

Análise de Desempenho Acadêmico Utilizando Redes Bayesianas: um estudo de caso¹

Danilo Raniery Alves Coutinho, Thereza P. P. Padilha, Vanessa F. Dantas

Departamento de Ciências Exatas - Universidade Federal da Paraíba (UFPB)
Campus IV – Rio Tinto – PB – Brasil

{danilo.coutinho, thereza, vanessa}@dce.ufpb.br

Abstract. *Dropout and failure rates in the initial programming courses at the UFPB-Campus IV are considered high. Out of 759 students enrolled in these courses, from 2011.1 semester to 2014.1, 251 did not a success and 303 dropout. In this context, this paper presents a case study involving the analysis of the students performance in the introduction to Programming courses in Computer Science and Information Systems Degrees, at UFPB-Campus IV, using the Bayesian networks technique (RB) to extract implicit knowledge. This, in order to identify possible causes of abandonment and reproof. The networks were modeled manually, by domain expert, in the Netica tool and also automatically generated in the GeNie 2.1 software. Among the discoveries, it was possible to identify the dropout student profile and that 89.9% of students who do not obtain score more than 70% in the first unit work also get low score in the first test.*

Resumo. *Os índices de evasão e reprovação das disciplinas introdutórias de programação, dos cursos de computação do Campus IV da UFPB são considerados elevados. Dos 759 educandos matriculados nestas disciplinas, entre os períodos de 2011.1 e 2014.1, 251 foram reprovados e 303 evadiram. Nesse contexto, este artigo apresenta um estudo de caso envolvendo a análise do desempenho dos alunos na disciplina de Introdução à Programação dos cursos de Licenciatura em Ciência da Computação e Bacharelado em Sistemas de Informação da UFPB-Campus IV, utilizando a técnica de Redes Bayesianas (RB) a fim de extrair conhecimento implícito. Esse, com o objetivo de identificar possíveis causas de abandono e reprovação. As redes foram modeladas manualmente pelo especialista do domínio na ferramenta Netica e, também, geradas automaticamente pelo software GeNie 2.1. Dentre as descobertas realizadas, foi possível identificar o perfil do aluno desistente e que 89,9% dos alunos que não obtém rendimento acima ou igual a 70% nos trabalhos da primeira unidade também não obtém tal rendimento na nota da primeira prova.*

¹ Trabalho de Conclusão de Curso apresentado pelo aluno Danilo Raniery Alves Coutinho sob a orientação da professora Thereza Patrícia Pereira Padilha e coorientação da professora Vanessa Farias Dantas como parte dos requisitos para obtenção do grau de Licenciado em Ciência da Computação na UFPB Campus IV.

1. Introdução

As instituições de ensino superior apresentam estruturas diferentes do ensino médio que podem influenciar indiretamente o sucesso do aluno no curso, como por exemplo, horários flexíveis, distanciamento dos professores, amizades com características heterogêneas e conteúdos menos apoiados em apostilas e livros [Cunha and Carrilho 2005]. Tudo isso, combinado com a exigência de um alto nível de autonomia e maturidade, podem contribuir para um baixo rendimento acadêmico, estimulando a reprovação ou a desistência dos discentes.

Segundo Silva Filho et al. (2007), a evasão e a reprovação na graduação é um objeto de estudo analisado por diferentes pesquisas, onde se pretende descobrir, principalmente, quais são os motivos que levam os alunos a desistirem do curso, pois essa prática gera desperdícios para o país, visto que, não existe retorno para este dos recursos financeiros e de pessoal investidos. Também foi identificado nos estudos realizados que as taxas de evasão são duas a três vezes maior no primeiro ano do curso, em relação aos outros anos.

Esse problema tende a se agravar nas disciplinas introdutórias de programação, pois, os graduandos não se sentem capazes de programar por não possuírem as habilidades de solucionar problemas, raciocínio lógico, habilidade matemática, capacidade de abstração, entre outros [Branco Neto and Schuvartz 2007]. Diante disso, os docentes precisam se esforçar para desenvolver e adequar técnicas de ensino para atender as necessidades dos educandos que necessitam obter essas habilidades.

Ao ministrarem as suas disciplinas, os professores armazenam diversos dados dos alunos como notas de trabalhos, notas de prova, frequência, média, entre outros. Analisar estes dados de forma manual com o intuito de verificar desempenhos individuais e da turma e indícios de possíveis alunos a evadir, por exemplo, é uma tarefa bastante trabalhosa e complexa. Diante disso, objetivando analisar dados educacionais para encontrar conhecimentos úteis, surgiu a mineração de dados educacionais (MDE), uma nova área de pesquisa que envolve as áreas de Mineração de Dados (MD) e Educação.

Segundo Marques (2014), a mineração de dados consiste na utilização de algoritmos no desenvolvimento de modelos e padrões para a obtenção de informações desconhecidas em grandes bases de dados. Nesta área, destaca-se a mineração de dados por meio de redes Bayesianas (RB), pois trabalha com relacionamentos probabilísticos de variáveis de domínios em que há incerteza. Assim, pode-se encontrar modelos que representem relacionamentos de causa-efeito e, sobretudo, explorar inferências probabilísticas entre as variáveis da rede.

Nesse contexto, percebe-se a necessidade da investigação para identificar informações que possam auxiliar os professores da disciplina introdutória de programação, bem como a gestão acadêmica. Logo, o presente trabalho tem como objetivo analisar e compreender probabilisticamente variáveis e/ou eventos que influenciaram o desempenho acadêmico dos alunos que cursaram a disciplina Introdução à Programação, dos cursos de Licenciatura em Ciência da Computação e Bacharelado em Sistemas de Informação, entre os semestres 2011.1 a 2014.1, da UFPB - Campus IV. Para isso, o foco está voltado para a tarefa de mineração de causas

utilizando redes Bayesianas desenvolvidas automaticamente por algoritmos de aprendizagem e manualmente por especialista do domínio.

Este artigo encontra-se organizado da seguinte forma: na seção 2 é realizada uma introdução sobre mineração de dados educacionais e o processo de descoberta de conhecimento em bancos de dados, enquanto que na seção 3 são apresentados conceitos básicos sobre redes Bayesianas. Na seção 4 são relatados trabalhos relacionados e na 5 a metodologia, apresentando os resultados dos experimentos realizados, bem como a descrição do conjunto de dados utilizado. Por fim, na seção 6, são apresentadas as considerações finais.

2. Mineração de Dados Educacionais

Han et al. (2012) afirmam que estamos vivendo na era dos dados, pois alimentamos diariamente a internet e vários dispositivos de armazenamento com terabytes ou petabytes de informações, sejam estas educacionais, empresariais, científicas ou de qualquer outro aspecto.

Nessa “sociedade dos dados” o mapeamento e a análise de informações, de forma automática, são necessidades que devem ser atendidas por ferramentas eficazes e versáteis, capazes de descobrir conhecimento oculto em grandes bases de dados. É dessa necessidade que então surge a mineração de dados, área da Inteligência Artificial, que objetiva converter grandes quantidades de dados em conhecimento [Han et al. 2012].

Williams (2011) configura a MD como sendo a busca pela construção de modelos capazes de fornecer *insights* sobre o mundo e como ele funciona, para que assim, seja possível realizar previsões. Witten et al. (2011) ressalta que a MD é utilizada para aquisição de conhecimento, pois, as estruturas de conhecimento e as suas descrições são tão valiosas quanto à capacidade de executar novos exemplos.

Para a obtenção destes modelos, existem cinco etapas a serem realizadas (ver Figura 1) que compõem o processo de descoberta de conhecimento em bancos de dados (*Knowledge Discovery in Database – KDD*), definido por Fayyad et al. (1996), que são:

- **seleção:** Nesta etapa é selecionado o conjunto de dados que será utilizado para a mineração. Este deve ser associado a um domínio e deve conter atributos (características ou variáveis) e registros. Por vezes é necessário acessar mais de uma fonte de dados para se obter essas informações;
- **pré-processamento:** A segunda etapa do processo, considerada a mais dispendiosa, consiste no trabalho realizado no conjunto de dados, para que sejam eliminados dados redundantes, inconsistentes e/ou incompletos. O objetivo desse refinamento é a obtenção de um conjunto de dados consistente que possa ser utilizado na mineração e, por isso, muitas vezes é necessária a realização da discretização dos dados contínuos (reduzir os valores por meio de intervalos);
- **transformação:** Após o pré-processamento do conjunto de dados, é realizado a sua conversão para formatos que sejam reconhecidos pelos algoritmos utilizados na mineração;

- **data mining (mineração de dados):** Nesta etapa o conjunto transformado é explorado e analisado por meio dos algoritmos de mineração de dados, objetivando a identificação de padrões (conhecimento);
- **interpretação/avaliação:** Por fim, é realizada a interpretação e avaliação dos padrões identificados, para que assim, se possa validar o conhecimento extraído na etapa anterior.

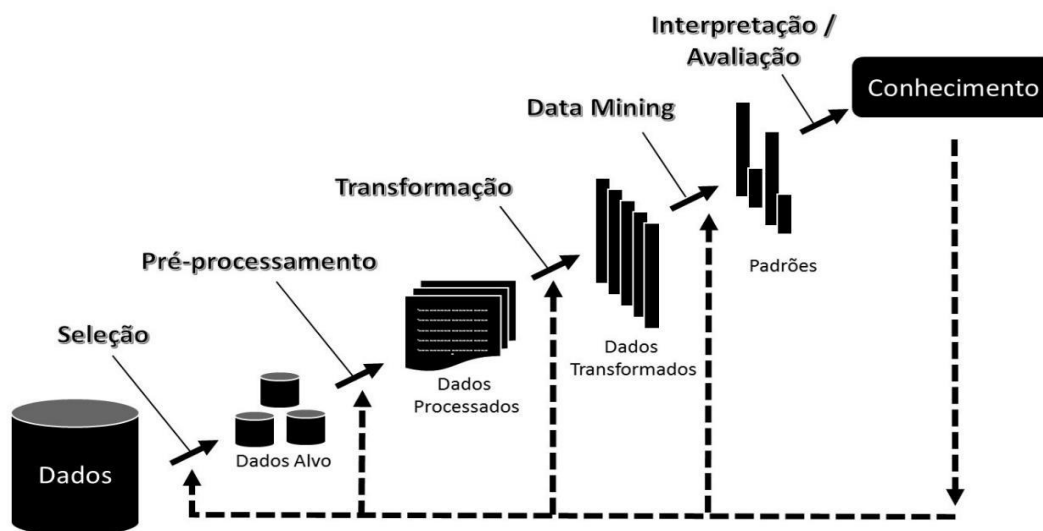


Figura 1. Etapas do processo KDD - Fayyad et al. (1996)

A Sociedade Internacional de Mineração de Dados Educacionais (2016) afirma que qualquer dado, seja extraído de ambientes interativos de aprendizagem ou dados administrativos retirados de escolas e universidades, sobre os alunos, é considerado um dado educacional.

Segundo Baker et al. (2011), MDE é uma área que busca encontrar informações implícitas em bases de dados aonde o domínio seja a Educação para, assim, auxiliar a tomada de decisões. As tarefas realizadas na MDE podem ser agrupamento, predição, mineração de relação, entre outras.

A tarefa de agrupamento busca identificar dados que se agrupem naturalmente, por meio de similaridades. Já a tarefa de predição tenta modelar padrões dedutivos através de análise dos dados. Já na mineração de relações é possível identificar relações entre as variáveis da base de dados.

A mineração de causas, subclasse da tarefa de mineração de relações, foco deste trabalho, busca verificar se um evento influencia a ocorrência de outro evento através de análise de padrões. Por exemplo, se um aluno apresenta performance ruim em uma prova isso pode decorrer por falta de domínio de conteúdos anteriores ou por dificuldade de aprendizagem. Assim, a mineração de causa tenta inferir eventos com variáveis buscando situações de causa-efeito.

3. Redes Bayesianas

Redes Bayesianas, segundo Korb and Nicholson (2010), são modelos gráficos para raciocínio sob incerteza, onde os nós representam variáveis (discreta ou contínua) e os arcos representam conexões diretas entre eles. A sua estrutura, ou topologia, é caracterizada pelos relacionamentos qualitativos entre as variáveis.

Quanto ao layout, as RBs são nomeadas seguindo uma metáfora de uma família, em que um nó é chamado de pai de outro nó, denominado filho, se existe um arco que afirme sua influência nele. Nesse contexto, também são utilizados conceitos de ancestrais e descendentes, como também, o de nó-raiz e o de nós-folhas, como em árvores [Korb and Nicholson 2010].

Assim, uma rede Bayesiana consiste nas causas entre as variáveis que formam sua estrutura (parte qualitativa) e as probabilidades dos estados de cada nó, influenciados pelo estado de seus nós-pais (quantitativa). As redes Bayesianas podem ser construídas de forma manual (considerando o conhecimento de especialistas sobre o domínio) ou automática (através de ferramentas que interpretam os conjuntos de dados e, utilizando algoritmos de aprendizado, geram a rede).

Com a rede Bayesiana construída e se conhecendo os relacionamentos entre os nós, é possível realizar a inferência Bayesiana que computa a probabilidade dos resultados dadas algumas entradas. A inferência Bayesiana, que é um tipo de inferência probabilística, pode ser utilizada para os propósitos de diagnóstico (verifica dos efeitos para as causas), causal (observa as causas sobre as evidências), inter-causal (analisa causas que têm um efeito em comum) e mista (combina dois ou mais tipos de propósitos) [Fernandes et al. 2012].

A inferência Bayesiana é apoiada pelo Teorema de Bayes, representado pela fórmula a seguir:

$$P(B|e) = \frac{P(e|B) P(B)}{P(e)}$$

Na fórmula, $P(B)$ é a probabilidade *a priori* do evento B ocorrer; $P(B|e)$ é a probabilidade a posteriori de B ocorrer após a evidência e , enquanto que $P(e|B)$ é a *verossimilhança* da evidência e dado B, e $P(e)$ é um fator de normalização. Assim, é possível calcular quais as chances de determinado evento ocorrer, como por exemplo a reprovação, aprovação ou desistência de um graduando, dadas as causas, como nota da primeira unidade baixa, notas boas nos trabalhos, entre outras.

4. Trabalhos Relacionados

Com relação aos trabalhos relacionados, destacam-se:

Baker et al. (2011): os autores fundamentam os conceitos da MDE e classificam suas subáreas por tarefas, exibindo os métodos e objetivos de pesquisa de cada uma. Também é realizada a comparação dos métodos da MDE com os da MD, como essa encontra-se no cenário brasileiro e quais são as vantagens que a área de pesquisa pode

trazer para o Brasil, em especial para a educação a distância. Para isso, é apresentada uma revisão de trabalhos na área que conseguiram beneficiar a educação em seu país, analisando seus métodos e aplicações.

Santos (2015): em sua dissertação “Uma abordagem temporal para identificação precoce de estudantes de graduação a distância com risco de evasão utilizando técnicas de mineração de dados”, apresenta uma abordagem temporal, onde se utiliza da mineração de dados educacionais para a predição do rendimento final dos alunos matriculados em um ambiente virtual de aprendizagem (AVA). Para isso, ele utiliza a técnica de classificação por regras de decisão, por meio dos algoritmos OneR, PART, JRip e Ridor. Em seu trabalho, foi possível identificar, inicialmente, quais foram as disciplinas e as notas que apresentaram mais impacto na evasão discente. Para que em um segundo momento, pudesse ser desenvolvido os modelos preditivos, que tornaram possível a previsão de evasão em cursos a distância de longa duração.

Afiune (2012): Na sua monografia “Mineração de Dados Educacionais: Predição Comportamental em Ambientes de Educação a Distância (EaD)”, Afiune (2012) desenvolve um sistema capaz de reconhecer alunos que possam vir a evadir do curso. Para isso, foram analisados logs de alunos em ambientes virtuais de aprendizagem através da mineração de dados educacionais e a criação de árvores de decisão para a predição. A pesquisa utilizou o software Weka, e o algoritmo J48, para realizar o desenvolvimento da descoberta de conhecimento nas bases de dados. Também foram consultados, por meio de pesquisa de campo, profissionais de EaD que apresentaram os motivos para a evasão. Estes, que basearam o sistema preditivo.

5. Metodologia

Para a realização do trabalho, além da revisão bibliográfica, foram seguidas as etapas do processo de descoberta de conhecimento em bancos de dados. Também foram elaborados três casos de testes, fundamentados em questões específicas levantadas pela professora da disciplina de introdução a programação, como por exemplo: “quais as chances de insucesso na prova da primeira unidade, dadas as notas dos testes parciais?”, “qual o impacto das notas parciais na reprovação dos alunos?”, dentre outras.

O caso 1 teve como foco responder à questão: “Caso se obtenha insucesso total na nota parcial da primeira unidade, qual a probabilidade de se obter sucesso, insucesso parcial e insucesso total na nota da prova da primeira unidade?”. O caso 2 teve por objetivo permitir inferências sobre as notas das provas e a situação final dos alunos. Enquanto que com o caso 3 buscou possibilitar a inferência entre a situação final e cada nota das três unidades. Um detalhamento de cada uma destas etapas é apresentado a seguir.

5.1 Seleção do conjunto de dados educacionais

A obtenção do conjunto de dados se deu por meio da professora de Introdução à Programação, que disponibilizou informações referentes ao desempenho dos alunos matriculados entre os semestres de 2011.1 a 2014.1 através de planilhas eletrônicas. Estas, compostas pelas notas de prova, notas parciais (formadas por trabalhos, mini testes, projetos e atividades online), notas finais de cada unidade, média e situação final dos educandos. A quantidade de registros coletados está discriminada na Tabela 1.

Tabela 1. Desempenho acadêmico nos semestres 2011.1 a 2014.1

Período	Curso	Aprovados	Reprovados	Desistentes	Total
2011.1	LCC	17	20	17	54
2011.1 (Turma Extra)	LCC e BSI	8	8	23	39
2011.2	LCC	11	11	32	54
2011.2 (Turma Extra)	LCC e BSI	22	11	28	61
2012.1	LCC	15	10	28	53
2012.1	BSI	18	13	23	54
2012.2	LCC	12	21	25	58
2012.2 (Turma Extra)	LCC e BSI	18	26	13	57
2013.1	BSI	13	22	16	51
2013.1	LCC	11	20	26	57
2013.1 (Turma Extra)	LCC e BSI	8	10	14	32
2013.2	LCC	8	25	12	45
2013.2 (Turma Extra)	LCC e BSI	2	11	2	15
2014.1	BSI	21	11	14	46
2014.1	LCC	12	14	20	46
2014.1 (Turma Extra)	LCC e BSI	9	18	10	37
Total		205	251	303	759

5.2 Pré-processamento dos dados coletados

Os dados selecionados na etapa anterior foram inicialmente analisados de forma manual, para que assim, fosse possível identificar quais eram as variáveis presentes nas planilhas. Destes registros foram descartadas a média, por ser apenas uma soma das notas de cada unidade, e os desistentes que se matricularam, mas, nunca compareceram na disciplina, por não acrescentar informações à pesquisa.

Identificamos algumas variáveis numéricas que apresentavam valores de texto, como por exemplo “reprovado”, “—”, ‘trancou’, e foi necessário tratar essa inconsistência, substituindo esses valores de texto pela nota 0,0 (zero). Assim, foi possível realizar a discretização das variáveis numéricas, onde as notas foram classificadas em “Sucesso”, caso a nota fosse maior ou igual a 70%, que é a média mínima para aprovação na disciplina sem a prova final, “Insucesso”, caso a nota fosse inferior a 70%. O “Insucesso” foi dividido ao meio, em “InsucessoParcial”, representando notas menores que 70% e maiores ou iguais a 35%, que são os valores próximos a média que classifica o aluno como apto a realizar a prova final, e “InsucessoTotal”, com notas abaixo dos 35% de aproveitamento, como exemplificado na Figura 2.

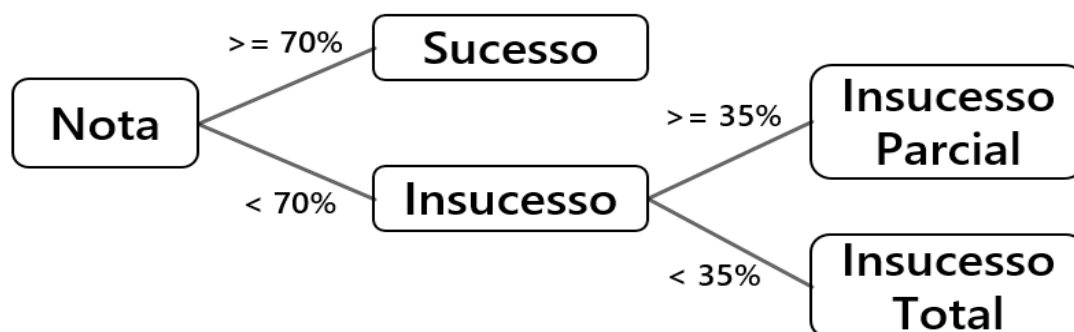


Figura 2. Representação das regras utilizadas na discretização dos dados

Após a discretização, as bases de dados foram reorganizadas para que fosse possível a obtenção de três conjuntos de dados que foram utilizados pelos casos de teste. A base de dados para o caso 1 consistia numa planilha contendo a nota parcial e a nota de prova da unidade 1 de todos os alunos, para o caso 2 foi formada uma base contendo a nota final das três unidades e a situação final e para o caso 3 foi organizado uma contendo todas as notas dos alunos e a situação final.

5.3 Transformação dos dados

As bases de dados foram convertidas de seu formato original (.xlsx) para o formato de texto (.txt), onde as variáveis e seus respectivos valores são separados por tabulação. Esta formatação foi necessária para que os valores pudessem ser lidos pelas ferramentas GeNIe 2.1 e Netica.

5.4 Extração de conhecimento utilizando redes Bayesianas e Interpretação e Avaliação do conhecimento obtido

Nesse momento da pesquisa, foram desenvolvidas duas etapas, em paralelo, do processo KDD, a que consiste na mineração de dados, onde foram elaboradas redes Bayesianas de forma manual e automática para que pudessemos adquirir modelos com a finalidade de se explorar ou buscar novos conhecimentos. E a avaliação e interpretação do conhecimento identificado, onde os resultados obtidos foram apresentados a especialista do domínio, que os validou como informações pertinentes e pertencentes à realidade da disciplina.

5.4.1 Rede gerada manualmente

Para respondermos questões pontuais, feitas pela professora da disciplina, e tentarmos obter novas informações por meio dos casos de teste, as redes Bayesianas foram criadas de forma manual através da ferramenta Netica [Norsys 2016], que nos possibilitou criar cada nó e arco considerando informações fornecidas pela especialista do domínio. Para isso, foi desenvolvida a rede apresentada na Figura 3.

Nessa rede, é possível visualizar os nós “NotaParcialUnidade1” e “ProvaUnidade1” e seus atributos “Sucesso”, “InsucessoParcial” e “InsucessoTotal”. Estes, até então, apresentando informações estatísticas, onde na nota parcial da primeira unidade é possível identificar um sucesso de 39.9% dos alunos, enquanto que na nota da prova o sucesso é de apenas 22.9%. Também é possível identificar o arco que liga o nó

“NotaParcialUnidade1” ao “ProvaUnidade1”, o que implica afirmar que o primeiro possui influência (causa) sobre o segundo (efeito).

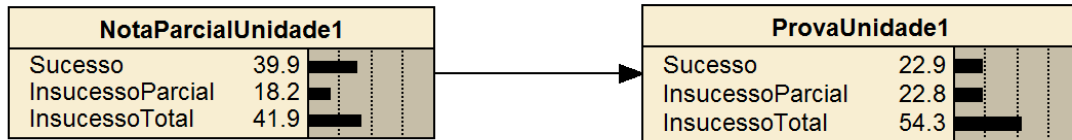


Figura 3. Rede Bayesiana gerada para o caso 1

A rede apresenta que 41.9% dos alunos obtiveram insucesso total na nota parcial da primeira unidade, logo, ao selecionarmos apenas esses alunos (100% dos alunos que obtiveram insucesso total), obtemos a rede apresentada na Figura 4. Nesta, é possível identificar que do conjunto de educandos selecionados no primeiro nó, 2,94% tem chance de obter sucesso, 7,19% de insucesso parcial e 89,9% de insucesso total na prova da primeira unidade.

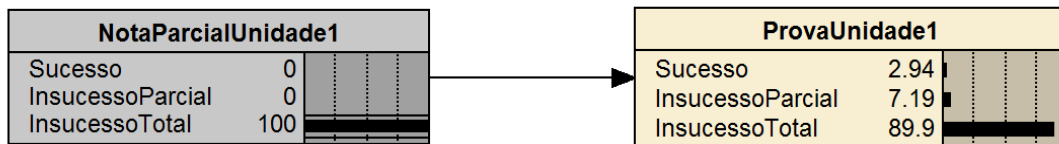


Figura 4. Inferência com os nós NotaParcial e ProvaUnidade1

Dessa forma, é possível identificar que a obtenção do conceito de insucesso total na nota parcial implica em uma chance de quase 90% de obtenção de conceito de insucesso total na primeira prova da disciplina.

Para o caso 2 foi elaborada uma rede Bayesiana contendo as três notas finais das unidades e a situação final dos alunos, conforme mostra a Figura 5. A intenção deste caso de teste era relacionar a situação final dos alunos com as notas das unidades e os arcos foram elaborados considerando que o conteúdo da disciplina é acumulativo, o que implica dizer que o assunto da primeira unidade influencia a compreensão das demais.

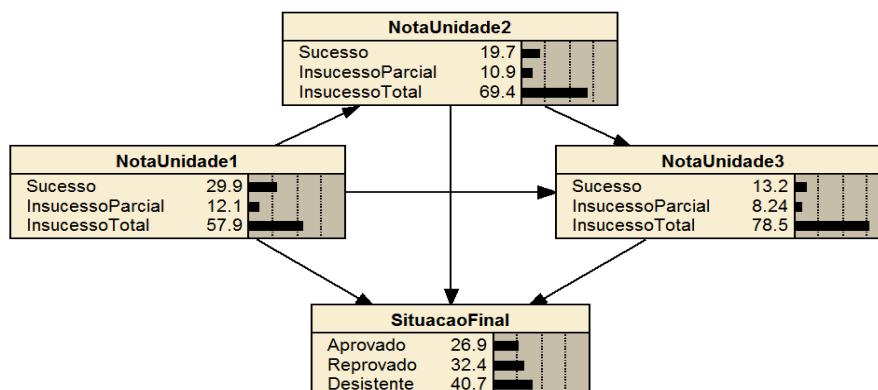


Figura 5. Rede Bayesiana gerada para o caso 2

Após realizarmos inferências com esta rede, identificamos que dos alunos reprovados, 64,5% adquiriram conceito de insucesso total e 21,2% de insucesso parcial na primeira unidade, como exposto na Figura 6.

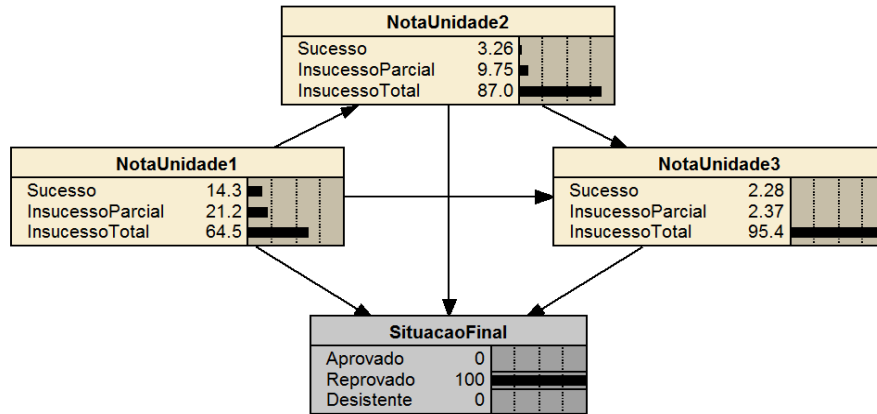


Figura 6. Inferência observando as causas da reprovação

Logo, com essa nova informação, podemos visualizar um perfil de aluno que, dada a nota da primeira unidade, possui tendência à reprovação.

Para o caso 3 utilizamos todos os dados disponíveis para elaborar uma rede Bayesiana em que pudéssemos relacionar as notas parciais, provas e a situação final da disciplina. A elaboração de seus arcos considerou que o conteúdo da disciplina, seja por unidade ou por semestre, é acumulativo. Ou seja, a nota parcial da primeira unidade influencia a nota de sua prova, que influenciará as demais notas da segunda e terceira unidade. A rede gerada é apresentada na Figura 7.

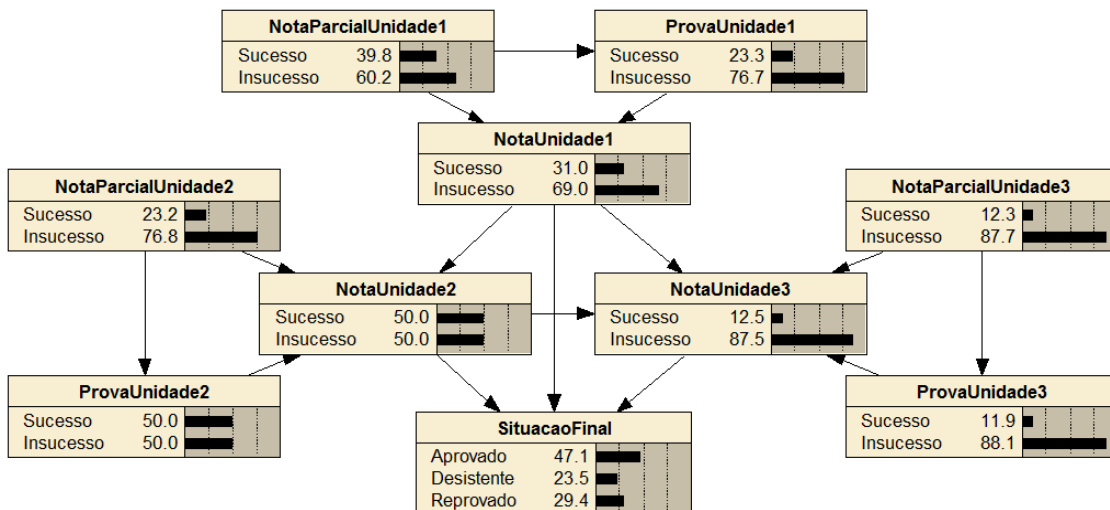


Figura 7. Rede Bayesiana gerada para o caso 3

Com a rede Bayesiana desenvolvida para o caso 3, foi possível analisar qual a influência das notas parciais no rendimento final dos alunos. Desse modo, como exposto na Figura 8, identificamos que os alunos que obtêm conceito de sucesso nas três notas parciais apresentam 71,3% de chance de serem aprovados. Evidenciando assim, o impacto do sucesso das notas parciais na aprovação do educando.

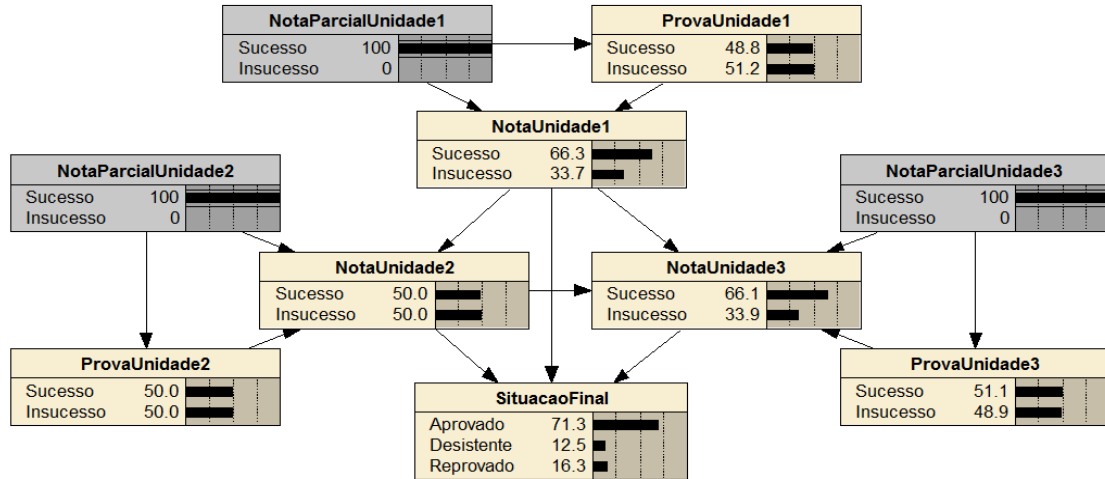


Figura 8. Inferência com os nós de notas parciais

Com a rede do caso 3 também foi possível identificar que 72,3% dos alunos que obtiveram sucesso na primeira unidade foram aprovados (Figura 9). Exemplificando assim, a importância da obtenção de sucesso na primeira unidade, pois, esta indica para uma provável aprovação na disciplina.

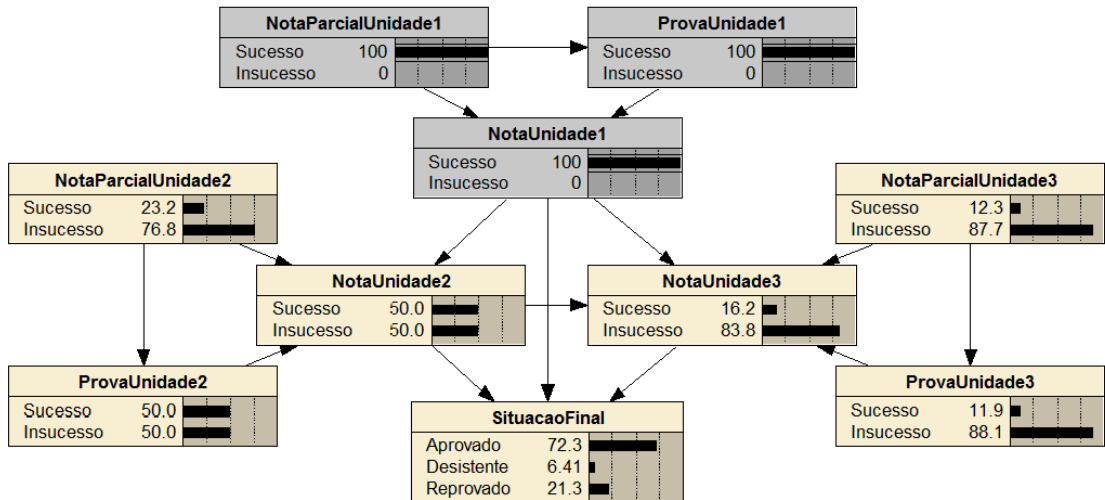


Figura 9. Inferência com os nós referentes a primeira unidade da disciplina

Outra inferência realizada foi a de observar os efeitos (aprovação, desistência ou reprovação) para tentar identificar as causas. Desse modo, chegando a rede Bayesiana apresentada na Figura 10.

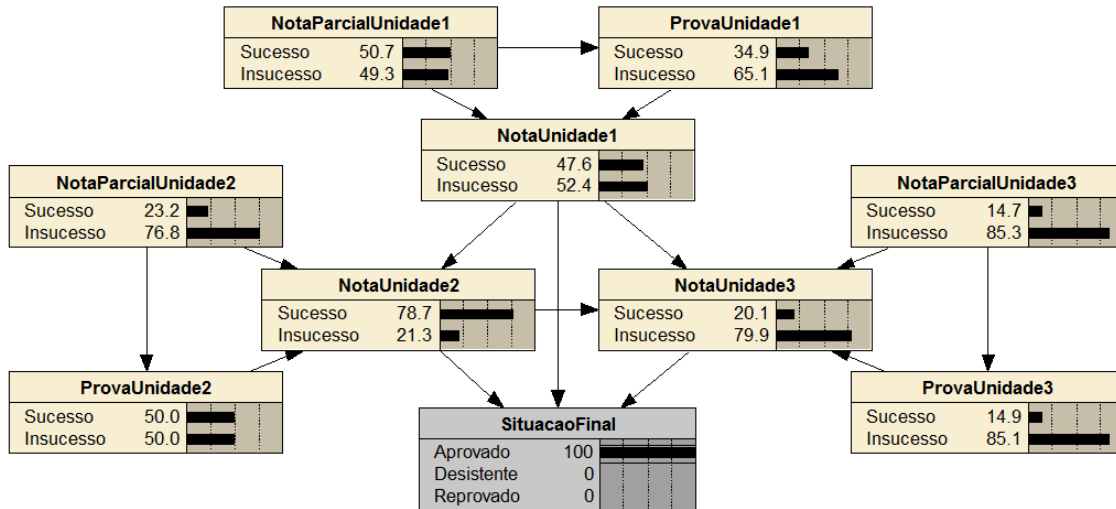


Figura 10. Inferência observando as causas da aprovação

Ao analisarmos a rede acima, identificamos que os alunos aprovados apresentam um perfil de 47,6% de chance de sucesso na primeira unidade e na segunda unidade se observa um aumento da probabilidade de sucesso para 78,7%. Entretanto, na terceira unidade essa cai para 20,1%.

5.4.2 Rede gerada automaticamente

Buscando encontrar informações ocultas, foi gerada uma rede Bayesiana com os dados existentes no caso 3, de forma automática por meio do software GeNIe 2.1 [Bayesfusion 2016]. Este, capaz de ler a base de dados em formato de texto (.txt) e gerar uma rede de acordo com o algoritmo de aprendizado escolhido.

Dentre os algoritmos presentes na ferramenta podemos encontrar o “Bayesian Search”, “PC”, “Essential Graph Search”, “Greedy Thick Thinning”, entre outros. Entretanto, o único que apresentou resultado satisfatório, onde as ligações eram mais próximas da rede modelada pelo especialista e que apresentou novas informações sobre o perfil dos alunos desistentes, foi o *Bayesian Search*, que desenvolveu a rede representada na figura 11. Os demais algoritmos não conseguiram gerar ligações entre os nós ou apresentaram um número muito pequeno de ligações, entre um ou três arcos para toda a rede.

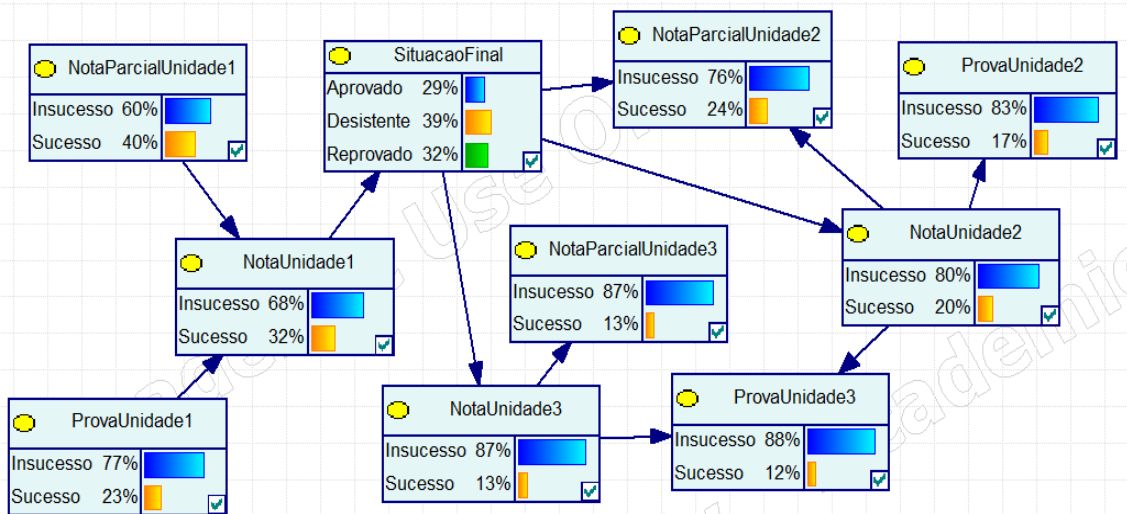


Figura 11. Rede Bayesiana gerada pelo algoritmo *Bayesian Search*

Através das inferências realizadas na rede foi possível constatar que dos alunos que obtinham insucesso nas três notas parciais apresentaram 56% de chances de evadir da disciplina e 38% de chance de reprovar (Figura 12). O que, mais uma vez, reafirma a importância dos trabalhos parciais para o sucesso na disciplina.

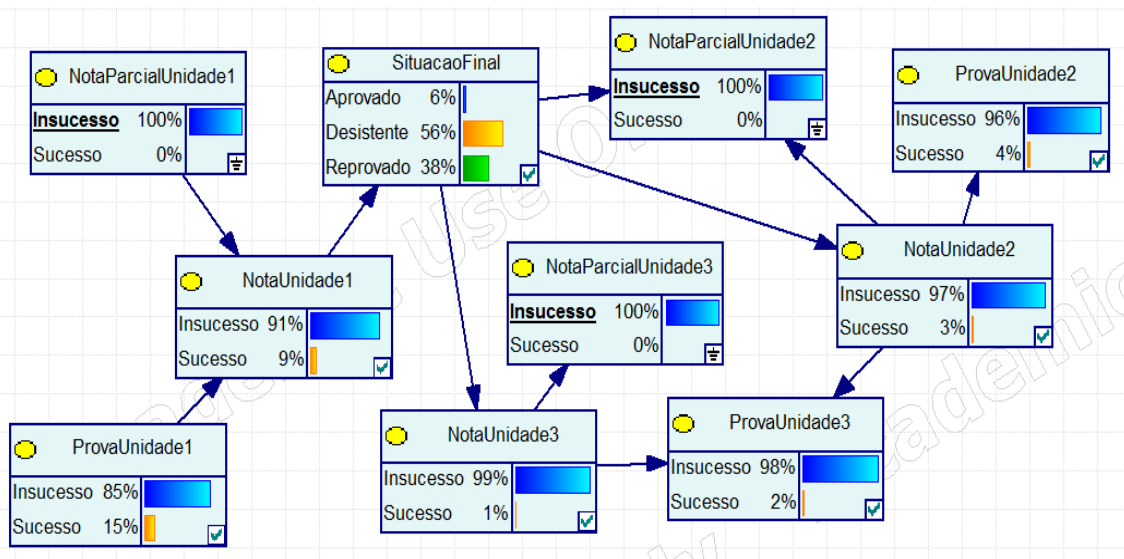


Figura 12. Inferência com os nós de notas parciais

Também foi possível identificar informações referentes ao perfil dos alunos desistentes, que apresentam grandes chances de insucesso nas notas das unidades, como representado na Figura 13.

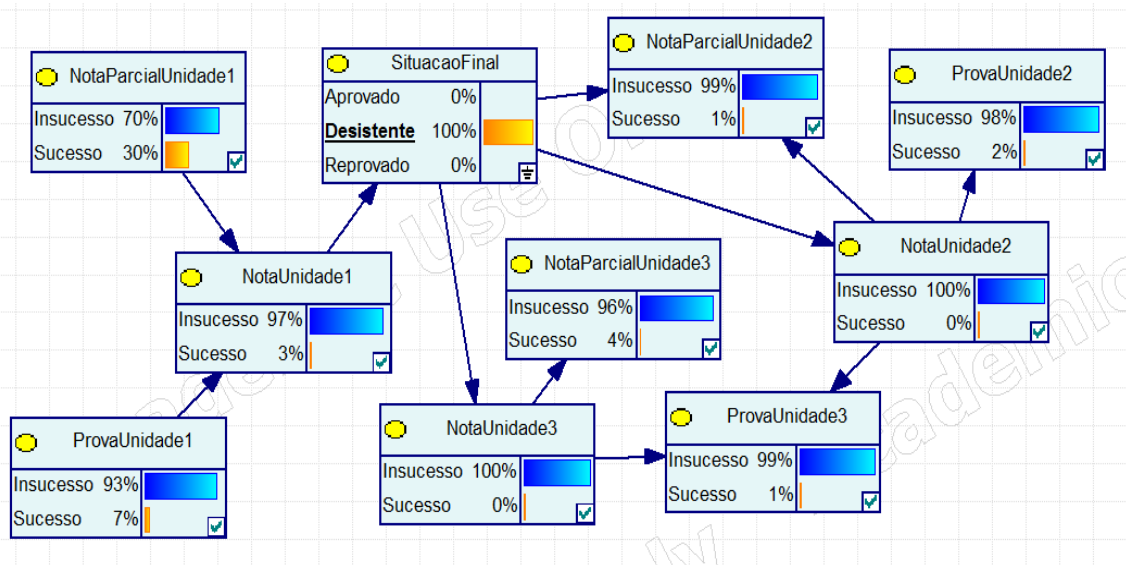


Figura 13. Inferência observando as causas da evasão

Por meio desta rede, podemos identificar que 70% dos alunos que obtém insucesso na nota parcial da primeira unidade tem chances de evadir da disciplina. O que nos aponta um grupo de educandos, nos primeiros momentos da unidade 1, que precisam de mais atenção e cuidados.

6. Considerações Finais

O presente trabalho, além de possuir o objetivo de analisar o desempenho dos alunos através de RBs, reafirma a importância de pesquisas com mineração de dados educacionais. MDE se configura como uma área promissora para auxiliar os gestores acadêmicos na resolução de problemas educacionais por meio da análise de dados, como por exemplo, a identificação de situações que levam a evasão ou reprovação dos graduandos.

Através das redes desenvolvidas e da análise dos dados, foi possível identificar, com as notas da primeira unidade, perfis de alunos que apresentam chances de reprovarem ou evadirem da disciplina de introdução a programação. Também foi observado qual a probabilidade de um aluno obter notas baixas na primeira prova da disciplina dada a nota de seus trabalhos. Informações essas, que podem auxiliar os professores e a gestão acadêmica na tomada de decisões para tentar diminuir os índices de reprovação e evasão.

Para encontrar essas informações, contamos apenas com as notas e a situação finais dos alunos da disciplina de introdução à programação, matriculados entre os semestres 2011.1 e 2014.1. O que, de certa forma, limitou a quantidade de novas informações obtidas. Logo, concluímos que vale a pena repetir os experimentos realizados na presente pesquisa utilizando dados de outros períodos, disciplinas ou dados sociais, como por exemplo, idade, situação civil, tipo de escola em que foi cursado o ensino médio, entre outras. Pois assim, será possível refinar os resultados já encontrados e descobrir novos conhecimentos ocultos.

Referências

- Afiune, Cally de Souza. (2012) “Mineração de Dados Educacionais: Predição Comportamental em Ambientes de Educação a Distância (EaD)”, 108 p. Monografia (Bacharelado em Sistemas de Informação) – Universidade Estadual de Goiás, Anápolis – GO.
- Baker, R., Isotani, S. and Carvalho, Adriana. (2011) “Mineração de Dados Educacionais: Oportunidades para o Brasil”, <http://www.br-ie.org/pub/index.php/rbie/article/view/1301/1172>, Fevereiro. In: *Revista Brasileira de Informática na Educação*. [S.l.], v. 19, n. 02, p. 03. ISSN 1414-5685.
- BayesFusion, LCC. (2016) “GeNIe”, <http://www.bayesfusion.com/>, Fevereiro.
- Branco Neto, Wilson Castello and Schuvartz, Aguinaldo Antonio. (2007) “Ferramenta Computacional de Apoio ao Processo de Ensino-Aprendizagem dos Fundamentos de Programação de Computadores”, In: *Simpósio Brasileiro de Informática na Educação*, p. 520-528.
- Cunha, Simone Miguez and Carrilho, Denise Madruga. (2005) “O processo de adaptação ao ensino superior e o rendimento acadêmico”, <http://www.scielo.br/pdf/pee/v9n2/v9n2a04.pdf>, Fevereiro. In: *Psicologia escolar e educacional*, v. 9, n. 2, p. 215-224.
- Fayyad, U., Piatetsky-shapiro, G. and Smyth, P. (1996) “From data mining to knowledge discovery: An overview”. In: *Advances in Knowledge Discovery and Data Mining*, AAAI Press / The MIT Press, MIT, Cambridge, Massachusetts, and London, England, p.1-34.
- Fernandes, Anita M. R., Zapelini, Clavison and Comuello, Eros. (2012) “Rede Bayesiana para Detecção da Intensidade da Dor”, In: *Simpósio de Excelência em Gestão e Tecnologia*, Tecnologia da Educação.
- Han, J., Kamber, M. and Pei, J. (2012), *Data Mining: Concepts and Techniques*, Elsevier, 3ª edição.
- Sociedade Internacional de Mineração de Dados Educacionais. (2016) “International Educational Data Mining Society”, <http://www.educationdatamining.org/>, Março.
- Korb, Kevin B., Nicholson, Ann E. (2010), *Bayesian Artificial Intelligence*, CRC Press, 2ª edição.
- Marques, Aramis Ferreira. (2014) “Aplicação de clusterização de dados na base de dados do zoneamento ecológico-econômico de Minas Gerais”. 88 p. Monografia (Bacharelado em Sistemas da Informação) - Universidade Federal de Lavras, Lavras – MG.
- Santos, Ramon Nóbrega. (2015) “Uma abordagem temporal para identificação precoce de estudantes de graduação a distância com risco de evasão utilizando técnicas de mineração de dados”. 151 p. Dissertação (Pós-Graduação em Informática – Sistemas de Computação) - Universidade Federal da Paraíba, João Pessoa – PB.

Silva Filho, R. L. L., Montejunas, P. R., Hipólito, O. and Lobo, M. B. C. M. (2007) “A evasão no ensino superior brasileiro”, <http://www.scielo.br/pdf/cp/v37n132/a0737132.pdf>, Março. In: Cadernos de Pesquisa, v. 37, n. 132, p. 641-659.

Norsys. (2016) “Netica Application”, <https://www.norsys.com/netica.html>, Junho.

Williams, Graham. (2011), Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery, Springer, 1ª edição.

Witten, I. H., Frank, E. and Hall, M. A. (2011), Data Mining: Practical Machine Learning Tools and Techniques, Elsevier, 3ª edição.