



21 A 23 DE NOVEMBRO DE 2025
XXX ENAPET

INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS:
DESAFIOS ÉTICOS PARA O SÉCULO XXI

O USO ÉTICO DA INTELIGÊNCIA ARTIFICIAL COMO FERRAMENTA DE DIAGNÓSTICO NA EDUCAÇÃO

PINTO, A. R.¹; ARAÚJO, J. I. S.¹; MIRANDA, M. E.¹; LEITE, M. M. A.¹; GOMES, V. A.¹; ALMEIDA, D. D. F.¹; COSTA, U. S.²

¹Grupo PET Ciência da Computação, UFRN, Campus Central; ²Tutor do Grupo PET Ciência da Computação, UFRN, Campus Central

artur.revoredo.116@ufrn.edu.br, jose.ivo.araujo.110@ufrn.edu.br, mariana.miranda.088@ufrn.edu.br,
marina.leite.018@ufrn.edu.br, victor.gomes.090@ufrn.edu.br, davi.almeida.135@ufrn.edu.br;
umberto.costa@ufrn.br, petcc@dimap.ufrn.br

RESUMO: Este trabalho demonstra o potencial do uso de algoritmos de Inteligência Artificial para realizar análises de dados de maneira ética. Ao empregar a clusterização com o algoritmo K-Means, uma técnica de aprendizado não supervisionado, em dados educacionais abertos da UFRN, a abordagem busca contornar vieses e conceitos pré-estabelecidos que poderiam contaminar uma análise humana. A aplicação da técnica revelou padrões objetivos, como a forte correlação entre turmas menores e alto rendimento acadêmico, e a descoberta contraintuitiva de que a maior titulação docente (doutorado) não se associou diretamente ao melhor desempenho dos discentes. Dessa forma, o estudo demonstra como a Inteligência Artificial pode ser uma ferramenta poderosa e confiável para extrair insights a partir de grandes coleções de dados, minimizando o impacto do viés dos pesquisadores, orientando intervenções pedagógicas e a otimização de recursos de forma justa e baseada em evidências.

Palavras-chave: Dados; Inteligência Artificial; Avaliação Institucional; Educação

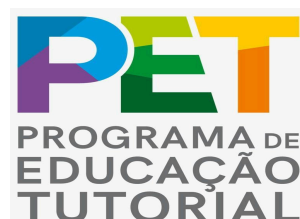
The Ethical Use of Artificial Intelligence as a Diagnostic Tool in Education

ABSTRACT: This work demonstrates the potential of using Artificial Intelligence algorithms for ethical data analysis. By employing K-Means clustering, an unsupervised learning technique, on open educational data from UFRN (Federal University of Rio Grande do Norte), this approach seeks to circumvent biases and preconceived notions that could contaminate human analysis. The application of the technique revealed objective patterns, such as a strong correlation between smaller class sizes and high academic performance, and the counterintuitive discovery that higher faculty qualifications were not directly associated with better student performance. Thus, this study demonstrates how Artificial Intelligence can be a powerful and reliable tool for extracting insights from large data collections, minimizing the impact of researcher bias, and guiding pedagogical interventions and resource optimization in a fair and evidence-based manner.

Keywords: Data; Artificial Intelligence; Institutional Evaluation; Education



XXX ENCONTRO NACIONAL DOS GRUPOS PET
UNIVERSIDADE DE BRASÍLIA (UnB)
Campus Darcy Ribeiro, Asa Norte
70910-900, Brasília - DF





21 A 23 DE NOVEMBRO DE 2025
XXX ENAPET

INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS: DESAFIOS ÉTICOS PARA O SÉCULO XXI

Introdução

A sociedade moderna é marcada pela conectividade e pela geração contínua e massiva de dados. Este fato exige aplicação de metodologias e tecnologias capazes de coletar, analisar e interpretar esses dados, extraindo informações muito valiosas para diversos setores da atividade humana. Dentre as muitas preocupações que surgem ao trabalhar com dados, uma das principais é como analisá-los, principalmente de forma ética, sem vieses e conceitos pré-estabelecidos que contaminariam as conclusões finais que se procuram estabelecer.

Nesse contexto, a Computação, através da Inteligência Artificial (IA), procura resolver essas preocupações por meio de ferramentas estatísticas que realizam a clusterização (agrupamento) (Jain, Murty & Flynn, 1999) de dados que sigam um mesmo padrão, trazendo à tona relações implícitas e de difícil percepção por meio da análise humana.

Este trabalho revela o potencial de algoritmos de IA aplicados à análise dos dados abertos de nossa Universidade, evidenciando pontos de melhoria e corroborando na busca de uma educação equitativa, um direito humano fundamental. As técnicas empregadas na análise de dados oferecem resultados objetivos e livres de pressupostos, oferecendo informações confiáveis e úteis para a tomada de decisões sobre intervenções e adoção de estratégias pedagógicas específicas para cada perfil de turma e para otimizar a alocação de recursos.

Método

Clusterização, na Computação, é uma tarefa de Aprendizado de Máquina Não Supervisionado que agrupa um conjunto de objetos baseando-se em similaridades entre os dados. O objetivo é descobrir estruturas ou padrões inerentes nos dados sem qualquer conhecimento prévio sobre as categorias existentes, pois o algoritmo não recebe exemplos rotulados de como os grupos devem ser formados. Neste trabalho, aplicamos a clusterização a dados oriundos do contexto educacional para a compreensão de fenômenos complexos, considerando múltiplas dimensões do desempenho acadêmico. Dessa forma, a partir de



INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS:
DESAFIOS ÉTICOS PARA O SÉCULO XXI

clusters (grupos) matemáticos obtidos, padrões ocultos no perfil das turmas são evidenciados, o que dificilmente se revelaria por meio de análises univariadas e bivariadas tradicionais.¹

No âmbito do Aprendizado Não Supervisionado, uma série de algoritmos se destacam por diferentes abordagens. Neste trabalho, o algoritmo *K-Means* (LLOYD, 1982), foi escolhido devido a sua eficiência computacional e interpretabilidade. A aplicação da técnica foi dividida em quatro etapas: (i) coleta e limpeza de dados; (ii) escolha de variáveis e seleção de números de *clusters*; (iii) validação dos *clusters*; (iv) análise dos *clusters*.

A coleta de dados foi realizada a partir de dados abertos disponibilizados de forma transparente pela Superintendência de Tecnologia da Informação da UFRN (ANDRESSA KROEFF E ANDRÉ LOURENÇO, 2025). Dentre as informações disponíveis, decidiu-se pela exploração de três bases de dados coletadas no período de 2013 a 2024: *Avaliação de Docência*, *Docentes* e *Turmas*. Estas bases de dados atendem a critérios fundamentais: possuem forte relevância para a comunidade acadêmica, possibilitando reflexões sobre a qualidade do Ensino e o desempenho dos docentes a partir da perspectiva dos discentes; são complexas o suficiente para que sejam extraídos *insights* relevantes. Após a seleção, foi realizado o processo de união e limpeza dos dados coletados, construindo uma base unificada, possibilitando a continuidade do trabalho para as etapas de escolha e variáveis e seleção de número de *clusters*.

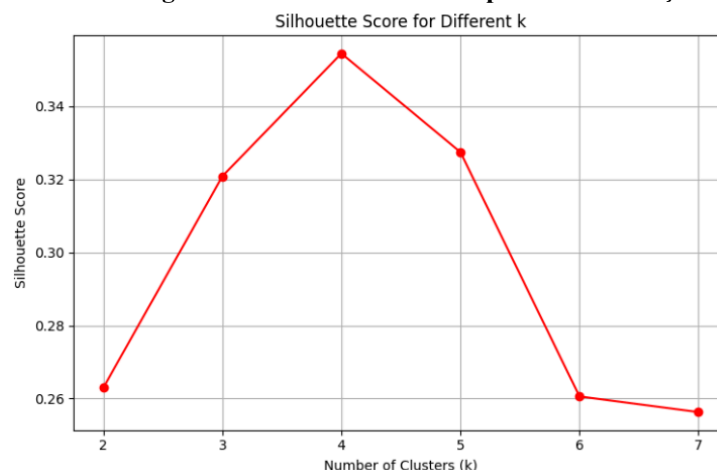
Para identificar características de turmas de sucesso, foram selecionadas cinco variáveis que capturam diferentes dimensões do desempenho acadêmico e profissional, sendo estas: (i) proporção de aprovados da turma; (ii) quantidade de discentes da turma; (iii) média final geral da turma; (iv) nota de atuação profissional do docente, referente à atuação didática; (v) nota de postura profissional do docente. As variáveis foram escolhidas por representarem aspectos qualitativos e quantitativos do processo educacional, permitindo uma caracterização abrangente das turmas. Como estas variáveis possuem escalas numéricas muito distintas, foi necessário realizar um ajuste estatístico para padronizar os dados. Tal ajuste permite que o

¹ A análise **univariada** refere-se à descrição de uma única variável por vez — por exemplo, cálculo de média, desvio-padrão ou histogramas. Já a análise **bivariada** investiga a relação entre duas variáveis, por meio de técnicas como correlação, regressão linear simples, tabelas de contingência, entre outras.

algoritmo compare diferentes dimensões de forma equilibrada, colocando todas as informações em uma mesma escala, garantindo que nenhuma variável influencie desproporcionalmente a formação dos *clusters* apenas por ter uma magnitude numérica maior.

A escolha do número de *clusters* (k) é uma escolha crítica que impacta diretamente a qualidade e a interpretabilidade dos resultados. Para auxiliar na tomada de decisão, foi utilizado o método da silhueta (SHUTAYWI, M et al., 2021). Este coeficiente mede a similaridade de cada amostra em relação ao seu *cluster* designado e aos demais, variando de -1 (classificação incorreta) a +1 (classificação correta). Para a base de dados estudada, a análise de silhueta indicou $k=4$ como configuração ótima, apresentando o maior score de separação entre *clusters*, como referenciado na Figura 1.

Figura 1 – Análise de Silhueta para Clusterização



Fonte: Autoria própria (2025).

Para validação dos resultados da clusterização, os *clusters* obtidos foram inspecionados visualmente através da Análise de Componentes Principais (PCA) (Hotelling, 1933). Esta análise revelou a formação de nuvens de pontos distintos e com baixa sobreposição, confirmando a geração de *clusters* coerentes. Tendo a estrutura dos *clusters* sido validada tanto quantitativamente (pela silhueta) quanto qualitativamente (pela inspeção visual), encerra-se a criação do modelo de dados. A seção a seguir se dedica à última etapa da metodologia adotada, voltada à análise de cada um dos perfis encontrados pela clusterização.

Resultados e Discussão

Com a estrutura dos *clusters* validada estatística e visualmente, a análise avança para a interpretação substantiva de cada *cluster* individualmente. Esta etapa consiste em examinar características de cada *cluster*, com base nos valores médios das cinco variáveis selecionadas. Ao analisarmos as médias das estatísticas de cada *cluster*, é possível extrair *insights* práticos sobre as diferentes dinâmicas de turma presentes no conjunto de dados, transformando valores numéricos dispersos em conhecimento aplicável ao contexto educacional.

Figura 2 – Estatísticas Descritivas dos Clusters

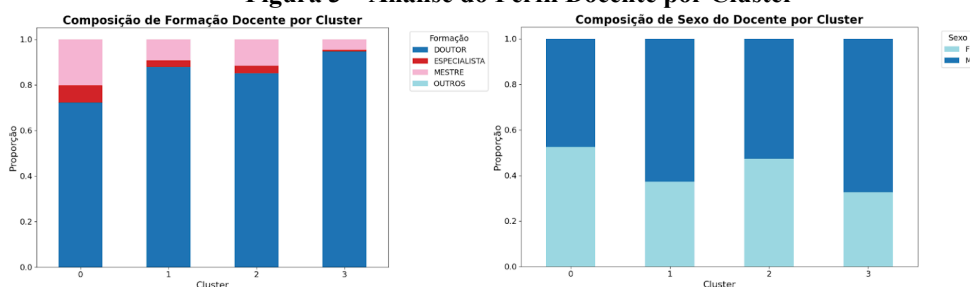
Cluster	0	1	2	3
Quantidade de Turmas	48.530,00	13.113,00	50.678,00	13.472,00
Proporção de Aprovados	0,972	0,869	0,943	0,581
Quantidade de Discentes	12,935	26,256	26,984	32,838
Média Final Geral	8,257	7,716	8,078	6,534
Atuação Didática	9,741	8,094	9,100	8,706
Postura Profissional	9,845	7,861	9,390	9,309

Fonte: Autoria própria (2025).

A análise inicial dos *clusters* obtidos, referenciados na Figura 2, revelou quatro perfis distintos de turmas: (0) *Perfil de Alto Rendimento*; (1) *Perfil de Desempenho Intermediário*; (2) *Perfil de Rendimento Sólido e Consistente*; (3) *Perfil de Dissonância Acadêmica*. O principal fator evidenciado é a forte correlação entre o tamanho da turma e o sucesso acadêmico. O *Perfil de Dissonância Acadêmica* (*cluster* 3) sugere que turmas com mais discentes estão diretamente associadas com menores taxas de aprovação e médias gerais. Em nítido contraste, o *Perfil de Alto Rendimento* (*cluster* 0) associa turmas menores a maiores taxas de aprovação e notas. Em outra perspectiva, o *Perfil de Desempenho Intermediário* e o *Perfil de Rendimento Sólido e Consistente* (*clusters* 1 e 2, respectivamente) evidenciam a diferença da abordagem profissional dos docentes em turmas com quantidade de discentes parecidas, visto que, sua maior distinção recai sobre as notas relativas às variáveis *Atuação Didática* e *Postura Profissional*, sugerindo abordagens de ensinos passíveis de melhoria.

Aprofundando essa análise, ao investigar o perfil docente por *cluster*, a pesquisa revelou correlações notáveis e contra intuitivas com o desempenho dos discentes, conforme dados apresentados na Figura 3. Surpreendentemente, a maior qualificação formal não se associou diretamente aos melhores resultados. O *cluster* 3 apesar de ser lecionado pela maior proporção de doutores, apresenta as menores médias de desempenho. Por outro lado, o *cluster* 0, embora possua as melhores médias de desempenho, detém a maior diversidade de qualificação de docentes, com mais mestres e especialistas. Adicionalmente, a composição de gênero demonstra que os *clusters* com melhor desempenho (0 e 2) possuem um maior equilíbrio entre a presença de professoras e professores. Em contrapartida, os *clusters* de menor desempenho (1 e 3) são predominantemente lecionados por docentes do sexo masculino. Embora essas relações não impliquem causalidade, elas levantam novas hipóteses a serem exploradas, sugerindo outros aspectos a investigar.

Figura 3 – Análise do Perfil Docente por Cluster



Fonte: Autoria própria (2025).

Estes achados, especialmente os contra intuitivos, reforçam o valor da abordagem ética de análise de dados proposta neste trabalho. Uma análise humana poderia facilmente partir do pressuposto de que maior qualificação formal levaria a melhores resultados, incorrendo em um viés de confirmação. O algoritmo, por sua vez, operou livre de tais preconceitos, revelando características implícitas nos *clusters* de dados. Dessa forma, a IA serve como uma ferramenta que oferece maior imparcialidade, ao não fornecer respostas definitivas, mas levanta hipóteses mais profundas e objetivas, permitindo que a instituição investigue os fatores de sucesso com base em evidências, e não em suposições.



21 A 23 DE NOVEMBRO DE 2025
XXX ENAPET

INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS: DESAFIOS ÉTICOS PARA O SÉCULO XXI

Conclusões

Os resultados encontrados ao longo deste trabalho dialogam diretamente com o tema da Inteligência Artificial e Direitos Humanos. Um ponto central é que a clusterização é realizada de forma puramente matemática, em que a intervenção humana se limita na escolha dos dados disponibilizados, se distanciando de vieses cognitivos e pré concepções humanas que podem influenciar a avaliação tradicional. Nesse sentido, a IA atua como uma ferramenta útil para possibilitar a garantia de um direito humano fundamental: o acesso a uma educação igualitária. Contudo, essa mesma capacidade de classificar introduz um dilema ético central: como utilizar as informações geradas para apoiar os alunos sem, erroneamente, criar rotulações para alunos ou professores. Portanto, os resultados reforçam que, enquanto a tecnologia oferece informações valiosas, a responsabilidade de traduzir esses dados em ações justas e humanizadas permanece sob responsabilidade humana.

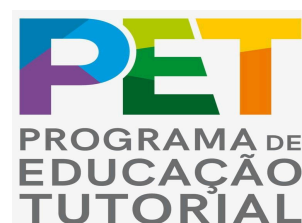
Ao final do trabalho, foi observado que os *clusters* criados são formados por milhares de turmas que podem ter variabilidade interna significativa não capturada pelas médias apresentadas nas estatísticas observadas. Além disso, a análise realizada é transversal e não captura evolução temporal ou trajetória estudantil longitudinal. Desse modo, trabalhos futuros podem se aprofundar em novas métricas de análises qualitativas, estudos das turmas diante da evolução temporal e composição das disciplinas lecionadas em cada *cluster*.

Agradecimentos

Agradecemos ao Fundo Nacional de Desenvolvimento da Educação (FNDE), pelo fomento e suporte ao Programa de Educação Tutorial (PET), e ao Departamento de Informática e Matemática Aplicada da UFRN (DIMAP).



XXX ENCONTRO NACIONAL DOS GRUPOS PET
UNIVERSIDADE DE BRASÍLIA (UnB)
Campus Darcy Ribeiro, Asa Norte
70910-900, Brasília - DF





21 A 23 DE NOVEMBRO DE 2025
XXX ENAPET

INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS:
DESAFIOS ÉTICOS PARA O SÉCULO XXI

Referências

SHUTAYWI, M.; et al. **Silhouette analysis for performance evaluation in machine learning clustering**. PMC – National Institutes of Health, 2021. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8234541/>. Acesso em: 6 out. 2025.

MINISTÉRIO DA EDUCAÇÃO (Brasil). UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE. **Plano de Dados Abertos**. Disponível em: https://ufrn.br/resources/documentos/planodedadosabertos/Plano_de_dados_abertos_da_UFRN_-_2022-2024.pdf. Acesso em: 6 out. 2025.

AHMED, M.; SERAJ, R.; ISLAM, S. M. S. **The k-means Algorithm: A Comprehensive Survey and Performance Evaluation**. *Electronics*, v. 9, n. 8, p. 1295, 2020. DOI: 10.3390/electronics9081295.

LLOYD, S. P. **Least squares quantization in PCM**. *IEEE Transactions on Information Theory*, v. 28, n. 2, p. 129-137, mar. 1982. DOI: 10.1109/TIT.1982.1056489.

A. K. Jain, M. N. Murty, and P. J. Flynn. 1999. **Data clustering: a review**. *ACM Comput. Surv.* 31, 3 (Sept. 1999), 264–323. <https://doi.org/10.1145/331499.331504>

ANDRESSA KROEFF E ANDRÉ LOURENÇO - STI. **Avaliação Institucional da UFRN**. Disponível em: <<https://www.avaliacao.ufrn.br/documento.php?id=117274266>>. Acesso em: 6 out. 2025.

Hotelling, H. (1933). **Analysis of a complex of statistical variables into principal components**. *Journal of Educational Psychology*, 24, 417-441. <http://dx.doi.org/10.1037/h0071325>



XXX ENCONTRO NACIONAL DOS GRUPOS PET
UNIVERSIDADE DE BRASÍLIA (UnB)
Campus Darcy Ribeiro, Asa Norte
70910-900, Brasília - DF

