

# Uma Análise de Algoritmos de Clusterização para Descoberta de Perfis de Engajamento

Pamella Letícia S. de Oliveira<sup>1</sup>, Rodrigo L. Rodrigues<sup>1</sup>, Jorge Luis Cavalcanti Ramos<sup>2</sup>, João Carlos Sedraz Silva<sup>2</sup>

<sup>1</sup>Universidade Federal Rural de Pernambuco (UFRPE)

<sup>2</sup>Universidade Federal do Vale do São Francisco (UNIVASF)

{pamella.lesilvao, rodrigomuribec, jorgeluiscavalcanti, jsedraz}@gmail.com

**Abstract.** *Distance education (EAD) has been taking great proportions in access to education, due to the flexibility of time and geographic location. However, despite the adoption by educational institutions, the number of graduates in this modality has low rates. How student engagement and interaction is directly related to academic performance factors. This article aimed to describe different grouping methods that were used in a database for a given course, which was divided into the first four and the last four periods. The results point to three engagement profiles, which showed no significant differences between the two phases of the course..*

**Resumo.** *A educação a distância (EAD) vem tomando grandes proporções no acesso à educação, devido a flexibilização do tempo e localização geográfica. No entanto, apesar da adoção pelas instituições de ensino, o número de concluintes nesta modalidade apresenta baixos índices. Como o engajamento e a interação entre alunos está relacionado diretamente aos fatores de desempenho acadêmico. Este artigo teve como objetivo descrever distintos métodos de agrupamento que foram usados em uma base de dados de um determinado curso, que foi dividido em quatro primeiros e quatro últimos períodos. Os resultados apontam para três perfis de engajamento, que não apresentaram diferenças significativas entre as duas fases do curso.*

## 1. Introdução

A tecnologia permitiu que o acesso a educação fosse modernizado, possibilitando vantagens como: flexibilização de horário, acesso geográfico e custos acessíveis, exemplos de plataformas que proporcionam isso são: Ambientes Virtuais de Aprendizagem (AVA), que possibilitam a interação e disponibilizam diversos recursos de comunicação [Rodrigues et al 2016]; Ensino mediado por aparelhos eletrônicos (do inglês, E-learning), que vem oferecendo o ensino à distância e o acesso a conteúdos online em tempo flexível [Haron et al 2017]; Cursos Online Abertos e Massivos (MOOC), que foi criada com a ideia de que as pessoas em todo o mundo conseguissem obter acesso a tópicos de alta qualidade em cursos profissionais e ensino superior, independentemente do seu país de residência, escolaridade e nível socioeconômico [Williams et al 2018].

No entanto, a qualidade do ensino superior teve um impacto negativo sobre a mídia nos últimos anos, quando são observados alguns pontos: evasão acadêmica, recém-formados com dificuldade de entrar no mercado de trabalho e a alta taxa de estudantes que não alcançam um bom nível de conhecimento. Esse conhecimento é avaliado no Brasil pelo Ministério da Educação por meio do Exame Nacional de Desempenho do Estudante (ENADE) [Coelho and Vega 2019]. Em específico as taxas de sucesso em cursos baseados em e-learning não estão como esperado, a maioria dos recursos de aprendizagem em e-learning é baseada em instrução individualizada, e isso tem ocasionado a diminuição da interação social, desdobrando-se em um aumento na reclusão dos estudantes [Haron et al 2017]. Além disso, pesquisas apontam que apenas 10% dos alunos matriculados na modalidade do tipo MOOC, conseguem finalizar o curso [Rodrigues et al 2016].

Um dos fenômenos educacionais que as pesquisas apontam como característica intrinsecamente ligada a taxa de sucesso nos cursos é o fenômeno do engajamento. Este fenômeno é considerado um dos principais componentes no ensino online eficaz, e é visto como importante para a diminuição da retenção e melhoria da qualidade da experiência geral do estudante [Haron et al 2017]. Além disso, a falta de engajamento pode ser visto como uma interação pobre [Capuano et al 2013]. Devido a isso, ao identificar o nível de engajamento, de qualidade da interação e de colaboração, é possível definir estratégias para reduzir o número de desistências dos alunos, contribuir para a maior qualidade da aprendizagem e aumentar a satisfação [Messias; Morgado and Barbas 2015]. Diante disso, o objetivo deste estudo é utilizar técnicas de mineração de dados educacionais (EDM), em específico a análise de agrupamento com o intuito de encontrar diferentes perfis de engajamento em plataformas de sistema de gestão de aprendizado (LMS).

Diante disso, para alcançar o objetivo deste estudo na metodologia será realizada técnicas de mineração de dados educacionais (EDM), em específico a análise de agrupamento com três diferentes tipos de algoritmos, para encontrar perfis de engajamento em plataformas de sistema de gestão de aprendizado (LMS) que ao analisá-los podem ser estudadas medidas para evitar reprovações e desistências dos alunos.

## **2. Engajamento em Plataformas de LMS**

O engajamento envolve alunos ativos, disposição, participação regularmente da colaboração e interação entre os participantes. Isso ocorre por meio das expressões, das idéias e das opiniões que surgem através das interações frequentes e das enriquecedoras experiências educacionais mediante o pensamento, as conversas e a interação entre os alunos, conteúdo e educadores [Haron et al 2017]. A compreensão de engajamento proposta por [Fredricks; Blumenfeld and Paris, 2004] presume três grandes categorias para o engajamento: o comportamental, que abraça a participação dos alunos em atividades acadêmicas e sociais ou co-curriculares e envolve ações observáveis, como a conclusão da tarefa; o cognitivo, que refere-se ao investimento dos alunos na aprendizagem, vontade de aceitar tarefas mentais desafiadoras e uso de estratégias de auto-regulação para controlar processos de pensamento; o afetivo, que envolve dimensões das emoções dos alunos e seus sentimentos sobre reações em relação aos outros, bem como a sua identificação com a escola e relacionamento com os outros [Pineda-Báez et al 2019].

Visto a importância do engajamento e da interação entre os alunos para a educação a distância, muitas plataformas ainda possuem dificuldade em proporcionar maior envolvimento entre os alunos. Os sistemas e ferramenta de e-learning atuais não contribuem para o aumento dos níveis de engajamento, isto se dá devido a uma interação pobre com os materiais de aprendizagem, falta de capacitação docente e discente para o uso adequado dos recursos, ocasionando em uma diminuição na interação social entre os sujeitos envolvidos [Capuano et al 2013]. Um LMS pode satisfazer uma série de requisitos de aprendizagem, mas a sua incapacidade para acomodar o lado social de aprendizagem é visto como uma desvantagem importante a ser mencionada [Whitty and Anane 2014]. No estudo de [Goh et al 2017] foi aplicado um questionário a 66 alunos, com o objetivo de explorar o processo de aprendizagem e percepção de alunos no uso de um MOOC. Ao analisar as respostas dos estudantes foi identificado que os facilitadores precisam treinar os alunos a usar MOOC, proporcionar designer interativo e possibilitar constantes feedbacks para assim melhorar o engajamento e o interesse dos alunos em utilizar o MOOC.

### **3. Trabalhos Relacionados**

O engajamento é um fenômeno multifacetado, por isto tem despertado o interesse de pesquisadores de diversas áreas [Rodrigues et al 2016]. O termo possui muitos fatores e muitas definições, embora muitas pesquisas relacionam o termo engajamento a interação entre os alunos. No estudo de [Goh et al 2017] eles citam que a interação dos aluno em ambientes de aprendizagem online tem implicações sobre o engajamento e a colaboração.

O engajamento em muitos estudos está relacionado com a performance, o desempenho e o sucesso do aluno no curso. Segundo [Oliveira, Souza and Rodrigues 2019] afirmam que geralmente ao buscar por pesquisas de engajamento em plataformas LMS, a performance sempre vem acompanhada, principalmente nos modelos de predição, que almejam saber qual será o desempenho do aluno e de qual forma poderia ajudar aqueles que não estão com notas boas. No de [Bergdahl et al 2019] afirmam que há um forte consenso de que o engajamento é necessário para os estudantes e isso tem uma forte implicação em relação ao sucesso na escola. Para saber os fatores que levam ao engajamento foi aplicado um questionário em 11 escolas da Suécia para 552 estudantes. Como resultados, compreenderam que os aspectos comportamentais, cognitivas, emocionais e sociais são fatores estritamente relacionados ao aumento no nível de engajamento dos estudantes. Além do mais, eles afirmam que as relações em trabalhos de grupo, a presença social dos professores em ambientes on-line e a capacidade de encontrar um equilíbrio digital são fatores benéficos para o engajamento e os resultados dos alunos, além de beneficiar dentro e fora da escola, agora e no o futuro.

O estudo de [Messias; Morgado and Barbas 2015] visa entender como a interação e a colaboração contribuem para o engajamento dos alunos em e-learning, principalmente em plataforma informal complementada de uma plataforma formal. Para isso, as técnicas utilizadas foram qualitativas, onde foram recolhidas mensagens dos estudantes em fóruns de discussão Moodle e em grupos no Facebook, além de questionários e experimentos. Ainda referente ao estudo de [Messias; Morgado and Barbas 2015] são citados algumas afirmações de outros autores que mostram a relação entre interação e engajamento, como: “o engajamento é desenvolvido através de interação” e “entre outros fatores o engajamento é centrado na interação do aluno e em

ambientes de aprendizagem online tem implicações sobre o envolvimento do aluno e a colaboração”.

Assim como os estudos mencionado acima, este artigo aborda o conceito do engajamento e sua importância para os alunos, professores e instituições. Além disso, o diferencial desse estudo em comparação aos demais é a utilização de técnicas de mineração de dados e análise de agrupamento para encontrar os diferentes perfis de engajamento.

#### **4. Metodologia**

A metodologia utilizada para o desenvolvimento desta pesquisa foi dividida em três fases: a primeira fase da caracterização da pesquisa, a segunda fase da aplicação de técnicas de análise de agrupamento e a terceira fase é direcionada a descrição dos perfis de engajamento.

##### **4.1 Sujeitos da Pesquisa**

Na primeira fase, foram extraídas informações do banco de dados da plataforma Moodle de uma universidade brasileira. Os dados são de 6527 alunos do curso de biologia do 1º ao 8º período, de 49 disciplinas. A maioria dos estudos relacionados ao nível de engajamento tem-se pautado pela medição da frequência ou interação dos alunos. No estudo de [Moubayed et al. 2018] utilizaram variáveis como notas do quiz, número de logins e nota do curso para medir o quão engajado está um aluno. No estudo de [Blanchett, J. 2012] diz que a interligação de mensagens é usado como um indicador de interação ou engajamento em discussões online. Então, para os objetivos deste estudo, definimos 8 variáveis avaliativas e três comportamentais. As variáveis avaliativas foram compostas por duas provas, quatro fóruns e duas atividades em formato de webquests. As variáveis comportamentais foram definidas como sendo a quantidade de mensagens enviadas, acesso ao ambiente e acesso ao quadro de notas.

Na segunda fase, os dados obtidos foram divididos em duas metades, a primeira composta pelos alunos do 1º ao 4º período e a segunda composta pelos alunos do 5º ao 8º período. Ambas fases foram testadas no algoritmo hierárquico, no DBSCAN e no K-Means para escolher qual dentre estes possuía a melhor formação dos grupos. Para essa escolha foram usadas métricas de validação que permitiram saber o quão bom está formado o grupo e o grau de similaridade entre eles, além disso, também foram consideradas características particulares dos algoritmos, tais como: a possibilidade de se operar com bases de dados maiores e de ser mais rápido.

Na terceira e última fase, a partir da análise resultante da aplicação dos algoritmos usando as variáveis avaliativas e comportamentais, foram encontrados grupos com comportamento similares, restando então três grupos que representaram as categorias de engajamento, onde os alunos puderem ser enquadrados em engajados, medianos e desengajados, as representações gráficas deles foram feitas por meio de gráfico de densidade e box plot.

##### **4.2 Técnicas de Agrupamento**

Nesta fase foi realizada análise de clusters (agrupamentos), essa técnica é uma das tarefas de mineração de dados, e tem como objetivo agrupar dados de tal forma que as características do grupo (denominado cluster) são mais semelhantes entre si do que com outros grupos [Ramos et al 2016]. Durante a realização do experimento, foram utilizadas duas formas de análises de agrupamento: a hierárquica e a não hierárquica. A

hierárquica consiste em uma série de sucessivos agrupamentos ou sucessivas divisões de elementos, nas quais os elementos são agregados ou desagregados visando construir uma hierarquia de clusters. A não hierárquica divide o conjunto de observações de uma base de dados em um número pré-definido de subconjuntos disjuntos e procuram iterativamente o melhor particionamento até atingir uma condição de parada [Ramos et al 2016].

Para saber qual o melhor algoritmo a ser utilizado nos dados disponíveis, foram avaliados três algoritmos: o hierárquico, o K-Means e o DBSCAN. O hierárquico tem sua principal vantagem a representação dos seus resultados por meio de dendrogramas. Dado que, o número de clusters não são definidos inicialmente no algoritmo, mas sim identificado por meio da análise do dendrograma, que infere o número adequado de clusters. Outro algoritmo que não necessita informar previamente o número desejado de grupos é o DBSCAN. Para isso, ele se baseia na classificação de cada objeto do conjunto de dados em uma dentre 3 categorias: objeto central, objeto de borda e ruído [Padilha and Carvalho 2017]. O DBSCAN assim como o Hierárquico descobre o número de clusters, no entanto, diferentemente do hierárquico ele é indicado para uma grande quantidade de dados. Assim como, o *K-means* que quando comparado com o método hierárquico é mais vantajoso nas aplicações que envolvam um grande número de conjuntos, pois não há a necessidade do dendrograma [Ramos et al 2016]. Devido a necessidade de estabelecer o número de clusters pelo usuário antes do processo foi usado o método *elbow*. Esse método testa a variância dos dados em relação ao número de clusters, até o momento que conforme o número de clusters aumenta não representa um valor significativo de ganho.

## 5. Resultados e Discussões

Nesta seção são apresentadas e interpretadas a análise de agrupamento e a descobertas dos grupos. Logo, serão mostrados os conceitos e os resultados dos algoritmos de agrupamento e qual deles mais se adequou aos dados da pesquisa. Os três perfis diferentes de engajamento formados ao analisar as variáveis avaliativas e após isso a análise das variáveis comportamentais.

### 5.1. Descoberta de Grupos

Visto as características dos algoritmos K-means, hierárquico e DBSCAN e a capacidade deles para manipular grandes quantidades de dados. Foi descartada a utilização do algoritmo hierárquico, que mesmo sendo dividido os dados em duas metades continuou sendo inviável o uso desse algoritmo. Então, foram usados o DBSCAN e o K-Means, os resultados dos grupos formados por eles estão apresentados na Tabela 1

**Tabela 1. Referente aos grupos formados por cada algoritmo.**

Nº Instâncias	Grupo1	Grupo2	Grupo3	Grupo4	Grupo5	Grupo6
DBSCAN (1ª Parte)	2294	23	881	919	655	84
DBSCAN (2ª Parte)	362	24	521	614	18	39
K-Means (1ª Parte)	1827	945	2137	-	-	-

K-Means (2ª Parte)	403	787	428	-	-	-
--------------------	-----	-----	-----	---	---	---

Ao analisar os resultados mostrados na Tabela 1 é notório que o K-Means apresentou uma divisão balanceada nos agrupamentos em comparado com o DBSCAN. Mesmo assim, para confirmar que ele possui o melhor particionamento foi necessário utilizar uma métrica de validação. Dado que, validações podem informar o quão bom está formado o agrupamento, exemplo é a Largura de Silhueta que analisa os grupos em que os elementos estão próximos entre si e distantes de elementos que estão em outros grupos. Na largura de silhueta é considerado valores no intervalo de  $[-1,1]$  e quanto mais próximo a 1 melhor é o agrupamento. Os valores encontrados da largura de silhueta para o algoritmo DBSCAN e o K-Means, respectivamente, foi de 0.10 e 0.55 na 1ª parte e 0.17 e 0.50 na 2ª parte. Apesar de ter usado a validação de Silhueta para ambos algoritmos, não é recomendado o uso para o DBSCAN, no entanto o valor da métrica para o K-Means apresentou resultado satisfatório. Então, é notório que ao comparar os agrupamentos formados por esses algoritmos o K-Means apresenta ter os melhores resultados.

## 5.2 Perfis de Engajamento

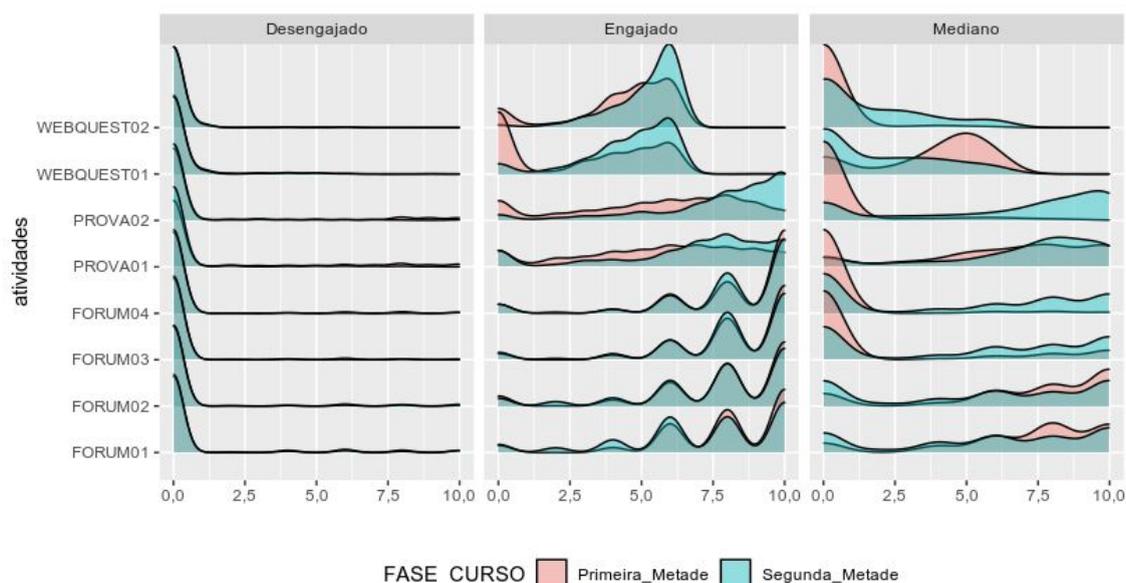
Os algoritmos foram utilizados para um total de 6527 alunos (um aluno uma entidade única por disciplina) de 49 disciplinas diferentes, com média de 60 alunos e desvio padrão de 70 ao longo de 8 períodos. As 8 variáveis avaliativas analisadas no K-Means, resultaram para as duas fases do curso três perfis de engajamento (desengajado, mediano e engajado). A tabela 2 mostra para a primeira e para a segunda metade do curso o quantitativo de alunos que compõe cada grupo e a média das variáveis que são refletidas neles.

**Tabela 2. Tabela referente a descritiva das médias para as duas fases do curso**

Características dos Alunos			Provas		Fóruns				Atividades WQ	
Fase do curso	Grupo	Total	I	II	I	II	III	IV	I	II
Primeira Metade	Desengajado	1827 (37,21%)	1,39	1,09	0,74	0,35	0,29	0,38	0,24	0,3
	Mediano	945 (18,93%)	6,46	0,58	7,11	7,19	1,86	0,42	3,74	0,22
	Engajado	2137 (42,82%)	5,63	5,50	7,95	7,82	8,26	8,27	2,71	4,20
Segunda Metade	Desengajado	428 (26,45%)	0,291	0,468	0,58	0,47	0,10	0,13	0,22	0,11
	Mediano	403 (24,90%)	6,73	6,26	5,79	5,47	4,95	4,37	1,76	1,52
	Engajado	787 (48,64%)	6,63	7,76	7,23	7,60	8,05	8,12	4,53	5,13

Ao analisar os três grupos formados pelo algoritmo K-Means, foi possível

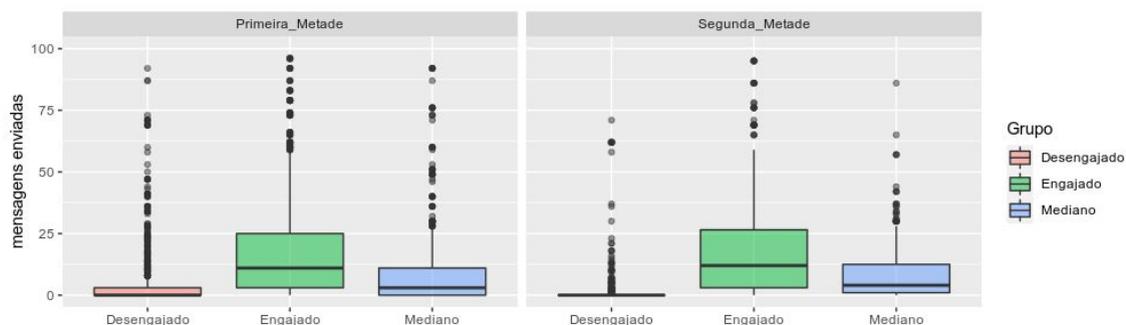
montar uma análise gráfica onde mostra o comportamento dos cluster com as notas referentes a cada variável, como apresentada na Figura 2.



**Figura 2. Análise da gráfica das variáveis e dos perfis de engajamento.**

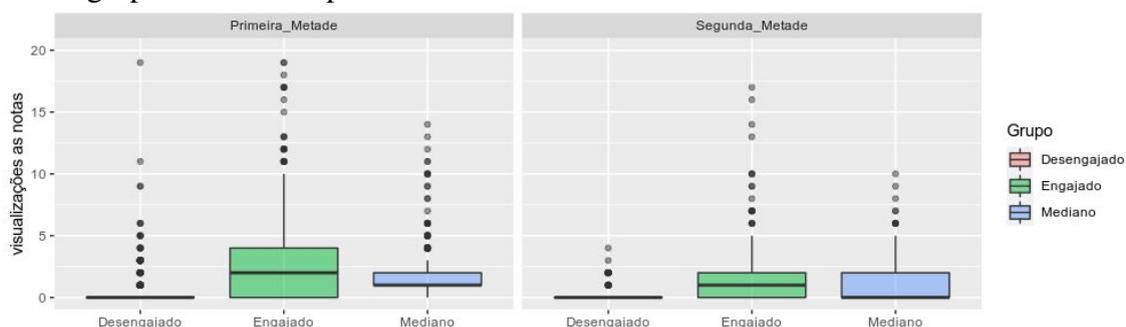
A Figura 2 mostra que existem três perfis de engajamento, o engajado, o mediano e o desengajado. O grupo engajado nas duas metades do curso é o que mais se destacou, apresentando picos nas notas mais altas das médias de provas, de fóruns e de webquest. Por isso, ele pode ser considerado o grupo de alunos que possuem as melhores notas, embora na primeira metade a webquest 1 apresenta pico nas notas baixas. O grupo mediano apresenta constantes variações, com picos em notas baixas, médias e altas, no entanto, a primeira metade possui mais notas baixas em comparação com a segunda metade do curso. Pode ser considerado o grupo que possuem alunos com grande variedade de notas. O grupo desengajado foi considerado o pior dos três grupos. Esse grupo possui pico nas notas próximas a zero em todas as variáveis para as duas metades do curso, o que representa que os alunos pertencentes a esse grupo possuem baixo desempenho.

Para ter uma ideia do comportamento dos grupos nas variáveis comportamentais foi necessário observar a variabilidade dos dados em torno da mediana. Utilizamos o Box plot como medida gráfica para plotar a medida de tendência central, a variabilidade dos dados e avaliar a existência de outliers (valores extremamente altos ou baixos) das três variáveis comportamentais: quantidade de mensagens enviadas no ambiente, visualização ao quadro de notas e acesso ao ambiente em dois momentos diferentes, na primeira parte e na segunda parte do curso.



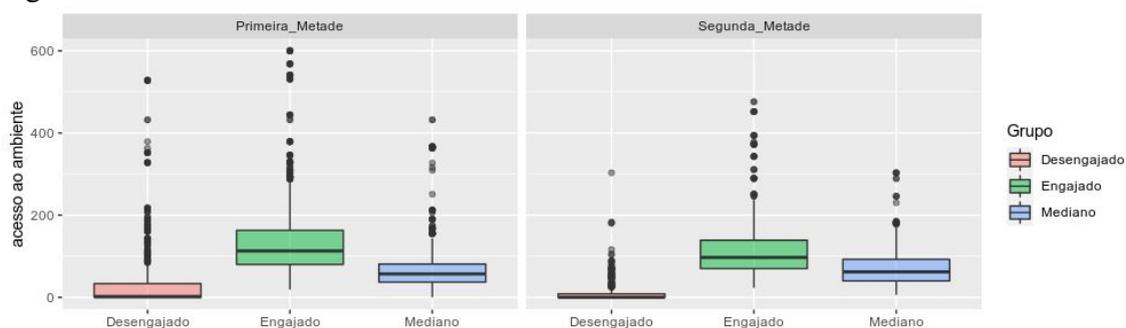
**Figura 3. Box plot referente a variável mensagens enviadas**

Na figura 3 o grupo desengajado nas duas metades do curso mostrou ser o grupo com a pior quantidade de mensagens enviadas devido ser que têm a mediana menor dentre eles, além disso é o que possui mais *outliers* (valores discrepantes). O grupo engajado, têm a mediana superior dentre os grupos nos dois momentos, mostrando ser o grupo que mais envia mensagens. O grupo mediano na segunda metade, graficamente, tem semelhanças ao grupo mediano da primeira metade, sua mediana está entre a dos outros grupos mostrando possuir valores centrais.



**Figura 4. Box plot referente a quantidade de visualizações ao quadro.**

Na Figura 4 o grupo desengajado nos dois momentos apresentou ser o que menos visualiza o quadro, tem mediana perto de zero e pouca variabilidade, embora a existência de outliers diminui no segundo momento. O grupo engajado é o que possui a maior mediana nos dois momentos, no entanto, a variabilidade diminui no segundo momento. Por outro lado, o grupo mediano apresenta uma maior variabilidade no segundo momento.



**Figura 5. Box plot referente a variável acesso ao ambiente.**

Na figura 5, o grupo desengajado nos dois momentos foi o que teve menos acessos ao ambiente, apresentou mediana em zero e existência de outliers. O grupo engajado foi o que mais teve acesso ao ambiente, a maior variabilidade e a mediana

superior. O grupo mediano nos dois momentos possui sua mediana entre os outros grupos, no segundo momento possui um pouco mais de variabilidade e menos outliers.

## **6. Implicações Educacionais**

A qualidade da aprendizagem on-line pode ser avaliada através da observação de engajamento dos alunos no ambiente de aprendizagem on-line [Beer; Clark and Jones, 2010, citada por Shukor et al 2014]. A importância de estudar o engajamento, está diretamente relacionada com o estudante ao longo do curso, dado que, detectar o engajamento do aluno é uma chave importante para melhorar um sistema de e-learning [Altuwairqi et al 2018]. Ao saber o nível de engajamento, a qualidade da interação e colaboração, pode ajudar a definir estratégias para reduzir o número de desistências dos alunos em contextos e-Learning, a contribuir com maior qualidade de aprendizagem e o aumentar a satisfação dos alunos [Messias; Morgado and Barbas 2015].

Analisar o engajamento em um ambiente de aprendizado online é fundamental para entender o desempenho do aluno e muitas vezes está relacionado com a interação. Dado que, a interação é crucial para garantir o sucesso dos alunos na aprendizagem [Haron et al 2017]. Durante a interação os alunos estão gerando explicações, justificativas, e em busca de novas informações o que pode encorajar o aluno, promover a aprendizagem dinâmica, dá autonomia aos alunos e eliminar barreiras como a timidez [Haron et al 2017]. Esta pesquisa por meio da mineração de dados e da análise de agrupamento analisou o engajamento e a interação entre os alunos em plataformas LMS e fornece implicação para os alunos, os educadores e as instituições. Devido que, este estudo mostra a significância desse tema e diferentes perfis de engajamento que existe em uma plataforma LMS, que ao analisá-los podem ser estudadas medidas para evitar reprovações e desistências de alunos. Diante disso, para trabalhos futuros planeja-se aplicar essas técnicas de agrupamento para entender os perfis de engajamento em outros cursos da mesma universidade.

## **7. Conclusões**

Como resultado principal dos agrupamentos realizados, verificou-se que o K-Means e o DBSCAN apesar de ambos serem usados para grandes quantidades de dados eles apresentaram resultados diferentes, formando grupos de tamanhos distintos. Com isso, foi necessário o uso de técnicas de validação o que conclui que o K-Means mostrou ter formado o melhor agrupamento.

Na formação dos grupos em relação tanto a variável avaliativa quanto a comportamental, percebeu-se que o grupo engajado é o que possui maior desempenho, devido ele ter a maior quantidade de notas altas, de mensagens enviadas, de acesso ao quadro de notas e de acesso ao ambiente. O grupo desengajado é o que têm o pior desempenho, com valores inferiores nas notas, na quantidade de mensagens enviadas, de acesso ao quadro e ao ambiente. O grupo mediano é o grupo que fica em constantes oscilações, possuindo notas altas e baixas, e valores centrais para a quantidade de mensagens enviadas e de acesso. De uma maneira geral, cada grupo formado pode oferecer alguma particularidade para que a coordenação do curso ou o professor possa promover ajustes no curso para evitar reprovações e desistências.

## **Agradecimentos**

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Pró-Reitoria de Pesquisa e Pós-Graduação da UFRPE -

(PRPPG) pela concessão da bolsa de iniciação científica - PIBIC.

## 8. Referências

- Altuwairqi, K., Jarraya, S. K., Allinjawi, A. *et al.* (2018) A new emotion-based affective model to detect student's engagement. *Journal of King Saud University - Computer and Information Sciences*.
- Bergdahl, N., Nouri, J. and Fors, U. (2019). Disengagement, engagement and digital skills in technology-enhanced learning. *Education and Information Technologies*, v. 25, n. 2, p. 957–983.
- Capuano, N., Mangione, G. R., Pierri, A. and Lin, E. (2013). Engaging e-learning for Risk Management: The ALICE Experience in Italian Schools. *2013 Seventh International Conference on Complex, Intelligent, and Software Intensive Systems*.
- Coelho, U. M. and Vega, I. S. (2019). The Pedagogical Formation and the Knowledge of Teachers in Computing in Teaching Strategies: Integration of Content, Didactic Material and Interdisciplinary or Integrator Project. *2019 XIV Latin American Conference on Learning Technologies (LACLO)*.
- Fredricks, J. A., Blumenfeld, P. C. and Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, v. 74, n. 1, p. 59–109.
- Goh, W., Ayub, E., Wong, S. Y. and Lim, C. L. (2017). The importance of teacher's presence and engagement in MOOC learning environment: A case study. *2017 IEEE Conference on e-Learning, e-Management and e-Services (IC3e)*.
- Haron, H., Aziz, N. H. N. and Harun, A. (2017). A Conceptual Model Participatory Engagement Within E-learning Community. *Procedia Computer Science*, v. 116, p. 242–250.
- Messias, I., Morgado, L. and Barbas, M. (2015). Students' engagement in Distance Learning: Creating a scenario with LMS and Social Network aggregation. *2015 International Symposium on Computers in Education (SIIE)*.
- Moubayed, A., Injadat, M., Shami, A. and Lutfiyya, H. (2018). Relationship Between Student Engagement and Performance in E-Learning Environment Using Association Rules. *2018 IEEE World Engineering Education Conference (EDUNINE)*.
- Padilha, V.A. and Carvalho, A.C.P.L.F. *Mineração de Dados em Python*. Instituto de Ciências Matemáticas e de Computação. Universidade de São Paulo. 2017.
- Oliveira, P. L. S. D., Souza, A. J. D. and Rodrigues, R. (2019). Identificação de pesquisas referentes ao engajamento de alunos em plataformas de LMS e suas relações com o desempenho acadêmico. *Anais do XXX Simpósio Brasileiro de Informática na Educação (SBIE 2019)*.
- Pineda-Báez, C., Manzuoli, C. H. and Sánchez, A. V. (2019). Supporting student cognitive and agentic engagement: Students' voices. *International Journal of Educational Research*, v. 96, p. 81–90.
- Ramos, J. L. C., Silva, R. E. D. E., Silva, J. C. S., Rodrigues, R. L. and Gomes, A. S. (2016). A Comparative Study between Clustering Methods in Educational Data Mining. *IEEE Latin America Transactions*, v. 14, n. 8, p. 3755–3761.
- Rodrigues, R., Ramos, J., Silva, J. and Gomes, A. (2016). Discovery engagement patterns MOOCs through cluster analysis. *IEEE Latin America Transactions*, v. 14, n. 9, p. 4129–4135.
- Shukor, N. A., Tasir, Z., Meijden, H. V. D. and Harun, J. (2014). A Predictive Model to Evaluate Students' Cognitive Engagement in Online Learning. *Procedia - Social and Behavioral Sciences*, v. 116, p. 4844–4853.
- Whitty, C. and Anane, R. (2014). Social Network Enhancement for Non-formal Learning. *2014 47th Hawaii International Conference on System Sciences*.
- Williams, K. M., Stafford, R. E., Corliss, S. B. and Reilly, E. D. (2018). Examining student characteristics, goals, and engagement in Massive Open Online Courses. *Computers & Education*, v. 126, p. 433.