

## Mineração de Dados Educacionais Orientada por Atividades de Aprendizagem

João Carlos Sedraz Silva – UNIVASF, joao.sedraz@univasf.edu.br  
Rodrigo Lins Rodrigues – UFRPE, rodrigo.linsrodrigues@ufrpe.br  
Jorge Luis Cavalcanti Ramos – UNIVASF, jorge.cavalcanti@univasf.edu.br  
Fernando da Fonseca de Souza – UFPE, fdfd@cin.ufpe.br  
Alex Sandro Gomes – UFPE, asg@cin.ufpe.br

**Resumo.** Este trabalho, por meio de um estudo de caso, apresenta uma abordagem de Mineração de Dados Educacionais orientada por atividades de aprendizagem, tendo como referência a Teoria da Atividade. O objetivo do estudo foi analisar as diferenças nos resultados dessa abordagem em relação a um processo de mineração holístico, no qual os modelos de predição permitem análises apenas no nível de disciplina, sem a observação de detalhes das atividades de aprendizagem. Os resultados desta pesquisa apontam vantagens da mineração orientada por atividade, que oferece informações em um contexto de interação significativa, com mais subsídios para monitorar e tratar contradições no processo de aprendizagem.

**Palavras-chave:** Mineração de Dados Educacionais, Atividades de Aprendizagem, Teoria da Atividade

### Educational Data Mining in Oriented Learning Activities

**Abstract.** This paper, through a case study, presents an approach Educational Data Mining oriented by learning activities, with reference to the Activity Theory. The aim of the study was to analyze the differences in the results of this approach to a holistic mining process, in which the prediction models allow analysis only in the level of discipline without the observation of details of the learning activities. The results of this research show the mining advantages driven by activity that provide information in a significant interaction context, with more subsidies to monitor and deals with contradictions in the learning process.

**Keywords:** Educational Data Mining, Learning Activities, Activity Theory.

#### 1. Introdução

A Mineração de Dados Educacionais (EDM - do inglês *Educational Data Mining*) é definida como uma área responsável pelo desenvolvimento de métodos para a extração de conhecimento a partir de base de dados educacionais, especialmente, de cursos a distância *online* (Kampff, 2008; Romero e Ventura, 2013).

Durante o curso, é desejável que a EDM obtenha dados de interação dos estudantes e interprete os seus significados, a fim subsidiar professores em eventuais ajustes nas estratégias pedagógicas. Nesse sentido, tem crescido o número de trabalhos que estudam modelos de predição do desempenho dos discentes (Peña-Ayala, 2014). Para Silva *et. al.* (2015), os fóruns de discussão são ferramentas relevantes para a interação na modalidade a distância e, em razão disso, sugerem pesquisas que analisem as relações entre a interação nessas ferramentas e o desempenho dos estudantes.

De acordo com Siemens e Baker (2012), a EDM precisa promover um diálogo entre a técnica e os domínios sociais e pedagógicos, para permitir intervenções que atendam as necessidades das partes envolvidas no processo educacional. Mas, uma barreira a ser enfrentada é a falta de bases teóricas para orientar esse diálogo (Peña-Ayala, 2014). Para superar essa barreira, alguns trabalhos têm adotado a estrutura proposta por Moore (1989), que classifica as interações a distância em três tipos: estudante-conteúdo, estudante-instrutor e estudante-estudante (Rodrigues *et al.*, 2014).

A estrutura de Moore (1989) não define o nível de observação das interações, permitindo que os pesquisadores adotem abordagens holísticas, em que as interações são mapeadas por disciplina, curso ou instituição. Em abordagens holísticas, as análises podem ser conduzidas em um nível indevidamente abstrato, sem informações suficientes para a precisa avaliação dos fatores que afetam o desempenho do estudante.

Uma proposta alternativa para evitar a perda de informações relevantes, é o mapeamento das interações dos estudantes por atividades de aprendizagem. O foco em atividades não é uma ideia nova em termos de concepção de aprendizagem, mas apenas recentemente instituições ligadas às pesquisas em análise de aprendizagem têm buscado especificações no nível de atividades de aprendizagem (Griffiths e Hoel, 2016).

Nesta pesquisa, com um estudo de caso envolvendo a predição do desempenho de estudantes por meio de dados das interações em fóruns, são avaliados os resultados de uma Mineração de Dados Educacionais orientada por atividades de aprendizagem. Diante disso, este trabalho teve como objeto de investigação as seguintes questões:

- *Para cada atividade de aprendizagem, qual o modelo que explica o desempenho dos estudantes?*
- *Em relação aos modelos obtidos por atividades de aprendizagem, uma abordagem holística produz um modelo que permite explicar melhor o desempenho dos estudantes?*

O objetivo geral deste artigo é, portanto, analisar as diferenças nos resultados da EDM orientada por atividade de aprendizagem em relação a uma abordagem holística<sup>1</sup>. Para tanto, utilizando como elemento norteador a Teoria da Atividade (Engeström, 2001), o estudo foi realizado a partir de dados de um curso de pós-graduação em Gestão Pública, oferecido a distância em uma universidade pública do Brasil. A escolha desse curso deve-se a sua presença na maioria das universidades públicas brasileiras que, desde 2010, já ofertaram mais de 100.000 vagas nessa pós-graduação<sup>2</sup>.

Sendo assim, além desta introdução, o trabalho está organizado com mais quatro seções, que apresentam a fundamentação teórica, o método utilizado, a discussão dos resultados e as considerações finais da pesquisa.

## 2. Teoria da Atividade

A Teoria da Atividade (TA) é composta por conceitos que explicam como e por que as atividades humanas são executadas. Essa teoria oferece elementos para a compreensão dos processos colaborativos mediados por computador, considerando a atividade como o contexto mínimo significativo para as ações dos indivíduos (Ramos, 2010).

Os conceitos iniciais da TA têm origem entre os séculos XVIII e XIX, com a filosofia clássica alemã, de Kant a Hegel, quando destacaram o papel da atividade mental na constituição da relação entre sujeito e objeto. Marx e Engels aprimoraram o

<sup>1</sup> Neste artigo, a abordagem holística corresponde ao mapeamento das interações dos estudantes apenas em nível de disciplina, sem a especificação das atividades de aprendizagem em que elas ocorrem.

<sup>2</sup> Disponível em: <http://sisrel.capes.gov.br/index.php/rel>. Acessado em: 20/01/2016.

conceito de atividade, representado também as raízes da TA. Mas, apenas a partir de 1920, com base na psicologia histórico-cultural, Vygotsky estabeleceu os fundamentos da teoria que, posteriormente, foi aperfeiçoada por Leont'ev, Luria e Engeström.

A TA pode ser classificada em três gerações (Batista *et al.*, 2010). A primeira geração está concentrada nos trabalhos de Vygotsky, que conceituou a atividade como mediação, gerando o modelo triangular da relação do sujeito com o objeto, mediada por artefatos materiais e culturais (Figura 1a). Segundo Vygotsky (1978), os indivíduos usam ferramentas para mudar o mundo e a eles mesmos. Embora essa representação aborde a aprendizagem como transformação e não como transmissão, é necessária uma articulação do indivíduo com o seu papel na estrutura social.

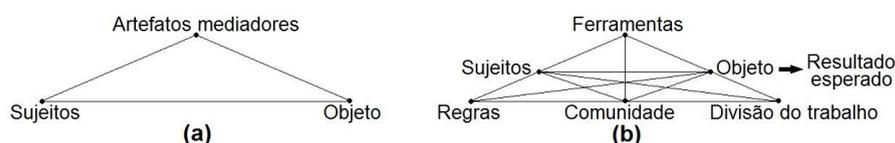


Figura 1. Modelos de representação da atividade (Engeström, 2001).

A segunda surge das contribuições de Leont'ev, avançando na distinção de ação coletiva e ação individual, estabelecendo uma estrutura mais abrangente da atividade. No entanto, nunca Leont'ev expandiu graficamente o modelo original de Vygotsky em um modelo de sistema de atividade coletiva. A terceira geração, proposta por Engeström em 1987, parte do modelo triangular de Vygotsky, expandindo-o para o modelo do sistema da atividade coletiva representado na Figura 1b.

Para Engeström (2001), participantes de uma atividade são sujeitos que interagem com o objeto para alcançar os resultados desejados. Nesse modelo, a relação entre o sujeito e objeto é mediada por ferramentas, a relação entre sujeito e comunidade é mediada por regras, e as relações entre objeto e a comunidade são mediadas pela divisão do trabalho. Entende-se por ferramenta qualquer artefato usado no processo de transformação, incluindo tanto as ferramentas materiais quanto as ferramentas de pensamento. As regras são as normas explícitas e implícitas, convenções e relações sociais dentro de uma comunidade. A divisão do trabalho se refere à organização da comunidade envolvida na transformação do objeto para o resultado.

No âmbito da educação a distância, a TA complementa o modelo conceitual de interações de Moore (1989), estruturado nas relações estudante-conteúdo, estudante-instrutor e estudante-estudante. Nas atividades de aprendizagem, quando observadas sob a perspectiva da Teoria da Atividade, as interações estudante-instrutor e estudante-estudante representam a relação entre o discente e a comunidade, enquanto a interação estudante-conteúdo a relação do discente com as ferramentas. Nesse sentido, o modelo de atividade proposto por Engeström corrobora Moore e, além disso, oferece outros elementos para subsidiar a compreensão dos significados das ações individuais, que são norteadas pelas regras, a divisão do trabalho e pelos resultados esperados.

### 3. Método

#### 3.1. Contexto do estudo

A pesquisa descrita neste artigo é um estudo de caso que, segundo Yin (2013), justifica-se em situações contemporâneas em que é necessária a investigação profunda de uma unidade de análise. Foram analisados dados de estudantes que participaram de quatro disciplinas no curso de pós-graduação em Gestão Pública, oferecido na modalidade a

distância pela Universidade Federal do Vale do São Francisco, no período de março a julho de 2015.

As disciplinas do curso foram realizadas de maneira sequencial e cada uma teve duração de três semanas. As atividades de aprendizagem aconteceram por meio do LMS Moodle, organizado em ambientes com uma estrutura semelhante a da Figura 2. Na ilustração, alguns elementos estão contornados em vermelho apenas para o esclarecimento da forma de apresentação das atividades Fórum de Notícias (FN), Fórum de Debates (FD) e Fórum da Galera (FG). Essas atividades foram consideradas como unidades de referência para a predição do desempenho dos estudantes.

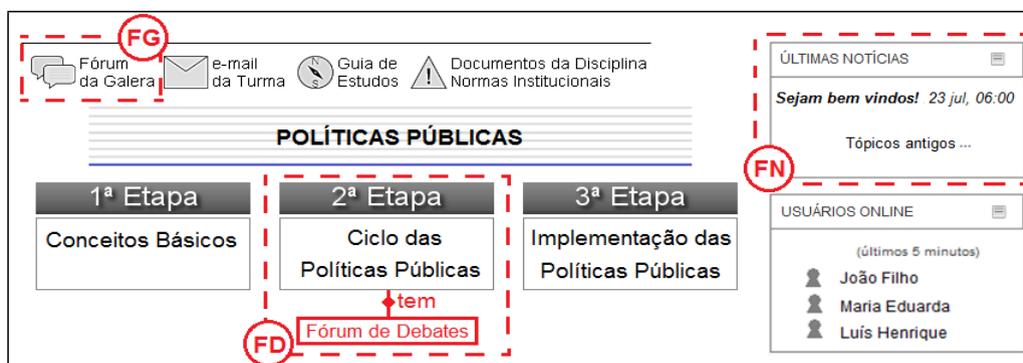


Figura 2. Estrutura típica do ambiente virtual de aprendizagem adotado no curso.

Tabela 1. Detalhamento das atividades na perspectiva da Teoria da Atividade.

Componente da atividade	Atividade		
	Fórum de Notícias	Fórum de Debates	Fórum da Galera
Sujeitos	Estudantes	Estudantes	Estudantes
Comunidade	Estudantes, professores e tutores	Estudantes, professores e tutores	Estudantes, professores e tutores
Ferramentas	Fórum de discussão	Fórum de discussão e material de apoio (apostila)	Fórum de discussão
Regras	A participação não compõe a nota do estudante na disciplina; Essa é a forma oficial de comunicação dos eventos da disciplina.	A participação vale de 15% da nota final do estudante na disciplina; As contribuições devem ser fundamentadas no material de apoio; Os critérios para a avaliação da participação são: aderência ao tema, interação com os colegas, qualidade dos textos e respeito aos prazos.	A participação não compõe a nota do estudante na disciplina; Respeitando a urbanidade, qualquer tema pode ser discutido.
Divisão do trabalho	Sempre que necessário, o professor cria tópicos sobre eventos importantes na disciplina; Os estudantes observam os tópicos criados pelo docente e, quando julgarem oportuno, fazem comentários ou solicitam esclarecimentos dos tutores ou professores.	O professor define o tema e o período das discussões; Cada estudante cria um tópico para expor sua opinião sobre o tema; Os estudantes comentam os tópicos criados pelos colegas; O professor atua para que as discussões não se afastem do tema; Os tutores estimulam os estudantes com baixa participação.	Estudantes criam tópicos e lideram as discussões, respondendo a perguntas uns dos outros; A participação dos professores e tutores é discreta e acontece apenas quando requerida pelos estudantes.
Objeto	Novos eventos da disciplina.	Conteúdo da segunda etapa da disciplina	Temas gerais relacionados à disciplina e a eventos sociais
Resultado esperado	Estudantes atualizados com informações de eventos importantes da disciplina.	Aquisição de conhecimento sobre um tema específico.	Esclarecimento de dúvidas gerais e fortalecimento de vínculos sociais

A partir dos fundamentos da Teoria da Atividade (Seção 2), na Tabela 1 é possível notar que, apesar das três atividades utilizarem fóruns de discussão como ferramenta, as especificações dos outros componentes previstos pela TA são diferentes. Essas diferenças produzem significados distintos e, conseqüentemente, afetam a forma de interação dos estudantes em cada uma das atividades.

### 3.2. Objetivo

Analisar as diferenças nos resultados da EDM orientada por atividade de aprendizagem em relação a uma abordagem holística.

### 3.3. Participantes

No período em que a pesquisa foi realizada, o curso possuía 150 estudantes matriculados. Dentre esses, foram analisados os dados de 122 estudantes, que realizaram atividades em todas as disciplinas selecionadas neste trabalho. Os dados do perfil dos participantes estão indicados na Tabela 2.

**Tabela 2. Dados do perfil dos estudantes que participaram da pesquisa**

Variáveis do perfil	Alternativas	N*	%
Experiência anterior em curso <i>on-line</i>	Sim	53	43.44
	Não	69	56.56
Faixa etária do estudante.	Entre 20 e 30 anos	63	51.63
	Entre 31 e 40 anos	42	34.43
	Entre 41 e 50 anos	13	10.66
	Mais de 50 anos	04	3.28
Sexo	Feminino	64	52.46
	Masculino	58	47.54

\* N = Número de estudantes para cada alternativa.

O perfil dos estudantes aponta uma heterogeneidade da turma, com destaque para um significativo grupo de discentes que não possuíam experiência anterior em cursos *on-line* (56.56 %).

### 3.4. Procedimentos

Para atender ao objetivo da pesquisa, os procedimentos metodológicos adotados seguiram as etapas da Mineração de Dados Educacionais (García *et al.*, 2011) que, dentro do contexto deste estudo, estão descritas a seguir.

#### 3.4.1. Dados coletados

Os dados analisados foram extraídos do LMS Moodle adotado nas disciplinas e de uma planilha com o perfil dos estudantes, elaborada pela coordenação do curso durante a matrícula dos discentes. Para as disciplinas e as atividades de interesse desta pesquisa, foram identificados 124.449 registros de interações dos estudantes no LMS.

#### 3.4.2. Pré-processamento

Os dados coletados foram integrados e filtrados, reunindo em uma mesma base apenas os registros de estudantes que participaram de todas as disciplinas, o que gerou 488 casos de análise (Tabela 3).

**Tabela 3. Distribuição dos casos de análise por disciplina**

Disciplina	Estudantes aprovados	Estudantes reprovados	Total de estudantes
O Estado e os Problemas Contemporâneos – EPC	104	18	122
O Público e o Privado na Gestão Pública – PPGP	105	17	122
Planejamento Estratégico Governamental – PEG	98	24	122
Políticas Públicas – PP	110	12	122
<b>Número de casos de análise</b>	<b>417</b>	<b>71</b>	<b>488</b>

Os registros das interações dos estudantes no LMS foram codificados em variáveis definidas a partir da estrutura conceitual desenvolvida por Moore (1989), que classifica as interações dos estudantes na modalidade a distância nas dimensões estudante-conteúdo, estudante-instrutor e estudante-estudante. Considerando as atividades delimitadas nesta pesquisa, foram observadas as variáveis apresentadas na Tabela 4, com dados descritivos indicados na Tabela 5.

**Tabela 4. Descrição das variáveis observadas no estudo**

Dimensão	Variável	Descrição
Estudante - Conteúdo	VAR01	Tempo de visualização de tópicos de discussão (em horas)
	VAR02	Quantidade de dias que acessou a tópicos de discussão
	VAR03	Quantidade de vezes que acessou a tópicos de discussão
	VAR04	Quantidade de vezes que acessou o material de apoio fornecido pelo professor
	VAR05	Quantidade de tópicos de discussão criados
	VAR06	Quantidade de mensagens enviadas
Estudante - Instrutor	VAR07	Quantidade de mensagens enviadas em resposta aos professores ou tutores
	VAR08	Quantidade de mensagens enviadas que foram respondidas por professores ou tutores
Estudante - Estudante	VAR09	Quantidade de mensagens enviadas em resposta a outros estudantes
	VAR10	Quantidade de mensagens enviadas que foram respondidas por outros estudantes
---	VARD	Desempenho final na disciplina

Pode-se observar que o Fórum de Debates, na maioria das variáveis, alcança médias superiores (Tabela 5). Esse maior nível de interação pode ser justificado pelas regras diferenciadas da atividade que, entre outros aspectos, motiva a participação dos estudantes por produzir impacto direto sobre a nota final na disciplina.

**Tabela 5. Dados descritivos das variáveis observadas no estudo**

Variável	Fórum da Galera				Fórum de Notícias				Fórum de Debates			
	Média	Máximo	Mínimo	Desvio	Média	Máximo	Mínimo	Desvio	Média	Máximo	Mínimo	Desvio
VAR01	1.78	39.75	0.00	5.05	0.97	48.46	0.00	3.61	3.75	68.58	0.00	7.66
VAR02	2.61	14.00	0.00	3.03	1.94	11.00	0.00	2.25	2.24	7.00	0.00	1.40
VAR03	8.14	145.00	0.00	13.32	4.03	34.00	0.00	5.75	16.18	100.00	0.00	15.16
VAR04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.94	11.00	0.00	1.61
VAR05	0.07	3.00	0.00	0.31	0.00	0.00	0.00	0.00	0.97	3.00	0.00	0.32
VAR06	0.43	7.00	0.00	0.95	0.18	3.00	0.00	0.46	2.16	12.00	0.00	1.90
VAR07	0.07	2.00	0.00	0.26	0.16	3.00	0.00	0.43	0.01	2.00	0.00	0.13
VAR08	0.06	2.00	0.00	0.26	0.03	3.00	0.00	0.21	0.07	2.00	0.00	0.29
VAR09	0.29	6.00	0.00	0.71	0.02	2.00	0.00	0.16	1.16	11.00	0.00	1.79
VAR10	0.29	13.00	0.00	1.00	0.02	1.00	0.00	0.14	1.16	8.00	0.00	1.54

A etapa de pré-processamento foi concluída com a transformação dos dados, que permitiu a aplicação da técnica de mineração selecionada na pesquisa e, além disso, simplificou a interpretação dos modelos obtidos. Com a transformação, os valores da variável VARD foram discretizados de forma binária, sendo os estudantes aprovados representados por 1 e os reprovados por 0. Para diminuir os impactos de valores discrepantes, nas variáveis VAR01, VAR02, VAR03, VAR04, VAR05, VAR06, VAR07, VAR08, VAR09 e VAR10 os dados atípicos (*outliers*) foram identificados por disciplina a partir do método do quartil [Torgo 2010] e substituídos pelo correspondente valor adjacente superior.

### 3.4.3. Mineração de dados

O propósito dessa etapa foi obter modelos de classificação do desempenho final dos estudantes, com base nas variáveis apresentadas na Tabela 4 e em dados de treinamento, provenientes das duas primeiras disciplinas (EPC e PPGP).

Os dados de treinamento estavam, originalmente, desbalanceados, com 209 casos de análise de estudantes aprovados e apenas 35 de reprovados. O desbalanceamento dos dados prejudica a geração de modelos de classificação, principalmente, em relação a identificação de casos da classe minoritária. Para balancear os dados de treinamento, foi utilizada a técnica SMOTE (*Synthetic Minority Over-sampling Technique*), que amplia a amostra de treinamento com instâncias sintéticas da classe minoritária (Gottardo *et al.*, 2013). Neste trabalho, a técnica SMOTE foi aplicada aos dados originais com uma taxa de sobreamostragem de 500 %, ampliando para 210 o número de casos de estudantes reprovados na amostra de treinamento.

Para a determinação dos modelos de classificação, foi utilizada a técnica de Regressão Logística, que estuda o comportamento de uma variável dependente binária com base em um conjunto de variáveis independentes numéricas ou categóricas.

$$\text{logit}(p_i) = \ln \left( \frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i} \quad (1)$$

Os modelos de Regressão Logística são formulados por (1), onde  $\beta_0, \beta_1, \dots, \beta_k$ , são os coeficientes das variáveis que explicam a ocorrência de um determinado evento. Existem algumas abordagens para auxiliar na seleção das variáveis explicativas. Neste trabalho, foi utilizada a abordagem *stepwise*, que adiciona ou retira variáveis independentes do modelo estatístico conforme o poder de discriminação que agregam ao grupo de variáveis explicativas (Hair *et al.*, 2009).

#### 3.4.4. Pós-processamento

Nessa etapa, os modelos obtidos pelos algoritmos de EDM devem ser compreensíveis e úteis para o processo de tomada de decisão. A validação dos modelos desenvolvidos foi realizada a partir de dados das disciplinas Planejamento Estratégico Governamental e Políticas Públicas, que nessa situação representaram a amostra de testes deste estudo. Para a avaliação dos modelos foram utilizados os valores da acurácia em cada classe (estudantes aprovados e estudantes reprovados) e da acurácia global.

## 4. Resultados

As análises estatísticas deste trabalho foram realizadas por meio do *software* estatístico R<sup>3</sup>, com o suporte do pacote DMwR<sup>4</sup>, a partir do processo de mineração de dados detalhado na Seção 3.4. A seguir, são apresentados os resultados obtidos, organizados de acordo com as questões de pesquisa formuladas na Seção 1.

### 4.1. Para cada atividade de aprendizagem, qual o modelo que explica o desempenho dos estudantes?

A Tabela 6 apresenta os modelos de Regressão Logística produzidos com a amostra de treinamento, separados por atividade, com destaque para as variáveis que demonstraram significativa relação com o desempenho dos estudantes, considerando uma significância de 0,05 ( $p\text{-valor} < 0,05$ ).

As variáveis VAR01 e VAR02 são as que melhor explicam o desempenho dos estudantes quando observados os dados da atividade Fórum da Galera, enquanto que para Fórum de Debates são as variáveis VAR02 e VAR05 e para Fórum de Notícias a variável VAR03.

O sinal dos coeficientes logísticos ( $\beta$ ) indica o sentido da relação (Hair *et al.*, 2009). No modelo da atividade Fórum de Notícias, a relação positiva da VAR03 significa que um aumento nessa variável é associado com o aumento na probabilidade do discente pertencer ao grupo de maior desempenho (classe dos estudantes aprovados).

Os coeficientes logísticos exponenciais ( $\text{Exp}(\beta)$ ) refletem diretamente a magnitude da variação no valor da probabilidade de um evento acontecer (Hair *et al.*, 2009). De todos os modelos, entre as variáveis significativas ( $p\text{-valor} < 0,05$ ), VAR05 na atividade Fórum de Debates representa a variável que mais contribuiu para explicar o desempenho superior dos estudantes. Isso, empiricamente, reforça a importância do monitoramento dos estudantes em relação ao atendimento da divisão do trabalho

<sup>3</sup> <https://cran.r-project.org>

<sup>4</sup> <https://cran.r-project.org/web/packages/DMwR/>

estabelecida nessa atividade (Tabela 1), que obriga cada discente a criar pelo menos um tópico de discussão ( $VAR05 > 0$ ).

**Tabela 6. Modelos para a classificação do desempenho dos estudantes.**

Modelo obtido a partir dos dados do Fórum da Galera.					
Variável	$\beta$	Exp( $\beta$ )	Std. Err.	Z	p> z
(Intercept)	-5,64E-01	0,57E+00	1,35E-01	-4,17E+00	3,11E-05 *
VAR01	-6,16E-01	0,54E+00	1,83E-01	-3,36E+00	7,69E-04 *
VAR02	3,38E-01	1,40E+00	1,01E-01	3,35E+00	8,05E-04 *
VAR03	3,46E-02	1,04E+00	2,70E-02	1,28E+00	2,00E-01
VAR04	---	---	---	---	---
VAR05	---	---	---	---	---
VAR06	9,37E-01	2,55E+00	5,91E-01	1,59E+00	1,13E-01
VAR07	---	---	---	---	---
VAR08	---	---	---	---	---
VAR09	-8,49E-01	0,43E+00	6,96E-01	-1,22E+00	2,22E-01
VAR10	---	---	---	---	---

Modelo obtido a partir dos dados do Fórum de Notícias.					
Variável	$\beta$	Exp( $\beta$ )	Std. Err.	Z	p> z
(Intercept)	-9,70E-01	0,38E+00	1,54E-01	-6,30E+00	2,99E-10 *
VAR01	4,99E-01	1,65E+00	2,97E-01	1,68E+00	9,24E-02
VAR02	1,02E-01	1,11E+00	1,37E-01	7,41E-01	4,58E-01
VAR03	1,99E-01	1,22E+00	6,99E-02	2,85E+00	4,43E-03 *
VAR04	---	---	---	---	---
VAR05	---	---	---	---	---
VAR06	1,38E+01	9,92E+05	6,11E+02	2,30E-02	9,82E-01
VAR07	-1,47E+01	4,22E-07	6,11E+02	-2,40E-02	9,81E-01
VAR08	---	---	---	---	---
VAR09	---	---	---	---	---
VAR10	---	---	---	---	---

Modelo obtido a partir dos dados do Fórum de Debates.					
Variável	$\beta$	Exp( $\beta$ )	Std. Err.	Z	p> z
(Intercept)	-6,77E+00	1,15E-03	1,39E+00	-4,87E+00	1,12E-06 *
VAR01	-3,35E-02	0,97E+00	7,48E-02	-4,48E-01	6,54E-01
VAR02	3,48E-01	1,42E+00	1,44E-01	2,43E+00	1,53E-02 *
VAR03	4,41E-04	1,00E+00	1,80E-02	2,40E-02	9,81E-01
VAR04	8,93E-02	1,09E+00	1,01E-01	8,83E-01	3,77E-01
VAR05	6,52E+00	6,81E+02	1,61E+00	4,05E+00	5,02E-05 *
VAR06	-2,82E-01	7,54E-01	6,92E-01	-4,08E-01	6,83E-01
VAR07	---	---	---	---	---
VAR08	---	---	---	---	---
VAR09	4,44E-01	1,56E+00	7,19E-01	6,17E-01	5,37E-01
VAR10	1,79E-01	1,20E+00	1,44E-01	1,25E+00	2,12E-01

\* Significância < 0,05,  $\beta$  = coef. logístico, Exp( $\beta$ ) coef. logístico exponenciais.

Como outro exemplo do que pode ser interpretado a partir dos valores de Exp( $\beta$ ) da Tabela 6, observando VAR02 (quantidade de dias que acessou aos tópicos de discussão) e mantendo constante as demais variáveis, o modelo da atividade Fórum de Debates indica que o aumento de uma unidade na frequência de acesso aos tópicos de discussão incrementa 1.42 vezes a probabilidade do estudantes ser aprovado.

#### **4.2. Em relação aos modelos obtidos por atividades de aprendizagem, uma abordagem holística produz um modelo que permite explicar melhor o desempenho dos estudantes?**

Com as tabelas 7 e 8, nota-se que a abordagem holística apresenta limitações qualitativas e quantitativas. Em relação às deficiências qualitativas, apesar de indicar variáveis que afetam o desempenho dos estudantes, o modelo obtido não permite inferir o impacto das interações em cada atividade de aprendizagem. Dessa maneira, não oferece ao professor elementos precisos para promover ajustes nas estratégias pedagógicas. Além disso, quantitativamente, neste estudo as taxas de acurácia da abordagem holística não foram as melhores em nenhuma das classes observadas.

**Tabela 7. Modelo para a classificação do desempenho dos estudantes obtido a partir de uma abordagem holística.**

Variável	$\beta$	Exp( $\beta$ )	Std. Err.	Z	p> z
(Intercept)	-6,96E+00	9,50E-04	1,43E+00	-4,88E+00	1,05E-06 *
VAR01	3,04E-03	1,00E+00	3,50E-02	8,70E-02	9,31E-01
VAR02	2,75E-01	1,32E+00	7,66E-02	3,59E+00	3,26E-04 *
VAR03	2,24E-03	1,00E+00	1,31E-02	1,71E-01	8,64E-01
VAR04	-1,19E-01	8,88E-01	1,19E-01	-9,96E-01	3,19E-01
VAR05	6,14E+00	4,65E+02	1,47E+00	4,19E+00	2,78E-05 *
VAR06	1,01E-01	1,11E+00	2,45E-01	4,12E-01	6,81E-01
VAR07	-1,17E+00	0,31E+00	4,65E-01	-2,52E+00	1,16E-02 *
VAR08	2,26E+00	9,62E+00	6,29E-01	3,60E+00	3,19E-04 *
VAR09	-1,79E-01	8,36E-01	3,27E-01	-5,50E-01	5,83E-01
VAR10	1,02E-01	1,11E+00	1,30E-01	7,80E-01	4,36E-01

\* p-valor&lt;0,05

Com 81,56% de classificações corretas, o Fórum de Debates alcançou a melhor acurácia global. O percentual de acerto do modelo gerado por essa atividade foi ainda maior em relação ao grupo de estudantes aprovados (85,58%). Com isso, observando as variáveis significativas do modelo associado a essa atividade (Tabela 6), fica evidente que os discentes de melhor desempenho são aqueles que criaram tópicos de discussões e que tiveram uma maior quantidade de dias de acesso ao fórum.

**Tabela 8. Taxas de acurácia, por classe, obtidas nos testes realizados nos modelos.**

Classes	Fórum da Galera	Fórum de Notícias	Fórum de Debates	Abordagem holística
Estudantes Aprovados	48.56 %	16.35 %	85.58 %	69.71 %
Estudantes Reprovados	83.33 %	86.11 %	58.33 %	80.56 %
Acurácia Global	53.69 %	26.64 %	81.56 %	71.31 %

O modelo associado ao Fórum de Notícias apresentou uma baixa acurácia global, com uma previsão do desempenho, significativamente, inferior aos outros modelos analisados. Mas, esse modelo tem uma elevada a acurácia relativa aos estudantes reprovados, indicando que a maioria absoluta desse grupo de discentes (86,11%) tende a acessar com menor frequência essa atividade.

## 5. Considerações Finais

Neste trabalho, com apoio da Teoria da Atividade (TA), buscou-se analisar as diferenças nos resultados da Mineração de Dados Educacionais (EDM) orientada por atividade de aprendizagem em relação a uma abordagem holística.

Os resultados deste estudo de caso fornecem evidências das vantagens da mineração orientada por atividade. Os modelos produzidos por essa abordagem, quando confrontados as especificações das atividades nos termos da TA (sujeitos, comunidade, ferramentas, regras, divisão do trabalho, objeto e resultado esperado), apresentaram informações em um contexto de interação significativa, que permite a avaliação precisa dos fatores que afetam o desempenho dos discentes em cada atividade de aprendizagem.

Com as informações da EDM orientada por atividade de aprendizagem, os professores têm subsídios para monitorar e tratar contradições no processo de aprendizagem, o que permite ajustes nas estratégias pedagógicas para elevar o desempenho dos discentes.

Ressaltamos que, por se tratar de um estudo de caso, os resultados deste trabalho estão limitados aos dados e a metodologia do curso analisado. Nesse sentido, para trabalhos futuros, propõe-se o desenvolvimento de um *plugin* para o LMS que automatize a obtenção de modelos de predição do desempenho dos estudantes por atividade de aprendizagem, assim como a aplicação do método utilizado nesta pesquisa em outros cenários educacionais para a validação dos resultados.

## Agradecimentos

Esta pesquisa é apoiada pela FACEPE-APQ No. 0525-1.03/14. Os autores também agradecem ao apoio da PRPPG/UFRPE (006999/2015-28) e PRPPGI/UNIVASF (Edital 18/2013). Alex Sandro Gomes é bolsista DT Nível 2/CNPq, processos nº 310466/2012-1 e nº 475634/2013-6.

## Referências bibliográficas

BATISTA, S., BEHAR, P., PASSERINO, L. **Contribuições da teoria da atividade para m-learning**. RENOTE: revista novas tecnologias na educação. Vol. 8, n. 2 (jul. 2010), 10 f., 2010.

ENGESTRÖM, Y. **Expansive learning at work: Toward an activity theoretical reconceptualization**. Journal of education and work. 2001.

GARCÍA, E., ROMERO, C., VENTURA, S., CASTRO, C. **A collaborative educational association rule mining tool**. The Internet and Higher Education. . 2011.

GOTTARDO, E. *et al.* **Aplicação de Técnicas de EDM para Estimativa de Desempenho Acadêmico de Estudantes em um AVA utilizando Dados com Classes Desbalanceadas**. In ICBL 2013 (pp. 24-28). 2013.

GRIFFITHS, D., HOEL, T. **Learning Analytics Interoperability**. Disponível em: <http://www.laceproject.eu>. Acessado em: 20/05/2016.

HAIR, J. *et al.* **Multivariate data analysis** (Vol. 6). Pearson. 2009.

KAMPPFF, A., REATEGUI, E., DE LIMA, J. **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.

MOORE, M. G. **Three types of interaction**. American Journal of Distance Education, 3(2), 1-7. 1989.

PEÑA-AYALA, A. **Educational data mining: A survey and a data mining-based analysis of recent works**. Expert systems with applications. 2014.

RAMOS, D. **Processos colaborativos mediados pelo computador e as contribuições da Teoria da Atividade**. RBIE. 2010.

RODRIGUES, R. *et al.* **A literatura brasileira sobre Mineração de Dados Educacionais**. In Anais CBIE. 2014.

ROMERO, C., VENTURA, S. **Data mining in education**. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2013.

SIEMENS, G., BAKER, R. **Learning analytics and educational data mining: towards communication and collaboration**. In Proceedings of international conference on LAK. 2012.

SILVA, J. C., BRITO, A. V., MEDEIROS, F. P. **Mapeamento Sistemático da Literatura acadêmico-científica sobre SNA aplicada em E-Learning**. RBIE. 2015.

TORGO, L. **Data mining with R: learning with case studies**. Chapman & Hall/CRC. 2010.

VYGOTSKY, L. **Mind in society: The development of higher psychological processes**. Harvard. 1978.

YIN, R. K. **Case study research: Design and methods**. Sage publications. 2013.