

Um Modelo Preditivo da Evasão dos Alunos na EAD a partir dos Construtos da Teoria da Distância Transacional

Jorge L. C. Ramos¹, João C. Sedraz Silva¹, Leonardo Cavalcante Prado¹, Alex Sandro Gomes², Fernando da F. de Souza², Erik de Gouveia Zambom², Rodrigo L. Rodrigues³

¹Universidade Federal do Vale do São Francisco
Juazeiro - BA – Brasil

²Universidade Federal de Pernambuco
Recife-PE, Brasil

³Universidade Federal Rural de Pernambuco
Recife-PE, Brasil.

Abstract. *The use of Transactional Distance Theory constructs [Moore, 1973; 1993; 2013] as predictors of student dropout in distance education (DE) was the main objective of this work. We used a set of representative variables of these constructs, obtained from a database of two higher courses offered by DE, through a process of multivariate data analysis. Next, supervised learning algorithms were used to define predictive models of student dropouts. The logistic regression model presented the best results, with success rates above 89%. These results pointed out that TDT constructs can establish a relationship with student dropout, confirming, for the courses and data analyzed, the assumptions found in the literature about this relation.*

Resumo. *O uso dos construtos da Teoria da Distância Transacional [Moore, 1973; 1993; 2013] como preditores da evasão de alunos na EAD foi o principal objetivo desse trabalho. Foi utilizado um conjunto de variáveis representativas desses construtos, obtido de uma base de dados de dois cursos superiores ofertados por EAD, por meio de um processo de análise multivariada de dados. Em seguida, foram utilizados algoritmos de aprendizagem supervisionadas para definição de modelos preditivos da evasão de alunos, sendo o modelo por regressão logística o que apresentou melhores resultados, com índices de acerto superiores a 89%. Esses resultados apontaram que os construtos da DT podem estabelecer uma relação com a evasão dos alunos, confirmando, para os cursos e dados analisados, as suposições encontradas na literatura acerca dessa relação.*

1. Introdução

O desenvolvimento da Educação a Distância (EAD) tem sido apoiado por teorias que apoiam o planejamento e a execução de cursos mais eficientes. As pesquisas na área—refletem o crescimento dessa modalidade, à medida que buscam atenuar problemas decorrentes dessa expansão, como os altos índices de evasão ainda verificados nesses cursos.

Para a maioria das instituições que participou do Censo Anual da EAD no Brasil em 2015, o grande obstáculo enfrentado foi a evasão nos cursos, quando para 40% das instituições pesquisadas, a taxa média de evasão foi entre 26% e 50% nos cursos totalmente ofertados a distância.

Partindo de uma necessidade da aplicação de teorias da EAD, no sentido de auxiliar no enfrentamento de um dos desafios da modalidade, esta pesquisa enfocou a **Teoria da Distância Transacional**, proposta por Moore (1973, 1993, 2013) como marco teórico principal do estudo,

com o objetivo de apresentar um modelo preditivo da evasão dos alunos em EAD, com base nas variáveis que compõem cada construto da Distância Transacional (DT).

Foi utilizada uma nova abordagem para determinação dos seus construtos da DT, descrita em Ramos *et al.* (2016) com o propósito de aplicá-los em um processo de detecção precoce de alunos com tendências a evasão, em cursos EAD. A utilização de técnicas de análise multivariada para a obtenção dos construtos da DT teve a intenção de buscar uma abordagem distinta das verificadas na literatura, que, na maioria dos casos, utilizam questionários aplicados a alunos e professores.

A partir da obtenção dos componentes dos construtos da DT em dois cursos de graduação ofertados a distância por uma universidade pública, foi definido e validado um modelo de previsão da evasão. Foram usados diversos algoritmos classificadores, sendo o classificador por regressão logística o que apresentou resultados mais relevantes quando comparados às pesquisas semelhantes verificadas na literatura na área.

Os resultados permitiram concluir que um conjunto de variáveis representativas da DT, quando incorporadas a um modelo preditivo de evasão de alunos, pode fornecer subsídios para o planejamento de intervenções, de modo a buscar reduzir o risco de evasão de estudantes na EAD.

2. Fundamentação Teórica

2.1 Teoria da Distância Transacional

Dentre as várias teorias que norteiam a EAD, a **Teoria da Distância Transacional** [Moore, 1973; 1993; 2013] é destacada por abrangência e atualidade, pois o autor tem feito renovações da mesma, considerando a própria evolução da EAD. Outros pesquisadores também a consideram como uma teoria seminal e importante para a modalidade [Peters, 2002; Zhang, 2003; Sandoe, 2005; Goel *et al.*, 2012; Huang *et al.*, 2015].

A separação geográfica e temporal existente na EAD conduz a padrões especiais de comportamento de alunos e professores, e afeta tanto o ensino quanto a aprendizagem. Com essa separação, surge um espaço psicológico e comunicacional a ser transposto, um espaço de potenciais mal-entendidos entre as intervenções do instrutor e as do aluno. Esse espaço psicológico e comunicacional é a **distância transacional** – DT [Moore, 1993].

De forma simplificada, essa teoria preconiza que os cursos a distância podem ser avaliados e planejados segundo uma medida derivada de três construtos: **autonomia do aluno, diálogo e estrutura do curso**.

A **autonomia** avalia a liberdade que os estudantes possuem para decidirem sobre o seu ritmo e percurso de aprendizagem. O **diálogo** é o construto usado para descrever a interação entre os participantes de um curso. A **estrutura** expressa o nível de flexibilidade das estratégias pedagógicas e quanto o curso pode ser personalizado aos interesses dos estudantes [Moore, 1993].

Essa é a principal teoria sob a qual esta pesquisa foi desenvolvida, embora outras teorias também tenham sua importância destacada na evolução da EAD. A sua escolha deveu-se ao fato da Teoria da Distância Transacional ser abrangente e incorporar conceitos essenciais de outras teorias da modalidade [Saba, 2007; Gokool-Ramdoe, 2008].

De fato, abordar a EAD por meio de uma visão sistêmica que subdivide todas as componentes da distância em vários grupos, facilita diversas intervenções, de tal forma que a modalidade possa se favorecer de um processo de melhoria contínua [Gokool-Ramdoe, 2008].

2.2 Evasão na EAD

De acordo com os censos anuais da Associação Brasileira de Educação a distância (ABED)¹, as taxas de evasão reportadas nos cursos a distância são maiores que as nos cursos presenciais. O Censo EAD.BR 2015 registrou uma evasão entre 26% e 50%, com 40% das ocorrências nas instituições que oferecem cursos regulamentados totalmente a distância. A Tabela 1 compara os índices dos 3 últimos censos feitos pela Associação Brasileira de Educação a Distância [ABED, 2014; 2015; 2016].

Tabela 1 – Taxas de evasão em cursos a distância.

Taxas de evasão declaradas	Percentuais de Instituições declarantes, por faixa.		
	2013	2014	2015
Até 25%	65%	50%	53%
Entre 26 e 50%	24%	38%	40%
Acima de 50%	2%	2%	7%
Não declararam	9%	10%	-

Fonte: [ABED, 2014; 2015; 2016].

Entre os motivos da evasão investigados pelas IES e declarados no censo, a falta de tempo dos alunos tem sido a principal causa. Curiosamente, uma das vantagens da modalidade é, exatamente, a flexibilidade que a mesma oferece em relação ao tempo para o estudo. Novas pesquisas podem confirmar ou contrapor essa declaração. Outros motivos apontados são: questões financeiras e a falta de adaptação à modalidade EAD ou à metodologia do curso. Todas essas questões reforçam a demanda por pesquisas e ações que busquem não somente estancar o crescimento, mas também reduzir essas taxas de evasão.

2.3 Relação entre a distância transacional e a evasão em cursos a distância

Pela sua definição, a Distância Transacional (DT) é um dos fatores que pode impedir ou dificultar estudantes no engajamento significativo e na comunicação no ambiente de aprendizagem [Goel *et al.*, 2012]. Além do próprio Moore (2013), outros autores afirmaram que quanto maior for a DT, maior a possibilidade de problemas que podem levar ao abandono de cursos (Zhang, 2003; Steinman, 2007; Horzum, 2011; Paul *et al.*, 2015).

Zhang (2003), demonstrou uma correlação negativa da DT com o envolvimento e a aprendizagem dos alunos, assim como a sensação de satisfação e a intenção do aluno em persistir no seu curso *on-line*. A autora também afirmou que o abandono do curso pelo estudante tem sido um problema sério da EAD. A teoria e a pesquisa em distância transacional podem lançar luzes sobre essa questão.

Na revisão do trabalho de Zhang (2003); e Paul *et al.* (2015) reforçaram a ideia de que os elementos da DT são preditores do engajamento e da satisfação dos alunos em cursos *on-line*, o que, de certa forma, são elementos relacionados com a decisão de permanecer ou abandonar o curso.

Para Steinman (2007), as percepções dos alunos de cursos *on-line* podem ser negativas se eles experimentam grande distância transacional com o instrutor e com os outros alunos, podendo influenciar sua decisão de permanecer no curso. Uma vez que a DT afeta a satisfação e retenção dos alunos, esse conceito é visto como um importante tópico de discussão sobre evasão em cursos *on-line*.

Os construtos da DT podem refletir um estado de um curso durante a sua realização, permitindo que professores e tutores percebam um distanciamento de determinados alunos e possam intervir no sentido de prever ou reverter situações de evasão [Horzum, 2011].

¹ www.abed.org.br

2.4 Mineração de Dados Educacionais

A Mineração de Dados Educacionais (EDM – do inglês *Educational Data Mining*) é a aplicação de técnicas de mineração de dados com dados provenientes de plataformas ou ambientes de educação on-line [García *et al.*, 2011].

Por um lado, o aumento tanto do *software* educativo instrumental, bem como de bases de dados de informação dos estudantes criaram grandes repositórios de dados que refletem como os alunos aprendem. Por outro lado, o *e-learning* tem gerado grandes quantidades de dados que, devidamente, explorados e classificados, podem fornecer importantes informações sobre os alunos e os cursos [Koedinger *et al.*, 2009].

O processo de EDM converte os dados brutos de sistemas educacionais em conhecimento que pode ser usada por desenvolvedores de *software* educacional, professores, pesquisadores educacionais, entre outros. Esse processo não difere muito de outras áreas de aplicação de mineração de dados, porque ele se baseia nos mesmos passos do processo de mineração de dados em geral.

2.5 Trabalhos relacionados

Algumas propostas para a predição da evasão na EAD, usando mineração de dados foram encontradas na literatura brasileira recente. Em dois desses trabalhos [Queiroga *et al.* 2015] e [Silva *et al.* 2015], avaliou-se a evasão somente em disciplinas.

O trabalho de Rigo *et al.* (2014) usou dados de algumas disciplinas em três semestres de três cursos de graduação para inferir sobre a evasão no curso. Apenas o trabalho de Dos Santos *et al.* (2014) utilizou dados de um curso superior completo nas suas análises, sem, entretanto, fundamentar a escolha dos dados usados no modelo preditivo. A carência de uma associação do processo de escolha das variáveis a uma teoria específica da EAD foi verificada em todos os trabalhos analisados.

Internacionalmente, a evasão na EAD também foi uma preocupação evidenciada em trabalhos como os de Levy (2007); Lykourantzou *et al.* (2009); Yasmin (2013); Woodley e Simpson (2014) e Yukselturk *et al.* (2014).

3. Percurso Metodológico

Este estudo usou dois tipos de análises quantitativas. A primeira técnica foi a análise multivariada de dados, baseado no trabalho de Ramos *et al.* (2016), no qual, a partir da técnica de Análise Fatorial Confirmatória (AFC), foram identificadas as variáveis mais relevantes que poderiam representar cada um dos três construtos da DT, em um banco de dados de um ambiente virtual de aprendizagem, usado por uma Instituição de Ensino Superior (IES), na oferta regular de cursos de graduação por EAD.

A outra técnica foi a Mineração de Dados Educacionais, a partir da adaptação e do instanciamento apropriado de um modelo de processo consolidado na mineração de dados e descoberta de conhecimento: o CRISP-DM (Acrônimo de *Cross-Industry Standard Process for Data Mining*)². Esse processo é bem, documentado e descreve abordagens comumente usadas em mineração de dados [Shearer, 2000].

3.1 CRISP-EDM

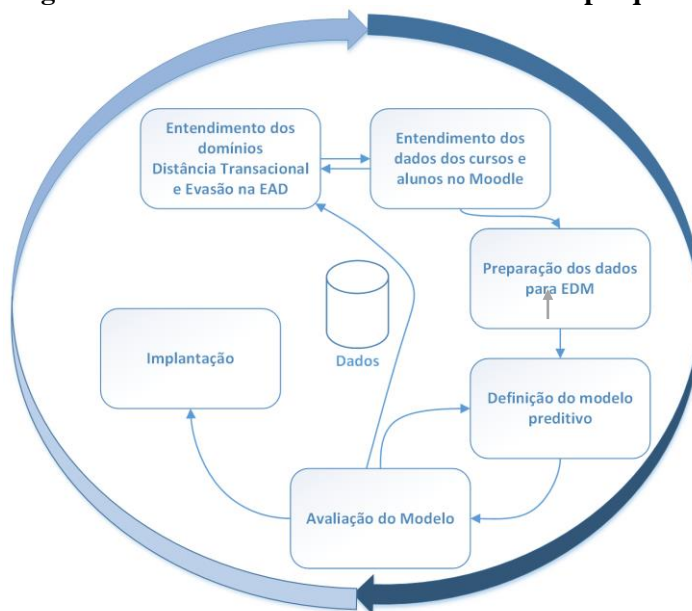
Em função da natureza e dos objetivos desta pesquisa indicarem o uso de técnicas de mineração de dados, como forma de determinação de um modelo preditivo baseado nos construtos da

² <http://www.crisp-dm.org/>

distância transacional, para alcançar os objetivos propostos, a estratégia metodológica deste estudo foi orientada pelo CRISP-DM.

O instanciamento desse processo para um contexto de EDM permitiu uma adaptação também na sua nomenclatura, para CRISP-EDM. A Figura 2 exhibe o modelo de CRISP-EDM, que foi aplicado nesta pesquisa.

Figura 2 – Modelo de CRISP-EDM usado na pesquisa.



Fonte: Adaptado de SHEARER (2000).

Cada etapa contemplou técnicas e abordagens devidamente adequadas ao domínio educacional:

1 – Entendimento dos domínios da Distância Transacional e Evasão na EAD – Análise da literatura especializada e da coleta de informações sobre a evasão da EAD, em particular os dados no Brasil, caracterizando-a como um problema a ser abordado.

2 – Entendimento dos dados dos cursos e dos alunos em uma base de dados do Moodle – Conhecimento aprofundado das tabelas, campos, registros e relacionamentos existentes no banco de dados de uma instância do Moodle usado em cursos de graduação em EAD. Foram identificadas as diversas variáveis no ambiente virtual que representam os construtos da DT, fazendo o mapeamento com base na literatura e consultas a especialistas.

3 – Preparação dos dados para EDM – Conforme o método descrito no trabalho de Ramos (2016), foram coletados os dados e feita a confirmação da lista de variáveis representativas dos construtos da DT, a partir de Análise Fatorial Confirmatória.

4 - Definição de modelos preditivos – A partir do uso de técnicas de aprendizagem supervisionada, foram definidos modelos preditivos da evasão do aluno, com base nas variáveis obtidas a partir dos construtos da DT.

5 – Avaliação dos modelos preditivos – Com base nas métricas de avaliação de modelos preditivos consagradas na literatura, cada modelo gerado foi avaliado, para se obter o que melhor representaria a previsão da evasão dos alunos, usando como fontes de dados as variáveis da DT coletadas.

Para a avaliação dos modelos preditivos, neste trabalho, foram selecionadas duas métricas de avaliação de classificadores: a **Acurácia** e a **área sob a curva ROC (Receiver Operating Characteristics)**. A **Acurácia** representa a taxa de acerto de todo o classificador e é obtida pela razão entre a soma dos acertos das duas classes e o número total de instâncias

classificadas. A **área sob a curva ROC** (AUC - *Area Under Curve*) é um valor escalar que representa o quanto um algoritmo é eficiente no processo de classificação. Seu valor irá sempre estar entre 0,0 (pior caso) e 1,0 (modelo ideal). Essa métrica é muito usada para comparar classificadores distintos.

6 – Implantação - Desenvolvimento de um protótipo de *software* que implemente o monitoramento preditivo da evasão em um ambiente virtual de aprendizagem.

Pelo objetivo deste artigo e pela limitação do espaço para a publicação, serão apresentados os resultados obtidos nas etapas 1 a 5 do CRISP-EDM, embora também tenha sido realizada a etapa de implantação e, inclusive, feitos testes de aceitação com usuários.

3.2 Bases de dados e tecnologias utilizadas

Para o desenvolvimento da pesquisa, a fonte primária e principal de dados foram os diversos bancos de dados do ambiente Moodle do Núcleo de Educação a Distância (NEAD) da Universidade de Pernambuco (UPE). Esse núcleo já conta com cerca de oito (08) anos de atuação na modalidade, ofertando cursos de graduação, especialização e extensão, com presença em diversas cidades do estado de Pernambuco.

A base de dados escolhida para ser explorada e analisada foi a de dois cursos de graduação por EAD: as licenciaturas em Pedagogia e em Ciências Biológicas. Os dados foram oficialmente cedidos pela coordenação do Núcleo, assim como toda a estrutura do ambiente virtual em uso por essa instituição. A escolha desses dois cursos foi em razão dos mesmos apresentarem os menores (**Pedagogia, Turmas 1 e 2**) e maiores (**Biologia, Turmas 3 e 4**) índices de evasão conforme informação da coordenação do Núcleo.

Esses dados foram divididos em duas bases diferentes para o processo de EDM: os dados de **Pedagogia – Turma 1 e Biologia – Turma 3**, por estarem completos, foram usados como bases para **treinamento** dos algoritmos de mineração e as outras duas foram usadas para base de **testes** dos modelos preditivos obtidos no processo. Assim, para todo o processo de mineração, foram usados dois conjuntos distintos de dados. Foram encontradas, nessas bases, tabelas com mais de 4 milhões de registros, o que dá uma ideia da sua dimensão e justifica a aplicação de um processo de EDM para busca de conhecimento nesse volume de dados.

Para a execução da pesquisa, diversas ferramentas computacionais foram utilizadas, destacando-se: 1) *SGBD MySQL e PostgreSQL* para extração e tratamento dos dados; 2) *Microsoft Excel*, para limpeza e preparação de tabelas, e 3) *R Project e RStudio*, com diversos *packages* associados, para a análise multivariada, mineração de dados e aprendizagem supervisionada.

4. Resultados

A etapa 1 do processo de CRISP-EDM apresentou, como principal resultado, o conhecimento apropriado da Teoria da Distância Transacional e das principais taxas e causas de evasão em cursos EAD no país, apresentados na fundamentação teórica.

Nas etapas 2 e 3, foram obtidas as variáveis representativas da DT nos dados dos cursos analisados. Inicialmente, professores e tutores da EAD apontaram uma lista de 39 variáveis representativas da distância transacional. Após coleta inicial de dados e aplicação da análise fatorial confirmatória, 21 variáveis foram confirmadas como relevantes para os três construtos da DT. Esse conjunto de variáveis obtido é listado no Quadro 01.

Essas variáveis foram coletadas nas bases listadas na Seção 3.2, formando dois conjuntos de dados com 6.654 instâncias para a base de treinamento e 11.440 instâncias para a base de testes.

Quadro 01 - Lista final de variáveis por construtos após a Análise Fatorial Confirmatória.

Variáveis	Construto	Carga Fatorial
Média semanal da quantidade de acessos do aluno ao ambiente no semestre.	AUTONOMIA	0,952
Quantidade de acessos do aluno ao ambiente por turno (Manhã), por semestre.	AUTONOMIA	0,771
Quantidade de acessos do aluno ao ambiente por turno (Tarde), por semestre.	AUTONOMIA	0,865
Quantidade de acessos do aluno ao ambiente por turno (Noite), por semestre.	AUTONOMIA	0,804
Tempo médio semanal de utilização da plataforma pelo aluno no semestre.	AUTONOMIA	0,650
Quantidade de acessos do aluno ao ambiente no semestre.	AUTONOMIA	0,952
Qtd de acessos do aluno aos diferentes tipos de atividades, por disciplina.	AUTONOMIA	0,783
Quantidade de acessos do aluno aos fóruns, por disciplina.	AUTONOMIA	0,828
Quantidade geral de postagens do aluno em fóruns, por disciplina.	DIÁLOGO	0,657
Quantidade geral de mensagens enviadas pelo aluno, por semestre.	DIÁLOGO	0,600
Quantidade geral de mensagens recebidas pelo aluno, por semestre.	DIÁLOGO	0,868
Qtd de colegas diferentes para quem o aluno enviou mensagens, por semestre.	DIÁLOGO	0,472
Quantidade de mensagens dos professores recebidas pelo aluno, por semestre.	DIÁLOGO	0,849
Quantidade de mensagens de colegas recebidas pelo aluno, por semestre.	DIÁLOGO	0,420
Qtd de mensagens enviadas pelo aluno para outros colegas, por semestre.	DIÁLOGO	0,474
Quantidade de mensagens enviadas pelo aluno aos professores, por semestre.	DIÁLOGO	0,604
Quantidade geral de recursos disponibilizados pelo professor por disciplina.	ESTRUTURA	0,824
Quantidade geral de atividades disponibilizadas pelo professor por disciplina.	ESTRUTURA	0,590
Qtd de atividades com prazos de envio definidos por professor, por disciplina.	ESTRUTURA	0,456
Qtd de fóruns de discussão disponibilizados sobre os conteúdos por disciplina.	ESTRUTURA	0,400
Qtd de sessões de web conferências disponibilizadas no curso, por disciplina.	ESTRUTURA	0,650

Como a tarefa de mineração de dados envolveu o modelo baseado em classificadores binários, cada instância foi rotulada como **0 (aluno não evadido)** ou **1 (aluno evadido)**, de acordo com as informações dos alunos fornecidas pelas coordenações dos respectivos cursos e, também, a partir da verificação de variáveis de acesso ao ambiente. Os resultados dessa rotulação são apresentados na Tabela 02.

Tabela 02 – Nº de casos (instâncias) de cada classe em ambas as bases.

BASE COMPLETA TESTES			BASE COMPLETA TREINAMENTO		
Classe	Nº de Casos	%	Classe	Nº de Casos	%
Não evadiu	9.502	83,1%	Não evadiu	4.941	74,3%
Evadiu	1.938	16,9%	Evadiu	1.713	25,7%
Total	11.440	100%	Total	6.654	100%

Essa rotulação serviu para que os fossem gerados os modelos preditivos na base de treinamento e que esses modelos pudessem ser avaliados, utilizando a base de testes.

Para o processo de aprendizagem supervisionada, foram utilizadas as seguintes técnicas de classificação binária, com seus respectivos algoritmos: Árvore de Decisão (TreeDecision), Máquina de Vetor de Suporte (SVM), Rede Neural Artificial (NeuralNet), *k-Nearest Neighbors* (KNN) e Regressão Logística (RegLog). A variável dependente do modelo foi o campo EVADIU presente nas bases de dados, conforme mostrado na Tabela 9. Para cada um dos classificadores, foi usado um mesmo conjunto de variáveis, baseado nos resultados obtidos na Análise Fatorial Confirmatória (Quadro 01). As métricas de avaliação de cada classificador são exibidas na Tabela 03.

Tabela 03 – Métricas de avaliação de cada classificador.

	RegLog	SVM	NeuralNet	TreeDecision	KNN
AUC	0,856	0,796	0,828	0,788	0,797
ACCURACY	0,894	0,882	0,879	0,869	0,896

Embora as técnicas de Aprendizagem de Máquina apresentadas e testadas no contexto desta pesquisa tenham, em diversas avaliações, apresentado resultados muito próximos, optou-se por escolher a **Regressão Logística** como a técnica a ser usada para gerar e testar os modelos preditivos.

A Regressão Logística estuda ou prevê o comportamento de uma variável dependente binária com base em um conjunto de variáveis independentes numéricas ou categóricas. Esse

método também pode acomodar diferentes tipos de variáveis independentes (métricas e não métricas). Os dados oriundos dos registros de banco de dados de ambientes virtuais de aprendizagem são heterogêneos e podem ser de diferentes tipos, sem necessidade de processos adicionais de transformação nos dados. A matriz de confusão para esse classificador é apresentada na Tabela 04:

Tabela 04 – Matriz de Confusão para o classificador por Regressão Logística

<i>RegLog</i>	Referência (Real)	
Modelo (Previsto)	NÃO EVADIU	EVADIU
NÃO EVADIU	9035	742
EVADIU	467	1196

As linhas dessa matriz representam os quantitativos obtidos em cada classe pelo modelo de previsão. As colunas indicam os valores reais de cada classe. A diagonal principal indica o número de classificações corretas em cada classe (Verdadeiros Negativos e Verdadeiros Positivos) e a secundária, as classificações incorretas (Falsos Negativos e Falsos Positivos).

A Regressão Logística, como técnica multivariada, é menos afetada quando suposições estatísticas básicas, como a normalidade dos dados, não são satisfeitas. Isso simplifica o processo inicial do CRISP-EDM de entendimento e preparação dos dados.

5. Discussões e considerações finais

A abordagem proposta para a obtenção dos componentes que podem obter os construtos da distância transacional, a partir do mapeamento das variáveis no banco de dados do Moodle que representam esses construtos, lança uma nova perspectiva para as pesquisas sobre a sua teoria. Embora não tenha sido estabelecida uma medida da distância transacional, o fato de se poder dispor de suas componentes de maneira mais automática e em momentos distintos de um curso, representa um passo importante para novas aplicações da teoria na EAD.

As variáveis obtidas na Análise Fatorial Confirmatória representam o modelo relacionado com o conjunto de dados analisados nesta pesquisa. Então, cabe esclarecer que as variáveis aqui definidas podem não ser as únicas a representar os construtos, mas o fazem dentro de padrões de validação aceitáveis, segundo as métricas de avaliação do processo de obtenção (Análise Fatorial Confirmatória) encontradas na literatura.

Em relação às taxas de evasão obtidas nos cursos analisados, destaca-se que o processo de obtenção dessas taxas também pode acontecer de maneira automática, a partir de consultas a determinadas variáveis na base de dados do Moodle. As variáveis descritas no Quadro 01 possibilitam identificar um aluno evadido ou com grande possibilidade de evadir-se do curso, já que registram suas principais interações no ambiente, inclusive todos os seus acessos.

A etapa de descoberta de conhecimento em bases de dados acrescentou o componente de inteligência computacional à pesquisa. Um conjunto de cinco técnicas e algoritmos de aprendizagem de máquina foi utilizado para a obtenção do modelo preditivo para evasão de alunos. Um fato relevante nesta etapa é que não houve a necessidade de divisão da base de dados em bases para treinamento e testes, como geralmente ocorre em processos de aprendizagem supervisionado. As bases distintas testaram e destacaram a capacidade preditiva dos modelos, evitando problemas como sub e superestimação de parâmetros nos modelos descobertos.

A literatura de previsão de evasão em EAD, usando técnicas de mineração de dados, ainda é limitada. A maioria dos trabalhos é sobre evasão em cursos presenciais e outros fazem predições de abandono de disciplinas on-line. Mesmo assim, foi possível comparar um dos resultados obtidos nesta pesquisa com outros trabalhos relacionados na literatura. Nesses trabalhos, a métrica de avaliação do classificador, que predomina, é a acurácia. A Tabela 05 exhibe os dados desses outros estudos, mostrando que esta pesquisa, com uma acurácia de

89,42%, obtida na Regressão Logística, apresentou resultado melhor que esses trabalhos relacionados.

Tabela 05 – Resultados de modelos preditivos de evasão em trabalhos relacionados.

Obra	Algoritmos/Técnicas	Acurácia
[Kotsiantis <i>et al.</i> , 2003]	Naive Bayes	83%
[Morris <i>et al.</i> , 2005]	Análise Discriminante Preditiva	74,5%
[Roblyer <i>et al.</i> , 2008]	Regressão logística	79,3%
[Lykourantzou <i>et al.</i> , 2009]	Combinação de técnicas	85%
[Kovacic, 2010]	Árvore de Decisão	60,5%
[Yasmin, 2013]	Árvore de Decisão	84,8%
[Yukselturk <i>et al.</i> , 2014]	Redes Neurais	87%
[Rigo <i>et al.</i> , 2014]	Redes Neurais	76,5%
[Cambruzzi, 2014]	Redes Neurais	75,7%
[Dos Santos <i>et al.</i> , 2014]	Árvore de Decisão	81,64%
[Silva <i>et al.</i> , 2015]	Árvore de Decisão	73,37%
[Queiroga <i>et al.</i> , 2015]	Vários	79,76%

É importante destacar também que esses trabalhos usaram diferentes tipos de dados, como informações demográficas, resultados de exames prévios de admissão, dados de desempenho dos alunos, questionários entre outros. Nenhum deles usou uma teoria como norteadora na definição das variáveis preditoras. Alguns utilizaram e analisaram dados para prever evasão em disciplinas. Assim, não se pôde estabelecer comparações mais detalhadas com os resultados desta pesquisa.

Essa lista de variáveis preditoras por cada construto estabelece uma relação entre os construtos da Teoria da Distância Transacional e a evasão dos alunos na EAD, confirmando, para os cursos e dados analisados, as suposições encontradas na literatura acerca dessa relação. A percepção dos níveis de cada uma dessas variáveis para cada aluno, a partir do modelo de predição definido, pode auxiliar o instrutor ou tutor na sua ação preventiva e assim buscar a reversão de possíveis alunos com tendências de evasão.

Ao ser observada lista das variáveis significativas do modelo, percebe-se que são indicativos triviais em cursos a distância, de fácil percepção e compreensão pelos profissionais da modalidade. Isso impacta diretamente nas ações que podem ser feitas no sentido de melhorar esses indicadores e, como consequência, reduzir a probabilidade de evasão dos alunos com esse risco.

Agradecimentos

Esta pesquisa é apoiada pela FACEPE-APQ No. 0525-1.03/14. Os autores também agradecem ao apoio do NEAD/UPE e à PRPPGI/UNIVASF. Alex Sandro Gomes é bolsista DT Nível 2/CNPq, processos nº 310466/2012-1 e nº 475634/2013-6.

Referências

- BROWN, T. A. Confirmatory factor analysis for applied research. Guilford Publications 2015.
- CAMBRUZZI, W. L. GVwise: uma aplicação de learning analytics para a redução da evasão na educação a distância. 2014. (Dissertação de Mestrado). Pós-Graduação em Computação Aplicada, Universidade do Vale do Rio dos Sinos (UNISINOS)
- DOS SANTOS, R. *et al.* Uma Abordagem Genérica de Identificação Precoce de Estudantes com Risco de Evasão em um AVA utilizando Técnicas de Mineração de Dados. XIX Congresso Internacional de Informática Educativa. Fortaleza-CE 2014.
- GARCÍA, E. *et al.* A collaborative educational association rule mining tool. The Internet and Higher Education, v. 14, n. 2, p. 77-88, 2011. ISSN 1096-7516.
- GOEL, L.; ZHANG, P.; TEMPLETON, M. Transactional distance revisited: Bridging face and empirical validity. Computers in Human Behavior, v. 28, n. 4, 2012.

- GOKOOL-RAMDOO, S. Beyond the theoretical impasse: Extending the applications of transactional distance education theory. *The International Review of Research in Open and Distributed Learning*, v. 9, n. 3, 2008.
- HORZUM, M. B. Developing Transactional Distance Scale and Examining Transactional Distance Perception of Blended Learning Students in Terms of Different Variables. *Educational Sciences: Theory and Practice*, v. 11, n. 3, p. 1582-1587. 2011
- HUANG, X. *et al.* Understanding transactional distance in web-based learning environments: An empirical study. *British Journal of Educational Technology* 2015. ISSN 1467-8535.
- KOEDINGER, K. R. *et al.* Educational software features that encourage and discourage “gaming the system”. *Proceedings of the 14th International Conference on Artificial Intelligence in Education*, 2009. p. 475-482.
- KOTSIANTIS, S. *et al.* Preventing student dropout in distance learning using machine learning techniques. *Knowledge-Based Intelligent Information and Engineering Systems*, 2003, Springer. p. 267-274.
- LEVY, Y. Comparing dropouts and persistence in e-learning courses. *Computers & education*, v. 48, n. 2, p. 185-204, 2007. ISSN 0360-1315.
- LYKOURENTZOU, I. *et al.* Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education*, v. 53, n. 3, p. 950-965, 2009. ISSN 0360-1315.
- MOORE, M. G. The theory of transactional distance. In: MOORE, M. G. (Ed.). *Handbook of distance education*. New York: Routledge 2013. cap. 5, p.66-85.
- MOORE, M. G. Theory of transactional distance. In: (Ed.). *Theoretical Principles of Distance Education*. New York: Routledge, 1993. p. 2-29.
- MOORE, M. G. Toward a theory of independent learning and teaching. *The Journal of Higher Education*, p. 661-679, 1973. ISSN 0022-1546.
- PAUL, R. C. *et al.* Revisiting Zhang’s scale of transactional distance: refinement and validation using structural equation modeling. *Distance Education*, v. 36, n. 3, p. 364-382 2015.
- QUEIROGA, E. *et al.* Um Estudo do Uso de Contagem de Interações Semanais para Predição Precoce de Evasão em Educação a Distância. *Anais dos WCBIE*, 2015. p.1074.
- RAMOS, J. L. C., Silva, J., Rodrigues, R., Gomes, A. S., & de Souza, F. D. F. Mapeamento de dados de um LMS para medida de construtos da distância transacional. In *Simpósio Brasileiro de Informática na Educação-SBIE*. (Vol. 27, No. 1, p. 1056), 2016.
- RIGO, S. J. *et al.* Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios. *Revista Brasileira de Informática na Educação*, v. 22, n. 01, p. 132, 2014.
- SABA, F. (2007). A systems approach in theory building. In M. G. Moore (Ed.) *Handbook of distance education*. (pp.43-57). Mahwah, NJ: Lawrence Erlbaum.
- SANDOE, C. Measuring transactional distance of online courses: The structure component. PhD. University of South Florida. 2005.
- SHEARER, C. The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, v. 5, n. 4, p. 13-22, 2000.
- SILVA, F. *et al.* Um modelo preditivo para diagnóstico de evasão baseado nas interações de alunos em fóruns de discussão. *Anais do Simpósio Brasileiro de Informática na Educação*, 2015. p.1187.
- STEINMAN, Debbie. Educational experiences and the online student. *TechTrends*, v. 51, n. 5, p. 46-52, 2007.
- WOODLEY, A.; SIMPSON, O. Student dropout: The elephant in the room. *On-line distance education: Towards a research agenda*, p. 459-484, 2014.
- YASMIN, D. Application of the classification tree model in predicting learner dropout behaviour in open and distance learning. *Distance Education*, v. 34, n. 2, p. 218-231, 2013. ISSN 0158-7919.
- YUKSELTURK, E. *et al.* Predicting dropout student: an application of data mining methods in an on-line education program. *European Journal of Open, Distance and E-learning*, v. 17, n. 1, p. 118-133, 2014.
- ZHANG, A. *Transactional Distance in Web-based College Learning Environments: Toward Measurement and Theory Construction*. PhD Thesis. VCU Retrospective. Richmond. 2003.