

# PREDIÇÃO DE DESEMPENHO DE DISCENTES NO ENADE

Rafael Danoski

## RESUMO

Este trabalho tem como problema de pesquisa analisar a qualidade do ensino na graduação dos cursos de computação e tecnologia a partir do desempenho dos discentes nas avaliações do ENADE. Como objetivo geral, se visa criar um modelo estatístico de avaliação e predição. A partir dos dados das avaliações de discentes fornecidos pelo INEP, tais como, informações socioeconômicas dos estudantes e de infraestrutura das instituições de ensino, o resultado esperado deste modelo será, após aplicado a uma base de alunos, predizer a nota dos mesmos na avaliação do ENADE, e portanto, permitir antever quais os pontos de melhorias necessários aos discentes da instituição.

Palavras-chave: inteligência artificial; modelo de predição; avaliação do ensino superior; ENADE.

## ABSTRACT

This work's research problem is to analyze the quality of teaching in undergraduate computing and technology courses based on students' performance in ENADE assessments. As a general objective, the aim is to create a statistical model for evaluation and prediction. Based on data from student assessments provided by INEP, such as socioeconomic information on students and the infrastructure of educational institutions, the expected result of this model will be, after being applied to a student base, to predict their grade in the assessment ENADE, and therefore, allow us to predict which points of improvement are needed by the institution's students.

Key-words: artificial intelligence; predictive models; ENADE.

## 1 INTRODUÇÃO

O percentual de domicílios com acesso à internet no Brasil tem aumentado nos últimos anos, ultrapassando a marca de 80% segundo o CETIC (Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação) em seu último estudo (CETIC, 2023). Além disso, a média de tempo gasta em dispositivos móveis do brasileiro nos últimos três anos é a maior dentre todos os países pesquisados (DATA.IA, 2023), ultrapassando a marca de 5h30m de acordo com a empresa Data.IA (antiga App Annie) em seu relatório anual State of Mobile. De acordo com dados da Anatel de março de 2023, os estados brasileiros com a maior densidade de celulares, ou seja, a quantidade de celulares a cada 100 habitantes, seguem sendo São Paulo (159,90), Mato Grosso (122,49) e Rio de Janeiro (120,45).

O relatório publicado em 2023 pela FGV (MEIRELLES, 2023) indica que o investimento no setor de TI no último ano no Brasil foi o maior apresentado historicamente, alcançando a marca de quase 9% do faturamento médio das empresas. A tendência é de subida nos próximos anos, no entanto, quando olhamos para os números de uso de tecnologia na educação em escolas não temos o mesmo crescimento. Segundo o TIC Educação de 2019 (dados pré pandemia) apenas 40%

dos alunos do ensino fundamental e médio possuem acesso à internet nas escolas, além disso apenas 40% das escolas possuem laboratórios funcionais de informática e menos de 9% disponibilizam tablets para os alunos.

Esses dados revelam um infeliz paradoxo de má alocação de investimento em tecnologia no país. Segundo Chaimovich, o interesse na formação de profissionais qualificados na área da computação na educação básica ou na graduação, seja pelo setor público ou privado, parece estar negligenciado frente ao setor industrial, e quando olhamos estritamente para o viés econômico vemos que o retorno financeiro é mais imediato, o investimento em educação tem por natureza um retorno a longo prazo. Apesar do cenário negativo para educação, como já evidenciado, há uma projeção esperançosa já que pelo menos o interesse e uso de tecnologia nos domicílios brasileiros tem aumentado, faltando apenas uma alocação maior de recursos na área.

Segundo Feenberg, a tecnologia é transformadora, influencia na maneira como vivemos, trabalhamos e aprendemos. Aqui vale ressaltar a tecnologia como sendo qualquer produto de ciência ou engenharia que envolva um conjunto de instrumentos, métodos e processos que visem a resolução de um problema. Este entendimento é de extrema importância para o desenvolvimento desse projeto.

Ainda segundo Feenberg, uma das formas de influência da tecnologia é como ela tem mudado a maneira como nos relacionamos. Antes, seria impossível se comunicar com alguém que está a milhares de quilômetros de distância, mas agora podemos nos conectar com pessoas de todo o mundo por meio de aplicativos de mensagem e redes sociais. Essas plataformas geram uma grande quantidade de dados que podem ser utilizados para analisar o comportamento humano. Com essas informações, podemos criar soluções que beneficiem os usuários e impactem positivamente sua rotina.

Segundo Silva, a ciência de dados têm permitido que empresas e organizações analisem grandes quantidades de dados e identifiquem tendências e padrões que antes passavam despercebidos. Algoritmos de machine learning e inteligência artificial estão sendo usados em várias áreas, como saúde, finanças, marketing e educação, para desenvolver soluções inovadoras e personalizadas que atendam às necessidades individuais dos usuários.

Ainda segundo Silva, na área de educação, a tecnologia também tem tido um impacto significativo. Ferramentas educacionais como plataformas de aprendizagem online e aplicativos de educação podem ajudar a melhorar a qualidade da educação e torná-la mais acessível a um número maior de pessoas.

Através das já mencionadas ferramentas de análise de dados, como ciência de dados, aprendizado de máquina e inteligência artificial, é possível explorar e trazer soluções para os questionamentos apontados. Uma das formas de exploração é analisar o nível de desempenho da graduação das áreas relacionadas à tecnologia no Brasil. Uma das principais ferramentas para avaliar a qualidade do nível universitário no país é o Exame Nacional de Desempenho dos Estudantes (ENADE), o ENADE é uma prova que avalia o desempenho dos estudantes em relação às competências, habilidades e conteúdos previstos nas diretrizes curriculares de cada curso.

Nesse contexto, o uso de técnicas de inteligência artificial e análise de dados pode ser uma ferramenta útil para ajudar instituições de ensino a identificar os fatores que mais impactam o desempenho dos estudantes no ENADE. Esse tipo de abordagem pode ser útil tanto para os próprios estudantes, que podem receber feedbacks personalizados e orientações para melhorar seu desempenho, quanto

para as instituições, que podem desenvolver políticas mais efetivas para melhorar a qualidade do seu ensino. Sendo assim, o produto tecnológico a ser criado será um modelo preditivo construído através de técnicas de aprendizado de máquina, uma sub-área da inteligência artificial, para prever o desempenho dos discentes que realizarão a avaliação do ENADE.

## **2 SISTEMA DE AVALIAÇÃO DO ENSINO SUPERIOR**

Para entender o sistema educacional brasileiro, mais especificamente a graduação, é preciso entender e se aprofundar na metodologia de avaliação do ENADE (Exame Nacional de Desempenho dos Estudantes) disponibilizada na página do INEP a respeito do tema.

O Sistema Educacional Brasileiro, vinculado ao Ministério da Educação, é estruturado e auto regulado a partir de diferentes bases de dados que fazem a leitura da realidade nas escolas e universidades brasileiras. As bases de dados apontam problemáticas, dificuldades, potencialidades que contribuem para a construção e reconstrução de um Sistema Educacional em congruência com a sociedade, economia e políticas no país.

A Educação Superior é auto regulada a partir da base de dados obtidos no Exame Nacional de Desempenho dos Estudantes (ENADE), realizado com estudantes no final dos cursos de graduação. O ENADE é composto por uma prova que abrange diversas áreas de conhecimento, e seu objetivo é medir o desempenho dos estudantes em relação aos conteúdos que foram ministrados ao longo do curso. A prova é construída a partir de uma matriz de referência, que é elaborada pelo MEC com a colaboração de especialistas em cada uma das áreas de conhecimento. Essa matriz define os conteúdos que serão avaliados.

A aplicação do ENADE é feita de forma presencial e os estudantes são convocados a participar do exame pela instituição de ensino na qual estão matriculados. A participação no exame é obrigatória para os estudantes selecionados. Além da prova, os estudantes também precisam responder a um questionário socioeconômico, que ajuda o MEC a traçar um perfil dos estudantes que estão concluindo a graduação no país.

## **3 TRABALHOS RELACIONADOS**

Existem vários trabalhos que já foram desenvolvidos buscando entender quais fatores estão relacionados ao desempenho dos discentes ou tentando prever seu desempenho na graduação. Nesta seção buscarei explorar alguns desses trabalhos, tanto em suas similaridades, quanto nas diferenças com este.

O intuito da pesquisa por trabalhos relacionados foi explorar as técnicas utilizadas em modelos de aprendizado de máquina em bases públicas da educação brasileira, e portanto, com foco mais técnico do que acadêmico, sendo assim, os artigos citados foram encontrados via Google Acadêmico a partir de palavras chave, tais como, “predição de desempenho ENADE”, “dados educação superior”, “aprendizado de máquina ENADE”.

No trabalho de Silva, Hoed e Saraiva (2019) foram utilizadas técnicas de mineração de regras de associação, com base no algoritmo Apriori e no método CRISP-DM, para descobrir quais fatores estão relacionados ao desempenho dos estudantes da área de computação. Para realizar isso, foi utilizado o questionário socioeconômico disponibilizado pelos microdados do ENADE de 2017. Foram

analisadas em torno de oitenta variáveis. Como resultado, chegaram em algumas conclusões como a de que 88% dos alunos que tiveram nota alta na parte de componente específico e que disse que seu curso disponibilizou monitores ou tutores e o fato da IES ter uma boa estrutura e boas políticas de acompanhamento obtiveram uma nota bruta geral alta.

No trabalho de Capelari (2022) foram utilizadas técnicas de aprendizado de máquina para prever o desempenho de alunos no ENADE. O autor utilizou cinco algoritmos diferentes e ao final avaliou qual obteve o melhor desempenho com os dados para aplicá-lo ao modelo. Como conclusão foi possível traçar o perfil do discente na avaliação, sendo de maneira geral uma pessoa que: faz curso no período noturno; fez o Ensino Médio em escola pública; tem entre 22 e 25 anos; é do sexo masculino; tem cor de pele branca; com o pai tendo estudado até o Ensino Fundamental e a mãe estudado até o Ensino Médio; possui uma renda familiar entre 3 e 10 salários mínimos e já possui um emprego.

Na tabela abaixo exemplifico as semelhanças e diferenças dos trabalhos citados anteriormente com este.

| Trabalho           | Semelhanças  | Diferenças  |
|--------------------|--|---|
| Silva et al., 2019 | Foram utilizados algoritmos para descrever quais as melhores variáveis que explicam o desempenho no ENADE. | Não foi utilizada uma base de dados muito extensa (apenas uma avaliação), implicando na qualidade do modelo.              |
| Capelari, 2022     | Foi utilizado técnicas de aprendizado de máquina para predição de desempenho no ENADE.                     | Os algoritmos utilizados são diferentes dos utilizados neste trabalho, implicando em resultados possivelmente diferentes. |

## 4 METODOLOGIA

A pesquisa a ser realizada se caracteriza como uma produção tecnológica e experimental quantitativa/qualitativa, tendo como base o referencial teórico e desenvolvimento prático de um modelo estatístico de predição de desempenho.

Como procedimento de pesquisa será adotada a coleta de dados de bases públicas, mais especificamente as fornecidas pelo INEP com dados de desempenho de discentes do curso de ciências da computação no ENADE. A partir de uma análise exploratória desses dados, será buscado um entendimento maior sobre suas informações para posteriormente ser construído um modelo de predição por meio de técnicas estatísticas e de aprendizado de máquina. O projeto será desenvolvido de maneira totalmente remota, utilizando ferramentas digitais durante todas as etapas. Para a realização do projeto será utilizado a linguagem de programação Python, a IDE (Ambiente de Desenvolvimento Integrado) Jupyter para utilização da linguagem, além de algoritmos fornecidos pela linguagem para implementação de técnicas estatísticas. Avaliação do projeto se dará a partir da comparação da acurácia dos modelos criados. Serão criados quatro modelos com princípios estatísticos diferentes e a comparação se dará por meio de indicadores de acurácia e eficiência do modelo para se entender qual o mais assertivo e vantajoso.

### 4.1 BASE DE DADOS

As bases de dados referentes ao ENADE estão dispostas no portal do INEP da seguinte forma: um diretório “LEIA-ME” com orientações sobre o uso dos dados, além de um dicionário das variáveis e o questionário sócio-econômico fornecido aos discentes; e um diretório “DADOS” contendo a base de dados segmentada em arquivos em extensão “.txt”. Cada um desses arquivos contém variáveis específicas sobre o exame aplicado em determinado ano.

Para obter um único conjunto de dados e utilizá-los para o modelo, se faz necessário unir os arquivos em uma base única, porém nem todos possuem alguma chave identificadora comum para que seja possível cruzá-los, portanto algumas variáveis que seriam essenciais ao modelo são perdidas.

No segundo arquivo do ano de 2021, por exemplo, existem as informações de início e conclusão do curso de todos os discentes que realizaram a prova naquele ano e no quinto arquivo a informação do sexo dos mesmos, porém a única chave em comum nos dois arquivos é o código do curso e o ano da realização da prova, sendo assim não seria possível atingir a granularidade a nível aluno necessária para formular o modelo.

Sendo assim, uma solução para resolver esse problema seria o próprio INEP disponibilizar a base de dados na íntegra sem a quebra por arquivos. Ressaltando que não há necessidade de nenhuma informação sensível dos discentes, apenas as informações que já estão disponibilizadas nos repositórios no INEP. Foi realizado o pedido ao serviço do SEDAP (Serviço de Acesso a Dados Protegidos) que oferece acesso a algumas bases de dados a pesquisadores da área educacional.

## **4.2 ANÁLISE EXPLORATÓRIA**

Antes de iniciar o tratamento dos dados e construção do modelo, foi realizada uma análise exploratória dos dados a fim de identificar as principais variáveis da base de dados e trazer insights significativos para o estudo. Para o estudo foram utilizadas apenas as informações dos cursos de tecnologia, sendo esses, “Tecnologia em análise e Desenvolvimento de Sistemas”, “Tecnologia em rede de computadores”, “Engenharia da Computação”, “Ciência da Computação (Bacharelado)”, “Ciência da Computação (Licenciatura)”, “Sistemas da Informação” e “Tecnologia em Gestão da Tecnologia da Informação”. Fazendo tal segmentação podemos perceber inicialmente o comparativo de nota geral dos cursos de tecnologia para os demais, enquanto no cenário geral temos uma nota média de 43,4, nos cursos de tecnologia esta nota é 42,6, uma distorção muito baixa.

Fazendo a abertura sobre o detalhamento dos cursos e instituições que compõem essa segmentação de tecnologia utilizada na análise encontramos algumas informações relevantes para o estudo. Quanto à organização acadêmica, aproximadamente 50% das instituições correspondem à Universidades, seguidas por Faculdades, Centro Universitários e uma pequena parte de Institutos e Centro Federais, como apresentado na imagem abaixo:

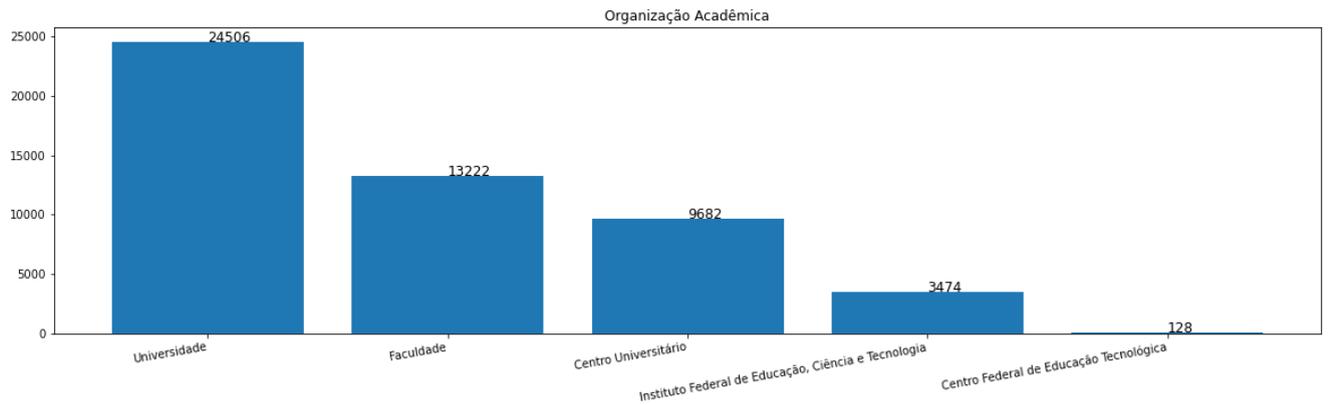


Imagem 1 - Organização Acadêmica.

A distribuição dos cursos segue uma distribuição mais linear, sendo liderado por “Sistemas da Informação”, “Tecnologia em análise de desenvolvimento de sistemas” e “Ciência da Computação (Bacharelado)”, conforme imagem abaixo:

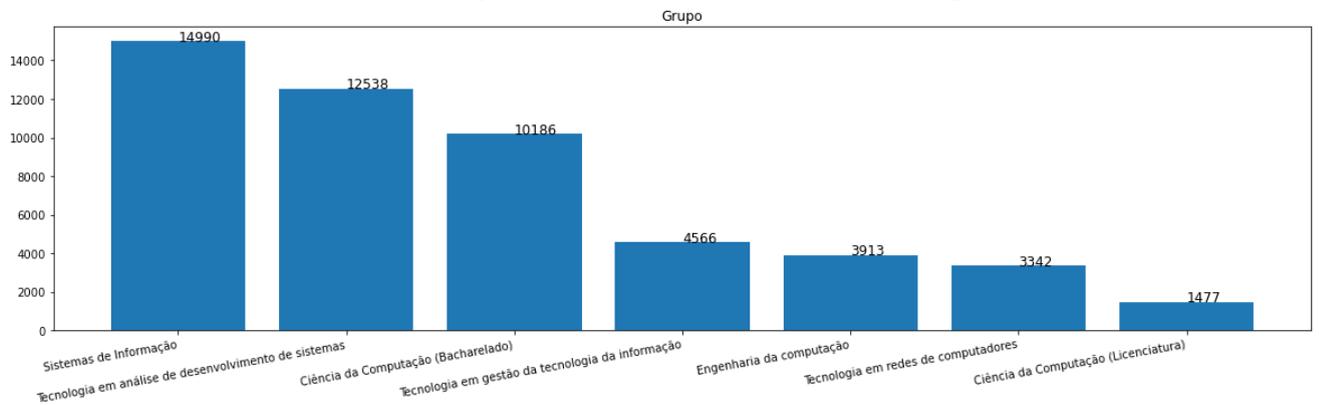


Imagem 2 - Grupo.

Outra variável interessante analisada é o tempo médio entre o término do ensino médio e início da graduação, para os discentes dos cursos de tecnologia há uma concentração bem alta no valor de um ano, causando assim, uma assimetria à esquerda no histograma, como podemos ver abaixo, isso indica um perfil mais jovem para os alunos destes cursos.

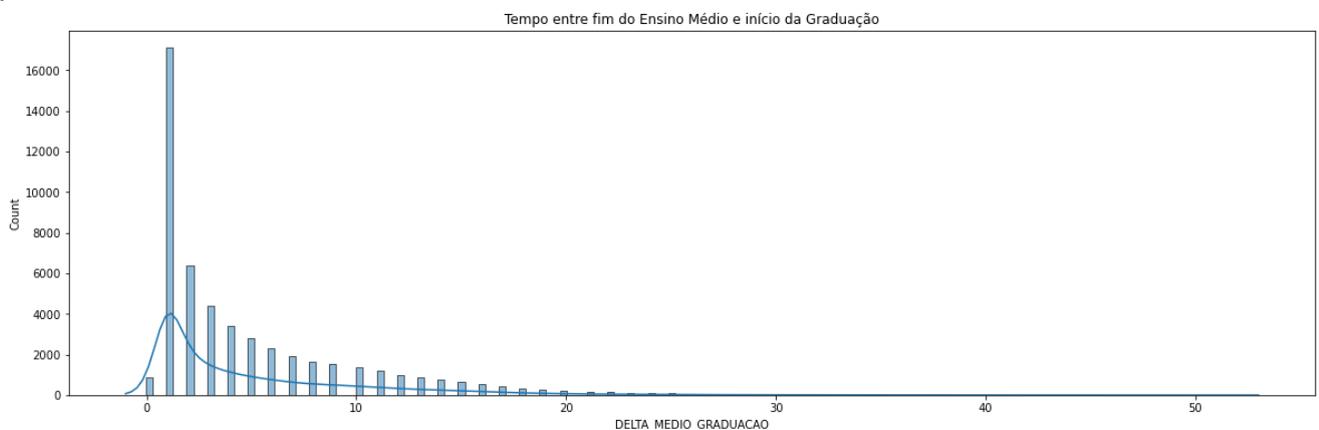


Imagem 3 - Tempo entre fim do ensino médio e início da graduação.

A distribuição da nota geral do ENADE entre os estudantes de cursos de tecnologia é simétrica tendo concentração maior próxima da nota 40, como podemos

ver abaixo, isso será importante no momento da construção do modelo, já que será preciso fazer o balanceamento da variável alvo.

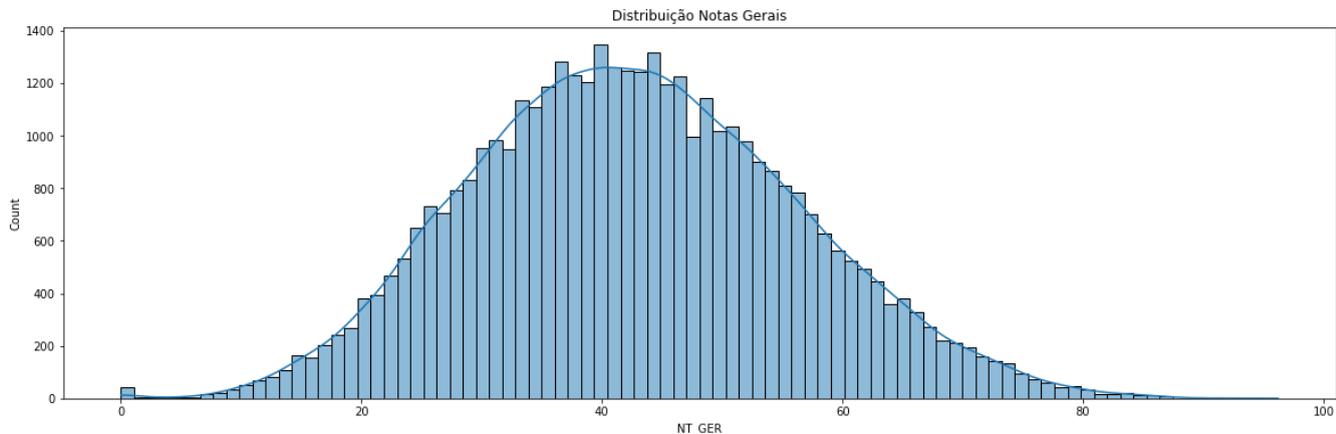


Imagem 4 - Distribuição Nota Geral.

Contribuindo para a variável de tempo médio entre o término do ensino médio e início da graduação, temos a variável de idade que aponta para a conclusão criada de uma faixa de idade menor para discentes dos cursos de tecnologia. Nota-se no gráfico abaixo uma concentração maior de idade na faixa dos vinte anos.

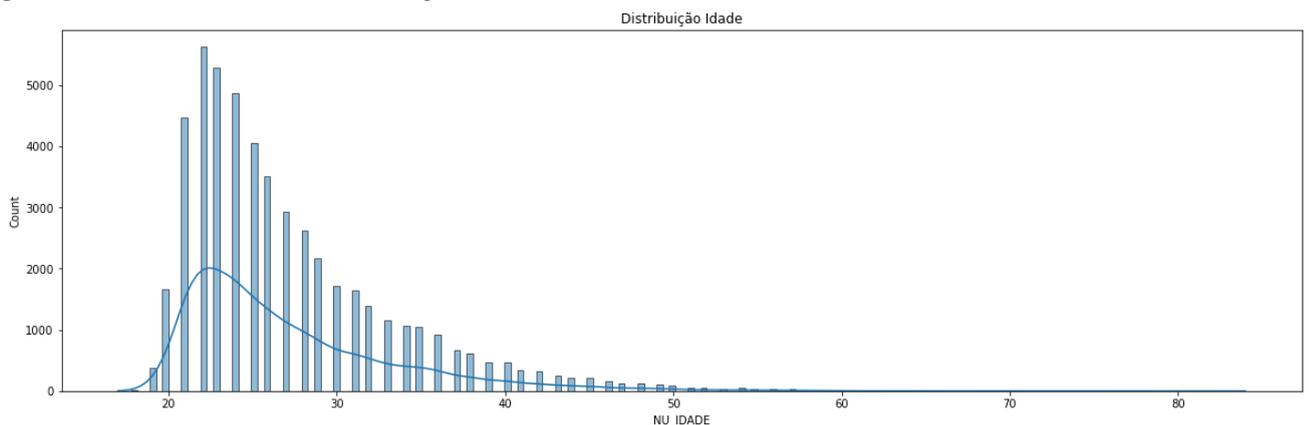


Imagem 5 - Distribuição de Idade.

Um dos indicadores mais importantes para o modelo é o nível de escolaridade dos pais, como podemos ver abaixo temos uma distribuição desproporcional entre alunos com pai apresentando nível superior de ensino versus apenas nível médio, no entanto, quando observamos o desempenho destes alunos no ENADE vemos uma inversão na distribuição de notas, isto quer dizer, alunos com pais com nível de escolaridade maior apresentam também desempenho melhor no ENADE.

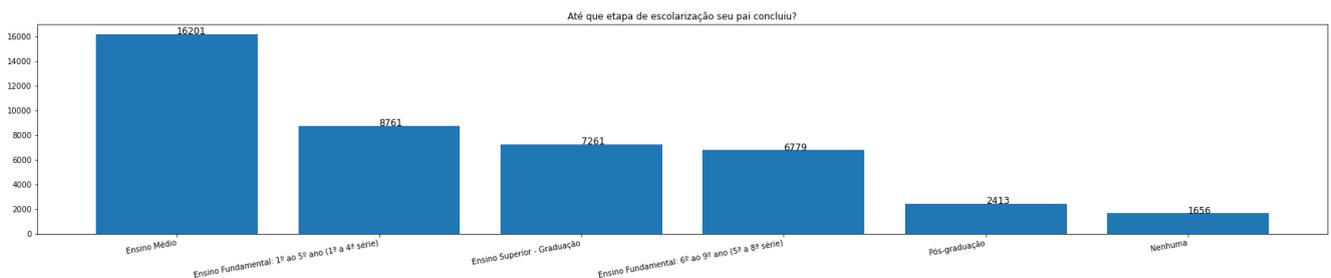


Imagem 6 - Nível de escolarização pai.

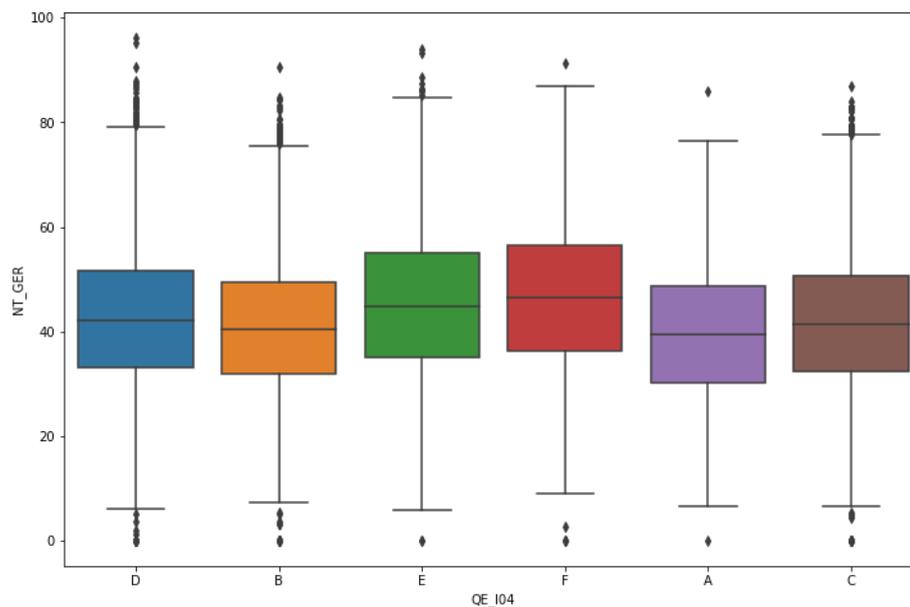


Imagem 7 - Distribuição de notas por escolarização do pai.

## 5 DESENVOLVIMENTO

Para o início do desenvolvimento do modelo, após a análise exploratória das variáveis, é necessário realizar a seleção das variáveis mais relevantes e que irão fazer a maior diferença na aplicação do algoritmo. Para o modelo em questão foram selecionadas as seguintes variáveis: tipo de grupo e organização, tempo entre saída do ensino médio e início da graduação, raça, escolarização do pai e da mãe, tipo de moradia atual, quantidade de integrantes na família, renda familiar, situação financeira, carga horária de trabalho, bolsa de estudos, tipo de escola no ensino médio, grupo de apoio durante a graduação, concluintes do ensino superior na família, quantidade de livros lidos no ano, horas semanais dedicadas ao estudo, aprendizado de idioma estrangeiro, motivo para escolha do curso e motivo para escolha da instituição.

Selecionadas as variáveis mais relevantes ao modelo, é necessário realizar a limpeza dos dados na base, para isso foi realizada a tratativa de dados nulos da seguinte forma, em torno de 15% dos registros possuíam mais de 80% das variáveis nulas, assim foi realizada a exclusão desses registros, no restante da base não foram realizadas mais alterações, com exceção dos dados nulos da variável alvo, neste caso, foram substituídos os nulos pela mediana da variável.

O próximo processo importante a ser realizado na base de dados antes de aplicar os algoritmos, é transformar as variáveis categóricas em variáveis numéricas para melhor compreensão do modelo, dessa forma todas as variáveis referentes à respostas de questões que se classificam originalmente como letras, são transformadas em números, assim a letra "A" se torna 0, "B" se torna 1 e assim sucessivamente.

Após a manipulação dos dados podemos verificar se ainda existe correlação forte entre as variáveis selecionadas, para que possamos analisar se é necessária a exclusão das mesmas, já que assim teríamos redundância na aplicação do modelo. A melhor forma de verificarmos graficamente a correlação das variáveis é através de um gráfico de correlação como podemos ver abaixo, aqui podemos verificar que não existe forte correlação positiva (correlação maior que 0,95) e nem negativa (correlação menor que -0,95) entre nenhuma das variáveis.

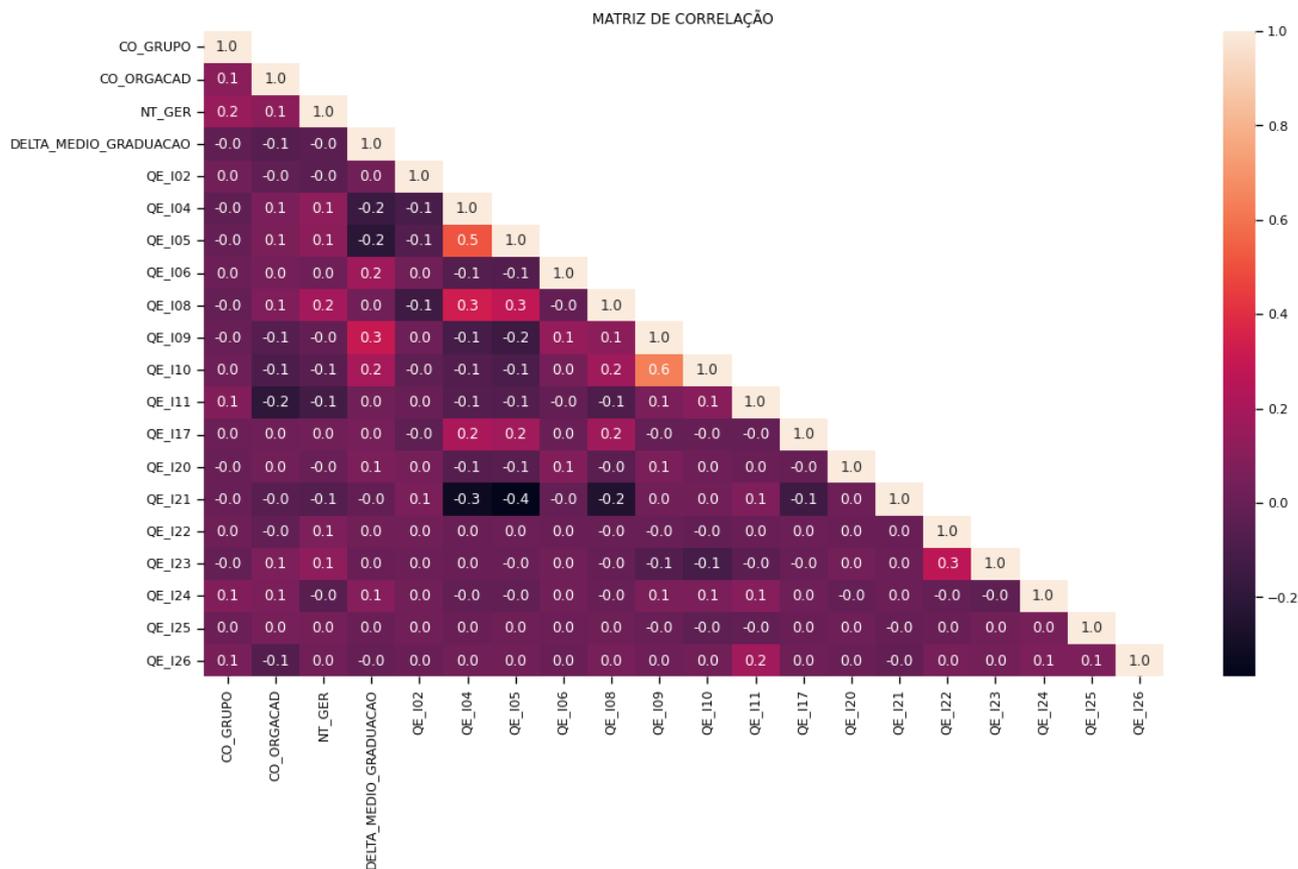


Imagem 8 - Correlação das variáveis.

Outra alteração importante para assertividade do modelo é a binarização da variável alvo, já que estamos tratando de uma variável numérica de ponto flutuante, ou seja, há infinitas possibilidades de números variando de 0 a 100, sendo assim, uma alteração interessante para auxiliar na assertividade do modelo é categorizar intervalos da variável alvo e assim diminuir o número de categorias a serem previstas pelo modelo, dessa forma, a variável NOTA GERAL foi categorizada em cinco categorias sendo notas entre 0 e 20 da categoria 1, entre 20 e 40 da categoria 2, entre 40 e 60 da categoria 3, entre 60 e 80 da categoria 4 e acima de 80 categoria 5. Dessa maneira o modelo não mais precisa prever a nota geral, mas sim a categoria a qual ela pertence.

Antes de testar os algoritmos ainda há um último ajuste a ser realizado na base de dados, o balanceamento das variáveis. Após a binarização as categorias não necessariamente se comportam de maneira simétrica já que a variável original não possuía distribuição simétrica, sendo assim, é necessário realizar o balanceamento da base de dados para que o modelo não possua mais registros específicos de uma única categoria e dessa forma enviesando o resultado do modelo.

Para aplicação dos algoritmos foi realizada a partição da base de dados em 70% para treinamento e o restante para teste. A métrica principal a ser analisada para escolha e tratamento do modelo é a acurácia. Para verificar o mínimo de acurácia aceitável para o modelo foi utilizado o método Dummy Classifier do Python, este método se utiliza do pior algoritmo possível para prever a variável alvo, ou seja, entrega a pior acurácia possível ao modelo, aplicando este classificador foi atingido a acurácia de 20%.

Foram escolhidos três algoritmos para teste no modelo: Regressão Logística, Árvore de Decisão e Random Forest Classifier. Aplicando o algoritmo de regressão logística com o parâmetro de 400 iterações máximas, a acurácia atingida foi de 94,44%, abaixo pode-se ver o resultado da matriz de confusão contabilizando os verdadeiros positivos e negativos, bem como os falsos positivos e negativos.

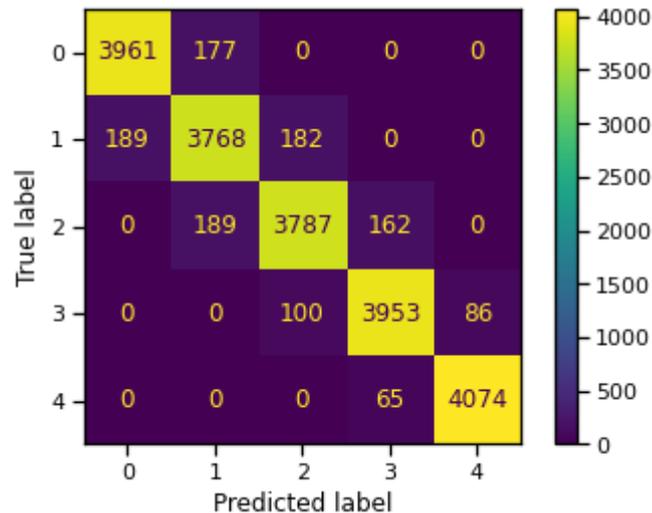


Imagem 9 - Matriz de confusão Regressão Logística.

Aplicando o algoritmo de árvore de decisão com o parâmetro de 2 nós máximos de profundidade, a acurácia atingida foi de 59,99%, abaixo pode-se ver o resultado da matriz de confusão contabilizando os verdadeiros positivos e negativos, bem como os falsos positivos e negativos.

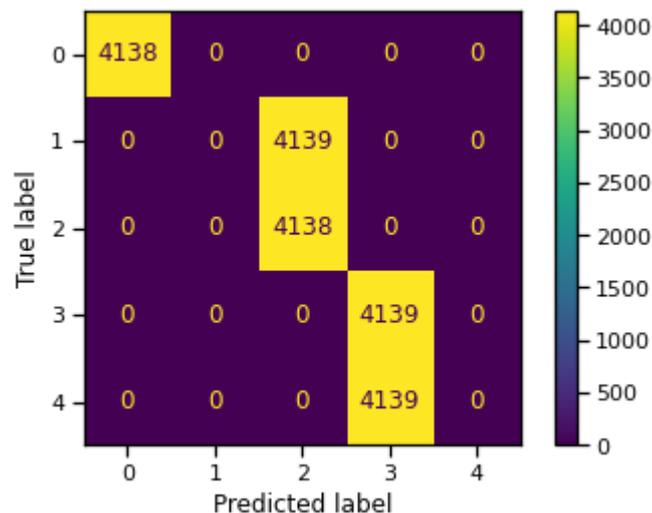


Imagem 10 - Matriz de confusão Árvore de Decisão.

Aplicando o algoritmo de random forest classifier com o parâmetro de 1 nó máximo de profundidade, a acurácia atingida foi de 97,18%, abaixo pode-se ver o resultado da matriz de confusão contabilizando os verdadeiros positivos e negativos, bem como os falsos positivos e negativos.

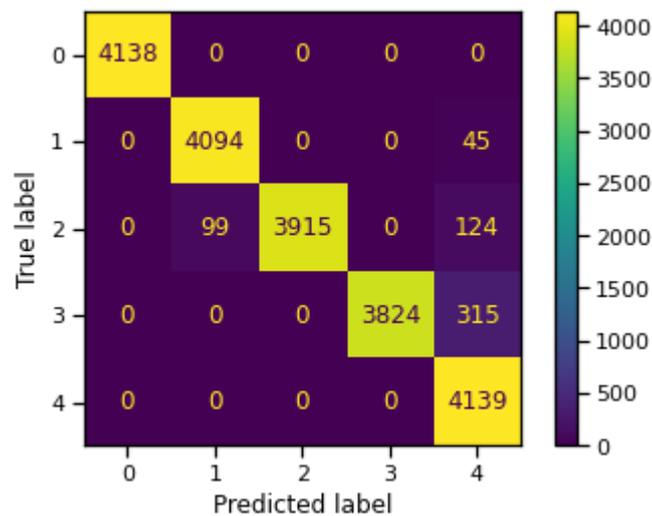


Imagem 11 - Matriz de confusão Random Forest Classifier.

Como pode-se perceber o modelo com maior assertividade é o de Random Forest Classifier atingindo 97,18% e poucas asserções falsas dentro das categorias.

## 6 CONSIDERAÇÕES FINAIS

O Brasil tem testemunhado um rápido crescimento no acesso à internet, ultrapassando 80% dos domicílios conectados, enquanto o tempo gasto em dispositivos móveis também aumenta. No entanto, a educação enfrenta desafios significativos, com apenas 40% dos alunos do ensino fundamental e médio tendo acesso à internet nas escolas, revelando uma alocação insuficiente de recursos na área. Embora o setor de tecnologia cresça, a formação de profissionais em computação na educação é negligenciada em comparação com o setor industrial. A tecnologia tem impactado a maneira como nos relacionamos e oferece oportunidades para melhorar a educação. O uso de inteligência artificial e análise de dados pode ser valioso para aprimorar o ensino e o aprendizado.

Com base nisso se desenvolveu um algoritmo de predição de desempenho do ENADE (Exame Nacional de Desempenho dos Estudantes), a partir de uma análise exploratória dos microdados fornecidos pelo INEP e do teste de três tipos de algoritmos. Ao final do desenvolvimento optou-se pelo algoritmo de melhor acurácia para o modelo, o de Random Forest Classifier, com acurácia de 97,18% de assertividade das notas dos discentes dos cursos de tecnologia.

## REFERÊNCIAS

Silva, Alexsander Figueiredo. Hoed, Raphael Magalhães. Saraiva, Pedro Fábio. ANÁLISE DO DESEMPENHO DOS ALUNOS DE CURSOS SUPERIORES EM COMPUTAÇÃO NO ENADE – UMA ABORDAGEM USANDO MINERAÇÃO DE DADOS. 2019. DOI: 10.33965/ciaca2019\_201914L026.

Capelari, Luís Otávio Oliveira. PREDIÇÃO DO DESEMPENHO NO ENADE DOS

DISCENTES DE COMPUTAÇÃO. 2022.

Leonardo Torres Marques, Bruno Torres Marques, Carlos Alexandre Morais Silva, Rayana Souza Rocha, Jesaias Carvalho Pereira Silva, Leonardo Chaves e Silva. A Evasão Escolar no Ensino Superior: Um Estudo de Caso do Curso de Ciência da Computação da UFERSA. 2020. DOI: <https://doi.org/10.34117/bjdv6n12-728>.

CETIC. Apresentação dos principais resultados - TIC EDUCAÇÃO. Disponível em: [https://cetic.br/media/analises/tic\\_educacao\\_2019\\_coletiva\\_imprensa.pdf](https://cetic.br/media/analises/tic_educacao_2019_coletiva_imprensa.pdf). Acesso em: 01/03/2023.

DATA.IA. State of Mobile 2023. Disponível em: <https://www.data.ai/en/go/state-of-mobile-2023/>. Acesso em: 01/04/2023.

Relatório anual de gestão ANATEL 2022. Disponível em: [https://sei.anatel.gov.br/sei/modulos/pesquisa/md\\_pesq\\_documento\\_consulta\\_externa.php?8-74Kn1tDR89f1Q7RjX8EYU46IzCFD26Q9Xx5QNDbqZG2Dq6f0ohSBFSvBZ6tX1N1DmIpl3xO7SRbFp\\_SpewnCN6J9b1EWrcW-GWwOG6gxCF05PJ2V3AdWj-CvW\\_MYkw](https://sei.anatel.gov.br/sei/modulos/pesquisa/md_pesq_documento_consulta_externa.php?8-74Kn1tDR89f1Q7RjX8EYU46IzCFD26Q9Xx5QNDbqZG2Dq6f0ohSBFSvBZ6tX1N1DmIpl3xO7SRbFp_SpewnCN6J9b1EWrcW-GWwOG6gxCF05PJ2V3AdWj-CvW_MYkw). Acesso em 01/05/2023.

MEIRELES, Fernando S. Relatório FGVcia Pes TI 2023. 34ª Edição Anual. FGVcia, 2023. Disponível em: [https://eaesp.fgv.br/sites/eaesp.fgv.br/files/u68/pesti-fgvcia-2023\\_0.pdf](https://eaesp.fgv.br/sites/eaesp.fgv.br/files/u68/pesti-fgvcia-2023_0.pdf) Acesso em: 01/05/2023.

Chaimovich, Hernan. Brasil, ciência, tecnologia: alguns dilemas e desafios. 2005.

Feenberg, Andrew. O que é a Filosofia da Tecnologia. 2010.

Silva A. Leandro, Silveira Frango Ismar, Silva Luciano, Ramos Luis Cavalcanti Jorge, Rodrigues Lins Rodrigo. Ciência de Dados Educacionais: definições e convergências entre as áreas de pesquisa. 2017.