

A Computational Strategy for Classification of Enem Issues Based on Item Response Theory

Uma Estratégia Computacional para Classificação de Questões do Enem Baseada na Teoria de Resposta ao Item

Gustavo Henrique Nunes¹, Bruno Alberto Soares Oliveira^{*2}, Ciniro Aparecido Leite Nametala¹

Resumo: The National High School Examination (ENEM) gains each year more importance, as it gradually, replacing traditional vestibular. Many simulations are done almost randomly by teachers or systems, with questions chosen without discretion. With this methodology, if a test needs to be reapplied, it is not possible to recreate it with questions that have the same difficulty as those used in the first evaluation. In this context, the present work presents the development of an ENEM Intelligent Simulation Generation System that calculates the parameters of Item Response Theory (TRI) of questions that have already been applied in ENEM and, based on them, classifies them. in groups of difficulty, thus enabling the generation of balanced tests. For this, the K-means algorithm was used to group the questions into three difficulty groups: easy, medium and difficult. To verify the functioning of the system, a simulation with 180 questions was generated along the ENEM model. It could be seen that in 37.7% of cases this happened. This hit rate was not greater because the algorithm confounded the difficulty of issues that are in close classes. However, the system has a hit rate of 92.8% in the classification of questions that are in distant groups.

Keywords: Item response theory — Classification algorithms — ENEM

Resumo: O Exame Nacional do Ensino Médio (ENEM) ganha a cada ano mais importância, conforme vem, aos poucos, substituindo vestibulares tradicionais. Muitos simulados são feitos de forma praticamente aleatória por professores ou sistemas, com questões escolhidas sem critério. Com essa metodologia, caso uma prova precise ser reaplicada, não é possível recriá-la com questões que possuam a mesma dificuldade das utilizadas na primeira avaliação. Neste contexto, o presente trabalho apresenta o desenvolvimento de um Sistema para Geração Inteligente de Simulados do ENEM que realiza os cálculos dos parâmetros da Teoria de Resposta ao Item (TRI) de questões que já foram aplicadas no ENEM e, com base neles, as classifica em grupos de dificuldade, possibilitando, assim, a geração de provas balanceadas. Para isso, foi utilizado o algoritmo *K-means*, para agrupar as questões em três grupos de dificuldade: fáceis, médias e difíceis. Para verificar o funcionamento do sistema, foi gerado um simulado com 180 questões nos moldes do ENEM. Pôde-se perceber que em 37,7% dos casos isso aconteceu. Essa taxa acertos não foi maior pelo fato de o algoritmo confundir a dificuldade de questões que estão em classes próximas. Entretanto, o sistema possui uma taxa de acertos de 92,8% na classificação de questões que estão em grupos distantes.

Palavras-Chave: Teoria de Resposta ao Item — Algoritmos de Classificação — ENEM

¹ Federal Institute of Minas Gerais, Campus Bambui, Brazil

² Federal University of Minas Gerais, Brazil

*Corresponding author: brunoalbertobambui@gmail.com

DOI: <http://dx.doi.org/10.22456/2175-2745.9240> • Received: 30/04/2019 • Accepted: 24/12/2019

CC BY-NC-ND 4.0 - This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

1. Introdução

Em 1998, o Exame Nacional do Ensino Médio (ENEM) foi criado para avaliar a qualidade de ensino no País. Com base nos resultados, o Ministério da Educação seria auxiliado na elaboração de novas políticas na educação expostas nos Parâmetros Curriculares Nacionais (PCNs). Durante o

período de dez anos (1998 - 2008), a prova foi composta por 63 questões baseadas nos conteúdos do ensino médio. Nesse período, o principal objetivo do exame era contribuir para a melhoria da educação no País [1].

O exame ganhou ainda mais importância em 2004, quando começou a ser utilizado para a concessão de bolsas de estudos integrais e parciais a estudantes de cursos de graduação e

de cursos sequenciais de formação específica em instituições privadas de educação superior. Para isso, é necessário que o candidato se inscreva no Programa Universidade para Todos (ProUni).

No ano de 2009, o ENEM adotou o sistema de Teoria de Resposta ao Item (TRI), que substituiu a Teoria Clássica dos Testes (TCT), o qual considerava somente a quantidade de acertos e erros para o cálculo da nota. Essa mudança foi motivada pelos benefícios que a TRI proporciona, sendo os principais, segundo [2], a melhor avaliação dos indivíduos e a possibilidade de permitir a reaplicação de provas com questões que possuam a mesma dificuldade da primeira.

Em 2010, foi criado o Sistema de Seleção Unificada (SiSU), que utiliza a nota do ENEM com o objetivo de selecionar candidatos para vagas ofertadas por instituições federais e estaduais [1].

Com o passar dos anos, o ENEM ganhou espaço em meio aos tradicionais vestibulares que eram aplicados por cada instituição. Em 1998, cerca de 157.221 pessoas se inscreveram na prova. Já em 2017, foram 7.603.209 pré-inscritos e 6.135.418 confirmados [3]. Ainda assim, atualmente existem poucas ferramentas que utilizam TRI para auxiliar na preparação dos estudantes. Grande parte dos simulados desenvolvidos em escolas e cursinhos preparatórios para o ENEM é feita por professores que montam as provas de maneira aleatória. Entretanto, essa pode não ser uma maneira eficiente de se construir um caderno de provas, já que os parâmetros da TRI de cada questão podem ser calculados, e, com base neles, é possível gerar provas com nível de dificuldade melhor balanceado e, posteriormente, fazer os cálculos dos escores de habilidade.

A TRI produz três parâmetros por questão, sendo (i) a dificuldade da questão, (ii) a chance de um aluno acertá-la ao acaso e, (iii) o seu poder de discriminação. Com isso, algoritmos de classificação podem ser usados para agrupar questões com parâmetros similares [4]. Dessa maneira, vão existir grupos de questões fáceis, médias e difíceis, possibilitando a geração de provas e simulados com base na dificuldade das questões, e não mais aleatoriamente.

Em um ranking sobre educação, divulgado em 2017, o Brasil ficou em penúltimo colocado, o que reforça a importância do desenvolvimento de aplicações voltadas à educação [5]. Além disso, o mercado de educação privada, no Brasil, vem crescendo e aumentando, sempre à procura de soluções para melhorar o ensino [6].

Neste contexto, o presente trabalho propõe o desenvolvimento de um Sistema para Geração Inteligente de Simulados do ENEM que classifique as questões com base nos parâmetros de TRI. Com isso, é possível realizar a criação de cadernos de questões para simulados do ENEM de forma balanceada, e não aleatoriamente. Para esta atividade, pretende-se utilizar como base os parâmetros de Teoria da Resposta ao Item (TRI). Estes serão calculados a partir das estruturas chamadas de “microdados” disponibilizadas pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

(INEP). São públicas as informações de aplicação dos exames de anos anteriores, desde 1998, que podem ser consultadas no site da instituição.

A princípio, pretendeu-se desenvolver e aplicar sistemas *crawlers* com sementes fixas, que tiveram por objetivo recuperar informações de bases de questões do ENEM na *web*. Essas questões levantadas foram associadas aos respectivos cálculos da TRI. Após a construção deste banco de dados, foram então avaliados algoritmos de reconhecimento de padrões, com o intuito de estimar o real nível de dificuldade de uma questão, de forma aglutinada, em um único parâmetro discreto.

Para se ter um indício de funcionamento do sistema, foi aplicado um simulado a todos os estudantes do terceiro ano do ensino médio dos cursos técnicos do IFMG - *Campus* Bambuí, totalizando 135 alunos. A prova seguiu a estrutura do ENEM, ou seja, 45 questões de cada área, somando 180. Logo após cada questão, foi inserida uma pergunta sobre a percepção do participante em relação ao seu nível de dificuldade. Com essas respostas, foi possível verificar se a percepção fornecida pelos alunos coincidia com a do sistema.

O Instituto Federal de Educação, Ciência e Tecnologia do estado de Minas Gerais é uma instituição pública de ensino que oferece educação superior, básica e profissional, de forma pluricurricular. O colégio foi fundado em 1968, na zona rural da cidade de Bambuí, no estado de Minas Gerais [7, 8].

O ENEM é utilizado por 1434 instituições de ensino superior no Brasil para selecionar os candidatos às vagas, sendo o maior vestibular do país [9]. Dado que muitos simulados são gerados de forma aleatória, este trabalho se justifica, pois (i) pode auxiliar na preparação dos alunos para o ENEM, (ii) facilita a geração de simulados, (iii) permite que análises de desempenho dos alunos possam ser geradas e (iv) contribui para os estudos de Teoria da Resposta ao Item aliada a Algoritmos de classificação.

Esta pesquisa está estruturada em cinco seções. Na primeira, foi contextualizado o problema de interesse, apresentada a proposta do pesquisa e especificados os objetivos.

Na segunda seção, fez-se a fundamentação teórica, com o objetivo explicar os conceitos envolvidos com o trabalho proposto. Após a fundamentação, uma revisão bibliográfica de trabalhos correlatos.

A terceira seção tratou da metodologia. Nela são descritos os passos seguidos para o desenvolvimento de cada item necessário para o trabalho. São apresentados detalhes da extração de questões; cálculo da TRI; tamanho das matrizes e convergência; geração da matriz de respostas; pacote *mirt*; análise de dados; aplicação dos algoritmos de classificação; fórmula para calcular a dificuldade de um grupo; e planejamento do experimento.

Na quarta seção, são apresentados os resultados obtidos com o desenvolvimento do trabalho sob a perspectiva da classificação das questões, do simulado desenvolvido e dos índices de desempenho do sistema.

Por fim, na quinta seção, encontram-se as conclusões obtidas durante o desenvolvimento do estudo. São apresentadas

sugestões de trabalhos futuros e as limitações identificadas. Ao final da quinta seção, é apresentada, adicionalmente, a lista de referências bibliográficas utilizadas.

2. Fundamentação Teórica

2.1 Teoria da Resposta ao Item

Segundo [10], a Teoria de Resposta ao Item (TRI) é um conjunto de modelos matemáticos que representam a chance do indivíduo que está respondendo à questão acertá-la com base nos parâmetros desta e da sua habilidade. Ou seja, o modelo é baseado em uma função crescente que indica que, quanto maior o traço latente (habilidade), maior a chance de acerto no item [11].

A TRI surgiu nos anos 1950, embora suas raízes sejam de uma década anterior. O trabalho de [12] comparou parâmetros obtidos pela teoria clássica com a TRI utilizada atualmente. [13] foi um dos pioneiros na utilização da expressão curva característica do item, que é um dos conceitos principais na TRI. [14] introduziu o traço latente, que também se tornou um dos principais conceitos da TRI. Mesmo com as grandes contribuições desses autores, quem mais contribuiu para a TRI moderna foi [15]. Ele elaborou modelos teóricos e métodos para estimar os parâmetros dos itens, utilizando o modelo da ogiva normal [16].

A TRI propõe resolver vários problemas existentes na Teoria Clássica dos Testes (TCT), como a impossibilidade da comparação de resultados de testes aplicados a grupos distintos [2]. Com a utilização da TRI, é possível que duas populações sejam comparadas, desde que as provas tenham itens em comum [11]. Isso é possível de ser feito graças ao modelo teórico baseado nos itens, e não nos testes, em que cada resposta correta apresenta uma informação sobre o traço latente. Assim, a probabilidade de um indivíduo dar determinada resposta depende das suas habilidades e das características da questão, não dependendo das outras questões do teste [2]. Mesmo com todas as vantagens, a TRI não veio para substituir toda a Psicometria Clássica, somente partes dela [16].

Segundo [17], “os primeiros modelos de resposta ao item surgiram na década de 50, e eram modelos em que se considerava que uma única habilidade, de um único grupo, estava sendo medida por um teste em que os itens eram corrigidos de maneira dicotômica”. Esses modelos começaram a ser desenvolvidos com a forma de uma função ogiva normal, sendo, posteriormente, descritos para as funções logísticas, que são mais convenientes e usadas até hoje.

Os modelos logísticos para itens binários são os modelos de resposta ao item mais utilizados. Existem três tipos, que se diferenciam pela quantidade de parâmetros utilizados para descrever o item. Os modelos de um, dois e três parâmetros consideram [17]:

1. somente a dificuldade do item;
2. a dificuldade e a discriminação;

3. a dificuldade, a discriminação e o índice de acerto casual.

O primeiro a desenvolver o modelo unidimensional de dois parâmetros, com base na distribuição normal acumulada foi [15]. Após alguns testes, ele sentiu a necessidade de incorporar um parâmetro que tratasse o problema de um aluno acertar uma questão “chutando”. Com isso, surgiu o modelo de três parâmetros. Posteriormente, substituiu a função ogiva normal pela função logística, que é mais conveniente do ponto de vista matemático, já que não envolve integração e é uma função explícita dos parâmetros e de habilidade.

As funções de probabilidade do tipo logística são as mais utilizadas e dão origem aos modelos logísticos [2]. No presente trabalho, será utilizado o modelo 3PL, que descreve a relação entre a habilidade do indivíduo examinado e a probabilidade de uma resposta correta, considerando três parâmetros: dificuldade, discriminação e casualidade [11]. Ele é um dos modelos mais completos, com isso, os outros podem ser obtidos a partir dele [17]. A Equação 1 apresenta o modelo logístico de três parâmetros.

$$P(X_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (1)$$

com $i = 1, 2, \dots, I$ (I representa os itens propostos para avaliar o traço latente) e $j = 1, 2, \dots, n$ (n representa os elementos que compõem a amostra, podendo ser indivíduos, empresas, etc), onde:

- X_{ij} é uma variável binária, assume o valor um caso o indivíduo j acerte o item i ou zero quando o indivíduo erre [17].
- θ_j refere-se à habilidade do j -ésimo indivíduo.
- $P(X_{ij} = 1|\theta_j)$ representa a chance do indivíduo j com habilidade θ_j acertar o item i .
- b_i mensura a dificuldade do item i , ou seja, itens com um alto valor de b exigem uma proficiência maior do respondente.
- a_i refere-se à capacidade que o item i possui para diferenciar os participantes que dominam a habilidade avaliada no item dos que não dominam.
- c_i mede a probabilidade de o indivíduo acertar o item i sem o conhecimento exigido (“chutando”), conhecido como acerto casual.
- D , segundo [17], “é um fator de escala. Utiliza-se o valor 1,7 quando se deseja que a função logística forneça os resultados semelhantes aos da função ogiva normal”.

No ano de 2009, a TRI foi implantada no cálculo das notas do ENEM, substituindo a Teoria Clássica dos Testes (TCT),

que faz o cálculo das notas levando em consideração a quantidade de acertos. Os fatores considerados pelo Ministério da Educação (MEC), para adoção da TRI no ENEM, foram a possibilidade de comparação de resultados entre os anos, a possibilidade de reaplicação do exame com questões com as mesmas dificuldades das utilizadas na primeira avaliação e o fato de as notas calculadas com TRI selecionarem de forma mais adequada os alunos. Em uma prova de 45 questões, por exemplo, com a utilização de TCT, existem 46 possibilidades de notas. No caderno de Matemática do ENEM de 2012, com a mesma quantidade de questões, foram obtidas 6406 notas distintas (sendo que 4.049.876 candidatos realizaram a prova) [18].

Diante das vantagens que a TRI proporciona, serão apresentados, na próxima seção, conceitos de traço latente e curva característica do item. Eles são importantes para o entendimento da probabilidade que um indivíduo possui de acertar um item.

2.2 Traço Latente

O traço latente trata das características de cada indivíduo, que não podem ser determinadas de forma direta. Existem vários exemplos de traço latente, como nível de satisfação, conhecimento, habilidade, grau de depressão, etc. Essas características são medidas através de fatores secundários que tenham relação com o traço latente estudado [19].

A relação existente entre o desempenho em um teste e o traço latente pode ser descrita por uma equação crescente, chamada Curva Característica do Item (CCI). Na Figura 1, é possível observar que indivíduos com uma habilidade (aptidão) maior terão uma probabilidade maior de acertar o item [20].

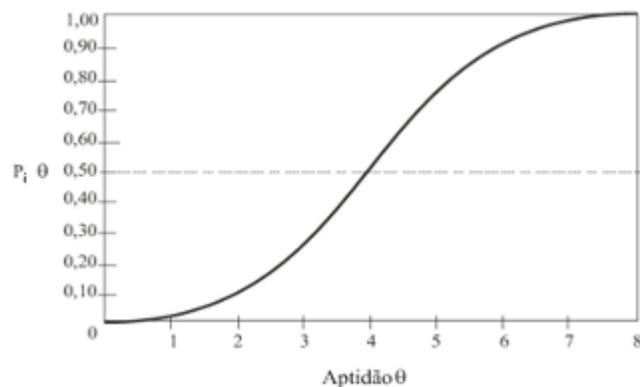


Figura 1. Curva característica do item

Segundo [10], a escala de habilidade é “uma escala arbitrária em que o importante são as relações de ordem existentes entre seus pontos, e não necessariamente sua magnitude”. O parâmetro b é medido na mesma unidade da habilidade, e o parâmetro c , que representa uma probabilidade, não depende de nenhuma escala, assumindo valores entre zero e um [10].

O parâmetro a é proporcional à derivada tangente da curva no ponto da inflexão. Com isso, questões com a negativo não

são esperadas, já que isso indicaria que a probabilidade de acertar a questão diminui com o aumento da habilidade. Pequenos valores de a apontam que a questão tem pouco poder de discriminação, ou seja, alunos com habilidades muito diferentes possuem a mesma probabilidade de acertar a questão [10]. Valores altos de c também não são esperados, uma vez que indicam a chance de o indivíduo acertar a questão sem o conhecimento necessário. Já o parâmetro b é referente à dificuldade da questão. A Figura 2 apresenta a relação entre os parâmetros da TRI com a CCI.

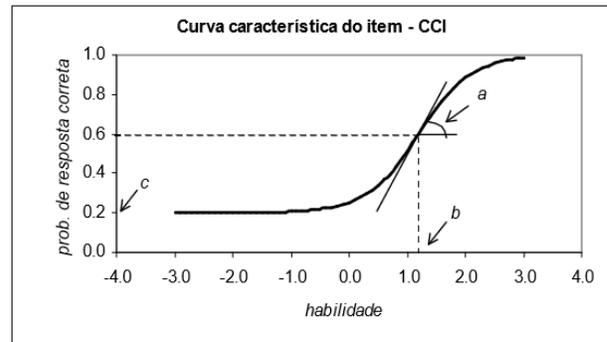


Figura 2. Curva característica do item

A função de informação do item é uma medida que é comumente utilizada em conjunto com a CCI. Ela permite analisar a capacidade que uma questão possui para medir a habilidade [10]. Esta função é dada por:

$$I_i(\theta) = D^2 a_i^2 \frac{Q_i(\theta)}{P_i(\theta)} \left[\frac{P_i(\theta) - c_i}{1 - c_i} \right]^2 \quad (2)$$

onde

- $I_i(\theta)$ é a informação do item i no nível de habilidade θ ;
- $P_i(\theta) = P(X_{ij} = 1|\theta)$, representa a probabilidade de o indivíduo j responder corretamente o item i , dado que ele tenha habilidade θ ;
- $Q_i(\theta) = 1 - P_i(\theta)$.

Esta função mostra a importância que cada parâmetro tem para a informação do item [10]. Isto é, a informação é maior:

- quando b_i se aproxima de θ ;
- quanto maior for o a_i ;
- e quanto mais c_i se aproximar de 0.

A informação que é fornecida pelo teste pode ser obtida com a soma das informações fornecidas por cada questão que o compõe, isto é:

$$\sum_{i=1}^I I_i(\theta) \quad (3)$$

Na próxima seção, é apresentado o conceito de reconhecimento de padrões. Ele será importante para o entendimento das classificações que serão feitas nas questões, após o cálculo da TRI.

2.3 Reconhecimento de padrões

O reconhecimento de padrões (RP) é uma tarefa da qual os humanos dependem ao longo dos anos para sobreviver. É algo que trouxe vantagem competitiva no meio natural, e, por isso, sabe-se que seu uso pode resolver diversos tipos de problemas e ser útil nas mais variadas situações. O reconhecimento de padrões, por suas vantagens, tem sido utilizado nos últimos anos como estratégia para melhorar o desempenho de diversas máquinas. É um subtópico da aprendizagem de máquina, que tem como objetivo classificar objetos em categorias. Para isso, análises estatísticas das aglomerações de dados são realizadas [21, 22]. RP pode ser aplicado a processamento de sinais e imagens, sistemas adaptativos, mineração de dados e outros contextos em que existam informações disponíveis e passíveis de análise [23].

Realizar o reconhecimento de padrões em determinados *datasets* pode não ser uma tarefa trivial, por isso foram criados os algoritmos de classificação. Eles têm como objetivo dividir os elementos da amostra em grupos (*clusters*), de forma que os elementos de um mesmo grupo tenham características semelhantes entre si [4, p.155].

Um dos principais obstáculos para a classificação é a escolha do melhor algoritmo e a quantidade de grupos. Não existe um valor correto para essa questão; no entanto, existem, na literatura, trabalhos que podem auxiliar na escolha [19]. Segundo [4], quando os dados já possuem uma partição no espaço em relação às variáveis selecionadas, os resultados de agrupamentos não são muito diferentes, mesmo com métodos distintos de classificação. Entretanto, quando os grupos sofrem muitas interseções entre si, os resultados se tornam diferentes, de acordo com os parâmetros e o método escolhido.

As redes neurais são um exemplo de algoritmo que pode ser utilizado na tarefa do reconhecimento de padrões. Elas se baseiam em um processo prévio de aprendizagem e nos dados fornecidos para encontrar um padrão ou resposta que corresponda ao esperado. As suas principais vantagens são: capacidade de generalização, adaptatividade e a tolerância a falhas [24, 25].

A matriz de confusão é uma forma de verificar o desempenho do classificador, verificando se o valor previsto pelo algoritmo corresponde ao esperado. As linhas da matriz representam os valores reais (*Actual*), e as colunas, os valores previstos (*Prediction*) pelo algoritmo [26].

Na próxima seção descreve-se o conceito de banco de dados, os quais serão úteis para fazer o gerenciamento das questões e dos dados relacionados a elas.

2.4 Banco de Dados

Segundo [27, p. 10], “um banco de dados é uma coleção de dados persistentes, usada pelos sistemas de aplicação de uma

determinada empresa. O termo empresa, aqui empregado, é simplesmente genérico para qualquer organização comercial, científica, técnica ou outra organização razoavelmente autônoma”. Nesse sentido, um banco de dados é um conjunto de dados que estão relacionados.

Os bancos de dados são importantes no funcionamento de várias tarefas do cotidiano. Realizar um depósito bancário, fazer reserva em hotéis, acessar o catálogo de uma biblioteca virtual ou comprar produtos pela *internet* são tarefas que vão envolver uma pessoa ou uma aplicação que vai utilizar-se de banco de dados [28].

As principais vantagens da sua utilização são:

1. Controle centralizado dos dados: os dados estão localizados em um único local, e isso proporciona maior controle [29].
2. Controle da redundância: faz o controle para que uma mesma informação não seja salva em vários registros [29].
3. Eliminação de inconsistências: sem um banco de dados, pode ocorrer a repetição de um arquivo, o qual apresenta valores divergentes. Nesse caso, os dados estão inconsistentes, e, se falta consistência, comumente falta integridade [29].

Existem bancos de dados que podem ser gerados e mantidos de forma manual ou computadorizada. Um catálogo de fichas bibliotecárias pode ser mantido manualmente por um funcionário. Entretanto, há grandes bancos que necessitam de um sistema computadorizado para fazer o seu gerenciamento. Foram criados, então, os Sistemas de Gerenciamento de Banco de Dados (SGBD). O SGBD é um grupo de programa que permite ao desenvolvedor a construção, a manipulação e a manutenção do banco de dados [28].

Neste trabalho, foi desenvolvido um banco de dados de questões do ENEM, com a utilização de um *crawler*, cujo conceito é explicado na seção a seguir.

2.5 Crawler

Um *crawler* é um sistema que navega em *websites* de forma sistêmica, sendo capaz de recuperar e inserir informações em páginas da *internet* segundo a necessidade do usuário [30]. Ele é utilizado pelos motores de busca para navegar em páginas *web*, criando cópias destas para realizar indexações, o que proporciona buscas mais rápidas aos usuários. É comumente utilizado, também, para tarefas de manutenção em *sites*, como verificação de *links* [31]. Existem *crawlers* de semente (lista de endereços a serem visitados) fixa e não fixa; isso significa que, caso seja fixa, o *crawler* navegará sobre um único site.

Grandes empresas utilizam *crawlers* para desempenhar as tarefas citadas acima. A Google, por exemplo, desenvolveu o Googlebot, que utiliza um conjunto de computadores para buscar páginas na *web* e incluí-las no banco de dados da empresa [32]. Neste trabalho, *crawlers* foram usados junto aos microdados do ENEM para cálculo dos parâmetros da TRI, explicados a seguir.

2.6 Microdados do ENEM

O INEP, visando dar publicidade e transparência à aplicação do ENEM, desde 1998, vem disponibilizando informações sobre a aplicação do exame, na forma de uma base de dados chamada de “microdados”.

Os microdados, entre outras informações, contém diversos conteúdos sobre variáveis econômicas, sociais, respostas de candidatos, notas obtidas e outras. Os arquivos estão disponíveis no formato ASCII, e, para fazer a sua leitura, é recomendado o uso dos programas SAS e SPSS. Eles podem ser acessados por meio da página do INEP na *internet* [33].

Para o desenvolvimento deste trabalho, existem alguns dados presentes nos microdados que possuem maior significância, sendo eles o gabarito de respostas do candidato, o código de identificação da prova e as informações sobre a presença do candidato nos dias de aplicação da prova. Esses são importantes para a geração da tabela de respostas dos participantes, ou seja, quais questões cada candidato acertou ou errou. Essa tabela será utilizada, posteriormente, para o cálculo dos parâmetros da TRI de cada questão.

2.7 Estado da Arte

Nesta seção, são apresentados trabalhos que contribuíram para o desenvolvimento e a fundamentação deste estudo. Foi realizado um levantamento dos pontos principais de cada um, assim como a conclusão obtida por cada autor.

No trabalho de [34], realizou-se um estudo sobre a quantidade mínima de amostras necessária para a utilização da Teoria de Resposta ao Item. Para isso, foram criadas nove amostras de tamanhos diferentes, a partir de uma base de 44 mil respostas de estudantes. Segundo os autores, é possível concluir que, amostras com 100 pessoas são instáveis, tornando-se estáveis com amostras a partir de 200 pessoas. Com um *dataset* de 500 sujeitos, é possível obter um resultado bem mais próximo em comparação com a amostra completa, sendo que, com 20 mil estudantes, são muito próximos aos resultados obtidos com a amostra total.

Existem alguns programas que podem ser utilizados para o cálculo da TRI, como o BILOG-MG [35], que é um sistema licenciado, e o R, que é um *software* livre. [36] desenvolveram uma comparação entre os resultados obtidos com o emprego desses sistemas. Para isso, utilizaram 25 questões de matemática e português, que já tinham seus parâmetros calibrados. Para a realização dos cálculos, foram escolhidos 364 alunos da 5ª e 9ª séries, de forma aleatória, que tinham suas habilidades conhecidas. Com essa habilidade, foi calculada a probabilidade de cada aluno acertar as questões, e, após esse processo, foi utilizada a distribuição de Bernoulli para gerar a provável resposta de cada aluno para cada questão. As habilidades calculadas pelos sistemas foram parecidas. Ocorreram algumas divergências na estimação de alguns parâmetros, mas, na grande maioria, as diferenças foram pequenas. Diante disso, o R foi considerado uma alternativa válida para a realização de análises de TRI.

Foi realizado por [2] o cálculo dos parâmetros de TRI

com a utilização do pacote *ltm* [37] do R, com a máxima verossimilhança e com o método “dos intervalos”, empregando os microdados disponibilizados pelo INEP. Segundo a conclusão dos autores, o método dos intervalos é o mais simples para ser implementado, entretanto é menos preciso. O método da máxima verossimilhança foi o mais eficiente para reproduzir as notas do ENEM. Por fim, o pacote *ltm* foi o que conseguiu o melhor resultado na estimação dos parâmetros da TRI.

No trabalho de [38] os autores selecionaram a base de dados da SARESP para realização de um estudo. Essa base possui uma amostra de três testes (manhã, tarde e noite) de Língua Portuguesa, que foram aplicados para alunos do ensino médio. No estudo, foi feita a análise com o pacote *ltm*, com o modelo de dois e três parâmetros. Os autores concluíram que, até o término do trabalho, ainda existiam algumas funções que poderiam ser desenvolvidas para facilitar os trabalhos com o R. Não existem, por exemplo, funções para estimação de parâmetros em modelos normais. Além dos pacotes utilizados no estudo, Andrade e Anjos ainda citam outro importante pacote que pode ser utilizado para análise de modelos multidimensionais, o *mirt* [39].

No trabalho de [40] os autores compararam criticamente as características da Teoria Clássica dos Testes (TCT) e da Teoria de Resposta ao Item (TRI), sendo discutidas as vantagens e as limitações de cada método. Os autores relataram, em suas conclusões, que o uso da TRI vem crescendo, mas a TCT ainda é muito utilizada, mesmo com as suas limitações. Disseram que isso pode ser explicado pelo fato de cursos de graduação e até de pós-graduação ainda utilizarem esse método para avaliação dos alunos, já que a TCT é mais fácil de ser utilizada e entendida.

O trabalho de [1] aplicou questionários a alunos do ensino médio para avaliar a dificuldade que eles possuem na resolução das questões da prova de Matemática e suas Tecnologias, do ENEM de 2015. O questionário foi aplicado em uma escola estadual na cidade de Manhuaçu, Minas Gerais. Contabilizou-se um total de 16 alunos dispostos a responder as perguntas, sendo que 13 classificaram as questões como sendo difíceis. Os autores concluem, a partir das análises feitas, que 81,25% dos alunos classificam as questões de matemática do ENEM de 2015 com um nível de dificuldade elevado.

O artigo de [41] avalia o desempenho dos alunos de escolas públicas e privadas da região sul fluminense no ENEM de 2015 por meio da técnica *K-means*. As escolas foram classificadas de acordo com o desempenho dos alunos em função das variáveis indicadoras de proficiência média em cada área da prova. Os resultados mostraram que os dois grupos separados foram caracterizados em função do porte da escola, município, dependência administrativa e indicador socioeconômico.

A pesquisa de [42] teve como objetivo apresentar e analisar algumas técnicas de clusterização como uma possibilidade de pesquisa na área do ensino. Portanto, foram apresentadas formas técnicas de análises de cluster bem como algumas de suas aplicações em pesquisas educacionais encontradas na

literatura. A análise de cluster é adequada para simplificar ou elaborar hipóteses sobre dados massivos, como pesquisas educacionais em larga escala. Os estudos na área da educação que usaram métodos de análise de cluster provaram ser frutíferos para obter resultados que colaboram com a área.

Diversos trabalhos na literatura relatam experimentos em relação ao TRI. Na pesquisa desenvolvida foi levantado aspectos em relação ao TRI aplicado em um exame avaliativo em nível nacional, o ENEM. A utilização de um algoritmo de clusterização, o *K-means*, possibilitou o desenvolvimento de um Sistema para Geração Inteligente de Simulados do ENEM, de modo que as instituições de ensino possam vir a utilizar o sistema proposto para a geração de provas balanceadas.

3. Metodologia

Nesta seção, são apresentados os métodos e procedimentos para o desenvolvimento de um Sistema para Geração Inteligente de Simulados do ENEM.

A sequência de apresentação da seção será a seguinte: modalidade de pesquisa, materiais utilizados, extração de questões, tamanho das matrizes e convergência, geração da matriz de respostas, cálculo da TRI com o pacote *mirt*, análise e classificação dos dados, desenvolvimento da fórmula para calcular a dificuldade de um grupo, planejamento do experimento e, por fim, coleta dos resultados. A Figura 3 apresenta os passos seguidos para o desenvolvimento do trabalho.

3.1 Modalidade de pesquisa

O presente trabalho é uma pesquisa quantitativa, de natureza aplicada. Ela é caracterizada por produzir conhecimento prático, que possa ser utilizado na sociedade. De acordo com [43] “a pesquisa aplicada é aquela em que o pesquisador é movido pela necessidade de conhecer para a aplicação imediata dos resultados. Contribui para fins práticos, visando a solução, mais ou menos imediata, do problema encontrado na realidade”.

Quanto ao objetivo, trata-se de uma pesquisa exploratória. Segundo [44], a pesquisa exploratória permite uma maior familiaridade entre o pesquisador e o tema pesquisado, visto que este ainda é pouco conhecido, pouco explorado. Em relação aos procedimentos técnicos, este estudo se configura como uma pesquisa-ação. Segundo [45], “é um estudo baseado em dados concretos, que procura resolver um problema”.

De acordo com [46] a coleta de dados é extremamente importante em pesquisas quantitativas, de modo que o pesquisador colete tais dados preocupando-se com os seguintes requisitos: validade, confiabilidade e precisão. Outro ponto citado em [46] é sobre o objetivo que a pesquisa quer alcançar, respondendo a certas hipóteses elaboradas. Uma vez que os dados foram coletados, trata-se de verificar se essas informações correspondem às hipóteses; ou seja, se os resultados observados correspondem aos resultados esperados pelas hipóteses ou questões da pesquisa.

3.2 Extração de questões

Para o desenvolvimento do trabalho proposto, foi necessária uma base de questões do ENEM, para realizar os respectivos cálculos de TRI e a classificação. Com isso, pôde-se gerar os cadernos inteligentes de prova. Entretanto, após buscas na *internet*, não foi encontrado nenhum banco de questões que atendessem às necessidades do projeto. Diante dessa dificuldade, realizaram-se pesquisas sobre como montar uma base própria de questões.

Primeiramente, foi desenvolvido um sistema para realizar a extração das questões de provas passadas do ENEM, que são disponibilizadas no formato *Portable Document Format* (PDF) pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Entretanto, após o desenvolvimento de uma parte do sistema extrator, ele estava apresentando problemas, como: dificuldade na captura das imagens e caracteres sem formatação. Diante disso, desenvolveu-se uma estratégia para extração das questões, o *crawler*. Essa metodologia foi adotada utilizando-se o site Olhonavaga [47], que contém dados sobre as questões do ENEM no período de 2005 a 2015. Para automatizar o uso do *crawler*, usou-se o Selenium [48], a linguagem de programação Python 3 [49] e o Geckodriver [50], que é de um driver responsável por fazer a comunicação entre o Selenium e o navegador.

O Selenium é um conjunto de ferramentas para automatizar navegadores. Ele possui métodos para fazer a captura de dados que estão presentes em uma página *web*, como o *Find_element_by_css_selector*. Para isso, ele recebe como parâmetro o seletor *HyperText Markup Language* (HTML) que o dado-alvo de extração possui. Então, foi feita uma análise no *site*, e todos os dados de interesse tiveram seus seletores capturados. De posse desses, foi necessário desenvolver a lógica do comportamento de mudanças das páginas do site, para que fosse possível percorrer todas as questões do período citado, de forma autônoma, fazendo a extração dos dados.

Verificando o *site*, foi possível observar que, quando um usuário clica no botão de próxima questão, é incrementado um contador no link da página. Com isso, foi implementada uma lógica que, com a finalização da extração dos dados presentes na página atual, um contador fosse incrementado. Este é concatenado com um *link*-padrão da página, ou seja, que não se altera, independentemente da página. O *link* gerado é passado para o método *GET* do Selenium, que é responsável por fazer a abertura de páginas; assim, uma nova página é carregada. A Figura 4 apresenta o diagrama de atividades do *crawler*.

Com a execução do *crawler* desenvolvido foi gerado um banco com 1634 questões do período de 2005 a 2015 do ENEM. Na próxima seção, é apresentada a metodologia de desenvolvimento do banco de dados para fazer o gerenciamento dessas questões.

3.3 Banco de dados

Com a captura dos dados das questões foi necessário fazer a sua inserção em um banco de dados. Para isto, empregou-se o

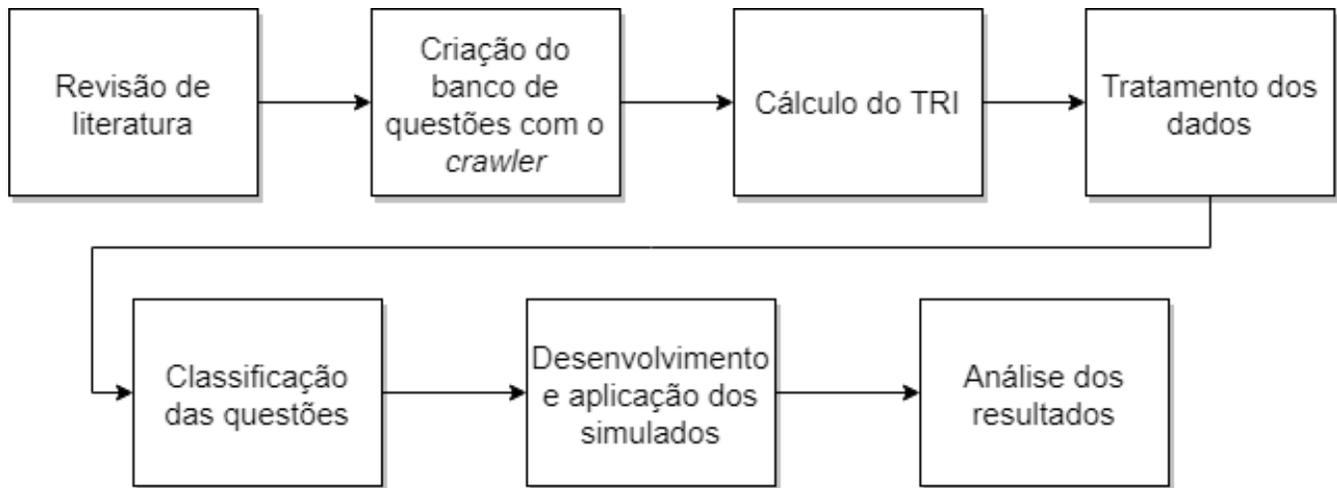


Figura 3. Esquema da metodologia utilizada

SQLAlchemy [51], essa é uma biblioteca desenvolvida em Python de mapeamento objeto-relacional. Com ela, os dados de cada questão podem ser salvos no banco de dados, após a sua extração. Nessa biblioteca, o banco é representado na forma de classes, sendo que cada tabela tem a sua. Foram criadas quatro classes: Alternativa, Área, Idioma e Questão.

As imagens presentes nos enunciados e alternativas foram salvas diretamente no arquivo de sistema, com a utilização da biblioteca Requests [52] salvando-se apenas o caminho relativo no banco de dados. Por fim, um passo importante no desenvolvimento do *crawler* foi a criação da *engine* responsável por realizar a conexão entre o *crawler* e o banco de dados. O Sistema de Gerenciamento de Banco de Dados (SGBD) selecionado para o trabalho foi o MySQL [53]. A Figura 5 mostra o diagrama de entidade-relacionamento correspondente.

3.4 Cálculo da TRI

Nesta seção, são demonstrados os procedimentos realizados para o desenvolvimento do cálculo da TRI. Essa etapa foi necessária após contatos com o Sistema Eletrônico de Serviço de Informação ao Cidadão (e-SIC), em que foi feita a requisição dos cálculos de TRI, realizados pelo INEP. Entretanto, em todas as solicitações, o acesso aos dados foi rejeitado, com a justificativa de que os eles pudessem ser acessados somente com acesso local na sede da instituição em Brasília, após requisição e aprovação pelos órgãos competentes. Diante das dificuldades desse deslocamento, optou-se pela realização dos cálculos.

Os microdados disponibilizados pelo INEP, referentes ao ENEM, possuem muitos campos de dados. Aliado a isso, milhões de estudantes fazem a prova, o que resulta em arquivos com um tamanho elevado. No ano de 2015, por exemplo, a planilha com os dados necessários para o trabalho possui aproximadamente cinco GB. Com isso, as funções tradicionais do R para leitura de arquivos não estavam demonstrando bom desempenho.

Dessa forma, foi feita uma nova tentativa, com o uso da biblioteca FF [54]. Ela é recomendada para fazer a leitura de grandes quantidades de dados, já que, após a leitura do arquivo, este se comporta quase que da mesma forma como se estivesse completamente na memória RAM; porém, uma parte ainda fica armazenada no disco.

Após a leitura do arquivo, foram removidas colunas que não eram de interesse do trabalho, como dados sociais e econômicos. Assim, o tamanho do arquivo diminuiu consideravelmente, o que gerou melhorias no desempenho dos *scripts*.

No trabalho desenvolvido por [34], foram realizados testes para averiguar o tamanho mínimo da amostra para convergência no sistema de TRI. Isso se torna importante pelo fato do custo computacional para o cálculo da TRI aumentar de acordo com tamanho da base. No ENEM de 2012, cerca de 5.791.065 participantes realizaram o exame, e caso fosse feito o cálculo levando em consideração toda a amostra, o tempo gasto seria grande. No trabalho de [2], foram necessárias 45 horas para calcular o TRI referente ao ano de 2012, com a utilização da função *tpm*.

No trabalho de [34] os autores demonstraram que não é preciso utilizar toda a amostra de dados para o cálculo da TRI. Eles concluíram que amostras a partir de 200 indivíduos já começam a se tornar estáveis. Então, para o presente trabalho, foram geradas três amostras de tamanhos diferentes de dados: 30, 40 e 50 mil indivíduos, no intuito de escolher aquela que atendessem melhor a esses critérios. O critério de exigência para escolha da amostra tratou-se da observação da oscilação entre o desvio padrão e a média de uma amostra com tolerância estipulada em valores menores que 0.01. Após os testes, verificou-se que a amostra de 30 mil já atendia aos requisitos.

Para evitar algum tipo de tendenciamento na escolha dos dados foi desenvolvido um código para realizar o sorteio de 30 mil candidatos. Em seguida verificou-se cada índice sorteado, para conferir se não existiam indivíduos repetidos e se o indivíduo compareceu em todos os dias da prova, além

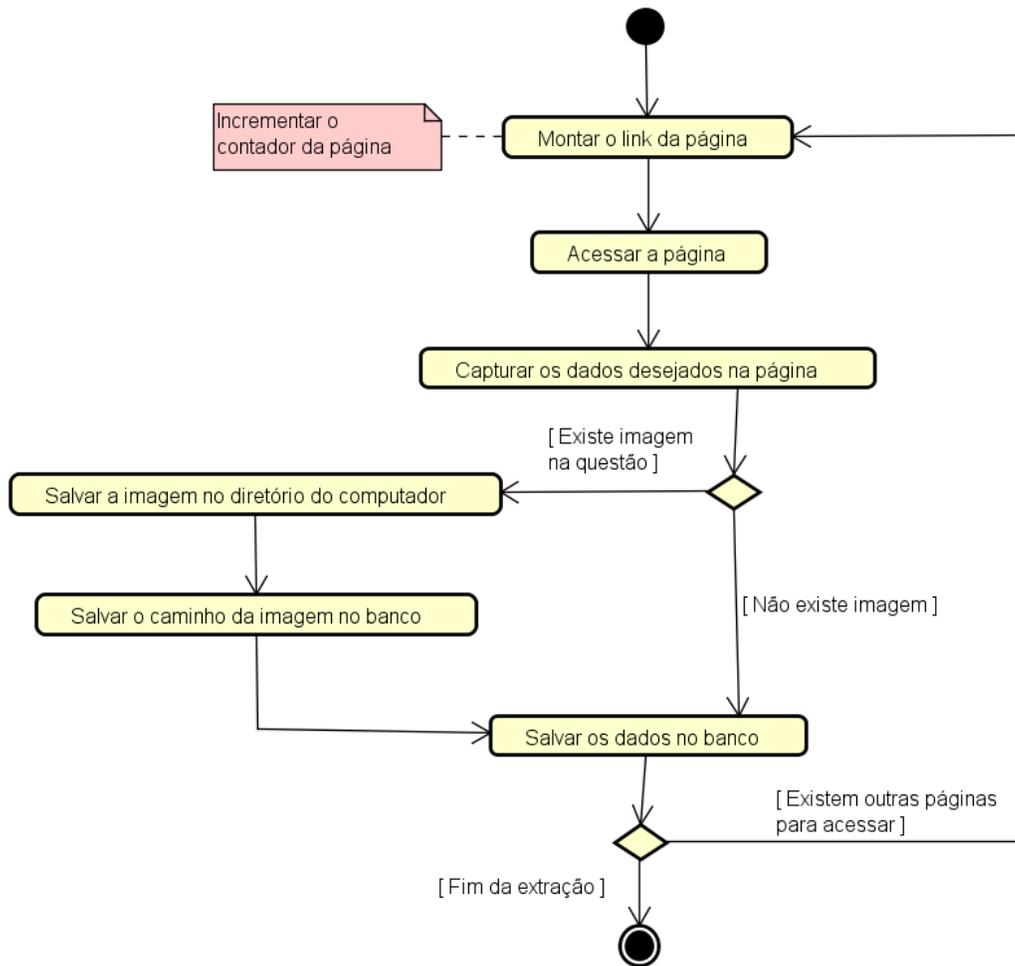


Figura 4. Diagrama de atividades do crawler

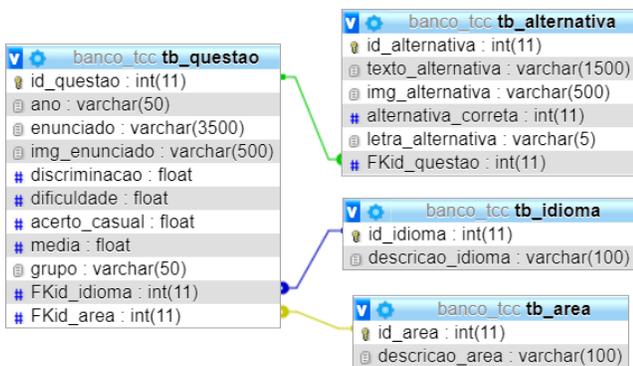


Figura 5. Estrutura do banco

de não ter sido desclassificado do exame. Caso alguma das premissas anteriores estivessem quebradas, um novo índice seria sorteado. A Figura 6 apresenta o diagrama de atividades para seleção da amostra de dados.

Definido o tamanho da amostra, será necessário montar a matriz de respostas dos indivíduos que fizeram a prova.

Os pacotes para cálculo de TRI existentes no R, atualmente na versão 3.4.3, exigem como parâmetro uma matriz de respostas com 0's e 1's, que representam, respectivamente, uma resposta incorreta e correta. Porém, nos microdados disponibilizados pelo INEP, as respostas dos indivíduos estão na forma alfabética, de A a E.

O ENEM possui quatro cadernos de provas, de cores e ordem de questões diferentes. Cada caderno possui um código, que é uma coluna nos microdados. Com isso, é necessário criar uma tabela de equivalência entre as questões, já que na matriz de respostas é necessário que, na coluna 1, por exemplo, estejam todas as respostas dessa questão. Sem essa equivalência, as respostas dos indivíduos ficariam desordenadas entre as colunas, inviabilizando os cálculos.

Então, de posse das equivalências, foi desenvolvido um script em R para fazer a conversão do gabarito presente nos microdados, da forma alfabética para a numérica. Para isso, foram comparadas as respostas dadas pelos indivíduos com os gabaritos das provas. Caso o indivíduo acertasse a questão, adicionava-se o valor um na matriz de respostas, na linha correspondente ao indivíduo e à coluna da questão. O pseu-

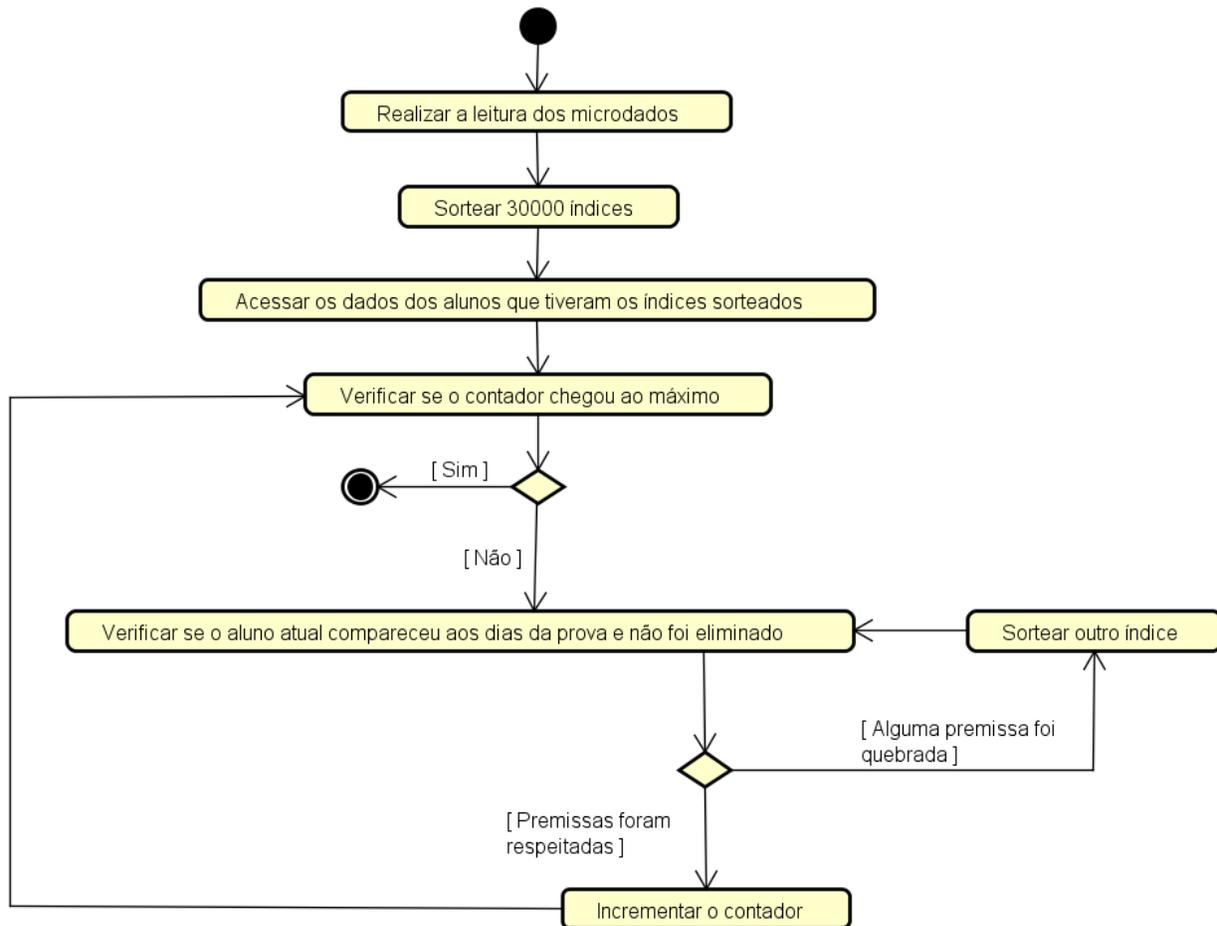


Figura 6. Diagrama de atividades para seleção da amostra

docódigo apresenta a metodologia utilizada para geração das matrizes de respostas.

Finalizada a etapa de matriz de respostas, em seguida, será definida a metodologia utilizada para o cálculo da TRI, com a utilização do pacote *mirt*.

O *mirt* é um pacote do R, para aplicação da TRI. Para isso, ele necessita da matriz de respostas e da quantidade de parâmetros a serem calculados. No caso deste trabalho, 3PL, foram calculados os parâmetros de discriminação, dificuldade e acerto casual das questões. Para realização dos cálculos, desenvolveu-se um *script* em R, que faz, basicamente, a chamada da função *mirt* e salva os resultados em um arquivo-texto. O ENEM, atualmente, possui quatro áreas de conhecimento, e, para cada uma, foi feito o cálculo de forma independente, inclusive para as questões de língua estrangeira.

Com a finalização dos cálculos dos parâmetros da TRI, foi possível fazer as classificações e a análise de dados.

3.5 Análise de dados

Nesta seção, são discutidas as metodologias utilizadas para tratamento e análise dos dados gerados a partir dos cálculos de TRI. Essa etapa de preparação é requisito para a classificação

das questões.

Segundo [55], *outliers* são dados que se diferenciam drasticamente dos outros, são pontos fora da curva. Ou seja, são valores que fogem da normalidade e podem causar problemas nas análises. Eles também são conhecidos como pontos fora da curva, valores atípicos, anomalias, etc.

Na análise de dados, após o cálculo dos parâmetros da TRI, foi possível observar a presença de alguns *outliers*, os quais se tratavam de parâmetros “fora do padrão” dos demais. Isso pode ser causado por uma questão muito fácil ou difícil, por exemplo, fazendo com que ela venha a apresentar um alto ou um baixo valor no parâmetro *b* (dificuldade da questão).

A remoção desses *outliers* pode acarretar problemas, visto que uma questão com alto índice de erros no seu ano de aplicação poderia ser removida, podendo esta ser de interesse dos usuários do sistema. Diante disso, não foi feita a exclusão desses pontos, sendo realizada apenas a normalização.

A normalização consistiu em padronizar os dados em valores entre zero e um. Isso foi importante durante a etapa de classificação, já que os dados originais de TRI, gerados pela biblioteca *mirt*, possuem escalas diferentes de magnitude, podendo acarretar problemas com o algoritmo de classificação.

Então, foi desenvolvida uma função na linguagem R para fazer a normalização dos dados. A fórmula utilizada na normalização é apresentada na Equação 4:

$$z_i = \frac{s_i - \min(s)}{\max(s) - \min(s)} \quad (4)$$

em que

- z_i é o número normalizado;
- s_i é o número a ser normalizado;
- $\min(s)$ retorna o menor valor do conjunto de dados;
- $\max(s)$ retorna o maior valor do conjunto de dados.

3.6 Aplicação de IA para clusterização

Para fazer a geração dos cadernos de prova, de forma inteligente, é necessário saber o grupo a que cada questão pertence. Para isso, foi utilizado um algoritmo de classificação, que com base nos parâmetros de TRI, calculados e normalizados, demonstrou capacidade de agrupar as questões com características semelhantes em um mesmo grupo.

Um dos algoritmos capazes de fazer essa classificação é o *K-means* [56]. Ele trabalha com centroides iniciais aleatórios, correspondentes a uma quantidade K de grupos esperados. Dados que estão mais próximos dos centroides gerados são relacionados a ele, sendo a distância entre eles calculada. É gerada, então, uma média dessas distâncias, e esse valor será a nova posição do centroide, podendo ocasionar em mudanças no grupo ao qual um ponto pertence. O algoritmo executa até não haver mais mudanças de centroide por parte dos pontos [57].

O *K-means* foi utilizado com o objetivo de realizar a tarefa da classificação das questões. O mesmo foi escolhido uma vez que se trata de um algoritmo de fácil implementação e não supervisionado. No trabalho, empregou-se um K igual a três, correspondente aos três grupos de dificuldade já discutidos.

Finalizadas as classificações, é importante conhecer a dificuldade das questões presentes em cada grupo. Na próxima seção, é apresentada a equação desenvolvida para essa finalidade.

3.7 Desenvolvimento da fórmula para calcular a dificuldade de um grupo

Com a classificação realizada pelo *K-means*, as questões foram divididas em três grupos. Entretanto, não tinha como saber qual era o de questões fáceis, médias ou difíceis. Para descobrir a dificuldade referente a cada grupo, desenvolveu-se uma fórmula que trabalha com o centróide de cada grupo de questões. Esses são pontos gerados aleatoriamente no início da execução do *K-means*, e, com a sua execução, são trocados de posição, que é a distância média dos pontos que se ligam ao centróide.

A fórmula desenvolvida para calcular a dificuldade de cada grupo de questões, com base nos parâmetros dos centróides, foi a seguinte:

$$Dificuldade = 20 * a + 60 * b - 20 * c \quad (5)$$

Os parâmetros a e b são proporcionais à dificuldade de uma questão. Já o parâmetro c é inversamente proporcional, já que valores altos de c representam uma chance maior de o indivíduo acertar a questão “chutando”.

Com a finalização da etapa de classificação das questões, é apresentada, na próxima seção, a metodologia utilizada para realização do experimento de validação dos grupos de questões.

3.8 Planejamento do experimento e aplicação do questionário de percepção

Para validação da classificação realizada pelo algoritmo foi aplicado um simulado preparatório para o ENEM aos estudantes dos cursos técnicos do IFMG – *Campus Bambuí*. As provas geradas passaram por uma análise prévia dos professores da instituição.

Hoje, o ENEM está dividido em quatro áreas do conhecimento: Linguagens, Códigos e suas Tecnologias; Ciências Humanas e suas Tecnologias; Ciências da Natureza e suas Tecnologias; e Matemática e suas Tecnologias. No simulado, constaram questões dessas áreas, totalizando 180. A prova foi aplicada em duas sextas-feiras, visando a um comparecimento maior dos alunos.

O Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), que é o órgão responsável pela elaboração da prova do ENEM possui um banco de 10 mil questões, as quais são desenvolvidas por professores de diversas universidades do país, sempre passando por uma revisão do INEP. As questões possuem um texto-base com cinco alternativas e uma resposta, relacionadas com conteúdos abordados no ensino médio [58].

As questões que compõem a prova do ENEM são escolhidas através de sorteio. Antes de serem aplicadas na prova, todas passam por um teste com alunos do 1º e 2º anos do ensino médio. Esse processo é sigiloso, e os alunos não sabem que estão testando possíveis questões que estarão no ENEM. Com esse teste, o INEP consegue mensurar a dificuldade das questões, as que possuem altos índices de acertos ou erros são descartadas, e classificando o restante em fáceis, médias e difíceis. A prova é composta por 25% de questões fáceis, 50% de médias e 25% de difíceis [58].

No simulado desenvolvido, não foi utilizada a metodologia seguida pelo ENEM no balanceamento das dificuldades das questões. Optou-se por colocar 33,33% de questões fáceis, 33,33% de médias e 33,33% de difíceis.

O trabalho de [1] utilizou questionários para verificar a dificuldade que os alunos possuem com a área de Matemática e suas Tecnologias no ENEM. Com isso, para analisar indícios sobre o funcionamento do sistema, foi acrescentado, após cada questão do simulado, um questionário de percepção, no

qual o aluno informou a sua opinião em relação à dificuldade da respectiva questão. Entretanto, é importante fazer algumas considerações sobre as limitações quanto à utilização desse questionário:

- Não é possível garantir que os indivíduos respondam ao questionário com sinceridade;
- a percepção de dificuldade é influenciada por: fatores fisiológicos e psicológicos; habilidade de cada indivíduo; motivação do aluno, etc;
- as grandezas comparadas são distintas.

Questão 156

Um estudante está pesquisando o desenvolvimento de certo tipo de bactéria. Para essa pesquisa, ele utiliza uma estufa para armazenar as bactérias. A temperatura no interior dessa estufa, em graus Celsius, é dada pela expressão $T(h) = -h^2 + 22h - 85$, em que h representa as horas do dia. Sabe-se que o número de bactérias é o maior possível quando a estufa atinge sua temperatura máxima e, nesse momento, ele deve retirá-las da estufa. A tabela associa intervalos de temperatura, em graus Celsius, com as classificações: muito baixa, baixa, média, alta e muito alta.

Intervalos de temperatura (°C)	Classificação
$T < 0$	Muito baixa
$0 \leq T \leq 17$	Baixa
$17 < T < 30$	Média
$30 \leq T \leq 43$	Alta
$T > 43$	Muito alta

Quando o estudante obtém o maior número possível de bactérias, a temperatura no interior da estufa está classificada como

- (A) muito baixa.
- (B) baixa.
- (C) média.
- (D) alta.
- (E) muito alta.

Qual a dificuldade dessa questão para você ?

- (A) Fácil
- (B) Média
- (C) Difícil

Figura 7. Exemplo de questão

Diante disso, os resultados obtidos com a percepção são apenas um indicativo do desempenho do sistema. Feitas essas considerações, estiveram presentes no simulado todas as questões que tiveram um valor mínimo ou máximo (0 ou 1), que, após a normalização em um de seus parâmetros da TRI (a , b e c), foram inseridas na prova. Assim, foi possível verificar, por exemplo, se uma questão com o parâmetro de dificuldade máxima foi realmente considerada difícil pelos alunos. A Figura 7 apresenta um exemplo de questão presente na prova.

Com a grande quantidade de alunos e questões presentes nas provas, a tabulação da percepção de dificuldade foi um obstáculo. Para facilitar essa tarefa, foi criado um formulário no Google. Este continha um campo onde o corretor informava seu nome, código e percepção de dificuldade informada por cada estudante. Para agilizar as tabulações, selecionaram-se 20 voluntários, que auxiliaram na tarefa de tabulação.

Feita a etapa das tabulações, foram coletados os dados, os quais foram utilizados para verificar se as classificações geradas pelo sistema coincidiram com as informadas dadas pelos alunos. Para facilitar essas visualizações de desempenho do sistema, foram geradas tabelas e gráficos.

Com base na metodologia apresentada, na próxima seção são apresentados os resultados obtidos.

4. Resultados e discussão

Nesta seção, apresentam-se os resultados obtidos com o desenvolvimento do Sistema para Geração Inteligente de Simulados do ENEM. A sequência de apresentação é a seguinte: curvas características dos itens, classificação das questões, centróides dos grupos de questões, prova desenvolvida, aplicação do simulado, percepções de dificuldades e análises do desempenho do sistema.

Os *scripts* desenvolvidos podem ser consultados na plataforma de hospedagem de código-fonte GitHub, no link: <https://github.com/ghnunes/TRI>.

4.1 Curvas características dos itens

As Figuras 8 e 9 apresentam as curvas características dos itens das questões de Matemática e suas Tecnologias, do ENEM de 2015. É possível perceber que as questões 16 e 33 não se comportam bem, já que indivíduos com baixa habilidade possuem uma probabilidade alta de acertá-las. Todas as questões aplicadas no ENEM são testadas antes da sua utilização na prova, então esse tipo de comportamento não é esperado.

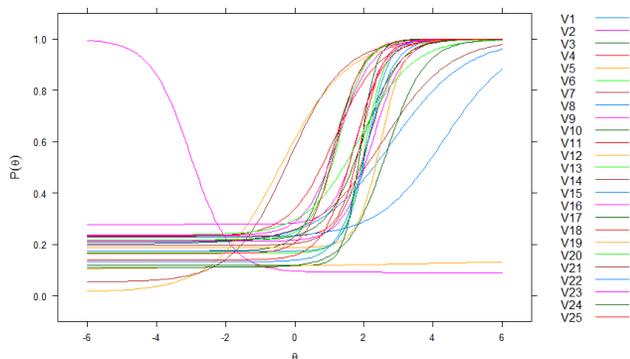


Figura 8. Curvas características dos itens das questões de Matemática e suas Tecnologias do ENEM de 2015 - Questões 1 a 25

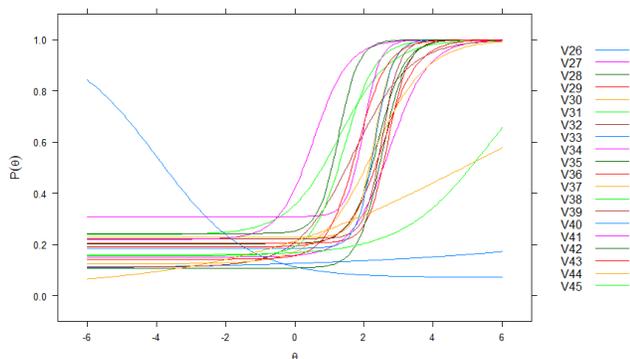


Figura 9. Curvas características dos itens das questões de Matemática e suas Tecnologias do ENEM de 2015 - Questões 26 a 45

4.2 Classificação de questões

As Tabelas 1, 2, 3, 4, 5 e 6 apresentam os centróides dos grupos de questões classificados pelo *K-means*, mostrados na Figura 10.

Tabela 1. Centróides dos grupos de questões de Ciências Humanas e suas Tecnologias

	a	b	c
1	0,280675	0,3928520	0,70481874
2	0,2253421	0,4249778	0,30893218
3	0,1719946	0,3142006	0,07937224

Tabela 2. Centróides dos grupos de questões de Ciências da Natureza e suas Tecnologias

	a	b	c
1	0,4571055	0,2313748	0,1166093
2	0,4449202	0,2388031	0,2882332
3	0,5961328	0,2331434	0,7445470

Tabela 3. Centróides dos grupos de questões de Linguagens, Códigos e suas Tecnologias

	a	b	c
1	0,6008155	0,4213811	0,6455132
2	0,5957857	0,4511615	0,3426516
3	0,4373236	0,3843769	0,0702844

Tabela 4. Centróides dos grupos de questões de Matemática e suas Tecnologias

	a	b	c
1	0,06900471	0,7189462	0,06118741
2	0,07136792	0,7154851	0,67054914
3	0,06737323	0,7150868	0,34395032

Tabela 5. Centróides dos grupos de questões de Espanhol

	a	b	c
1	0,11241939	0,8044363	0,20759011
2	0,95883851	0,6283279	0,06569395
3	0,08326018	0,4159264	0,05164764

Tabela 6. Centróides dos grupos de questões de Inglês

	a	b	c
1	0,43901208	0,3583714	0,71614450
2	0,05896505	0,1065844	0,06051471
3	0,13207069	0,4224048	0,26738191

As questões presentes na Figura 10 foram classificadas em três grupos de dificuldade, cada um representado por uma cor. É possível perceber que as questões de Matemática e suas Tecnologias (Figura 10 (d)) e Ciências da Natureza e suas Tecnologias (Figura 10 (b)) estão mais agrupadas, variando mais no parâmetro *c* (acerto casual).

4.3 Prova desenvolvida

O simulado gerado conteve 180 questões e uma proposta de redação. Em reunião feita com a Diretoria de Ensino do Instituto Federal de Minas Gerais - *Campus Bambuí*, foi requerida uma prova de dificuldade balanceada. Diante disso, como as questões foram classificadas em três grupos de dificuldade, e cada área de conhecimento do ENEM é composta por 45 questões, foram colocadas 15 de cada grupo. Isso não foi possível de ser realizado com as questões de Matemática e suas Tecnologias, pois poucas questões se enquadraram no grupo dois (questões fáceis). Essa metodologia pode ter gerado uma prova com dificuldade maior do que a balanceada, e influenciando na dificuldade geral da prova.

As Figuras 11 e 12 apresentam as questões classificadas, sendo possível observar que os pontos (questões) estão bem distribuídos, algo que era esperado, já que foram adicionadas questões de vários grupos de dificuldade.

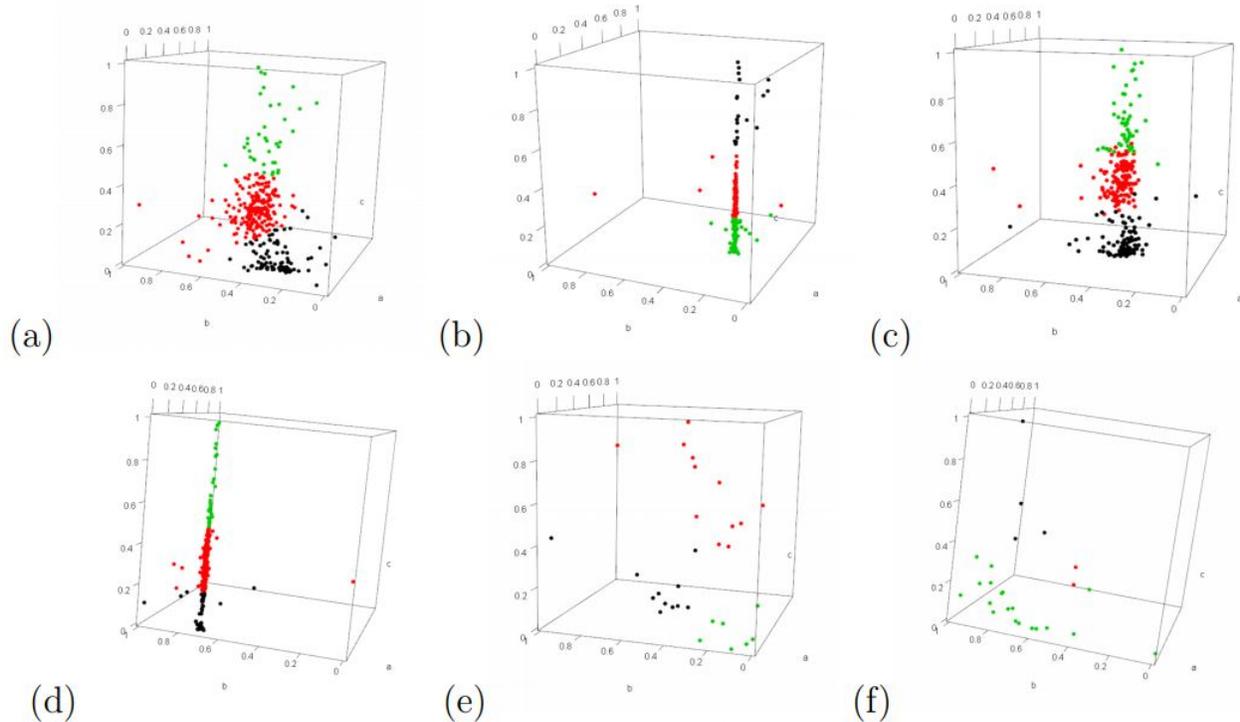


Figura 10. Sequência de classificação das questões de (a) Ciências Humanas e suas Tecnologias, (b) Ciências da Natureza e suas Tecnologias, (c) Linguagens, Códigos e suas Tecnologias, (d) Matemática e suas Tecnologias, (e) Inglês e (f) Espanhol

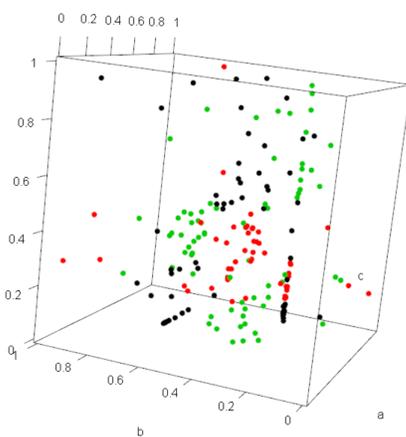


Figura 11. Questões da prova classificadas - prova de inglês

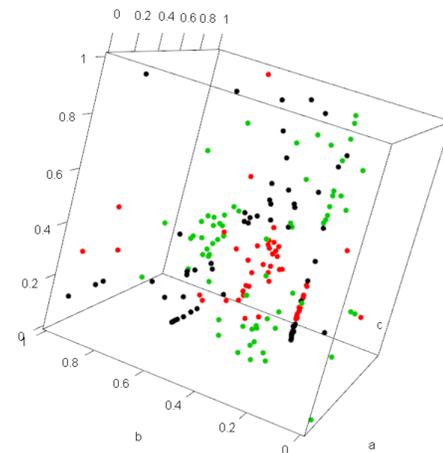


Figura 12. Questões da prova classificadas - prova de espanhol

4.4 Aplicação do simulado

O objetivo do simulado organizado pela Diretoria de Ensino do IFMG - *Campus* Bambuí foi preparar os alunos para a realização do ENEM. Então, ele foi aplicado nos dias 21/09 e 05/10, para que fosse proveitoso para os estudantes, antes da prova oficial.

Seguindo o padrão do ENEM, no primeiro dia, 118 alunos fizeram a prova de Linguagens; Códigos e suas Tecnologias; Ciências Humanas e suas Tecnologias; e a redação. Já no segundo dia, 94 estudantes compareceram para fazer as questões de Ciências da Natureza e suas Tecnologias, e Matemática e suas Tecnologias.

4.5 Tabulação da percepção de dificuldade

Além da preparação dos estudantes para o ENEM, o simulado tinha o objetivo de fornecer a percepção dos alunos sobre a dificuldade de cada questão, para fins de comparação com as classificadas pelo sistema.

Com a finalização do simulado, foram feitas análises nas provas antes da correção, constatando-se que algumas pessoas marcaram a mesma dificuldade em todas questões, ou deixaram todas as alternativas em branco. Diante disso, essas provas foram tabuladas de forma separada, sendo desconsideradas. No primeiro dia, foram utilizadas 73 provas, e, no segundo, 41.

A Figura 13 apresenta a percepção média de dificuldade por área relatada pelos alunos que fizeram o simulado. É possível observar que a dificuldade da prova está balanceada. Com isso, o objetivo de gerar um simulado balanceado foi alcançado. Isso mostra a capacidade que o algoritmo possui para gerar provas com uma dificuldade calibrada pelo usuário.

4.6 Avaliação do sistema

Aplicado o questionário foi obtido as percepções dos estudantes em relação à dificuldade da questão e a fornecida pelo sistema desenvolvido. O número de respondentes é diferente de uma questão para outra pelo fato de alguns alunos não terem respondido determinada percepção. Notou-se um caso atípico com as questões de Espanhol, devido ao número baixo de estudantes que fizeram a opção por essa língua.

As Tabelas 7, 8, 9, 10 e 11 são matrizes de confusão que mostram o desempenho do sistema na tarefa de classificar as questões. É importante ressaltar novamente a subjetividade da comparação, uma vez que estão sendo comparadas as classificações fornecidas pelo *K-means* com as percepções fornecidas pelos alunos. Feitas as considerações, analisando-se a diagonal principal da Tabela 11, é possível observar que a taxa de acertos é de 37,22%, no geral. Isoladamente, a área de Matemática e suas Tecnologias obteve o maior índice de acertos, com 51,11%, e o pior foi na de Linguagens, Códigos e suas Tecnologias, com 28,88%.

Essa porcentagem baixa de acertos foi causada pelo fato de o sistema confundir questões que estão em escalas próximas, ou seja: Fácil com Média, Média com Difícil, Média com Fácil, Média com Difícil e Difícil com Média. O conceito de Fácil e Médio, ou Médio e Difícil, é algo complexo de se definir, já que vai estar relacionado com as habilidades do indivíduo que está realizando o teste. Por exemplo, uma questão pode ser tornar fácil para um estudante que viu o conteúdo poucos dias antes da prova, e ser média para alguém que o viu há mais tempo.

Tabela 7. Matriz de confusão Linguagens, Códigos e suas Tecnologias

	Fácil	Média	Difícil
Fácil	2	9	0
Média	5	9	3
Difícil	5	11	1

Tabela 8. Matriz de confusão Ciências Humanas e suas Tecnologias

	Fácil	Média	Difícil
Fácil	0	9	2
Média	4	12	1
Difícil	0	15	2

Tabela 9. Matriz de confusão Ciências da Natureza e suas Tecnologias

	Fácil	Média	Difícil
Fácil	2	11	1
Média	1	13	1
Difícil	0	13	3

Tabela 10. Matriz de confusão Matemática e suas Tecnologias

	Fácil	Média	Difícil
Fácil	2	0	0
Média	4	19	1
Difícil	5	12	2

Tabela 11. Matriz de confusão geral

	Fácil	Média	Difícil
Fácil	6	29	3
Média	14	53	6
Difícil	10	51	8

A Figura 14 mostra um gráfico do somatório de acertos do sistema por área.

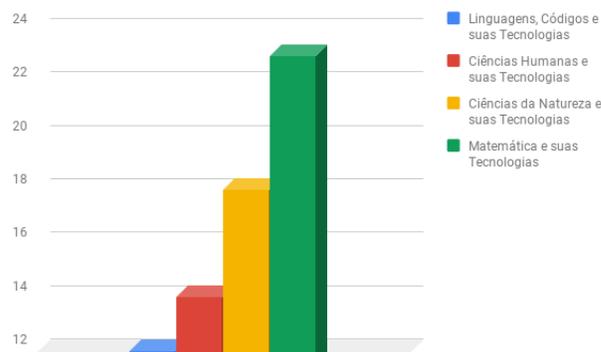


Figura 14. Gráfico do somatório de acertos por área

O sistema classificou 13 questões (7,22%) consideradas fáceis pelos alunos como difíceis, ou o inverso. As Figuras 15 e 16 apresenta um comparativo de desempenho do sistema para diferentes definições para erro.

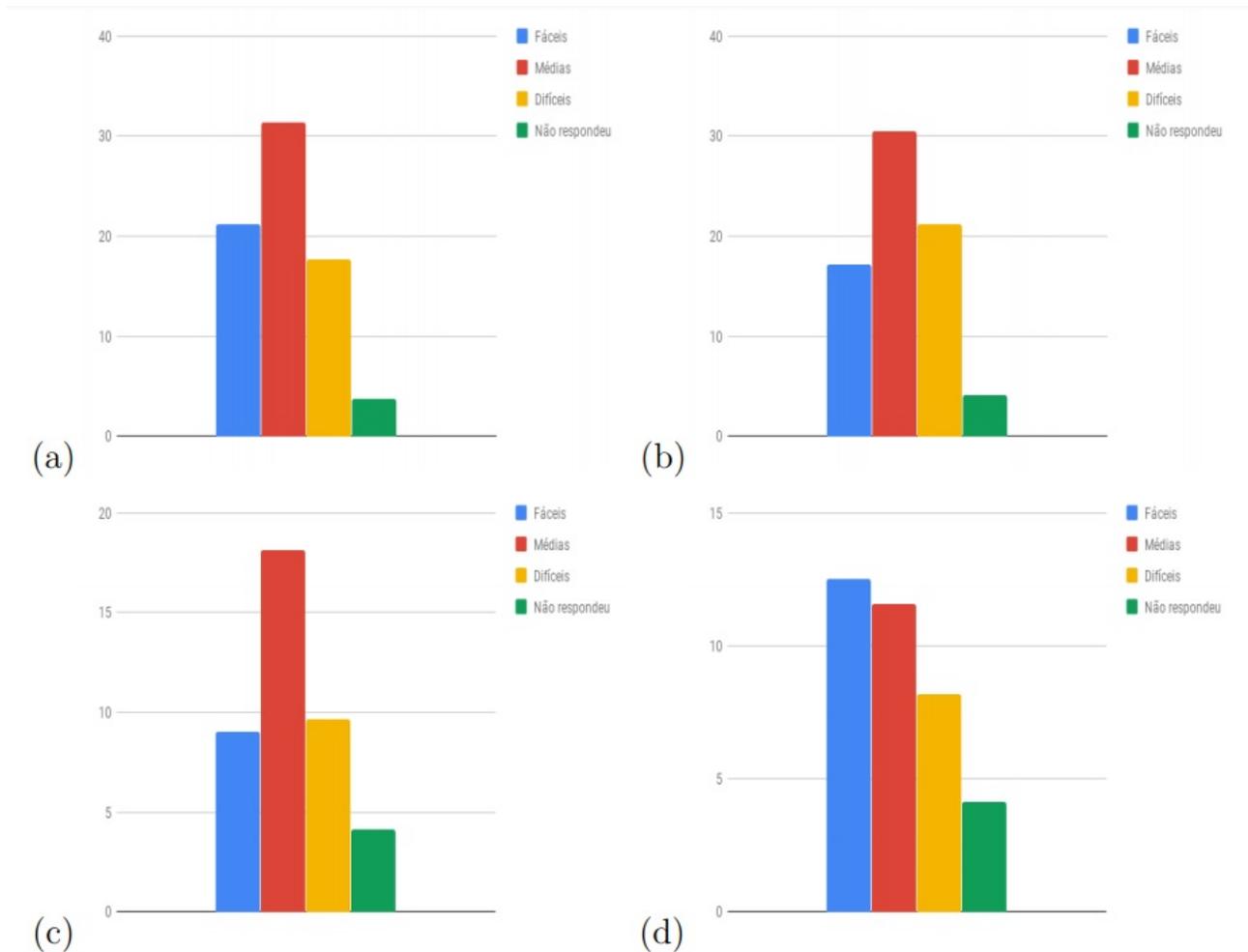


Figura 13. Dificuldade média por área (a) Linguagens, Códigos e suas Tecnologias (b) Ciências Humanas e suas Tecnologias (c) Ciências da Natureza e suas Tecnologias e (d) Matemática e suas Tecnologias

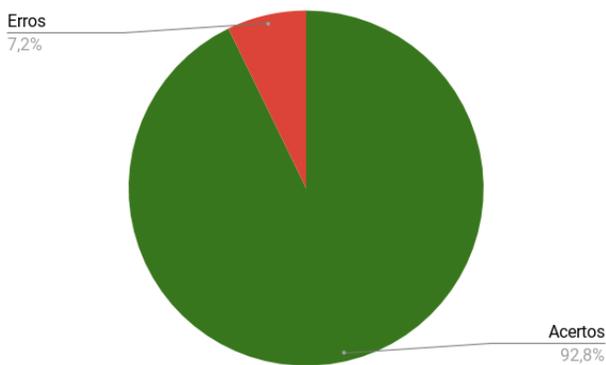


Figura 15. Taxa de acerto do sistema - considerando erro apenas questões Fáceis classificadas como Difíceis ou o inverso

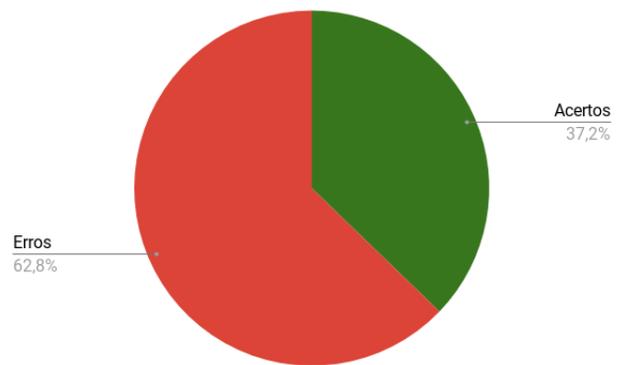


Figura 16. Taxa de acerto do sistema - considerando erro dificuldade diferente entre a fornecida pelo sistema e pela percepção

A Figura 17 demonstra a questão que apresentou o maior valor no parâmetro de dificuldade. Ela foi classificada como uma questão difícil, ou seja, está no Grupo um de questões, na área de Matemática e sua Tecnologias.

A forma do vidro foi concebida de modo que o eixo x sempre divida ao meio a altura h do vidro e a base do vidro seja paralela ao eixo x . Obedecendo a essas condições, o engenheiro determinou uma expressão que fornece a altura h do vidro em função da medida n de sua base, em metros.

A expressão algébrica que determina a altura do vidro é

- (A) $\log\left(\frac{n+\sqrt{n^2+4}}{2}\right) - \left(\log\frac{n-\sqrt{n^2+4}}{2}\right)$
- (B) $\log\left(1 + \frac{n}{2}\right) - \log\left(1 - \frac{n}{2}\right)$
- (C) $\log\left(1 + \frac{n}{2}\right) + \log\left(1 - \frac{n}{2}\right)$
- (D) $\log\left(\frac{n+\sqrt{n^2+4}}{2}\right)$
- (E) $2\log\left(\frac{n+\sqrt{n^2+4}}{2}\right)$

Qual a dificuldade dessa questão para você ?

- (A) Fácil
- (B) Média
- (C) Difícil

Figura 17. Questão considerada a mais difícil da área de Matemática e suas Tecnologias

Através de análises da Figura 18, é possível verificar que a questão 144 provavelmente se trata de uma questão difícil. Das 41 pessoas que responderam ao questionário de percepção de dificuldade, nenhuma classificou a questão como fácil. Outro fator que corrobora essa hipótese foi o baixo número de alunos que acertaram a questão.

Na seção seguinte, são apresentadas as conclusões sobre este trabalho.

5. Conclusão

Nesta pesquisa, apresentou-se um trabalho que visou o desenvolvimento de um Sistema para Geração Inteligente de Simulados do ENEM que possibilitasse a produção de simulados com base nos parâmetros de cada questão. Os objetivos foram cumpridos, uma vez que o sistema desenvolvido permite a geração de simulados com base na dificuldade de cada questão estabelecida pelo sistema.

Como foi discutido na seção de Resultados e discussões, a comparação das classificações realizadas pelo *K-means* com a percepção fornecida pelos alunos é apenas um indicativo do funcionamento do sistema, já que grandezas distintas estão sendo comparadas. Além disso, a percepção de dificuldade fornecida por um aluno é diretamente influenciada por suas habilidades, fatores emocionais, motivação na prova, etc. Diante dessas ressalvas, o sistema apresentou um baixo índice de acertos, uma vez que acaba confundindo questões de classes

Questão 144

Um engenheiro projetou um automóvel cujos vidros das portas dianteiras foram desenhados de forma que suas bordas superiores fossem representadas pela curva de equação $y = \log(x)$, conforme a figura.

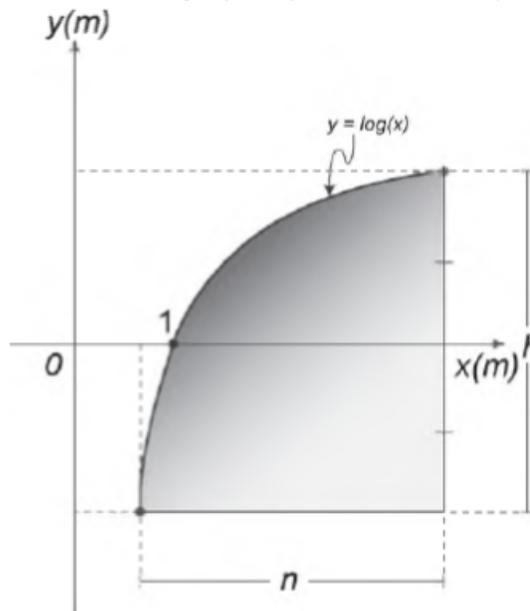


Figura 18. Questão considerada difícil da área de Matemática e suas Tecnologias

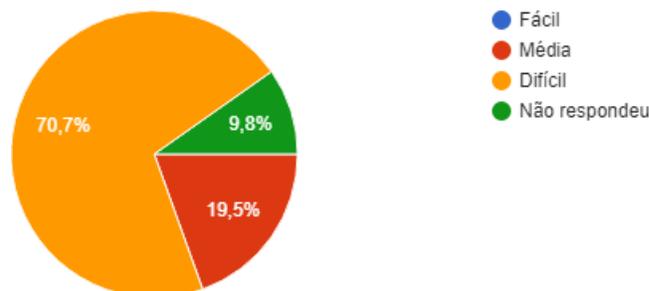


Figura 19. Percepção de dificuldade da questão 144

próximas. Entretanto, é importante ressaltar que o sistema classificou apenas 7% de questões difíceis como fáceis, ou o inverso. Isso mostra que ele funciona em classes distantes. Com a finalidade de melhorar a acurácia, podem ser válidas estratégias com o uso de outros algoritmos de classificação, outras fórmulas para calcular a dificuldade, com base nos parâmetros dos centróides, ou, até mesmo, utilizar quantidades diferentes de grupos de dificuldade.

Uma vez que não foram encontrados trabalhos na literatura que classifiquem as questões em grupos de dificuldade, com base nos parâmetros da TRI, todo o conteúdo desenvolvido ao longo desta pesquisa dá o primeiro passo na utilização de algoritmos de classificação para mensurar a dificuldade de

uma questão.

Como proposta de continuidade, pode ser desenvolvido um sistema para capturar as questões diretamente das provas disponibilizadas pelo INEP, em formato PDF. Com isso, quebra-se a dependência dos *sites*-sementes utilizados pelo *crawler* para montar o banco de questões. Além disso, pode ser desenvolvido um sistema *web* para facilitar a geração dos simulados.

Também é uma proposta de continuidade recalculando os parâmetros da TRI das questões que foram aplicadas no simulado, com a utilização da matriz de respostas. Com isso, é possível comparar se os parâmetros calculados antes da aplicação do simulado, com a utilização dos microdados, coincidem com o novo cálculo feito a partir das respostas dos alunos do IFMG - *Campus Bambuí*.

Outro trabalho futuro que pode ser considerado é a possibilidade da comparação entre o resultado de um aluno que realizou o uso da solução proposta com o seu resultado no Enem, com o objetivo de verificar se o simulado do sistema proposto tende a gerar questões com dificuldades parecidas com a da avaliação principal do ano recorrente. É possível também avaliar o sistema desenvolvido para outros tipos de testes diferentes do ENEM como, por exemplo, para a POS-COMP.

Contribuições dos autores

Todos os três autores participaram efetivamente no desenvolvimento do trabalho e posteriormente na construção do artigo científico. O primeiro autor executou a maior parte de seus experimentos, sendo auxiliado pelo segundo autor que ficou responsável por grande parte da elaboração do artigo científico. O papel da orientação em todas as etapas do trabalho ficou sob responsabilidade do terceiro autor que contribuiu diretamente do início ao fim desta pesquisa.

Referências

- [1] RIBEIRO, A. G.; MENDES, A. A. *A dificuldade de resolução das questões de matemática do exame nacional do ensino médio: ineficiência matemática ou interpretativa?* 2016. Disponível em: <http://pensaracademico.facig.edu.br/index.php/semiariocientifico/article/viewFile/127/106>.
- [2] CASTRO, L. G. M. de. *Análise dos microdados do enem a partir da teoria da resposta ao item*. UFMG, Belo Horizonte, 2017.
- [3] INEP. *Enem 2017 tem 7,6 milhões de inscritos*. 2017. Disponível em: http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/enem-2017-tem-7-6-milhoes-de-inscritos/21206.
- [4] MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada*. Belo Horizonte: Editora UFMG, 2005.
- [5] FUENTES, A. *Em ranking da educação com 36 países, Brasil fica em penúltimo*. 2017. Disponível em: <https://veja.abril.com.br/blog/impavido-colosso/>.
- [6] CONSULTADORIA, C. *Escolas particulares tem lucro até 30 % maior em 2017 devido a queda da inflação*. 2017. Disponível em: <https://castromarketing.com.br/escola-particular/>.
- [7] OLIVEIRA, B. A. S.; ASSIS, S.; NOLLI, C. Development of a prototype electrical energy monitoring system via internet/desenvolvimento de um protótipo de sistema de monitoramento de energia elétrica via internet. *Revista de Engenharia da Universidade Católica de Petrópolis*, v. 12, n. 1, p. 48–61, 2019.
- [8] CASTORIL, B. R. Previsão de demanda de energia elétrica no setor de produção do laticínio do ifmg campus bambuí utilizando regressão de vetores de suporte. *Engenharia de Computação*, 2018.
- [9] MATIAS, L.; TOLEDO, S. *Enem: veja como 1.434 instituições usam a nota do exame*. 2016. Disponível em: <https://guiadoestudante.abril.com.br/enem/enem-como-1-434-instituicoes-usam-a-nota-do-exame-2/>. Acesso em: 20 jul. 2018.
- [10] ANDRADE, D. F. de; TAVARES, H. R.; VALLE, R. da C. Teoria de resposta ao item: Conceitos e aplicações. Caxambu, p. 1–164, 2000.
- [11] MOREIRA, F. G. Cálculo das proficiências dos alunos de mg na prova de matemática do enem de 2011 através de um modelo da teoria de resposta ao item. UFMG, Belo Horizonte, 2014.
- [12] RICHARDSON, M. W. The relation between the difficulty and the differential validity of a test. *Psychometrika*, Springer, v. 1, n. 2, p. 33–49, 1936.
- [13] TUCKER, L. R. Maximum validity of a test with equivalent items. *Psychometrika*, Springer, v. 11, n. 1, p. 1–13, 1946.
- [14] LAZARSELD, P. F. The logical and mathematical foundation of latent structure analysis. *Studies in Social Psychology in World War II Vol. IV: Measurement and Prediction*, Princeton University Press, p. 362–412, 1950.
- [15] LORD, F. A theory of test scores. *Psychometric monographs*, 1952.
- [16] PASQUALI, L.; PRIMI, R. Fundamentos da teoria da resposta ao item: Tri. *Avaliação Psicológica: Interamerican Journal of Psychological Assessment*, Instituto Brasileiro de Avaliação Psicológica (IBAP), v. 2, n. 2, p. 99–110, 2003.
- [17] VALLE, R. da C. Teoria de resposta ao item. *Estudos em avaliação educacional*, n. 21, p. 7–92, 2000.
- [18] ROSSI, F. *O Enem e a Teoria de Resposta ao Item (TRI)*. 2016. Disponível em: <https://manualdoestudo.deaaz.com.br/2016/10/07/o-enem-e-a-teoria-de-resposta-ao-item/>.

- [19] MILLER, K.; JORDAN, M. I.; GRIFFITHS, T. L. Nonparametric latent feature models for link prediction. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2009. p. 1276–1284.
- [20] PASQUALI, L. Psychometrics. *Revista da Escola de Enfermagem da USP*, SciELO Brasil, v. 43, n. SPE, p. 992–999, 2009.
- [21] WANG, J.-L.; CHIOU, J.-M.; MÜLLER, H.-G. Functional data analysis. *Annual Review of Statistics and Its Application*, Annual Reviews, v. 3, p. 257–295, 2016.
- [22] BISHOP, C. M. *Pattern recognition and machine learning*. [S.l.]: springer, 2006.
- [23] ASSIS, S. S. de. Desenvolvimento de um protótipo para classificação de perfis de pulverização utilizando processamento digital de sinais e redes neurais artificiais. *Engenharia de Computação*, 2017.
- [24] OLIVEIRA, B. A. S. et al. Avaliação de uma rede neural artificial como estimador temporal pluviométrico no sistema de abastecimento cantareira. *Revista de Informática Aplicada*, v. 14, n. 1, 2018.
- [25] SANTOS, F. H. W. dos et al. Multicritério e reconhecimento de imagens em avaliação de embarcações.
- [26] TOWNSEND, J. T. Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, Springer, v. 9, n. 1, p. 40–50, 1971.
- [27] DATE, C. J. *Introdução a sistemas de bancos de dados*. [S.l.]: Elsevier Brasil, 2004.
- [28] ELMASRI, R. et al. *Sistemas de banco de dados*. Pearson Addison Wesley, 2005.
- [29] MAIER, D. *The theory of relational databases*. [S.l.]: Computer science press Rockville, 1983. v. 11.
- [30] HEYDON, A.; NAJORK, M. Mercator: A scalable, extensible web crawler. *World Wide Web*, Springer, v. 2, n. 4, p. 219–229, 1999.
- [31] AD, G. *O que é Crawler?* 2017. Disponível em: <https://support.google.com/webmasters/answer/182072>.
- [32] GOOGLE. *Googlebot*. 2017. Disponível em: <https://support.google.com/webmasters/answer/182072>.
- [33] INEP. *MICRODADOS*. 2018. Disponível em: <http://inep.gov.br/microdados>.
- [34] NUNES, C. H. Sancineto da S.; PRIMI, R. Impacto do tamanho da amostra na calibração de itens e estimativa de escores por teoria de resposta ao item. *Avaliação Psicológica*, Instituto Brasileiro de Avaliação Psicológica, v. 4, n. 2, 2005.
- [35] TOIT, M. D. *IRT from SSI: Bilog-MG, multilog, parscale, testfact*. [S.l.]: Scientific Software International, 2003.
- [36] BARROSO, N. C.; RODRIGUES, M. A. *Comparações dos softwares bilog-mg e r no uso da teoria de resposta ao item*. 2012.
- [37] LTM, R. D. an r package for latent variable modeling and item response theory analyses. *J Stat Softw*, v. 17, n. 5, p. 1–25, 2006.
- [38] ANJOS, A. dos; ANDRADE, D. F. de. *Teoria da resposta ao item com uso do r*. 2012.
- [39] CHALMERS, R. P. mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, v. 48, n. 6, p. 1–29, 2012.
- [40] SARTES, L. M. A.; SOUZA-FORMIGONI, M. L. O. d. Avanços na psicometria: da teoria clássica dos testes à teoria de resposta ao item. *Psicologia: Reflexão e Crítica*, Curso de Pós-Graduação em Psicologia da Universidade Federal do Rio Grande do Sul, 2013.
- [41] LEONI, R. C.; SAMPAIO, N. A. de S. Desempenho das escolas públicas e privadas da região do vale do paraíba: Uma aplicação da técnica de agrupamentos kmeans com base nas variáveis do enem 2015. *Cadernos do IME-Série Estatística*, v. 42, p. 31, 2017.
- [42] OCAMPO, D. M.; TOLENTINO-NETO, L. C. B. de. Cluster analysis for data processing in educational research. *Acta Scientiae*, v. 21, n. 4, p. 34–48, 2019.
- [43] BARROS, A. J. da S.; LEHFELD, N. A. de S. *Fundamentos de metodologia científica*. [S.l.]: Pearson, 2007.
- [44] PINSONNEAULT, A.; KRAEMER, K. Survey research methodology in management information systems: an assessment. *Journal of management information systems*, Taylor & Francis, v. 10, n. 2, p. 75–105, 1993.
- [45] MASCARENHAS, S. *Metodologia científica*. [S.l.]: Pearson, 2017.
- [46] GERHARDT, T. E.; SILVEIRA, D. T. *Métodos de pesquisa*. [S.l.]: Plageder, 2009.
- [47] OLHONAVAGA. *Olho na vaga*. 2018. Disponível em: <https://olhonavaga.com.br>.
- [48] SELENIUMHQ. *SeleniumHQ*. 2018. Disponível em: <https://www.seleniumhq.org>.
- [49] PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, v. 12, n. Oct, p. 2825–2830, 2011.
- [50] MOZILLA. *Geckodriver*. 2018. Disponível em: <https://github.com/mozilla/geckodriver/releases>.
- [51] COPELAND, R. *Essential sqlalchemy*. [S.l.]: ” O’Reilly Media, Inc.”, 2008.
- [52] REITZ, K. *Requests*. 2018. Disponível em: <http://docs.python-requests.org/en/master/>.
- [53] DUBOIS, P.; BY-WIDENIUS, M. F. *MySQL*. [S.l.]: New riders publishing, 1999.
- [54] ADLER, D. et al. Package ‘ff’. *Universitaet Goettingen, Goettingen, Germany*, p. 2–2, 2018.
- [55] HAWKINS, D. M. *Identification of outliers*. [S.l.]: Springer, 1980. v. 11.

- [56] HARTIGAN, J. A.; WONG, M. A. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, JSTOR, v. 28, n. 1, p. 100–108, 1979.
- [57] KANUNGO, T. et al. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, IEEE, n. 7, p. 881–892, 2002.
- [58] ENEM, B. do. *Prova do Enem – Saiba melhor como é estruturado o exame*. 2013. Disponível em: <https://blogdoenem.com.br/prova-do-enem-exame/>.