



APLICAÇÃO DO PROCESSO DE DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS NUM PROCESSO SELETIVO DE EDUCAÇÃO A DISTÂNCIA

Anderson Alex de Souza Sales – asales@iff.edu.br
Instituto Federal Fluminense – campus Bom Jesus do Itabapoana
Avenida Dário Vieira Borges, 235, Parque do Trevo
28360-000 – Bom Jesus do Itabapoana – RJ

Georgia Regina Rodrigues Gomes – georgia@ucam-campos.br
Universidade Candido Mendes
Rua Anita Peçanha, 100, Parque São Caetano
28030-335 – Campos dos Goytacazes – RJ

***Resumo:** O presente artigo apresenta uma aplicação prática do processo de descoberta de conhecimento em bases de dados com a empregabilidade do algoritmo de data mining Apriori por meio da ferramenta WEKA. O objetivo é de delinear o perfil dos candidatos que se submeteram a um processo seletivo na modalidade de educação a distância e analisar o impacto que as ações afirmativas geraram na classificação final. O estudo em questão fora realizado a partir da base de dados proveniente do sistema de inscrição online e da lista de divulgação do resultado final de um processo seletivo ocorrido no 2º semestre de 2012 numa instituição federal de ensino. A partir das informações obtidas foi possível gerar novos conhecimentos a respeito das características do perfil do público que pleiteia cursos dessa natureza o que permite, portanto, aos gestores do instituto de educação conhecerem-nas e terem subsídios na tomada de decisões estratégicas para futuros processos de seleção.*

***Palavras-chave:** Descoberta de conhecimento em bases de dados, Data mining, WEKA, Educação a distância*

1. INTRODUÇÃO

A revolução tecnológica a qual se presencia atualmente traz à tona mudanças em diversas áreas como medicina, economia, cultura e, certamente, a educação (SENE, 2012). Dessa forma, é observada uma forte tendência que constitui-se na valorização da educação formal. Essa propensão está relacionada à expansão da visão de comunidade, propiciada pelos avanços tecnológicos que se tornam disponíveis para uma gama cada vez maior das atividades



humanas. Sendo assim, uma parte da sociedade, carente de formação e/ou devido ao anseio de inserção e até mesmo objetivando melhores alocações no competitivo mercado de trabalho pode utilizar a seu favor um grande aliado: a Educação a Distância (MARTINHAGO, 2005).

A educação a distância corresponde a uma modalidade de ensino já utilizada há algumas décadas e é caracterizada basicamente pelo distanciamento, seja ele físico ou temporal, do professor e do aluno (DE SOUZA, 2007). Outra peculiaridade desse modo de ensino consiste no uso por parte dos educadores de ambientes de aprendizagem virtuais como um meio para disponibilizar acesso e informações pertinentes aos cursos ministrados. Entretanto, essas ferramentas virtuais oferecem pouco suporte para avaliar as atividades dos estudantes e discriminar as diferenças nos comportamentos *online* dos mesmos (ZAIANE & LUO, 2001).

Em contrapartida, nota-se que a empregabilidade de ferramentas tecnológicas no âmbito educacional pode auxiliar na obtenção de informações relevantes e a subsidiar o desenvolvimento de ações, tais como a minimização da reprovação ou evasão escolar e na adoção de estratégias específicas voltadas ao processo de ensino-aprendizagem (GOTTARDO *et al.*, 2012). Dessarte, a linha de pesquisa KDD (*Knowledge Discovery in Databases*), conceituada genericamente como o processo de descoberta de padrões não triviais de um conjunto de dados, pode ser utilizada para esse fim, auxiliando a tomada de decisões estratégicas (WITTEN *et al.*, 2005).

De acordo com a conjuntura apresentada este estudo visa a aplicar o processo de KDD na base de dados de um sistema de inscrição *online* e na lista que contem o resultado final de um processo seletivo de educação a distância de uma instituição federal de ensino. O objetivo principal se concentra nas inferências que a mineração de dados pode propiciar no delineamento do perfil dos candidatos inscritos e na correspondência que as ações afirmativas (pontuação extra para os candidatos oriundos do ensino público e trabalhadores em efetivo exercício de suas funções) geraram na classificação final do referido processo.

2. DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS

O processo de descoberta de conhecimento baseia-se na exploração de dados em bases que os armazenam para oportunizar a obtenção de informações, ora implícitas e desconhecidas, em conhecimento utilizável e compreensível que permita assistir a tomadas de decisões relevantes (GALVÃO & MARIN, 2009).

Esse processo de descoberta usa alguns princípios como métodos estatísticos e técnicas de inteligência artificial, e se subdivide nas seguintes etapas: seleção, pré-processamento, transformação, mineração de dados e interpretação (FAYYAD *et al.*, 1996). Tais etapas são ilustradas na Figura 1 e descritas a seguir:

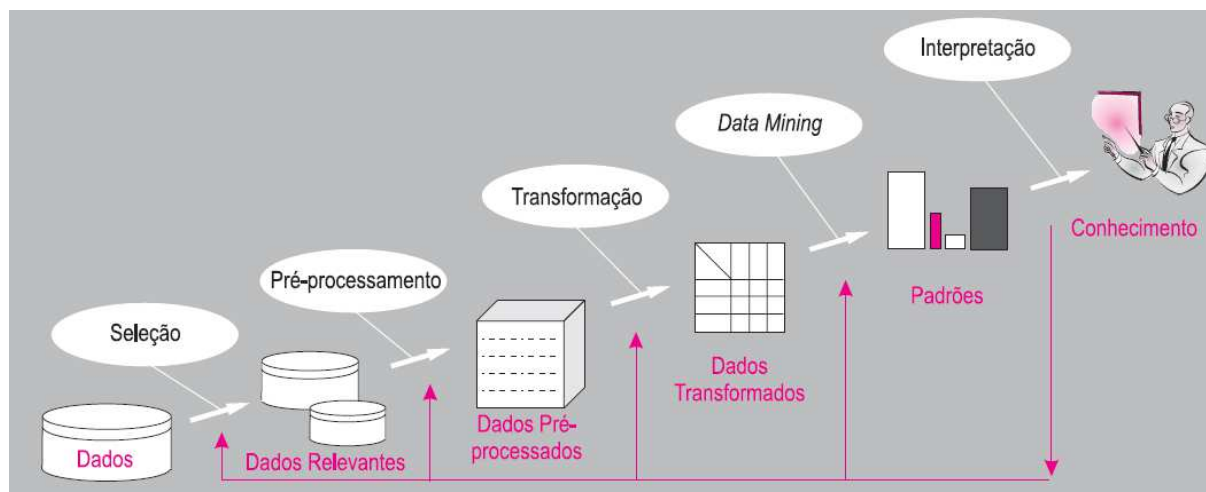


Figura 1 – Processo de KDD. Fonte: Fayyad *et al.* (1996).

1. Seleção: estágio onde são estabelecidos os objetivos almejados visando à aplicação dos algoritmos de mineração (DILLY, 1996; REZENDE, 2003);
2. Pré-processamento: nessa etapa, os dados considerados desnecessários são removidos, reduzindo o tamanho da base de dados e qualificando-os de forma que não sejam apresentados erros ou inconsistências (DILLY, 1996; GONÇALVES, 2000);
3. Transformação: nesse passo os dados são convertidos num modelo para que possam ser submetidos à etapa posterior (GONÇALVES, 2000);
4. *Data Mining*: é considerada a parte mais importante, sendo o núcleo do processo. Nela, são definidas as tarefas e técnicas e aplicados os algoritmos escolhidos sobre o modelo obtido. Durante esse procedimento, pode ser preciso acessar dados adicionais e/ou executar outras mudanças nos dados originalmente selecionados (BERRY & LINOFF, 1997);
5. Interpretação: consiste em validar o conhecimento extraído da base de dados, identificando e interpretando padrões em conhecimentos utilizáveis (DILLY, 1996).

Cabe ressaltar que, muitas vezes, os conceitos de descoberta de conhecimento em bases de dados e mineração de dados são empregados como sinônimos, o que é errôneo, pois como fora visto, o KDD compreende todo o processo de extração de conhecimento enquanto mineração de dados caracteriza-se por ser uma de suas etapas (CARVALHO, 2005).

O desenvolvimento da técnica de mineração de dados envolve a escolha de tarefas específicas de acordo com o objetivo a ser alcançado. Dentre as várias tarefas existentes, destaca-se a de associação, que consiste na geração de regras que descrevem o quanto a ocorrência de um conjunto de atributos existentes numa base de dados acarreta a presença de outro conjunto de atributos na mesma base. Essas regras são geralmente exibidas no formato $E \rightarrow D$ (E associa-se a D), que correspondem respectivamente ao lado esquerdo e direito da regra e indicam que se E existe em alguma transação, há uma probabilidade de R também ocorrer. O algoritmo Apriori, que será o utilizado nesse trabalho, é um dos algoritmos que implementa a tarefa de descoberta de associações na mineração de dados (REZENDE, 2003).



3. METODOLOGIA

A pesquisa em questão possui caráter quantitativo e exploratório, e as seções abaixo descrevem detalhadamente como fora o desenvolvimento da mesma.

3.1. Panorama do processo seletivo

O processo seletivo objeto de estudo ocorreu no 2º semestre de 2012 e contemplou somente cursos técnicos associados à modalidade de educação a distância de uma instituição federal de ensino. Tal processo possuiu 4664 inscrições constantes no sistema de inscrição *online*, sendo que 2550 efetivamente foram confirmadas posteriormente. Os cursos oferecidos foram os de Eventos, Guia de Turismo e Segurança do Trabalho, sendo distribuídos em 8 polos. Para uma melhor exemplificação, a Tabela 1 apresenta o número de vagas oferecidas por curso e polo e o número de inscrições confirmadas para cada um deles:

Tabela 1 – Relação de vagas ofertadas e candidatos inscritos por curso e polo

Curso	Polo	Nº de vagas	Nº de inscrições confirmadas
Eventos	Cabo Frio	75	53
	Campos - Centro	75	180
Guia de Turismo	Cabo Frio	75	56
	Itaperuna	75	34
Segurança do Trabalho	Bom Jesus do Itabapoana	50	247
	Cabo Frio	100	415
	Casimiro de Abreu	100	206
	Miracema	100	231
	Quissamã	50	175
	São João da Barra	100	953

O resultado final do processo seletivo indicou a seguinte relação: 626 candidatos convocados para matrícula, 419 candidatos aguardando reclassificação, 569 candidatos reprovados e 936 candidatos faltantes.

3.2. Procedimentos de coleta, limpeza e transformação dos dados



Os dados dos candidatos, provenientes do sistema de inscrição online, foram importados separadamente para cada curso no formato .csv (comma-separated values) da própria interface gráfica de administração desse sistema e, posteriormente, agrupados num único arquivo.

Para que a manipulação dos dados fosse realizada de maneira otimizada, foi desenvolvido um algoritmo na linguagem de programação Python™ que coletou, especificamente, os dados do arquivo que continha o resultado final do processo seletivo, pois o mesmo possuía a extensão .pdf (*portable document format*) e tamanho considerável. Com isso, foi possível integrar e salvar os dados do resultado final na mesma base obtida anteriormente.

Iniciando-se a fase de pré-processamento dos dados, foram avaliados e retirados os campos que pudessem identificar os candidatos e outros considerados desnecessários, sendo eles: Nome completo, Filiação, Telefone, E-mail, CPF, N° de documento de identificação com foto, Endereço, UF, ID (n° de inscrição), Inscrição confirmada, Local e endereço para a realização da prova e Pontuação na prova.

Ainda na referida etapa, os valores nominais do campo “Cidade” foram submetidos a uma correção ortográfica e padronizados, pois eram informados durante o preenchimento do formulário de inscrição e foram escritos das mais diversas formas pelos candidatos.

Seguindo a ideia de dinamismo, foi necessário adequar os dados relativos aos campos “ano de nascimento” e “idade” num único denominado “faixa_etaria”, e sua distribuição se apresenta da seguinte forma: abaixo de 24 anos, entre 25 e 34 anos, entre 35 e 64 anos e acima de 64 anos.

Finalizando a etapa de pré-processamento, os atributos considerados mais significativos foram confirmados, estando demonstrados a seguir: faixa_etaria, sexo, cidade, curso, polo, estudou_em_escola_publica, trabalha e resultado.

No que concerne a etapa de transformação, realizou-se o processo de conversão do arquivo .csv para o .arff (*Attribute-Relation File Format*), que consiste no padrão aceito pelo *software* de mineração. Para essa finalidade, usou-se uma ferramenta de conversão *online* denominada csv2arff (TKALCIC, 2014).

O *software* escolhido para realizar a mineração de dados foi o WEKA (*Waikato Environment for Knowledge Analysis*), que se constitui numa suíte de *data mining* que engloba um conjunto de implementações de algoritmos que aplicam diversas tarefas de mineração de dados. Foi desenvolvido na universidade de Waikato e escrito por meio da linguagem de programação Java™, e é mantido sob domínio da licença GPL (*General Public License*) (DAMASCENO, 2010).

4. RESULTADOS

Primeiramente, são elencadas as regras que apresentam algumas informações do perfil dos candidatos, sendo elas a faixa etária e o sexo, e suas correlações com as ações afirmativas. A métrica disponibilizada pelo WEKA e escolhida para análise foi o grau de confiança, que é exibido junto a cada uma das mesmas conforme explicitado na tabela 2:

Tabela 2 – Regras associativas do perfil dos candidatos e ações afirmativas



Regra 1	Confiança	99%
	Correlação	faixa etária abaixo de 24 anos E sexo feminino E não estudou em escola pública E não trabalha
	Descrição	considerando todos os cursos: de um total de 454 candidatos do sexo feminino que se encontravam abaixo de 24 anos e que não estudaram em escola pública, 449 não trabalhavam
Regra 2	Confiança	99%
	Correlação	faixa etária abaixo de 24 anos E sexo masculino E não estudou em escola pública E não trabalha
	Descrição	considerando todos os cursos: de um total de 289 candidatos do sexo masculino que se encontravam abaixo de 24 anos e que não estudaram em escola pública, 285 não trabalhavam
Regra 3	Confiança	97%
	Correlação	faixa etária entre 25 e 34 anos E sexo feminino E não estudou em escola pública E não trabalha
	Descrição	considerando todos os cursos: de um total de 535 candidatos do sexo feminino que se encontravam na faixa etária entre 25 e 34 anos e que não estudaram em escola pública, 520 não trabalhavam
Regra 4	Confiança	97%
	Correlação	faixa etária entre 25 e 34 anos E sexo masculino E não estudou em escola pública E não trabalha
	Descrição	considerando todos os cursos: de um total de 473 candidatos do sexo masculino que se encontravam na faixa etária entre 25 e 34 anos e que não estudaram em escola pública, 461 não trabalhavam
Regra 5	Confiança	98%
	Correlação	faixa etária entre 35 e 64 anos E sexo feminino E não estudou em escola pública E não trabalha
	Descrição	considerando todos os cursos: de um total de 320 candidatos do sexo feminino que se encontravam na faixa entre 35 e 64 anos e que não estudaram em escola pública, 314 não trabalhavam
Regra 6	Confiança	97%
	Correlação	faixa etária entre 35 e 64 anos E sexo masculino E não estudou em escola pública E não trabalha
	Descrição	considerando todos os cursos: de um total de 242 candidatos do sexo masculino e que se encontravam na faixa entre 35 e 64 anos e que não estudaram em escola pública, 234 não trabalhavam

Analisando as informações da Tabela 2 com o auxílio dos dados dispostos na interface *Explorer* do WEKA, que exhibe os dados numéricos de um atributo, observa-se uma preponderância dos candidatos do sexo feminino para qualquer faixa de idade. Já a faixa etária que compreende a maior parcela de candidatos corresponde a que se encontra entre 25 e 34 anos (43,8%), seguida da abaixo de 24 anos (31,72%). Para os candidatos acima dos 64 anos, apenas 11, não houve a geração de regras significantes.



Outra característica expressiva é a alta relação dos candidatos que não eram oriundos do ensino público (91,1 %) com aqueles que não possuíam ofício à época do processo seletivo (93,5%).

Nas tabelas a seguir são expostas informações pertinentes a cada um dos cursos ofertados no processo seletivo, iniciando-se pelo curso de Eventos, que abrange 9,13% (233 candidatos) da amostra total. As correspondências dos candidatos desse curso às ações afirmativas são vistas na Tabela 3:

Tabela 3 – Regras associativas do curso de Eventos e ações afirmativas

	Confiança	92%
Regra 7	Correlação	curso Eventos E não estudou em escola pública
	Descrição	de um total de 233 candidatos que escolheram o curso de Eventos, 215 deles não estudaram em escola pública
	Confiança	94%
Regra 8	Correlação	curso Eventos E não trabalha
	Descrição	de um total de 233 candidatos que escolheram o curso de Eventos, 220 deles não trabalhavam
	Confiança	89%
Regra 9	Correlação	curso Eventos E não estudou em escola pública E não trabalha
	Descrição	de um total de 233 candidatos que escolheram o curso de Eventos, 207 deles não estudaram em escola pública e não trabalhavam
	Confiança	94%
Regra 10	Correlação	curso Eventos E resultado aguardando reclassificação E não estudou em escola pública E não trabalha
	Descrição	de um total de 34 candidatos que escolheram o curso de Eventos e que aguardavam reclassificação, 32 deles não estudaram em escola pública e não trabalhavam
	Confiança	77%
Regra 11	Correlação	curso Eventos E resultado convocado E não estudou em escola pública E não trabalha
	Descrição	de um total de 104 candidatos que escolheram o curso de Eventos e que foram convocados, 80 deles não estudaram em escola pública e não trabalhavam

A maior correlação observada para o curso de Eventos é a de candidatos que não trabalhavam (94%), enquanto que os alunos provenientes do ensino público abrangiam uma taxa ligeiramente menor (92%). Associando-se todos os candidatos inscritos às duas ações afirmativas, 26 dos 233 (11%) atendiam as mesmas.

Em se tratando da análise da classificação no processo seletivo, 2 dos 34 candidatos que aguardavam reclassificação (6%) trabalhavam e eram originários de escola pública. Em



relação aos convocados para matrícula esse índice acresce para 23%, ou seja, 24 dos 104 no total foram favorecidos.

A Tabela 4, demonstrada abaixo, expressa as informações do curso de Guia de Turismo, que engloba a menor parcela dos candidatos da amostra total, contendo cerca de 3,52% dos mesmos (90 candidatos):

Tabela 4 – Regras associativas do curso de Guia de Turismo e ações afirmativas

	Confiança	94%
Regra 12	Correlação	curso Guia de Turismo E não estudou em escola pública
	Descrição	de um total de 90 candidatos que escolheram o curso de Guia de Turismo, 85 deles não estudaram em escola pública
	Confiança	99%
Regra 13	Correlação	curso Guia de Turismo E não trabalha
	Descrição	de um total de 90 candidatos que escolheram o curso de Guia de Turismo, 89 deles não trabalhavam
	Confiança	94%
Regra 14	Correlação	curso Guia de Turismo E não estudou em escola pública E não trabalha
	Descrição	de um total de 90 candidatos que escolheram o curso de Guia de Turismo, 85 deles não estudaram em escola pública e não trabalhavam
	Confiança	90%
Regra 15	Correlação	curso Guia de Turismo E resultado convocado E não estudou em escola pública E não trabalha
	Descrição	de um total de 49 candidatos que escolheram o curso de Guia de Turismo e que foram convocados, 44 deles não estudaram em escola pública e não trabalhavam

A correlação mais acentuada para o curso de Guia de Turismo corresponde aos candidatos que não trabalhavam (89 dos 90 que escolheram o curso), produzindo, portanto um grau de confiança de 99%. Com uma proporção um pouco menor, de 94%, ressaltam-se aqueles que não estudaram em escola pública, sendo 85 no quantitativo geral. Associando-se as duas ações afirmativas a todos os candidatos, 5 dos 90 inscritos (6%) as atendiam.

Quanto a observação da relação da classificação dos candidatos às ações afirmativas do curso em questão aponta-se que 10% deles atendiam as duas premissas: 5 de 49 convocados para matrícula. Cabe salientar que para esse curso não houve candidatos aguardando reclassificação.

A seguir, a Tabela 5 apresenta as informações pertinentes aos candidatos que escolheram o curso técnico de Segurança do Trabalho, que totalizaram 87,35% da amostra total (2227 candidatos):



Tabela 5 – Regras associativas do curso de Segurança do Trabalho e ações afirmativas

	Confiança	91%
Regra 16	Correlação	curso Segurança do Trabalho E não estudou em escola pública
	Descrição	de um total de 2227 candidatos que escolheram o curso de Segurança do Trabalho, 2024 deles não estudaram em escola pública
	Confiança	93%
Regra 17	Correlação	curso Segurança do Trabalho E não trabalha
	Descrição	de um total de 2227 candidatos que escolheram o curso de Segurança do Trabalho, 2076 deles não trabalhavam
	Confiança	89%
Regra 18	Correlação	curso Segurança do Trabalho E não estudou em escola pública E não trabalha
	Descrição	de um total de 2227 candidatos que escolheram o curso de Segurança do Trabalho, 1982 deles não estudaram em escola pública e não trabalhavam
	Confiança	87%
Regra 19	Correlação	curso Segurança do Trabalho E resultado aguardando reclassificação E não estudou em escola pública E não trabalha
	Descrição	de um total de 385 candidatos que escolheram o curso de Segurança do Trabalho e que aguardavam reclassificação, 335 deles não estudaram em escola pública e não trabalhavam
	Confiança	59%
Regra 20	Correlação	curso Segurança do Trabalho E resultado convocado E não estudou em escola pública E não trabalha
	Descrição	de um total de 473 candidatos que escolheram o curso de Segurança do Trabalho e que foram convocados, 278 deles não estudaram em escola pública e não trabalhavam

Iniciando-se a análise das informações existentes, visualiza-se que boa parte dos candidatos não eram provenientes de ensino público, onde 2024 possuíam essa característica, gerando um grau de confiança de 91%. De forma um pouco mais acentuada, com 93% de confiança, 2076 candidatos não estavam empregados à época do concurso. Correlacionado as duas ações afirmativas a todos os inscritos, 11% atendia as mesmas.

Quando examinado o relacionamento da classificação dos candidatos com as ações afirmativas, o curso de Segurança do Trabalho expressa a maior proporção dos três cursos ofertados: 13% para os candidatos que aguardavam reclassificação e 41% para os candidatos aprovados.

5. CONCLUSÃO



Diante do exposto a pesquisa apresentou a partir dos resultados obtidos características determinantes do perfil dos candidatos que se submeteram ao processo seletivo de educação a distância e o impacto real que a pontuação das ações afirmativas tiveram no resultado final.

Sendo assim, é factível afirmar que a pontuação extra não foi tão significativa na classificação dos candidatos convocados que optaram pelo curso de Guia de Turismo, que possuiu um índice 10% de beneficiados. Já para os cursos de Eventos e Segurança do Trabalho os índices foram mais avultados tanto para os candidatos que aguardavam reclassificação quanto para os convocados para matrícula, contando, respectivamente para cada curso e status de classificação, com 6% e 23% e 13% e 41% que usufruíram dessa bonificação.

Avaliando as informações do perfil dos candidatos, conjecturando-se todos os cursos, ratifica-se o predomínio do sexo feminino, que correspondeu a 65,6% do total de inscritos. Outro atributo considerável é a jovialidade em se tratando da faixa etária dos candidatos. Reitera-se ainda a alta correlação de todas as faixas de idade com a não procedência do ensino público e por não exercerem ofício, o que corrobora a manutenção e até mesmo a ampliação dentro da instituição de ensino de políticas de assistência estudantil a fim de atrair e incentivar públicos desfavorecidos economicamente a iniciarem, permanecerem e concluírem seus estudos.

Em se tratando do curso de Segurança do Trabalho, evidencia-se que sua procura é muito superior aos outros ofertados, inclusive por candidatos residentes fora do estado do Rio de Janeiro, como Espírito Santo, Minas Gerais, Distrito Federal, Goiás, Rio Grande do Norte e Rio Grande do Sul. Pelo fato do público interessado nesse curso ser jovem, denota-se que o mesmo é visto como uma excelente oportunidade de se ingressar numa área do mercado de trabalho muito promissora. O mesmo não se pode afirmar devido a baixa procura para os demais cursos, o que pode salientar aos gestores do instituto de educação a possibilidade de interrupção em determinados polos da oferta dos cursos de Eventos e Guia de Turismo em processos seletivos subsequentes.

6. REFERÊNCIAS

BERRY, Michael J.; LINOFF, Gordon. Data mining techniques: for marketing, sales, and customer support. 3.ed. Indianapolis: John Wiley & Sons, 1997. 888 p.

CARVALHO, Deborah Ribeiro. UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, Coordenação de Programas de Pós-Graduação em Engenharia. Árvore de decisão/ algoritmo genético para tratar o problema de pequenos disjuntos em classificação de dados, 2005. 173p, il. Tese (Doutorado).

DAMASCENO, M. Introdução à Mineração de Dados usando o Weka. Anais: V - Congresso Norte Nordeste de Pesquisa e Inovação. Maceió: IFAL, 2010.



DE OLIVEIRA, A. G.; GARCIA, D. F. Mineração de Base de Dados de um Processo Seletivo Universitário. INFOCOMP Journal of Computer Science, v.3, n.2, p. 38-43, 2004.

DE SOUZA, E. P. Avaliação Formativa em Educação a Distância via web. Anais: XIII – Associação Brasileira de Educação a Distância. Curitiba: CIETEP, 2007.

DIAS, M. M. *et al.* Aplicação de Técnicas de Mineração de dados no Processo de Aprendizagem na Educação a Distância. Anais: XIX – Simpósio Brasileiro de Informática na Educação. Ceará: UFC, 2008.

DILLY, Ruth. **Data Mining: An Introduction.** Disponível em: < http://www.adt.unipd.it/corsi/Bianco/www.pcc.qub.ac.uk/tec/courses/datamining/ohp/dm-OHP-final_1.html > Acesso em: 15 abr. 2014.

FAYYAD, Usama M. *et al.* Advances in knowledge discovery and data mining. Massachusetts: AAAI Press/MIT Press, 1996. 626p.

GALVÃO, N. D.; MARIN, H. F. Técnica de mineração de dados: uma revisão da literatura. Acta Paulista de Enfermagem, São Paulo, v.22, n.5, p. 686-690, 2009.

GONÇALVES, Alexandre L. UNIVERSIDADE FEDERAL DE SANTA CATARINA, Programa de Pós-graduação em Engenharia de Produção. Utilização de técnicas de mineração de dados em bases de C&T: uma análise dos grupos de pesquisa no Brasil, 2000. 132p. Dissertação (Mestrado).

GOTTARDO, E.; KAESTNER, C.; NORONHA, R. V. Avaliação de Desempenho de Estudantes em Cursos de Educação a Distância Utilizando Mineração de Dados. Anais do Workshop de Desafios da Computação Aplicada à Educação, v. 0, n. 0, p. 30–39, 17 jul. 2012.

GOTTARDO, E.; KAESTNER, C.; NORONHA, R. V. Previsão de Desempenho de Estudantes em Cursos EAD Utilizando Mineração de Dados: uma Estratégia Baseada em Séries Temporais. Anais do Simpósio Brasileiro de Informática na Educação, v. 23, n. 1, 2012a.

MARTINHAGO, Sergio. UNIVERSIDADE FEDERAL DO PARANÁ, Programa de Pós-Graduação em Métodos Numéricos em Engenharia. Descoberta de Conhecimento sobre o Processo Seletivo da UFPR, 2005. 125p, il. Dissertação (Mestrado).

PENEDO, J. R.; CAPRA, E. P. Mineração de Dados na Descoberta do Padrão de Usuários de um Sistema de Educação a Distância. Anais: VIII – Simpósio Brasileiro de Sistemas de Informação. São Paulo: EACH-USP, 2012.



REZENDE, Solange Oliveira. Sistemas inteligentes: fundamentos e aplicações.1.ed. Barueri: Editora Manole Ltda, 2003. 550 p.

SENE, J. E. DE. A SOCIEDADE DO CONHECIMENTO E AS REFORMAS EDUCACIONAIS. Revista Geotemas, v. 2, n. 1, p. 129–143, 24 jul. 2012.

TKALCIC, M. **csv2arff: Online CSV --> ARFF conversion tool**. Disponível em: <<http://slavnik.fe.uni-lj.si/markot/csv2arff/csv2arff.php>> Acesso em: 12 abr. 2014

WITTEN, Ian H.; FRANK, Eibe. Data Mining: Practical Machine Learning Tools and Techniques.2.ed. Boston, MA: Morgan Kaufmann, 2005. 525p.

ZAIANE, O. R.; LUO, J. Towards evaluating learners' behaviour in a Web-based distance learning environment. Anais: IEEE International Conference on Advanced Learning Technologies. Madison: University of Wisconsin, 2001.

APPLICATION OF THE KNOWLEDGE DISCOVERY IN DATABASES PROCESS ON A SELECTION PROCESS OF DISTANCE EDUCATION

Abstract: *This paper presents a practical application of the knowledge discovery in databases process with the employability of the Apriori data mining algorithm by WEKA tool. The objective is to delineate the candidate's profile who have submitted to a selective process in the modality of distance education and analyzing the impact that affirmative actions generated in the final classification. The study in question was realized on the databases of the online registry system and from the list that contains the final result of a selection process occurred in 2nd half of 2012 on a federal educational institution. From the obtained informations it was possible to generate new knowledge about the characteristics of public profile's that pleads study in courses of this nature which allows, however, managers from education's institute knows them and have subsidies in strategics decision-taking for future selection processes.*

Key-words: *Knowledge discovery in databases, Data mining, WEKA, Distance education*