Classificação curricular das questões do ENADE em Engenharia de Computação: uma mineração de texto

Katia Emanuelly de Souza[1,2], **Milton Miranda Neto**[3], **Cristiane Aparecida Lana**[1,2]

[1]Centro Universitário de Viçosa (Univiçosa) – Viçosa – MG – Brasil

[2]Universidade Federal de Viçosa (UFV) – Viçosa – MG – Brasil

[3]Universidade de São Paulo (USP) – São Paulo – SP – Brasil

katiaemanuelly60,milton.cm.miranda@gmail.com,cristiane.lana@alumni.usp.br

Resumo. *A mineração de texto e algoritmos de aprendizado de máquina têm sido amplamente utilizados para classificar textos e extrair padrões ocultos em dados educacionais. No entanto, poucas pesquisas exploram o relacionamento entre o conteúdo das questões do ENADE, componentes curriculares e Taxonomia de Bloom. Este estudo analisou e classificou as questões do ENADE do curso de Engenharia da Computação, utilizando o framework CRISP-DM e dados dos últimos cinco anos do ENADE. Os resultados mostraram que os núcleos "Núcleo Computação"e "Núcleo Desenvolvimento Pessoal"foram os mais abordados, com ênfase nos níveis da Taxonomia de Bloom de conhecimento, compreensão e aplicação.*

Abstract. *Text mining and machine learning algorithms have been widely used to classify texts and extract hidden patterns in educational data. However, few studies have explored the relationship between ENADE questions, curricular components, and Bloom's Taxonomy. This study analyzed and classified ENADE questions for the Computer Engineering program using the CRISP-DM framework and data from the last five years of ENADE. The results indicated that the "Computing Core"and "Personal Development Core"were the most addressed areas, with an emphasis on Bloom's Taxonomy levels of knowledge, comprehension, and application.*

1. Introdução

A educação superior desempenha um papel fundamental no desenvolvimento econômico regional e na formulação de políticas públicas nacionais [Gallagher 2016]. Contudo, para que a educação superior seja eficaz e responda às demandas do mercado de trabalho, é necessário submetê-la a avaliações que validem sua qualidade e adequação às constantes transformações no mercado [Tambling 2020]. Nesse contexto, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) tem implementado diversas formas de avaliar a educação no Brasil. Para a educação superior, em particular, foi instituído em 2004 o Sistema Nacional de Avaliação da Educação Superior (SINAES), do qual o Exame Nacional de Desempenho dos Estudantes (ENADE) é um dos principais instrumentos de avaliação [Lopes e Vendramini 2015].

DOI: 10.5753/sbie.2024.244889

O ENADE avalia o desempenho dos estudantes em relação às competências adquiridas ao longo do curso [Pissaia 2018, FDV 2021], fornecendo relatórios que auxiliam os coordenadores de curso na adequação da grade curricular [Sousa e Sousa 2012, Alvarenga et al. 2017]. A prova é composta por 40 questões, sendo 10 voltadas para a formação geral e 30 específicas para cada área [BRASIL 2020]. As questões são frequentemente elaboradas com base na Taxonomia de *Bloom* [Olivera 2011], que organiza objetivos educacionais em uma hierarquia de complexidade cognitiva, desde habilidades mais básicas, como o conhecimento memorizado, até as mais complexas, como análise e avaliação [Olivera 2011, Oliveira et al. 2016].

Com as mudanças no cenário educacional, o conceito de Mineração de Dados Educacionais (*Educational Data Mining*) [Ramos et al. 2020] e Mineração de Texto [Abdusalomovna et al. 2023] tem ganhado destaque por sua capacidade de extrair conhecimento de dados não estruturados, como as questões do ENADE disponibilizadas em formato *Portable Document Format* (PDF) [Algarni 2016, Tavares et al. 2020]. Automatizar a análise dessas questões pode facilitar a classificação dos conteúdos e seu alinhamento com os componentes curriculares, impactando a formação dos estudantes em consonância com as mudanças do mercado [Charão et al. 2020]. Para essa automatização, algoritmos de mineração podem ser empregados para extrair conhecimento desses textos, auxiliando na tomada de decisões dos coordenadores de cursos, como, por exemplo, a identificação de componentes curriculares que necessitam de ajustes. Além disso, a análise das questões do ENADE, alinhada aos componentes curriculares e à Taxonomia de *Bloom*, pode possibilitar a reavaliação da disposição das informações oferecidas aos discentes, dos métodos avaliativos e da grade curricular. A falta de uma análise técnica e detalhada dos resultados do ENADE pode prejudicar a formação dos estudantes e comprometer sua competitividade no mercado [Lira Silva et al. 2024, Oliveira et al. 2024].

Neste contexto, o objetivo deste estudo é classificar as questões do ENADE com base nos componentes curriculares, utilizando técnicas de mineração de texto. Para isso, foi adotado o *framework* CRISP-DM [Saltz e Hotz 2022], e seis algoritmos foram testados. Os melhores resultados foram obtidos com *Random Forest* e *K-Nearest Neighbors* (KNN), embora com métricas ainda insatisfatórias, como um F1-score de aproximadamente 52Essas análises preliminares, conduzidas por meio do método *data split* [Pires 2023], indicaram que os núcleos mais frequentemente abordados nas questões do ENADE são Computação e Desenvolvimento Pessoal, incluindo componentes como redes de computadores e teoria da computação. Os níveis mais utilizados da Taxonomia de *Bloom* foram conhecimento, compreensão e aplicação. Esses resultados permitem que os gestores dos cursos avaliados, compreendam melhor quais componentes curriculares estão sendo avaliados no ENADE e como esses se alinham às exigências do mercado. Quando necessário, essas informações podem orientar adaptações nos currículos dos cursos.

A baixa *performance* dos algoritmos indica a necessidade de novas abordagens analíticas. Assim, como trabalhos futuros, serão aplicados o método de validação cruzada estratificada (*stratified k-fold cross validation*), que preserva a distribuição dos dados entre os conjuntos de treino e teste [Zhang e Liu 2023], e o modelo *BERT*, amplamente utilizado na literatura para análise de contexto de palavras [NETO 2023], visando melhorar a classificação dos componentes curriculares. Com essas novas estratégias, espera-se

aumentar a precisão das classificações, melhorar a confiabilidade dos resultados, reduzir a variabilidade das previsões e promover uma compreensão mais profunda dos dados analisados, contribuindo para a melhoria do processo de tomada de decisão.

O restante do documento está organizado da seguinte forma: Na seção 2 é tratado sobre os trabalhos relacionados, enquanto na seção 3 está descrita a metodologia adotada na condução deste trabalho. Na seção 4 são abordados os resultados e a avaliação. Por fim, na seção 5 é apresentada a conclusão do mesmo.

2. Trabalhos relacionados

Embora a análise de dados do ENADE não seja um tema novo, a maioria dos estudos se concentra em instituições ou áreas específicas. Por exemplo, [Santos 2022] utiliza aprendizado de máquina para analisar projetos pedagógicos em Ciência da Computação e Sistemas de Informação, alcançando uma acurácia de cerca de 80% na previsão do Conceito ENADE Faixa e um erro percentual absoluto de aproximadamente 11% no Conceito ENADE Contínuo. Similarmente, [Araujo 2021] explora o desempenho de algoritmos como Redes Neurais, *Naive Bayes*, *Support Vector Machine (SVM)* e *Random Forest* com microdados do ENADE, constatando que a rede neural simples não foi eficiente e os outros algoritmos apresentaram desempenho mediano, sugerindo a necessidade de mais pesquisas.

O estudo de [Neto 2021] desenvolveu um modelo para auxiliar gestores e coordenadores na melhoria da tomada de decisão ao identificar características institucionais que influenciam o desempenho dos discentes no ENADE. A mineração de dados dos cursos de Tecnologia da Informação de 2014 e 2017 revelou que estratégias como oportunidades de estágio, acesso a computadores e aulas práticas são eficazes para alcançar objetivos pedagógicos. Os resultados sugerem que, para cursos como o de Sistemas de Informação, investir em planejamento didático e treinamento acadêmico pode ter um impacto significativo. Por fim, [Gerab et al. 2014] aplicaram Análise de Componentes Principais e Análise por Agrupamento Hierárquico para explorar correlações entre variáveis como desempenho acadêmico, pontuação em exames de admissão e duração do curso. A análise revelou a formação de nove grupos distintos de disciplinas, com o grupo ”Básicas”sendo o mais relevante em relação à extensão do curso.

Embora essa pesquisa seja aplicada em um único curso, ele difere dos demais por considerar os anos de 2008 a 2019 de aplicação do ENADE para o curso de Engenharia da Computação. A base é composta das questões do ENADE, Taxonomia de *Bloom*, componentes curriculares e ementas de cada componentes. A inclusão da Taxonomia de *Bloom* e a consideração dos componentes curriculares agregam valor à pesquisa, possibilitando uma análise. Para realizar a análise foram adotados seis algoritmos diferentes e está sendo considerados dois métodos de divisão da base de dados em treino e teste, o *data split* e *stratified k-fold cross validation*. Neste artigo é descrito a plicação do primeiro e segundo método será aplicado no futuro bem como o algoritmo *Bert* como descrito na Seção 1.

3. Metodologia

Para alcançar os objetivos desta pesquisa, foram implementados e analisados algoritmos de aprendizado de máquina com foco na identificação e extração de conhecimento a partir das questões do ENADE dos anos de 2008, 2011, 2014, 2017 e 2019. Esses dados foram

correlacionados com os componentes curriculares e a Taxonomia de *Bloom*, permitindo uma análise aprofundada das relações entre o conteúdo curricular e o desempenho dos alunos. O estudo foi conduzido como um estudo de caso [Ventura 2007], centrado no curso de Engenharia de Computação (ECO) de uma instituição de ensino superior.

A metodologia adotada seguiu o *framework* CRISP-DM [Saltz e Hotz 2022], um modelo amplamente utilizado para estruturação de projetos de mineração de dados, baseado no processo de *Knowledge Discovery in Databases* (KDD) [Fayyad et al. 1996]. Composto por seis etapas iterativas, o CRISP-DM provou ser eficaz na resolução de problemas específicos de negócio e na estruturação da análise de dados [IBM 2021], conforme detalhado em [Saltz e Hotz 2022]. O primeiro passo foi compreender o problema e os dados disponíveis, que estavam armazenados em arquivos PDF. A análise concentrou-se nos enunciados das questões do ENADE e nas descrições dos componentes curriculares, complementadas pelos planos de ensino para um entendimento mais profundo. A preparação dos dados seguiu o processo de ETL (*Extract, Transform, Load*), integrando informações de diferentes fontes em uma única base de dados. Esta fase de preparação de dados é conhecida por consumir entre 50% e 70% do tempo total do projeto [Pérez et al. 2015, Sivabalan e Minu 2021], sendo essencial para garantir a qualidade dos dados a serem utilizados nas etapas seguintes.

Para assegurar a precisão e confiabilidade dos modelos, os dados foram divididos em duas partes: uma base de treino e uma base de teste [Fávero e Belfiore 2017, Dhawan 2023, IBM 2022]. A divisão foi realizada com o método *Data Split*, alocando 80% dos dados para o treinamento e 20% para o teste [Pires 2023, Rácz et al. 2021]. A base de treino foi usada para ajustar os modelos, enquanto a base de teste foi empregada para avaliar seu desempenho na previsão dos componentes curriculares [Escovedo e Koshiyama 2020]. Na sequência, foram desenvolvidos modelos supervisionados de classificação utilizando algoritmos como *Decision Tree*, *Random Forest*, *Naive Bayes*, KNN e *Logistic Regression*. A implementação dos modelos foi realizada em Python, com o auxílio das bibliotecas *pandas*, *seaborn* e *matplotlib*, além das plataformas *Google Colaboratory* e *Jupyter Lab*. O treinamento dos modelos ocorreu em um *Sony Vaio* com processador Intel Core i5-3337U e 12 GB de RAM.

4. Resultados e avaliação

Nesta seção, são apresentados os resultados obtidos com os algoritmos, utilizando o método de divisão de dados (*data split*). Cada algoritmo foi treinado e posteriormente validado com a base de teste, e as métricas de desempenho foram avaliadas. As métricas adotadas incluíram acurácia, *recall*, *F1-score* e precisão. A acurácia mede a proporção de previsões corretas em relação ao total de previsões, mas pode ser enganosa em conjuntos de dados desbalanceados [Souza 2019]. Por isso, também foram utilizados *recall*, que reflete a proporção de instâncias relevantes corretamente identificadas [Rodrigues 2019], e *F1-score*, que combina precisão e *recall* em uma única métrica, variando de 0 a 1 [Souza 2019]. O *Random Forest* e o KNN apresentaram o melhor desempenho e foram avaliados adicionalmente com a *matriz de confusão* e a *curva ROC*, incluindo a área sob a curva (AUC[1]) [Silva 2020].

Inicialmente, foram considerados os 33 componentes curriculares presentes na

[1] *area under the ROC curve*

matriz curricular de 2023. Contudo, muitos componentes não foram identificados corretamente devido à grande quantidade e diversidade de dados, tornando as disciplinas menos frequentes (classes minoritárias) mais difíceis de prever. Para melhorar a previsão, os componentes foram agrupados em quatro núcleos com base na matriz curricular do curso de Engenharia de Computação (ECO). Esses núcleos agrupam componentes curriculares similares definidos como: (i) Núcleo Exatas e Comunicação; (ii) Núcleo Computação; (iii) Núcleo Eletrônica; e (iv) Empreendedorismo, Inovação e Desenvolvimento Pessoal, sendo este último simplificado como "Desenvolvimento Pessoal."

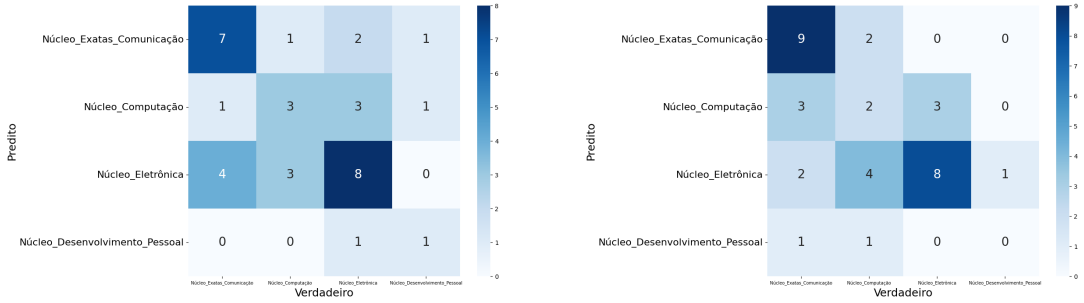
Para o algoritmo *Decision Tree*, foi utilizado o método *DecisionTreeClassifier* da biblioteca *scikit-learn*, que cria um modelo de árvore de decisão que divide os dados de entrada em subconjuntos com base em regras de decisão para inferir a classe dos componentes curriculares. O modelo obteve uma acurácia de 44,44%, precisão de 42,21%, *recall* de 44,44% e um *F1-score* de 42,41%. O modelo de regressão logística, utilizando o método *LogisticRegression*, alcançou uma acurácia de 50,00%, precisão de 47,81%, *recall* de 50,00% e um *F1-score* de 45,00%. Apesar de os modelos terem mostrado capacidade de previsão dos componentes curriculares, os algoritmos *Random Forest* e KNN apresentaram um melhor desempenho no processo de previsão.

Por outro lado, os algoritmos *Naive Bayes* e *SVM* obtiveram as piores avaliações. O *Naive Bayes* apresentou uma acurácia de 33,33%, precisão de 51,94%, *recall* de 33,33% e *F1-score* de 37,36%. Da mesma forma, o *SVM*, utilizando o método *SVC (Support Vector Classifier)*, alcançou uma acurácia de 44,44%, precisão de 38,05%, *recall* de 44,44% e *F1-score* de 35,35%. Ambos apresentaram um desempenho insatisfatório em comparação com os demais algoritmos, evidenciando limitações na classificação correta dos componentes curriculares.

Por fim, os melhores modelos foram os gerados pelo *Random Forest* e KNN. O *Random Forest*, utilizando o método *RandomForestClassifier* da biblioteca *scikit-learn*, cria várias árvores de decisão e combina suas previsões por meio de votação para determinar a classe final prevista [Müller 2019]. O modelo obteve uma acurácia de 52,78%, precisão de 52,78%, *recall* de 52,78% e um *F1-score* de 52,78%. Similarmente, o KNN, utilizando o método *KNeighborsClassifier*, baseia-se na proximidade para classificar os componentes curriculares, alcançando uma acurácia de 52,78%, precisão de 53,57%, *recall* de 52,78% e um *F1-score* de 52,02%. Esses resultados indicam que o KNN tem um desempenho semelhante ao do *Random Forest*, tornando-os candidatos fortes no processo de classificação utilizando o método *data split*. A Figura 1 compila os resultados de cada algoritmo, apresentando as métricas na ordem: acurácia, precisão, *recall* e *F1-score*.

	Model	Accuracy	Precision	Recall	F1 Score
0	KNN	52.78	53.57	52.78	52.02
1	Decision Tree	44.44	42.21	44.44	42.41
2	Naive Bayes	33.33	51.94	33.33	37.36
3	Logistic Regression	50.00	47.81	50.00	45.00
4	Random Forest	52.78	52.78	52.78	52.78
5	SVM	44.44	38.05	44.44	35.35

Figura 1. Comparação das métricas feitas por modelos
Fonte: Elaborado pela autora



(a) Matriz de confusão do *Random Forest* **(b) Matriz de confusão do *KNN***

Figura 2. Matriz de confusão dos algoritmos *Random Forest* e *KNN*
Fonte: Elaborado pela autora.

4.1. Matriz de confusão dos modelos *Random Forest* e KNN

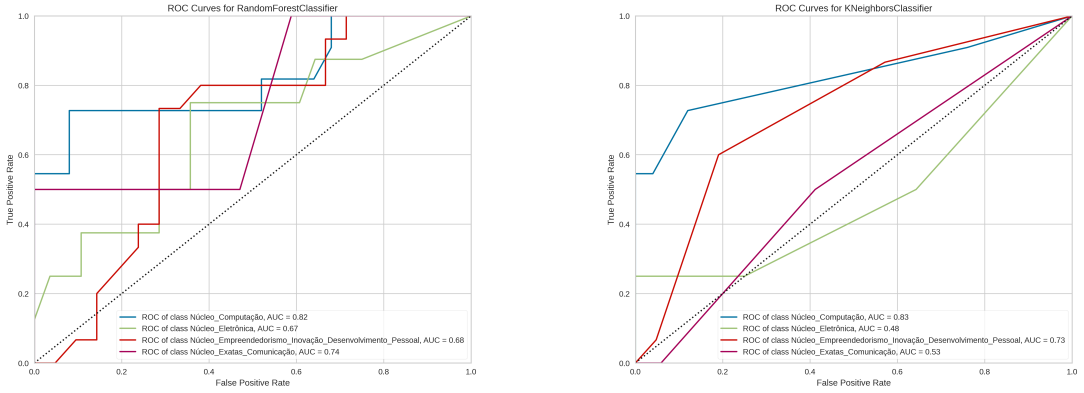
A matriz de confusão, apresentada na Figura 2a, ilustra o desempenho dos modelos de classificação. A diagonal principal da matriz representa as previsões corretas para cada componente curricular, sendo essencial para a avaliação da precisão do modelo. As células fora da diagonal principal indicam previsões incorretas: as células abaixo da diagonal principal correspondem a falsos negativos, onde o modelo falha em classificar corretamente a classe de destino; por outro lado, as células acima da diagonal principal representam falsos positivos, onde o modelo incorretamente classifica componentes como pertencentes a uma classe diferente da real. Esta análise detalhada permite identificar e entender as áreas de falha do modelo, facilitando a melhoria contínua na precisão das previsões.

Na Figura 2a, a análise da primeira coluna, correspondente ao ”Núcleo Exatas e Comunicação”, revela que dos 11 componentes avaliados, o modelo conseguiu classificar corretamente sete. No entanto, foram observadas classificações incorretas: um componente foi erroneamente classificado como ”Núcleo Computação”e quatro como ”Núcleo Eletrônica”. Além disso, alguns componentes do ”Núcleo Exatas e Comunicação”foram incorretamente atribuídos ao ”Núcleo Computação”.

Em contraste, a Figura 2b ilustra a matriz de confusão para o modelo KNN. Para o ”Núcleo Exatas e Comunicação”, o modelo acertou nove previsões, mas cometeu erros significativos ao classificar componentes deste núcleo como pertencentes ao ”Núcleo Computação”. No ”Núcleo Computação”, o KNN fez apenas duas previsões corretas, enquanto duas classificações incorretas foram atribuídas ao ”Núcleo Eletrônica”. O ”Núcleo Eletrônica”obteve oito previsões corretas, mas um componente foi erroneamente classificado como ”Núcleo Computação”. Finalmente, o ”Núcleo Desenvolvimento Pessoal”não teve previsões corretas, com todos os componentes incorretamente classificados como ”Núcleo Eletrônica”.

4.2. Curva ROC dos modelos *Random Forest* e KNN

A Figura 3a ilustra o desempenho do modelo *Random Forest* para os quatro núcleos curriculares analisados. O modelo apresentou uma precisão de 67% para o ”Núcleo Eletrônica”e 68% para o ”Núcleo Empreendedorismo, Inovação e Desenvolvimento Pessoal”, com margens de erro de 33% e 32%, respectivamente. Em contraste, os núcleos ”Núcleo Computação”e ”Núcleo Exatas e Comunicação”demonstraram um desempenho



(a) Curva *ROC* do *Random Forest* **(b) Matriz de confusão do *KNN***

Figura 3. Curva *ROC* dos algoritmos *Random Forest* e *KNN*
Fonte: Elaborado pela autora.

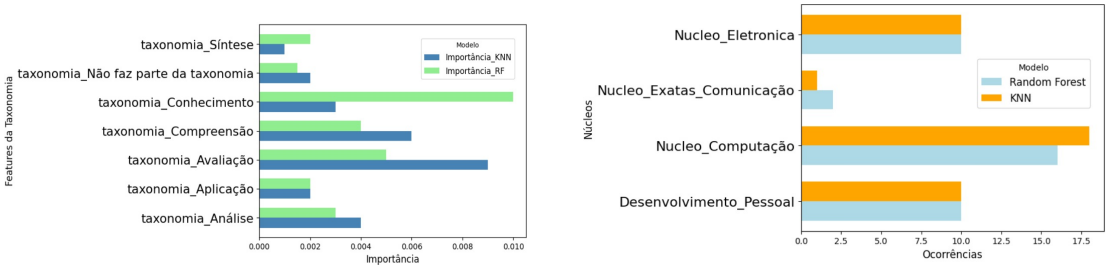
superior. Entretanto, a análise da matriz de confusão identificou erros de classificação, possivelmente atribuídos ao desequilíbrio dos dados, com menor número de componentes curriculares em algumas classes. Embora a Área sob a Curva (AUC) sugira um desempenho geral satisfatório do modelo, a classificação de determinadas classes ainda enfrenta desafios, indicando áreas para aprimoramento na precisão das previsões.

A Figura 3b ilustra o desempenho do modelo KNN para os quatro núcleos de disciplinas. A classificação das componentes do ”Núcleo Eletrônica”revelou-se insatisfatória, com uma precisão de 48,0% e uma margem de erro de 52%, indicando dificuldades significativas na correta distinção dos componentes desse núcleo. Em contraste, os núcleos ”Núcleo Computação”, ”Núcleo Desenvolvimento Pessoal”e ”Núcleo Exatas e Comunicação”apresentaram um desempenho superior. O ”Núcleo Computação”obteve uma precisão de 83%, enquanto o ”Núcleo Desenvolvimento Pessoal”alcançou 73%, com margens de erro de 17% e 27%, respectivamente. Esses resultados sugerem que o modelo KNN foi mais eficaz na classificação das componentes desses núcleos, evidenciando uma maior capacidade de diferenciação e precisão nas previsões.

4.3. Análise dos resultados com Random Forest e KNN

Nesta seção, procurou-se responder à questão de pesquisa: ”Como classificar as questões do ENADE conforme os componentes curriculares, utilizando técnicas de mineração de texto?”. Os resultados obtidos com os algoritmos *Random Forest* e KNN são ilustrados nas figuras 4a e 4b.

A Figura 4a revela que, ao longo dos últimos cinco anos, o ENADE para o curso de Engenharia da Computação tem focado predominantemente nos níveis de compreensão, conhecimento e aplicação da Taxonomia de *Bloom*, os quais são os mais frequentes nas questões. Ao comparar os modelos *Random Forest* e KNN, observa-se uma leve diferença: o *Random Forest* apresenta uma maior incidência do nível de aplicação em comparação ao KNN, embora a variação seja pequena. Essa discrepância pode ser atribuída ao desempenho ligeiramente superior do *Random Forest* nas previsões. Apesar dessas pequenas diferenças, ambos os modelos se mostram adequados para a previsão de componentes curriculares. Além disso, a Figura 4a indica que o modelo KNN não registrou ocorrências no nível de análise da Taxonomia de *Bloom*, enquanto o *Random Forest*



(a) Comparação dos resultados da Taxonomia de Bloom no *Random Forest* e no *KNN*

(b) Comparação dos resultados dos núcleos das componentes no *Random Forest* e no *KNN*

Figura 4. Comparação dos resultados da taxonomia de bloom e dos núcleos das componentes nos algoritmos *Random Forest* e do *KNN*
Fonte: Elaborado pela autora.

apresentou uma pequena ocorrência desse nível. Essas diferenças podem ser atribuídas ao fato de o *Random Forest* ter um F1-score ligeiramente maior ao do KNN, refletindo uma melhor capacidade de avaliação correta das componentes curriculares em comparação com o KNN.

A Figura 4b revela que os núcleos ”Computação”e ”Desenvolvimento Pessoal”foram mais frequentemente classificados pelos algoritmos, sugerindo que essas categorias predominaram nas provas do ENADE para o curso de Engenharia da Computação nos últimos cinco anos. Em comparação, os núcleos ”Eletrônica”e ”Exatas e Comunicação”apresentaram menor ocorrência. Essa observação sugere que, nos últimos cinco anos, o ENADE focou mais nos componentes dos núcleos ”Computação”e ”Desenvolvimento Pessoal”, avaliando habilidades da taxonomia de bloom relacionadas à compreensão, conhecimento, síntese e aplicação. Ambos os modelos, *Random Forest* e KNN, demonstraram um desempenho semelhante, sendo considerados adequados para a previsão de componentes curriculares utilizando o método *data split*.

5. Conclusão

Neste trabalho, foi realizada uma análise classificatória utilizando modelos de *machine learning* para prever os componentes curriculares aplicados nas provas do ENADE para Engenharia da Computação. Seis algoritmos foram avaliados: *Decision Tree*, *Random Forest*, *Naive Bayes*, KNN, *Logistic Regression* e SVM, com *Random Forest* e KNN se destacando pelos melhores resultados usando o método *data split*.

O estudo seguiu as cinco primeiras etapas do *framework* CRISP-DM e revelou que os núcleos ”Núcleo Computação”e ”Núcleo Desenvolvimento Pessoal”foram mais frequentes nas provas do ENADE, enquanto os níveis da Taxonomia de *Bloom* mais utilizados foram Conhecimento, Compreensão e Aplicação.

As limitações do trabalho incluem a análise restrita a um subconjunto dos anos do ENADE e a utilização do método *data split*, que pode introduzir variações. Além disso, uma análise mais profunda dos dados textuais das questões poderia oferecer *insights* mais detalhados. Por outro lado, os trabalhos futuros deverão incorporar o método *stratified k-fold cross validation* e o algoritmo *BERT* para melhorar a precisão da classificação dos componentes curriculares e a análise dos dados textuais, visando resultados mais consistentes e generalizáveis.

Referências

Abdusalomovna, T. D. et al. (2023). Text mining. *European Journal of Interdisciplinary Research and Development*, 13:284–289.

Algarni, A. (2016). Data mining in education. *International Journal of Advanced Computer Science and Applications*, 7(6).

Alvarenga, A. M., Tauchen, G., and Alvarenga, B. T. (2017). A interdisciplinaridade nos componentes curriculares de cursos de licenciatura da área de ciências exatas e da terra. *Revista Thema*, v. 14(n. 3):151–166.

Araujo, L. R. D. (2021). Classificação automática de questões de provas: análise comparativa de algoritmos e aplicação ao enade. Trabalho de conclusão de curso, Universidade Federal de Santa Maria, Santa Maria, Brasil.

BRASIL (2020). Exame nacional de desempenho dos estudantes (Enade). Disponível em: `https://www.gov.br/inep/pt-br/acesso-a-informacao/perguntas-frequentes/exame-nacional-de-desempenho-dos-estudantes-enade`. [Online] Acesso em 26 de março de 2023.

Charão, A. S., Wiechork, K., Rodrigues, M. L., and Barbosa, F. P. (2020). Explorando resultados por questao no enade em ciência da computaçao para subsidiar revisao de projeto pedagógico de curso. In *Anais do XXVIII Workshop sobre Educaçao em Computaçao*, Santa Maria.

Dhawan, S. (2023). Divisão de dados para modelos de aprendizado de máquinageeksforgeeks. GeeksforGeeks. [Online] Acesso em 22 de outubro de 2023.

Escovedo, T. and Koshiyama, A. (2020). *Introdução a Data Science: Algoritmos de Machine Learning e métodos de análise*. Casa do Código.

Fávero, L. P. and Belfiore, P. (2017). *Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®*. Elsevier Brasil.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37–37.

FDV (2021). Diretoria de avaliação da educação superior (daes). Technical report, FDV - Faculdade de Viçosa, SINAES,Viçosa.

Gallagher, S. R. (2016). *The future of university credentials: New developments at the intersection of higher education and hiring*. Harvard Education Press.

Gerab, F., Bueno, I. A. M., and da Silva Gerab, I. F. (2014). Análise das interações curriculares em um curso de ciência da computação: buscando subsídios para aprimoramento curricular. *Revista Brasileira de Informática na Educação*, 22(01):30.

IBM (2021). CRISP-DM Help Overview. Disponível em: `https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=dm-crisp-help-overview`. Acesso em: 10 de março de 2022.

IBM (2022). Transformação de dadosdocumentação da ibm. IBM. [Online] Acesso em 22 de outubro de 2023.

Lira Silva, L. G., Barros, A. N., and Falcão, T. P. (2024). O impacto da nova matriz curricular da licenciatura em computação no desempenho dos discentes. In *Anais do IV Simpósio Brasileiro de Educação em Computação*, pages 256–265. SBC.

Lopes, F. L. and Vendramini, C. M. M. (2015). Propriedades psicométricas das provas de pedagogia do enade via tri. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 20(1):27–47.

Müller, R. A. (2019). Aplicação de aprendizado de máquina para identificar o meio de transporte baseado em localizações de gps.

NETO, A. P. D. L. (2023). Estudo comparativo entre modelos baseados em bert na classificação estática de malware.

Neto, W. R. C. (2021). O uso de mineração de dados educacionais sob o enade como apoio ao processo de tomada de decisão de gestores do ensino superior. Trabalho de conclusão de curso, Universidade Federal do Ceará, Quixadá, Brasil.

Oliveira, A. P. S. B., de Aguiar Pontes, J. N., and Marques, M. A. (2016). O uso da taxionomia de bloom no contexto da avaliação por competência. *Revista Pleiade*, 10(20):12–22.

Oliveira, T. et al. (2024). Educação e ensino no exame nacional de desempenho dos estudantes (enade) de licenciatura em ciências biológicas-2014, 2017 e 2021.

Olivera, S. (2011). Taxonomia de bloom. *Universidad Cesar Vallejo*, 4.

Pérez, J., Iturbide, E., Olivares, V., Hidalgo, M., Almanza, N., and Martínez, A. (2015). A data preparation methodology in data mining applied to mortality population databases. *New Contributions in Information Systems and Technologies: Volume 1*, pages 1173–1182.

Pires, L. C. (2023). Modelo de propensão: Como identificar os clientes com maior chance de compra?

Pissaia, L. e. a. E. (2018). Exame nacional de desempenho dos estudantes – enade e o desenvolvimento de competências no ensino superior. In *Research, Society and Development*, volume 7. Acesso em 26 de março de 2023.

Rácz, A., Bajusz, D., and Héberger, K. (2021). Effect of dataset size and train/test split ratios in qsar/qspr multiclass classification. *Molecules*, 26(4):1111.

Ramos, J. L. C., Rodrigues, R. L., Silva, J. C. S., and de Oliveira, P. L. S. (2020). Crisp-edm: uma proposta de adaptação do modelo crisp-dm para mineração de dados educacionais. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, pages 1092–1101. SBC.

Rodrigues, V. (2019). Métricas de Avaliação: acurácia, precisão, recall... quais as diferenças? `https://vitorborbarodrigues.medium.com/m%C3%A9tricas-de-avalia%C3%A7%C3%A3o-acur%C3%A1cia-precis%C3%A3o-recall-quais-as-diferen%C3%A7as-c8f05e0a513c`.

Saltz, J. and Hotz, N. (2022). Data science project management. `https://www.datascience-pm.com/crisp-dm-2/`. [Online, Acesso em 26 de março de 2023].

Santos, C. A. P. D. (2022). Uma análise exploratória da influência dos projetos pedagógicos dos cursos superiores no resultado do enade por meio de mineraçãoo de textos e aprendizado de máquina. Master's thesis, CAPES, Campo Grande, Brasil.

Silva, A. R. d. (2020). Uma visão geral sobre machine learningclassificaçãostatplace a estatística ao alcance de todos. cursos e consultoria. [Online] Acesso em 20 de outubro de 2023.

Sivabalan, S. and Minu, R. (2021). Heterogeneous data integration with elt and analytical mpp database for data analysis application. In *2021 Innovations in Power and Advanced Computing Technologies (i-PACT)*, pages 1–5. IEEE.

Sousa, B. P. B. d. and Sousa, J. V. d. (2012). Resultados do enade na gestão acadêmica de cursos de licenciaturas: um caso em estudo. *SOU.*, v. 23(n. 52):232–253.

Souza, E. G. d. (2019). Entendendo o que É matriz de confusão com python. [Online] Acesso em 22 de outubro de 2023.

Tambling, P. (2020). Can education and skills development be more aligned locally reflecting local work patterns, business growth Disponível em: `https://abrir.link/PwfgE`. [Online] Acesso em 26 de março de 2023.

Tavares, L. A., Meira, M. C., and do Amaral, S. F. (2020). Inteligência artificial na educação: Survey. *Brazilian Journal of Development*, 6(7):48699–48714.

Ventura, M. M. (2007). O estudo de caso como modalidade de pesquisa. *Revista SoCERJ*, 20(5):383–386.

Zhang, X. and Liu, C.-A. (2023). Model averaging prediction by k-fold cross-validation. *Journal of Econometrics*, 235(1):280–301.