



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS DE QUIXADÁ
CURSO DE GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

CARLOS SAMUEL SOUSA SANTOS

**IDENTIFICAÇÃO DE PONTOS CRÍTICOS NA ESTRUTURA CURRICULAR DO
CURSO DE CIÊNCIA DA COMPUTAÇÃO DA UFC UTILIZANDO MINERAÇÃO DE
PROCESSOS**

QUIXADÁ
2025

CARLOS SAMUEL SOUSA SANTOS

IDENTIFICAÇÃO DE PONTOS CRÍTICOS NA ESTRUTURA CURRICULAR DO CURSO
DE CIÊNCIA DA COMPUTAÇÃO DA UFC UTILIZANDO MINERAÇÃO DE PROCESSOS

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Ciência da Computação
da Universidade Federal do Ceará, como
requisito parcial à obtenção do grau de bacharel
em Ciência da Computação.

Orientador: Prof. Dr. Paulo de Tarso
Guerra Oliveira.

QUIXADÁ

2025

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- S234i Santos, Carlos Samuel Sousa.
Identificação de pontos críticos na estrutura curricular do curso de ciência da computação da UFC utilizando mineração de processos / Carlos Samuel Sousa Santos. – 2025.
57 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Quixadá, Curso de Ciência da Computação, Quixadá, 2025.
Orientação: Prof. Dr. Paulo de Tarso Guerra Oliveira.
1. Mineração de processos. 2. Evasão acadêmica. 3. Mineração de dados educacionais. 4. Rede de petri. 5. Análise de conformidade. I. Título.

CDD 004

CARLOS SAMUEL SOUSA SANTOS

IDENTIFICAÇÃO DE PONTOS CRÍTICOS NA ESTRUTURA CURRICULAR DO CURSO
DE CIÊNCIA DA COMPUTAÇÃO DA UFC UTILIZANDO MINERAÇÃO DE PROCESSOS

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Ciência da Computação
da Universidade Federal do Ceará, como
requisito parcial à obtenção do grau de bacharel
em Ciência da Computação.

Aprovada em: 25/02/2025.

BANCA EXAMINADORA

Prof. Dr. Paulo de Tarso Guerra
Oliveira (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Davi Romero de Vasconcelos
Universidade Federal do Ceará (UFC)

Prof. Dr. Críston Pereira de Souza
Universidade Federal do Ceará (UFC)

À minha mãe e minhas irmãs, minhas maiores
inspirações e motivações. Por elas, sigo em
frente, superando desafios e buscando sempre o
melhor de mim. Tudo o que faço é por elas, e é
nelas que encontro a força para continuar.

AGRADECIMENTOS

Agradeço, primeiramente, ao Professor Paulo de Tarso por ter aceitado me orientar e por ser uma referência, não apenas como profissional, mas também como ser humano. O mundo precisa de mais pessoas como você!

Sou grato a Deus e a todas as circunstâncias que me trouxeram até aqui e me permitiram viver este momento.

Ao meu avô, peço desculpas pela ausência nos seus últimos momentos. Lamento não ter conseguido me formar antes da sua partida, mas carrego comigo tudo o que aprendi com você.

À minha mãe Patrícia e às minhas irmãs Letícia, Luciana e Mirella agradeço por existirem. Amo vocês.

Ao Jordão, pelo apoio nesses últimos meses e pela ajuda na execução deste trabalho, meu sincero obrigado.

À Yasmin e à Ana Vitória, pela parceria, irmandade e por serem meu lar e acolhimento nos momentos em que mais precisei.

A todos os amigos que fiz ao longo desse período, vocês tornaram essa caminhada mais leve. Obrigado por cada momento compartilhado – pelos rachas no campinho da UECE, pelas caminhadas na madrugada pelas ruas de Quixadá, pelas idas à praça, pelos pastéis com cabaré ice. E também pelos momentos difíceis – pelas noites mal dormidas estudando para provas, pelas longas esperas nas filas quilométricas do RU e dos ônibus. Não citarei nomes para não correr o risco de esquecer alguém, mas saibam que cada lembrança estará sempre comigo.

À Universidade Federal do Ceará, por ter transformado minha vida.

Obrigado a todos que, de alguma forma, fizeram parte dessa jornada.

“O que vale na vida não é o ponto de partida e
sim a caminhada. Caminhando e semeando, no
fim, terás o que colher.”

(Cora Coralina)

RESUMO

A evasão acadêmica é um dos principais desafios enfrentados pelas Instituições de Ensino Superior, impactando tanto a trajetória dos alunos quanto a eficiência institucional. Neste trabalho, utilizou-se a técnica de *Mineração de Processos*, especificamente a *análise de conformidade via replay*, para investigar a estrutura curricular do curso de Ciência da Computação da UFC - Campus de Quixadá. A modelagem da grade curricular foi realizada utilizando *Redes de Petri*, permitindo uma representação formal dos fluxos acadêmicos dos alunos. A partir da execução do *replay* sobre os dados acadêmicos, foram extraídas métricas quantitativas que possibilitaram a identificação de padrões de progressão estudantil e pontos críticos que podem influenciar a evasão. Além disso, foi desenvolvida uma ferramenta *web* para visualização interativa dos resultados, permitindo visualizações por turma e período. Os resultados obtidos destacam as disciplinas com maior acúmulo de inconformidades na estrutura curricular.

Palavras-chave: mineração de processos; análise de conformidade; redes de petri; mineração de dados educacionais; evasão acadêmica.

ABSTRACT

Academic dropout is one of the main challenges faced by Higher Education Institutions, impacting both students' trajectories and institutional efficiency. In this study, the *Process Mining* technique, specifically *conformance analysis via replay*, was used to investigate the curricular structure of the Computer Science program at UFC - Quixadá Campus. The curriculum modeling was conducted using *Petri Nets*, enabling a formal representation of students' academic flows. By executing the *replay* on academic records, quantitative metrics were extracted, allowing the identification of student progression patterns and critical points that may influence dropout rates. Additionally, a *web* tool was developed for interactive visualization of the results, enabling detailed analyses by cohort and period. The findings highlight courses with the highest accumulation of academic non-conformities.

Keywords: process mining; conformance analysis; petri nets; educational data mining; academic dropout.

LISTA DE FIGURAS

Figura 1 – Evolução dos indicadores de trajetória dos estudantes no curso de ingresso em 2014.	16
Figura 2 – Diagrama hierárquico detalhando as sub-áreas da Mineração de Dados Educacionais (MDE)	18
Figura 3 – Tipos de Mineração de Processos com relação às entradas e saídas.	21
Figura 4 – Exemplo de visualização do modelo de processo usando Rede de Petri.	23
Figura 5 – Representação de Rede de Petri com <i>token</i> no <i>place start</i>	24
Figura 6 – Representação de consumo e produção de <i>tokens</i> na Rede de Petri.	24
Figura 7 – Estrutura curricular do curso de Ciência da Computação da Universidade Federal do Ceará (UFC) do Campus de Quixadá.	31
Figura 8 – Conjunto de transições e places que representam a disciplina QXD0001.	32
Figura 9 – Conjunto de transições e places que representam a disciplina QXD0029.	32
Figura 10 – Estrutura curricular modelada em Rede de Petri.	34
Figura 11 – Dados brutos dos estudantes do curso de Bacharelado em Ciência da Computação.	35
Figura 12 – <i>Log</i> de eventos criado nesta etapa.	36
Figura 13 – <i>Log</i> de eventos de um aluno que foi aprovado em todas as disciplinas obrigatórias.	37
Figura 14 – Comparativo de Estatísticas do <i>Replay</i> entre alunos formados e aluno não formados	38
Figura 15 – Conjunto de transições e places que representam a disciplina QXD0001 com tokens acumulados no <i>place p2</i>	39
Figura 16 – Fluxograma da Estrutura Curricular Representado como um Mapa de Calor.	41
Figura 17 – Gráfico de pizza representando a distribuição de <i>Tokens</i> por Área de Conhecimento.	42
Figura 18 – Tela inicial da aplicação.	44
Figura 19 – Tela de Visualização por Turma.	45
Figura 20 – Tela de Visualização por Período.	47
Figura 21 – Tela de Visualização de Mineração de Processos.	49

LISTA DE TABELAS

Tabela 1 – Um exemplo de um <i>log</i> de eventos, descrevendo o comportamento relacionado ao processo de solicitação de compensação para ingresso de concerto.	20
Tabela 2 – Número de Tokens por Disciplina	40

LISTA DE QUADROS

Quadro 1 – Quadro comparativo das abordagens dos trabalhos relacionados e deste trabalho.	29
---	----

LISTA DE ABREVIATURAS E SIGLAS

AVA	Ambiente Virtual de Aprendizagem
CRISP-DM	Cross Industry Standard Process for Data Mining
EAD	Ensino à distância
IES	Instituição de Ensino Superior
INEP	Instituto Nacional de Estudos e Pesquisas
KNN	K-Nearest Neighbors
MDE	Mineração de Dados Educacionais
PM2	Process Mining Project Methodology
SVM	Support Vector Machine
UFC	Universidade Federal do Ceará
UFSM	Universidade Federal de Santa Maria
UNIVESP	Universidade Virtual do Estado de São Paulo

SUMÁRIO

1	INTRODUÇÃO	13
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	Evasão acadêmica	15
2.2	Mineração de Dados Educacionais	17
2.3	Mineração de Processos	19
2.4	Modelagem de processos	22
3	TRABALHOS RELACIONADOS	26
3.1	Mineração de Dados Educacionais para Predição da Evasão em Cursos de Graduação Presenciais no Ensino Superior	26
3.2	Análise da Trajetória de Aprendizagem do aluno no ensino a distância por meio da Mineração de Processos	27
3.3	Mineração de dados para identificação do perfil de evasão de alunos da UFC - Campus Quixadá	28
3.4	Quadro comparativo	29
4	MINERAÇÃO DE PROCESSOS: ANÁLISE DE CONFORMIDADE . .	30
4.1	Modelagem Formal da Estrutura Curricular	30
4.1.1	<i>Estrutura da Modelagem</i>	31
4.1.2	<i>Conexões entre as Disciplinas</i>	32
4.1.3	<i>Modelo Completo da Estrutura Curricular</i>	33
4.2	Preparação dos Dados e Criação do <i>Log</i> de Eventos	35
4.3	Análise de Conformidade via <i>Replay</i>	37
4.4	Identificação de Pontos Críticos	39
4.4.1	<i>Análise dos Resultados</i>	40
5	FERRAMENTA DE VISUALIZAÇÃO	44
5.1	Visualização Por Turma	45
5.2	Visualização Por Período	46
5.3	Visualização de Mineração de Processos	48
6	CONCLUSÕES E TRABALHOS FUTUROS	52
	REFERÊNCIAS	53

1 INTRODUÇÃO

A evasão escolar no ensino superior representa um desafio para a gestão entidades educacionais e a busca por compreender suas causas continua sendo um tema central em diversas pesquisas e estudos na área da educação (Filho *et al.*, 2007). Esse fenômeno é um problema global que compromete os resultados dos sistemas educacionais, provocando graves consequências sociais, acadêmicas e econômicas (Baggi; Lopes, 2011).

Existem vários fatores que podem motivar o aluno a evadir, os quais podem ser classificados em três grandes grupos: a percepção de uma escolha de curso inadequada aos interesses do estudante, a impossibilidade de manutenção do vínculo por questões socioeconômicas e o abandono por inadequação aos métodos de ensino ou insucessos nas disciplinas (Barroso; Falcão, 2004). Para este último, torna-se importante a identificação de pontos críticos nas estruturas curriculares dos cursos para entender e mitigar o fenômeno da evasão (Santos *et al.*, 2021).

No contexto deste trabalho, um ponto crítico na matriz curricular de um curso é definido como um obstáculo que dificulta ou impede o progresso dos alunos. Os obstáculos podem ser representado por uma disciplina específica ou uma sequência de disciplinas que são caracterizadas por altas taxas de reprovação, seja devido a notas insuficientes ou baixa frequência dos estudantes.

O estudo de Ambiel *et al.* (2021) identificou preditores de evasão no ensino superior entre estudantes trabalhadores e não trabalhadores, utilizando uma abordagem quantitativa e a Escala de Motivos para Evasão do Ensino Superior (M-ES) para analisar dados de 782 estudantes. De modo similar Alves e Mantovani (2016), focaram nos cursos de Engenharia de uma instituição privada, aplicando um questionário a 265 alunos para avaliar o perfil dos estudantes e identificar motivos de evasão, como a dificuldade de conciliar estudos e trabalho, o grande número de reprovações e dificuldades financeiras.

Para tentar prever a evasão acadêmica, diversos autores empregam técnicas de mineração de dados e aprendizado de máquina. Em Brito *et al.* (2020) é utilizado o algoritmo de classificação *Random Forest* para identificar os atributos mais relevantes na previsão da evasão. Já em (Pinheiro *et al.*, 2018) é analisada a capacidade preditiva de classificadores gerados por algoritmos de aprendizado de máquina na identificação de estudantes em risco de evasão.

A Mineração de Processos é uma área que combina conceitos de ciência de dados e modelagem de processos para analisar registros de eventos e extrair informações sobre o funcionamento de um sistema (Aalst, 2022). Essa abordagem permite a descoberta de padrões,

a identificação de gargalos e a verificação de conformidade entre processos reais e modelos esperados. Uma das principais técnicas utilizadas para análise de conformidade é o *replay*, que compara a execução real dos eventos registrados com um modelo formal, destacando desvios e pontos críticos. As técnicas de Mineração de Processos podem ser aplicadas em diferentes contextos, incluindo negócios, saúde e educação, auxiliando na tomada de decisões e na otimização de fluxos operacionais. No contexto acadêmico, elas possibilitam uma análise da trajetória dos alunos, permitindo identificar fatores que influenciam seu progresso ou contribuem para a evasão (Aalst, 2022).

Um fluxo acadêmico pode ser visto como um processo que os alunos percorrem ao longo de seu curso na instituição de ensino. Isso inclui a sequência de disciplinas que eles devem cursar, os pré-requisitos necessários para cada disciplina, os períodos de matrícula, os exames e avaliações, e a progressão até a conclusão do curso. Nesse contexto, a Mineração de Processos pode permitir uma análise dos fluxos acadêmicos e a identificação de gargalos que podem afetar o desempenho dos alunos e contribuir para a evasão (Aalst, 2012).

O estudo de Unger *et al.* (2022) destaca que a aplicação de Mineração de Processos no ensino a distância pode ser utilizada para avaliar o engajamento dos alunos, analisar seus perfis de aprendizagem em diversos cursos, verificar a qualidade da educação a distância e aprimorar as estratégias de ensino e aprendizagem. Já Umer *et al.* (2017) mostra a eficácia da Mineração de Processos para prever precocemente alunos que correm maior risco de reprovação, usando dados de interação em ambientes de ensino *online*.

Diante disso, este trabalho analisa a estrutura curricular do curso de Ciência da Computação da UFC do Campus de Quixadá como um processo, aplicando a técnica de *replay* da Mineração de Processos para verificar a conformidade das trajetórias acadêmicas dos alunos. O objetivo é identificar pontos críticos na estrutura curricular, ou seja, disciplinas que atuam como gargalos e podem estar relacionadas à evasão. Para isso, foram utilizados dados dos históricos acadêmicos dos alunos, referentes ao período de 2013 a 2024, transformados em um *log* de eventos, conforme detalhado nas seções subsequentes.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo introduz os conceitos centrais que permeiam este estudo. Na Seção 2.1, será apresentado o conceito de evasão acadêmica. A Seção 2.2 explora os princípios da mineração de dados educacionais. Na Seção 2.3, são apresentados os conceitos e técnicas de Mineração de Processos. Por fim, a Seção 2.4 trata dos conceitos e técnicas de modelagem de processos.

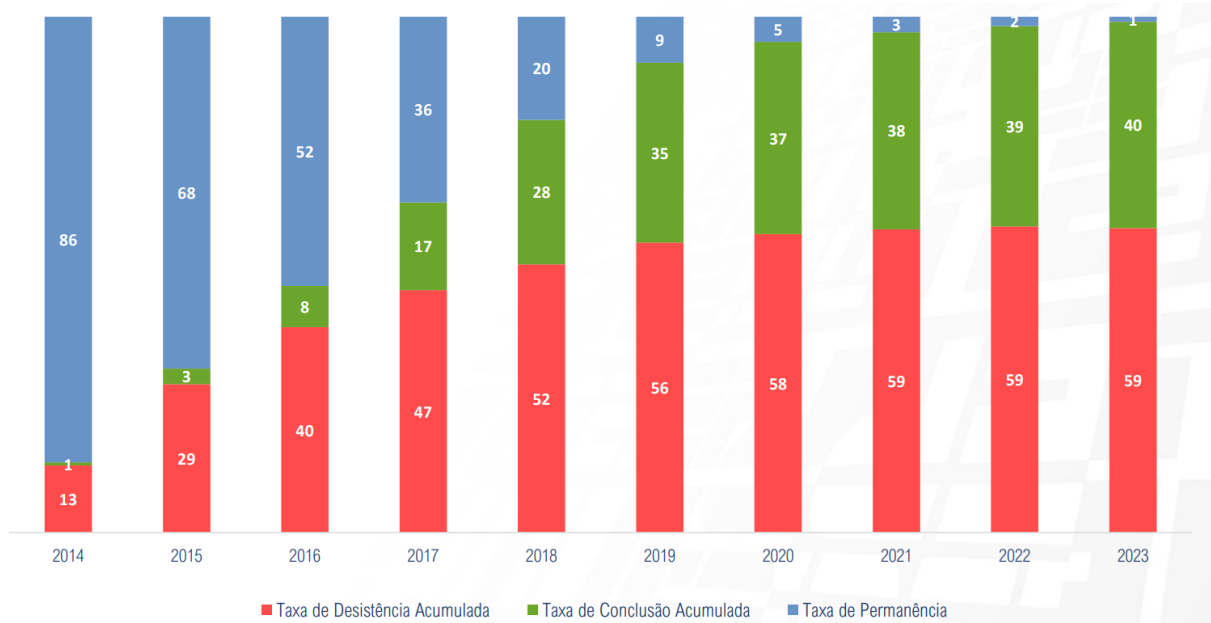
2.1 Evasão acadêmica

A evasão acadêmica é reconhecida como um desafio que afeta as instituições de ensino em todos os níveis, com particular incidência no ensino superior, onde o primeiro ano registra as taxas mais altas de desistência (Filho *et al.*, 2007). Cardoso (2008) distingue entre a evasão "aparente", quando há a troca de curso ou instituição, e a evasão "real", que representa o abandono total do ensino superior.

No contexto atual, a evasão é um fenômeno universal que demonstra comportamentos relativamente similares entre áreas de conhecimento e países, apesar das variadas nuances socioeconômicas e culturais (Veloso; Almeida, 2013). MARTINHO (2014) ressaltam que as implicações da evasão são vastas, afetando dimensões acadêmicas, econômicas, políticas, financeiras e sociais, e envolvendo desde o aluno individual até órgãos governamentais.

De acordo com os dados do levantamento do Instituto Nacional de Estudos e Pesquisas (INEP), apresentados na Figura 1, é possível acompanhar a trajetória dos estudantes ingressantes em 2014 ao longo do período de 2014 a 2023 em todas as instituições de ensino superior do Brasil, tanto públicas quanto privadas. O gráfico permite analisar a evolução dos percentuais de desistência, conclusão e permanência, proporcionando uma visão abrangente do comportamento dos estudantes no ensino superior em nível nacional.

Figura 1 – Evolução dos indicadores de trajetória dos estudantes no curso de ingresso em 2014.



Fonte: Adaptado de Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) (2023).

Os dados indicam que, em 2023, entre os alunos que ingressaram no curso em 2014, 59% dos alunos haviam desistido do curso, enquanto 40% concluíram e apenas 1% permaneciam matriculados. Essas taxas são definidas da seguinte forma:

- Taxa de Desistência Acumulada: percentual de estudantes que abandonaram o curso ao longo dos anos, seja por evasão definitiva, transferência para outra instituição ou outro motivo.
- Taxa de Conclusão Acumulada: percentual de alunos que efetivamente concluíram o curso dentro do período analisado.
- Taxa de Permanência: percentual de alunos que, apesar do tempo decorrido, continuam matriculados no curso.

Diversos modelos têm sido propostos para explicar e abordar a evasão escolar, com foco nas diversas categorias de evasão identificadas pelo SESU/MEC e ABRUEM (1996), que incluem: evasão de curso, que ocorre quando o estudante se desliga do curso superior em situações diversas, como abandono, desistência, transferência ou exclusão por norma institucional; evasão da instituição, quando o estudante se desliga da instituição na qual está matriculado; e evasão do sistema, quando o estudante abandona de forma definitiva ou temporária o ensino superior. Da mesma forma, Filho *et al.* (2007) contribuem com a definição de "evasão anual média" e "evasão total", conceitos que quantificam respectivamente a não continuidade anual de matrícula e a não obtenção de diploma num intervalo de tempo estipulado.

Neste estudo, a evasão acadêmica é examinada como a saída definitiva de um aluno de uma Instituição de Ensino Superior (IES) ou de seu curso, seja por desistência, transferência, interrupção de matrícula ou exclusão institucional. A análise é particularmente centrada na evasão ocorrida no curso de Ciência da Computação do Campus da UFC em Quixadá.

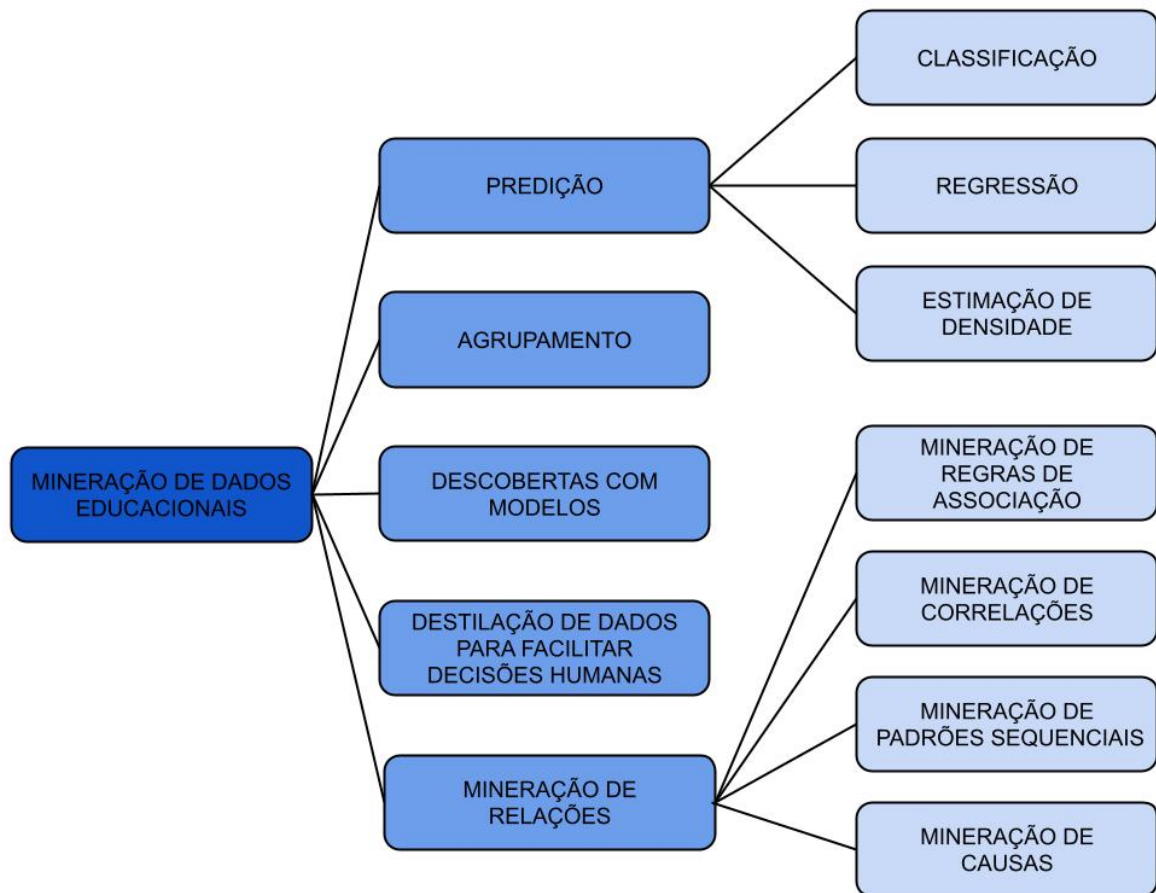
2.2 Mineração de Dados Educacionais

A Mineração de Dados Educacionais(MDE) é definida como uma área de pesquisa focada no desenvolvimento de métodos para explorar conjuntos de dados coletados em ambientes educacionais. A natureza diversa desses dados exige adaptações e o desenvolvimento de novas técnicas analíticas para lidar com as particularidades dos dados educacionais (Carvalho, 2011).

O volume crescente de dados educacionais impulsionou o uso da mineração de dados para responder a questões sobre processos de aprendizagem, desenvolvimento de materiais instrucionais e acompanhamento de alunos. A Mineração de Dados Educacionais (MDE) busca compreender melhor o comportamento de estudantes e fatores associados à aprendizagem, permitindo, por exemplo, a identificação de padrões relacionados à evasão acadêmica (Costa *et al.*, 2012).

A Figura 2 apresenta uma classificação de técnicas de MDE por sub-áreas, conforme proposto por Carvalho (2011).

Figura 2 – Diagrama hierárquico detalhando as sub-áreas da MDE



Fonte: Elaborada pelo autor.

A predição envolve a criação de modelos que prevêm resultados futuros com base em dados históricos. O agrupamento consiste em identificar grupos ou categorias dentro dos dados educacionais com características semelhantes. A mineração de relações foca na descoberta de associações entre variáveis dentro dos dados. A destilação de dados para facilitar decisões humanas apresenta dados complexos de forma compreensível, facilitando a tomada de decisões por meio de técnicas de visualização. Por fim, a descoberta com modelos utiliza modelos predefinidos para revelar novos padrões e conhecimentos nos dados educacionais (Carvalho, 2011).

A sub-área destilação de dados para facilitar decisões humanas tem como principal objetivo a apresentação de dados de forma legível, para facilitar a compreensão humana e apoiar decisões baseadas em dados. Ela é uma área de interesse crescente dentro da MDE, pois, a análise dos dados realizada por agente humanos, em sua maioria, só pode ser realizada se os dados forem apresentados de forma apropriada. O principal método dessa área da mineração de

dados educacionais é o de visualização da informação (Carvalho, 2011).

2.3 Mineração de Processos

A Mineração de Processos é uma disciplina que une a ciência dos processos e a mineração de dados, permitindo que dados de registros de eventos (*logs*) sejam utilizados para extrair conhecimento sobre os processos dentro de uma organização (Aalst, 2022). Estes dados são especialmente valiosos porque descrevem as sequências reais de atividades, proporcionando *insights* sobre o funcionamento efetivo dos processos (Aalst *et al.*, 2012).

Para a aplicação eficaz de Mineração de Processos, é essencial a existência de um *log* de eventos detalhado, como mostrado na Tabela 1. Este *log* deve incluir elementos como o *CaseId*, *Timestamp*, descrição das atividades e os recursos envolvidos. O *CaseId* é um identificador único que agrupa todos os eventos relacionados a uma única instância de processo, facilitando o rastreamento e a análise de caminhos específicos dentro do processo. Os *Timestamp* são registros de data e hora que marcam exatamente quando cada evento ocorreu, necessários para a análise temporal do processo, como a duração das atividades. A descrição das atividades fornece detalhes sobre o que cada evento representa, permitindo um entendimento de cada passo do processo. Finalmente, a identificação dos recursos envolvidos, que podem incluir pessoal, equipamentos, ou qualquer outro ativo necessário e é crucial para avaliar como os recursos são alocados e utilizados ao longo do processo (Aalst, 2016).

Tabela 1 – Um exemplo de um *log* de eventos, descrevendo o comportamento relacionado ao processo de solicitação de compensação para ingresso de concerto.

<i>Case Id</i>	<i>Timestamp</i>	<i>Activity</i>	<i>Resource</i>
...			
12373	30-7-2019 11.02	Register request	Barbara
12373	30-7-2019 11.12	Register request	Barbara
12374	30-7-2019 11.32	Register request	Jan
12374	30-7-2019 11.44	Register request	Jan
12373	30-7-2019 12.12	Check ticket	Hajo
12374	30-7-2019 14.16	Examine casually	Jorge
12375	30-7-2019 14.32	Register request	Josep
12374	30-7-2019 14.16	Examine casually	Jorge
12373	30-7-2019 14.42	Check ticket	Hajo
12375	30-7-2019 15.32	Register request	Josep
12375	30-7-2019 15.42	Examine thoroughly	Marlon
12373	03-8-2019 11.18	Examine thoroughly	Barbara
12375	03-8-2019 12.42	Examine thoroughly	Marlon
12373	03-8-2019 15.18	Examine thoroughly	Barbara
...			

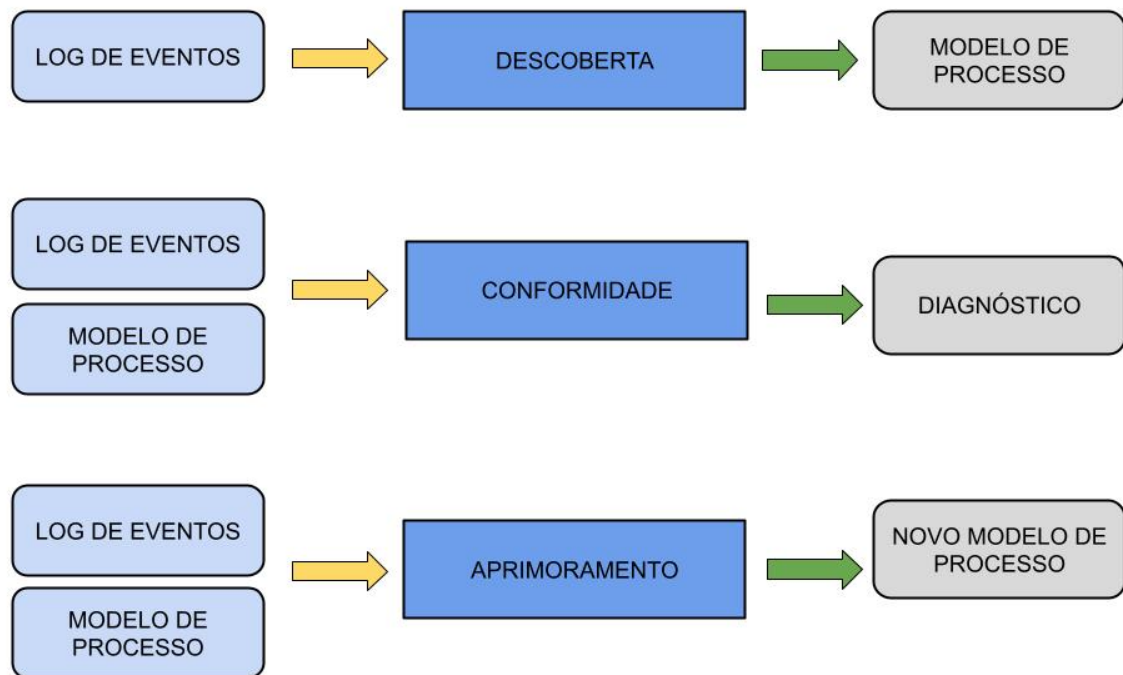
Fonte: Aalst (2016)

A Mineração de Processos oferece três tipos básicos de análise conforme destacado por Aalst *et al.* (2012):

- **Descoberta do Modelo de Processo:** envolve a criação de um modelo baseado nos dados dos registros de eventos, capturando assim o processo tal como ele é executado;
- **Verificação de Conformidade:** permite detectar, localizar, medir e explicar as discrepâncias entre as regras de negócios e o processo conforme ocorre na prática;
- **Aprimoramento:** visa complementar ou aprimorar um modelo existente para refletir melhor a realidade operacional.

Cada um dos tipos de Mineração de Processos está vinculado a objetivos específicos e pode ser aplicado em diferentes etapas do ciclo de vida dos processos. Contudo, independentemente do tipo de mineração praticado, todos eles mantêm uma relação intrínseca com os modelos de processos e os *logs* de eventos. Esta relação é ilustrada na Figura 3, que detalha os tipos básicos de Mineração de Processos em termos de suas entradas e saídas.

Figura 3 – Tipos de Mineração de Processos com relação às entradas e saídas.



Fonte: Adaptado de Aalst *et al.* (2012).

O tipo de Mineração de Processos de descoberta recebe como entrada um *log* de eventos e tem como saída um modelo de processo que processo. A conformidade tem como entrada tanto um *log* de eventos quanto um modelo de processo e retorna um diagnóstico que identifica as diferenças e semelhanças entre as duas entradas. Por fim, o aprimoramento também recebe duas entradas, o *log* de eventos e o modelo de processo, e retorna um novo modelo de processo otimizado.

Replay é uma técnica de verificação de conformidade que permite verificar a aderência dos eventos reais ao modelo de processo estabelecido (Aalst, 2022). Esta técnica envolve a simulação dos *logs* de eventos sobre o modelo de processo, possibilitando a observação direta de como as atividades reais se alinham, ou divergem, do modelo teórico. O *replay* ajuda a detectar, quantificar e entender as discrepâncias entre o processo modelado e sua execução real (Aalst, 2022).

Especificamente em contextos educacionais, a Mineração de Processos emerge como uma área de pesquisa promissora, permitindo análises detalhadas sobre a aprendizagem baseada em dados. Este enfoque não só ajuda a entender como os alunos aprendem, mas também suporta a tomada de decisões estratégicas para melhorar a educação (Aalst *et al.*, 2012).

2.4 Modelagem de processos

As técnicas de Mineração de Processos necessitam de uma modelagem formal para serem eficazmente aplicadas. Isso se deve à complexidade e à variabilidade dos processos reais, que exigem representações precisas para análise e melhoria (Aalst, 2016).

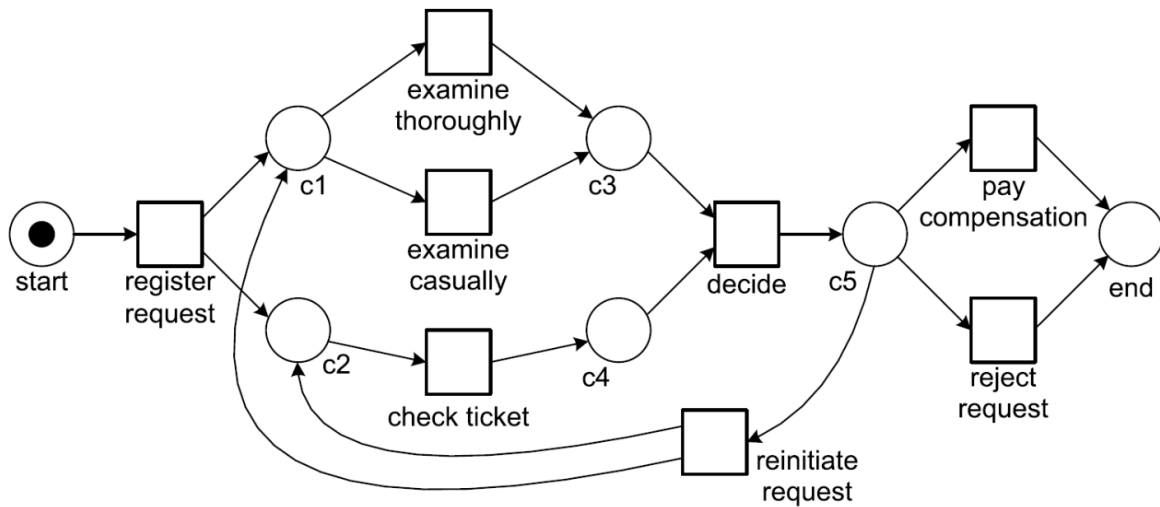
Uma modelagem de processos fornece uma estrutura que captura todos os aspectos essenciais do processo, desde a sequência de atividades até as interações entre recursos e dados. Além disso, facilita a verificação de conformidade e a simulação de cenários hipotéticos, proporcionando um entendimento mais profundo e uma maior capacidade de otimização dos processos. Sem uma base formal robusta, as análises podem ser imprecisas e os *insights* obtidos menos confiáveis, comprometendo a eficácia das intervenções propostas (Aalst, 2016).

A rede de Petri é um modelo matemático para modelagem de sistemas que consegue capturar tanto a estrutura estática quanto o comportamento dinâmico, conforme descrito por Desel e Reisig (1996). Esta abordagem tem sido adotada para a modelagem de processos de negócio desde sua introdução no contexto por Aalst (1998).

As Redes de Petri são representadas com estados desenhados como círculos e transições como retângulos. Os estados, também conhecidos como lugares, representam as condições ou situações no sistema, enquanto as transições representam os eventos que podem ocorrer, alterando essas condições. As setas conectam lugares a transições e vice-versa, indicando a direção do fluxo de controle (Aalst, 2016).

A Figura 4 ilustra uma rede de Petri correspondente ao *log* de eventos descrito na Tabela 1. Na figura, cada círculo representa um estado ou condição do processo (por exemplo, "*c1*", "*c2*", "*c3*"), e cada retângulo representa uma transição ou evento (por exemplo, "*register request*", "*check ticket*"). As setas mostram a sequência em que as atividades ocorrem e como o estado do processo muda de um lugar para outro através das transições. O processo começa no estado "*start*", indicado pelo círculo preenchido, e progride através de vários estados e transições até alcançar o estado "*end*", que representa a conclusão do processo.

Figura 4 – Exemplo de visualização do modelo de processo usando Rede de Petri.



Fonte: Aalst (2016).

A representação por Redes de Petri viabiliza a aplicação da técnica de *Replay* baseado em *tokens* (*Token-based Replay*). Essa abordagem consiste no consumo e produção de *tokens* dentro da Rede de Petri, permitindo avaliar a conformidade entre um *log* de eventos e um modelo de processo.

Durante o *replay*, algumas métricas são utilizadas para mensurar o grau de aderência entre a execução real e o modelo teórico, incluindo:

- *Token-Fitness*: Mede a porcentagem de eventos e trajetórias do *log* que podem ser reproduzidos pelo modelo, variando de 0 (completamente desalinhado) a 1 (perfeitamente aderente).
- *Missing Tokens*: Representam *tokens* esperados no modelo para a execução de uma atividade registrada no *log*, mas que estavam ausentes.
- *Consumed Tokens* e *Produced Tokens*: Indicadores da quantidade total de transições corretamente processadas dentro do modelo.
- *Remaining Tokens*: Quantidade de *tokens* remanescentes no modelo sem serem consumidos, sugerindo possíveis interrupções ou desvios no processo.

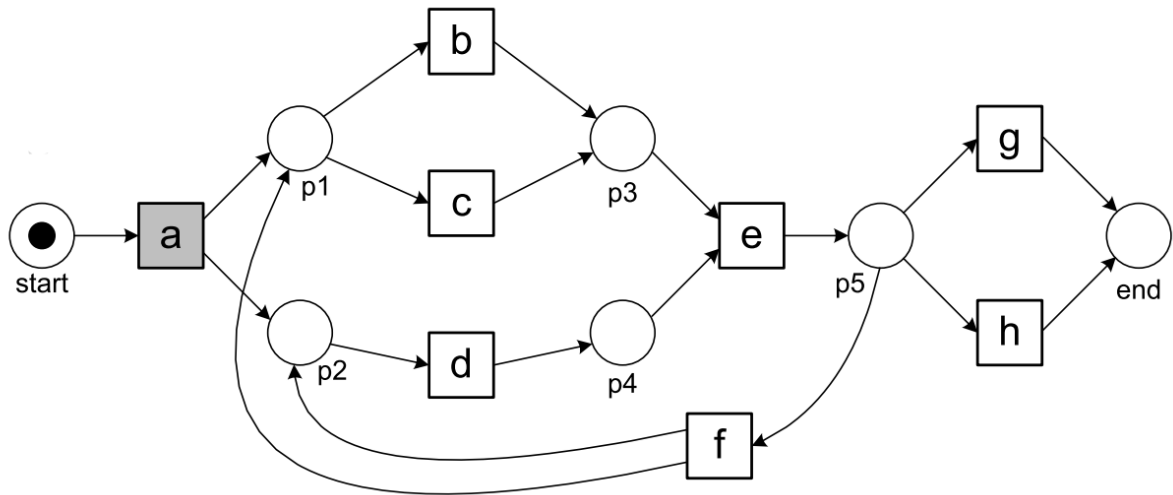
A execução do *replay* consiste em reproduzir cada evento registrado no *log* dentro do modelo de processo, movimentando *tokens* pelos lugares da Rede de Petri conforme as regras definidas. O fluxo de *tokens* segue a seguinte dinâmica:

- Consumo de *tokens*: Quando uma transição é disparada, os *tokens* presentes nos lugares de entrada são consumidos, permitindo sua execução.

- Produção de *tokens*: Após o disparo da transição, novos *tokens* são gerados e depositados nos lugares de saída, representando a progressão do processo.

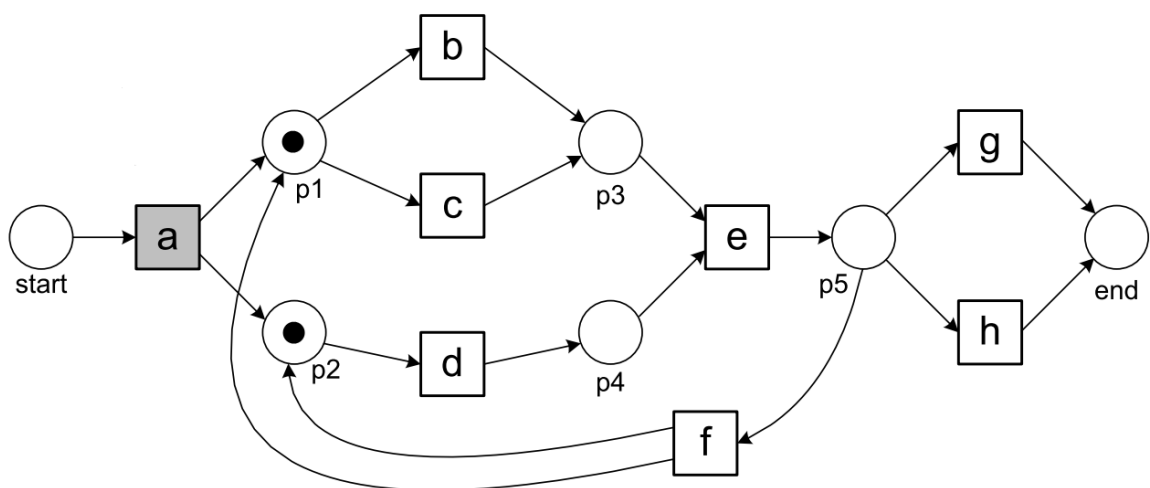
A Figura 5 ilustra a configuração inicial de uma Rede de Petri, onde um *token* está presente no *place start*, que é um lugar de entrada para a transição *a*. Quando essa transição é disparada, o *token* em *start* é consumido, e dois novos *tokens* são produzidos e depositados nos lugares de saída *p1* e *p2*, conforme mostrado na Figura 6.

Figura 5 – Representação de Rede de Petri com *token* no *place start*.



Fonte: Aalst (2016).

Figura 6 – Representação de consumo e produção de *tokens* na Rede de Petri.



Fonte: Aalst (2016).

Se uma atividade ocorrer no *log* sem que os *tokens* necessários estejam disponíveis, *tokens* ausentes (*missing tokens*) são registrados, indicando um desvio. Da mesma forma,

se *tokens* permanecerem no modelo sem serem consumidos ao final da execução, eles são classificados como *tokens* remanescentes (*remaining tokens*), sugerindo trajetórias incompletas (Aalst, 2016).

A modelagem de processos oferece uma maneira de representar a estrutura curricular de um curso. Ao mapear disciplinas e pré-requisitos como elementos e fluxos, essas ferramentas ajudam a visualizar o progresso dos alunos através de *tokens* ou etapas, garantindo que os requisitos sejam atendidos de forma organizada, facilitando a aplicação das técnicas de Mineração de Processos.

3 TRABALHOS RELACIONADOS

Este capítulo apresenta alguns trabalhos relacionados destacando as semelhanças e diferenças com este trabalho.

3.1 Mineração de Dados Educacionais para Predição da Evasão em Cursos de Graduação Presenciais no Ensino Superior

Este estudo conduzido por Kantorski *et al.* (2023) investiga o fenômeno da evasão acadêmica no ensino superior na Universidade Federal de Santa Maria (UFSM). Os autores propõem uma abordagem que combina técnicas de MDE e aprendizado supervisionado para analisar e prever o abandono estudantil em cursos presenciais da UFSM. O foco do estudo é explorar como diferentes métodos de classificação, quando combinados com técnicas de ciência de dados e ferramentas de inteligência de negócios, podem contribuir para um modelo preditivo robusto e eficaz.

Os autores utilizaram uma variedade de algoritmos de aprendizado supervisionado juntamente com a metodologia Cross Industry Standard Process for Data Mining (CRISP-DM) para prever a evasão. A abordagem incluiu quatro fases: compreensão da evasão, análise e preparação de dados, modelagem, e avaliação. Eles desenvolveram modelos preditivos usando dados de mais de 79.000 estudantes de 106 cursos durante 3 períodos letivos, aplicando técnicas como árvores de decisão, Naïve Bayes, e perceptrons multicamadas.

O estudo conseguiu identificar estudantes com alto risco de evasão, alcançando um *recall* médio de 92%. Apesar de uma precisão relativamente baixa, a capacidade de prever evasões permitiu intervenções proativas para mitigar o abandono escolar. Os resultados destacam a eficácia de combinar técnicas de mineração de dados educacionais com métodos de aprendizado de máquina para prever a evasão.

O trabalho de Kantorski *et al.* (2023) assemelha-se a este por utilizar dados de históricos acadêmicos dos discentes para identificar perfis de evasão em cursos de graduação presencial. Contudo, este estudo se distingue pela utilização de técnicas de Mineração de Processos.

3.2 Análise da Trajetória de Aprendizagem do aluno no ensino a distância por meio da Mineração de Processos

O estudo de Unger *et al.* (2022) analisa a interação dos alunos com o Ambiente Virtual de Aprendizagem (AVA) Canvas da Universidade Virtual do Estado de São Paulo (UNIVESP), utilizando técnicas de Mineração de Processos para mapear a trajetória de aprendizagem. A pesquisa foca especificamente na trajetória de aprendizagem em quatro disciplinas iniciais, que são comuns a dez cursos. O objetivo principal é identificar padrões de comportamento e pontos de engajamento e desengajamento dos alunos em cursos de educação a distância.

Para realizar esta análise, o estudo adota a metodologia Process Mining Project Methodology (PM2), que é dividida em seis fases distintas: planejamento, que define o escopo e os objetivos; extração de dados, que envolve coletar os registros de log do AVA; processamento de dados, onde os dados são limpos e preparados; mineração de dados, aplicando algoritmos de Mineração de Processos para extrair modelos de descoberta de processos; análise desses modelos para entender os comportamentos e padrões; e, por fim, a fase de avaliação e refinamento, onde os resultados são validados e propostas melhorias.

A técnica de Descoberta do Modelo de Processo foi aplicada para analisar os *logs* de eventos do AVA, que incluem informações como identificadores de alunos, atividades realizadas (como acesso a conteúdos, *quizzes* e fóruns) e *timestamps* (data e hora das interações). Esses *logs*, totalizando 15.820 casos e 703.425 eventos, foram processados e analisados utilizando a ferramenta Celonis. Os modelos gerados identificaram as trajetórias mais comuns de interação dos alunos, além de gargalos e padrões de comportamento, permitindo uma análise comparativa entre alunos aprovados e reprovados.

A análise dos *logs* revelou que as atividades de conteúdo assíncrono, como recursos de *wiki* e *quizzes*, são as mais frequentadas pelos estudantes. No entanto, foi observado que, apesar da alta inicialização nessas atividades, existe um decréscimo significativo na frequência de acessos ao longo das semanas, sugerindo períodos de desengajamento. Além disso, a comparação entre alunos aprovados e reprovados mostrou que os aprovados interagem com o sistema de maneira mais consistente e contínua, evitando assim os gargalos identificados nos percursos dos alunos reprovados, que apresentam maior irregularidade no acesso e menor interação com atividades-chave.

Embora o estudo de Unger *et al.* (2022) e este trabalho assemelhem-se pelo uso de técnicas de Mineração de Processos para analisar dados educacionais, eles se distinguem

significativamente em termos de foco e aplicabilidade. Enquanto a análise de Unger *et al.* (2022) se concentra na interação individual dos alunos com disciplinas específicas dentro de um ambiente virtual, buscando entender o engajamento dos alunos em cursos online, este trabalho direciona-se à identificação de pontos críticos estruturas curriculares dos cursos.

3.3 Mineração de dados para identificação do perfil de evasão de alunos da UFC - Campus Quixadá

O estudo conduzido por LIMA JÚNIOR, A. A. de (2019) explora a utilização da mineração de dados para identificar perfis de alunos propensos à evasão no curso de Ciência da Computação do Campus da UFC em Quixadá. A pesquisa destaca a crescente preocupação com as altas taxas de evasão escolar em cursos de graduação, representando significativas perdas para alunos, instituições e para o sistema educacional como um todo. O objetivo central é aplicar técnicas avançadas de mineração de dados para extrair padrões e perfis de alunos que possam indicar riscos iminentes de evasão, visando fornecer subsídios para ações preventivas por parte da administração acadêmica.

LIMA JÚNIOR, A. A. de (2019) utiliza a ferramenta de análise de dados *Orange* para aplicar algoritmos de aprendizado de máquina, incluindo *Decision Trees*, *Random Forests*, *Support Vector Machine (SVM)* e *K-Nearest Neighbors (KNN)*. O estudo é realizado em dois cenários distintos: o primeiro com o conjunto de dados original, que apresenta um desequilíbrio natural entre as classes de alunos ativos, evadidos e concluintes; e o segundo, uma versão balanceada desses dados para verificar a consistência dos modelos preditivos. Foram coletados registros históricos dos alunos do ano de 2013 a 2018. A seleção de dados foi focada nos alunos do primeiro ano do curso, baseando-se na premissa de que a maioria dos casos de evasão ocorrem nesse período.

Os resultados obtidos por LIMA JÚNIOR, A. A. de (2019) demonstram uma alta eficácia dos modelos aplicados, com acurácias alcançando até 99% na identificação de alunos com risco de evasão no cenário de dados desbalanceados. No segundo cenário, os resultados foram semelhantes, com acurácias próximas a 95,5%. Os desempenhos satisfatórios dos classificadores nos dois cenários indica que os atributos utilizados são suficientes para realizar a previsão dos estudantes com risco de evasão. Ademais, perfis detalhados de risco foram identificados, destacando fatores como baixo Índice de Rendimento Acadêmico (IRA), número de reprovações e pouca integração curricular como indicativos de maior probabilidade de

abandono do curso.

O estudo de LIMA JÚNIOR, A. A. de (2019) é similar ao presente trabalho ao investigar o fenômeno da evasão acadêmica no curso de Ciência da Computação do Campus da UFC em Quixadá, utilizando registros de histórico acadêmico. No entanto, diferenciam-se pelo uso de técnicas distintas de mineração de dados.

3.4 Quadro comparativo

No Quadro 1, é realizada uma comparação entre as abordagens deste trabalho e as dos trabalhos relacionados. Essa comparação considera as técnicas utilizadas, a origem e o tipo dos dados coletados, bem como a modalidade de ensino.

Quadro 1 – Quadro comparativo das abordagens dos trabalhos relacionados e deste trabalho.

Trabalho	Técnicas utilizadas	Origem dos dados	Modalidade de ensino
<i>Esse trabalho</i>	Análise de conformidade	Histórico acadêmico	Presencial
<i>Kantorski et al. (2023)</i>	Aprendizado supervisionado e classificação	Histórico acadêmico e dados socioeconômicos e socioculturais	Presencial
<i>Unger et al. (2022)</i>	Descoberta de processos	<i>Logs</i> de interações em ambientes virtuais de ensino	EAD
<i>LIMA JÚNIOR, A. A. de (2019)</i>	Classificação	Histórico acadêmico	Presencial

Fonte: Elaborado pelo autor.

4 MINERAÇÃO DE PROCESSOS: ANÁLISE DE CONFORMIDADE

Neste capítulo, é apresentada a aplicação de Mineração de Processos no contexto da trajetória acadêmica dos estudantes do curso de Bacharelado em Ciência da Computação da UFC - Campus Quixadá, com o objetivo de avaliar a conformidade do percurso dos alunos em relação à estrutura curricular formalmente estabelecida.

A Seção 4.1 descreve a modelagem da estrutura curricular por meio de uma Rede de Petri. Em seguida, a Seção 4.2 detalha a preparação dos dados e a construção do *log* de eventos. Na Seção 4.3, é realizada a análise de conformidade por meio do *replay*. Por fim, a Seção 4.4 apresenta a identificação de pontos críticos, destacando as disciplinas com maior acúmulo de inconformidades.

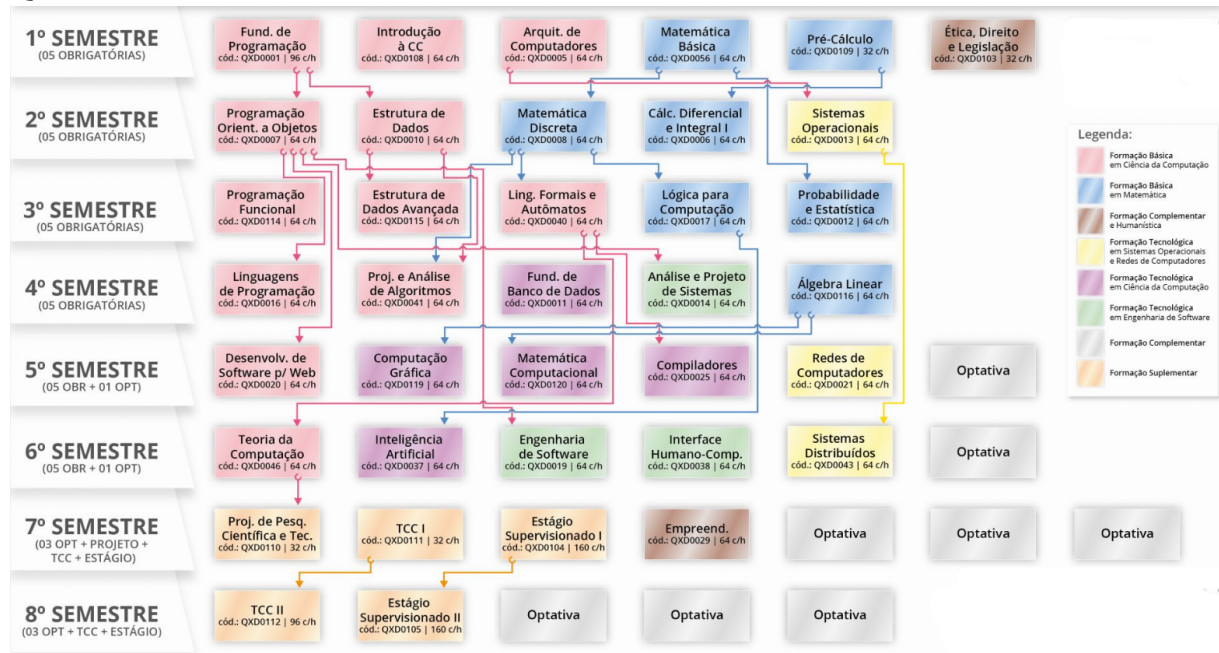
4.1 Modelagem Formal da Estrutura Curricular

A estrutura curricular do curso de Ciência da Computação é apresentada na Figura 7. Essa estrutura refere-se ao currículo de 2013.1 e está organizada em oito semestres, onde cada linha representa as disciplinas correspondentes a um semestre. A matriz curricular é composta por disciplinas obrigatórias, optativas (representadas na imagem como formação complementar) e atividades curriculares (denominadas formação suplementar), conforme descrito no Projeto Pedagógico do Curso ¹.

As disciplinas estão organizadas por áreas do conhecimento, diferenciadas por cores distintas, conforme indicado no campo de legenda da Figura 7. Por exemplo, as disciplinas destacadas em azul — Matemática Básica, Pré-Cálculo, Matemática Discreta, Cálculo Diferencial e Integral, Lógica para Computação, Probabilidade e Estatística e Álgebra Linear — pertencem à área de conhecimento Formação Básica em Matemática. Da mesma forma, as demais cores representam diferentes áreas do conhecimento, cada uma agrupando disciplinas relacionadas.

¹ Disponível em: <https://cc.quixada.ufc.br/wp-content/uploads/2024/04/20221013-Projeto-Pedagogico-do-Curso-Curriculo2013.1-22.10.13.pdf>

Figura 7 – Estrutura curricular do curso de Ciência da Computação da UFC do Campus de Quixadá.



Fonte: Disponível em <https://cc.quixada.ufc.br/estrutura-curricular/fluxograma-de-disciplinas/>

A organização da estrutura curricular também envolve relações de pré-requisitos entre disciplinas. Nesse contexto, uma disciplina é considerada pré-requisito quando há uma seta partindo dela em direção a outra disciplina. Da mesma forma, uma disciplina possui pré-requisitos quando está conectada a outra por uma seta. Esse comportamento pode ser observado na Figura 7, onde, por exemplo, a disciplina Arquitetura de Computadores é um pré-requisito para Sistemas Operacionais. Dessa forma, Sistemas Operacionais tem como pré-requisito a disciplina de Arquitetura de Computadores.

Neste estudo, foram modeladas somente as disciplinas obrigatórias. As disciplinas optativas (representadas na imagem como formação complementar) e as atividades curriculares (denominadas formação suplementar) não foram incluídas na modelagem.

Para a modelagem formal da estrutura curricular, optou-se pela notação de Rede de Petri, devido à sua maior adaptabilidade na representação da estrutura.

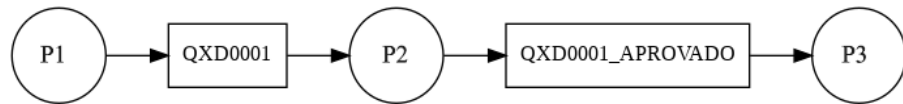
4.1.1 Estrutura da Modelagem

Cada disciplina foi representada como uma estrutura composta por três lugares e duas transições, conforme ilustrado na Figura 8:

- **Lugar inicial (P1):** Representa a disciplina antes de ser cursada.

- **Transição com o código da disciplina:** Indica a realização da disciplina.
- **Lugar intermediário (P2):** Representa já cursada e ainda sem a aprovação.
- **Transição de aprovação** (ex: *QXD001_APROVADO*): Representa a conclusão bem-sucedida da disciplina.
- **Lugar final (P3):** Indica que a disciplina foi aprovada.

Figura 8 – Conjunto de transições e places que representam a disciplina QXD0001.



Fonte: Elaborado pelo autor.

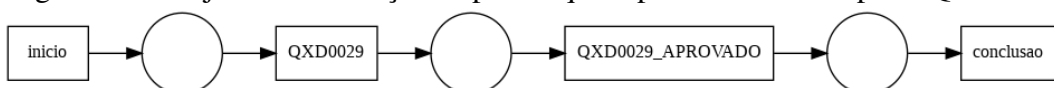
A modelagem inicia-se com a adição de uma transição denominada "**início**", responsável por disparar *tokens* para os lugares iniciais (**P1**) das disciplinas que não possuem pré-requisito. Além disso, há a transição "**conclusao**", que controla a finalização do processo, garantindo que só seja disparada quando todas as disciplinas tiverem sido aprovadas, conforme apresentado na Figura 10.

4.1.2 Conexões entre as Disciplinas

As disciplinas que **não possuem pré-requisitos** são conectadas diretamente pela transição "**início**", como mostrado no lugar inicial **P1** da Figura 9. Já as disciplinas que possuem pré-requisitos têm seu **P1** conectado ao **P3** das disciplinas que devem ser cursadas previamente.

Por outro lado, as disciplinas que **não são pré-requisitos** para nenhuma outra têm seu **P3** conectado à transição "**conclusao**", conforme a Figura 9. Essa transição só pode ser disparada quando todos os *places* que a antecedem estiverem marcados com *tokens*, garantindo assim que todas as disciplinas necessárias tenham sido aprovadas antes da conclusão do processo.

Figura 9 – Conjunto de transições e places que representam a disciplina QXD0029.

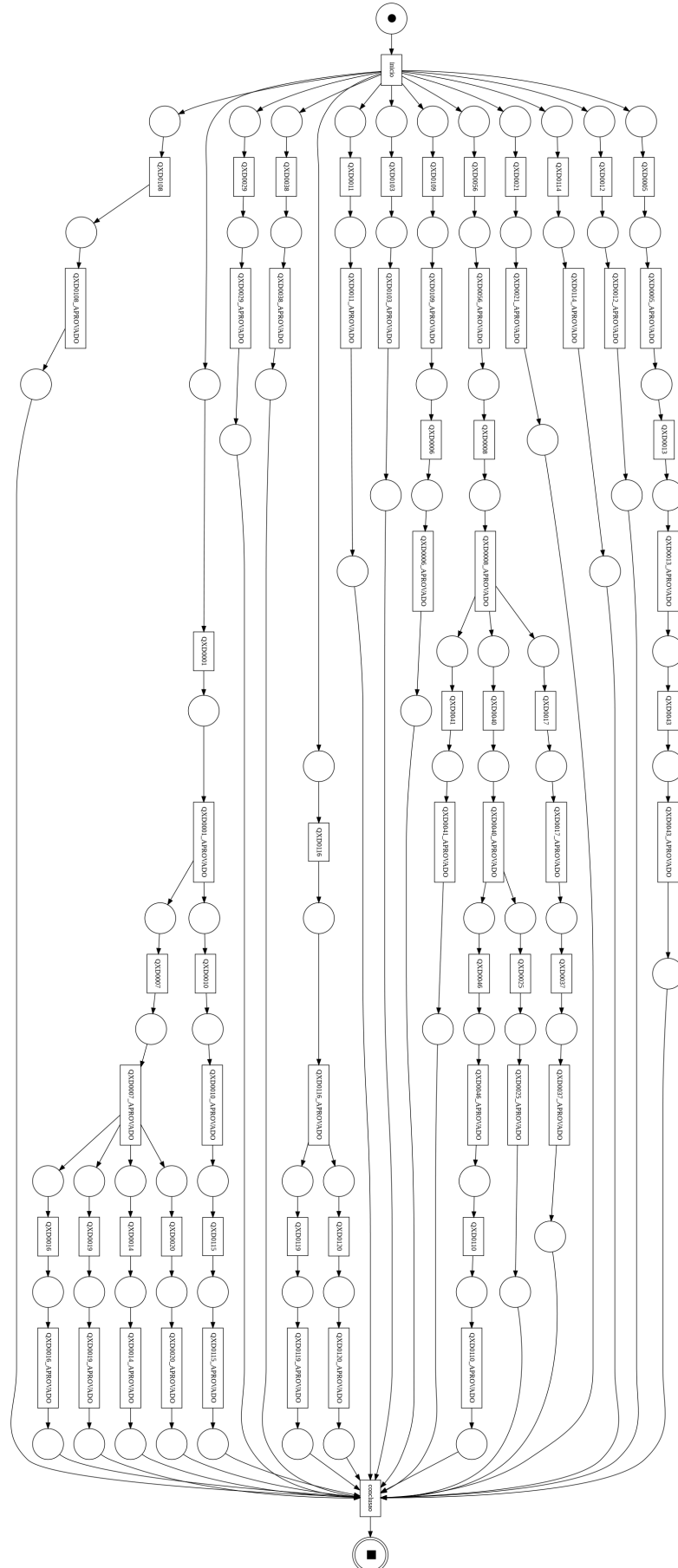


Fonte: Elaborado pelo autor.

4.1.3 Modelo Completo da Estrutura Curricular

A Figura 10 apresenta uma visão geral do modelo completo, mostrando a relação entre as disciplinas e suas dependências.

Figura 10 – Estrutura curricular modelada em Rede de Petri.



Fonte: Elaborado pelo autor.

4.2 Preparação dos Dados e Criação do *Log* de Eventos

Neste estudo, foram utilizados dados históricos dos estudantes do curso de Bacharelado em Ciência da Computação do Campus da UFC em Quixadá, fornecidos pela direção do campus. Os registros, que abrangem o período de 2013 a 2024, incluem informações de alunos com matrículas ativas e inativas, englobando aqueles que concluíram o curso, abandonaram ou ainda estão cursando. A privacidade dos alunos foi assegurada por meio da anonimização de seus nomes e matrículas. A base de dados contém um total de 550 registros de alunos.

Os dados incluem informações detalhadas sobre cada disciplina cursada pelos alunos. Estas informações abrangem o semestre e o ano em que a disciplina foi cursada, a carga horária, o código da disciplina e o status da turma (indicando se a turma está consolidada ou aberta, representando turmas ativas no semestre atual). Além disso, são registrados os resultados dos alunos em cada disciplina, especificando se foram aprovados, reprovados, suprimiram a disciplina ou foram reprovados por falta. Os dados são apresentados na Figura 11.

Figura 11 – Dados brutos dos estudantes do curso de Bacharelado em Ciência da Computação.

	ano	periodo	codigo_turma	codigo	nome_disciplina	id_tipo_componente	descricao	id_situacao_turma	id_discente	id_situacao_matricula	resultado	id_curriculo	ano_ingresso	periodo_ingresso	id_forma_ingresso
0	2023	1	05A	QXD0005	ARQUITETURA DE COMPUTADORES	2	DISCIPLINA	3	999000	4	APROVADO	13730020	2023	1	45
1	2023	1	8	QXD0103	ÉTICA, DIREITO E LEGISLAÇÃO	2	DISCIPLINA	3	999000	4	APROVADO	13730020	2023	1	45
2	2023	1	12A	QXD0001	FUNDAMENTOS DE PROGRAMAÇÃO	2	DISCIPLINA	3	999000	4	APROVADO	13730020	2023	1	45
3	2023	1	4	QXD0108	INTRODUÇÃO A CIÊNCIA DA COMPUTAÇÃO	2	DISCIPLINA	3	999000	4	APROVADO	13730020	2023	1	45
4	2023	1	6	QXD0056	MATEMÁTICA BÁSICA	2	DISCIPLINA	3	999000	4	APROVADO	13730020	2023	1	45
5	2023	1	3	QXD0109	PRÉ-CÁLCULO	2	DISCIPLINA	3	999000	4	APROVADO	13730020	2023	1	45
6	2023	2	02A	QXD0006	CÁLCULO DIFERENCIAL E INTEGRAL I	2	DISCIPLINA	3	999000	4	APROVADO	13730020	2023	1	45
7	2023	2	02A	QXD0010	ESTRUTURA DE DADOS	2	DISCIPLINA	3	999000	4	APROVADO	13730020	2023	1	45
8	2023	2	03A	QXD0008	MATEMÁTICA DISCRETA	2	DISCIPLINA	3	999000	4	APROVADO	13730020	2023	1	45
9	2023	2	05A	QXD0007	PROGRAMAÇÃO ORIENTADA A OBJETOS	2	DISCIPLINA	3	999000	4	APROVADO	13730020	2023	1	45

Fonte: Elaborado pelo autor.

Para preparar esses dados para a análise de Mineração de Processos, realizou-se uma etapa de pré-processamento. Nessa fase, colunas consideradas irrelevantes para a análise foram removidas, incluindo *id_curso*, *forma_de_ingresso*, *periodo_de_ingresso*, *id_curriculo* e *ch_dedicada_periodo*. Além disso, foram excluídos registros com valores ausentes, disciplinas optativas e disciplinas com o *status* "MATRICULADO". Com os dados limpos, eles foram transformados em um *log* de eventos, conforme ilustrado na Figura 12.

A criação do *log* de eventos apresentado na Figura 12 envolveu os seguintes passos:

- **Seleção do *CaseId*:** Definiu-se a coluna *id_discente* para representar o *CaseId*.
- **Combinação de *ano* e *periodo* para criar o *Timestamp*:** Foi criada uma nova coluna

de *Timestamp* combinando os valores das colunas *ano* e *periodo*. O início do primeiro semestre foi padronizado como 01/02 e o do segundo semestre como 01/08.

- **Combinação de *codigo* e *resultado* para criar a *Activity***: Foi criada uma nova coluna de *Activity* combinando os valores das colunas *codigo* e *resultado*, separados por um sublinhado (_).

Figura 12 – *Log* de eventos criado nesta etapa.

	<i>id_discente</i>	<i>codigo</i>	<i>resultado</i>	<i>timestamp</i>
0	999000	QXD0005	APROVADO	01/02/2023
1	999000	QXD0103	APROVADO	01/02/2023
2	999000	QXD0001	APROVADO	01/02/2023
3	999000	QXD0108	APROVADO	01/02/2023
4	999000	QXD0056	APROVADO	01/02/2023
...	...	...	...	...
25855	999549	QXD0040_APROVADO		- 01/06/2023
25856	999549	QXD0120_APROVADO		- 01/06/2023
25857	999549	QXD0038_APROVADO		- 01/12/2023
25858	999549	QXD0043_APROVADO		- 01/12/2023
25859	999549	QXD0046_REPROVADO		- 01/12/2023

Fonte: Elaborado pelo autor.

Para cada aluno, foram adicionadas duas atividades na coluna *activity*: “início” e, quando aplicável, “conclusao”. A atividade “início” foi registrada um mês antes do primeiro registro de disciplina, garantindo que durante a execução do *replay*, ela seja sempre a primeira a ser executada. Já a atividade “conclusao” foi adicionada apenas para os alunos que foram aprovados em todas as disciplinas obrigatórias, sendo registrada um dia após a última disciplina cursada, conforme ilustrado na Figura 13.

Figura 13 – *Log* de eventos de um aluno que foi aprovado em todas as disciplinas obrigatórias.

	id_discente	activity	timestamp
101	999002	inicio	2013-01-01
102	999002	QXD0005	2013-01-02
103	999002	QXD0103	2013-01-02
104	999002	QXD0001	2013-01-02
105	999002	QXD0108	2013-01-02
...	...	...	...
164	999002	QXD0029	2016-01-02
165	999002	QXD0110	2016-01-02
166	999002	QXD0110_APROVADO	2016-01-06
167	999002	QXD0029_APROVADO	2016-01-06
168	999002	conclusao	2016-01-07

Fonte: Elaborado pelo autor.

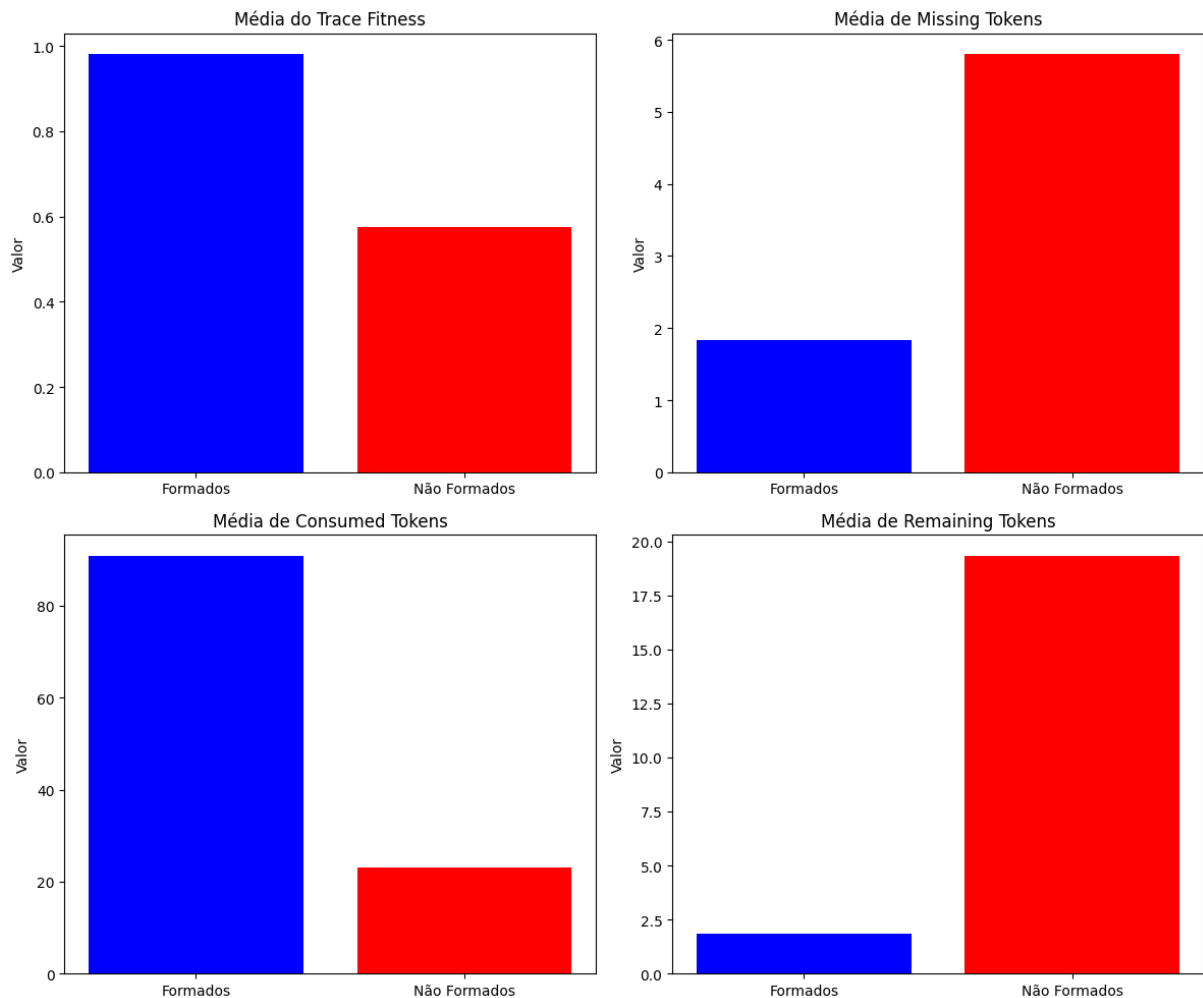
Estes procedimentos garantem que os dados estejam estruturados de maneira adequada para a aplicação do *replay* que será explorada na próxima Seção.

4.3 Análise de Conformidade via *Replay*

A análise de conformidade foi conduzida por meio da técnica de *replay*, aplicada ao *log* de eventos e à Rede de Petri modelada. Esse processo permitiu avaliar o alinhamento dos percursos acadêmicos dos alunos com a estrutura curricular esperada, quantificando os desvios por meio de métricas como *trace fitness*, *missing tokens* (tokens ausentes), *remaining tokens* (tokens remanescentes), entre outras.

Para a execução do *replay*, foram considerados exclusivamente os registros dos alunos ingressantes entre 2013 e 2019, garantindo que a análise contemplasse apenas aqueles que tiveram tempo suficiente para concluir o curso no prazo padrão. A partir dos resultados, foram geradas estatísticas que permitiram comparar o comportamento dos alunos que concluíram o curso (formados) com aqueles que não o concluíram (não formados). A Figura 14 apresenta um resumo dessas estatísticas.

Figura 14 – Comparativo de Estatísticas do *Replay* entre alunos formados e aluno não formados



Fonte: Elaborado pelo autor.

A análise comparativa revelou diferenças entre os dois grupos:

- O *trace fitness* médio dos alunos formados foi de 0.98, indicando que suas trajetórias acadêmicas seguiram melhor a estrutura curricular modelada. Em contraste, os não formados apresentaram um *fitness* médio de 0.57, demonstrando desvios consideráveis em seus percursos.
- A métrica de *missing tokens*, que representa disciplinas esperadas mas não registradas, foi menor nos alunos formados (1.83) do que nos não formados (5.80), sugerindo que os formados seguiram a progressão curricular esperada com menos interrupções.
- A quantidade de *remaining tokens*, que indica eventos não finalizados dentro do modelo, foi maior nos alunos não formados (19.32) em comparação com os formados (1.83), refletindo a interrupção dos percursos acadêmicos antes da conclusão do curso.
- Os alunos formados consumiram e produziram mais *tokens*, indicando que avançaram

mais no fluxo acadêmico e concluíram mais transições esperadas.

Os resultados indicam que os alunos formados apresentaram trajetórias mais alinhadas ao modelo curricular, com menos interrupções e maior aderência. Já os alunos não formados mostraram trajetórias mais fragmentadas, refletindo possíveis reprovações, desistências ou mudanças no percurso acadêmico.

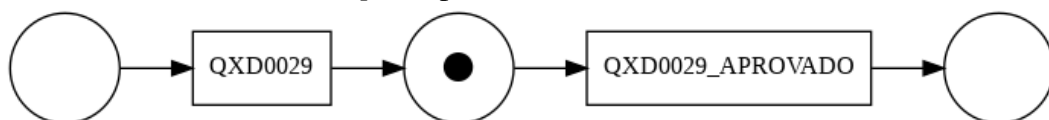
No entanto, é importante ressaltar que a análise de uma métrica isoladamente pode levar a conclusões imprecisas. Um aluno pode, por exemplo, apresentar um *trace fitness* relativamente alto, mas ainda assim não ter concluído o curso, o que significaria que ele teve alta taxa de aprovação até o momento em que ele evadiu. Da mesma forma, um número elevado de *tokens* remanescentes pode indicar tanto uma evasão precoce quanto um prolongamento do tempo de curso por fatores como reprovações. Assim, a análise de conformidade deve ser complementada por outras abordagens.

4.4 Identificação de Pontos Críticos

Nesta etapa, foram identificados os principais pontos críticos da estrutura curricular do curso. Para essa análise, considerou-se a métrica *remaining_tokens* do *replay*, que representa a quantidade de *tokens* remanescentes em cada *place* da Rede de Petri. Essa métrica responde quais disciplinas acumulam mais *tokens* devido a inconformidades no percurso acadêmico dos alunos, como reprovações, trancamentos e supressões.

Como descrito anteriormente, no modelo proposto, cada disciplina é representada por uma sequência de transições e *places*. Assim, foi necessário analisar quais *places* estavam acumulando *tokens* e indicando eventuais dificuldades dos alunos. Durante essa análise, observou-se que, quando um aluno possui atividades que não seguem a progressão esperada, os *tokens* tendem a se acumular no *place* **P2**. Conforme destacado na Figura 15.

Figura 15 – Conjunto de transições e places que representam a disciplina QXD0001 com tokens acumulados no *place* p2.



Fonte: Elaborado pelo autor.

Com base nessa observação, foram aplicados filtros para identificar, em cada disciplina, o número de *tokens* remanescentes em P2, permitindo determinar quais disciplinas

apresentavam os maiores índices de inconformidade. A Tabela 2 apresenta as 10 disciplinas com maior acúmulo de *tokens* remanescentes.

Tabela 2 – Número de Tokens por Disciplina

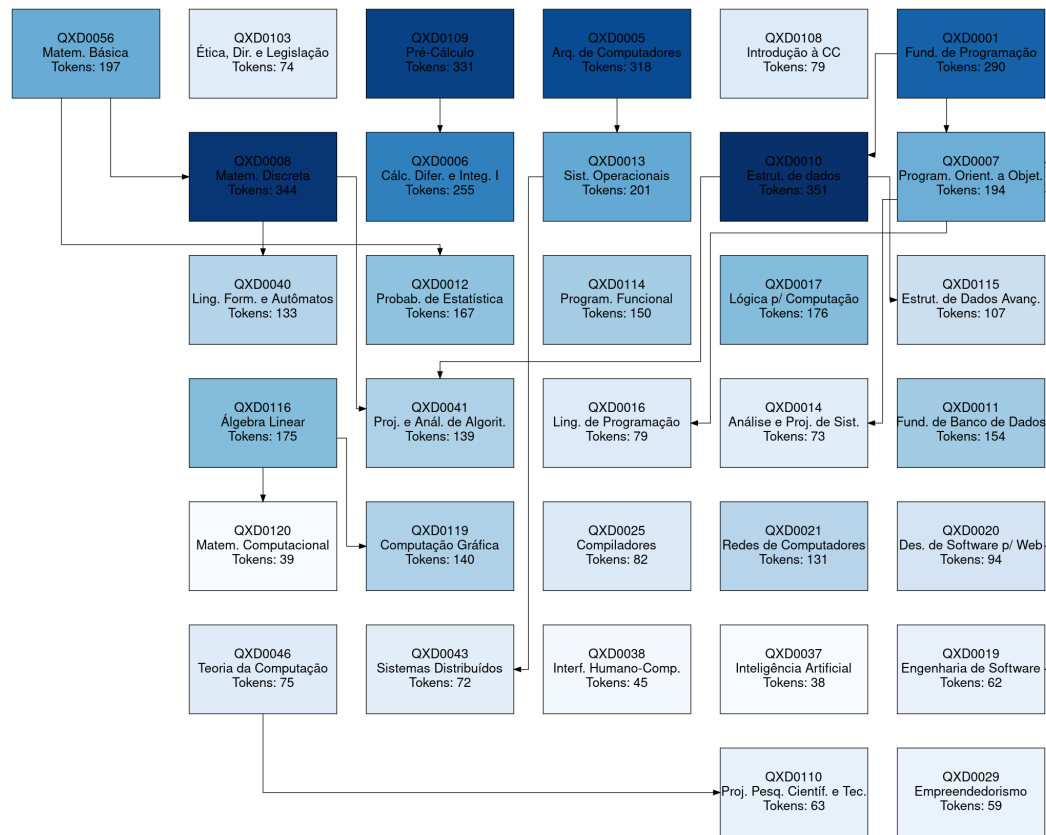
Código	Nome	Tokens restantes
QXD0010	Estrutura de Dados	351
QXD0008	Matemática Discreta	344
QXD0109	Pré-Cálculo	331
QXD0005	Arquitetura de Computadores	318
QXD0001	Fundamentos de Programação	290
QXD0006	Cálculo Diferencial e Integral I	255
QXD0013	Sistemas Operacionais	201
QXD0056	Matemática Básica	197
QXD0007	Programação Orientada a Objetos.	194
QXD0017	Lógica para Computação	176

4.4.1 Análise dos Resultados

A análise dos dados revela que, entre as dez disciplinas com maior acúmulo de *tokens* remanescentes, nove pertencem ao primeiro e segundo semestres do curso. Esse resultado indica que os primeiros semestres concentram a maior parte das reprovações, trancamentos e supressões, sugerindo que essa fase do curso pode estar associada a uma maior taxa de evasão.

A Figura 16 apresenta um mapa de calor que ilustra essa distribuição, onde tons de azul mais escuros representam um maior acúmulo de *tokens* ao final da execução do *replay*. Essa visualização reforça a concentração de dificuldades acadêmicas nos primeiros semestres, evidenciando as disciplinas que representam os principais pontos críticos ao longo do curso.

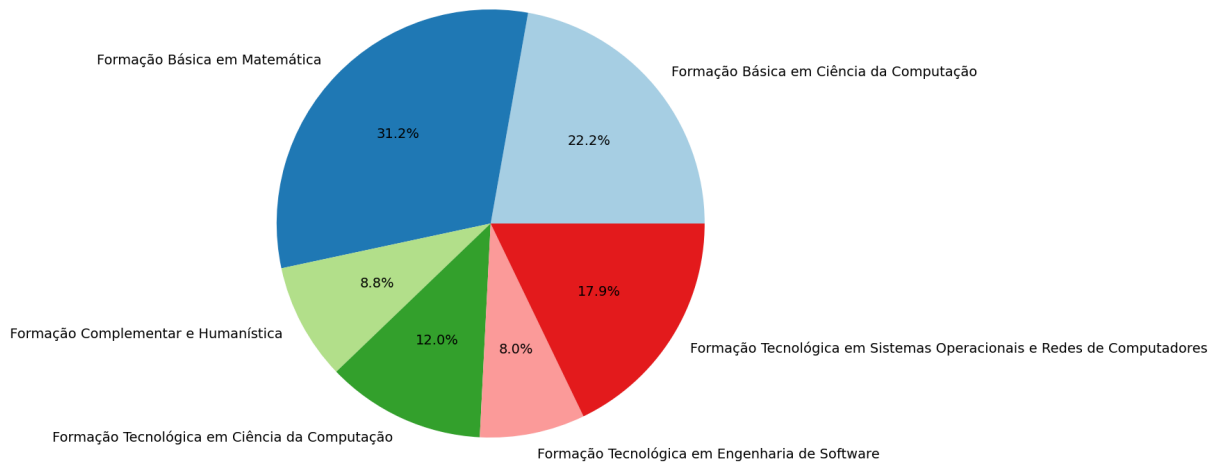
Figura 16 – Fluxograma da Estrutura Curricular Representado como um Mapa de Calor.



Fonte: Elaborado pelo autor.

A Figura 17 apresenta a média de *tokens* acumulados por área de conhecimento. As áreas com maiores médias são Formação Básica em Ciência da Computação e Formação Básica em Matemática. Esse resultado indica que os primeiros semestres concentram a maior parte das reprovações, trancamentos e supressões, sugerindo que essa fase do curso pode estar associada a uma maior taxa de evasão.

Figura 17 – Gráfico de pizza representando a distribuição de *Tokens* por Área de Conhecimento.



Fonte: Elaborado pelo autor.

A Formação Básica em Ciência da Computação inclui disciplinas como Fundamentos de Programação, Introdução à Ciência da Computação, Arquitetura de Computadores, Programação Orientada a Objetos, Estrutura de Dados, Programação Funcional, Estruturas de Dados Avançadas, Linguagens Formais e Autômatos, Linguagens de Programação, Projeto e Análise de Algoritmos, Desenvolvimento de Software para WEB e Teoria da Computação.

A Formação Básica em Matemática é composta por Matemática Básica, Pré-Cálculo, Matemática Discreta, Cálculo Diferencial e Integral, Lógica para Computação, Probabilidade e Estatística e Álgebra Linear.

A predominância de *tokens* nessas áreas reforça a hipótese de que os desafios acadêmicos estão mais concentrados no início do curso, uma vez que algumas disciplinas são do período básico do curso e servem como pré-requisitos para disciplinas mais avançadas. As dificuldades enfrentadas nelas podem afetar a progressão dos alunos, tornando-as pontos críticos para a retenção e evasão no curso.

Por outro lado, áreas como Formação Complementar e Humanística, composta pelas disciplinas Ética, Direito e Legislação e Empreendedorismo, e Formação Tecnológica em Engenharia de Software, que inclui Análise e Projeto de Sistemas, Engenharia de Software e Interface Humano-Computador, apresentam médias de *tokens* menores. Isso pode indicar que essas disciplinas são menos críticas para a retenção dos alunos ou possuem uma taxa de aprovação mais elevada.

No entanto, é importante considerar que, à medida que os alunos avançam nos

semestres, há uma redução no número total de estudantes ativos. Isso pode gerar um viés na análise, fazendo com que disciplinas críticas em semestres mais avançados passem despercebidas quando se analisa o número absoluto de *tokens*.

Para mitigar esse efeito e obter uma visão mais equilibrada, próximo capítulo traz novas visualizações que permitem investigar as disciplinas críticas ao longo de todo o percurso acadêmico.

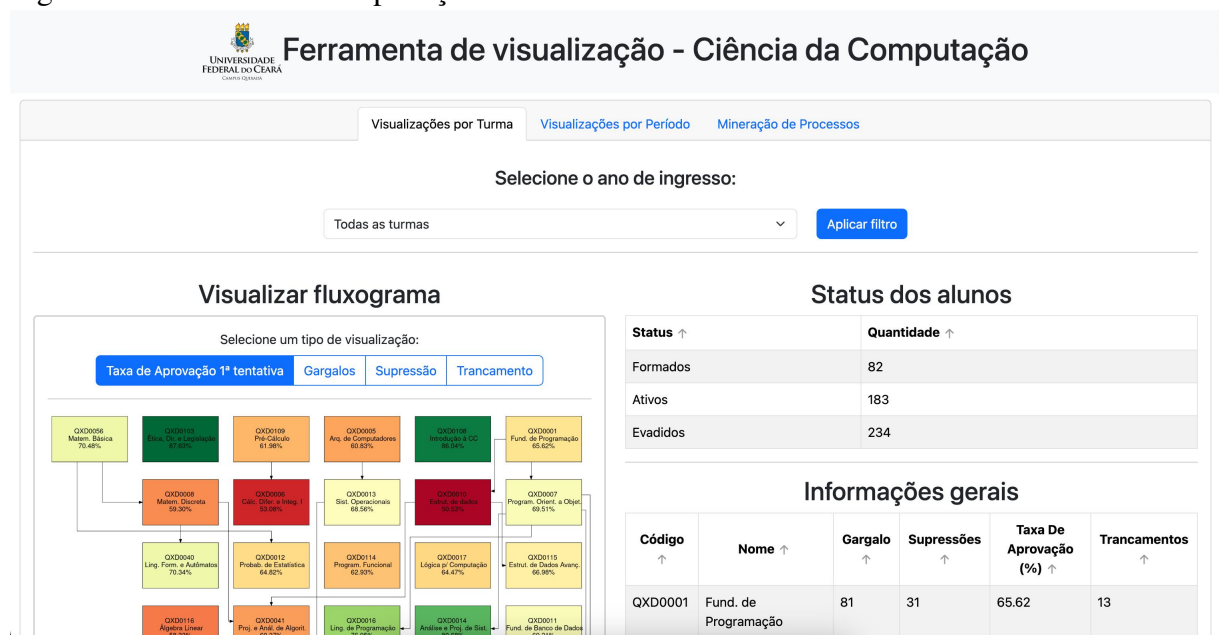
5 FERRAMENTA DE VISUALIZAÇÃO

Este capítulo apresenta a ferramenta desenvolvida durante este trabalho, cujo objetivo é proporcionar visualizações interativas e dinâmicas para facilitar a análise dos dados acadêmicos do curso.

Optou-se por desenvolver a ferramenta como uma aplicação *web* responsiva, garantindo acessibilidade, escalabilidade e flexibilidade. Diferentemente de aplicações *desktop*, uma plataforma *web* permite que usuários possam acessar as visualizações e análises de qualquer dispositivo com um navegador, sem necessidade de instalações locais. Além disso, a abordagem *web* possibilita futuras integrações com outras bases de dados e aprimoramentos contínuos.

A implementação foi realizada utilizando o *Flask*¹ no *back-end*, responsável pelo processamento dos dados e pela comunicação com a camada de apresentação. No *front-end*, foi adotado *Vue.js*², um *framework* para interfaces. A interface da aplicação pode ser vista na Figura 18.

Figura 18 – Tela inicial da aplicação.



Fonte: Elaborado pelo autor.

A ferramenta oferece três tipos de visualização, cada uma projetada para atender a diferentes necessidades de análise:

- **Visualização por turma:** filtra uma turma de acordo com o ano de ingresso.

¹ Disponível em: <https://flask.palletsprojects.com/en/stable/>

² Disponível em: <https://vuejs.org/>

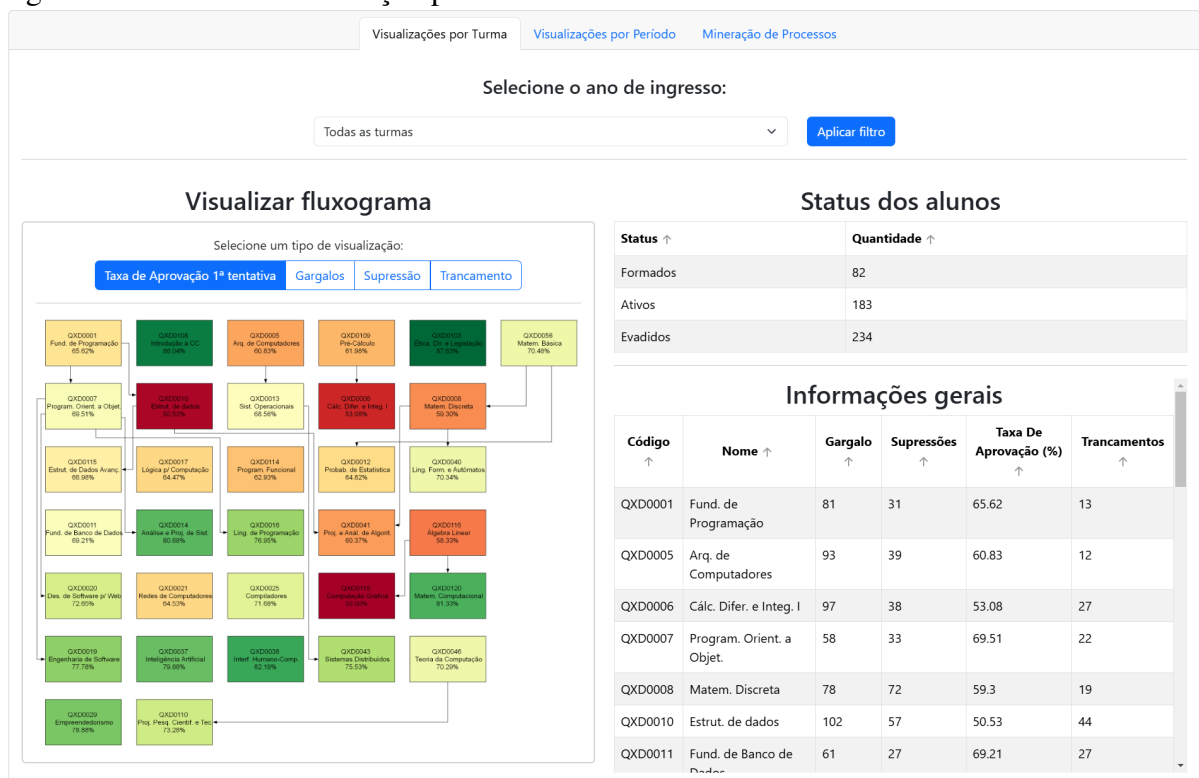
- **Visualização por período:** Filtra um período selecionado.
- **Visualização de Mineração de Processos:** apresenta os resultados da análise de conformidade (*replay*) de acordo com o período selecionado.

Cada uma dessas visualizações será detalhada nas seções a seguir.

5.1 Visualização Por Turma

A Visualização Por Turma permite filtrar os alunos com base no ano de ingresso ou selecionar todas as turmas. Com essa seleção, a ferramenta apresenta três tipos principais de informações, conforme apresentado na Figura 19.

Figura 19 – Tela de Visualização por Turma.



Fonte: Elaborado pelo autor.

Os três tipos de informações são:

1. **Status dos Alunos:** Exibe a distribuição dos alunos daquela turma, categorizando-os como ativos, formados ou evadidos.
2. **Informações Gerais:** Apresenta uma tabela com informações sobre as disciplinas cursadas pela turma selecionada. A tabela contém as seguintes colunas:
 - **Código:** Identificação da disciplina.
 - **Nome:** Nome da disciplina.

- **Gargalo:** Indica quantos alunos cursaram a disciplina e nunca conseguiram aprovação.
- **Supressões:** Número de alunos que suprimiram a disciplina.
- **Taxa de Aprovação na 1ª Tentativa:** Percentual de alunos que foram aprovados na disciplina na primeira vez em que a cursaram.
- **Trancamentos:** Quantidade de alunos que trancaram a disciplina.

3. **Visualização de Fluxograma:** Exibe um fluxograma de mapa de calor, sendo possível escolher entre quatro tipos de visualizações definidos anteriormente:

- Taxa de Aprovação na 1ª Tentativa.
- Disciplinas consideradas gargalos.
- Quantidade de supressão.
- Quantidade de trancamentos.

Essa visualização permite ao coordenador do curso acompanhar o desempenho de cada turma ao longo do tempo, identificando quais disciplinas tiveram baixas taxas de aprovação, alto número de supressão e trancamento, de além de disciplinas consideradas gargalos, que nunca foram aprovadas e podem indicar pontos de evasão.

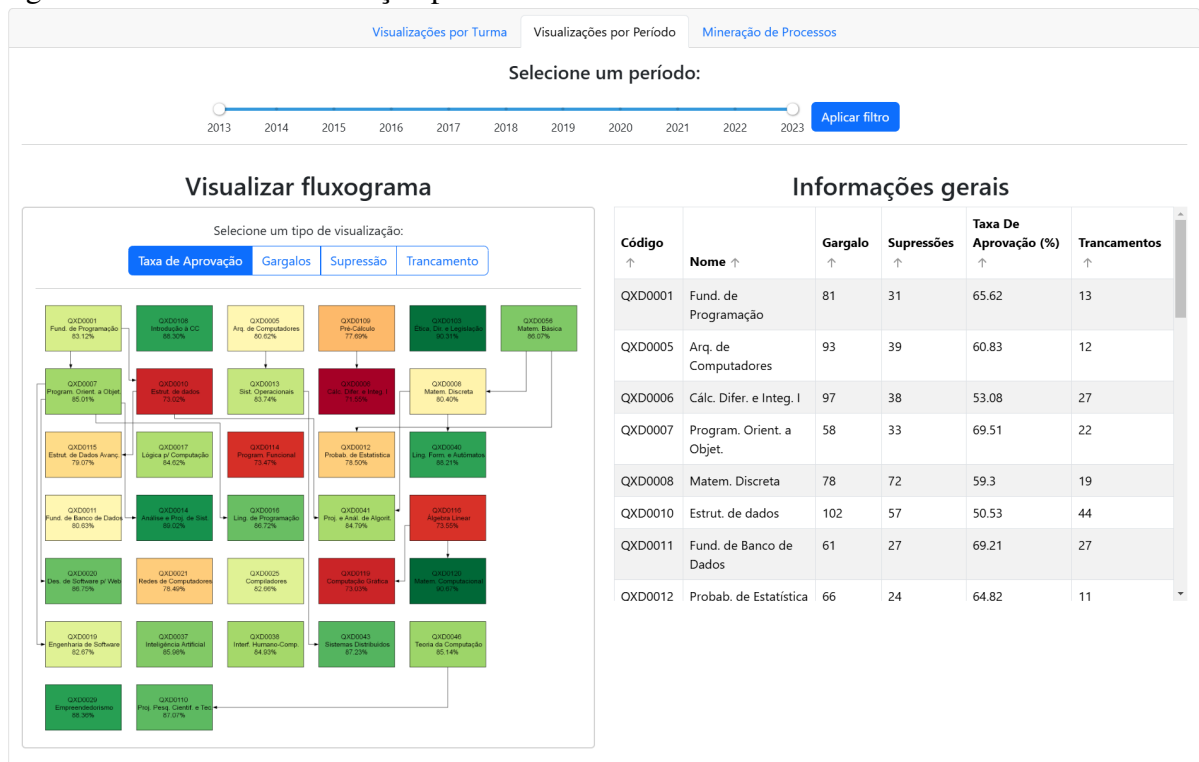
Além disso, ela oferece uma visão sobre o status dos alunos, categorizando-os em ativos, evadidos ou formados. Essas informações permitem compreender a dinâmica de cada turma, como a taxa de evasão e o percentual de alunos que concluíram o curso.

Ao comparar turmas de diferentes anos, o coordenador pode identificar disciplinas que permaneceram problemáticas ao longo do tempo ou que deixaram de ser críticas, permitindo por exemplo avaliar o impacto de possíveis políticas acadêmicas implementadas. A ferramenta também possibilita a análise comparativa do desempenho e do avanço de turmas que ingressaram em períodos distintos, como antes, durante e após a pandemia. Isso permite investigar como eventos externos impactaram o progresso acadêmico das turmas.

5.2 Visualização Por Período

A Visualização Por Período diferencia-se da anterior pelo filtro principal, que agora é baseado em um período. Ao selecionar um período, a ferramenta exibe todas as atividades e dados que os alunos realizaram nesse intervalo. A Figura 20 apresenta a interface dessa visualização.

Figura 20 – Tela de Visualização por Período.



Fonte: Elaborado pelo autor.

A ferramenta apresenta dois tipos principais de informações:

1. **Informações Gerais:** Exibe uma tabela com detalhes sobre as disciplinas cursadas no período selecionado, contendo as seguintes colunas:
 - **Código:** Identificação da disciplina.
 - **Nome:** Nome da disciplina.
 - **Gargalo:** Quantidade de alunos que cursaram a disciplina e nunca conseguiram aprovação nesse período.
 - **Supressões:** Número de alunos que suprimiram a disciplina durante o período.
 - **Taxa de Aprovação:** Percentual de alunos que foram aprovados na disciplina no período selecionado.
 - **Trancamentos:** Número de alunos que trancaram a disciplina.
2. **Visualização de Fluxograma:** Permite visualizar os dados da tabela em formato de mapa de calor, facilitando a análise comparativa entre diferentes períodos. O usuário pode escolher entre as seguintes opções de visualização:
 - Taxa de Aprovação.
 - Disciplinas consideradas gargalos.
 - Quantidade de supressão.

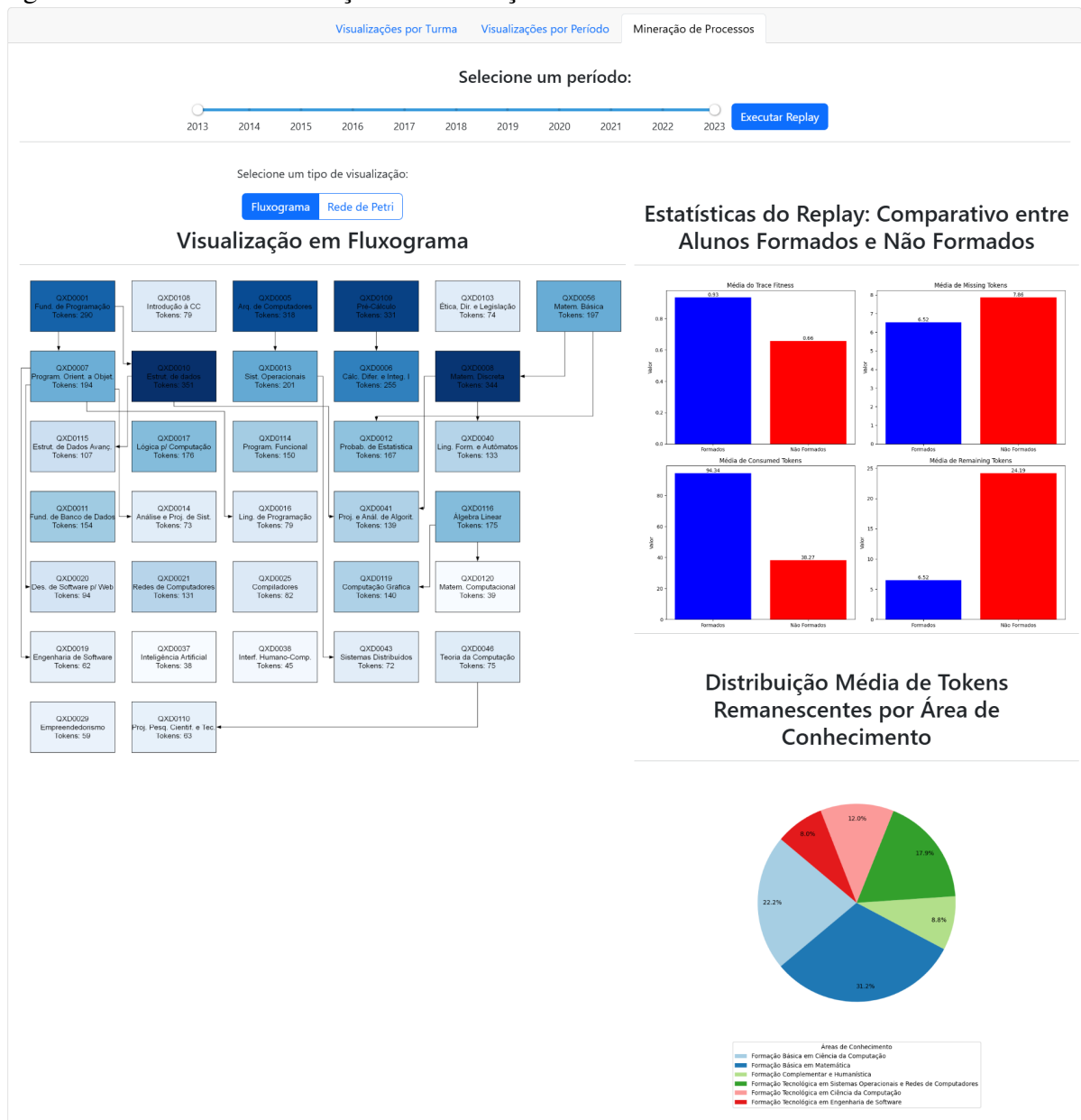
- Quantidade de trancamentos.

A Visualização por Período permite que o coordenador acompanhe o que aconteceu em um ano ou período específico. Com ela, é possível ver, por exemplo, a porcentagem de alunos reprovaram em Estrutura de Dados no primeiro semestre de 2022 ou quantos trancaram Matemática discreta durante a pandemia. Esses dados ajudam a identificar problemas pontuais, como uma disciplina que teve um pico de reprovação em um determinado período, e a entender se isso foi causado por algo específico, como a adaptação ao ensino remoto, uma mudança de docente durante o semestre ou até mesmo a retomada do ensino presencial.

5.3 Visualização de Mineração de Processos

A Visualização de Mineração de Processos permite analisar o alinhamento das trajetórias acadêmicas dos alunos em relação à estrutura curricular modelada na Rede de Petri. Essa funcionalidade é baseada na execução do *replay* dos eventos registrados no *log*, considerando um período específico selecionado pelo usuário. A Figura 21 apresenta sua interface.

Figura 21 – Tela de Visualização de Mineração de Processos.



Fonte: Elaborado pelo autor.

A interface dessa visualização contém um seletor de período e um botão para executar o *replay*. Após a seleção do período e a execução do *replay*, são geradas três representações visuais principais:

1. **Mapa de Calor de Tokens Restantes:** Essa visualização permite visualizar um mapa de calor de *tokens* remanescentes. O usuário pode escolher entre duas formas de exibição:
 - **Visualização da Rede de Petri:** Apresenta a estrutura curricular modelada em Rede de Petri, exibindo a distribuição dos *tokens* ao final da execução do *replay*. A representação utiliza um de mapa de calor, onde a intensidade da cor indica a

concentração de *tokens* em cada *place*.

- **Visualização do Fluxograma:** Exibe a estrutura curricular no formato de fluxograma, destacando apenas os locais onde os *tokens* representam desvios, como discutido no 4. Assim como na visualização anterior, o fluxograma também utiliza um esquema de mapa de calor, facilitando a identificação visual dos principais gargalos do curso.

2. **Estatísticas do *Replay*: Comparativo entre Alunos Formados e Não Formados:** A Segunda visualização apresenta as estatísticas do *replay*, trazendo um gráfico de barras que compara a média das métricas geradas pelo *replay* entre alunos formados e não formados. As métricas apresentadas são: *Fitness Missing Tokens*, *Consumed Tokens* e *Remaining Tokens*.

3. **Distribuição Média de Tokens Remanescentes por Área de Conhecimento:** Essa visualização exibe um gráfico de pizza onde cada área de conhecimento é representada por uma cor distinta. A média de *tokens* remanescentes foi calculada para cada área, permitindo identificar quais delas concentram em média maior acúmulo de tokens.

A Visualização de Mineração de Processos permite analisar o progresso acadêmico dos alunos em relação à estrutura curricular do curso. Utilizando a modelagem em Rede de Petri e o *replay*, ela permite identificar pontos críticos que podem levar à evasão. O Mapa de Calor de *Tokens* Remanescentes destaca visualmente as disciplinas que concentram dificuldades, enquanto a análise estatística do *replay* compara métricas, como *Fitness*, *Missing Tokens*, *Consumed Tokens* e *Remaining Tokens*, diferenciando alunos formados e não formados. Além disso, a Distribuição Média de *Tokens* Remanescentes por Área de Conhecimento possibilita a identificação das áreas que, em média, apresentam maiores desvios.

A ferramenta desenvolvida neste trabalho oferece um meio para analisar a trajetória acadêmica dos alunos por meio de visualizações interativas. Sua implementação como uma aplicação *web* garante acessibilidade, escalabilidade e facilita a implementação de possíveis novas funcionalidades.

Com três tipos principais de visualização – por turma, por período e por mineração de processos – a ferramenta permite diferentes análises sobre o desempenho acadêmico. A Visualização por Turma auxilia na identificação de padrões de evasão e dificuldades persistentes. A Visualização por Período permite diagnosticar problemas específicos em determinados intervalos de tempo, avaliando impactos causados por fatores externos, como greves ou a pandemia. Já a Mineração de Processos utiliza a modelagem da estrutura curricular com a Rede de Petri,

destacando pontos críticos e desvios da trajetória.

No geral, a ferramenta pode contribuir para a gestão acadêmica ao fornecer dados que auxiliam na tomada de decisões voltadas para a melhoria da retenção e da taxa de conclusão dos alunos. Seu uso pode ser ampliado para outros cursos que desejam aprimorar a análise do desempenho estudantil, servindo como um apoio para coordenadores, gestores e pesquisadores.

6 CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho utilizou Mineração de Processos, especificamente a análise de conformidade via *replay*, para avaliar a estrutura curricular do curso de Ciência da Computação da UFC. A modelagem da grade curricular como uma Rede de Petri permitiu representar o fluxo acadêmico dos alunos e identificar disciplinas que atuam como gargalos, impactando a progressão e, possivelmente, levando à evasão.

A partir da execução do *replay* sobre os dados acadêmicos, foram extraídas métricas como *trace fitness*, *missing tokens* e *remaining tokens*, possibilitando a comparação entre alunos formados e não formados. Os resultados demonstraram que os alunos não formados apresentaram maior acúmulo de *tokens* remanescentes, indicando trajetórias interrompidas ou prolongadas. Além disso, a análise revelou que os maiores gargalos estão concentrados nos primeiros semestres do curso, onde há maior incidência de reprovações, supressões e trancamentos.

Para facilitar a visualização e interpretação dos dados, foi desenvolvida uma ferramenta *web* que permite analisar o desempenho dos alunos e identificar padrões na progressão acadêmica em diferentes momentos do curso. Essa ferramenta pode ser utilizada pela coordenação para monitorar o avanço dos estudantes, e embasar a adoção de estratégias que minimizem a evasão e melhorem a eficiência do percurso acadêmico.

Os resultados deste trabalho podem contribuir para a formulação de estratégias institucionais voltadas à redução da evasão e melhoria do fluxo acadêmico, possibilitando intervenções direcionadas nas disciplinas mais críticas. Como trabalhos futuros, sugere-se a aplicação de outras técnicas de análise de conformidade, como o alinhamento de processos, que permite uma comparação mais detalhada entre os percursos acadêmicos dos alunos e a estrutura curricular modelada.

Além disso, a ferramenta desenvolvida pode ser aprimorada com a inclusão de novas análises que aprofundem a compreensão da progressão acadêmica dos estudantes. Outra possibilidade é expandir seu uso para os outros cursos do Campus de Quixadá, permitindo avaliar padrões entre os cursos, possibilitando a adoção de estratégias institucionais para melhorar a formação dos alunos e reduzir a evasão.

REFERÊNCIAS

- AALST, V. D. **Data science in action**. [S. l.]: Springer, 2016.
- AALST, W. M. Van der. **The application of Petri nets to workflow management**. Journal of circuits, systems, and computers, World Scientific, v. 8, n. 01, p. 21–66, 1998.
- AALST, W. M. van der. **Process mining: a 360 degree overview**. In: Process Mining Handbook. [S. l.]: Springer, 2022. p. 3–34.
- AALST, W. V. D.; ADRIANSYAH, A.; MEDEIROS, A. K. A. D.; ARCIERI, F.; BAIER, T.; BLICKLE, T.; BOSE, J. C.; BRAND, P. V. D.; BRANDTJEN, R.; BUIJS, J. *et al.* **Process mining manifesto**. In: SPRINGER. Business Process Management Workshops: BPM 2011 International Workshops, Clermont-Ferrand, France, August 29, 2011, Revised Selected Papers, Part I 9. [S. l.], 2012. p. 169–194.
- AALST, W. van der. **Process Mining: Overview and Opportunities**. ACM Trans. Manage. Inf. Syst., Association for Computing Machinery, New York, NY, USA, v. 3, n. 2, jul 2012. ISSN 2158-656X. Disponível em: <https://doi.org/10.1145/2229156.2229157>.
- ALVES, M. F. S.; MANTOVANI, K. L. **Identificação do perfil dos acadêmicos de engenharia como uma medida de combate à evasão**. Revista de Ensino de Engenharia, Associação Brasileira de Educação em Engenharia (ABENGE), v. 35, n. 2, p. 1–11, 2016. Disponível em: <https://revista.abenge.org.br/index.php/abenge/article/view/464>. Acesso em: 9 mar. 2024.
- AMBIEL, R. A. M.; CORTEZ, P. A.; SALVADOR, A. P. **Predição da Potencial Evasão Acadêmica entre Estudantes Trabalhadores e Não Trabalhadores**. Psicologia: Teoria e Pesquisa, Instituto de Psicologia, Universidade de Brasília, v. 37, p. e37305, 2021. ISSN 0102-3772. Disponível em: <https://doi.org/10.1590/0102.3772e37305>. Acesso em: 9 mar. 2024.
- BAGGI, C. A. D. S.; LOPES, D. A. **Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica**. Avaliação: Revista da Avaliação da Educação Superior (Campinas), Publicação da Rede de Avaliação Institucional da Educação Superior (RAIES), da Universidade Estadual de Campinas (UNICAMP) e da Universidade de Sorocaba (UNISO), v. 16, n. 2, p. 355–374, Jul 2011. ISSN 1414-4077. Disponível em: <https://doi.org/10.1590/S1414-40772011000200007>. Acesso em: 29 mar. 2024.
- BARROSO, M. F.; FALCÃO, E. B. **Evasão universitária: o caso do Instituto de Física da UFRJ**. IX Encontro Nacional de Pesquisa em Ensino de Física, v. 9, p. 1–14, 2004.
- BRITO, B.; MELLO, R.; ALVES, G. **Identificação de atributos relevantes na evasão no Ensino Superior Público Brasileiro**. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. Anais [...]. Porto Alegre: SBC, 2020. p. 1032–1041. ISSN 0000-0000. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/12859>. Acesso em: 3 jun. 2024.
- CARDOSO, C. B. **Efeitos da política de cotas na Universidade de Brasília: uma análise do rendimento e da evasão**. Universidade Federal de Brasília. Faculdade de Educação. Dissertação, 2008.
- CARVALHO, R. B. e Seiji Isotani e A. **Mineração de Dados Educacionais: Oportunidades para o Brasil**. Revista Brasileira de Informática na Educação, v. 19, n. 02, p. 03, 2011. ISSN 2317-6121. Disponível em: <http://milanesa.ime.usp.br/rbie/index.php/rbie/article/view/1301>. Acesso em: 9 mai. 2024.

COSTA, E.; BAKER, R. S.; AMORIM, L.; MAGALHÃES, J.; MARINHO, T. **Mineração de dados educacionais**: conceitos, técnicas, ferramentas e aplicações. Jornada de Atualização em Informática na Educação, v. 1, n. 1, p. 1–29, 2012.

DESEL, J.; REISIG, W. **Place/transition Petri nets**. In: SPRINGER. Advanced Course on Petri Nets. [S. l.], 1996. p. 122–173.

FILHO, R. L. L. e. S.; MOTEJUNAS, P. R.; HIPÓLITO, O.; LOBO, M. B. C. M. **A evasão no ensino superior brasileiro**. Cadernos de Pesquisa, Fundação Carlos Chagas, São Paulo, v. 37, n. 132, p. 641–659, 2007. Disponível em: <https://publicacoes.fcc.org.br/cp/article/view/346>. Acesso em: 9 abr. 2024.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). **Censo da Educação Superior 2023 - Resultados**. 2023. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-da-educacao-superior/resultados>. Acesso em: 8 jun. 2024.

KANTORSKI, G.; MARTINS, R.; BALEJO, A.; FRICK, M. **Mineração de Dados Educacionais para Predição da Evasão em Cursos de Graduação Presenciais no Ensino Superior**. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. Anais do XXXIV Simpósio Brasileiro de Informática na Educação. Porto Alegre: SBC, 2023. p. 1133–1142. ISSN 0000-0000. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/26741>. Acesso em: 1 mai. 2024.

LIMA JÚNIOR, A. A. de. **Mineração de dados para identificação do perfil de evasão de alunos da UFC - Campus Quixadá**. 37 p. Trabalho de Conclusão de Curso (Graduação em Sistemas de Informação), Quixadá, 2019.

MARTINHO, V. R. d. C. **Sistema inteligente para a predição de grupo de risco de evasão discente**. Dissertação (Dissertação (Mestrado)), São Paulo, 2014. Disponível em: <https://acervodigital.unesp.br/handle/11449/100340>. Acesso em: 8 abr. 2024.

PINHEIRO, M. A. L.; SILVA, J. C. da; SOUZA, B. F. de. **Aprendizado de máquina aplicado à análise de evasão no ensino superior**. In: Anais do Computer on the Beach. Univali, 2018. p. 512–521. Disponível em: <https://periodicos.univali.br/index.php/acotb/article/view/12810>. Acesso em: 8 mai. 2024.

SANTOS, C. H.; MARTINS, S.; PLASTINO, A. **É Possível Prever Evasão com Base Apenas no Desempenho Acadêmico?** In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. Anais do XXXII Simpósio Brasileiro de Informática na Educação. Porto Alegre: SBC, 2021. p. 792–802. ISSN 0000-0000. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/18107>. Acesso em: 3 abr. 2024.

SESU/MEC, A.; ABRUEM. **Diplomação, retenção e evasão nos cursos de graduação em IES públicas: Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras**. [S. l.]: s. n., 1996.

UMER, R.; SUSNJAK, T.; MATHRANI, A.; SURIADI, S. **On predicting academic performance with process mining in learning analytics**. Journal of Research in Innovative Teaching & Learning, Emerald Publishing Limited, v. 10, n. 2, p. 160–176, 2017. Disponível em: <https://doi.org/10.1108/JRIT-09-2017-0022>. Acesso em: 3 abr. 2024.

UNGER, A.; JUNIOR, D. L.; LIMA, F.; GERALDO, I.; VENERO, S.; AMBROSIO, R. **Análise da trajetória de aprendizagem do aluno no ensino a distância por meio da mineração de processos.** In: Anais do XXXIII Simpósio Brasileiro de Informática na Educação. Porto Alegre, RS, Brasil: SBC, 2022. p. 1163–1172. ISSN 0000-0000. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/22490>. Acesso em: 15 abr. 2024.

VELOSO, T. C. M. A.; ALMEIDA, E. P. d. **Evasão nos cursos de graduação da Universidade Federal de Mato Grosso, campus universitário de Cuiabá – um processo de exclusão.** Série-Estudos - Periódico do Programa de Pós-Graduação em Educação da UCDB, Universidade Católica Dom Bosco (UCDB), [S. l.], n. 13, 2013. Disponível em: <https://www.serie-estudos.ucdb.br/serie-estudos/article/view/564>. Acesso em: 19 mar. 2024.