



GESTÃO DA INFORMAÇÃO DO PERFIL DOS PROFESSORES E INFRAESTRUTURA DAS ESCOLAS PÚBLICAS DO ESTADO DO PARÁ: UM MODELO BASEADO EM ÁRVORE DE DECISÃO

Yuri Vidal Mendonça
yurisantiago@ufpa.br
Universidade Federal do Pará

Thalita Carvalho Bezerra
thalitacarvalhobezerra@gmail.com
Universidade da Amazônia

RESUMO

As avaliações de larga escala têm sido o principal indicador de qualidade da educação. No Brasil, a literatura da área tem apontado diversos fatores com efeitos preditivos no índice de desenvolvimento da educação básica - IDEB. Nesse sentido, este estudo tem por objetivo analisar o efeito dos atributos educacionais, criados a partir de características das unidades de ensino da rede pública do estado do Pará, como a estrutura física e tecnológica das escolas e o perfil dos professores no IDEB. Tendo como diferencial de análise o treinamento de um modelo baseado em árvore de decisão, o que permite a compreensão dos fatores que efetivamente estão contribuindo com a qualidade educacional no estado. A amostra adotada tem característica não probabilística, composta a partir de três bases de dados oficiais, com 450 escolas da rede estadual de ensino. Os resultados apontam que o número de computadores por aluno, a existência de espaços pedagógicos que envolvem o uso de tecnologia, como laboratórios de informática e a idade dos professores tiveram maior efeito no IDEB.

Palavras-chave: IDEB, Perfil dos professores, infraestrutura das escolas, árvore de decisão.

Eixo Temático 1: Gestão Pública e do Desenvolvimento



1 INTRODUÇÃO

A qualidade da educação básica aferida principalmente pelas avaliações de larga escala, tem promovido várias pesquisas, as quais evidenciam que o desempenho educacional está diretamente relacionado à diversos fatores, tais como: a infraestrutura das escolas (Barrett, Treves, Shmis, Ambasz, & Ustinova, 2019), a escassez de recursos humanos e materiais (Talis, 2018), bem como a qualificação e formação continuada dos professores (Fauth, Decristan, Decker, Büttner, Hardy, Klieme, & Kunter, 2019). Além disso, o acesso às tecnologias digitais de informação e comunicação (TDIC's) (Machado, 2016) e a quantidade de alunos em sala de aula (Monks & Schmidt, 2011), também tem se demonstrado determinantes importantes no que concerne aos resultados avaliativos dos estudantes, bem como ao desempenho dos professores atuantes neste processo.

Considerando esse contexto, emerge ainda a necessidade de (re) construção dos modelos educacionais, em específico, quando os indicadores gerais da qualidade do ensino apontam para uma necessidade de reavaliação de estratégias pedagógicas e das políticas educacionais voltadas para a melhoria da qualidade da educação básica. Um exemplo disso é o estado do Pará, localizado na região norte do Brasil, que apresentou o segundo pior índice de desenvolvimento da educação básica (IDEB) do país em 2019, para o ensino médio, com nota 3,4 (Portal G1-Pará, 2020).

Segundo o site da Secretaria Estadual de Educação do Pará (SEDUC/PA) em 2020, a rede atendeu cerca de 549.761 alunos, distribuídos em 22 unidades regionais de ensino (URE's), presentes nos 144 municípios que compõem o estado. Esses números demonstram a complexidade e a necessidade de analisar, planejar e implementar ações mais eficazes, que possibilitem não só a melhoria dos indicadores da qualidade educacional aferidos nas avaliações externas, mas também possibilitem o melhor desenvolvimento e/ou desempenho dos sujeitos envolvidos neste cenário.

Apesar de um número crescente de estudos demonstrarem os efeitos positivos na qualidade do ensino e do sucesso na aprendizagem (Pianta, Downer, & Hamre, 2016), a maioria dos pesquisadores buscou demonstrar implicações de uma única dimensão (da Silva, & Ciasca, 2020) e/ou de um conjunto de características combinadas de um aspecto educacional específico (Pangeni, 2014), que possam explicar como ocorre o sucesso do desenvolvimento escolar (Wibowo, Budiantara & Hidayanto, 2018; Zhang, Shao & Johnston, 2019). Contudo, grande parte das pesquisas até o momento apresenta uma visão mais genérica, usando apenas estatísticas descritivas e/ou inferenciais com o objetivo de testar características determinantes no processo do desenvolvimento educacional (Ajayi, & Ekundayo, 2011; Pangeni, 2014). Assim, ainda são poucos os estudos que utilizam técnicas de classificação avançadas e mineração de dados (Khanna, Singh, & Alam, 2016), empregando um grande número de atributos para prever, com precisão o desempenho escolar.

Esses aspectos sugerem um gap na literatura constante dos fatores promotores da qualidade educacional, podendo ser utilizados de forma a contribuir na



**Engajando ciência,
gestão e sociedade**

09 a 11 de novembro de 2021

XII CODS
COLÓQUIO ORGANIZAÇÕES,
DESENVOLVIMENTO & SUSTENTABILIDADE

PPAD
Programa de Pós-graduação em Administração

UNAMA
UNIVERSIDADE
DA AMAZÔNIA

ser
educacional

melhoria dos resultados dos estudantes em avaliações de larga escala. Portanto, a pesquisa atual tem por objetivo analisar o efeito dos atributos educacionais, criados a partir de características das unidades de ensino, como a estrutura física e tecnológica das escolas e o perfil dos professores no indicador de desenvolvimento da educação básica (IDEB), especificamente, dos anos iniciais.

Nesse sentido, esse estudo justifica-se pela perspectiva gerencial das instituições educativas, pois pode contribuir com a análise dos efeitos das características propostas para o desenvolvimento de estratégias e políticas públicas, com vistas à melhoria dos resultados educacionais. Também se justifica pelo ponto de vista metodológico, em função do uso de técnicas de classificação avançada para previsão dinâmica do efeito das características educacionais nos indicadores de qualidade da educação. Nesse caso, foi utilizada a modelagem de *machine learning* (Shahbazi, 2020), para compreensão dos efeitos desses elementos no construto alvo: o Índice da Educação Básica (IDEB).

Dessa forma, este trabalho tem caráter quantitativo, pois utiliza dados de escolas estaduais de 7 (sete) municípios paraenses, retirados de três bases oficiais: Secretaria de Estado de Educação do Pará - SEDUC/PA, Qedu.org.br e DashBoard da plataforma *Google For Education*. A população investigada foi escolhida por conveniência e portanto, de caráter não probabilístico (Fávero, & Belfiore, 2017), visando utilizar o método de aprendizado de máquina que inclui tarefas de classificação e regressão, denominado de Árvore de Decisão (Alsagheer, Alharan & Al-Haboobi, 2017).

2 REFERENCIAL TEÓRICO

Nessa seção, serão analisadas literaturas relevantes, as quais os pesquisadores observam uma ou mais características e/ou perfil do professor, a infraestrutura das escolas bem como o perfil tecnológico das mesmas e de que forma tais atributos interferem no desempenho dos alunos, bem como no Índice de Desenvolvimento da Educação Básica (IDEB).

2.1 ATRIBUTOS DO PERFIL DOCENTE

As avaliações de larga escala aplicadas no âmbito nacional e internacional, têm contribuído para a emergência de estudos acerca do perfil dos professores que atuam nas escolas, procurando identificar características ou atributos desses profissionais que poderiam estar associados ao sucesso escolar dos alunos. Assim, ao longo dos últimos anos, alguns autores vêm se dedicando a analisar o perfil do professor da educação básica no Brasil (Carvalho, 2018; Souza & Gouveia, 2011), uma vez que se constitui um passo muito importante para uma melhor compreensão acerca do papel destes sobre o rendimento dos alunos.

Informações sobre o perfil sociodemográfico docente podem ser obtidas através do Censo Escolar que é realizado anualmente no Brasil e que apresenta



**Engajando ciência,
gestão e sociedade**

09 a 11 de novembro de 2021

XII CODS
COLÓQUIO ORGANIZAÇÕES,
DESENVOLVIMENTO & SUSTENTABILIDADE

PPAD
Programa de Pós-graduação em Administração

UNAMA
UNIVERSIDADE
DA AMAZÔNIA

ser
educacional

características dos professores quanto à distribuição por etapas de ensino, sexo, idade, etnia, formação inicial, localização da escola que leciona entre outras informações. O Censo Escolar compõe hoje a base de dados mais ampla disponível no país, uma vez que obtém informações de todas as escolas em funcionamento, em todos os níveis e modalidades de ensino (municipal, estadual ou privada). Outro meio de obtenção de dados sobre o perfil docente, o ambiente de ensino e o rendimento educacional é a Pesquisa Internacional sobre Ensino e Aprendizagem (Teaching and Learning International Survey – TALIS). No entanto, o censo e o relatório de TALIS não coletam, como outras pesquisas, o levantamento de características, percepções e significados específicos que os docentes atribuem a diferentes aspectos e contextos de sua atuação (Fleuri, 2015).

Neste sentido, a importância do professor no processo educativo pode ser contextualizada através de algumas características e/ou atributos como: boa comunicação, criatividade, organização, comprometimento, planejamento, domínio do conteúdo, bem como atributos pessoais como idade, gênero e a crença, onde todos esses interagem com fatores ambientais (Klassen et al., 2018), como a infraestrutura do local de trabalho, a quantidade de alunos em sala, a quantidade de escolas que o docente leciona e a qualidade da formação deste, influenciando na sua prática cotidiana e refletindo ainda, no processo de aprendizado dos alunos. Nesta pesquisa, foram testadas algumas características as quais, percebe-se direta relação no desempenho do professor e conseqüentemente, no desempenho dos alunos.

2.1.1 Jornada de trabalho

Alguns professores assumem trabalho docente em mais de uma turma, escola, ou disciplina, demandando deslocamentos e mais disposição desses profissionais. Além disso, com o objetivo de suprir a necessidade de professores para a população em idade escolar, algumas redes de ensino aumentam a jornada média de trabalho desses profissionais, dificultando a atuação dos que têm a jornada fragmentada entre diversos vínculos, diferentes redes e escolas (Elacqua & Marotta, 2020), sendo estas, condições que impactam diretamente na capacidade de organização e de trabalho dos docentes, afetando sua saúde, seu rendimento profissional, e subsequente a qualidade do aprendizado do aluno (Ronfeldt et al., 2013).

2.1.2 Gênero

A perspectiva de gênero é um aspecto importante, que deve ser considerado na reflexão e na compreensão das características das relações que se constituem entre professores e alunos no contexto escolar, sendo também um aspecto importante na construção de políticas para a carreira docente. O censo Escolar de 2017 apontou que 96% dos professores que atuam nos anos iniciais do ensino fundamental são do sexo feminino (Carvalho, 2018). Alguns dos motivos para o magistério ser predominantemente feminino é o fato de permitir que as mulheres façam uma



conciliação entre família, trabalho doméstico e o mercado de trabalho (Bruschini & Amado, 1988). Nessa perspectiva, professores do sexo feminino tendem a ter um desgaste maior no que se refere ao seu desempenho profissional.

2.1.3. Participação nas formações

A formação de professores é uma temática amplamente discutida na política educacional brasileira, principalmente nas últimas décadas, quando o Brasil iniciou a participação em avaliações internacionais, a exemplo do Programme for International Student Assessment (PISA) e também a partir do desenvolvimento de avaliações de larga escala como o Índice de Desenvolvimento da Educação Básica (IDEB), que tem por objetivo mensurar variáveis relacionadas ao fluxo escolar e ao desempenho dos alunos nas disciplinas de Língua Portuguesa e Matemática nos 5º e 9º anos do ensino fundamental, referindo-se, portanto, à avaliação de alunos, meio oportuno para analisar o protagonismo docente nos resultados de aprendizagem.

Nessa perspectiva, infere-se que professores motivados para aprender e ensinar tendem a potencializar sua prática educativa, além de conseguirem estimular o engajamento dos alunos no processo de aprendizagem. Assim sendo, a formação docente tem sido entendida como elemento crucial para a garantia da qualidade do ensino e da aprendizagem dos alunos (Soares et al., 2012; Christophersen; Elstad; Turmo, & Solhaug, 2016), na qual analisa-se como fator positivo a inovação formativa através da utilização de recursos tecnológicos, como as Tdic's, que têm sido cada vez mais incorporadas na prática docente com a potencialidade de vim a contribuir com o processo de ensino do professor, promovendo uma aprendizagem mais significativa, apoiando na construção e implementação de metodologias mais ativas, despertando maior interesse e empenho dos alunos (Soares, et al., 2018; Aldunate & Nussbaum, 2013).

2.1.4 Idade do professor/Tempo de docência

A idade do docente que atua na educação básica pode indicar algumas das características do seu perfil profissional (Amadi & Allagoa, 2017). Assim, a distribuição das idades dos professores em cada um dos ciclos da educação básica reflete de certa forma, o que acontece na trajetória desses profissionais e na carreira docente: profissionais mais jovens nas etapas iniciais de ensino e profissionais mais experientes nas etapas finais (Carvalho, 2018). Portanto, a faixa etária dos professores pode estar associada à maior experiência profissional, considerando que aqueles que atuam há mais tempo na carreira, têm maior amadurecimento em suas ações pedagógicas e no enfrentamento de desafios (Souza & Gouveia, 2011), enquanto aqueles que são mais jovens na docência têm menos experiência e em alguns casos, menos domínio do conteúdo, porém mais disposição e criatividade no processo de ensino.



Ressalta-se ainda, que a idade do docente pode estar relacionada também com a postura que este assume em relação à aceitação e uso de novas tecnologias, ou seja, professores mais jovens, tendem a ter maior familiaridade na utilização desses recursos, bem como maior facilidade em assimilá-las em sua prática profissional. Professores mais velhos, por outro lado, mais acostumados às práticas tradicionais de ensino, podem mostrar uma postura mais passiva diante do uso de ferramentas tecnológicas, demandando maior esforço de capacitação na assimilação do uso de tais tecnologias no processo de mediação de ensino - aprendizagem (Cantini et al., 2006; Shepherd & Devers, 2017).

2.2 A INFRAESTRUTURA DA ESCOLA

Acredita-se que uma escola com a infraestrutura adequada proporciona uma sensação de bem-estar, não somente para os alunos, como também para professores e demais funcionários, abrindo espaço para oportunidades de aprendizado e relações sociais significativas. Nesse contexto, o ambiente escolar deve ser desde a estrutura física, um ambiente que favoreça a produção de saberes e experiências afetivas, sociais e culturais (Serôdio 2015), refletindo ainda no desempenho e aprendizado do aluno.

Priorizar uma infraestrutura escolar ideal nos tempos atuais é respeitar e propor as múltiplas possibilidades do desenvolvimento intelectual, social, motor e profissional que esse ambiente pode proporcionar aos estudantes e professores.

Uma infraestrutura escolar de qualidade pode representar significativamente uma evolução no desempenho escolar do estudante, podendo-se perceber vários aspectos positivos como por exemplo, uma melhor frequência, o interesse e a motivação dos estudantes e professores pelo ensino e aprendizado.

Assim, infere-se que escolas que apresentam, em sua infraestrutura, salas de aulas bem arejadas, com iluminação adequada podem trazer uma sensação de conforto e bem-estar para aqueles que delas usufruem. Da mesma forma, mais ambientes adequados com recursos pedagógicos como salas de informática, laboratórios de ciências, quadra poliesportiva coberta, entre outros, potencializam e engajam o aprendizado dos estudantes.

2.3 O PERFIL TECNOLÓGICO DAS ESCOLAS

Atualmente, um desafio que se impõe é tornar o uso das TDIC'S nas escolas tão significativo quanto o uso cotidiano que se faz desses recursos, dada a importância da inserção e uso desses recursos tecnológicos (Miranda & Russell, 2012). Dessa forma, não se trata de criar laboratórios de informática com horários estanques e pontuais, mas de incorporar as tecnologias na rotina da sala de aula, como recursos à disposição de alunos e professores, além de propor estratégias pedagógicas as quais desenvolvam no aluno a capacidade de analisar, avaliar e decidir sobre a resolução de problemas (Silva & Ross, 2019).



Em uma época em que o acesso à informação e as possibilidades de educação se tornam ubíquas e na qual, variados recursos midiáticos influenciam os modos de leitura das pessoas, cabe à escola exercer um papel proativo fomentando novas ideias sobre aprendizagem, comunicação e tecnologias. Contudo, não basta que a instituição escolar tenha acesso à tecnologia, é necessário que o ambiente escolar esteja preparado e disposto para inovar suas estratégias de aprendizagem (Batistello, 2020), compreendendo todas as inovações tecnológicas como um importante recurso que auxilia na prática pedagógica (Modelski; Giraffa, & Casartelli, 2019).

3 METODOLOGIA

3.1 CENÁRIO E PARTICIPANTES

O conjunto de dados escolhido para o treinamento do modelo de *machine learning* inclui o cruzamento de informações de três bases:

Base 1: A Secretaria de Estado de Educação (SEDUC) responsável pelo fornecimento das informações dos professores;

Base 2: O site qedu.org.br que organiza e disponibiliza de maneira didática as informações do Ministério da Educação sobre as escolas públicas;

Base 3: O acesso ao painel de controle da plataforma Google For Education permitiu identificar o acesso por ID dos professores, suas características e a escola que de fato utilizou ou não a plataforma.

Sendo que dessas bases foram retiradas as informações dos atributos escolhidos para verificação dos seus efeitos no alvo (IDEB das escolas referente ao ano de 2019). Foram tomados como parâmetros de filtro dos dados, aspectos como: a participação total ou parcial nas formações continuadas promovidas pela plataforma *Google For Education*, ser professor das séries iniciais da educação básica e por fim, estar lotado em escolas da região metropolitana da capital do estado. Dessa forma, foram selecionados dados de 450 escolas de sete municípios. Sendo que a base de análise após refinamento dos dados apresentou uma amostra com 617 professores.

3.2 PRESSUPOSTOS DO MÉTODO

Este estudo adotou o uso de *Decision Trees* (DT), que são algoritmos multivariados de aprendizagem supervisionada, utilizados em problemas de classificação e predição de alvos. Modelos baseados em DT são não paramétricos, significando que não assumem qualquer tipo pré-definido de distribuição probabilística nos dados de entrada (Han et al., 2011; Tan & Steinbach, 2006). Através do uso de DT, é possível entender a relação existente entre as variáveis independentes e dependentes. Uma variedade de modelos clássicos baseados em DT pode ser encontrada na literatura, como Chi-Square Automatic Interaction Detector (CHAID; Kass, 1980) e Classification and Regression Trees (CART; Breiman et al., 1984). Os modelos de regressão utilizados neste estudo são algoritmos do tipo CART.



Para as modelagens de *machine learning* da etapa de análise de dados, utilizou-se o pacote PyCaret, uma biblioteca para modelagem de *machine learning* em linguagem Python que requer poucas linhas de código para modelar dados das mais variadas complexidades, possibilitando ainda a comparação de resultados para diferentes algoritmos baseados em DTs de maneira simultânea, fazendo com que escolhamos apenas os modelos que mais se destacam positivamente.

As modelagens propostas neste estudo foram desenvolvidas em quatro diferentes instâncias (modelos). Na primeira (M1), treinou-se um modelo para prever o índice de Desenvolvimento da Educação Básica (IDEB) para escolas de ensino fundamental, com base nas seguintes informações listadas abaixo, referente ao perfil dos professores de cada instituição.

- X1 = Gênero
- X2 = Idade
- X3 = Tempo de serviço
- X4 = Número de escolas em que leciona
- X5 = Frequência de participação na capacitação

Portanto, foi medido o impacto que as variáveis atreladas exclusivamente ao perfil dos professores tinham no IDEB. Na segunda etapa de modelagem (M2), treinou-se um modelo para prever o IDEB das escolas de ensino fundamental, com base nas seguintes informações listadas abaixo, referentes ao perfil de infraestrutura física das escolas.

- X6 = Presença de sala de leitura
- X7 = Abastecimento de água
- X8 = Presença de laboratório de informática

Como na modelagem anterior, analisou-se o impacto que as mudanças na infraestrutura física das escolas causam no IDEB. Na terceira etapa da modelagem (M3), foi treinado um modelo para prever o IDEB das escolas de ensino fundamental, com base nas seguintes informações listadas abaixo, referentes ao perfil tecnológico das escolas.

- X9 = Presença de DVD
- X10 = Presença de projetor
- X11 = Presença de impressora
- X12 = Presença de copiadora
- X13 = Presença de TV
- X14 = Presença de computadores para alunos
- X15 = Existência de internet banda larga

Como nas modelagens anteriores, foi analisado o impacto que mudanças no perfil tecnológico das escolas causam no IDEB. Finalmente, treinou-se o modelo (M4), com as variáveis de melhor performance nos modelos M1, M2 e M3.



Para a primeira instância de modelagem foi utilizado um modelo do tipo *Light Gradient Boosting Machine* (LightGBM), para as segunda e terceira instância de modelagem foram utilizados modelos do tipo *Extra Tree Regressor* (ETR). Estes são algoritmos do tipo *ensemble*, ou seja, a predição final desse tipo de modelo é baseada na combinação de N algoritmos base, chamados de Base *Learners* (BL), que ficam dentro do modelo *ensemble*. Em problemas de classificação, os BL decidem, por voto da maioria, qual classe/categoria será o *output* do modelo. Já em problemas de regressão, a média das predições dos BL é o *output* final do modelo.

Todos os experimentos de modelagem da etapa de análise dos dados foram padronizados, tendo como variável alvo o IDEB do ensino fundamental menor, um conjunto de treino correspondente a 90% dos dados disponíveis, restando 10% dos dados para teste em validação cruzada. Todos os preditores foram normalizados através do método Z-score, o qual também é utilizado para remoção de *outliers* de maneira padrão.

Como métrica de performance dos *LightGBM* e ETR foi utilizado o R-quadrado, ou coeficiente de determinação. Se trata de uma medida estatística que indica quanto da variação da variável dependente é explicada pelas variáveis independentes em um modelo de regressão.

$$R^2 = 1 - (\text{Variação não explicada} / \text{Variação total}).$$

Assim, o R-quadrado explica até que ponto a variação de uma dada variável explica a variação de uma segunda variável. E, finalmente, após as três etapas de modelagem, foi realizada uma última análise com as características de melhor desempenho de cada perfil verificado anteriormente, continuando com o foco no IDEB.

4 RESULTADOS

As modelagens propostas foram desenvolvidas em duas fases. Na primeira fase, desenvolveu-se três modelos preditivos diferentes (M1, M2, M3 e M4) para prever o resultado das dimensões (atributos) no IDEB. Esses três modelos preditivos são ilustrados na Figura 1.

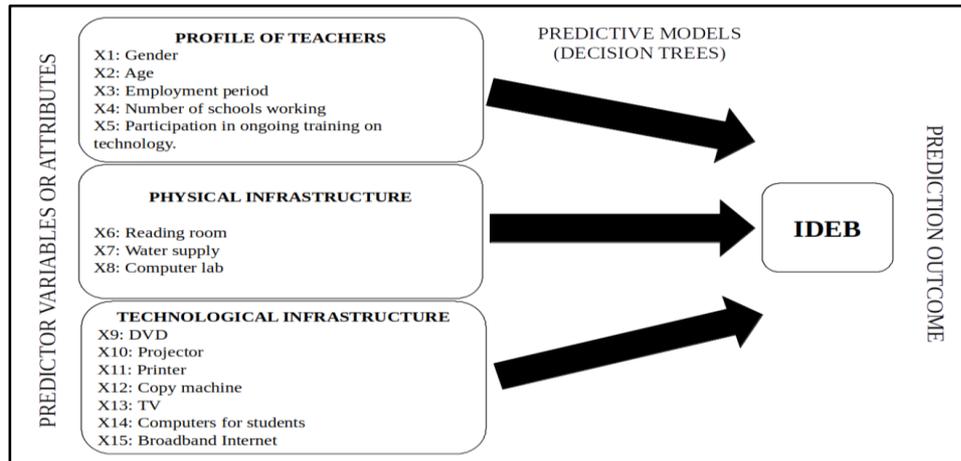


Figura 1. Fase 1, treinamento dos Modelos M1, M2 e M3 com alvo no IDEB.

Conforme ilustrado na Figura 2, na Fase 2 utilizamos as variáveis preditoras de melhor desempenho de cada atributo e repetimos com sucesso o desenvolvimento do modelo preditivo para prever o desempenho do IDEB. A modelagem na Fase 2 foi uma tentativa de examinar conjuntamente os atributos e assim compreender como as dimensões (físicas, estruturais e humanas) interferem no resultado.

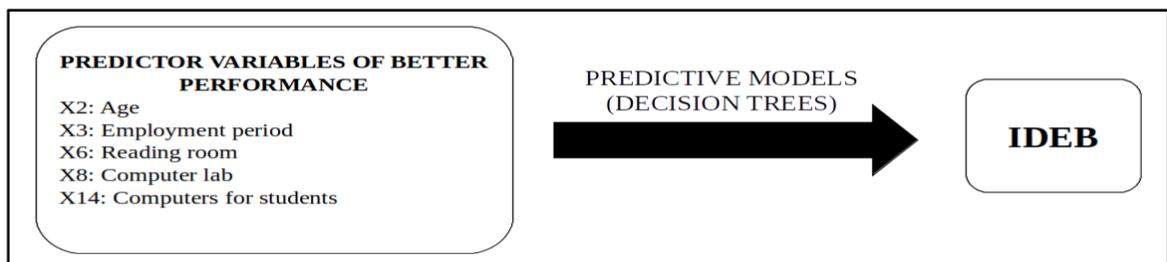


Figura 2. Fase 2, classificação dos preditores de melhor desempenho no resultado.

Em todas as instâncias de modelagem foram comparados os seguintes modelos baseados em árvore de decisão:

- *Decision Tree Regressor*
- *Random Forest Regressor*
- *Extra Trees Regressor*
- *AdaBoost Regressor*
- *Light Gradient Boosting Machine (LightGBM)*

Desses modelos considerados na etapa de modelagem, apenas o que mais se destacou na tarefa de regressão foi selecionado para otimização e consequentemente, para gerar os resultados apresentados neste trabalho.



Para cada experimento (ou instância de modelagem) deste trabalho, foram replicados os mesmos parâmetros de modelagem, para que um resultado confiável fosse atingido.

- I. Conjunto de treino = 90%
- II. Normalização via Z-score
- III. Remoção de *outliers* via Z-score
- IV. K-fold = 8

Os 10% de dados não utilizados no conjunto de treino foram separados para validação do modelo, através do método de validação cruzada (*cross-validation*), utilizado durante o treinamento para evitar problemas relacionados à variância nos dados. Para interpretação dos modelos foram usados o *Shapley Additive Explanations* (SHAP), gráfico da taxa de ganho de informação e os indicadores de performance da regressão. O método SHAP nos ajuda a interpretar modelos de *machine learning* com maior facilidade, através das relações diretas e indiretas, existentes entre os preditores (Batunacun et al., 2021). Valores SHAP são medidas de contribuição que cada preditor tem em um determinado modelo, porém, não só quantificando a importância de cada preditor na tarefa de regressão, como também a direção na relação de causalidade (Batunacun et al., 2021). No caso deste trabalho, a importância se resume ao aumento ou diminuição do valor do IDEB das escolas de ensino fundamental.

Assim, serão apresentados os resultados das etapas de modelagem com base nas fases de treinamento, ou seja, inicialmente será demonstrado os resultados por modelos preditivos.

4.1 MODELO PREDITIVO (M1)

Para a modelagem do perfil dos professores, cinco preditores foram utilizados, tendo dois destes maiores destaques para a tarefa de regressão. A tabela 1, mostra o comparativo de performance entre os modelos baseados em regressão DTs considerados no estudo.

Tabela 1:
Performance dos Modelos DTS de M1

MODELO	MAE	MSE	RMSE	R2	RMSLE	MAPE
Light Gradient Boosting Machine	0.4264	0.3169	0.5602	0.1522	0.0937	0.0870
Random Forest Regressor	0.4120	0.3374	0.5742	0.1039	0.0955	0.0836
AdaBoost Regressor	0.4632	0.3453	0.5856	0.0877	0.0986	0.0955



**Engajando ciência,
gestão e sociedade**

09 a 11 de novembro de 2021

PPAD
Programa de Pós-graduação em Administração

UNAMA
UNIVERSIDADE
DA AMAZÔNIA

ser
educacional

Extra Trees Regressor	0.4216	0.4029	0.6258	-0.0678	0.1040	0.0856
Decision Tree Regressor	0.4364	0.4401	0.6567	-0.1736	0.1086	0.0882

Nota: Células em negrito indicam o melhor valor para cada indicador de performance dos modelos.

A tabela 1, demonstra que o modelo *LightGBM* apresenta melhores resultados para os indicadores de performance no modelo M1, $R^2 = 0.1522$. Em seguida foi realizado o método *Bayesian Optimization* (ver tabela 2), método de otimização de modelos que levam muito tempo até a sua convergência, sendo recomendado para modelos com menos de 20 preditores (Frazier, 2018). Esse procedimento, visa o refinamento dos resultados obtidos com o modelo de melhor desempenho, dessa forma, foi possível a obtenção do valor de $R^2 = 0.2450$, ou seja, as variáveis treinadas no modelo M1, explicam cerca de 24.5% do alvo.

Tabela 2:

Otimização pelo método Bayesiano para M1

MODELO	MAE	MSE	RMSE	R2	RMSLE	MAPE
Light Gradient Boosting Machine	0.4051	0.2854	0.52914	0.2450	0.0883	0.0827

4.2 MODELO PREDITIVO M2

Nesse modelo foram treinados os efeitos de X6, X7 e X8 no alvo (IDEB). A tabela 3 demonstra os indicadores gerais de performance dos modelos DT para a regressão preditiva M2.

Tabela 3:

Performance dos modelos DTs de M2

MODELO	MAE	MSE	RMSE	R2	RMSLE	MAPE
Extra Trees Regressor	0.4786	0.3475	0.5886	0.1588	0.0981	0.0968
Light Gradient Boosting Machine	0.4786	0.3475	0.5886	0.1588	0.0981	0.0968
Decision Tree Regressor	0.4786	0.3475	0.5886	0.1588	0.0981	0.0968
Random Forest Regressor	0.4784	0.3478	0.5889	0.1581	0.0982	0.0966
AdaBoost Regressor	0.4803	0.3516	0.5921	0.1489	0.0988	0.0976

Nota: Células em negrito indicam o melhor valor para cada indicador de performance dos modelos.



Todos os indicadores de ajuste e desempenho foram mais bem previstos pelo modelo *Extra Trees Regressor* - ETRs, com um valor inicial de $R^2 = 0.1588$. O procedimento de otimização do ETRs também foi realizado (tabela 4), obtendo-se $R^2 = 0.2221$.

Tabela 4:
Otimização pelo método Bayesiano para M2

MODELO	MAE	MSE	RMSE	R2	RMSLE	MAPE
Extra Trees Regressor	0.4031	0.2850	0.5302	0.2221	0.0884	0.0814

4.3 MODELO PREDITIVO M3

Para modelagem do perfil tecnológico das escolas foram utilizados sete preditores (X9, X10, X11, X12, X14 e X15), tendo dois deles os maiores destaques (X10 e X14). Outro aspecto importante é que a variável X15 não obteve importância preditiva. A tabela 5, mostra o comparativo de performance entre os modelos de DTs.

Tabela 5:
Performance dos modelos DTs de M3T

MODELO	MAE	MSE	RMSE	R2	RMSLE	MAPE
Extra Trees Regressor	0.2877	0.1884	0.4286	0.4658	0.0719	0.0588
Decision Tree Regressor	0.2876	0.1894	0.4296	0.4618	0.0721	0.0588
Random Forest Regressor	0.2956	0.1900	0.4305	0.4612	0.0722	0.0605
Light Gradient Boosting Machine	0.3773	0.2351	0.4815	0.3327	0.0809	0.0770
AdaBoost Regressor	0.4058	0.2693	0.5171	0.2313	0.0868	0.0833

Nota: Células em negrito indicam o melhor valor para cada indicador de performance dos modelos.

Após a otimização do modelo ETRs, observou-se que houve pouca melhora dos indicadores de performance, conforme mostra a tabela a seguir.

Tabela 6:
Otimização pelo método Bayesiano para M3

MODELO	MAE	MSE	RMSE	R2	RMSLE	MAPE
Extra Trees Regressor	0.2896	0.194	0.4391	0.483	0.0735	0.0591
		9		5		

4.4 MODELO PREDITIVO M4

Nesta última etapa, para a análise de modelagem preditiva M4, foram utilizados os dois melhores preditores de cada um dos atributos das modelagens anteriores, sendo utilizadas as variáveis (X2, X3, X6, X8, X10 e X14). Finalmente, a tabela 7 mostra o comparativo de performance entre os modelos de DTs resultantes das variáveis de melhor performance dos atributos.

Tabela 7:
Performance dos modelos DTs de M4

MODELO	MAE	MSE	RMSE	R2	RMSLE	MAPE
Random Forest Regressor	0.3337	0.237	0.4840	0.366	0.0800	0.0662
		3		0		
Extra Trees Regressor	0.3008	0.241	0.4844	0.356	0.0800	0.0596
		4		6		
Light Gradient Boosting Machine	0.4074	0.271	0.5182	0.278	0.0857	0.0816
		1		5		
Decision Tree Regressor	0.3095	0.272	0.5166	0.271	0.0853	0.0611
		5		6		
AdaBoost Regressor	0.2126	0.238	0.4870	0.243	0.0830	0.0553
		0		2		

Nota: Células em negrito indicam o melhor valor para cada indicador de performance dos modelos.

Como o modelo *Random Forest Regressor* – RFR obteve os melhores resultados em todos os indicadores de desempenho, foi feita a otimização pelo método de Otimização Bayesiana.

Tabela 8:
Otimização pelo método Bayesiano para M4

MODELO	MAE	MSE	RMSE	R2	RMSLE	MAPE
Random Forest Regressor	0.3098	0.1720	0.4102	0.3771	0.0696	0.0633



Os resultados mostrados na tabela 8, demonstram o valor de $R^2 = 0.3771$, portanto, as variáveis treinadas para o modelo de predição com alvo final sendo o IDEB, explicam 37.71% dos efeitos desejados.

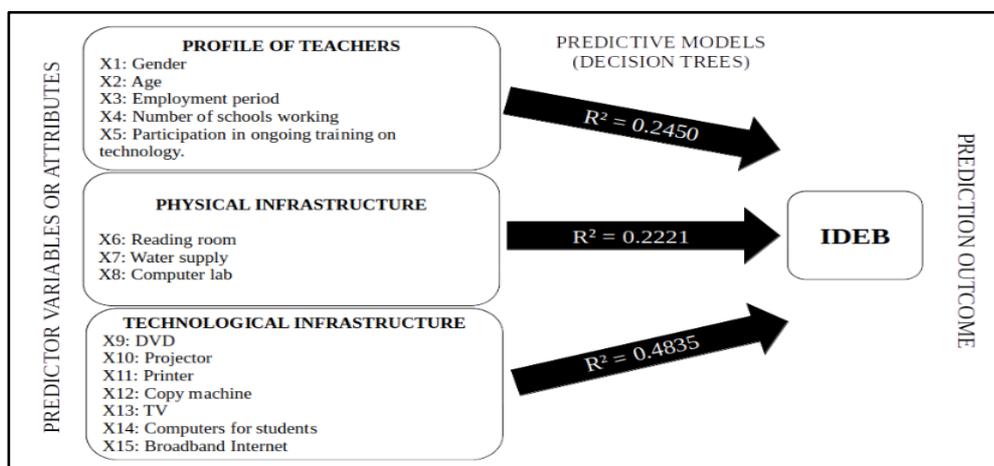


Figura 2. Resultados de R^2 do treinamento dos Modelos M1, M2 e M3 com alvo no IDEB.

Os resultados apresentados na figura 2, demonstram que a infraestrutura tecnológica teve maior performance de explicação das variáveis no alvo. Sendo que os outros atributos analisados obtiveram desempenho semelhante na explicação do IDEB.

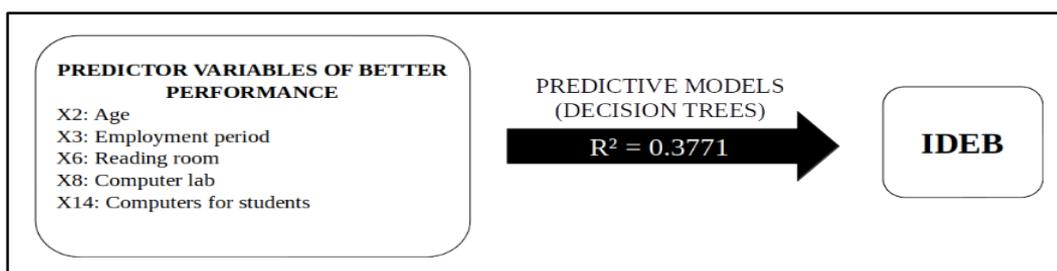


Figura 3. Resultados de R^2 do treinamento do modelo M4.

Por fim, a figura 3 apresenta a performance das variáveis treinadas no alvo, sendo que o modelo final M4, apresentou um grau de explicação de 37,71% do total de efeitos possíveis no IDEB - 2019.

4.5 DISCUSSÃO DOS RESULTADOS

Inicialmente, o atributo perfil dos professores representado por 5 (cinco) variáveis, explicou 24,50% dos efeitos no IDEB. Esse resultado é congruente aos estudos de Zukowski-Tavares; Fernandes e Luz (2017), os quais inferem que compreender a qualidade da educação básica, bem como o desenvolvimento da



Engajando ciência, gestão e sociedade

09 a 11 de novembro de 2021

XII CODS
COLÓQUIO ORGANIZAÇÕES,
DESENVOLVIMENTO & SUSTENTABILIDADE



aprendizagem dos alunos perpassa pela discussão acerca do perfil pessoal, social e cultural dos docentes envolvidos nesse processo.

As variáveis de melhor performance dentro de M1, foram idade dos professores e tempo de serviço vinculado à secretaria estadual de educação do Pará, sendo que a idade demonstrou efeito difuso quanto à direção no alvo estabelecido. Já o tempo de serviço, o modelo SHAP demonstrado no gráfico 2, comprova que os professores com o menor tempo de serviço causaram efeito positivo no IDEB, esse resultado é compatível com os estudos de Souza e Gouveia (2011), que ressaltam que professores mais jovens tendem a ter mais disposição e criatividade na atuação e ainda que os professores com maior tempo de docência tendem a ter maior desgaste físico e emocional, interferindo dessa forma no seu desempenho e consequentemente, no processo de aprendizagem do aluno.

O atributo infraestrutura física das escolas foi avaliado no modelo M2 por meio de 6 (seis) variáveis, tendo como efeito preditivo otimizado o valor de $R^2 = 0.2221$. Esse resultado é condizente com diversos estudos como os de Barrett et al. (2019), os quais apontam o efeito determinante desse atributo no desempenho dos estudantes, visto que uma infraestrutura adequada proporciona uma sensação de bem-estar, não somente para o estudante, mas também para os docentes, criando assim um ambiente para oportunidades de aprendizado e relações sociais significativas. Nesse atributo, as variáveis com melhor performance foram a presença de sala de leitura e do laboratório de informática. Na análise do modelo SHAP, ficou demonstrado que para ambas as variáveis, a intensidade tem efeito positivo no alvo. Isso significa que essas variáveis influenciam positivamente no IDEB das escolas, sendo esse um aspecto também comprovado em estudos anteriores como os de Borges e Castro (2020), os quais informam que os espaços pedagógicos da escola contribuem com a qualidade da educação, e ainda, que os laboratórios de informática, da mesma forma, contribuem com a inclusão digital dos educandos (Soares; Santos & Rela, 2019).

A infraestrutura tecnológica foi avaliada no modelo preditivo M3, tendo como performance explicativa para o melhor modelo treinado o $R^2 = 0.4835$, esse é um resultado bem contundente e fortemente amparado pela literatura pregressa, como os estudos de Tondeur et al. (2018) e Napal, Peñalva-Vélez & Mendióroz Lacambra (2018), os quais inferem que a inclusão digital e tecnológica dos estudantes tem efeitos significativos na qualidade do ensino.

Finalmente, o modelo preditivo M4 demonstrou que as variáveis de melhor performance dos atributos envolvidos nos modelos preditivos anteriores (M1, M2 e M3), efetivamente constituem variáveis chave para geração de efeito no IDEB das escolas, onde o resultado otimizado apresentou um modelo com $R^2 = 0.3771$, portanto escolas com professores atuando a pouco tempo, com laboratórios de informática e computadores suficientes para atender os alunos, podem ter bons desempenhos no que tange os indicadores de qualidade educacional, como o IDEB. Outros estudos, como os de Carvalho (2018) e Talis (2018), apontam esses fatores como importantes para o desenvolvimento da qualidade educacional nas escolas. Porém, Fauth et al. (2019), apontam que outros aspectos são igualmente importantes, como a



qualificação profissional dos professores, as formações continuadas oferecidas e o suporte pedagógico e tecnológico oferecido nestas.

5 CONCLUSÃO

O objetivo deste estudo foi alcançado à medida que foram analisadas as performances de cada dimensão educacional (atributos) relevantes para o desenvolvimento de uma educação de qualidade. Nesse sentido, as variáveis de melhor desempenho no modelo final, como a idade, demonstra que é necessário manter os professores engajados e atualizados para que este efeito, o qual pressupõe a experiência do docente, tenha efeito positivo no IDEB; por sua vez a variável número de computadores por aluno, âmbito das escolas pesquisadas, demonstrou ter alta performance no alvo, o que aponta a importância e a necessidade de investimento em tecnologias que possibilitem o educando ter papel ativo na construção de seu próprio conhecimento.

As limitações deste estudo estão relacionadas principalmente com a coleta de dados, pois mesmo tendo sido usado três bases de informação que continham dados de todas as escolas do estado, ainda assim devido a falhas no preenchimento de informações dessas bases, foi necessário limitar a amostra final, adotando como critério central escolas estaduais que oferecem o ensino fundamental, anos iniciais. Sendo que a rede estadual de educação tem como principal modalidade de atendimento o ensino médio.

As recomendações estão relacionadas ao uso de outras metodologias, como a qualitativa, a fim de analisar como as variáveis de maior efeito no IDEB podem estar sendo usadas (modelo pedagógico) para potencializar o desempenho dos alunos. Finalmente, evidencia-se que é necessário analisar os efeitos de outras variáveis igualmente importantes, como no caso do perfil dos professores, a situação socioeconômica do docente e por sua vez, no contexto da infraestrutura das escolas, outros espaços pedagógicos, como a biblioteca e laboratórios de ciências; e ainda no caso da infraestrutura tecnológica, são necessários estudos experimentais que comprovem que as capacitações em tecnologia desenvolvidas pelo estado associadas à espaços bem equipados geram melhores resultados no IDEB.

REFERÊNCIAS

- Ajayi, I. A., & Ekundayo, H. T. (2011). Factors determining the effectiveness of secondary schools in Nigeria. *The anthropologist*, 13(1), 33-38
- Aldunate, R., & Nussbaum, M. (2013). Teacher adoption of technology. *Computers in Human Behavior*, 29(3), 519-524.
- Amadi, E. C., & Allagoa, I. C. (2017). Demographic Variables as Determinants of Teachers' Effectiveness in Classroom Management in Secondary Schools in Rivers



State, Nigeria. *International Journal of Innovative Development & Policy Studies*, 5(6), 65-70

Barrett, P., Treves, A., Shmis, T., Ambasz, D., & Ustinova, M. (2019). *The impact of school infrastructure on learning: A synthesis of the evidence*. The World Bank.

Batistello, K. (2020, August). TECNOLOGIAS DIGITAIS DE INFORMAÇÃO E COMUNICAÇÃO (TDICS): INFLUÊNCIAS POSITIVAS E NEGATIVAS NO CONTEXTO ESCOLAR. In *Anais do CIET: EnPED: 2020-(Congresso Internacional de Educação e Tecnologias| Encontro de Pesquisadores em Educação a Distância)*.

Batunacun, Wieland, R., Lakes, T., and Nendel, C.: Using Shapley additive explanations to interpret extreme gradient boosting predictions of grassland degradation in Xilingol, China, *Geosci. Model Dev.*, 14, 1493–1510, <https://doi.org/10.5194/gmd-14-1493-2021>, 2021.

Borges, E. M., & Castro, M. C. P. S. (2020). Qualidade da educação: os desafios de uma escola justa e eficaz. *Educação em Foco*, 23(39), 8-26.

Bruschini, M. C. A., & Amado, T. (1988). Estudos sobre mulher e educação: algumas questões sobre o magistério. *Cadernos de pesquisa*, (64), 4-13.

Cantini, M. C., BORTOLOZZO, A. R. S., FARIA, D. D. S., FABRÍCIO, F. B. V., BASZTABIN, R., & MATOS, E. (2006). O desafio do professor frente as novas tecnologias. In *CONGRESSO NACIONAL DE EDUCAÇÃO* (Vol. 6, pp. 875-883).

Christophersen, K. A., Elstad, E., Turmo, A., & Solhaug, T. (2016). Teacher education programmes and their contribution to student teacher efficacy in classroom management and pupil engagement. *Scandinavian Journal of Educational Research*, 60(2), 240-254.

Codo, W. (Coor.) *Educação: carinho e trabalho*. 4 ed. Petrópolis, Rj: Vozes/ Brasília, Confederação Nacional do Trabalhadores em Educação, Universidade de Brasília, Laboratório de Psicologia do Trabalho, 2006.

da Silva, L. M., & Ciasca, M. I. F. L. (2020). Estrutura física escolar como fator determinante da qualidade na educação em escolas profissionais do Ceará: entre a realidade e o mito. *Research, Society and Development*, 9(7), e642974634-e642974634.

de Carvalho, M. R. V. (2018). Perfil do professor da educação básica. *Relatos de Pesquisa*, (41), 68-68.



**Engajando ciência,
gestão e sociedade**

09 a 11 de novembro de 2021

PPAD
Programa de Pós-graduação em Administração

UNAMA
UNIVERSIDADE
DA AMAZÔNIA

ser
educacional

de Souza, Â. R., & Gouveia, A. B. (2011). Os trabalhadores docentes da educação básica no Brasil em uma leitura possível das políticas educacionais. *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas*, 19, 1-22.

Fauth, B., Decristan, J., Decker, A. T., Buettner, G., Hardy, I., Klieme, E., & Kunter, M. (2019). The effects of teacher competence on student outcomes in elementary science education: The mediating role of teaching quality. *Teaching and Teacher Education*, 86, 102882.

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

Fillipin, G. G., Lobato, B. C., Jacobi, L. F., & Zanini, R. R. (2020). Uma visão sobre o IDEB, suas aplicações e resultados. *Ciência e Natura*, 42, 25.

Fleuri, R. M. (2015). Perfil Profissional Docente no Brasil. *Relatos de Pesquisa*, (40), 84-84.

Frazier, P. I. (2018). A tutorial on Bayesian optimization. *arXiv preprint arXiv:1807.02811*.

G1 Pará. 2021. Disponível em: <https://g1.globo.com/pa/para/noticia/2020/09/15/para-tem-2o-pior-ideb-para-ensino-medio-e-nao-atinge-meta-desde-2009.ghtml>, acesso em 03/10/2020.

Herout, L. (2016). Use of the information and communication technologies with the internet access by the pupils of elementary schools. In *EDULEARN2016: 8th International Conference on Education and New Learning Technologies*.

Hlasna, P., Klímová, B., & Poulova, P. (2017). Use of information and communication technologies in primary education—A case study of the Czech Republic. *International Electronic Journal of Elementary Education*, 9(3), 681-692.

INEP, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Relatório nacional; pesquisa internacional sobre ensino e aprendizagem: Talis 2018: primeira parte. – Brasília; Inep, 2019.

Khanna, L., Singh, S. N., & Alam, M. (2016, August). Educational data mining and its role in determining factors affecting students academic performance: A systematic review. In *2016 1st India international conference on information processing (IICIP)* (pp. 1-7). IEEE.

Klassen, R. M., Durksen, T. L., Al Hashmi, W., Kim, L. E., Longden, K., Metsäpelto, R.-L., et al. (2018). National context and teacher characteristics: Exploring the critical



non-cognitive attributes of novice teachers in four countries. *Teaching and Teacher Education*, 72, 64–74.

Laffey, J. (2004). Appropriation, mastery and resistance to technology in early childhood preservice teacher education. *Journal of research on technology in education*, 36(4), 361-382.

Machado, S. C. (2016). Análise sobre o uso das tecnologias digitais da informação e comunicação (TDICS) no processo educacional da geração internet. *RENOTE-Revista Novas Tecnologias na Educação*, 14(2).

Miranda, H. P., & Russell, M. (2012). Understanding factors associated with teacher-directed student use of technology in elementary classrooms: A structural equation modeling approach. *British Journal of Educational Technology*, 43(4), 652-666.

Modelski, D., Giraffa, L. M., & Casartelli, A. D. O. (2019). Tecnologias digitais, formação docente e práticas pedagógicas. *Educação e Pesquisa*, 45.

Monks, J., & Schmidt, R. M. (2011). The impact of class size on outcomes in higher education. *The BE Journal of Economic Analysis & Policy*, 11(1).

Napal Fraile, M., Peñalva-Vélez, A., & Mendióroz Lacambra, A. M. (2018). Development of digital competence in secondary education teachers' training. *Education Sciences*, 8(3), 104.

Nóvoa, A. (1992). Formação de professores e profissão docente. In: *NÓVOA, A. Os professores e a sua formação*. Lisboa: Dom Quixote, p. 13-33.

Pangeni, KP (2014). Fatores que determinam a qualidade educacional: Desempenho dos alunos em matemática no Nepal. *International Journal of Educational Development*, 34, 30-41.

Pianta, R., Downer, J., & Hamre, B. (2016). Quality in early education classrooms: Definitions, gaps, and systems. *The Future of Children*, 119-137.

Ronfeldt, M., Loeb, S. and Wyckoff., 2013. How teacher turnover harms student achievement. *American Educational Research Journal*, 50(1), pp. 4-36.

Rudin, C., & Wagstaff, K. L. (2014). Machine learning for science and society.

Saviani, D. (2011). Formação de professores no Brasil: dilemas e perspectivas. *Póiesis Pedagógica*, 9(1), 07-19.



Seduc. 2021. Disponível em: http://www.seduc.pa.gov.br/portal/escola/consulta_matricula/RelatorioMatriculas.php, acesso em: 04/10/2020.

Shepherd, D., & Devers, C. J. (2017). Principal perceptions of new teacher effectiveness. *Journal of Education*, 197(2), 37-47.

Soares, D. M. R., Teles, G., Sena, T. B. Q. L., Loureiro, R. C., & De Lima, L. (2018). As tecnologias digitais da informação e comunicação (tdics) na prática docente. *CIET: EnPED*.

Soares, E. M. S., Santos, A. S., & Rela, E. (2019). Práticas docentes mediadoras da aprendizagem: laboratório de informática e dispositivos móveis. *Revista Diálogo Educacional*, 19(61), 739-754.

Soares, J. F., Fonseca, I. C., Álvares, R. P., & GUIMARÃES, R. R. M. (2012). Exclusão intraescolar nas escolas públicas brasileiras: um estudo com dados da Prova Brasil 2005, 2007 e 2009. *Debates ED*, 4, 1-77.

Tondeur, J., Aesaert, K., Prestridge, S., & Consuegra, E. (2018). A multilevel analysis of what matters in the training of pre-service teacher's ICT competencies. *Computers & Education*, 122, 32-42.

Zukowski-Tavares, C., Fernandes, A. A. P., & Luz, E. L. (2017). Perfil docente brasileiro e relatório talis: O caso de uma rede confessional de ensino. *Revista e-Curriculum*, 15(3), 693-712.

Silva, P. V. T., & Ross, P. R. (2019). Dificuldades, dilemas e pontos de tensão no uso da tecnologia: pela formação docente e inclusão sociodigital permanente. *Debates em Educação*, 11(23), 19-35.