

Análise da trajetória de aprendizagem do aluno no ensino a distância por meio da mineração de processos

**Adriana J. Unger¹, Daniel A. L. Junior¹, Felipe O. Lima¹, Iago C. Geraldo¹,
Sheila K. Venero¹, Rosana R. A. Ambrosio¹**

¹Universidade Virtual do Estado de São Paulo (UNIVESP)
São Paulo, SP, Brasil

adriana.unger@cursos.univesp.br, daniel.lino.junior@cursos.univesp.br,
felipe.lima@cursos.univesp.br, iago.geraldo@cursos.univesp.br,
skvenero@ic.unicamp.br, rosana.ambrosio@univesp.br

Abstract. *Thanks to the use of virtual learning environments in distance education, user actions can be recorded in event logs, capturing student activities at different levels of granularity. In this scenario, process mining applications can reveal students' learning paths, helping to discover, monitor, and improve educational processes. This article presents the mining of educational processes in a Brazilian virtual university to analyze the processes carried out by the students during their learning path. The analysis of interactions between students and content modules showed which activities generated more student engagement and helped to understand how students learn.*

Resumo. *Graças ao uso de ambientes virtuais de aprendizagem na educação a distância, ações de usuários podem ser registradas em logs de eventos, capturando as atividades dos estudantes em diferentes níveis de granularidade. Nesse cenário, aplicações de mineração de processos podem revelar a trajetória dos estudantes, ajudando a descobrir, monitorar e melhorar os processos educacionais. Este artigo apresenta a mineração de processos educacionais de uma universidade virtual brasileira para analisar os processos executados pelos estudantes durante sua trajetória de aprendizagem. A análise das interações entre os estudantes e os módulos de conteúdos mostrou quais atividades geram mais engajamento dos alunos e ajudou a compreender como os alunos aprendem.*

1. Introdução

A Educação a Distância (EaD) teve um crescimento vertiginoso nos últimos vinte anos, apresentando ao longo de seu processo histórico várias características e formas distintas de conceber o processo de aprendizado mediado pela tecnologia. No Brasil, o número de instituições de ensino públicas e privadas que oferecem cursos nesta modalidade tem crescido bastante, principalmente depois da publicação da Lei de Diretrizes e Bases – LDB em 1996 [Mugnol 2009].

Uma das principais vantagens do ensino a distância é proporcionar formação profissional a indivíduos que estão impedidos de continuar seus estudos por diferentes motivos e que não têm a oportunidade de frequentar o ensino tradicional. Assim, o ensino a

distância por sua flexibilidade almeja suprir as carências do ensino tradicional e facilitar o acesso à educação a um número maior de pessoas.

Nesse contexto, foi fundada em 2012 a Universidade Virtual do Estado de São Paulo (UNIVESP), uma instituição de ensino superior pública com cursos oferecidos na modalidade educacional a distância, tornando-se a quarta universidade pública do Estado de São Paulo. A UNIVESP conta com mais de 46 mil alunos e está presente em 374 polos, localizados em 325 municípios, ou seja, seu alcance abrange mais de 50% do território paulista e 90% da população estadual.

A UNIVESP oferece cursos semipresenciais que se realizam no Ambiente Virtual de Aprendizagem (AVA), onde ocorre a interação do estudante com o tutor. Dentro do AVA são fornecidos os conteúdos pedagógicos, videoaulas, fóruns, atividades e bibliotecas digitais. Graças aos ambientes virtuais de aprendizagem, as ações de usuários podem ser registradas em um *log* de eventos, capturando as atividades dos estudantes em diferentes níveis de granularidade. Esses *logs* refletem a interação do estudante com os conteúdos no AVA e podem se tornar suporte para as ações pedagógicas e de gerenciamento de ensino por professores, tutores, gestores e outros profissionais da área.

Um dos problemas observados no EaD é o baixo engajamento dos alunos com as atividades fornecidas nos ambientes virtuais de ensino [de Oliveira et al. 2019]. Analisar quais são as atividades que geram mais engajamento dos alunos e quais não são tão atrativas para eles permitiria melhorar os instrumentos utilizados no ensino e, dessa forma, melhorar a estratégia de ensino e aprendizagem.

Nesse cenário, aplicações da mineração de processos podem revelar a trajetória dos estudantes no ambiente de aprendizagem, ajudando a descobrir, monitorar e melhorar os processos educacionais. Assim, este artigo tem como objetivo analisar os processos educacionais executados pelos estudantes da UNIVESP durante sua trajetória de aprendizagem usando técnicas de mineração de processos [Van Der Aalst 2012]. Esses processos abrangem as interações entre os estudantes e os módulos de conteúdos e atividades propostos pela instituição. As interações mostram quais são as atividades mais utilizadas pelos estudantes e ajudam a compreender como os alunos aprendem, identificando o caminho que eles realizam dentro do AVA. Desse modo, é possível verificar ainda as diferentes trajetórias de aprendizagem de acordo com o perfil do aluno, propondo assim soluções para potencializar a estratégia de ensino na instituição.

Este artigo está estruturado da seguinte forma: a Seção 2 apresenta a fundamentação teórica, a Seção 3 descreve os materiais e métodos usados na pesquisa, a Seção 4 apresenta os resultados encontrados por meio da mineração de processos e a Seção 5 encerra o artigo com as considerações finais.

2. Fundamentação Teórica

2.1. Mineração de Processos

Para [Van Der Aalst 2012], a mineração de processos é uma importante ferramenta para análise e interpretação de eventos nos sistemas de uma organização. A mineração de processos educacionais é uma área de pesquisa emergente que permite a análise de aprendizagem baseada em dados (*learning analytics*) por meio do conhecimento dos processos educacionais. A mineração de processos está na interface entre a mineração de dados

(*data mining*) e a ciência de processos (*process science*), permitindo que dados de registros de eventos (*logs*) de sistemas de informação possam ser usados para extrair conhecimento sobre os processos de negócio – cadeia de atividades executadas – de uma organização.

[Bogarín et al. 2018] explicam que, ao se utilizar dados de registros coletados no ambiente educacional, viabilizam-se a descoberta, a análise e o fornecimento de representação visual do processo educacional. Segundo [Juhaňák et al. 2019], com as técnicas de Mineração de Processos – *Process Mining (PM)*, torna-se possível a identificação do comportamento estratégico dos alunos. [Grigorova et al. 2017] destacam que o *PM* auxilia no fornecimento de *feedback* aos tutores de plataformas virtuais, ao se afirmar como suporte para a tomada de decisões que envolvam a melhoria da aprendizagem dos alunos. Com os resultados das análises de determinado conjunto de processos, torna-se possível desenvolver ações preventivas que minimizem fatores negativos que impactam nos processos do saber. Portanto, a utilização de ferramentas como *PM* na educação permitiria a extração de informações úteis e a obtenção de uma melhor compreensão da aprendizagem, gerando recomendações e promovendo melhorias contínuas no ensino [dos Santos Neto et al. 2022, Maita et al. 2017]

2.2. Metodologia de projeto de mineração de processos

Dentre os métodos de aplicação da mineração de processos, destaca-se o *Process Mining Project Methodology (PM²)* de [Eck et al. 2015]. O *PM²* orienta as organizações que executam projetos de mineração de processos com o objetivo de melhorar o desempenho do processo ou a conformidade com regras e regulamentos. O método é composto por seis estágios, a saber: planejamento, extração, processamento de dados, mineração e análise, avaliação e refinamento (como ilustrado na Figura 1).

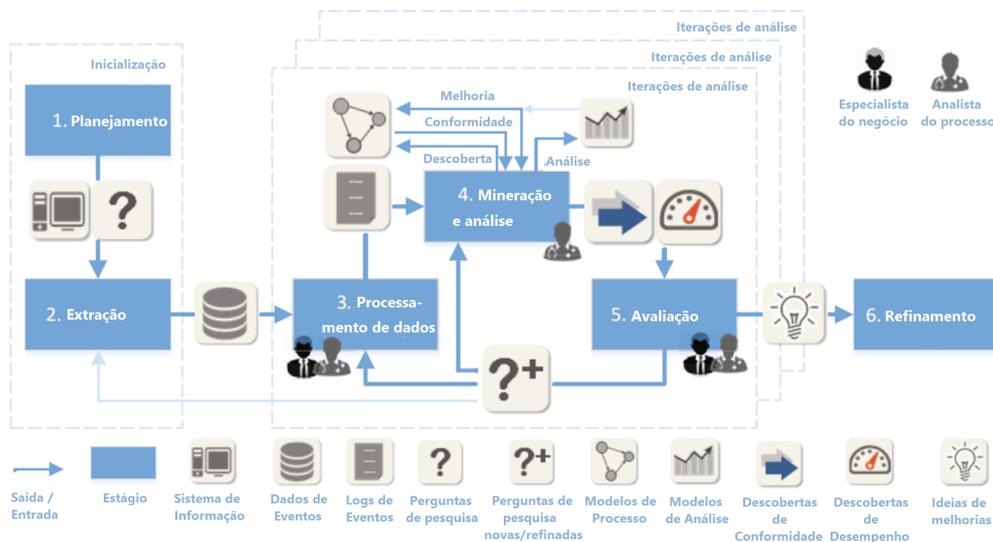


Figura 1. Visão global da metodologia *PM²* [Eck et al. 2015].

Durante os dois primeiros estágios do método, planejamento e extração, são definidas as perguntas de pesquisa iniciais e extraídos os dados dos eventos. Logo, uma ou mais iterações de análise são realizadas, possivelmente em paralelo. Por sua vez, cada

iteração de análise executa uma ou mais vezes as etapas de processamento de dados, mineração e análise, e avaliação. O tempo que leva para completar uma iteração pode variar de minutos a dias, dependendo da complexidade do processo de mineração e análise. Se os resultados forem satisfatórios, eles podem ser usados na etapa de refinamento.

A seguir, os seis estágios da metodologia PM^2 são explicados, incluindo suas entradas e saídas, e as atividades por eles abrangidas:

- **Planejamento:** Nesta etapa o projeto é estruturado e são definidas as perguntas de pesquisa, que são as saídas deste estágio.
- **Extração:** Nesta etapa são extraídos os dados dos eventos. As entradas desta etapa são as perguntas de pesquisa e os sistemas de informação que suportam a execução dos processos de negócio que foram selecionados para serem analisados. As saídas desta etapa são os dados dos eventos.
- **Processamento de dados:** O objetivo principal desta etapa é criar os *logs* de eventos, processando os dados de maneira que possam ser usados na etapa de mineração e análise.
- **Mineração e análise:** Nesta etapa são aplicadas as técnicas de mineração de processos nos *logs* de eventos a fim de responder às perguntas da pesquisa e obter, desse modo, uma visão sobre o desempenho e a conformidade dos processos. As entradas são os *logs* de eventos. As saídas são as descobertas que respondem as perguntas de pesquisa referentes às métricas de desempenho e conformidade.
- **Avaliação:** Nesta etapa os resultados da análise são relacionados com as ideias de melhoria que alcançam os objetivos do projeto. As entradas deste estágio são os modelos de processos e as descobertas em termos de desempenho e conformidade do estágio anterior. As saídas são as ideias de melhoria ou novas questões de pesquisa.
- **Refinamento:** O objetivo desta etapa é usar as percepções obtidas para modificar a execução real do processo. As entradas desta fase são as ideias de melhoria da fase de avaliação. As saídas desta etapa são as modificações do processo.

3. Materiais e Métodos

Para analisar a trajetória de aprendizagem do aluno no ensino a distância por meio da mineração de processos, foi adotado o método indutivo. Dessa forma, partimos da observação particular de 4 disciplinas em 7 cursos durante um período de duração de 4 bimestres para chegarmos à afirmação de um princípio geral. Além disso, aplicamos o método de pesquisa PM^2 .

Para que a pesquisa no tema da mineração de processos educacionais pudesse ser realizada, foi viabilizado o acesso aos dados dos *logs* de atividades dos alunos no AVA da UNIVESP. O uso de algoritmos de mineração de processos requer no mínimo três informações:

- **Id do caso:** identificação do aluno, p.ex. Registro de Aluno (RA) ou e-mail institucional (anonimizado por meio de *hash* para fins de proteção de dados pessoais);
- **Atividade:** descrição da ação realizada no AVA, p.ex. acesso a conteúdo, interação no fórum, entrega de atividade;
- **Timestamp:** data e hora em que a atividade foi realizada.

Para a etapa de extração de dados, foram consideradas as disciplinas listadas na Tabela 1, cursadas durante o ano letivo de 2020 no AVA Canvas. A escolha das disciplinas considerou a preferência por disciplinas iniciais dos cursos (1º e 2º ano), com alunos presumidamente mais interessados e ativos no AVA. Considerou-se também a preferência por disciplinas que são ministradas para alunos de vários cursos, para permitir a comparação entre as trajetórias de aprendizagem desses alunos. Adicionalmente, de modo a possibilitar análises de correlação da trajetória do aluno com seu desempenho nas disciplinas cursadas, foi solicitada a extração de dados de conceitos e resultados (aprovação ou reprovação) dos alunos no Sistema de Gestão Acadêmica (SGA).

Tabela 1. Disciplinas e cursos para extração de dados educacionais.
Fonte: Autoria própria.

Disciplina	Curso
Didática	Letras Licenciatura em Matemática Pedagogia
Algoritmos e Programação de Computadores I	Engenharia de Computação Bacharelado em Tecnologia da Informação Bacharelado em Ciência de Dados
Sistemas de Informação	Engenharia de Computação Engenharia de Produção
Administração I	Engenharia de Computação Engenharia de Produção

Os dados extraídos do AVA Canvas foram processados para sua conversão no formato de *log* de eventos. O processamento de dados incluiu a junção dos casos e eventos das disciplinas, a anonimização do RA e a eliminação dos dados pessoais como o nome de alunos. O *log* de eventos resultante, contendo 15.820 casos (alunos) e 703.425 eventos (atividades de aprendizagem), foi então importado na ferramenta de mineração de processos Celonis.¹

4. Resultados

O resultado da mineração de processos educacionais contendo o percurso dos discentes nessas disciplinas pode ser visualizado na Figura 2. A trajetória dos alunos pode ser filtrada por disciplina cursada, categoria de atividade realizada no AVA, faixa de valores de nota e resultado de aprovação nas disciplinas.

A Figura 2 apresenta o modelo de processos representativo da trajetória de aprendizado dos alunos, considerando o escopo de observação desta pesquisa, sob a perspectiva da frequência de atividades, ou seja, do número de vezes que cada atividade de aprendizagem foi executada. Os nós e transições em cor destacada no modelo indicam, respectivamente, as atividades mais frequentemente executadas pelos alunos e as sequências mais frequentes entre atividades no AVA. Uma breve descrição das atividades é apresentada na Tabela 2.

¹A Celonis é uma empresa multinacional com sede na Alemanha que é líder no mercado de soluções em mineração de processos.

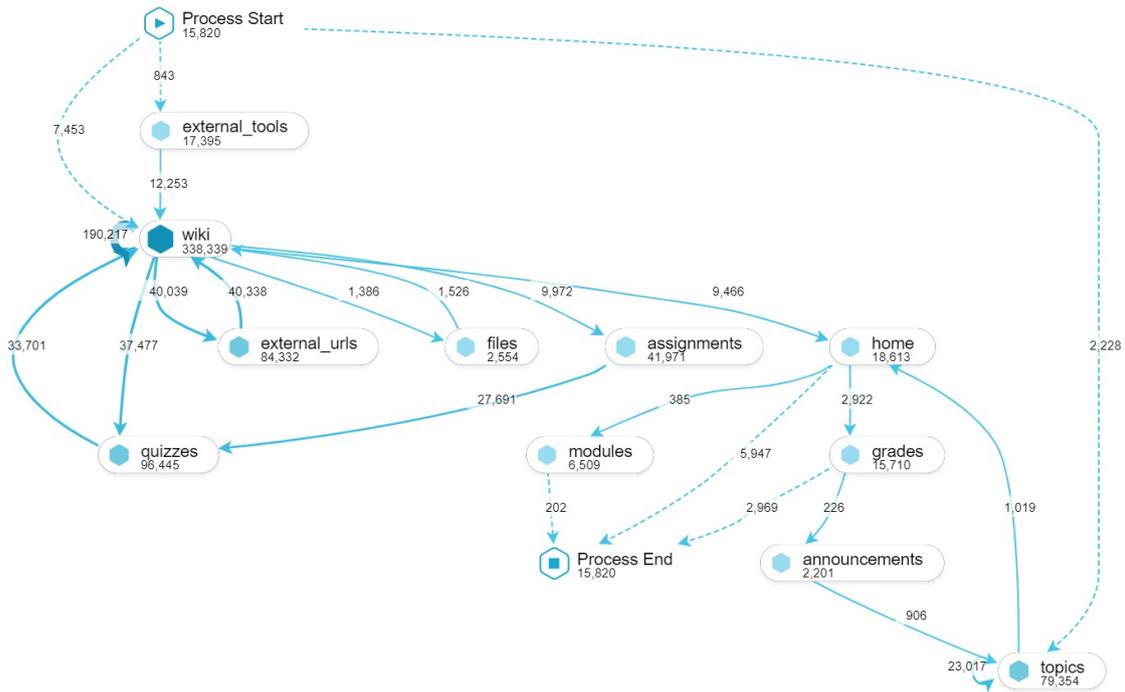


Figura 2. Mineração de processos educacionais sob a perspectiva de frequência da atividade. Fonte: Autoria própria.

A mineração de processos permite ainda análises comparativas entre as trajetórias típicas de alunos de diferentes disciplinas. A Figura 3 ilustra os modelos de processo educacional de aprendizagem na disciplina de Didática (ministrada nos cursos da área de Humanas) e nas disciplinas Algoritmos e Programação de Computadores I, Sistemas de Informação e Administração I (ministradas nos cursos da área de Exatas), considerando a perspectiva da frequência de atividades da mineração de processos.

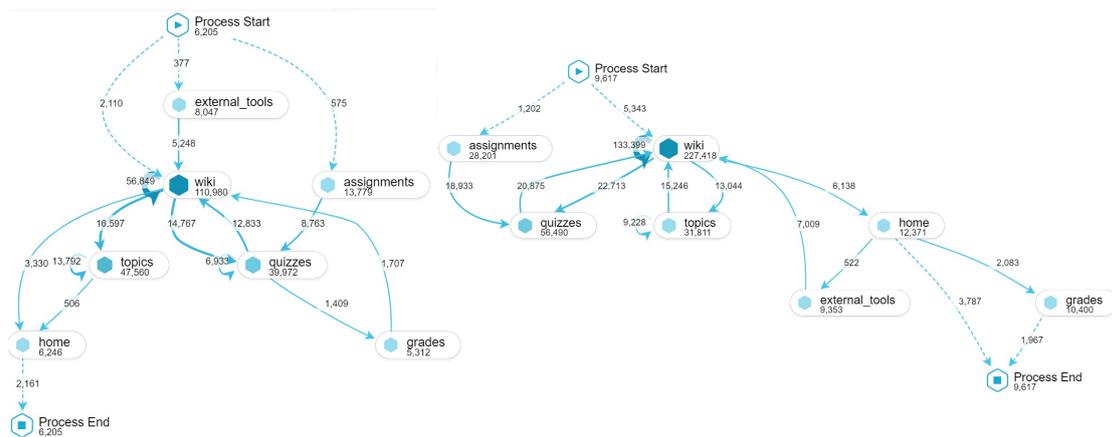


Figura 3. Mineração do processo educacional em cursos de Humanas (à esquerda) e Exatas (à direita). Fonte: Autoria própria.

Tabela 2. Atividades de aprendizagem efetuadas pelos alunos no AVA. Fonte: Autoria própria.

Atividade	Descrição	Número de alunos que executaram a atividade	Número de vezes que a atividade foi executada
<i>wiki</i>	Acesso a conteúdos assíncronos das semanas, como videoaulas e textos	14.990	338.339
<i>quizzes</i>	Execução de atividades avaliativas da disciplina no formato de quiz	14.730	96.445
<i>external_urls</i>	Acesso a recursos indicados em websites externos	8.223	84.332
<i>topics</i>	Acesso aos fóruns de discussão	11.440	79.354
<i>assignments</i>	Execução de atividades avaliativas da disciplina no formato de tarefas	11.259	41.971
<i>home</i>	Acesso a homepage da disciplina	15.772	18.613
<i>external_tools</i>	Acesso a recursos externos ao AVA, como biblioteca	9.180	17.395
<i>grades</i>	Acesso às notas	13.524	15.710
<i>modules</i>	Acesso aos menus de módulos da disciplina	6.081	6.509
<i>files</i>	Acesso a arquivos anexados	2.552	2.554
<i>announcements</i>	Acesso a avisos da disciplina	2.197	2.201

A comparação entre as métricas de ambos os modelos de processo permite verificar que a taxa de acessos aos conteúdos das disciplinas (atividade *wiki*) por aluno é mais de 30% maior nos cursos de exatas: $227.418 \text{ acessos} / 9.617 \text{ alunos} = \text{taxa de } 23,6 \text{ acessos/aluno}$ nos cursos de exatas, comparado a $110.980 \text{ acessos} / 6.205 \text{ alunos} = \text{taxa de } 17,9 \text{ acessos/aluno}$ nos cursos de humanas. A Tabela 3 apresenta o comparativo detalhado das taxas de acesso nos cursos de Exatas e Humanas.

Outra grande discrepância diz respeito à quantidade de acessos aos fóruns de discussão (atividade *topics*): tipicamente um aluno de um curso de Humanas usa duas vezes mais o fórum do que um aluno de um curso de Exatas. Esse tipo de análise permite responder à pergunta de pesquisa, no que diz respeito às atividades que geram mais engajamento dos alunos e quais não são tão atrativas, considerando a trajetória típica e perfil dos alunos de cada curso.

Comparando ainda as trajetórias de aprendizagem e métricas do processo edu-

Tabela 3. Taxas de acesso das atividades de aprendizagem no AVA, por área de curso. Fonte: Autoria própria.

Atividade	Taxa de acesso nas disciplinas dos cursos de Humanas	Taxa de acesso nas disciplinas dos cursos de Exatas
wiki	17,9	23,6
quizzes	6,4	5,9
topics	7,7	3,3
assignments	2,2	2,9
home	1,0	1,3
external_tools	1,3	1,0
grades	0,9	1,1

cional de alunos aprovados e reprovados, é possível obter *insights* para melhorar os instrumentos utilizados no ensino e, dessa forma, melhorar a estratégia de ensino e aprendizagem. As Figuras 4 e 5 apresentam os modelos e métricas de processo educacional dos alunos aprovados e reprovados, respectivamente, em todas as disciplinas.

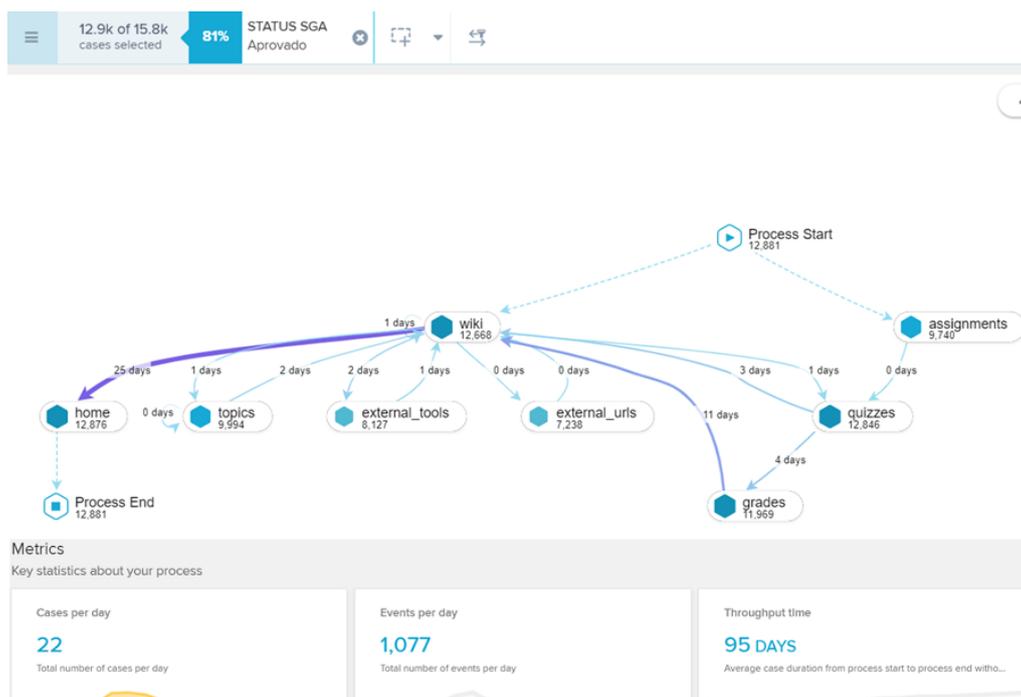


Figura 4. Modelo e métricas do processo educacional de alunos aprovados nas disciplinas. Fonte: Autoria própria.

O baixo engajamento dos alunos reprovados nas atividades do AVA pode ser observado nitidamente: por meio do menor número de eventos por dia proporcionalmente ao número de alunos; pela menor duração do processo (30 dias a menos do que no caso dos alunos aprovados), o que indica um abandono da disciplina mesmo antes da reprovação; e pela maior presença de gargalos no processo (transações lentas indicadas nas Figuras 4 e 5 pelas setas em cor destacada), revelando grandes períodos sem nenhum acesso ao AVA por parte desses alunos.

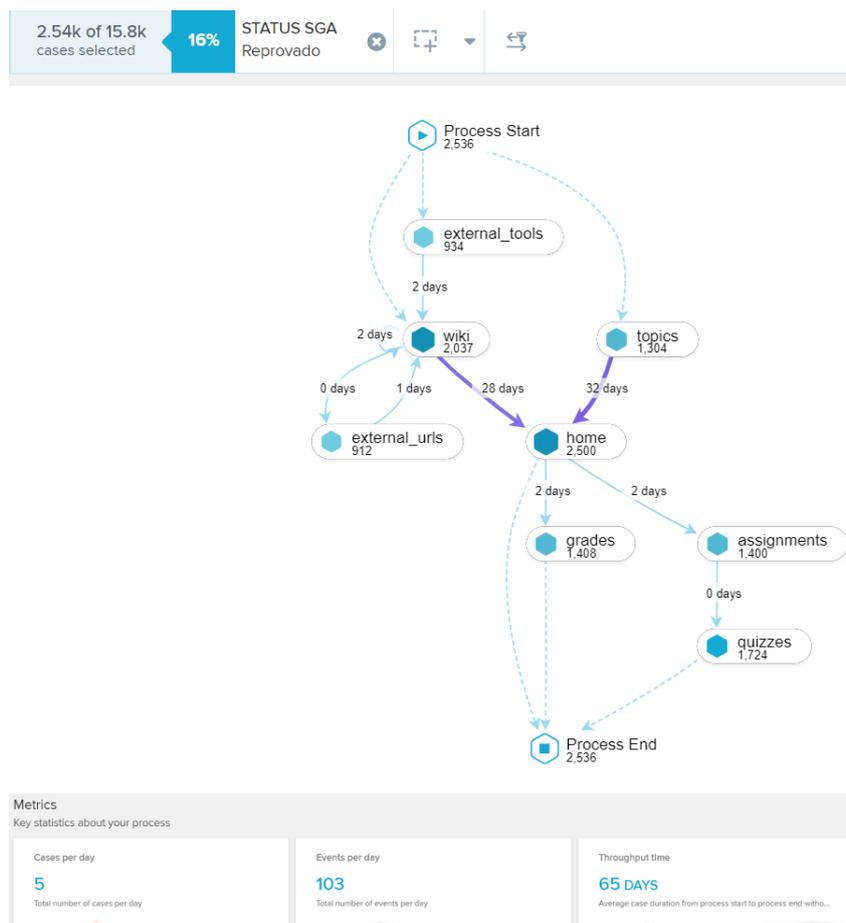


Figura 5. Modelo e métricas do processo educacional de alunos reprovados nas disciplinas. Fonte: Autoria própria.

5. Considerações Finais

Este estudo teve como objetivo analisar a trajetória de aprendizagem dos alunos no ensino a distancia. Para atingir esse objetivo foi aplicada a mineração de processos no *log* de eventos do AVA Canvas da UNIVESP.

O conjunto de dados analisados advém de quatro disciplinas iniciais (1º e 2º ano) que foram ministradas durante o ano letivo de 2020 no AVA Canvas a alunos de vários cursos. Estas disciplinas foram escolhidas com o intuito de observar alunos com diferentes perfis e pelo fato destes serem alunos iniciantes, supostamente mais engajados e ativos no ambiente virtual. Assim, as disciplinas escolhidas foram: Didática, Algoritmos e Programação de Computadores I, Sistemas de Informação e Administração I.

Os principais achados em relação ao engajamento dos alunos mostram que:

1. a trajetória mais comum dos alunos é de *home* para *wiki*, que representa os conteúdos assíncronos das semanas, como videoaulas e textos;
2. a quantidade média de atividades acessadas por dia é de 1,17 atividades, sendo que a maioria dos alunos chega a passar quase 1 mês sem acessar nenhuma atividade;
3. a atividade mais utilizada pelos alunos é a *wiki* (conteúdos assíncronos);
4. a segunda atividade mais acessada pelos alunos são os *quizzes* (atividades avaliativas da disciplina no formato de questionário);

5. os avisos das disciplinas são os menos acessados.

Os alunos aprovados, em sua maioria, iniciam seu trajeto da página inicial do curso para *wiki*, e acessam com frequência *quizzes*, tópicos, notas, tarefas, ferramentas e websites externos. Em relação aos alunos reprovados, embora tenham uma trajetória similar à dos alunos aprovados, observou-se baixo engajamento, ou seja, menor frequência de acesso às atividades do AVA.

Assim, podemos concluir que a mineração de processos pode ajudar a analisar tanto o engajamento dos alunos nas diferentes atividades propostas pela instituição como a avaliar o perfil de aprendizagem do aluno nos diferentes cursos. Os resultados obtidos demonstram que a mineração de processos oferece uma importante contribuição para avaliar a qualidade do EaD e melhorar a estratégia de ensino e aprendizagem.

Como trabalhos futuros, sugere-se realizar diferentes análises com outras disciplinas e com alunos de diferentes anos. Outro aspecto interessante a analisar é a trajetória de alunos que repetidamente apresentam desempenho ruim no decorrer do curso.

Referências

- Bogarín, A., Cerezo, R., and Romero, C. (2018). A survey on educational process mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(1):e1230.
- de Oliveira, P. L. S., de Souza, A. J., and Rodrigues, R. (2019). Identificação de pesquisas referentes ao engajamento de alunos em plataformas de lms e suas relações com o desempenho acadêmico. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 30, page 1631.
- dos Santos Neto, J. F., Marques Peres, S., Correia, P., and Fantinato, M. (2022). Is my classroom flipped? using process mining to avoid subjective perception. *ELearn*, 2021(12).
- Eck, M. L. v., Lu, X., Leemans, S. J., and Van Der Aalst, W. M. (2015). Pm²: a process mining project methodology. In *International conference on advanced information systems engineering*, pages 297–313. Springer.
- Grigorova, K., Malysheva, E., and Bobrovskiy, S. (2017). Application of data mining and process mining approaches for improving e-learning processes. In *3rd International Conference on Information Technology and Nanotechnology*, pages 25–27.
- Juhaňák, L., Zounek, J., and Rohlíková, L. (2019). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Computers in Human Behavior*, 92:496–506.
- Maita, A. R. C., Fantinato, M., Peres, S. M., Thom, L. H., and Hung, P. C. K. (2017). Mining unstructured processes: An exploratory study on a distance learning domain. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 3240–3247.
- Mugnol, M. (2009). A educação a distância no brasil: conceitos e fundamentos. *Revista Diálogo Educacional*, 9(27):335–349.
- Van Der Aalst, W. (2012). Process mining: Overview and opportunities. *ACM Transactions on Management Information Systems (TMIS)*, 3(2):1–17.