

A interdisciplinaridade no desempenho da nota de matemática: um olhar para evolução do processo de ensino por meio de modelos regressivos*

Mirley Bitencourt Ferreira, Myrna Amorim, Eduardo Ogasawara, Rafael Barbastefano

¹Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ)

mirley.ferreira@aluno.cefet-rj.br, myrna.amorim@cefet-rj.br,

eogasawara@ieee.org, rafael.barbastefano@cefet-rj.br

Abstract. *Access to a large volume of open data has expanded the possibilities for improving public systems management. One of these bases is the National Secondary Education Examination (ENEM). It contains relevant information from test performance to socioeconomic and cultural characteristics of candidates. In this scenario, this work aimed to analyze the relationship of math grades with other grades in different areas, including writing, using ENEM 2019 as the basis for analysis. The results obtained show that other grades influence the math grade (MT). Among the eight models applied, the Gradient Boosting was the best, with a 7.4% error in the TM prediction. This analysis is relevant because we can guide public policies that can improve overall academic performance.*

Resumo. *O acesso a um grande volume de dados abertos ampliou as possibilidades de melhoria da gestão dos sistemas públicos. Uma dessas bases é o Exame Nacional do Ensino Médio (ENEM). Nela há informações relevantes do candidato e de seu desempenho nas provas. Nesse cenário, este trabalho teve como objetivo analisar a relação das notas de matemática com outras notas em diferentes áreas, incluindo a escrita, da base do ENEM 2019. Os resultados obtidos apresentam que a nota de matemática (MT) é influenciada pelas demais notas e dentre os oito modelos aplicados o melhor foi o Gradient Boosting com 7,4% de erro na previsão da MT. Essa análise é relevante porque podemos orientar políticas públicas que podem melhorar o desempenho acadêmico geral.*

1. Introdução

No Brasil, a partir do Decreto Presidencial de 03 de abril de 2000, tem-se o objetivo de revisar e propor políticas e diretrizes com base nas interações eletrônicas e incorporando frequentemente o uso das tecnologias da informação [Wiedenhöft et al., 2020]. A importância dessas ações para a educação aparece no ENEM. O ENEM é um programa para avaliar os resultados dos candidatos para seleção para a graduação e possui seus dados públicos. A partir dele, é possível fazer diversos estudos sobre as bases de informações a fim de discutir e avaliar políticas públicas e pedagógicas com fatos e dados. No ENEM 2019, por exemplo, houve em torno de 5 milhões de inscritos. Os dados e as notas são

*Os autores agradecem à FAPERJ, à CAPES (código 001) e ao CNPq pelo financiamento do projeto.

disponibilizados publicamente no portal do Instituto Nacional de Estudos e Pesquisas Educacionais (INEP) [INEP, 2020].

Como o ENEM está associado à etapa final da educação básica, em que a grande maioria dos candidatos estão se preparando para entrada na graduação, é importante monitorar o desempenho dos estudantes ao longo da educação básica. Iniciativas como o Programa Internacional de Avaliação de Estudantes (PISA) possibilitam trazer um retrato intermediário e geral dos alunos. Este programa é realizado pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE) a cada três anos. A partir dele, tem-se um comparativo de desempenho de alunos, na faixa etária de 15 anos, entre os países nos seguintes domínios: leitura, matemática e ciências [Sjøberg, 2019]. Sendo um programa internacional, os países podem comparar os resultados buscando referências de melhores práticas e políticas de ensino e aprendizagem externas.

O Brasil tem participado de todos os exames do PISA desde a criação em 2006. Em 2018 a média OCDE mundial para as notas de Matemática e Ciência foram de 489 e para a nota de Leitura foi de 487. O desempenho do Brasil em 2018 foi abaixo da média mundial em 105 pontos em matemática, 85 pontos em Ciências e 74 pontos em leitura. Além disso, em todos os resultados neste Programa, a nota de matemática é a mais baixa. Já no *ranking* mundial do PISA 2018 o país teve novamente um dos dez piores desempenhos em matemática [OECD, 2021].

Além deste cenário, onde há um destaque de notas baixas em matemática, alguns autores apontam uma grande desmotivação dos alunos pela disciplina [Simpkins et al., 2006; Lazowski and Hulleman, 2016; Oppermann and Lazarides, 2021]. O estudo de Lazowski and Hulleman [2016] reforça que a motivação é o aspecto mais importante no processo de aprendizagem e que está relacionada diretamente com o desempenho dos alunos e conseqüentemente o êxito escolar. Da mesma forma, observa-se que baixos resultados em matemática podem ser devidos a fatores de ordem pessoal e motivacional [Oppermann and Lazarides, 2021].

É importante destacar que as habilidades esperadas ao estudar matemática, de acordo com os Parâmetros Curriculares Nacionais (PCN) para o ensino fundamental e para o ensino médio, envolvem o desenvolvimento de raciocínios que auxiliam em estratégias como resolução de problemas e tomada de decisões, competências fundamentais na formação do indivíduo, na capacidade da própria gestão financeira, além de outras competências que o ajudarão em pleno exercício de um cidadão responsável [MEC, 2021].

Neste contexto, é importante analisar o desempenho dos alunos para entender a evolução ao longo do tempo e a relação entre as notas das diversas áreas do conhecimento. A avaliação diagnóstica é fundamental nos processos de ensino e aprendizagem, tanto para os órgãos responsáveis pelo desenvolvimento de políticas de aprendizagem, como para alunos e professores, que precisam identificar as habilidades e competências para decidir quais ações são essenciais na melhoria do aprendizado. Contudo é necessário que os gestores das políticas públicas educacionais divulguem amplamente esses resultados e invistam na interdisciplinaridade, a fim de potencializar o conhecimento matemático. Portanto, para fortalecer o desempenho na matemática, faz-se necessário também elevar o índice de aprendizagem de todas as demais disciplinas. Esta recomendação, da interdisciplinaridade, vem sendo indicada por vários teóricos da educação e pelos PCNs

[Santos et al., 2017].

Diante do exposto, este trabalho tem como hipótese avaliar se o desempenho dos alunos na matemática pode ser estimado a partir das demais áreas do conhecimento. Isso pode ser traduzido pela relação entre as notas em matemática e as demais notas das áreas do conhecimento aplicadas. O objetivo desta pesquisa é avaliar a nota de matemática no ENEM com base nas demais notas, inclusive de redação. Para tanto, foram aplicados modelos regressivos baseados em aprendizado de máquina sobre a base de dados do ENEM 2019. A relevância da metodologia é detectar a relação entre as áreas de conhecimento, resultando que um melhor desempenho nas notas de matemática pode estar associado ao melhor desempenho nas demais áreas. Assim, a técnica é um recurso a mais para monitorar a interdisciplinaridade.

Este trabalho está organizado em cinco seções incluindo esta introdução. Na Seção 2 são apresentados outros trabalhos relacionados, já na Seção 3 é apresentada a metodologia utilizada neste estudo, enquanto na Seção 4 são apresentados os resultados e discussões. Por fim, a Seção 5 descreve as considerações finais.

2. Trabalhos relacionados

Na literatura existem muitas interações das técnicas de Mineração de Dados (MD) para descobrir novos conhecimentos aplicados ao contexto educacional motivados por diferentes frentes: educação *offline* para análises em dados de desempenho do aluno; aprendizado eletrônico (*e-learning*) e Sistema de Gestão da Aprendizagem; Sistemas Tutores Inteligentes e Sistemas Hiper mídias Adaptativos Educacionais [Romero and Ventura, 2010]. Como Chiheb et al. [2017] que analisam alunos de graduação e de pós-graduação, classificando-os de acordo com os resultados, que podem ser utilizados pela gestão para indicar os discentes em risco de evasão. Já Shayan and van Zaanen [2019] verificam o comportamento do aluno no Sistema de Gerenciamento de Aprendizado a fim de identificar diferentes desempenhos durante o curso e, assim, apresentar os resultados aos docentes que poderão usar as informações para aprimorar os conteúdos das disciplinas e distinguir aqueles que necessitem de maior atenção.

Outras abordagens utilizam a base de dados pública do ENEM as quais exploram o nível social médio dos alunos, o índice da formação do docente e as médias das notas das áreas do conhecimento da aplicação desta Prova. O trabalho de Simon and Cazella [2017] pautado nesta base relaciona fatores socioeconômicos das escolas do ensino médio de todo o território nacional para gerar um modelo preditivo no indicador de desempenho médio dos alunos em ciências da natureza e suas tecnologias. Outro exemplo sobre MD nos dados públicos do ENEM é o estudo de Stearns et al. [2017] para prever o desempenho de estudantes baseando-se apenas em informações socioeconômicas. Os resultados apontam que existe um viés significativo das características socioculturais dos discentes em relação às notas.

Este estudo difere-se por aplicar técnicas de mineração de dados na base do ENEM sobre as notas de matemática em relação as demais áreas, sem utilizar os dados socioeconômicos. Entende-se que avaliar a nota de matemática em relação as demais, possibilita que avaliações possam ser feitas tendo a matemática como variável *proxy*. Isso viabiliza, um maior número de testes de acompanhamento sem incorrer no aumento significativo de custo operacional, possibilitando ações mais ágeis de políticas pedagógicas públicas

focadas no processo ensino-aprendizagem dos alunos ao longo da educação básica.

3. Metodologia

O presente estudo utiliza aprendizados de máquinas na tarefa de regressão das notas de matemática a partir das demais notas. O processo de mineração de dados adotado neste trabalho é composto pelas etapas, conforme apresentado na Figura 1. Este processo é chamado de Descoberta de Conhecimento em Bases de Dados [Fayyad et al., 1996]. É importante destacar que para gerar conhecimento ao final, este processo é iterativo, onde as etapas podem ocorrer mais de uma vez. Os parágrafos e subseções seguintes detalham cada uma delas.



Figura 1. Etapas deste trabalho (adaptado de Fayyad et al. [1996])

A **coleta de dados** dos candidatos do ENEM foi a partir do portal do INEP do concurso realizado em 2019. O conjunto de dados tem 136 atributos, sendo 52 dados categóricos, 53 *booleanos* e 31 numéricos. Toda a composição destes atributos é apresentada e divulgada em um dicionário de dados. Nesta grande base de dados há informações socioeconômicas e notas dos estudantes. É importante destacar que esta base apresentou 9,1% de valores faltantes. A visão geral dos dados apresentados foi obtida por meio da biblioteca *Profiling Report* presente no *Python*.

O **Pré-processamento de dados** foi subdividido em três etapas. Na primeira foi feita uma análise exploratória para conhecimento da base, na segunda a limpeza de dados e na terceira o tratamento de *outliers*.

a) Análise exploratória: o universo amostral contém os 27 estados brasileiros que representa o número de alunos de acordo com a Unidade Federativa (UF) de residência. Pode-se verificar que a maioria dos alunos são do estado de São Paulo. Verificou-se a distribuição das idades dos candidatos e observou-se que há uma concentração de candidatos em torno de 18 anos, algo que é esperado já que é a idade em que a maioria dos alunos completa o terceiro ano do ensino médio. A idade mínima é 13 anos e a máxima é de 82 anos. Ainda nesta fase inicial de exploração da base de dados verificou-se que a maioria dos candidatos se autodeclararam do gênero feminino.

Na Figura 2, buscou-se entender como as quatro notas de múltiplas escolhas estão distribuídas. O primeiro histograma apresenta as notas de Ciências Naturais (CN), com média 477,9 e desvio padrão 76,3. O segundo são as notas de Ciências Humanas (CH) com as médias de notas em torno de 507,3 pontos. O terceiro histograma, com as notas de Linguagens e Códigos (LC), tem média 520 e menor desvio em relação aos outros, ou seja, 64,5. E por último, as notas de Matemática (MT), que é a variável de resposta do modelo, com a maior média de notas 523,5 e maior desvio padrão de 109,4. Em todas as notas observou-se valores mínimos iguais a zero. A nota máxima, em relação às outras, foi de matemática com 985 pontos.

Na **limpeza de dados**, considerando-se o objetivo do trabalho, é preciso levar em conta apenas os alunos que fizeram realmente a prova. O banco de dados apresenta o

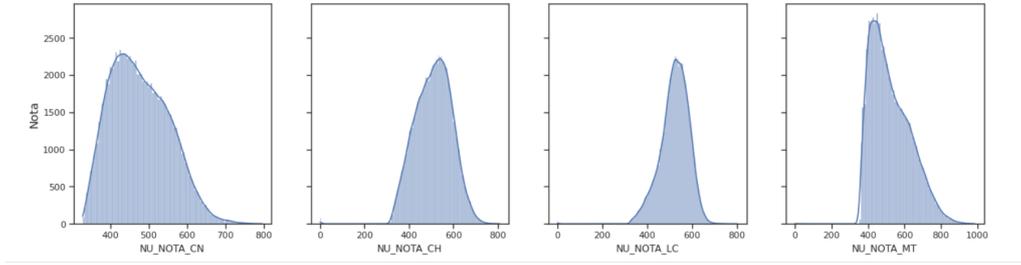


Figura 2. Histograma das notas dos candidatos.

atributo TP_PRESENCA onde o valor 0 (zero) significa que o aluno não realizou o exame naquele dia. Esses registros foram removidos durante o pré-processamento. Da mesma forma, a base de dados é preparada com apenas os atributos das notas de CN, CH, LC, Redação e a nota de MT. Finalmente, o conjunto de dados foi separado em treino e teste. Para o treinamento foram destinados 80% dos registros e os demais 20% para teste.

O **tratamento de outliers** ocorreu nos valores extremos que estão fora do intervalo esperado, visando a preparação dos dados para o treinamento. Tais valores foram removidos do conjunto de treinamento, uma vez que podem influenciar a modelagem de aprendizado de máquina.

Na etapa de **mineração de dados**, para verificar a força dos modelos a serem construídos, foi realizada a análise de correlação de *Pearson* entre as notas, conforme apresentado na Figura 3. Foi observada a correlação forte de 0,75 entre as notas de LC e CH e a menor correlação foi de 0,49 entre as notas de MT e Redação. Ou seja, este é um indicativo que essas variáveis têm um bom grau de correlação entre si. Neste trabalho foram explorados os seguintes modelos: (i) Regressão Linear Multivariada, (ii) Regressão Linear Lasso; (iii) *Random Forest*, (iv) *Gradient Boosting* e (v) Redes Neurais. Os dois primeiros são de aprendizado estatístico e capturam as relações lineares entre as variáveis, enquanto os demais são de aprendizado de máquina. Embora os modelos de aprendizado estatísticos tenham sido incluídos para efeitos de comparação, dada as correlações observadas, espera-se que eles tenham um bom desempenho.

Para a **avaliação dos modelos** foram empregadas as seguintes métricas para analisar e avaliar o desempenho da variável de resposta: *Root Mean Squared Error* (RMSE) e *R-Squared* (R^2). O RMSE, apresentado na Equação 1, calcula a raiz quadrada da média das distâncias ao quadrado entre o valor correto (y_i) e a predição (\hat{y}_i) para todas as notas.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

Já o R^2 calcula a razão entre a variância do estimador e a variância das observações. Seu valor está no intervalo $[0, 1]$, onde 1 representa o melhor resultado. Na Equação 2, o R^2 é apresentado e \bar{y} é o valor médio da variável y .

$$R^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

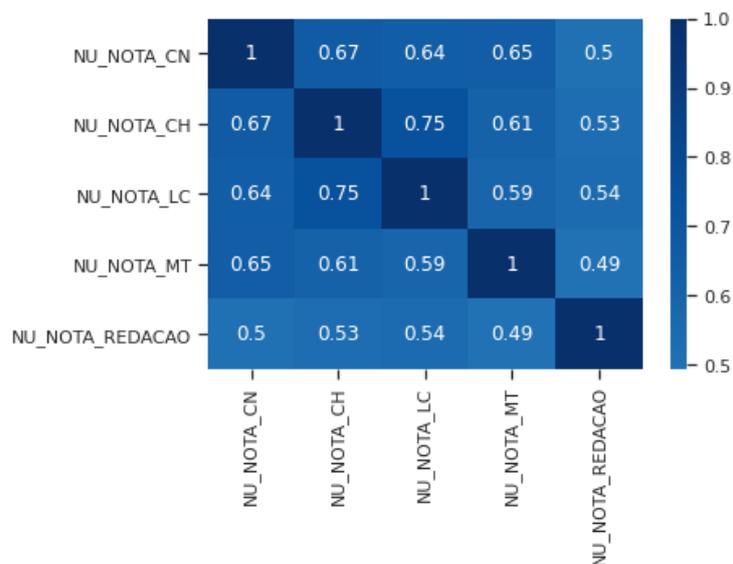


Figura 3. Correlação de *Pearson* entre das notas dos candidatos.

Para implementação dessas etapas foi utilizada a linguagem Python e várias bibliotecas como *Pandas*, *Seaborn* e *Scikit Learn*. A plataforma de desenvolvimento foi o *Jupyter Notebook* no COLAB, a qual é *on-line* e gratuita do Google.

4. Resultados e Discussões

Durante a etapa de treinamento, foram explorados alguns hiper parâmetros. Nas redes neurais, variou-se o número de camadas (1 a 3) e número de neurônios por camadas (1 a 30). Os melhores resultados foram com 2 camadas com 20 neurônios cada uma. O *Random Forest* foi variado o número de árvores de 1 a 100. O melhor resultado foi obtido com 47. Já no *Gradient Boosting* o hiper parâmetro de profundidade foi configurado entre 1 e 6, onde o melhor resultado foi obtido com a profundidade 3.

É preciso destacar que os resultados obtidos foram baseados no conjunto de teste. Foram desconsiderados os candidatos faltosos, mas sem remoção de *outliers*. A Tabela 1 apresenta os resultados das métricas de cada modelo. O menor RMSE foi do *Gradient Boosting* com 74,00 e com coeficiente de regressão R^2 igual a 0,74. O RMSE apresenta que o modelo tem erro igual a 74 pontos. Ou seja, para um universo onde os pontos da prova vão de 0 a 1.000, isto corresponde a 7,4% de erro.

Vale destacar ainda que a rede neural obteve bons desempenhos perante os algoritmos escolhidos. A rede neural possui duas camadas com 20 neurônios cada uma. Além disso, esta última obteve um tratamento de normalização com uma função de quartil associada, onde as variáveis são medidas em escalas diferentes, porém diretamente comparáveis. Esta parametrização da Rede Neural foi obtida por meio da própria documentação da biblioteca, *Scikit Learn* do *Python*, de Regressão de Redes MLP.

Observa-se que os modelos regressivos apontaram que a nota de matemática tem relação com as demais notas. Para fins de análise da qualidade, foi levantado uma *baseline* de comparação, fundamentada na média das notas de matemática da base de treino. Este *baseline* é interessante, pois caracterizaria o cenário de não haver ligação entre as

notas. Lembrando que a média de matemática foi de 523,5. Então, foi calculado o RMSE da *baseline* com as notas de matemática do conjunto de teste, obtendo o valor 109,79 (dispersão de 11.0%). Este valor é bem maior do que os RMSEs calculados. Logo, essa avaliação reitera a existência da relação entre a matemática e as demais notas. Isso pode ser confirmado com *p-value* do teste de hipótese *t-student* apresentado na Tabela 1. A hipótese nula foi refutada em todos os casos e observou-se que os modelos construídos foram superiores ao *baseline*.

Tabela 1. Métricas de Avaliação

Modelo	RMSE	R^2	p-value
Regressão Linear Multivariada	76,89	0,71	0
Regressão Linear Lasso	76,89	0,71	0
Random Forest	76,08	0,71	0
Gradient Boosting	74,00	0,74	0
Rede Neural	75,99	0,72	0

O resultado obtido através desse processo aplicado de mineração de dados indica que a matemática pode ser utilizada como variável *proxy* para as demais disciplinas. Esta pesquisa ao utilizar-se desta técnica propicia a oportunidade de orientar políticas públicas no sentido de melhorar o desempenho das disciplinas avaliadas no PISA, ENEM entre outros. Reforça-se assim, a necessidade de um acompanhamento sistemático das notas dos discentes ao longo do educação básica, bem como realça a importância interdisciplinaridade como indica os PCN's.

5. Considerações Finais

Conclui-se que através dos modelos regressivos encontrados pelo processo de mineração de dados foi possível prever a nota de matemática a partir das demais disciplinas com 7,4% de erro. Este valor é inferior à dispersão de 11.0% da nota de matemática, o que reitera a existência da relação entre a nota de matemática e das notas de ciências da natureza, ciências humanas, linguagens e códigos e redação do ENEM. Observou-se que o melhor modelo, baseado na métrica RMSE, foi o *Gradient Boosting*. Nota-se ainda, que mesmo os modelos de aprendizado estatísticos tiveram desempenhos aceitáveis. Embora a relevância deste resultado alcançado percebe-se que é possível aprimorar a construção destes modelos através da inclusão, por exemplo, de variáveis sócias econômicas dos discentes.

O trabalho abordou um problema educacional brasileiro com relação ao baixo desempenho da matemática e uma grande desmotivação dos alunos nesta disciplina. Sendo assim, o resultado obtido pode auxiliar os gestores, que aplicam políticas de ensino-aprendizagem, no intuito de promover ações que estimulem o estudo da matemática de modo interdisciplinar ao longo da educação básica. O que fortalece a interdisciplinaridade para elevar o grau de aprendizagem do aluno e é uma motivação a mais para dialogar e gerar maior cumplicidade entre as áreas do conhecimento. Tendo como finalidade o aumento das notas das disciplinas que compõem os testes avaliativos nacionais e internacionais, como ENEM e PISA. Para tanto é preciso que a mesma seja divulgada e levada em consideração pelos órgãos competentes.

Referências

- F. Chiheb, F. Boumahdi, H. Bouarfa, and D. Boukraa. Predicting students performance using decision trees: Case of an Algerian University. In *Proceedings of the 2017 International Conference on Mathematics and Information Technology, ICMIT 2017*, volume 2018-January, pages 113–121, 2017.
- UM Fayyad, G Piatetsky-Shapiro, and P Smyth. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In *International Conference on Knowledge Discovery and Data Mining*, pages 82–88, 1996. ISBN 1-57735-004-9. URL <http://www.aaai.org/Papers/KDD/1996/KDD96-014>.
- INEP. Microdados do ENEM 2019. Technical report, <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>, 2020.
- R.A. Lazowski and C.S. Hulleman. Motivation Interventions in Education: A Meta-Analytic Review. *Review of Educational Research*, 86(2):602–640, 2016.
- MEC. Parâmetros Curriculares Nacionais. Technical report, <http://basenacionalcomum.mec.gov.br/implementacao>, 2021.
- OECD. Programme for International Student Assessment. Technical report, <https://www.oecd.org/pisa/>, 2021.
- E. Oppermann and R. Lazarides. Elementary school teachers’ self-efficacy, student-perceived support and students’ mathematics interest. *Teaching and Teacher Education*, 103, 2021.
- C. Romero and S. Ventura. Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 40(6):601–618, 2010.
- Carla Madalena Santos, Rubia Amanda Franco, Diego Leon, Daniel Bovolenta Ovigli, and Pedro Donizete Colombo. Interdisciplinarity in Education: Overcoming Fragmentation in the Teaching-Learning Process. *International Education Studies*, 10(10): 71–77, 2017. ISSN 1913-9020.
- P. Shayan and M. van Zaanen. Predicting student performance from their behavior in learning management systems. *International Journal of Information and Education Technology*, 9(5):337–341, 2019.
- Augusto Simon and Sílvio Cazella. Mineração de dados educacionais nos resultados do ENEM de 2015. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 6, page 754, October 2017.
- S.D. Simpkins, P.E. Davis-Kean, and J.S. Eccles. Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology*, 42(1):70–83, 2006.
- Svein Sjøberg. PISA: A success story? In Roman Langer and Thomas Brüsemeister, editors, *Handbuch Educational Governance Theorien*, Educational Governance, pages 653–690. Springer Fachmedien, Wiesbaden, 2019. ISBN 978-3-658-22237-6.
- B. Stearns, F. Rangel, F. Rangel, F.F. De Faria, and J. Oliveira. Scholar performance prediction using boosted regression trees techniques. In *ESANN 2017 - Proceedings, 25th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pages 329–334, 2017.
- G.C. Wiedenhöft, E.M. Luciano, and G.V. Pereira. Information Technology Governance Institutionalization and the Behavior of Individuals in the Context of Public Organizations. *Information Systems Frontiers*, 22(6):1487–1504, 2020.