

Utilizando o *Learning Analytics* com o K-Means para Análise de Dificuldades de Aprendizagem na Educação Básica

Arienne S. Torcate^{1,2}, José Carlos F. Barbosa³, Cleyton Mário de O. Rodrigues^{1,2,3}

¹ Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil

² Pós-graduação em Engenharia de Computação, Escola Politécnica de Pernambuco, Pernambuco, Brasil

³ Universidade de Pernambuco, Garanhuns, Pernambuco, Brasil

{ariannesarmento0, josecarlosfelix8, cleyton.rodrigues}@gmail.com

Abstract. *This article refers to an experience report of an intervention project in Basic Education, contemplating the use of digital technologies registered in Data Mining in the teaching-learning process. In particular, the project highlighted here has the means of identifying among the mathematical contents which are the students who have the most difficulties. Therefore, techniques of Learning Analysis, Educational Data Mining and Unsupervised Learning Strategies were used together. The results achieved so far reinforce the potential to collect, use and analyze data for more personalized and immersive learning.*

Resumo. *O presente artigo refere-se a um relato de experiência de um projeto de intervenção na Educação Básica, contemplando o uso de tecnologias digitais associadas à Mineração de Dados no processo de ensino-aprendizagem. Em particular, o projeto aqui destacado tem como finalidade identificar dentre oito conteúdos matemáticos quais são os que os alunos apresentam mais dificuldades. Portanto, técnicas de Análise de Aprendizagem, Mineração de Dados Educacionais e estratégias de Aprendizagem Não Supervisionada foram utilizadas em conjunto. Os resultados alcançados até então reforçam as potencialidades de coletar, utilizar e analisar dados para uma aprendizagem mais personalizada e imersiva.*

1 Introdução

Learning Analytics (LA) (Análise de Aprendizagem ou Avaliação da Aprendizagem) é uma área de pesquisa relacionada com a Mineração de Dados Educacionais [Costa *et al.*, 2012]. Por meio da LA é possível, por exemplo, entender o processo de aprendizagem no ambiente ao qual o aluno está imerso, possibilitando que haja a identificação de problemáticas no processo de ensino-aprendizagem, como também a otimização deste processo. Seu principal objetivo é utilizar técnicas para coleta e análise de dados referente aos alunos e seus contextos de aprendizagem [Brasil *et al.*, 2018].

Desta forma, o propósito deste trabalho foi utilizar técnicas de LA combinadas com estratégias de aprendizagem de máquina não supervisionadas [Candão e Real, 2018] para identificar dificuldades de aprendizagem entre conteúdos matemáticos do 6º ano do Ensino Fundamental de um colégio público. Para levantar esses dados, foram desenvolvidos jogos

digitais referente a cada conteúdo; assim, através das informações individuais da interação de cada aluno foi possível montar um conjunto de dados (do inglês, *Dataset*) para análise.

Para guiar a aplicação da *Education Data Mining* (EDM - traduzido do inglês, Mineração de Dados Educacionais), a fim de lidar com o grande volume de dados obtidos, optou-se por utilizar a metodologia *Knowledge Discovery in Database* - KDD (Extração de conhecimento em base de dados) [Candão e Real, 2018] e o algoritmo de aprendizagem de máquina não supervisionada K-Means. Esses métodos foram utilizados em conjunto com o objetivo de encontrar padrões e similaridades entre os dados para transformá-los em informações úteis para o contexto pesquisado.

Este artigo está estruturado da seguinte forma: além da introdução, a Seção 2 e suas respectivas subseções apontam os referenciais teóricos das tecnologias utilizadas neste projeto; a Seção 3 apresenta alguns trabalhos relacionados à pesquisa desenvolvida; na Seção 4 é exposto o percurso metodológico adotado; já a Seção 5 descreve detalhadamente os resultados obtidos e na Seção 6 destacamos as considerações finais, além das perspectivas de trabalhos futuros.

2 Fundamentação Teórica

Nesta seção, apresentamos as principais tecnologias e abordagens utilizadas para a realização do projeto, com destaque a Mineração de Dados Educacionais, a Análise de Aprendizagem e ao algoritmo de aprendizagem K-Means.

2.1 Mineração de Dados Educacionais

A descoberta de conhecimentos em grandes bases de dados (ou KDD) é “um processo, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis” [Fayyad *et al.*, 1996]. O KDD é composto por cinco fases, as quais estão melhor descritas no Quadro 1, conforme exposto por Candão e Real (2018). Como principal atividade dentro do KDD, a Mineração de Dados é uma área de pesquisa interdisciplinar que envolve Bancos de Dados, Estatística, e Algoritmos da Computação Inteligente, e que permite descobrir padrões, tendências, ou mesmo relações entre os dados disponíveis.

Quadro 1. Fases do KDD e suas respectivas descrições.

Fases	Descrição
Seleção	Esta fase é responsável por definir quais as variáveis/atributos que constituirão os dados.
Pré-processamento (ou “limpeza”)	É a fase que trata os dados que contêm algum tipo de problema, tais como os dados com valores ausentes ou discrepantes.
Transformação (ou “formatação”)	É onde acontece a conversão dos dados em um formato comum, para a aplicação dos algoritmos.
Mineração de dados	Aplica técnicas de mineração (ex: Algoritmos de Aprendizado de Máquina de classificação, agrupamento, regra de associação, etc)
Interpretação/Avaliação	Esta fase é responsável por utilizar variadas técnicas de interpretação, visualização e avaliação dos dados.

Fonte: Própria

Neste contexto, a área de Mineração de Dados Educacionais ou (*Educational Data Mining*) é voltada, particularmente, ao desenvolvimento estratégico de métodos que têm o objetivo de extrair conhecimento através de *datasets* educacionais [Silva *et al.*, 2016]. Ou seja, a EDM aplica métodos para detectar padrões em grandes volumes de dados educacionais [Costa *et al.*, 2012], com o intuito de obter informações úteis que tenham impacto positivo na aprendizagem dos alunos.

2.2 Análise de Aprendizagem

A Análise de Aprendizagem (do inglês, *Learning Analytics* - LA) tem como objetivo principal coletar e analisar dados gerados pelas interações de estudantes nos seus contextos de aprendizagem. Ou seja, LA refere-se à interpretação de um grande volume de dados produzidos pelos alunos, a fim de avaliar o seu progresso acadêmico, e assim prever o desempenho e detectar possíveis problemas de aprendizagem [Fonseca e de Araújo Jr, 2019].

Na literatura, alguns estudos [Nogueira *et al.*, 2018; Gomes *et al.*, 2019; Moissa *et al.*, 2014] destacam a LA como uma área de pesquisa ainda em desenvolvimento, mas com potencial para proporcionar benefícios no campo da Informática na Educação (IE). Isto é possível pela relação de LA com outras áreas, como Mineração de dados Educacionais, Ciência de dados, Psicologia, Pedagogia, dentre outras. Os mesmos estudos apontam que, apesar da LA surgir nos espaços virtuais de aprendizagem (AVAs), a mesma vem ganhando cada vez mais adesão pelas instituições de ensino físicas.

Isto decorre dos benefícios proporcionados pela LA, que Segundo Campos e Cazella (2018) podem ser vistos sob duas perspectivas. A primeira é para os estudantes, onde os benefícios residem na atenção personalizada, ou seja, uma aprendizagem focada em seu contexto e necessidade. A segunda é para os docentes, onde os benefícios concentram-se na obtenção de *feedbacks* valiosos que possibilitam uma melhor tomada de decisão, além do estabelecimento de ações estratégicas para intervenções que visem conceber melhorias na aprendizagem dos alunos.

2.3 K-Means

Dentro da Inteligência da Artificial, a subárea da Aprendizagem de Máquina estuda se (e como) é possível fazer com que computadores consigam aprender a classificar entradas de dados, muitas vezes baseado em outras classificações pré-estabelecidas, e sem que sejam explicitamente programados para tal [COPPIN, 2010]. De maneira geral, a área envolve o desenvolvimento de sistemas que possam extrair padrões dos dados de forma automática, para que estes possam ser aplicados à novos casos. Do ponto de vista mais técnico, a Aprendizagem de Máquina decorre da dificuldade em tratar manualmente o volume de dados disponíveis, propondo sistemas inteligentes que possam aprender padrões ou regularidades nos dados. De acordo com Mitchell (1997), podemos destacar os seguintes tipos: a aprendizagem supervisionada (onde são utilizados dados já rotulados e conhecidos para novas análises), por reforço (ações são tomadas analisando as circunstâncias presentes), e não-supervisionada (onde se busca extrair características similares a partir da estrutura dos dados disponíveis).

Em particular, a fim de encontrar agrupamentos nos dados, optou-se por utilizar o K-Means, que é uma estratégia de aprendizado de máquina não supervisionada. Geralmente o K-Means é utilizado para classificar um conjunto de valores de uma amostra em grupos que possuam características mais semelhantes possíveis entre si [Novack *et al.*, 2013]. De

maneira breve, dispondo os dados em um espaço n-dimensional, os agrupamentos no K-Means são identificados a partir dos seguintes passos:

1. Selecione K-Centróides iniciais;
2. Formule K *Clusters* (ou seja, grupos) associando ao centróide mais próximo;
3. Recalcule a posição dos centróides com base no centro de gravidade do grupo;
4. Repita itens (2) e (3) até que os centróides não se movimentem mais.

3 Trabalhos Relacionados

Os autores Moore e Kearsley (2007) afirmam que a utilização de EDM em conjunto com conhecimentos adicionais pode ser uma ferramenta em potencial para personalização e otimização do ensino e aprendizagem. Isso fica evidente na pesquisa realizada por Da Costa *et al.* (2014) onde os autores utilizaram um *dataset* de cursos de especialização à distância e aplicaram técnicas de EDM para identificar o perfil dos alunos que estavam propícios à evasão. Para a análise dos dados, regras de classificação foram aplicadas através de árvores de decisão e, os autores obtiveram 97,6% de assertividade nos resultados.

A potencialidade do uso de LA pode ser evidenciada na pesquisa de Souza Brito *et al.* (2020), onde foi realizada uma análise de indicadores sociais, cognitivos e comportamentais da aprendizagem de alunos através de dados fornecidos por um AVA. Assim, utilizando técnicas de LA, os autores implementaram um *plugin* no AVA do tipo *report* que auxilia professores e coordenadores na identificação de alunos que possuem risco de evasão. Por fim, a pesquisa realizada por Vivas *et al.* (2019) também aplica técnicas de LA, mas num curso de Língua Brasileira de Sinais (LIBRAS), já que este possui um alto índice de reprovações. Com o grande volume de dados gerados por este curso, os autores desenvolveram gráficos e realizaram mineração de regras de classificação para auxiliar na seleção de objetos de aprendizagem específicos e relevantes para se aprender LIBRAS no formato *online*.

O trabalho realizado por Fonseca e Namen (2016) analisa a base de dados disponibilizada pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) a fim de identificar informações importantes que contribuam para ações voltadas a melhorias no processo de ensino e aprendizagem de Matemática na Educação Básica. Tendo em vista que essa base possui um grande volume de dados, a metodologia KDD foi aplicada para identificar fatores e conhecimentos importantes para o contexto pesquisado. Os autores pontuam que utilizar esta metodologia foi de grande importância para analisar informações de grande potencial que muitas vezes não são exploradas devido ao seu grande volume de dados que está além da capacidade humana para realizar interpretações.

Com os trabalhos apresentados acima, fica evidente que a área de LA e EDM podem ser aplicadas em contextos educacionais com objetivos diversos. Diferentemente dos estudos apresentados, neste projeto propomos desenvolver jogos digitais para coleta de informações referente as interações dos alunos e com isso, criar nosso próprio *dataset*. Outro diferencial é que, para a análise destes dados, tecnologias como LA e EDM foram utilizadas em conjunto com o algoritmo de aprendizagem de máquina não supervisionada (K-Means) para identificar agrupamentos que traduzam a realidade do aprendizado dos alunos em um domínio de conhecimento específico.

4 Percurso Metodológico

Os procedimentos metodológicos utilizados na presente pesquisa se encaixam nas teorias experimentais de Gerhardt e Silveira (2009), pois são realizadas coletas de dados juntamente com observações. Normalmente, isso ocorre para verificar o efeito de um objeto em relação a um indivíduo. Para melhor condução, esta pesquisa foi dividida em quatro etapas, sendo elas:

1) Planejamento: Nessa etapa foi verificada a viabilidade do projeto junto a professora titular da disciplina de Matemática. Após isso, foi decidido os oito conteúdos matemáticos que seriam trabalhados no projeto, sendo eles: As Quatro Operações, Potenciação, Expressões Numéricas, Frações, Operações com Frações, Porcentagem, Transformações de Unidades de Medidas e Operações com Números Decimais.

2) Desenvolvimento dos jogos: Para desenvolvimento dos jogos, o software JClíc Author foi utilizado. Dentre as dezesseis (16) possibilidades para criação de atividades e jogos lúdicos que o software oferece, para manter o mesmo nível de dificuldade, optou-se por criar dois jogos por conteúdo, sendo eles do tipo Associação Simples (Jogo 1) e Palavras Cruzadas (Jogo 2). A Figura 1 ilustra os dois tipos de jogos desenvolvidos neste projeto.

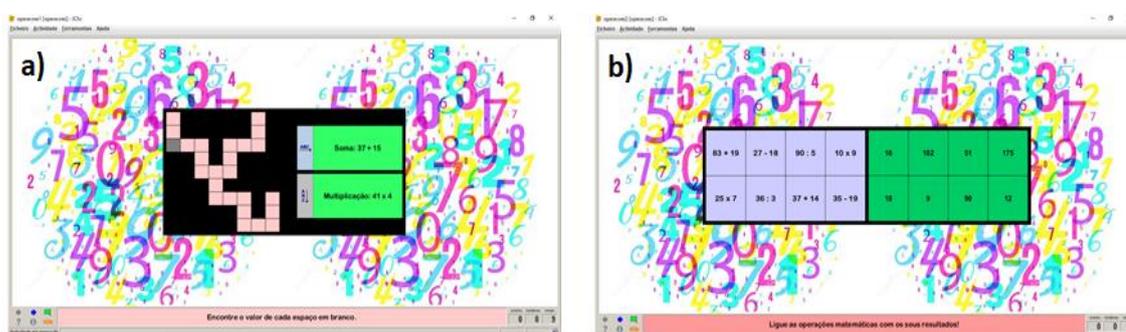


Figura 1. Tipo dos jogos desenvolvidos: a) Palavras Cruzadas e b) Associação simples.
Fonte: Própria.

3) Aplicação: Os jogos foram executados no software JClíc Player e aplicados com 36 alunos do 6º ano, em um Laboratório de Computação (Figura 2), com uma infraestrutura com suporte a um aluno por máquina. O intuito da aplicação foi coletar dados individuais dos alunos durante a interação com os jogos. Isso foi possível através do JClíc Reports, que é uma funcionalidade do JClíc para armazenar informações referente às interações dos jogadores. Com isto, as informações fornecidas pelo JClíc Reports foram coletadas e transcritas de forma manual para uma tabela no Excel e o *dataset* foi construído.

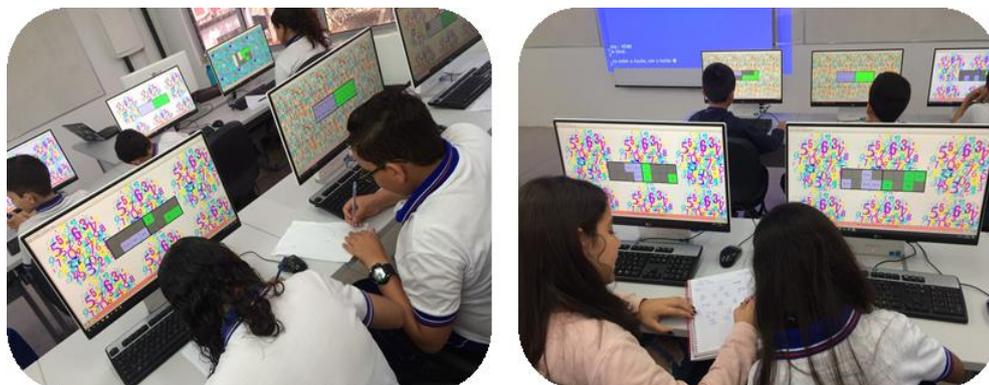


Figura 2. Momento de interação com os jogos. Fonte: Própria.

4) Aplicação das fases do KDD: As fases do KDD foram aplicadas conforme a perspectiva de Candão e Real (2018), mas adaptadas para o contexto da presente pesquisa. O Quadro 2 apresenta cada fase com suas respectivas ações.

Quadro 2. Aplicação das fases do KDD.

Fases do KDD	Ações
Seleção	As variáveis selecionadas para criar o <i>dataset</i> foi referente ao i) número de tentativas, ii) tempo, iii) pontuação e iv) status da entrega do jogo (finalizado corretamente ou não concluído).
Pré-processamento	Dados do tipo <i>string</i> foram convertidos para valores binários e valores ausentes substituídos por zero. Além disso, foi realizada a junção do resultado dos dois tipos de jogos por conteúdo.
Transformação	A planilha no excel foi convertida ao formato CSV.
Mineração dos dados	Para encontrar padrões de agrupamentos, o algoritmo K-Means foi utilizado através do software Orange Data Mining.
Interpretação dos resultados	Para auxiliar na interpretação das informações obtidas, os dados foram exibidos através dos visualizadores Scatter Plot e Silhouette Plot, disponibilizados também pelo software Orange Data Mining.

Fonte: Própria.

De maneira geral, seguindo as quatro etapas que estão organizadas e expostas acima, foi possível obter conhecimentos a partir dos dados coletados. Isto é evidenciado na seção a seguir.

5 Análise dos Resultados

Após inserir o *dataset* no software Orange Data Mining e fazer as configurações necessárias, com o intuito de visualizar os dados classificados e agrupados pelo algoritmo K-Means, utilizou-se dois visualizadores distintos, como já foi mencionado, são eles: Scatter Plot e Silhouette Plot. Na Figura 3 é possível identificar os dois parâmetros que foram utilizados na análise; enquanto o eixo X destaca as atividades, o eixo Y refere-se as pontuações alcançadas.

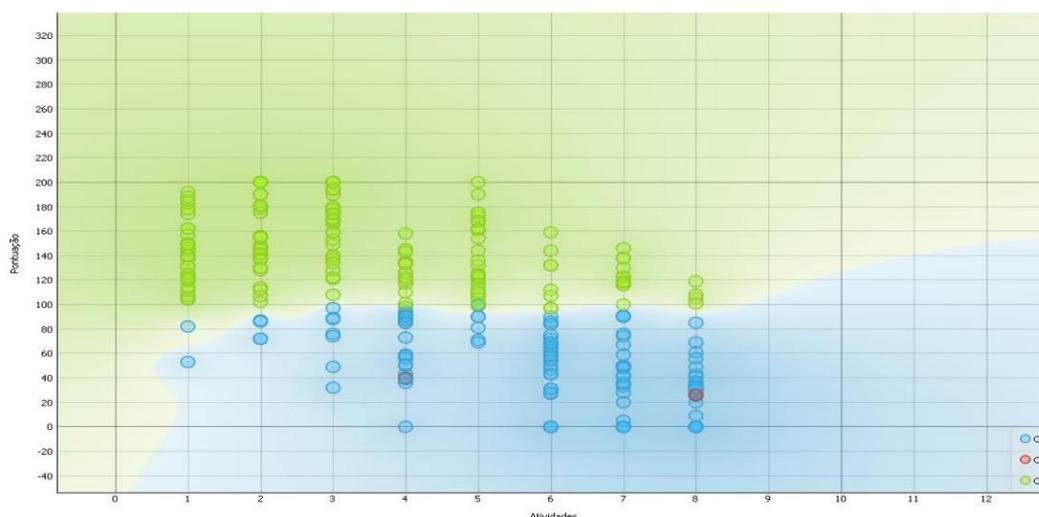


Figura 3. Visualização dos dados através do Scatter Plot. Fonte: Própria.

Como apresenta a Figura 3, as pontuações variaram de zero (0) a duzentos (200); além disso, oito (8) jogos (atividades) foram analisados. É válido lembrar que para cada conteúdo foi desenvolvido dois tipos de jogos (associação simples e palavras-cruzadas). Entretanto, após a aplicação dos jogos, na etapa de pré-processamento dos dados, houve a junção dos resultados desses dois (2) jogos referente a cada conteúdo, diminuindo de dezesseis (16) para oito (8), no intuito de analisar o conteúdo trabalhado.

A Figura 4 apresenta a visualização dos dados através do Silhouette Plot, que nos fornece uma interface voltada a dados numéricos. Mais uma vez, no eixo X temos as atividades e no eixo Y temos as pontuações, que estão claramente classificadas acima de 0 (positivas) e abaixo de 0 (negativas).

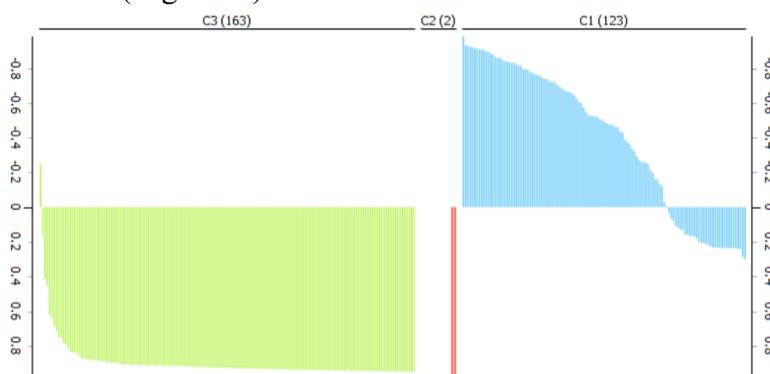


Figura 4. Visualização dos dados através do Silhouette Plot. Fonte: Própria.

Ambas as Figuras 3 e 4 apresentam três *clusters* (grupos) que possuíam entre si maior similaridade; grupos estes criados automaticamente pelo K-Means. Esses *clusters* são denominados C1, C2 e C3. Através dos visualizadores é possível enxergar os dados de cada *cluster* sob perspectivas diferentes. Dessa forma, fica nítido que o C1 é composto pelos alunos que tiveram uma pontuação baixa, variando de 0 a 97 (Scatter Plot) ou grande parte abaixo de 0 (Silhouette Plot). Isso indica que este grupo de alunos tiveram dificuldades nos jogos referente aos conteúdos que estão expostos na Figura 5.

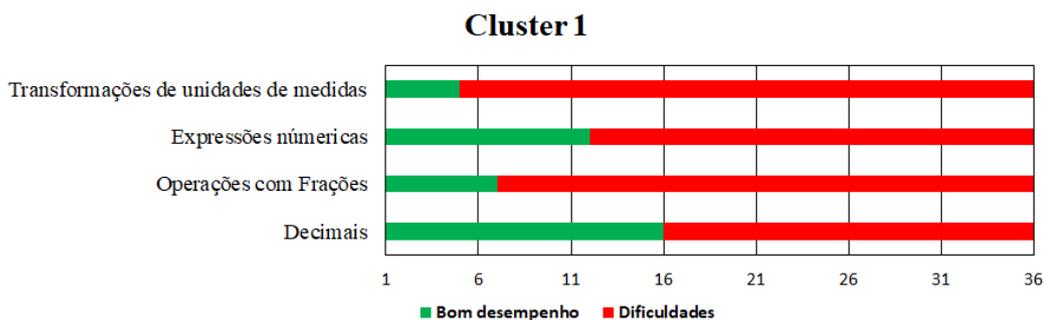


Figura 5. Resultado do Cluster 1. Fonte: Própria.

Com este resultado referente ao grupo C1, fica evidente que dos 36 alunos que participaram do projeto, 31 deles apresentam dificuldades nos jogos de Transformações de unidades de medidas. O conteúdo de Expressões Numéricas teve 28 de 36 alunos com dificuldades. Nos jogos referente às Operações com Frações 28 de 36 alunos também apresentaram dificuldades. E por fim, nos jogos de Decimais 20 de 36 alunos tiveram dificuldades para resolver os desafios propostos nos jogos.

Os parâmetros utilizados para chegar a esses resultados foi a pontuação dos alunos por jogo. Mas, para reforçar ainda mais este resultado, no decorrer das análises foi possível identificar que os alunos do C1 tiveram um número de tentativas maior e passaram mais tempo para resolver as questões propostas nos jogos. Ou seja, o número de tentativas e o tempo são dois parâmetros que também colaboram na união e correlação desse *cluster*.

Durante a análise do *cluster 2* foi identificado que o único parâmetro que une esse grupo é o número exacerbado de tentativas, como apresenta a Tabela 1. Porém, esse fator não é tão preocupante porque essa exceção só aconteceu em relação a dois alunos.

Tabela 1. Resultado C2.

Cluster 2		
Conteúdos	Aluno	Tentativas
Transformações de unidades de medidas	29	743
Decimais	30	774

Fonte: Própria.

O *Cluster 3* é composto por alunos que tiveram uma pontuação consideravelmente alta, variando de 100 a 200. Os alunos deste *Cluster* apresentaram um ótimo desempenho nos jogos referente aos conteúdos expostos na Figura 6.

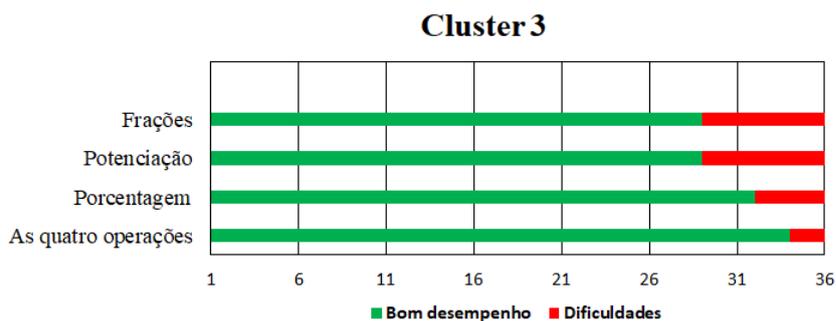


Figura 6. Resultados C3. Fonte: Própria.

Com os resultados apresentados na Figura 6 é possível identificar que 29 de 36 alunos tiveram um bom desempenho em Frações. Em seguida, 29 de 36 alunos conseguiram resolver os desafios dos jogos referente a Potenciação. Também houve um desempenho altamente significativo no que se refere aos jogos de Porcentagem, onde 32 de 36 alunos conseguiram resolver os problemas propostos. Por fim, nos jogos referente as quatro operações, 34 de 36 alunos obtiveram êxito no desempenho e conseguiram entregar corretamente os desafios dos jogos. Além disso, observando por outros parâmetros, os alunos deste *Cluster* não demoraram para resolver os desafios (tempo) e o número de tentativas com erros foi significativamente baixa (tentativas).

Diante do que foi exposto no decorrer desta seção, fica evidente que o principal objetivo do projeto foi alcançado com sucesso. Ou seja, foi identificado quais os conteúdos que os alunos apresentaram mais dificuldades e quais eles obtiveram um melhor desempenho, utilizando um processo de descoberta de conhecimento alinhada com estratégias de aprendizagem de máquina inteligente. Isto foi possível através das informações fornecidas pelos três *clusters* apresentados acima.

Portanto, utilizar os visualizadores Scatter Plot e Silhouette Plot em conjunto foi importante para verificar a relação entre os dados que estavam sendo analisados. Além disso, enxergar outros parâmetros na análise foi totalmente relevante para confirmar a veracidade das informações de cada *cluster*, visto que ambos foram criados de forma automática pelo algoritmo K-Means no software Orange Data Mining.

6 Conclusão

Este artigo teve como objetivo demonstrar através de um relato de experiência o potencial de se utilizar a área de Mineração de Dados Educacionais em paralelo com *Learning Analytics* para proporcionar benefícios ao âmbito educacional. Isto fica evidente nos resultados obtidos, onde foi possível oferecer um *feedback* através de dados concretos para a professora de matemática, de forma que ela possa refletir e reavaliar sua metodologia para suprir as dificuldades encontradas, assim obtendo melhores resultados quanto ao processo de ensino e aprendizagem dos alunos participantes da disciplina.

Além disso, outro fator significativo foi que os alunos não demonstraram estar entediados na aplicação dos jogos desenvolvidos, mas sim interessados e motivados a resolver os problemas matemáticos para progredir de fase. Isso foi notório através das interações com a professora para tirar dúvidas e das anotações realizadas no caderno, como mostra a Figura 2 da Seção 4.

Como perspectivas de trabalhos futuros, destaca-se sugerir uma intervenção com base nos dados expostos neste projeto, com estratégias pedagógicas assertivas e que trabalhem diretamente nas dificuldades encontradas. O intuito é acompanhar a intervenção e realizar uma segunda aplicação do projeto levando em consideração os conhecimentos construídos acerca desses discentes. Para isso, pretende-se desenvolver novos jogos utilizando técnicas de GLA (*Game Learning Analytics*) juntamente com as demais tecnologias já abordadas nesta pesquisa para criação de um novo *dataset* que será analisado a fim de identificar a efetividade da intervenção.

Referências

- Brasil, P., C.; Medeiros, T., J.; Nunes, I., D. (2018). “O uso de Learning Analytics no Brasil: uma revisão sistemática da literatura”. Revista Tecnologias na Educação – Ano 10 – Número/Vol.26.
- Candão, J., P.; Real, E., D. (2018). “A Mineração de Dados Educacionais como Apoio na Análise e Compreensão do Processo de Aprendizagem”. Editor. Nuevas Ideas en Informática Educativa, Volumen 14, p. 545 - 550. Santiago de Chile.
- Campos, A.; Cazella, S. C. (2018). Learning Analytics em processos de personalização de aprendizagem: uma revisão sistemática de literatura. Revista Novas Tecnologias na Educação (RENTE), v. 16, n. 1.
- Coppin, Ben. (2010). Inteligência artificial. Rio de Janeiro: LTC.
- Costa, E.; Baker, R. S. J. D.; Amorim, L.; Magalhães, J.; Marinho, T. (2012). “Mineração de dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações”. Jornada de Atualização em Informática na Educação – JAIE.

- Da Costa, S. S., Cazella, S., & Rigo, S. J. (2014). “Minerando Dados sobre o desempenho de alunos de cursos de educação permanente em modalidade EAD: Um estudo de caso sobre evasão escolar na UNA-SUS”. *RENOTE*.
- De Souza Brito, M. T., Medeiros, F. P. A. de., Bezerra, E. P., & Barbosa, A. S. R. (2020). “Contributions of the report-type plugin for dropout risk identification in Moodle VLE based on data visualization”. *Brazilian Journal of Computers in Education (Revista Brasileira de Informática na Educação - RBIE)*, 28, 01-29.
- Fayyad, U., G. Piatetsky-Shapiro e P. Smyth. (1996). “From data mining to knowledge discovery: An overview”. *Advances in knowledge discovery and data mining. AI Magazine*, p. 1–34.
- Fonseca, E. S.; de Araújo Júnior, C. F. (2019). “Learning Analytics para avaliar o desempenho discente em uma disciplina de cursos a distância”. *REnCiMa*, v. 10, n. 4, p. 137-151.
- Fonseca, S. O.; Namen, A. A. (2016). “Mineração em bases de dados do Inep: Uma Análise Exploratória para nortear melhorias no Sistema Educacional Brasileiro. *Educ. Rev. Belo Horizonte*, V. 32, N. 1, P. 133-157.
- Gerhardt, T. E.; Silveira, D. T. (2009). “Métodos de pesquisa”. Editora UFRGS.
- Gomes, M. B. S.; Sandeski, M. S.; Matiello, S.; Padula, D. (2019). “LEARNING ANALYTICS: O uso de ferramentas tecnológicas no processo de ensino e aprendizagem”. *Revista Mundi Engenharia, Tecnologia e Gestão*. v.4, n.1.
- Mitchell, T. M. (1997). “Machine Learning”. McGraw-Hill, New York, 1.1.
- Moore, M.; Kearsley, G. (2007). “Educação a Distância: uma visão integrada”. São Paulo: Thomson Learning.
- Moissa, B.; Gasparini, I.; Kemczinski, A. (2014). “Learning Analytics: um mapeamento sistemático”. *Nuevas Ideas en Informática Educativa (TISE)*.
- Novack *et al.* (2013). “Distribuição de subgrupos com base nas respostas fisiológicas em Jogadores Profissionais de Futebol pela técnica K-Means Cluster”. *Rev Bras Med Esporte – Vol. 19, Nº 2*.
- Nogueira, S., Francisco, R., Silva, L., Fernandes, M., & Júnior, C. (2018). Big Data com Learning Analytics para apoiar o Planejamento Pedagógico Acadêmico. In *Anais do VII Workshop de Desafios da Computação aplicada à Educação*. Porto Alegre: SBC.
- Silva, J., C., S.; Rodrigues, R., L.; Ramos, J., L., C.; Souza, F., F.; Gomes, A., S. (2016). “Mineração de Dados Educacionais Orientada por Atividades de Aprendizagem”. *CINTED-UFRGS*, V. 14 Nº 1.
- Vivas, A.; Perini, C. A. S.; Pitangui, C. G.; Silva, C. M.; Assis, L. P. (2019). “Learning Analytics aplicado ao curso online da Língua Brasileira de Sinais de uma Instituição de Ensino Superior”. In *Anais dos Workshops do VIII Congresso Brasileiro de Informática na Educação (WCBIE 2019)*.