

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**ANÁLISE DOS FATORES RELACIONADOS
AO DESEMPENHO DAS ESCOLAS NO IDEB:
ESTUDO DE CASO NO ESTADO DO PARÁ**

Vitor Hugo Macedo Gomes

DM 07/2022

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil

2020

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Vitor Hugo Macedo Gomes

**ANÁLISE DOS FATORES RELACIONADOS
AO DESEMPENHO DAS ESCOLAS NO IDEB:
ESTUDO DE CASO NO ESTADO DO PARÁ**

DM 07/2022

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil

2020

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Vitor Hugo Macedo Gomes

**ANÁLISE DOS FATORES RELACIONADOS AO
DESEMPENHO DAS ESCOLAS NO IDEB: ESTUDO
DE CASO NO ESTADO DO PARÁ**

Dissertação submetida à Banca Examinadora do Programa de Pós-Graduação em Engenharia Elétrica da UFPA para a obtenção do Grau de Mestre em Engenharia Elétrica na área de Computação Aplicada.

Orientador: Prof. Dr. Marcelino Silva da Silva

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil

2020

**Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)**

G633a Gomes, Vitor Hugo Macedo.
Análise dos Fatores Relacionados ao Desempenho das Escolas
no IDEB: Estudo de Caso no Estado do Pará / Vitor Hugo Macedo
Gomes. — 2022.
72 f. : il. color.

Orientador(a): Prof. Dr. Marcelino Silva da Silva
Dissertação (Mestrado) - Universidade Federal do Pará,
Instituto de Tecnologia, Programa de Pós-Graduação em
Engenharia Elétrica, Belém, 2022.

1. IDEB, SAEB, Censo Escolar, Mineração de Dados,
Regressão Linear. I. Título.

CDD 006.312



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**“ANÁLISE DOS FATORES RELACIONADOS AO DESEMPENHO DAS
ESCOLAS NO IDEB: ESTUDO DE CASO NO ESTADO DO PARÁ”**

AUTOR: VITOR HUGO MACEDO GOMES

DISSERTAÇÃO DE MESTRADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

APROVADA EM: 11/02/2022

BANCA EXAMINADORA:

Prof. Dr. Marcelino Silva da Silva
(Orientador – PPGEE/UFPA)

Prof. Dr. Carlos Renato Lisboa Francês
(Avaliador Interno – PPGEE/UFPA)

Prof. Dr. Jorge Antônio Moraes de Souza
(Avaliador Externo – UFRA)

Prof.^a Dr.^a Liviane Ponte Rego
(Avaliadora Externa – UFOPA)

VISTO:

Prof. Dr. Carlos Tavares da Costa
(Coordenador do PPGEE/ITEC/UFPA)

*Dedico este trabalho aos meus pais que sempre
me deram todo o apoio para que eu trilhasse o caminho
dos estudos.*

Agradecimentos

Agradeço a Deus que sempre me ajudou a superar todas as dificuldades.

Agradeço ao meu Orientador Prof. Dr. Marcelino Silva da Silva pelos ensinamentos, paciência, orientação, inspiração e pela contribuição para minha formação.

À minha mãe, Maria de Fátima Macedo Gomes, pelo exemplo de pessoa, pela determinação, pelo incentivo aos estudos.

Ao meu Pai Walter Sampaio Gomes, pelos incentivos e apoio.

As minhas irmãs Jaqueline e Suany pela amizade e companheirismo.

Aos amigos e em especial Priscila Siqueira, Jonatã Paulino e Áurea Milene, que sempre me incentivaram na elaboração desta pesquisa. Obrigado pelo apoio e amizade!

Agradeço pela amizade e por todas as contribuições dos meus amigos do Laboratório de Planejamento de Redes de Alto Desempenho da Universidade Federal do Pará.

À Universidade Federal do Pará (UFPA), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)¹ e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)² pelo apoio financeiro em forma de bolsa.

À todos os professores do programa de pós-graduação em Engenharia Elétrica (PPGEE³) pelos ensinamentos.

À Universidade Federal do Pará (UFPA), pela oportunidade dada a mim para a realização desta dissertação.

¹<http://www.cnpq.br/>

²<http://http://www.capes.gov.br/>

³<http://www.ppg ee.ufpa.br>

*“Eu não tenho talentos especiais
Estou apenas apaixonadamente curioso
(Albert Einstein)”*

Sumário

1	Introdução	1
1.1	Motivação e caracterização do problema	1
1.2	Objetivos	3
1.3	Organização do Trabalho	4
2	Referencial teórico	5
2.1	Considerações iniciais	5
2.2	Índices Educacionais	5
2.2.1	Histórico	5
2.2.2	Exame Nacional do Ensino Médio - ENEM	6
2.2.3	Censo Escolar	7
2.2.4	Sistema de Avaliação da Educação Básica - SAEB	7
2.2.5	O Índice de Desenvolvimento da Educação Básica - IDEB	8
2.3	Índice de Desenvolvimento Humano Municipal - IDHM	10
2.4	Índice de Vulnerabilidade Social - IVS	11
2.5	Prosperidade Social	11
2.6	Mineração de Dados	12
2.6.1	Mineração de Dados Educacionais - MDE	13
2.7	Randon Florest - RF	14
2.8	Coeficiente de Correlação Linear	15
2.9	Regressão Linear Simples	16
2.10	Considerações finais	18
3	Trabalhos Correlatos	19
3.1	Considerações Iniciais	19
3.2	Correlatos	19
3.2.1	Análise dos trabalhos correlatos e contribuições desta dissertação	21
3.2.2	Considerações Finais	24
4	Metodologia	25
4.1	Considerações iniciais	25
4.2	Metodologia empregada	25
4.3	Pré-processamento	26
4.4	Extração de Padrões	29
4.4.1	Randon Florest - RF	29
4.4.2	Regressão Linear	30

4.5	Pós-processamento	31
4.6	Considerações finais	32
5	Estudo de caso e Discussão dos resultados	33
5.0.1	Considerações iniciais	33
5.0.2	Estudo de Caso	33
5.0.2.1	Metas para educação	33
5.0.2.2	Renda Familiar dos Estudantes	35
5.0.2.3	Infraestrutura Escolar	38
5.0.2.4	Auxílio dos Pais na preparação dos filhos	40
5.0.2.5	IDEB, IDHM e IVS	41
5.1	Precisão do Modelo de Regressão Linear	49
5.2	Considerações finais	50
6	Conclusão	52
6.1	Trabalhos futuros	53
6.2	Dificuldades encontradas	53
	Referências	55

Lista de ilustrações

Figura 1	Nota média de proficiência em Português e Matemática no SAEB na Região Norte.	2
Figura 2	Meta a ser atingida pelo Pará.	3
Figura 3	Meta IDEB	9
Figura 4	Faixas do IDHM	11
Figura 5	Faixas do IVS	11
Figura 6	Construção das faixas do IDHM e IVS	12
Figura 7	Mineração de Dados Educacionais	13
Figura 8	Randon Florest	15
Figura 9	Correlação linear	16
Figura 10	Representação dos parâmetros β_0 e β_1	17
Figura 11	Diagrama com a metodologia empregada	26
Figura 12	Importância das Variáveis	30
Figura 13	Índices educacionais e socioeconômicos no estado do Pará.	35
Figura 14	Situação Socioeconômica dos Participantes do ENEM	36
Figura 15	Média das faixas de renda pela média das notas em matemática de cada município	37
Figura 16	Infraestrutura escolar	38
Figura 17	Nível escolar dos pais	40
Figura 18	IDHM 2010	41
Figura 19	Índices globais do IDHM e IVS	42
Figura 20	IVS - 2010	42
Figura 21	Evolução da nota de proficiência x IDHM x IVS	43
Figura 22	Abandono escolar x IDHM x IVS	45
Figura 23	Professores do Grupo 1 x IDHM x IVS	46
Figura 24	Professores do Grupo 3 x IDHM x IVS	47
Figura 25	Professores do Grupo 5 x IDHM x IVS	49

Lista de tabelas

Tabela 1	Diferenças encontradas entre os artigos publicados na literatura e este trabalho	22
Tabela 2	Estrutura básica dos atributos estudados e sua fonte	27
Tabela 3	Categorias de adequação da formação dos docentes em relação à disciplina que leciona	28
Tabela 4	Nível de escolaridade dos pais	28
Tabela 5	Métricas de avaliação.	50

Lista de abreviaturas e siglas

AG	Algoritmo Genéticos
ANEB	Avaliação Nacional da Educação Básica
DM	Data Mining
ENEM	Exame Nacional do Ensino Médio
IDEB	Índice de Desenvolvimento da Educação Básica
IDH	Índice de Desenvolvimento Humano
IDHM	Índice de Desenvolvimento Humano Municipal
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
IVS	Índice de Prosperidade Social
KDD	Knowledge Discovery in Databases
MD	Mineração de Dados
MDE	Mineração de Dados Educacionais
MEC	Ministério da Educação
PIB	Produto Interno Bruto
PISA	Programa Internacional de Avaliação dos Estudantes
RF	Radiofrequência
RF	Randon Florest
RL	Regressão Linear
RNB	Renda Nacional Bruta
RQNP	Algoritmo de Regressão Quântica
SAEB	Sistema de Avaliação da Educação Básica
SISU	Sistema de Seleção Unificada
TRI	Teoria de Resposta ao Item

Resumo

A complexidade em identificar todos os fatores que estão relacionados ao desempenho das escolas no Índice de Desenvolvimento da Educação Básica (IDEB) é enorme. Neste estudo foram analisados três bases de dados com o objetivo de identificar diversos fatores em que há correlação com o baixo desempenho nas escolas estaduais no estado do Pará. Inicialmente observou-se por meio das análises que 142 municípios do estado estavam com risco de descumprimento da meta no que se refere a diminuição da evasão escolar e, conseqüentemente, afetando o desempenho das escolas. Este estudo utilizou técnicas de mineração de dados educacionais para, primeiramente, selecionar variáveis com características estruturais no ambiente de ensino, comparando as escolas com maior e menor desempenho no IDEB identificando possíveis relações com a evasão escolar. Em seguida, foi utilizado o algoritmo Random Forest (RF), para selecionar as variáveis mais importantes e que impactam de forma direta ou indireta no índice do IDEB. Após a fase de seleção, as variáveis foram submetidas ao algoritmo de Regressão Linear (RL). Os resultados revelam que no grupo de escolas abaixo da média no IDEB, 60,6% residem em famílias com rendimentos até um salário-mínimo, enquanto que 37,5% possuem rendimentos acima de um salário-mínimo. No grupo de escolas acima da média no IDEB, 42,4% residem em famílias com rendimentos até um salário-mínimo, enquanto que 51,6% residem em famílias com rendimentos acima de um salário-mínimo. Evidenciando que a renda das famílias está relacionada à melhores notas no IDEB e, conseqüentemente, melhores condições de infraestrutura. Os resultados também apontam que os rendimentos das famílias dos estudantes estão relacionados a renda média das famílias nos municípios analisados. Em seguida foram utilizadas variáveis relacionadas ao rendimento dos pais para identificar possível relação entre a escolaridade dos pais e desempenho dos alunos. Por fim, as análises se encerram com a análise do impacto do Índice de Desenvolvimento Humano Municipal (IDHM) nas variáveis que relacionadas à nota dos alunos, qualificação dos professores e experiência vivenciada pelos professores no ambiente escolar. Os resultados revelam que há correlação entre o índice e o aprendizado dos estudantes em sala de aula. Por outro lado, melhores notas no IDEB, estão diretamente relacionadas a adequação do currículo à disciplina ministrada, além de boas condições de trabalho para os professores.

Palavras-chave: IDEB, SAEB, Censo Escolar, Mineração de Dados, Regressão Linear.

Abstract

The complexity of identifying all the factors that are related to the performance of schools on the Basic Education Development Index (IDEB) is enormous. In this study, three databases were analyzed with the objective of identifying several factors that correlate with low performance in state schools in the state of Pará. Initially, it was observed through the analysis that 142 municipalities in the state were at risk of not meeting the goal regarding the reduction of school dropouts and, consequently, affecting the performance of schools. This study used educational data mining techniques to, first, select variables with structural characteristics in the teaching environment, comparing the schools with higher and lower performance in IDEB, identifying possible relationships with school dropouts. Then, the Random Forest (RF) algorithm was used to select the most important variables that directly or indirectly impact the IDEB index. After the selection phase, the variables were submitted to the Linear Regression (LR) algorithm. The results reveal that in the group of schools below average in IDEB, 60.6% reside in families with incomes up to one minimum wage, while 37.5% have incomes above one minimum wage. In the group of schools above average in IDEB, 42.4% live in families with incomes up to one minimum wage, while 51.6% live in families with incomes above one minimum wage. Evidencing that family income is related to better IDEB scores and, consequently, better infrastructure conditions. The results also indicate that the income of students' families is related to the average family income in the analyzed municipalities. Next, variables related to parents' income were used to identify a possible relationship between parents' schooling and students' performance. Finally, the analysis ends with the analysis of the impact of the Municipal Human Development Index (HDI) on the variables related to the students' grades, the teachers' qualifications, and the teachers' experience in the school environment. The results reveal that there is a correlation between the index and student learning in the classroom. On the other hand, better IDEB scores are directly related to the adequacy of the curriculum to the subject taught, in addition to good working conditions for teachers.

Keywords: IDEB, SAEB, Censo Escolar, Mineração de Dados, Regressão Linear.

1 Introdução

1.1 Motivação e caracterização do problema

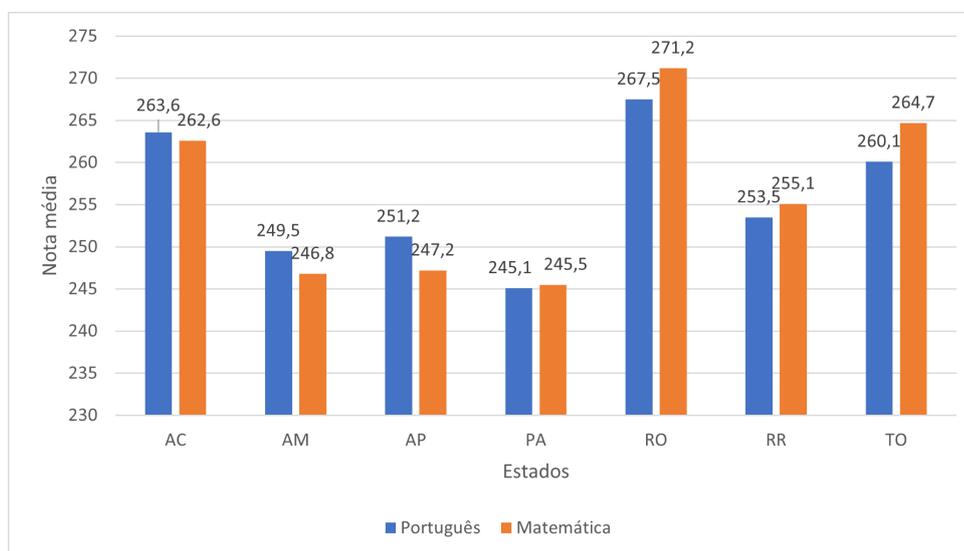
O INEP, vinculado ao Ministério da Educação (MEC), é responsável pela aplicação do Pisa no Brasil, realizado a cada três anos, o Pisa tem o objetivo de mensurar até que ponto os jovens adquiriram conhecimento e habilidades essenciais para a vida social e econômica. Em questão territorial, os índices nas regiões são: Sul com índice 419; Centro-Oeste com 415; Sudeste 414; Norte com 384; Nordeste com índice de 383. Norte e nordeste possuem os piores números (FERNANDES; GREMAUD, 2009). De acordo com o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira INEP (2019). O programa Internacional de Avaliação de Estudantes (PISA), apontou que entre os 78 países participantes da avaliação, o Brasil tem baixa proficiência em leitura, matemática e ciências. A edição de 2018 aponta que 68,1% dos estudantes com 15 anos de idade não possuem nível básico de matemática, o mínimo para o exercício pleno da cidadania. Em ciências, o número chega a 55%, em leitura 50%. Tais índices estão estagnados desde 2009.

Além dos dados de proficiência, outros fatores associados podem estar impactando nos resultados. Situações em que quanto mais rico social, cultural e economicamente o estudante for, maiores são as oportunidades de acesso à educação, e por consequência, o melhor desempenho escolar. Um em cada dez estudantes pobres, na faixa etária de 15 anos, acreditam que não vão concluir o ensino superior. Já nos casos dos ricos, a média é de 1 em cada 25. Os resultados apontados pelo PISA são importantes para avaliar o país como um todo, mas não são aplicáveis para medir o desempenho das escolas e dos estudantes (FERNANDES; GREMAUD, 2009).

O Sistema de Avaliação da Educação Básica (SAEB) permite ao INEP realizar um diagnóstico da educação básica brasileira e de fatores que podem influenciar no desempenho do estudante. Por meio de testes e questionários que são aplicados a cada dois anos na rede pública e por amostra da rede privada, o SAEB reflete os níveis de aprendizagem demonstrados pelos estudantes avaliados, possibilitando as escolas, avaliar a qualidade da educação oferecida aos estudantes. Nesse sentido, o resultado é um indicativo da qualidade do ensino e fornece subsídios para elaboração de políticas públicas educacionais com base em evidências (MEC, 2020b). Conforme ilustrado na Figura 1, o estado do Pará tem menor proficiência média em português e matemática na região norte. Apenas Rondônia com 271,2 pontos, tem índice acima da média nacional, que é 269,74 em matemática. Há uma diferença de 38 pontos entre o estado com maior e menor proficiência: Espírito Santo

com 283,7 pontos e Pará com 245,1 pontos; Em matemática, a diferença é de 46,1 pontos, Espírito Santo com 291,6 pontos e Pará com 245,5 pontos (MEC, 2020b).

Figura 1 – Nota média de proficiência em Português e Matemática no SAEB na Região Norte.



Fonte: (INEP, 2017)

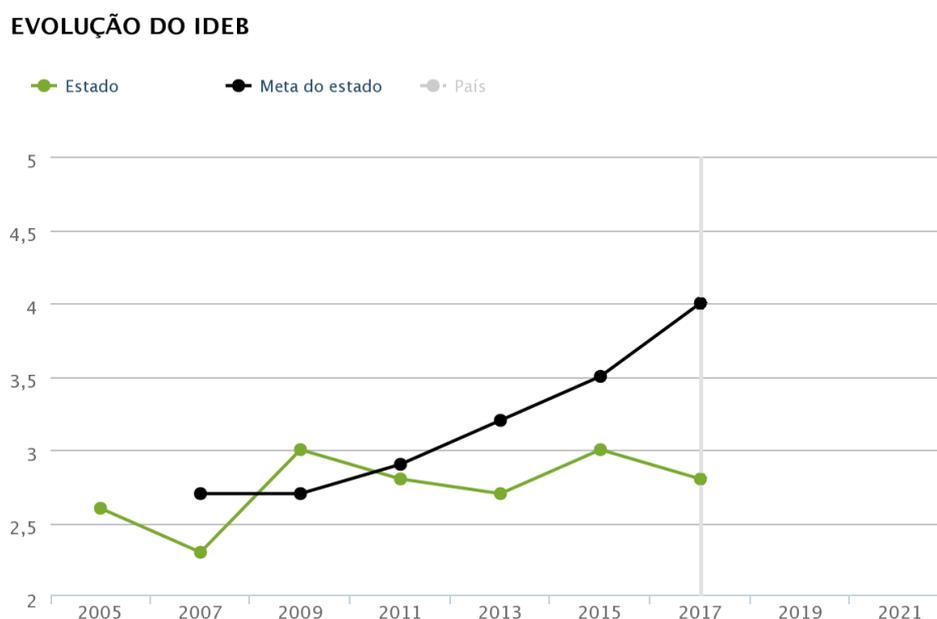
Para o entendimento das questões que são determinantes para melhoria do sistema educacional, é importante observar alternativas que possam detectar escolas com baixo desempenho e analisar os processos que levam ao aluno ter baixos índices de aprendizagem. Deste modo, O INEP, vem produzindo indicadores que possibilitam a análise da qualidade do ensino e aprendizagem. No intuito de fiscalizar e acompanhar a evolução desses indicadores educacionais, foi criado o Índice de Desenvolvimento da Educação Básica (IDEB), cujo objetivo é a avaliação do sistema educacional de acordo com as metas preestabelecidas pelo governo. (ALVES; SOARES, 2013). Segundo Fernandes e Gremaud (2009), o IDEB é o principal indicador utilizado para o monitorar a qualidade da educação básica no país, e tem entre outros, o objetivo garantir que os alunos que passam de ano, realmente passaram por um processo de aprendizagem e, dessa forma, prosseguir para próxima etapa de ensino. De acordo com o MEC (2020a), o IDEB é calculado a cada dois anos, a partir dos dados sobre aprovação escolar obtidos por meio do Censo Escolar e das médias de desempenho no Sistema de Avaliação da Educação Básica (SAEB). O índice é um importante subsídio para a adoção de política pública em prol da qualidade da educação. Tornando-se uma importante ferramenta para acompanhamento das metas estabelecidas para a educação básica.

Dentro desse contexto, uma forma eficaz de contribuir com essa questão é identificar variáveis que possam dar subsídios à bons resultados no IDEB, analisando características como: infraestrutura escolar, perfil socioeconômicos dos alunos, desempenho e

outros fatores, por meio do uso de ferramentas computacionais para uso em Mineração de Dados.

Na literatura existem alguns trabalhos que analisam o IDEB no ensino fundamental das escolas brasileiras. Entretanto, pesquisas no Estado do Pará são inexistentes tanto para o ensino fundamental quanto para o ensino médio, mesmo possuindo um dos piores desempenhos no ensino do país, pois a meta a ser atingida para o estado era de 4,0, no entanto, alcançou média de 2,8 em 2017 (MEC, 2020), conforme mostrado pela Figura 2. À vista disso, o uso da Mineração de Dados Educacionais (MDE), constitui em uma importante abordagem para a obtenção de diversas informações referentes à situação da educação no estado do Pará.

Figura 2 – Meta a ser atingida pelo Pará.



Fonte: QEDu.org.br. Dados do Ideb/Inep (2019)

Os dados supracitados indicam a necessidade de se traçar estratégias utilizando mineração de dados para apontar os diferentes fatores que podem estar contribuindo com essa questão.

1.2 Objetivos

A presente pesquisa tem por objetivo geral utilizar ferramentas de mineração de dados, visando identificar variáveis relacionadas ao desempenho do IDEB nas escolas estaduais públicas do estado do Pará e associar ao Índice de Desenvolvimento Humano Municipal (IDHM) de 2010, último ano de divulgação. Dessa forma, revelando um pouco

da complexidade que é, identificar a diversidade de fatores e de que maneira estão relacionados, esta dissertação apresenta os seguintes objetivos específicos:

- Realizar análise exploratória em diferentes dados públicos;
- Identificar os padrões relacionados à diferentes fatores, e que estão de forma intrínseca e extrínseca relacionadas ao ambiente escolar;
- Utilizar algoritmo para seleção de variáveis a fim de identificar quais estão fortemente relacionadas para, a seguir, submeter ao algoritmo de regressão linear revelando a força da correlação entre essas variáveis;

1.3 Organização do Trabalho

Esta dissertação encontra-se organizada em 6 capítulos. O capítulo 1 relatou a motivação e a caracterização do problema:

- Capítulo 2: Neste capítulo traz uma explicação teórica a cerca dos índices utilizados e das tecnologias utilizadas para o desenvolvimento desta dissertação .
- Capítulo 3: Apresentam-se os trabalhos relacionados ao estudo de caso desta dissertação.
- Capítulo 4: Apresenta-se a metodologia empregada, começando pela aquisição, preparação dos dados e métodos para análise estatística e identificação de padrões mediante utilização do algoritmo.
- Capítulo 5: Neste capítulo, apresenta-se os resultados obtidos do estudo de caso proposto.
- Capítulo 6: Por fim, são apresentadas as conclusões obtidas por meio da realização deste trabalho, dificuldades encontradas e possíveis desdobramentos para trabalhos futuros.

2 Referencial teórico

2.1 Considerações iniciais

Este capítulo tem por finalidade conceituar teoricamente os principais elementos necessários para desenvolvimento deste trabalho.

2.2 Índices Educacionais

2.2.1 Histórico

Nas últimas décadas, políticas de avaliação educacional tem avançado de forma abrangente no país. Tendo como foco, uma análise vasta e em larga escala, visando o desempenho do aluno no sistema de educação básica pública. A política de avaliação vem sendo implementada em diferentes níveis educacionais com programas de avaliação específicos, possibilitando os pesquisadores, analisar a eficiência do sistema de ensino e divulgar os resultados. Seja em publicações científicas ou em mídias para que os resultados que porventura sejam insatisfatórios, possam nortear políticas de aprimoramento de ações de melhoria na aprendizagem (COELHO, 2008).

Vale ressaltar que nos anos 80 já se discutia propostas para avaliar a educação no Brasil, mas foi na década de 90 que, de fato, o país teve um sistema de avaliação de aprendizagem. Até meados da década de 90, não se tinha um indicador que avaliasse a aprendizagem no país. Segundo Guimarães et al. (2009), o que havia antes, era uma escola pública excludente, cuja qualidade era acessível para poucos eleitos. Atualmente o problema enfrentado no país, é a disponibilização de um sistema de ensino público de qualidade e democrático para todos.

Ainda nos anos 70, havia outras abordagens sobre a forma de avaliação do ensino, onde eram analisadas características sociais e organizacionais das escolas, conhecidas como “Clima Escolar”, cujo objetivo era encontrar características comuns nessas escolas, onde o aluno obtinha resultado superior ao que era esperado com base nas características socioeconômicas (COELHO, 2008). Segundo Brookover et al. (1978), o conceito de “Clima Escolar” tem sido usado de diversas maneiras. Uma delas, é fazendo análise do corpo docente de uma determinada instituição: situação econômica, raça, cor e outros indicadores. Pesquisas analisando a personalidade do discente também foram realizadas; alguns professores sugerem que, cada aula tem personalidade distinta, “Clima” que pode influenciar

na eficácia da aprendizagem de seus membros.

Partindo para a consolidação de um sistema de avaliação de aprendizagem no país. Em meados dos anos 90, foi criado o Sistema Nacional de Avaliação Básica (SAEB), um sistema de avaliação para aferição do rendimento escolar dos alunos nos ensinos fundamental e médio das redes de ensino de todas as regiões do Brasil (GUIMARA et al., 2009). O SAEB se consolidou como um importante subsídio para que se possa determinar políticas públicas voltadas para o desenvolvimento educacional. A partir de 1995, além da rede particular, outras análises foram incorporadas ao exame; características socioeconômicas e culturais dos alunos (COELHO, 2008).

Nesse sentido, a qualidade da educação passou a ser pautada pelas notas obtidas por meio dos resultados do SAEB e Prova Brasil, que diferentemente do SAEB, é o exame aplicado aos alunos nos anos iniciais e finais do ensino fundamental. Deste modo, dando um indicativo de uma possível qualidade educacional ofertada pelas escolas, por meio de exames de proficiência realizado nas disciplinas de Português e Matemática com o objetivo de mensurar habilidades cognitivas dos participantes (CHIRINÉA; BRANDÃO, 2015).

E nesse processo de evolução dos índices para mensurar a qualidade da educação, que é criado o Índice de Desenvolvimento da Educação Básica IDEB em 2007, pelo INEP, onde as notas são obtidas por meio do SAEB e as taxas de aprovação escolar. Além de ser um importante indicador da qualidade da educação brasileira, identificando escolas que estão mais frágeis e assim propor um plano estratégico para a melhoria da qualidade de ensino. (HADDAD, 2008).

2.2.2 Exame Nacional do Ensino Médio - ENEM

O ENEM, foi criado em 1998 para avaliar o desempenho dos alunos ao término da educação básica. Em 2009 passou a ser utilizado como mecanismo de acesso ao ensino superior, por meio do Sistema de Seleção Unificada (SISU).

O exame é dividido em quatro áreas de conhecimento:

- Linguagens, códigos e suas tecnologias.
- Ciências humanas e suas tecnologias.
- Ciências da natureza e suas tecnologias.
- Matemática e suas tecnologias.

Os itens supracitados somam o total de 180 questões, além da redação que exige do aluno um texto dissertativo-argumentativo.

Os participantes do ENEM recebem um questionário socioeconômico onde são feitas perguntas com o objetivo de coletar informações sobre: opiniões dos participantes sobre assuntos gerais, membros da família, situação socioeconômica, e planos para o futuro.

2.2.3 Censo Escolar

O principal meio para coleta de informações sobre a educação básica utilizado pelo INEP, é o Censo Escolar. Sendo uma colaboração entre união, estados e municípios. De acordo com a Portaria MEC nº 316, de 4 de abril de 2007. O Censo abrange diferentes etapas de ensino (INEP, 2007):

- Ensino regular (educação infantil, ensino fundamental e médio).
- Educação especial – modalidade substitutiva.
- Educação de Jovens e Adultos (EJA).
- Educação profissional (cursos técnicos e cursos de formação inicial continuada ou qualificação profissional).

Os dados coletados do censo são referentes as características das instituições de ensino: situação das turmas, alunos, formação dos professores, condições físicas do prédio, equipamentos existentes, pessoal técnico e administrativo. Deste modo, sendo de fundamental importância ao fornecer dados para melhor compreensão da situação do ensino do país. Dando subsídio ao planejamento de metas para a educação que devem ser cumpridas até 2024 (INEP, 2007). Todas as escolas respondem o Censo Escolar, que é feito por meio de um questionário padronizado.

2.2.4 Sistema de Avaliação da Educação Básica - SAEB

O SAEB, foi realizado pela primeira vez em 1990, com o objetivo de avaliar a qualidade da educação oferecida aos estudantes. Por ser um indicativo da qualidade da educação brasileira, pode-se utilizá-lo para subsidiar políticas públicas que visam corrigir falhas no processo de aprendizagem dos estudantes, com base em evidências. Os resultados variam de 0 a 500, variando em uma escala de desempenho capaz de descrever as habilidades dos estudantes em cada nível (INEP, 2020).

Ao longo dos anos o SAEB vem aprimorando seu processo de avaliação. Em 1990, quando o exame foi aplicado pela primeira vez, o público-alvo era somente estudantes do 1^a, 3^a, 5^a e 7^a séries do ensino fundamental em uma amostra das escolas públicas. Tendo como disciplina alvo: Língua portuguesa, matemática, ciências naturais e redação.

Em 1995, na terceira edição, foi adotada uma nova metodologia chamada de Teoria de Resposta ao Item (TRI). Deste modo, foi possível a comparação de resultados ao longo do tempo. Sendo aplicado na 4^a, 8^a do ensino fundamental e 3^a série do ensino médio. Em 1997, o exame começou a avaliar os estudantes por meio de escalas de proficiência, abrangendo as disciplinas de Língua portuguesa, matemática, ciências (física, química e biologia)(INEP, 2020). Em 1999, a novidade se dava por meio da adição da disciplina de geografia. Em 2001, o SAEB passa a aplicar os testes somente em língua portuguesa e matemática. Em 2003, de forma consolidada o SAEB mantém o mesmo formato da edição anterior. Mas em 2005, o exame sofre uma reestruturação pela Portaria Ministerial n^o 931, de 21 de março. Sendo composto por duas avaliações: Avaliação Nacional da Educação Básica (ANEB) e Avaliação Nacional do Rendimento Escolar(ANRESC), mais conhecida como Prova Brasil. O formato lançado em 2005, permitiu ao INEP combinar as médias de desempenho dos estudantes que participaram do SAEB, com as taxas de aprovação, reprovação e abandono, identificadas por meio do Censo Escolar, e deste modo, calcular o IDEB (INEP, 2020). Segundo o (MEC, 2020), os resultados do exame variam de 0 a 500. Os resultados são categorizados em uma escala de nível habilidade que o estudante demonstra. Os critérios para a divulgação dos resultados são: 1 – a participação de no mínimo 10 alunos presentes; 2 – alcançar participação de 80% (oitenta por cento) dos estudantes matriculados, conforme declarado pela escola.

2.2.5 O Índice de Desenvolvimento da Educação Básica - IDEB

O IDEB, foi criado em 2007 com o objetivo de avaliar a qualidade da educação brasileiras, utilizando indicadores de desempenho dos estudantes e sintetizado em resultados globais de qualidade da educação, tornando-se a única forma de avaliar a educação brasileira. Melhores resultados no IDEB estão de acordo com bons rendimentos dos estudantes e diminuem quando as taxas de aprovação também diminui. O IDEB se tornou um importante e respeitável indicador pelo fato de agregar em um único índice, dimensões fundamentais para relevantes análises no sistema da educação básica (SOARES; XAVIER, 2013).

De acordo com o INEP (2020), para calcular o IDEB de uma determinada escola, é preciso obter as notas de proficiências das disciplinas de português e matemática que estão padronizadas na escala de 0 (zero) a 10 (dez). Após a obtenção das notas, multiplica-se pelas taxas de aprovação (ensino médio), que vão de um percentual de 0 (zero) a 100 (cem). A equação 1 ilustra como é calculado as notas do IDEB.

$$IDEB = N_{ij}P_{ij} \quad (1)$$

Onde:

i = ano do exame (Saeb e Prova Brasil) e Censo Escolar;

N_{ij} = Média da proficiência em Língua Portuguesa e Matemática, padronizada para o indicador de 0 a 10, dos alunos da unidade j , obtida em determinada edição do exame realizado ao final da etapa de ensino;

P_{ij} = Indicador de rendimento baseado na taxa de aprovação da etapa de ensino dos alunos a unidade j ;

O IDEB tem como objetivo identificar se as metas estabelecidas pelo Termo de Adesão ao Compromisso Todos pela Educação, do Ministério da Educação foram cumpridas. Através dos exames de proficiência, mede-se a qualidade educacional por meio dos rendimentos dos alunos. Deste modo, espera-se que após o lançamento das notas, o Brasil alcance a meta de 6,0. Atingindo o nível de qualidade educacional estabelecido (INEP, 2020). Para o cálculo da meta a ser atingida, assume-se que as trajetórias têm uma função logística.

Figura 3 – Meta IDEB



Fonte: (INEP, 2020)

A Figura 3, ilustra a trajetória para o alcance do índice por meio das quatro informações apontadas: IDEB inicial observado; valor da meta para o IDEB; tempo para atingir a meta e "esforço" ou velocidade empregada. A partir dessas informações, pode-se calcular as métricas intermediárias para o IDEB, a partir da equação 2.

$$IDEB_{it} = \frac{1}{1 + e^{-\left(\ln\left(\frac{ideb_{i0}}{10 - ideb_{i0}}\right) + \gamma_i \cdot t\right)}} \quad (2)$$

Onde o t : 0,...16 para metas da 1ª fase do ensino fundamental; 0,...20 para metas

da 2ª fase do ensino fundamental 0,...23 para metas do ensino médio; tempo em anos, desde o ano do IDEB inicial.

- i : município, UF, Brasil, rede de ensino ou escola
- $idebit$: valor do Ideb no ano t para determinado i
- $idebi0$: IDEB inicial ($t=0$) para determinado i
- γ_i : esforço individual

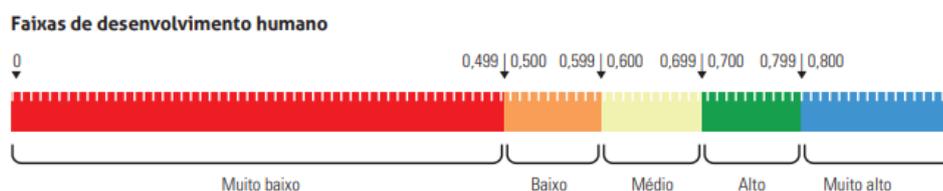
2.3 Índice de Desenvolvimento Humano Municipal - IDHM

Em (IPEA, 2017), o IDH foi criado para contrapor o Produto Interno Bruto (PIB), que tem como objetivo mensurar a dimensão econômica do desenvolvimento. O IDH é uma medida que é constituída por três pilares:

- Vida longa e saudável (longevidade): Ter uma vida longa e saudável é fundamental para uma vida plena. A ampliação do desenvolvimento humano requer melhores oportunidades de para que as pessoas possam evitar morte prematura, e de garantir a elas um ambiente saudável, com acesso à saúde de qualidade, para que possam atingir o padrão mais elevado de saúde física e mental.
- Acesso ao conhecimento (educação): O acesso ao conhecimento é importante para o bem-estar e importante para o exercício das atividades individuais. A educação é fundamental para expandir habilidades das pessoas para que elas possam decidir seu futuro, lhes dando confiança, dignidade e ampliação dos horizontes e as perspectivas de vida.
- Padrão de vida (renda): A renda é essencial para que as pessoas possam ter acesso a necessidades básicas como água, comida, abrigo, mas também ir além dessas necessidades lhes dando uma vida de escolhas. A renda nos possibilita diversos fins e sua ausência pode limitar as oportunidades de vida

O IDHM é feito a partir de um ajuste do IDH para avaliar a situação socioeconômica dos municípios brasileiros, mantendo as três dimensões do IDH, acrescentando três componentes: IDHM Longevidade; IDHM Educação; IDHM Renda (PNUD, 2014). A faixa do IDHM varia entre 0 e 1. Quanto mais próximo de 1, maior o desenvolvimento humano de um município. Conforme ilustrado na Figura 4 (IPEA, 2017).

Figura 4 – Faixas do IDHM



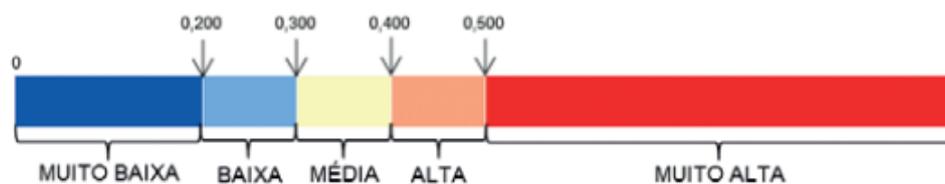
Fonte: (IPEA, 2017)

2.4 Índice de Vulnerabilidade Social - IVS

O IVS pode ser utilizado de forma complementar ao que o IDHM retrata. Dando destaque a indicadores que refletem condições menos favoráveis de inserção social: capital humano; mercados de trabalho; condições de moradia; infraestrutura urbana (IPEA, 2017).

De acordo com (IPEA, 2017) o IVS varia entre 0 e 1. Quanto mais próximo de 1, maior é a vulnerabilidade social de um município: Municípios com IVS entre 0 e 0,200 estão na faixa muito baixa; Valores entre 0,201 e 0,300 estão na faixa de vulnerabilidade baixa; IVS entre 0,301 e 0,400 são de média vulnerabilidade; Valores entre 0,401 e 0,500 são considerados de alta vulnerabilidade; Os que apresentam valores entre 0,501 e 1 são considerados municípios com muito alta vulnerabilidade social. Conforme ilustrado na Figura 5

Figura 5 – Faixas do IVS



Fonte: (IPEA, 2017)

2.5 Prosperidade Social

O cruzamento entre as faixas do IDHM e IVS oferece a oportunidade de analisar o que se denomina *Prosperidade social*. Uma sociedade considerada próspera possui índices de desenvolvimento humano alto com baixa vulnerabilidade social. Quando há essa relação inversa entre os dois índices em uma sociedade, ocorre um desenvolvimento humano menos vulnerável e socialmente mais próspera (IPEA, 2017).

A Figura 6 ilustra o cruzamento entre as faixas do IDHM e IVS.

Figura 6 – Construção das faixas do IDHM e IVS

		IDHM		
		Baixo/muito baixo	Médio	Alto/muito alto
IVS	Baixa/muito baixa	Média	Alta	Muito alta
	Média	Baixa	Média	Alta
	Alta/muito alta	Muito baixa	Baixa	Média

Fonte: (IPEA, 2017)

2.6 Mineração de Dados

A prática de se encontrar padrões em uma grande quantidade de dados surgiu da preocupação em extrair conhecimento de grandes quantidades de dados armazenados nas empresas e organizações. Apesar de ter recebido outros nomes, o termo mineração de dados - *Data Mining* (DM) é comumente utilizado pelos estatísticos, analistas de dados, entre outros (AMO, 2004; FAYYAD et al., 1996). De acordo com (FAYYAD et al., 1996), a Mineração de Dados é apenas uma parte específica de um processo global chamado *Knowledge Discovery and Data Mining* (KDD). Dentro do KDD, a mineração de dados se caracteriza pelo uso de algoritmos para reconhecimento de padrões que podem ser de classificação, regressão, associação, segmentação, clusterização e outros. A classificação, que tem como objetivo mapear um conjunto de dados, identificando padrões e adicionar à um conjunto de classes pré-definidas. Deste modo, para qualquer outra variável nova que surja, poderá ser feita a previsão com base nos padrões encontrados para definir à qual classe irá pertencer. Como exemplo de algoritmos de classificação: Redes Neurais, *Back-Propagation*, Classificadores Bayesianos, Árvore de Decisão e Algoritmos Genéticos. Nos algoritmos de regressão utilizam funções lineares ou não, que são aplicadas em variáveis com valores contínuos (GALVÃO; MARIN, 2009).

Em vista disso, o processo de mineração de dados tem como característica a seleção das variáveis que serão utilizadas pelo modelo preditivo, sendo uma importante estratégia para definição dos dados que serão utilizados. A dificuldade está em selecionar as variáveis corretamente para reduzir a dimensionalidade dos dados e o custo computacional para analisar grandes quantidades de variáveis. A evolução dos sistemas que fazem a coleta

destes dados e o grande volume armazenado faz com que este processo se torne cada vez mais complexo, demandando diferentes estratégias para melhor aproveitamento dos dados. (NISBET; ELDER; MINER, 2009).

2.6.1 Mineração de Dados Educacionais - MDE

A MDE, está preocupada em resolver problemas relacionados a detecção de padrões estudantis em grandes repositórios de dados. Pois um dos maiores problemas é que o crescimento exponencial dos dados educacionais torna difícil a detecção e consequentemente a tomada de decisão. A mineração de dados educacionais, (do inglês *Educational data mining* (EDM)), se tornou nos últimos anos uma área de pesquisa para analisar dados relacionados a educação, que são gerados nas diferentes instituições de ensino no país, envolvendo disciplinas como: Ciência da computação, educação e estatística. Através da intercessão essas diferentes áreas é que surge a MDE (ROMERO; VENTURA, 2013). Conforme ilustrado na Figura 7.

Figura 7 – Mineração de Dados Educacionais



Fonte: Adaptado de (ROMERO; VENTURA, 2013)

Nos últimos anos a MDE se tornou uma importante área de pesquisa independente que utiliza técnicas de classificação, agrupamento, modelagem bayesiana, descoberta de relacionamento por meio de modelos de aprendizagem. Deste modo, sendo um importante subsídio para tomada de decisão na prática educacional. Alguns dos problemas a serem resolvidos estão relacionados abaixo, de acordo com principal ator envolvido (ROMERO; VENTURA, 2013):

- Alunos: Fornecer feedback, atendendo suas necessidades para melhorar o desempenho na aprendizagem.

- Educadores: Compreender o processo de aprendizagem dos alunos para reflexão dos métodos adotados no ensino, melhoria do docente, entender como os aspectos sociais influenciam no aprendizado.
- Pesquisadores: Utilização de técnicas para identificar fatores que podem ajudar a resolver algum problema educacional específico, avaliar a eficácia da aprendizagem ao utilizar diferentes métodos.
- Administradores: Organizar os recursos (humanos e materiais) e sua oferta.

2.7 Randon Florest - RF

O algoritmo RF, proposto por Breiman (2001), é utilizado para solucionar problemas de classificação mediante uso de árvore aleatória, que por sua vez, é composto por diversos classificadores estruturados como árvores identificando e selecionando as variáveis mais importantes. Existem diversos algoritmos de RF, sendo que o mais famoso é o algoritmo C4.5, que utiliza recursividade para a seleção dos dados de treinamento em uma base de dados e em seguida realiza testes para identificar o potencial de cada variável. Este processo é o mais importante, pois, o algoritmo precisa identificar qual variável será colocada à frente. Contudo, será necessário a utilização de meios que possam medir o ganho de informação de cada variável para identificar qual característica terá maior influência (NOVAKOVIC, 2009). Na literatura existem diversos trabalhos que utilizam RF para seleção de variáveis devido a baixa complexidade de implementação. O algoritmo 1, gera um *rank* com as variáveis mais importantes identificadas (TORGO, 2016).

Algoritmo 1 Algoritmo Random Florest para Classificação

```

1: for b = 1 to B: do
2:   Faça uma amostra bootstrap com tamanho  $N$  dos dados de treinamento
3:   Crie uma random-forest  $T_b$  com os dados selecionados, repetindo
4:   recursivamente para cada nó terminal da árvore até que o nó mínimo  $n_{min}$ 
5:   seja encontrado.
6:   I. Selecione  $m$  variáveis aleatoriamente a partir das variáveis  $p$ .
7:   II. Escolha a melhor variável/ponto-dividido entre  $m$ .
8:   III. Divida o nó em dois nós filhos.
9:   Saída do conjunto de árvores  $\{T_b\}_1^B$ .
10:  Para fazer a predição de um novo ponto  $x$ .
11:  Para classificação: Que  $C_b(x)$  seja a classe predita da  $b$ th
12:  random-forest. Então  $\hat{C}_{rf}^B(x)$  votos majoritários  $\{\hat{C}_b(x)\}_1^B$ .
13: end for

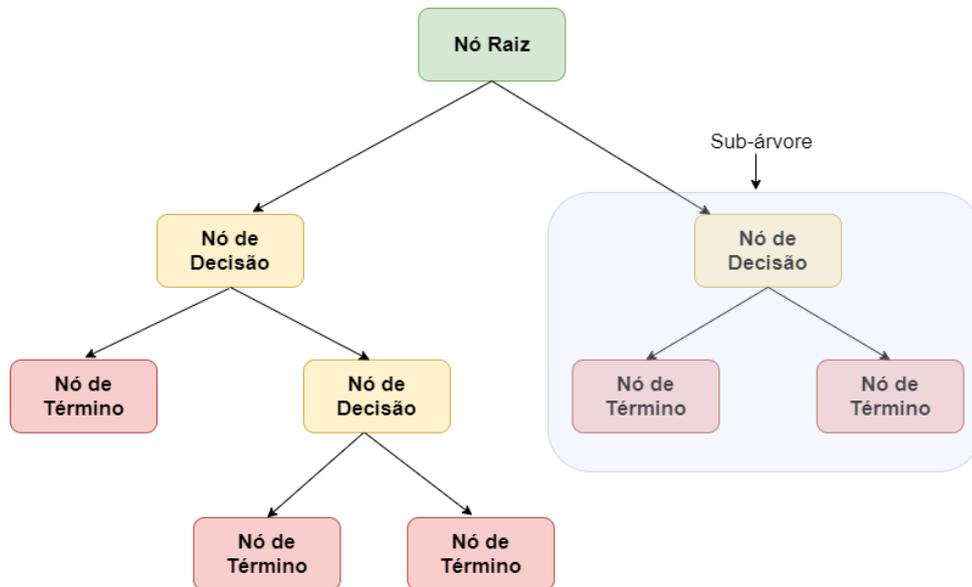
```

Fonte: (TORGO, 2016)

O algoritmo é uma melhoria da técnica *Bagging*. Onde, a partir da repetição do algoritmo, é produzido árvores a partir de amostras dos indivíduos que foram deixados

de fora *out-of-bag* do treinamento, e estes serão utilizados para a verificação do erro de predição e importância dos atributos. A execução repetitiva do algoritmo produz uma floresta de árvores, que é a combinação de muitas árvores binárias de decisão. (REIF et al., 2006). Conforme ilustrado na Figura 8.

Figura 8 – Random Florest

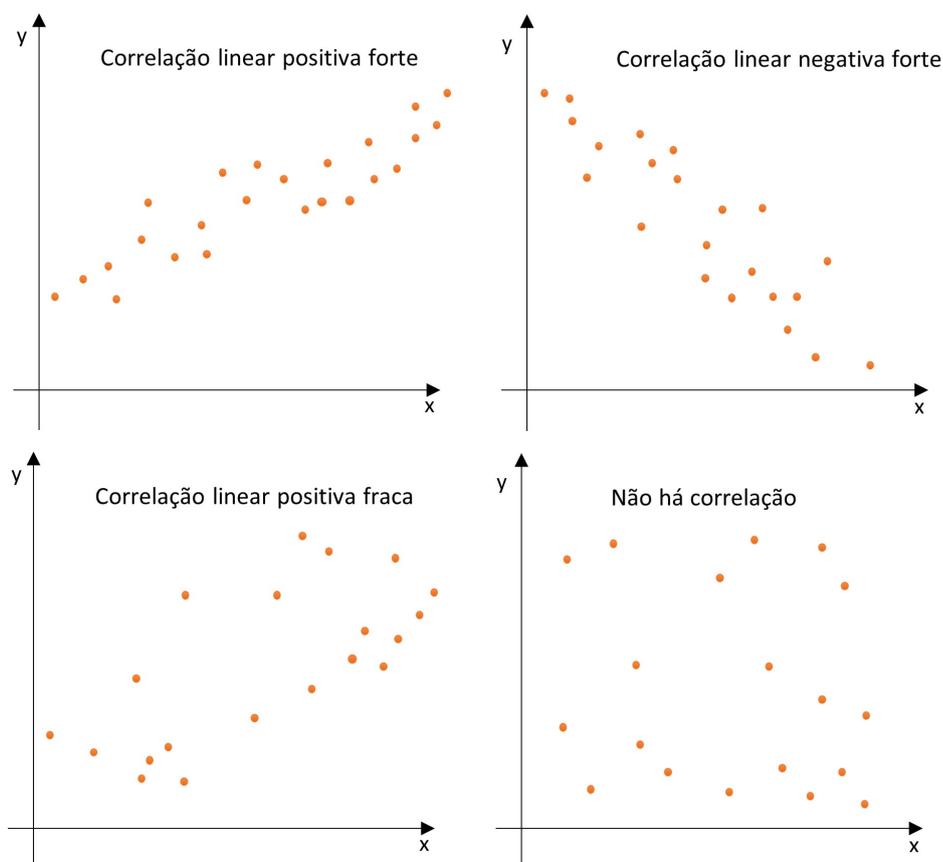


Fonte: Autor 2021

2.8 Coeficiente de Correlação Linear

A relação entre duas variáveis pode-se chamar correlação, que podem ser representados por pares ordenados (x,y) , onde x é a variável independente (ou explanatória) e y é a variável dependente (ou resposta) (FARBER; LARSON, 2010). Um gráfico de dispersão pode ser usado para determinar se existe ou não uma correlação linear (linha reta) entre duas variáveis. A Figura 9, ilustra os tipos de correções que podem ser identificados (FARBER; LARSON, 2010).

Figura 9 – Correlação linear



Fonte: Adaptada de (RODRIGUES, 2012)

O objetivo da correlação é avaliar o grau de associação entre duas variáveis que muitas vezes pode ser subjetiva. Para resolver este problema, uma maneira mais precisa é realizar o cálculo do coeficiente de correlação linear por meio da equação 3, onde r representa o coeficiente de correlação amostral (FARBER; LARSON, 2010). A amplitude do coeficiente é -1 para 1. Se (x,y) tem uma correlação forte, o valor de r estará próximo de 1. Se (x,y) tem uma correlação negativa forte, o valor de r estará próximo de -1. Se caso não houver correlação linear ou for fraca, com valor de r próximo a zero, não significa na haja correlação entre (x,y) . Neste caso, não há somente uma relação linear. Conforme ilustrado na Figura 9 (FARBER; LARSON, 2010).

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}} \quad (3)$$

2.9 Regressão Linear Simples

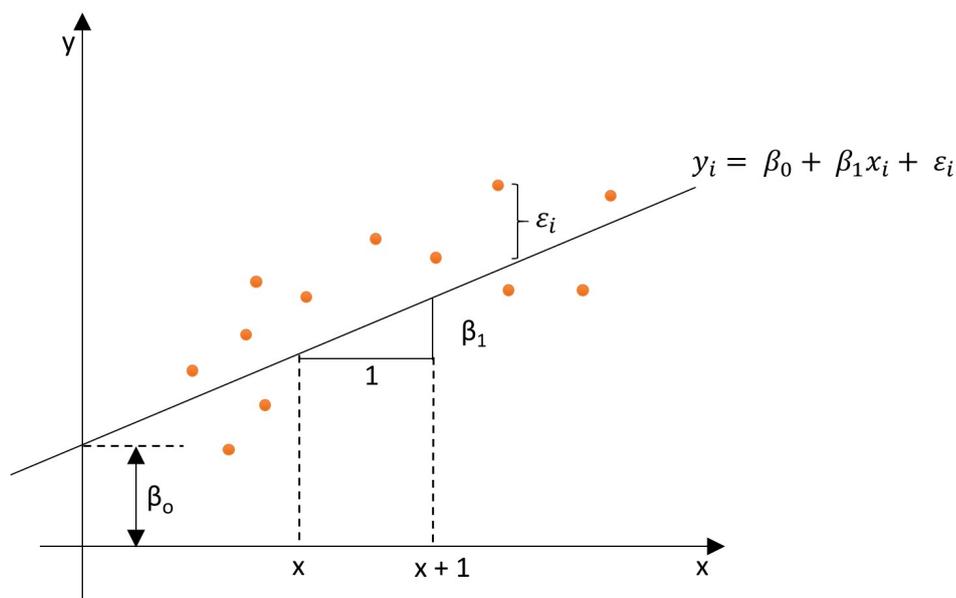
Conforme descrito anteriormente. O grau de relacionamento entre duas ou mais variáveis pode ser medido por intermédio da correlação. A regressão compreende a análise

de dados amostrais para saber como duas ou mais variáveis estão relacionadas em uma população. O modelo foi proposto pela primeira vez em 1885 em um estudo que demonstrou que a altura dos filhos tende a refletir a altura dos pais, mas com tendência a se manter na média da população (RODRIGUES, 2012). Por meio da equação 4, pode-se estimar valores futuros de uma variável com base em dados atuais em outra variável. Conforme descrito a seguir: (SELL, 2005).

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (4)$$

A saída y_i é a variável dependente ou variável resposta, no qual sofre influência da variável independente x_i . A variável β_0 e β_1 representam os parâmetros do modelo. A variável y_i retorna a saída do modelo, cujo resultado pode ser comparado com os dados reais, onde a diferença entre esses valores é chamada de erro de predição (RODRIGUES, 2012).

Figura 10 – Representação dos parâmetros β_0 e β_1



Fonte: Adaptada de (RODRIGUES, 2012)

A probabilidade de y_i é representada por β_1 que indica mudança na distribuição média em relação y_i , ou seja, a inclinação da reta conforme o aumento da variável x (RODRIGUES, 2012).

Para identificar se a reta está bem ajustada aos dados e identificar como o modelo de regressão está errado, pode-se utilizar as unidades da variável resposta. O erro padrão de estimativa S_e , é o erro médio utilizado após o ajuste da equação para se obter uma estimativa. Quanto menor for o erro, melhor será o ajuste da equação e, conseqüentemente

os pontos estão mais próximos da reta. Em vista disso, sendo uma importante estatística que pode ser utilizada (FARBER; LARSON, 2010).

$$S_e = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - 2}} \quad (5)$$

Onde n é o número de pares obtidos no conjunto de dados.

2.10 Considerações finais

O objetivo deste capítulo foi desvelar todos os componentes utilizados durante o processo de descoberta do conhecimento aplicado neste estudo. Ao descrever diferentes bases de dados, fica evidenciado a importância da disponibilização de dados abertos para que se possa acompanhar a evolução da educação. Além de abordar conceitos de mineração de dados educacionais, algoritmos de aprendizagem e métricas de avaliação para os algoritmos.

3 Trabalhos Correlatos

3.1 Considerações Iniciais

Nesta seção serão apresentados alguns trabalhos que de alguma forma irão compor esta pesquisa, com a implementação de técnicas de mineração de dados para descoberta de conhecimento no âmbito educacional, com diferentes abordagens, mas que estão em acordo com os objetivos deste trabalho. As principais fontes de pesquisa para elaborar esta sessão: biblioteca online, Scielo; Revista, Faesef e Revista novas tecnologias na educação; Conferências, IEEE e Sbie; Workshop, Workshop de desafios da computação aplicada à educação; Site, INEP, IPEA e MEC.

Palavras-chave: Infraestrutura, Renda, Mineração de dados, IDEB, SAEB, ENEM, Regressão Linear.

3.2 Correlatos

O trabalho de (COSTA, 2010), tem como objetivo de utilizar o índice do IDEB como indicador de qualidade na educação nos estados do sudeste desde sua concepção em 2007, para denotar a importância da qualidade da educação. Utilizando as notas técnicas disponibilizadas pelo INEP. O estudo, aborda as métricas para cálculo do IDEB e o cálculo esforço necessário para que uma determinada região atinja sua meta estabelecida no índice para o ano estabelecido. Segundo o estudo, o acesso à educação não é mais um problema, pois, grande parte dos estudantes tem acesso ao sistema educacional. O maior problema está relacionado as altas taxas de repetência e abandono escolar. Os resultados do estudo mostram a série histórica dos anos de 2005, 2007 e 2009, com as respectivas notas no IDEB.

O trabalho de (PINTO; JÚNIOR; COSTA, 2019), realizou uma pesquisa de cunho quantitativa e exploratória com a finalidade de identificar fatores que impactam o desempenho escolar (IDEB) das escolas municipais do município de Teotônio Vilela. Utilizando dados da Prova Brasil, o estudo utiliza técnicas de seleção de variáveis para identificar fatores que impactam no desempenho. O estudo utiliza o software Weka para identificação das variáveis que causam impacto na avaliação escolar. Os resultados apontam a descoberta de 18 atributos que influenciam fortemente o desempenho escolar nas disciplinas de língua portuguesa e matemática.

O trabalho de (COSTA et al., 2018) Investiga como a má qualidade de ensino e infraestrutura pode interferir na motivação das aulas de educação física e o IDEB. A metodologia empregada foi de cunho descritivo, por meio de questionários de necessidades psicológicas básicas para apurar opiniões e atitudes, utilizando software estatístico. Participaram da pesquisa 339 alunos (184 do sexo masculino e 155 do feminino), da rede estadual de ensino, 205 de escola com IDEB maior e 134 de escolas com IDEB menor. Os resultados apontam que houve uma diferença significativa entre os índices de autonomia, competência e relacionamento entre os sexos masculino e feminino.

O trabalho de (LIMA; OLIVEIRA, 2018), tem como objetivo analisar o índice do IDEB como instrumento de avaliação em larga escala, utilizando conceitos teóricos para explicar importância do exame e o processo que subsidia o cálculo para a obtenção das notas, realizando análises da evolução ao longo do tempo e ressaltando a importância para o alcance da meta. Segundo o autor, o exame é mais que um indicador que por meios estatísticos para avaliação, mas sim, um condutor de políticas públicas para melhoria da qualidade da educação. O trabalho ainda mostra a evolução histórica do índice do IDEB ressaltando a evolução das metas a serem atingidas e o importante papel do professor na superação dos desafios no meio socio-educativo para administrar as situações didáticas de aprendizagem.

A pesquisa de (SANTOS et al., 2016), apresenta uma análise descritiva sobre fatores que prejudicam o aprendizado dos estudantes em sala de aula, como: sócio-econômicos, educacionais e político-cultural. Trazendo questionamentos sobre as dificuldades da classe trabalhadora para ter acesso à educação de qualidade. O trabalho conclui que as desigualdades encontradas na sociedade e principalmente no âmbito escolar, prejudicam o aprendizado do aluno. Deste modo, propondo uma análise do sistema educacional a fim de diminuir as desigualdades sociais na escola.

O trabalho de (PINTO et al., 2019) utiliza técnicas de seleção de atributos para identificar quais fatores impactam no IDEB das escolas municipais de Maceió. O estudo utilizou dados do SAEB de 13 escolas para identificação desses fatores. O estudo utiliza alguns algoritmos que podem ser utilizado por meio do software Weka. Os resultados apontam quais atributos estão relacionados com as disciplinas de português e matemática.

A pesquisa de (JUNIOR et al., 2019) utiliza dados do INEP e modelo de previsão para prever a aprovação e reprovação escolar no ensino médio. As variáveis utilizadas pelo modelo foram selecionadas por meio do algoritmo RF. O estudo apresenta como resultado duas abordagens para previsão de aprovação e reprovação escolar no ensino médio, utilizando dados de 2016. Estes dados foram submetidos ao Algoritmo de Regressão Quântica (RQNP) e RQNP otimizada por Algoritmo Genéticos (AG).

O estudo de (ADEODATO et al., 2014) utiliza os dados do ENEM 2011 para avaliar a qualidade de ensino nas escolas particulares, utilizando a técnica *forwardstepwise*

para descobrir as variáveis mais influentes pelo critério de máxima verossimilhança, para a seguir, submeter ao modelo de regressão logística. Os resultados apontam o modelo de regressão logística foi capaz de gerar uma pontuação a partir das características das escolas, para avaliar a propensão de sucesso da escola.

No contexto da evasão escolar, o trabalho de (CALIXTO et al., 2017) realizou um estudo para identificar variáveis concernentes à evasão escolar, utilizando dados do Censo Escolar de 2014, 2015 e 2016 dos estados do Ceará e Sergipe, aplicando técnicas de mineração de dados para modelagem, a seguir, submeter a um modelo de Regressão Logística. Os resultados apontam que a idade, etapa de ensino, modalidade de ensino, existência de laboratórios e Localização da escola se destacam como variáveis influentes na evasão escolar.

O estudo de (DIGIAMPIETRI et al., 2016) analisa a evasão escolar no ensino superior utilizando como base o histórico escolar de 1896 alunos, de 2005 até início de 2015. Nos resultado, o estudo apresenta cinco disciplinas do primeiro ano do curso de Bacharelado em Sistemas de Informação. As análises se concentram nas notas alcançadas por esses alunos e, por fim, a submissão desses dados ao classificador *Rotation Forest* com 90% de acurácia.

Ainda se tratando da evasão escolar, o trabalho de (RIGO et al., 2012) faz uma contextualização do problema e justifica a importância de se realizar um levantamento das questões relacionadas por meio do uso da MDE que pode realizar análise de diversos fatores relacionados à evasão, compondo cenários coerentes com os conhecimentos adquiridos por meio da MDE. O estudo conclui que existem um conjunto complexo de causas e situações que podem levar à evasão escolar, através de estudos utilizando algoritmo de redes neurais, que conseguiu identificar com 90% de acerto, perfis ligados à evasão escolar.

Já a pesquisa de (AGAOLU, 2016) utiliza questionário para avaliar o desempenho dos docentes do ensino superior, utilizando técnicas de classificação, como por exemplo, Árvore de decisão, Máquinas vetoriais de suporte, Redes Neurais e análises discriminatórias, são utilizadas. Com objetivo de identificar as perguntas mais importantes no questionários. O trabalho conclui que os algoritmos se mostram eficazes em classificar o desempenho do docente em "satisfatório" e "não satisfatório", com base na percepção dos estudantes.

3.2.1 Análise dos trabalhos correlatos e contribuições desta dissertação

O levantamento realizado neste capítulo revela a preocupação de diversos autores em identificar fatores que possam auxiliar políticas públicas que visam a melhoria do sistema educacional no país, utilizando técnicas de mineração de dados. Nesse sentido, foram listados alguns dos importantes estudos que visam de alguma forma dar contribuições:

identificação de fatores relacionados à evasão escolar; predição de desempenho utilizando base de dados do ENEM; estudos sobre fatores que possam prejudicar o aprendizado em sala de aula; relação da má qualidade de ensino e infraestrutura na motivação dos alunos em educação física; análise de fatores que podem impactar no desempenho no índice do IDEB.

Esta pesquisa contribui com análise de forma mais ampla, analisando diversas bases de dados para que se tenha uma visão mais ampla do problema, o que não é visto nos demais trabalhos. Assim, diferentes correlações que podem estar relacionadas ao desempenho do IDEB nas escolas estaduais do estado do Pará, podem ser desveladas, possibilitando a adoção de estratégia mais robusta em relação as descritas abaixo.

Todos os trabalhos apresentados possuem contribuições na área desta dissertação e, em alguma medida, colaboraram para a construção desta dissertação. A Tabela 1 abaixo descreve a síntese dos principais trabalhos abordados neste capítulo.

Tabela 1 – Diferenças encontradas entre os artigos publicados na literatura e este trabalho

(COSTA, 2010)	O trabalho enfatiza o uso do índice do IDEB como indicador de qualidade da educação, abordando métodos para cálculo para obtenção do índice. Considerando fatores intrínsecos que podem afetar o rendimento dos alunos.
(PINTO; JÚNIOR; COSTA, 2019)	O estudo utiliza alguns algoritmos na ferramenta Weka para determinar as variáveis diretamente ligadas às melhorias no IDEB, retornando uma tabela com a precisão das variáveis para as disciplinas de português e matemática versus atributos socioeconômicos.
(COSTA et al., 2018)	O estudo se dá por meio de uma pesquisa de cunho descritivo para identificar como o desempenho no IDEB, interferem na motivação do estudante em educação física.
(LIMA; OLIVEIRA, 2018)	O estudo trata da metodologia empregada para o cálculo do IDEB e reforça seu uso como norteador de política públicas em larga escala.

(SANTOS et al., 2016)	O trabalho fala sobre a importância das questões socioeconômicas para a continuidade do bom desempenho escolar dos estudantes, no final, é elaborado uma síntese que discute sobre a importância de se considerar tais fatores e a sugestão de uma reforma educacional.
(PINTO et al., 2019)	O estudo utiliza algoritmos para determinar as variáveis diretamente ligadas às melhorias no IDEB, utilizando algoritmos disponibilizados por meio da ferramenta Weka utilizando as disciplinas que subsidiam a obtenção da nota do IDEB e algumas características vivenciadas pelos estudantes no ambiente familiar.
(JUNIOR et al., 2019)	O estudo se propõe a identificar variáveis que estão relacionadas aos índices de reprovação no ensino médio, por meio de regressão quantílica e otimização por algoritmos genéticos e regressão quântica não parametrizada. O estudo se propõe a mostrar a eficiência da utilização desses algoritmos.
(ADEODATO et al., 2014)	O trabalho utiliza técnicas de MDE para realizar predições de desempenho das escolas por meio do exame do ENEM, mediante uso de Regressão Logística. O experimento utiliza escolas particulares.
(CALIXTO et al., 2017)	O estudo utiliza o método CRISP-DM para descoberta de conhecimento e, mediante uso de algoritmo de Regressão Logística identifica variáveis que estão relacionadas a evasão escolar. Considerando variáveis relacionadas ao contexto escolar no qual o aluno está inserido.

(DIGIAMPIETRI et al., 2016)	O estudo tem como proposta, analisar aspectos associados à evasão escolar no ensino superior, utilizando o histórico escolar dos alunos, tendo como principal objetivo, analisar as notas obtidas pelos alunos nas disciplinas cursadas ao longo dos semestres. O estudo não considera aspectos como: características dos docentes, rendimento familiar, estrutura e outras variáveis que podem estar atreladas ao abandono.
(RIGO et al., 2012)	O trabalho faz uma análise à cerca da evasão escolar, utilizando algoritmo de Redes Neurais. O algoritmo consegue identificar com eficácia os perfis ligados à evasão. As análises se limitam ao ensino superior.
(AGAUGLU, 2016)	O estudo utiliza técnicas de mineração de dados para medir a performance dos professores no ensino superior. Se baseando em questionários respondidos pelos alunos.

Fonte: Autor

3.2.2 Considerações Finais

Os trabalhos apresentados nesta seção evidenciam a dificuldade vivenciada pelos pesquisadores no Brasil, com a demora na liberação dos dados causando transtornos e dificultando muito o trabalho dos pesquisadores. Com demora do governo em publicar dados atuais, as políticas públicas que poderiam ser tomadas a partir de diversos estudos, acabam se tornando pouco eficazes perante a necessidade atual. Já que a maioria dos trabalhos se baseiam de dados de dois ou três anos antes. A situação se agravou ainda mais com a não realização do Censo 2020, prejudicando muito a área da pesquisa.

4 Metodologia

4.1 Considerações iniciais

Neste capítulo é apresentada a metodologia empregada no presente estudo, seguindo todas as etapas de descoberta de conhecimento.

4.2 Metodologia empregada

O trabalho inicia com a obtenção das bases de dados, públicas disponibilizadas pelo Governo Federal (SAEB, ENEM, IDEB, IPEA e Censo Escolar), com o objetivo de identificar variáveis diretamente relacionadas ao desempenho no IDEB. A seguir, variáveis que não estão diretamente ligadas a fonte de dados no site do INEP, mas que podem de alguma maneira contribuir para melhores índices (PNE, Atlas Brasil). Para os experimentos, foi seguido as etapas do processo de Mineração de Dados (MD), proposta por (REZENDE et al., 2003), com o objetivo de extrair conhecimentos implícitos mas que sejam importantes para tomada de decisão.

O fluxo de trabalho foi realizado obedecendo as três grandes etapas que compõe o processo MD: Pré-processamento; Extração de padrões e Pós-processamento. Conforme ilustrado na figura 11

Figura 11 – Diagrama com a metodologia empregada



(REZENDE et al., 2003)

4.3 Pré-processamento

Para o melhor entendimento das causas que levam a um determinado desempenho no IDEB, é necessário coletar dados de diversas fontes, que podem estar em diversos formatos. Esses dados foram manualmente coletados do portal do INEP, Plano Nacional de Educação (PNE), Instituto de Pesquisa Econômica Aplicada (IPEA) e Atlas Brasil. Os dados estão acessíveis para uso público e as análises são referentes ao ano de 2017, por conta da não divulgação do SAEB para o ano de 2019 a tempo da pesquisa.

Dentro do portal do INEP, foram selecionadas as bases de dados do IDEB, SAEB, ENEM e Censo Escolar. A seguir, foram analisados dados coletados do Sistema de Monitoramento dos Planos da Educação. Neste portal, é disponibilizado uma ferramenta para verificar se as ações previstas nos planos da educação estão sendo atendidas, em termos percentuais, dentro do prazo definidos pela legislação. O sistema contempla informações

referente a META 3B – Taxa de Atendimento (15 a 17 anos) dos alunos no ensino médio, determinando que 85% deles estejam matriculados no ensino médio até 2024. (ATRICON-IRB, 2017). Com esta meta, possível analisar o percentual dos alunos que frequentam o ensino médio com classificador de risco de descumprimento, que é fornecido pelo portal. Foram utilizados dados de IDHM e IVS, que podem ser obtidos por meio dos dados disponibilizados pelo IPEA. Os dados podem ser filtrados por estado e deste modo o seu download pode ser executado. Pode ser analisado também a quantidade de matriculados e o total de matrícula por município.

Foram coletados dados de 10038 alunos que participaram do exame do SAEB e de 7446 alunos que participaram do exame do ENEM, responderam a pergunta sobre rendimento familiar e possuem identificador de suas respectivas instituições de ensino, cada uma com a devida particularidade referente a infraestrutura e características dos docentes que nelas atuam. Os dados correspondem aos alunos e escolas de todos os municípios do estado do Pará que possuem nota divulgada no IDEB.

A Tabela 2 ilustra as variáveis utilizadas por categoria.

Tabela 2 – Estrutura básica dos atributos estudados e sua fonte

Categoria	Descrição	Fonte
Capital Social	Rendimentos, Convívio, formação e atitude dos pais ou responsáveis na educação do aluno.	SAEB/ENEM
Estrutura Escolar	Estruturas existentes nas escolas, salas e equipamentos disponíveis.	Censo Escolar
Trajetória escolar	Tempo de permanência na escola, reprovação e abandono.	SAEB
Atitudes em relação a estudos específicos	Atitudes do aluno e do professor em relação ao estudo de português e matemática, somados a um item sobre uso da biblioteca ou sala de leitura.	IDEB/SAEB
Formação dos docentes	Docentes com graduação nas áreas que lecionam ou não e com ou sem proficiência acadêmica	Censo Escolar
Taxa de Atendimento	População de 15 a 17 anos que frequenta o ensino médio	PNE/TCeduca
Índice de Desenvolvimento Humano Municipal	IDHM 2010 do estado do Pará	Atlas Brasil

Fonte: Autor

O presente estudo também utilizou algumas variáveis que são categorizadas pelo INEP e possuem dados em formato *string*. Essas variáveis, são descritas de acordo com a 3 e 4.

Tabela 3 – Categorias de adequação da formação dos docentes em relação à disciplina que leciona

Grupo	Descrição
Grupo 1	Docentes com formação superior de licenciatura na mesma disciplina que lecionam, ou bacharelado na mesma disciplina com curso de complementação pedagógica concluído.
Grupo 2	Docentes com formação superior de bacharelado na disciplina correspondente, mas sem licenciatura ou complementação pedagógica.
Grupo 3	Docentes com licenciatura em área diferente daquela que leciona, ou com bacharelado nas disciplinas da base curricular comum e complementação pedagógica concluída em área diferente daquela que leciona.
Grupo 4	Docentes com outra formação superior não considerada nas categorias anteriores.
Grupo 5	Docentes que não possuem curso superior completo.

Fonte: (INEP, 2014)

Com o objetivo de identificar outros fatores que podem está diretamente relacionados ao baixo desempenho dos alunos. Foram analisados por meio do algoritmo RF, o nível de escolaridade dos pais dos alunos, separados por sete níveis de escolaridade, conforme descrito na Tabela 4. Esses níveis serão confrontados diretamente com o desempenho obtidos na prova do SAEB.

Tabela 4 – Nível de escolaridade dos pais

Nível	Escolaridade
1	Nunca estudou.
2	Não completou a 4. ^a série/5. ^o ano.
3	Completou a 4. ^a série/5. ^o ano, mas não completou a 8. ^a série/9. ^o ano.
4	Completou a 8. ^a série/9. ^o ano, mas não completou o Ensino Médio.
5	Completou o Ensino Médio, mas não completou a Faculdade.
6	Completou a Faculdade.
7	Não sei.

Fonte: (INEP, 2014)

Com o objetivo de identificar os alunos que estudam em escolas com IDEB abaixo e acima da média estadual. Foi criado uma nova variável chamada *classe*. Essa nova variável classifica cada aluno nas duas possíveis condições.

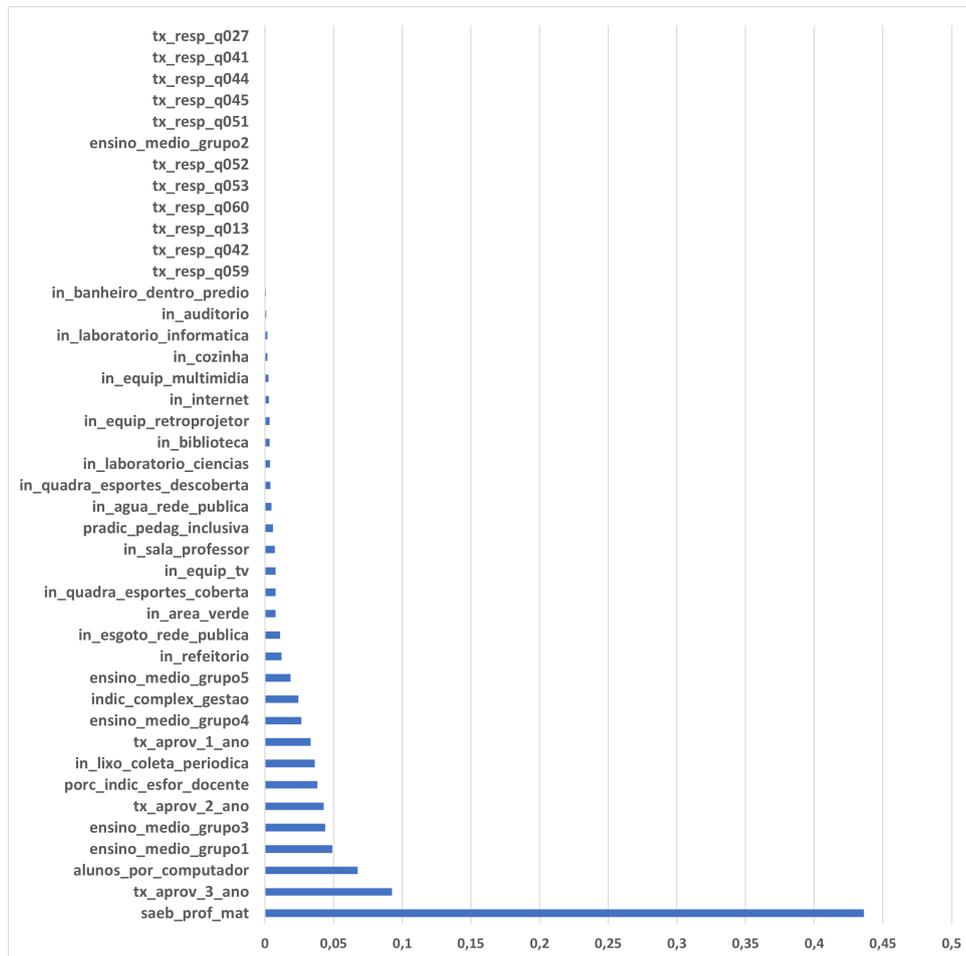
As bases que foram obtidas por meio do portal do INEP, IBGE e IPEA e possuem informações de todos os estados e uma grande quantidade de variáveis que podem ser analisadas. Para facilitar a manipulação destes dados, foi utilizado a plataforma Anaconda onde foi realizado operações para remoção de dados faltantes, verificação de inconsistências, redução da quantidade de campos em cada registro, eliminação de valores nulos e duplicados. Variáveis relacionadas ao questionário socioeconômico, contém valores no formato de *string* que passaram pelo processo de categorização para o formato *numeric*, tal medida se faz necessário, pois alguns algoritmos de classificação trabalham somente neste formato.

4.4 Extração de Padrões

4.4.1 Randon Florest - RF

Conforme descrito anteriormente no *capítulo 2.7*, o algoritmo RF é utilizado para solucionar tarefas de classificação. Porém, neste estudo, foi utilizado para quantificar a importância das variáveis, em tal caso, se torna um fator crucial para auxiliar na interpretação dos dados, possibilitando a compreensão dos fenômenos subjacentes. A figura 12, ilustra o resultado obtido por meio do Algoritmo 1. Onde é elaborado um *rank* com as variáveis selecionadas para, a seguir, submeter ao algoritmo de Regressão linear. As variáveis que melhor se adéquam ao modelo serão utilizadas neste estudo.

Figura 12 – Importância das Variáveis



Fonte: Autor 2021

Neste modelo, utilizou a validação cruzada dividindo a base em 10 partes com *10 fold* para testes. Após aplicação do método de validação, a precisão média obtida foi de 96% utilizando o parâmetro do modelo, com 7% para treino e 25% para teste. Após a aquisição dos valores referentes ao nível de importância, o próximo passo é dar início aos teste com algoritmo de regressão.

4.4.2 Regressão Linear

O modelo de Regressão linear (RL) é um algoritmo de aprendizado supervisionado que consiste em utilizar uma equação linear, o modelo de implementado está disponível no pacote *scikit-learn*. Neste experimento, as variáveis mais bem pontuadas foram adicionadas ao conjunto de dados para testes, onde este conjunto foi dividido em 75% para treino e 25% para teste. Ao final desta etapa, além dos resultados obtidos pelo algoritmo, serão coletados valores referente ao desempenho destas variáveis por meio de equações que serão descritas na próxima sessão.

4.5 Pós-processamento

Neste estudo foram utilizadas algumas das medidas de desempenho utilizadas em modelos de regressão e que estão relacionadas entre si. A seguir, serão descritas todas as métricas utilizadas no modelo. Onde y_i é o valor real da variável dependente; \hat{y}_i valor real da variável dependente e n é o tamanho total da amostra:

O *Mean Absolute Error* (MAE), é uma medida de desempenho comumente utilizada em valores numéricos em modelos de Regressão Linear. Sendo considerado um modelo robusto na tarefa de reproduzir a realidade. É uma medida similar ao MSE, porém, mais sensível ao grandes erros. O MSE é sempre positivo e a simulação perfeita é indicada quando o MSE é igual a zero. O modelo é implementado por meio da equação 7 (HALLAK; FILHO, 2011).

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (7)$$

A *Mean Squared Error*, é a métrica é utilizada na equação 8 serve para medir a diferença entre os valores reais e preditos. Quanto mais próximo de zero, mais preciso é o resultado (CHICCO; WARRENS; JURMAN, 2021).

$$MSE = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2 \quad (8)$$

A equação 9 traz resultados calculados pela métrica o *Mean Square Error* (MSE) levados ao quadrado. De forma similar a equação supracitada, a equação expressa o erro da variável de interesse, e quanto mais próximo de zero, melhor. Porém, por elevar ao quadrado os erros, resultados elevados podem serão potencializados pela equação (CHICCO; WARRENS; JURMAN, 2021).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2} \quad (9)$$

O *Mean Absolute Percentage Error* (MAPE), calculados pela equação 10, é outra métrica de desempenho para modelos de regressão que trata o erro médio com uma porcentagem, sendo possível a comparação com outro modelos, porém, se os valores foram próximos a zero, o resultado da métrica pode tender ao infinito (CHICCO; WARRENS; JURMAN, 2021).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - x_i}{x_i} \right| \quad (10)$$

Por fim, outra métrica utilizada e a *Symmetric Mean Absolute Percentage Error* (SMAPE) que veio como alternativa para tratar os inconvenientes da equação MAPE. Po-

rém, há pouco consenso sobre a equação definitiva. A equação utilizada possui resultados de 0% a 200%, e podem ser calculados pela equação 11.

$$SMAPPE = \frac{100\%}{n} \sum_{i=1}^m \frac{|x_i - y_i|}{(|x_i| + |y_i|) / 2} \quad (11)$$

4.6 Considerações finais

Nesta sessão foi utilizada a ferramenta Jupyter Notebook para o carregamento e manipulação dos dados. Tal ferramenta se mostrou bastante eficaz nas análises dos dados, dispensando uso de softwares estatísticos proprietários. Em seu uso, é possibilitado o acesso à biblioteca *scikit-learn*, muito utilizada na literatura por disponibilizar algoritmos de aprendizado de máquina, em código aberto na linguagem python.

5 Estudo de caso e Discussão dos resultados

5.0.1 Considerações iniciais

Os resultados apresentados nesta sessão utilizam dados das escolas públicas estaduais no estado do Pará e seus respectivos alunos. Começando com a apresentação de análise exploratória dos dados, a seguir, serão apresentados os resultados obtidos por meio dos algoritmo de regressão.

5.0.2 Estudo de Caso

Para efeito prático da aplicação da metodologia proposta nesta dissertação, este estudo de caso se inicia com a utilização de variáveis que não aparecem no *ranking* do algoritmo de seleção de variáveis, por pertencerem à bases distintas e sem conexão direta à outras bases. No entanto, essas variáveis podem estar relacionadas à diversas situações dentro do contexto escolar. Os resultados iniciais foram obtidos por meio da coleta e análise exploratória dos dados cuja a obtenção se dá por meio de bases de dados públicas descritas anteriormente. Neste momento, objetivo principal é apontar diferentes correlações entre a situação socioeconômica dos alunos e do ambiente no qual estão inseridos sem deixar de lado variáveis que estão dentro do contexto escolar. Deste modo, tal análise deve revelar um cenário mais amplo da educação no estado e eventuais dificuldades impostas a esses estudantes. Algumas variáveis utilizadas inicialmente na análise exploratória também passaram por testes no algoritmo de Regressão linear, bem como as variáveis já previamente selecionadas pelo algoritmo Random forest. Neste caso, os esforços se concentram em identificar possíveis correlações entre essas variáveis e outras que não estão diretamente ligadas ao contexto escolar, como por exemplo, subíndices do IDEB e IVS.

5.0.2.1 Metas para educação

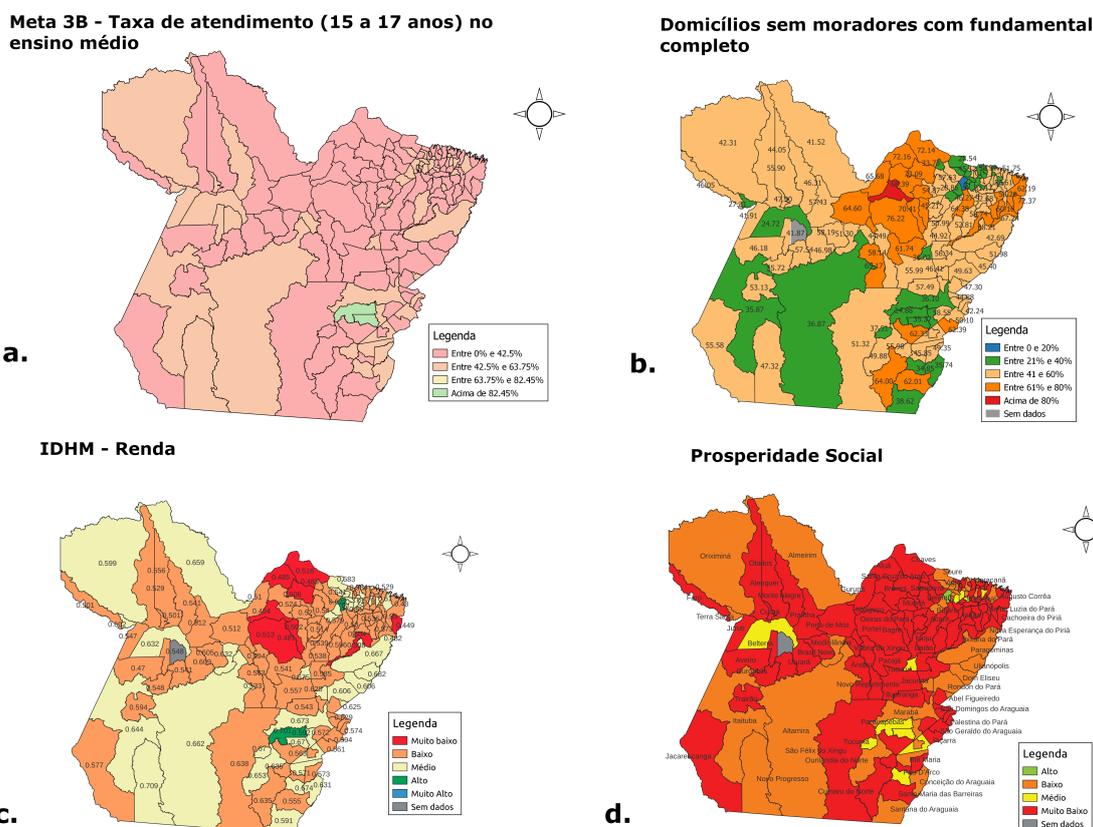
O (PNE, 2014) define diretrizes e metas para a educação no período de 2014 até 2024. Dentre outras, pode-se citar a “META 3 – Ensino Médio” com estratégias como a reformulação do ensino médio implementando abordagens interdisciplinar com conteúdo de diversas áreas: ciências, tecnologia, cultura e esporte; Garantindo infraestrutura necessária para a execução dessas atividades de forma atrativa para que a população jovem entre 15 e 17 anos, e adultos que estejam fora da escola, possam se sentir estimulados a participar de cursos na área da ciência e tecnologia, garantindo o acesso gratuito a esses cursos; Observando-se também todas as peculiaridades existentes que podem elevar os

índices de evasão escolar, adotando políticas públicas para garantir que filhos e filhas de famílias itinerantes possam ter a oportunidade de cursar alternativas ao ensino médio garantindo proteção contra qualquer forma de exclusão. Essas são algumas das diversas diretrizes definidas pelo PNE para cumprimento no período de vigência.

Dentro desse contexto. A Figura 13(a.), ilustra o percentual de alunos matriculados em escolas estaduais no ensino médio nos municípios do estado do Pará. A taxa de atendimento mede a porcentagem de vagas ocupadas em relação ao total disponível em cada município. De acordo com as categorias definidas pelo portal, apenas 2 municípios (Parauapebas e Canaã dos Carajás) ficaram acima de 82.45% e estão fora do risco de descumprimento da meta. A taxa de atendimento no ensino médio está muito abaixo do que seria o ideal para o cumprimento da meta para o estado, cujo indicador é de 85% até 2024. O cálculo para obtenção da taxa, é realizado da seguinte forma: População de 15 a 17 anos de idade que frequenta o ensino médio/População de 15 a 17 anos*100 (ATRICON-IRB, 2017). Os resultados evidenciam problemas relacionados à evasão ou até mesmo o abandono escolar.

Um dos fatores que estão relacionados ao baixo desempenho, pode ser observado na Figura 13(b). Onde a porcentagem de domicílios se moradores com fundamental completo nos municípios paraenses é ilustrado com menor e maior percentual: São Belém com 16,61% e Melgaço com 80,91%. Resultados ruins sobre a escolaridade dos pais se reflete nas oportunidades de emprego. A vulnerabilidade de renda e trabalho é medida por meio de fatores: insegurança de renda; baixa escolaridade; pessoas que dependem de pessoas idosas; presença de trabalho infantil (IPEA, 2015). Segundo Pereira e Cabral (2019), houve uma queda de 11,8% no desemprego, porém, o que chama a atenção é o aumento significativo do trabalho informal. Segundo dados da Pesquisa por Amostra de Domicílios Contínua - PNAD, o Brasil apresentou no segundo trimestre de 2019 o percentual de 41,3% de trabalhadores informais ocupados. A Figura 13(c.), ilustra o subíndice IDHM - Renda nos municípios paraenses, onde 91 ou 64% estão na faixa baixa. Já na Figura 13(d), é ilustrada a Prosperidade Social nos municípios paraenses. Este índice é obtido por meio do cruzamento de faixas do IDHM e IVS, e nota-se na região do Marajó e nordeste do estado o número maior de municípios com índices muito baixos, o que corresponde a 66,43%. Apenas dois estão na faixa alta, Belém e Ananindeua. Nenhum município paraense possui índice de prosperidade social muito alto.

Figura 13 – Índices educacionais e socioeconômicos no estado do Pará.



Os resultados apontados até aqui, evidenciam que a baixa escolaridade e o baixo rendimento nos domicílios podem ser um importante indicador para outras situações vivenciadas pelos estudantes. De um ponto de vista mais amplo, a pesquisa realizada pelo (IPEA, 2022), aponta que 57% dos domicílios na região norte estão com insegurança alimentar. Dados revelados pela pesquisa, mostram a gravidade da situação em mais da metade dos domicílios na região norte. À vista disso, fica evidenciado que a implementação de políticas públicas: transferência de renda; escolaridade; emprego; saneamento básico; alimentação escolar. São de extrema importância para melhorar o bem-estar social dessas famílias, e no contexto da educação, dar melhores condições para que os alunos possam se dedicar aos estudos. Conforme descrito anteriormente, 66,43% dos municípios estão com baixa prosperidade social. Nesse contexto, a seguir, será revelado o impacto dos rendimentos no desempenho dos alunos.

5.0.2.2 Renda Familiar dos Estudantes

Aspectos socioeconômicos dos estudantes, é um importante limitador à melhoria do desempenho escolar. Medidas paliativas foram criadas pelo governo federal, estadual e municipal para que famílias carentes possam manter o aluno na escola e garantir a frequên-

cia na sala de aula. No entanto, a frequência por si só, não garante que esse estudante terá bom desempenho nos estudos. O aluno precisa estar em sintonia com os componentes básicos que norteiam uma boa aprendizagem, sem interferência das mazelas que são escancaradas pela situação econômica precária (SANTOS et al., 2016). A Figura 14 reúne dados de 7446 alunos que possuíam registro de suas respectivas escolas na base de dados com nota no IDEB.

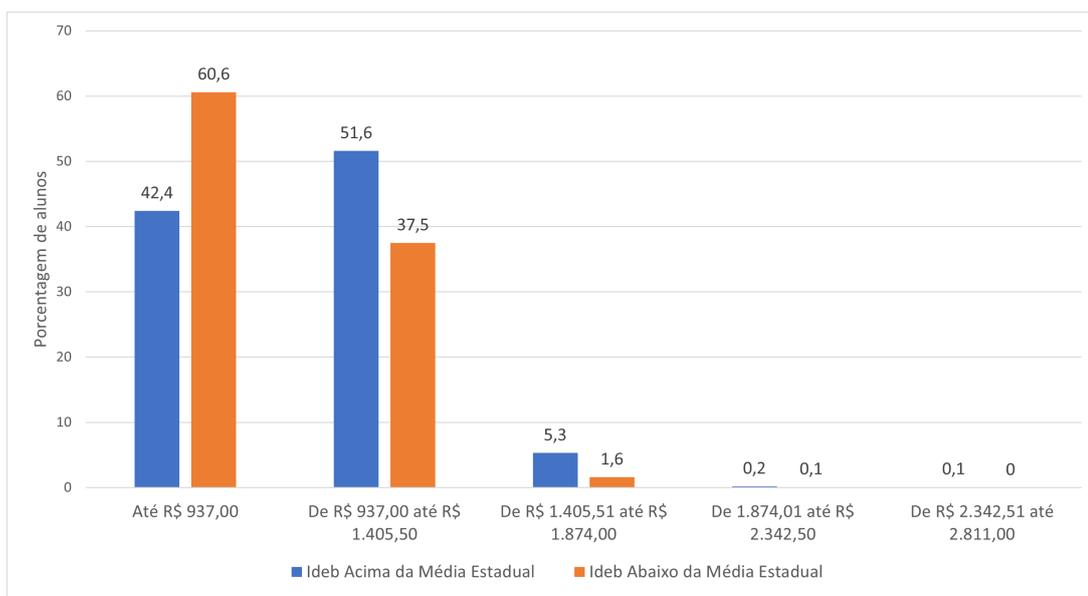


Figura 14 – Situação Socioeconômica dos Participantes do ENEM

Fonte: Autor

A Figura 14, ilustra a situação da maioria dos estudantes de ensino médio que responderam ao questionário socioeconômico do ENEM em 2017, em tal caso, foram considerados o quartil das escolas com melhor e pior desempenho no IDEB. Foram utilizadas as respostas do ENEM porque o exame do SAEB não possui questionário socioeconômico. Vale ressaltar que os alunos que fizeram a prova do SAEB provavelmente fizeram a prova do ENEM, pois ambos são aplicados para os alunos que estão no ensino médio. A relação direta entre renda e melhores desempenho no IDEB é apontada na Figura 14, onde os estudantes de famílias que possuem rendimento até um salário, em sua maioria estão situados em escolas com IDEB abaixo da média com 60,6%, porém, ao olhar para faixa seguinte, onde os rendimentos partem de um salário ou mais, a relação se inverte, a maioria dos estudantes estão matriculados em escolas com IDEB acima da média, com 51,6%, revelando que melhores rendimentos estão relacionados à melhor desempenho no IDEB. À vista disso, Parte-se do pressuposto de que o aluno necessita de forças e “capacidade” de concentração para “aprender a aprender” o mínimo necessário no seu dia-dia escolar. Deste modo, o fator socioeconômico é um dentre outros que podem levar o aluno ao fracasso escolar e por consequência até o abandono.

Bancos os gastos do aluno ou da família que não tem condições para que o estudante possa dar continuidade nos estudos, é algo essencial para que o ensino básico possa ser finalizado de maneira satisfatória. No entanto, ainda é um gasto que não é tratado pelo governo com o empenho necessário para que esses alunos possam de fato ter subsídios para alcançar melhor desempenho (SANTOS et al., 2016).

Na Figura 15, foram utilizados dados de 239252 estudantes para compor a média por municípios. As faixas de renda apontadas na Figura 14 foram categorizadas de 1 a 6 conforme descrito anteriormente para facilitar as análises. Nenhuma renda - valor 1; R\$ 937,00 - valor 2; De R\$ 937,00 até R\$ 1.405,50 - valor 3; De R\$ 1.405,51 até R\$ 1.874,00 - valor 4; De R\$ 1.874,01 até R\$ 2.342,50 - valor 5; De R\$ 2.342,51 até R\$ 2.811,00 - valor 6. Os resultados obtidos no eixo horizontal da figura, são referentes as médias das faixas de renda.

A Figura 15, denota sobre um outra perspectiva, a relação direta entre o rendimento médio familiar dos estudantes e o desempenho médio em matemática, uma das disciplinas utilizadas para avaliar a proficiência dos alunos na prova do SAEB.

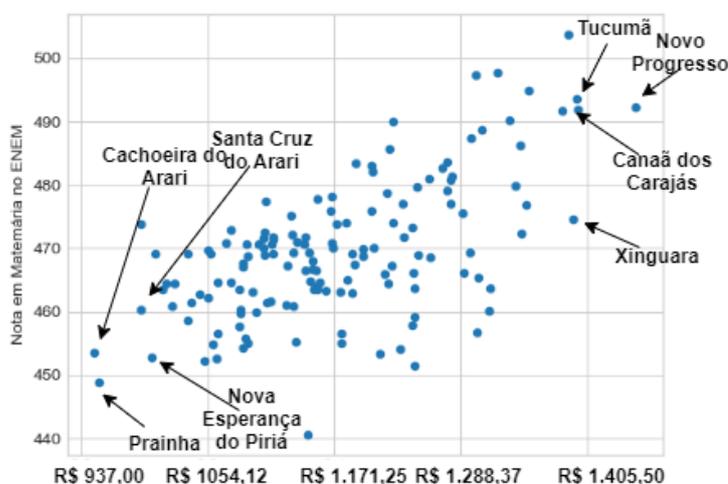


Figura 15 – Média das faixas de renda pela média das notas em matemática de cada município

Fonte: Autor

Com uma visão mais ampla a Figura 15, aponta que é importante a elaboração de políticas públicas: aumento do emprego, capacitação profissional, incentivo ao microempreendedorismo, auxílio e outros, por parte dos gestores para que as famílias possam dar suporte adequado para que seus filhos tenham a capacidade de se concentrar e se dedicar exclusivamente aos estudos. Pois, para aqueles que almejam ingressar em uma instituição pública superior, terão muitas dificuldades ao disputar uma vaga com os mais ricos, já que a renda é um fator que está diretamente ligado melhores condições de infraestrutura escolar e consequentemente, a maiores notas, potencializando as diferenças entre ricos e

pobres. Nesse contexto, a seguir, esta pesquisa trás uma relação entre melhores condições de infraestrutura escolar e seu desempenho no IDEB.

5.0.2.3 Infraestrutura Escolar

A Figura 16 aponta para as características que mais se destacaram nesta análise. São dados de 87 escolas públicas estaduais de todos os municípios do estado do Pará que obtiveram resultados do IDEB. As informações foram retiradas da base dados do Censo Escolar, onde valor é 0 para “não” e 1 para “sim”. As barras apresentam a porcentagem de escolas que possui determinada característica, que é descrita no eixo horizontal da figura, possibilitando a comparação entre as escolas com maior e menor desempenho no exame. Essas escolas foram divididas nas duas condições possíveis: 46 escolas com IDEB acima da média e 41 escolas com IDEB abaixo da média.

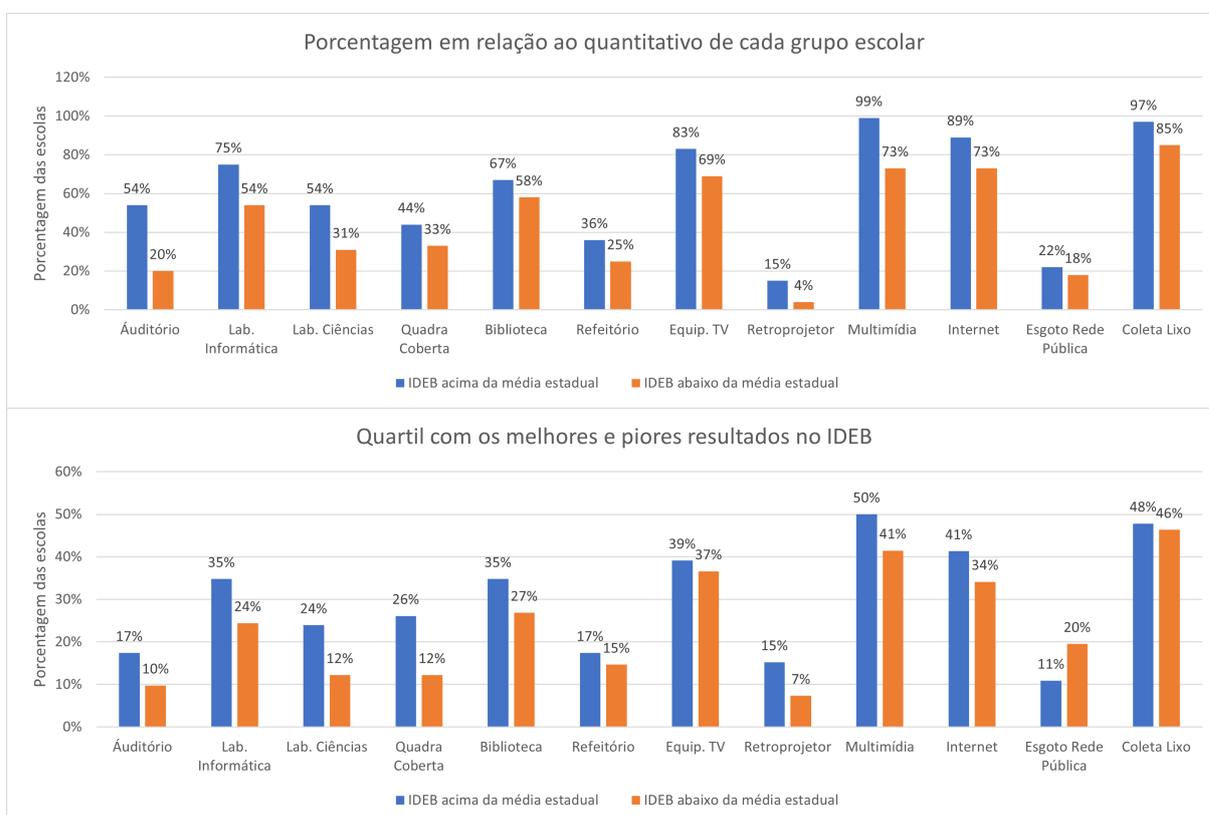


Figura 16 – Infraestrutura escolar

Fonte: Autor

A Figura 16(a.), aponta que a infraestrutura das escolas está diretamente ligada à melhor desempenho no IDEB. No grupo das escolas acima da média do IDEB, 54% possuem auditório, enquanto que no grupo de escolas abaixo da média no IDEB, apenas 20% possuem auditório. Nas escolas acima da média do IDEB, 75% possuem laboratório de informática, enquanto que no grupo de escolas abaixo da média no IDEB, 54% possuem laboratório. Nas escolas com média acima do IDEB, 54% possuem laboratório de ciências,

enquanto que nas escolas abaixo da média no IDEB, apenas 31% possuem laboratório de ciências. Nas escolas acima da média do IDEB, 44% possuem quadra coberta, enquanto que nas escolas abaixo da média do IDEB, 33% possuem quadra coberta. Nas escolas acima da média do IDEB, 67% possuem biblioteca, enquanto que as escolas abaixo da média do IDEB, 58% possuem biblioteca. Nas escolas acima da média do IDEB, 36% possuem refeitório, enquanto que nas escolas abaixo da média do IDEB, 25% possuem refeitório. Nas escolas acima da média do IDEB, 83% possuem equipamento de TV, enquanto que nas escolas abaixo da média do IDEB, 69% possuem equipamento de TV. Nas escolas acima da média do IDEB, 15% possuem retroprojektor, enquanto que nas escolas abaixo da média do IDEB, 4,5% possuem retroprojektor. Nas escolas acima da média do IDEB, 99% possuem multimídia, enquanto que nas escolas abaixo da média do IDEB, apenas 73% possuem multimídia. Nas escolas acima da média do IDEB, 89% possuem internet, enquanto que nas escolas abaixo da média, 73% possuem internet. Nas escolas acima da média do IDEB, 22% possuem rede pública de esgoto, enquanto que nas escolas abaixo da média no IDEB, 18% possuem rede pública de esgoto. Nas escolas acima da média do IDEB, 97% possuem coleta de lixo, enquanto que nas escolas abaixo da média no IDEB, 85% possuem coleta de lixo. Já na Figura 16(b.), foram considerados os quartil de escolas com melhor e pior desempenho no IDEB. No grupo de escolas acima da média do IDEB, 16% possuem auditório, enquanto que no grupo de escolas abaixo da média no IDEB, apenas 8% possuem auditório. Nas escolas acima da média do IDEB, 32% possuem laboratório de informática, enquanto que no grupo de escolas abaixo da média no IDEB, 22% possuem laboratório de informática. Nas escolas acima da média no IDEB, 22% possuem laboratório de ciências, enquanto que nas escolas abaixo da média no IDEB, 10% possuem laboratório de ciências. Nas escolas acima da média no IDEB, 24% possuem quadra coberta, enquanto que nas escolas abaixo da média 10% possuem quadra coberta. Nas escolas acima da média no IDEB, 32% possuem biblioteca, enquanto que nas escolas abaixo da média no IDEB, 24% possuem biblioteca. Nas escolas acima da média no IDEB, 16% possuem refeitório, enquanto que as escolas abaixo da média no IDEB, 12% possuem refeitório. Nas escolas acima da média no IDEB, 38% possuem equipamento de TV, enquanto que as escolas abaixo da média no IDEB, 32% possuem refeitório. Nas escolas acima da média no IDEB, 14% possuem retroprojektor, enquanto que nas escolas abaixo da média no IDEB, 6% possuem retroprojektor. Nas escolas acima da média no IDEB, 46% possuem multimídia, enquanto que nas escolas abaixo da média, 36% possuem multimídia. Nas escolas acima da média no IDEB, 38% possuem internet, enquanto que as escolas com IDEB abaixo da média, 28% possuem internet. Nas escolas acima da média no IDEB, 10% possui rede pública de esgoto, enquanto que nas escolas abaixo da média no IDEB, 16% não possuem rede pública de esgoto. Nas escolas acima da média no IDEB, 44% possuem coleta de lixo, enquanto que nas escolas abaixo da média no IDEB, 40% possuem coleta de lixo. Desta forma, revela a importância de investimentos por parte do

estado. A falta desses recursos leva parte desses alunos a uma outra situação, que será tratada a seguir.

5.0.2.4 Auxílio dos Pais na preparação dos filhos

Além dos fatores supracitados, outros certamente contribuem para o sucesso dos estudantes. Segundo (CHECHIA; ANDRADE, 2005), os pais relatam diversos problemas que são vivenciados pelos estudantes, que vão desde a adaptação escolar até problemas conjugais vivenciados pelos pais, outros problemas estão relacionados ao comportamento dos filhos. O estudo aponta também, problemas com adaptação escolar e consequentemente a dificuldade dos filhos em aprender.

Um fator desvelado nos dados analisados nesta pesquisa, aponta que o sucesso dos alunos na escola está de alguma forma atrelado à relação que os pais têm com o ensino. Nesta etapa, são utilizados dados de 8287 alunos de todo o estado que responderam no questionário a pergunta sobre a trajetória dos pais, a seguir, foram calculadas as médias das notas e do nível escolar em que os pais se encaixam por município, totalizando 52 municípios, para por fim, submeter ao algoritmo de regressão. Os resultados ilustrado na Figura 17, aponta que o sucesso dos estudantes está diretamente relacionado com a escolaridade dos pais, deste modo, quanto maior a escolaridade dos pais, melhor será o rendimento dos seus filhos na escola. Tanto os pais de filhos com sucesso quanto de filhos com insucesso, são bastante ativos na vida escolar dos filhos, porém, esse auxílio ocorre de maneira pouco eficiente por falta de compreensão do conteúdo pedagógico ou orientações específicas. Ou seja, os pais se sentem perdidos diante das dificuldades dos filhos, pois se baseiam na experiência escolar vivida por eles que na maioria das vezes é a mesma dos alunos. Deste modo, outra forma de fazer com que os filhos se interessem pelas atividades escolares é o acompanhamento do material escolar como forma de disciplinar o comportamento dos filhos (CHECHIA; ANDRADE, 2005).

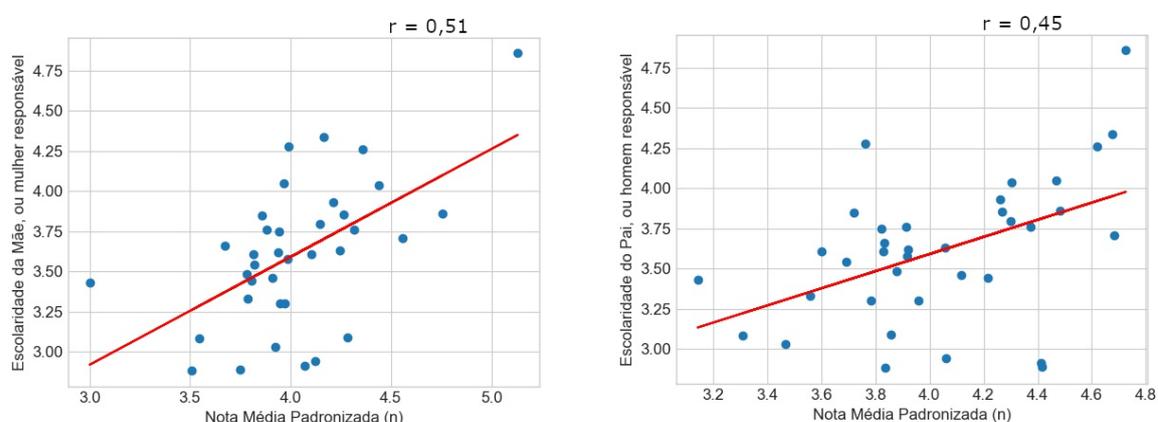


Figura 17 – Nível escolar dos pais

Fonte: Autor

Os resultados obtidos por meio desta pesquisa, e ilustrados na Figura 17, estão de acordo com a pesquisa de campo realizada pelo estudo de (CHECHIA; ANDRADE, 2005). Onde os resultados são obtidos com o preenchimento de questionários por partes dos pais.

5.0.2.5 IDEB, IDHM e IVS

Dados preocupantes relacionados ao IDHM são revelados; dos 143 municípios: 8 com IDHM muito baixo; 88 com IDHM baixo; 44 com IDHM médio; 3 com IDHM alto. Conforme ilustrado na Figura 18, a faixa com maior quantidade de municípios, corresponde a 62% do total.

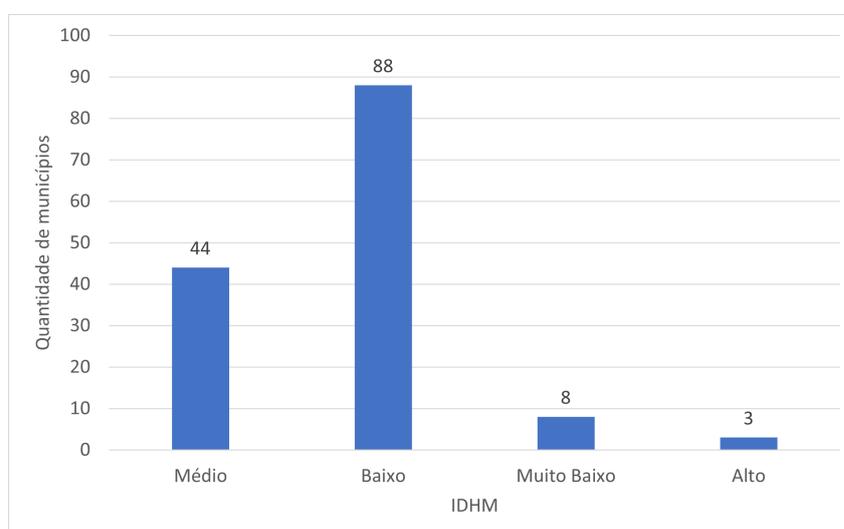
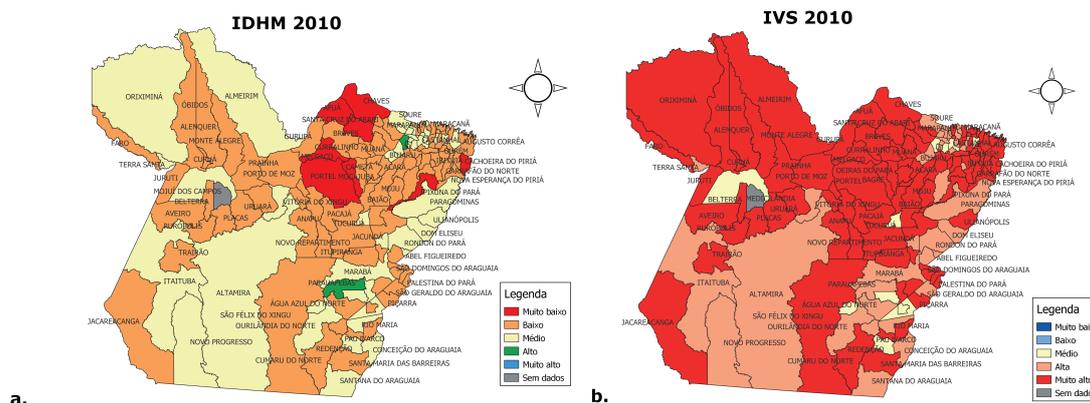


Figura 18 – IDHM 2010

Fonte: Autor

Os três municípios com maior IDHM são: Belém e Ananindeua, na região metropolitana de Belém; Paragominas, município que é situado na região de carajás. De acordo com o (IBGE, 2010a), Paragominas no ano de 2010, situavam grandes empreendimentos na área de mineração, deste modo, se beneficiando com a alta de preços no mercado internacional. Segundo (IBGE, 2010b), o índice mais alto no IDHM-Educação, pertence ao município de Belém com índice de 0,673, onde 39,48% das pessoas entre 18-20 anos possuem ensino médio completo; o índice mais baixo no IDHM-Educação, pertence ao município de Melgaço com índice de 0,207, onde 5,63% das pessoas entre 18-20 anos possuem ensino médio, situando-se entre os oito municípios com IDHM muito baixo. A Figura 19(a), ilustra as faixas do IDHM no estado do Pará

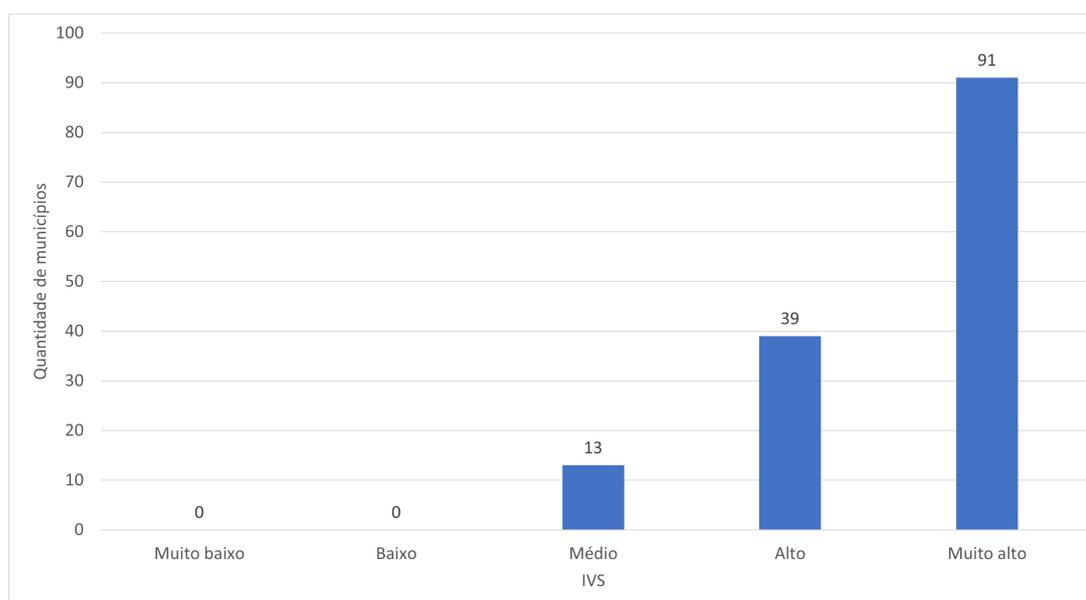
Figura 19 – Índices globais do IDHM e IVS



Fonte: Adaptado de (IBGE, 2010a)

A Figura 19(b.) aponta que dos 143 municípios: 13 com IVS médio; 39 com IVS alto; 91 com IVS muito alto. Em (IPEA, 2015), o Pará, assim como os demais estados da região norte, não possuem índice IVS muito baixo. Com IVS muito alto, o Pará fica atrás apenas do estado do Amazonas, com 63% ou 91 municípios nessa faixa, conforme ilustrado na Figura 20. Belém e outros municípios na região norte fazem parte do percentual de 24,1% dos que estão da faixa média do IVS. O município de Aveiro-PA possui o pior índice da região norte com 0,769.

Figura 20 – IVS - 2010



Fonte: Autor

Esta pesquisa traz uma comparação parcial entre IDHM, IDEB e IVS utilizando dados de 87 escolas e 10371 estudantes do ensino médio, revelando a morfologia dos fatores

que podem dar subsídio ou não à evolução do índice. Analisando variáveis relacionadas ao desempenho no SAEB, experiência de uso no meio digital e preparo dos docentes no exercício das atividades escolares. Conforme observado nas Figura 21,22,23 e 24.

A Figura 21(a.) ilustra a correlação existente entre a nota média padronizada obtida por meio da prova do SAEB e o índice de IDHM das 87 escolas participantes do IDEB. Os pontos em amarelo e azul, pertencem à alunos que estudam em escolas com IDEB abaixo e acima da média, respectivamente.

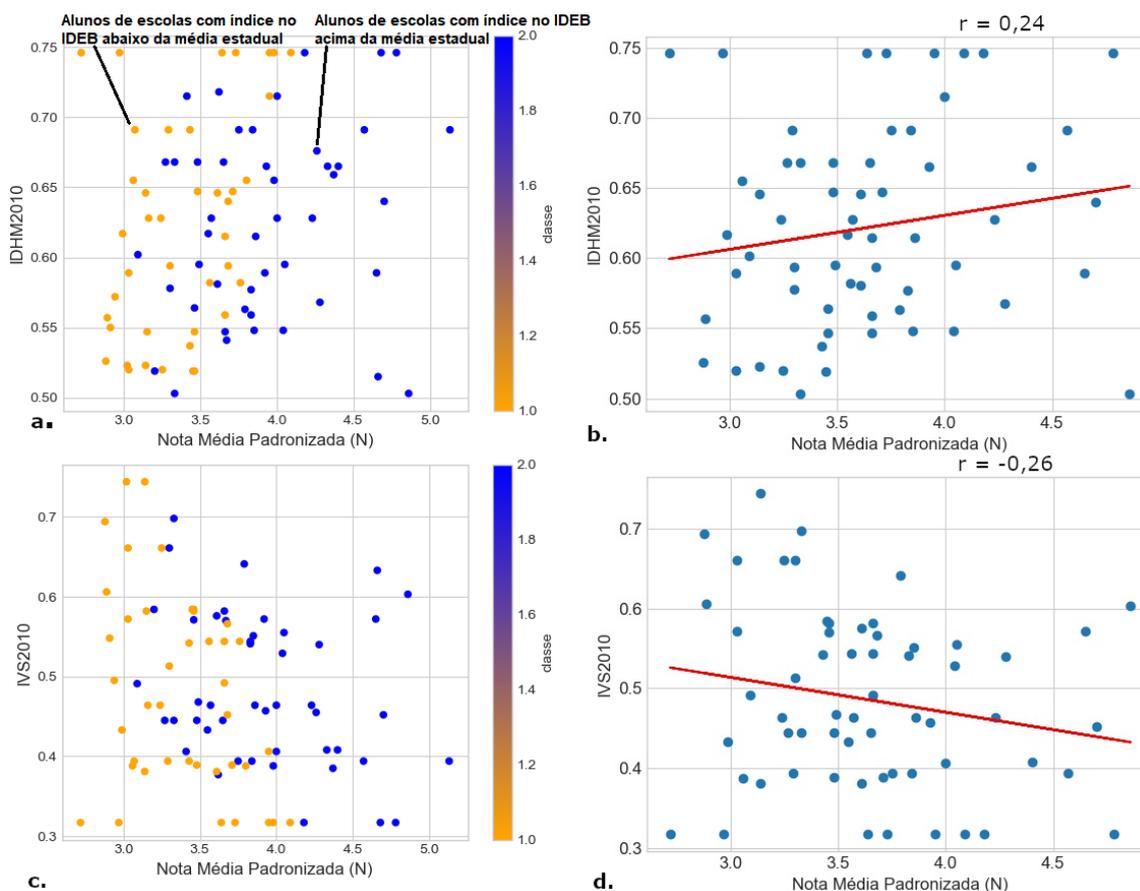


Figura 21 – Evolução da nota de proficiência x IDHM x IVS

Fonte: Autor

A Figura 21(a.) revela de modo claro que os alunos que tiveram aproximadamente 4,0, são em sua maioria alunos que estudam em escolas com IDEB abaixo da média. A partir 4,0 nota-se que a relação direta entre o melhor desempenho escolar e alto IDEB. No entanto, a relação entre essas notas e o IDHM não podem ser identificadas. Para isso, essas duas variáveis foram analisadas pelo algoritmo de regressão linear, Figura 21(b.). Em vista disso, fica evidenciado por meio do algoritmo que os melhores resultados estão diretamente ligados à municípios mais bem avaliados no IDHM. Do mesmo modo, a variável que compõe a média padronizada dos estudantes, foi ilustrado no gráfico de dispersão com o IVS no eixo vertical, Figura 21(c.), a seguir, a força da correlação entre

ambas foi medido por meio do algoritmo conforme ilustrado no Figura 21(d.), onde a tendência aponta para melhor desempenho em municípios com baixo IVS. Nota-se a relação inversamente proporcional das variáveis IDHM e IVS, com grau de correlação $r=0,24$ e $r=-0,26$, respectivamente, quando relacionadas a nota média padronizada dos alunos.

A Figura 22 ilustra a tendência entre números mais altos de abandono escolar com piores índices, no eixo horizontal com os índices IDHM e IVS, eixo vertical. Nesta análise, foram utilizadas respostas do questionário socioeconômico de 8180 alunos, calculou-se as médias para que seja possível a análise por meio do algoritmo de regressão, com o total de 48 registros. A quantidade de alunos que responderam ao questionário socioeconômico do SAEB, sugere dificuldade maiores para os alunos que moram em áreas com alto índice de vulnerabilidade social e baixo desenvolvimento humano. Em vista disso (FILHO; ARAÚJO, 2017), traz a visão de que o problema da evasão escolar no Brasil é crônico e passivelmente tolerado por escolas, que ao ofertam matrícula acima do esperado em determinada turma, já contando com o abandono de alguns. Problemas intrínsecos e extrínsecos ao ambiente escolar estão contribuindo fortemente para que esse problema não se resolva em um curto prazo. Fatores como: dependência química; falta de valores; falta de incentivo da família; necessidade de trabalhar. Afastam os alunos das escolas, e por fim, acabam engrossando a mão de obra barata. Além destes, outros fatores também podem estar associados, como o desinteresse em continuar os estudos ou a transferência para outra escola. Uma forma eficaz de se chegar às causas do abandono, é investigar a trajetória escolar desses alunos, junto com questões familiares, mas sem deixar de se levar em conta o ambiente escolar e a comunidade onde esse estudante está inserido. (FILHO; ARAÚJO, 2017).

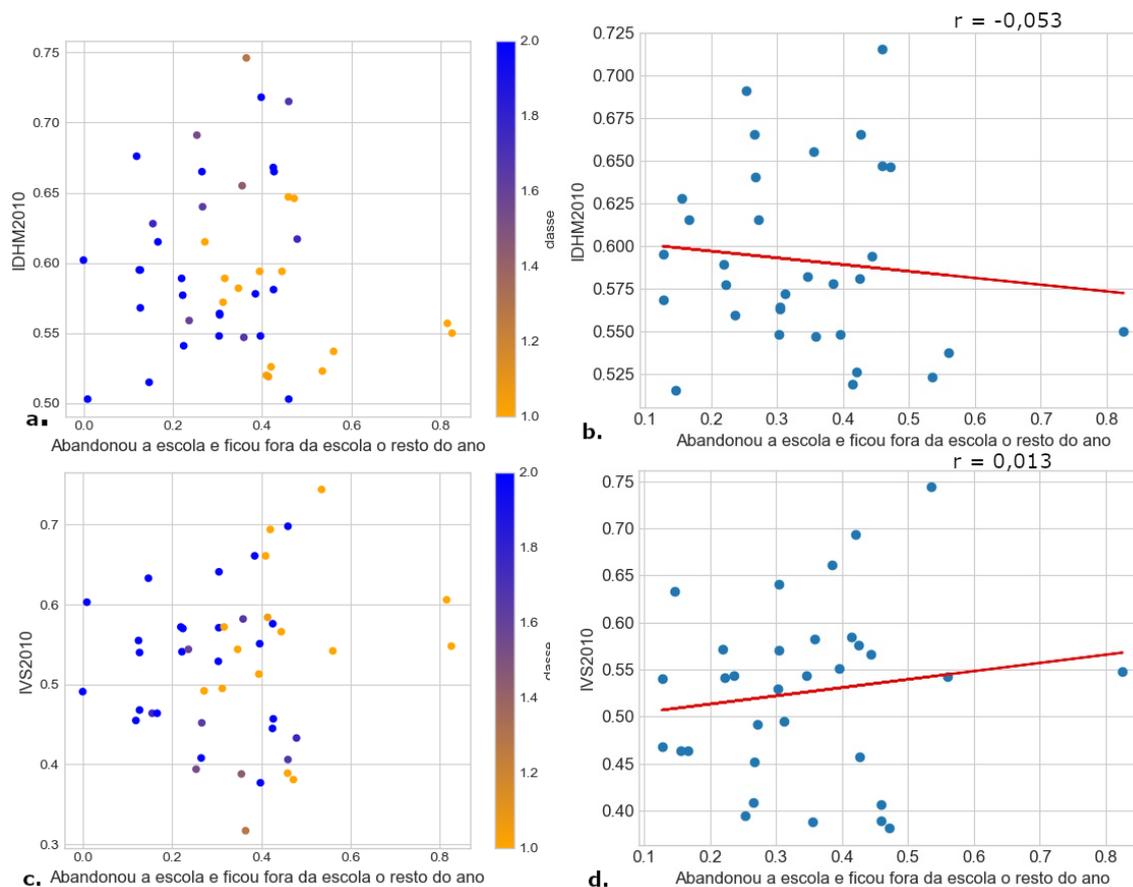


Figura 22 – Abandono escolar x IDHM x IVS

Fonte: Autor

A Figura 22(a.), revela dados difusos entre escolas com índice acima e abaixo da média, com tendência em menor quantidade de alunos, para aquelas que possui melhor desempenho no IDEB. A Figura 22(b.) aponta uma discreta correlação desse perfil de estudantes à melhor pontuação no IDHM. Da mesma forma, as Figura 22(c.) e Figura 22(d.), para os índices IVS. Onde maiores índices de vulnerabilidade estão relacionados com escolas com números maiores de evasão escolar. A força da correlação entre as variáveis descritas anteriormente com grau de $r=-0,053$ e $r=0,013$ para os valores de IDHM e IVS, respectivamente. A correlação muito baixa entre a variável abandono e os municípios paraenses, que de modo geral possuem baixos índices de prosperidade social, pode estar relacionada a eficácia de políticas públicas de transferência de renda para famílias em extrema pobreza, como Bolsa família. No contexto da educação, a família ao receber esse benefício, deve obedecer algumas condicionalidades, dentre elas: os responsáveis devem matricular crianças de 6 a 17 anos na escola e a frequência deve ser de, pelo menos, 75%; levar os menores de 7 anos para tomar as vacinas recomendadas; as gestantes devem fazer pré-natal e ir as consultas nas unidades de saúde (FACIROLI; SIRQUEIRA, 2017). Em 2017, no estado do Pará, 39% dos residentes de todas as faixas etárias eram beneficiárias do bolsa família (SENARC, 2017). Na região do marajó, onde se concentram os muni-

cípios com pior renda per capita: Melgaço com 3489 famílias beneficiadas; Portel com 7683 famílias beneficiadas. O valor médio recebido por essas famílias foi de R\$ 244,49 (NASCIMENTO et al., 2017).

Um fator muito importante para a eficiência do aprendizado do aluno é a adequação do professor em relação a disciplina ministrada em sala de aula. De acordo com o (INEP 2017), as disciplinas de português e matemática utilizadas para medir a proficiência dos alunos no exame do SAEB possui percentuais acima de 70% de professores no indicador (Grupo 1), professores pertencentes nessa categoria possuem formação mais adequada para a disciplina que lecionam. No entanto, a maioria desses profissionais estão concentrados em municípios com maior IDHM, conforme observado na Figura 23. A Figura 23(a.), aponta que 19% dos alunos no estado estão concentrados nessas escolas e com índice do IDEB acima da média. Apenas 5% estão nesse grupo, estão com índices abaixo da média. De acordo com os resultados evidenciados pela presente pesquisa, estudantes pertencentes à essas escolas estão com os melhores índices de proficiência. A Figura 23(b), evidencia a correlação entre a maior porcentagem de docentes pertencentes ao (Grupo 1) e IDHM alto com índice de correlação de $r=0,29$.

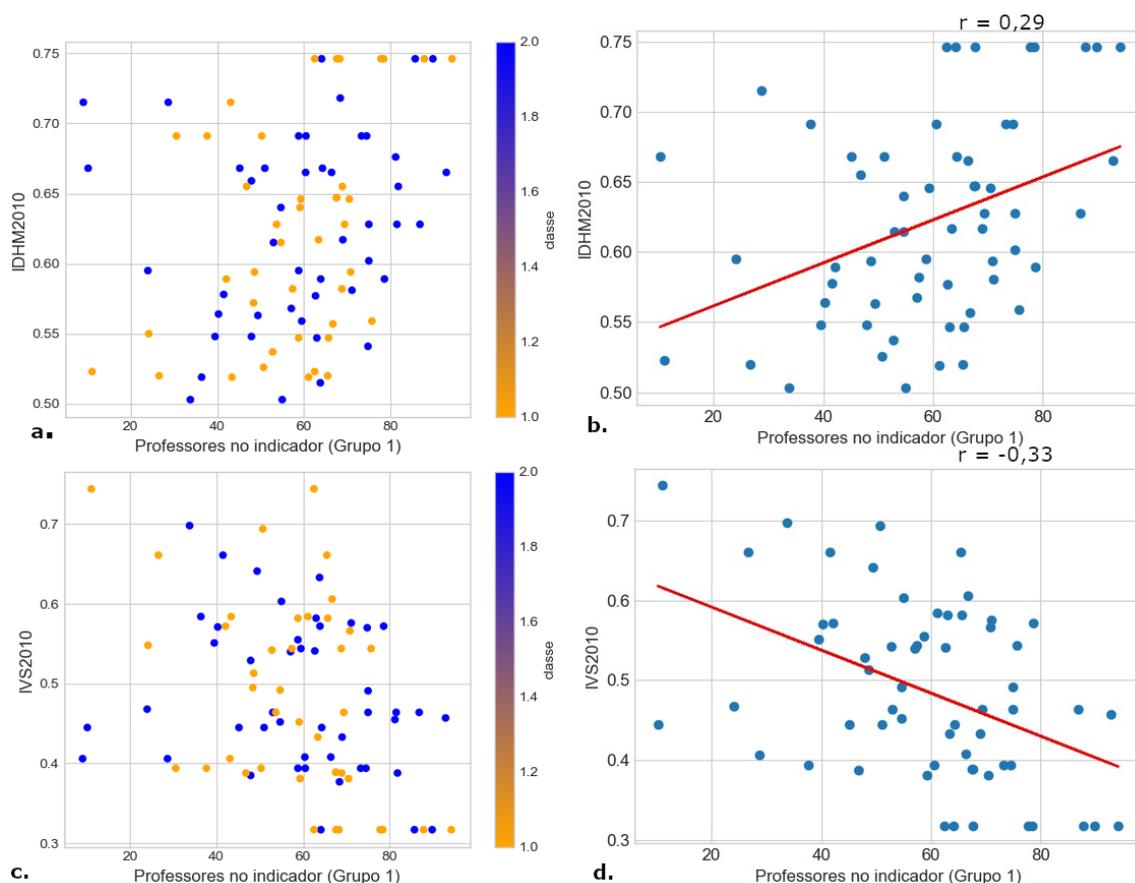


Figura 23 – Professores do Grupo 1 x IDHM x IVS

Fonte: Autor

Nas Figura 23(c.) e Figura 23(d.), é notada a relação inversa apresentada nos gráficos anteriores, conforme apontado pelo índice de correlação $r=-0,33$. Os resultados demonstram que regiões com queda da vulnerabilidade possuem educadores mais adequados à disciplina que lecionam. Isso corresponde a apenas 9% dos municípios paraenses que estão na faixa média de vulnerabilidade, conforme descrito anteriormente, nenhum município paraense possui municípios na faixa baixa e muito baixa.

Ao analisar o perfil dos docentes do (Grupo 2), que pertencem a professores com formação superior de bacharelado na disciplina correspondente, mas sem licenciatura ou complementação pedagógica. Foi identificado que os valores não obedecem os requisitos exigidos pelo algoritmo de regressão, contendo 99% dos dados com valor igual a zero, impossibilitando também a análise de dispersão dos dados.

Da mesma forma, foram analisados docentes do (Grupo 3), que pertence a professores que possuem uma graduação em licenciatura e complementação pedagógica diferente da matéria que leciona. A Figura 24, ilustra o percentual de docentes pertencentes a esse grupo no eixo horizontal, e no eixo vertical, novamente são apresentados os valores de IDHM.

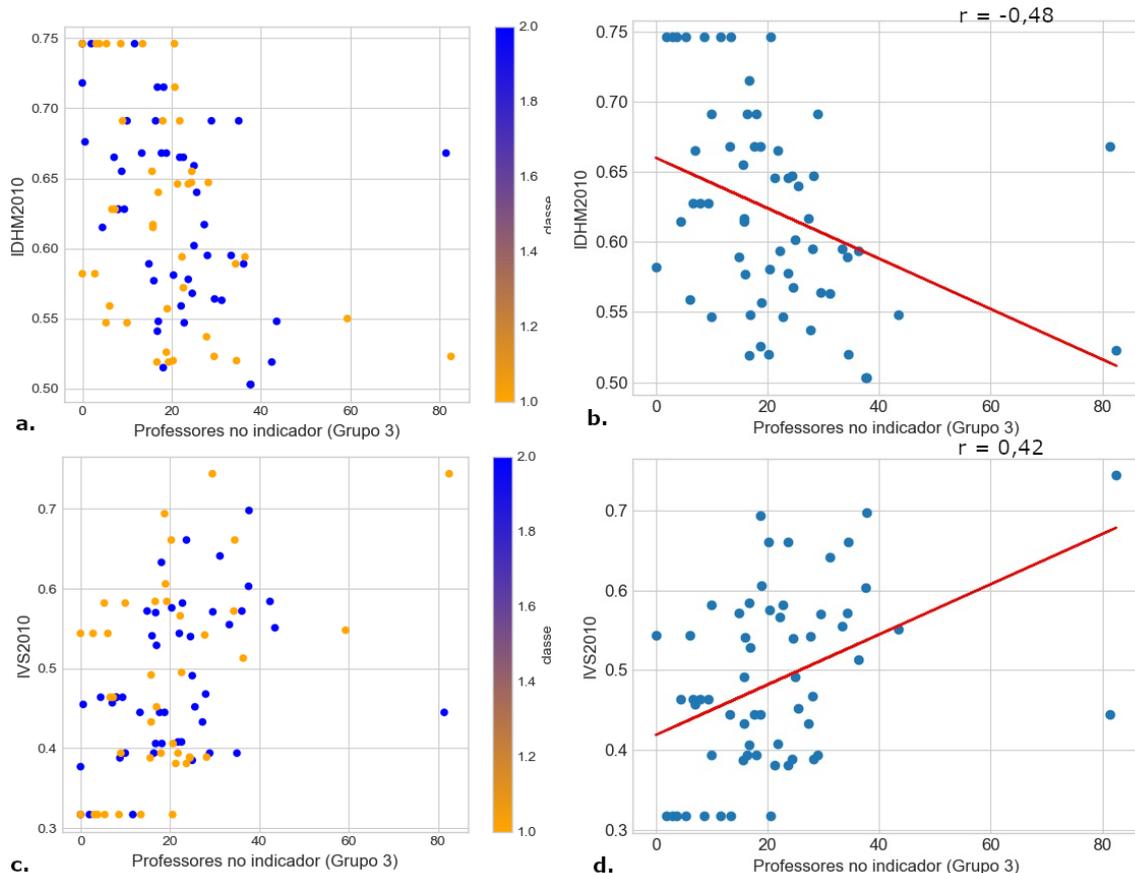


Figura 24 – Professores do Grupo 3 x IDHM x IVS

Fonte: Autor

A Figura 24(a.), ilustra a distribuição dos estudantes pertencentes a escolas com IDEB abaixo e acima da média. Nota-se que não há uma relação direta entre esses alunos, 88% estão concentrados em escolas com IDEB acima da média. Essas escolas possuem de 20 a 40% do seu corpo docente pertencente ao (Grupo 3). No entanto, há uma correlação entre o perfil desses docentes e o índice IDHM. A Figura 23(b.), aponta de fato, por meio do algoritmo de regressão linear, que a maior concentração desse profissional se dá em municípios com menor IDHM. Denotando que a concentração de profissionais com nível de formação menos adequada, se encontra em municípios menos estruturados. Nas Figura 24(c.) e Figura 24(d.), os resultados são semelhantes e mostram a maior concentração desse perfil em municípios com resultados ruins no IVS, evidenciado a carência de profissionais mais adequados para atender justamente aquelas microrregiões onde a população possui alto risco de exclusão social. Isso é preocupante pois 63% dos municípios paraenses estão com alto índice de vulnerabilidade social.

Ao analisar o perfil dos docentes do (Grupo 4), que pertencem a professores com outra formação superior não considerada nas categorias superiores, observou-se a correlação de $r=0,087$ ao cruzar com dados do IDHM, e $r=0,038$ ao cruzar dos dados com IVS, a descrição desta variável não explica de forma objetiva a característica dos dados contidos nas bases de dados. Em vista disso, optou-se por não utilizar nos resultados do presente estudo.

Já o (Grupo 5), pertencem aos docentes que não possuem curso superior completo, porém, ministram aulas nas escolas do estado. Os resultados são apresentados na Figura 25.

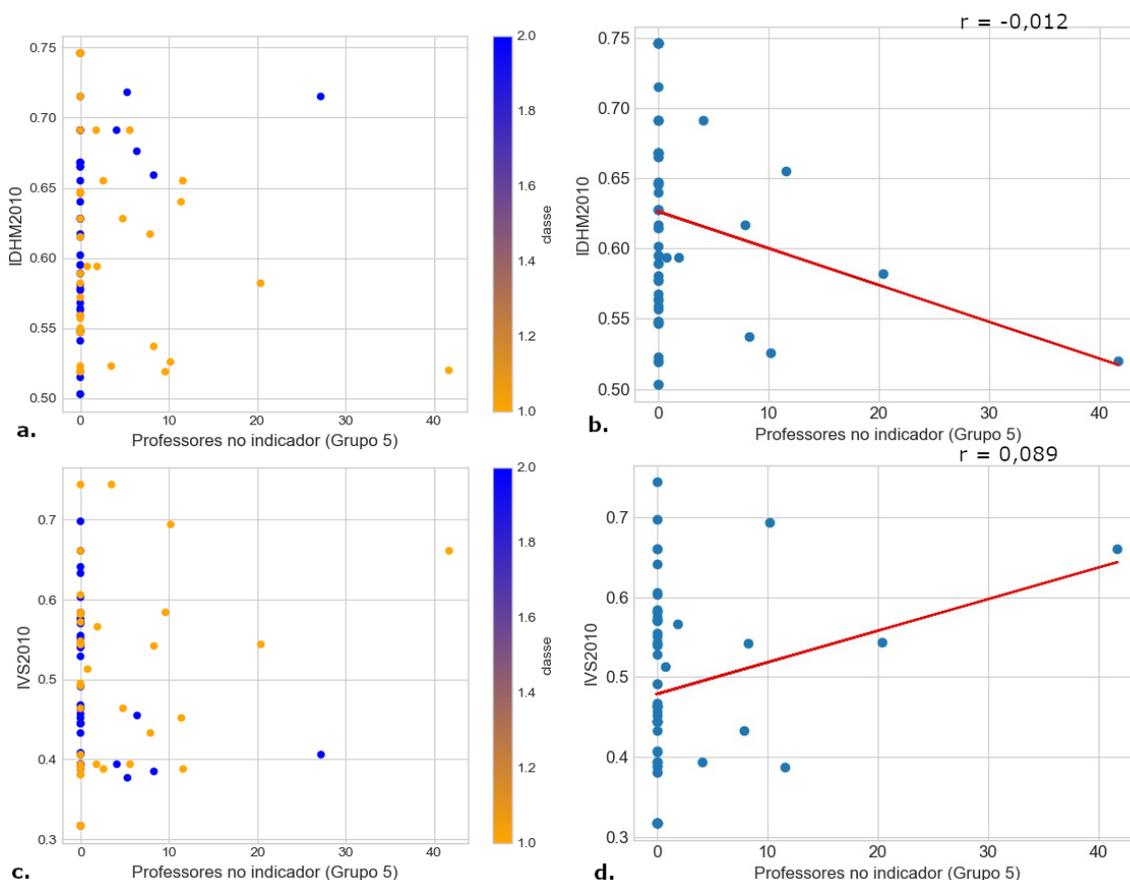


Figura 25 – Professores do Grupo 5 x IDHM x IVS

Fonte: Autor

A Figura 25(a.) e (b) ilustra a correlação entre desenvolvimento humano no eixo vertical, e o perfil de docentes pertencente ao (Grupo 5) no eixo horizontal, o índice de correlação $r=-0,012$, sugere a baixa relação entre desenvolvimento humano e docentes pertencentes a este grupo. Do mesmo modo, com índice de correlação $r=0,089$, os resultados evidenciados na Figura 25(c.) e (d), evidencia a baixa correlação entre vulnerabilidade social e docentes pertencentes ao Grupo supracitado. O resultados apontados na Figura 25, correspondem a 39% das escolas do estado, e neste percentual, 75% possui IDEB abaixo da média, conforme observado na Figura 25(a.) e (c), onde há grande concentração de escolas com porcentagem em torno de zero, porém, o espalhamento dos alunos em maioria são do grupo de escolas abaixo da média.

5.1 Precisão do Modelo de Regressão Linear

O algoritmo RL retornou como resultado uma reta com os valores preditos a posição da reta é calculada pela equação de forma que ela consiga acompanhar a tendência de observada nos dados. A distância entre cada ponto e a reta contém um valor de erro,

quanto maior for a distância maior será o erro. O cálculo do erro foi feito pela equação (2), cujo resultado está descrito na Tabela 5.

Tabela 5 – Métricas de avaliação.

Variáveis	MAE	MSE	RMSE	MAPE	SMAPE
Escolaridade do Pai, ou homem responsável x Nota Média Padronizada (n)	0,25	0,12	0,35	0,06	6,8
Escolaridade da Mãe, ou mulher responsável x Nota Média Padronizada (n)	0,24	0,12	0,35	0,06	6,4
IDHM 2010 x Nota Média Padronizada (n)	0,05	0,004	0,06	0,08	8,8
IVS 2010 x Nota Média Padronizada (n)	0,008	0,009	0,09	0,17	16,5
IDHM 2010 x Abandono escolar	0,05	0,005	0,07	0,09	9,5
IVS 2010 x Abandono escolar	2,0	1,1	3,3	8,6	8,6
IDHM 2010 x Porcentagem de professores pertencentes ao Grupo 1	0,06	0,005	0,07	0,10	10,6
IVS 2010 x Porcentagem de professores pertencentes ao Grupo 1	0,09	0,01	0,11	0,20	19,4
IDHM 2010 x Porcentagem de professores pertencentes ao Grupo 3	0,05	0,004	0,06	0,09	9,3
IVS 2010 x Porcentagem de professores pertencentes ao Grupo 3	0,08	0,009	0,09	0,17	16,9
IDHM 2010 x Porcentagem de professores pertencentes ao Grupo 5	0,07	0,006	0,07	0,11	11
IVS 2010 x Porcentagem de professores pertencentes ao Grupo 5	0,08	0,010	0,10	0,19	18,8

Fonte: Autor

Os resultados apresentados na Tabela 5, evidenciam algumas tendências. Ao verificar a métrica SMAPE, que é uma medida de desempenho onde o algoritmo pode ser analisado sem a necessidade de haver alguma comparação com outros modelos. Notou-se que as variáveis relacionadas à escolaridade dos pais, possuem menor porcentagem de erro quando comparada às demais. Revelando que a trajetória escolar dos pais possui uma relação importante com o bom desempenho dos filhos.

5.2 Considerações finais

Neste capítulo, conclui-se que avaliar o desempenho faz parte do processo do ciclo de vida de qualquer sistema e exige uma atenta definição da metodologia e ferramentas, para que a escolha seja funcional e intuitiva. A apresentação das diferentes técnicas de

modelagem fornece um amplo conjunto de opções, que dependem do estágio do desenvolvimento do sistema alvo.

6 Conclusão

Este estudo buscou por meio da MDE, desvelar diferentes fatores que estão relacionados ao baixo desempenho das escolas estaduais no estado do Pará. Sinalizando que a situação socioeconômica de risco das famílias que possuem filhos em idade escolar, prejudica não só o desempenho desses alunos, mas também a permanência em sala de aula. Por isso, deve haver um maior engajamento das escolas para estimular o interesse dos alunos pela persistência no ambiente escolar. Além de subsídios por parte do governo para que os professores possam exercer seu papel na sociedade de forma plena.

Embora o índice do IDEB seja um ponto de partida para subsidiar políticas públicas, não deve ser considerado como único meio para aferição da qualidade educacional. Pois deixa de fora diversos elementos importantes e que estão diretamente ligados à qualidade da educação. Nesse sentido, vem a importância do procedimento supracitado. Nesta etapa, as fases do processo de mineração de dados foram importantes para nortear de forma mais assertiva o processo de descoberta de conhecimento, afim de desvendar diferentes padrões. Os resultados apontados por este estudo, revelaram algumas dessas correlações.

Nessa perspectiva, a situação das famílias que possuem filhos em idade escolar e estão em situação socioeconômica de risco, tem desempenho abaixo do esperado. Pois, elas precisam participar de programas sociais que de fato estimulem não somente o ingresso e permanência na escola, mas também incentivos para que os alunos possam se sentir interessados em aprender. Dessa forma, criando uma consciência social para que as famílias de fato olhem para a educação como algo de altíssima importância na vida de um indivíduo.

Ao mesmo tempo, o governo deve propor meios para que os docentes tenham cada vez mais acesso facilitado a formação continuada e adequada à disciplina ministrada, salários dignos condizentes com o importante papel na sociedade e outras ações para que o profissional se sinta estimulado no seu ambiente escolar, e dessa forma, contribuir com a educação de qualidade.

A escola tem um importante papel social e deve estar alinhada com boas práticas no que se refere a planejamento e ações para que haja um engajamento maior da sociedade ao seu redor, estando atenta à diversas particularidades que possam de alguma forma interferir na adaptação dos alunos no processo de aprendizagem, garantindo infraestrutura adequada para que de alguma forma possam por em prática o que se aprende em sala de

aula.

Resolver o problema da qualidade da educação no estado do Pará é algo complexo, pois são diversos fatores relacionados ao desempenho escolar e, muitos destes, não são mensurados ou não há disponibilidade para que se possa incorporar nas análises que se pode realizar atualmente. É crucial que diferentes análises fora do ambiente escolar possam ser realizadas por diferentes algoritmos de classificação ou regressão, para que haja compreensão dos diversos fatores que podem ser utilizados em futuras políticas educacionais.

A contribuição deste estudo está em reunir diferentes bases para compor um arcabouço de dados que podem gerar informações amplas à respeito da qualidade da educação no estado do Pará, sob diferentes perspectivas. Possibilitando certa compreensão do problema à ser tratado, por meio de algoritmos de previsão. Neste estudo, ficou evidente a importância da utilização de regressores para identificar padrões relacionados ao baixo desempenho, que antes eram identificados somente em pesquisa de campo por meio de entrevista. Contribuindo de forma relevante na identificação e aprimoramento de aspectos relevantes à melhoria do ensino básico.

6.1 Trabalhos futuros

Como possíveis desdobramentos deste trabalho, seguem os principais objetivos futuros:

- Explorar novas técnicas de previsão a fim de obter maior eficiência ao encontrar a solução ótima com um menor custo computacional;
- Realizar análise multifatorial para identificar de forma ainda mais assertiva os diferentes aspectos envolvidos no processo de aprendizagem;
- Explorar diferentes cenários com um número ainda mais amplo de dados e possivelmente relacioná-los ao contexto da pandemia de Covid-19, afim de estimar como os prejuízos causados pela emergência sanitária influenciaram nos índices educacionais.

6.2 Dificuldades encontradas

- A principal dificuldade que norteou o trabalho foi a pouca disponibilidade de conteúdo em algumas bases de dados, tornando quase que obrigatório o relacionamento entre bases distintas para compensar determinada informação. A ausência importante notada foi a falta de dados de rendimentos dos alunos nas bases do SAEB;

- Outro ponto oneroso foi justamente a disponibilização dos microdados que nem sempre estão de fácil acesso. Em alguns momentos os dados tiveram que ser coletados diretamente da página web por conta dessa ausência ou dificuldade em localizar;

Referências

- ADEODATO, P. J. et al. Predição de desempenho de escolas privadas usando o enem como indicador de qualidade escolar. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2014. v. 25, n. 1, p. 891. Citado 2 vezes nas páginas 20 e 23.
- AGAOGLU, M. Predicting instructor performance using data mining techniques in higher education. *IEEE Access*, 2016. IEEE, v. 4, p. 2379–2387, 2016. Citado 2 vezes nas páginas 21 e 24.
- ALVES, M. T. G.; SOARES, J. F. Contexto escolar e indicadores educacionais: condições desiguais para a efetivação de uma política de avaliação educacional. *Educação e pesquisa*, 2013. SciELO Brasil, v. 39, n. 1, p. 177–194, 2013. Citado na página 2.
- AMO, S. de. Técnicas de mineração de dados. *Jornada de Atualização em Informática*, 2004. 2004. Citado na página 12.
- ATRICON-IRB. Sistema de monitoramento dos planos da educação. 2017. 2017. Disponível em: <<https://tceduca.irbcontas.org.br/pne//public/inicio>>. Citado 2 vezes nas páginas 27 e 34.
- BREIMAN, L. Random forests. *Machine learning*, 2001. Springer, v. 45, n. 1, p. 5–32, 2001. Citado na página 14.
- BROOKOVER, W. B. et al. Elementary school social climate and school achievement. *American educational research journal*, 1978. Sage Publications Sage CA: Los Angeles, CA, v. 15, n. 2, p. 301–318, 1978. Citado na página 5.
- CALIXTO, K. et al. Mineração de dados aplicada a educação: um estudo comparativo acerca das características que influenciam a evasão escolar. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2017. v. 28, n. 1, p. 1447. Citado 2 vezes nas páginas 21 e 23.
- CHECHIA, V. A.; ANDRADE, A. d. S. O desempenho escolar dos filhos na percepção de pais de alunos com sucesso e insucesso escolar. *Estudos de Psicologia (Natal)*, 2005. SciELO Brasil, v. 10, p. 431–440, 2005. Citado 2 vezes nas páginas 40 e 41.
- CHICCO, D.; WARRENS, M. J.; JURMAN, G. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, 2021. PeerJ Inc., v. 7, p. e623, 2021. Citado na página 31.
- CHIRINÉA, A. M.; BRANDÃO, C. d. F. O ideb como política de regulação do estado e legitimação da qualidade: em busca de significados. *Ensaio: Avaliação e Políticas Públicas em Educação*, 2015. SciELO Brasil, v. 23, n. 87, p. 461–484, 2015. Citado na página 6.

- COELHO, M. I. d. M. Veinte años de la evaluación de la educación básica en brasil: aprendizajes y desafíos. *Ensaio: Avaliação e políticas públicas em Educação*, 2008. SciELO Brasil, v. 16, n. 59, p. 229–258, 2008. Citado 2 vezes nas páginas 5 e 6.
- COSTA, C. C. R. C. O ideb como indicador da qualidade da educação brasileira: o caso da região sudeste. *e-xacta*, 2010. v. 3, n. 2, 2010. Citado 2 vezes nas páginas 19 e 22.
- COSTA, L. C. A. d. et al. Tecendo relações entre a motivação para as aulas de educação física e o ideb. *Revista Brasileira de Ciências do Esporte*, 2018. SciELO Brasil, v. 40, n. 4, p. 370–373, 2018. Citado 2 vezes nas páginas 20 e 22.
- DIGIAMPIETRI, L. A. et al. Mineração de dados para identificação de alunos com alto risco de evasão: Um estudo de caso. *Revista de Graduação USP*, 2016. v. 1, n. 1, p. 17–23, 2016. Citado 2 vezes nas páginas 21 e 24.
- FACIROLI, J.; SIRQUEIRA, T. F. M. Análise espacial da taxa de cobertura do programa bolsa família para os municípios brasileiros. *Desenvolvimento Regional em debate: DRd*, 2017. Universidade do Contestado, v. 7, n. 2, p. 118–137, 2017. Citado na página 45.
- FARBER, B.; LARSON, R. Estatística aplicada. 4ª edição. *São Paulo*, 2010. 2010. Citado 3 vezes nas páginas 15, 16 e 18.
- FAYYAD, U. et al. From data mining to knowledge discovery in databases. *AI magazine*, 1996. v. 17, n. 3, p. 37–37, 1996. Citado na página 12.
- FERNANDES, R.; GREMAUD, A. P. Qualidade da educação: avaliação, indicadores e metas. *Educação básica no Brasil: construindo o país do futuro*. Rio de Janeiro: Elsevier, 2009. v. 1, p. 213–238, 2009. Citado 2 vezes nas páginas 1 e 2.
- FILHO, R. B. S.; ARAÚJO, R. M. de L. Evasão e abandono escolar na educação básica no brasil: fatores, causas e possíveis consequências. *Educação por escrito*, 2017. v. 8, n. 1, p. 35–48, 2017. Citado na página 44.
- GALVÃO, N. D.; MARIN, H. d. F. Técnica de mineração de dados: uma revisão da literatura. *Acta Paulista de Enfermagem*, 2009. SciELO Brasil, v. 22, n. 5, p. 686–690, 2009. Citado na página 12.
- GUIMARA, M. H. et al. A consolidação da política de avaliação da educação básica no brasil. *Revista Meta: Avaliação*, 2009. v. 1, n. 3, p. 271–296, 2009. Citado 2 vezes nas páginas 5 e 6.
- HADDAD, F. *O Plano de Desenvolvimento da Educação: razões, princípios e programas*. [S.l.]: MEC–Ministério da Educação, INEP–Instituto Nacional de Estudos e Pesquisas . . . , 2008. Citado na página 6.
- HALLAK, R.; FILHO, A. J. P. Metodologia para análise de desempenho de simulações de sistemas convectivos na região metropolitana de são paulo com o modelo arps: sensibilidade a variações com os esquemas de advecção e assimilação de dados. *Revista Brasileira de Meteorologia*, 2011. SciELO Brasil, v. 26, p. 591–608, 2011. Citado na página 31.

- IBGE. Produto interno bruto dos municípios. 2010. 2010. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv62930.pdf>>. Citado 2 vezes nas páginas 41 e 42.
- IBGE. Síntese do Índice de desenvolvimento humano municipal - idhm para o estado do Pará. 2010. 2010. Disponível em: <<http://www.fapespa.pa.gov.br/upload/Arquivo/anexo/234.pdf?id=1479216410>>. Citado na página 41.
- INEP. Censo escolar. 2007. 2007. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar>>. Citado na página 7.
- INEP. Indicador de adequação da formação do docente da educação básica. 2014. 2014. Disponível em: <https://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2014/docente_formacao_legal/nota_tecnica_indicador_docente_formacao_legal.pdf>. Citado na página 28.
- INEP. Sistema de avaliação da educação básica (saeb). 2017. 2017. Disponível em: <<https://medium.com/@inep/resultados-do-saeb-2017-f471ec72168d>>. Citado na página 2.
- INEP. Pisa 2018 revela baixo desempenho escolar em leitura, matemática e ciências no Brasil. 2019. 2019. Disponível em: <http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/pisa-2018-revela-baixo-desempenho-escolar-em-leitura-matematica-e-ciencias-no-brasil/21206>. Citado na página 1.
- INEP. Histórico. 2020. 2020. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb/historico>>. Citado 3 vezes nas páginas 7, 8 e 9.
- IPEA. Atlas da vulnerabilidade social nos municípios brasileiros. 2015. 2015. Disponível em: <http://ivs.ipea.gov.br/images/publicacoes/Ivs/publicacao_atlas_ivs.pdf>. Citado 2 vezes nas páginas 34 e 42.
- IPEA. *Atlas do Desenvolvimento Humano nas Regiões Metropolitanas Brasileiras*. [S.l.: s.n.], 2017. Citado 3 vezes nas páginas 10, 11 e 12.
- IPEA. Pof 2017-2018. 2022. 2022. Disponível em: <<https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/28896-pof-2017-2018-proporcao-de-domicilios-com-seguranca-alimentar-fica-abaixo-do-resultado-de-2004>>. Citado na página 35.
- JUNIOR, R. N. et al. Estimação de índices de aprovação e reprovação escolar do ensino médio. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2019. v. 30, n. 1, p. 339. Citado 2 vezes nas páginas 20 e 23.
- LIMA, F. R.; OLIVEIRA, G. A. O índice de desenvolvimento da educação básica (ideb) como instrumento de avaliação em larga escala: Expedições pelo ensino. *Revista da FAESF*, 2018. v. 2, n. 3, 2018. Citado 2 vezes nas páginas 20 e 22.
- MEC. Ideb - resultados e metas. 2020. 2020. Disponível em: <<http://ideb.inep.gov.br/resultado/>>. Citado na página 2.

- MEC. Sistema de avaliação da educação básica (saeb). 2020. 2020. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb>>. Citado 2 vezes nas páginas 1 e 2.
- MEC. Sistema de avaliação da educação básica (saeb). 2020. 2020. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb>>. Citado na página 8.
- NASCIMENTO, E. C. et al. O papel do programa bolsa família na segurança alimentar das famílias do território do marajó, pa. *Interações (Campo Grande)*, 2017. SciELO Brasil, v. 18, p. 59–70, 2017. Citado na página 46.
- NISBET, R.; ELDER, J.; MINER, G. *Handbook of statistical analysis and data mining applications*. [S.l.]: Academic Press, 2009. Citado na página 13.
- NOVAKOVIC, J. Using information gain attribute evaluation to classify sonar targets. In: CITESEER. *17th Telecommunications forum TELFOR*. [S.l.], 2009. p. 1351–1354. Citado na página 14.
- PEREIRA, S. d. O. G.; CABRAL, J. P. C. Informalidade e crise do emprego no brasil. *Humanidades & Inovação*, 2019. v. 6, n. 18, p. 92–102, 2019. Citado na página 34.
- PINTO, G.; JÚNIOR, O. d. G. F.; COSTA, E. de B. 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-Revista Novas Tecnologias na Educação*, 2019. v. 17, n. 3, p. 183–193, 2019. Citado 2 vezes nas páginas 19 e 22.
- PINTO, G. da S. et al. Identificação dos fatores de melhorias no ideb pelo uso de mineração de dados: Um estudo de caso em escolas municipais de maceió. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2019. v. 30, n. 1, p. 1828. Citado 2 vezes nas páginas 20 e 23.
- PNE. Pne em movimento. 2014. 2014. Disponível em: <<http://pne.mec.gov.br/18-planos-subnacionais-de-educacao/543-plano-nacional-de-educacao-lei-n-13-005-2014>>. Citado 2 vezes nas páginas 33 e 35.
- PNUD. O que é idh. 2014. 2014. Disponível em: <<https://www.br.undp.org/content/brazil/pt/home/idh0/conceitos/o-que-e-o-idh.html>>. Citado na página 10.
- REIF, D. M. et al. Feature selection using a random forests classifier for the integrated analysis of multiple data types. In: IEEE. *2006 IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology*. [S.l.], 2006. p. 1–8. Citado na página 15.
- REZENDE, S. O. et al. Mineração de dados. *Sistemas inteligentes: fundamentos e aplicações*, 2003. v. 1, p. 307–335, 2003. Citado 2 vezes nas páginas 25 e 26.
- RIGO et al. Minerando dados educacionais com foco na evasão escolar: oportunidades, desafios e necessidades. In: *Anais do Workshop de Desafios da Computação Aplicada à Educação*. [S.l.: s.n.], 2012. p. 168–177. Citado 2 vezes nas páginas 21 e 24.

- RODRIGUES, S. C. A. *Modelo de regressão linear e suas aplicações*. Tese (Doutorado) — Universidade da Beira Interior, 2012. Citado 2 vezes nas páginas 16 e 17.
- ROMERO, C.; VENTURA, S. Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2013. Wiley Online Library, v. 3, n. 1, p. 12–27, 2013. Citado na página 13.
- SANTOS, E. B. S. et al. *Faculdade São Luís de França*, 2016. 2016. Disponível em: <<https://portal.fslf.edu.br/wp-content/uploads/2016/12/FATORES-SOCIO-ECONOMICOS.pdf>>. Citado 4 vezes nas páginas 20, 23, 36 e 37.
- SELL, I. Utilização da regressão linear como ferramenta de decisão na gestão de custos. In: *Anais do Congresso Brasileiro de Custos-ABC*. [S.l.: s.n.], 2005. Citado na página 17.
- SENARC. Bolsa família: Visão geral. 2017. 2017. Disponível em: <<https://aplicacoes.mds.gov.br/sagirms/bolsafamilia/painel.html>>. Citado na página 45.
- SOARES, J. F.; XAVIER, F. P. Pressupostos educacionais e estatísticos do ideb. *Educação & Sociedade*, 2013. SciELO Brasil, v. 34, p. 903–923, 2013. Citado na página 8.
- TORGO, L. *Data mining with R: learning with case studies*. [S.l.]: CRC press, 2016. Citado na página 14.