



III Encontro de Iniciação Científica e Tecnológica
III EnICT
ISSN: 2526-6772
IFSP – Câmpus Araraquara
19 e 20 de Setembro de 2018



Mineração de Dados para Caracterização dos Estudantes de Educação Profissional de Nível Médio do Instituto Federal de São Paulo

MARIA REGINA RAMALHO¹
CRISTIANE AKEMI YAGUINUMA²
FÁBIO JOSÉ JUSTO DOS SANTOS³
CINTIA MAGNO BRAZOROTTO⁴

1 Graduando em Análise e Desenvolvimento de Sistemas, Bolsista PIBIFSP, IFSP Câmpus Araraquara, em maria.regina@aluno.ifsp.edu.br.

2 Docente do Instituto Federal de São Paulo – Câmpus Araraquara – cristiane.yaguinuma@ifsp.edu.br (Orientadora)

3 Docente do Instituto Federal de São Paulo – Câmpus Araraquara - fabiojjs@ifsp.edu.br (Colaborador)

4 Técnico administrativo do Instituto Federal de São Paulo – Câmpus Araraquara cbrazorotto@ifsp.edu.br (Colaboradora)

Área de conhecimento (Tabela CNPq): Banco de Dados – 1.03.03.03-0

RESUMO: Considerando as transformações sociais recentes e a necessidade do fortalecimento do ensino é importante desenvolver e analisar os perfis dos cursos e dos estudantes por meio de novas tecnologias para um profundo aproveitamento dos dados obtidos, principalmente no setor acadêmico. Existe um grande interesse em se avaliar os dados disponíveis que envolvam os cursos do Ensino Médio Integrado ao Técnico (EMI) pelo fato de serem importantes no processo de qualificação dos jovens. Em base disso, a mineração de dados educacionais (EDM, *Educational Data Mining*) proporciona maneira inovadora de tratar e explorar este imenso universo de informações e dados, facilitando o mapeamento dos dados em novas estruturas de análise para que sirvam de suporte nos avanços do Ensino Médio Integrado. Assim, esta pesquisa visa delimitar novas estratégias, onde as equipes de educadores e de gestores possam identificar a realidade do perfil dos cursos e dos estudantes por meio de regras de associação. Os padrões obtidos devem apoiar decisões mais pontuais e fundamentadas para o Ensino Médio Integrado ao Técnico do IFSP ao caracterizar o perfil dos estudantes desse segmento educacional.

PALAVRAS-CHAVE: *Mineração de Dados, Regras de Associação, Dados Educacionais.*

INTRODUÇÃO

Com a incorporação da tecnologia nos mais variados setores, o volume de informações disponíveis aumentou consideravelmente, assim como capacidade de armazenamento de dados. O volume atual de dados é extenso e contém informações valiosas para diversas áreas. A análise de uma quantidade grande de dados tem se mostrado objeto de interesse e, para tanto, faz-se necessário o uso de ferramentas capazes de fazer esse trabalho. Segundo Fayyad, Piatetsky-Shapiro e Smyth (1996), o modelo tradicional para transformação de dados em informação consiste em um processo manual, em que especialistas analisam as informações e produzem relatórios. Na maioria das situações, por conta do grande volume de dados, esse processo manual se torna impraticável. É nesse contexto que a Mineração de Dados (do inglês, Data Mining ou DM) surgiu, da necessidade de análise de grandes volumes de dados com o intuito de extrair informações úteis. De acordo com Han, Kamber e Pei (2011), "... a área da mineração de dados pode ser vista como um resultado da evolução

natural da tecnologia de informação, permitindo extrair conhecimento a partir do processamento de volumes abundantes de dados”. Assim, DM é um recurso de busca de informações que tem se destacado e que oferece resultados interessantes.

Na área da Educação, observa-se também uma quantidade enorme de dados passíveis de análise para extração de informações e, a partir disso, originou-se a área de mineração de dados educacionais (do inglês, *Educational Data Mining*) (PEÑA-AYALA, 2014). Esta área é útil, portanto, para explorar o ambiente educacional com o intuito de compreender os fatores que influenciam o processo de ensino-aprendizagem.

Neste contexto, este trabalho visa aplicar a mineração de dados para caracterização do perfil dos estudantes de Educação Profissional de Nível Médio do Instituto Federal de São Paulo por meio de regras de associação, com o intuito de identificar aspectos para promover um processo de ensino-aprendizagem efetivo e de qualidade.

FUNDAMENTAÇÃO TEÓRICA

O processo de mineração de dados consiste na aplicação de ferramentas que através da implementação de algoritmos específicos que diagnosticam correlação entre os dados, apresentando ligações entre estes estabelecendo assim padrões frequentes. Este processo é realizado por cálculos matemáticos e podem ser executado de diversas maneiras: extração de regras de associação, classificação, regressão e agrupamento (HAN; KAMBER; PEI, 2011). Em especial, a *mineração de regras de associação* é bastante utilizada na literatura por permitir a descoberta de associações e correlações a partir da análise de padrões frequentes presentes nos dados.

Considerando um banco de dados de uma loja de artigos eletrônicos, suponha que clientes que compram computadores também tendem a comprar software antivírus em uma mesma cesta de compras. Esta associação pode ser representada pela seguinte regra de associação (HAN; KAMBER; PEI, 2011):

$$\text{computador} \rightarrow \text{software_antivirus} [\text{suporte: } 2\%, \text{confiança: } 60\%]$$

Uma regra de associação descreve que, quando há ocorrência dos itens ao lado esquerdo da seta, há uma tendência para que ocorram os itens ao lado direito da seta nas mesmas transações ou, no caso do exemplo, cestas de compras. *Suporte* e *confiança* são duas medidas de interesse para classificar as regras de associação. No exemplo, o valor de *suporte* significa que 2% de todas as cestas de compras analisadas mostram que computadores e software antivírus são comprados simultaneamente. O valor de *confiança* determina que 60% dos clientes que compraram computador também compraram software antivírus na mesma cesta de compras. De forma mais geral, dada uma regra de associação $A \rightarrow B$, o cálculo de suporte e confiança é definido da seguinte forma:

$$\text{suporte}(A \rightarrow B) = P(A \cup B) = \frac{\text{suporte}(A \cup B)}{|T|}$$

$$\text{confiança}(A \rightarrow B) = P(B|A) = \frac{\text{suporte}(A \cup B)}{\text{suporte}(A)}$$

onde $\text{suporte}(A \cup B)$ é a quantidade de transações que contêm A e B simultaneamente, $|T|$ é a quantidade total de transações e $\text{suporte}(A)$ é a quantidade de transações que contêm A. Tipicamente, as regras de associação são consideradas interessantes se elas satisfazem um limite de suporte mínimo (*minsup*) e um limite de confiança mínimo (*minconf*), que podem ser definidos ou ajustados por especialistas do domínio dos dados (HAN; KAMBER; PEI, 2011).

Dentre as abordagens clássicas para mineração de regras de associação, estão os algoritmos *Apriori* (AGRAWAL; SRIKANT, 1994), *Frequent Pattern Growth (FP-Growth)* (HAN; PEI; YIN, 2000) e *Eclat* (ZAKI, 2000).

Na revisão bibliográfica de Peña-Ayala (2014), se verifica uma forte utilização da mineração de dados usando regras de associação no âmbito educacional para identificar características dos estudantes e análise na

estrutura do ensino aprendizagem. Romero *et. al.* (2010) implementaram variantes do algoritmo *Apriori* para descobrir regras de associação que descrevem comportamentos raros ou infrequência dos estudantes. Zorrilla, García-Saiz e Balcázar (2011) compararam diversos algoritmos de mineração de regras de associação sobre dados educacionais para avaliar a ferramenta de mineração Yacaree (BALCÁZAR, 2011). Buldu e Üçgün (2010) aplicaram a mineração de regras de associação para a descoberta de padrões de recorrência de reprovação de estudantes em cursos ofertados em diferentes períodos letivos. Além disso, De Baker, Isotani e De Carvalho (2011) apontam que é crescente o potencial para pesquisa, desenvolvimento e aplicação de EDM considerando o cenário da educação brasileira. Assim, utilizando a disponibilidade das ferramentas tecnológicas, uma delas sendo a mineração de regras de associação aplicada a dados educacionais, pretende-se obter informações com significância para o desenvolvimento nesta área.

METODOLOGIA

Diante deste contexto, esta pesquisa realizou uma análise dos dados dos alunos de cursos Técnicos Concomitantes e Integrados do Ensino Médio do Instituto Federal de São Paulo, a fim de identificar suas características a partir de questionário socioeconômico. Os dados foram coletados em todos os câmpus do IFSP, no contexto de um projeto de doutorado que trata a caracterização da educação profissional no Brasil. O objetivo desta pesquisa consiste da extração de regras de associação para proporcionar padrões relevantes e frequentes para o projeto de doutorado vinculado. O intuito é oferecer suporte e resultados para detectar os comportamentos e o perfil dos estudantes na educação profissional do ensino médio e alimentar com informações úteis os programas de acesso, permanência e êxito de estudantes no IFSP.

A pesquisa baseou-se no processo KDD (do inglês, *Knowledge Discovery in Databases*), que tem como objetivo descobrir novas informações através da análise de grandes quantidades de dados (HAN; KAMBER; PEI, 2011). O processo KDD foi realizado da seguinte maneira:

- 1. Pré-processamento dos dados:** consiste da limpeza e eliminação de dados incompletos, conflitantes e inconsistentes nos arquivos. Compreende também seleção de atributos e transformação dos dados para adaptação das nomenclaturas e formato dos arquivos. Esta etapa foi realizada a partir de uma planilha de dados contendo 24.604 linhas e 78 colunas, onde cada linha representa uma inscrição de candidato a algum curso de nível médio do IFSP e as colunas contêm informações socioeconômicas dos candidatos, tais como: dados cadastrais básicos, formação escolar, formação escolar dos pais, motivo da escolha do curso, se possui equipamentos eletrônicos, aspectos de deficiências, entre outros. Para melhor resultado foram padronizadas as nomenclaturas dos dados e posteriormente realizada transformação para formatos compatíveis com as ferramentas de mineração de dados (formatos ARRF e CSV). Na sequência, foram selecionados atributos para iniciar os estudos, com a supervisão de especialista no domínio. Os atributos selecionados foram: gênero, curso, tipo de curso (Concomitante ou Integrado), raça e renda.
- 2. Mineração:** consiste no processo de aplicação de algoritmos para extrair padrões de dados e suas correlações. No estudo, foi utilizado o software Weka (HALL et al., 2009), uma ferramenta para mineração de dados sob a licença GPL (*General Public License*). Para esta pesquisa foi aplicado especificamente o algoritmo Apriori (AGRAWAL; SRIKANT, 1994) que se adequa mais facilmente ao processo, sendo bastante utilizado na literatura para verificar conjuntos de itens frequentes. Os detalhes da organização dos experimentos serão apresentados na seção de resultados e discussão.
- 3. Avaliação e interpretação dos padrões obtidos:** para avaliar e interpretar os padrões foram aplicados parâmetros diferentes em cada experimento, obtendo-se, assim, diversos conjuntos de padrões. Estabeleceu-se um comparativo crescente de confiança (valores 0.1, 0.5, 0.9), restringindo a 10, 50 e 100 regras, sempre com suporte mínimo entre 0.1 e 1.0, com variação dos atributos.

4. **Apresentação e visualização do conhecimento:** No trabalho, os diversos resultados foram organizados e estruturados de maneira sequencial, tabulados e legendados. Os detalhes das regras extraídas serão apresentados na seção de resultados e discussão.

RESULTADOS E DISCUSSÃO

Os experimentos foram organizados conforme citado na seção de metodologia. No total, os experimentos obtiveram 25 conjuntos de regras de associação, com destaque para três conjuntos de regras:

Experimento 1: limite de 10 regras, $minsup = 0.2$, $minconf = 0.5$, atributos: gênero, curso, tipo de curso, raça e renda.

1. Curso=INFORMATICA 5207 ==> Tipo=INTEGRADO 5207 <conf:(1)> lift:(2) lev:(0.11) [2609] conv:(2609.04)
2. Tipo=CONCOMITANTE 12004 ==> Genero=Masculino 7337 <conf:(0.61)> lift:(1.08) lev:(0.02) [529] conv:(1.11)
3. Tipo=INTEGRADO 11953 ==> Raca=Branca 7034 <conf:(0.59)> lift:(1.06) lev:(0.02) [421] conv:(1.09)
4. Raca=Branca 13254 ==> Genero=Masculino 7577 <conf:(0.57)> lift:(1.01) lev:(0) [60] conv:(1.01)
5. Genero=Masculino 13587 ==> Raca=Branca 7577 <conf:(0.56)> lift:(1.01) lev:(0) [60] conv:(1.01)
6. Genero=Feminino 10370 ==> Tipo=INTEGRADO 5703 <conf:(0.55)> lift:(1.1) lev:(0.02) [529] conv:(1.11)
7. Genero=Feminino 10370 ==> Raca=Branca 5677 <conf:(0.55)> lift:(0.99) lev:(-0) [-60] conv:(0.99)
8. Genero=Masculino 13587 ==> Tipo=CONCOMITANTE 7337 <conf:(0.54)> lift:(1.08) lev:(0.02) [529] conv:(1.08)
9. Raca=Branca 13254 ==> Tipo=INTEGRADO 7034 <conf:(0.53)> lift:(1.06) lev:(0.02) [421] conv:(1.07)
10. Tipo=INTEGRADO 11953 ==> Genero=Masculino 6250 <conf:(0.52)> lift:(0.92) lev:(-0.02) [-529] conv:(0.91)

Estão destacadas em tonalidade cinza as regras 2, 3 e 4, onde se observa que das instâncias totais (23.957), foi analisado suporte mínimo de 2% ($minsup$) representando 4791 instâncias. Assim, a regra 2 apresenta que o tipo de curso Concomitante, os alunos em 61% de confiança são do gênero masculino e na regra 3 o tipo de curso integrado em 59% com confiança são alunos da raça branca. A regra 4 apresenta que dos candidatos de raça branca 57% são do gênero masculino. Nestas análises, com as medidas de interesse testadas, resultou como indicador que o gênero masculino e raça branca tem correlações nos dois tipos de curso (Concomitante e Integrado).

Experimento 2: limite de 100 regras, $minsup = 0.1$, $minconf = 0.5$, atributos: gênero, curso, tipo de curso, raça e renda. Melhores regras obtidas:

1. Curso=INFORMATICA 5207 ==> Tipo=INTEGRADO 5207 <conf:(1)> lift:(2) lev:(0.11) [2609] conv:(2609.04)
2. Curso=INFORMATICA Raca=Branca 3121 ==> Tipo=INTEGRADO 3121 <conf:(1)> lift:(2) lev:(0.07) [1563] conv:(1563.82)
3. Genero=Masculino Curso=INFORMATICA 2620 ==> Tipo=INTEGRADO 2620 <conf:(1)> lift:(2) lev:(0.05) [1312] conv:(1312.79)
4. Genero=Feminino Curso=INFORMATICA 2587 ==> Tipo=INTEGRADO 2587 <conf:(1)> lift:(2) lev:(0.05) [1296] conv:(1296.25)
5. Genero=Masculino Renda=1 A 2 SM 3960 ==> Tipo=CONCOMITANTE 2630 <conf:(0.66)> lift:(1.33) lev:(0.03) [645] conv:(1.48)
6. Renda=3 A 5 SM 4521 ==> Genero=Masculino 2855 <conf:(0.63)> lift:(1.11) lev:(0.01) [290] conv:(1.17)
7. Renda=3 A 5 SM 4521 ==> Raca=Branca 2791 <conf:(0.62)> lift:(1.12) lev:(0.01) [289] conv:(1.17)
8. Tipo=CONCOMITANTE Raca=Parda 4150 ==> Genero=Masculino 2543 <conf:(0.61)> lift:(1.08) lev:(0.01) [189] conv:(1.12)
9. Tipo=CONCOMITANTE Raca=Branca 6220 ==> Genero=Masculino 3804 <conf:(0.61)> lift:(1.08) lev:(0.01) [276] conv:(1.11)
10. Tipo=CONCOMITANTE 12004 ==> Genero=Masculino 7337 <conf:(0.61)> lift:(1.08) lev:(0.02) [529] conv:(1.11)
11. Renda=1 A 2 SM 7582 ==> Tipo=CONCOMITANTE 4586 <conf:(0.6)> lift:(1.21) lev:(0.03) [786] conv:(1.26)
12. Genero=Masculino Tipo=INTEGRADO 6250 ==> Raca=Branca 3773 <conf:(0.6)> lift:(1.09) lev:(0.01) [315] conv:(1.13)
13. Curso=INFORMATICA 5207 ==> Raca=Branca 3121 <conf:(0.6)> lift:(1.08) lev:(0.01) [240] conv:(1.11)
14. Curso=INFORMATICA Tipo=INTEGRADO 5207 ==> Raca=Branca 3121 <conf:(0.6)> lift:(1.08) lev:(0.01) [240] conv:(1.11)
15. Curso=INFORMATICA 5207 ==> Tipo=INTEGRADO Raca=Branca 3121 <conf:(0.6)> lift:(2.04) lev:(0.07) [1592] conv:(1.76)
16. Genero=Masculino Raca=Parda 4275 ==> Tipo=CONCOMITANTE 2543 <conf:(0.59)> lift:(1.19) lev:(0.02) [400] conv:(1.23)
17. Tipo=INTEGRADO 11953 ==> Raca=Branca 7034 <conf:(0.59)> lift:(1.06) lev:(0.02) [421] conv:(1.09)
18. Renda=2 A 3 SM 5543 ==> Genero=Masculino 3213 <conf:(0.58)> lift:(1.02) lev:(0) [69] conv:(1.03)
19. Genero=Feminino Raca=Branca 5677 ==> Tipo=INTEGRADO 3261 <conf:(0.57)> lift:(1.15) lev:(0.02) [428] conv:(1.18)
20. Tipo=CONCOMITANTE Renda=1 A 2 SM 4586 ==> Genero=Masculino 2630 <conf:(0.57)> lift:(1.01) lev:(0) [29] conv:(1.01)
21. Renda=3 A 5 SM 4521 ==> Tipo=INTEGRADO 2592 <conf:(0.57)> lift:(1.15) lev:(0.01) [336] conv:(1.17)
22. Genero=Feminino Tipo=INTEGRADO 5703 ==> Raca=Branca 3261 <conf:(0.57)> lift:(1.03) lev:(0) [105] conv:(1.04)
23. Raca=Branca 13254 ==> Genero=Masculino 7577 <conf:(0.57)> lift:(1.01) lev:(0) [60] conv:(1.01)
24. Genero=Masculino 13587 ==> Raca=Branca 7577 <conf:(0.56)> lift:(1.01) lev:(0) [60] conv:(1.01)

Alterando as regras para 100, dez vezes mais que na amostragem anterior, com suporte mínimo de 1% e com confiança mínima de 50%, verifica-se a correlação entre três atributos: curso de informática, raça branca e gênero masculino, de acordo com as regras 2 e 3 destacadas.

A regra 4 aponta o gênero feminino e curso de informática no antecedente e como consequente o tipo do curso do Integrado, com confiança de 100%. Nesta análise é possível identificar um comportamento em que o gênero feminino aparece novamente somente na regra 19, com a raça branca, mostrando que o gênero feminino tem um comportamento diferente do masculino.

A partir da regra 5, a confiança atinge níveis de 60% a 55%. Esta regra indica como antecedente gênero masculino e renda de 1 a 2 salários mínimos e consequente tipo do curso concomitante, com 60% de confiança, mostrando que alunos do gênero masculino com rendas baixa optam pelo curso do tipo concomitante. A regra 16 descreve que alunos do gênero masculino e raça parda tendem a escolher o tipo de curso concomitante. Analisando essas regras conjuntamente, é possível delinear o perfil de alunos de cursos do tipo concomitante com renda baixa (1 a 2 salários mínimos) e raça parda.

Experimento 3: limite de 10 regras, $minsup = 0.1$, $minconf = 0.1$, atributos: raça e renda. Melhores regras obtidas:

1. Renda=3 A 5 SM 4521 \implies Raca=Branca 2791 <conf:(0.62)> lift:(1.12) lev:(0.01) [289] conv:(1.17)
2. Renda=2 A 3 SM 5543 \implies Raca=Branca 3074 <conf:(0.55)> lift:(1) lev:(0) [7] conv:(1)
3. Renda=1 A 2 SM 7582 \implies Raca=Branca 3762 <conf:(0.5)> lift:(0.9) lev:(-0.02) [-432] conv:(0.89)
4. Renda=1 A 2 SM 7582 \implies Raca=Parda 2767 <conf:(0.36)> lift:(1.13) lev:(0.01) [326] conv:(1.07)
5. Raca=Parda 7710 \implies Renda=1 A 2 SM 2767 <conf:(0.36)> lift:(1.13) lev:(0.01) [326] conv:(1.07)
6. Raca=Branca 13254 \implies Renda=1 A 2 SM 3762 <conf:(0.28)> lift:(0.9) lev:(-0.02) [-432] conv:(0.95)
7. Raca=Branca 13254 \implies Renda=2 A 3 SM 3074 <conf:(0.23)> lift:(1) lev:(0) [7] conv:(1)
8. Raca=Branca 13254 \implies Renda=3 A 5 SM 2791 <conf:(0.21)> lift:(1.12) lev:(0.01) [289] conv:(1.03)

Neste experimento podemos observar que a correlação renda e raça não são as mesmas das de raça e renda nas instâncias analisadas que somam 2396, observa-se nas três primeiras regras que a renda não ultrapassa a 3 a 5 salários mínimos que como consequente tem a confiança de 62% e nas regras 2 e 3 diminui a renda e diminui o percentual de confiança para entre 50% e 55%. Na sequência nas regras 6,7 e 8 a raça branca tem como consequente a renda em ordem crescente com confiança decrescente.

CONCLUSÕES

Este estudo nos indica que existe um perfil de aluno com características específicas em determinados cursos. O grau de interesse nos cursos é acentuado pela renda e raça. Há ainda cursos que são preferidos pelo gênero masculino e outros pelo feminino. As ferramentas utilizadas apresentam uma vasta gama de resultados com potencial de informações relevantes. Dentro de todo o contexto de dados e possíveis combinações de seleções, a ferramenta Weka abre um extenso leque de resultados. Iniciamos o caminho com estes estudos e desenvolvendo mais experimentos torna-se possível buscar informações até então desconhecidas.

Como trabalhos futuros, espera-se realizar mais experimentos, além de uma análise comparativa dos alunos inscritos no vestibular versus alunos aprovados, a fim de identificar o perfil dos alunos ingressos quando comparado com os inscritos.

AGRADECIMENTOS

Agradecimentos ao Programa Institucional de Bolsas de Iniciação Científica (PIBIC) do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Edital nº 262/2016 - por financiar esta pesquisa.

REFERÊNCIAS

AGRAWAL, R.; SRIKANT, R. **Fast Algorithms for Mining Association Rules in Large Databases**. Proceedings of the 20th International Conference on Very Large Data Bases. Anais...: VLDB '94. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1994.

- BALCÁZAR, J. L. **Parameter-free Association Rule Mining with Yacaree**. Extraction et gestion des connaissances (EGC'2011), Brest, France. **Anais...**: Revue des Nouvelles Technologies de l'Information.Hermann-Éditions, 2011
- BULDU, A.; ÜÇGÜN, K. Data mining application on students' data. **Procedia - Social and Behavioral Sciences**, v. 2, n. 2, p. 5251–5259, 2010.
- DE BAKER, R. S. J.; ISOTANI, S.; DE CARVALHO, A. M. J. B. Mineração de dados educacionais: oportunidades para o Brasil. **Revista Brasileira de Informática na Educação**, v. 19, n. 2, 2011.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, Palo Alto, USA, v. 17, n. 3, p.37-54, Set. 1996. Disponível em: <<https://www.aaai.org/ojs/index.php/aimagazine/article/viewFile/1230/1131>>. Acesso em: 03 ago. 2018.
- HALL, M. et al. The WEKA Data Mining Software: An Update. **SIGKDD Explor. Newsl.**, v. 11, n. 1, p. 10–18, 2009.
- HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques**. 3rd. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011.
- HAN, J.; PEI, J.; YIN, Y. **Mining Frequent Patterns Without Candidate Generation**. Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. **Anais...**: SIGMOD '00.New York, NY, USA: ACM, 2000.
- PEÑA-AYALA, A. Educational data mining: A survey and a data mining-based analysis of recent works. **Expert Systems with Applications**, v. 41, n. 4, p. 1432– 1462, 2014.
- ROMERO, C. et al. **Mining Rare Association Rules from e-Learning Data**. (R. S. J. de Baker, A. Merceron, P. I. P. Jr., Eds.)Educational Data Mining 2010, The 3rd International Conference on Educational Data Mining, Pittsburgh, PA, USA, June 11-13, 2010. Proceedings. **Anais...**2010.
- ZAKI, M. J. Scalable Algorithms for Association Mining. **IEEE Transactions on Knowledge and Data Engineering**, v. 12, n. 3, p. 372–390, 2000.
- ZORRILLA, M. E.; GARCÍA-SAIZ, D.; BALCÁZAR, J. L. **Towards Parameter-free Data Mining: Mining Educational Data with Yacaree**. (M. Pechenizkiy et al., Eds.)Proceedings of the 4th International Conference on Educational Data Mining, Eindhoven, The Netherlands, July 6-8, 2011. **Anais...**2011.