

Análise de Dados dos Estágios Obrigatórios da Escola Politécnica de Pernambuco (POLI)

Data Analysis of Mandatory Internships at the Polytechnic School of Pernambuco (POLI)

André Silva¹

orcid.org/0000-0002-1323-061X

Anthony Militão¹

orcid.org/0009-0009-7832-771X

Lucas Henrique¹

orcid.org/0009-0000-3677-1497

¹Escola Escola Politécnica de Pernambuco,
Universidade de Pernambuco, Recife, Brasil.
E-mail: lhsn@ecomp.poli.br

DOI: [10.25286/repav11i1.3178](https://doi.org/10.25286/repav11i1.3178)

Esta obra apresenta Licença Creative Commons Atribuição-Não Comercial 4.0 Internacional.

Como citar este artigo pela NBR 6023/2018:
André Silva; Anthony Militão; Lucas Henrique. Análise de Dados dos Estágios Obrigatórios da Escola Politécnica de Pernambuco (POLI) Revista de Engenharia e Pesquisa Aplicada, v.11, n. 1, p. 1-11, 2026.

RESUMO

Este artigo investigou o impacto do estágio obrigatório no desempenho acadêmico de alunos dos cursos de engenharia da Escola Politécnica de Pernambuco (POLI), aplicando técnicas de mineração de dados, como o algoritmo K-Means, para identificar padrões e variáveis associadas. A análise segmentou os estudantes em clusters antes e depois do estágio, possibilitando uma compreensão detalhada das mudanças no desempenho, frequência, aprovações e posicionamento no ranking. Os resultados demonstraram que o estágio influencia significativamente a trajetória acadêmica, com alguns grupos apresentando melhorias e outros enfrentando desafios relacionados à frequência e desempenho.

PALAVRAS-CHAVE: Estágio obrigatório, desempenho acadêmico, mineração de dados, K-Means, engenharia.

ABSTRACT

This paper investigated the impact of mandatory internships on the academic performance of engineering students at the Polytechnic School of Pernambuco (POLI), applying data mining techniques, such as the K-Means algorithm, to identify patterns and associated variables. The analysis segmented students into clusters before and after the internship, enabling a detailed understanding of changes in performance, attendance, approvals, and ranking. The results demonstrated that the internship significantly influences academic trajectory, with some groups showing improvements and others facing challenges related to attendance and performance.

KEY-WORDS: Mandatory internship, academic performance, data mining, K-Means, engineering.

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

O estágio obrigatório é essencial na formação de estudantes, especialmente em áreas técnicas como engenharia, ao aplicar conhecimentos teóricos na prática e desenvolver competências como resolução de problemas e adaptação [1]. Na Escola Politécnica de Pernambuco (POLI), ele conecta academia e mercado, preparando os alunos para desafios profissionais e demandas reais.

Apesar disso, a gestão dos estágios enfrenta desafios, como a fragmentação de dados entre sistemas [2]. Destaca-se que alinhar teoria e prática exige estruturas robustas para monitorar o progresso dos alunos e avaliar os impactos dessas experiências, algo dificultado pela falta de integração de informações [3].

Técnicas como mineração de dados oferecem soluções promissoras para esses desafios. Aponta-se que essas ferramentas ajudam a identificar padrões e variáveis de sucesso, permitindo à POLI melhorar a gestão dos estágios e consolidar sua posição como referência na formação de engenheiros para um mercado dinâmico e exigente [4].

1.2 DESCRIÇÃO DO PROBLEMA

O principal problema dos estágios obrigatórios na POLI é a falta de acompanhamento sistemático e de uma análise integrada para avaliar o impacto dessas experiências na formação acadêmica e profissional dos alunos. A ausência de ferramentas adequadas para monitorar a relação entre estágio e desempenho acadêmico compromete tanto a experiência dos estudantes quanto a eficiência da POLI em otimizar os processos de estágio, prejudicando a integração entre teoria e prática, essencial para a preparação ao mercado de trabalho [5].

Outro desafio é a fragmentação de informações em diferentes plataformas, dificultando uma análise abrangente dos estágios nos cursos de engenharia. Técnicas de mineração de dados, poderiam extrair informações valiosas para aprimorar a gestão dos estágios [4]. Além disso, ressalta-se a importância da análise de dados educacionais para prever desempenho acadêmico e melhorar políticas educacionais, algo crucial em cursos onde os estágios desempenham papel essencial na formação dos estudantes [6].

1.3 OBJETIVOS

Aplicar técnicas de mineração e análise de dados para identificar padrões, causas e variáveis relacionadas aos estágios dos alunos nos cursos de engenharia da POLI, com o intuito de compreender e proporcionar informações que apoiem o processo de tomada de decisão futura quanto a formação acadêmica e profissional dos estudantes, bem como melhorar a eficiência da gestão dos processos de estágio na instituição.

Os objetivos específicos deste estudo visam compreender o impacto dos estágios na vida acadêmica dos alunos utilizando técnicas de mineração de dados:

- Consolidar dados de estágio e históricos acadêmicos para criar uma base unificada para análise.
- Identificar padrões e correlações entre estágio obrigatório e desempenho acadêmico por meio de algoritmos de clustering.
- Determinar variáveis que influenciam o desempenho e avaliar mudanças comportamentais relacionadas aos estágios.
- Gerar relatório com insights para melhorar a gestão dos estágios e integrar teoria e prática acadêmica.

1.4 JUSTIFICATIVA

O estágio acadêmico obrigatório desempenha um papel essencial na formação dos alunos dos cursos de engenharia da POLI, ao conectar a teoria aprendida em sala de aula com a prática profissional. Essa experiência promove o desenvolvimento de habilidades técnicas e comportamentais, fundamentais para o mercado de trabalho [7]. No entanto, a gestão desses estágios enfrenta desafios como a fragmentação de dados e a falta de integração entre sistemas, dificultando o monitoramento e a otimização das experiências acadêmicas.

A aplicação de técnicas de mineração de dados pode transformar a gestão dos estágios na POLI, permitindo identificar padrões e variáveis que impactam o sucesso dos alunos. Essas análises auxiliam na adequação das oportunidades de estágio às demandas do mercado e do currículo, potencializando o desenvolvimento de competências-chave e aumentando em até 25% o desempenho acadêmico dos estudantes [8]. Além disso, estágios bem gerenciados podem elevar em

até 30% as chances de empregabilidade dos alunos na área de formação, trazendo benefícios significativos para os graduados e fortalecendo a reputação da instituição [9].

Com uma gestão mais eficiente, a POLI pode consolidar sua posição como referência em inovação educacional no Brasil. Ao melhorar a retenção e a satisfação dos alunos, além de fortalecer parcerias com empresas, a instituição se torna mais competitiva no cenário educacional. Isso reflete diretamente na qualidade da formação dos engenheiros, aumentando em até 30% as taxas de satisfação e empregabilidade, e posicionando a POLI como líder na formação de profissionais altamente qualificados [10].

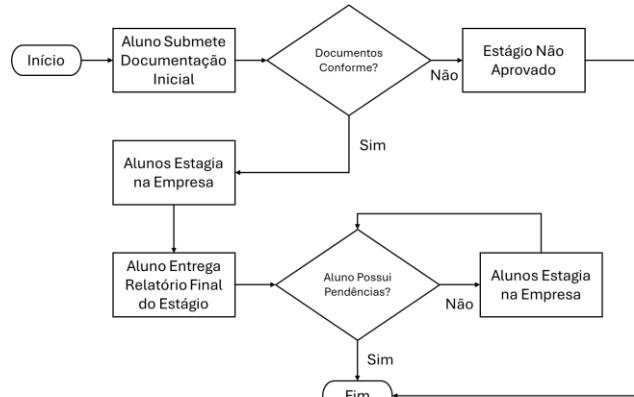
2 FUNDAMENTAÇÃO TEÓRICA

2.1 GESTÃO DOS ESTÁGIOS OBRIGATÓRIOS

A gestão dos estágios obrigatórios é essencial para conectar teoria e prática, promovendo o desenvolvimento de competências técnicas e comportamentais nos alunos, especialmente em cursos que demandam aplicação prática [5][8][9]. Essa gestão exige coordenação eficiente entre instituições de ensino, empresas e alunos, garantindo que as atividades realizadas estejam alinhadas ao currículo e às demandas do mercado. Contudo, a falta de sistemas robustos de monitoramento e avaliação pode comprometer o pleno aproveitamento dessa experiência, dificultando a identificação de lacunas e oportunidades de melhoria [11]. Uma gestão eficaz assegura experiências enriquecedoras para os alunos e contribuições significativas das empresas parceiras na formação de profissionais qualificados.

A Figura 1 apresenta o fluxograma que descreve de forma geral o processo de estágio obrigatório na POLI. O esquema ilustra as etapas principais, desde a submissão inicial da documentação pelo aluno, passando pelas verificações necessárias, até a conclusão do processo com a aprovação final, após o cumprimento de todos os requisitos.

Figura 1 – Fluxo do Processo de Estágio Obrigatório



Fonte: Autores

2.2 MINERAÇÃO DE DADOS

Para este estudo foi utilizado a base de dados dos alunos dos cursos de engenharia da POLI e o algoritmo de agrupamento escolhido foi o K-Means. A escolha do K-Means se justifica por sua simplicidade e eficiência na segmentação de grandes conjuntos de dados, como é o caso dos registros de desempenho e participação em estágios dos alunos de engenharia. O K-Means requer a definição prévia do número de clusters, o que permite agrupar alunos com base em padrões claros de desempenho acadêmico e experiência profissional, facilitando a análise das trajetórias estudantis de forma estruturada [12][13]. No contexto dos alunos de engenharia, onde há diversidade nas trajetórias acadêmicas e experiências em estágios, o K-Means ajuda a identificar grupos homogêneos de estudantes com perfis semelhantes, fornecendo informações sobre padrões de sucesso acadêmico e desafios ao longo da formação.

2.3 FUNDAMENTAÇÃO TEÓRICA DO ALGORITMO

O K-Means é amplamente utilizado como um algoritmo eficiente para segmentação em dados de grande escala, buscando minimizar a variância dentro dos clusters ao agrupar pontos próximos de um centróide. Através de um processo iterativo, o algoritmo ajusta os centróides até alcançar uma configuração otimizada, atribuindo cada ponto ao cluster mais próximo. Embora assuma clusters de formato esférico, o K-Means é eficaz para dados com estrutura uniforme, como observado em análises de desempenho acadêmico [12][14].

Na análise dos alunos da POLI, o K-Means foi aplicado para identificar agrupamentos de estudantes com características de desempenho similares, como alto desempenho, desempenho intermediário e dificuldades acadêmicas. Essa segmentação permitiu uma base sólida para investigar trajetórias diferentes e identificar perfis acadêmicos específicos, fornecendo insights valiosos para intervenções educacionais e estratégias de suporte direcionadas [15].

A aplicação do K-Means na segmentação dos alunos de engenharia da POLI mostrou-se eficaz na organização e análise das trajetórias acadêmicas, permitindo uma compreensão mais clara dos padrões de desempenho. No futuro, a combinação do K-Means com técnicas de detecção de outliers pode complementar essa abordagem, oferecendo uma visão mais detalhada sobre perfis atípicos e aprimorando ainda mais as estratégias de intervenção educacional.

2.4 TRABALHOS RELACIONADOS

A aplicação de técnicas de mineração de dados em ambientes educacionais tem se mostrado promissora na otimização de processos e na melhoria da gestão acadêmica, no Quadro 1 abaixo são apresentados os trabalhos relacionados.

Destaca-se a mineração de dados como uma ferramenta eficaz para extrair conhecimento a partir de grandes volumes de informações fragmentadas, enquanto também se demonstra como essas técnicas podem identificar padrões de comportamento estudantil, prever desempenho e personalizar experiências de aprendizagem [4][6]. Essas abordagens têm contribuído para a gestão mais eficiente das instituições de ensino, proporcionando suporte estratégico à tomada de decisões.

Reforça-se essa perspectiva ao revisar o uso de EDM e learning analytics, evidenciando como essas tecnologias ajudam a identificar alunos em risco de evasão e a melhorar metodologias de ensino [16]. Por meio de ajustes baseados em dados, é possível aumentar o engajamento e o sucesso acadêmico, impactando positivamente tanto a administração acadêmica quanto a pedagogia. Essa abordagem transforma a forma como instituições educacionais gerenciam suas operações e orientam intervenções educacionais.

No Brasil, explorar-se a aplicação de mineração de dados na gestão de estágios, mostrando como a análise de dados históricos pode prever demandas

futuras, alocar recursos de forma eficiente e fortalecer parcerias alinhadas com as necessidades dos alunos [17]. De forma similar, destaca-se como a integração de dados dispersos pode resolver problemas de inconsistência e redundância, facilitando análises precisas e decisões estratégicas em ambientes educacionais que utilizam múltiplos sistemas de informação [18][19].

Quadro 1 – Resumo dos Trabalhos Relacionados

Autor	Ideia Central	Técnica/Modelo Matemático de Avaliação	Uso de Recursos Computacionais
[4]	Mineração de dados para extrair conhecimento educacional.	Mineração de Dados	Não especificado
[6]	EDM para identificar padrões e personalizar o aprendizado.	Análise de padrões, predição de desempenho	Não especificado
[16]	EDM e Learning Analytics para melhorar gestão e pedagogia.	Learning Analytics, mineração de dados	Não especificado
[17]	Mineração de dados para tendências e demandas de estágios.	Mineração de Dados, análise histórica	Processamento de dados históricos de estágios
[18]	Integração de dados para resolver inconsistências em sistemas.	Integração de Dados, Mineração de Dados	Técnicas para unificação de dados fragmentados

Fonte: Autores

3 MATERIAIS E MÉTODOS

3.1 SISTEMA DE INFORMAÇÕES E GESTÃO ACADÊMICA (SIGA)

O Sistema de Informações e Gestão Acadêmica (SIGA) é uma plataforma digital que centraliza

informações acadêmicas, facilitando a gestão para alunos, professores e coordenadores. Por meio do sistema, os alunos podem acessar dados sobre disciplinas, notas, faltas e histórico acadêmico, além de realizar matrículas, promovendo um acompanhamento contínuo de seu desempenho [19]. Para professores e coordenadores, o SIGA permite monitorar o progresso dos alunos, verificando o cumprimento de requisitos como carga horária e atividades relacionadas ao curso, otimizando a comunicação e a eficiência administrativa [17].

A integração do SIGA na gestão acadêmica e de estágios proporciona à UPE maior controle sobre as atividades dos alunos, organizando dados essenciais para a avaliação da qualidade dos estágios e o alinhamento com as diretrizes curriculares. O sistema não apenas facilita o cumprimento das exigências acadêmicas, mas também apoia o desenvolvimento acadêmico e profissional dos alunos, contribuindo para uma gestão mais eficaz de suas trajetórias e experiências educacionais [20].

3.2 DESCRIÇÃO DA BASE DE DADOS

O sucesso deste projeto depende da coleta, organização e análise de um conjunto abrangente de dados que possibilitaram uma compreensão profunda do cenário dos estágios nos cursos de engenharia na POLI. O inventário de dados inclui várias fontes críticas que serão fundamentais para a implementação das melhorias na gestão dos estágios e orientar a tomada de decisões estratégicas.

As bases de dados extraídas do SIGA incluem informações sobre a situação acadêmica, como status dos alunos (ativo, inativo, trancado), disciplinas cursadas, notas e desempenho, vínculos institucionais e o histórico completo dos estudantes. Esses dados possibilitam a análise detalhada do desempenho acadêmico em relação às experiências de estágio, identificando fatores que influenciam diretamente a formação e o sucesso dos alunos.

O SIGA também fornece um contexto amplo para compreender como os estágios se integram ao currículo acadêmico e impactam o progresso dos estudantes nos cursos. Essa visão integrada ajuda a correlacionar dados de diferentes fontes, permitindo uma melhor compreensão do papel dos estágios na trajetória acadêmica.

A combinação dessas bases possibilitou uma análise abrangente e multifacetada, fundamental para avaliar o impacto dos estágios na formação dos

alunos da POLI. Além disso, foram desenvolvidos dashboards interativos e ferramentas de visualização que permitem monitoramento em tempo real e embasar decisões estratégicas fundamentadas.

O dicionário de dados, descrito no Quadro 2, detalha as informações principais utilizadas, especificando nome, descrição, tipo de dado, tamanho e valores permitidos.

Quadro 2 - Dicionário de Dados do Projeto

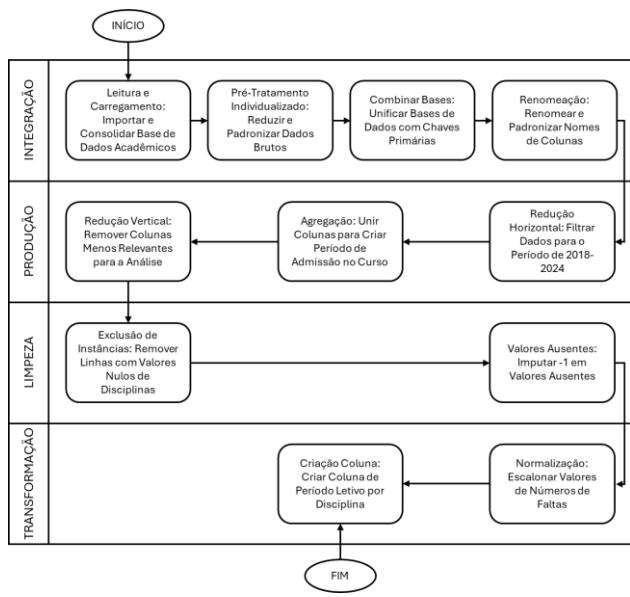
Nome	Descrição
codigo_estudante	Código para identificação do estudante
ranking_antes	Média das médias do estudante nas disciplinas cursadas antes do estágio
ranking_depois	Média das médias do estudante nas disciplinas cursadas depois do estágio
discs_antes	Número total de disciplinas cursadas pelo estudante antes do estágio
discs_depois	Número total de disciplinas cursadas pelo estudante depois do estágio
reps_falta_antes	Número total de reprovações por falta em disciplinas pelo estudante antes do estágio
reps_falta_depois	Número total de reprovações por falta em disciplinas pelo estudante depois do estágio
reps_media_antes	Número total de reprovações por média em disciplinas pelo estudante antes do estágio
reps_media_depois	Número total de reprovações por falta em disciplinas pelo estudante depois do estágio
aprovs_geral_antes	Número total de aprovações em disciplinas pelo estudante antes do estágio
aprovs_geral_depois	Número total de aprovações em disciplinas pelo estudante depois do estágio
Cluster	Número identificador do agrupamento

Fonte: Autores

3.3 PRÉ-PROCESSAMENTO DOS DADOS

O processo de pré-processamento foi aplicado de forma independente a cada uma das bases de dados envolvidas, a fim de preparar e adequar as informações para análises subsequentes, conforme Figura 2.

Figura 2 - Pré-processamento dos dados



Fonte: Autores.

O pré-processamento dos dados seguiu quatro etapas principais: Integração, Redução, Limpeza e Transformação, cada uma com um conjunto de tarefas específicas.

Primeiramente, as bases foram lidas e carregadas no sistema. Após isso, foi feito um pré-tratamento individualizado, onde cada base foi ajustada conforme as suas particularidades. Em seguida, ocorreu a unificação das bases utilizando duas chaves primárias: `codigo_estudante` e `CD_ATIV_ACAD`. Isso permitiu a junção correta dos registros. Além disso, houve a renomeação de colunas: as colunas `CD_PERD LETV_x` e `CD_PERD LETV_y` foram renomeadas para `CD_PERD LETV_COMPT_CURR` e `CD_PERD LETV_estudante`, respectivamente, para deixar mais claro o seu propósito.

Na etapa de Redução, foram realizadas tanto a eliminação de colunas quanto a fusão de informações relevantes, eliminando colunas desnecessárias como `DT_COL_GRAU`, `NU_ANO_CONC`, `NU_ANO_ADMINIS`, `NU_SEMTR_ADMINIS` e `NM_PROGR_FORM`, que não acrescentavam valor para a análise. Depois, foi feita uma agregação, onde as colunas `NU_ANO_ADMINIS` e `NU_SEMTR_ADMINIS` foram combinadas em uma nova coluna, `NU_PERIODO_ADMINIS`, consolidando os dados do período de admissão. Além disso, houve uma redução horizontal, que segmentou a coluna `NM_PROGR_FORM` para focar nos alunos do curso de Engenharia da Computação. Outra redução horizontal foi feita para segmentar os dados da

coluna `NU_ANO_ADMINIS`, mantendo apenas registros dos anos de 2018 a 2024, que são mais relevantes para o estudo.

Visou corrigir valores ausentes e remover registros incompletos. Primeiramente, os valores ausentes na coluna `NU_MEDIA_FIM` foram tratados imputando-se o valor -1, indicando ausência de nota final. Da mesma forma, na coluna `CD_TIPO_COTA`, valores ausentes também foram preenchidos com -1. Após isso, ocorreu a exclusão de instâncias: foram removidas todas as linhas em que a coluna `CD_COMPN_CURR` apresentava valores nulos, eliminando registros incompletos que poderiam prejudicar a análise.

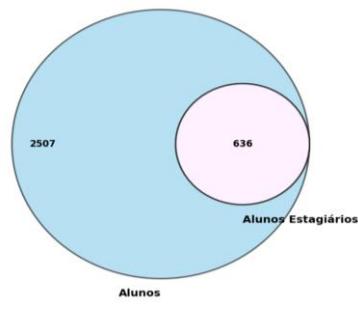
Na última etapa foram feitas algumas modificações nos dados para melhorar sua qualidade e usabilidade. Foi criada uma nova coluna chamada `PERIODO LETIVO`, que informa o período letivo em que cada aluno está cursando a disciplina. Por fim, os dados da coluna `NU_TOT_FALTAS` passaram por um processo de normalização utilizando o método Min-Max, o que permitiu padronizar os valores de faltas dos alunos, facilitando comparações e análises futuras.

3.4 ANÁLISE DESCRIPTIVA DOS DADOS

A análise descritiva dos dados explora variáveis relacionadas ao desempenho acadêmico, como código do estudante, número de faltas, nota final e situação acadêmica. Também considera fatores como ano de admissão e uso de cotas, permitindo identificar padrões e entender o progresso dos alunos ao longo do tempo.

O Gráfico 1 apresenta a distribuição dos 3143 alunos da POLI nos cursos de engenharia entre 2018 e 2024. Desses, 80% (2507 alunos) ainda não realizaram o estágio obrigatório, enquanto 20% (636 alunos) já concluíram ou estão em andamento.

Gráfico 1 – Base de Alunos de 2018 a 2024

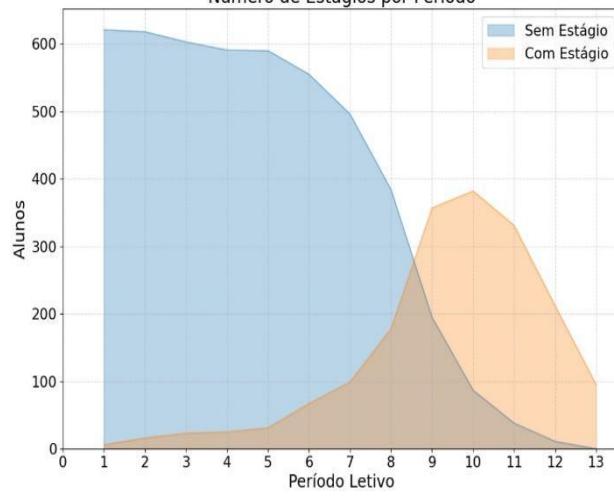


Fonte: Autores

O Gráfico 2 do Número de Estágio por Período, representa o número de alunos com estágio e sem estágio ao longo dos diferentes períodos letivos. Nos primeiros períodos (1º ao 6º), a grande maioria dos alunos está sem estágio, com o número atingindo o pico em torno de 600 alunos. A partir do 7º período, há uma inversão: o número de alunos sem estágio começa a diminuir drasticamente, enquanto o número de alunos com estágio aumenta significativamente, com a maior concentração ocorrendo nos períodos entre o 9º e o 11º.

Esse gráfico ilustra uma tendência natural de que a maioria dos alunos começa a realizar estágios nos períodos finais do curso, entre o 8º e o 12º período, quando as oportunidades de estágio são mais comuns ou obrigatórias para a conclusão do curso.

Gráfico 2 – Número de Estágio por Período
Número de Estágios por Período



Fonte: Autores

3.5 METODOLOGIA EXPERIMENTAL

Na Metodologia Experimental do nosso projeto, seguimos uma abordagem semelhante à proposta no artigo de Mohamed Nafuri et al. (2022), mas com ajustes específicos para a análise de desempenho acadêmico e participação em estágios dos estudantes. Optou-se pelo algoritmo de clustering K-Means devido à sua eficiência e simplicidade e por ser mais adequado para bases de dados de formato aproximadamente esférico e com densidade uniforme.

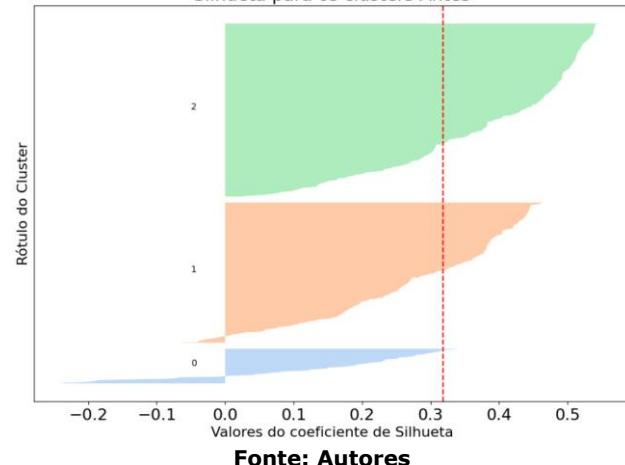
Na preparação dos dados, foram coletados mais de 90 mil registros acadêmicos, incluindo variáveis como a média geral dos alunos, número total de faltas e os cursos de formação. Utilizamos dados do SIGA para filtrar e estruturar as informações, removendo variáveis irrelevantes e normalizando os dados para garantir que todas as variáveis

estivessem na mesma escala. Após essa etapa, o conjunto foi reduzido para 3.143 registros relevantes, concentrando-se nas variáveis mais significativas para o agrupamento.

Foi aplicado o algoritmo K-Means para realizar o agrupamento dos alunos, com o objetivo de identificar grupos com comportamentos acadêmicos semelhantes. Como o algoritmo K-Means requer que o número de clusters seja definido previamente, testamos diferentes valores de K utilizando métodos como o "elbow method" (método do cotovelo) para encontrar o número ideal de clusters. O agrupamento foi baseado nas variáveis de: ranking do aluno, número de períodos cursados, disciplinas matriculadas, número de reprovação por faltas, número de reprovações por média e número de aprovações, com relação ao desempenho antes do estágio e depois do estágio.

Após a aplicação da análise de coeficiente de silhueta para avaliar a qualidade dos agrupamentos antes do estágio, conforme Gráfico 3, verificou-se que o agrupamento com 3 clusters apresenta uma separação consistente entre os grupos, com valores médios de silhueta superiores a 0.3. Isso sugere que a divisão em 3 clusters é uma escolha apropriada para capturar bem a estrutura dos dados.

Gráfico 3 - Análise do Método da Silhueta para determinar o K-Means Antes do Estágio
Silhueta para os clusters Antes

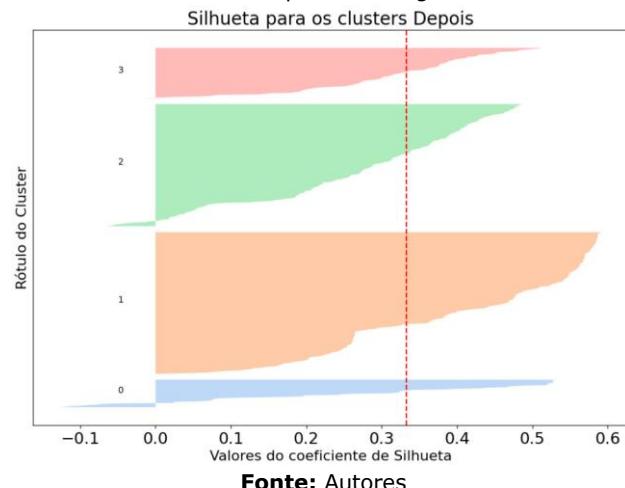


Fonte: Autores

Após o estágio, a análise dos coeficientes de silhueta, conforme Gráfico 4, indica que o agrupamento com 4 clusters apresenta uma qualidade geral ligeiramente melhor em comparação à situação anterior. O valor médio de silhueta ultrapassa 0.3 para a maioria dos clusters, destacando uma boa separação e coesão dentro dos grupos.

Análise de Dados dos Estágios Obrigatórios da Escola Politécnica de Pernambuco (POLI)

Gráfico 4 – Análise do Método da Silhueta para determinar o K-Means Depois do Estágio

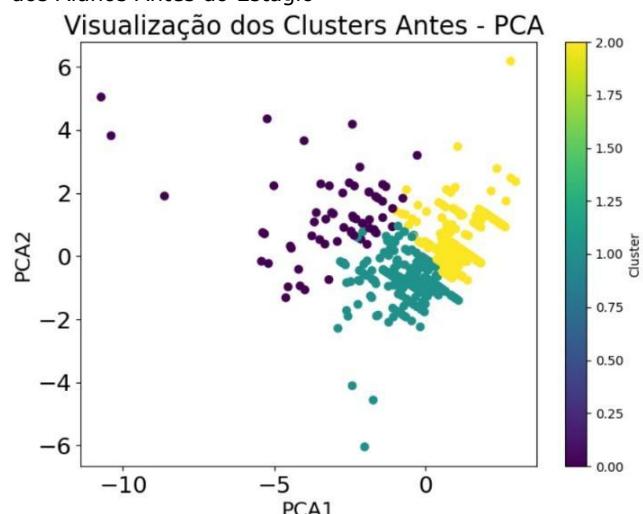


Fonte: Autores

A análise de clusterização, revelou padrões distintos de desempenho por área de formação, destacando mudanças e aprimorando a segmentação dos dados para uma interpretação mais precisa.

O Gráfico 5 apresenta a clusterização dos alunos antes do estágio utilizando o método K-means, com a redução dimensional aplicada via Análise de Componentes Principais (PCA). O gráfico evidencia três clusters distintos, representados por cores diferentes, que agrupam os alunos com base em características de desempenho similares. Cada ponto no gráfico corresponde a um aluno, posicionado em relação às duas principais componentes que explicam a maior variação nos dados.

Gráfico 5 – Clusterização por K-means do Desempenho dos Alunos Antes do Estágio



Fonte: Autores

Os clusters identificados antes do estágio revelam padrões específicos de desempenho dos alunos, categorizando-os em grupos homogêneos com base em características como ranking acadêmico, número de períodos cursados, quantidade de disciplinas cursadas, reprovações por falta ou média, e aprovações. O Cluster 0, representando 10,06% dos alunos, apresenta o menor ranking médio e menos aprovações, indicando possíveis dificuldades de desempenho. Já o Cluster 1, com 40,25% dos alunos, destaca-se pelo maior número médio de períodos cursados, sugerindo maior progresso acadêmico. O Cluster 2, representando 49,69% dos alunos, evidencia o maior ranking médio e maior número de aprovações, refletindo um desempenho mais consistente. Essa análise permite identificar áreas de desafio e pontos fortes em diferentes grupos, fornecendo subsídios para intervenções pedagógicas mais direcionadas e eficazes, conforme a Tabela 1.

Tabela 1 - Características dos Clusters Antes do Estágio

Fonte: Autores

Variáveis	0	1	2
No. Alunos	64	256	316
MC%	10,06%	40,25%	49,69%
Ranking	7,345	7,845	8,175
No. Disciplinas	5	5	6
Reps/Faltas	1	0	0
Reps/Média	1	0	0
Aprovações	3	4	6

O Gráfico 6 apresenta a clusterização do desempenho dos alunos após o estágio, utilizando o método K-means, neste caso, são identificados quatro clusters distintos, representados por diferentes cores, refletindo grupos homogêneos de alunos com características de desempenho similares.

A distribuição dos clusters após o estágio reflete alterações significativas nos padrões de desempenho dos alunos, indicando maior diversidade nos resultados acadêmicos. O Cluster 0, que representa 8,18% dos alunos, possui o menor ranking médio e ausência de aprovações, evidenciando um grupo com maiores desafios. O Cluster 1, com 41,35% dos alunos, apresenta um alto número médio de disciplinas cursadas e aprovações, sugerindo consistência acadêmica. O

Cluster 2, abrangendo 35,69% dos alunos, demonstra um desempenho equilibrado, enquanto o Cluster 3, com 14,78% dos alunos, apresenta desempenho intermediário, com algumas reprovações e aprovações moderadas. Essas diferenças entre os grupos reforçam o impacto do estágio na formação, permitindo identificar áreas de maior ou menor desenvolvimento, e oferecendo insumos para a melhoria contínua do processo pedagógico, conforme Tabela 2.

Gráfico 6 – Clusterização por K-means do Desempenho dos Alunos Depois do Estágio

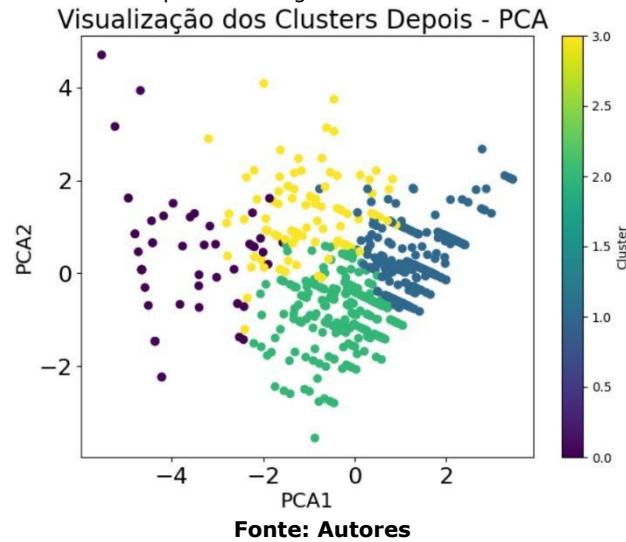


Tabela 2 - Características dos Clusters Depois do Estágio
Fonte: Autores

	Cluster			
Variáveis	0	1	2	3
No. Alunos	52	263	227	94
MC%	8,18%	41,35%	35,69%	14,78%
Ranking	2,24	8,18	7,67	6,275
No. Disciplinas	4,75	6	4,5	5,5
Reps/Faltas	4	0	1	1
Reps/Média	0	0	0	1
Aprovações	0	5,5	3,5	3

4 RESULTADOS

4.1 RESULTADOS OBTIDOS

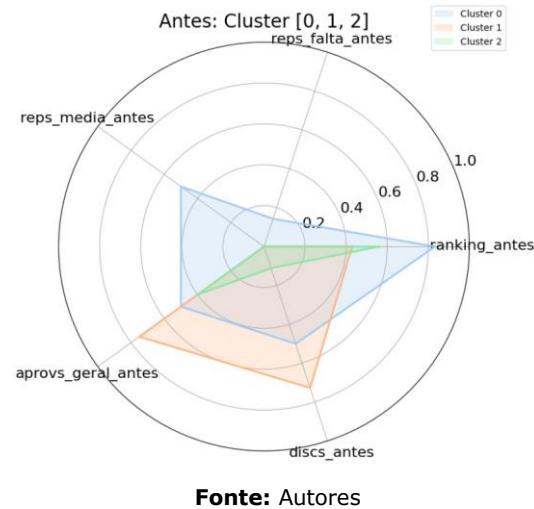
A análise dos três clusters demonstra perfis distintos de desempenho acadêmico dos alunos antes do estágio, conforme ilustrado no Gráfico 7. O Cluster 0, com 64 indivíduos, apresenta um perfil

acadêmico equilibrado, com bom desempenho no ranking geral (0.8) e um número moderado de disciplinas cursadas (0.6), refletindo um envolvimento significativo no ambiente acadêmico. As reprovações por média e por falta são baixas (entre 0.2 e 0.4), indicando boa frequência e esforço para atingir as médias mínimas. No entanto, as taxas de aprovação geral são moderadas (entre 0.4 e 0.6), sugerindo que, embora os alunos sejam assíduos e engajados, enfrentam dificuldades em transformar esse esforço em maior aproveitamento.

O Cluster 1, com 256 indivíduos, reflete um desempenho acadêmico mais mediano. Os alunos desse grupo cursaram um número alto de disciplinas cursadas (entre 0.6 e 0.8) e apresentam um posicionamento no ranking geral em torno de 0.4, o que indica um engajamento baixo. As taxas de aprovação geral são (0.3), refletindo dificuldades em alcançar o sucesso nas disciplinas. Apesar disso, as reprovações por falta e por média permanecem muito baixas.

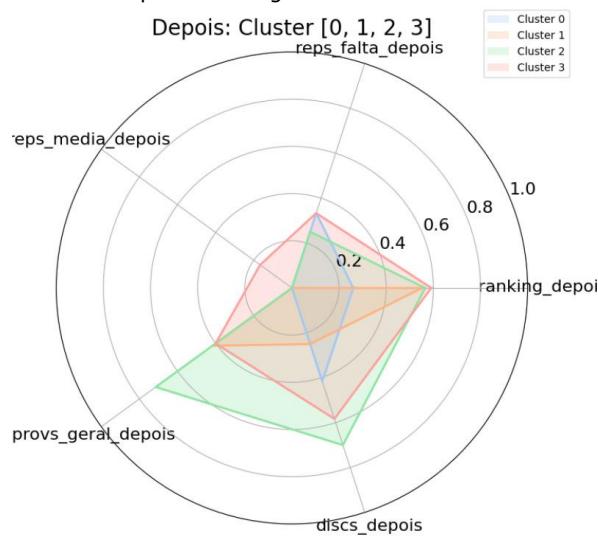
Já o Cluster 2, composto por 316 indivíduos, apresenta um perfil com desempenho geral equilibrado, embora com indicadores que apontam para um envolvimento moderado. Os alunos desse grupo possuem um posicionamento razoável no ranking geral (próximo de 0.6) e cursaram um número intermediário de disciplinas (próximo de 0.2), sugerindo um engajamento acadêmico consistente, mas não excepcional. As taxas de reprovação por média e por falta são baixas (abaixo de 0.2), indicando que, apesar de uma participação moderada, esses alunos mantêm uma frequência adequada às atividades acadêmicas.

Gráfico 7 – Clusterização por K-means do desempenho dos alunos antes do estágio



Em outra perspectiva, o Gráfico 8 apresenta a análise dos clusters após o estágio em perfis distintos de desempenho acadêmico entre os alunos. O Cluster 0, com uma população de 52 indivíduos, apresenta o menor número de alunos entre os grupos analisados após o estágio. Esse cluster demonstra desempenho muito limitado em todas as variáveis analisadas. As taxas de aprovação geral (0.2) e o número de disciplinas cursadas (0.2) são baixos, indicando baixo engajamento acadêmico e dificuldades consistentes no progresso nas disciplinas. As reprovações por falta próximas a 0.4, o que pode indicar uma participação limitada, mas sem grandes faltas. Esse grupo reflete alunos que precisam de suporte acadêmico intensivo e estratégias personalizadas para melhorar o engajamento e a retenção de conteúdo.

Gráfico 8 – Clusterização por K-means do desempenho dos alunos depois do estágio



Fonte: Autores

O Cluster 1, composto por 263 indivíduos, reflete um desempenho mediano, com valores intermediários em várias métricas. O número de disciplinas cursadas (acima de 0.2) e o posicionamento no ranking geral (próximo a 0.6) sugerem um engajamento moderado. As reprovações por falta e por média também estão abaixo de 0.2, o que indica boa frequência. No entanto, as taxas de aprovação geral são moderadas (próximo 0.6) são alunos que enfrentam dificuldades em traduzir o engajamento em sucesso acadêmico. Este grupo pode se beneficiar de iniciativas voltadas à melhoria da eficiência no

aprendizado, como programas de reforço e monitoria.

O Cluster 2, com 227 indivíduos, apresenta características mais positivas em comparação aos outros grupos. Os alunos deste cluster cursaram um número razoável de disciplinas (acima de 0.6), tiveram boa aprovação geral (próximo a 0.8) e um posicionamento consistente no ranking geral (0.6). As reprovações por falta e por média são muito baixas (entre 0.2 e 0.4), indicando que esses alunos participam ativamente e apresentam engajamento acadêmico acima da média. Este grupo representa um perfil de alunos com bom desempenho geral, mas que ainda podem trabalhar na consolidação de resultados.

O Cluster 3, formado por 94 indivíduos, reflete um perfil equilibrado, com desempenhos intermediários em todas as métricas. O número de disciplinas cursadas (0.4), as taxas de aprovação geral (0.4) e o posicionamento no ranking geral (0.6) sugerem alunos que mantêm um desempenho constante, mas sem grandes destaques. As reprovações por falta e por média (entre 0.2 e 0.4) são baixas, indicando que a frequência não é um problema para esse grupo. Este cluster pode melhorar ainda mais com apoio direcionado ao fortalecimento das habilidades acadêmicas para transformar a constância em resultados superiores.

5 CONCLUSÕES

O estudo alcançou seus objetivos ao aplicar técnicas de mineração de dados para identificar padrões e variáveis no desempenho acadêmico dos alunos de engenharia da POLI, antes e depois do estágio obrigatório. Utilizando o algoritmo K-Means, foi possível agrupar os alunos em clusters distintos, fornecendo uma análise detalhada sobre o impacto do estágio. A pesquisa consolidou dados fragmentados, correlacionou desempenho acadêmico com estágios e gerou insights relevantes para a gestão acadêmica e os processos de estágio.

Os resultados mostram que o estágio obrigatório influencia diretamente o desempenho dos alunos, trazendo avanços e desafios. Antes do estágio, os clusters apresentaram padrões consistentes de engajamento e desempenho acadêmico. Após o estágio, houve maior diversidade nos resultados, com alguns grupos melhorando em aprovações e rankings, enquanto outros enfrentam dificuldades, como reprovações por falta. Esses achados destacam a importância do

estágio na formação, mas também apontam a necessidade de intervenções para minimizar os desafios enfrentados por parte dos alunos.

O estudo sugere que futuras pesquisas explorem algoritmos mais avançados, como DBSCAN e Random Forest, para identificar padrões mais complexos e outliers. Análises sobre tipos de estágio, como remunerados ou em diferentes setores, também podem aprofundar o entendimento do impacto no desempenho acadêmico e na empregabilidade. Além disso, investigações longitudinais, acompanhando os alunos desde a entrada na instituição até sua inserção no mercado de trabalho, podem oferecer uma visão abrangente do papel do estágio, orientando políticas educacionais mais alinhadas às demandas do mercado.

REFERÊNCIAS

- [1] KOLB, D. A. Experiential Learning: Experience as the Source of Learning and Development. Prentice-Hall, 1984.
- [2] DAVENPORT, T. H.; HARRIS, J. G. Competing on Analytics: The New Science of Winning. Harvard Business Review Press, 2007.
- [3] SILVA, M. R. A integração teoria-prática nos cursos superiores: desafios e oportunidades. Editora Universitária, 2015.
- [4] HAN, J.; KAMBER, M.; PEI, J. Data Mining: Concepts and Techniques. 3rd ed. Morgan Kaufmann, 2012.
- [5] CUNHA, A. C.; CAVALCANTI, A. G. Integração entre Teoria e Prática no Ensino de Engenharia: Desafios e Oportunidades. Revista Brasileira de Educação em Engenharia, v. 42, n. 1, p. 45-60, 2020.
- [6] ROMERO, C.; VENTURA, S. Educational Data Mining and Learning Analytics: An Updated Survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, v. 10, n. 3, p. e1355, 2020.
- [7] SILVA, M. R.; BORGES, R. F. O papel do estágio na formação de engenheiros: desafios e oportunidades. Editora Universitária, 2016.
- [8] FREIRE, P. A.; MIRANDA, G. R. A Influência do Estágio na Formação Acadêmica: Uma Análise sobre o Impacto no Desempenho dos Alunos de Engenharia. Revista Brasileira de Educação e Tecnologia, v. 34, n. 2, p. 56-70, 2021.
- [9] RIBEIRO, R.; OLIVEIRA, L. A importância do estágio na formação profissional: uma análise de impacto. Revista Brasileira de Educação, 2018.
- [10] JACKSON, D. The contribution of work-integrated learning to graduate employability. Asia-Pacific Journal of Cooperative Education, v. 14, n. 2, p. 99-115, 2013.
- [11] MIRANDA, C. A.; ALMEIDA, R. A. A Supervisão de Estágios: Desafios e Potenciais no Ensino Superior. Editora Vozes, 2019.
- [12] LLOYD, S. P. Least squares quantization in PCM. IEEE Transactions on Information Theory, v. 28, n. 2, p. 129-137, 1982.
- [13] MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, v. 1, n. 14, p. 281-297, 1967.
- [14] TAN, P. N.; STEINBACH, M.; KUMAR, V. Introduction to Data Mining. Pearson Addison Wesley, 2006.
- [15] JAIN, A. K. Data clustering: 50 years beyond K-means. Pattern Recognition Letters, v. 31, n. 8, p. 651-666, 2010.
- [16] PEÑA-AYALA, A. Educational Data Mining: A Survey and a Data Mining-Based Analysis of Recent Works. Expert Systems with Applications, v. 138, p. 112816, 2019.
- [17] FERREIRA, T. L.; ANDRADE, P. C. A Gestão de Estágios Obrigatórios no Ensino Superior: Desafios e Soluções com o Uso de Plataformas Digitais. Estudos em Administração Educacional, v. 39, n. 1, p. 75-89, 2022.
- [18] BICHLER, S.; KALGOTRA, P.; SHARDA, R. Data Mining for Educational Institutions: Addressing Fragmented Data Issues. Journal of Decision Systems, v. 28, n. 1, p. 89-105, 2019.
- [19] ALVES, R. G.; SILVA, M. F. Sistemas de Gestão Acadêmica no Ensino Superior: Impactos na Administração Universitária. Revista de Educação e Tecnologia, v. 45, n. 3, p. 102-118, 2021.
- [20] PEREIRA, J. M.; COSTA, A. F. Mineração de Dados Educacionais: Uma Análise da Eficiência dos Sistemas de Gestão Acadêmica em Instituições Públicas. Journal of Educational Data, v. 18, n. 2, p. 55-70, 2023.