

A qualidade da educação para além do IDEB: Um estudo através de técnicas de Mineração de Dados

Renê P. Gusmão¹, Cleonides S. D. Gusmão², Mateus S. S. Dias³

¹ Programa de Pós-Graduação em Ciência da Computação
Universidade Federal de Sergipe (UFS) – São Cristóvão, SE – Brazil

² Departamento de Fundamentação da Educação
Universidade Federal da Paraíba (UFPB) – João Pessoa, PB – Brazil

³ Departamento de Ciência da Computação
Universidade Estadual da Paraíba (UEPB) – Patos, PB – Brazil

rene@dcomp.ufs.br, cleonides_silva@hotmail.com, mateuussilvapb@gmail.com

Abstract. *The Basic Education Development Index (IDEB) refers to national educational performance. The objective of this research is to investigate the IDEB considering a broader context, using Data Mining for this. Data from 96 high schools in Sergipe were considered. The results revealed that the models obtained from the Linear Regression and Support Vector Machine achieved the best accuracy among the considered classifiers. Furthermore, the variables related to physical structure, professional qualification, working conditions and teacher valuation are the most important in explaining the IDEB, according to which ones should be considered in the basis of educational public policies.*

Resumo. *O Índice de Desenvolvimento da Educação Básica (IDEB) refere-se ao desempenho educacional nacional. O objetivo desta pesquisa é investigar o IDEB considerando um contexto mais amplo, utilizando para isso a Mineração de Dados. Foram considerados dados de 96 escolas do Ensino Médio de Sergipe. Os resultados revelaram que os modelos obtidos a partir da Regressão Linear e Máquina de Vetor de Suporte apresentaram maior acurácia entre os classificadores considerados. Ademais, as variáveis relacionadas à estrutura física, qualificação profissional, condições de trabalho e valorização docente são as mais importantes na explicação do IDEB, as quais devem ser consideradas na formulação de políticas públicas educacionais.*

1. Introdução

O Índice de Desenvolvimento da Educação Básica (IDEB) refere-se ao desempenho educacional nacional, que é calculado através da nota dos alunos em exames padronizados (de língua portuguesa e matemática) e do índice de aprovação. Para que o IDEB alcance melhores resultados, essas duas variáveis devem ser levadas em consideração [INEP 2021, Fernandes 2007].

Através desse índice, metas para a educação também são estabelecidas. Essas metas são projetadas de acordo com o desempenho passado da escola e, portanto, são diferentes para cada instituição, exigindo, assim, diferentes esforços para alcançá-las [INEP 2021]. As avaliações utilizadas para chegar ao resultado do IDEB restringem-se às

habilidades em Português e Matemática. Esse índice busca, assim, revelar a qualidade do ensino e pautar as decisões relacionadas à gestão escolar, além de impactar em políticas públicas [Chirinéa and Brandãos 2015].

Apesar disso, apenas os dois elementos levados em consideração para se chegar ao resultado do IDEB são insuficientes para revelar a qualidade da educação. Outras variáveis não incluídas nessa equação devem ser consideradas, tais como: “nível socioeconômico e cultural dos alunos; formação docente; valorização do magistério; condições materiais e imateriais de trabalho; gestão escolar; infraestrutura da escola, e insumos, entre outros” [Chirinéa and Brandãos 2015, p. 461]. Deve-se destacar o papel do Estado no que diz respeito ao investimento nessas instituições para que os aspectos acima citados possam ser melhorados (Ibidem).

A sobrecarga de trabalho docente também deve ser considerada quando o assunto é qualidade da educação. O acúmulo de atividades, o alto número de turmas, disciplinas e alunos interfere diretamente no trabalho do professor, na qualidade das suas aulas, no acompanhamento individualizado dos estudantes e na eficiência do seu ensino [Paiva et al.].

Portanto, este trabalho foi realizado com o intuito de incluir algumas dessas variáveis acima citadas e problematizar à qualidade na educação, levando-se em consideração que essa é uma questão complexa e multideterminada, não relacionada apenas a notas obtidas em testes padronizados, como apontado por [Chirinéa and Brandãos 2015]. Para isso, buscou-se utilizar a Mineração de Dados Educacionais (MDE) para prever o cumprimento da meta projetada do IDEB no estado de Sergipe através de um conjunto de variáveis que se voltam ao contexto mais amplo da educação, como o número de professores, número de alunos por professor, infraestrutura escolar, tipo de contrato de trabalho estabelecido entre a instituição e o docente, entre outras variáveis.

O presente trabalho está estruturado da seguinte forma: a próxima seção apresenta trabalhos relacionados; a seção 3 trata da metodologia do trabalho; a seção 4 trata sobre os resultados encontrados por meio da pesquisa empírica e discute esses resultados com base na literatura existente; e, por fim, a última seção versa sobre as considerações finais.

2. Trabalhos relacionados

Para realizar a revisão de literatura, utilizou-se a base de dados do Google Acadêmico aplicando-se a string de busca: IDEB and “mineração de dados”. As dissertações e teses encontradas não foram analisadas.

A pesquisa realizada por [Pinto et al. 2020] buscou prever as variáveis relacionadas ao bom desempenho no IDEB em escolas públicas municipais de ensino fundamental da cidade de Teotônio-Vilela (AL). Os seguintes algoritmos de classificação foram usados: NaiveBayes, J48, JRip, LibSVM, RandomForest, IBK, OneR e REPTree. As técnicas que apresentaram melhor desempenho foram: OneR, LibSVM e J48. Os autores concluíram que é possível identificar o perfil de alunos que contribui para resultados positivos no IDEB. Vale a pena destacar que os autores não trabalharam diretamente com o valor do IDEB, mas com as notas dos alunos em Português e Matemática, as quais são utilizadas para o cálculo do IDEB.

Uma pesquisa que buscou estudar as características educacionais e a desigualdade na educação nos diversos municípios do Rio Grande do Norte foi realizada por [Diniz et al. 2012]. Para isso, os autores utilizaram o IDEB e, para análise de dados, o algoritmo de indução de Árvore de decisão C4.5. Para os autores, esse estado precisa melhorar em relação a qualidade educacional que está abaixo da meta, focando na estrutura das escolas e na redução das desigualdades sociais.

O estudo realizado por [Júnior et al. 2019] investigou as variáveis que influenciam o IDEB das escolas públicas de Maceió (AL), por meio de técnicas de Mineração de Dados (MD): regressão linear simples e árvore de decisão (J48). A escolaridade dos pais, o incentivo para que o aluno estude e o compromisso do professor foram variáveis que se destacaram em relação ao desempenho do aluno. Nesse estudo, ao aplicar a árvore de decisão, os autores perceberam que a infraestrutura não se destacou em relação ao IDEB, sendo mais importantes nessa análise as ações dos docentes e da gestão escolar. Para eles, a infraestrutura pode não ter se destacado devido a forma como ela é utilizada, sendo importante não apenas ter o recurso material, mas também usá-lo de forma adequada.

A pesquisa realizada por [Pinto et al. 2019b] com alunos do ensino fundamental de escolas municipais de Teotônio Vilela (AL) buscou investigar, por meio de MD, as variáveis que influenciam o desempenho dos estudantes na Prova Brasil. Os autores utilizaram as seguintes técnicas: NaiveBayes, J48, JRip, LibSVM, RandomForest, IBK, OneR e REPTree. Desses algoritmos, o OneR, LibSVM e J48 apresentaram maior acurácia para classificação ($> 98\%$). Atributos que se destacaram nessa pesquisa foram: notas em português, a atenção do professor, a utilização da biblioteca/sala de leitura da escola, a prática de leitura da mãe, a prática de leitura do aluno e fatores socioeconômicos (a casa possuir TV, banheiro, máquina de lavar) [Pinto et al. 2019b].

O estudo realizado por [Pinto et al. 2019a] utilizou técnicas de seleção de atributos para identificar as variáveis relacionadas ao desempenho no IDEB em escolas públicas municipais de Maceió (AL). Os algoritmos utilizados foram: NaiveBayes, J48, JRip, LibSVM, RandomForest, IBK, OneR e REPTree. Para o caso dessa pesquisa, as técnicas que se destacaram e obtiveram 100% de acurácia foram J48, OneR, JRip e LibSVM (p. 1834). As variáveis que se destacaram em relação ao desempenho dos alunos foram: Dependência Administrativa, proficiência em Português e matemática, incentivo dos pais para o estudo, sexo, dever de casa, gosta de português, frequência à biblioteca, localização da escola, escolaridade do pai, se o pai sabe ler e escrever, tempo que o estudante gasta em tarefas domésticas. É importante destacar que, os autores não utilizaram diretamente o IDEB, mas sim as notas dos alunos que são usadas para calculá-lo.

[Verica et al. 2015] realizaram uma pesquisa buscando encontrar perfis municipais no Estado do Paraná no que diz respeito aos dados educacionais. A análise dos dados foi realizada por meio de Agrupamentos de Dados. Os autores não encontraram diferenças em relação ao IDEB entre os municípios. As variáveis que se destacaram quanto a esse índice foram: taxa de analfabetismo, taxa de aprovação e taxa de distorção idade/série. Além disso, nessa pesquisa, o pouco investimento em educação também se destacou como variável associada aos menores IDEBs.

[Santos and Medeiros 2020] investigaram, através da MD, a relação entre os gastos com educação e o IDEB das escolas na Paraíba. Para isso, utilizou-se os algoritmos

de correlação e regressão linear simples. Através das análises realizadas, não foi possível detectar relação linear direta entre o gasto com a educação e o IDEB. Como foi utilizado o IDEB do município e não de escolas separadamente, os autores acreditam que esse agrupamento de dados pode ter gerado distorções e dificultado a identificação da relação com o investimento em educação.

Outros trabalhos acerca de temas relacionados à MDE foram encontrados envolvendo temas como: evasão, identificação do perfil de estudantes, desempenho, entre outros. O diferencial desta pesquisa em relação à literatura encontrada é: a região em que o fenômeno foi estudado (Sergipe), o escopo do estudo que volta-se para o Ensino Médio (EM), o foco do estudo voltado para a qualidade da educação e IDEB, bem como o contexto mais amplo (como condições de trabalho dos professores, estrutura das instituições) incluído para a análise do fenômeno educacional.

3. Metodologia

As etapas deste trabalho foram realizadas seguindo a metodologia *Cross-Industry Standard Process for Data Mining* (CRISP-DM) [Shearer 2000]. Segundo essa metodologia, o processo de MD deve conter seis fases: (1) Entendimento do negócio, (2) Entendimento dos Dados, (3) Preparação dos Dados, (4) Modelagem, (5) Avaliação e (6) Implantação.

3.1. Fase 1 - Entendimento do Negócio

A primeira etapa do CRISP-DM consiste em definir o problema de Mineração de Dados a partir do entendimento do contexto do problema. Tendo o objetivo principal de investigar quais são as variáveis associadas a previsão do IDEB das escolas de Sergipe do Ensino Médio, esta pesquisa teve as seguintes questões norteadoras:

1. Quais modelos de classificação possuem melhor desempenho ao predizer se a escola atingirá a meta do IDEB?
2. Quais as variáveis mais importantes identificadas na predição da meta do IDEB?
3. Há diferença significativa no IDEB de escolas públicas e privadas no EM?
4. Há diferenças na estrutura das escolas públicas e privadas?
5. Há diferenças entre escolas que atingiram a meta e escolas que não atingiram a meta em relação à estrutura das instituições escolares?

3.2. Fase 2 - Entendimento dos Dados

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), vinculado ao Ministério da Educação (MEC), realiza pesquisas e avaliações nos vários níveis de Ensino e disponibiliza¹ diversas bases de dados em formato de Dados Abertos. Neste trabalho, foram utilizados os Microdados do Censo da Educação Básica 2017. Os Microdados da Educação Básica são compostos por Tabelas no formato CSV referentes as Turmas, Escolas, Docentes e Matrículas.

3.3. Fase 3 - Preparação dos Dados

Este trabalho incluiu as escolas de Sergipe que obtiveram nota observada do IDEB em 2017, somando 96 escolas. Cada escola é descrita por 188 atributos. Além das variáveis

¹<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados>

disponibilizadas na Tabela contendo os dados das Escolas, as variáveis derivadas descritas na Tabela 1 foram produzidas. Também foram coletados os dados do IDEB dessas escolas, onde a meta projetada para o Estado em 2017 foi de 4,6. Por fim, uma variável binária de rótulo foi criada para informar se as escolas atingiram a meta em 2017. Portanto, 201 atributos foram usados para descrever as escolas após adicionar as variáveis derivadas.

Variável	Descrição
VD1	Quantidade de Docentes com Licenciatura na Escola
VD2	Quantidade de Docentes com Mestrado na Escola
VD3	Quantidade de Docentes com Doutorado na Escola
VD4	Quantidade de Docentes com Especialização
VD5	Quantidade de Docentes sem Pós-Graduação
VD6	Quantidade de Docentes com Formação Continuada para Ensino Médio
VD7	Quantidade de Docentes Efetivos
VD8	Quantidade de Docentes CLT
VD9	Quantidade de Docentes Temporários
VD10	Quantidade de Docentes Terceirizados
VD11	Quantidade total de docentes na escola
VD12	Quantidade média de alunos por professor na escola
Meta	Informa se o IDEB observado da escola atingiu a meta projetada para Sergipe

Tabela 1. Variáveis derivadas produzidas

A plataforma RapidMiner² foi utilizada como ambiente para limpeza e análise dos dados. Após a construção da base de dados, ocorreu a etapa de limpeza dos dados para remover atributos com alta estabilidade de valores, estes possuem pouca importância para modelos de aprendizagem de máquina, e também atributos com muitos valores ausentes. Após a etapa de limpeza, o tamanho do conjunto de atributos foi reduzido para 89. As variáveis relacionadas ao ensino fundamental, infantil, de jovens e adultos também foram retiradas. Por fim, um subconjunto contendo trinta e três variáveis foi considerado para os experimentos realizados.

3.4. Fase 4 - Modelagem

Foram considerados cinco métodos de máquina supervisionada voltados para o problema de classificação, a saber: Regressão Logística, Árvore de Decisão, Regressão Linear, Naive Bayes e Máquina de Vetor de Suporte (SVM). Na primeira parte dos experimentos, considerou-se o método Holdout [Han et al. 2012], em que 70% dos dados foram usados para o treinamento e 30% para a realização de testes. Na segunda parte dos experimentos, utilizou-se a técnica Validação Cruzada k-folds [Han et al. 2012]. Nos experimentos realizados, foi definido um total de 10 iterações para cada modelo de classificação. A avaliação dos resultados obtidos, referente a Fase 5 do CRISP-DM, será feita na seção 4.

4. Resultados

Os modelos utilizados neste trabalho foram analisados em termos de Acurácia (ACC) [Han et al. 2012]. Além da acurácia, as variáveis identificadas como mais relevantes

²<https://rapidminer.com/>

também são apresentadas. A subseção 4.1 apresenta algumas estatísticas descritivas que servem para responder as questões de pesquisa 3, 4 e 5. A pergunta de pesquisa 1 é respondida na subseção 4.2. A questão de pesquisa 2 é respondida na subseção 4.3. Por fim, a discussão dos resultados é apresentada na subseção 4.4.

4.1. Estatísticas Descritivas

A Tabela 2 apresenta o IDEB médio das escolas que atingiram a meta e das escolas que não atingiram a meta, sendo a diferença significativa entre essas médias ($t = 13,07$; $p \leq 0,01$). Também é possível observar a diferença estatisticamente significativa ($t = 10,59$; $p \leq 0,01$) entre o IDEB médio das escolas privadas e das públicas.

	Atingiram a meta	Não atingiram a meta	IDEB Médio
Escolas privadas	17	3	5,24
Escolas públicas	1	75	3,28
IDEB Médio	5,46	3,28	

Tabela 2. IDEB médio

Para as escolas que atingiram a meta, a média de professores com licenciatura é 21,33 (dp = 13,19), a média de docentes com especialização é 6,56 (dp = 5,32), a média de docentes com mestrado é 2,39 (dp = 5,12) e a média de docentes com doutorado é de 0,50 (dp = 2,12). Para as escolas que não atingiram a meta, a média de professores com licenciatura é 20,44 (dp = 8,70), a média de docentes com especialização é 9,26 (dp = 5,68), a média de docentes com mestrado é 1,60 (dp = 1,54) e a média de docentes com doutorado é de 0,19 (dp = 0,60).

Das 18 escolas que atingiram a meta, 10 (55,56%) delas possuem sala de leitura, 17 (94,44%) delas possuem biblioteca e/ou sala de leitura, 12 (66,67%) delas possuem biblioteca, 10 (55,56%) delas possuem laboratório de informática, 7 (38,89%) delas possuem laboratório de ciências. Das 78 escolas que não atingiram a meta, 27 (34,62%) delas possuem sala de leitura, 62 (79,49%) delas possuem biblioteca e/ou sala de leitura, 49 (62,82%) delas possuem biblioteca, 68 (87,18%) delas possuem laboratório de informática, 27 (34,62%) delas possuem laboratório de ciências.

4.2. Comparação entre modelos de classificação

A Tabela 3 apresenta o desempenho dos métodos nos dois experimentos. Considerando-se os resultados obtidos usando validação cruzada, a Máquina de Vetor de Suporte obteve desempenho superior em comparação aos demais, tendo uma acurácia média de 95.89% e desvio padrão de 7.11%. O modelo com menor desempenho foi o Naive Bayes com 90.67% de acurácia média e 12.65% de desvio padrão.

Para o método Holdout, os métodos de Regressão Linear e Máquina de Vetor de Suporte obtiveram um desempenho superior em termos de acurácia quando comparado aos demais métodos. Por sua vez, os demais algoritmos alcançaram o mesmo desempenho, com o mesmo nível de acurácia. No entanto, as variáveis mais influentes identificadas pelos modelos foram distintas.

Modelo	Validação Cruzada		Holdout
	Acurácia média	Desvio padrão	
Naive Bayes	90.67%	12.65%	89.29%
Regressão Logística	92.89%	10.63%	89.29%
Regressão Linear	93.78%	8.86%	92.86%
Árvore de Decisão - Gain Ratio	90.89%	11.02%	89.29%
Árvore de Decisão - Gini Index	93.78%	10.04%	89.29%
Árvore de Decisão - Information Gain	92.67%	9.89%	89.29%
Máquina de Vetor de Suporte	95.89%	7.11%	92.86%

Tabela 3. Acurácias com Validação Cruzada k-Folds e Holdout

4.3. Variáveis relevantes identificadas

Na Tabela 4, observa-se a frequência das variáveis com maior peso entre os algoritmos. É possível perceber que as variáveis quantidade de docentes com mestrado e quantidade de docentes com especialização apresentaram frequência igual a quatro e que, portanto, foram apresentadas por quatro modelos entre as variáveis com maiores pesos. É importante frisar as condições de trabalho e valorização do professor, visto que o tipo de contrato de trabalho, formação continuada em mestrados e especializações são atributos que se destacaram em quase todos os modelos preditivos. Outro importante aspecto é a infraestrutura, destacando-se, especialmente, a disponibilidade de alimentação escolar para os estudantes, de laboratório de ciências e o número de computadores. A existência de Biblioteca e Sala de Leitura nas escolas, apesar de serem citadas com menor frequência, possuem destaque em alguns modelos avaliados.

Variável	Frequência	Descrição	Contexto
QtdeDocentesMest	4	No. de Docentes com Mestrado	Formação Docente
QtdeDocentesEsp	4	No. de Docentes com Especialização	Formação Docente
QtdeDocentesSemPós	3	No. de Docentes sem Especialização	Formação Docente
QtdeDocentesEfetivos	3	No. de Docentes efetivos	Recursos Humanos
QtdeDocentesCLT	2	No. de Docentes com contrato CLT	Recursos Humanos
RelAlunProf	1	No. de alunos por professor	Condições de trabalho
QtdeAlunosTotal	1	No. de alunos total estudando	Condições de trabalho
QtdeTurmasProf	1	No. de turmas por professor	Condições de trabalho
IN-Alimentação	4	Alimentação escolar para os alunos	Infraestrutura da escola
Lab. de Ciências	3	Lab. de Ciênc. disponível	Infraestrutura da escola
NU-COMPUTADOR	2	No. de computadores	Infraestrutura da escola
NU-SALAS-EXISTENTES	1	No. de Salas	Infraestrutura da escola
Sala de Leitura	1	Sala de leitura disponível	Infraestrutura da escola
Lab. de Informática	1	Lab. de Info. disponível	Infraestrutura da escola
IN-Biblioteca	1	Existência de biblioteca na escola	Infraestrutura da escola

Tabela 4. Variáveis relevantes

A Figura 1 apresenta os pesos dos atributos de acordo com o modelo Máquina de Vetor de Suporte, tendo em vista que esse algoritmo foi o que obteve melhor desempenho entre os demais. Em suma, variáveis relacionadas à valorização docente, formação docente, condições de trabalho e infraestrutura institucional foram as mais relevantes, considerando-se todos os modelos utilizados.

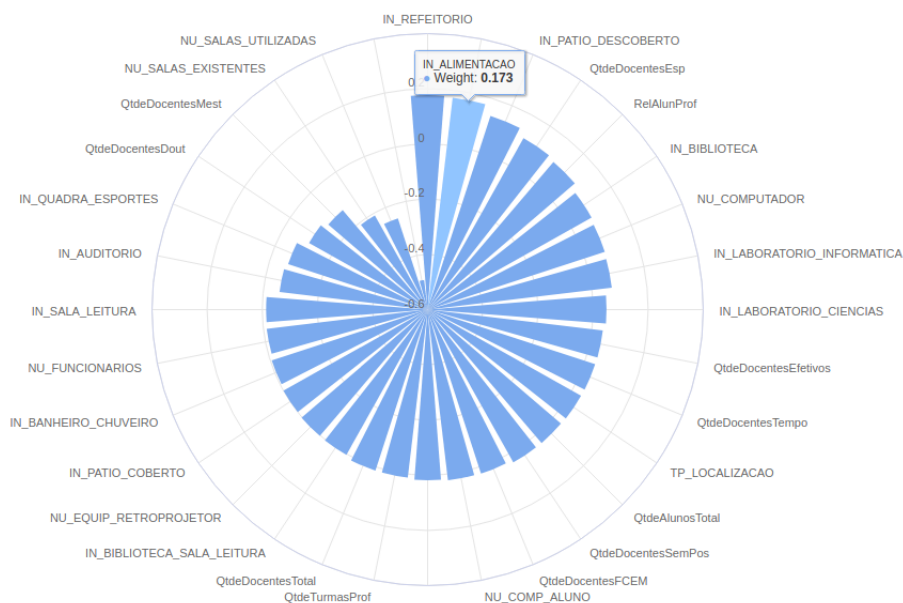


Figura 1. Pesos dos atributos apresentados pela Máquina de Vetor de Suporte

4.4. Discussões

No geral, para a base de dados utilizada no presente trabalho, os algoritmos alcançaram melhor desempenho com a utilização do método de validação cruzada. É possível afirmar que o algoritmo mais adequado para prever o IDEB em escolas do terceiro ano do ensino médio é o Máquina de Vetor de Suporte, independente do método utilizado. Este resultado vai ao encontro da pesquisa realizada por [Pinto et al. 2020] que destacou o desempenho do algoritmo LibSVM para realizar a previsão das notas de português e matemática. O mesmo aconteceu na pesquisa realizada por [Pinto et al. 2019b] que buscou prever o desempenho dos estudantes na Prova Brasil; e na pesquisa realizada por [Pinto et al. 2019a] que buscou prever o desempenho dos alunos por meio dos dados do teste Saeb e Prova Brasil.

Em relação às diferentes árvores de decisão utilizadas na presente pesquisa, a árvore com índice Information Gain também obteve bom desempenho utilizando-se o método de validação cruzada, apesar de não ter sido o melhor entre os algoritmos. Já para o método Holdout, todas as árvores de decisão, junto à regressão logística e Naive Bayes, obtiveram o mesmo desempenho e não se destacaram como melhores. [Diniz et al. 2012] utilizou o algoritmo de indução de árvore de decisão C4.5 em seu trabalho com o objetivo de caracterizar a educação e analisar as desigualdades em relação a essa área em diferentes municípios e também obteve bom desempenho (87,5%). [Júnior et al. 2019, Paiva et al. 2013] também utilizaram a árvore de decisão nos seus respectivos trabalhos, mas não discutiram o desempenho dessa técnica.

A regressão logística e a regressão linear, para o método de validação cruzada, alcançaram um bom desempenho. Para o método Holdout, a Regressão Linear, apesar de ter se destacado como um dos melhores desempenhos, não conseguiu alcançar a acurácia obtida no método de validação cruzada, considerando-se o mesmo algoritmo. De forma semelhante, [Calixto et al. 2017] obtiveram bom desempenho ao utilizar a regressão logística para investigar variáveis associadas a evasão escolar. [Júnior et al. 2019]

e [Santos and Medeiros 2020] também utilizaram a regressão linear em suas pesquisas de MDE, mas não analisaram o desempenho desse algoritmo.

Em relação ao algoritmo Naive Bayes ele obteve o pior desempenho para o método de Validação Cruzada e não se destacou no método Holdout. Resultados semelhantes foram obtidos por [Pinto et al. 2020]; [Pinto et al. 2019a] e [Pinto et al. 2019b].

Ao se comparar as escolas que atingiram a meta com as que não atingiram, é possível notar a diferença em relação à estrutura das intuições. Aquelas possuem uma maior porcentagem de salas de leitura, biblioteca e/ou sala de leitura e laboratório de ciências. Já as que não atingiram a meta possuem uma porcentagem maior de bibliotecas e laboratórios de informática. No entanto, é necessário investigar a qualidade desses espaços, bem como a forma como eles são utilizados, como já apontado por [Júnior et al. 2019].

Em relação as variáveis que mais se destacaram nas análises, é possível afirmar que o investimento em educação no que diz respeito a estrutura material, qualificação profissional, condições de trabalho e valorização docente são questões de fundamental importância, impactando na qualidade da educação. Nesse sentido, a estrutura física da instituição (quantidade de computadores e salas existentes, possuir laboratórios de informática e ciências, biblioteca, sala de leitura), a qualificação profissional (quantidade de docentes com mestrado e especialização) e as condições de trabalho as quais o profissional é submetido (como o tipo de contrato de trabalho) são variáveis importantes no que diz respeito ao IDEB.

5. Considerações finais

Os resultados deste estudo revelaram a importância da estrutura material e imaterial, e do investimento em educação. Nesse sentido, políticas públicas que se voltem para a melhoria desses aspectos são de extrema importância quando a qualidade da educação está em discussão.

Em concordância com o que foi apontado por [Chirinéa and Brandãos 2015], é importante que uma variável dessa complexidade não seja problematizada levando-se em consideração apenas variáveis individuais, sendo de vital importância a ampliação do papel do Estado em relação ao investimento na estrutura escolar e nos profissionais da educação no que diz respeito a valorização profissional, salarial e de qualificação.

Além disso, ao encontro de [Paiva et al.], a sobrecarga de trabalho somada a desvalorização social da profissão docente são questões que não podem ser ignoradas quando se trata de educação e, especialmente, da qualidade da educação.

Referências

- Calixto, K., Segundo, C., and Gusmão, R. (2017). Mineração de dados aplicada a educação: um estudo comparativo acerca das características que influenciam a evasão escolar. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, 28(1):1447.
- Chirinéa, A. M. and Brandãos, C. F. (2015). O ideb como política de regulação do estado e legitimação da qualidade: em busca de significados. *Ensaio: Avaliação e Políticas Públicas em Educação*, 23:461–484.

- Diniz, F. A., Silva, F. R. M., Costa, R. D., Silva, T. R., Moura, I. B. G., and Castro, A. F. (2012). Análise espacial dos Índices educacionais do rio grande do norte com o uso de técnicas de mineração de dados. In *Anais VI Simpósio Regional de Geoprocessamento e Sensoriamento Remoto - Geonordeste*.
- Fernandes, R. (2007). Índice de desenvolvimento da educação básica (ideb). Technical report, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.
- Han, J., Kamber, M., and Pei, J. (2012). *Data mining concepts and techniques, third edition*. Morgan Kaufmann Publishers.
- INEP (2021). Resultados do Índice de desenvolvimento da educação básica: 2019: resumo técnico. Technical report, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.
- Júnior, O. G. F., Rodrigues, W. R. M., Barbirato, J. C. C., and Costa, E. B. (2019). Melhoria da gestão escolar através do uso de técnicas de mineração de dados educacionais: um estudo de caso em escolas municipais de maceió. *RENOTE*, 17:296–305.
- Paiva, R., Bittencourt, I., Pacheco, H., Silva, A. A. D., Jacques, P., and Isotani, S. Mineração de dados e a gestão inteligente da aprendizagem: desafios e direcionamentos. In *Anais do Workshop de Desafios da Computação Aplicada à Educação (desafie) 2012*.
- Paiva, R., Bittencourt, I. I., and Silva, A. P. (2013). Uma ferramenta para recomendação pedagógica baseada em mineração de dados educacionais. *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, 2(1).
- Pinto, G., Júnior, O. F., and Costa, E. (2020). Mineração de dados educacionais: Um modelo de predição do perfil do aluno para melhoria do ideb. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, pages 1172–1182, Porto Alegre, RS, Brasil. SBC.
- Pinto, G. S., Júnior, O. F., Costa, E., Barbirato, J. C. C., and Rodrigues, W. R. M. (2019a). Identificação dos fatores de melhorias no ideb pelo uso de mineração de dados: Um estudo de caso em escolas municipais de maceió. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, 30(1):1828.
- Pinto, G. S., Júnior, O. G. F., and Costa, E. B. (2019b). Identificação dos fatores de melhorias no ideb pelo uso de mineração de dados: Um estudo de caso em escolas municipais de teotônio vilela-alagoas. *RENOTE*, 17:183–193.
- Santos, A. and Medeiros, F. P. A. (2020). Relação do financiamento federal com os resultados do ideb em um estado do brasil: uma abordagem baseada em mineração de dados educacionais. In *2020 15th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–4.
- Shearer, C. (2000). The crisp-dm model: the new blueprint for data mining. *Journal of data warehousing*, 5(4):13–22.
- Verica, W. R., Villwock, R., and Johann, J. A. (2015). Uso de técnicas de mineração de dados para agrupamento e espacialização de dados educacionais no estado do paraná. *Anais XVII Simpósio Brasileiro de Sensoriamento Remoto - SBSR*, pages 1753–1760.