

## Identificação de gargalos em e-learnings gamificados e indicação dos erros mais frequentes para viabilizar e priorizar melhorias

Tarcísio H. M. Hazin<sup>1</sup>, Dênis Leite<sup>1</sup>, Pedro H. R. Macêdo<sup>1</sup>,  
Daniel V. Pires<sup>1</sup>, Alexandre M. A. Maciel<sup>1</sup>, Mêuser J. S. Valença<sup>1</sup>

<sup>1</sup> Escola Politécnica de Pernambuco – Universidade de Pernambuco (POLI / UPE)  
Caixa Postal 50720-001 – Recife – PE – Brasil

**Abstract.** *Due to the growth of e-learning and the recent changes on student learning preferences and behavior, the scientific community has proposed new approaches to support the application of active learning strategies and gamification. This work presents a solution to identify “bottlenecks” in courses, with the purpose of reducing effort of professors and students on the learning process. The solution includes data mining and clustering using Self-Organizing Map (SOM) to verify the course tasks with the highest error rates and then to group its errors by similarity. Thus it was possible to prioritize interventions at the points that most hinder students’ progress. The approach was successfully applied in an industrial automation e-learning.*

**Resumo.** *Diante do novo perfil dos estudantes e do crescimento do e-learning, a comunidade científica vem propondo métodos para produção de cursos gamificados e que utilizam metodologias ativas. Nesse contexto, este trabalho oferece uma contribuição para identificar “gargalos” em cursos gamificados, com propósito de reduzir o esforço de professores e alunos no processo de ensino e aprendizagem. A solução contemplou mineração de dados e clusterização através do Self-Organizing Map (SOM) para identificar as atividades com maior incidência de erros e agrupar os tipos de erros por similaridade. Assim, foi possível priorizar intervenções nos pontos que mais atrapalham o progresso dos estudantes. A abordagem foi aplicada com sucesso em um e-learning de automação industrial.*

### 1. Introdução

Entre os principais desafios do *e-learning* está oferecer e executar cursos em formatos que sejam envolventes, engajadores, desafiadores e divertidos. Isso se torna ainda mais urgente ao passo em que os novos estudantes são nativos digitais. Ou seja, pessoas que ainda nos primeiros meses de vida são expostos a tecnologias que lhes permitem tanto acessar conteúdos que são do seu interesse quanto mudar de conteúdo quando seu interesse muda. Esses indivíduos se habitam a aprender a partir da experimentação, motivados pela simples curiosidade e chegam a aprender a ler e até soletrar palavras antes mesmo de serem alfabetizados formalmente, pois descobrem um novo propósito para a leitura e a escrita: acessar o que lhes interessa.

O resultado dessas mudanças é que o processo tradicional de ensino e aprendizagem, orientado em conteúdos e centrado no professor, vem se tornando cada vez mais

ineficaz e entediante para essa nova geração [Dicheva et al. 2015]. Por essa razão, os processos de ensino e aprendizagem ativos e centrados no estudante, como a aprendizagem baseada em problemas, e o uso de mecânicas e dinâmicas de jogos, através da gamificação, vem recebendo cada vez mais atenção da comunidade acadêmica em razão do seu potencial de aumentar o engajamento e promover aprendizagem ativa.

Pela perspectiva dos docentes, um dos problemas práticos é que apesar desses novos paradigmas fazerem todo o sentido, seu uso no planejamento e na execução de cursos não lhes é natural, pois toda sua experiência como estudante foi orientada a conteúdos, com pouca evidência da aplicabilidade prática e nenhuma preocupação com a diversão.

O ponto é que seja qual for o método utilizado para a concepção de qualquer curso, sempre será possível aperfeiçoá-lo. Entretanto para identificar as oportunidades de melhoria mais importantes ou urgentes, é necessário colher dados e fazer análises. Nesse contexto, o *e-learning* oferece uma vantagem em relação ao ensino presencial uma vez que por ser intermediado por tecnologia, é possível coletar automaticamente diversas informações sobre o comportamento, as atividades e as interações entre estudantes, professores, tutores.

Diante disso, este artigo apresenta uma abordagem de mineração de dados, para verificação de quais são as atividades que mais demandam esforço dos estudantes e agrupar os erros mais frequentes, aqui chamados de “gargalos”. Com isso espera-se oferecer subsídio aos tutores e conteudistas para a execução de melhorias nos cursos que promovam a redução do tempo e do esforço dedicado pelos estudantes e tutores, sem prejuízo.

A abordagem foi desenvolvida e aplicada em uma disciplina do curso de engenharia de controle e automação, da Universidade de Pernambuco, que é realizada de forma semi-presencial. Nessa disciplina não há aulas expositivas. O estudante tem acesso a uma plataforma online que combina realidade virtual e gamificação, na qual recebe demandas para desenvolver projetos, juntamente com os conteúdos, especificações, orientações e dicas. Nas interações presenciais, o professor tira dúvidas dos alunos, discute as soluções apresentadas e assiste apresentações de projetos de grupos da turma.

## **2. Fundamentação Teórica**

Com o crescimento do uso dos Ambientes Virtuais de Aprendizagem (AVA), surge também uma grande quantidade de dados gerados por eles. Com a utilização de técnicas de mineração de dados, é possível obter informações e saber a real situação do curso nos ambientes e entender o comportamento dos alunos e como eles reagem às atividades. Neste trabalho, serão aplicadas técnicas para detectar as atividades em que os alunos mais cometem erros e os tipos de erro mais frequentes, que são considerados “gargalos” no processo de ensino e aprendizagem.

### **2.1. Mineração de Dados Educacionais**

Dados gerados a partir do processo de ensino e aprendizagem são minerados de forma diferente dos que são gerados em outras indústrias. Portanto, Mineração de Dados Educacionais foca em entender e resolver problemas relacionados exclusivamente à educação [Dutt et al. 2017]. Para [Abuteir e El-Halees 2012], há uma grande demanda no uso de dados educacionais e sua mineração pode favorecer as instituições de ensino e melhorar cursos ministrados em seus ambientes.

O tipo de dado comumente utilizado em Mineração de Dados Educacionais são os *logs* gerados a partir do uso das plataformas de ensino. A partir deles, é possível coletar e prever informações como nota dos alunos [Minaei-Bidgoli et al. 2003], participação [Beer 2010] ou reprovação [K. Looi et al. 2005]. Porém, analisar esses *logs* não é uma atividade fácil, pois existe uma grande quantidade de informação disponível, mas de forma não tratada [Soares et al. 2016]. Sendo assim, é preciso coletar e preparar os dados, para obter informações úteis aos educadores e estudantes.

## 2.2. Clusterização

Uma das técnicas de clusterização mais utilizada em Mineração de Dados Educacionais é Self-Organizing Map (SOM) de [Kohonen 1990]. O algoritmo é um tipo de Redes Neurais Artificiais com aprendizagem não-supervisionada e tem a finalidade de agrupar vetores que possuem similaridade entre si. Cada neurônio da rede possui um peso que representa um vetor com a mesma dimensão dos atributos dos dados. Durante o treinamento, o algoritmo possui uma abordagem baseada na competição entre os neurônios e um dos critérios para definir o vencedor é a distância euclidiana dos pesos calculada em relação a cada vetor de entrada. No final do treinamento, cada neurônio irá representar um *cluster*, a partir do deslocamento dos seus respectivos pesos durante o aprendizado, em um mapa bidimensional que englobará conjuntos distintos dos dados.

A literatura contempla diversos trabalhos nos quais o SOM foi aplicado. Em [Sahu et al. 2014], a técnica foi utilizada para inspirar novos algoritmos para mineração de sequência usando *web logs*. Foi utilizada uma nova abordagem chamada “Frequent Sequential Traversal Pattern Mining” ou FSTSOM, para tratamento de problemas ordenados. Em [Bara et al. 2018], foi utilizado o SOM para agrupar alunos de acordo com suas similaridades na interação com a plataforma e foi realizada análises sobre a correlação entre os comportamentos de aprendizagem dos alunos e o grau de sucesso que cada um alcançou.

Além disso, outras técnicas de clusterização foram utilizadas para minerar dados educacionais. No trabalho de [Kausar et al. 2018], por exemplo, foi aplicada uma abordagem adaptativa, que consiste em agrupar estudantes com base em seu comportamento de aprendizagem, bem como detectar que tipo de conteúdo é melhor para cada perfil identificado. Além de testar técnicas já existentes de clusterização, foi demonstrado que a abordagem implementada foi eficiente em processar grandes dados e fornecer informações que serviram de subsídio para administradores e professores tornar plataformas de *e-learning*s mais robustas e adaptáveis.

Em outro trabalho, [Benmakrelouf et al. 2015] utilizou técnicas de clusterização para identificar perfis dos jogadores dentro de um jogo. Com a identificação e seleção de variáveis, foi estudada a correlação entre as características dos jogadores para o agrupamento. O método mostrou três perfis de participação dos jogadores: iniciante, intermediário e avançado, que interagiram com o jogo de acordo com sua experiência.

## 2.3. O Erro como Recurso Didático

O erro pode ser utilizado como uma ferramenta pedagógica. É importante que o estudante não veja a sua falha como algo permanente [De Paula e Valente 2016]. O erro deve ser considerado algo transitório, um passo para o aprendizado, e que será necessário mais

estudos para que possa chegar ao resultado esperado. [De Paula e Valente 2016] ainda afirmam que é possível aprender a partir de erros já cometidos, desde que haja uma análise daquela falha em questão. O que significa que se for avaliado o como e o porquê de sua falha, o estudante poderá adquirir conhecimentos preciosos.

[Chandra e Nandhini 2010] afirmam que o fracasso do aluno em determinada atividade não aponta apenas para a capacidade do estudante em cumprir aquele objetivo. Isso indica que o erro não é apenas uma informação útil para o estudante, mas pode ser utilizado também para discutir as estratégias pedagógicas utilizadas e quais deverão ser utilizadas no futuro [Cury e Konzen 2007]. Sendo assim, caso diversos alunos estejam errando em determinado ponto do exercício ou até mesmo não conseguindo concluir, podemos ter uma indicação que houve uma falha no ensino. Então os professores precisarão avaliar qual o principal motivo do erro do aluno, para que possa identificar onde está a possível falha.

Essas falhas e gargalos que podem aparecer durante o processo também afetam a qualidade do ambiente virtual. Uma grande quantidade de falhas representa uma falha no curso [Chandra e Nandhini 2010]. Isso significa que os administrados dos Ambientes Virtuais de Aprendizado devem ficar atentos às falhas cometidas pelos seus estudantes, para que possam aprimorar o ambiente.

### 3. Metodologia

Tendo em vista o problema de pesquisa (reduzir o esforço dedicado por estudantes e professores em cursos gamificados na modalidade *e-learning*) e o artefato (abordagem para indicação das etapas com maiores índices de erros e as formas mais frequentes como os alunos erram) optou-se pela aplicação do CRISP-DM (Cross Industry Standard Process for Data Mining) para orientar a execução do projeto com clusterização através de SOM. A aplicação do CRISP-DM em suas várias etapas está descrita nos subitens a seguir.

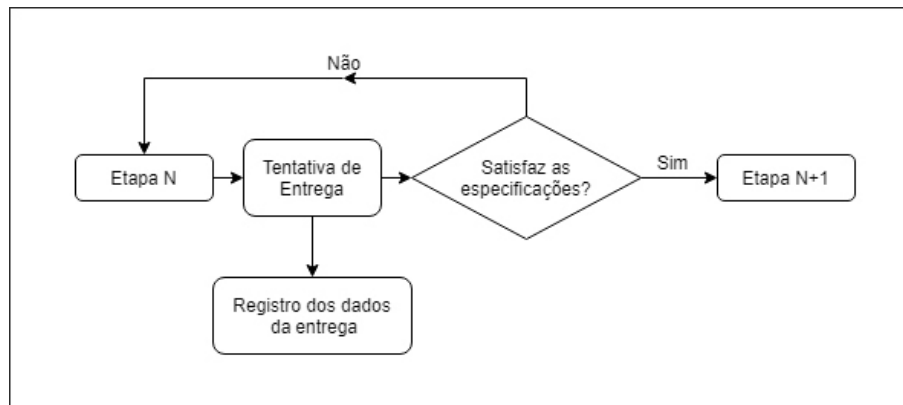
#### 3.1. Entendimento do Negócio

A base de dados utilizada neste projeto foi extraída da Fábricasim: uma plataforma de *e-learning* gamificada, com ambientes industriais virtuais, para ensino de desenvolvimento de sistemas de automação industrial. Ela foi desenvolvida pela *startup* Mekatronik, apoiada pela FACEPE, pela FINEP e pelo CNPQ com recursos de subvenção econômica à inovação. Atualmente ela é utilizada em disciplinas da graduação e da pós-graduação lato-sensu da Universidade de Pernambuco.

Nos cursos realizados na Fábricasim o estudante é imerso em uma história em que ele é *trainee* da equipe de engenharia da empresa (a própria Fábricasim). Lá ele aprende a desenvolver sistemas de automação para máquinas e processos a partir de desafios, que são pequenos projetos. Em cada projeto estão contempladas a especificação do que deve ser feito, as orientações, os conteúdos e as dicas para o desenvolvimento. Além disso, os projetos são divididos em etapas menores, com entregas funcionais tangíveis.

Sempre que entende que concluiu a implementação de uma etapa, o estudante solicita ao “operador virtual” que teste a solução entregue. Nesse momento, o “operador virtual” testa em tempo real se a máquina virtual está funcionando de acordo com as especificações e apresenta o *check-list* do teste. Em seguida ele informa ao estudante se a entrega foi realizada com sucesso ou se ele precisa de revisão. A Figura 1 ilustra esse

processo. A partir dela pode-se verificar que sempre que o aluno entende que implementou a solução de acordo com as especificações do professor, ele pode tentar entregá-la, e cada tentativa é registrada no sistema.



**Figura 1. Processo de submissão de Entregas do estudante.**

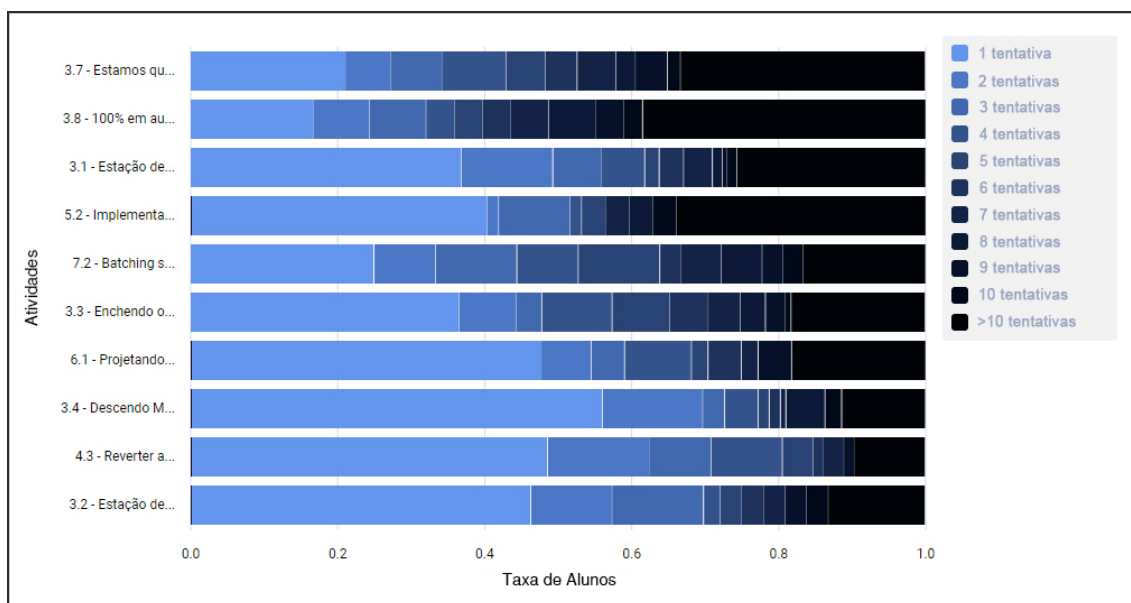
O problema de negócio que é objeto desta pesquisa consiste em reduzir o esforço dos estudantes na implementação das soluções para cada etapa, bem como dos professores no esclarecimento de dúvidas. A saída do processo de mineração é, portanto, a recomendação sobre quais etapas dos projetos envolvem mais tentativas de entrega dos estudantes até alcançarem a solução correta e quais os erros mais frequentemente cometidos por eles. Com isso, entende-se que o professor poderá priorizar as ações de melhoria nos cursos e promover o avanço mais rápido dos estudantes sem prejuízo à aprendizagem.

### 3.2. Entendimento dos Dados e Análise Descritiva

O Banco de Dados da Fábricasim é relacional e é composto por 51 tabelas. De todas as tabelas, serão utilizadas apenas 4 para este trabalho:

- Experiência - registro de todas as atividades dos alunos: acesso a cursos, acesso a projetos, etapas e tentativas de entrega. Ela possui chave estrangeira de todas as tabelas seguintes;
- Soluções - soluções corretas para cada etapa;
- Entregas - lista de eventos que representam cada tentativa de entrega realizada pelos estudantes;
- Atividades - cadastro das atividades dos projetos propostos aos estudantes.

Através da análise do gráfico da Figura 2, pode-se verificar as dez primeiras atividades com maior incidência de erros cometidos pelos estudantes. Nesta etapa, o propósito é identificar e priorizar a atividade em que os alunos realizam um maior número de tentativas antes de entregar da forma correta. Na atividade 3.8, por exemplo, verifica-se que um pouco mais de 20% dos alunos conseguiram entregar com sucesso em apenas 1 tentativa, enquanto um pouco menos de 40% deles precisaram tentar mais de 10 vezes. Sendo assim, a partir da hipótese de que a atividade 3.8 é a que exige o maior número de tentativas dos estudantes, decidiu-se estudar a forma como os estudantes erram com propósito de identificar o maior “gargalo” da atividade para realizar melhorias.



**Figura 2. Porcentagem de Estudantes por Tentativas de Entrega nas Etapas dos Projetos.**

### 3.3. Pré-processamento e Modelagem

O objetivo desta etapa do trabalho é realizar a clusterização de tipos diferentes de erros na atividade analisada e priorizada, a fim de descobrir “gargalos” que impedem os estudantes de entregá-la. A atividade “3.8 - 100% em Automático” foi a priorizada para aplicação da abordagem. Essa é a última atividade do desafio proposto para automatizar o processo completo de uma máquina de dosagem de gesso. Dentre os dados coletados de cada entrega da atividade, está a sequência de eventos que são registrados no processo de funcionamento da máquina e servirá de subsídio para a clusterização.

O desafio do pré-processamento foi adequar os dados relacionados às entregas dos alunos para que pudessem ser agrupados pelo algoritmo. Para isso foi extraída do banco uma tabela com todas as entregas sem sucesso dos alunos, associadas a atividade. A entrega, composta por uma sequência de eventos relacionados ao funcionamento da máquina virtual, é representada por dados categóricos que precisam ser transformados em dados numéricos, através de uma codificação para que sirvam de entrada para o algoritmo de clusterização.

Para melhor entendimento, assumamos  $S = \{ A, B, C, D, E, F, G, X \}$  como a sequência da solução correta para uma etapa e  $E = \{ A, B, C, E, X, F \}$ , a sequência da entrega do estudante. Essas informações categóricas foram transformadas em numéricas a partir da construção de um vetor  $Ev$  cujo número de dimensões é a quantidade de eventos de  $S$ . Nesse vetor, cada dimensão contempla a distância entre as posições dos eventos em  $S$  e em  $E$ . Quando um evento de  $S$  não existe em  $E$  a dimensão correspondente assume valor -1. Sendo assim,  $Ev = [ 0, 0, 0, -1, 1, 0, -1, 3 ]$ , uma vez que os eventos  $A, B, C$  ocorreram na sequência correta, o evento  $D$  não ocorreu, o evento  $E$  estava defasado em 1 posição, o evento  $F$  ocorreu na posição correta, o evento  $G$  não ocorreu e o evento  $X$  ocorreu defasado em 3 posições. A interpretação da transformação é que o vetor  $Ev$  contempla a distância euclidiana entre  $S$  e  $E$ .

Uma vez que as entregas dos alunos passaram a ser representadas por vetores, foi possível agrupá-los de acordo com sua proximidade, através da utilização do SOM.

#### 4. Resultados

A Figura 3 contempla o mapa de *clusters* gerado no treinamento para todas as tentativas de entrega da etapa priorizada. Nele, é possível observar que o maior *cluster* (7, 2) tem ocorrência 62,85% maior que o segundo lugar (0, 2). Além disso os três maiores *clusters* somados correspondem a mais de 50% dos erros cometidos pelos alunos.

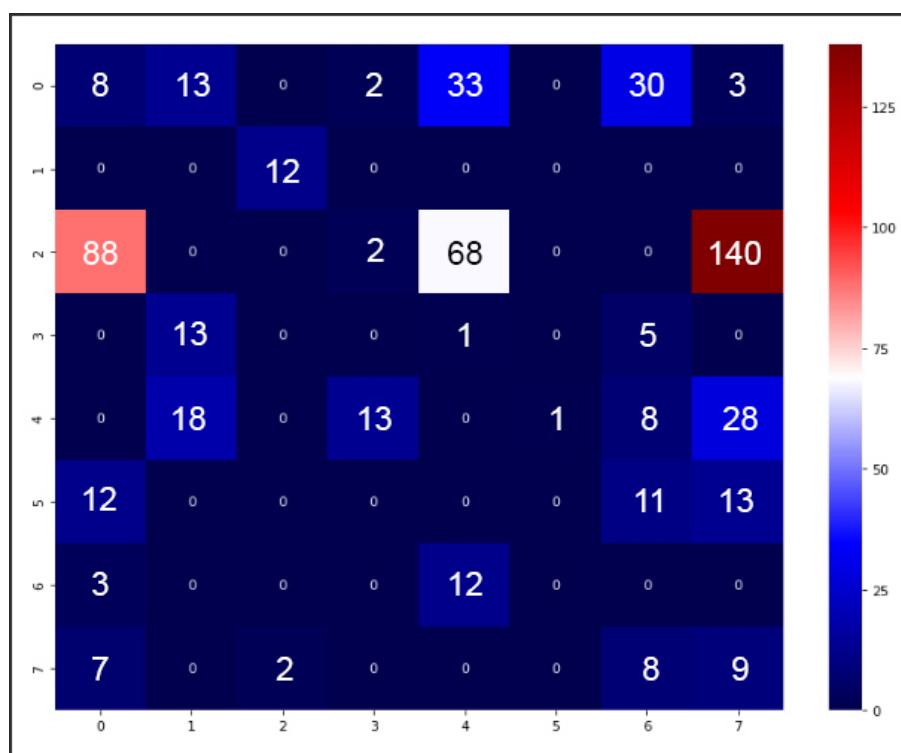


Figura 3. Mapa gerado pelo algoritmo SOM.

Diante disso, foi recomendada a revisão da atividade 3.8 do curso, com foco no problema descrito pela entrega do *cluster* (7, 2) e entregue as sequências de eventos correspondente a esse *cluster*. Ao analisar a sequência, o professor verificou que o erro cometido pelos estudantes era decorrente de uma especificação que não estava sendo corretamente interpretada pelos alunos. O problema foi encontrado no conteúdo da atividade, que havia uma orientação para fazer o aproveitamento da implementação da atividade anterior, mas que deveria fazer algumas alterações antes de implementar a solução nova. Essa orientação estava sendo confundida por vários estudantes e atrasando a entrega da atividade. A ação realizada pelo professor foi melhorar a especificação da etapa, para que não houvesse mais equívocos na interpretação.

#### 5. Conclusão e Trabalhos Futuros

Este trabalho propôs o desenvolvimento de uma abordagem de mineração para ajudar professores a identificarem atividades-gargalos em cursos gamificados e realizar intervenções de melhorias. A partir dos resultados apresentados, verifica-se que o objetivo

foi alcançado uma vez que a recomendação feita com base nas informações mineradas, efetivamente ajudou o professor a intervir na atividade que demandava maiores esforços dos alunos.

Esse trabalho se limitou apenas em identificar “gargalos” em cursos propostos em *e-learning*s através dos erros dos estudantes, não focando em outras informações, como por exemplo, perfil de aprendizagem. Oportunidades para trabalhos futuros incluem associar os tipos de erro à outras informações dos estudantes, com o propósito de personalizar dicas, recomendações e conteúdos. Além disso, pode-se explorar outros algoritmos de Mineração de Dados Educacionais, com o intuito de comparar o desempenho e obter melhores informações sobre as atividades dos estudantes.

## Referências

- Abuteir, M. and El-Halees, A. (2012). Mining educational data to improve students' performance: A case study. *International Journal of Information and Communication Technology Research*, 2:140–146.
- Bara, M. W., Ahmad, N. B., Modu, M. M., and Ali, H. A. (2018). Self-organizing map clustering method for the analysis of e-learning activities. In *2018 Majan International Conference (MIC)*, pages 1–5.
- Beer, C. (2010). *Online student engagement: New measures for new methods*. PhD thesis, Central Queensland University.
- Benmakrelouf, S., Mezghani, N., and Kara, N. (2015). Towards the identification of players' profiles using game's data analysis based on regression model and clustering. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1403–1410.
- Chandra, E. and Nandhini, K. (2010). Knowledge mining from student data. *European Journal of Scientific Research*, 47(1):156–163.
- Cury, H. and Konzen, B. (2007). Uma aplicação de jogos na análise de erros em educação matemática. *Revemat: Revista Eletrônica de Educação Matemática*, 2(1):107–117.
- De Paula, B. and Valente, J. (2016). Errando para aprender: a importância dos desafios e dos fracassos para os jogos digitais na educação. *RENOTE*, 13.
- Dicheva, D., Dichev, C., Agre, G., and Angelova, G. (2015). Gamification in education: A systematic mapping study. *Educational Technology & Society*, 18:75–88.
- Dutt, A., Ismail, M. A., and Herawan, T. (2017). A systematic review on educational data mining. *IEEE Access*, 5:15991–16005.
- K. Looi, C., McCalla, G., Bredeweg, B., and Breuker, B. (2005). Artificial intelligence in education: Supporting learning through intelligent and socially informed technology. *IEEE Transactions on Pattern Analysis and Machine Intelligence - PAMI*.
- Kausar, S., Huahu, X., Hussain, I., Wenhao, Z., and Zahid, M. (2018). Integration of data mining clustering approach in the personalized e-learning system. *IEEE Access*, 6:72724–72734.
- Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480.



- Minaei-Bidgoli, B., Kashy, D. A., Kortemeyer, G., and Punch, W. F. (2003). Predicting student performance: an application of data mining methods with an educational web-based system. In *33rd Annual Frontiers in Education, 2003. FIE 2003.*, volume 1, pages T2A–13.
- Sahu, S., Saurabh, P., and Rai, S. (2014). An enhancement in clustering for sequential pattern mining through neural algorithm using web logs. In *2014 International Conference on Computational Intelligence and Communication Networks*, pages 758–764.
- Soares, F., José Richard Machado, C., Diniz, D., Maciel, A., and Rodrigues, R. (2016). Educational data mining to support distance learning students with difficulties in the portuguese grammar. *Brazilian Symposium on Computers in Education*, 27(1):956.