

MINERAÇÃO DE DADOS EDUCACIONAIS PARA IDENTIFICAÇÃO DE BARREIRAS NA UTILIZAÇÃO DA EDUCAÇÃO A DISTÂNCIA

Brasília – DF – 04/2015

Vinícius Coutinho G. Coelho – Universidade de Brasília – vinicius.coelho.unb@gmail.com

João Paulo C. Lustosa da Costa – Universidade de Brasília – jpdacosta@unb.br

Daniel da Cunha R. de Sousa – Universidade de Brasília – daniel.cunha@gmail.com

Edna Dias Canedo – Universidade de Brasília – ednacanedo@unb.br

Daniel Guerreiro e Silva – Universidade de Brasília – danielgs@ene.unb.br

Rafael Timóteo de Sousa Júnior – Universidade de Brasília – desousa@ene.unb.br

Investigação Científica (IC): Pesquisa

Educação Continuada em Geral

Características de Aprendizizes

Descrição de Projeto em Andamento

RESUMO

Este artigo apresenta os resultados obtidos a partir da análise dos dados de um ambiente virtual de aprendizagem na Escola Nacional de Administração Pública (Enap), com o objetivo de identificar as dificuldades e barreiras que os alunos dos cursos a distância registram, nas ferramentas de comunicação com a instituição. A literatura aponta diversos registros em relação a barreiras à Educação a Distância (EAD), barreiras como as dificuldades de adaptação pessoal e da necessidade de habilidades para utilização de recursos tecnológicos no dia a dia e para a efetiva utilização dessa modalidade de aprendizagem. A partir da mineração e análise dos dados dos canais de comunicação utilizados nesse ambiente virtual de aprendizagem foi possível identificar padrões que realçam a necessidade de avaliação contínua dos ambientes com o intuito de melhoria em questões de acessibilidade e ergonomia. Tais padrões podem alavancar uma possível evasão de alunos que possuam dificuldades que não estejam diretamente relacionadas ao conteúdo ou a qualidade dos cursos oferecidos, mas com dificuldade de conseguir acessar e utilizar os ambientes onde esses cursos são ofertados.

Palavras chave: Educação a Distância, Mineração de Dados Educacionais, Canais de Comunicação, Barreiras à EAD, Alfabetização Digital.

Os autores agradecem à Escola Nacional de Administração Pública - ENAP - do Ministério do Planejamento, Orçamento e Gestão, bem como a CAPES, o CNPq e a FINEP, pelo suporte a este trabalho.

1. INTRODUÇÃO

A crescente utilização da Educação a Distância (EAD) é destaque em todas as suas áreas de aplicação devido à sua grande amplitude no atendimento das demandas por aprendizagem. A partir desse crescimento, a utilização da EAD tem ganhado destaque na educação corporativa e vem sendo adotada em programas de qualificação e formação profissional.

Neste contexto, diversas instituições na Europa, Canadá e também no Brasil adotaram essa modalidade como ferramenta para o desenvolvimento e capacitação dos servidores públicos. No Brasil, a Escola Nacional de Administração Pública (Enap) tem como missão o desenvolvimento de competências dos servidores públicos para aumentar a capacidade de governo na gestão de políticas públicas (Abbad, 2007).

No entanto, a utilização da EAD atualmente está relacionada a um processo de ensino e aprendizagem mediado diretamente por tecnologias. Isto significa que é necessário que os atores envolvidos nesse processo possuam uma alfabetização tecnológica básica para que possa haver interação com os ambientes de estudo. (Abbad, 2007)

Em outro contexto, diversos estudos apresentam perspectivas de tratamento e mineração dos dados de Ambientes Virtuais de Aprendizagem (AVA), com objetivos de identificação de padrões, indicadores preditivos de evasão e de rendimento, suporte a sistemas de tutoria inteligente, *Intelligent Tutor Systems (ITS)* (Kampff *et al.*, 2008, Costa *et al.*, 2013, Kowalski *et al.*, 2014). Nota-se daí a existência de poucos trabalhos relacionados à utilização destas tecnologias para identificar possíveis padrões de perfis que apresentem dificuldades de acesso aos ambientes tecnológicos presentes na modalidade EAD.

Este trabalho, portanto, aplica tecnologias de mineração de dados educacionais, *Educational Data Mining (EDM)*, para avaliação e tratamento dos dados provenientes dos sistemas utilizados como canal de comunicação entre a Enap e seu público-alvo. Como resultado foi possível identificar as dificuldades registradas pelos usuários, como barreiras relacionadas à efetiva utilização e disseminação da EAD.

Este artigo está organizado como segue. A Seção 2 apresenta os objetivos que motivaram a realização do presente estudo. A Seção 3 apresenta

o embasamento e o referencial teórico relacionado à mineração de dados educacionais (EDM). A Seção 4 trata dos procedimentos metodológicos que foram aplicados na Enap, além de apresentar as limitações e restrições em relação às bases de dados disponibilizadas para o estudo. A Seção 5 apresenta os resultados da análise dos dados que foram agrupados e quantificados. Por fim a Seção 6 apresenta as conclusões e as propostas para trabalhos futuros.

2. OBJETIVOS

A realização deste trabalho se deu aplicando metodologias de EDM para tratar dados registrados nos sistemas que são usados como canal de comunicação de um AVA em uma Escola de Governo. Os dados eram gerados e armazenados por duas ferramentas distintas: (i) o sistema “Fale Conosco” e (ii) uma caixa de e-mails específica, a Caixa EAD.

A partir daí o objetivo foi realizar o levantamento e avaliação dos dados dos sistemas envolvidos no atendimento dos usuários do AVA, no período de 2010 a 2013. O foco principal foi a identificação e a quantificação de registros relacionados a dificuldades de utilização das tecnologias presentes nos cursos na modalidade EAD, ofertados por essa instituição.

A partir da análise qualitativa e quantitativa de ambos os segmentos do canal de comunicação foi possível identificar as incidências, nos registros dos usuários, que estão relacionadas a dificuldades de utilização do ambiente virtual onde foram ofertados os cursos. Como resultado, podem ser feitas sugestões de melhorias nos meios de comunicação entre a escola virtual e seu público-alvo com o intuito de viabilizar a acessibilidade e inclusão de todos os usuários.

3. REFERÊNCIAL TEÓRICO

EDM é definida como *“uma disciplina emergente, preocupada com o desenvolvimento de métodos para exploração dos tipos de dados únicos provenientes dos ambientes educacionais e, utilizar esses métodos para entender melhor os alunos e as características de como eles aprendem”* (Baker e Yacef, 2009)

A literatura relacionada à EDM aborda a aplicação de técnicas, entre elas, classificação, regressão e agrupamento de dados, para o tratamento da grande quantidade de dados que são gerados a partir da utilização da EAD, ou seja, os dados gerados pelos usuários ao acessarem ambientes educacionais e utilizarem as ferramentas e os meios de iteração aplicados (Romero e Ventura, 2010; Baker *et al.*, 2011).

O uso da EAD está baseado na utilização de Tecnologias da Informação e Comunicação (TIC), que requerem que seus usuários possuam uma devida alfabetização tecnológica, como, por exemplo, o manuseio de editores de texto, planilhas, e-mail, participação em chats e utilização de buscas na internet (Almeida *et al.* 2013).

Conforme estudos relacionados com fatores que afetam o desempenho de estudantes, percebe-se que algumas barreiras ainda precisam ser vencidas para que a EAD seja realmente acessível e para que sejam levadas em consideração as limitações e dificuldades dos usuários, os quais podem apresentar brechas para o processo inclusivo da aprendizagem. A falta de uma alfabetização tecnológica pode gerar um impacto negativo nos indivíduos que têm seus primeiros contatos com a EAD e também para os que já são usuários dessa modalidade, mas que ainda possuem dificuldades de adaptação tecnológica. (Ramos *et al.* 2014).

A necessidade do uso das TIC pode gerar a evasão de alunos que possuam um grande potencial de aprendizagem, mas com limitações relacionadas aos conhecimentos tecnológicos. Essa situação pode excluí-los das oportunidades e vantagens que o modelo de ensino e aprendizagem da EAD disponibiliza (Albertin e Brauer, 2012).

4. METODOLOGIA APLICADA

O presente artigo utilizou a pesquisa qualitativa e quantitativa, baseada em um estudo de caso, com o objetivo de identificar e quantificar as incidências de termos relacionados às dificuldades de acesso ao AVA, nos chamados registrados nos canais de comunicação, no período de 2010 a 2013.

Inicialmente, o trabalho foi realizado utilizando abordagens tecnológicas diferenciadas, tendo em vista a complexidade e diversidade dos dados coletados nos sistemas envolvidos nos canais de comunicação.

Os dados do sistema “Fale Conosco” eram armazenados por um sistema específico, empregado pela Enap como ferramenta para controle de chamados. Esse sistema armazenava as informações em um banco de dados e disponibilizava uma interface para acesso e utilização. A Caixa EAD, que se tratava de uma conta de e-mail, ead@enap.gov.br, armazenava informações de forma aleatória e não era exclusivamente direcionada para o atendimento de chamados relacionados ao ambiente virtual.

Os dados foram disponibilizados da seguinte forma: uma base de dados SQL, *Structure Query Language*, referente ao sistema Fale Conosco e um arquivo em formato PST, referente aos e-mails registrados na Caixa EAD. Após uma análise inicial da integridade dos resultados, decidiu-se não considerar as informações da ferramenta Caixa EAD nesse estudo, tendo em vista que esta não era a ferramenta adotada como canal formal para a comunicação com a Enap.

Dado que o foco deste artigo é o processo da mineração e não a avaliação de ferramentas, o emprego de apenas uma ferramenta não invalida os resultados, visto que, conforme argumenta (Yin, 2001), os dados e o contexto da situação representam um estudo de caso suficientemente rico e complexo para validar a análise. Além disso, a exploração de dados provenientes de outras ferramentas é uma extensão natural do trabalho, estando validada a metodologia e o processo.

Deste modo, serão apresentados somente os estudos realizados nos dados do sistema “Fale Conosco”, onde foram realizadas a estruturação dos dados em um *Data Warehouse (DW)* e a construção de um modelo de dados multidimensional (Cubo), com o objetivo de identificar e agrupar os termos relacionados às dificuldades de acesso ao AVA (Kimball e Ross, 2011).

A Figura 1 apresenta o sistema Fale Conosco, que disponibiliza uma interface para preenchimento de informações cadastrais dos usuários e campos específicos para seleção de valores iniciais para a Categoria e o Assunto. O campo Categoria refere-se a qual área da Enap o usuário deseja contatar e o campo Assunto refere-se ao motivo ao qual o chamado está relacionado. O campo Assunto possui os seguintes valores para seleção: Acesso, Cadastro, Dúvidas, Reclamações, Senha do Sítio Enap, Senha da Escola Virtual, Sugestões e Outros.

A estruturação dos dados foi realizada a partir de um processo de engenharia e análise, composto por atividades específicas, conforme apresentado na Figura 2.

Fale Conosco - Enviar Mensagem

Nome Completo*

E-mail*

Telefone

Categoria*
 Cursos a Distância

Assunto*
 - Selecione -

PARA UMA RESPOSTA MAIS ASSERTIVA, INFORME SEU CPF NO CORPO DA MENSAGEM.
 Mensagem*:

(*)campos de preenchimento obrigatório

Figura 1. – Formulário do sistema Fale Conosco

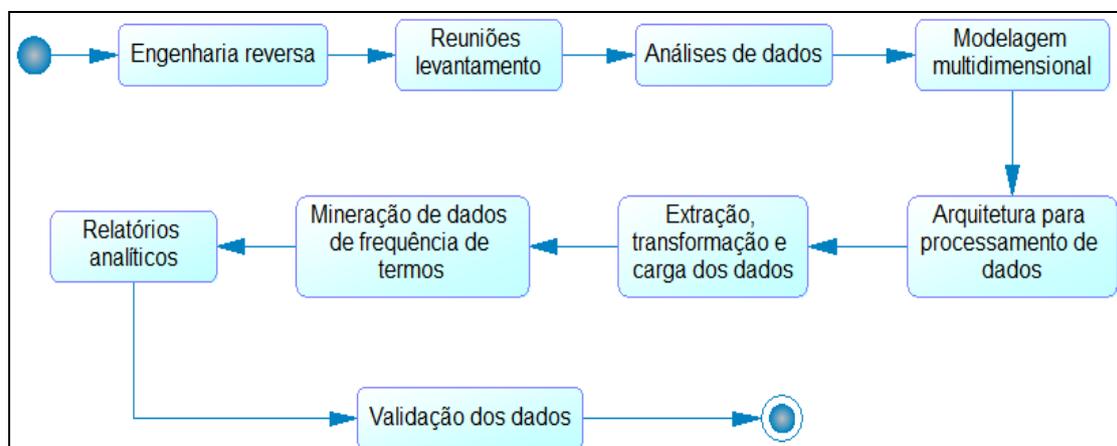


Figura 2. Apresentação do processo de construção do modelo de dados multidimensional para análise das informações

A Engenharia reversa foi realizada com base nos dados disponibilizados e das informações coletadas em Reuniões de levantamento, onde foi possível haver o entendimento do cenário em que se registravam os chamados, conforme apresentado na Figura 2. A Engenharia reversa permitiu a visualização da base de dados de todo o sistema Fale Conosco da Enap, o que demonstrou que tal sistema atende não somente a Escola Virtual, mas também todas as outras áreas da Enap. Com a realização das Reuniões para

levantamento das informações das regras de negócio, foi possível delimitar as mensagens que eram exclusivamente direcionadas para a Escola Virtual.

O processo de Modelagem multidimensional foi realizado a partir da Análise dos dados de forma padronizada, intuitiva e de rápido acesso. Esse processo foi realizado a partir das informações coletadas nos processos de Engenharia reversa e Reunião de levantamento. Um Modelo multidimensional é composto basicamente por dimensões, fatos e suas métricas.

Após a modelagem dimensional, foi construído uma Arquitetura Computacional que possibilitasse o devido processamento dos dados para a realização da etapa de extração, transformação e carga dos dados. Com o ambiente devidamente preparado, foi realizado o processo de mineração dos dados com objetivo de identificar e quantificar a frequência dos termos presentes nos chamados.

A mineração de dados utilizou o resultado do modelo apresentado na Figura 2, no processo de Arquitetura para processamento dos dados, como base para montagem do Cubo que viabilizou a construção dos relatórios analíticos. Conforme o modelo proposto na Figura 3, os registros de mensagens no sistema Fale Conosco geram um Fato (f_mensagem), e a partir de uma mensagem são adicionadas as informações relacionadas (Dimensões) no modelo multidimensional, sendo elas: Assunto (d_assunto), Categoria (d_categoria), Palavras (d_palavras), UF (d_uf), Tempo (d_tempo) e Frase (d_frase). A Figura 3 apresenta o modelo multidimensional resultante da etapa 4 do processo apresentado na Figura 2.

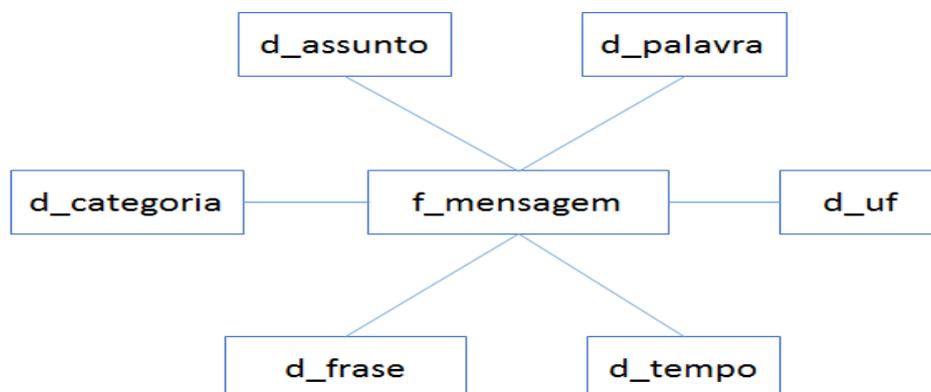


Figura 3. Representação do modelo de dados multidimensional

O processo foi finalizado com a construção de relatórios analíticos que embasaram os estudos apresentados nesse artigo. Como resultado final do

processo de construção do modelo multidimensional apresentado na Figura 2 foi possível identificar a incidência de termos relacionados às dificuldades de acesso ao AVA, os quais puderam ser associados a características apresentadas em estudos que investigam as barreiras no uso da EAD.

5. APRESENTAÇÃO DOS RESULTADOS

A partir do processo de engenharia reversa, apresentado na Seção de Metodologia, Figura 1, foram criadas análises distintas dos dados obtidos no modelo de dados multidimensional, apresentado na Seção 4, Figura 2. Para fins de estruturação do relatório, as análises foram realizadas individualmente, a partir dos dados agrupados e detalhados.

A Tabela 1 possibilita a visualização da quantidade de termos registrados por ano e constitui uma análise sintética dos dados de chamados registrados, agrupados por assunto e ano. Essa análise consolida os chamados registrados entre os anos de 2010 e 2013, de forma que seja possível analisar a evolução dos valores ao longo dos anos.

Assunto	2010	2011	2012	2013	Total por assunto
Acesso	2.133	1.875	1.569	1.039	6.616
Dúvidas	1.339	1.417	983	868	4.607
Outros	1.227	1.139	822	933	4.121
Senha do Sítio Enap	611	526	514	330	1.981
Senha da Escola Virtual	604	486	377	141	1.608
Cadastro	560	451	405	464	1.880
Reclamações	400	330	266	238	1.234
Sugestões	63	52	33	37	185
Total por ano	6.937	6.276	4.969	4.050	

Tabela 1. Visão analítica da quantidade de mensagens por assunto e ano – Fale Conosco

Conforme apresentado na Tabela 1, houve uma variação entre a quantidade dos chamados contendo os assuntos diretamente relacionados a possíveis problemas com acesso ao ambiente virtual.

A redução de mensagens registradas na Tabela 1 variou de 2010 a 2013, de acordo com cada termo monitorado, em diferentes proporções, a saber: o número de mensagens referentes ao Acesso reduziu em 51,3%; Dúvidas, 35,1%; Outros 23,9%; Senha do Sítio da Enap, 45,9%; Senha da Escola Virtual; 76,6%, Cadastro, 17,1%; Reclamações 40,5% e Sugestões, 41,2%.

É possível identificar o uso excessivo de assuntos genéricos como “Outros” e “Dúvidas”. As análises dos textos dos chamados indicaram que, em alguns casos, um assunto diferente poderia classificar melhor o chamado.

Em quase todas as categorias houve uma redução de aproximadamente metade, ou mais, do quantitativo no envio de mensagens durante os quatro anos, sendo que a quantidade de alunos aumentou gradativamente durante esse período. Como exceção, as categorias dúvidas, outros e cadastro, não registraram tal redução e tal fato pode ser explicado devido a estas categorias não estarem diretamente relacionadas a problemas de acesso dos usuários ao ambiente virtual.

No decorrer do período entre 2010 e 2013, foram realizadas diversas ações para redução de chamados de acordo com as incidências mapeadas, ações como: Contato telefônico direto; Disponibilização de funcionalidade para recuperação de senhas no portal; envio de e-mails para os alunos ausentes ou sem atividades registradas no sistema e pelas melhorias presentes na evolução da versão de software do AVA e de seu layout.

6. CONCLUSÕES E RECOMENDAÇÕES

As ações realizadas no ambiente virtual de aprendizagem no decorrer dos anos de 2010 a 2013 refletiram em um decréscimo considerável nos registros de chamados relacionados a dificuldades de acesso.

A partir do cenário do estudo, é possível afirmar que as ações realizadas pela Escola Nacional de Administração Pública vinculadas à inclusão e acessibilidade em sua escola virtual alcançaram resultados positivos, tendo em vista que a quantidade de alunos registrados na escola aumentou gradativamente todos os anos.

Como recomendação, propõe-se a integração entre os canais de comunicação, o sistema de gestão acadêmica e o AVA, para geração de indicadores como por exemplo: Chamados por turma; Chamados por visitantes (Usuário sem cadastro); Chamados por alunos não matriculados e Indicadores de desistência de alunos.

Esses indicadores possibilitarão uma melhor visualização e monitoramento do AVA e assim, a criação de ações direcionadas e mais

efetivas com o intuito de favorecer a usabilidade e acessibilidade ao público da Escola Virtual da Enap.

REFERÊNCIAS BIBLIOGRÁFICAS

ABBAD, Gardênia da Silva. Educação a distância: O estado da arte e o futuro necessário. **Revista do Serviço Público**, v. 58, n. 3, p. 351-374, 2007.

ALBERTIN, Alberto Luiz; BRAUER, Marcus. Resistência à educação a distância na educação corporativa. **Rev. Adm. Pública**, v. 46, n. 5, p. 1367-389, 2012.

ALMEIDA, Onília Cristina de Souza de; ABBAD, Gardênia; MENESES, Pedro Paulo Murce; ZERBINI, Thaís. Evasão em cursos a distância: fatores influenciadores. **Revista Brasileira de Orientação Profissional**, v. 14, n. 1, p. 19-33, 2013.

BAKER, Ryan SJD; YACEF, Kalina. The state of educational data mining in 2009: A review and future visions. **JEDM-Journal of Educational Data Mining**, v. 1, n. 1, p. 3-17, 2009.

BAKER, Ryan SJD; ISOTANI, Seiji; CARVALHO, Adriana. Mineração de Dados Educacionais: Oportunidades para o Brasil. **Revista Brasileira de Informática na Educação**, v. 19, n. 02, p. 03, 2011.

COSTA, Evandro; BAKER, Ryan SJD; AMORIM, Lucas; MAGALHÃES, Jonathas; MARINHO, Tarsis. Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações. **Jornada de Atualização em Informática na Educação**, v. 1, n. 1, p. 1-29, 2013.

KAMPPFF, Adriana Justin Cerveira; REATEGUI, Eliseo Berni; DE LIMA, José Valdeni. Mineração de dados educacionais para a construção de alertas em ambientes virtuais de aprendizagem como apoio à prática docente. **RENOTE**, v. 6, n. 1, 2008.

KIMBALL, Ralph; ROSS, Margy. **The data warehouse toolkit: the complete guide to dimensional modeling**. John Wiley & Sons, 2011.

KOWALSKI, John; ZHANG, Yanhui; GORDON, Geoffrey J., Statistical Modeling of Student Performance to Improve Chinese Dictation Skills with an Intelligent Tutor. **JEDM-Journal of Educational Data Mining**, v. 6, n. 1, p. 3-27, 2014.

RAMOS, Jorge Luis Cavalcanti; RODRIGUES, Rodrigo Lins; SILVA, João Carlos Sedraz; GOMES, Alex Sandro. Analisando Fatores que Afetam o Desempenho de Estudantes Iniciantes em um Curso a Distância. **Anais do Simpósio Brasileiro de Informática na Educação**. 2014. p. 99-108.

ROMERO, Cristóbal; VENTURA, Sebastián. Educational data mining: a review of the state of the art. **Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on**, v. 40, n. 6, p. 601-618, 2010.

YIN, Robert. **Estudo de Caso: Planejamento e Métodos**. 2. Ed. Bookman, 2001.