

FORMA DE INGRESSO X SITUAÇÃO DOS ALUNOS NO CURSO SUPERIOR DE TECNOLOGIA EM SISTEMAS PARA INTERNET DO IFRS

Denise Luzia Wolff*

Cristina Almeida da Silva**

Mara Rosane Noble Tavares***

Resumo: Este artigo tem como objetivo identificar e analisar a relação da forma de ingresso com a situação inativa dos discentes no curso *Tecnologia em Sistemas Para Internet* do Instituto Federal do Rio Grande do Sul-Campus Porto Alegre. Para isso, utilizou-se o *software* RapidMiner e a base de dados do curso, disponível no Sistema Acadêmico da Instituição. Foram realizadas as cinco etapas do processo de KDD na base de dados do Sistema Acadêmico do curso, que foram: seleção, pré-processamento, transformação, mineração e avaliação dos dados. O KDD é o processo de descoberta de conhecimentos, que envolve etapas que vão desde a preparação da base de dados até a apresentação do conhecimento encontrado através das técnicas de mineração de dados. Atualmente, há a preocupação com o número crescente de alunos na ‘situação’ em recuperação, alunos que cursam semestres em atraso no curso. A descoberta de conhecimentos na base de dados do Sistema Acadêmico permitiu chegar à conclusão de que os alunos ingressados na instituição por vestibular, no segundo semestre de 2013 e pertencentes ao turno da manhã, apresentam uma probabilidade maior para entrar na situação de recuperação, indicando a necessidade de novas pesquisas para investigar o motivo.

Palavras-chave: Mineração de Dados Educacionais. KDD. Ingresso. Situação. RapidMiner.

1 Introdução

O presente artigo é o resultado de uma pesquisa sobre a descoberta de conhecimentos em base de dados educacionais, realizada no período de abril a julho de 2016. Apresenta a proposta de mineração na base de dados do Sistema Acadêmico¹ (SCA) do curso *Tecnologia em Sistemas Para Internet* do IFRS-Campus Porto Alegre, com o objetivo de identificar e

* Mestranda em Informática na Educação – Mestrado Profissional em Informática na Educação. Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul - *Campus* Porto Alegre - RS / Brasil. Email: <denise.wolff@poa.ifrs.edu.br>.

** Mestranda em Informática na Educação – Mestrado Profissional em Informática na Educação. Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul - *Campus* Porto Alegre - RS / Brasil. Email: <crisalm.silva@gmail.com>.

*** Mestranda em Informática na Educação – Mestrado Profissional em Informática na Educação. Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul - *Campus* Porto Alegre - RS / Brasil. Email: <maramestradoifrs@gmail.com>.

¹ O sistema acadêmico do IFRS-Campus Porto Alegre (SCA) é um sistema próprio, criado na década de 1980, pela equipe de Tecnologia da Informação da antiga Escola Técnica da Universidade Federal do Rio Grande do Sul. Este sistema passou por adaptações e atualizações ao longo destes anos.



analisar a relação das diferentes formas de ingresso com a situação inativa dos discentes no curso.

O curso tem a duração de três anos ou seis semestres e, de acordo com a Resolução nº 046, de 08 de maio de 2015², o aluno tem no máximo seis anos, ou doze semestres, para sua conclusão. É possível ingressar no curso de cinco formas: por vestibular, transferência externa, SISU³, exame de seleção, ingresso de diplomados e transferência interna. A pesquisa usou operadores do RapidMiner⁴ para realizar o processo de descoberta de conhecimentos em base de dados educacionais e responder à questão de pesquisa: “*qual a relação da forma de ingresso com a situação dos alunos no curso de Tecnologia em Sistemas para Internet?*”, pois há, atualmente na instituição, uma preocupação com o número crescente de alunos que não integralizam o curso.

Segundo Barth (2005), o RapidMiner é uma ferramenta que fornece a implementação de algoritmos. O algoritmo é uma sequência de raciocínios ou operações que se realiza para solucionar certo problema. No programa, eles são utilizados em problemas de aprendizagem de máquina com uma interface gráfica para um entendimento mais fácil e o desenvolvimento rápido de projetos.

A descoberta de conhecimentos em base de dados compreende várias etapas de depuração e investigação para identificar padrões de informações compreensíveis e úteis a partir de bases de dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996), chamado de *Knowledge Discovery in Databases* ou KDD. Em português, esse processo chama-se Descoberta de Conhecimento em Base de Dados.

A mineração de dados educacionais procura desenvolver ou adaptar métodos e algoritmos de mineração existentes, de tal modo que se prestem a compreender melhor os dados em contextos educacionais, produzidos principalmente por estudantes e professores, considerando os ambientes nos quais eles interagem (COSTA et al., 2012).

Ou seja, através das informações encontradas por meio da aplicação de processos de filtragem, agrupamentos, combinações, conversões de informações em algoritmos e análise

² Cf. <http://www.ifrs.edu.br/site/midias/arquivos/20154149565553od_versao_final_2.pdf>. Acesso em: 13 jul. 2016.

³ O Sistema de Seleção Unificada (Sisu) é o sistema informatizado, gerenciado pelo Ministério da Educação (MEC), pelo qual instituições públicas de educação superior oferecem vagas a candidatos participantes do Exame Nacional do Ensino Médio (Enem).

⁴ O RapidMiner é um premiado *software opensorce*, de mineração de dados, desenvolvido em Java, com licença *Affero General Public License* (AGPL). Há diversas extensões disponíveis como, integração com as ferramentas R e Weka, suporte à linguagem *Predictive Model Markup Language* (PMML), execução em paralelo em máquinas multi-core, suporte à mineração de texto e séries temporais (CAVALCANTI e PIRES, 2011).

dos resultados encontrados a partir de um banco de dados é possível prever padrões de comportamento e de aprendizagem.

A base de dados do SCA ofereceu dados como ‘modalidade de ingresso’; ‘forma de ingresso’; ‘situação do aluno’: em recuperação, alunos que não renovaram a matrícula no semestre e não formalizaram a situação de matrícula, encontrando-se em semestre diferente do previsto, de acordo com o ano de ingresso; em curso, semestre que o aluno se encontra; cancelado, alunos que formalizaram o desligamento do curso e trancado, alunos que formalizaram o afastamento do curso no semestre.

O processo de Descoberta de Conhecimento em Base de Dados (KDD) iniciou com a etapa de seleção dos dados, seguida do pré-processamento e limpeza da base, realizada com a seleção de operadores que permitiram filtrar apenas a ‘situação’ recuperação, sendo aplicada a técnica de árvore de decisão⁵ para categorizar os turnos manhã e noite na etapa final de mineração dos dados⁶. A avaliação dos resultados permitiu concluir que os alunos ingressantes por vestibular, no semestre 2013/2, do turno da manhã, pertencem ao grupo que apresenta maior risco de evasão ou de não conclusão do curso, por encontrar-se em ‘situação’ de recuperação. As seções a seguir apresentam a explicação dos conceitos envolvidos, as etapas, aplicações das técnicas e os resultados obtidos com o uso do *software* RapidMiner.

2 Conceitos envolvidos

De acordo com Costa et al. (2012), “uma grande quantidade de dados é produzida atualmente no setor educacional, necessitando ser analisada urgentemente”. O desafio que se apresenta, portanto, é “como explorar adequadamente esses dados, visando obter informação valiosa, considerando principalmente requisitos de qualidade de consistência e correção, de rápido tempo de obtenção e o seu caráter oportuno” (COSTA et al., 2012).

A Descoberta de Conhecimento em Base de Dados ou *Knowledge Discovery in Databases* (KDD), segundo Fayyad, Piatetsky-Shapiro e Smyth. (1996), “é um processo, de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de bases de dados”. O processo de KDD, aplicado na Mineração de Dados Educacionais (MDE), consiste das etapas de seleção das informações contidas na base de dados, pré-processamento dessas informações, transformação

⁵ Árvore de Decisão é um modelo prático, assim como uma tabela, usado para a inferência indutiva de informações.

⁶ Minerar significa extrair. A mineração de dados consiste em aplicar alguma funcionalidade para agregar e organizar informações, encontrando padrões, associações, mudanças e anomalias relevantes nessas informações.



dos dados em valores para sua padronização, mineração de dados, que significa a extração de informações dos dados, e avaliação do resultado encontrado, conforme explicação de cada etapa do processo de KDD, descrita a seguir:

- a) Seleção de Dados: Para iniciar o processo de descoberta do conhecimento, é necessário ter o domínio da aplicação sobre a qual é efetuado o trabalho, no caso a aplicação é educacional, a mineração de dados foi aplicada para se descobrir qual a relação entre a forma de ingresso dos alunos com a situação em que se encontram no curso. Essa delimitação é importante, porque é a partir do conhecimento prévio do assunto que se fará uma definição clara dos objetivos a serem atingidos. A partir do entendimento do domínio da aplicação, foi realizada a seleção de um conjunto de dados de acordo com os objetivos do processo;
- b) Pré-Processamento: Geralmente, os dados disponíveis não estão em um formato adequado para a análise final, podendo apresentar inconformidades tais como: duplicidade, falta de consistência, ruídos, que são dados faltando ou no formato incorreto, entre outros. Dessa forma, torna-se necessária a aplicação de métodos para tratamento, limpeza, integração, discretização. Aqui, discretização significa dividir ou particionar um todo em partes que apresentem menor complexidade para facilitar a realização de cálculos e a redução do volume de dados, a fim de preparar os dados selecionados para as próximas etapas de análise. Nesta etapa, na experiência relatada, realizou-se o agrupamento dos dados;
- c) Transformação: transformar os dados para formatos que sejam aceitos pelos algoritmos do programa a ser usado na etapa de mineração de dados;
- d) Mineração de dados: aplicação das técnicas e algoritmos adequados ao objetivo proposto para descobrir padrões, ou seja, para chegar à descoberta de conhecimento;
- e) Avaliação: interpretação do conhecimento extraído dos dados através da etapa de mineração de dados, com o intuito de verificar a relevância do conhecimento obtido.

A mineração de dados para a descoberta de conhecimentos tem origem nas áreas da Estatística, Inteligência Artificial e Aprendizagem de Máquina, sendo a MDE uma área recente de pesquisa, com o principal objetivo de desenvolver métodos para explorar conjuntos de dados produzidos por professores e estudantes e coletados em ambientes educacionais

(COSTA et al., 2012). Na seção a seguir, são descritos os procedimentos adotados na seleção dos dados do SCA.

3 Etapa de seleção dos dados

Durante o processo de importação da base de dados para o RapidMiner, o programa acusou erro na importação de alguns atributos (colunas), pois todos apresentavam o mesmo nome dos atributos (*label*), impedindo que o processo fosse completado. O erro acusado resultou na escolha entre duas alternativas - alterar manualmente o nome dos atributos ou excluí-los da base. Optou-se por excluí-los da base. Foram retirados os atributos: unidade da federação ('UF'), 'nome', 'nível', 'data de nascimento', 'estado civil', 'naturalidade', 'nacionalidade', 'número de identidade', 'data de expedição', 'órgão emissor', 'título de eleitor', 'zona/seção', 'data de emissão', 'CPF', 'código INEP', 'endereço residencial', 'número', 'complemento', 'bairro', 'CEP', 'cidade', 'e-mail', 'telefone', 'curso', 'modalidade' e 'turma'. Considerou-se que essas informações não eram relevantes para a resolução do problema proposto.

Após a manipulação dos dados, a importação da base foi efetuada com sucesso para o RapidMiner, como apresentado na Figura 1. Foram importados 187 registros, ainda com alguns ruídos, como valores desconhecidos nos atributos: 'cor raça', 'forma de ingresso' e 'modalidade de ingresso'.

Figura 1 – Resultado da segunda importação da base de dados com 187 registros após as correções dos valores, mas apresentando ainda ruídos a serem eliminados ou corrigidos.

ExampleSet (187 examples, 0 special attributes, 9 regular attributes)

Filter (187 / 187 examples):

Row No.	Ingresso	Sexo	Escolaridade	CorRaca	Turno	Curriculo	Formalngre...	ModalidadeL...	Situacao
1	2013/2	MASCULINO	ENSINO MÉD...	BRANCA	MANHA	2010/2	VESTIBULAR	ENSINO PUB...	RECUPERA...
2	2015/1	MASCULINO	ENSINO MÉD...	BRANCA	NOITE	2014/1	TRANSFERE...	ACESSO UNI...	RECUPERA...
3	2015/1	MASCULINO	ENSINO MÉD...	BRANCA	NOITE	2014/1	TRANSFERE...	ACESSO UNI...	3º SEMESTRE
4	2015/1	MASCULINO	ENSINO MÉD...	BRANCA	NOITE	2014/1	TRANSFERE...	ACESSO UNI...	3º SEMESTRE
5	2014/2	FEMININO	ENSINO MÉD...	BRANCA	NOITE	2014/1	VESTIBULAR	ENSINO PUB...	5º SEMESTRE
6	2014/1	MASCULINO	ENSINO MÉD...	BRANCA	MANHA	2010/2	TRANSFERE...	ACESSO UNI...	RECUPERA...
7	2016/1	MASCULINO	ENSINO SUP...	BRANCA	MANHA	2014/1	TRANSFERE...	ACESSO UNI...	2º SEMESTRE

Fonte: imagens extraídas do *software* RapidMiner, dados da base do SCA.

As informações dos atributos 'forma de ingresso' e 'modalidade de ingresso' eram muito relevantes para o processo, não sendo possível atribuir um valor padrão significativo para os mesmos, porém excluir todos os exemplos que continham valores desconhecidos



nesses atributos significava a exclusão de 24 linhas. Recorreu-se, então, a uma nova exportação de dados do sistema, dessa vez, com sucesso. Esses ruídos foram eliminados e na nova importação não havia nenhum atributo com valor desconhecido. Apesar da exclusão de vários atributos antes da importação da base, ainda foi necessário selecionar apenas os atributos relevantes para a pesquisa, como ‘ingresso’, ‘sexo’, ‘escolaridade’, ‘turno’, ‘currículo’, ‘forma de ingresso’, ‘modalidade de ingresso’ e ‘situação do aluno’. Os dados apresentados na seção a seguir resultam da etapa de pré-processamento, realizado para a extração de informações da base de dados do SCA.

4 Etapa de pré-processamento dos dados

A etapa do pré-processamento dos dados foi realizada com três técnicas. A primeira teve por objetivo a limpeza da base de dados antes da importação para o *software* RapidMiner, a segunda teve o objetivo de converter os dados de texto para valores numéricos e a terceira, objetivou gerar uma discretização para categorizar a relação da situação dos alunos no curso.

A primeira técnica consistiu em retirar manualmente os atributos irrelevantes para o problema proposto. Como havia alguns valores desconhecidos foi efetuada uma nova exportação da base de dados que, dessa vez, foi efetuada com 187 exemplos sem ruídos.

Na segunda técnica, converteram-se as informações de alguns atributos como ‘ingresso’, ‘currículo’, ‘escolaridade’, ‘forma’, ‘modalidade’, ‘sexo’, ‘situação do aluno’ e ‘turno’ em valores numéricos, para dar maior clareza na escolha daqueles atributos relevantes, cujas informações eram de interesse para responder qual a relação entre a forma de ingresso e a situação dos alunos no curso. Após a conversão para dados numéricos, apresentada no Quadro 1, as informações da base gerada pela grade do curso foram agrupadas por categorias.

Quadro 1 – Informações geradas pelo SCA, agrupadas por categoria e convertidas para dados numéricos para dar maior clareza na escolha dos atributos usados na mineração dos dados.

Atributo	Tipo de Dado	Valores Possíveis	Descrição
Ingresso	Polinomial	2010/2 2011/1 2011/2 2012/1 2012/2 2013/1 2014/1 2014/2 2015/1 2015/2	Ano e semestre de ingresso do aluno no curso.
Sexo	Polinomial	1: FEMININO 2: MASCULINO	Gênero do aluno.
Escolaridade	Polinomial	1: ENSINO MÉDIO 2: ENSINO MÉDIO INCOMPLETO 3: ENSINO SUPERIOR INCOMPLETO 4: GRADUACAO 5: MESTRADO	Escolaridade no momento do ingresso no curso.
CorRaca	Polinomial	1: BRANCA 2: INDÍGENA 3: NEGRA 4: PARDA 5: NÃO DECLARADA	Cor/raça a qual a pessoa se classifica.
Turno	Polinomial	1: MANHÃ 2: NOITE	Turno no qual o aluno cursa.
Currículo	Polinomial	2010/2 2013/1	Currículo e semestre em que o aluno ingressou.
FormalIngresso	Polinomial	1: DIPLOMADO 2: EXAME DE SELEÇÃO 3: SISU 4: TRANSFERÊNCIA EXTERNA 5: TRANSFERÊNCIA INTERNA 6: VESTIBULAR	Forma de ingresso do aluno no curso.
Situacao	Polinomial	1: 1º SEMESTRE 2: 2º SEMESTRE 3: 3º SEMESTRE 4: 4º SEMESTRE 5: 5º SEMESTRE 6: 6º SEMESTRE 7: RECUPERAÇÃO 8: REINGRESSO	Situação atual do aluno no curso

Fonte: Construção das autoras a partir da base do SCA do curso *Tecnologia em Sistemas Para Internet*.

A terceira técnica foi aplicada para remover outros atributos, irrelevantes para a solução do problema, através da clusterização. Clusterização é a atividade de agrupar coisas ou atividades semelhantes ou que se desenvolvem conjuntamente. Traduz a ideia de agrupamento, junção, união, agregação, integração. A clusterização dividiu os alunos em agrupamentos, conforme os padrões apresentados na situação do curso, gerando uma discretização (divisão) de quatro *clusters* (blocos de agrupamento) dos exemplos. Em mineração de dados:

Clusterização é a classificação não-supervisionada de dados, formando agrupamentos ou clusters. A análise de clusters envolve, portanto, a organização de um conjunto de padrões, usualmente representados na forma de vetores de atributos ou pontos em um espaço multidimensional – espaço de atributos, em clusters, de acordo com alguma medida de similaridade (MOSCATO e VON ZUBEN, 2002).

O objetivo dessa técnica foi categorizar a situação dos alunos, em recuperação ou não, a fim de prever, através dos resultados, quais alunos se encontravam em risco de evasão ou de não conclusão do curso. Também na discretização dos exemplos, os campos que continham duas informações ‘ingresso’ (ano) e ‘semestre’ (/1 ou /2), foram transformados em valores numéricos para facilitar o uso de alguns algoritmos que só trabalham com esse tipo de atributo.

Figura 2. Tabela com o resultado do agrupamento dos alunos em quatro blocos ou *clusters*.

Attribute	cluster_0	cluster_1	cluster_2	cluster_3 ↓
Curriculo = 2010/2	0	0	0	1
Turno = MANHA	1	0.027	0	1
SituacaoAluno = RECUPERAC...	0.226	0	0.127	0.952
ModalidadeIngresso = ACESS...	0.547	0.514	0.636	0.738
Sexo = MASCULINO	0.736	0.811	0.836	0.714
FormalIngresso = VESTIBULAR	0.472	0.486	0.564	0.571

Fonte: imagens extraídas do *software* RapidMiner, processo aplicado à base de dados do SCA.

Conforme apresenta a Figura 2, no *cluster* 0 (zero) os exemplos selecionados apresentaram alunos que estavam cursando no ‘currículo’ 2014/1, com a ‘situação do aluno’ no 1º semestre. Aqui há uma incidência de 0.226 alunos com a ‘situação’ recuperação. No *cluster* 1 (um), os exemplos selecionados são na maioria pertencente ao ‘turno’ da manhã e do ‘currículo’ 2014/1. Aqui não há alunos com a ‘situação’ recuperação. No *cluster* 2 (dois), os exemplos selecionados são, na maioria, do ‘turno’ da noite e do ‘currículo’ 2014/1. Aqui, há uma incidência de 0.127 alunos com a ‘situação’ recuperação. No *cluster* 3 (três) os exemplos selecionados são todos do ‘turno’ da manhã, do ‘currículo’ 2010/2 e aqui há uma incidência de 0.952 alunos com a ‘situação’ recuperação. São estudantes que apresentam a possibilidade de não completarem o curso, visto que a normativa prevê o dobro do tempo para a integralização. Na seção a seguir, serão apresentadas as informações extraídas no processo de KDD da base de dados do SCA.

5 Etapa de mineração de dados

Com o intuito de reduzir os exemplos, mantiveram-se os registros de interesse para a pesquisa, filtrando apenas os exemplos que estavam com a ‘situação’ recuperação, que foram reduzidos de 187 para 59 exemplos. A partir desse filtro, foram utilizadas algumas técnicas de

mineração para comparar os resultados obtidos. Empregou-se um operador para definir o *label* do atributo 'FormaIngresso' antes de utilizar o operador *DecisionTree* (árvore de decisão) para gerar o detalhamento apresentado no Quadro 2.

Quadro 2 – Resultado da ÁRVORE DE DECISÃO 1 apresentada em forma de quadro, mostra a combinação entre a modalidade de ingresso e a situação do aluno no curso.

MODALIDADE INGRESSO	VESTIBULAR	TRANSFERÊNCIA EXTERNA	SISU	EXAME DE SELEÇÃO	DIPLOMADO	TRANSFERÊNCIA INTERNA
ACESSO UNIVERSAL	20	5	13	0	0	1
ENSINO PÚBLICO S. M. < 1,5	4	0	2	0	0	0
ENSINO PÚBLICO S. M. = < 1,5 AD	4	0	0	0	0	0
ENSINO PÚBLICO S. M. > 1,5	4	0	4	0	0	0
ENSINO PÚBLICO S. M. > 1,5 AD	2	0	0	0	0	0

Fonte: Dados extraídos da árvore de decisão gerada pelo RapidMiner e convertida em quadro no Excel.

Analisando a árvore gerada, foi possível montar um quadro comparativo. Identificou-se que, dos 59 exemplos que estão na 'situação' recuperação, 34 ingressaram pelo vestibular, enquanto 19 ingressaram pelo SISU, 5 por transferência externa e apenas 1 por transferência interna. A tendência apresentada nesta árvore é de que a maioria dos exemplos que estão nessa situação ingressaram pelo vestibular. A mesma lógica foi aplicada para descobrir as tendências (Quadro 3) de oferta do curso nos turnos da manhã e da noite.

Utilizando-se dos mesmos operadores da Árvore de Decisão 1, acrescentou-se, ao processo, o turno. O resultado mostra uma concentração maior de exemplos no turno da manhã, contudo, deve-se levar em consideração que o turno diurno representa o turno de oferta original do curso, acontecendo a oferta para o turno noturno apenas a partir de 2013/2.

Quadro 3 – Resultado da ÁRVORE DE DECISÃO 2 e 3 apresenta, em forma de quadro, a combinação entre a modalidade de ingresso e a situação dos alunos que frequentam o curso no turno da manhã e no turno da noite.

Turno = MANHÃ						
MODALIDADE INGRESSO	VESTIBULAR	TRANSFERÊNCIA EXTERNA	SISU	EXAME DE SELEÇÃO	DIPLOMADO	TRANSFERÊNCIA INTERNA
ACESSO UNIVERSAL	20	4	12	0	0	0
ENSINO PÚBLICO S. M=< 1,5	3	0	2	0	0	0
ENSINO PÚBLICO S. M. =< 1,5 AD	3	0	0	0	0	0
ENSINO PÚBLICO S. M. > 1,5	3	0	3	0	0	0
ENSINO PÚBLICO S. M. > 1,5 AD	2	0	0	0	0	0
Turno = NOITE						
MODALIDADE INGRESSO	VESTIBULAR	TRANSFERÊNCIA EXTERNA	SISU	EXAME DE SELEÇÃO	DIPLOMADO	TRANSFERÊNCIA INTERNA
ACESSO UNIVERSAL	0	1	1	0	0	1
ENSINO PÚBLICO S. M=< 1,5	1	0	0	0	0	0
ENSINO PÚBLICO S. M. =< 1,5 AD	1	0	0	0	0	0
ENSINO PÚBLICO S. M. > 1,5	1	0	0	0	0	0
ENSINO PÚBLICO S. M. > 1,5 AD	1	0	1	0	0	0

Fonte: Dados extraídos da árvore de decisão gerada pelo RapidMiner e convertida em quadro no Excel.

Aqui, a tendência de maior incidência de alunos em recuperação pelo ingresso por vestibular, com frequência no turno da manhã foi de 31 alunos, seguido pelo SISU, 17 alunos e por transferência externa, 4 (quatro) alunos, não houve incidências nas modalidades por exame de seleção, diplomados e transferência interna. No turno da noite, a incidência foi menor, 4 (quatro) alunos por vestibular, 1 (um) por transferência externa, 2 (dois) por SISU e 1 (um) por transferência interna.

Utilizando-se ainda desses mesmos operadores, foi acrescentado, na pesquisa, o período de ingresso combinado com a modalidade de ingresso para encontrar o período com maior incidência da ‘situação’ recuperação e, portanto, com maior probabilidade de não integralização do curso, conforme apresentado no Quadro 4.

Quadro 4 – Resultado da ÁRVORE DE DECISÃO 4. O quadro representa o período de ingresso (Ano/Semestre) combinado com a forma de ingresso do aluno.

INGRESSO	VESTIBULAR	TRANSFERÊNCIA EXTERNA	SISU	EXAME DE SELEÇÃO	DIPLOMADO	TRANSFERÊNCIA INTERNA
2010/2	3	0	0	0	0	0
2011/1	0	1	0	0	0	0
2011/2	5	0	1	0	0	0
2012/1	0	1	0	0	0	0
2012/2	3	0	5	0	0	0
2013/2	12	0	7	0	0	0
2014/1	1	2	0	0	0	0
2014/2	4	0	4	0	0	1
2015/1	2	1	2	0	0	0
2015/2	4	0	0	0	0	0

Fonte: Dados extraídos da árvore de decisão gerada pelo RapidMiner e convertida em quadro no Excel.

Foi possível identificar que o período de ingresso 2013/2, na modalidade por vestibular, apresentou a maior incidência de alunos em recuperação, com 12 alunos ingressados por vestibular e 7 (sete) por SISU, seguido com quantidades menores, principalmente nas segundas ofertas de cada período.

6 Descoberta de conhecimento: considerações finais

Os conhecimentos descobertos através das cinco etapas do processo de KDD na base de dados do SCA podem, futuramente, ser utilizados em aplicações práticas, como investigações sobre os motivos da evasão dos alunos e na identificação do perfil desses alunos.

As duas primeiras etapas do processo de KDD, o trabalho de seleção e de pré-processamento dos dados, foram fundamentais para que os objetivos pretendidos na pesquisa fossem alcançados. As etapas de transformação e mineração dos dados permitiram uma maior acurácia, indicando a modalidade de ingresso dos alunos, o período e turno em que apresentaram maior índice da ‘situação’ recuperação, oferecendo uma avaliação clara do processo.

As três técnicas utilizadas no programa RapidMiner para a descoberta de conhecimentos na base de dados do SCA, durante a etapa de pré-processamento, permitiram realizar o processo de organização dos registros com velocidade e confiabilidade, pois o



software forneceu um ambiente integrado para a mineração e a análise preditiva do conjunto de dados fornecido.

Como resultado do processo de KDD, identificou-se que o maior grupo em ‘situação’ recuperação encontrava-se entre os alunos ingressados no curso, no semestre 2013/02, na modalidade por vestibular e no turno da manhã, permitindo responder à questão de pesquisa: “qual a relação da forma de ingresso com a situação dos alunos no curso de Tecnologia em Sistemas para Internet?”.

A relação encontrada, portanto, foi que a maioria dos alunos ingressados por vestibular no segundo semestre de 2013, para cursarem no turno da manhã, não integralizaram o curso por diversas razões. Razões essas que precisam ser investigadas no futuro, mas que, durante a pesquisa levantaram algumas hipóteses, como a de os alunos terem sido admitidos em outras instituições, de os alunos desistirem para trabalhar no horário da manhã, período original em que o curso foi oferecido, entre outras possibilidades. Os dados encontrados sugerem que esses alunos não renovaram a matrícula no semestre seguinte ou não formalizaram a situação de matrícula, encontrando-se em um semestre diferente do previsto, ou ainda que não permaneceram no curso, evadindo.

Foi possível concluir que todos os grupos de alunos ingressados por vestibular, no turno da manhã, com exceção dos períodos 2011/1 e 2012/1, são grupos de alunos que apresentam maior probabilidade de evasão ou de não conclusão do curso. Isso indica a necessidade de realizar pesquisas futuras, a partir dos conhecimentos encontrados na presente pesquisa, através da extração de informações na base de dados do SCA, para investigar os motivos da evasão e da não integralização desses alunos no curso.

ENTRY FORM “X” SITUATION OF STUDENTS IN THE HIGHER EDUCATION COURSE OF TECHNOLOGY SYSTEMS FOR THE INTERNET AT IFRS

Abstract: This article aims to identify and analyze the relation between the entry way and the inactive status of students in the course of Technology Systems for the Internet at the Federal Institute of Rio Grande do Sul - Campus Porto Alegre. Thus, we used the RapidMiner software and the course database of Academic System of the Institution. The five steps of the KDD process were followed in the Academic System database of the course, which were selection, pre-processing, processing, mining and data evaluation. KDD is the process of knowledge discovery, which involves steps ranging from the preparation of the database up until presentation of knowledges found through data mining techniques. Currently there is a concern about the increasing number of students in the recovery 'situation', students who attend semesters with delay in the course. The discovery of knowledges in the Academic

System database has led to the conclusion that students who entered in the institution through college entrance in the second half of 2013 and belonging to the morning shift are more likely to enter the recovery situation, indicating the need of further research to investigate the reason.

Keywords: Educational Data Mining. KDD. Entrance. Situation. RapidMiner.

Referências

BARTH, Fabrício J. **Mineração de regras de associação em servidores Web com RapidMiner**. 2005. Disponível em: <<http://fbarth.net.br/materiais/webMining/webUsageMining.pdf>>. Acesso em: 09 de jul. 2016.

CAVALCANTI, Elmano R.; PIRES, Carlos Eduardo. **Uma solução baseada em software livre para detecção e avaliação de cola em provas escolares**. 2011. Disponível em: <https://www.researchgate.net/profile/Elmano_Cavalcanti/publication/270510756_Uma_Solucao_Baseada_em_Software_Livre_para_Deteco_e_Avaliao_de_Cola_em_Provas_Escolares/links/54ac5caf0cf23c69a2b7be67.pdf>. Acesso em: 23 maio 2016.

COSTA, Evandro et al. **Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações. Jornada de Atualização em Informática na Educação**, v. 1, n. 1, p. 1-29, 2012. Disponível em: <<http://br-ie.org/pub/index.php/pie/article/view/2341>>. Acesso em: 09 de jul. 2016.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, v. 17, n. 3, p. 37, 1996. Disponível em: <<http://www.csd.uwo.ca/faculty/ling/cs435/fayyad.pdf>>. Acesso em: 23 maio 2016.

MOSCATO, P.; VON ZUBEN, F. J. **Uma visão geral de clusterização de dados**. DCA/FEEC/Unicamp–Campinas, 2002. Disponível em: <ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia368_02/topico5_02.pdf>. Acesso em: 09 de jul. 2016.