

Uma Proposta para Identificação de Causas da Evasão na Educação a Distância através de Mineração de Dados

Warley Barbosa¹, Douglas Máximo¹, Anthony Jatobá¹, Anderson Leite¹ e Elvys Soares¹

¹Instituto Federal de Alagoas (IFAL) - Campus Palmeira dos Índios
Av. Alagoas, s/n - Palmeira de Fora - Palmeira dos Índios - AL

Abstract. *This work describes a proposal for using Data Mining Techniques to identify evasion causes and to assess the evading student profile on a distance learning project for teaching the English Language at the Federal Institute of Science, Education and Technology of Alagoas. Causes and necessities are described, as well as more elaborated descriptions about the project and its objectives, methodology and the expected results.*

Resumo. *Este trabalho descreve uma proposta para uso de técnicas de Mineração de Dados para identificação de causas de evasão e de levantamento do perfil do aluno evadido num projeto de ensino de língua inglesa à distância do Instituto Federal de Educação Profissional, Científica e Tecnológica de Alagoas. São apontadas causas e necessidades, assim como uma descrição mais ampla do projeto, seus objetivos, metodologia e os resultados esperados.*

1. Introdução

Recentemente, com o advento de novas tecnologias, como a Internet, o processo de ensino/aprendizagem deixou de ser limitado à sala de aula ou o contexto físico e/ou temporal entre o professor e o aluno [Moran 2008]. A essa modalidade, que já foi implementada com sucesso no século XX na britânica Open University, na Fern Universität¹ alemã, bem como há poucos anos na Universidade Aberta do Brasil (UAB)², foi dado o nome de educação a distância.

Especialmente no Brasil, apesar de não haver registros históricos precisos do surgimento da EAD, há indícios de que seu desenvolvimento começou no início do século passado [Alves 2005]. A primeira experiência do Brasil na EaD veio, principalmente, pelas ondas do rádio [Kenski 2009, Saraiva 1996]. A princípio com a Fundação da Rádio Sociedade do Rio de Janeiro (1923), depois surgiu o Instituto Rádio Monitor (1939), o Movimento de Educação de Base - MEB (1956) e o Projeto Minerva (1970). A partir do início da década de 70, com o suporte das televisões, surgiram o Projeto Saci (1973), o Telecurso de 1o e 2o grau (1978) (posteriormente, em 1995, passou a ser chamado de Telecurso 2000) e diversas outras iniciativas. Em 1996 foi criado outro projeto nacional, a TV Escola, objetivando, essencialmente, a formação de professores [Kenski 2009]. [Saraiva 1996] apresenta dados e conclusões a respeito desses e outros programas, como SENAC, ABT e UnB.

¹Fern Universität: <https://www.fernuni-hagen.de/>

²UAB: <http://www.uab.capes.gov.br/>

Assim como a educação tradicional é de suma importância para a construção do caráter do indivíduo, tornando-o um cidadão capacitado para poder tomar decisões no decorrer de sua vida, a educação a distância vem demonstrando seu papel indispensável na sua formação e capacitação profissional. O número de matrículas no ensino superior subiu de 5,3 mil para 1,1 milhão no período de 2001 a 2011. É importante ressaltar que entre 2011 e 2012 as inscrições em cursos EAD tiveram aumento de 12,2%, enquanto nos cursos presenciais o aumento foi de 3,1% [Mercadante 2012]. A atuação da educação a distância no Brasil demonstra que, apesar de seu progresso ser lento, é a modalidade que mais cresce, abrindo espaço para o surgimento de mais pesquisas e discussões. Essa trajetória revela a existência de problemas que não facilitam a ampliação da EAD que, por suas características, se apresenta como modelo extremamente necessário e capaz de atender as demandas de ampliação do ensino no país [Mugnol 2009].

Segundo [Alves 2005], a educação a distância é uma modalidade de imenso valor no que diz respeito ao atendimento de grandes contingentes de alunos, efetivamente melhor que outras modalidades sem riscos de reduzir a qualidade dos serviços oferecidos em decorrência da expansão do número de participantes. As vantagens com essa modalidade de ensino são inúmeras, especialmente no que concerne à flexibilidade que o aluno tem para o seu estudo. No entanto, problemas oriundos da educação tradicional também encontraram seu espaço na educação a distância. Um bom exemplo disso é a evasão – ou desistência de alunos –, que vem atormentando as instituições educacionais na procura de explicações factíveis e soluções efetivas para este empecilho [Jorge et al. 2010]. Tais obstáculos já são objetos de pesquisas, na qual é demonstrado que o perfil do evadido, tal como no caso particular do estudo em [Jorge et al. 2010], é predominantemente feminino, e que a faixa etária com maior taxa de evasão é entre 18 e 30 anos. O estudo de [Santos and OLIVEIRA NETO 2008] identificou que a maior parte das evasões correspondiam a causas extrínsecas ao curso, como, por exemplo, a não adequação ao modelo de aprendizagem da EaD; dentre as causas intrínsecas, menos predominantes, podemos citar a falta de acompanhamento do professor-tutor e falta de apoio/incentivo institucional.

Há uma variedade significativa de técnicas voltadas à identificação de problemas em EaD sem, no entanto, se apresentarem como solução definitiva. Por exemplo, estudos feitos por meio da aplicação de questionários podem deixar a desejar na coleta de dados para análise, pois geralmente tomam porções muito reduzidas da realidade. Isto significa que os dados coletados, devido a um número muito limitado de observações, podem não contemplar todos os aspectos importantes na avaliação de determinado estudo [André 2001]. O problema do espaço amostral, num cenário onde dados são escassos, atinge a maior parte das técnicas observadas na Educação a Distância.

2. O Projeto IngRede

No Instituto Federal de Alagoas, a EAD está presente através de duas iniciativas: a Universidade Aberta do Brasil (UAB) e o projeto IngRede, que é resultado do consórcio entre UFG, UFMG, UFMT, UFRJ, UFSJ, UFSM, UFU, UFJF, UFPA e UFPEL como parte das atividades da UNIREDE – Universidade Virtual Pública do Brasil – mais tarde denominada como Universidade Aberta do Brasil – UAB [de Oliveira et al. 2013].

O IngRede tem como objetivo capacitar em inglês instrumental alunos desde o

nível técnico ao doutorado e sua presença nas instituições de ensino é de fundamental importância, devido ao atual cenário globalizado, que cobra o conhecimento da língua inglesa e também pela escassez de professores da matéria nas instituições.

Como diferencial, o curso propõe o ensino de inglês voltado às necessidades do usuário: os alunos são agrupados em 6 grandes áreas, de acordo com sua área de atuação. Cada grande área possui atividades específicas, com uso de vocabulário e conteúdo típicos de tal área. Dessa maneira, o aluno se sente motivado a seguir o curso, pois supõe-se que os assuntos tratados são de seu interesse. As grandes áreas do curso são: Engenharias; Ciências Agrárias, Exatas e da Terra; Ciências Sociais Aplicadas; Ciências da Saúde e Biológicas; Ciências Humanas; Linguística, Letras e Artes e Funcionários IFAL;

Os cursos do IngRede (ver Figura 1) são desenvolvidos na plataforma Moodle, uma plataforma de gerenciamento de aprendizagem desenvolvida por Martin Dougiamas, que possui uma grande diversidade de ferramentas que auxiliam na construção do conhecimento. Atualmente com milhares de utilizadores e desenvolvedores, e traduzido para mais de 73 línguas, o Moodle tem-se revelado um importante LMS (Learning Management System) devido à flexibilidade, valor educativo e facilidade de utilização, graças à interface simples e amigável mesmo para usuários menos experientes [Legoinha et al. 2006].

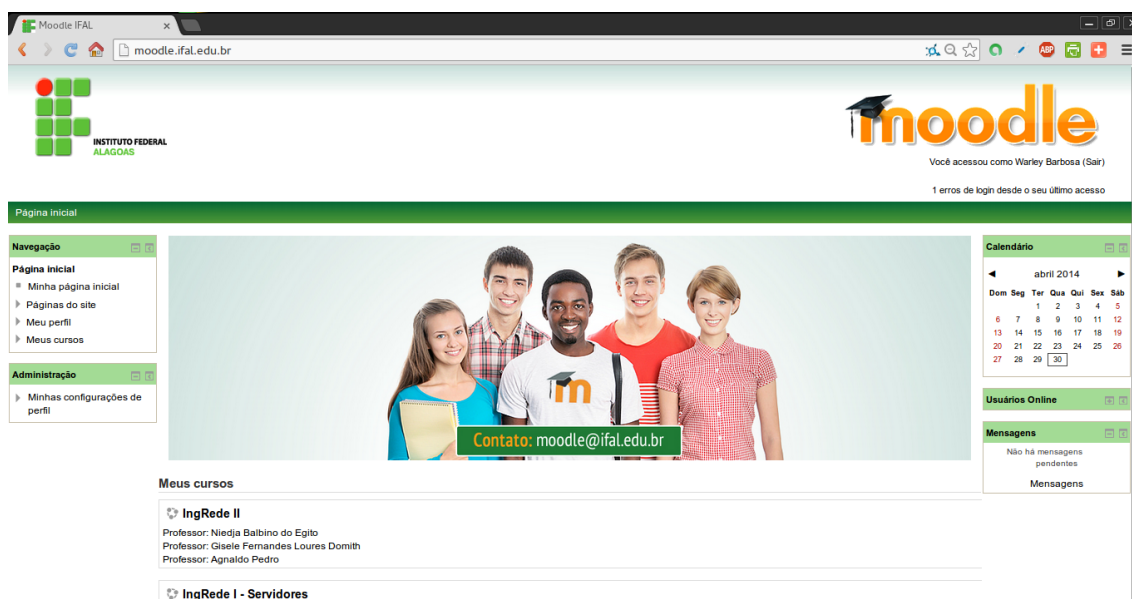


Figura 1. Tela Inicial do Moodle do projeto IngRede

O Moodle como sistema de gestão de ensino e aprendizagem apresenta funcionalidades com forte componente de participação, comunicação e colaboração entre formandos, formadores e pares. Enquanto software educativo, a componente de avaliação (assessment and inquiry) não poderia ser esquecida: são oferecidas ferramentas de avaliação específicas de diversas atividades, como a possibilidade de classificar (pelos professores ou pares), através de escala elaborada para o efeito, discussões de fórum, trabalhos enviados ou realizados online, lições com questões e entradas de glossário [Legoinha et al. 2006]. Com essas e outras funcionalidades, o Moodle se mostra detentor de uma gama muito rica de informações sobre os usuários, que podem ser utilizadas por técnicas de prospecção de

dados para fins de identificação de perfis de usuários e seus comportamentos.

O projeto vem sendo executado desde 2012, recebendo mais de 1000 inscrições de alunos e servidores para as duas turmas, em breve iniciando sua terceira. O curso, porém, enfrenta o problema da evasão: do total de inscrições, pouco mais de 100 alunos concluíram o curso. É necessário, portanto, procurar meios para diminuir ao mínimo possível o número de alunos evadidos. Para isso escolhemos a mineração de dados, visto ser bem conhecida e eficiente no que diz respeito a análise de dados.

3. Mineração de Dados Educacionais

Atualmente, em termos mundiais, o volume de dados armazenado é gigantesco e continua crescendo rapidamente. Infelizmente, devido a incapacidade do ser humano de interpretar tamanha quantidade de dados, muita informação e conhecimento, possivelmente úteis, podem estar sendo desperdiçados, ficando ocultos dentro das Bases de Dados espalhadas pelo mundo. Em consequência disso, a necessidade de se desenvolver novas ferramentas e técnicas de extração de conhecimento a partir de dados armazenados também vem crescendo e se mostrando cada vez mais indispensável [Rezende et al. 2003]. A partir dessa necessidade de extrair informação útil de grandes conjuntos de dados, surge a mineração de dados.

De forma clássica, pode-se definir mineração de dados como um processo não-trivial de identificar, em dados, padrões válidos, novos, potencialmente úteis e compreensíveis [Fayyad et al. 1996]. [Ng and Han 2002] define como processo de descoberta do conhecimento em grandes quantidades de dados armazenados em bases de dados, datawarehouses ou outros repositórios de informações.

A Mineração de Dados pode ser aplicada de duas formas: como um processo de verificação e como um processo de descoberta. No processo de verificação o sistema é limitado a verificar as hipóteses formuladas pelo usuário. No processo de descoberta, o sistema autonomamente encontra novos padrões [Fayyad et al. 1996]. De acordo com os objetivos da aplicação e da natureza dos dados, uma determinada tarefa é utilizada. Na mineração existem dois tipos de tarefas básicas: descritiva (ou não-supervisionada) e preditiva (ou supervisionada). As descritivas se concentram em encontrar padrões que descrevam os dados de forma interpretável pelos seres humanos, como, por exemplo, identificar dentro de um conjunto de dados grupos de pessoas que possuem características parecidas. As preditivas realizam inferência nos dados para construir modelos que serão usados para previsões do comportamento de novos dados, como, por exemplo, descobrir a previsão do tempo através de dados coletados por informações de satélite para prever se em uma determinada região vai chover ou não [Fayyad et al. 1996, Ng and Han 2002].

A mineração de dados educacionais (EDM, em inglês) é uma área relativamente recente de pesquisa [Baker and Yacef 2009], que tem por objetivo o desenvolvimento de métodos para explorar conjuntos de dados com o intuito de fornecer informações úteis no processo de tomada de decisão no âmbito da educação [Baker et al. 2011]. Esses dados provêm de uma grande variedade de fontes diretamente relacionadas a um contexto educacional, sejam software educacionais, cursos online, ou diversas outras [Romero and Ventura 2010].

Com a recente ampliação dos cursos de educação a distância e daqueles com suporte computacional, surgem perguntas que impulsionam as pesquisas no ramo da

EDM, por exemplo, como utilizar os dados gerados por AVAs para aperfeiçoar o ensino a distância? Ou como prever quais estudantes tem uma maior tendência a evadir? [Gottardo et al. 2012]. A mineração de dados educacionais visa prover respostas que apoiem os problemas propostos utilizando e adaptando técnicas já existentes de mineração, como pode ser analisado em [Manhães et al. 2011]. [Costa et al. 2013] apresentam um apanhado de técnicas e aplicações para mineração de dados educacionais.

Com o aumento da utilização de softwares educacionais tornou-se mais fácil a utilização de seus dados para análise de pesquisadores da área de EDM, melhorando, assim, a capacidade de os possíveis resultados serem aplicados no contexto escolar [Baker et al. 2011]. Exemplos como em [Gottardo et al. 2012], o qual mostra resultados da utilização de determinadas técnicas com o intuito de avaliar o desempenho dos alunos; [Manhães et al. 2011] e [Rigo et al. 2012] na previsão da evasão em cursos EAD; [Kampff 2009] identifica - por meio de técnicas de mineração - características de alunos com risco de evasão ou reprovação para alertar o professor a respeito; e muitos outros apresentados por [Baker and Yacef 2009] e [Romero and Ventura 2010] constataam que os estudos na área de EDM vêm oferecendo contribuições significativas para a teoria e a prática da educação.

No Brasil, as pesquisas nessa área ainda estão em seu estágio inicial, contudo, as oportunidades para seu desenvolvimento vem crescendo muito, principalmente por causa da ampliação de cursos a distância [Baker et al. 2011]. Um caso especial que pode se beneficiar (ou possivelmente já se beneficie) com isso seria a UAB, visto a imensa quantidade de dados produzidos, já tratada nesse artigo anteriormente. Apesar de poucos trabalhos terem sido publicados internacionalmente com participação brasileira, lentamente estamos conseguindo nosso espaço. O artigo de Prata e colegas [Baker et al. 2011], bem como de [Pimentel and Omar 2006] são exemplos disso.

Nossa proposta neste artigo é demonstrar a metodologia de mineração de dados a ser utilizada para tentar descobrir causas e fatores que influenciem na evasão, bem como traçar um perfil do evadido. É importante deixar claro que nosso trabalho está em fase inicial, se tratando, apenas, de uma possibilidade para alcançar os resultados almejados. Explicamos, na próxima seção, qual será a nossa metodologia para mineração de dados educacionais.

4. Trabalhos Relacionados

Em [Manhães et al. 2011] a mineração de dados é usada com o objetivo identificar precocemente alunos em risco de evasão nos cursos de graduação. A técnica de mineração de dados oferece diversos algoritmos que podem ser empregados para identificar alunos em risco de abandono. Este trabalho apresenta um estudo utilizando dados acadêmicos de alunos de graduação de uma universidade brasileira (UFRJ).

[Kampff 2009] descreve em seu artigo o estágio atual de uma pesquisa que busca, através de mineração de dados, identificar perfis de alunos com risco de evasão ou reprovação, bem como boas práticas pedagógicas de acompanhamento e intervenção, de forma a promover alterações nos AVA (Ambientes virtuais de aprendizagem) que facilitem a implementação de alertas. Um estudo interessante é o de [Baker et al. 2008], na qual são utilizadas técnicas de Mineração de Dados para detectar comportamentos inadequados dos estudantes.

No trabalho de [Pimentel and Omar 2006] é descrito a utilização de técnicas de Mineração de Dados para identificar as relações entre as medidas cognitivas (real desempenho do aluno) e as metacognitivas (grau de consciência do aluno acerca do que ele estudou). [Prata et al. 2009] demonstram que conflitos interpessoais são benéficos para o aprendizado, concluindo que atividades colaborativas podem facilitar e agilizar o trabalho de pesquisadores e educadores.

A principal diferença que se apresenta entre a proposta deste trabalho e os trabalhos relacionados tem base na utilização de Mineração de Dados Educacionais. Enquanto a maioria se concentra em levantar dificuldades de aprendizagem do aluno para, por exemplo, alimentar sistemas de recomendação de conteúdo, nosso trabalho foca em identificar perfis de alunos que estão prestes a evadir com base em fatores (baixa interatividade no curso, notas baixas, perda de interesse no curso, e outros) que possam indicar comportamentos de risco no ambiente virtual de aprendizagem do projeto em questão. Dessa forma, ações pedagógicas por parte da equipe responsável poderão ser tomadas de forma preventiva.

5. Proposta

A metodologia CRISP-DM (Cross Industry Standard Process for Data Mining) propõe um compreensivo modelo de processo para projetos de mineração de dados. É importante dizer que este modelo é semelhante, em estrutura, a outros existentes, no entanto, escolhemos por sua vasta literatura [Azevedo 2008] e por mostrar-se completo, organizado e com uma estrutura bem definida, o que permite um fácil entendimento ou revisão de todo o projeto. Uma comparação entre KDD, SEMMA e CRISP-DM pode, também, ser encontrada no artigo citado previamente. O CRISP-DM é dividido em seis fases de maneira cíclica (ver Figura 2): business understanding, data understanding, data preparation, modeling, evaluation, and deployment [Shearer 2000]. Definimos, abaixo, de acordo com [Wirth and Hipp 2000, Shearer 2000], estes conceitos e o que faremos em cada um:

- **Business understanding (Entendimento do negócio):** foca no conhecimento dos objetivos do projeto e, então, converte-se esse conhecimento numa definição de problema de mineração de dados e um plano de projeto preliminar com a intenção alcançar os objetivos.
Nesta fase, além de definirmos os objetivos a serem alcançados, determinaremos os critérios de sucesso do projeto. Também serão definidos os recursos necessários à execução de todo o plano de mineração.
- **Data understanding (Entendimento dos dados):** os dados podem vir de diversas fontes e ter diversos formatos. Assim, a partir de uma coleção inicial, já livre de possíveis problemas, é necessário descrevê-los, explorá-los e, por fim, verificar sua qualidade.
Aqui os dados do consórcio serão analisados, separados e selecionados de acordo com os critérios necessários à execução dos algoritmos posteriores. Nesse caso, podemos, também, utilizar técnicas de classificação para definir qual a melhor coleção. Pode-se voltar à fase anterior, caso o entendimento dos dados não se adeque a definição dos negócios.
- **Data preparation (Preparação dos dados):** a fase de preparação tem o objetivo de construir o conjunto final de dados que serão utilizados na ferramenta de mo-

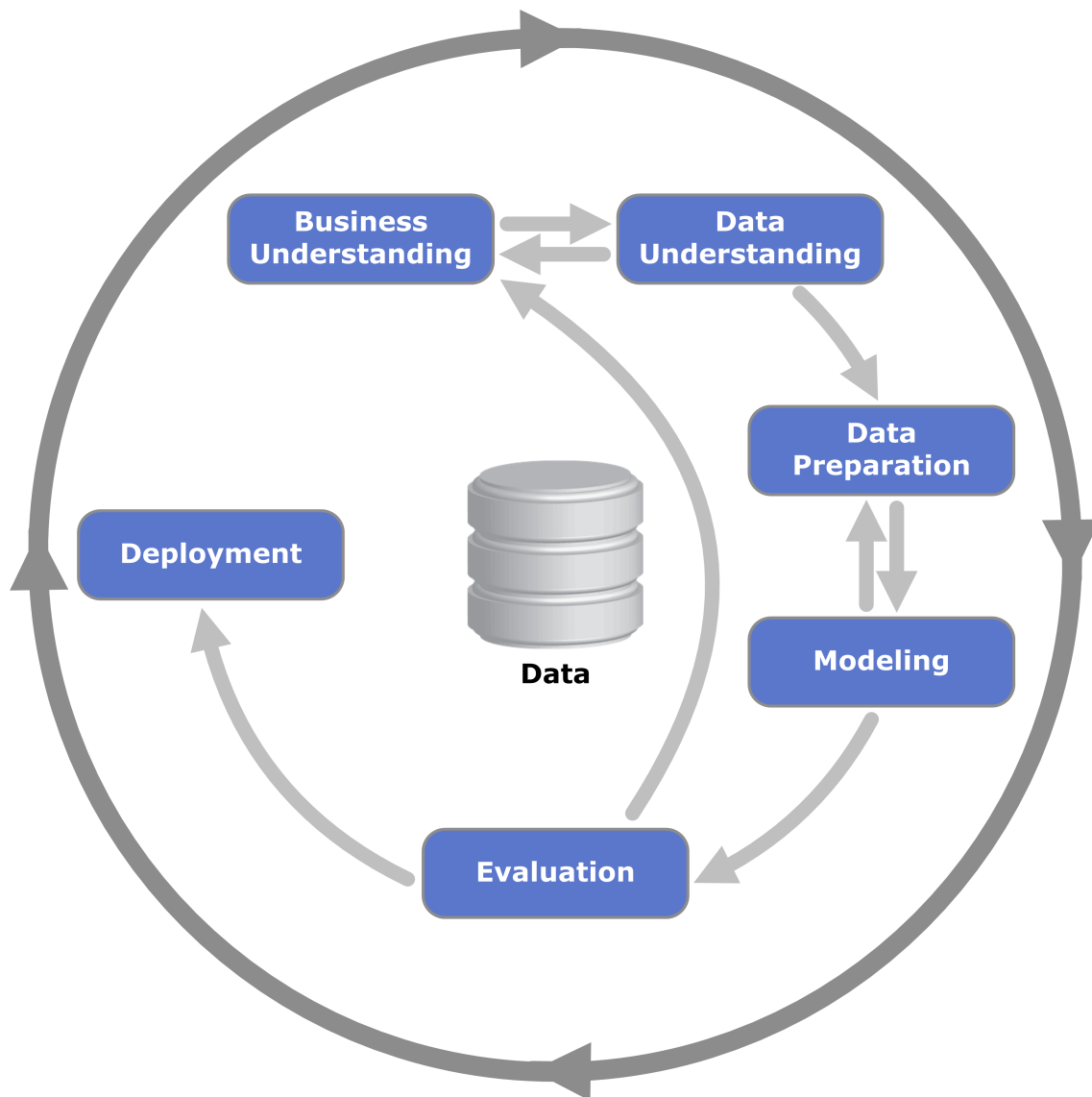


Figura 2. Diagrama dos processos da metodologia CRISP-DM

delagem. As tarefas de preparação podem ser realizadas muitas vezes, e sem uma ordem predeterminada.

Adequaremos, nessa fase, os dados selecionados (provindos de bancos de dados relacionais) à ferramenta a ser utilizada. Como é dito acima, esta tarefa pode ser repetida diversas vezes até um resultar no conjunto que melhor se adapte às técnicas de modelagem.

- **Modeling (Modelagem):** nessa fase, várias técnicas de modelagem são selecionadas e aplicadas. Tipicamente, existem diversas técnicas para o mesmo tipo de problema de mineração, no entanto, há algumas que dependem do objetivo desejado.

Aplicação de técnicas (previamente avaliadas) de modelagem no conjunto de dados já selecionado. Nosso objetivo aqui é encontrar atributos que, associados à frequência no curso, por exemplo, possam indicar uma tendência a evadir, e quais aspectos importantes esses atributos revelam sobre tal aluno.

- **Evaluation (Avaliação):** nesse estágio, tem-se construído um ou mais modelos que aparentam ter alta qualidade. Ao final será tomada uma decisão a partir dos resultados da mineração, sem, entretanto, desconsiderar alguma questão que seja importante.

Os melhores modelos serão aplicados no contexto escolar. Pode haver questões a serem reconsideradas, o que consiste na revisão de todo o projeto. Assim pode acontecer novos planos de mineração, mas já com mais informações a respeito.

- **Deployment (Desenvolvimento):** a criação do modelo não é o fim do projeto. Em alguns casos, o conhecimento adquirido precisará ser organizado e apresentado de uma forma que possa ser usado. Geralmente a aplicação do conhecimento é responsabilidade do cliente.

Conforme dito antes, o resultado do projeto pode ser aplicado tantos em outros novos projetos, que por ventura podem aparecer, como nos objetivos inicialmente traçados. São os educadores os responsáveis pelo projeto que decidem se e de que maneira o conhecimento será utilizado.

Nosso objetivo com essa proposta é aplicar a metodologia apresentada acima em um projeto para mineração de dados, visto que não há, entre as referências consultadas, trabalhos relacionados a identificação de causas de desistência e perfis dos evadidos, simultaneamente. O propósito desse projeto é obter resultados que auxiliem a identificação de circunstâncias que levam o aluno a evadir, ou seja, quais fatores o levam a tal ato. Da mesma maneira, buscaremos traçar seu perfil para melhor compreensão e adaptação de ambientes educacionais, visando possibilitar respostas proativas por parte dos educadores.

6. Considerações Finais

A próxima etapa do trabalho será a execução das técnicas de mineração nos dados obtidos do Projeto IngRede e, com a experiência adquirida, poderemos partir para um trabalho com os dados de todo o consórcio. O trabalho com os dados do consórcio será uma extensão do trabalho atual, porém, seus resultados terão uma melhor precisão e o conhecimento obtido terá uma maior importância para a tentativa de combater a evasão, já que o consórcio tem um volume de dados muito maior ao disponível no Projeto IngRede e, para a mineração de dados, quanto maior o volume de dados, mais preciso e útil será o conhecimento obtido. Nele, esperamos analisar tanto o cenário nacional de EaD, encontrando causas universais de evasão, quanto trabalhar com os dados de uma ou mais regiões, áreas de ensino ou grupos específicos de alunos. Assim, seriam encontradas informações voltadas à necessidade de um grupo ou região específica que enfrente o problema da evasão.

Referências

- Alves, J. R. M. (2005). Educação a distância e as novas tecnologias de informação e aprendizagem. *línea*, <http://www.engenheiro2001.org.br/programas/980201a1.htm>.
- André, M. (2001). Pesquisa em educação: buscando rigor e qualidade. *Cadernos de pesquisa*, 113:51–64.
- Azevedo, A. I. R. L. (2008). Kdd, semma and crisp-dm: a parallel overview.
- Baker, R., Isotani, S., and Carvalho, A. (2011). Mineração de dados educacionais: Oportunidades para o brasil. *Revista Brasileira de Informática na Educação*, 19(02):03.

- Baker, R. S. and Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *JEDM-Journal of Educational Data Mining*, 1(1):3–17.
- Costa, E., Baker, R. S., Amorim, L., Magalhães, J., and Marinho, T. (2013). Mineração de dados educacionais: Conceitos, técnicas, ferramentas e aplicações. *Jornada de Atualização em Informática na Educação*, 1(1):1–29.
- d Baker, R. S., Corbett, A. T., Roll, I., and Koedinger, K. R. (2008). Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction*, 18(3):287–314.
- de Oliveira, V. L. M., Braga, P. J. d. C. F., Carneiro, M. M., Racilan, M., Junior, R. C. G., Lima, L. A., et al. (2013). Leitura em inglês na rede: a trajetória do projeto ingrede. *Educação & Tecnologia*, 17(3).
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37.
- Gottardo, E., Kaestner, C., and Noronha, R. V. (2012). Avaliação de desempenho de estudantes em cursos de educação a distância utilizando mineração de dados. In *Anais do Workshop de Desafios da Computação Aplicada à Educação*, pages 30–39.
- Jorge, B. G., Martins, C. Z., Carniel, F., Lazilha, F. R., Vieira, M. C., Goi, V. M., and Pesquisa, C. (2010). Evasão na educação a distância: um estudo sobre a evasão em uma instituição de ensino superior. *Maringá: abril*.
- Kampff, A. J. C. (2009). Mineração de dados educacionais para geração de alertas em ambientes virtuais de aprendizagem como apoio à prática docente.
- Kenski, V. M. (2009). O desafio da educação a distância no brasil.
- Legoinha, P., Pais, J., and Fernandes, J. (2006). O moodle e as comunidades virtuais de aprendizagem. VII Congresso Nacional de Geologia 2006. Sociedade Geológica de Portugal.
- Manhães, L. M. B., da Cruz, S. M. S., Macário Costa, R. J., Zavaleta, J., and Zimbrão, G. (2011). Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 1.
- Mercadante, A. (2012). Censo da educação superior.
- Moran, J. M. (2008). O que é educação a distância.
- Mugnol, M. (2009). A educação a distância no brasil: conceitos e fundamentos. *Revista Diálogo Educacional*, 9(27):335–349.
- Ng, R. T. and Han, J. (2002). Clarans: A method for clustering objects for spatial data mining. *Knowledge and Data Engineering, IEEE Transactions on*, 14(5):1003–1016.
- Pimentel, E. P. and Omar, N. (2006). Descobrimos conhecimentos em dados de avaliação da aprendizagem com técnicas de mineração de dados. In *Anais do Workshop de Informática na Escola*, volume 1.
- Prata, D. N., d Baker, R. S., Costa, E. d. B., Rosé, C. P., Cui, Y., and de Carvalho, A. M. (2009). Detecting and understanding the impact of cognitive and interpersonal conflict in computer supported collaborative learning environments. *International Working Group on Educational Data Mining*.

- Rezende, S. O., Pugliesi, J., Melanda, E., and Paula, M. d. (2003). Mineração de dados. *Sistemas inteligentes: fundamentos e aplicações*, 1:307–335.
- Rigo, S. J., Cazella, S. C., and Cambuzzi, W. (2012). Minerando dados educacionais com foco na evasão escolar: oportunidades, desafios e necessidades. In *Anais do Workshop de Desafios da Computação Aplicada à Educação*, pages 168–177.
- Romero, C. and Ventura, S. (2010). Educational data mining: a review of the state of the art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40(6):601–618.
- Santos, E. M. d. and OLIVEIRA NETO, J. D. d. (2008). Evasão na educação a distância: identificando causas e propondo estratégias de prevenção. In *Congresso Internacional de EaD*, volume 23.
- Saraiva, T. (1996). Educação a distância no brasil: lições da história. pages 17–27.
- Shearer, C. (2000). The crisp-dm model: the new blueprint for data mining. *Journal of data warehousing*, 5(4):13–22.
- Wirth, R. and Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, pages 29–39. Citeseer.