Modelo de Regressão Linear aplicado à previsão de desempenho de estudantes em ambiente de aprendizagem

Rodrigo Lins Rodrigues[1,3], Francisco P. A. de Medeiros[2], Alex Sandro Gomes[3]

[1]Universidade Federal Rural de Pernambuco, Instituto Federal da Paraíba[2], Universidade Federal de Pernambuco[3]

rlr@ded.ufrpe.br, fcopetroniomedeiros@gmail.com, asg@cin.ufpe.br

Abstract. *Extracting relevant information in collaborative supported computer learning to enable an effective monitoring of students and fostering learning management is a challenge. Several works have been developed from data mining techniques for the purpose of educational decision making. The main objective of this study was to investigate the feasibility of using the linear regression model to obtain inferences in the early stages of conducting online courses as a way to support decision making by teachers and administrators. We propose the use of linear regression to estimate the performance of students based on their interactions within the learning management system, taking into account behavioral variables. The results showed that it's possible to use the linear regression technique to obtain inferences with good accuracy rates.*

Resumo. *Extrair informações relevantes que auxiliem a gestão da aprendizagem e viabilizem o acompanhamento efetivo de estudantes em cursos mediados por tecnologia é um desafio. Diversos trabalhos vem sendo desenvolvidos a partir de técnicas de mineração de dados educacionais para fins de tomada de decisão. O objetivo principal deste trabalho foi investigar a viabilidade da utilização do modelo de regressão linear para a obtenção de inferências em etapas iniciais da realização de cursos online, como forma de apoiar a tomada de decisão por parte de professores e gestores. Propomos a utilização da técnica de regressão linear para estimar o desempenho de alunos baseados em suas interações dentro da plataforma virtual de aprendizagem, levando em consideração variáveis comportamentais. Os resultados obtidos demonstraram que é possível utilizar a técnica de regressão linear para obter inferências com boas taxas de precisão.*

1. Introdução

Interação é um conceito complexo e multifacetado em todas as formas de educação. Diversas taxonomias foram propostas ao longo das últimas décadas. Moore (1989) descreveu três formas de interação em educação à distância: interação entre o professor e o estudante, interação entre os estudantes e interação entre o estudante e o conteúdo. Anderson e Garrison (1998) expandiram a taxonomia proposta por Moore acrescentando três novas formas de interação: interação entre os professores, interação entre o professor e o conteúdo e interação entre os conteúdos. Hirumi (2006)

DOI: 10.5753/CBIE.SBIE.2013.607

complementou as taxonomias anteriores, considerando as interações entre o professor e o aprendiz com as ferramentas.

Conforme apresentado nas taxonomias, há inúmeras possibilidades de interação nos AVAs que suportam o processo de aprendizagem online. Há na literatura trabalhos que associam a relação entre o nível de interação dos participantes de cursos *online* com indicadores de desempenho acadêmico. Picciano (2002) observou uma correlação entre interações e aprendizagem e concluiu que esse desempenho e o sucesso de um curso dependem da habilidade de interagir dos participantes. Russo e Benson (2004) encontraram uma correlação significativa entre o desempenho dos estudantes e a percepção da própria presença social *online*, a qual depende das interações estabelecidas entre os participantes do curso. Watson (2008) associa às interações em ferramentas assíncronas de comunicação - como fóruns de discussão - com a maneira como os estudantes experimentam um maior sentido de engajamento. Johnson *et al*. (2010) enfatiza a importância das interações online na melhoria dos resultados de aprendizagem.

Trabalhos que objetivam mediar as relações entre os níveis de colaboração e engajamento com o desempenho acadêmico utilizam, em sua maioria, a técnica de análise de correlação; muito útil em trabalhos exploratórios, quando o interesse está no grau ou força do relacionamento.

Neste estudo utilizamos uma análise de regressão, cujo foco principal está na busca da natureza do relacionamento entre interação e desempenho.

De acordo com Souza (2001) a análise de regressão compreende analisar dados amostrais para saber se e como duas ou mais variáveis relacionam-se umas com as outras. A análise de regressão tem como resultado uma equação matemática que descreve o relacionamento entre essas variáveis e pode ser usada para estimar ou predizer, valores futuros de uma variável quando se conhecem ou se supõem conhecidos valores da outra variável.

Nosso trabalho apresenta um estudo preliminar do uso de algumas técnicas de Mineração de Dados Educacionais (MDE), especificamente a Análise de Regressão Linear, em uma base de dados do AVA Moodle[1]. O objetivo principal é a obtenção de um modelo de previsão de desempenho de estudantes a partir de dados de interação durante o período de realização de um curso na modalidade à distância.

Este artigo está organizado da seguinte forma. Na seção 2 é apresentada a revisão da literatura sobre a área de mineração de dados educacionais, em particular a técnica de regressão linear. Na seção 3 apresenta-se o experimento realizado, objetivos e procedimentos. A seção 4 mostra os resultados. A seção 5 mostra as discussões e contribuições e por fim a seção 6 apresenta as conclusões do trabalho.

2. Revisão da Literatura

Nesta seção iniciamos a revisão de literatura na qual são apresentados os conceitos necessários à fundamentação deste trabalho. Entre os conceitos a serem explorados estão: Mineração de dados educacionais (EDM) e Regressão Linear simples.

2.1 Mineração de dados educacionais (MDE, do Inglês EDM)

[1] www.moodle.org

A Mineração de dados tem sido aplicada em diversas áreas do conhecimento. Recentemente, com a expansão dos cursos à distância e com suporte computacional, muitos pesquisadores da área de Informática na Educação (em particular, Inteligência Artificial Aplicada à Educação) têm mostrado interesse em utilizar mineração de dados para investigar perguntas científicas na área de educação (e.g., quais são os fatores que afetam a aprendizagem? Ou como desenvolver sistemas educacionais mais eficazes?). Dentro deste contexto, surgiu uma nova área de pesquisa conhecida como "Mineração de Dados Educacionais".

A EDM é definida como a área de pesquisa que tem como principal foco o desenvolvimento de métodos para explorar conjuntos de dados coletados em ambientes educacionais. Assim, é possível compreender de forma mais eficaz e adequada os alunos, como eles aprendem, o papel do contexto na qual a aprendizagem ocorre, além de outros fatores que influenciam a aprendizagem [GOTTARDO, 2012]. Por exemplo, é possível identificar em que situação um tipo de abordagem instrucional (e.g., aprendizagem individual ou colaborativa) proporciona melhores benefícios educacionais ao aluno.

Atualmente, as tendências na área indicam um grande crescimento de pesquisas em EDM no cenário internacional. Este grande crescimento é resultado direto de fatores como, a difusão e utilização de softwares educacionais que produzem grande quantidade de dados educacionais bem estruturados [Baker, 2011]. Um dos fatores que vem promovendo o crescimento da EDM é o uso de ambientes virtuais de aprendizagem (AVA). Esses ambientes vêm sendo adotados por muitos professores, escolas e universidades em todo mundo e permitem que pesquisadores possam obter grandes quantidades de dados reais e relevantes para realizar análises utilizando técnicas provindas da área de EDM.

2.1.1 Regressão Linear Simples

A regressão linear é uma tentativa de modelar uma equação matemática linear que descreva o relacionamento entre duas variáveis [CURRAL, 1994].

Há diversas maneiras de uso dessas equações de regressão, em situações envolvendo duas variáveis. Por exemplo, um pesquisador em EDM pode tentar explicar as variações do desempenho de alunos em função do aumento da carga horária diária de estudos. Deve-se notar, entretanto, que a lógica de uma relação causal deve advir de fenômenos externos ao âmbito da estatística. A análise estatística de regressão apenas modela qual relacionamento matemático pode existir, se existir algum [HOFFMANN, 2006].

Ainda como aplicação da regressão pode-se estar interessado em predizer valores futuros de uma variável. Por exemplo, costuma-se aplicar testes a empregados ou estudantes potenciais, para avaliar o potencial de sucesso na escola ou no emprego. É de presumir que haja um relacionamento matemático entre o resultado do teste e o potencial futuro.

Embora tais relações possam assumir uma grande diversidade de formas, discutiremos apenas as equações lineares. As equações lineares são importantes porque servem para aproximar muitas relações da vida real, e porque são relativamente fáceis de lidar e de interpretar. Outras formas da análise de regressão, tais como regressão múltipla (mais de duas variáveis) e regressão curvilínica (não-linear) envolvem extensões dos mesmos conceitos usados na regressão linear simples.

2.1.2 Equação Linear

Duas importantes características da equação linear são: o coeficiente angular e o coeficiente linear da reta em determinado ponto. Uma equação linear tem a forma:

$$\gamma i = \alpha + \beta x i + €i, \qquad i=1,2,...,n$$

α é coeficiente linear da reta, β é o coeficiente angular e n é o tamanho da amostra. A variável γ é a variável que deve ser predita (variável resposta ou dependente), e x é o valor preditor (variável independente ou explicativa) isto é, preferencialmente deve-se utilizar x como a variável mais acessível (mais fácil de ser obtida) e *Y* a variável menos acessível (mais difícil de ser obtida). É importante ter em mente que nem todas as situações são bem aproximadas por uma equação linear [HOFFMANN, 2006]. Por isso, em geral é necessário desenvolver um trabalho preliminar para determinar se um modelo linear é adequado. O processo mais simples consiste em fazer o diagrama de dispersão dos dados, isto é, verificar se os valores de x e de γ, mostram uma tendência visual linear.

3. Realização do experimento

3.1 Objetivo

O objetivo deste experimento foi estimar a relação existente entre a quantidade de interações com os artefatos educacionais disponibilizados na disciplina (vídeo, fórum, textos, slides) com o desempenho acadêmico (média final) dos alunos a partir da abordagem de análise de regressão linear.

3.2 Procedimentos

A realização do experimento se deu através da disciplina de Jogos Educacionais da graduação em Licenciatura em Computação da Universidade Federal Rural de Pernambuco. O curso contou com a participação de 98 alunos do sétimo período na modalidade de EAD.

A disciplina teve duração de dois meses durante o primeiro semestre de 2013, divididos em seis módulos de conteúdos. Dentre os módulos que foram abordados, tivemos: Semana 1 (27 a 31/05) - Nativos digitais, Geração Y e imigrantes digitais, Semana 2 (03 a 07/06) - Estilos de aprendizagem, Semana 3 (10 a 14/06) - Design de jogos educacionais, Semana 4 (17 a 28/06) - Projeto conceitual e fundamentos de Game Design e por fim nas Semanas 5 e 6 (01 a 12/07) - Técnicas e Ferramentas de Implementação de jogos.

Durante a disciplina diversos artefatos de interação foram utilizados, tais como: utilização de fóruns, vídeo-aulas, material em texto, disponibilização de slides e envio de atividades. As avaliações da disciplina foram compostas por duas unidades, a primeira unidade foi composta por três atividades virtuais e uma prova escrita presencial. A segunda unidade foi composta por duas atividades virtuais e uma prova escrita, ao final o aluno teve seu desempenho calculado através da média das duas unidades.

3.3 Coleta e tratamento dos dados

Para a construção do banco de dados, utilizou-se a base de dados do AVA Moodle com informações sobre a interação dos participantes com os diversos artefatos educacionais disponibilizados durante o curso. Durante o experimento não foram considerados estudantes desistentes, tendo em vista a indisponibilidade de resultado ou nota final nesses casos.

A Tabela 1 a seguir exibe as principais variáveis coletadas para realizar a aplicação da técnica de regressão:

Tabela 1. variáveis de interação coletadas

Variáveis	Sobre
Desemp_final	Esta variável foi composta por dados das 5 atividades virtuais e das duas provas presenciais, totalizando uma média representativa do desempenho final do aluno.
Int_forum_sem1, Int_forum_sem2, Int_forum_sem2, Int_forum_sem4, Int_forum_sem5,	Esta variável foi composta pelo somatório das quantidades de interações nos seis fóruns disponibilizados na disciplina durante todo o curso.
Vídeo_sem1, Vídeo_sem2, Vídeo_sem3 Vídeo_sem4, Vídeo_sem5, Vídeo_sem6	Essa variável foi composta pelo somatório das quantidades de interações nas vídeo aulas disponibilizadas pelo professor da disciplina.
Mat_sem1, Mat_sem2, Mat_sem3, Mat_sem4, Mat_sem5, Mat_sem6	Esta variável foi composta através dos dados de acesso dos materiais em formato de textos, apostilas disponibilizadas para leitura.
Ppt_sem1, Ppt_sem2, Ppt_sem3, Ppt_sem4, Ppt_sem5, Ppt_sem6	Variável que mostra a quantidade de acesso ao material do tipo slide.
Media_horas_acesso	Esta variável foi composta pela hora média que os alunos costumam acessar o ambiente virtual.

Estas variáveis foram coletadas através do próprio painel de logs do Moodle, no qual foi possível extrair dados no formato .xls que serviram pra quantificar o número de acessos para cada artefato disponibilizado durante a disciplina. Estes dados passaram por um processo de tratamento para posteriormente ser feito as possíveis análises.

4. Resultados

Os resultados obtidos na construção deste experimento foram realizados através do *software estatístico R*[2], modelos lineares podem ser acessados no ambiente R a partir da função `lm()`, que gera um objeto correspondente a uma lista com vários componentes do modelo linear em questão (coeficientes, resíduos, valores preditos, dentre outros). A estes objetos encontram-se associados também modos específicos das funções `plot()` e `summary()`, mostrando tabelas de análise de variância para o modelo linear. Nas seções seguintes veremos com mais detalhes cada uma das análises realizadas.

[2] http://www.r-project.org/

4.1 Análise descritiva

Antes da aplicação da análise de regressão, foi necessário realizar algumas estatísticas descritivas com o intuito de ter uma visão mais ampla sobre o comportamento dos dados coletados.

A primeira análise realizada foi com relação às médias de acesso por artefatos durante os dois meses de curso, verificou-se que os alunos utilizaram a ferramenta fórum em uma média de 24,6 vezes durante todo o curso e com uma média de 11 interações por fórum semanalmente. Com relação ao artefato vídeo-aula verificou-se que a média de interação foi de 5,6 acessos nos 6 vídeos disponibilizados. Em relação aos textos disponibilizados, obtivemos uma média de 7,3 acessos nos seis textos disponibilizados. Os materiais em formatos de slides, tiveram uma média de 6,5 acessos, e por fim tivemos uma média de horários que os alunos mais acessam o ambiente, e obtivemos uma valor médio em horas de 12:26 h. A próxima seção aborda a parte inferencial do trabalhando, modelando uma regressão linear de acordo com os dados de interação dos alunos.

4.2 Estimando o modelo linear e aplicando a técnica de regressão

Para ajustar um modelo de regressão a um conjunto de dados, antes é necessário fazer a verificação dos pressupostos do modelo tais como: (1) A relação entre x *e* y deve ser linear, (2) Os valores de x são fixos, isto é, x não é uma variável aleatória, (3) A média dos erros é nula, isto é, $E(ei)=0,\ i=1,2,...,n$; (4) O erro em uma observação é não correlacionado com o erro em qualquer outra observação e por fim (5) os erros devem seguir uma distribuição normal.

4.3 Teste de normalidade e pressupostos do modelo

Na seção anterior verificamos que um dos pressupostos é a normalidade dos resíduos das variáveis. No nosso estudo temos como variável dependente o desempenho final do aluno na disciplina (média final) e como variáveis independentes: (1) media da quantidade de interações em fóruns *(MediaIntForum)*, (2) Média das interações com vídeo-aulas *(MediaIntVideo)*, (3) Média das interações com textos *(MediaIntMat)* e por fim a (4) média das interações com slides em formato .ppt *(MediaIntPtt).*

Ao aplicarmos os testes de normalidade foi verificado que apenas a variável *MediaIntForum* segue uma distribuição normal. A análise do modelo quanto a normalidade dos termos do erro é geralmente mais difícil do que outras pressuposições. Em primeiro lugar variações do acaso pode levar a conclusões erradas quando é estudado a natureza da distribuição de probabilidade, especialmente no caso de pequenas amostras. Vários testes estatísticos são encontrados na literatura para testar a normalidade de variáveis, como os testes de aderência de *Qui-quadrado*, *Kolmogorov-Smirnov* e *Lilifors*. Outros testes para normalidade são o *Shapiro-Wilks* e *Shapiro-Francia*. No nosso estudo utilizamos o teste de *Shapiro-Wilks*. A Tabela 2 mostra os resultados do teste.

Tabela 2. Testes de normalidade.

Variável	Resultado do teste *Shapiro-Wilks*
MediaIntForum	*p-value = 0.1886*
MediaIntVideo	*p-value = 0.0003031*

MediaIntMat	*p-value = 0.000419*
MediaIntPtt	*p-value = 0.01826*

Como podemos observar, apenas a variável *MediaIntForum* obteve um resultado satisfatório quanto ao teste de normalidade, desta forma nosso modelo de regressão se tornou um modelo de regressão linear simples, como tendo a variável dependente *Desemp_Final* e a variável *MediaIntForum* como sendo a variável independente. Abaixo temos o modelo de regressão simples que mostra a inclinação da reta de regressão ajustada na relação entre as duas variáveis.

```
> dados <- lm(Desemp_Final~MediaIntForum)
> dadosReg
  Call:
  lm(formula = Desemp_Final ~ MediaIntForum)
  Coefficients:
   (Intercept)  MediaIntForum
     13.110          3.626
```

O coeficiente chamado *Intercept* corresponde ao α da reta de regressão, ao passo que o coeficiente identificado como MediaIntForum corresponde ao β ou coeficiente angular.

4.4 Plotando o gráfico de dispersão

Gráficos de dispersão mostram as relações entre os valores numéricos em varias sequencias de dados ou plotam dois grupos de números como uma sequencia de coordenadas *XY*. Neste experimento plotamos as variáveis *Desemp_Final* contra *MediaIntForum*, através do gráfico podemos verificar visualmente que existe uma relação de linearidade entre as duas variáveis.

```
> plot <- (Desemp_Final~MediaIntForum)
```

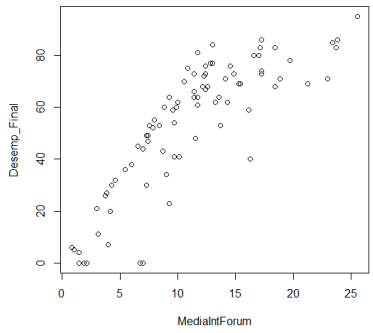

Figura 1: plotagem do gráfico de dispersão

No entanto o gráfico é apenas uma medida descritiva, o que não nos permite confirmar uma afirmação concreta, logo se utilizou medidas de inferência para comprovar os resultados explicitados pelo gráfico.

4.5 Teste de significância

A partir da decomposição da soma de quadrados total, definimos o coeficiente de determinação, ou explicação do modelo, que é dado por:

$$r^2 = \frac{SQreg}{SQtot} = \frac{bSxy}{Sxx}$$

Este valor de R^2 nos indica a proporção da variação de γi que é explicada pela regressão, no caso deste experimento, é o valor que indica a proporção do desempenho dos alunos, em função da quantidade de interações via fórum no ambiente de aprendizagem. A função abaixo gera o teste de significância, bem como nos mostra o valor do R^2 através do `(R-squared)`.

```
> summary(dadosReg)
      Call:
      lm(formula = Desemp_Final ~ MediaIntForum)
      Residuals:
      Min       1Q  Median      3Q     Max
      -38.491 -10.284   2.693    10.095  25.416
      Coefficients:
             Estimate Std. Error t value Pr(>|t|)
      (Intercept)    13.1097     3.3396   3.926 0.000179 ***
      MediaIntForum   3.6259     0.2653  13.668  < 2e-16 ***
      ---
      Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
      Residual standard error: 14.04 on 82 degrees of freedom
      Multiple R-squared:  0.795,     Adjusted R-squared:  0.7912
      F-statistic: 186.8 on 1 and 82 DF,  p-value: < 2.2e-16
```

O valor do coeficiente de determinação (`Adjusted R-squared`) encontrado no nosso caso correspondeu a `0.7912`. É uma medida que dá indícios sobre a qualidade do ajuste do modelo obtido.

4.6 Análise de resíduos e testes de diagnósticos

A análise de resíduos é de extrema importância para subsidiar a confiança no ajuste do modelo encontrado. Na Figura 2 temos quatro gráficos que nos possibilite a verificação das premissas do modelo.

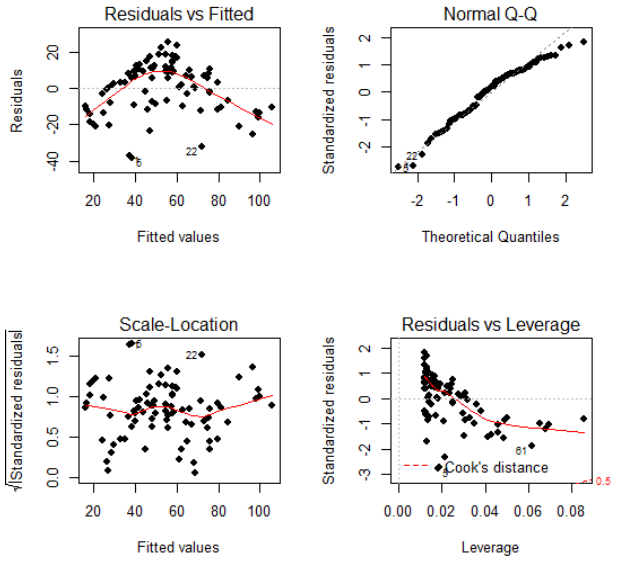

Figura 2. Gráficos de testes de diagnósticos

O gráfico do canto superior esquerdo (*Residuals vs Fitted*) corresponde aos resíduos comparados aos valores preditos pelo modelo. Neste experimento, não encontramos nenhuma tendência clara e a distribuição dos resíduos parece simétrica em torno da

linha. O segundo gráfico (normal Q-Q plot) mostrou o gráfico de quantis para os resíduos, no qual foi possível verificar a normalidade dos resíduos. Os pontos se mostraram distribuídos seguindo uma tendência de linearidade, dando indícios que os resíduos apresentam uma distribuição normal. O terceiro gráfico (*Scale-Location Plot*) mostra a distribuição da raiz do módulo dos resíduos em comparação aos valores preditos. Este gráfico permite a verificação da tendência de heteroscedasticidade em que a magnitude da variação dos resíduos está correlacionada com os valores preditos.

5. Discussão

Alguns pontos importantes a destacar na fase de análise descritiva foram os indícios de que em relação aos materiais didáticos, os alunos costumam acessar apenas uma vez, tendo em vista a possibilidade de baixar para si o material e não mais ser necessário participar de interação com esses artefatos. Ao contrário da ferramenta fórum, nela obteve-se uma média de 11 interações semanais por aluno.

Com relação ao valor médio de acesso ao ambiente, foi possível justificar que existe um pico de acesso nos horários próximos ao meio dia, devido a grande quantidade de alunos que trabalham e cursam graduação na modalidade a distância, sendo este um horário de disponibilidade para a realização das atividades e acesso ao ambiente.

Sobre a parte inferencial do experimento, podemos afirmar que o modelo linear simples explicou a relação entre dados de interação via fórum e o desempenho dos alunos. Este modelo linear foi considerado um bom modelo para explicar que existe uma relação entre a quantidade de interações via fórum de discussão e o desempenho dos alunos.

Podemos ainda afirmar, com uma confiança de 95%, que 79,12% dos alunos que interagem via fórum de discussão, a medida que aumenta a quantidade de interação, tende a aumentar o despenho acadêmico. Este resultado é satisfatório, se pensarmos que podemos aplicar modelos desta natureza para tentar prever a proporção de alunos que terão desempenhos satisfatórios em um curso, baseado nas suas interações iniciais.

6. Considerações finais e trabalhos futuros

Diante deste trabalho pode-se concluir que a análise de regressão é útil em áreas nas quais atividades de planejamento do futuro é fundamental. Esse tipo de condição ocorre em cursos à distância, nos quais a tomada de decisões durante o curso por parte de tutores, professores e gestores é um desafio.

Neste trabalho investigou-se a viabilidade da aplicação de técnicas de modelagem de regresão linear para realizar inferências. Os resultados obtidos com o experimento confirmam a viabilidade de se realizar inferências relativas ao desempenho de estudantes. Isso indica que a proporção da variação da variável desempenho, pode ser inferida em função da quantidade de interações via fórum no ambiente de aprendizagem. Desta forma podemos afirmar que é possível estimar o desempenho de alunos baseado na quantidade de interações em ferramentas do tipo fórum de discussão.

Como trabalhos futuros pretendemos aplicar uma modelagem envolvendo mais de uma variável independente e modelos não lineares, utilizando a análise de regressão linear generalizada, por ser uma tecnica mais robusta e levar em consideração diversas

variáveis que afetam direta ou indiretamente o desempenho de alunos em ambientes de cursos online.

7. Referências

Anderson, T. e Garrison, D. Learning in a networked world: New roles and responsibilities. In C. Gibson (Ed.), *Distance Learners in Higher Education*. p. 97-112. Madison, WI.: Atwood Publishing, 1998.

Baker, R.S.J.d. Data Mining for Education. McGaw, B., Peterson, P., Baker, E. (Eds.) International Encyclopedia of Education (3rd edition). Oxford, UK: Elsevier. 2010.

Baker, R.S.J.d., Isotani, S., de Carvalho, A. (2011) Mineração de Dados Educacionais: Oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, 19 (2), 3-13.

Curral, J. *Statistics Packages: A General Overview*. Universidade de Glasgow, 1994.

Gottardo, E., Kaestner, C., Noronha, R.V. Previsão de Desempenho de Estudantes em Cursos EAD Utilizando Mineração de Dados: uma Estratégia Baseada em Séries Temporais. In *Proceedings of Simpósio Brasileiro de Informática na Educação – SBIE 2012*, 26 a 30 de novembro de 2012. Rio de Janeiro, RJ.

Hoffmann, R. *Análise de Regressão: Uma Introdução a Econometria*. Editora Hucitec, 2006.

Hirumi, A. Analyzing and designing e-learning interactions. In Charles Juwah (Ed.), *Interactions in online education implications for theory e practice*, 46 – 71, New York: Routledge Publishing, 2006.

Johnson, G. M. e Johnson, J. A. Dimensions of Online Behavior: Implications for Engineering E-Learning. In: Magued, I.; Vikram, K.; Mohammed, K. (Org.), *Technological Developments in Education and Automation*, Springer, 2010.

Kear, K. Social presence in online learning communities. *Proceedings of the 7th International Conference on Networked Learning*, Aalborg, Denmark, 2010.

Moore, M. Three types of interaction. *American Journal of Distance Education*, 3(2), 1 – 6, 1989.

Picciano, A. G. "Beyond student perceptions: Issues of interaction, presence, and performance in an online course." *Journal of Asynchronous learning networks*, 21-40, 2002.

Russo, T. C. e Scott C. Perceptions of mediated presence in an asynchronous onlinecourse: Interplay of communication behaviors and medium. *Distance Education*, 25(2), 215-232, 2004.

Souza, S.G.; *Introdução aos Modelos de Regressão Linear e Não-Linear*. Embrapa Produção de Informação, 2001.

Watson, A. Developing teaching practice for more effective use of synchronous discussion: A preliminary investigation. In Hello! Where are you in the landscape of educational technology? *Proceedings ascilite*, Melbourne, 2008.