



Minerando dados sobre o desempenho de alunos de cursos de educação permanente em modalidade EAD: Um estudo de caso sobre evasão escolar na UNA-SUS

Susane Santos da Costa¹, Silvio Cazella², Sandro José Rigo¹

¹UNISINOS – Universidade do Vale do Rio dos Sinos

²UFCSPA – Universidade Federal Ciência e Saúde Porto Alegre

sarah_mei_suzi@hotmail.com, silvio.cazella@gmail.com, rigo@unisinobr

Resumo. Um dos problemas que as universidades brasileiras vêm enfrentando é o crescente número da evasão escolar. Pesquisas foram realizadas sobre este assunto afim de identificar padrões que expliquem este comportamento visando contribuir como uma forma de minimizar o impacto da evasão escolar. Este artigo apresenta uma pesquisa onde se buscou, através da aplicação do processo de Descoberta de Conhecimento em Bases de Dados, explicitar padrões de evasão nos cursos de educação permanente em modalidade EAD para profissionais da saúde, promovidos pela Universidade Aberta do SUS (UNA-SUS). Os dados foram analisados aplicando a tarefa de regras de classificação utilizando a técnica de árvores de decisão e como resultado obteve-se, um modelo preditivo de fácil entendimento com 97,6% de acerto na classificação do conjunto de treinamento. Palavras-chaves: Mineração de Dados Educacionais, Educação a Distância, Evasão.

Abstract. The data mining have been analyzed and we obtained a predictive model easy to understand, which ranked a potential of 97.3% accuracy, with the use of decision trees. One of the problems that Brazilian universities are facing is the increasing number of school dropouts. Much research has been conducted on this subject in order to identify patterns that explain this behavior and to minimize the impact of the dropout. This work presents a research with data mining techniques on data from the specialization courses in distance education promoted by Universidade Aberta do SUS (UNA-SUS), seeking to generate a model to identify the profile of students related to dropout. The data mining have been analyzed and we obtained a predictive model easy to understand, which ranked a potential of 97.6% accuracy, with the use of decision trees. Keywords: Educational Data Mining, Distance Education, Dropout.

1. Introdução

O tema evasão escolar tem aparecido com frequência nas discussões acerca da universidade, pois se tornou um fenômeno complexo e que está interferindo na gestão universitária por todo Brasil, seja em Instituições de Nível Superior (IES) públicas, seja IES privadas (RIBEIRO, 2005). Anualmente, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) realiza a coleta de dados sobre a educação superior com o objetivo de oferecer à comunidade acadêmica e à sociedade em geral informações detalhadas sobre a situação e as grandes tendências da área (INEP, 2013). O Ministério da Educação (MEC) investe em ações que facilitam o acesso às vagas de ensino no nível superior, através de programas como Universidade para todos (ProUni), Universidade Aberta do Brasil (UAB) e o Fundo de Financiamento ao Estudante do Ensino Superior (Fies), ou na expansão da rede

públicade ensino, por meio do Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (Reuni). As ações do programa contemplam o aumento de vagas nos cursos de graduação, tendo alcançado a marca de 200 mil vagas em 2011, ampliando a oferta de cursos noturnos, a promoção de inovações pedagógicas e o combate à evasão, entre outras ações, com o propósito de diminuir as desigualdades sociais no país (MEC, 2014). A Universidade Aberta do Sistema Único de Saúde (UNA-SUS), foi criada para atender às necessidades de capacitação e educação permanente dos profissionais do Sistema Único de Saúde (SUS), através do desenvolvimento da modalidade de educação à distância (EAD) na área da saúde, tendo observado um aumento na evasão escolar dos cursos ofertados, o que ocasiona uma perda considerável do investimento feito pelo governo nestes profissionais.

Buscando explicitar padrões de evasão a partir de base de dados, pode-se utilizar o processo de Descoberta de Conhecimento em Bases de Dados (DCDB), o qual se constitui, conforme definido por Fayyad (1996), em: "...um processo não trivial de identificação de novos padrões válidos, úteis e compreensíveis...". Segundo Cardoso e Machado (CARDOSO e MACHADO, 2008) o objetivo da mineração de dados, enquanto etapa do processo DCBD, é descobrir, de forma automática ou semiautomática, o conhecimento disponível em grandes massas de dados armazenadas nos bancos de dados das organizações, permitindo, desta forma, agilidade na tomada de decisão. Uma organização que emprega o processo de DCBD na análise de seus dados é capaz de: 1) criar parâmetros para entender o comportamento dos dados, os quais podem ser referentes a pessoas envolvidas com a organização; 2) identificar afinidades entre dados que podem ser, por exemplo, entre pessoas e produtos e ou serviços; 3) prever hábitos ou comportamentos das pessoas e analisar hábitos para se detectar comportamentos fora do padrão entre outros. De acordo com Rigo et al (2014), a comunidade de Informática na Educação tem realizado pesquisas que auxiliam na identificação de atributos componentes de grandes massas de dados educacionais com maior potencial para apoiar a análise de perfis de alunos evasores, desta forma contribuindo com o tratamento deste problema na Educação.

Este artigo descreve a aplicação da tecnologia de mineração de dados (etapa do processo de Descoberta de Conhecimento em Bases de Dados) para identificar o perfil destes alunos com potencial de evasão. O artigo descreve um estudo de caso com os dados fornecido pela Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSIPA), a fim de determinar o perfil do aluno propenso à evasão no que tange a cursos de especialização de educação permanente em saúde.

O artigo encontra-se organizado em sete seções incluindo a introdução e conclusão. A seção 2 aborda o referencial teórico utilizado na construção do conhecimento necessário para o desenvolvimento deste artigo. A seção 3 aborda os trabalhos relacionados. A seção 4 descreve os materiais e métodos. A seção 5 apresenta o estudo de caso e os dados estudados. A seção 6 apresenta a análise dos resultados obtidos. Por fim, a seção 7 descreve a conclusão do estudo realizado e sugere trabalhos futuros.

2. Referencial teórico

2.1. Evasão escolar no ensino superior

A evasão escolar no ensino superior é um problema não somente brasileiro, mas internacional que afeta o resultado dos sistemas educacionais. As perdas de



estudantes que iniciam, mas não concluem seus cursos, são desperdícios sociais, acadêmicos e econômicos. No setor público, são recursos públicos investidos sem o devido retorno. No setor privado, é uma importante perda de receitas. Em ambos os casos, a evasão é uma fonte de ociosidade de professores, funcionários e equipamentos (FILHO et al, 2007).

As ações do programa Federal denominado Reuni, contemplam o aumento de vagas nos cursos de graduação, batendo a marca de 200 mil vagas em 2011, a ampliação da oferta de cursos noturnos, a promoção de inovações pedagógicas e o combate à evasão, entre outras metas, com o propósito de diminuir as desigualdades sociais no país (MEC, 2014). O monitoramento da evasão é uma tarefa imprescindível para investigação das situações-limite e avanços do ensino-aprendizagem na modalidade a distância (OLIVEIRA et al, 2014).

2.2 UNA-SUS

O UNA-SUS foi criado pelo MEC em 2010 para atender às necessidades de capacitação e educação permanente dos profissionais de saúde que atuam no SUS. O Sistema é composto por três elementos: a) rede colaborativa de instituições de ensino superior que atualmente conta com 16 instituições de ensino superior (UNA-SUS a, 2013); b) acervo de Recursos Educacionais em Saúde (ARES) acervo público de materiais, tecnologias e experiências educacionais, construído de forma colaborativa, de acesso livre pela rede mundial de computadores (UNA-SUS a, 2013); c) plataforma Arouca: base de dados nacional, interligada ao sistema nacional de informação do SUS, contendo o registro histórico dos trabalhadores do SUS, seus certificados educacionais e experiência profissional (UNA-SUS a, 2013).

2.3 Descoberta de Conhecimento em Base de Dados

A mineração de dados constitui-se em uma das etapas de um processo denominado de Descoberta de Conhecimento em Base de Dados (DCBD). De acordo com Camilo e Silva (2009), a Mineração de Dados é uma das tecnologias mais promissoras da atualidade uma vez que permite extrair conhecimento a partir de bases de dados. Um dos fatores que reforçam a oportunidade no uso desta tecnologia, consiste na grande quantidade de dados atualmente disponível, ainda com pouca utilização efetiva (LAROSE, 2005). Han (2006) destaca essa situação e enfatiza a necessidade de ampliação do uso de Técnicas de Mineração de Dados em diversas áreas. Além da iniciativa privada, o setor público e o terceiro setor também são apontados como possíveis beneficiários da Mineração de Dados (WANG, 2013).

Segundo Cardoso e Machado (2008) o objetivo da mineração de dados é descobrir, de forma automática ou semi-automática, o conhecimento que está “escondido” nas grandes quantidades de dados armazenados nos bancos de dados da organização, permitindo agilidade na tomada de decisão.

3. Trabalhos Relacionados

Com base na disponibilidade de dados que atualmente é abundante quando são considerados os ambientes virtuais de ensino e aprendizagem, existem trabalhos já desenvolvidos com objetivo de aproveitar esta disponibilidade de dados para diversas

atividades. Alguns destes trabalhos são analisados brevemente a seguir, para enfatizar a linha de atuação adotada neste trabalho aqui apresentado.

Uma delas é a identificação de perfis de alunos que possam ser relacionados com a evasão e também possam ser utilizados para a geração de informações de uso pelos professores, em atividades preventivas. O trabalho de Kampff (2009) foi desenvolvido com este propósito, analisando possibilidades de utilização de informações diversas geradas em Ambientes virtuais de ensino e aprendizagem para a identificação de perfis de alunos que pudessem estar associados com evasão do curso. Além desta identificação, o trabalho também foi voltado para a geração de alertas para que professores e alunos pudessem se beneficiar destas informações em ações preventivas. O trabalho de Manhães, et al (2011) apresenta bons resultados com esta abordagem, em uma abordagem também identificada com a atividade de previsão do risco de evasão em cursos de Educação a Distância.

Também são conhecidos trabalhos de análise mais ampla de possibilidades que detalha e estuda as diferentes vantagens e possibilidades com a utilização de técnicas diversas de Mineração de Dados Educacionais no contexto da Educação, em especial a Educação a Distância, devido à sua ampla utilização de mediação digital. Um destes trabalhos é o trabalho de Baker, Isotani e Carvalho (2011) que apresenta uma análise muito positiva de possibilidades nesta área, que são comprovadas posteriormente, conforme será observado. O trabalho de Romero e Ventura (2013), aborda este mesmo contexto e amplia a análise de possibilidades, trazendo para a discussão aspectos mais específicos das técnicas de Mineração de Dados e suas adaptações.

Por fim, o trabalho de Cambruzzi et al (2012) defende ser possível a construção de sistemas voltados para a predição de evasão e o seu uso em Educação a Distância a partir de um contexto mais amplo, com o uso e integração de diversas fontes de dados, o que se torna cada vez mais uma necessidade, tendo em vista a grande quantidade de dados disponíveis atualmente em diversos sistemas de informação voltados para o tratamento de aspectos complementares, tais como o aspecto de interação dos alunos, observado em ambientes virtuais de ensino e aprendizagem, ou então, como o aspecto histórico, geralmente disponibilizado a partir de sistemas de controle acadêmico.

Esta breve revisão de aspectos de trabalhos na área corrobora a tendência de utilização destes recursos para a área de Mineração de dados Educacionais e sua associação com metodologias de prevenção da evasão.

4. Materiais e Métodos

Esta pesquisa caracteriza-se por ser uma pesquisa exploratória de acordo com seu objetivo, de natureza aplicada, com abordagem quantitativa uma vez que métricas são assumidas para avaliar os modelos obtidos. O método de pesquisa aplicado foi estudo de caso.

Os materiais utilizados na pesquisa foram provenientes dos cursos de especialização em EAD do UNA-SUS disponibilizados pela coordenação dos cursos junto a Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA).

5. Estudo de Caso

O estudo de caso foi realizado junto aos cursos de especialização a distância do UNASUS, ofertados pela UFCSPA. O objetivo da aplicação do Processo de Descoberta

de Conhecimento em bases de dados, com foco no resultado da etapa de Mineração de dados, foi a geração de um modelo preditivo que representasse a evasão escolar referente ao estudo de caso e tendo o intuito de apresentar o conhecimento explicitado de forma inteligível aos especialistas do negócio.

Para realizar a mineração dos dados provenientes destes cursos, realizou-se primeiramente a extração dos dados, que foram estudados e analisados. Para este estudo de caso a UFCSPA forneceu uma apresentação em PowerPoint que explicava o Sistema de notas e avaliações do Curso de Especialização em Saúde da Família – UNASUS/PROVAB(Programa de Valorização da Atenção Básica), também foi fornecido um arquivo Excel com planilhas contendo dados sobre avaliação de desempenho de 249 alunos e sua situação no curso. Estes dados eram referentes às turmas de 2013.

5.1 Sistema de notas e avaliações

O sistema de notas e avaliações aplicado pela UNA-SUS é dividido em três eixos formadores, sendo o eixo um a introdução à Educação a Distância, apresentando o curso, seus objetivos e sua organização e tem a duração de dez horas. O eixo dois é formado por três unidades: 1) Organização da Atenção à saúde com a duração de quarenta horas, 2) Epidemiologia aplicada com duração de sessenta horas e 3) Interações em APS com duração de oitenta horas. Ao todo o eixo 2 também chamado de Campo de Saúde Coletiva contém cento e oitenta horas e seu processo avaliativo consiste em avaliação das atividades somativas à distância e avaliações presenciais com médias separadas para cada uma das três unidades. Exige-se média superior a setenta para aprovação em cada unidade (ou sessenta caso o aluno passe por exame), a figura 1 apresenta o fluxo do processo avaliativo do eixo 2.

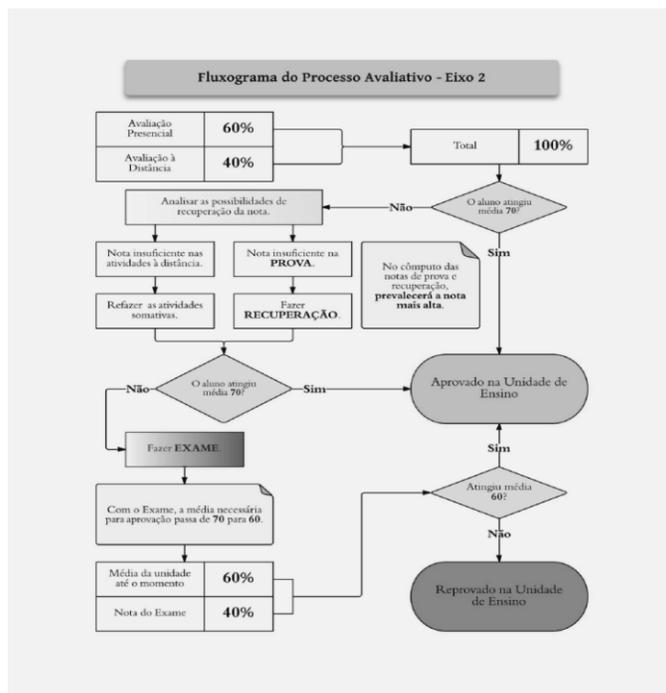


Figura 1. Fluxograma do Processo Avaliativo – Eixo 2



O eixo 3 também chamado de Núcleo Profissional é Composto por 25 Casos Complexos avaliados, 5 Casos Complexos complementares e 5 itens do Portfólio com duração de cento e oitenta horas. O processo avaliativo constitui-se na média entre a avaliação das atividades somativas dos Casos Complexos avaliados e a avaliação dos itens do Portfólio. A nota resultante deverá ser superior a 70. A equação 1 apresenta como o processo avaliativo do eixo três é calculado.

$$\text{NotaEixoTrês} = (\text{CC} + \text{P})/2 \quad (1)$$

Onde:

CC: Soma das avaliações dos casos complexos;

P: Soma das notas dos portfólios.

Por fim o trabalho de conclusão do curso (TCC) que é constituído pelo escrito final que condensa atividades desenvolvidas durante o Eixo 3 (itens do portfólio) e pela apresentação oral avaliada pela Comissão Avaliadora com carga horária de vinte horas e processo avaliativo a média entre a nota do Portfólio entregue e da apresentação oral. A nota resultante deverá ser superior a 70. A equação 2 apresenta como o processo avaliativo do TCC é calculado.

$$\text{NotaTCC} = (\text{P} + \text{AP})/2 \quad (2)$$

Onde:

P: Soma das notas dos portfólios;

AP: Nota da apresentação oral.

5.2. Dados Minerados

Os dados fornecidos pela UFCSPA para a mineração foram divididos em planilhas contendo o nome, a sede e o tutor de cada aluno. As planilhas de avaliações continham as notas presenciais, EAD e recuperação. Planilhas de acesso informavam quantas vezes por mês o aluno acessou as salas virtuais e o apoio acadêmico informava se o aluno havia evadido ou não, assim como se o desempenho do aluno havia sido satisfatório ou insatisfatório.

Atualmente os dados estão disponíveis em extensas planilhas dificultando a visualização dos mesmos. A tabela 1 apresenta parcialmente as variáveis contidas nas planilhas e suas tipagens. A coluna Variável representa as variáveis que foram selecionadas na etapa de pré-processamento no processo de DCBD. As variáveis Nota final UE1, UE2 e UE3 foram escolhidas, pois possuem o valor da avaliação EAD mais a avaliação presencial no Eixo 2, Casos Complexos e Portifólio são avaliações do Eixo 3, Apresentação Oral e avaliações do TCC, e a variável Status que informa se o aluno é ou não evasor. Este é o formato do arquivo de treinamento para geração do modelo preditivo, onde UE1, UE2, UE3, Casos Complexos, Portifólio, Apresentação Oral e avaliações do TCC são entendidos como atributos preditivos e Status constitui-se em atributo objetivo.

Tabela 1. Variáveis e tipos

Variável	Tipo
UE1	Real
UE2	Real
UE3	Real
CasosComplexos	Real



Portfolio	Real
Apresentação Oral	Real
TCC	Real
Status	String

5.3. Aplicando o Processo DCBD

A seguir será descrita a aplicação detalhada do processo de Descoberta de Conhecimento em Base de Dados (DCBD) nos dados alvo de análise.

5.3.1. Seleção da base de dados

Seguindo o processo DCDB, a primeira etapa realizada foi a seleção de dados. Os dados utilizados nesta pesquisa foram dados amostrais referentes a 249 alunos da turma de 2013, de pós graduação lato sensu, que foram disponibilizados pela coordenação dos cursos. Os alunos não foram identificados nestes dados, garantindo-se a privacidade completa dos mesmos. O fluxo da seleção de dados aplicados nesta pesquisa envolveu os dados referentes a avaliação do curso e os dados sobre avaliação de desempenho dos alunos, que compuseram a base de dados utilizada na mineração de dados.

5.3.2. Pré-Processamento

A base de dados foi analisada, e no primeiro momento optou-se por usar os seguintes dados de cada aluno (atributos a serem analisados): 1) a soma das avaliações; 2) as avaliações finais e 3) o status de desempenho. Logo percebeu-se que não era necessário uso da soma das avaliações, pois este atributo era derivado das avaliações finais. Em seguida adaptou-se os dados do status de desempenho para que este fosse o atributo objetivo da mineração (os demais configuraram os atributos preditivos). Nesta etapa foram excluídos os dados dos alunos que possuíam status “licença saúde”, pois os mesmos não eram relevantes para o objetivo da mineração de dados.

5.3.2.1 Transformação

Durante a etapa de transformação os dados selecionados foram convertidos em um arquivo tipo ARFF e CSV para serem operados na ferramenta WEKA (ferramenta na qual foi executada a mineração dos dados).

5.3.3. Mineração de Dados

Nesta etapa foi utilizada tarefa de classificação de dados através do uso do algoritmo J48 para minerar os dados. A saída deste algoritmo permite visualizar o modelo preditivo, em formato de árvore de decisão. Foram utilizadas 248 instâncias no conjunto de treinamento.

5.3.4.Pós-Processamento

Ao final da mineração de dados, obteve-se um modelo preditivo que foi analisado através das métricas oferecidas pela própria ferramenta WEKA. No que tange a número de instâncias corretamente classificadas, obteve-se um modelo com 97,6% de classificações corretas. A análise de resultados será detalhada na seção 6.

6. Análise de Resultados

O treinamento do modelo foi realizado com um conjunto de 248 instâncias, e neste treinamento foi utilizado o método de teste denominado cross-validation, assumindo o valor de 10 folds. O modelo preditivo gerado classificou corretamente 242 (97,6%) instâncias e incorretamente 6 (2,4%) instâncias. A figura 2 apresenta o resultado obtido confirmando a percepção inicial de que o desempenho do aluno está relacionado à tendência de evasão.

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      242          97.5806 %
Incorrectly Classified Instances     6           2.4194 %
Kappa statistic                    0.9225
Mean absolute error                 0.0324
Root mean squared error             0.155
Relative absolute error             10.1783 %
Root relative squared error         38.9265 %
Coverage of cases (0.95 level)     97.5806 %
Mean rel. region size (0.95 level) 50.2016 %
Total Number of Instances          248

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
          0,990   0,082   0,980     0,980   0,980     0,923  0,927   0,963   Ativo
          0,918   0,010   0,957     0,957   0,957     0,923  0,927   0,910   Evasor
Weighted Avg.   0,976   0,067   0,976     0,976   0,976     0,923  0,927   0,952

=== Confusion Matrix ===

  a  b  <-- classified as
197  2 | a = Ativo
  4 45 | b = Evasor
```

Figura 2. Resultado obtido

Analisando-se a Matriz de Confusão (confusionmatrix) observa-se que das 248 instâncias do conjunto de treinamento que representavam os alunos ativos, 242 instâncias foram classificadas corretamente pelo algoritmo J48, e 06 instâncias foram incorretamente classificadas (recaindo em classe de Evasor, sendo um falso positivo para este classificador). Outras 02 instâncias são falsos positivos, e foram classificadas na classe Ativo. Das 49 instâncias que representam o conceito de Evasor, 04 instâncias foram corretamente classificadas, e duas foram incluídas nesta classe como falso positivo. Quanto à acurácia do modelo preditivo, percebe-se que tanto a precisão (precision) como a abrangência (recall) apresentaram percentuais satisfatórios, conforme apresentado na figura 2.

A figura 3 apresenta a árvore de decisão gerada pela mineração dos dados. Onde atributo UE3 representa o primeiro nodo da árvore de decisão gerada, onde caso o aluno tenha obtido nota final menor ou igual a 24.65 este aluno pode ser classificado como evasor ou ativo, porém com desempenho insatisfatório, quando a nota final é maior que 24.65 o aluno é considerado ativo. Assim sendo o modelo evidencia que alunos com desempenho insatisfatório são candidatos a evasão.

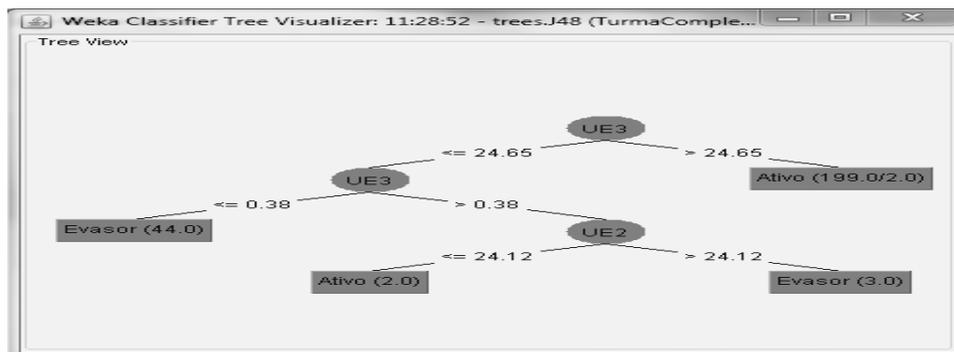


Figura 3 – Árvore de Decisão

7. Conclusões

A evasão escolar é um problema social. Na UNA-SUS nos cursos na modalidade EAD este fato vem sendo monitorado através do armazenamento de dados com informações sobre as avaliações e as iterações no Moodle dos alunos. A fim de subsidiar o desenvolvimento deste trabalho, que teve como objetivo principal gerar um modelo que auxilie os profissionais da educação na UNA-SUS a identificar alunos candidatos à evasão. A mineração de dados foi aplicada para classificação de perfis de alunos com potencial de evasão. O processo de DCDB foi utilizado para permitir a mineração dos dados fornecidos pela UFSCPA. Iniciou-se com o pré-processamento da base de dados definindo o quanto da base de dados seria utilizado e quais dados seriam minerados. Em seguida realizou-se a mineração e análise dos dados obtidos. Para conclusão deste trabalho os dados obtidos da mineração, foram analisados e foi obtido um modelo preditivo de fácil entendimento e simples aplicação, que classificou um potencial de evasão de 97,6% de precisão.

Este artigo não teve como objetivo apresentar um método para solucionar o problema da evasão. Ele se limitou a apresentar um modelo resultante da mineração dos dados fornecidos pela UFSCPA, que se restringiu apenas a universidades que estão autorizadas a trabalhar EAD para extensão e turmas de 2013, baseando se apenas nos resultados obtidos pelos alunos de pós graduação lato sensu.

Como trabalhos futuros pretende-se ampliar a base de dados (amostra) e utilizar as informações sobre as iterações destes alunos na plataforma Moodle, buscando gerar um modelo que compare o desempenho nas avaliações e o comportamento no Moodle.

Agradecimentos

Agradecemos a coordenação do UNA-SUS da UFSCPA pelo apoio e disponibilização dos dados, bem como a toda equipe técnica do UNA-SUS pelo apoio.

Referências

- BAKER, R. ISOTANI, S. CARVALHO, A. **Mineração de Dados Educacionais: Oportunidades para o Brasil**, Revista Brasileira de Informática na Educação v. 19, n.2, 2011.
- CAMBRUZZI, W. L.; MORAES, R. de; LEITHARDT, V. R. Q.; MENDES, C.; GEYER, C. F. R.; COSTA, C. A. da; BARBOSA, J. L. V.; RIGO, S. J. **Um Modelo para Gerenciamento de Múltiplas Trilhas Aplicado a Sistemas de Apoio à Educação**. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 2012, Rio de Janeiro. Anais. . . [S.l.: s.n.], 2012
- CAMILO, C. O., SILVA, J. C. - **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas- Mestrando em Ciência da Computação** - INF/UFMG Orientador - INF/U - RTINF_001-09- Relatório Técnico 2009
- CARDOSO, O. N. P., MACHADO, R. T. M. **Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras**. Revista de Administração Pública – Rio de Janeiro, 2008.
- FAYYAD, U. M., G. PIATETSKY-SHAPIO, P. SMYTH, R. UTHURUSAMY (1996), **Advances in Knowledge Discovery and Data Mining**, AAAI/MIT Press. FEELDERS, A., H. Daniels e M. Holsheimer (2000), “Methodological and practical aspects of data mining”, Information & Management, Vol. 37, pp. 271-281
- FILHO, R., M. HIPÓLITO, P. R. LOBO, O, Maria Beatriz C.M. **A Evasão no ensino superior brasileiro** Recebido em: janeiro 2007, Aprovado para publicação em: fevereiro 2007
- HAN, J; KAMBER, M. **Data Mining: Concepts and Techniques**. Elsevier, 2006.
- INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). Censo da Educação Superior Brasília DF, [2011?]. Disponível em: <<http://portal.inep.gov.br/web/censo-da-educacao-superior>> Acesso em: 10 ago. 2014
- KAMPFF, A. J. C. **Mineração de Dados Educacionais para Geração de Alertas em Ambientes Virtuais de Aprendizagem como Apoio à Prática Docente**. RENOTE, v. 6, n.2, 2009.
- LAROSE, D. **Discovering Knowledge in Data: An Introduction to Data Mining** John Wiley and Sons, Inc, 2005
- MANHÃES, L. M. B.; CRUZ, S. M. S. da; COSTA, R. J. M.; ZAVALETA, J.; ZIMBRÃO, G. **Previsão de Estudantes com Risco de Evasão Utilizando Técnicas de Mineração de Dados**. In: XXII SBIE - XVII WIE, 2011, Aracaju. Anais. . . [S.l.: s.n.], 2011. p. 150–159.
- Ministério da Educação e Cultura (MEC) Educação ___, ___ Disponível em: <<http://portal.mec.gov.br> > Acesso em: 15 ago. 2014
- OLIVEIRA, A.E.F, FRANÇA, R.M, FERREIRA, E.B., JUNIOR, M.H. **Avaliação da Produtividade no Processamento dos Dados em um Curso de Pós-graduação Lato Sensu da UNASUS/UFMA na Modalidade a Distância utilizando o Sistema de Monitoramento (SIM)** Disponível em: <<http://www.telessaude.uerj.br/resource/goldbook/pdf/27.pdf>> Acesso em 28 jul 2014
- RIBEIRO, M. A. - **O Projeto Profissional Familiar como Determinante da Evasão Universitária- Um Estudo Preliminar** – Revista Brasileira de Orientação Profissional, 2005, 6(2), pp. 55 – 70
- RIGO, S., BARBOSA, J., CAMBRUZZI, W., CAZELLA, S. **Minerando Dados Educacionais com foco na evasão escolar: oportunidades, desafios e necessidades**. Revista Brasileira de Informática na Educação, v.33, n. 01, 2014.



ROMERO, C.; VENTURA, S. **Data mining in education**. Wiley Interdisciplinary y Reviews:Data Mining and Knowledge Discovery, [S.l.], v. 3, n. 1, p. 12–27, 2013. Universidade Aberta do SUS (UNA-SUS). O que é a UNA-SUS? Brasília DF, [2013?]. Disponível em: <<http://www.unasus.gov.br/node/1>>Acesso em:12 out. 2013.