

## Análise comparativa de classificadores de predição em data mining para geração de informações

Igor Kühn<sup>1</sup>, Leandro Krug Wives<sup>2</sup>, Mariele de Almeida Lanes<sup>3</sup>, Francisco Dutra dos Santos Junior<sup>4</sup>

*Universidade Federal do Rio Grande do Sul (UFRGS), Programa de Pós-Graduação em Informática na Educação (PPGIE)<sup>1</sup>, Universidade Federal do Rio Grande do Sul (UFRGS)<sup>2</sup>, Programa de Pós-Graduação em Computação (PPGC)<sup>3</sup>, Universidade Federal do Rio Grande (FURG), Campus Santo Antônio da Patrulha<sup>4</sup>*

*e-mail: kuhnigor@gmail.com, Leandro.wives@ufrgs.br, infomariele@gmail.com, prof.chicosantosjr@gmail.com*

**Resumo.** Esse artigo buscou a partir da comparação de diferentes classificadores o melhor modelo de predição que gere informações que possibilitem a tomada de decisão quanto a uma trajetória de aprendizado mais adequada para alunos em Ambientes Virtuais de Ensino e Aprendizado (AVEA). Para isso foram utilizadas notas de 384 estudantes que tiveram como fonte a plataforma Moodle de um curso de pós-graduação a distância, oferecido pelo Centro Interdisciplinar de Tecnologias da Educação (CINTED/UFRGS) em parceria com a Universidade Aberta do Brasil. São também apontados artigos que se relacionam ao tema de pesquisa, bem como a metodologia utilizada. A partir dos resultados encontrados é possível verificar que, com o conjunto de dados selecionados, o classificador bayesiano apresentou melhor desempenho quando comparado com outros tipos de classificadores. As informações obtidas servem como subsídio para que professores e tutores possam indicar aos alunos trajetórias de aprendizagem em que o aluno possa estar utilizando o AVEA da melhor forma. A tarefa de predição realizada pode servir para que já nas primeiras avaliações possam ser realizados ajustes corretivos em relação ao percurso dos alunos com desempenhos não satisfatórios, evitando possíveis desistências, abandono de curso e evasões.

**Palavras Chave:** Classificadores de Predição, Ambientes Virtuais de Ensino e Aprendizado, Trajetórias de Aprendizagem.

### Introdução

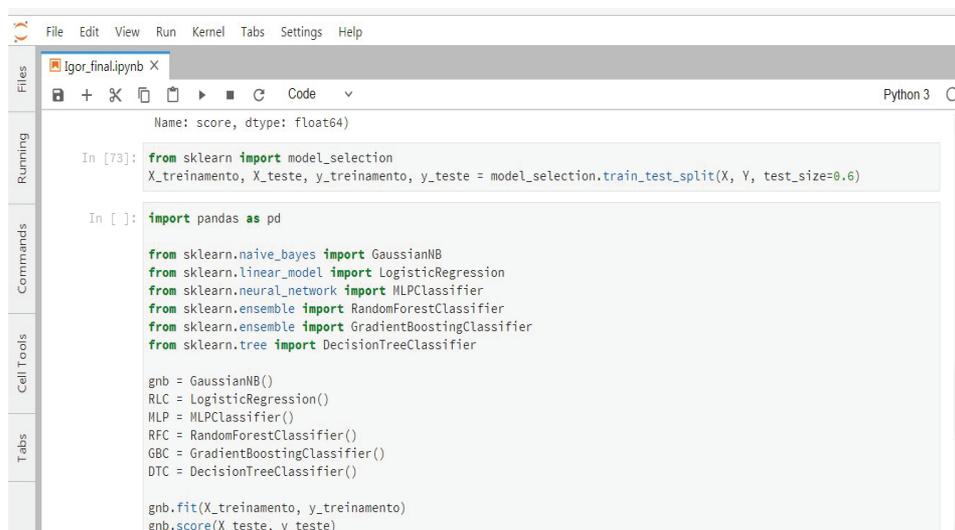
A sociedade passa por profundas e constantes transformações dentre as quais é possível destacar o desenvolvimento tecnológico. Dentre essas mudanças, uma das mais significativas envolve o processo de comunicação e de troca de informações. O processo de utilização de dispositivos, sejam eles móveis ou não, está modificando a capacidade de interação e das relações entre as pessoas. Segundo Castells (2002) “as redes interativas de computadores estão crescendo exponencialmente, criando novas formas e canais de comunicação, moldando a vida e, ao mesmo tempo, sendo moldadas por elas”.

Os processos de ensino e aprendizagem não estão à parte dessa mudança, também estão sendo modificados quando são introduzidos em ambientes escolares elementos que até algumas décadas não faziam parte deste cenário. O crescente uso de computadores, bem como a utilização de dispositivos móveis como ferramentas de acesso a Ambientes Virtuais de Ensino e Aprendizado (AVEAs), tem ocasionado um aumento significativo no volume de dados disponibilizados a partir desses ambientes. Um AVEA fornece uma diversidade de dados interacionais que, quando recebem um tratamento adequado através de um processo de mineração de dados, pode sinalizar comportamentos e padrões que em um primeiro momento podem passar despercebidos. No presente artigo o principal elemento motivador é a utilização de alguns desses dados para gerar informações relevantes que possam auxiliar professores e tutores a indicar uma trajetória de aprendizado mais adequada ao aluno.

Para esta descoberta de comportamentos e padrões, foi utilizado o processo de Descoberta de Conhecimento em Bancos de Dados, ou KDD, do inglês *Knowledge Discovery in Databases*, que segundo Fayad (1996) "é um processo não trivial de identificação de novos padrões válidos, úteis e compreensíveis.

Alguns exemplos desses dados gerados a partir das interações são: notas, número de acessos, número de exercícios respondidos, frequência de interação com o recurso educacional, interação via fóruns e interações via chat. Para se descobrir relações entre os dados, e como consequência informações que possam ser posteriormente analisadas, será utilizada Mineração de Dados Educacionais (do inglês, *Educational Data Mining*, ou MDE).

A MDE é definida por Baker (2011) como a área de pesquisa que tem como principal foco o desenvolvimento de métodos para explorar conjuntos de dados coletados em ambientes educacionais. Esse tipo de ferramenta proporciona ao seu usuário filtrar e selecionar dados que possam fornecer alguma informação sobre o aluno ou grupo de alunos. Nesse artigo, a ferramenta selecionada para este fim foi a plataforma Microsoft Azure Notebooks por apresentar algumas vantagens em relação a outras ferramentas, entre elas a grande quantidade de bibliotecas disponíveis e a versatilidade da linguagem (Python). Essa também possui a característica de ser multiplataforma, livre, de código aberto e utiliza a modalidade de computação em nuvem. Na Figura 01 abaixo é possível visualizar o uso da ferramenta na realização do treinamento e teste com o conjunto de dados selecionados, com o objetivo de comparação entre diferentes classificadores.



```
File Edit View Run Kernel Tabs Settings Help
igor_final.ipynb X Python 3
Name: score, dtype: float64
In [73]: from sklearn import model_selection
X_treinamento, X_teste, y_treinamento, y_teste = model_selection.train_test_split(X, Y, test_size=0.6)
In []: import pandas as pd
from sklearn.naive_bayes import GaussianNB
from sklearn.linear_model import LogisticRegression
from sklearn.neural_network import MLPClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.tree import DecisionTreeClassifier
gnb = GaussianNB()
RLC = LogisticRegression()
MLP = MLPClassifier()
RFC = RandomForestClassifier()
GBC = GradientBoostingClassifier()
DTC = DecisionTreeClassifier()
gnb.fit(X_treinamento, y_treinamento)
gnb.score(X_teste, y_teste)
```

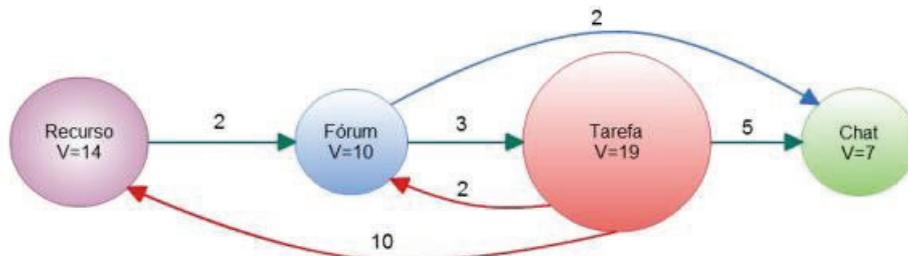
**Figura 01: Plataforma Microsoft Azure Notebooks Fonte: Adaptado pelo autor.**

O que se busca ao utilizar a ferramenta é realizar um estudo comparativo entre diferentes tipos de algoritmos classificadores. Os classificadores são algoritmos que podem ser utilizados para as tarefas automáticas de classificação de dados, podendo gerar informações para tomada de decisões, utilizando para isso informações pretéritas de desempenho de alunos em ambientes de aprendizagem. Alguns tipos de classificadores são: *Gaussian NB*, *Logistic Regression*, *MLP Classifier*, *Randon Forrest Classifier*, *Gradient Boosting Classifier*, *Decision Tree Classifier*.

O principal elemento a ser analisado entre os classificadores é o grau de precisão para predição, ou seja, o objetivo é selecionar o melhor algoritmo de classificação que possa, através de dados coletados da plataforma Moodle, prever qual o desempenho de um aluno em AVEA. Com o uso dessas informações é possível traçar uma trajetória educacional mais adequada a característica de aprendizado do aluno.

Por trajetórias de aprendizagem, nesse trabalho é utilizado o conceito de Confrey (2012) que afirma que aprendizagem é um caminho através de um corredor conceitual, onde existem obstáculos previsíveis e pontos de referência e, dessa forma, o caminho particular de um estudante nada mais é do que uma questão de probabilidades e probabilidades esperadas. As trajetórias de aprendizagem permitem especificar, em um nível adequado e mensurável de detalhes, quais conceitos os estudantes precisam

dominar durante o desenvolvimento e a evolução de um dado conceito ao longo do tempo, bem como quais tarefas, avaliações, fóruns e chats o aluno acessa. A Figura 02 a seguir exemplifica um modelo visual de trajetória de aprendizado com suas etapas.



**Figura 02: Modelo de Trajetória de Aprendizagem. Fonte: Ramos, D. B., Oliveira, E. H. T., Ramos, I. M. M. & Oliveira, K. M. T. (2015).**

Na seção seguinte são apresentados alguns trabalhos relacionados a temática abordada nesse trabalho, com a finalidade de contextualização do trabalho quanto a relevância desse tipo pesquisa tanto em âmbito nacional quanto internacionalmente.

## Trabalhos Relacionados

Nos últimos anos diversos trabalhos utilizaram as ferramentas de classificação para diferentes finalidades. Nesta seção, foram selecionados trabalhos que também utilizaram a técnica de MDE para tentar compreender melhor o comportamento dos alunos em um AVEA, ou seja, a utilização de ferramentas de classificação direcionadas para o ambiente virtual educacional.

No estudo de Mayilvaganan e Kalpanadevi (2014), foi realizada uma comparação entre classificadores com a função específica de predição de desempenho de alunos em seu ambiente de aprendizado. Para isso foram utilizados atributos como: presença em sala de aula, idade do aluno, horas disponibilizadas para estudo, acesso a recursos como biblioteca e internet, notas de testes, participação em seminários e trabalhos realizados em laboratórios. De posse desses dados foram utilizadas as seguintes técnicas de classificação: Árvore de Decisão, Redes Bayesianas e K-Vizinho mais Próximo. Este último apresentou um melhor desempenho para o conjunto de dados utilizados.

De forma semelhante, Amornsinalaphachai (2014) propôs a comparação de sete algoritmos de classificação para predição de desempenho de aprendizado na disciplina de programação. Foram comparados os seguintes classificadores: Redes Neurais, K-Vizinho mais Próximo, Redes Bayesianas, JRIP, ID3 e C4.5 obtendo um melhor desempenho o classificador C4.5 para o conjunto de dados selecionados, apresentando uma precisão de 74,5%, ou seja, o algoritmo conseguiu com quase setenta e cinco por cento de precisão determinar o desempenho dos alunos de forma individual. Já em Amrieh, Hamtini e Aljarah (2015) foram utilizados dados acadêmicos para através dos classificadores gerar modelos de predição baseados em características comportamentais dos alunos. Para isso foram utilizados os seguintes classificadores: Redes Neurais, Árvore de Decisão e Redes Bayesianas. Para esse estudo e com conjunto de dados selecionados, o classificador que utiliza Redes Neurais apresentou um desempenho acima dos demais. Com a utilização desse modelo de predição foi possível uma melhora de 25% a 29% no desempenho acadêmico dos alunos.

No estudo de Gottardo, Kaestner e Noronha (2012), foram utilizados dados gerados a partir de um AVEA, para através de técnicas de MDE obter informações que possam servir para melhorar o desempenho dos alunos. Buscou-se principalmente descobrir características de aprendizado dos alunos, sejam elas individuais ou coletivas. Para isso se fez uso de séries temporais, que são os dados gerados a partir de iterações ao longo de um período específico. Os autores utilizaram algoritmos de classificação

*Random Forrest* e *MultiplayerPerceptron* que apresentaram precisão semelhante quanto a capacidade de predição de desempenho em AVA, aproximadamente em 74% de precisão.

Weber, Zat e Lima (2013), diferentemente dos artigos anteriores que utilizaram modelos de classificação, fizeram uso da técnica de agrupamento (clusterização) para a partir de três fontes diferentes de dados tentar descobrir padrões e evidências em relação aos estudantes e suas diferentes formas de aprendizagem. Outra diferença em relação os artigos anteriores foi a separação dos alunos em três grupos distintos a partir de notas de pré-testes realizados, sendo essa uma característica da técnica utilizada.

A seleção dos artigos acima tem como intenção demonstrar a importância de técnicas de MDE para as grandes quantidades de dados geradas durante o processo de aprendizado em um AVEA. A semelhança entre o desempenho dos classificadores (aproximadamente 75% de precisão) indica a necessidade de uma busca por melhores ferramentas para que seja possível seu uso, para com um grau maior de precisão prever desempenhos dentro de um AVEA. Na seção seguinte é abordada a metodologia utilizada no presente trabalho.

## Metodologia

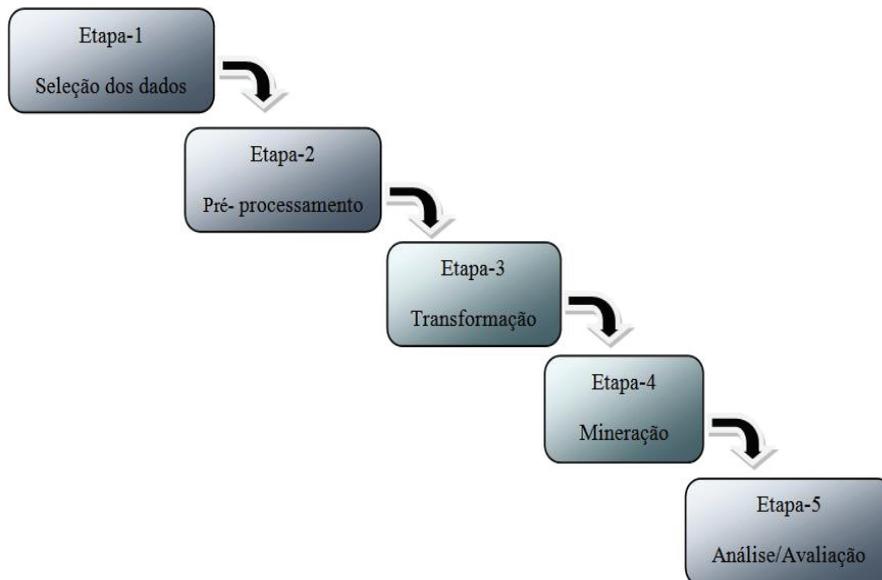
O conceito de metodologia, segundo Alves (2003, p.49), a considera um instrumento do pesquisador, uma vez que é através da especificação dos caminhos a serem adotados que se torna possível delimitar a criatividade e definir o **como, onde, com quem, com quê, quanto e de que maneira** se pretende captar a realidade e seus fenômenos.

O estudo aqui apresentado foi realizado por meio da coleta de dados do Curso de Pós-Graduação a Distância que possui como público alvo professores da educação básica, oferecido pelo Centro Interdisciplinar de Tecnologias da Educação (CINTED/UFRGS), fazendo parte do programa Universidade Aberta do Brasil (UAB). Foram coletados dados provenientes do Ambiente EAD Moodle, com extração através da emissão de relatório em planilha eletrônica, do período de 28 de Agosto a 22 de Dezembro de 2017, referente a notas de 384 alunos em 7 disciplinas (organizadas em 3 módulos), conforme ilustrado no Quadro 01.

Módulo	Disciplina
Módulo I	Treinamento Inicial com o Moodle
	Uso de email
Módulo II	Introdução ao Computador
	Operação de Computadores
Módulo III	Editores de Texto
	Editores de Apresentação
	Editores de Imagens

**Quadro 01: Módulos do Curso Fonte: CINTED/UFRGS.**

A metodologia deste artigo é constituída de cinco etapas (Figura 03), e é fundamentada no processo de mineração de dados educacionais (MDE). Esta seção será melhor detalhada nas subseções descritas a seguir.



**Figura 03: Processo de MDE. Fonte: Elaborado pelo autor.**

## Seleção de Dados

Na primeira etapa é realizada a seleção e coleta dos dados, tendo como principal cuidado garantir a qualidade da fonte do material, bem como garantir o sigilo dos dados pessoais fornecidos pelos alunos. Os dados utilizados nesse artigo (*dataset*) tem como origem a Plataforma Moodle do curso já referenciado anteriormente.

## Pré-processamento

Na segunda etapa são realizadas tarefas que tenham como objetivo a eliminação de dados repetidos e dados inconclusivos. Também é possível nesse momento de o processo recuperar dados incompletos. Devido ao grande número de dados utilizados, pode ser necessária a redução do tamanho da amostra com a finalidade de também diminuir o número de variáveis que compõe o algoritmo.

## Transformação

Na terceira etapa, após serem selecionados, limpos e pré-processados, os dados necessitam ser armazenados e formatados adequadamente para que os algoritmos possam ser aplicados. A transformação e a combinação também são utilizadas para obter “dados derivados” que são os dados faltantes.

## Mineração

Na quarta etapa do processo é realizada a mineração de dados (*data mining*), que segundo alguns

especialistas é a principal fase do processo, claro que é possível sempre fazer a ressalva que, se as etapas anteriores não forem realizadas com um rigor técnico elevado, a qualidade da mineração acaba sendo comprometida levando a resultados equivocados. A mineração de dados pode ser conceituada como segue:

*A mineração de dados é o processo de descoberta de informações acionáveis em grandes conjuntos de dados. A mineração de dados usa análise matemática para derivar padrões e tendências que existem nos dados. Normalmente, esses padrões não podem ser descobertos com a exploração de dados tradicional pelo fato de as relações serem muito complexas ou por haver muitos dados. (Microsoft Developer Network ,2006)*

Nesse artigo, na quarta etapa foram testados os classificadores citados anteriormente. Como já mencionado, o objetivo da comparação entre os classificadores é a partir de dados originários da plataforma Moodle, indicar uma trajetória de aprendizado.

## **Análise/Avaliação**

Como quinta e última etapa, foi realizada a análise e avaliação dos dados já transformados em informação. A informação que se busca nesse artigo é um processo no qual um conjunto de dados possa gerar informações que possam ser utilizados afim de indicar uma trajetória de aprendizado para o aluno ou grupo de alunos.

Ainda sobre esta etapa, é possível destacar que em alguns casos existe a necessidade de um especialista para que esse possa realizar a correta leitura e análise das informações obtidas. Na próxima seção são apresentados os resultados bem como a análise dos mesmos.

## **Resultados Encontrados**

Nessa seção são apresentados resultados obtidos na simulação realizada com a intenção de avaliar o desempenho de algoritmos de predição. Os dados foram coletados e armazenados em planilha formato Excel, posteriormente realizado o tratamento (pré-processamento) e os dados salvos em formato CSV (*Comma Separated Values*), que é o formato utilizado pela plataforma selecionada. Nesse experimento foram selecionados grupos de variáveis 'notas médias' que para a utilização no algoritmo de classificação se tornaram os atributos 'MOD1\_Media', 'MOD2\_Media', 'MOD3\_Media', os quais serão chamados atributos preditivos. Os atributos preditivos são chamados de precedentes, ou seja, a partir deles é que teremos os atributos consequentes. Nesse artigo estes atributos (antecedentes) são constituídos pelas notas médias dos módulos e o atributo Conceito (consequente). Ainda sobre a constituição dos grupos de variáveis (atributos preditivos), é resultado da média das notas individuais obtidas em cada módulo cursado pelos 384 alunos, sendo que cada módulo é constituído de duas ou três disciplinas. Os atributos selecionados constituirão a variável dependente X, como variável independente Y foi realizada a criação de um atributo chamado 'Conceito', onde foram criadas quatro faixas (intervalos) de notas a qual foram atribuídos conceitos A, B, C, D e E. O conceito A corresponde à nota mais alta (10) e E a nota mais baixa (0). Este último atributo foi utilizado como objetivo/alvo.

De acordo com Soares (2005), a tarefa de classificação tem como objetivo encontrar algum tipo de relacionamento entre os atributos preditivos e o atributo objetivo, de modo a obter um conhecimento que possa ser utilizado para prever a classe de um determinado registro que ainda não possui classe definida. A acurácia de um modelo em um determinado conjunto de dados de treinamento é a porcentagem de amostras de teste que são corretamente classificadas pelo modelo.

Dos classificadores anteriormente mencionados na seção introdutória deste artigo, quatro foram utilizados, sendo eles *Gaussian NB* (Modelo Bayesiano), *MLP Classifier* (Modelo Neural), *Random Forrest Classifier* (Modelo Random Forest) e *Decision Tree Classifier* (Modelo de Árvore de Decisão).

Os principais critérios utilizados para avaliar a precisão de um modelo são (HAN *et al.*, 2001): precisão da predição, velocidade, robustez, escalabilidade, interpretabilidade. Nesse artigo, que possui um caráter exploratório dessa base de dados, o critério de ordenamento foi a precisão da predição.

Para o conjunto específico de dados selecionados e com o uso da plataforma Microsoft Azure Notebooks, foi possível verificar (Quadro 02) um melhor desempenho do classificador Bayesiano, mesmo variando o número de atributos. Cabe ainda destacar a pouca diferença entre o desempenho dos classificadores, uma possível explicação para essa proximidade seja o tamanho da amostra utilizada (384 alunos).

Classificador	Precisão %		
	1 Atributo	2 Atributos	3 Atributos
Bayesiano	<b>63.6</b>	<b>84.8</b>	<b>91.0</b>
Random Forest	61.9	82.2	88.7
Árvore de Decisão	63.2	81.4	85.2
Modelo Neural	61.9	73.2	72.7

**Quadro 02. Atributos. Fonte: Elaborado pelo autor**

O modelo apresenta um alto grau de precisão quando comparado com o desempenho de modelos e conjunto de dados utilizados em trabalhos relacionados. Quanto a diminuição do número de dados, reduzindo de 3 até 1 atributo, a intenção foi testar a capacidade preditiva dos modelos, que apresentaram um desempenho satisfatório mesmo para um volume reduzido de notas. A análise que pode ser feita é que, de acordo com resultado obtido pelo modelo com melhor desempenho, a partir de um conjunto de notas obtidas durante os primeiros períodos do aluno no curso, é possível prever com mais de 90% (3 atributos) e 63.63 % (com 1 atributo) de precisão qual conceito final que o aluno ou grupo de alunos atingirá.

## Considerações Finais

O uso da tecnologia no âmbito do ensino e aprendizagem tem se tornado um elemento fundamental para a melhoria da qualidade da educação. O uso de plataformas virtuais de aprendizagem como, por exemplo, o Moodle proporciona um número elevado de dados interacionais gerados nesses ambientes, oriundos das interações dos alunos dentro desses espaços. O presente artigo teve como objetivo apresentar a utilização alguns desses dados conjuntamente com uma ferramenta de mineração de dados para gerar informações relevantes que contribuam para um melhor desempenho do aluno. Essas informações servem como subsídio para que professores e tutores possam indicar para aos alunos trajetórias de aprendizagem em que o aluno possa estar utilizando o AVEA da melhor forma.

Partindo ainda da hipótese que o aluno que atingiu um bom conceito final possui uma trajetória adequada e deva-se manter a proposta inicial de trajetória e contrariamente, o aluno que não obteve um bom desempenho, a sua trajetória deva ser ajustada de forma mais adequada. Para isso é interessante professores ou tutores disponibilizar tarefas/atividades específicas a cada aluno ou grupo de alunos, bem como uma melhor sequência a ser seguida. Cabe ainda com um exemplo demonstrar como a inserção de uma tarefa extra pode significar uma mudança de trajetória. A simples realização de uma avaliação extra

ou um objeto de aprendizagem<sup>18</sup> com características diferentes para um aluno pode caracterizar uma modificação na trajetória de aprendizagem proposta no início do curso, pois a trajetória passa a possuir uma característica individual ou uma característica específica de um grupo limitado de alunos.

A mineração de dados de forma mais específica a ferramenta de predição, possibilita diagnosticar em tempo hábil alguma dificuldade apresentada pelo aluno podendo essa ser dirimida pelos professores ou tutores. A ferramenta apresentada possui como uma de suas características uma maior acurácia quando utilizado em um grande conjunto de dados. No presente trabalho foram utilizadas notas de 384 alunos, em simulações realizadas anteriormente a mesma ferramenta apresentou um desempenho mais elevado com o uso de um conjunto de dados de mais de mil alunos.

Cabe ainda salientar o caráter introdutório do artigo quanto ao uso dos dados. Foram considerados apenas os atributos notas para a realização da simulação ficando como proposta de novos trabalhos o incremento de atributos que possuam um caráter significativo quanto a tarefa de predição para indicação de trajetória em um AVEA.

## Referências

AMORNSINLAPHACHAI, Pensri. **Efficiency of data mining models to predict academic performance and a cooperative learning model**. 8<sup>th</sup> International Conference of Knowledge and Smart Technology. 2016.

BASSANI, P. S. **Mapeamento das interações em ambiente virtual de aprendizagem: uma possibilidade para avaliação em educação a distância**. Tese de Doutorado. Porto Alegre: PPGIE/UFRGS, 2006.

BAKER, R. S. J.; ISOTANI, S.; CARVALHO, A. M. J. B. **Mineração de Dados Educacionais: Oportunidades para o Brasil**. Revista Brasileira de Informática na Educação. v.19, n.2, 2011.

CAMILO, C. O.; SILVA, J. C. **Mineração de Dados: conceitos, tarefas, métodos e ferramentas**. Relatório Técnico. Instituto de Informática – UFG, 2009.

CARINE, G. Webber. Daline Zat. Maria de Fátima Webber do Prado Lima. **UTILIZAÇÃO DE ALGORITMOS DE AGRUPAMENTO NA MINERAÇÃO DE DADOS EDUCACIONAIS**. Revista Novas Tecnologias na Educação. V. 11 N° 1, julho, 2013.

CASTELLS, M. **A sociedade em rede**. Rio de Janeiro: Bertrand Brasil, 2002.

CASTILLO G., DESCALÇO L., Diogo S., Millán E., Oliveira P., Anjo B. **Computerized Evaluation and Diagnosis of Student's Knowledge Based on Bayesian Networks**. In: Wolpers M., Kirschner P.A., Scheffel M., Lindstaedt S., Dimitrova V.(eds) Sustaining TEL: From Innovation to Learning and Practice. EC-TEL 2010. Lecture Notes in Computer Science, vol 6383. Springer, Berlin, Heidelberg, 2010,

CHEN, G.; LIU, C.; OU, K.; LIU, B. B. **Discovering Decision Knowledge from Web Log Portfolio for Managing Classroom Processes by Applying Decision Tree and Data Cube Technology**. Journal of Educational Computing Research, 23(3), pp.305–332,2000.

CONFREY, J. **Articulating a Learning Sciences Foundation for Learning Trajectories in the CCSS-M**. PMENA, 2012.

---

<sup>18</sup> Consideramos aqui Objetos de Aprendizado como um recurso educacional digital que tem como característica o reuso e o seu objetivo principal é o apoio a aprendizagem.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. **Advances in knowledge discovery and data mining**. MIT PRESS. 1996.

GOTTARDO, Ernani. **Estimativa de desempenho acadêmico de estudantes em um AVA utilizando técnicas de mineração de dados**. 84 f. Dissertação – Programa de Pós-Graduação em Computação Aplicada, Universidade Tecnológica Federal do Paraná. Curitiba, 2012.

HAN, Jiawei; KAMBER, Micheline. **Data Mining Concepts and Techniques**. San Francisco, EUA: Morgan Kaufmann, 2001. 550 p.

KAMPFF, A. J. C. **Mineração de dados educacionais para geração de alertas em ambientes virtuais de aprendizagem como apoio à prática docente**. Tese de Doutorado. Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Informática na Educação, 2009.

KING, R. D., STERNBERG, M. J. E., SRINIVASAN, A. **Probabilistic reasoning in intelligent systems-networks of plausible inference**. Morgan Kaufmann, pp 1– 552, 1989.

Microsoft Developer Network(2006) <[https://msdn.microsoft.com/pt-br/library/ms174949\(v=sql.120\).aspx](https://msdn.microsoft.com/pt-br/library/ms174949(v=sql.120).aspx)>. Acesso em: 20 Maio 2018.

MILLÁN, E., PÉREZ-DE-LA CRUZ, J.-L., GARCIA, F. **Dynamic versus Static Student Models Based on Bayesian Networks: An Empirical Study**. In: Palade, V., Howlett, R.J., Jain, L. (eds.) KES 2003. LNCS, v. 2774, pp. 1337–1344. Springer, Heidelberg, 2003.

MITCHELL, T. **Machine Learning**. Boston: McGraw Hill, 1997.

M. Mayilvaganan. D. Kalpanadevi. **Comparison of Classification Techniques for predicting the performance of Students Academic Environment**. International Conference on Communication and Network Technologies (ICCNT).2014.

NETTO, O. P. **Um filtro interativo de utilizando árvores de decisão**. Dissertação de Mestrado. Programa Interunidades de Bioinformática. Universidade de São Paulo, 2013.

Nguyen Thai Nghe, Paul Janecek, and Peter Haddawy. **A COMPARATIVE ANALYSES OF TECHNIQUES FOR PREDICTING ACADEMIC PERFORMANCE**. Computer Science and Information Management Program, Asian Institute of Technology (AIT), Thailand 12120.

PEARL, J. **Bayesian networks, causal inference and knowledge discovery**. TechnicalReport R-281, UCLA Cognitive Systems Laboratory, 2001.

RAMOS, D. B., Oliveira, E. H. T., Ramos, I. M. M. & Oliveira, K. M. T. (2015) **Trilhas de Aprendizagem em Ambientes Virtuais de Ensino-aprendizagem: Uma Revisão Sistemática da Literatura**. In: XXVI Simpósio Brasileiro de Informática na Educação, Maceió. RIGO, S. J.; CAMBRUZZI, W.; BARBOSA, J. L. V.; CAZELLA, S.C. **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.1, 2014.

Soares de França, Roselma. José Costa Amaral, Haroldo. **UTILIZAÇÃO DE ALGORITMOS DE AGRUPAMENTO NA MINERAÇÃO DE DADOS EDUCACIONAIS**. Revista Novas Tecnologias na Educação, V. 11 Nº 1,P n...n, julho, 2013

SOARES, Silviane L. **Aplicação de técnicas de Mineração de Dados na Gestão de Sistemas de Energia Elétrica**. 2005. 106 f. Dissertação (Mestrado em Engenharia Elétrica) – Programa de Pós-Graduação em Engenharia Elétrica, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2005.

YIN, R. K. **Estudo de Caso: planejamento e métodos**. 4ª Edição. Porto Alegre: Bookman, 2010.