



1999 - 2019

# Challenges 2019

Desafios da Inteligência Artificial  
Artificial Intelligence Challenges

António José Osório  
Maria João Gomes  
António Luís Valente

Livro de atas  
XI Conferência Internacional de TIC na Educação - Challenges 2019  
13, 14 e 15 de maio, Braga, Universidade do Minho

Universidade do Minho. Centro de Competência  
Braga, Portugal

# Challenges 2019: Desafios da Inteligência Artificial, Artificial Intelligence Challenges

(Atas da XI Conferência Internacional de Tecnologias de Informação e Comunicação na Educação – Challenges 2019, realizada em Braga de 13 a 15 de maio de 2019)

## ORGANIZADORES

António José Osório  
Maria João Gomes  
António Luís Valente

## PRODUÇÃO

Centro de Competência em Tecnologias de Informação e Comunicação na Educação  
(CCTIC-IEUM)

*Font typeface:* Open Sans

U: [www.nonio.uminho.pt](http://www.nonio.uminho.pt)  
E: [centrodecompetencia@ie.uminho.pt](mailto:centrodecompetencia@ie.uminho.pt)

## ISBN

978-989-97374-8-8

## EDIÇÃO

Universidade do Minho. Centro de Competência  
Campus de Gualtar  
4710-057 Braga, Portugal  
1.ª edição  
Maio, 2019



Este trabalho está publicado com uma licença Creative Commons Attribution-ShareAlike 4.0 International.

# Análise de desempenho de estudantes utilizando técnicas de agrupamento com mineração aplicado a curso de pós-graduação à distância

Igor Kühn, kuhnigor@gmail.com  
Universidade Federal do Rio Grande do Sul

Paulo Santana Rocha, rochap01@gmail.com  
PPGIE – Universidade Federal do Rio Grande do Sul

José Valdeni de Lima, valdeni@inf.ufrgs.br  
PPGIE – Universidade Federal do Rio Grande do Sul

Leandro Krug Wives, leandro.wives@ufrgs.br  
PPGIE – Universidade Federal do Rio Grande do Sul

Raquel Salcedo Gomes, raquel.salcedo@ufrgs.br  
PPGIE – Universidade Federal do Rio Grande do Sul

Rafaela Jardim, rafa.rjardim@gmail.com  
PPGIE – Universidade Federal do Rio Grande do Sul

**Resumo:** Este artigo apresenta um estudo do uso de técnicas de mineração de dados para avaliação de desempenho de 385 estudantes 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. Para tanto, foi utilizada a ferramenta *WEKA*, identificando *clusters* específicos de alunos de acordo com seu desempenho. Esta análise aponta para uma reflexão sobre a importância do acompanhamento da evolução dos alunos, indicando, no cenário estudado, um padrão de comportamento de notas obtidas entre módulos.

**Palavras-chave:** mineração de dados; agrupamento; desempenho; análise de aprendizado; educação

**Abstract:** This paper presents a study of the use of data mining techniques for the evaluation of grades of 385 students from a distance-learning postgraduate course offered by the Interdisciplinary Center for New Technologies in Education (CINTED / UFRGS) and Universidade Aberta do Brasil. In this context, WEKA tool was used, identifying specific clusters of students according to their performance. This analysis points to a reflection on the importance of monitoring student evolution, indicating, in the scenario studied, a behavioral pattern of grades obtained among the study units.

**Keywords:** data mining; clustering; performance; learning analytics; education

## Introdução

Com a popularização das Tecnologias da Informação e da Comunicação (TIC), surgiram os cursos a distância por meio da *Web*, em diversos níveis de ensino, favorecidos por uma infraestrutura com custos decrescentes e potencial de crescimento, quanto à qualidade e à abrangência geográfica, atendendo a diversos segmentos sociais e regiões, antes não alcançados. Por meio da convergência das mídias utilizadas até então, ampliou-se a possibilidade de interação entre os agentes envolvidos no processo de ensino-aprendizagem. Com as novas ferramentas da Internet, um número maior de indivíduos pode compartilhar suas aprendizagens em tempo real, mantendo interações com o professor ou com outro estudante, independentemente do local onde esteja (Unesco, 2002).

Nesse cenário, a educação a distância (EAD), conforme aponta Giolo (2008), é um forte instrumento de apoio à formação de professores, sobretudo por ampliar a oferta e massificar o acesso aos cursos de aprimoramento docente.

Por outro lado, os índices de desempenho nos cursos EAD são objeto de várias pesquisas, conforme explicita Fornari (2017), que aponta o alto grau de reprovação e o baixo desempenho nas disciplinas como fatores de evasão em cursos à distância.

Dessa forma, o acompanhamento e a análise do desempenho dos alunos são estratégias cada vez mais utilizadas por instituições de ensino na busca pela garantia de bons níveis de aprendizagem, conforme observa-se nos trabalhos de (Detoni, Cechinel, & Araújo, 2015), (Ramos, 2017) e (Queiroga, 2018).

Nesse sentido, o uso de técnicas de mineração de dados pode ser forte aliada nos processos de tomada de decisão visando a qualificação da aprendizagem, uma vez que a mineração de dados é capaz de gerar informações e métricas baseadas em algoritmos e métodos estatísticos que permitem identificar possíveis fragilidades no cenário apresentado. O uso da mineração de dados em cursos de educação à distância é fonte de trabalhos na literatura, com estudos das mais variadas naturezas, como análise de evasão escolar (Portal, 2016), colaboração entre estudantes (Vasconcelos & Neto, 2018) e desempenho acadêmico (Rodrigues & Siqueira, 2015).

Considerando importância da modalidade EAD e o potencial da mineração de dados para contextos educacionais, apresenta-se aqui uma avaliação de dados coletados de alunos do curso de Pós-Graduação a distância, oferecido pelo Centro Interdisciplinar de Tecnologias da Educação (CINTED/UFRGS), no âmbito da Universidade Aberta do Brasil (UAB).

A análise realizada leva em conta notas dos estudantes em 3 módulos do curso (distribuídos por 7 disciplinas). Os resultados indicam um padrão de comportamento dos alunos no decorrer dos módulos, em que indivíduos com

baixas notas no primeiro módulo acabam apresentando um menor rendimento no restante do curso.

A partir da identificação deste padrão, discute-se a aplicação do processo de mineração em um ambiente virtual de aprendizagem, analisando-se as relações entre as diferentes notas dos alunos obtidas nas disciplinas do curso.

## **Trabalhos Relacionados**

Nesta seção, a fim de contextualizar brevemente a área de pesquisa a discutir alguns de seus processos e contextos, são apresentados trabalhos em que foram utilizadas ferramentas de mineração em dados educacionais.

O artigo de Rodrigues & Ramos (2016) buscou identificar padrões de engajamento de alunos participantes de um Curso Online Aberto e Massivo (MOOC) em Língua Portuguesa. A análise de clusters, que é um processo que busca descobrir as relações entre os elementos com características semelhantes em um conjunto de dados. Nesse experimento foi usada especificamente para o estudo dos diferentes agrupamentos de alunos, identificando padrões de engajamento de alunos em cursos naquela modalidade. Os dados utilizados na análise são publicações que os participantes do curso realizaram durante as atividades on-line. Todos os envios foram extraídos do ambiente de aprendizagem OPENEDU Openredu (2019), que é um ambiente de aprendizagem *online* concebido para proporcionar formas intensas de colaboração com o uso de mídias dispostas em suas páginas. Durante a realização do experimento, foram utilizadas duas formas de análises de agrupamento: a hierárquica e a não hierárquica. A análise de agrupamento hierárquico teve como objetivo encontrar, através da geração de um dendrograma, uma quantidade  $k$  de grupos, até então desconhecidos. Um dendrograma é um diagrama de árvore que apresenta os grupos formados por agrupamento de observações a cada passo e em seus níveis de similaridade. No referido trabalho foi utilizado o algoritmo *Kmeans* na aplicação da clusterização não-hierárquica.

As técnicas de agrupamento, são técnicas capazes de identificar conjuntos de dados similares, que podem corresponder a um padrão ou comportamento típico observado nos dados. A análise de agrupamento, possibilitou identificar grupos com características comportamentais distintas que muitas vezes não são percebidas por professores e gestores de MOOCs.

Já no trabalho de Webber & Lima (2013) são utilizadas ferramentas de mineração aplicadas a três conjuntos de dados públicos: Geometry composto por dados retirados de um ano de curso de geometria, Chinese Tone Study que possui informações sobre os estudantes que estão aprendendo uma segunda língua, Álgebra I 2006 que armazena uma quantidade relativa de números e equações matemáticas. Para realizar a análise dos dados foram selecionados os algoritmos de agrupamento denominados *K-média*, que forma grupos visando minimizar a

distância entre os elementos do grupo em relação ao centro, EM que possui o objetivo de encontrar o melhor ajuste de um modelo para um conjunto de dados através da estimativa da máxima verossimilhança, imunológico Essa ferramenta (Machado, 2011) utiliza conceitos com origem nos Sistemas Imunológicos Artificiais, seu funcionamento está baseado na formação de grupos levando em consideração uma maior homogeneidade ente os elementos do mesmo grupo e uma maior heterogeneidade entre elementos de grupos distintos. E por fim o algoritmo hierárquico que agrupa os elementos em uma estrutura de árvore, organizando os grupos em formato hierárquico, resultando assim uma sequência aninhada de subdivisões.

Nesse sentido, para realizar a análise do conjunto de dados Geometry foram selecionados 12 atributos considerados mais relevantes. O estudo de caso para o conjunto de dados da língua Chinesa iniciou com a seleção de 6 atributos considerados mais relevantes. Já para o conjunto de dados Álgebra, foram considerados 3 atributos relevantes. Os algoritmos foram testados nas ferramentas WEKA e R. O trabalho avaliou os grupos formados pelos algoritmos através dos índices de homogeneidade e separação. Em uma análise preliminar foi possível concluir que a ferramenta imunológica alcançou os melhores índices de homogeneidade e separação.

O estudo de Kampff (2009) teve o objetivo geral de propor uma arquitetura para sistemas de alertas em Ambientes Virtuais de Aprendizagem (AVA), baseada em informações extraídas por processos de Mineração de Dados (MD), buscando identificar estudantes com características e comportamentos que podem levar à evasão ou à reprovação. Foram considerados dados de 1564 alunos de edições de cursos anteriores para extração das regras por mineração. As regras de classificação foram aplicadas para gerar alertas durante o acompanhamento de 230 alunos. A mineração foi realizada no ambiente RapidMiner com o uso do algoritmo RuleLearner, com cerca de 230 atributos selecionados. O trabalho desenvolvido na tese buscou definir uma arquitetura para sistemas de alertas em AVA, alertas que devam dar suporte á atuação pedagógica do professor no acompanhamento das situações de aprendizagem, com o monitoramento de perfis de alunos e eventos no ambiente, notificando o professor sobre grupos de alunos que compartilham necessidades específicas.

## **Metodologia**

O estudo aqui apresentado foi realizado através da coleta de dados do Curso de Pós-Graduação a Distância, 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 385 alunos em 7 disciplinas (organizadas em 3 módulos), conforme ilustrado na Tabela 1.

Tabela 1. Lista de disciplina analisadas.

<b>Módulo</b>	<b>Disciplina</b>
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

Fonte: CINTED/UFRGS.

Os alunos foram submetidos a avaliações em cada disciplina, além de uma prova cumulativa para medir o desempenho ao final de cada módulo, momento onde todas as disciplinas daquele módulo eram levadas em conta. Dessa forma, os dados coletados apresenta 10 diferentes notas para cada aluno, sendo 7 referentes a disciplinas e 3 correspondentes a prova final de cada módulo.

Foram excluídas da pesquisa os dados de alunos desistentes, cujo as notas eram parciais, sem conclusão de todos os módulos. Com tal refinamento, passou para 336 o número total de alunos.

Para o processo de Mineração de Dados (MD), foi definida como a primeira etapa a realização de um pré-processamento, onde ao analisar os 35 atributos existentes em planilha eletrônica optou-se por manter 10 atributos (todos do tipo numérico), por entender que esses eram mais interessantes para os objetivos do processo de mineração. Ainda nesta etapa, foi realizada uma limpeza para excluir dados inconsistentes e caracteres especiais para facilitar a leitura em software de mineração.

Em seguida os dados foram convertidos para arquivo com o formato CSV para serem carregados na ferramenta WEKA (*Waikato Environment for Knowledge Analysis*) Weka (2019) versão 3.9.2 que é uma ferramenta aberta e gratuita. Uma vez no WEKA, os dados foram tratados e submetidos a regras de Associação e Clusterização, com utilização dos algoritmos *Apriori* e *SimpleKMeans*, respetivamente.

O *Apriori* é um dos algoritmos mais conhecidos para mineração por regras de associação, empregando busca em profundidade e gerando conjuntos de itens candidatos (padrões) de  $K$  elementos a partir de conjuntos de itens de  $K-1$  elementos, onde os padrões não frequentes são eliminados. Toda a base de dados é rastreada e os conjuntos de itens frequentes são obtidos a partir dos conjuntos de itens candidatos.

De acordo com Silva (2016) o algoritmo *SimpleKMeans* por sua vez tem como objetivo encontrar partições no conjunto de dados de forma que  $k$  grupos separados de exemplares sejam descobertos, sendo que  $k$  é um parâmetro de entrada para o algoritmo. Na próxima seção são apresentados os resultados obtidos neste experimento.

## Resultados

A identificação de grupos ou clustering trata-se de um modelo de mineração de dados que procura encontrar exemplares ou objetos de um conjunto de dados com atributos semelhantes. Portanto, a divisão da base em grupos é realizada a partir de medidas de similaridade. Esta técnica é útil de ser aplicada em problemas que envolvem agrupamento não se encontra disponível o atributo especial classe ou rótulo (Han & Kamber, 2001).

Ainda na etapa de pré-processamento optou-se por discretizar as notas de alunos em categorias. Tal distribuição possibilita uma visão macro do desempenho dos estudantes além de permitir sua execução com algoritmos de clusterização, conforme é possível observar na Tabela 2.

Tabela 2. Distribuição de notas por categorias

<b>Categoria</b>	<b>Descrição</b>
0	Notas de 0 até 6 (nota < 6)
1	Notas de 6 a 8 (nota $\geq 6$ E nota < 8)
2	Notas de 8 a 10 (nota $\geq 8$ E $\leq 10$ )

Fonte: elaborada pelos autores.

Dessa forma, ainda na Tabela 2, alunos com notas na categoria 0 (zero) implica em um desempenho baixo, já alunos com conceito na categoria 1 (um) corresponde a um desempenho mediano e notas na categoria 2 são consideradas, neste estudo, como sendo de alto desempenho. Foram considerados alunos com notas de 0 a 6, como reprovados e de 6 a 10 como aprovados.



Nesse contexto, de posse dos dados categorizados, utilizou-se inicialmente das técnicas de Clusterização do WEKA para identificar a quantidade de clusters adequados para avaliar a distribuição de notas por categorias de desempenho de estudantes e posteriormente, a análise de associação.

### Resultado da utilização das técnicas de agrupamento

A análise de agrupamento ou clusterização é um processo que busca descobrir as relações entre os elementos de um conjunto de dados com características semelhantes. No presente artigo, o elemento que se busca é a similaridade entre as notas dos alunos. O grau de similaridade entre os elementos, conforme ilustra a Figura 1, onde no eixo horizontal é possível verificar o número total de alunos e no eixo vertical o desempenho obtido durante os módulos. O que buscou foi a observação através do nível de proximidade entre eles, ou seja, o que se busca é minimização das medidas de distância intragrupos (pontos verdes, vermelhos e azuis).

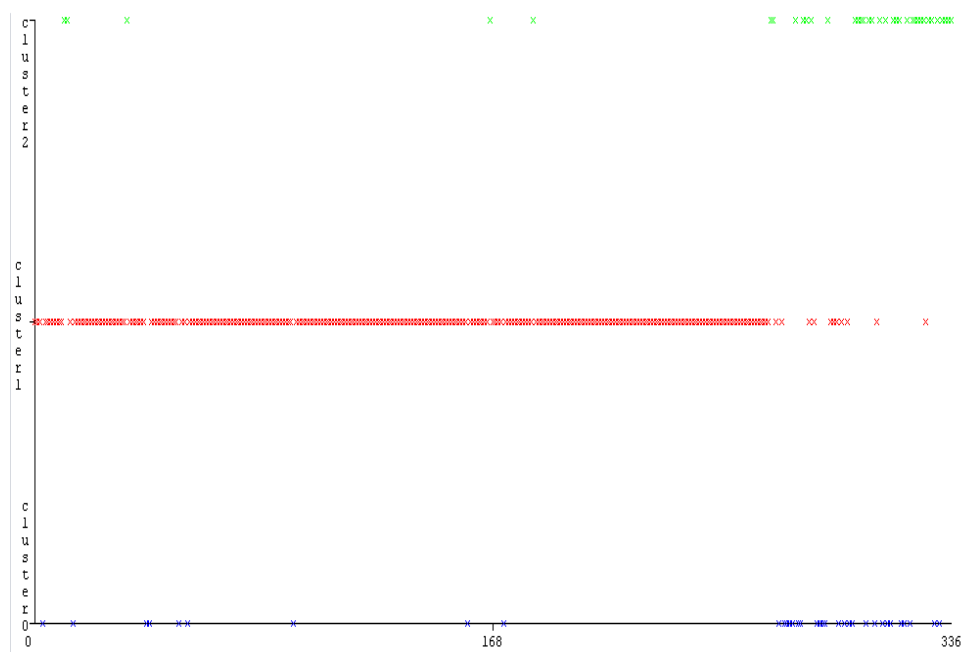


Figura 1: Visualização dos Clusters

A escolha de 3 (três) clusters, conforme ilustrado na Figura 2, foi adotada por considerar que em divisões superiores a esse número, as características passam a não ser tão claras, ou seja, os clusters não mais representam corretamente o grupo de alunos agrupados pelo critério notas.

```

Final cluster centroids:

Attribute      Full Data      Cluster#
                (337.0)        0          1          2
                (34.0)        (267.0)    (36.0)
-----
M1_TOTAL      2              2          2          0
M2_D1         2              1          2          0
M2_D2         2              2          2          0
M3_D1         2              0          2          0
M3_D2         2              0          2          0
M3_D3         2              0          2          0

Time taken to build model (full training data) : 0.04 seconds

=== Model and evaluation on training set ===

Clustered Instances

0          34 ( 10%)
1          267 ( 79%)
2          36 ( 11%)

```

Figura 2: Agrupamento dos Clusters.

É possível verificar uma concentração das notas dos alunos (79%) no cluster número 1 com uma distribuição equilibrada entre os clusters 0 (10%) e 2 (11%). Sinalizando uma distribuição normal em torno da média. Este resultado serve como um indicativo para as análises realizadas na seção seguinte.

### Resultados da utilização das técnicas de associação

A tarefa de associação segundo Schneider (2002) tem o intuito de identificar associações entre registros de dados que, de alguma maneira, estão ou devem estar relacionados. Sua premissa básica é encontrar elementos que implicam na presença de outros em uma mesma transação.

Podemos ainda afirmar que as regras de associação são usadas para descobrir elementos que ocorrem em comum dentro de um determinado conjunto de dados. Foram verificados elementos como o *lift*, que pode ser compreendido como uma regra de associação entre  $A \text{ } P \text{ } B$  indica o quanto mais frequente torna-se B, quando A ocorre. É medido por  $Lift(A \text{ } P \text{ } B) = Conf(A \text{ } P \text{ } B) \div Sup(B)$ .

Outro elemento testado foi o *leverage* que mede a diferença entre a probabilidade conjunta observada e esperada de XY, assumindo que X e Y são independentes. Se dá pela fórmula:

$$leverage(X \rightarrow Y) = P(XY) - P(X) \cdot P(Y) = r_{sup}(XY) - r_{sup}(X) \cdot r_{sup}(Y)$$

Por fim foi medida a *conviction*, que é a razão da frequência esperada que X ocorrer sem Y (a frequência que a regra faz uma predição incorreta) se X e Y forem divididos pela frequência das predições incorretas.

Dessa forma, após submeter os dados ao algoritmo de associação (*Apriori*) foi possível identificar a relação entre as notas obtidas nos 3 módulos do curso, em que se observa relação entre as disciplinas do módulo 1 (M1) e do módulo 2 (M2), assim como relação das disciplinas do módulo 1 (M1) com o módulo 3 (M3), conforme é ilustrado na Tabela 3.

Tabela 3: Relação entre módulos.

<b>Relação entre módulo 1 e disciplinas do módulo 2</b>
<b>Relação entre módulo 1 e disciplinas do módulo 3</b>
m1_total=2 m2_d1=2 216 ==> m2_d2=2 196 conf:(0.91) < lift:(1.22)> lev:(0.1) [35] conv:(2.62)
m1_total=2 m2_d2=2 219 ==> m2_d1=2 196 conf:(0.89) < lift:(1.21)> lev:(0.1) [34] conv:(2.38)
m1_total=2 m3_d2=2 227 ==> m3_d1=2 215 conf:(0.95) < lift:(1.22)> lev:(0.11) [38] conv:(3.89)
m1_total=2 m3_d1=2 220 ==> m3_d2=2 215 conf:(0.98) < lift:(1.22)> lev:(0.11) [38] conv:(7.18)

Fonte: elaborada pelos autores.

Na relação entre M1 e M2, o grau de confiança é respectivamente 0.91 e 0.89, indicando o percentual de ocorrência da mesma, com alto índice de confiança (em média 90%). Neste sentido, é possível confirmar, que os alunos que possuem bom desempenho no módulo 1 tendem a ter boas notas nas disciplinas no módulo 2 do curso.

Da forma similar, a relação também ocorre entre M1 e M3, com média de confiança de 0.95. Sugere-se que tal fator de dependência se deve ao conteúdo das disciplinas do Módulo 1, tendo sobretudo um viés introdutório, mostrando aspectos gerais do uso correto do ambiente tecnológico e conceitos iniciais do curso, se verifica forte indicativo de sua importância para bom rendimento dos estudantes e da necessidade de acompanhamento de notas, mesmo em fases iniciais do curso.

Nesse contexto, Bambrick-Santoyo (2018) cita alguns princípios do ensino orientados a dados, onde relaciona bons resultados na aprendizagem com 2 pontos fundamentais: avaliação e análise. Dessa forma, a avaliação do desempenho dos alunos em um estágio inicial de evolução no curso torna-se fator estratégico para

identificar os estudantes potencialmente abaixo da média no decorrer do conteúdo proposto.

Tal análise torna possível a realização de planejamento mais preciso para uma atuação individualizada e pormenorizada, de acordo com as falhas ou lacunas de aprendizado encontradas, o que faz da mineração de dados um instrumento com potencial aplicação educacional, conforme aponta a análise do cenário aqui apresentado.

### **Considerações Finais**

O crescimento das tecnologias da informação permitiram o surgimento de cursos EAD, sendo aplicadas em diversos níveis de ensino, garantindo mecanismos tecnológicos para interações entre os sujeitos envolvidos no processo de ensino e aprendizagem, eliminando o aspecto geográfico como barreira limitadora, e fazendo com que a EAD tenha características diferentes da educação presencial.

Um fator relevante peculiar dos ambientes de educação a distância é a quantidade de registros gerados a partir das ações dos atores educacionais em um ambiente de ensino e aprendizado virtual (AVEA), que torna esses espaços digitais fortes candidatos ao uso de técnicas de mineração de dados.

Este artigo procurou utilizar alguns desses registros e, por meio de técnicas e ferramentas de mineração de dados, obter informações úteis para auxiliar na tomada de decisão. Esse tipo de análise pode contribuir para a prevenção da evasão, uma vez que permite uma avaliação do desempenho dos alunos, possibilitando a detecção de pontos frágeis e o delineamento de estratégias de reforço da motivação e da aprendizagem àqueles alunos que tendam a apresentar baixo desempenho.

Ficou visível, neste experimento, um padrão de comportamento dos alunos no cenário estudado, que aponta para uma forte relação de dependência entre os conteúdos iniciais do curso e o desenvolvimento do aluno no decorrer dos módulos, fator que serve de análise para criação de mecanismos de acompanhamento e monitoramento dos estudantes, com ênfase nas fases iniciais do aprendizado.

É possível de posse dos resultados obtidos, adotar medidas para que o grupo identificado com menor desempenho (cluster 0) contando com 34 alunos, 10 % do total de alunos melhore sua performance. Dentre as práticas pedagógicas e administrativas algumas são possíveis de serem adotadas, dentre elas podemos recomendar: o reforço do conteúdo relativo ao módulo 1(M1), apresentar de forma diferenciada os conteúdos do Módulo 2(M2) com o uso de hiperdocumentos, elevar a interação entre alunos e professores/ tutores, disponibilização de um maior número de exercícios não avaliativos para que os alunos possam verificar o nível de seu conhecimento, informar ao aluno em que nível o seu aprendizado se encontra

e qual a probabilidade de não concluir o módulo, e por último aumentar o número de canais de comunicação entre a secretaria do curso e os alunos por meio de troca de emails, ligações telefônicas, mensagens de texto, etc.

De forma complementar a esta pesquisa, sugerem-se a realização de novas análises, na forma de trabalhos futuros, para identificar outros padrões de comportamento, bem como aplicar outras técnicas de mineração de dados dentre elas a mineração de textos para que seja possível a verificação dos alunos em fóruns e chats. Cabe ainda como sugestão a continuidade do uso do mesmo dataset para realizar investigações complementares dos atributos disponíveis, que possam ter relevância na tarefa de caracterizar grupos de alunos-alvo.

## Referências

- Bambrick-Santoyo, P. (2018). *Leadership 2.0: A Practical Guide to Building Exceptional Schools*. John Wiley & Sons.
- Daline, M. Z., G., C., & do Prado Lima, W. (Julho de 2013). Utilização de Algoritmos de Agrupamento na Mineração de Dados Educacionais. *Revista Novas Tecnologias na Educação*.
- Detoni, D., Cechinel, C., & Araújo, R. M. (2015). Modelagem e Predição de Reprovação de Acadêmicos de Cursos de Educação a Distância a partir da Contagem de Interações. *Revista Brasileira de Informática na Educação*.
- Fornari, A. (2017). Cálculo Diferencial e Integral e Geometria Analítica e Álgebra Linear na educação a distância. *Ciência & Educação*, pp. 475-492.
- Giolo, J. (2008). A educação a distância e a formação de professores. *Educação & Sociedade*, pp. 1211-1234.
- Han, J., & Kamber, M. (2001). *Data Mining Concepts and Techniques*. San Francisco: Morgan Kaufmann.
- Kampff, A. C. (2009). *Mineração de Dados Educacionais para Geração de Alertas em Ambientes Virtuais de Aprendizagem como Apoio à Prática Docente*. Porto Alegre, RS, Brasil: PPGIE.Universidade Federal do Rio Grande do Sul.
- Machado, L. R. (2011). *Desenvolvimento de um Algoritmo Imunológico para Agrupamento de Dados*. Universidade de Caxias do Sul. Caxias do Sul, RS, Brasil.
- Openredu. (1 de Março de 2019). Openredu. Obtido de <http://openredu.org/Portal>.
- Portal, C. (2016). Estratégias para minimizar a evasão na educação a distância: o uso de um sistema de mineração de dados educacionais e learning analytics.
- Queiroga, E. M. (2018). Modelo de Predição da Evasão de Estudantes em Cursos Técnicos a Distância a partir da Contagem de Interações. *Revista Thema*, pp. 425-438.
- Ramos, J. C. (2017). Um Modelo Preditivo da Evasão dos Alunos na EAD a Partir dos Construtos da Teoria da Distância Transacional. *Simpósio Brasileiro de Informática na Educação-SBIE*.

- Rodrigues, J., & Siqueira, R. F. (2015). Mineração de dados do desempenho acadêmico na educação a distância. Universidade Tecnológica Federal do Paraná.
- Rodrigues, R., & Ramos, J. (2016). Discovery Engagement Patterns MOOCs Through Cluster Analysis. IEEE Latin America Transactions.
- Schneider, L. F. (2002). Mineração de Dados - Conceitos. Universidade Federal do Rio Grande do Sul.
- Silva, L. (2016). Introdução á mineração de dados: com aplicações em R. Rio de Janeiro: Elsevier.
- Unesco. (2002). Open and Distance Learning:Trends,Policy and Strategy Considerations. Division of Higher Education. Unesco.
- Vasconcelos, C. R., & Neto, T. L. (2018). Proposta de Um Modelo de Mineração de Dados Educacionais para Identificar a Colaboração entre Estudantes da EAD. CIET: EnPED.
- Webber, C. G., & Lima, M. W. (2013). Utilização de Algoritmos de Agrupamento na Mineração de Dados Educacionais. Revista Novas Tecnologias na Educação.
- Weka. (1 de Março de 2019). Weka. Obtido de <https://www.cs.waikato.ac.nz/ml/weka/>