

Modelo preditivo do desempenho acadêmico: um estudo para a disciplina de cálculo diferencial e integral para Engenharias

Predictive model of academic performance for the discipline of Differential and Integral Calculus for Engineering

Engenharia de Computação

Diogo Martins Gonçalves de Moraes (pro7113@cefsa.edu.br)
Doutor em Administração pela Universidade Municipal de São Caetano do Sul e professor da Faculdade Engenheiro Salvador Arena.

Heloisa Rades de Sousa (061210035@faculdade.cefsa.edu.br)
Graduanda em Engenharia de Controle e Automação pela Faculdade Engenheiro Salvador Arena.

Vinícius Benevides da Silva (081210030@faculdade.cefsa.edu.br)
Graduando em Engenharia de Computação pela Faculdade Engenheiro Salvador Arena.

FTT Journal of Engineering and Business

- SÃO BERNARDO DO CAMPO, SP DEZ. 2022
- ISSN 2525-8729

Submissão: 01 jun. 2022 Aceitação: 14 dez.2022

Sistema de avaliação: às cegas dupla (double blind review)

FACULDADE ENGENHEIRO SALVADOR ARENA, p. 47-68

FTT JOURNAL
of Engineering and Business



Resumo

O presente estudo propõe um sistema computacional (modelo preditivo) capaz de prever o desempenho escolar dos estudantes das áreas de engenharia nas disciplinas de Cálculo, a partir da análise do desempenho escolar obtido em anos anteriores pelos estudantes na educação básica, além de incluir o perfil socioeconômico desses estudantes. Para tanto, foi utilizada uma abordagem quantitativa em nível explicativo, pois o estudo investigou o grau de associação existente entre o perfil dos estudantes ingressantes de dois cursos de engenharia distintos e o rendimento constatado em suas notas do primeiro bimestre de 2022; ademais, com o uso do Algoritmo Genético, escrito em linguagem Python, foi elaborado um modelo preditivo desse rendimento. Além de oferecer uma ferramenta estratégica para os gestores educacionais, no que se refere à elaboração de propostas para o nivelamento dos estudantes ingressantes, o resultado desta pesquisa pretende auxiliar o corpo docente no acompanhamento pedagógico de suas turmas. Concluiu-se nesse estudo que o modelo preditivo é viável e conseguiu prever com sucesso o desempenho dos estudantes a partir de 7 variáveis explicativas, com taxa de acerto de aproximadamente 96%, constatando que o algoritmo genético se mostrou adequado para este tipo de aplicação. Assim sendo, este estudo confere um percurso metodológico adequado para a proposição futura de modelos preditivos do desempenho dos estudantes em outras disciplinas.

Palavras-chave: Algoritmo Genético; Cálculo Diferencial e Integral. Desempenho Escolar. Modelo Preditivo.

Abstract

This research project proposes the development of a computational system (predictive model) capable of obtaining the performance of Engineering students in Calculus, based on data obtained from school performance in previous years by the students during basic schooling, besides the socioeconomic profile from these students. To achieve this goal,, it was used a quantitative approach at an explanatory level, since the study investigated the existing degree of association between the profile from incoming students of two diferente Engineering courses and the educational attainment found in their grades during the first two months of 2022, and with the use of Genetic Algorithm, written in Python language, it was elaborated a predictive model of this performance. In addition to offering a strategic tool for educational management, regarding the elaboration of a proposal to levelling incoming students, the result of this research aims to assist the college in pedagogical monitoring of their classes. This study concluded that the predictive model is viable and it could successfully predict the performance of the students, based on 7 explanatory variables, with a success rate of 96%, showing that the genetic algorithm was adequate for this type of application. Furthermore, this study provides an adequate methodological approach for the future proposition of predictive models of students performance in other disciplines.

Keywords: Genetic Algorithm; Calculus; School Performance; Predictive Model.

Introdução

Muitos cursos do Ensino Superior, principalmente aqueles voltados à área de Ciências Exatas, possuem a disciplina de Cálculo Diferencial e Integral em sua grade curricular. Sua compreensão é de suma importância para a continuidade do curso, visto que é uma disciplina básica e seus conceitos estão presentes em diversas outras disciplinas e aplicações, não só na área acadêmica.

Apesar de sua relevância, observa-se que há altos índices de reprovação dos estudantes nessa disciplina, fazendo com que as relações de ensino-aprendizagem do Cálculo Diferencial e Integral tenham sido alvo de diversas pesquisas que buscam quantificar e especificar as causas de tal realidade (ROSA et al., 2019; SENA, SOUZA e SOUZA, 2014; WISLAND, FREITAS e ISHIDA, 2014; ZARPELON, RESENDE e REIS, 2017).

Os resultados de Rosa et al. (2019) obtidos pela análise dos desempenhos acadêmicos do primeiro semestre de 2010 ao segundo semestre de 2016 dos alunos na Universidade Federal de Goiás demonstraram que o percentual de reprovação em Cálculo Diferencial e Integral foi de 65%. Por sua vez, Sena, Souza e Souza (2014) apontam que na Universidade Federal de Alagoas (UFAL), nos períodos de 2010.1 a 2012.1, o percentual de desistência (trancamento) e de reprovação dos alunos que se matricularam na disciplina de CDI foi em média de 52%. Além desses, outros autores, tais como Wisland, Freitas e Ishida (2014) e Zarpelon, Resende e Reis (2017) também indicaram que o índice de reprovação nessa disciplina é elevado.

Dos dados apresentados emerge a questão central desse estudo, que é a possibilidade de se reduzir a reprovação dos estudantes de Engenharias na disciplina de Cálculo Diferencial e Integral, por meio de ferramentas da inteligência artificial que ofereçam um cenário de possibilidades palpável para os gestores educacionais e professores de Cálculo, antes mesmo que esses estudantes sejam submetidos às avaliações cotidianas da disciplina.

Nesse contexto, o estudo justifica-se por apresentar a proposta de um modelo preditivo do desempenho acadêmico para a disciplina de Cálculo Diferencial e Integral I no âmbito dos cursos de Engenharia, a partir de dados sobre o desempenho escolar obtidos anteriormente pelos estudantes na educação básica, além do perfil socioeconômico desses estudantes.

Diante disso, a proposta destina-se ao desenvolvimento de um algoritmo genético, produzido em linguagem Python e que, a partir de variáveis qualitativas e quantitativas, busca analisar e

antecipar a provável situação final do aluno na disciplina. Dessa forma, esse sistema proveria informações relevantes para a efetiva tomada de decisão acadêmica, tanto por parte do corpo discente quanto do corpo docente e administrativo.

Para a concretização deste estudo, foram realizados os seguintes objetivos específicos:

- a) Pesquisa na literatura científica para determinar quais seriam os fatores de influência do desempenho acadêmico dos estudantes, ou seja, que passariam a ser as variáveis válidas para serem incorporadas como parâmetros de predição do desempenho acadêmico;
- b) Estudar a codificação e a aplicabilidade do algoritmo genético;
- c) Obter testagem dos parâmetros influenciadores estudados.

A apresentação do estudo está dividida em cinco partes: a primeira delas introdutória, em que são expostos o contexto e o objetivo do estudo; a seção 2, que apresenta o referencial teórico e os resultados dos principais estudos já realizados; a seção 3, que indica os procedimentos metodológicos utilizados; a apresentação e a análise dos resultados aparecem na seção 4; e a seção 5, contendo as considerações finais.

Referencial teórico

A fim de satisfazer o objetivo geral deste estudo, que é o desenvolvimento de um modelo preditivo do desempenho acadêmico para a disciplina de Cálculo Diferencial e Integral I das Engenharias, com o intuito de auxiliar a tomada de decisão em gestão educacional, buscou-se identificar na literatura científica o embasamento teórico necessário para ser aplicado nas variáveis que serão utilizadas no projeto.

Os estudos de Araújo et. al (2013) visavam identificar amostras do desempenho acadêmico dos alunos do curso de Ciências Contábeis de uma IES privada em Belo Horizonte. Constatou-se que o desempenho é explicado pelas variáveis de idade, gênero, frequência às aulas, natureza da disciplina, tipo de disciplina, período letivo, existência ou não de um campus, situação final da disciplina e período letivo, levando em conta o dimensionamento da nota final das disciplinas como base. Dessa maneira, foi realizada uma pesquisa descritiva e quantitativa dos dados disponíveis dos anos de 2001 até 2009, utilizando o método de regressão *cross-section*. Finalmente, após a análise de cada amostra, concluiu-se que algumas variáveis são

correlacionadas e que seis das variáveis analisadas podem ser utilizadas para a explicação do desempenho dos alunos, sendo que a frequência às aulas, idade, gênero e tipo de disciplina tiveram uma presença maior de influência.

Os estudos de Teles et. al (2019), por sua vez, tiveram como objetivo a análise dos fatores de principal relevância no desempenho acadêmico final dos estudantes da Universidade Federal de Santa Catarina. Para tal, as variáveis selecionadas foram: a nota do vestibular de ingresso à universidade, fatores econômicos (bolsa de estudos e trabalho), idade e gênero, habilidades acadêmicas, estudo prévio e fatores institucionais (corpo docente e horários das aulas). O método de investigação escolhido foi o de análise e correlação de variáveis por procedimentos estatísticos descritivos, testes inferenciais (ex.: correlação linear de Pearson), dentre outros testes lineares (matrizes) e modelos estatísticos. Assim, percebeu-se o maior desempenho em variáveis como gênero e idade, sendo que não somente a nota de ingresso é um fator relevante como diz o material de base do estudo.

As análises de Rocha, Leles e Queiroz (2019) tratam dos aspectos interpostos entre o desempenho estudantil de Nutrição em relação aos principais testes relacionados às universidades, como o Exame Nacional de Desempenho de Estudantes (Enade). Para tanto, verificou-se a performance discente em relação aos fatores socioeconômicos, trajetória acadêmica e perfil da instituição. Como metodologia, após a coleta dos dados do Enade no portal do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), do Ministério da Educação (MEC), aplicou-se uma análise descritiva dos dados baseados em Estatística Descritiva. Por fim, chegou-se à conclusão de que a classe social tem grande relação com o desempenho estudantil universitário em exames como o Enade.

Os estudos de Nogueira (2013) referem-se à identificação dos fatores de estilo de aprendizagem, número de faltas, idade e gênero e suas relações com o desempenho acadêmico dos discentes de Ciências Contábeis. Por meio de uma coleta de dados obtidos por questionários, aplicados em uma população de 279 alunos, e pelo processamento das informações recolhidas através de gráficos Box plot e análises de outliers, foram postos em prática conceitos estatísticos analíticos como base de pesquisa. Concluiu-se que, de todas as variáveis analisadas, somente o número de faltas foi significativo em relação ao desempenho acadêmico.

Os estudos de Souto-Maior, Borba e Knupp (2015) tratam de fatores pertinentes ao desempenho de estudantes de Administração e Contabilidade. Metodologicamente, foram selecionadas 556 notas finais de universitários da graduação e levantaram-se hipóteses sobre a correlação dos elementos influentes, posteriormente testadas pelo teste *t bicaudal*, que acabou por confirmar a maioria delas, como a significância do gênero, a classificação no vestibular, o horário do curso (maior destaque para aqueles que estudam no período matutino) e a frequência às aulas.

Fagundes, Luce e Rodriguez Espinar (2014) propuseram um modelo avaliativo do desempenho acadêmico dos alunos que passam da Educação Básica (Ensino Médio) para o Ensino Superior (Universidades) a partir de um tratamento teórico, focado no êxito durante essa transição. Para tanto, realizou-se um estudo preditivo e desenvolvimento a longo prazo por meio da pesquisa *ex post facto*, da regressão múltipla *stepwise* e de abordagens quantitativas validadas nos cursos de licenciatura da Universidade Federal do Rio Grande do Sul (UFRGS). Dessa maneira, observou-se que os itens formação na educação básica (maior influência), características sociodemográficas e antecedentes socioculturais constituem fatores de influência do desempenho acadêmico dos alunos nesse contexto.

Amaro e Beuren (2018) averiguaram a influência de fatores contingenciais no desempenho acadêmico de Ciências Contábeis. Adotou-se nesse estudo uma abordagem quantitativa da pesquisa descritiva realizada para encontrar os fatores de destaque da amostra disponível de 295 contribuintes, aplicando o teste de Kruskal-Wallis sobre ela. Assim, foram considerados o nível de instrução dos pais, as horas de estudo semanal extraclasse e a experiência profissional como fatores influentes e, que no final do estudo, tiveram alta relevância em relação à conduta acadêmica dos discentes.

Os estudos de Magalhães, Andrade (2006) tiveram por objetivo analisar, por meio de regressão múltipla e correlação linear de Pearson, a importância de alguns fatores na explicação da variação do desempenho acadêmico na disciplina Contabilidade Introdutória e no primeiro semestre do curso de graduação em Ciências Contábeis. As pesquisas apontaram que os resultados obtidos na primeira fase do vestibular não auxiliam na previsão do desempenho do aluno. Por outro lado, as variáveis idade, gênero e os resultados na segunda fase do vestibular tiveram influência no rendimento dos estudantes.

Martins et al (2015) buscaram enumerar as principais variáveis preditoras para o sucesso acadêmico. Para tal, utilizou-se uma amostra de 1318 alunos dos subsistemas politécnico e universitário, em Portugal, e recorreu-se a uma regressão hierárquica (3 modelos). Os resultados evidenciam um peso moderado (47%) do modelo mais integrador, destacando-se variáveis do sujeito e institucionais. Os dados utilizados foram: gênero, idade, nível socioeconômico, nota de candidatura, provas de ingresso, classificações obtidas no curso secundário, prioridade de escolha do curso, ano curricular (calouros e finalistas), participação em atividades extracurriculares, regularidade de estudo, adaptação ao ensino superior, área de estudos (Psicologia, Letras e Ciências neste caso) e raciocínio numérico e verbal.

Em sua pesquisa de natureza bibliográfica, Silva e Júnior (2016) buscaram mapear e sintetizar os fatores determinantes que estão relacionados ao desempenho acadêmico dos alunos no Ensino Superior, analisando seis produções acadêmicas. Dessa forma, foi concluído que variáveis como a escolaridade dos pais, o fato de trabalhar concomitantemente ao curso, a renda familiar, as expectativas acadêmicas e as habilidades sociais possuem grande relevância para o aproveitamento do estudante.

Complementando tais pesquisas, Martins e Marinho (2019) analisaram a relação entre as variáveis referentes às qualificações do corpo docente, do corpo discente e da IES em relação à variável desempenho acadêmico. Para tal, realizaram uma pesquisa de levantamento documental analisada com técnicas estatísticas univariadas, como o Alfa de Cronbach, a prova Kolmogorov-Smirnov para o teste de normalidade e os coeficientes de correlação ponto-bisserial e de Spearman. Como principal resultado, foi identificado que fatores como a estratégia de ensino, status socioeconômico, absenteísmo, desempenho escolar anterior, horas de estudo, motivação, ambiente de estudo e forma de ingresso são os que apresentam maior relação com a variável desempenho acadêmico.

Ressalta-se ainda o trabalho de Manhães e Cruz (2019), que tiveram por objetivo fornecer aos gestores educacionais das universidades uma abordagem que ofereça acompanhamento semestral sobre o desempenho acadêmico dos graduandos e a previsão dos que estão em risco de abandonar o sistema, apresentando a arquitetura WAVE baseada em Mineração de Dados Educacionais (EDM), que utilizaram dados reais de inúmeros cursos de graduação da Universidade Federal do Rio de Janeiro (UFRJ) durante um período de 16 anos. Concluiu-se que

o ano e o período em que o estudante ingressou na faculdade, o curso e a instituição em que ele está matriculado, o atual Rendimento Acadêmico Acumulado durante essa graduação (sendo o fator mais relevante a nota na disciplina de Cálculo Diferencial e Integral I), o número de disciplinas cursadas e o número de disciplinas já reprovadas por falta e/ou média do aluno foram de considerável importância no modelo.

O trabalho realizado por Abbas e Lopes (2020) investigou a influência dos fatores pessoais, institucionais e dos estilos de aprendizagem no desempenho acadêmico dos discentes de Ciências Contábeis nas disciplinas com o maior índice de reprovação. Para testar as hipóteses, foi utilizada a estatística descritiva e multivariada (regressão). Os resultados indicam que alguns dos fatores pessoais, tais como turno, escola do ensino médio, estado civil, experiência na área, horas de estudo extraclasse, faixa etária, jornada de trabalho, membros da família, moradia, participação econômica familiar e renda familiar apresentaram relação estatisticamente significativa com o desempenho.

Sousa, Ferreira e Miranda (2019), tendo como base a Teoria da Atribuição de Causalidade, tiveram como objetivo confrontar as percepções de professores e de estudantes de Ciências Contábeis sobre a relevância dos fatores determinantes do desempenho acadêmico. Os resultados evidenciaram que as variáveis consideradas mais importantes por alunos e professores foram a motivação, esforço pessoal, quantidade de horas de estudo, frequência às aulas, conhecimento prévio, formação pedagógica do corpo docente e a infraestrutura da IES.

Miranda et al (2015) buscaram identificar variáveis apresentadas na literatura que afetam o desempenho acadêmico no ensino superior na área de Negócios. Para tanto, foi realizada uma revisão de 52 artigos relativos ao tema. Como principais resultados, pode-se constatar que as variáveis relacionadas ao corpo discente são as que mais fortemente explicam o desempenho acadêmico (status socioeconômico, absenteísmo, desempenho escolar anterior, conhecimento prévio do conteúdo, área de especialização do discente, horas de estudo, motivação e tipo de aprendizagem). Em segundo lugar, estão as variáveis relacionadas ao corpo docente (estratégias ou método de ensino, experiência profissional do docente e credenciais profissionais). O quadro 1 apresenta a síntese da revisão da literatura.

Quadro 1 – Síntese da revisão da literatura

Autores e ano	Objetivo do estudo	Revista	Variáveis relevantes
ABBAS, K.; LOPES, A. K. (2020).	Impacto dos fatores pessoais, institucionais e estilos de aprendizagem no desempenho acadêmico: uma análise com estudantes de contabilidade	Revista Catarinense da Ciência Contábil, Florianópolis, SC, v. 19, 1-31, e3020, 2020	Turno, escola do ensino médio, estado civil, experiência na área, horas de estudo extraclasse, faixa etária, jornada de trabalho, membros da família, moradia, participação econômica familiar e renda familiar.
AMARO, H. D.; BEUREN, I. M. (2018)	Influência de Fatores Contingenciais no Desempenho Acadêmico de Discentes do Curso De Ciências Contábeis	Revista de Educação e Pesquisa em Contabilidade (REPeC), v. 12, n. 1, p. 21–42, 29 mar. 2018.	Fatores contingenciais externos, Nível de instrução do pai, Horas de estudo semanal extraclasse, Experiência profissional
ARAÚJO, E. A. T. et al. (2013)	Desempenho Acadêmico de Discentes do Curso de Ciências Contábeis: uma Análise dos seus Fatores Determinantes em uma IES Privada.	Contabilidade Vista & Revista, v. 24, n. 1, p. 60–83, 2013.	Idade, Gênero, Frequência, Tipo de disciplina (básica ou específica).
FAGUNDES, C. V.; LUCE, M. B.; RODRIGUEZ ESPINAR, S. (2014)	O desempenho acadêmico como indicador de qualidade da transição Ensino Médio-Educação Superior.	Ensaio: Avaliação e Políticas Públicas em Educação, v. 22, n. 84, p. 635–669, set. 2014.	Características sociodemográficas – gênero, idade e tipologia da instituição de origem; Antecedentes familiares e econômicos; Antecedentes escolares.
MAGALHÃES, F. A. C.; ANDRADE, J. X. (2006).	Exame vestibular, características demográficas e desempenho na universidade: em busca de fatores preditivos.	Congresso USP de Controladoria e Contabilidade. Volume 6.	Idade, Gênero, resultados na segunda fase do vestibular.
MANHÃES, L. M. B.; CRUZ, S. M. S. (2019).	Predição do desempenho acadêmico de alunos da graduação utilizando mineração de dados	XIX Simpósio de Pesquisa Operacional e Logística da Marinha. Rio de Janeiro, RJ, Brasil, 06 a 08 de novembro de 2019.	Período de ingresso na faculdade, o curso, a instituição, o atual Rendimento Acadêmico Acumulado, o número de disciplinas cursadas e o número de disciplinas já reprovadas por falta e/ou média do aluno.

Continua

MARTINS, E. C. et al. (2015).	Um estudo dos preditores de sucesso acadêmico no ensino superior	Revista de Psicología y Educación, 2015, 10(1),77-100.	Gênero, idade, nível socioeconômico, nota de candidatura, provas de ingresso, classificações do secundário, prioridade de escolha do curso, ano curricular, participação em atividades extracurriculares, regularidade de estudo, adaptação ao ensino superior, área de estudos e raciocínio numérico e verbal.
MARTINS, Z. B.; MARINHO, S. V. (2019).	Relação das variáveis concernentes ao desempenho acadêmico: um estudo com alunos de graduação em ciências contábeis	Revista Universo Contábil, FURB, Blumenau, v. 15, n. 1, p. 27-48, jan./mar., 2019.	Estratégia de ensino, status socioeconômico, absenteísmo, desempenho escolar anterior, horas de estudo, motivação, ambiente de estudo e forma de ingresso.
MIRANDA, G. J. et al. (2015)	Determinantes do Desempenho Acadêmico na Área de Negócios	Meta: Avaliação Rio de Janeiro, v. 7, n. 20, p. 175-209, maio/ago. 2015	Status socioeconômico, absenteísmo, desempenho escolar anterior, conhecimento prévio do conteúdo, área de especialização do discente, horas de estudo, motivação, tipo de aprendizagem, estratégia ou método de ensino, experiência profissional do docente e credenciais profissionais.
NOGUEIRA, D. R. et al. (2013)	Fatores que impactam o desempenho acadêmico: Uma análise com discentes do Curso de Ciências Contábeis no Ensino Presencial.	Revista de Informação Contábil, v. 7, n. 3, p. 51-62, 2013.	Número de faltas.
ROCHA, A.; LELES, C.; QUEIROZ, M. (2019)	Fatores associados ao desempenho acadêmico de estudantes de Nutrição no Enade.	Revista Brasileira de Estudos Pedagógicos, v. 99, n. 251, 18 jun. 2019.	Fatores Socioeconômicos – gênero, idade, renda familiar, grau de escolaridade dos pais.

Continua

SILVA, H. G.; JÚNIOR, A. P. O. (2016).	Fatores determinantes do desempenho acadêmico no ensino superior: estado da arte	Revista Plurais – Virtual, Anápolis - Go, vol.6, n. 2 – jul./dez. 2016.	Escolaridade dos pais, trabalhar concomitantemente ao curso, a renda familiar, as expectativas acadêmicas e as habilidades sociais.
SOUTO-MAIOR, C.; BORBA, J. A.; KNUPP, P. (2015)	Como Melhorar a Aprendizagem em Pesquisa Operacional? Uma Análise do Desempenho de Estudantes de Administração de Empresas e de Contabilidade.	Sociedade, Contabilidade e Gestão, v. 9, n. 3, 4 maio 2015.	Gênero, Nota no vestibular, Período do curso (matutino/noturno), Ausência nas aulas.
SOUZA, Z. A. S.; FERREIRA, M. A.; MIRANDA, G; J. (2019).	Teoria da atribuição de causalidade: percepções docentes e discentes sobre os determinantes do desempenho acadêmico	Advances in Scientific and Applied Accounting, São Paulo v.12, n.2 p. 040-058 maio / ago. de 2019	Motivação, quantidade de horas de estudo, frequência às aulas, conhecimento prévio, formação pedagógica do corpo docente e a infraestrutura da IES.
TELES, J. et al. (2019)	Desempenho acadêmico dos estudantes: uma análise dos fatores preditivos.	FACEF Pesquisa - Desenvolvimento e Gestão, v. 22, n. 1, 17 dez. 2019.	Idade, Nota de ingresso à universidade, Gênero.

Fonte: elaborado pelos autores (2022).

A revisão da literatura evidenciou que as variáveis de influência do desempenho acadêmico de um estudante de graduação podem ser distribuídas em três categorias, a saber: influência do corpo docente, influência do corpo discente e influência institucional.

As variáveis mais relevantes, que se repetiram pelo menos uma vez nos estudos avaliados, foram:

- Influência do corpo docente: método e estratégias de ensino.
- Influência do corpo discente: idade; antecedentes, fatores socioeconômicos; gênero; frequência às aulas; horas de estudo extraclasse; desempenho escolar anterior; conhecimento prévio do conteúdo; experiência profissional na área; grau de escolaridade dos pais; nota de ingresso à faculdade (vestibular); período do curso, turno; tipo de

instituição do ensino médio proveniente; trabalhar concomitantemente ao curso; área de especialização; motivação.

- Influência institucional: infraestrutura, ambiente de estudo.

Com o intuito de adequar o estudo para o contexto de aplicação da instituição de ensino escolhida, assim como da disciplina Cálculo Diferencial e Integral I, oferecida em dois cursos de engenharia distintos, optou-se por excluir do modelo preditivo as variáveis: experiência profissional na área, trabalhar concomitantemente ao curso, área de especialização, método e estratégias de ensino, motivação, infraestrutura e ambiente de estudo, tipo de instituição do ensino médio proveniente, conhecimento prévio, frequência às aulas, nota de ingresso à faculdade (vestibular) e período do curso e turno. O quadro 2 apresenta o quantitativo das variáveis explicativas do desempenho na revisão da literatura.

Quadro 2 – Estudo de variáveis

Variáveis explicativas do desempenho	Número de artigos referenciados
Idade	7
Fatores socioeconômicos	7
Gênero	7
Horas de estudo extraclasse	6
Desempenho escolar anterior	5
Grau de escolaridade dos pais	3
Tipo de escola do ensino médio	3

Fonte: elaborado pelos autores (2022).

Desta maneira, as variáveis utilizadas no modelo preditivo do desempenho acadêmico foram definidas para a elaboração do modelo preditivo.

Procedimentos metodológicos

O estudo desenvolvido para a concretização do projeto utilizou uma abordagem quantitativa em nível explicativo, pois investigou o grau de associação existente entre o perfil acadêmico e socioeconômico dos estudantes ingressantes de dois cursos de engenharia distintos, assim como os seus desempenhos na disciplina de Cálculo Diferencial e Integral I, em uma faculdade privada, porém gratuita, pertencente a uma Fundação, do município de São Bernardo do Campo, constatados no período compreendido entre janeiro de 2022 e junho do mesmo ano.

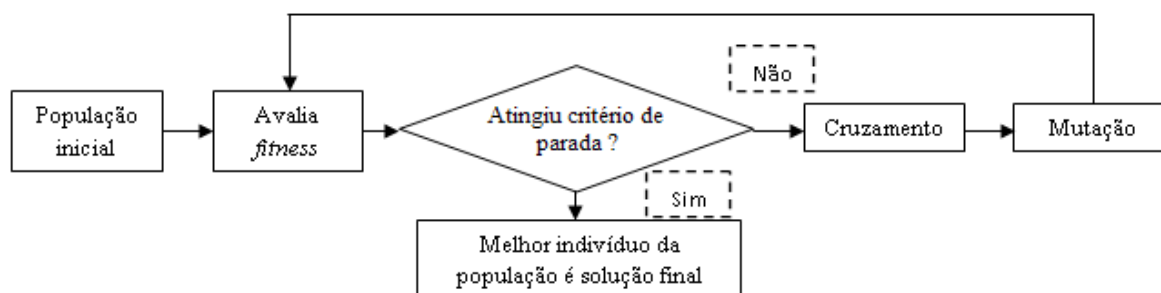
A amostra utilizada no estudo é não probabilística e intencional, formada por 63 alunos que ingressaram nos cursos de Engenharia de Alimentos (1 turma – desde o semestre/2022) e Engenharia de Controle e Automação (1 turma – desde o primeiro semestre de 2022). Em cada uma das turmas ingressaram 40 alunos, que serão classificados entre “Acima da média” e “Abaixo da média”, tendo como padrão de média 5 pontos (média necessária para a aprovação na disciplina em uma escala de 0 a 10).

Para a elaboração do modelo utilizaram-se as variáveis mais relevantes, que se repetiram pelo menos uma vez nos estudos avaliados. Além disso, excluíram-se as variáveis que não se relacionaram com o contexto de aplicação da instituição de ensino escolhida, assim como da disciplina Cálculo Diferencial e Integral I, oferecida em dois cursos de engenharia distintos. O referencial teórico apresenta as variáveis excluídas e as variáveis escolhidas para a composição do modelo.

Inicialmente foram consideradas as 7 variáveis que emergiram na revisão da literatura especializada, a saber: idade, gênero, fatores socioeconômicos, horas de estudo extraclasse, desempenho escolar anterior, grau de escolaridade dos pais e tipo de escola que concluiu o ensino médio, coletadas por meio de entrevista e pelo desempenho acadêmico desses mesmos estudantes. Duas variáveis foram diferenciadas em outros indicadores: a variável de escolarização dos pais diferiu-se para cada um, mãe e pai; a variável de fator socioeconômico foi estimada a partir de uma proxy, representada pela média de quatro questões que, no final, tornaram-se dados de influência para o algoritmo genético. As demais variáveis foram separadas em intervalos a fim de facilitar os cálculos do modelo.

Na elaboração do modelo, utilizou-se a técnica denominada algoritmo genético, escrita com a linguagem de programação denominada Python. O algoritmo genético foi inspirado na Biologia, e está representado na Figura 1.

Figura 1 – Fluxograma do algoritmo genético



Fonte: elaborado pelos autores (2022).

Assim como estudado e descrito por Morais, Souza e Cassoni (2019), os algoritmos genéticos são cálculos de busca e otimização global, baseados nos procedimentos de seleção natural e da genética. O desenvolvimento desta técnica está relacionado às pesquisas de John Holland (1975), inspirado nos estudos de Charles Darwin e Gregor Mendel sobre evolução e genética, respectivamente. A princípio, é gerada uma população inicial, formada por diversos indivíduos (cromossomos, ou cadeia de bits onde cada bit é um gene) aleatórios, que podem ou não ser a solução do problema.

Cada cromossomo tem por objetivo fornecer uma função matemática cujos parâmetros de entrada são os dados coletados, e seu resultado expressa o desempenho esperado do aluno na disciplina, para este estudo. Com isso, o cromossomo possuirá, no final, um *fitness* (performance) baseado na porcentagem de acertos e erros em relação à variável independente dos alunos, neste caso, seu resultado em Cálculo.

Diante desses resultados, interpolam-se pares de cromossomos, selecionados mediante suas taxas de acerto, para gerar outros indivíduos (“filhos”). Esse processo é chamado de Reprodução e visa avaliar a eficiência e a exatidão do algoritmo final.

Dentro das populações analisadas durante o processo, armazenam-se os grupos com melhores resultados os quais estão sujeitos a modificações em seus genes, por meio de cruzamento (crossover) e mutações, assim gerando descendentes para a próxima geração, que podem ter maiores potenciais de exatidão.

Esse processo acontece repetidamente e só cessará quando um índice de acerto desejável ou favorável ao estudo for alcançado. Para isso, recomendou-se pelo menos 1 hora de realização desta atividade.

Para o desenvolvimento do modelo, utilizaram-se os seguintes equipamentos: VISAGE PC BLEU 7500 445A (Windows 10 Pro Versão 21H1 – 64 bits, com memória RAM de 4GB, Processador Intel® Core™2 Duo CPU E7500 @ 2.93GHz 2.93 GHz). Em relação à programação, foi utilizado o ambiente de desenvolvimento PyCharm, versão 2021.3.0.0, e a linguagem Python, versão 3.10, para o algoritmo genético, e a IDE (do inglês *Integrated Development Environment*) Visual Studio, versão 16.11.31729.503, em linguagem C#, versão 8.0.

Apresentação e discussão dos resultados

Para decidir as variáveis a serem utilizadas, realizou-se uma revisão da literatura especializada acerca dos fatores de influência do desempenho acadêmico, e foram escolhidos aqueles com maiores índices de repetições presentes nos estudos acadêmicos sobre o tema (3 aparições ou mais), resultando em 7 variáveis para o desenvolvimento.

Posteriormente, adequou-se a variável “fatores socioeconômicos” para quatro questões que indicaram essa condição. O quadro 3 apresenta os detalhes das variáveis utilizadas.

Quadro 3 – Especificação das variáveis utilizadas

Natureza da variável	Variável	Respostas possíveis
Atributo Individual	Idade	De 17 a 19, de 20 a 22, Maior que 23
	Gênero	Masculino e Feminino
	Desempenho escolar anterior	Intervalos: [0,2];]2,4];]4,6];]6,8];]8,10]
	Horas de estudo extraclasse	Nenhuma, De uma à 3, De 4 à 7, De 8 à 12, Mais que 12 horas
	Tipo de escola no Ensino Médio	Todo em escola pública, Todo em escola privada, Todo no exterior, Maior parte em escola pública, Maior parte em escola privada
Contexto Familiar	Escolaridade da mãe	Nenhuma, Fundamental I, Fundamental II, Ensino Médio Completo, Ensino Superior - Graduação, Ensino Superior - Pós
	Escolaridade do pai	
Fatores Socioeconômicos	Possui um smartphone	Sim, com acesso à internet; Sim, sem internet própria; Não
	Possui um notebook	
	Possui um ambiente de estudos em casa	Sim, local exclusivo; Local adequado mas compartilhado; Não
	Trabalha concomitante ao curso	Sim, todos os dias; Sim, nem todos os dias; Não

Fonte: elaborado pelos autores (2022).

Figura 2 – Declaração de variáveis e constantes

```
1 var
2     Inteiro: qtdCromossomos = 14
3     Inteiro: qtdVariaveis = 11
4     Inteiro: qtdAlunos
5
6
7     Vetor: vetorAlunos = Vetor[1..qtdAlunos]
8     Vetor: PrimeiraPopulacao = Vetor[1.. qtdCromossomos]
9     Vetor: PopulacaoAtual = Vetor[1.. qtdCromossomos]
10    Vetor: fitnessAtual = Vetor[1.. qtdCromossomos]
```

Fonte: elaborado pelos autores (2022).

A quantidade inicial de cromossomos definida para o projeto foi determinada por conveniência após testagens com valores menores, e chegou-se à conclusão de que um número maior de cromossomos iniciais poderia convergir mais rapidamente à resposta do problema. Na Figura 3, mostra-se o método que origina as populações iniciais.

Figura 3 – Construção da população inicial

```
1 Metodo GeraPrimeiraPopulacao ( ): Vetor[1.. qtdCromossomos]
2
3 Inicio
4     Vetor: primeira_populacao = Vetor[1.. qtdCromossomos]
5
6     Para (Inteiro: n = 1) ate (qtdCromossomos) faca
7         Vetor: NovoCromossomo = [0] * qtdVariaveis
8
9         Para (Inteiro: m = 1) ate (qtdVariaveis) faca
10            NovoCromossomo [m] = numeroAleatorio(-10000, 10000) / 10000.0
11
12            primeira_populacao.Insere(NovoCromossomo)
13
14     Retorne primeira_populacao
15 FimMetodo
```

Fonte: elaborado pelos autores (2022).

Para o cálculo do *fitness* é utilizada a seguinte fórmula matemática:

$$\text{Fitness} = \left(\frac{\text{n}^\circ \text{ desempenhos favoráveis encontrados}}{\text{n}^\circ \text{ favoráveis no banco de dados}} \right) \times \left(\frac{\text{n}^\circ \text{ desempenhos não favoráveis encontrados}}{\text{n}^\circ \text{ não favoráveis do banco de dados}} \right)$$

Na Figura 4, apresenta-se o modo como o algoritmo realiza a operação anterior para cada um dos cromossomos existentes.

Figura 4 – Cálculo do *fitness*

```
Metodo ObtemFitness (Vetor: PopulacaoAtual): Vetor[1.. qtdCromossomos]
Inicio
  Vetor: fitnessAtual = Vetor[1.. qtdCromossomos]

  Para (Inteiro: c = 1) ate (qtdCromossomos) faca
    Inteiro: acertos_acima = 0
    Inteiro: acertos_abaixo = 0

    Para (Inteiro: n = 1) ate (qtdAlunos) faca
      Real: resultado = 0

      Para (Inteiro: m = 1) ate (qtdVariaveis - 1) faca
        resultado += vetorAlunos[n][m] * PopulacaoAtual[c][m]
      resultado += PopulacaoAtual [c][qtdVariaveis]

      Se (vetorAlunos[n][qtdVariaveis] >= 5) E (resultado >= 0)
        acertos_acima += 1
      Se (vetorAlunos[n][qtdVariaveis] < 5) E (resultado < 0)
        acertos_abaixo += 1

      Real: fitness = (acertos_acima / qtdAlunosAcima) * (acertos_abaixo / qtdAlunosAbaixo )
      fitnessAtual.Insere(fitness)

  Retorne fitnessAtual
FimMetodo
```

Fonte: elaborado pelos autores (2022).

Finalmente, o programa desenvolvido, que gerencia as populações, escolhe os pais, realiza cruzamentos e mutações, está descrito na Figura 5. Ele funciona repetidamente (loop) até que alcance um *fitness* desejado ou definitivo (taxa de acerto estacionária).

Figura 5 – Programação: cruzamentos e mutações

```
1 Metodo GeraFilhosEMutantes (Vetor: PopulacaoAtual): Vetor[1.. qtdCromossomos]
2
3 Inicio
4   Vetor: fitnessAtual = ObtemFitness(PopulacaoAtual)
5   Vetor: filhosEMutantes = Vetor[1..qtdCromossomos]
6
7   Para (Inteiro: v = 1) ate (qtdCromossomos / 2) faca
8     Real: fitTotal = Soma(fitnessAtual)
9     Inteiro: indicePai = numeroAleatorio (0, 100 * fitTotal) / 100
10
11     Para (Inteiro: i = 1) ate (qtdCromossomos) faca
12       indicePai = indicePai - fitnessAtual[i]
13       Se (indicePai <= 0)
14         indicePai = i
15         fitTotal = fitTotal - fitnessAtual [i]
16       FimPara
17
18     Inteiro: indiceMae = numeroAleatorio (0, 100 * fitTotal) / 100
19
20     Para (Inteiro: j = 1) ate (qtdCromossomos) faca
21       Se (indicePai <> j)
22         indiceMae = indiceMae - fitness_atual[j]
23         Se (indiceMae <= 0)
24           indiceMae = j
25         FimPara
26
27     Vetor: pai = PopulacaoAtual[indicePai]
28     Vetor: mae = PopulacaoAtual[indiceMae]
29
30     Inteiro: indiceDivisao = numeroAleatorio (0, qtdVariaveis - 1)
31     Vetor: filho = pai[0:indiceDivisao] + mae[indiceDivisao:qtdVariaveis]
32     filhosEMutantes.Insere(filho)
33
34     Inteiro: indiceMutacao = numeroAleatorio (0, qtdVariaveis - 1)
35     Inteiro: indiceOrigem= numeroAleatorio (0 qtdCromossomos - 1)
36     Real: valorMutacao = PopulacaoInicial[indiceOrigem , indiceMutacao]
37
38     filho[numeroAleatorio(0, qtdVariaveis - 1)] = valorMutacao
39     filhosEMutantes.Insere(filho)
40
41   Retorne filhosEMutantes
42 FimMetodo
```

Fonte: elaborado pelos autores (2022).

Anota-se que o modelo preditivo elaborado dispõe de 11 variáveis, presentes no Quadro 3, e chegou a 96,97% de acerto dos alunos com alto e baixo riscos de reprovação, durante o período de 2 horas de execução do programa.

Considerações finais

O modelo preditivo desenvolvido foi capaz de identificar o provável desempenho dos estudantes da disciplina de Cálculo Diferencial e Integral I, com alto e baixo riscos de reprovação. As informações fornecidas, acadêmicas e pessoais, dos alunos influenciaram na taxa de acerto desejada.

Dessa forma, o modelo preditivo se configura como uma ferramenta estratégica para gestores e docentes, visto que apresenta um cenário antecipado do desempenho dos alunos, possibilitando realizar ações para a alteração da realidade em curso ou mitigar os impactos dessa realidade.

O algoritmo genético revelou-se ideal para se enfrentar esse tipo de problema, uma vez que sua eficiência de cálculo e formulação de constantes para a função final, dentro de um lapso favorável, trouxeram uma proporção de acertos satisfatória.

A linguagem de programação Python foi determinante para a elaboração do projeto, visto que possui uma sintaxe de fácil compreensão, interpretação, interação, acesso e destaca-se por seu código aberto. Além disso, a facilidade de integração com outras linguagens, neste caso C#, também é um de seus benefícios.

Em razão da alta importância de se identificar rendimentos insuficientes em disciplinas do ensino superior, o modelo preditivo aqui formulado torna-se uma ferramenta estratégica para a aplicação em gestão educacional da instituição ofertante dos dados. Antecipar possíveis reprovações mostra-se de fundamental relevância a fim de se evitar despesas e prejuízos, tanto por parte da academia quanto dos alunos. O sistema produzido é, também, replicável em outras faculdades que incorporem em seus currículos a disciplina de Cálculo Diferencial e Integral I, desde que disponibilizem de um banco de dados compatível com as mesmas variáveis.

Em relação a futuros estudos e aplicações, primeiramente, recomenda-se a adição de novos dados de alunos e a atualização do modelo para obter-se uma convergência que se alinhe com os fins da análise de desempenho provável. Além disso, sugere-se a implantação de outras variáveis apontadas na revisão da literatura especializadas e de acordo que se adequam a esta proposta, que, porém, não foram utilizadas.

Referências

- ABBAS, K.; LOPES, A. K., Impacto dos fatores pessoais, institucionais e estilos de aprendizagem no desempenho acadêmico: uma análise com estudantes de contabilidade. **Revista Catarinense da Ciência Contábil**, Florianópolis, SC, v. 19, 1-31, e3020, 2020.
- AMARO, H. D.; BEUREN, I. M., Influência de Fatores Contingenciais no Desempenho Acadêmico de Discentes do Curso De Ciências Contábeis. **Revista de Educação e Pesquisa em Contabilidade (REPeC)**, v. 12, n. 1, p. 21–42, 29 mar. 2018.
- ARAÚJO, E. A. T. et al., Desempenho Acadêmico de Discentes do Curso de Ciências Contábeis: uma Análise dos seus Fatores Determinantes em uma IES Privada. **Contabilidade Vista & Revista**, v. 24, n. 1, p. 60–83, 2013.
- FAGUNDES, C. V.; LUCE, M. B.; RODRIGUEZ ESPINAR, S., O desempenho acadêmico como indicador de qualidade da transição Ensino Médio-Educação Superior. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 22, n. 84, p. 635–669, set. 2014.
- MAGALHÃES, F. A. C.; ANDRADE, J. X., Exame vestibular, características demográficas e desempenho na universidade: em busca de fatores preditivos. **Congresso USP de Controladoria e Contabilidade.**, volume 6. 2006.
- MANHÃES, L. M. B.; CRUZ, S. M. S., Predição do desempenho acadêmico de alunos da graduação utilizando mineração de dados. **XIX Simpósio de Pesquisa Operacional e Logística da Marinha**. Rio de Janeiro, RJ, Brasil, 06 a 08 de novembro de 2019.
- MARTINS, E. C. et al. Um estudo dos preditores de sucesso acadêmico no ensino superior. **Revista de Psicología y Educación**, 2015, 10(1),77-100.
- MARTINS, Z. B.; MARINHO, S. V., Relação das variáveis concernentes ao desempenho acadêmico: um estudo com alunos de graduação em ciências contábeis. **Revista Universo Contábil, FURB, Blumenau**, v. 15, n. 1, p. 27-48, jan./mar., 2019.
- MIRANDA, G. J. et al., Determinantes do Desempenho Acadêmico na Área de Negócios. **Meta: Avaliação / Rio de Janeiro**, v. 7, n. 20, p. 175-209, maio/ago. 2015.
- MORAIS, D. M. G. DE; SOUZA, A. A. M. DE; CASSONI, V. Um modelo preditivo da evasão de estudantes no ensino superior. **FTT Journal of Engineering and Business**, v. 1, n. 5, 2 fev. 2020.
- NOGUEIRA, D. R. et al., Fatores que impactam o desempenho acadêmico: Uma análise com discentes do Curso de Ciências Contábeis no Ensino Presencial. **Revista de Informação Contábil**, v. 7, n. 3, p. 51–62, 2013.
- ROCHA, A.; LELES, C.; QUEIROZ, M., Fatores associados ao desempenho acadêmico de estudantes de Nutrição no Enade. **Revista Brasileira de Estudos Pedagógicos**, v. 99, n. 251, 18 jun. 2019.
- ROSA, C. M. et al. Desempenho Acadêmico em Cálculo Diferencial e Integral: um Estudo de Caso. **Rev. Inter. Educ. Sup. Campinas, SP**, v.5, 1-16, 2019.
- SENA, T. T. O; SOUZA, A. A; SOUZA, A. G. Cálculo diferencial e integral: analisando os fatores que interferem no desempenho dos acadêmicos do curso de matemática da Ufal – campus de Arapiraca. **I Simpósio de Pesquisa e Extensão em Grupos Colaborativos e Cooperativos**. 26 e 27 de novembro, 2014.
- SILVA, H. G.; JÚNIOR, A. P. O., Fatores determinantes do desempenho acadêmico no ensino superior: estado da arte. **Revista Plurais – Virtual**, Anápolis - Go, vol.6, n. 2 – jul./dez. 2016.
- SOUTO-MAIOR, C.; BORBA, J. A.; KNUPP, P., Como Melhorar a Aprendizagem em Pesquisa Operacional? Uma Análise do Desempenho de Estudantes de Administração de Empresas e de Contabilidade. **Sociedade, Contabilidade e Gestão**, v. 9, n. 3, 4 maio 2015.

SOUZA, Z. A. S.; FERREIRA, M. A.; MIRANDA, G; J., Teoria da atribuição de causalidade: percepções docentes e discentes sobre os determinantes do desempenho acadêmico. **Advances in Scientific and Applied Accounting**, São Paulo v.12, n.2 p. 040-058 maio / ago. de 2019.

TELES, J. et al., Desempenho acadêmico dos estudantes: uma análise dos fatores preditivos. **FACEF Pesquisa - Desenvolvimento e Gestão**, v. 22, n. 1, 17 dez. 2019.

WISLAND, B; FREITAS, M, C, D; ISHIDA, C, Y. Desempenho acadêmico dos alunos em curso de Engenharia e Licenciatura na disciplina de Cálculo I. **Iberoamerican Journal of Industrial Engineering**, Florianópolis, SC, Brasil, v. 6, n. 11, p. 94-112, 2014.

ZARPELON, E; RESENDE, L. M. M; REIS, E, F. Análise do desempenho de alunos ingressantes de Engenharia na disciplina de Cálculo Diferencial e Integral I. **Interfaces da Educ., Paranaíba**, v.8, n.22, p.303-335, 2017.