12/0)
| | | Estado_civil = Solteiro(a) : Insatisfatorio (34/0)
| | Ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | Ano_conclusao = De2000a2010 : Satisfatorio (12/0)
| Polo = Esperantina : Insatisfatorio (18/0)
| Polo = Gilbues
| | Escola_publica = Sim
| | | Raca = Branco : Satisfatorio (0/0)
| | | Raca = Negro : Insatisfatorio (5/0)
| | | Raca = Pardo
| | | | Ano_conclusao = Depoisde2010 : Satisfatorio (8/0)
| | | | Ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | | | Ano_conclusao = De2000a2010 : Insatisfatorio (3/0)
| | Escola_publica = Nao : Insatisfatorio (2/0)
| Polo = Jaicos
| | Raca = Branco : Satisfatorio (7/0)
| | Raca = Negro : Satisfatorio (0/0)
| | Raca = Pardo
| | | Escola_publica = Sim
| | | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (7/0)
| | | | Estado_civil = Solteiro(a)
| | | | | Ano_conclusao = Depoisde2010 : Satisfatorio (22/0)
| | | | | Ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | | | | Ano_conclusao = De2000a2010 : Insatisfatorio (2/0)
| | | Escola_publica = Nao : Insatisfatorio (6/0)
| Polo = Oeiras
| | Raca = Branco : Insatisfatorio (3/0)
| | Raca = Negro : Insatisfatorio (14/0)
| | Raca = Pardo
| | | Estado_civil = NaoSolteiro(a)
| | | | Ano_conclusao = Depoisde2010 : Satisfatorio (0/0)
| | | | Ano_conclusao = Antesde2000 : Insatisfatorio (2/0)
| | | | Ano_conclusao = De2000a2010 : Insatisfatorio (3/1)
| | | Estado_civil = Solteiro(a) : Satisfatorio (13/1)
| Polo = Regeneracao
| | Ano_conclusao = Depoisde2010 : Insatisfatorio (15/0)
| | Ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | Ano_conclusao = De2000a2010
| | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (6/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte IV (Base SI_UAB5).

```

| | | Estado_civil = Solteiro(a)
| | | | Raca = Branco : Satisfatorio (0/0)
| | | | Raca = Negro : Insatisfatorio (6/0)
| | | | Raca = Pardo : Satisfatorio (1/0)
| Polo = Simoes
| | Ano_conclusao = Depoisde2010
| | | Escola_publica = Sim : Insatisfatorio (6/2)
| | | Escola_publica = Nao : Insatisfatorio (1/0)
| | Ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | Ano_conclusao = De2000a2010
| | | Raca = Branco
| | | | Estado_civil = NaoSolteiro(a) : Satisfatorio (12/0)
| | | | Estado_civil = Solteiro(a) : Insatisfatorio (1/0)
| | | Raca = Negro : Satisfatorio (0/0)
| | | Raca = Pardo
| | | | Estado_civil = NaoSolteiro(a) : Satisfatorio (11/4)
| | | | Estado_civil = Solteiro(a) : Satisfatorio (1/0)
| Polo = Simplicio Mendes : Insatisfatorio (25/0)
| Polo = Valenca do Piaui
| | Estado_civil = NaoSolteiro(a) : Satisfatorio (4/0)
| | Estado_civil = Solteiro(a)
| | | Ano_conclusao = Depoisde2010
| | | | Raca = Branco : Satisfatorio (2/0)
| | | | Raca = Negro : Satisfatorio (0/0)
| | | | Raca = Pardo : Insatisfatorio (15/5)
| | | Ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | | Ano_conclusao = De2000a2010 : Insatisfatorio (10/0)
| Polo = Uniao
| | Ano_conclusao = Depoisde2010
| | | Escola_publica = Sim : Satisfatorio (1/0)
| | | Escola_publica = Nao
| | | | Raca = Branco : Insatisfatorio (6/0)
| | | | Raca = Negro : Satisfatorio (0/0)
| | | | Raca = Pardo : Satisfatorio (4/0)
| | Ano_conclusao = Antesde2000
| | | Escola_publica = Sim : Satisfatorio (2/0)
| | | Escola_publica = Nao : Insatisfatorio (20/0)
| | Ano_conclusao = De2000a2010
| | | Estado_civil = NaoSolteiro(a) : Satisfatorio (17/1)
| | | Estado_civil = Solteiro(a) : Satisfatorio (2/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte V (Base SI_UAB5).

```

Raca = Branco
| Polo = Redencao do Gurgueia : Satisfatorio (0/0)
| Polo = Santa Cruz do Piaui
| | Segundo_grau_ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | Segundo_grau_ano_conclusao = Depoisde2010 : Satisfatorio (4/0)
| | Segundo_grau_ano_conclusao = de2000a2010 : Insatisfatorio (2/0)
| Polo = Sao Jose do Peixe : Satisfatorio (0/0)
| Polo = Castelo do Piaui : Satisfatorio (3/0)
| Polo = Corrente : Satisfatorio (1/0)
| Polo = Esperantina : Insatisfatorio (32/0)
| Polo = Floriano : Insatisfatorio (2/0)
| Polo = Inhumas : Satisfatorio (37/0)
| Polo = Jaicos : Insatisfatorio (6/0)
| Polo = Urucui : Insatisfatorio (2/0)
Raca = Negro : Satisfatorio (46/0)
Raca = Pardo
| Polo = Redencao do Gurgueia
| | Escola_publica = Nao : Insatisfatorio (5/0)
| | Escola_publica = Sim
| | | Sexo = F : Satisfatorio (38/0)
| | | Sexo = M
| | | | Segundo_grau_ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | | | Segundo_grau_ano_conclusao = Depoisde2010 : Insatisfatorio (11/4)
| | | | Segundo_grau_ano_conclusao = de2000a2010 : Satisfatorio (7/0)
| Polo = Santa Cruz do Piaui
| | Segundo_grau_ano_conclusao = Antesde2000 : Satisfatorio (10/0)
| | Segundo_grau_ano_conclusao = Depoisde2010 : Insatisfatorio (14/0)
| | Segundo_grau_ano_conclusao = de2000a2010 : Satisfatorio (5/0)
| Polo = Sao Jose do Peixe
| | Segundo_grau_ano_conclusao = Antesde2000 : Satisfatorio (2/0)
| | Segundo_grau_ano_conclusao = Depoisde2010 : Satisfatorio (27/11)
| | Segundo_grau_ano_conclusao = de2000a2010 : Satisfatorio (15/0)
| Polo = Castelo do Piaui
| | Sexo = F
| | | Segundo_grau_ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | | Segundo_grau_ano_conclusao = Depoisde2010 : Satisfatorio (8/0)
| | | Segundo_grau_ano_conclusao = de2000a2010 : Satisfatorio (5/1)
| | Sexo = M
| | | Escola_publica = Nao : Satisfatorio (6/0)
| | | Escola_publica = Sim : Insatisfatorio (5/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte I (Base HI_UAB5).

```

| Polo = Corrente
| | Sexo = F
| | | Segundo_grau_ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | | Segundo_grau_ano_conclusao = Depoisde2010 : Insatisfatorio (4/0)
| | | Segundo_grau_ano_conclusao = de2000a2010 : Satisfatorio (1/0)
| | Sexo = M : Satisfatorio (4/0)
| Polo = Esperantina
| | Segundo_grau_ano_conclusao = Antesde2000 : Insatisfatorio (1/0)
| | Segundo_grau_ano_conclusao = Depoisde2010 : Satisfatorio (6/0)
| | Segundo_grau_ano_conclusao = de2000a2010 : Satisfatorio (11/0)
| Polo = Florianopolis
| | Segundo_grau_ano_conclusao = Antesde2000 : Satisfatorio (18/0)
| | Segundo_grau_ano_conclusao = Depoisde2010 : Satisfatorio (0/0)
| | Segundo_grau_ano_conclusao = de2000a2010 : Insatisfatorio (4/0)
| Polo = Inhumas
| | Estado_civil = NaoSolteiro(a) : Satisfatorio (6/0)
| | Estado_civil = Solteiro(a)
| | | Sexo = F
| | | | Segundo_grau_ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | | | Segundo_grau_ano_conclusao = Depoisde2010 : Insatisfatorio (17/0)
| | | | Segundo_grau_ano_conclusao = de2000a2010 : Satisfatorio (4/0)
| | | Sexo = M : Satisfatorio (6/0)
| Polo = Jaicós
| | Estado_civil = NaoSolteiro(a) : Satisfatorio (24/0)
| | Estado_civil = Solteiro(a)
| | | Escola_publica = Nao : Satisfatorio (5/0)
| | | Escola_publica = Sim
| | | | Segundo_grau_ano_conclusao = Antesde2000 : Satisfatorio (0/0)
| | | | Segundo_grau_ano_conclusao = Depoisde2010 : Satisfatorio (9/0)
| | | | Segundo_grau_ano_conclusao = de2000a2010 : Insatisfatorio (11/3)
| Polo = Urucui
| | Estado_civil = NaoSolteiro(a)
| | | Segundo_grau_ano_conclusao = Antesde2000 : Insatisfatorio (4/0)
| | | Segundo_grau_ano_conclusao = Depoisde2010 : Insatisfatorio (16/4)
| | | Segundo_grau_ano_conclusao = de2000a2010
| | | | Sexo = F : Satisfatorio (14/0)
| | | | Sexo = M : Insatisfatorio (22/10)
| | Estado_civil = Solteiro(a) : Satisfatorio (13/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte II (Base HI_UAB5).

```

Raca = Branco
| Polo = Agua Branca : Insatisfatorio (1/0)
| Polo = Alegrete do Piaui : Satisfatorio (11/0)
| Polo = Barras
| | Estado_civil = Solteiro(a) : Insatisfatorio (5/0)
| | Estado_civil = NaoSolteiro(a) : Satisfatorio (5/0)
| Polo = Bom Jesus : Insatisfatorio (7/0)
| Polo = Buriti dos Lopes : Insatisfatorio (9/0)
| Polo = Cajazeiras do Piaui : Satisfatorio (0/0)
| Polo = Campo Alegre de Lourdes : Satisfatorio (11/4)
| Polo = Campo Maior
| | Conclusao_escola_publica = Sim : Satisfatorio (11/0)
| | Conclusao_escola_publica = Nao : Insatisfatorio (24/0)
| Polo = Canto do Buriti : Satisfatorio (11/0)
| Polo = Castelo do Piaui : Insatisfatorio (28/0)
| Polo = Elesbao Veloso : Satisfatorio (1/0)
| Polo = Esperantina
| | Estado_civil = Solteiro(a) : Satisfatorio (6/0)
| | Estado_civil = NaoSolteiro(a) : Insatisfatorio (2/0)
| Polo = Gilbues : Insatisfatorio (1/0)
| Polo = Inhumas
| | Sexo = F : Satisfatorio (8/0)
| | Sexo = M : Insatisfatorio (7/0)
| Polo = Jaicos
| | Estado_civil = Solteiro(a) : Satisfatorio (3/0)
| | Estado_civil = NaoSolteiro(a) : Satisfatorio (6/2)
| Polo = Marcos Parente
| | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | Ano_conclusao_segundo_grau = De2000a2010 : Insatisfatorio (6/0)
| | Ano_conclusao_segundo_grau = Depoisde2010 : Satisfatorio (1/0)
| Polo = Oeiras : Insatisfatorio (6/0)
| Polo = Paes Landim
| | Curso = Computacao : Satisfatorio (15/0)
| | Curso = Sistemas de Informacao
| | | Estado_civil = Solteiro(a) : Insatisfatorio (7/0)
| | | Estado_civil = NaoSolteiro(a) : Satisfatorio (3/0)
| Polo = Picos : Insatisfatorio (6/0)
| Polo = Pio IX
| | Sexo = F
| | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | Ano_conclusao_segundo_grau = De2000a2010
| | | | Estado_civil = Solteiro(a) : Insatisfatorio (1/0)
| | | | Estado_civil = NaoSolteiro(a) : Satisfatorio (9/0)
| | | Ano_conclusao_segundo_grau = Depoisde2010 : Insatisfatorio (7/0)
| | Sexo = M : Insatisfatorio (10/0)
| Polo = Piripiri : Insatisfatorio (10/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte I (Base LC_SI_UAB4_UAB5).

```

| Polo = Regeneracao : Insatisfatorio (18/0)
| Polo = Santa Cruz do Piaui : Satisfatorio (0/0)
| Polo = Simplicio Mendes : Satisfatorio (7/0)
| Polo = Simoes
| | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | Ano_conclusao_segundo_grau = De2000a2010 : Insatisfatorio (2/0)
| | Ano_conclusao_segundo_grau = Depoisde2010
| | | Estado_civil = Solteiro(a) : Satisfatorio (7/0)
| | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (8/0)
| Polo = Sao Joao do Piaui : Satisfatorio (0/0)
| Polo = Teresina : Insatisfatorio (1/0)
| Polo = Uniao
| | Curso = Computacao : Insatisfatorio (8/0)
| | Curso = Sistemas de Informacao
| | | Conclusao_escola_publica = Sim : Insatisfatorio (7/0)
| | | Conclusao_escola_publica = Nao : Satisfatorio (4/0)
| Polo = Valenca do Piaui
| | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | Ano_conclusao_segundo_grau = De2000a2010 : Satisfatorio (4/0)
| | Ano_conclusao_segundo_grau = Depoisde2010 : Satisfatorio (12/4)
Raca = Negro
| Polo = Agua Branca : Satisfatorio (2/0)
| Polo = Alegrete do Piaui : Satisfatorio (11/0)
| Polo = Barras : Insatisfatorio (10/0)
| Polo = Bom Jesus : Insatisfatorio (9/0)
| Polo = Buriti dos Lopes : Insatisfatorio (5/0)
| Polo = Cajazeiras do Piaui : Insatisfatorio (28/0)
| Polo = Campo Alegre de Lourdes : Satisfatorio (1/0)
| Polo = Campo Maior
| | Sexo = F : Satisfatorio (4/0)
| | Sexo = M : Insatisfatorio (13/0)
| Polo = Canto do Buriti
| | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | Ano_conclusao_segundo_grau = De2000a2010 : Insatisfatorio (4/0)
| | Ano_conclusao_segundo_grau = Depoisde2010 : Satisfatorio (16/0)
| Polo = Castelo do Piaui : Satisfatorio (5/0)
| Polo = Elesbao Veloso : Satisfatorio (0/0)
| Polo = Esperantina : Insatisfatorio (1/0)
| Polo = Gilbues : Satisfatorio (0/0)
| Polo = Inhumas : Satisfatorio (24/0)
| Polo = Jaicos : Insatisfatorio (10/0)
| Polo = Marcos Parente
| | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | Ano_conclusao_segundo_grau = De2000a2010 : Satisfatorio (12/0)
| | Ano_conclusao_segundo_grau = Depoisde2010 : Insatisfatorio (5/0)
| Polo = Oeiras : Insatisfatorio (2/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte II (Base LC_SI_UAB4_UAB5).

```

| Polo = Paes Landim
| | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (9/0)
| | Ano_conclusao_segundo_grau = De2000a2010 : Insatisfatorio (6/0)
| | Ano_conclusao_segundo_grau = Depoisde2010 : Insatisfatorio (3/0)
| Polo = Picos : Insatisfatorio (1/0)
| Polo = Pio IX : Insatisfatorio (3/0)
| Polo = Piripiri : Satisfatorio (0/0)
| Polo = Regeneracao
| | Estado_civil = Solteiro(a)
| | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | Ano_conclusao_segundo_grau = De2000a2010 : Insatisfatorio (19/0)
| | | Ano_conclusao_segundo_grau = Depoisde2010
| | | | Sexo = F : Insatisfatorio (3/0)
| | | | Sexo = M : Satisfatorio (29/6)
| | Estado_civil = NaoSolteiro(a) : Insatisfatorio (14/0)
| Polo = Santa Cruz do Piaui : Insatisfatorio (3/0)
| Polo = Simplicio Mendes : Satisfatorio (0/0)
| Polo = Simoes : Satisfatorio (5/0)
| Polo = Sao Joao do Piaui
| | Estado_civil = Solteiro(a) : Satisfatorio (11/0)
| | Estado_civil = NaoSolteiro(a) : Insatisfatorio (19/0)
| Polo = Teresina : Insatisfatorio (7/0)
| Polo = Uniao
| | Curso = Computacao : Insatisfatorio (6/0)
| | Curso = Sistemas de Informacao
| | | Conclusao_escola_publica = Sim
| | | | Estado_civil = Solteiro(a)
| | | | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | | | Ano_conclusao_segundo_grau = De2000a2010 : Satisfatorio (14/2)
| | | | | Ano_conclusao_segundo_grau = Depoisde2010 : Satisfatorio (6/0)
| | | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (6/0)
| | | Conclusao_escola_publica = Nao : Satisfatorio (6/0)
| Polo = Valenca do Piaui : Satisfatorio (1/0)
Raca = Pardo
| Polo = Agua Branca
| | Estado_civil = Solteiro(a)
| | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | Ano_conclusao_segundo_grau = De2000a2010 : Insatisfatorio (6/0)
| | | Ano_conclusao_segundo_grau = Depoisde2010
| | | | Sexo = F : Satisfatorio (7/0)
| | | | Sexo = M : Insatisfatorio (18/7)
| | Estado_civil = NaoSolteiro(a) : Insatisfatorio (23/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte III (Base LC_SI_UAB4_UAB5).

```

| Polo = Alegrete do Piaui
| | Sexo = F
| | | Estado_civil = Solteiro(a)
| | | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | | Ano_conclusao_segundo_grau = De2000a2010 : Satisfatorio (17/8)
| | | | Ano_conclusao_segundo_grau = Depoisde2010 : Insatisfatorio (6/0)
| | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (12/0)
| | Sexo = M : Satisfatorio (12/0)
| Polo = Barras
| | Sexo = F
| | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (6/3)
| | | Ano_conclusao_segundo_grau = De2000a2010 : Insatisfatorio (7/0)
| | | Ano_conclusao_segundo_grau = Depoisde2010 : Insatisfatorio (4/0)
| | Sexo = M
| | | Conclusao_escola_publica = Sim
| | | | Estado_civil = Solteiro(a)
| | | | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | | | Ano_conclusao_segundo_grau = De2000a2010 : Insatisfatorio (9/1)
| | | | | Ano_conclusao_segundo_grau = Depoisde2010 : Satisfatorio (11/4)
| | | | Estado_civil = NaoSolteiro(a)
| | | | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (10/3)
| | | | | Ano_conclusao_segundo_grau = De2000a2010 : Satisfatorio (18/8)
| | | | | Ano_conclusao_segundo_grau = Depoisde2010 : Satisfatorio (0/0)
| | | Conclusao_escola_publica = Nao
| | | | Ano_conclusao_segundo_grau = Antesde2000 : Insatisfatorio (2/0)
| | | | Ano_conclusao_segundo_grau = De2000a2010
| | | | | Estado_civil = Solteiro(a) : Satisfatorio (3/0)
| | | | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (9/0)
| | | | Ano_conclusao_segundo_grau = Depoisde2010 : Satisfatorio (0/0)
| Polo = Bom Jesus
| | Sexo = F : Insatisfatorio (13/0)
| | Sexo = M
| | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (8/0)
| | | Ano_conclusao_segundo_grau = De2000a2010 : Insatisfatorio (22/0)
| | | Ano_conclusao_segundo_grau = Depoisde2010 : Satisfatorio (0/0)
| Polo = Buriti dos Lopes
| | Conclusao_escola_publica = Sim
| | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | Ano_conclusao_segundo_grau = De2000a2010 : Satisfatorio (2/0)
| | | Ano_conclusao_segundo_grau = Depoisde2010 : Insatisfatorio (25/0)
| | Conclusao_escola_publica = Nao : Insatisfatorio (8/0)
| Polo = Cajazeiras do Piaui
| | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | Ano_conclusao_segundo_grau = De2000a2010
| | | Estado_civil = Solteiro(a) : Insatisfatorio (10/0)
| | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (5/2)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte IV (Base LC_SI_UAB4_UAB5).

```

| | Ano_conclusao_segundo_grau = Depoisde2010 : Insatisfatorio (21/0)
| Polo = Campo Alegre de Lourdes
| | Estado_civil = Solteiro(a)
| | | Sexo = F : Insatisfatorio (7/2)
| | | Sexo = M : Satisfatorio (10/4)
| | Estado_civil = NaoSolteiro(a) : Insatisfatorio (20/0)
| Polo = Campo Maior
| | Curso = Computacao
| | | Sexo = F : Satisfatorio (9/0)
| | | Sexo = M
| | | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | | Ano_conclusao_segundo_grau = De2000a2010
| | | | | Estado_civil = Solteiro(a) : Satisfatorio (5/0)
| | | | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (10/1)
| | | | | Ano_conclusao_segundo_grau = Depoisde2010 : Insatisfatorio (7/0)
| | Curso = Sistemas de Informacao
| | | Sexo = F
| | | | Estado_civil = Solteiro(a) : Satisfatorio (2/0)
| | | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (4/0)
| | | Sexo = M : Insatisfatorio (20/0)
| Polo = Canto do Buriti
| | Sexo = F : Insatisfatorio (1/0)
| | Sexo = M : Satisfatorio (20/0)
| Polo = Castelo do Piaui
| | Conclusao_escola_publica = Sim
| | | Estado_civil = Solteiro(a)
| | | | Sexo = F
| | | | | Curso = Computacao : Insatisfatorio (15/5)
| | | | | Curso = Sistemas de Informacao : Satisfatorio (4/0)
| | | | | Sexo = M
| | | | | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | | | | Ano_conclusao_segundo_grau = De2000a2010
| | | | | | | Curso = Computacao : Satisfatorio (4/0)
| | | | | | | Curso = Sistemas de Informacao : Insatisfatorio (3/0)
| | | | | | | Ano_conclusao_segundo_grau = Depoisde2010
| | | | | | | | Curso = Computacao : Satisfatorio (18/0)
| | | | | | | | Curso = Sistemas de Informacao : Satisfatorio (15/4)
| | | | | Estado_civil = NaoSolteiro(a)
| | | | | | Curso = Computacao
| | | | | | | Sexo = F : Insatisfatorio (2/0)
| | | | | | | Sexo = M : Satisfatorio (3/0)
| | | | | | | Curso = Sistemas de Informacao : Insatisfatorio (9/0)
| | Conclusao_escola_publica = Nao : Satisfatorio (11/0)
| Polo = Elesbao Veloso
| | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | Ano_conclusao_segundo_grau = De2000a2010

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte V (Base LC_SI_UAB4_UAB5).

```

| | | Estado_civil = Solteiro(a) : Satisfatorio (3/0)
| | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (2/0)
| | Ano_conclusao_segundo_grau = Depoisde2010
| | | Sexo = F : Satisfatorio (1/0)
| | | Sexo = M
| | | | Estado_civil = Solteiro(a) : Satisfatorio (32/4)
| | | | Estado_civil = NaoSolteiro(a) : Satisfatorio (10/0)
| Polo = Esperantina
| | Curso = Computacao
| | | Sexo = F : Satisfatorio (1/0)
| | | Sexo = M
| | | | Estado_civil = Solteiro(a) : Insatisfatorio (2/0)
| | | | Estado_civil = NaoSolteiro(a) : Satisfatorio (9/3)
| | Curso = Sistemas de Informacao
| | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | Ano_conclusao_segundo_grau = De2000a2010 : Insatisfatorio (27/0)
| | | Ano_conclusao_segundo_grau = Depoisde2010
| | | | Sexo = F : Insatisfatorio (3/0)
| | | | Sexo = M : Insatisfatorio (29/13)
| Polo = Gilbues : Insatisfatorio (7/0)
| Polo = Inhuma
| | Sexo = F
| | | Estado_civil = Solteiro(a)
| | | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | | Ano_conclusao_segundo_grau = De2000a2010 : Satisfatorio (8/0)
| | | | Ano_conclusao_segundo_grau = Depoisde2010 : Insatisfatorio (2/0)
| | | Estado_civil = NaoSolteiro(a)
| | | | Ano_conclusao_segundo_grau = Antesde2000 : Insatisfatorio (6/0)
| | | | Ano_conclusao_segundo_grau = De2000a2010 : Satisfatorio (2/0)
| | | | Ano_conclusao_segundo_grau = Depoisde2010 : Satisfatorio (0/0)
| | Sexo = M
| | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | Ano_conclusao_segundo_grau = De2000a2010
| | | | Estado_civil = Solteiro(a) : Satisfatorio (17/4)
| | | | Estado_civil = NaoSolteiro(a) : Satisfatorio (2/0)
| | | Ano_conclusao_segundo_grau = Depoisde2010 : Insatisfatorio (5/1)
| Polo = Jaicos : Insatisfatorio (71/0)
| Polo = Marcos Parente
| | Estado_civil = Solteiro(a)
| | | Sexo = F : Satisfatorio (10/2)
| | | Sexo = M
| | | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | | Ano_conclusao_segundo_grau = De2000a2010 : Insatisfatorio (2/0)
| | | | Ano_conclusao_segundo_grau = Depoisde2010 : Satisfatorio (5/0)
| | Estado_civil = NaoSolteiro(a)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte VI (Base LC_SI_UAB4_UAB5).

```

| | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | Ano_conclusao_segundo_grau = De2000a2010
| | | | Conclusao_escola_publica = Sim : Insatisfatorio (5/0)
| | | | Conclusao_escola_publica = Nao : Satisfatorio (4/0)
| | | Ano_conclusao_segundo_grau = Depoisde2010 : Insatisfatorio (6/0)
| Polo = Oeiras
| | Estado_civil = Solteiro(a)
| | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | Ano_conclusao_segundo_grau = De2000a2010 : Insatisfatorio (5/0)
| | | Ano_conclusao_segundo_grau = Depoisde2010 : Satisfatorio (32/7)
| | Estado_civil = NaoSolteiro(a)
| | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (4/0)
| | | Ano_conclusao_segundo_grau = De2000a2010 : Insatisfatorio (8/0)
| | | Ano_conclusao_segundo_grau = Depoisde2010 : Satisfatorio (0/0)
| Polo = Paes Landim
| | Sexo = F
| | | Estado_civil = Solteiro(a)
| | | | Curso = Computacao : Satisfatorio (1/0)
| | | | Curso = Sistemas de Informacao : Insatisfatorio (5/0)
| | | Estado_civil = NaoSolteiro(a)
| | | | Curso = Computacao : Satisfatorio (4/0)
| | | | Curso = Sistemas de Informacao : Insatisfatorio (4/0)
| | Sexo = M
| | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | Ano_conclusao_segundo_grau = De2000a2010
| | | | Estado_civil = Solteiro(a)
| | | | | Conclusao_escola_publica = Sim
| | | | | | Curso = Computacao : Satisfatorio (4/0)
| | | | | | Curso = Sistemas de Informacao : Insatisfatorio (18/0)
| | | | | Conclusao_escola_publica = Nao : Insatisfatorio (4/0)
| | | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (1/0)
| | | Ano_conclusao_segundo_grau = Depoisde2010 : Insatisfatorio (11/0)
| Polo = Picos
| | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | Ano_conclusao_segundo_grau = De2000a2010
| | | Conclusao_escola_publica = Sim : Insatisfatorio (1/0)
| | | Conclusao_escola_publica = Nao : Satisfatorio (9/0)
| | Ano_conclusao_segundo_grau = Depoisde2010 : Insatisfatorio (4/0)
| Polo = Pio IX
| | Estado_civil = Solteiro(a) : Insatisfatorio (19/0)
| | Estado_civil = NaoSolteiro(a)
| | | Sexo = F : Insatisfatorio (9/0)
| | | Sexo = M : Satisfatorio (29/7)
| Polo = Piripiri : Insatisfatorio (31/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte VII (Base LC_SI_UAB4_UAB5).

```

| Polo = Piripiri : Insatisfatorio (31/0)
| Polo = Regeneracao
| | Ano_conclusao_segundo_grau = Antesde2000 : Insatisfatorio (4/0)
| | Ano_conclusao_segundo_grau = De2000a2010
| | | Estado_civil = Solteiro(a)
| | | | Conclusao_escola_publica = Sim
| | | | | Sexo = F : Insatisfatorio (4/0)
| | | | | Sexo = M : Satisfatorio (1/0)
| | | | Conclusao_escola_publica = Nao : Insatisfatorio (5/0)
| | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (23/0)
| | Ano_conclusao_segundo_grau = Depoisde2010 : Insatisfatorio (30/0)
| Polo = Santa Cruz do Piaui : Satisfatorio (13/0)
| Polo = Simplicio Mendes : Insatisfatorio (3/0)
| Polo = Simoes
| | Sexo = F
| | | Estado_civil = Solteiro(a) : Satisfatorio (2/0)
| | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (1/0)
| | Sexo = M : Insatisfatorio (21/0)
| Polo = Sao Joao do Piaui
| | Ano_conclusao_segundo_grau = Antesde2000
| | | Sexo = F : Insatisfatorio (15/0)
| | | Sexo = M : Satisfatorio (11/0)
| | Ano_conclusao_segundo_grau = De2000a2010
| | | Estado_civil = Solteiro(a)
| | | | Sexo = F : Insatisfatorio (23/11)
| | | | Sexo = M : Insatisfatorio (35/3)
| | | Estado_civil = NaoSolteiro(a) : Insatisfatorio (4/0)
| | Ano_conclusao_segundo_grau = Depoisde2010
| | | Sexo = F : Satisfatorio (2/0)
| | | Sexo = M
| | | | Conclusao_escola_publica = Sim
| | | | | Estado_civil = Solteiro(a) : Insatisfatorio (7/0)
| | | | | Estado_civil = NaoSolteiro(a) : Satisfatorio (6/0)
| | | | Conclusao_escola_publica = Nao : Satisfatorio (6/0)
| Polo = Teresina
| | Estado_civil = Solteiro(a)
| | | Conclusao_escola_publica = Sim : Satisfatorio (3/0)
| | | Conclusao_escola_publica = Nao : Insatisfatorio (2/0)
| | Estado_civil = NaoSolteiro(a) : Insatisfatorio (22/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte VIII (Base LC_SI_UAB4_UAB5).

```

| Polo = Uniao
| | Estado_civil = Solteiro(a)
| | | Ano_conclusao_segundo_grau = Antesde2000 : Insatisfatorio (14/0)
| | | Ano_conclusao_segundo_grau = De2000a2010
| | | | Curso = Computacao : Satisfatorio (13/5)
| | | | Curso = Sistemas de Informacao : Satisfatorio (2/0)
| | | Ano_conclusao_segundo_grau = Depoisde2010
| | | | Curso = Computacao : Satisfatorio (4/0)
| | | | Curso = Sistemas de Informacao : Insatisfatorio (1/0)
| | Estado_civil = NaoSolteiro(a)
| | | Sexo = F : Satisfatorio (10/0)
| | | Sexo = M
| | | | Ano_conclusao_segundo_grau = Antesde2000
| | | | | Conclusao_escola_publica = Sim : Satisfatorio (14/0)
| | | | | Conclusao_escola_publica = Nao : Insatisfatorio (12/0)
| | | | Ano_conclusao_segundo_grau = De2000a2010
| | | | | Conclusao_escola_publica = Sim : Satisfatorio (9/4)
| | | | | Conclusao_escola_publica = Nao : Satisfatorio (25/0)
| | | | Ano_conclusao_segundo_grau = Depoisde2010 : Satisfatorio (0/0)
| Polo = Valenca do Piaui
| | Sexo = F : Insatisfatorio (3/0)
| | Sexo = M
| | | Estado_civil = Solteiro(a)
| | | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | | Ano_conclusao_segundo_grau = De2000a2010 : Insatisfatorio (20/0)
| | | | Ano_conclusao_segundo_grau = Depoisde2010 : Satisfatorio (22/11)
| | | Estado_civil = NaoSolteiro(a)
| | | | Ano_conclusao_segundo_grau = Antesde2000 : Satisfatorio (0/0)
| | | | Ano_conclusao_segundo_grau = De2000a2010 : Insatisfatorio (8/2)
| | | | Ano_conclusao_segundo_grau = Depoisde2010 : Insatisfatorio (2/0)

```

Regras de produção geradas pelo algoritmo *RandomTree* - Parte IX (Base LC_SI_UAB4_UAB5).

A.2 Algoritmo SimpleCart

```

Polo=(Piripiri)|(Bom Jesus)|(Pio IX)|(Teresina)|(Inhuma)
| Sexo=(M)
| | Escola_publica=(Sim): Insatisfatorio(129.0/5.0)
| | Escola_publica!=(Sim)
| | | Polo=(Teresina)|(Bom Jesus)|(Marcos Parente)|(Pio IX)|(Piripiri)|(Sao Joao do Piaui): Insatisfatorio(7.0/0.0)
| | | Polo!=(Teresina)|(Bom Jesus)|(Marcos Parente)|(Pio IX)|(Piripiri)|(Sao Joao do Piaui): Satisfatorio(7.0/0.0)
| Sexo!=(M)
| | Ano_conclusao=(Depoisde2010): Insatisfatorio(16.0/0.0)
| | Ano_conclusao!=(Depoisde2010)
| | | Polo=(Piripiri)|(Bom Jesus)|(Pio IX)
| | | | Raca=(Pardo): Insatisfatorio(11.0/0.0)
| | | | Raca!=(Pardo)
| | | | Estado_civil=(Solteiro(a)): Insatisfatorio(5.0/2.0)
| | | | Estado_civil!=(Solteiro(a)): Satisfatorio(8.0/0.0)
| | | Polo!=(Piripiri)|(Bom Jesus)|(Pio IX): Satisfatorio(34.0/0.0)
Polo!=(Piripiri)|(Bom Jesus)|(Pio IX)|(Teresina)|(Inhuma)
| Estado_civil=(NaoSolteiro(a))
| | Ano_conclusao=(Antesde2000): Insatisfatorio(4.0/0.0)
| | Ano_conclusao!=(Antesde2000)
| | | Sexo=(M)
| | | | Polo=(Sao Joao do Piaui): Insatisfatorio(6.0/0.0)
| | | | Polo!=(Sao Joao do Piaui)
| | | | | Ano_conclusao=(Depoisde2010): Insatisfatorio(3.0/0.0)
| | | | | Ano_conclusao!=(Depoisde2010): Satisfatorio(6.0/1.0)
| | | Sexo!=(M): Satisfatorio(4.0/0.0)
| Estado_civil!=(NaoSolteiro(a))
| | Ano_conclusao=(Depoisde2010)
| | | Sexo=(M): Insatisfatorio(2.0/0.0)
| | | Sexo!=(M): Satisfatorio(2.0/0.0)
| | Ano_conclusao!=(Depoisde2010): Satisfatorio(38.0/0.0)

```

Regras de produção geradas pelo algoritmo *SimpleCart* (Base LC_UAB4).

As regras de produção geradas pelo algoritmo *SimpleCart* para as bases LC_UAB5, SI_UAB5, HI_UAB5 e LC_SI_UAB4_UAB5 não puderam ser mostradas em virtude de ter uma largura muito grande e as regras não aparecerem de forma completa.

```

Polo=(Piripiri)|(Bom Jesus)|(Pio IX)|(Teresina)|(Inhuma)
| Sexo=(M)
| | Escola_publica=(Sim): Insatisfatorio(129.0/5.0)
| | Escola_publica!=(Sim)
| | | Polo=(Teresina)|(Bom Jesus)|(Marcos Parente)|(Pio IX)|(Piripiri)|(Sao Joao do Piaui): Insatisfatorio(7.0/0.0)
| | | Polo!=(Teresina)|(Bom Jesus)|(Marcos Parente)|(Pio IX)|(Piripiri)|(Sao Joao do Piaui): Satisfatorio(7.0/0.0)
| | Sexo!=(M)
| | Ano_conclusao=(Depoisde2010): Insatisfatorio(16.0/0.0)
| | Ano_conclusao!=(Depoisde2010)
| | | Polo=(Piripiri)|(Bom Jesus)|(Pio IX)
| | | | Raca=(Pardo): Insatisfatorio(11.0/0.0)
| | | | Raca!=(Pardo)
| | | | | Estado_civil=(Solteiro(a)): Insatisfatorio(5.0/2.0)
| | | | | Estado_civil!=(Solteiro(a)): Satisfatorio(8.0/0.0)
| | | | Polo!=(Piripiri)|(Bom Jesus)|(Pio IX): Satisfatorio(34.0/0.0)
Polo!=(Piripiri)|(Bom Jesus)|(Pio IX)|(Teresina)|(Inhuma)
| Estado_civil=(NaoSolteiro(a))
| | Ano_conclusao=(Antesde2000): Insatisfatorio(4.0/0.0)
| | Ano_conclusao!=(Antesde2000)
| | | Sexo=(M)
| | | | Polo=(Sao Joao do Piaui): Insatisfatorio(6.0/0.0)
| | | | Polo!=(Sao Joao do Piaui)
| | | | | Ano_conclusao=(Depoisde2010): Insatisfatorio(3.0/0.0)
| | | | | Ano_conclusao!=(Depoisde2010): Satisfatorio(6.0/1.0)
| | | | Sexo!=(M): Satisfatorio(4.0/0.0)
| | Estado_civil!=(NaoSolteiro(a))
| | | Ano_conclusao=(Depoisde2010)
| | | | Sexo=(M): Insatisfatorio(2.0/0.0)
| | | | Sexo!=(M): Satisfatorio(2.0/0.0)
| | | Ano_conclusao!=(Depoisde2010): Satisfatorio(38.0/0.0)

```

Regras de produção geradas pelo algoritmo *SimpleCart* (Base SI_UAB4).

```

Sexo=(M)
| Polo=(Inhuma)|(Urucui)|(Castelo do Piaui)|(Jaicos): Insatisfatorio(63.0/11.0)
| Polo!=(Inhuma)|(Urucui)|(Castelo do Piaui)|(Jaicos): Satisfatorio(18.0/0.0)
Sexo!=(M)
| Polo=(Inhuma): Insatisfatorio(13.0/0.0)
| Polo!=(Inhuma)
| | Polo=(Esperantina)
| | | Estado_civil=(NaoSolteiro(a)): Insatisfatorio(8.0/0.0)
| | | Estado_civil!=(NaoSolteiro(a)): Satisfatorio(13.0/0.0)
| | | Polo!=(Esperantina): Satisfatorio(95.0/3.0)

```

Regras de produção geradas pelo algoritmo *SimpleCart* (Base HI_UAB4).