



ARTIGO

# MINERAÇÃO DE PROCESSOS EDUCACIONAIS

POR

*Petrônio Cândido de Lima*  
[petronio.candido@ifnmg.edu.br](mailto:petronio.candido@ifnmg.edu.br)

Muito antes da pandemia da COVID-19 ser declarada, em fevereiro de 2020, a Educação à Distância já era uma tendência mundial. Contudo, foi durante o período de isolamento social da COVID-19 que vários atores do processo pedagógico - professores, tutores, alunos, gestores e pais - foram subitamente forçados a adotar as tecnologias da Educação à Distância sem o devido preparo, evidenciando que as particularidades dessa modalidade de ensino podem ser desafiadoras para aqueles vindos diretamente da sala de aula tradicional.

Diversas tarefas corriqueiras do dia a dia

do professor, como o acompanhamento do comportamento do aluno, a percepção da sua evolução nas disciplinas e suas dificuldades ficam comprometidas no ambiente do ensino à distância, sendo muitas vezes relegadas a relatórios pouco intuitivos e demasiadamente abstratos. Por outro lado, ter o poder de acompanhar a forma como os alunos interagem com as atividades do curso on-line, como eles se movimentam entre os materiais didáticos disponíveis e as atividades avaliativas, pode revelar padrões de comportamento que auxiliam os professores, tutores e gestores educacionais na produção de estratégias de ensino que maximizem a aprendizagem e redu-

zam a evasão e insucesso.

Nesse contexto, a Mineração de Processos Educacionais - EPM (*Educational Process Mining*) [2] surge como uma alternativa para a extração de conhecimento em bases de dados de logs gerados pelas plataformas de ensino à distância. A EPM é uma sub-área da Mineração de Dados Educacionais [1], e pode ser realizada sob a perspectiva de múltiplos atores do processo de ensino-aprendizagem on-line: gestores e técnicos-administrativos, professores e tutores, alunos, pais e responsáveis, enfim, todos os que interagem com o sistema de ensino virtual. Cada ator representa uma perspectiva/visão do processo, um nível de acesso a diferentes ferramentas, dado que nem todos os recursos da plataforma de ensino estão acessíveis para todos os atores, logo os objetivos da mineração de processos variam de acordo com o perfil dos usuários. Para os perfis administrativos, como os professores e tutores, a conformidade de processos pode exercer um papel de destaque, mas certamente o perfil discente é o principal foco de interesse da literatura e também objeto do presente trabalho.

Os Ambientes Virtuais de Aprendizagem (*Virtual Learning Environments* - VLE) são os softwares responsáveis principalmente por auxiliar na gestão dos objetos de aprendizagem (os recursos pedagógicos) pelos professores e tutores estruturar logicamente esses objetos em unidades de ensino (cursos, módulos, disciplinas, etc) e gerenciar o acesso dos alunos, registrando e monitorando suas interações. Dentre os exem-

plos de VLEs encontram-se o software de código aberto Moodle<sup>1</sup> e a plataforma Google Classroom<sup>2</sup>.

Um curso on-line, em sua estrutura, é uma composição de vários objetos de aprendizagem ordenados logicamente de acordo com um conteúdo programático gerenciado pelo VLE. Já os objetos de aprendizagem são os recursos didáticos do ambiente virtual que incluem páginas de texto, downloads de arquivos, atividades, provas, salas de bate-papo, fóruns, vídeos, etc., que auxiliam o professor na confecção do conteúdo do curso e em suas avaliações.

As interações dos alunos com os objetos de aprendizagem do curso geram os eventos registrados pelo VLE nos logs, produzindo os traços ou variantes do processo de negócio. Via de regra, cada aluno gera uma única variante do processo, a não ser que ele repita o curso. Cada variante desse processo deverá ser associada ao status de término, que corresponde às situações finais do aluno no curso: aprovado, reprovado, desistente, etc.

Partindo desses componentes - VLEs, cursos e logs - é possível delinear uma metodologia de EPM [3], sintetizada na Figura 1 a partir de uma simplificação do processo tradicional de descoberta de conhecimento em bases de dados. Essa metodologia constitui-se de quatro etapas: Aquisição de Dados, Pré-Processamento de Dados, Rotulação de Casos e Mineração de Processos. As etapas e suas atividades são detalhadas a seguir.

1 <https://moodle.com/pt/>

2 <http://classroom.google.com>

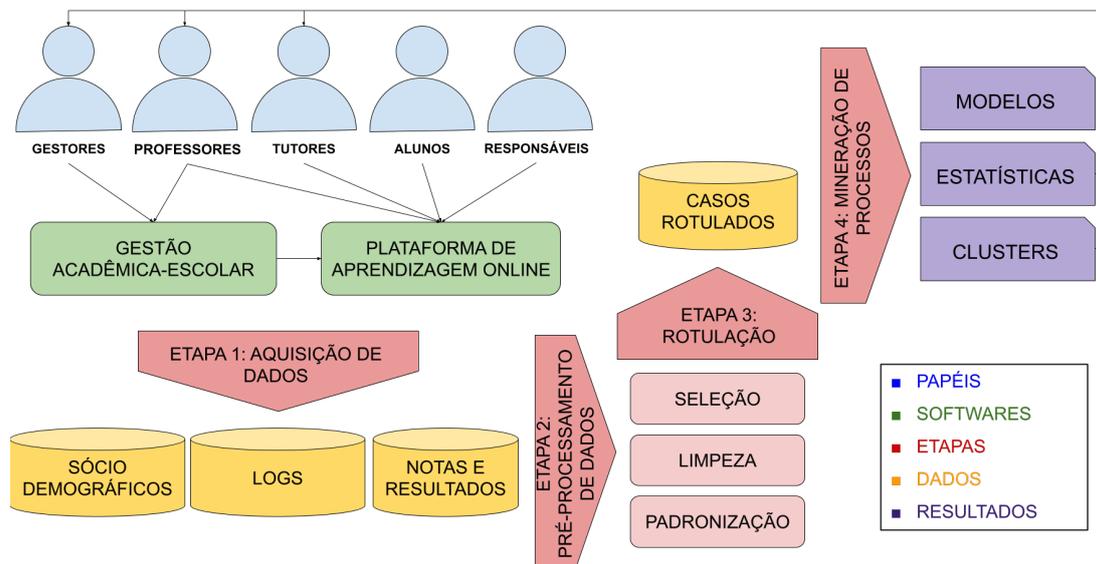


FIG. 01 | SÍNTESE DE UMA METODOLOGIA PARA MINERAÇÃO DE PROCESSOS EDUCACIONAIS

A etapa inicial, **Aquisição de Dados**, compreende a obtenção de todos os logs do VLE referentes à unidade de ensino sob estudo. A Aquisição de Dados deve ser enriquecida com dados auxiliares como as notas e resultados finais e, opcionalmente, com dados sócio-demográficos dos alunos, ampliando as possibilidades de descoberta de conhecimento. No entanto, é importante salientar que os dados podem ser classificados como informações sensíveis pela legislação vigente de proteção de dados pessoais e sua utilização indevida pode levar a sanções penais. Diante do exposto, é salutar que os dados coletados no VLE ou nos Sistemas de Gestão Acadêmica sejam pré-processados para a anonimização e remoção de informações sensíveis. Em particular, as notas e situações dos alunos nos cursos são importantes para classificar cada caso do processo, e posteriormente distinguir os comportamentos típicos nos casos que podem ser associados a cada situação final.

A etapa de **Pré-Processamento de Dados** compreende as atividades de sele-

ção, limpeza e padronização dos logs. Na etapa de seleção, os logs devem ser filtrados selecionando apenas as atividades e usuários correspondentes aos perfis desejados, nesse caso o dos alunos. Nessa etapa é importante também decidir qual a granularidade de análise dos eventos, que implicará a granularidade de representação dos objetos de aprendizagem. Isso implica em escolher se os eventos dos objetos de aprendizagem serão tratados individualmente, gerando uma atividade específica, ou se eles serão agrupados por tipo de objeto, aumentando a granularidade, simplificando o processo final, porém perdendo um pouco de poder de análise.

A etapa de **Rotulação dos Traços** é caracterizada pelo cruzamento de dados entre os casos e os dados de notas e resultados para a rotulação do processo entre as situações finais. Essa etapa é crucial para que se possa discriminar aspectos específicos nos modelos de processo que sejam característicos de uma situação final de curso. Essa etapa é igualmente mandatória quando houver a demanda

de aplicação de algoritmos supervisionados de aprendizado de máquina, particularmente visando tarefas de classificação e predição. Além das situações finais, os logs podem ser rotulados com dados sócio-demográficos dos alunos, se disponíveis, que auxiliarão na identificação de comportamentos correlacionados com cada rótulo.

Na quarta etapa, a de **Mineração de Processos**, os logs já pré-processados e rotulados são combinados e destilados para extrair um modelo de processos de negócio que representa o curso on-line e as estatísticas relacionadas a esse modelo. Diversas atividades de mineração de processos podem ser empregadas [4], dentre elas a Descoberta e Simplificação de Processos, responsável por extrair o processo a partir dos logs, Conformidade de Processos, que valida a adequação de um caso ao modelo de processo gerado, etc. As atividades dessa etapa pode ser inteiramente desempenhadas utilizando ferramentas computacionais tradicionais da Mineração de Processos, como o PRoM<sup>3</sup>, pm4py<sup>4</sup>, Apromore<sup>5</sup>, Celonis<sup>6</sup>, dentre outras.

O modelo do processo, descoberto a partir dos logs, deve ser objeto de estudo dos professores e gestores educacionais. É importante contrastá-lo com a estrutura do curso no VLE, que ainda pode ser realizado a partir dos métodos de *Compliance*. Essa análise é uma ferramenta útil para detectar disparidades entre o

planejamento do professor e a real forma como os alunos consomem o curso. O Agrupamento de Casos (clustering) é empregado na identificação de perfis de comportamento dos alunos. Para investigações mais detalhadas sobre ocorrências em casos isolados, o Replay de Casos é uma ferramenta valiosa para os professores e tutores compreenderem a dinâmica e decisões do estudante.

Outro resultado do processo de mineração são as estatísticas de processo [3] [5], que incluem as distribuições de frequência de eventos (a frequência em que uma atividade do processo é acessada, considerando o processo geral ou cada caso), a distribuição de tempos gastos pelos alunos em cada atividade do processo, entre outras. Essas distribuições nos permitem identificar, em suas caudas, tanto os objetos de aprendizagem irrelevantes (que nunca ou quase nunca são acessados) quanto os gargalos, aquelas atividades que prendem o aluno por muito tempo, bloqueando a continuidade do curso. A identificação desses objetos é crucial para auxiliar o professor a otimizar o curso on-line.

Por fim, conclui-se que a metodologia da EPM leva à extração de uma forma de conhecimento valiosa no contexto educacional: o comportamento. Cada traço de um aluno é uma amostra da dinâmica de aprendizagem, que será representada em um modelo de processo, com suas estatísticas e agrupamentos. Embora seja uma metodologia valorosa, o seu emprego ainda não é difundido, existindo ainda oportunidades de pesquisa em aberto.

---

3 <https://promtools.org/>

4 <https://pm4py.fit.fraunhofer.de/>

5 <https://apromore.com/>

6 <https://www.celonis.com/>

**Referências:**

1. ALEEM, Abdul; GORE, Manoj Madhava. Educational data mining methods: A survey. In: 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT), pp. 182-188. IEEE, 2020.
2. BOGARÍN, Alejandro; CERESO, Rebeca; ROMERO, Cristóbal. A survey on educational process mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, v. 8, n. 1, p. e1230, 2018.
3. DAYRELL, Marcella Souza; SILVA, Petrônio Cândido de Lima. Análise do Comportamento Discente em Cursos à Distância Baseada em Mineração de Processos. In: Anais do Encontro Nacional de Modelagem Computacional, Encontro de Ciência e Tecnologia de Materiais, Conferência Sul em Modelagem Computacional e Seminário e Workshop em Engenharia Oceânica. Anais...Pelotas(RS) UFPel / FURG / UNIPAMPA, 2022.
4. GARCIA, Cleiton dos Santos et al. Process mining techniques and applications—A systematic mapping study. Expert Systems with Applications 133 (2019): 260-295.
5. NETO, José F. S.; PERES, Sarajane M., CORREIA, Paulo; FANTINATO, Marcelo. Is my classroom flipped? using process mining to avoid subjective perception. eLearn, 2021(12).



**PETRÔNIO CÂNDIDO DE LIMA E SILVA** é professor no Instituto Federal do Norte de Minas Gerais. Doutor em Engenharia Elétrica pela Universidade Federal de Minas Gerais. Atua na área de Ciência de Dados e Inteligência Computacional. Atua nos grupos de pesquisa {cidic}/IFNMG, MINDS/UFMG e no PPGMCS/Unimontes.