

Uma Revisão Sistemática da Literatura para Identificação de Lacunas de Pesquisa Relacionadas a Mineração de Processos Educacionais

Lucielton Manoel da Silva¹, Rodrigo Lins Rodrigues¹, João Carlos Sedraz Silva², Jorge Luis Cavalcanti Ramos², Gibran Medeiros Chaves de Vasconcelos²

¹Universidade Federal Rural de Pernambuco - UFRPE, ²Universidade Federal do Vale do São Francisco - UNIVASF

Abstract. *This paper has as objective present the state of art of Educational Process Mining(EPM) and the application of yours results in Learning Management Systems (LMS), to do this was made a systematic review of literature(SRL), where through a well defined and objective process was selected articles to serve as base to the SRL. After the definition of the protocol and the realization of the methodological process we got nine articles that contributed satisfactorily to the responses of research questions related with the studied phenomenon, where was analisade many aspects of them to get informations about the current state of arte of EPM.*

Palavras-chave: Mineração de Processos, Visualização de Dados, LMS.

Resumo. *Este artigo teve como finalidade apresentar o estado da arte em Mineração de Processos Educacionais (MPE) e a aplicação de seus resultados em Learning Management System (LMS), para tal foi realizada uma Revisão Sistemática da Literatura (RSL), onde através de um processo bem definido e rigoroso foram selecionados artigos que contribuíram para o entendimento do fenômeno de pesquisa abordado neste trabalho. Após a definição do protocolo e a realização do processo metodológico foram obtidos nove artigos que contribuíram de forma satisfatória para responder às questões de pesquisa relacionadas ao fenômeno estudado, onde foram analisados diversos aspectos dos mesmos a fim de obter informações acerca do atual estado da MPE.*

Keywords: Process Mining, Data Visualization, LMS.

1. Introdução

Atualmente tem sido frequente a utilização dos sistemas de gestão de ensino, conhecidos como LMS (do inglês “*Learning Management System*”), possibilitando capturar dados da interação dos usuários com o próprio ambiente. Essa quantidade de dados é gerada de forma massiva, impossibilitando a realização de análises manuais por parte de professores e gestores, para isso se aplicam técnicas de mineração de dados. Este tipo de técnica é comumente utilizado para gerar modelos preditivos e descritivos de uma grande quantidade de dados oriundas de sistemas LMS(do inglês “*Learning Management System*”) e Ambiente Virtuais de Aprendizagem (Manhães et. al., 2011), porém as técnicas comuns de mineração de dados tendem a utilizar apenas dados brutos quantificados, não permitindo compreender o

processo educacional por completo, para isso um subcampo da mineração de dados foi criado, intitulado como mineração de processos educacionais (Bogarín, Cerezo and Romero, 2018).

A área de mineração de dados educacionais vem despertando a atenção de pesquisadores de diversas áreas, contudo grande parte das pesquisas desenvolvidas na área não abordam resultados obtidos utilizando técnicas de mineração de processos, como também não discorrem sobre as implicações que os resultados poderiam levar, como por exemplo, a possibilidade de um aluno visualizar seu comportamento dentro do sistema e o comportamento dos demais alunos da mesma turma, assim permitindo-o melhorar seu próprio comportamento e obtendo um aumento de desempenho educacional.

Este trabalho tem como intuito a realização de uma revisão sistemática da literatura, buscando o entendimento, com profundidade, de um fenômeno de interesse que produz resultados específicos e detalhados por meio da análise de conteúdo e qualidade do material pesquisado (Kitchenham et. al., 2009), onde buscou-se artigos em que fossem gerados formas visuais dos resultados da mineração de processos educacionais.

2. Trabalhos relacionados

Nesta seção definimos processos de aprendizagem e mineração de processos educacionais, e em seguida é apresentado o método utilizado no protocolo desta RSL, apresentando às questões de pesquisa, critérios de inclusão e exclusão e por fim o processo da extração dos dados.

2.1. Processos de Aprendizagem em Plataformas LMS

Com o auxílio de plataformas LMS e Ambientes Virtuais de Aprendizagem, ficou viável a captura da interação dos usuários com este tipo de ferramenta, estas interações são armazenadas em *logs* de dados que posteriormente são analisados e interpretados para se formar o entendimento do processo educacional de um aluno. Este processo refere-se a maneira ao qual um aluno se comporta dentro de um ambiente de aprendizagem, seja uma sala de aula ou um ambiente virtual.

“Usando métodos de mineração de processos, identificamos tipos específicos de sequências de interações que lançam uma nova luz sobre as estratégias para resolução de questionário dos alunos no LMS [...] usando métodos de mineração de processos que podem ser muito benéficos para o monitoramento e análise de processos de aprendizagem (cognitivos)” (Libor Juhanak et al. 2017).

Através da análise de como o aluno está se comportando, é possível observar suas diferenças comportamentais em relação demais alunos da classe, bem como identificar o que está prejudicando sua aprovação, assim a análise dos processos educacionais fornece a possibilidade do professor intervir e ajudar os alunos com baixa taxa de aprovação para que estes atinjam um crescimento satisfatório, apenas modificando seu processo.

2.2. Mineração de Processos Educacionais

A mineração de processos educacionais é um subcampo da mineração de dados educacionais (MDE). Como a MDE analisa os dados brutos gerados por ambientes virtuais de aprendizagem, sobretudo sistemas LMS, a MPE é um campo mais especializado que trata de analisar os *logs* de dados de sistemas LMS com o intuito de construir o processo educacional, seja este processo a estruturação de um curso à distância ou até mesmo a maneira que o aluno realizou as atividades.

Segundo Aalst (2012, p.11) “A Mineração de Processos é uma ferramenta importante para organizações modernas que necessitam gerenciar processos operacionais não triviais”.

As técnicas de MPE tem a finalidade de: (1) Descobrir modelos de processos educacionais completos, capazes de expressar o comportamento dos usuários; (2) Analisar a conformidade, verificando se um dado comportamento modelado corresponde ao comportamento observado; (3) Melhorar os processos, extraindo conhecimento para facilitar a compreensão dos problemas e das oportunidades de otimização (Trcka e Pechenizkiy 2009).

3. Método Utilizado

Uma Revisão Sistemática da Literatura é uma maneira de buscar, extrair, catalogar, analisar e sintetizar dados dos artigos de congressos, jornais e periódicos (Silva, Medeiros e Aranha 2014) às pesquisas relacionadas à uma determinada área de pesquisa, e com isto entender o que está sendo produzido de conhecimento científico na área.

O objetivo principal desta RSL foi analisar os estudos relevantes dentro da Mineração de Processos Educacionais, tendo como objetivo secundário verificar técnicas de MPE, sobretudo relacionado à variáveis de comportamento, e também técnicas de visualização de MPE, a aplicação e também os resultados destas técnicas no contexto educacional tentando observar como são utilizados seus resultados, visando extrair seus possíveis benefícios dentro de *Learning Management Systems* (Sistemas de Gestão de Aprendizado).

3.1 Questões de pesquisa

A RSL visou responder às Questões de Pesquisa (QP) descritas na tabela 1. A busca de artigos nesta RSL foi realizada no período de tempo entre 1º janeiro de 2008 e 1º de janeiro de 2019. A questão de pesquisa principal foi: *A visualização dos comportamentos de estudantes dentro de sistemas LMS impactam na melhoria acadêmica?*

Foram definidas 9 questões derivadas ao tema, que podem ser observadas na tabela 1, cujo o propósito foi entender as técnicas de construção de formas visuais para MPE, as técnicas de análise de dados da MPE, de onde os dados da MPE são comumente extraído.

Questões	Descrição
QP1	Quais são as formas de visualização disponíveis para <i>logs</i> de eventos em plataformas educacionais?
QP2	Quais variáveis são relevantes para construir o caminho de aprendizagem dos alunos?
QP3	Quais são as técnicas mais utilizadas na mineração de processos educacionais?
QP4	Quais as formas de visualização são mais utilizadas?
QP5	Quais os instrumentos de coleta dos dados?
QP6	Qual o público alvo da pesquisa?
QP7	Com qual objetivo o estudo foi realizado?
QP8	Como a visualização do processo educacional dos alunos impactam sua taxa de aprovação?

Tabela 1. Questões de Pesquisa

Estas questões foram geradas com o intuito de facilitar o entendimento do atual estado da MPE, pode-se ver na tabela 1, que além de questões relacionadas aos problemas, temos questões relacionadas quanto a própria pesquisa, visando melhorar a compreensão da finalidade das mesmas.

3.2 Critério de Exclusão e Inclusão

Os critérios foram definidos para retirar artigos que não abordassem técnicas de MPE, mas sim frameworks, softwares já prontos ou propostas de software, artigos que fugissem da área da educação ou não tivessem alguma relação com educação ou sistemas LMS, não tivessem relação alguma com mineração de dados ou mineração de processos, não fossem estudos primários, como outras revisões, surveys e manuscritos, e também retirar os estudos que estivessem em línguas diferentes do inglês ou português, para evitar vieses de tradução, e foram selecionados artigos que estivessem no período estabelecido entre 1º de janeiro de 2008 e 1º de janeiro de 2019 e fossem estudos primários com artigo completo. A aplicação destes critérios foi inicialmente feita no *abstract* e título dos artigos e após este filtro inicial, os critérios passaram a ser aplicados no artigos completos.

3.3 Processo de Extração de Dados

O processo de busca dos estudos primários se deu pela utilização dos principais engines de busca, os quais foram utilizados: *IEEE Explorer*, *Springer Link* e *Science Direct*. As palavras chaves utilizadas na *string* de busca para a obtenção dos estudos primários foram: *Process mining* e *Data mining* que representam o fenômeno de pesquisa, *Education* e *Learning management system* representando a área de aplicação, *Visualization* para se obter técnicas relacionadas a visualização em MPE e *Behaviour* representando a variável estudada. A *string* de busca também abrangia sinônimos de cada uma das palavras, para se obter o máximo de artigos possíveis, estes sinônimos foram oriundos dos próprios resultados e reformulações. Assim, a *string* de busca utilizada foi:

```
("Process mining" OR "Paths of learning" OR "Path of learning" OR "EPM" OR  
"Educational Data Mining" OR "EDM") AND ("Education" OR "Educational") AND(  
"Learning Management System" OR "LMS") AND "Visualization" AND ("Behaviour"  
OR "Behavior")
```

Quadro 1: String de busca

Com esta *string* de busca foram realizadas a procura pelos artigos que serão utilizados como base para a revisão sistemática.

4. Resultados e Discussões

Nesta seção são apresentados os resultados obtidos desde o início do processo da RSL, passando pelo filtro dos critérios de exclusão, os motivos pelo qual artigos foram retirados e em seguida são abordadas as questões de pesquisa com suas respostas já extraídas e devidamente discutidas em cada tópico.

4.1 Resultado das Buscas

Após a aplicação da *string* de busca, inicialmente obtivemos 254 artigos, sendo a contribuição de cada engine dada por IEEE explorer contribuiu com 0 artigos, Springer

Link contribuiu com 147 e Science Direct contribuiu com 107 artigos.

Após uma rápida verificação, percebeu-se a existência de artigos duplicados dentro dos engenhos e após a retirada destes, o resultado final foi de 203 artigos, após isto, foi aplicando os critérios de inclusão e exclusão, onde inicialmente foi aplicado no *abstract* dos artigos para se ter uma diminuição dos artigos totais a serem analisados por completo, após a aplicação dos critérios já descritos na seção 3.2, obtivemos 178 artigos excluídos, ficando com o total de 25 artigos para análise completa.

Nos passos seguintes, foi aplicado novamente os critérios de inclusão e exclusão, após uma leitura minuciosa dos artigos obtidos, ao final desta aplicação foram extraídos 16 artigos, ficando no final com 9 artigos que contribuíram totalmente com esta revisão sistemática. O fluxograma x apresenta uma ilustração do processo realizado e os respectivos resultados.

A seguir um gráfico mostrando a percentagem da inclusão dos artigos excluídos em cada critério de exclusão correspondente.

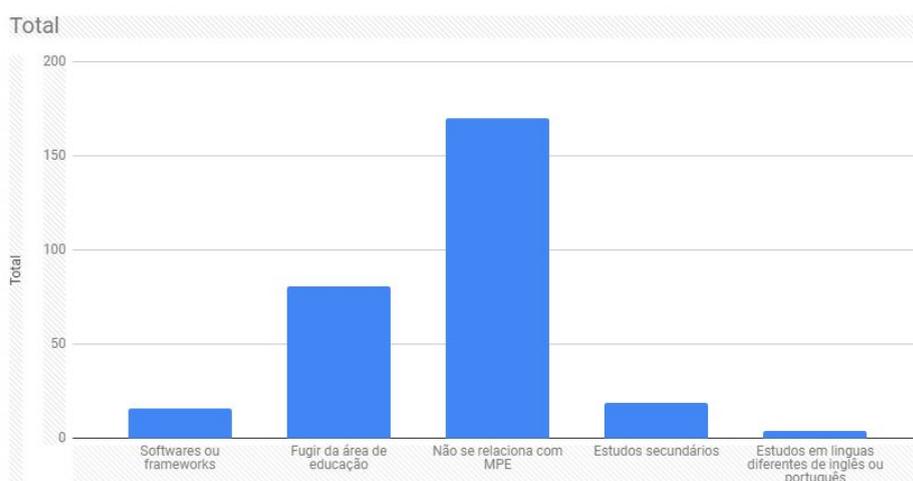


Gráfico 1. Total de artigos excluídos por categoria

Com os resultados do gráfico 1 pode-se observar os motivos que levaram a exclusão do restantes dos artigos resultados pela string de busca

4.2 Resultados relacionados às questões de pesquisa

Após os processos de captura e aplicação dos critérios de exclusão, foram lidos os 9 artigos e aqui são discutidas as respostas das questões de pesquisa definidas na seção 2.1 e devidamente são discutidos seus resultados.

QP1: Quais são as formas de visualização disponíveis para logs de eventos em plataformas educacionais?

Nesta pergunta buscamos às técnicas de construção de gráficos de mineração de processos mais utilizadas na literatura, tentando entender quais técnicas são as mais tradicionais e quais técnicas estão sendo utilizadas recentemente. Dos 9 artigos observados apenas 4 forneceram alguma forma de visualização dos dados obtidos pelos logs, os estilos de gráficos utilizados foram gráficos em barras, que foi observado em (Konomu Dobashi 2015), gráficos em linhas observados em (Arnon Hershkovitz, Rafi Nachmias 2011), gráficos em séries temporais utilizados em (Konomu Dobashi 2016) e (Konomu Dobashi 2017).



Gráfico 2. Técnicas de visualização ao longo do tempo

Nem todos os artigos se apropriaram de técnicas gráficas em sua base de dados, porém dos artigos que se utilizaram de tais técnicas percebeu-se que para a análise dos logs de dados está em ascensão o uso de gráficos de séries temporais e que os gráficos em linhas e em barras, aparecem apenas uma vez, mostrando que os gráficos de séries temporais são os mais recomendados para se mostrar os dados de logs.

QP2: Quais variáveis são relevantes para construir o caminho de aprendizagem dos alunos?

Aqui nós queremos obter às variáveis dentro dos logs de dados que mais são utilizadas para a construção do processo de aprendizagem do aluno e que expressam bem esta característica. Para cada um dos 9 artigos lidos teve um conjunto de variáveis diferentes, dentre estes artigos presentes nesta revisão não houveram muitas semelhanças dentre as variáveis utilizadas, às variáveis utilizadas estão desde aquelas relacionadas às atividades, como atividades totais e acumuladas, apresentadas em (Arnon Hershkovitz, Rafi Nachmias 2011), data e tempo e histórico de visualização dentro do sistema LMS como em (Konomu Dobashi 2015), transição das páginas e tempo de navegação apresentados em (Konomu Dobashi 2016) e demais variáveis disponibilizadas em logs de dados.

Um dos fatores que influenciam a escolha das variáveis é a finalidade como será visto adiante às finalidades eram bem diferente entre si, portanto isto representa um viés a escolha das variáveis.

QP3: Quais são as técnicas mais utilizadas na mineração de processos educacionais?

Dentro da mineração de processos existem diversos algoritmos a serem utilizados, estes algoritmos não são próprios da mineração de processos, eles já vem da mineração de dados educacionais, porém é interessante observar a categoria de algoritmos que são utilizados. Foram percebidos grupos de técnicas relacionados a clusterização em (Libor Juhaňák, Jiří Zounek, Lucie Rohlíková 2017), (Shinichi Oeda, Genki Hashimoto 2017), outro relacionados a séries temporais em (Konomu Dobashi 2015), (Konomu Dobashi 2016) e (Konomu Dobashi 2017), outro relacionados a regressão em (Ya-Han Hu, Chia-Lun Lo, Sheng-Pao Shih 2014) e (Maria Theobald, Henrik Bellhäuser, Margarete Imhof 2018) e por último um grupo de algoritmos sobre classificação em (Arnon Hershkovitz, Rafi Nachmias 2011), (Ya-Han Hu, Chia-Lun Lo, Sheng-Pao Shih 2014), (Kingsley Okoye et. al. 2014) e (Libor Juhaňák, Jiří Zounek, Lucie Rohlíková 2017).

No gráfico abaixo observamos uma linha do tempo dos algoritmos utilizados, os algoritmos estão agrupados por categorias, onde cada linha mostra uma categoria diferente, no eixo Y está a quantidade de vezes que dado grupo de algoritmos apareceu nos artigos e no eixo X está o ano do artigo em que o algoritmo foi utilizado.

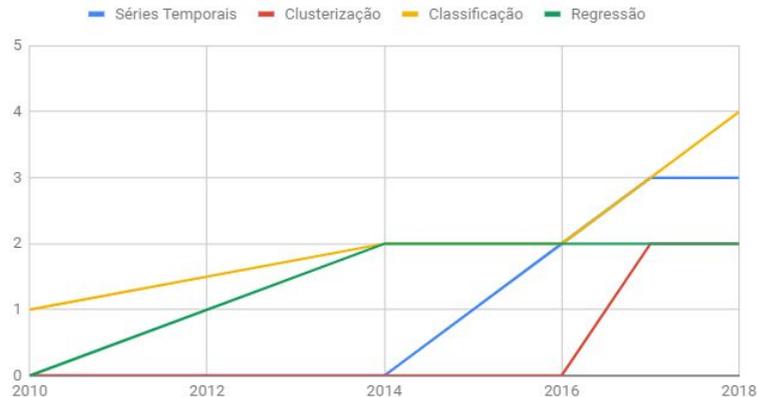


Gráfico 3. Técnicas utilizadas ao longo do tempo

Podemos observar que a categoria de algoritmos mais utilizados é a de classificação, isto também recebe forte influência do objetivo dos artigos, visto que a maioria deles visa fazer uma predição sobre os alunos.

QP4: Nível de reprodutibilidade do estudo?

Os artigos analisados receberam uma classificação quanto a sua reprodutibilidade, sendo classificados com reprodutibilidade baixa os artigos que apenas discorriam sobre o algoritmo utilizado e os resultados obtidos como foi percebido ao se observar os artigos (Kingsley Okoye et. al. 2014), (Konomu Dobashi 2015), (Konomu Dobashi 2016) e (Libor Juhaňák, Jiří Zounek, Lucie Rohlíková 2017), reprodutibilidade média foram os artigos que além de falar sobre o algoritmo utilizado e às técnicas, também eram bem detalhados quanto ao método utilizado na pesquisa isto foi percebido nos artigos (Arnon Hershkovitz, Rafi Nachmias 2011), (Ya-Han Hu, Chia-Lun Lo, Sheng-Pao Shih 2014), (Konomu Dobashi 2017), (Shinichi Oeda, Genki Hashimoto 2017) e (Maria Theobald, Henrik Bellhäuser, Margarete Imhof 2018), já reprodutibilidade alta, além do que foi falado anteriormente, foram os artigos que disponibilizaram a base de dados para se reproduzir a pesquisa totalmente.

QP5: Quais os instrumentos de coleta dos dados?

Como se trata de mineração de processos educacionais, não houve alguma variação percebida quanto o local de onde os dados eram obtidos, a mineração de processos em todos os artigos observados pegaram seu dados do log de dados de um sistema LMS o que não difere de forma alguma dos métodos utilizados na literatura, visto que este método de mineração comumente obtém seus dados de logs, a única diferença foi de qual plataforma esse log pertence, alguns artigos como (Arnon Hershkovitz, Rafi Nachmias 2011), (Konomu Dobashi 2015), (Konomu Dobashi 2016) e (Konomu Dobashi 2017) explicitaram que o sistema LMS utilizado foi o moodle, já sobre os demais artigos não explicitaram nem foi possível observar qual ambiente LMS foi utilizado.

QP6: Qual o público alvo da pesquisa?

Nos artigos relacionados a mineração de processos há um foco muito grande no estudante (Arnon HersHKovitz, Rafi Nachmias 2011), (Ya-Han Hu, Chia-Lun Lo, Sheng-Pao Shih 2014), (Kingsley Okoye et. al. 2014), (Libor Juhaňák, Jiří Zounek, Lucie Rohlíková 2017), (Shinichi Oeda, Genki Hashimoto 2017) e (Maria Theobald, Henrik Bellhäuser, Margarete Imhof 2018) e os artigos (Konomu Dobashi 2015), (Konomu Dobashi 2016) e (Konomu Dobashi 2017) tem como foco o professor, porém através desta questão nós queremos obter algum outro possível foco da pesquisa. No gráfico abaixo temos, além do alvo da pesquisa, estudantes e professores, o tipo de técnica utilizada quando se está tratando de algum alvo específico.

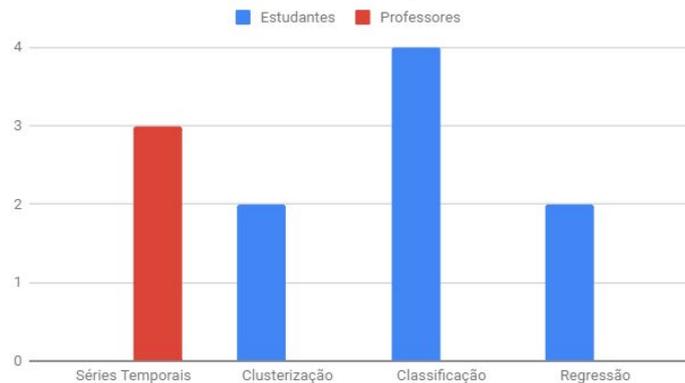


Gráfico 4. Relação entre o público alvo e técnicas

Vemos que séries temporais são muito utilizadas quando a intenção da mineração de processo se trata de estruturar o processo de ensino do professor, por outro lado vemos uma gama de categorias de técnicas utilizadas quando o intuito é montar o processo de educacional do estudante, com uma ressalva à classificação que está liderando na categoria de técnicas neste tipo de alvo, isto deve se relacionar aos artigos terem a finalidade de realizarem predição visto que a maioria dos artigos se tratam com este foco.

QP7: Com qual objetivo o estudo foi realizado?

Através desta variável analisamos o intuito das pesquisas analisadas. O intuito foi entender quais problemas estão tentando ser resolvidos com a mineração de processos educacionais, além disso houve a construção do gráfico 5, relaciona as metas encontradas nos artigos, com às técnicas utilizadas para a solução dos mesmos e a quantidade de vezes que estes aparecem.



Gráfico 5. Técnicas de MPE de acordo com o objetivo

Aqui apresentamos uma descrição de cada finalidade encontrado nos artigos lidos, junto com sua respectiva abreviação, esta que foi utilizada no gráfico anterior. Os objetivos desta RSL foram descritos abaixo:

1. Analisar a persistências dos usuários de LMS
2. Predição de performance
3. Visa guiar o professor para executar um bom blended learning
4. Melhoria no método de ensino
5. Comportamento e interações
6. Predição de evasão
7. Correlacionar notas dos alunos com outras variáveis

É possível perceber nos objetivos descritos acima a diferença entre os objetivos e que isto foi um fator que influenciou bastante às questões desta revisão, visto que algumas questões possuíam uma correlação, onde o objetivo tem forte influência sobre os algoritmos e público alvo analisado.

5. Conclusões

Esta RSL teve como objetivo mostrar a atual situação da MPE e também procurar nos artigos encontrados às aplicações desta técnica e seu uso dentro de ambientes educacionais. A busca ao todo retornou 254 artigos onde, ao se aplicar os critérios de exclusão estabelecidos, restaram apenas 9 artigos que foram devidamente analisados e seus dados foram a base para construção desta RSL. Tratamos dos artigos no período de 1º de Janeiro de 2008 a 1º de Janeiro de 2019, porém o menor ano encontrado nos artigos foi 2010 e o maior ano foi o de 2018, às buscas foram realizadas no engenho *IEEE explorer*, *Springer Link* e *Science Direct*, sendo o *Science Direct* o que mais contribuiu para a pesquisa.

A MPE é um subcampo da mineração de dados educacionais, ainda é um campo recentemente criado, contudo nota-se a importância desta técnica que pode possibilitar a construção de processos educacionais a partir de log de dados. Após a análise dos artigos foi possível perceber a categoria de algoritmos mais utilizada, além de poder relacionar estes com os intuitos e público alvo tratado, além disso, nota-se que a aplicação da MPE se dá em 2 metodologias, tanto completamente à distância, quanto em ensino híbrido, suas finalidades são bem diversas, mas seu público alvo são os agentes da sala de aula (professores e alunos), porém nenhum artigo tratou de utilizar ambos como alvo ao mesmo tempo e por fim, vemos

que a categoria de algoritmos mais utilizados para este tipo de técnica é a de algoritmos de classificação, e para técnicas visuais do log de dados, nota-se uma preferência por gráficos de séries temporais.

6. Agradecimentos

Os autores agradecem a Pró-Reitoria de Pesquisa e Pós-Graduação (PRPPG) e ao Núcleo de Inovação Tecnológica (NIT) pela oportunidade na disponibilização dos recursos de Iniciação em Desenvolvimento Tecnológico e Inovação, os quais foram essenciais para o desenvolvimento deste trabalho.

7. Referências

- Arnon Hershkovitz, Rafi Nachmias, Online persistence in higher education web-supported courses, *The Internet and Higher Education*, Volume 14, Issue 2, 2011, Pages 98-106.
- Aalst, W. V. D. (2012) "Process Mining: Overview and Opportunities". *ACM Transactions on Management Information Systems – TMIS*, v.3, n.2.
- Bogarín, A., Cerezo, R. and Romero, C. (2018), A survey on educational process mining. *WIREs Data Mining Knowl Discov*, 8: e1230.
- Kingsley Okoye, Abdel-Rahman H. Tawil, Usman Naeem, Rabih Bashroush, Elyes Lamine, A Semantic Rule-based Approach Supported by Process Mining for Personalised Adaptive Learning, *Procedia Computer Science*, Volume 37, 2014, Pages 203-210.
- Kitchenham, B.; Brereton, P.; Budgen, D.; Turner, M.; Bailey, J.; Linkman, S. (2009) "Systematic literature reviews in software engineering – A systematic literature review". In: *Information and Software Technology*, v.51, p.7-15.
- Konomu Dobashi, Time Series Analysis of the in class Page View History of Digital Teaching Materials using Cross Table, *Procedia Computer Science*, Volume 60, 2015, Pages 1032-1040.
- Konomu Dobashi, Development and Trial of Excel Macros for Time Series Cross Section Monitoring of Student Engagement: Analyzing Students' Page Views of Course Materials, *Procedia Computer Science*, Volume 96, 2016, Pages 1086-1095.
- Konomu Dobashi, Automatic data integration from Moodle course logs to pivot tables for time series cross section analysis, *Procedia Computer Science*, Volume 112, 2017, Pages 1835-1844.
- Libor Juhaňák, Jiří Zounek, Lucie Rohlíková, Using process mining to analyze students' quiz-taking behavior patterns in a learning management system, *Computers in Human Behavior*, 2017.
- Manhães, L. M. B.; Cruz, S. M. S.; Costa, R. J. M.; Zavaleta, J.; Zimbrão, G. (2011) "Previsão de Estudantes com Risco de Evasão Utilizando Técnicas de Mineração de Dados". In: *XXII Simpósio Brasileiro de Informática na Educação – SBIE*, p.150- 159.
- Maria Theobald, Henrik Bellhäuser, Margarete Imhof, Identifying individual differences using log-file analysis: Distributed learning as mediator between conscientiousness and exam grades, *Learning and Individual Differences*, Volume 65, 2018, Pages 112-122, ISSN 1041-6080.
- Shinichi Oeda, Genki Hashimoto, Log-Data Clustering Analysis for Dropout Prediction in Beginner Programming Classes, *Procedia Computer Science*, Volume 112, 2017, Pages 614-621.
- Silva, T. R.; Medeiros, T. J.; Aranha, E. H. S. (2014) "Jogos Digitais para Ensino e Aprendizagem de Programação: uma Revisão Sistemática da Literatura". In: *XXV Simpósio Brasileiro de Informática na Educação - SBIE*, p.692-701.
- Trčka, N.; Pechenizkiy, M. (2009) "From Local Patterns to Global Models: Towards Domain Driven Educational Process Mining". In: *Ninth International Conference on Intelligent Systems Design and Applications – ISDA*, Pisa, pp.1114-1119.
- Ya-Han Hu, Chia-Lun Lo, Sheng-Pao Shih, Developing early warning systems to predict students' online learning performance, *Computers in Human Behavior*, Volume 36, 2014, Pages 469-478.