



**UNIVERSIDADE FEDERAL DE ALAGOAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM
COMPUTACIONAL DO CONHECIMENTO- PPMCC**

WANDERSON RUBIAN MARTINS RODRIGUES

**UMA ABORDAGEM HUMANO COMPUTACIONAL PARA IDENTIFICAÇÃO DOS
FATORES QUE INFLUENCIAM O IDEB NA REDE DE ENSINO BÁSICO NA
VISÃO DOS GESTORES EDUCACIONAIS**

**MACEIÓ-ALAGOAS
2020**

WANDERSON RUBIAN MARTINS RODRIGUES

**UMA ABORDAGEM HUMANO COMPUTACIONAL PARA IDENTIFICAÇÃO DOS
FATORES QUE INFLUENCIAM O IDEB NA REDE DE ENSINO BÁSICO NA
VISÃO DOS GESTORES EDUCACIONAIS**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Modelagem Computacional do Conhecimento do Instituto de Computação da Universidade Federal de Alagoas.

Orientador: Prof. Dr. Evandro de Barros Costa

Coorientador: Prof. Dr. João Carlos Cordeiro Barbirato

MACEIÓ-ALAGOAS
2020

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecária: Taciana Sousa dos Santos – CRB-4 – 2062

R696a Rodrigues, Wanderson Rubian Martins.

Uma abordagem humano computacional para identificação dos fatores
que influenciam o IDEB na rede de ensino básico na visão de gestores educacionais /
Wanderson Rubian Martins Rodrigues. – 2020.

90 f. il. : figs. ; graf. ; tabs. color.

Orientador: Evandro de Barros Costa.

Coorientador: João Carlos C. Barbirato.

Dissertação (Mestrado em Modelagem Computacional do
Conhecimento) – Universidade Federal de Alagoas. Instituto de Computação.
Maceió, 2020.

Bibliografia: f. 90-93.

Apêndices: f. 94-98.

Anexos: f. 99-105.

1. Gestão pública. 2. Índice de Desenvolvimento da Educação Básica
(IDEB). 3. Mineração de dados educacionais. I. Título.

CDU: 37.057: 004.8

Folha de Aprovação

Wanderson Rubian Martins Rodrigues

Uma Abordagem Humano Computacional para Identificação dos Fatores que influenciam o IDEB na Rede de Ensino Básico na Visão dos Gestores Educacionais

Dissertação submetida ao corpo docente do Programa de Pós-Graduação em Modelagem Computacional de Conhecimento da Universidade Federal de Alagoas e aprovada em 15 de julho de 2020.



Prof. Dr. Evandro de Barros Costa

Instituto de Computação - UFAL

Orientador



Prof. Dr. João Carlos Cordeiro Barbirato

Centro de Tecnologia- UFAL

Coorientador

Banca Examinadora:



Prof. Dr. Marcus de Melo Braga

Instituto de Computação - UFAL

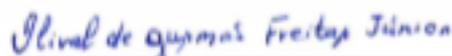
Examinador interno



Prof. Dr. Bruno Almeida Pimentel

Instituto de Computação- UFAL

Examinador externo



Prof. Dr. Olival de Gusmão Freitas Junior

Instituto de Computação- UFAL

Examinador externo

AGRADECIMENTOS

Serei sempre grato a Jesus Cristo autor e redentor da nossa fé, a Deus que proporcionou força nesta jornada!

Aos melhores professores do mundo para mim: Dr. Olival Freitas, Dr. João C. Barbirato, Dr. Evandro Costa, Dra. Roberta Vilhena, Dr. Marcus Braga e ao mestre e amigo Giseldo Neo que acreditaram e depositaram confiança e empenho ao longo desta trajetória acadêmica, aos meus filhos Ester Alves Rodrigues e Israel Alves Rodrigues que compreenderam que esse sonho poderia se transformar em realidade, que estes sejam os próximos a contemplarem este caminho de pesquisa e conhecimento em prol da nossa sociedade e da ciência. Aos meus pais que sempre me ajudaram com palavras de conforto e otimismo acerca deste processo acadêmico.

Aos demais colegas que surgiram em etapas importantes deste trabalho, entre eles, especial gratidão a Edjane Ferreira, pelas palavras de ânimo, ao mestre em Meteorologia, Anselmo Santos, pelo apoio durante o desenvolvimento deste trabalho. Ao mestre Glevson da Silva Pinto, que ajudou com valiosas críticas para melhoria deste trabalho, principalmente na coleta, tratamento e análise dos dados.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, pela qual esse trabalho foi apoiado e concluído.

*Em nosso mundo, conhecimento é poder, e a educação empodera.
Ela é parte indispensável da equação do desenvolvimento.
Ela tem valor intrínseco – que se estende muito além do
econômico – para empoderar as pessoas a determinar o
próprio destino. É por isso que a oportunidade de ser educado é
central para o avanço do desenvolvimento humano.*

*Helen Clark
Administradora do PNUD*

RESUMO

A busca pela alocação efetiva dos recursos a fim de alcançar o maior número de pessoas é o desafio constante do gestor público. Desta forma, o presente trabalho tem como objetivo identificar quais fatores, percebido pelos gestores educacionais, impacta no IDEB das escolas municipais de Maceió-Alagoas. Os dados iniciais trabalhados nesta pesquisa já existem, porém precisam ser analisados frente à realidade cotidiana do âmbito escolar, estabelecendo uma relação entre a percepção dos gestores e os resultados divulgados pelo INEP/MEC. Assim, surge a pergunta central que norteia esse estudo: como integrar a visão computacional (máquina) com a visão dos gestores educacionais (humanos) para identificar os fatores que possam influenciar positivamente a melhoria do IDEB, visando apoiar a tomada de decisão dos gestores educacionais? Para tanto, utilizou-se dados do teste SAEB das escolas públicas de Maceió. A presente pesquisa tem caráter descritivo e o desenvolvimento metodológico da investigação advém da revisão da literatura, da análise e compilação dos dados do rendimento dos alunos dos anos iniciais do ensino fundamental da rede de ensino municipal de Maceió-Alagoas no IDEB do ano de 2015, e da realização de entrevistas com os diretores escolares, visando compreender as impressões da gestão escolar sobre os resultados do IDEB, conduzindo um estudo experimental, obtendo resultados na tarefa de identificação dos atributos relevantes, como os extraídos pela visão computacional principalmente nas dimensões: violência na escola; políticas, ações e programas na escola; gestão escolar e recursos financeiros. Na percepção dos gestores escolares, observa-se os seguintes atributos como relevantes para melhoria do IDEB: o nível de escolaridade do gestor, o processo de elaboração do projeto pedagógico da escola, medidas para redução das taxas de reprovação, medidas para minimizar as faltas dos alunos, melhoria na comunicação escola-família, a questão da merenda escolar, apoio das instâncias superiores, recursos financeiros para financiar os gastos, o guia didático do MEC e as escolhas dos livros didáticos pelos professores.

Palavras-chave: Tomada de Decisão, Mineração de Dados, IDEB, Gestão Pública.

ABSTRACT

The search for effective resource allocation in order to reach the largest number of people is the constant challenge of the public manager. Thus, this work aims to identify which factors, perceived by educational managers, impact on the IDEB of the municipal schools of Maceió-Alagoas. The initial data worked on in this research already exist, but they need to be analysed in view of the daily reality of the school environment, establishing a relationship between the perception of managers and the results disseminated by INEP/MEC. Thus, the central question that guides this study arises: how to integrate the computational vision (machine) with the vision of educational managers (human) in order to identify the factors that may positively influence the improvement of the IDEB, aiming at supporting the decision making of educational managers? For this, data from the SAEB test of the public schools of Maceió were used. This research has a descriptive character and the methodological development of the research comes from the literature review, the analysis and compilation of data on the performance of students of the initial years of primary schools in the Maceió-Alagoas municipal school network in the IDEB of 2015, and from the carrying out of interviews with school headmasters, in order to understand the impressions of school management on the results of the IDEB, conducting an experimental study, obtaining results in the task of identifying the relevant attributes, such as those extracted by computer vision mainly in the dimensions: violence at school; policies, actions and programs at school; school management and financial resources. In the perception of school managers, the following attributes are observed as relevant for the improvement of IDEB: the level of schooling of the manager, the process of elaboration of the pedagogical project of the school, measures to reduce failure rates, measures to minimise student absences, improvement in school-family communication, the issue of school meals, support from higher levels, financial resources to finance the expenses, the MEC didactic guide and the choices of textbooks by teachers.

Keywords: Decision Making, Data Mining, IDEB, Public Management.

LISTA DE FIGURAS

Figura 1 - Fluxo esquemático dos fundamentos utilizados na pesquisa.....	45
Figura 2 - Etapas específicas do desenvolvimento da pesquisa.	50

LISTA DE TABELAS

Tabela 1 - Relação das escolas pesquisadas e regiões administrativas.	44
Tabela 2 - Dimensão descritas no questionário do IDEB aplicado para os gestores em 2015.	69
Tabela 3 - Análise da percepção dos gestores.	71

LISTA DE QUADROS

Quadro 1 - Metas para a educação básica de Alagoas.....	21
Quadro 2 - Dimensões do questionário do diretor escolar.	53
Quadro 3 - Abordagens de seleção de atributos utilizados na pesquisa.	54
Quadro 4 - Seleção de atributos extraídos da máquina através da abordagem de seleção de atributos por categoria.	56
Quadro 5 - Acurácia de predição para o conjunto de dados IDEB 2015 (Gestores) obtida por meio de conjunto de treinamento e teste.	57
Quadro 6 - Relação das escolas escolhidas para a realização da pesquisa (INEP 2015).	67
Quadro 7-Atributos selecionados por meio da EDM e através dos especialistas (Gestor escolar e Secretário de Educação do município estudado).	68

LISTA DE GRÁFICOS

Gráfico 1 - Palavras mais citadas na entrevista.	77
---	----

LISTA DE ABREVIATURAS

ANEB	Avaliação Nacional da Educação Básica
ANRESC	Avaliação Nacional do Rendimento Escolar
HTPC	Horário de Trabalho Pedagógico Coletivo
HTPI	Horário de Trabalho Pedagógico Individual
IBGE	Instituto Brasileiro de Geografia e Estatística
IDEB	Índice de Desenvolvimento da Educação Básica
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
MDE	Mineração de Dados Educacionais
SEMED	Secretaria Municipal de Educação

SUMÁRIO

1	INTRODUÇÃO.....	16
1.1	Contextualização do Tema.....	16
1.2	Motivação	18
1.3	Problemática	19
1.4	Objetivos.....	19
1.4.1	Objetivo Geral	19
1.4.2	Objetivos Específicos	19
1.5	Relevância da Pesquisa.....	20
1.6	Estrutura da Dissertação	21
2	FUNDAMENTAÇÃO TEÓRICA E REVISÃO DA LITERATURA.....	23
2.1	Gestão da Educação Pós-LDB	23
2.2	Avaliação da Educação Básica	25
2.3	IDEB.....	27
2.4	Mineração de Dados Educacionais.....	29
2.5	Seleção de Atributos	33
2.5.1	Abordagem Filtro.....	33
2.5.2	Abordagem Wrapper (Embrulhamento)	34
2.5.3	Abordagem <i>Embedded</i> (Embutida)	34
2.6	Validação cruzada (Cross-validation).....	35
2.7	Técnicas de Classificação de Mineração de Dados Educacionais	35
2.7.1	Classificadores Baseados em Regras	36
2.7.2	Árvores de Decisão	36
2.7.3	Redes Bayesianas.....	37
2.7.4	Máquina de Vetores de Suporte.....	38
2.7.5	Agrupamento	38
2.7.6	Regressão Linear	38

2.8	Trabalhos Correlatos.....	39
3	METODOLOGIA.....	44
4	UMA ABORDAGEM HUMANO COMPUTACIONAL PARA IDENTIFICAÇÃO DOS FATORES QUE INFLUENCIAM O IDEB NA REDE DE ENSINO BÁSICO DE MACEIÓ-ALAGOAS.....	52
4.1	Introdução	52
4.2	Visão Computacional.....	52
4.3	Visão Humana	62
4.4	Abordagem Integradora	66
4.5	Análise da percepção dos fatores dos gestores escolares.....	71
4.6	Análise da fala dos gestores	73
5	CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS.....	79
6	REFERÊNCIAS	81
	APÊNDICES	85
	APÊNDICE A – Solicitação de Aplicação de Questionário Para Fins de Trabalho Acadêmico Em Escolas da Rede Municipal de Maceió	85
	APÊNDICE C – Script Escrito em Python 3 utilizado para gerar nuvens de palavras das entrevistas dos diretores.....	88

1 INTRODUÇÃO

1.1 Contextualização do Tema

A busca por melhoria na educação e pela alocação efetiva dos recursos, tanto humano como financeiro, a fim de alcançar o maior número de pessoas é o desafio constante do gestor público, sendo condição imperativa para otimizar tais recursos. Para tanto, informações de qualidade são fundamentais para que o gestor possa tomar decisões.

O aceleramento da qualidade educacional por meio de análises minuciosas de dados bem como a experiência prévia do responsável pela interpretação desses dados pode ser facilitada através de técnicas e ferramentas computacionais, que auxiliam o tomador de decisões. É neste contexto, que surge a visão computacional alinhada com a visão humana dos gestores educacionais, compondo assim, uma harmonia para o melhoramento da educação de qualidade, ponto alvo deste trabalho.

Segundo Cantero *et al.* (2013), a educação é fundamental para a transformação de uma nação. Segundo os autores, os municípios que priorizam a educação apresentam rendimentos superiores e isso acaba se refletindo em uma melhoria da expectativa e qualidade de vida.

Destaca-se que a melhora da qualidade da educação no Brasil tem sido um dos maiores objetivos dos sistemas de ensino e dos pesquisadores em educação. Nos últimos anos, os resultados das avaliações da educação (básica e superior) passaram a ser consideradas como principais parâmetros de qualidade da educação no Brasil. Nesse contexto, existem três pontos que estão diretamente ligados à avaliação educacional de larga escala: a qualidade da educação, indicadores educacionais e a prestação de contas dos recursos investidos (DA SILVA, 2012).

As transformações ocorridas na gestão e organização dos sistemas educacionais vêm demandando a tomada de decisões no sentido de melhorar o processo de ensino e aprendizagem nas instituições educacionais públicas do Brasil. A utilização de indicadores de avaliação de larga escala tem norteado os governos em âmbito federal, estadual e municipal, tais avaliações possibilitam orientar políticas públicas educacionais. Esses resultados, assim, sintetizam indicadores globais de qualidade da educação, entre os quais o mais importante é o Índice de Desenvolvimento da Educação Básica (IDEB), introduzido pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP).

O IDEB tem sido influenciado por vários fatores educacionais, presentes nas escolas oriundos de avaliações sobre o aproveitamento escolar dos alunos, por meio do censo escolar e as médias de desempenho nas avaliações do Sistema de Avaliação da Educação Básica (IDEB,

2016).

Atualmente, a qualidade da educação de uma escola, de uma rede de ensino, de um município, de um estado e, até mesmo do país, é mensurada quantitativamente através do IDEB, justamente, com o propósito de aferir a qualidade da educação básica. O IDEB é um indicador sintético que tem o resultado baseado na taxa de aprovação dos alunos, obtido do Censo Escolar realizado anualmente, e no desempenho na Prova Brasil - Avaliação Nacional do Rendimento Escolar – que é um teste padronizado aplicado de forma censitária e bianual, no 5º ano e 9º ano do ensino fundamental e foca nas disciplinas de Língua Portuguesa, com ênfase na leitura, e Matemática, com ênfase na resolução de problemas (INEP, 2016).

Esse índice tem sido influenciado por vários fatores educacionais, presentes nas escolas oriundos de avaliações sobre o aproveitamento escolar dos alunos, por meio do censo escolar e as médias de desempenho nas avaliações do Sistema de Avaliação da Educação Básica, a Prova Brasil, a Avaliação Nacional de Alfabetização entre outras.

Dentre as avaliações está a Prova Brasil cuja nota estima a qualidade de ensino em escolas municipais e estaduais. Tal prova é aplicada nos anos iniciais (5º ano) do ensino fundamental, nos anos finais (9º Ano) do ensino fundamental e no ano final (3º ano) do ensino médio. Além disso, a Prova Brasil também é composta por um conjunto de questionários que são aplicados para o corpo discente, docente e diretoria; a fim de identificar os aspectos socioeconômicos que afetam a qualidade da educação.

Com base nas informações obtidas por esses questionários, o MEC e as secretarias estaduais e municipais de educação podem definir ações voltadas ao aprimoramento da qualidade da educação no país e a redução das desigualdades existentes.

Pesquisadores têm produzido estudos para analisar a influência das condições e das estruturas educacionais e econômicas na qualidade do ensino ministrado nas redes municipais da educação fundamental pública no Brasil. Segundo levantamento bibliográfico realizado no presente estudo, não se pode afirmar que uma variável ou mesmo uma combinação de variáveis (HAIR *et al.*, 2009), teriam a primazia na explicação do desempenho dos alunos em sistemas de avaliação geral.

Os estudos sugerem que há uma considerável combinação de variáveis com distintas descrições para cada contexto e momento. Fatores associados às características pessoais e familiares dos alunos, comprometimento dos docentes e gestores educacionais, das estruturas macroeconômicas e sociais até aquelas que se podem identificar com as estruturas de oferta do serviço público, presentes nas localidades do desenvolvimento das atividades da educação

fundamental, estão entre as mais indicadas pela literatura com influenciadores dos resultados educacionais.

A mineração de dados educacionais (MDE) é um campo de pesquisa que busca descobrir padrões ou evidências sobre alunos e formas de aprendizagem. Nos últimos anos diversos trabalhos têm explorados os benefícios que a MDE traz ao ambiente educacional.

Na atualidade as escolas públicas e privadas estão investindo mais na automação do *software*, pelo fato de agilizar os níveis estratégicos para as tomadas de decisões, pois esses *softwares* armazenam os dados em *data warehouse* para gerar relatórios complexos na construção de horários de aulas, serviços de secretaria, controle de estoques, e no acompanhamento do ensino e da aprendizagem. Nos estudos de Baker *et al.* (2011):

A mineração de dados educacionais (EDM) é uma área recente de pesquisa que tem como principal objetivo o desenvolvimento de métodos para explorar conjuntos de dados coletados em ambientes educacionais. Atualmente ela vem se estabelecendo como uma forte e consolidada linha de pesquisa que possui grande potencial para melhorar a qualidade do ensino (BAKER *et al.*, 2011, p.1).

Segundo Webber *et al.* (2013, p.2), “a área denominada Mineração de Dados Educacionais (MDE) é um campo de pesquisa que busca extrair informações novas e úteis com o intuito de desenvolver e fortalecer as teorias cognitivas de ensino-aprendizagem”. A mineração de dados educacionais é uma tarefa importante visto o grande volume de informações produzidas pelos softwares educacionais bem como dos dados coletados pelas avaliações de larga escala. Minerar significa buscar padrões relevantes que possam ser usados para aprimorar os processos de ensino e aprendizagem.

1.2 Motivação

A sociedade atual, marcadamente orientada pela economia baseada no conhecimento e pela tecnologia da informação e comunicação, apresenta intensa dinâmica social, relações e influências globalizadas que, ao mesmo tempo, constituem-se em oportunidades culturais estimulantes e interessantes a todas as pessoas e organizações, assim como desafios e exigências extraordinários. Nesse contexto, a educação se torna imprescindível como ação contínua e permanente, demandando das instituições de ensino, a necessidade de reinventar-se e melhorar suas competências continuamente.

Novos desafios e exigências são apresentados à escola, que recebe a missão de formar cidadãos com capacidade de não só enfrentar esses desafios, mas também de superá-los. Como consequência, para trabalhar em educação, de modo a atender essas demandas, torna-se imprescindível que se conheça a realidade e que se tenha as competências necessárias para

realizar nos contextos educacionais os ajustes e mudanças de acordo com as necessidades e demandas emergentes no contexto da realidade dos ambientes externo e interno da escola. No contexto dessa sociedade, a natureza da educação e as finalidades da escola ganham uma dimensão mais abrangente, complexa e dinâmica e, em consequência, o trabalho daqueles que atuam nesse meio.

Essa premissa subsidia a afirmação de que a criação, o aprimoramento e a evolução das avaliações educacionais externas e as políticas públicas educacionais estão entrelaçadas, uma vez que, os índices resultantes destas avaliações, sejam elas nos parâmetros nacionais, como o IDEB, ou internacionais como o PISA (*Programme for International Student Assessment – Programa Internacional de Avaliação de Alunos*), balizam e orientam as diretrizes governamentais.

1.3 Problemática

O problema de estudo dessa pesquisa pode ser definido através da seguinte pergunta: Como integrar a visão computacional (máquina) com a visão dos gestores educacionais (humanos) para identificar os fatores que possam influenciar positivamente a melhoria do IDEB visando apoiar tomada de decisão dos gestores educacionais?

1.4 Objetivos

1.4.1 Objetivo Geral

Este trabalho tem como objetivo geral desenvolver uma abordagem metodológica que integre a área computacional com a perspectiva dos gestores educacionais, visando identificar os fatores relevantes que possam contribuir para a melhoria do IDEB.

1.4.2 Objetivos Específicos

Para alcançá-lo, foram definidos os seguintes objetivos específicos:

- a) Determinar os algoritmos de mineração de dados educacionais (MDE) que possam ser aplicados ao conjunto de dados educacional;
- b) Definir técnicas de seleção de atributos, nos algoritmos de MDE, que possam reduzir os fatores sem perda de informações;
- c) Identificar os fatores relevantes da perspectiva computacional (Máquina);
- d) Desenvolver um questionário simplificado com auxílio de especialista do domínio para pesquisa *in loco*.
- e) Investigar os fatores relevantes da perspectiva do gestor (Humano);
- f) Determinar os fatores relevantes resultantes das duas perspectivas.

1.5 Relevância da Pesquisa

Segundo INEP (2019), o Brasil gasta 5,8% do PIB em educação pública por ano. Esse índice é superior à média dos países da OCDE, que é de 5,5%, e acima de países como Chile (4,8%) e Estados Unidos (5,4%). Verifica-se assim, que a baixa qualidade da educação não se deve à insuficiência de recursos. Embora tenha havido aumento da oferta de vagas, a qualidade da educação básica brasileira ainda é muito precária, quando comparado internacionalmente.

Em 2015 no PISA, o Brasil ficou na 63^o posição em ciências, na 59^o em leitura e na 66^o colocação em matemática entre 70 países. Em 2018 no PISA, o Brasil ficou na 66^o posição em ciências, na 57^o em leitura e na 70^o colocação em matemática entre 79 países. Ou seja, o modelo de ensino adotado no Brasil se mostra pouco eficaz e ineficiente, comprometendo o desenvolvimento econômico futuro (INEP, 2019).

Diante dessa realidade, percebe-se cada vez mais a necessidade de informações de qualidade por parte dos atores do sistema educacional (gestores, diretores, coordenadores pedagógicos, docentes e discentes), visando à efetividade na tomada de decisões no sentido de melhorar o processo de ensino e aprendizagem nas instituições educacionais (privadas e públicas) do Brasil.

Pensando neste índice, resolveu-se utilizar a mineração de dados educacionais (MDE) para analisar o IDEB das escolas municipais de Maceió-Alagoas e trabalhar com diferentes atributos com que possam na visão dos gestores educacionais fornecer resultados significativos para a melhoria da gestão nas escolas públicas de Alagoas.

Este estudo avalia as escolas da cidade municipal de Maceió. De acordo como o Instituto Brasileiro de Geografia e Estatística (IBGE, 2019), Maceió tem uma população total estimada em pouco mais de 1 milhão de habitantes com uma taxa de escolarização de 95% para crianças de 6 a 14 anos, de acordo com o último censo realizado em 2010. O QEDu (2019) anuncia que o IDEB médio do município de Maceió em 2017 foi de 3,8 enquanto a média nacional foi de 4,3.

A Secretaria Municipal de Educação (SEMED) é um órgão público da prefeitura que formula e garante a política municipal da educação. Dentre seus deveres está o gerenciamento escolar e pesquisa educacional. Para a finalidade desse estudo, informações sobre as escolas municipais foram disponibilizadas pela SEMED.

Segundo o MEC (2018), o IDEB no ano de 2017 para os anos iniciais da rede pública de Maceió cresceu e atingiu a meta, alcançando 5,6 (**Quadro 1**). O IDEB da rede pública de ensino municipal de Maceió cresceu, porém ainda não atingiu a meta 6,0. O município de

Maceió possui 55 instituições de Educação Infantil (sendo 32 creches) e 91 escolas de anos iniciais do Ensino Fundamental e 29 de anos finais do Ensino Fundamental. Essas escolas oferecem o ensino fundamental na modalidade regular e na Educação de Jovens e Adultos, além de Programas Educacionais do tipo: Brasil Alfabetizado, Projovem Urbano, Mais Educação e outros (PME, 2014).

Quadro 1 - Metas para a educação básica de Alagoas.

	IDEB Observado						Metas (MEC)				
	2009	2011	2013	2015	2017	2007	2009	2011	2013	2015	2017
Total	4.6	5.0	5.2	5.5	5.8	3.9	4.2	4.6	4.9	5.2	5.5
Estadual	4.9	5.1	5.4	5.8	6.0	4.0	4.3	4.7	5.0	5.3	5.6
Municipal	4.4	4.7	4.9	5.3	5.6	3.5	3.8	4.2	4.5	4.8	5.1
Privada	6.4	6.5	6.7	6.8	7.1	6.0	6.3	6.6	6.8	7.0	7.2
Pública	4.4	4.7	4.9	5.3	5.5	3.6	4.0	4.4	4.7	5.0	5.2

Fonte: MEC (2018)

O interesse da investigação neste município fundamenta-se principalmente, pelo fato que até o tempo presente ainda existem poucos trabalhos na área. Além de apresentar relevância acadêmica, que possibilita aos docentes e discentes maior familiaridade sobre a qualidade do ensino básico na cidade de Maceió, proporcionando à sociedade em geral, principalmente, aos gestores municipais, professores e famílias, uma noção sobre o desenvolvimento de políticas educacionais no município, gerando subsídios que contribuam para a ampliação da qualidade do ensino público, podendo servir como base de apoio para o monitoramento e direcionamento dos programas e políticas educacionais. Assim disseminando conhecimentos, servindo como apoio para trabalhos futuros no meio acadêmico.

1.6 Estrutura da Dissertação

O presente trabalho está organizado da seguinte forma: o **Capítulo 1** possui caráter introdutório e apresenta uma contextualização geral sobre o tema, também justificando e definindo a relevância da pesquisa e apresentando seus objetivos.

O **Capítulo 2** aborda a fundamentação teórica bem como os trabalhos relacionados a esta temática, tentando mostrar a originalidade do presente trabalho. Este capítulo apresenta uma revisão de literatura de forma a constituir um referencial teórico para a pesquisa que será realizada.

No **Capítulo 3** é definida a metodologia empregada para que sejam alcançados os objetivos da pesquisa, definindo sua abordagem, os métodos e as etapas de execução.

No **Capítulo 4** é apresentado o modelo de integração abordando a visão computacional (MDE) e a visão humana (gestores educacionais). Apresentam-se, ainda, os resultados e as discussões, a partir da análise dos dados obtidos por meio da aplicação da pesquisa de campo com os gestores educacionais e análise exploratória da base de dados do INEP/MEC por meio das técnicas de MDE.

No **Capítulo 5** de conclusão, são reapresentados os principais resultados de forma resumida, procurando-se demonstrar como estes resultados auxiliaram a alcançar os objetivos definidos para esta pesquisa, as limitações e propostas de pesquisas futuras.

2 FUNDAMENTAÇÃO TEÓRICA E REVISÃO DA LITERATURA

2.1 Gestão da Educação Pós-LDB

Para melhor regulamentar a educação no Brasil, foi instituída a Lei de Diretrizes e Bases (LDB), que se configurou ao longo da promulgação de diversas leis, dentre as quais se destaca, a Lei nº 9.394, de 20 de dezembro de 1996, onde se encontram as diretrizes para o ensino em todos os níveis no Brasil. É a mais importante lei brasileira relacionada à educação, sua composição compreende 92 artigos, que apresentam as diretrizes para todos os níveis de educação, do ensino infantil ao ensino superior. Entre outras características, a LDB apresenta: o direito de acesso gratuito ao ensino fundamental; as funções dos governos, em todas as instâncias, no que se refere a área de educação; as obrigações das instituições de ensino; a carga horária dos níveis de ensino; diretrizes curriculares e os deveres e competências dos profissionais da educação.

Com a publicação da LDB de 1996 a concepção de descentralização dos serviços educacionais foi fortalecida. A gestão da educação brasileira se caracteriza como uma ação compartilhada pelos três entes federados – União, Estados e Municípios – prevista no artigo 211 da Constituição Federal de 1988. As competências e obrigações da União em relação ao sistema educacional brasileiro estão expressas nos artigos 8º e 9º da LDB; as do Estado estão estabelecidas no artigo 10, enquanto as obrigações do Município com a educação estão explicitadas no artigo 11 da mesma lei.

No sistema de delegação de competências para oferta de educação pública, cabe ao Município a responsabilidade exclusiva pela oferta de educação infantil, sendo o ensino fundamental responsabilidade compartilhada entre Município e Estado e o ensino médio de responsabilidade do Estado.

A implantação do Fundo de Manutenção e Desenvolvimento do Ensino Fundamental e de Valorização do Magistério (Fundef), em 1998, criou um ambiente favorável a que os municípios assumissem a oferta integral do ensino fundamental, uma vez que a grande maioria deles, especialmente os do Nordeste, tem uma expressiva parcela de suas receitas dependente de repasses constitucionais ou transferências legais. Ao assumir a oferta do ensino fundamental o município aumentaria sua receita com o recebimento de recursos provenientes do Fundef.

Assim, a municipalização do ensino fundamental foi efetivada, independentemente das condições de ofertados serviços educacionais. A motivação financeira propiciada pelo Fundef criou um clima político que arrastou consigo toda e qualquer possibilidade de crítica mais

efetiva sobre as deficiências de natureza técnica e pedagógica das equipes municipais. Naquele momento, qualidade da educação foi tema pouco presente no processo de expansão.

No entanto, desde os anos 1990, aspectos relacionados à eficácia da gestão educacional no âmbito municipal têm sido abordados pela literatura em política educacional no Brasil. Preocupações relativas à eficiência dos serviços educacionais, a qualidade dos gastos com educação e indicadores de desempenho escolar não ocupam o foco das pesquisas de então.

Com a questão da municipalização do ensino fundamental posta, os pesquisadores mostram interesse pelo tema, e resultados de investigações revelam que a contribuição da esfera local, explicitada por meio da gestão educacional no município, tem sido importante marco de inovações que contribuem para a melhoria da qualidade da educação.

Já é amplamente reconhecido que a qualidade da educação se assenta sobre a competência de seus profissionais em oferecer para seus alunos e a sociedade em geral experiências educacionais formativas e capazes de promover o desenvolvimento de conhecimentos, habilidades e atitudes necessárias ao enfrentamento dos desafios vivenciados em um mundo globalizado, tecnológico, orientado por um acervo cada vez maior e mais complexo de informações e por uma busca de qualidade em todas as áreas de atuação.

Somente a partir dos anos 1990 o Brasil passa a se preocupar com a criação de uma política nacional de avaliação, implantando o Sistema de Avaliação da Educação Básica (SAEB), o Exame Nacional do Ensino Médio (Enem) e o Exame Nacional de Curso (ENC), conhecido como “Provão”, para o ensino superior.

Esses instrumentos permitiram a construção de uma valiosa base de dados para a elaboração de diagnósticos precisos e detalhados sobre problemas relativos ao desempenho de alunos brasileiros e sobre a política educacional em geral. Os resultados mostraram que há muito a ser feito para que o Brasil possa alinhar-se aos países que têm obtido êxito na promoção do sucesso escolar de seus alunos.

Uma das consequências da divulgação dos resultados das avaliações foi o aumento do processo de responsabilização pelos resultados, que ora elege os professores ora os gestores, ou ambos pelo sucesso ou fracasso dos alunos e a criação de mecanismos de incentivos simbólicos ou monetários que vêm se acentuando cada vez mais na elaboração de programas e planos governamentais.

Importante destacar que no primeiro momento de criação dos sistemas de avaliação de larga escala, os resultados obtidos e divulgados não encontraram eco nos municípios, muito menos nas escolas, uma vez que eles não se viam nos resultados. É a partir dessa constatação

que em 2005, o SAEB vai passar por uma significativa reestruturação, permitindo a implantação de forma definitiva de uma política de *accountability* no país.

O termo *accountability* educacional é proveniente da língua inglesa e vem sendo traduzido como prestação de contas e responsabilização. Neste caso, a prestação de contas se refere ao processo pelo qual os sujeitos informam e/ou exigem informações acerca do uso dos recursos – financeiros, humanos, materiais, etc. – para a obtenção de um determinado objetivo. A responsabilização está intimamente associada à prestação de contas, uma vez que ser responsável por um processo ou um resultado implica – sobretudo no setor público em uma sociedade democrática – em ter que prestar contas dos resultados e do uso dos recursos aplicados.

Para se implantar um sistema de *accountability* educacional são necessários, pelo menos, quatro elementos: informações, padrões, autoridade e consequências, sendo o último o elemento mais importante e complexo do sistema, uma vez que o objetivo principal é melhorar a educação. A política educacional brasileira mostra que se avançou muito nos últimos anos no que diz respeito a esses quatro elementos e que vem se implantando, de forma gradativa e sutil, um sistema de *accountability*.

A experiência do SAEB foi extremamente positiva e serviu para revelar os sérios problemas de qualidade enfrentados pela educação pública brasileira. Embora a metodologia de aplicação obedecesse a todos os padrões internacionais para o tipo de avaliação de sistema que este representava, o processo de publicização dos resultados não tinha impacto significativo junto às redes de ensino e escolas.

A reformulação do SAEB com a aplicação da Prova Brasil permite a discriminação dos resultados por escola e faz com que se possa dispor de um conjunto expressivo de informações a partir de padrões. As consequências podem ser observadas pelos programas, projetos e benefícios que o próprio MEC cria, selecionando as prioridades de atendimento a partir dos resultados de desempenho alcançados pelas escolas e redes de ensino no SAEB.

2.2 Avaliação da Educação Básica

Os avanços na criação de uma política de avaliação da educação brasileira, fez com que, na década 2000, o interesse pela temática da educação municipal passasse a incorporar outros indicadores quantitativos provenientes de avaliação de larga escala, especialmente os que se referem a desempenho dos alunos, e princípios da administração pública como eficiência, eficácia e transparência.

Com a universalização do acesso, começaram debates em torno da qualidade da educação e a avaliação passou a ser um importante elemento, tendo se expandido em todo o Brasil. Essa mobilização em torno das avaliações motivou o desenvolvimento de indicadores que pudessem dar conta de analisar essa qualidade.

Em 1990, o Ministério da Educação criou o Sistema de Avaliação da Educação Básica (SAEB). O SAEB é uma avaliação em larga escala que acontece bienalmente no Brasil e objetiva principalmente diagnosticar a situação da educação brasileira a fim de buscar alternativas para melhorá-la e sanar deficiências em todas as esferas – federal, estadual e municipal (SAEB, 2019).

Na primeira edição, o SAEB avaliou escolas com 1^a, 3^a, 5^a e 7^a séries do Ensino Fundamental, atendo-se apenas a escolas públicas e localizadas em áreas urbanas. Ao longo dos anos, o SAEB se aperfeiçoou e em 1995 começou a utilizar a Teoria de Resposta ao Item (TRI), ferramenta que permite a comparação de dados ao longo dos tempos (SAEB, 2019).

Segundo INEP (2007), nos anos de 1997 e 1999, os testes foram aplicados a alunos da 4^a e 8^a séries, com questões de Língua Portuguesa, Matemática e Ciências. Já os alunos da 3^a série do Ensino Médio responderam ainda questões de História e Geografia. Na edição seguinte (2001) as avaliações passaram a ter questões apenas de Língua Portuguesa e Matemática e se mantém assim até hoje. Em 2003, as provas foram aplicadas em apenas algumas escolas, tendo um caráter amostral. Em 21 de março de 2005, a Portaria n^o 931 reestruturou o SAEB, que passou a ser composto por duas avaliações: a Avaliação Nacional da Educação Básica (ANEB) e a Avaliação Nacional do Rendimento Escolar (ANRESC), popularmente conhecida como Prova Brasil. Dois anos mais tarde, o MEC criou o IDEB, que objetiva primordialmente aferir o nível das redes de ensino e das escolas em particular.

De acordo com INEP (2017), o Brasil tem quase 50 milhões de alunos matriculados em pouco mais de 180 mil escolas particulares e públicas, da Educação Infantil ao Ensino Médio, incluindo EJA (Ensino de Jovens e Adultos) e Educação Especial. Desse total de educandos, 80% estão em escolas públicas. Os índices educacionais de hoje são alarmantes e apontam que na rede pública de ensino somente 30% dos alunos aprendem o adequado na competência de leitura e interpretação de textos até o 9^o ano. Já em Matemática, apenas 14% deles aprendem o adequado em resolução de problemas até o último ano do Ensino Fundamental.

No Brasil, nas últimas décadas, a Educação Básica vem sendo monitorada por meio de índices que avaliam o desempenho dos alunos em testes padronizados, utilizando também as taxas de aprovação e reprovação da escola.

Um estudo publicado pela McKinsey e Company (2010) mostra que na busca pela melhoria do desempenho dos alunos, as escolas focalizam suas iniciativas em três vertentes – estrutura, recursos e processos. Embora as mudanças na estrutura e nos recursos sejam os mais visíveis socialmente, são nos processos que os sistemas aportam maiores recursos, aperfeiçoando a forma de aprendizagem, muito mais do que modificando o conteúdo ensinado.

Outro aspecto que o estudo chama a atenção é na combinação entre “exigência x persuasão” no que tange ao cumprimento das reformas propostas, sendo que, na maioria das situações, prevalece a primeira em detrimento da segunda. O estudo constata ainda que seis intervenções são comuns a todos os estágios de desempenho ao longo da jornada de melhoria: capacitação dos professores (em termos de habilidades de ensino) e de gestão dos diretores; avaliação dos alunos; aperfeiçoamento dos sistemas de dados; facilitação das melhorias mediante a introdução de documentos de política e leis sobre educação; revisão de normas e currículos; e garantia de uma estrutura adequada de recompensa e remuneração para professores e diretores.

2.3 IDEB

Em 2007, o Ministério da Educação cria o Índice de Desenvolvimento da Educação Básica, um indicador de qualidade educacional que combina informações de desempenho em exames padronizados (Prova Brasil ou SAEB) – obtido pelos alunos ao final das etapas de ensino (5ª e 9ª séries do ensino fundamental e 3ª série do ensino médio) – com informações sobre rendimento escolar (FERNANDES, 2007).

IDEB reúne em um só indicador, dois importantes conceitos na qualidade da educação: fluxo escolar e as médias de desempenho nas avaliações. Esse indicador é calculado a partir dos dados sobre aprovação escolar, obtido no Censo Escolar, e médias de desempenho nas avaliações do INEP, o SAEB – para as unidades da federação e para o país, e a Prova Brasil – para os municípios (INEP, 2007).

Em nível nacional, o IDEB, criado pelo Decreto nº 6094 de 24 de abril de 2007, foi implantado para acompanhar e monitorar a qualidade das escolas de ensino fundamental, focando atualmente no 5º e 9º ano. Tal índice combina o desempenho dos alunos nas disciplinas de Português e Matemática e o fluxo escolar.

A criação do IDEB representa uma iniciativa precursora no Brasil, pois a partir dele passou a ser possível acompanhar a qualidade da educação sob dois aspectos igualmente importantes: o rendimento escolar, traduzido pelas taxas de aprovação, reprovação e abandono,

e a aprendizagem, captada por meio das médias de desempenho na Prova Brasil e no SAEB (INEP, 2007).

De acordo com INEP (2007), o IDEB procura ampliar as possibilidades de mobilização da sociedade em prol da educação, pois o indicador é comparável nacionalmente por meio dos resultados que reúnem aprendizagem e fluxo. Trata-se de uma política pública que busca a qualidade da educação, ou seja, uma ferramenta para o acompanhamento das metas de qualidade do Plano de Desenvolvimento da Educação que estabelece como meta que, em 2021, o IDEB do Brasil será de 6,0 (Meta média comparável a um sistema educacional de qualidade dos países desenvolvidos).

O IDEB, considerado mais do que simplesmente um indicador estatístico, é um condutor de políticas para a melhoria da qualidade da educação das escolas, em nível nacional, estadual e municipal. Ele permite o estabelecimento de metas individuais intermediárias para as escolas, possibilitando a busca pela qualidade do ensino. Tais metas são os percursos estabelecidos de evolução individual dos índices, para que o Brasil atinja o patamar educacional dos países da Organização para a Cooperação e Desenvolvimento Econômico (OCDE).

A ideia central é que cada escola melhore seus indicadores, contribuindo para que o Brasil chegue à meta 6,0 em 2022. Mesmo para aqueles estabelecimentos de ensino que já atingiram a meta, está prevista a continuidade na melhoria. E para aquelas redes de ensino e escolas que apresentarem problemas está previsto apoio específico a fim de auxiliá-las em seu desempenho.

O IDEB varia de zero até 10 pontos em uma escala de qualidade, e quanto maior a nota melhor o desempenho dos alunos e maior a regularidade no fluxo escolar. Neste caso, o resultado da Prova Brasil, que é apresentado em pontos em uma escala de proficiência (Escala SAEB), é posicionado e distribuído em quatro níveis: Insuficiente, Básico, Proficiente e Avançado.

O IDEB apresenta, segundo Ronca (2013), alguns pontos positivos, como a articulação entre os conceitos de fluxo e de desempenho, favorecendo o entendimento da qualidade das escolas e dos sistemas, auxiliando os gestores e os professores na identificação de competências e dificuldades dos alunos nas disciplinas avaliadas – Português e Matemática.

Não há dúvidas sobre a importância de um indicador como o IDEB para monitorar a qualidade da educação, no entanto a sua simples existência não assegura nenhum tipo de utilização, muito menos compromisso de melhorias. Os países que conseguiram melhorias substantivas nos seus sistemas educacionais o fizeram através da criação de uma política de *accountability* educacional consistente e permanente.

A criação do IDEB, associada a um conjunto de evidências relacionado a mecanismos de controle interno e social dos recursos da educação, aponta na direção da implantação de uma política de *accountability* educacional no país, embora os documentos oficiais pouco se pronunciem sobre o assunto.

Brooke *et al.* (2011) elencaram sete tipos de usos de resultados do IDEB como instrumento da gestão educacional nos estados e municípios, quais sejam: avaliação de programas, monitoramento, planejamento e pesquisa; comunicar as escolas sobre o aprendizado dos alunos e definir estratégias de melhorias; prestar contas à sociedade; alocar recursos de forma eficiente e eficaz; como componente da política de avaliação docente e para certificação de alunos e escolas.

O IDEB auxilia também no diagnóstico, no monitoramento e no agrupamento de informações para a criação de políticas públicas, estabelecendo metas individualizadas para cada escola, sistema de ensino, municípios e estados brasileiros. Tal situação tem o objetivo, entre outros, de controlar e elaborar ações para combater a evasão escolar. No entanto, o IDEB também apresenta várias limitações. De fato, não é possível assumir o IDEB como dispositivo de compreensão da qualidade da educação básica. O IDEB utilizando somente dois fatores, o fluxo escolar e o desempenho dos alunos, associados diretamente com a questão da leitura e da matemática, não são capazes de captar a complexidade da escola como um todo.

2.4 Mineração de Dados Educacionais

Pensando no IDEB, resolveu-se utilizar ferramentas de mineração de dados educacionais, para analisar os fatores que impactam o IDEB das escolas de Maceió e trabalhar com diferentes atributos com a utilização de algoritmos que podem fornecer resultados significativos para a melhoria da gestão nas escolas públicas de Maceió.

Técnicas de mineração de dados já vêm sendo utilizadas com certa frequência e apresentando resultados satisfatórios na busca de soluções para essa problemática tão presente no meio educacional. Através de meios computacionais identifica-se não apenas padrões de comportamento similares como prever se determinado aluno se encaixa ou não neste perfil.

Assim, utilizando destes artifícios busca-se não apenas identificar possíveis padrões e fatores que levaram o aprimoramento dos aspectos do ensino e aprendizagem, como também auxiliar a gestão escolar com um plano de ações que visam minimizar a evasão e a reprovação escolar.

Segundo Costa *et al.* (2012), Mineração de Dados Educacionais (MDE) procura desenvolver ou adaptar métodos e algoritmos de mineração existentes de tal modo que se

prestem a compreender melhor os dados em contextos educacionais, produzidos principalmente por alunos e professores, considerando os ambientes nos quais eles interagem, tais como: AVAs, Sistemas Tutores Inteligentes (STIs), entre outros. Com tais métodos visam-se, por exemplo, entender melhor o aluno no seu processo de aprendizagem, analisando-se sua interação como ambiente. Assim há a necessidade, por exemplo, de adequação dos algoritmos de mineração de dados existentes para lidar com as especificidades inerentes aos dados educacionais, tais como a não independência estatística e a hierarquia dos dados.

De acordo Costa *et al.* (2012), apesar de algumas iniciativas primeiras com *workshops* específicos dentro das conferências sobre *Artificial Intelligence in Education* (AIED) e sobre *Intelligent Tutoring Systems* (ITS), foi somente em 2005, em Pittsburgh, EUA, que foi organizado o primeiro *Workshop on Educational Data Mining*, como parte do *20th National Conference on Artificial Intelligence* (AAAI 2005). Daí em diante, houve mais algumas realizações deste workshop entre 2006 e 2007. Seguindo-se, em 2008 lança-se, em Montreal, Canadá, a primeira conferência em MDE “*First International Conference on Educational Data Mining*”, evento este que se estabeleceu e ganhou regularidade de realização anual, estando agora em 2020 na sua décima oitava edição. Em 2009, essa sociedade investiu na criação de um periódico e publicou o seu primeiro volume JMDE – *Journal of Educational Data Mining*. Em 2011 constituiu-se a sociedade científica para MDE (*International Educational Data Mining Society*).

Quanto ao Brasil, de acordo com Baker *et al.* (2011), os primeiros trabalhos publicados ocorreram por volta de 2006 e mesmo assim, até hoje, o número de publicações nos principais periódicos é mínimo.

Segundo Costa *et al.* (2012), há diversas tarefas envolvidas em MDE, notadamente as que decorrem diretamente da análise de dados gerados nas interações dos alunos com os ambientes de aprendizagem. Dessa análise surgem demandas para responder questões relacionadas como melhorar a aprendizagem do aluno, como desenvolver ambientes educacionais mais eficazes que contribuam efetivamente para os alunos aprenderem mais e em menos tempo? Em outra perspectiva, pretende-se saber quais métodos de mineração de se adequam às necessidades presentes na área de MDE? Quais ajustes devem ser feitos nas técnicas de forma a suprir a necessidade de MDE? Do ponto de vista computacional, alguns desafios práticos que se apresentam em vários contextos educacionais estão relacionados, por exemplo, a falta de padronização dos dados, o que acaba exigindo grande esforço de pré-processamento.

A partir dos métodos tradicionais de análise de dados, a mineração de dados abriu oportunidades excitantes para explorar novos tipos de dados e para analisar antigos tipos de dados de maneiras diferentes. Assim, a mineração de dados educacionais surge com o recente aumento dos cursos à distância e do suporte computacional aos cursos presenciais.

Os pesquisadores da área da educação descobriram que os métodos de mineração de dados precisam ser constantemente modificados para atender às particularidades dos dados educacionais (BAKER, 2010; ROMERO e VENTURA, 2013):

- Há uma variedade de contextos educacionais, onde os dados podem, por exemplo, ser organizados em termos de estrutura do material de aprendizagem (habilidades, problemas, unidades, aulas) e da estrutura de contexto de aprendizagem (alunos, professores, pares de colaboração, classes e escolas) (COSTA *et al.*, 2012);
- A variedade de níveis hierárquicos dos dados sugere que a informação pode estar entre vários níveis;
- Existe uma constante falta de independência estatística entre os dados educacionais, provavelmente pela variedade de níveis hierárquicos. A estatística aponta que dados não independentes devem ser analisados em conjunto (VICINI e SOUZA, 2005). Dizer que um dado é dependente significa que ele pode ser usado para inferir informação sobre outro. A estatística também aponta que a falta de independência pode ocasionar interpretações equivocadas nas análises de inferência estatística.

O principal objetivo da MDE é antigo e estudado há muito tempo, que é conseguir compreender como ocorre o processo da aprendizagem. A diferença é que agora os pesquisadores possuem uma grande escala de dados e conseguem analisar uma aprendizagem prática e real (sem a antiga necessidade de realizar experimentos para obter dados). Sendo possível, por exemplo, observar alunos em um curso durante oito meses e descobrir quais atividades geram melhor aprendizado a longo prazo, verificar o desempenho do aluno em sala de aula de acordo com o tempo que ele inicia a tarefa de casa ou se revisões do conteúdo visto em sala trazem algum benefício ou não (ROMERO *et al.*, 2011).

A MDE pode ser aplicada para resolver muitas tarefas (ROMERO e VENTURA, 2010 e 2013), dentre as quais destacam-se:

- Criação de alertas. Os alertas podem ser sobre comportamentos indesejáveis, como baixa motivação, uso indevido dos recursos, probabilidade de evasão, etc. As técnicas mais frequentes são mineração de processo e análise exploratória dos dados, através de análise estatística e visualizações/relatórios.

- Manutenção e melhoria dos cursos. Envolve tarefas de pesquisa científica, construção de material didático, planejamento e programação, onde se deve analisar como ocorre a aprendizagem do aluno, verificando, por exemplo, o que foi e como foi utilizado. Sendo importante a realização de testes de teorias sobre a aprendizagem baseada em novas tecnologias, para a formulação de novas hipóteses científicas, que poderão apoiar a construção ou adequação dos materiais didáticos, o planejamento de futuros cursos, quais disciplinas um aluno deve cursar em um período, alocação de recursos, etc. As técnicas mais utilizadas são associação, agrupamento e classificação.
- Geração de recomendação. Verifica as necessidades do aluno em um dado momento, gerando uma recomendação, que pode lhe proporcionar aprofundamento em um determinado domínio ou auxiliá-lo em uma dúvida. As recomendações podem ser links para visitas, dicas ou a resolução para um problema, a indicação de algum recurso ou material, um curso que pode ser feito, etc. A maioria das técnicas para esta tarefa envolvem associação, sequenciamento, classificação e agrupamento.
- Previsão de notas e resultados de aprendizagem. Essa tarefa é considerada como a mais antiga e popular da MDE. Consiste em utilizar os dados de atividades do curso para prever as notas finais do aluno ou algum outro tipo de resultado de aprendizagem, como uma possível evasão ou futura capacidade de aprender algo. As técnicas utilizadas podem ser regressão linear, classificação, agrupamento e associação.
- Criação de perfis de alunos. Utilizada para detectar o estado e as características dos alunos, como a satisfação, motivação, progresso de aprendizagem, estilo de aprendizagem, preferências e assim por diante. Para a criação desses perfis, também é considerado alguns problemas que possam impactar negativamente nos resultados de aprendizagem, como a apresentação de muitos erros na realização de uma tarefa, a má utilização ou subutilização de tutores inteligentes, manipulação dos sistemas, exploração dos recursos de forma ineficiente, etc. As técnicas mais frequentes além de agrupamento, classificação e análise de associação, são análises estatísticas, redes Bayesianas, modelos psicométricos e aprendizado por reforço.
- Análise da estrutura de domínio. Realizada para determinar a qualidade do conteúdo apresentado e a sequência em que ele foi dado, através da previsão do desempenho dos alunos, descrevendo o domínio de instrução em termos de conceitos, habilidades, itens de aprendizagem e suas inter-relações. As técnicas mais utilizadas são regras de associação, agrupamento e algoritmos *spacesearching*.

Em algumas aplicações, as tarefas apresentadas poderão estar relacionadas entre si, como: (i) geração de recomendação de acordo com a criação do perfil do aluno, (ii) previsão de

notas e a criação de alertas, com as possíveis evasões, (iii) manutenção e melhoria dos cursos, através dos perfis dos alunos, da previsão de notas e da estrutura de domínio; etc. Essas possibilidades de relacionamento, também justificam a segregação da mineração de dados como MDE e caracterizam o quão complexa e desafiante essa área de pesquisa.

2.5 Seleção de Atributos

A seleção de atributos pode ser entendida como a tentativa de selecionar o menor subconjunto de atributos de forma que a precisão dos classificadores não seja significativamente reduzida (DASH e LIU, 2003). Em muitos bancos de dados existem atributos que são irrelevantes ou redundantes, sendo assim, faz-se necessário o uso de técnicas de seleção de atributos. Dentre os benefícios destas técnicas está à facilitação da visualização de dados, a compreensão dos mesmos, a redução dos requisitos computacionais para a análise dos dados e a redução de tempo de treinamento e resposta dos algoritmos aplicados. Além disso, a redução de atributos também pode aumentar a precisão dos algoritmos de predição (GUYON e ELISSEEFF, 2003).

Técnicas de seleção de atributos, também conhecidas como seleção de *features*, podem ser aplicadas de várias formas. Os principais métodos serão utilizados neste trabalho a fim de reduzir o modelo completo de questões. Tais métodos são descritos a seguir.

2.5.1 Abordagem Filtro

A abordagem de seleção de atributos filtro introduz um processo separado, o qual ocorre antes da aplicação do algoritmo de aprendizado propriamente dito. A ideia é filtrar atributos irrelevantes, segundo algum critério, antes do aprendizado ocorrer. Esse passo de pré-processamento considera características gerais do conjunto de dados para selecionar alguns atributos e excluir outros.

Dessa forma, métodos de filtros são independentes do algoritmo de aprendizado que, simplesmente, receberá como entrada o conjunto de exemplos descrito usando somente o subconjunto de atributos importantes fornecido pelo filtro. Métodos de filtro analisam características intrínsecas dos dados, tais como: informação, consistência, dependência, entre outros (DASH e LIU, 2003). Porém, esses métodos são suscetíveis a seleção de atributos redundantes. Exemplos de algoritmos de filtragem são:

- Teste qui-quadrado. Esse método avalia os atributos individualmente usando a medida χ^2 com relação à classe. Quanto maior o valor de χ^2 , mais provável é a correlação das variáveis (atributo e classe).

- Ganho de informação. Esse método ranqueia os atributos através do ganho de informação. Nesse processo o algoritmo realiza a redução da entropia, porém tem uma desvantagem esse método se torna impraticável quando o número de atributos é muito grande.
- Seleção baseada em correlação (CSF). É um método em que um subconjunto de atributos é selecionado através da alta relação com a classe e no qual um atributo individual é verificado também em relação aos demais atributos.

2.5.2 Abordagem Wrapper (Embrulhamento)

Métodos de embrulhamento (do inglês, *wrapper methods*) consideram a seleção como um problema de busca, no qual avaliam o modelo preparando vários subconjuntos de atributos, comparando os resultados das combinações. Exemplos de métodos dessa abordagem são: Wrapper com o Naive Bayes, *Forward Backward Selection*.

Essa abordagem também ocorre externamente ao algoritmo básico de aprendizado, porém usando tal algoritmo como uma caixa preta para analisar, a cada iteração, o subconjunto de atributos em questão. Ou seja, métodos de embrulhamento geram um subconjunto candidato de atributos selecionado do conjunto de treinamento, e utilizam a precisão resultante do classificador induzido para avaliar o subconjunto de atributos em questão. Esse processo é repetido para cada subconjunto de atributos até que o critério de parada determinado seja satisfeito.

Um argumento para justificar o uso da abordagem *wrapper* é que o mesmo algoritmo de aprendizado que usará o subconjunto de atributos selecionado deve prover uma estimativa de precisão melhor que um outro algoritmo, o qual pode possuir bias de aprendizado totalmente diferente (KOHAVI e JOHN, 1997). Todavia, a maior desvantagem desse método é o custo computacional o qual resulta da execução do algoritmo de aprendizado para avaliar cada subconjunto de atributos a ser considerado.

2.5.3 Abordagem *Embedded* (Embutida)

Na abordagem embutida a tarefa de seleção de atributos é realizada internamente pelo próprio algoritmo de aprendizado. Ou seja, a seleção de atributos está embutida no algoritmo de aprendizado de máquina.

A abordagem embutida consiste na busca por encontrar o conjunto de atributos ótimo enquanto constrói um modelo preditivo. Seu uso é menos custoso computacionalmente do que os métodos de empacotamento, porém são mais complexos conceitualmente. Além disso,

modificações no algoritmo de classificação podem resultar em uma fraca performance. Um exemplo para esta técnica é o método *RecursiveFeatureElimination*.

2.6 Validação cruzada (Cross-validation)

A validação cruzada (*cross-validation*) é uma técnica amplamente usada em algoritmos de seleção de atributos. Essa técnica consiste na divisão dos dados em subgrupos, onde parte destes subgrupos são destinados a amostra de treinamento do algoritmo de seleção ou predição utilizado, enquanto um subgrupo é usado como amostra de teste na validação cruzada, ou também chamados de amostras de validação, e são utilizados para estimar o erro do algoritmo. Então, a validação cruzada seleciona o(s) grupo(s) de menor risco. Esse método evita o sobreajuste de dados (*overfitting*) devido ao fato dos dados de treinamento serem independentes dos dados de validação (WITTEN *et al.*, 2011).

Com o objetivo de reduzir qualquer viés que esteja presente nos algoritmos aplicados foi usado o método *K-Fold Cross-Validation*. Essa técnica divide aleatoriamente os dados em k subconjuntos de tamanho aproximadamente iguais, onde $k-1$ subconjuntos são usados no treinamento do algoritmo e a estimativa de erro é calculada usando o subconjunto restante que aplicado como amostra de teste. Em seguida, os resultados de cada grupo são comparados entre eles, sendo possível a visualização da performance média dos algoritmos (RODRIGUEZ *et al.*, 2010).

De acordo com Wong (2015), há 4 (quatro) fatores que impactam a precisão do método *K-Fold Cross-Validation*: o número de partições (*folds*); o número de instâncias em cada partição; a distribuição de dados por partição e a quantidade de repetição. Para este trabalho foi utilizado o método com dez grupos, ou seja, *10-fold Cross-validation*, que para encurtamento será chamado de *10-fold*, executado 30 vezes.

2.7 Técnicas de Classificação de Mineração de Dados Educacionais

Entre as tarefas e métodos de mineração de dados educacionais a serem discutidos no presente texto, incluem-se: classificação e regressão, agrupamento de dados, mineração de regras de associação. Outras abordagens, entretanto, serão comentadas.

A tarefa de classificação diz respeito ao processo de encontrar um modelo que descreve e distingue classes de dados ou conceitos. Os modelos são derivados com base nas análises de coleções de dados, denominadas conjuntos de treinamentos, os quais correspondem a objetos de dados para os quais os rótulos de classes são conhecidos. O modelo é usado para prever o rótulo de classe de objetos para os quais o rótulo de classe é desconhecido. Ele associa um item de dado a uma ou várias classes pré-definidas.

Enquanto na classificação, a predição é feita para um atributo classificador que assume valores discretos, em modelos de regressão a variável alvo é contínua, ou seja, associa um item de dado a uma ou mais variáveis de predição de valores reais. Por sua vez, a análise de agrupamento de dados procura associar um item de dado com um ou vários agrupamentos determinados pelos dados, valendo-se principalmente de medidas de similaridades. Já a abordagem de mineração de regras de associação busca encontrar possíveis relações interessantes entre atributos de uma base de dados. Essas abordagens serão discutidas em detalhes na próxima seção deste capítulo, onde se discutem outras abordagens.

Segundo Costa *et al.*(2012), grande parte das técnicas utilizadas na área de MDE são provindas da área de mineração de dados. Entretanto, na maioria das vezes há a necessidade de adaptá-las devido às particularidades existentes em ambientes educacionais e seus dados. As técnicas estão apresentadas conforme sua categorização nas subáreas de MDE, seguindo-se o que consta na taxonomia proposta por (Baker *et. al.*, 2011), tal como segue.

2.7.1 Classificadores Baseados em Regras

Esse tipo de técnica deve expor o conhecimento "escondido" nos dados, mostrando possíveis inconsistências, evitando dessa forma resultados que não podem ser previstos. Este tipo de técnica é eficaz em situações que não é necessário um grande conjunto de regras para descrever os dados (DUCH, 2013).

O classificador baseado em regra, também conhecido como método separar e conquistar é um processo iterativo no qual são criadas regras do tipo *SE-ENTÃO (IF-THEN)* que definem um subconjunto de amostras de treinamento. Após essas regras serem definidas, o classificador é limpo e então uma nova interação ocorre preenchendo o classificador com outro subconjunto que também irá gerar novas regras até não haver mais subconjuntos (MAHAJAN e GANPATI, 2014).

No modelo de regras *SE-ENTÃO* gerados, as condições *SE* são conhecidas como regras antecedentes ou pré-condições e as regras *ENTÃO* são chamadas de regras consequentes.

2.7.2 Árvores de Decisão

A árvore de decisão é um modelo estatístico que utiliza treinamento supervisionado para construir uma estrutura de árvore, onde cada nó interno pode ser entendido como um teste de atributo e as folhas da árvore representam um determinado rótulo. A árvore irá classificar a instância de acordo com o caminho percorrido na árvore que termina em um nó folha, atribuindo a instância o rótulo do mesmo (COSTA *et al.*, 2012).

Classificadores de árvores de decisão quebram problemas com um conjunto de regras complexas em um modelo multicamada (estrutura de árvore) que utiliza a união de regras mais simples (SAFAVIAN e LANDGREBE, 1991). Os algoritmos mais populares de árvore de decisão são: C 4.5, C 5.0 e CART.

O algoritmo C 4.5 tem como entrada um conjunto de treinamento de dados já rotulados e gera um modelo, a partir de um conjunto de dados semelhante, porém não rotulado (conjunto de teste), na forma de uma árvore de decisão, usando o conceito de entropia. A cada nó da árvore é selecionado um atributo dos dados que divide o conjunto de amostras de maneira mais efetiva em subconjuntos destinados a uma classe ou outra. O critério usado para realizar a divisão é o ganho de informação que resulta na escolha do atributo divisor (COSTA *et al.*, 2012).

Uma das vantagens de modelos baseados em árvore é a fácil visualização e compreensão de dados. Muitas ferramentas, como Weka, que é utilizado neste trabalho, possuem a funcionalidade de gerar diagramas visuais de árvores geradas pelos algoritmos.

2.7.3 Redes Bayesianas

Redes Bayesianas são classificadores estatísticos que podem prever a relação entre classes. Essa rede produz um modelo casual das relações, o qual é usado para a aprendizagem e para a classificação (HAN e KAMBER, 2001).

O modelo construído na Rede Bayesiana pode ser descrito como um grafo direcionado acíclico que permite a representação eficiente da distribuição de probabilidade de junção, dado um conjunto de variáveis. Cada vértice representa uma variável e cada aresta representa a correlação direta entre as variáveis (FRIEDMAN *et al.*, 1997).

Essas redes são modelos que trabalham com conhecimento incerto ou incompleto através do Teorema de Bayes, criado pelo matemático Thomas Bayes em 1773, que define a probabilidade de um evento dado em um conhecimento prévio. Tal teorema é conhecido por sua famosa fórmula:

$$P(A|B) = P(B|A).P(A) \text{ (Eq. 2.1)}$$

$P(B)$ na qual $P(A)$ significa a probabilidade de A ser verdade e $P(A/B)$ a probabilidade de A ser verdade caso B seja verdadeiro.

2.7.4 Máquina de Vetores de Suporte

Máquina de vetores de suporte (*Support Vector Machine* ou SVM) é um algoritmo supervisionado para classificação que procura separar instâncias de rótulos diferentes usando um hiperplano, o qual é descoberto usando vetores de suporte (COSTA *et al.*, 2012).

A utilização de SVM tem se tornado popular, pois tem demonstrado bons resultados quando comparado a outros algoritmos de aprendizagem em diversas áreas, tais como, a detecção de faces em imagens e a categorização de textos. Lorena e Carvalho (2002) descrevem as principais vantagens da SVM: boa capacidade de generalização; robustez diante de objetos com grandes dimensões, como imagens; convexidade da função objetivo, pois contém apenas um mínimo local para funções quadráticas e teoria bem definida dentro da matemática e estatística.

Todavia, Zeng *et al.* (2008) apresentam um desafio para as SVM que é o espaço em memória para problemas quadráticos com um grande conjunto de dados de treinamento, pois este espaço de memória aumenta em $O(n^2)$, onde n é o número de dados de treinamento.

2.7.5 Agrupamento

O agrupamento é uma técnica que consiste em dividir o conjunto de dados em grupos, de maneira que os objetivos contidos nos dados fiquem agrupados naturalmente segundo a semelhança entre eles. Os algoritmos de agrupamento são técnicas de aprendizado não-supervisionado.

Os algoritmos de agrupamento podem tanto começar sem nenhuma hipótese a priori sobre os grupos de dados (tal como o algoritmo *k-means* com reinício aleatório), ou começar a partir de uma hipótese específica, gerada possivelmente por pesquisa anterior em outros dados. Um algoritmo de agrupamento pode gerar grupos do tipo hard, no qual cada elemento pertence a somente um grupo (p.e., algoritmo *k-means*) ou do tipo soft (p.e., algoritmo *fuzzy*), no qual um elemento pode pertencer a mais de um grupo (p.e., modelos de Mistura Gaussiana).

2.7.6 Regressão Linear

Regressão linear é uma técnica de predição que envolve uma variável preditiva y e uma única variável preditora x , onde y é modelada em função linear de x (HAN e KAMBER, 2000), segundo a equação abaixo,

$$y = b + wx \quad (\text{Eq. 2.2})$$

na qual, assume-se que a variância de y é constante e b e w são coeficientes de regressão. Esses coeficientes podem ser resolvidos pelo método dos mínimos quadrados, que estima qual reta melhor representa os dados, como sendo aquela que minimiza o erro entre os dados atuais e a estimativa da reta.

2.8 Trabalhos Correlatos

Essa seção abordará os principais trabalhos da literatura relacionados no âmbito da MDE para educação e fatores que contribuem para melhoria do ensino, os quais conceberam a base para essa abordagem e produção desta pesquisa de mestrado. Foram destacados 8 (oito) trabalhos de natureza acadêmicas na perspectiva da educação, avaliação externa e IDEB e 3 (três) artigos de relevância sob as perspectivas das técnicas computacionais de *Machine Learning* com dados em educação.

Vidal e Vieira (2011) analisaram o comportamento do IDEB do ensino fundamental em 10 municípios cearenses, tendo como pressuposto a análise da gestão educacional municipal e sua visão sistêmica nas duas etapas do ensino fundamental, fazendo uso de dados quantitativos obtidos em bases de dados nacionais e em dados qualitativos provenientes de pesquisa de campo. Constatou-se que o fortalecimento da dimensão pedagógica criou condições para o aperfeiçoamento da instrução ministrada, no entanto perceberam-se também mudanças no conteúdo ensinado, com a redução do currículo e a supervalorização das matrizes da Prova Brasil. Os resultados mostraram que a gestão educacional municipal deve se concentrar no sistema como um todo (visão sistêmica) e não apenas em uma etapa da educação fundamental.

Chirinéa e Brandão (2014) procuraram discutir a qualidade da educação básica, legitimada pelo Índice de Desenvolvimento da Educação Básica (IDEB). Abordaram conceitos do IDEB explicando que combina duas variáveis, o desempenho dos alunos na avaliação e o fluxo escolar, culminando no estabelecimento de uma nota. Segundo os autores, as argumentações partem da premissa de que o IDEB não leva em consideração outros fatores que incidem sobre a qualidade, tais como: nível socioeconômico e cultural dos alunos; formação docente; valorização do magistério; condições materiais e imateriais de trabalho; gestão escolar; infraestrutura da escola, e insumos. Dessa forma, esse trabalho evidenciou uma reflexão que conduz ao debate sobre a insuficiência de variáveis para estabelecer a qualidade da educação no Brasil por meio do IDEB.

Tosta (2015) apresentou um recorte histórico após a Constituição de 1988 sobre política educacional brasileira, ressaltando que a baixa qualidade do ensino público ainda é um grande obstáculo para que se alcance maior equidade social. Apresentou o IDEB como um dos instrumentos utilizados para que se alcance a melhoria da qualidade da educação. O trabalho de pesquisa consistiu em conhecer a percepção dos professores da rede estadual de Campos dos Goytacazes/RJ sobre o IDEB. A autora utilizou o método da triangulação, aplicando questionários de amostragem probabilística e realizando entrevistas semiestruturadas de amostragem não probabilística por conveniência, em sete escolas estaduais da zona urbana do município. Os dados e as entrevistas mostraram que os professores acreditam que deve haver processos avaliativos da política educacional, porém criticam a metodologia do IDEB, alegando que o instrumento não gera dados fiéis, estimula disputas entre escolas (e professores), contribuindo muito pouco para a melhoria da política educacional.

Neves (2012) analisou o caso de sucesso de uma escola estadual do interior do estado do Rio de Janeiro, que obteve 7.4 e 7.7 no IDEB 2009 e 2011, respectivamente. O trabalho observou com ênfase o modelo de gestão e nas mudanças pedagógicas que fizeram parte desta transformação. A autora fez uso de pesquisa de campo, observação direta, aplicação de entrevistas e análise de documentos a fim de responder as questões de pesquisa. Como conclusão verificou-se que o tipo de gestão, a busca de maior entrosamento com alunos, professores e comunidade e a instituição de colegiados foram os responsáveis pelo êxito alcançado pela escola. Ainda segundo o trabalho estrutura física da escola, mesmo que singelas, contribuíram para a melhoria do sentimento de bem estar do aluno com relação à escola. A bonificação foi outro elemento importante. Os alunos se sentiram motivados a participar, na medida em que sabiam que seriam premiados pelos bons resultados.

Sanchez (2017) apresentou um estudo de levantamento bibliográfico de teses e dissertações que abordaram o índice no período de 2007 a 2015 no Brasil, considerando as contribuições dos estudos que traziam à época elementos para a compreensão de como o IDEB vinha operando na configuração e na implementação de políticas públicas educacionais. Após leitura integral das teses e dissertações, segundo a autora, resultou na identificação de 5 (cinco) maneiras de abordagem ao IDEB, a saber: a) referência ao IDEB como indutor de políticas e programas; b) investigação de fatores para obtenção de alto ou de baixo IDEB por escolas ou redes públicas de ensino; c) referência ao IDEB como indicador de resultado de programas/políticas/gestão; d) utilização de dados do IDEB no delineamento da pesquisa e /ou na análise de dados, cenário em que o índice é usualmente tratado como uma das variáveis

explicativas dos resultados do estudo; e) e menção ao IDEB como um dos critérios de seleção de estados, municípios e /ou escolas a serem pesquisados ou como elemento de caracterização do contexto de realização do estudo.

Crozatti (2019) identificou as variáveis que mais influenciaram o IDEB das redes do ensino fundamental gerido pelos municípios do estado de São Paulo no ano de 2017. A partir de trinta e três indicadores das redes educacionais e das condições econômicas dos municípios, citados pela literatura e disponibilizados por órgãos públicos, o estudo faz uso da regressão múltipla pelos Mínimos Quadrados Ordinários (MQO) com o procedimento *stepwise* como método de análise. O trabalho apresentou as variáveis que mais influenciaram o IDEB: a remuneração dos profissionais da educação, a distorção idade série dos alunos, a existência do conselho municipal de educação e a taxa de aprovação dos alunos nas redes. Segundo o autor esta combinação de variáveis tem o poder de explicar quase 70% da nota do IDEB das redes municipais em foco.

Calixto *et al.* (2017) identificaram as variáveis concernentes à evasão escolar, utilizando os dados do censo educacional de 2014, 2015 e 2016 dos estados de Ceará e Sergipe. A metodologia CRISP-DM foi usada e a separação dos dados foi utilizada a ferramenta SPSS. As análises utilizaram técnicas de indução de Regras e Regressão Logística. Os modelos criados apresentaram acurácia em torno de 87%. Além disso, a idade, etapa de ensino, modalidade de ensino, existência de laboratórios e localização da escola se destacaram como variáveis influentes na evasão escolar.

Júnior *et al.* (2019) aplicaram uma ferramenta de mineração de dados, para analisar o Índice de Desenvolvimento da Educação Básica (IDEB) das escolas públicas do município de Maceió, anos finais, visando auxiliar no processo decisório dos gestores educacionais pela adoção de medidas de melhoria da gestão escolar. Utilizam-se os dados educacionais das escolas públicas de Maceió coletados do portal do INEP, principalmente, das informações dos questionários aplicados aos alunos que realizaram a Prova Brasil. O trabalho utilizou duas técnicas de mineração de dados: regressão linear e árvore de decisão, visando identificar fatores que influenciam no desempenho do IDEB. Os resultados obtidos fortalecem estudos realizados, que apontam diversos fatores como influenciadores ao desempenho do aluno, tais como: a escolaridade dos pais do aluno, o incentivo aos estudos e o compromisso do docente, segundo os autores.

Alves e Soares (2013) analisaram a relação entre o indicador e o contexto escolar, considerando para tal o perfil de seus alunos e as características do estabelecimento de ensino.

A análise empírica utiliza dados da Prova Brasil, do Censo Escolar e do próprio IDEB, por meio de modelos de regressão linear múltipla. Os resultados mostraram que as escolas que atendem a alunos de menor nível socioeconômico têm piores resultados. Para essas escolas, é muito mais difícil elevar o valor do indicador. Além disso, as condições de infraestrutura e de complexidade da instituição também guardam relação com o IDEB. Por fim, os resultados indicaram que são necessárias políticas de superação dessas limitações e que tais condições não podem ser ignoradas na análise do IDEB.

Wahyuddin (2018) examinou as percepções e ações dos formuladores de políticas educacionais em relação ao envolvimento dos pais na educação das crianças em um distrito na província de Sulawesi Ocidental, na Indonésia. O estudo empregou um estudo qualitativo, utilizando uma abordagem fenomenológica. O autor coletou dados de cinco participantes através de uma entrevista semiestruturada, envolvendo o chefe do distrito, o departamento de educação chefe, um subdistrito, um chefe de aldeia e um diretor de escola. Os dados coletados foram analisados aplicando os métodos analíticos de Moustakas (1994) consistindo em três etapas, nomeadamente *bracketing*, criação de grupos de significado e descrição textual. Os resultados mostraram que os formuladores de políticas do distrito tinham uma percepção sobre o incremento do envolvimento dos pais na educação, não apenas para alunos, mas também para escolas, pais e governos locais. No entanto, percepções positivas dos formuladores de políticas educacionais no distrito não eram consistentes com suas ações, comprovadas por nenhuma política específica que pretenda o envolvimento dos pais.

Fraser e Killen (2003) apresentaram as respostas de uma recente investigação empírica na Universidade de Pretória que tentou identificar os fatores pós-matrícula que professores e alunos consideraram ter influências importantes na educação dos alunos sucesso nos estudos da universidade. O estudo investigou as diferentes expectativas dos alunos do primeiro ano, dos estudantes seniores e dos professores e identificaram inúmeras semelhanças e inconsistências importantes. Para os autores as percepções dos alunos do primeiro ano não estavam fortemente correlacionadas com os professores ou alunos seniores. Os resultados da pesquisa podem ser usados de três maneiras. Os gestores universitários poderiam oferecer um aprendizado mais favorável no ambiente para aumentar as chances de sucesso dos alunos e os professores também podem usar as informações para aumentar as influências de fatores positivos na aprendizagem dos alunos. A influência de fatores negativos também pode ser inimiga. Finalmente, os alunos também poderiam ser assistidos e apoiados para abordar estudos universitários de uma maneira que aumente suas chances de sucesso.

Márquez-Vera *et al* (2013) aplicou técnicas de mineração de dados, investindo em seleção de atributos, a um *data set* de 670 alunos do ensino médio de Zacatelas (México) para descrever o insucesso escolar através da identificação de quais alunos poderiam evadir, considerando um modelo preditivo aplicado a uma coleção de atributos selecionados. Com isso, algumas ações preventivas poderiam ser tomadas para evitar a evasão escolar desses alunos.

Ferreira *et al.* (2018) analisaram como o IDEB é percebido pelos gestores de escolas públicas na cidade de Salgueiro-PE. Os dados iniciais foram confrontados frente à realidade cotidiana do ambiente escolar, estabelecendo uma relação entre a percepção dos gestores e os resultados divulgados. Resultados das avaliações externas são vistos como dados produzidos pela hierarquia do sistema educacional como algo alheio ao âmbito escolar, ora são vistos como qualificadores educacionais e condutores de políticas públicas. Os resultados apontaram que a percepção dos gestores e os dados apresentados do IDEB divergem, e que os gestores veem este índice como insuficiente para representar e subsidiar, de forma eficaz, as práticas das instituições educacionais municipal, porém é um importante instrumento de monitoramento e avaliação da rede de ensino.

3 METODOLOGIA

Nesta seção são apresentados os procedimentos metodológicos que viabilizaram o desenvolvimento da pesquisa. A pesquisa desenvolvida teve caráter descritivo (CERVO *et al.*, 2007), uma vez que visou uma investigação sobre um grupo representativo de escolas de ensino básico na rede municipal em todas as 8 (oito) regiões administrativas da capital Maceió.

Há também um enquadramento com a pesquisa de campo, visando a compreensão de diversos fatores e aspectos no que diz respeito a opinião (percepção) dos diretores escolares frente aos fatores potencialmente averiguados. O estudo de caso se concentrou em um grupo contendo 16 escolas pertencentes a rede de educação básica anos iniciais 1º ao 5º. Essas escolas tiveram como critério de seleção pertencer a áreas da Região Administrativa (RA), atendendo esse critério, foram selecionadas 02 (duas) escolas: a que tem o maior IDEB e a com menor IDEB por RA (Região Administrativa). A seguir é apresentado o rol de escolas (**Tabela 1**) que compuseram a amostra.

Tabela 1 - Relação das escolas pesquisadas e regiões administrativas.

Nome Escola	Reg. Adm (RA)
Orlando Araújo	1
José Carrascosa	1
Almeida Leite	2
Élio Lemos (Nosso Lar)	2
Higino Belo	3
Padre Brandão Lima	3
Zilka De Oliveira	4
Marcos Soriano	4
Paulo Freire	5
Audival Amélio	5
Elma Marques Curti	6
Maria De Fátima Lira	6
Denisson L. C. Menezes	7
Pedro Suruagy	7
Marechal Floriano Peixoto	8
Pedro Barbosa Júnior	8

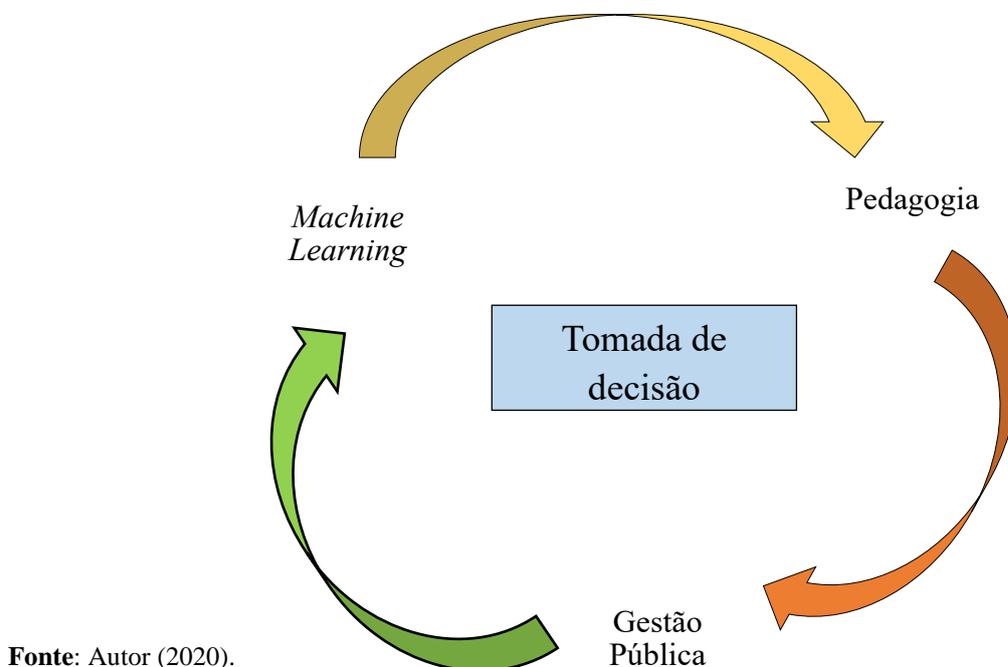
Fonte: Autor (2020).

Admite-se que a presente pesquisa tenha cunho de trabalho aplicado, uma vez que as informações pertinentes à análise de indicadores para a melhoria do IDEB das escolas públicas do Município de Maceió podem contribuir para ajustes nas políticas públicas e melhor o gasto público (APPOLINÁRIO, 2009).

Considerou-se indispensável para formulação deste trabalho a pesquisa bibliográfica específica, para suporte, desenvolvimento, aplicação das ideias e a formulação do instrumento de pesquisa aplicado (questionário).

Este trabalho tem fundamento nos seguintes conhecimentos: (i) *machine leaning* onde foi utilizando métodos de análise de dados que automatizam a construção de modelos analíticos; (ii) pedagogia no tocante a avaliação educacional, a análise dos discursos dos entrevistados, entendimento da área que se concentra os dados; (iii) gestão pública neste trabalho é identificada na dinâmica de ações antecipadas para melhoria de processos. A **Figura 1** embasa metodologicamente este estudo.

Figura 1 - Fluxo esquemático dos fundamentos utilizados na pesquisa.



Para tanto, foram levantados artigos em periódicos indexados através do Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), diversas outras bases de periódicos e livros focados nas temáticas mencionadas, seguindo a noção essencial da pesquisa bibliográfica de levantar dados ou categorias teóricas já trabalhadas por outros autores (SEVERINO, 2008).

A pesquisa teve abordagem quali-quantitativa, uma vez que coletou dados de natureza variada (ora dados qualitativos ora dados quantitativos) através de um questionário, para serem analisados segundo métodos estatísticos específicos (APPOLINÁRIO, 2009).

Para o desenvolvimento da pesquisa de campo foi elaborado um questionário embasado em achados teóricos de artigos sobre aspectos ligados ao IDEB bem como auxílio de um especialista do domínio.

O problema de estudo da pesquisa pode ser definido através da seguinte pergunta: Como integrar a visão computacional (máquina) com a visão dos gestores educacionais (humanos) para identificar os fatores que possam influenciar positivamente a melhoria do IDEB visando apoiar tomada de decisão dos gestores educacionais?

A seguir, serão detalhados o planejamento e os métodos que foram aplicados na pesquisa de campo.

Este estudo fez a análise dos dados educacionais de 82 escolas públicas municipais de Maceió ano iniciais do ensino fundamental (5º ano). O foco da pesquisa é contribuir para o melhoramento da gestão educacional e do processo de ensino e aprendizagem em escolas públicas municipais de ensino fundamental (recorte nos anos iniciais), por meio da utilização da ferramenta de mineração de dados Weka 3.8. Trata-se de um software de aprendizado de máquina de código aberto experimentado e testado que pode ser acessado por meio de uma interface gráfica do usuário, aplicativos de terminal padrão ou uma API Java. É amplamente utilizado para ensino, pesquisa e aplicações industriais, contém uma infinidade de ferramentas internas para tarefas padrão de aprendizado de máquina e, além disso, fornece acesso transparente a caixas de ferramentas conhecidas, como scikit-learn, R e DeepLearning4j (WEKA, 2019).

Pretende-se criar uma abordagem metodológica que integre a área computacional com a perspectiva dos gestores educacionais, visando identificar os fatores relevantes que possam contribuir para a melhoria do IDEB.

Analisando as diversas abordagens dos trabalhos correlatos, torna-se mais evidente que o foco exclusivo nas instituições educacionais municipais de ensino básico ainda não havia sido explorado, nesta perspectiva humano-computacional motivando, desta forma, a elaboração desta pesquisa.

A identificação das variáveis chaves é a entrada para o estágio de transformação dos dados em informações úteis, usando a Mineração de Dados. Conforme mencionado em Bartolomeu (2002, p.10), deve-se reavaliar as variáveis selecionadas anteriormente, baseando-se nos resultados obtidos com as análises estatísticas. Feito isso, será possível identificar as variáveis que se mostram mais relevantes, para que possam ser levadas às próximas etapas para a simulação de cenários favoráveis aos indicadores de qualidade do ensino, expressos nos

descritores do IDEB. Os dados foram obtidos do sítio do INEP/Microdados, os arquivos em formato .CSV, contendo todos os IDEB no âmbito nacional.

Em seguida, foram selecionadas apenas as escolas municipais de Maceió, de ensino fundamental, por meio de aplicação de filtro de dados, utilizando o software Excel, obtendo-se o recorte das 82 escolas que totalizam toda a rede municipal, daquele ano 2015, no escopo dos anos iniciais. O arquivo foi salvo em formato .CSV, formato reconhecido pelo Weka, para tratamento posterior.

As bases de dados usadas nesse estudo foram disponibilizadas abertamente pelo INEP em seu portal. Coletou-se os dados educacionais das 82 escolas selecionadas no portal do INEP. Não houve muito a ser feito em relação à pré-processamento e transformação de dados, pois os dados já haviam sido limpos e validados pelo próprio portal.

Utilizou-se a ferramenta Anaconda (Distribuição) para visualizar os dados e conferir os tipos de dados. Esta etapa teve como objetivo o de construir o conjunto final de dados que foram utilizados nas ferramentas de modelagem. As tarefas de preparação foram realizadas muitas vezes, sem uma ordem pré-determinada. Esta etapa envolve operações como tratar a falta de dados em alguns campos, limpeza de dados, como a verificação de inconsistências, redução da quantidade de campos em cada registro, o preenchimento ou a eliminação de valores nulos, e remoção de dados duplicados. Inicialmente, devido ao fato de o INEP disponibilizar apenas os dados nacionais, foi necessário um filtro para selecionar apenas os diretores de Maceió.

Realizou-se a preparação dos dados na ferramenta Weka onde, para efeito de maior precisão, foram excluídos ruídos da base original (informações que não seriam sensíveis aos algoritmos utilizados nesta pesquisa, como por exemplo: ID. região, ID. caderno de provas e extrato ANEB).

A etapa de seleção de atributos tem como objetivo excluir atributos redundantes e que não são úteis para a criação do modelo de predição. Ao usar a seleção de atributos, busca-se um melhor desempenho e também a simplificação do modelo, reduzindo o custo computacional Tasca (2015). Ao concluir o pré-processamento de dados, aplicou-se à base de dados os algoritmos de seleção de atributos e *Ranking ChiSquared-AttributeEval*, *Filtered-AttributeEval*, *GainRatio-AttributeEval*, *InfoGain-AttributeEval*, *OneRAttributeEval* do Weka.

Neste trabalho, utilizou-se a metodologia do CRISP-DM (CHAPMAN *et al.*, 2000). Essa metodologia define uma sequência não rígida de seis fases, que permite a construção e implementação de um modelo de mineração para ser usado em um ambiente real, auxiliando as decisões de negócio (MORO *et al.*, 2011). Dessa forma, o desenvolvimento desse trabalho seguiu as fases do CRISP-DM (LEE, 2005) descritas a seguir.

1ª Etapa: Compreensão do domínio. Nesta etapa se definiram os objetivos a serem alcançados do projeto e, então, converte-se esse conhecimento numa definição de problema de mineração de dados e um plano de projeto preliminar com a intenção de alcançar os objetivos. Nesta fase, além de definir os objetivos a serem alcançados, determinam-se os critérios de sucesso do projeto. Também serão definidos os recursos necessários à execução de todo o plano de mineração.

Este estudo fez a análise dos dados educacionais de 82 escolas públicas municipais de Maceió ano iniciais do ensino fundamental (5º ano). O foco da pesquisa é contribuir para o melhoramento da gestão educacional e do processo de ensino e aprendizagem em escolas públicas municipais de ensino fundamental (recorte nos anos iniciais), por meio da utilização da ferramenta de mineração de dados Weka 3.8 (WEKA, 2019).

Pretende-se criar uma abordagem metodológica que integre a área computacional com a perspectiva dos gestores educacionais, visando identificar os fatores relevantes que possam contribuir para a melhoria do IDEB. Analisando as diversas abordagens dos trabalhos correlatos, torna-se mais evidente que o foco exclusivo nas instituições educacionais municipais de ensino básico ainda não havia sido explorado, nesta perspectiva humano-computacional motivando, desta forma, a elaboração desta pesquisa.

2ª Etapa: Entendimento dos dados. Os dados podem vir de diversas fontes e ter diversos formatos. Assim, a partir de uma coleção inicial, já livre de possíveis problemas, é necessário descrevê-los, explorá-los e, por fim, verificar sua qualidade. As bases de dados usadas nesse estudo foram disponibilizadas abertamente pelo INEP em seu portal. Coletou-se os dados educacionais das 82 escolas selecionadas no portal do INEP. Os dados são a principal fonte no processo de descoberta de conhecimento em bases de dados, mas normalmente as informações não estão disponíveis de uma forma pronta para mineração.

A identificação das variáveis chaves é a entrada para o estágio de transformação dos dados em informações úteis, usando a Mineração de Dados. Conforme mencionado em Bartolomeu (2002), deve-se reavaliar as variáveis selecionadas anteriormente, baseando-se nos resultados obtidos com as análises estatísticas. Feito isso, será possível identificar as variáveis que se mostram mais relevantes, para que possam ser levadas às próximas etapas para a simulação de cenários favoráveis aos indicadores de qualidade do ensino, expressos nos descritores do IDEB. Foram obtidos do sítio do INEP/Microdados, os arquivos em formato .CSV, contendo todos os IDEB no âmbito nacional. Em seguida, foram selecionadas apenas as escolas municipais de Maceió, de ensino fundamental, por meio de aplicação de filtro de dados,

utilizando o software Excel, obtendo-se o recorte das 82 escolas que totalizam toda a rede municipal, daquele ano 2015, no escopo dos anos iniciais. O arquivo foi salvo em formato .CSV, formato reconhecido pelo Weka, para tratamento posterior.

3ª Etapa: Preparação dos dados. Essa fase tem o objetivo de construir o conjunto final de dados que serão utilizados nas ferramentas de modelagem. As tarefas de preparação podem ser realizadas muitas vezes, e sem uma ordem pré-determinada. Essa etapa envolve operações como tratar a falta de dados em alguns campos, limpeza de dados como a verificação de inconsistências, redução da quantidade de campos em cada registro, o preenchimento ou a eliminação de valores nulos, remoção de dados duplicados. Realizou-se a preparação dos dados na ferramenta Weka onde, para efeito de maior precisão, foram excluídos ruídos da base original (informações que não seriam sensíveis aos algoritmos utilizados nesta pesquisa, como por exemplo: ID. região, ID. caderno de provas e extrato ANEB).

4ª Etapa: Modelagem. Nessa etapa, várias técnicas de modelagem são selecionadas e aplicadas. Nesta fase buscou-se priorizar a aplicação direta de algoritmos de classificação, com e sem a seleção de atributos, em uma base de dados extraídas de repositórios abertos de avaliações do INEP/MEC.

Os algoritmos selecionados para a predição foram: NaiveBayes, J48, JRip, LibSVM, RandomForest, IBK, OneR e REPTree, todos “caixa branca. Tais algoritmos foram selecionados pelo fato de estarem largamente presentes em outros estudos relevantes na comunidade acadêmica, e além disso, representam as duas principais classes de algoritmos de classificação.

A etapa de seleção de atributos tem como objetivo excluir atributos redundantes e que não são úteis para a criação do modelo de predição. Ao usar a seleção de atributos, busca-se um melhor desempenho e também a simplificação do modelo, reduzindo o custo computacional Tasca (2015). Ao concluir o pré-processamento de dados, aplicou-se à base de dados os algoritmos de seleção de atributos e *Ranking ChiSquared-AttributeEval*, *Filtered-AttributeEval*, *GainRatio-AttributeEval*, *InfoGain-AttributeEval*, *OneRAttributeEval* do Weka. Os cinco métodos de seleção do tipo Ranker aplicados, mostraram-se eficientes e conseguiram reduzir a base de dados de 88 atributos para apenas 21, uma redução de 78% em relação ao número total de atributos.

Severino (2008) conceitua a pesquisa de campo como aquela em que o objeto de estudo é abordado dentro do seu próprio meio ambiente. Como a pesquisa foi aplicada com escolas,

foi preciso que os dados necessários para a pesquisa fossem coletados de forma que eles representassem de forma mais fidedigna possível a realidade dessas escolas, isto é, que o fenômeno que é também alvo do estudo – a experiência de cada diretor – fosse observada de forma natural no próprio meio ambiente destas escolas, optando por meio de entrevistas semiestruturadas *in-loco*.

Destaca-se que para a elaboração das entrevistas utilizadas neste trabalho foram observados aspectos como planejamento da coleta de informações, as variáveis que afetariam os dados de coleta e bem como a análise conforme preconizou Manzini (2003).

Ainda segundo o autor, no tocante a entrevista semiestruturada está possui como característica questionamentos básicos que são apoiados em teorias e hipóteses que se relacionam ao tema da pesquisa. Tais questionamentos dariam frutos a novas hipóteses surgidas a partir das respostas dos entrevistados no presente estudo os gestores educacionais.

Segundo Triviños (1987, p. 152) a entrevista semiestruturada “[...] favorece não só a descrição dos fenômenos sociais, mas também sua explicação e a compreensão de sua totalidade [...]”. Diante deste processo a adequação e análise do roteiro ganha importância para evitar algum viés que comprometa a riqueza das informações que serão obtidas diante deste processo de interação social em sua interpretação.

Figura 2 - Etapas específicas do desenvolvimento da pesquisa.



Fonte: Autor (2020).

1. Primeira Rodada. Nesta etapa objetivou avaliar através dos 16 diretores escolares o que eles consideravam das perguntas existentes no formulário SAEB 2015: Sistema de Avaliação da Educação Básica Questionário Diretor que se encontra no apêndice B deste trabalho, e que eles escolhessem as 10 questões mais relevantes para influenciar o IDEB da sua escola.

2. Filtragem. Nessa etapa o especialista da área educacional, selecionou do questionário SAEB 2015 as questões que julgou mais relevantes para impactar positivamente o IDEB, resultando essa filtragem em 31 questões dentre as 111. Essa tarefa foi importante para elaborar um novo questionário mais sintético para ser aplicado numa segunda rodada junto às escolas. As questões selecionadas encontram-se dispostas no **Quadro 2** - Atributos selecionados pelo especialista.

3. Segunda rodada. Conforme seleção de 21 questões filtradas pelo especialista da área educacional motivou-se uma análise para elaboração de um novo questionário desta vez, dimensionado pelas áreas temáticas provenientes do questionário SAEB 2015, a fim de obter junto aos diretores suas percepções dos fatores que contribuem para uma melhoria no IDEB. Esse novo questionário teve 6 (seis) perguntas.

4. Análise. Mediante a aplicação junto aos 16 diretores escolares do questionário derivado do especialista através de entrevista semiestruturada, foram analisados os conteúdos por meio de técnicas de nuvem de palavras e feito um ranking de 30 palavras mais citadas e confrontadas com as saídas da parte computacional e discutidas.

5. Técnicas de MDE. Processo realizado para obtenção das questões mais relevantes pelos algoritmos por parte da máquina, perspectiva computacional e confrontado pelos gestores escolares conforme **Quadro 4**.

6. Validação. Essa etapa consistiu em apresentar os resultados e validação pelo especialista da área e gestores.

4 UMA ABORDAGEM HUMANO COMPUTACIONAL PARA IDENTIFICAÇÃO DOS FATORES QUE INFLUENCIAM O IDEB NA REDE DE ENSINO BÁSICO DE MACEIÓ-ALAGOAS

4.1 Introdução

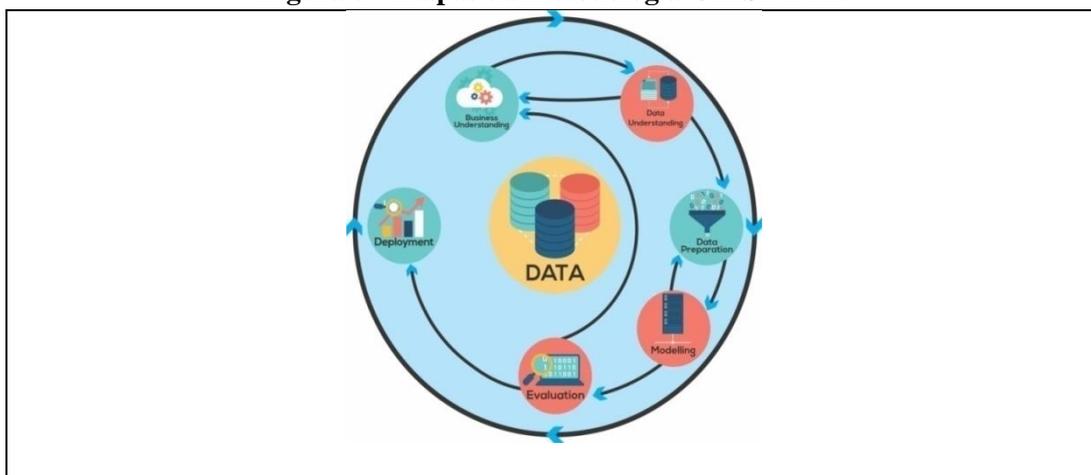
Este capítulo tem como objetivo geral desenvolver uma abordagem metodológica que integre a visão computacional com a perspectiva dos gestores educacionais (visão dos diretores escolares), visando identificar os fatores relevantes que possam contribuir para a melhoria do IDEB. Para tanto, selecionou-se um conjunto de 16 escolas da rede municipal de ensino fundamental de Maceió (séries iniciais), sendo 8 com maiores IDEB e 8 com os menores IDEB por região. Já na visão computacional foram observadas todas as 82 escolas da rede municipal de ensino básico relativas ao ano letivo 2015 para efeito de modelagem por meio das técnicas de seleção de atributos e mineração de dados educacionais utilizados nesta pesquisa.

4.2 Visão Computacional

Na visão computacional utilizou-se a metodologia CRISP-DM (CHAPMAN *et al.*, 2000). Essa metodologia define uma sequência não rígida de seis etapas, que permite a construção e implementação de um modelo de mineração para ser usado em um ambiente real, auxiliando as decisões de políticas públicas (negócio, dentro do linguajar de TIC).

Dessa forma, no desenvolvimento do trabalho seguiram as etapas do CRISP-DM, apresentadas na **Figura 3** descrita a seguir.

Figura 3 - Etapas da Metodologia CRISP-DM.



Fonte: Adaptado de Chapman *et al.* (2000).

1ª Etapa: Compreensão do domínio. Nesta etapa, além de definir os objetivos a serem alcançados, determinam-se os critérios de sucesso do projeto. Também, serão definidos os recursos necessários à execução de todo o plano de mineração. O objetivo deste projeto é desenvolver uma abordagem metodológica que integre a visão computacional com a perspectiva dos gestores educacionais (visão humana), visando identificar os fatores relevantes que possam contribuir para a melhoria do IDEB na rede municipal de Maceió-Alagoas.

2ª Etapa: Entendimento dos dados. Os dados iniciais trabalhados nesta pesquisa já existem, porém necessitam ser analisados frente à realidade do sistema educacional, estabelecendo uma relação direta entre a percepção dos gestores escolares e os resultados divulgados. As bases de dados usadas nesse estudo foram disponibilizadas abertamente pelo INEP em seu portal. Coletou-se os dados educacionais das escolas municipais de Maceió no portal do INEP. Segundo o INEP (2019), o questionário do diretor dos anos iniciais do ensino fundamental consiste de 111 itens, distribuídos em 8 (oito) categorias ou dimensões: informações básicas; características da equipe escolar; políticas, ações e programas escolares; visão sobre a merenda escolar; visão sobre a escola e a gestão; recursos financeiros e livros didáticos; violência na escola e ensino religioso (**Quadro 2**).

Quadro 2 - Dimensões do questionário do diretor escolar.

Visão do ator	Dimensões	Características das dimensões
Gestor	Formação e experiência profissional	Formação, experiência profissional e características funcionais
	Equipe escolar	Atitudes e composição da equipe escolar
	Políticas, ações e programas	Políticas usadas para admissão e alocação de alunos
	Escola e gestão	Problemas da escola e dificuldade de gestão
	Recursos financeiros e livros didáticos	Fontes de recursos que contribuem para o funcionamento da escola e sobre os livros didáticos
	Merenda escolar	Merenda oferecida pela escola
	Violência na escola	Ocorrências de fatos que afetam a segurança na escola
	Ensino religioso	Atividades de ensino religioso que ocorreram na escola

Fonte: Adaptado do INEP (2019)

3ª Etapa: Preparação dos dados. Esta etapa tem o objetivo de construir o conjunto final de dados que será utilizado nas ferramentas de modelagem. As tarefas de preparação podem ser realizadas muitas vezes, e sem uma ordem pré-determinada. Nesta etapa foi utilizada a ferramenta Pandas com a linguagem em PYTHON, envolvendo operações para tratar a falta de dados em alguns campos, limpeza de dados como a verificação de inconsistências, redução da

quantidade de campos em cada registro, o preenchimento ou a eliminação de valores nulos, remoção de dados duplicados. Inicialmente, devido ao fato de o INEP disponibilizar apenas os dados nacionais, foi necessário um filtro para selecionar apenas os alunos das escolas públicas do município de Maceió.

4ª Etapa: Modelagem. Por meio de processos de Mineração de Dados Educacionais é possível encontrar padrões de comportamentos de um conjunto de dados e por meio desses padrões usar algoritmos com métodos indutivos com objetivo de validar uma hipótese no espaço de possíveis questionamentos dentro dos conjuntos de dados dispostos. Neste trabalho, tem-se um problema de aprendizagem de máquina para um conjunto de dados educacionais, onde para solucionar os questionamentos realizados se utilizou os diversos algoritmos de aprendizagem de máquina, respeitando o processo de CRISP-DM onde na primeira etapa se utilizou algoritmos de seleção de atributos conforme disponível no **Quadro 3**. Em seguida aplicou-se algoritmos de aprendizagem de máquina do grupo de aprendizagem supervisionada na qual para mensurar a qualidade de predição se obtém os valores de saída dos algoritmos por meios das seguintes métricas: acurácia, Fmeasure e Curva ROC.

Quadro 3 - Abordagens de seleção de atributos utilizados na pesquisa.

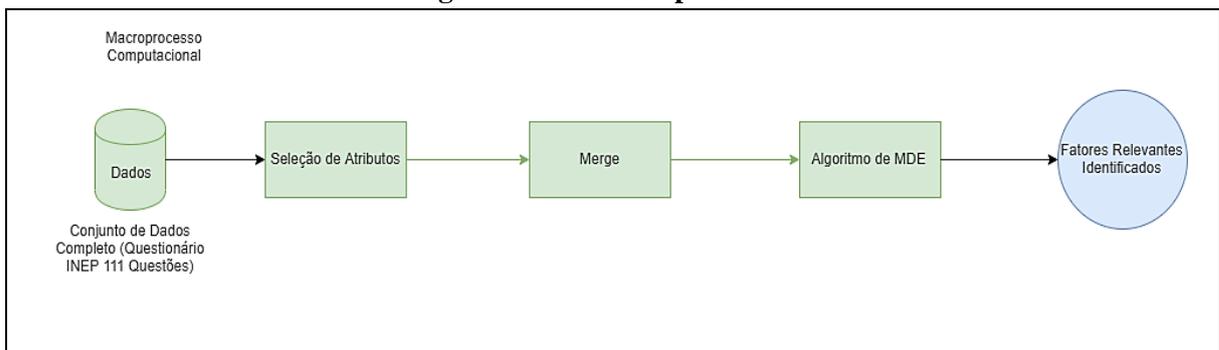
Abordagem	Algoritmo
Embutida	REPTree e J48
Filtro	CfsSubsetEval, CorrelationAttribute, ChiSquaredSubsetEval, GainRatio-AttributeEval, InfoGain-AttributeEval, OneRAttributeEval, SymmetricalUncertAttributeEval e ReliefFAttributeEval
Embrulhamento	WrapperSubsetEval com o NaiveBayes

Fonte: Autor (2020).

Na **Figura 4**, têm-se as etapas da visão computacional utilizada com base no processo CRISP-DM, na qual é representado no primeiro passo a compreensão dos dados. No segundo passo, é representada a seleção de atributos utilizando algoritmo do Filtro, Embrulhamento e Embutida. No terceiro passo, é apresentada a utilização de um método de junção de atributos para formar uma base de dados com os melhores atributos pós-processos de seleção de atributos. No quarto procedimento, utiliza-se o processo de mineração de dados educacionais por meio dos seguintes algoritmos: NaiveBayes, J48, JRip, LibSVM, RandomForest, IBK,

OneR e REPTree. Todos os algoritmos usados fazem parte da categoria de algoritmos de aprendizagem supervisionada. Sendo NaiveBayes, LibSVM, RandomForest e IBK caixa preta e J48, JRip, OneR e REPTree caixa branca. A última fase do processo é a verificação dos valores obtidos com as fases anteriores, denominada de reconhecimento de fatores relevantes usados na tomada de decisão. As informações escolhidas por meio dos procedimentos realizados podem ajudar aos especialistas na tomada de decisão para melhoria do IDEB do município de Maceió.

Figura 4 - Visão computacional.



Fonte: Autor (2020).

A etapa de pré-processamento dos dados é realizada porque na maioria das vezes os dados encontrados da base de dados não são utilizáveis para os algoritmos de MDE uma vez que é necessário realizar as seguintes etapas para que os dados sejam ajustáveis aos algoritmos: Eliminação manual de atributos; Integração de dados; Amostragem de dados; Balanceamento de dados; Limpeza de dados; Redução de dimensionalidade; e Transformação de dados (CARVALHO, 2011).

Durante a etapa de processamento de dados e demais processos dessa dissertação, utilizou-se a API Java Weka com implementações em linguagem de programação Java para encontrar os melhores resultados durante o as tarefas anteriormente mencionadas. No final desse processo, construiu-se um arquivo “.csv” contendo o conjunto de dados utilizado como entrada para realizar as tarefas seguintes o arquivo “diretores-base-maceio.csv”.

Ao utilizar a seleção de atributos, busca-se um melhor desempenho e a simplificação do modelo, reduzindo com isso o custo computacional (MÁRQUEZ-VERA *et al*, 2013). Para selecionar os dados mais significativos para este trabalho foram utilizados algoritmos de cada grupo de método de seleção que são: filtro, embrulhamento e embutida. A abordagem “Todos” representa a combinação das três abordagens anteriores (**Quadro 4**).

Quadro 4 - Seleção de atributos extraídos da máquina através da abordagem de seleção de atributos por categoria.

Abordagem	Algoritmo	Questionário Diretor	
		Atributos	Quantidade
Embutida	REPTree e J48	23,31,35,83,93,103 e 111	07
Filtro	CfsSubsetEval, CorrelationAttribute, ChiSquaredSubsetEval, GainRatio-AttributeEval, InfoGain-AttributeEval, OneRAttributeEval, SymmetricalUncertAttributeEval e ReliefFAttributeEval	5,13,14, 24,31,32,35, 36,39,41,53,62,64,66,70, 71,72,75,81,84,91,94,100 ,101,102,103 e 104	27
Embrulhamento	WrapperSubsetEval com o NaiveBayes	111	01
Todos		5,13,14,23,24,31,32,35,3 6,39,41,53,62,64,66,70,7 1,72,75,81,83,84,91,93,9 4,100,101,102,103, 104 e 111	31

Fonte: Autor (2020)

Além das abordagens de seleção de atributos tradicionais apresentadas, utiliza-se um método, denominado de Merge, que combina os atributos mais frequentes nos melhores conjuntos de seleção (LIMA, 2016). Nesse método para cada atributo será gerado um *score*. Ordenando esses *scores* será possível obter um *ranking* da mesma forma que qualquer técnica individual de uma das abordagens apresentadas anteriormente.

Fatores relevantes são os resultados obtidos por meio do processo de MDE, onde nessa etapa, avaliam-se os resultados obtidos na etapa de seleção de atributos e as saídas dos algoritmos caixa branca a exemplo o algoritmo J48 onde se obteve os melhores resultados tem-se a visualização da árvore gerada que pode ser usado no processo de análise de resultados na

próxima etapa através do especialista.

A partir do **Quadro 4** é possível observar que a estratégia de seleção de atributos usada por categorias como embutida, filtro e embaralhada, permitiu evidenciar os melhores atributos do conjunto de 111 para 31, redução de 72,08% desses atributos. De acordo com os atributos selecionados que compuseram o modelo *Todos*, 31 atributos foram identificados como fortemente impactantes pelas abordagens de seleção de atributos, quais sejam: 5, 13, 14, 23, 24, 31, 32, 35, 36, 39, 41, 53, 62, 64, 66, 70, 71, 72, 75, 81, 83, 84, 91, 93, 94, 100, 101, 102, 103, 104 e 111. Esses 31 atributos foram selecionados pelos diretores escolares (visão computacional) de todas as escolas municipais de Maceió/AL com base nos algoritmos de seleção como os mais relevantes.

O conjunto de dados foi dividido em Treino e Teste com *Fold* de tamanho 10 e executado 30 vezes para gerar um ranking para obter resultado de acurácia de classificação conforme **Quadro 5** que apresenta a acurácia de predição para o conjunto de dados IDEB 2015 (Gestores) obtida por meio de conjunto de treinamento e teste, por fim, realizado o teste estatístico de Friedman e Nemenyi com objetivo de verificar a eficiência do conjunto de dado gerado pela técnica de EDM em relação aos atributos escolhidos pelos especialista.

Quadro 5 - Acurácia de predição para o conjunto de dados IDEB 2015 (Gestores) obtida por meio de conjunto de treinamento e teste.

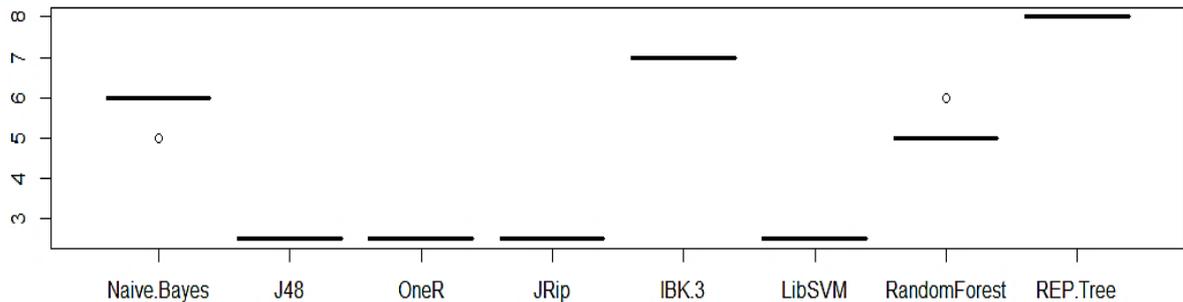
Algoritmo	Base de Diretores	
	Completa	Todos
LibSVM	85,37%	85,37%
RandomForest	85,37%	85,33%
REPTree	84,43%	84,84%
<i>NaiveBayes</i>	82,80%	91,75%
OneR	82,48%	84,80%
<i>JRip</i>	80,24%	79,76%
<i>J48</i>	78,46%	81,95%
IBK	78,25%	85,57%
Precisão Média	82,64%	85,35%

Fonte: Autor (2020).

Na **Figura 7**, tem-se o gráfico estatístico de Friedman e Nemenyi que foi gerado utilizando o *software R Studio*. É importante lembrar que para gerar a diferença estatística foi considerada a exaustão da base de dados “Todos” e “Completa”; em ambos os casos foi executado o processo de mineração de dados educacionais com *cross* validação e *fold10*, onde

se executou um *script* com *apijavaWeka* para cada um dos algoritmos usados o processo de classificação afim de classificar o IDEB acima da média ou a abaixo da média, executado 30 vezes. E, por fim, realizada a operação de média dos números de execuções, nesse caso 30 vezes com sementes diferentes para cada uma das execuções dos algoritmos com os 8 algoritmos das diferentes modalidades. Processo realizado por duas vezes conforme as duas bases existentes.

Figura 5 - Aplicação do método de avaliação estatística de Friedman e Nemenyi no R para comparar as saídas dos classificadores.



Fonte: Autor (2020).

A partir da **Figura 5**, pode-se perceber que NaiveBayes obteve o melhor resultado e *OneR*, *RandomForest*, *LibSVM*, *REPTree*, *IBK* empatarão, estando próximos estatisticamente em relação aos seus resultados. Já os algoritmos *JRip* e *J48* têm resultados bons, mas também são diferentes estatisticamente entre eles e, além disso, em relação aos demais. Por fim, é importante lembrar que o comportamento de predição da base de dados definidos como “muito bom” estão no intervalo compreendido entre 70% a 91% de acurácia de predição, assim como são também bons nas demais métricas.

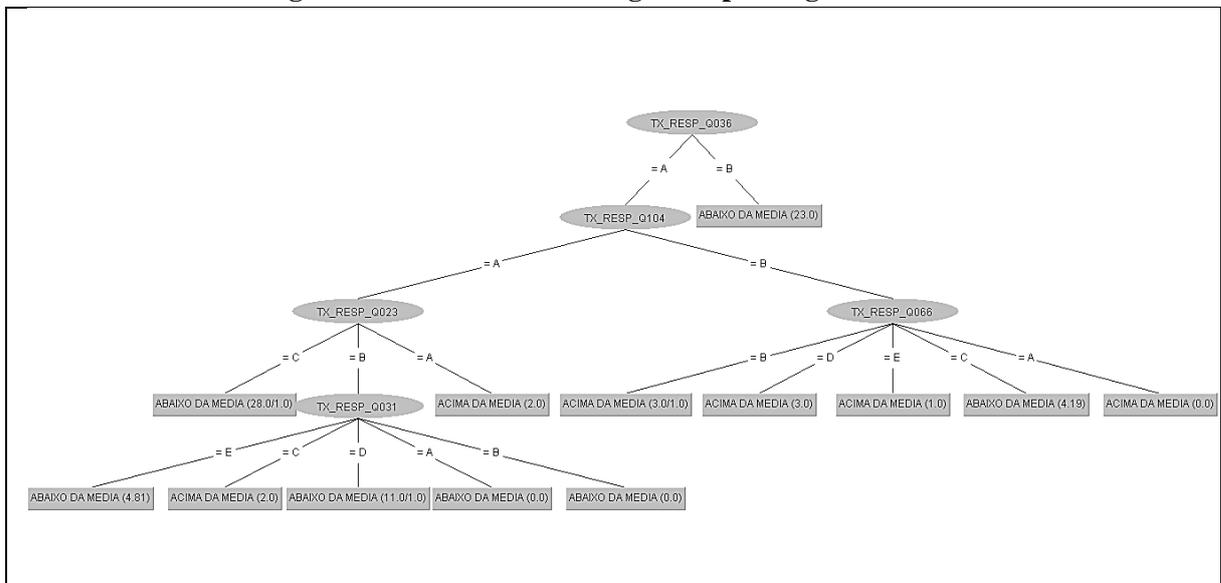
Na **Figura 6**, tem-se a árvore de decisão obtida pelo algoritmo *J48* com a base de dados “Todos” onde se apresentam como principais atributos:

- **Questão Q-23** (Se você gostaria de ter participado de mais atividades de desenvolvimento profissional, os seguintes motivos impediram sua participação? Houve conflito com o meu horário de trabalho). Os gestores que possuem dedicação exclusiva com carga horário maior dedicada a escola possuem melhor condições de evoluem os resultados do IDEB da escola;
- **Questão Q-31** (O conselho de classe é um órgão formado por todos os professores que lecionam em cada turma/série. Neste ano e nesta escola, quantas vezes se reuniu o conselho de classe?). O envolvimento do conselho escolar da unidade de ensino é bastante importante para a evolução dos resultados positivos para a escola possivelmente isso ocorre porque a unidade escolar que possuem uma gestão democrática com foco em atender as demandas da comunidade escola como um todo

podem de forma coletiva resolver lacunas existentes providenciar as melhores infraestrutura e aparatos pedagógicos que podem melhorar nos resultados do IDEB;

- **Questão Q-36** (Você conhece os resultados do SAEB?). Quando o diretor da unidade escolar conhece a estrutura do sistema de avaliação ele tem bem mais chance de evoluir nos resultados de sua escola;
- **Questão Q-66** (Em relação à merenda escolar, como você avalia os seguintes aspectos: Disponibilidade de funcionários). A merenda, nas escolas públicas, sempre foi um vetor de presença e satisfação. Nesse ponto, pode-se compreender que a alimentação dos alunos é muito importante para a tentativa de melhoria dos resultados do IDEB por parte da gestão escolar. Para isso, há uma demanda de responsabilidade administrativa que a equipe gestora deve tomar para garantir a boa qualidade de merenda e manter um número de funcionários suficientes para atender aos alunos, mantê-los motivados a frequentarem à escola, fator, portanto, de permanência de fluxo;
- **Questão Q-104** (Nesta escola, há projetos nas seguintes temáticas: Bullying). Nesse caso os gestores que trabalham com a temáticas relacionada ao emocional dos alunos possuem melhores condições para avançar no melhor resultado do exame. Quando a escola se preocupa com o bullying, é porque a gestão está mais próxima dos alunos, tomando providências, dando segurança.

Figura 6 - Árvore de decisão gerado pelo algoritmo J48.



Fonte: Autor (2020).

Conforme informações que constam no questionário do IDEB a questão **Q-36** (Você conhece os resultados do SAEB (Prova Brasil e/ou ANEB) do seu estado?) Refere-se ao conhecimento e utilização do gestor dos resultados da prova SAEB do ano anterior. O item A. (Sim) representa o conjunto de dirigentes escolares que conhecem os resultados do IDEB de

suas respectivas escolas e B (Não) são os que não conhecem. Questão **Q-104** (Escolas com projetos temáticos na escola), item A (Sobre Bullying) essa folha da árvore mostra que ao se importar por temáticas iguais a essa as chances aumentam na busca de melhores resultados e desempenho dos alunos. No item B (Não) se refere aos gestores que não trabalham essas temáticas os ramos no levam a direita da árvore para questão **Q-66** (Aspectos da merenda escolar) com os itens (A= Inexistente, B= Ruim, C= Razoável, D= Bom e E = Ótimo), na folha em que a qualidade da merenda está baixa tem-se também uma diminuição do desempenho da escola. Na folha da árvore que leva a questão **Q-23** (Conflito de carga horária do gestor), tem-se (A= Não gostaria de ter participado, B= Não, C= Sim) nesse caso fica entendido que a existência de conflito de cargo horário por parte do gestor escolar leva a instituição a resultados baixo no IDEB. A questão **Q-31** (Encontros realizados pelo conselho de classe da escola), tem-se (B= nenhuma vez, C= uma vez e A= três vezes) é possível notar que a participação efetiva do conselho de classe aumenta a possibilidade de melhoria dos resultados da escola uma vez que o órgão colegiado da escola oportuniza a correção de problemas existente com alunos ou com profissionais durante o ano letivo, contribuindo na melhoria dos resultados da entidade.

5ª Etapa: Avaliação. Nesta etapa, tem-se construído um ou mais modelos que aparentam ter alta qualidade. Ao final será tomada uma decisão a partir dos resultados da mineração, sem, entretanto, desconsiderar alguma questão que seja importante.

A partir do **Quadro 5** é possível observar que a estratégia de seleção de atributos usada por categorias como embutida, filtro e embrulhamento, permitiu evidenciar os melhores atributos do conjunto de 111 para 31, redução de 72,08% dos atributos. De acordo com os atributos selecionados que compuseram o modelo *Todos*, 31 atributos foram identificados como fortemente impactantes pelas abordagens de seleção de atributos, quais sejam: 5, 13, 14, 23, 24, 31, 32, 35, 36, 39, 41, 53, 62, 64, 66, 70, 71, 72, 75, 81, 83, 84, 91, 93, 94, 100, 101, 102, 103, 104 e 111. Esses 31 atributos foram selecionados pelos diretores escolares de todas as escolas municipais de Maceió/AL com base nos algoritmos de seleção como os mais relevantes. Portanto, na visão computacional se obteve os seguintes atributos mais relevantes:

- **Questão Q-5:** Há quantos anos você obteve o nível de escolaridade assinalado anteriormente (até a graduação)?
- **Questão Q-13:** Qual é a sua carga horária de trabalho semanal, como Diretor, nesta escola?
- **Questão Q-14:** Você assumiu a direção desta escola por meio de que?
- **Questão Q-23:** Houve conflito com o meu horário de trabalho para participar de mais atividades de desenvolvimento profissional?
- **Questão Q-24:** Não tinha disponibilidade de tempo para participar de mais atividades de desenvolvimento profissional?

- **Questão Q-31:** Quantas vezes se reuniu o conselho de classe?
- **Questão Q-32:** Neste ano e nesta escola, como se deu a elaboração do projeto pedagógico?
- **Questão Q-35:** Você conhece os resultados do SAEB (Prova Brasil e/ou ANEB) do seu município?
- **Questão Q-36:** Você conhece os resultados do SAEB (Prova Brasil e/ou ANEB) do seu estado?
- **Questão Q-39:** Neste ano, qual foi o principal critério utilizado para a formação das turmas nesta escola?
- **Questão Q-41:** Nesta escola, há alguma ação para redução das taxas de abandono?
- **Questão Q-53:** Neste ano, a escola promoveu eventos para a comunidade?
- **Questão Q-62:** Merenda escolar – recursos financeiros.
- **Questão Q-64:** Merenda escolar – qualidade dos alimentos.
- **Questão Q-66:** Merenda escolar – disponibilidade de funcionários.
- **Questão Q-70:** Carência de pessoal de apoio pedagógico.
- **Questão Q-71:** Falta de recursos pedagógicos.
- **Questão Q-72:** Interrupção das atividades escolares.
- **Questão Q-75:** Alta rotatividade do corpo docente.
- **Questão Q-81:** Esta escola recebeu neste ano apoio financeiro do governo federal?
- **Questão Q-83:** Esta escola recebeu neste ano apoio financeiro do governo municipal?
- **Questão Q-84:** Esta escola recebeu neste ano apoio financeiro de empresas ou doadores individuais.
- **Questão Q-91:** Agressão verbal ou física de alunos a outros alunos da escola.
- **Questão Q-93:** Você foi ameaçado por algum aluno.
- **Questão Q-94:** Você foi vítima de furto (sem uso de violência).
- **Questão Q-100:** Violência.
- **Questão Q-101:** Os malefícios do uso de drogas.
- **Questão Q-102:** Racismo.
- **Questão Q-103:** Machismo e homofobia.
- **Questão Q-104:** Bullying.
- **Questão Q-111:** Nesta escola há atividades para os estudantes que não queiram participar das aulas de ensino religioso.

Dessa forma, podem-se agrupar os atributos mais relevantes da visão computacional nas suas respectivas dimensões. Observa-se que 25,8% dos atributos se referem à dimensão Violência na escola, seguida da dimensão Informações básicas com 16,12% e da dimensão Políticas, ações e programas com os mesmos 16,12%. Em seguida vem a dimensão Escola e a gestão com 12,9%, a dimensão Visão sobre a merenda escola com 9,67% e a dimensão Recursos financeiros e livros didáticos com os mesmos 9,67%. Por último, vem a dimensão Características da equipe escolar com 6,45%.

Esses atributos permitem verificar de imediato a tendência dos alunos para o sucesso ou não do resultado no IDEB. À medida que se realizou a seleção de atributos, teve-se os atributos com melhor *ranking* e também melhoria do tempo de processamento da base de dados.

4.3 Visão Humana

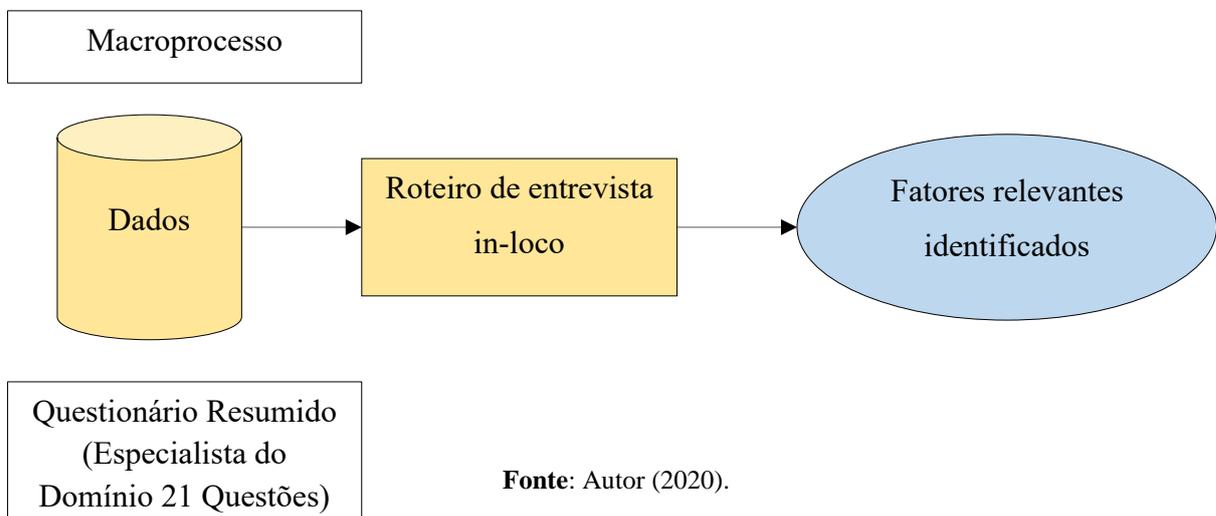
É necessário se fazer algumas considerações para a introdução adequada deste item. O ponto central da pesquisa é definir fatores que influenciariam o IDEB de escolas de ensino fundamental. Nesse universo, são detectados três atores principais: estudantes, professores e diretores. As definições junto ao orientador para levar a cabo a presente pesquisa, foram complementadas por trabalhos obtidos na pesquisa bibliográfica realizada.

Vários trabalhos relatados em artigos utilizaram dados de estudantes e professores, sem se referir ao diretor para entender a formação do IDEB nas escolas. Por outro lado, viu-se que no sítio do INEP existem relatórios com dados socioeconômicos oriundo dos atores citados, os quais poderiam ser utilizados para uma busca e análise de detalhes para entender melhor a gestão da escola e, conseqüentemente, o alcançar de metas educacionais. Elegeram-se esses atores e não os demais, por conta desse contexto.

Ressalte-se que a pesquisa contou, também, com a análise de um especialista na área que emitiu suas impressões sobre as ações do gestor educacional na Rede Pública das escolas municipais de Maceió. Dentro de uma escola, diante do grande número de ações, o corre-corre do dia a dia, a gestão possui a visão geral da escola, desde as ações pedagógicas, as administrativas, bem como as políticas para com a comunidade interna, comunidade externa e a Secretaria Municipal de Educação (sua mantenedora). Permitiria, portanto, maiores subsídios para analisar o comportamento da escola em seu desempenho frente à avaliação do SAEB no indicador IDEB.

A **Figura 7** mostra um esquema do roteiro simplificado da pesquisa com os diretores das escolas que participaram da presente pesquisa.

Figura 7 - Esquema simplificado da pesquisa com os diretores das escolas.



Os diretores, portanto, foram os atores escolhidos, por que foi levada em consideração a visão que os mesmos têm dos aspectos administrativos, da gestão escolar e todos eles são professores, e passaram pelas salas de aula. Portanto, a visão desse ator é a mais completa ao entendimento das ações nas escolas, suas dificuldades e suas soluções. A sua experiência, por conta dessa visão ampla, ajudaria bastante nas respostas ao questionário e, como se espera na pesquisa, levar aspectos e informações novas ao que se encontra na literatura técnica sobre o assunto.

Inicialmente, foram selecionadas as 111 questões do formulário IDEB. Houve uma primeira rodada de entrevista onde o pesquisador se dirigiu às 16 escolas, nas 8 Regiões Administrativas da cidade de Maceió, e foram deixadas essas questões para os diretores responderem e entregarem. Depois dessa visão, primeira rodada, isso ficou arquivado para futuras análises. Passou-se, então, para a análise de seleção de atributos, e foram reduzidas de 111 para 31 questões, via aplicação de algoritmos em máquina.

Feita essa redução, essas foram encaminhadas para as considerações do especialista que levou em conta a experiência gerencial e as definições do INEP para o IDEB. O questionário do SAEB já agrupa as questões em 8 dimensões: formação e experiência profissional; equipe escolar; políticas, ações, programas da escola; recursos financeiros; livro didático; merenda; violência; ensino religioso. Das 31 questões, o especialista conjuntamente com o pesquisador, conseguiu reduzir para 6 questões, de forma a cobrir as 8 dimensões e as 31 questões originais. Isso foi feito para dar maior tranquilidade para os gestores escolares responderem. Ressalte-se que as 6 questões trouxeram em seu bojo a necessidade de resposta abarcando as 31 questões originais.

Figura 8 - Considerações das questões SAEB relevantes pelo especialista.



UNIVERSIDADE FEDERAL DE ALAGOAS/UFAL
Programa de Pós-Graduação em Modelagem Computacional de Conhecimento
Avenida Lourival Melo Mota, Km 14, Bloco 09, Cidade Universitária
CEP 57.072-900 – Maceió – AL – Brasil
Telefone: (082) 3214-1364



Questionário derivado do especialista

Questões Pessoais e Funcionais

1. Você acha que alguma característica pessoal do diretor tais como: sexo, formação acadêmica e estado civil podem influenciar o seu desempenho profissional? Justifique sua resposta. (Ref. 1,2,4 e 8)

Questões Profissionais

2. O seu tempo como diretor e/ou professor na escola em que atua pode influenciar na sua política administrativa? Justifique sua resposta (Ref. 13,15, 17)
3. Você entende que o desempenho acadêmico da sua escola é fruto do trabalho cooperativo entre docentes, família do aluno e sociedade cível? Justifique sua resposta. (Ref. 29 ,31,32)
4. O tempo e as atividades desempenhadas pelos professores na escola são trabalhados de maneiras integradas? Por exemplo; projeto de laboratório, aulas de reforço, concomitante ao conteúdo dado em sala de aula, projetos junto à comunidade (Ex. reciclagem). Justifique sua resposta. (Ref. 28, 40, 42, 50 a 55)
5. Como se processa a comunicação entre os atores da escola (direção, professor, aluno e responsáveis) com foco no ensino e aprendizagem e relações interpessoais (respeito, simpatia e ISE (Inteligência Sócio Emocional) (Ref. 44,45 a 49)

Questão de Gestão

6. Quais as carências mais frequentes em sua escola e como elas influenciam no desenvolvimento das atividades escolares (merenda, espaço físico, material de consumo, didático e humano). (Ref. 65, 68 a 76)

Fonte: Autor (2020).

Na primeira rodada da pesquisa *in loco* foram entrevistados 16 diretores, aplicando-se o mesmo questionário com 111 questões. Uma data foi estabelecida para o retorno, em quatro dias no máximo, com o recolhimento nas escolas pelo pesquisador. No primeiro contato, o pesquisador explicou o questionário, a intenção, dentre outros aspectos. Na segunda rodada da pesquisa *in loco*, com o questionário com 6 questões, as respostas foram imediatas. Apenas 2 (duas) escolas, devido ausência das diretoras na escola, ficou para serem encaminhadas por meio eletrônico, o que aconteceu. Portanto, do total das 16 escolas foram obtidas respostas 100% em tempo hábil.

Como houve duas etapas da pesquisa em campo, no primeiro momento para responder a 111 questões, os diretores criticaram, acharam um excesso de trabalho em perguntas similares ou sem sentido, e muito cansativo. A reflexão do pesquisador foi relacionar esse modelo ao já utilizado por todas as gestoras no INEP. Não houve, portanto, uma resistência ao questionário, mas reclamações pelo número de questões e o tempo para respondê-las. Mas, houve respostas. Já na segunda fase, com 6 questões, o questionário foi muito bem aceito e já entenderam a ideia

da pesquisa de mestrado, no batimento com os dados do SAEB no INEP, garimpados via máquina. Elogiaram a proposta do trabalho e ficaram bem à vontade com o formulário reduzido.

Em um número de 5 diretores responderam às perguntas de forma mais rápida, sem adicionar detalhamentos. Mas, a maioria das 16 escolas, procurou responder as questões com mais explicações, chegando-se a um caso onde foram gastos 30 min para responder a uma única questão. Houve, portanto, boa participação dos diretores entrevistados com respostas detalhadas. O tempo de entrega do questionário da primeira rodada, levou 4 dias. Na segunda rodada, as respostas foram praticamente imediatas.

É importante falar sobre a ligação das respostas às questões do SAEB já respondidas ao INEP (no Censo Escolar INEP) e as obtidas na pesquisa de campo. Na primeira rodada, as críticas ao formulário completo do SAEB foram contundentes por parte dos diretores, por não entenderem a necessidade de tantas perguntas, e que não tinham a ver com o dia a dia da escola, inclusive ao IDEB. Já na segunda etapa, com as 6 perguntas no novo formulário, o questionário foi melhor entendido pelos diretores que entenderam bem as questões e as consideraram importantes para o resultado no IDEB, de modo nítido. Portanto, o segundo formulário propiciou ao diretor participar com mais foco nos ideais da presente pesquisa de mestrado.

Na aplicação da pesquisa em campo, ouvindo todos os diretores, a pergunta de tempo de serviço nos cargos diretor/professor influencia nos resultados positivos para a escola. Foi forte, também, que o trabalho cooperativo na escola e com os familiares dos estudantes influencia sobremaneira para a melhoria das respostas da escola às suas metas educacionais. Na infraestrutura, também, houve concordância em reclamações nas condições de espaços com conforto térmico não os ideais e tamanho de espaços, bem como a falta deles, para todas as atividades inerentes à escola. Esses detalhes faltantes na infraestrutura diziam negativamente para alcançar as metas positivas das escolas.

Para o melhor IDEB na escola, haver harmonia no meio ambiente, nas relações gestor professor, pessoal auxiliar, técnicos administrativos, estudantes e comunidade foi colocado como fundamental para o bom funcionamento das escolas e referencial positivo no IDEB. As questões, portanto, de meio ambiente nas escolas, traz condições para se obter melhores resultados pedagógicos, conseqüentemente, melhor IDEB.

A metodologia utilizada foi interessante e inovadora, na medida que se pensou em obter o que o INEP coletava de informações, através de seu formulário padrão, e que não funcionou tão bem como se esperava. Mas, na medida em que foi feita a seleção das questões de maneira computacional e pelo especialista em gestão educacional, o engajamento, a disponibilidade dos atores diretores foram muito positivos, o que contribuiu para se ter respostas mais detalhadas,

com melhores qualidades nas mesmas. Para a pesquisa esse ajuste foi importante, lidando com o humano, o que deixa uma reflexão sobre o que o formulário do INEP está retirando de fato, se são todas as informações com fidedignidade para os censos do País.

Em dados mais determinísticos, o perfil dos diretores escolares segundo os atributos e características pessoais de todas as escolas públicas municipais de Maceió, segundo os dados do SAEB 2015 (censo escolar 2015), pode-se constatar que a expressiva maioria dos diretores era do gênero feminino (90,24%); quando à cor/raça, que a maioria se declarou pardo (60,97) seguida por brancos (24,39%) e negros (9,75%). Quanto às demais denominações, a porcentagem não ultrapassou 4,98%. Quanto à idade, 65,85% se encontraram na faixa etária compreendida entre 40 e 55 anos.

Apresentaram-se, também, cinco opções para que os diretores apontassem a alternativa mais condizente com a sua formação. Das alternativas elencadas, 54,87% dos diretores se referiram como possuindo “superior – pedagogia”; 37,80% como “superior – licenciatura”; 2,43% com formação superior em outros cursos; 4,87% não possuindo curso de nível superior ou com “normal – antigo magistério”.

Entre os diretores, 1,21% possuem pós-graduação em caráter de aperfeiçoamento – mínimo de 180 horas; 79,26% em caráter de aperfeiçoamento – mínimo de 360 horas; e apenas 2,43% possuíam mestrado ou doutorado, por fim, 17,07% não fez ou não completou nenhum curso de pós-graduação.

À maioria dos diretores, 65,85% tinha alguma formação específica na área de administração escolar e, entre esses, 18,29% possuía graduação com habilidade em administração escolar e 84,13%, especialização nesta área.

O cargo de diretor em 82,92% dos casos ocorreu por meio de eleições; 0% eleitos pelo colegiado; 2,43% indicação, 14,63% por meio de processo seletivo. Relacionou-se, ainda, quantos anos os diretores estavam ocupando o cargo de direção. Observou-se que 15,85% dos diretores ocupam esse cargo por mais de 10 anos, enquanto 48,78% ocupam esse cargo por menos de 5 anos.

4.4 Abordagem Integradora

A abordagem integradora consiste no primeiro momento em realizar um procedimento estatístico, com objetivo de exibir o conjunto de dados e suas particularidades. Em seguida dividem-se os dados em treino e teste e retira-se uma amostra equivalente a 16 escolas divididas por regiões e características definidas pelo gestor (especialista).

Quadro 6 - Relação das escolas escolhidas para a realização da pesquisa (INEP 2015).

REG. ADM		UNIDADES ESCOLARES DO 1º AO 5º ANO	IDEB 2015
RA 01	02	Esc. Mun. Dr. Orlando Araújo – 1º ao 5º	5.3
	03	Esc. Mun. Mª José Carrascosa – 1º ao 5º	3.5
RA 02	05	Esc. Mun. Almeida Leite – 1º ao 5º	4.8
	11	Esc. Mun. Élio Lemos (Nosso Lar) – 1º ao 5º	2.9
RA 03	14	Esc. Mun. Higino Belo – 1º ao 5º	4.9
	19	Esc. Mun. Padre Brandão Lima – 1º ao 5º	2.8
RA 04	24	Esc. Mun. Zilka de Oliveira – 1º ao 5º	2.7
	25	Esc. Mun. Marcos Soriano – 1º ao 5º	5.1
RA 05	38	Esc. Mun. Paulo Freire – 1º ao 5º	5.3
	41	Esc. Mun. Audival Amélio – 1º ao 5º	3.8
RA 06	44	Esc. Mun. Elma Marques Curti – 1º ao 5º	5.7
	46	Esc. Mun. Mª de Fátima Lira – 1º ao 5º	3.9
RA 07	59	Esc. Mun. Dr. Denisson L. C. Menezes – 1º ao 5º	4.0
	70	Esc. Mun. Pedro Suruagy – 1º ao 5º	5.8
RA 08	76	Esc. Mun. Marechal Floriano Peixoto – 1º ao 5º	4.9
	77	Esc. Mun. Pedro Barbosa Júnior – 1º ao 5º	3.1

Fonte: Autor (2020).

Na pesquisa realiza-se o agrupamento das questões por dimensões com objetivo de encontrar os atributos que influenciam no IDEB e compreender por meio das dimensões que os mesmo estão enquadrados abaixo, tem-se as dimensões e no **SAEB 2015** pode-se ver com mais detalhes quais o questionário utilizado como base para aplicação da entrevista em **Áudio e Questionário Físico**.

Segundo o INEP (2019), o questionário do diretor dos anos iniciais do ensino fundamental consiste de 111 itens, distribuídos em 8 (oito) categorias ou dimensões: informações básicas; características da equipe escolar; políticas, ações e programas escolares; visão sobre a merenda escolar; visão sobre a escola e a gestão; recursos financeiros e livros didáticos; violência na escola e ensino religioso.

- Dimensão pessoal (questões 1-27).
- Dimensão profissional /equipe escolar (questões 28-32).
- Dimensão políticas, ações e programas (questões 33- 61).
- Dimensão merenda escolar (questões 62-66).
- Dimensão visão sobre os problemas da escola e dificuldade de gestão (questões 67- 80).
- Dimensão recursos financeiros e livros didáticos (questões 81-89).
- Dimensão violência na escola (questões 90-108).
- Dimensão ensino religioso (questões 109-111).

Quadro 7-Atributos selecionados por meio da EDM e através dos especialistas (Gestor escolar e Secretário de Educação do município estudado).

Campo de Visão	Atributos Selecionados	Estratégias de busca dos artefatos
Visão computacional	5,13,14,23,24,31,32,35,36,39,41,53,62,64,66,70,71,72,75,81,83,84,91,93,94,100,101,102,103,104,111	Seleção de atributos por categorias; Método de ranking de tributos Merge; Classificação do conjunto de dados por meio da base de dados treinamento e teste (Dados IDEB 2015).
Visão do especialista (Secretário da Semed e Diretor escolar)	1,2,4,8,13,15,17,28,29,31,32,34,35,36,40,42,44-45 até 49; 50 até 55; 65 e 68 até 76.	Por meio da análise do questionário e utilizando a expertise de gestão como Secretário Municipal de Educação foi realizada a escolha de atributos que mais influenciam na melhoria dos resultados do IDEB segundo o gestor.

Diretor escolar	28,29,31,32,34,35,36,40,42,44-45 até 49; 50 até 55; 65 e 68 até 76.	Por meio de entrevista (Áudio) e questionário aplicado com gestores de 16 escolas selecionadas um conjunto de dados no conjunto de dados de Maceió IDEB 2015.
-----------------	---	---

Fonte: Autor (2020).

A **Tabela 2** apresenta os artefatos encontrados da intersecção do processo de visão computacional com a visão humana (do especialista).

Tabela 2 - Dimensão descritas no questionário do IDEB aplicado para os gestores em 2015.

Dimensão	Fator	Questão selecionada	Descrição
Pessoal	Escolaridade do diretor	Q-05	Há quantos anos você obteve o nível de escolaridade assinalado anteriormente (até a graduação)?
		Q-13	Qual é a sua carga horária de trabalho semanal, como diretor, nesta escola?
		Q-14	Você Assumiu A Direção Desta Escola Por Meio De
		Q-23	Houve conflito com o meu horário de trabalho para participar de mais atividades de desenvolvimento profissional.
		Q-24	Não tinha disponibilidade de tempo para participar de mais atividades de desenvolvimento profissional.
Profissional /equipe escolar		Q-31	Quantas vezes se reuniu o conselho de classe?
		Q-32	Neste ano e nesta escola, como se deu a elaboração do projeto pedagógico?
Políticas, ações e programas		Q-35	Você conhece os resultados do saeb (prova brasil e/ou aneb) de 2013 do seu município?
		Q-36	Você conhece os resultados do saeb (prova brasil e/ou aneb) de 2013 do seu estado?
		Q-39	Neste ano, qual foi o principal critério utilizado para a formação das turmas nesta escola?

		Q-41	Nesta escola, há alguma ação para redução das taxas de abandono?
		Q-53	Neste ano, a escola promoveu eventos para a comunidade.
Merenda escolar		Q-62	Recursos financeiros.
		Q-64	Qualidade dos alimentos.
		Q-66	Disponibilidade de funcionários.
Problemas da escola e dificuldade de gestão		Q-70	Carência de pessoal de apoio pedagógico (supervisor, coordenador, orientador educacional).
		Q-71	Falta de recursos pedagógicos.
		Q-72	Interrupção das atividades escolares.
		Q-75	Alta rotatividade do corpo docente.
Recursos financeiros e livros didáticos		Q-81	Esta escola recebeu neste ano apoio financeiro do governo federal?
		Q-83	Esta escola recebeu neste ano apoio financeiro do governo municipal?
		Q-84	Esta escola recebeu neste ano apoio financeiro de empresas ou doadores individuais?
Violência na escola		Q-91	Agressão verbal ou física de alunos a outros alunos da escola.
		Q-93	Você foi ameaçado por algum aluno.
		Q-94	Você foi vítima de furto (sem uso de violência).
		Q-100	Violência.
		Q-101	Os malefícios do uso de drogas.
		Q-102	Racismo.
		Q-103	Machismo e homofobia.
		Q-104	<i>Bullying.</i>
Ensino religioso		Q-111	Nesta escola há atividades alternativas para os estudantes que não queiram participar das aulas de ensino religioso.

Fonte: Autor (2020).

4.5 Análise da percepção dos fatores dos gestores escolares

Em conformidade com entrevista realizada *in loco* junto aos diretores escolares, serão descritos através da **Tabela 3**, a síntese dos resultados captados (percepções) e os fatores abordados nas respectivas dimensões.

Tabela 3 - Análise da percepção dos gestores.

Dimensão	Percepção	Fator	Questões identificadas
Pessoal	Pode ser considerado pela amostra que a questão de formação acadêmica exerceu uma forte influência no exercício da boa gestão.	Formação acadêmica Experiência profissional	Q-04, Q-08 e Q-19
Profissional /equipe escolar	O Conselho Escolar e o Conselho de Classe e o foram apontados como colegiados que tem um forte papel para condução da gestão e melhoramento das ações da escola. Um ponto abordado por algumas gestoras acerca do time escolar foram: a forma de trabalho cooperativo com a comunidade e o trabalho colaborativo com os servidores para que a escola consiga ser um ambiente harmônico e agradável isso reflete num conjunto de ações que podem refletir no IDEB.	Conselhos Colegiados Trabalho Colaborativo Clima Organizacional	Q-29 e Q-31
Políticas, ações e programas	Ações para reforço escolar à aprendizagem dos alunos foram destacadas na amostra, juntamente com discursão com os professores com o objetivo de melhorar o ensino e a aprendizagem dos alunos, inclusive com jogos didáticos (reforçando a questão de trabalhar em equipe e saber perder e ganhar) bem como interação com os pais por meio de comunicação (convites) são estratégias que dão resultados favoráveis para o IDEB segundo as entrevistadas. Ações executadas pela secretaria de educação são citadas em mais de 90 % dos entrevistados a saber: HTPC e HTPI essa constatação apresenta um resultado importantíssimo para essa pesquisa. A pesquisa conseguiu captar acerca dos processos escolares que para algumas diretoras tomam muito tempo e poderiam ser investidos em processos pedagógicos. Cabe neste aspecto	Reforço escolar Interação com os pais Formação Continuada	Q-43, Q-44 e Q-46

	<p>uma investigação mais profunda para utilização de novos processos se forem possíveis.</p>		
Merenda escolar	<p>Aspectos de quantidade e qualidade das merendas não foram problemas para 100% da amostra, contudo, no que se refere aos espaços para preparação das merendas foi relatado por algumas como insuficientes. Houve relatos da significância desta dimensão para evolução do aluno frente aos estudos e por consequência melhoramento acadêmico (IDEB)</p>	Alimentação escolar	Q-62, Q-63, Q-64, Q-65 e Q-66
Problemas da escola e dificuldade de gestão	<p>Para 75 % dos entrevistados o fator espaço das escolas é insuficiente para práticas das atividades cotidianas e principalmente na hora do recreio, houve relatos que para cozinhar em algumas escolas era muito difícil devido ao espaço físico, quanto a acessibilidade algumas escolas citaram não ter. Neste fator houve a percepção desafiadora para a gestão.</p> <p>Algumas escolas da amostra relataram carência de pessoal de apoio pedagógico (supervisor, coordenador, orientador educacional)</p> <p>*No questionário SAEB/2015 do diretor não fica evidenciado a questão espaço físico.</p>	Espaço Físico Acessibilidade	Q-57 e Q-70
Recursos financeiros e livros didáticos	<p>Não houve por parte das escolas pesquisadas críticas acerca dos livros didáticos e nem falta dos recursos para manter as atividades. A participação da SEMED foi destacada por alguns diretores como importante neste sentido.</p>	Recursos Financeiros Livros didáticos	
Violência na escola	<p>Na amostra pesquisada o tema violência ficou evidenciado em relatos que casos pontuais de alunos precisavam do acompanhamento do conselho tutelar para aconselhamento. Identificar o contexto familiar foi citado como estratégia para mitigar ações violentas através de projetos, estudos apontam que o papel da escola com a comunidade local reduz índices de violência projetos nas temáticas: Violência, os malefícios do uso de drogas, Bullying dentre outros são importantes.</p>	Violência	

Ensino religioso	Não foi captado pela entrevista fatores relacionados a temática.
-------------------------	--

Fonte: Autor (2020).

4.6 Análise da fala dos gestores

Nesta seção serão analisadas e destacadas as falas mais pertinentes dos gestores de cada unidade de ensino sobre a temática, pois, apesar de seguir uma sequência de perguntas, os entrevistados tinham a liberdade de discorrerem sobre demais aspectos do contexto escolar enriquecendo desta forma a visão humana dos fatores que exercem influência junto ao IDEB.

É importante salientar que ao analisar um discurso mesmo que o documento considerado seja a reprodução de um simples ato de fala individual, conforme esclarece Fischer (2001):

“ não estamos diante da manifestação de um sujeito, mas sim nos defrontamos com um lugar de sua dispersão e de sua descontinuidade, já que o sujeito da linguagem não é um sujeito em si, idealizado, essencial, origem inarredável do sentido: ele é ao mesmo tempo falante e falado, porque através dele outros ditos se dizem”.

A tarefa é desafiadora e deve-se considerar a interdiscursividade, isto é, deixar que aflorem as contradições as diferenças, alguns apagamentos, os esquecimentos; enfim, significa deixar aflorar a heterogeneidade que subjaz a todo discurso ainda conforme Fischer (2001). No contexto da análise do discurso reforça Foucault (1986) acerca do enunciado:

...[não há] enunciado livre, neutro e independente; mas sempre um enunciado fazendo parte de uma série ou de um conjunto, desempenhando um papel nomeio dos outros, neles se apoiando e deles se distinguindo: ele se integra sempre em um jogo enunciativo, onde tem sua participação, por ligeira e ínfima que seja. [...] Não há enunciado que não suponha outros; não há nenhum que não tenha, em torno de si, um campo de coexistências. (FOUCAULT, 1986, p.114)

Realizada entre os dias 29 de outubro à 05 de novembro de 2019a amostra teve 16 (dezesesseis) gestores de escolas municipais de Maceió, conforme citadas na **Tabela 1** deste estudo, pode-se perceber que alguns aspectos são caracterizados e abordados da mesma forma em todas as escolas participantes. Dentre os aspectos abordados estão:

- Formação e experiência profissional;
- Equipe escolar;
- Políticas, ações e programas;
- Escola e gestão;
- Recursos financeiros e livros didáticos;
- Merenda escolar;

- Violência na escola;
- Ensino religioso.

Questionado sobre a questão 1: Você acha que alguma característica pessoal do diretores tais como: sexo, formação acadêmica e estado civil podem influenciar o seu desempenho profissional?

Observa-se nas falas dos gestores acerca da questão 1: a ênfase na formação acadêmica sim, corroborando com questões fortemente identificadas pela perspectiva computacional questão **Q-05** (Há quantos anos você obteve o nível de escolaridade assinalado anteriormente (até a graduação)?), questão **Q-23** (Houve conflito com o meu horário de trabalho para participar de mais atividades de desenvolvimento profissional) e questão **Q-24** (Não tinha disponibilidade de tempo para participar de mais atividades de desenvolvimento profissional). Sendo assim, investimentos na capacitação educacional dos gestores educacionais podem apresentar bons resultados quando direcionados e alinhados com programas estratégicos da rede de ensino.

Questionado sobre a questão 2: O seu tempo como diretor e/ou professor na escola em que atua pode influenciar na sua política administrativa?

Observa-se nas falas dos gestores acerca da questão 2 que o tempo de direção na escola pode contribuir no que se refere a dinâmica junto aos alunos e comunidade, e está ligada a questão de equipe escolar **questão Q-32** (Neste ano e nesta escola, como se deu a elaboração do projeto pedagógico? Extraída pela visão computacional, como existe um prazo determinado para a “mandato” como diretora. Uma proposta seria um plano de ação contemplando as ações durante o período do exercício, não se recomenda muitos anos à frente da direção, tem que existir um rodízio para se evitar saturação e conflitos desestimulando novos olhares frente ao cargo de direção.

Questionado sobre a questão 3: Você entende que o desempenho acadêmico da sua escola é fruto do trabalho cooperativo entre docentes, família do aluno e sociedade civil?

Quando questionado sobre o desempenho acadêmico da sua escola é fruto de um trabalho cooperativo, percebe-se que sim, é fruto de um trabalho cooperativo, principalmente, com a família. Tudo que acontece de positivo na escola é fruto de um trabalho em equipe. O papel da família é fundamental, devendo está integrada ao ambiente escolar. O protagonismo do trabalho da equipe escolar é evidenciado na questão **Q-41** (relativo à taxa de abandono

escolar) que indaga se a escola faz algo para combater o abandono escolar e na **Q-53** (eventos junto à comunidade). Isso é importante e tem relevância para influenciar o IDEB da escola visto que, a composição do índice é o fluxo e a aprendizagem. Portanto, reduzindo-se a evasão escolar há melhoria no indicador. Vale ressaltar que as questões **Q-29** e **Q-31**, que remete aos conselhos de classe e escolar, são identificadas pelos gestores escolares como importantes no trabalho de equipe.

Questionado sobre a questão 4: O tempo e as atividades desempenhadas pelos professores na escola são trabalhados de maneiras integradas? Por exemplo; projeto de laboratório, aulas de reforço, concomitante ao conteúdo dado em sala de aula, projetos junto à comunidade (p.e., reciclagem)

Observa-se nas falas dos gestores acerca da questão 4 que o tempo e as atividades desempenhadas pelos professores na escola são trabalhados de maneira integrada. A reunião de planejamento adotada como ação pela SEMED teve ampla aceitação junto aos diretores escolares revelou a pesquisa, contribuindo para a melhoria dos indicadores do IDEB. Essa informação é muito rica visto que diretamente pela visão computacional não foi capturada pelos algoritmos. O destaque na integração deve-se, dentre outras, às ações ao Horário de Trabalho Pedagógico Coletivo- HTPC e Horário de Trabalho Pedagógico Individual – HTPI, que consiste dentro da carga horária do professor, 1/3 é destinado por lei para planejamento, que foi observado e cumprido pela secretaria de educação. Essa implantação foi fruto da atual secretária que ao implantar resultou em melhorias aos salários da categoria docente. Houve, assim, monitoramento e efetividades dessas reuniões, alinhamento pedagógico, a partir de reuniões onde gestores (diretores, coordenadores pedagógicos) embasam as ações juntos à toda equipe.

Quando a gestão maior da rede de ensino municipal, verificou-se que o planejamento das atividades nas unidades escolares não estava sendo executado com a intensidade necessária, de pronto foram tomadas medidas para que esse item importante de gestão não fosse negligenciado. E isso foi bem citado nas respostas dos diretores das escolas pesquisadas. Já a realização de projetos específicos e fora da sala de aula, com práticas, parece ainda longe de ser um ponto corriqueiro nas escolas. As respostas traduzem a falta de implantação desse item importante nas atividades pedagógicas. Entretanto, verifica-se a mudança de paradigmas na medida em que passam a utilizar as reuniões periódicas de planejamento e discussão das atividades pedagógicas e não pedagógicas.

Questionado sobre a questão 5: Como se processa a comunicação entre os atores da escola (direção, professor, aluno e responsáveis) com foco no ensino e aprendizagem e relações interpessoais (respeito, simpatia e ISE (Inteligência Sócio Emocional))?

De acordo com falas dos gestores acerca da questão 5 que pela visão computacional não foi captada pelos algoritmos as questões norteadoras dessa questão, mas pela entrevista, obteve-se informações que o diálogo entre gestores e pais de alunos se dá por reuniões e comunicados, a ISE (Inteligência Sócio Emocional) foi citada fortemente como habilidades para resolução de conflitos e para o trabalho em equipe com harmonia. Aqui há necessidade de contextualizar a localização das escolas, geralmente, na franja do tecido urbano, em regiões de população com vulnerabilidade econômica, com problemas de infraestrutura pública e segurança. Os conflitos nessas escolas ocorrem e é preciso que a equipe esteja bem preparada, no que tange ao ECA e à inteligência emocional para contornar todas as situações; no fundo, acolher adequadamente à criança ao jovem.

Questionado sobre a questão 6: Quais as carências mais frequentes em sua escola e como elas influenciam no desenvolvimento das atividades escolares (merenda, espaço físico, material de consumo, didático e humano)?

De acordo com falas dos gestores acerca da questão **Q-06** a dimensão problemas da escola e dificuldade de gestão fica evidenciado pela visão computacional e pelos gestores as questões relativas à carência de pessoal de apoio pedagógico (supervisor, coordenador, orientador educacional) presente na questão **Q-70**, a ausência de recursos pedagógicos indicado na questão **Q-71**, interrupção das atividades escolares apresentada na questão **Q-72** e fechando a dimensão alta rotatividade do corpo docente referente a questão **Q-75**.

Cabe ressaltar que mesmo não estando contemplado no questionário SAEB diretor 2015 questões específicas acerca de infraestrutura, pois essa análise poderá ser identificada em questionário próprio sobre a escola que também é realizado pelo SAEB, foi cuidadoso e pertinente destacar esse fator, mediante a importância dada pelos diretores e que constava no Questionário Derivado do Especialista, item da questão **Q-06**.

A seguir, no **Gráfico 1** são apresentadas as palavras mais destacadas pelos diretores escolares que nortearam essa pesquisa.

a graduação?), **Q-23** (Houve conflito com o meu horário de trabalho para participar de mais atividades de desenvolvimento profissional), **Q-24** (Não tinha disponibilidade de tempo para participar de mais atividades de desenvolvimento profissional.), **Q-32** (Neste ano e nesta escola, como se deu a elaboração do projeto pedagógico?), **Q-41** (relativo à taxa de abandono escolar) a questão indaga se a escola faz algo para combater tal abandono e a **Q-53** (eventos junto à comunidade) isso é importante e tem relevância para influenciar o IDEB da escola, visto que, a composição do índice é o fluxo e aprendizagem. E, encerrando os fatores comuns pela visão humano computacional, as questões da dimensão problemas da escola e dificuldade de gestão ficam evidenciadas relativas à carência de pessoal de apoio pedagógico (supervisor, coordenador, orientador educacional) **Q-70**, a ausência de recursos pedagógicos indicado na **Q-71**, interrupção das atividades escolares apresentada na **Q-72** e fechado a dimensão alta rotatividade do corpo docente referente a **Q-75**.

5 CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS

Ao longo desta pesquisa de mestrado, apresentou-se uma contextualização geral sobre o tema: fatores que contribuem para melhoria do IDEB, também justificando e definindo a relevância deste estudo e apresentando seus objetivos, antecipando-se às necessidades de gestores educacionais frente a tomada de decisões, para obter a melhor alocação de recursos não apenas financeiros como também humanos, possibilitando desta forma, uma percepção na melhoria da qualidade do ensino básico no município de Maceió.

Uma boa gestão traz um clima favorável no ambiente escolar. A harmonia é contagiante e consegue agrupar todos para as metas das escolas. Não houve comentários negativos com relação à merenda escolar, considerada por muitos como fator importantíssimo para a permanência dos estudantes nas escolas, levando a melhorias nos indicadores de fluxo nos anos praticados nas escolas. O conselho de classe foi tido como muito importante, para os debates e engajamentos dos atores nas escolas. A premiação das escolas com melhores desempenhos é fator importante de motivação para todos, principalmente no alcançar de metas no IDEB. Não houve nenhuma citação sobre que a dificuldade de obtenção de índices no IDEB em relação a salários baixos. Algumas escolas tinham em seu interior cartazes falando de sua atuação no IDEB anterior, como resultado de sucesso. Nessas escolas, a participação de todos os atores parecia mais harmoniosa, com engajamentos e metas para sempre superar a anterior.

Em suma, na percepção dos gestores escolares, observa-se os seguintes atributos como relevantes para melhoria do IDEB: o nível de escolaridade do gestor, o processo de elaboração do projeto pedagógico da escola, medidas para redução das taxas de reprovação, medidas para minimizar as faltas dos alunos, melhoria na comunicação escola-família, a questão da merenda escolar, apoio das instâncias superiores, recursos financeiros para financiar os gastos, o guia didático do MEC e as escolhas dos livros didáticos pelos professores.

Pelos dados apresentados no item de reconhecimento da Rede de Educação de Maceió, verifica-se que, após tomadas de decisões importantes, o índice IDEB passa a ter outra inclinação na curva de desempenho, aproximando das metas estabelecidas pelo MEC e ultrapassando-as nas edições mais atuais. Alguma coisa parece ter acontecido. E isso foi detectado tanto pelo procedimento Máquina quanto o procedimento Humano. O planejamento no ambiente escolar parece ter sido uma ação diferencial para esse fim. Adicionam-se as ações relativas ao acolhimento dos estudantes e de suas famílias, das ações de provimento de merenda escolar, das ações de recuperação de prédios, dentre outras, e se determinam os requisitos que levaram à melhoria do IDEB. Desta forma, consolidada a metodologia de máquina para o

diagnóstico bem aproximado da realidade das unidades escolares, pode-se partir para a confecção de cenários, modificando-se aspectos dos descritores escolares, nos diversos seguimentos que compõem as escolas, e predizendo qual IDEB seria o indicado para essa nova e HTPI, isto sob uma perspectiva humana complementado por uma camada computacional repleta de 31 atributos fortes que complementa a visão sistêmica dos atores.

Apesar de todo esforço e apoio incondicional dos atores desta pesquisa, esse trabalho possui limitações e necessidades de melhorias que devem ser expressas para conduzir novos avanços em tempo oportuno, e ser estendido para um trabalho de doutorado mediante vastas possibilidades da pesquisa. São passíveis de melhoria a metodologia para criação de um questionário mais aderente aos diretores para evitar dubiedade de pensamentos em relação ao que se pretende investigar, observar o contexto educacional internacional e conseguir abster para a realidade brasileira e local, para dá melhores alternativas aos gestores, como se trata de uma abordagem de caráter inédita é preciso que o entrevistador detenha um arcabouço de conhecimentos relativos a área educacional e computacional para facilitar a identificação de novos fatores possíveis.

Diante do que foi alcançado no presente trabalho, sugere-se com intuito de aprimoramento dos resultados obtidos:

- i. Aprimorar os questionários para pesquisa *in loco*;
- ii. Acrescentar mais uma rodada junto aos diretores com os achados para validação e;
- iii. Criar um programa que possa através da IA e MDE apoiar os gestores educacionais.

6 REFERÊNCIAS

- ALVES e SOARES, **Contexto escolar e indicadores educacionais**: condições desiguais para a efetivação de uma política de avaliação educacional. Educação e Pesquisa, São Paulo, 2013.
- APPOLINÁRIO, F. **Metodologia Científica**: filosofia e prática da pesquisa. São Paulo: Cengage Learning, 2009.
- BAKER, R. S. J., Isotani, S., Carvalho, A. de. "**Mineração de dados educacionais: Oportunidades para o Brasil**" Revista Brasileira de Informática na Educação, 2011, v. 12, n. 2,
- BAKER, Ryan. **Data mining for education**. Internationalencyclopediaofeducation, v. 7, n. 3, p. 112-118, 2010.
- BARTOLOMEU, T.A. **Modelo de Investigação de acidentes do trabalho baseado na aplicação de tecnologias de extração de conhecimento**. Universidade Federal de Santa. 2002.
- BROOKE, N & CUNHA, M. A. A. **A avaliações externa como instrumento de gestão educacional nos estados**. Estudos & Pesquisas Educacionais Fundação Victor Civita, v. 2, p. 3-64, 2011.
- BUENO, Marina Fernandes. **A educação básica na política do mec: o plano de Desenvolvimento da Educação e o Índice de Desenvolvimento da Educação Básica em questão**. 2013. Disponível em: <http://www.joinpp.ufma.br/jornadas/joinpp2013/JornadaEixo2013/anais-eixo15-impasesedesafiosdaspoliticadeeducacao/pdf/aeducacaobasicanapoliticadomec-.pdf>. Acessado em: 10 mai. 18.
- CAFEZEIRO, I. **Ciência da Computação, Ciência da Informação, Sistemas de Informação: uma reflexão sobre o papel da informação e daab interdisciplinaridade na configuração das tecnologias e das ciências Computer Science, Information Science, Information Systems**: A., p. 111–133, 2016.
- CALIXTO et al., **Mineração de dados aplicada a educação: um estudo comparativo acerca das características que influenciam a evasão escolar**. Anais do XXVIII Simpósio Brasileiro de Informática na Educação, 2017.
- CANTERO, A. S; CARBONERA, T. D; FERREIRA, J. L. D. "**Educação e desenvolvimento: uma análise dos gastos com educação pública nos municípios do Paraná**" In: Anais do VIII Encontro de Produção Científica e Tecnológica, Campo Mourão, 21 a 25 de out. de 2013.
- CATARINA. **O QEdu é**. Disponível em: <<http://www.fundacaolemann.org.br/QEdu/>>. Acesso em: 10 de setembro de 2016. p. 3 – 13.
- CERVO, A. L.; BERVIAN, P. A.; SILVA, R. **Metodologia Científica**. 6. ed., São Paulo: Pearson Prentice Hall, 2007
- CHAPMAN, P. *et al.* (2000). **CRISP-DM 1.0 step-by-step data mine guide**. CRISP- DM Consortium.

CHIRINÉA e BRANDÃO, **O IDEB como política de regulação do Estado e legitimação da qualidade**: em busca de significados. Ensaio: Avaliação e Políticas Públicas em Educação, 2014.

COSTA, Evandro et al. **Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações**. Jornada de Atualização em Informática na Educação, v. 1, n. 1, p. 1-29, 2012.

CROZATTI, **Variáveis de influência no IDEB do ensino fundamental dos municípios paulista em 2017**. XXVI Congresso Brasileiro de Custos, 2019.

DA SILVA, J. L. M.; ALMEIDA, J. C. L. **Eficiência no gasto público com educação: uma análise dos municípios do Rio Grande do Norte**. Planejamento e Políticas Públicas, n. 39, 2012.

DASH, M., LIU, H. **Consistency-based search in feature selection**. Artificial Intelligence 151 (2003).

FACELI, A. C. Lorena, J. Gama, and A. C. P. L. F. de Carvalho, **Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina**. 2011.

FERREIRA, Stéfanny Barbara J. et al. **Avaliação educacional em larga escala: uma análise sobre os resultados do ideb e a visão dos gestores das escolas públicas do município de Salgueiro/PE**. Revista Opara, v. 8, n. 2, p. 26-41, 2018.

FISCHER, Rosa Maria Bueno. **Foucault e a análise do discurso em educação**. Cadernos de pesquisa, n. 114, p. 197-223, 2001.

FOUCAULT, M. **A Arqueologia do saber**. Rio de Janeiro: Forense, 1986.

FRASER, William J.; KILLEN, Roy. **Factors influencing academic success or failure of first-year and senior university students: do education students and lecturers perceive things differently?**. South African journal of education, v. 23, n. 4, p. 254-263, 2003.

FRIEDMAN, N.; GEIGER, D.; GOLDSZMIDT, M. **Bayesian network classifiers**. *Machinelearning*, Springer, 1997.

GUYON, I., ELISSEEFF, A. **An introduction to variable and feature selection**. *Journal of Machine Learning Research* 3 (2003).

IDEB. **Índice de Desenvolvimento da Educação Básica**. Disponível em: <http://ideb.inep.gov.br/>>. Acesso em: 10 de setembro de 2016.

IDEB. **Índice de Desenvolvimento da Educação Básica**. Disponível em: <http://ideb.inep.gov.br/>>. Acesso em: 10 de setembro de 2018.

INEP. **Censo escolar da educação básica 2018.**, 2018.

INEP. Prova Brasil. **Sistema de Avaliação da Educação Básica (Saeb)**. Disponível em: <http://prova-brasil.inep.gov.br/>>. Acessado em: 10 de setembro de 2018.

INEP/MEC. **Indicadores da Qualidade da Educação**. São Paulo: ação educativa, 2007.

HAIR, J. F., BLACK, W. C., BABIN, B. J., ANDERSON, R. E. e TATHAN, R. L. **Análise multivariada de dados**. Porto Alegre: 2009, Bookman Ed.

JANNUZZI, Paulo de Martino. **Indicadores socioeconômicos na gestão pública**. – 3. ed. rev. atual. – Florianópolis: Departamento de Ciências da Administração / UFSC.CAPES: UAB, 2014.

JÚNIOR, Olival de Gusmão Freitas et al. **Melhoria da gestão escolar através do uso de técnicas de mineração de dados educacionais: um estudo de caso em escolas municipais de Maceió**. RENOTE-Revista Novas Tecnologias na Educação, v. 17, n. 1, p. 296-305, 2019.

LIMA, R. A. F. et al. **Estratégias de seleção de atributos para detecção de anomalias em transações eletrônicas**. Dissertação (Dissertação em Ciência da Computação), Universidade Federal de Minas Gerais, p. 25. 2016.

LORENA, A. C. & CARVALHO, A. C. P. L. F. **As máquinas de vetores suporte, relatório interno**, Laboratório de Inteligência Computacional, ICMC/USP, São Carlos – SP, junho, 2002.

LEE, H. D. Seleção de atributos importantes para a extração de conhecimento de bases de dados. **Tese (Doutorado) do Curso de Pós-Graduação em Ciências de Computação e Matemática Computacional - ICMC - USP**, p. 182, 2005.

MARCONI, M. A.; LAKATOS, E. M. **Técnicas de Pesquisa**, 7. ed., São Paulo: Atlas, 2010.

MÁRQUEZ-VERA, C.; Morales, C. R.; Soto, S. V. **Predicting School Failure and Dropout by Using Data Mining Techniques**. IEEE Journal of Latin-American Learning Technologies, Vol. 8, no. 1, February, 2013.

MANZINI, Eduardo José. Considerações sobre a elaboração de roteiro para entrevista semi-estruturada. Colóquios sobre pesquisa em educação especial. Londrina: Eduel, v. 2010, 2003.

MOUSTAKAS, C. **Phenomenological research methods**. Thousand Oaks: Sage Publications, 1994.

NEVES, **IDEB: o caso de sucesso de uma escola do interior do estado do rio de janeiro**. Orientador: Prof. Dr. Lourival Batista de Oliveira Júnior. Juiz de Fora, 2012.

ROMERO, Cristobal et al. **Handbook of educational data mining**. CRC Press, 2011.

ROMERO, Cristobal; VENTURA, Sebastian. **Data mining in education**. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, v. 03, n. 1, p. 12-27, 2013.

ROMERO, Cristóbal; VENTURA, Sebastián. **Educational data mining: a review of the state of the art**. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), v. 40, n. 6, p. 601-618, 2010.

SANCHEZ, **O Ideb em pesquisas acadêmicas brasileiras entre 2007 e 2015**. Orientador: Sandra Maria Zákia Lian Sousa. São Paulo: s.n., 2017.

SEVERINO, A. J. **Metodologia do Trabalho Científico**. 23 ed., São Paulo: Cortez editora, 2008

SOUZA, A. A. **Organização, processos e tomada de decisão** – Florianópolis: Departamento de Ciências da Administração / UFSC; Brasília: CAPES: UAB, 2015.158p.

TASCA, M. **Contribuições ao Problema de Seleção de Atributos**. Diss. Ph. D. thesis, Universidade Federal Fluminense, Niterói, Rio de Janeiro, 2015.

TRIVIÑOS, A. N. S. **Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em educação**. São Paulo: Atlas, 1987.

VIDAL, Eloisa Maia; VIEIRA, Sofia Lerche. **Gestão educacional e resultados no Ideb: um estudo de caso em dez municípios cearenses**. Estudos em Avaliação Educacional, São Paulo, v. 22, p. 419-434, 2011.

WAHYUDDIN, **Perceptions and Actions of Educational Policy Makers regarding Parental Engagement in Education**. IRJE, 2018.

WEKA. **Waikato Environment for Knowledge Analysis**. Disponível em: <<https://www.cs.waikato.ac.nz/ml/weka/>>, Acesso em: jul. 2019.

WITTEN, I. H., FRANK, E.; HALL, M. A. **Data Mining: Practical Machine Learning Tools and Techniques**. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3rd edition. 2011.

APÊNDICES

APÊNDICE A – Solicitação de Aplicação de Questionário Para Fins de Trabalho Acadêmico Em Escolas da Rede Municipal de Maceió

Maceió, AL dezembro de 2018.

Prezado(a) Diretor(a),

Cumprimentando V.Sa., vimos solicitar autorização para proceder aplicação de questionário visando pesquisa sobre dados escolares em algumas Unidades Escolares da Rede Municipal de Maceió, especificamente a Escola _____ . Trata-se de pesquisa inserida no projeto de mestrado em Modelagem Computacional – Universidade Federal de Alagoas, o qual pertence ao curso de Ciências da Computação da citada universidade. O questionário é baseado **SAEB – 2015: Sistema de Avaliação da Educação Básica** **Questionário Diretor**.

O trabalho de pesquisa (mestrado) tem título provisório: *Predição de fatores que exercem influência no desempenho do IDEB: Aplicação de Mineração de Dados para auxiliar no Processo Decisório dos gestores em escolas da Rede Municipal de Educação Básica em Maceió*. Tem a orientação do Prof. Evandro de Barros Costa e coorientação do Prof. João Carlos Cordeiro Barbirato.

Contando com a autorização desta instituição, colocamo-nos à disposição para qualquer esclarecimento.

RG.1238160 SSP /AL



João Carlos Cordeiro Barbirato – Coorientador
SIAPE 1121793
CTEC/UFAL

APÊNDICE B – Sistema de Avaliação da Educação Básica



UNIVERSIDADE FEDERAL DE ALAGOAS/UFAL
 Programa de Pós-Graduação em Modelagem Computacional de Conhecimento
 Avenida Lourival Melo Mota, Km 14, Bloco 09, Cidade Universitária
 CEP 57.072-900 – Maceió – AL – Brasil
 Telefone: (082) 3214-1364



SAEB _ 2015: Sistema de Avaliação da Educação Básica Questionário Diretor

O presente questionário tem como objetivo coletar dados dos diretores da rede municipal de Maceió para confrontar com as saídas de algoritmos de aprendizagem de máquinas e devidas correlações com o IDEB.

ID n° _____

Questões	Assinale com (X) o grau de importância que atribui as questões				
1. QUAL É O SEU SEXO?	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro ()	Concordo Parcialmente ()	Concordo Totalmente ()
2. VOCÊ PODERIA NOS DIZER A SUA FAIXA ETÁRIA?	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro ()	Concordo Parcialmente ()	Concordo Totalmente ()
3. COMO VOCÊ SE CONSIDERA?	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro ()	Concordo Parcialmente ()	Concordo Totalmente ()
4. QUAL É O MAIS ALTO NÍVEL DE ESCOLARIDADE QUE VOCÊ CONCLUIU (ATÉ A GRADUAÇÃO)?	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro ()	Concordo Parcialmente ()	Concordo Totalmente ()
5. HÁ QUANTOS ANOS VOCÊ OBTVEU O NÍVEL DE ESCOLARIDADE ASSINALADO ANTERIORMENTE (ATÉ A GRADUAÇÃO)?	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro ()	Concordo Parcialmente ()	Concordo Totalmente ()
6. INDIQUE EM QUE TIPO DE INSTITUIÇÃO VOCÊ OBTVEU SEU DIPLOMA DE ENSINO SUPERIOR.	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro ()	Concordo Parcialmente ()	Concordo Totalmente ()
7. DE QUE FORMA VOCÊ REALIZOU O CURSO DE ENSINO SUPERIOR?	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro ()	Concordo Parcialmente ()	Concordo Totalmente ()
8. INDIQUE O CURSO DE PÓS-GRADUAÇÃO DE MAIS ALTA TITULAÇÃO QUE VOCÊ POSSUI.	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro ()	Concordo Parcialmente ()	Concordo Totalmente ()
9. INDIQUE A ÁREA TEMÁTICA DO CURSO DE PÓS-GRADUAÇÃO DE MAIS ALTA TITULAÇÃO QUE VOCÊ POSSUI.	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro ()	Concordo Parcialmente ()	Concordo Totalmente ()
10. COMO DIRETOR, QUAL É, APROXIMADAMENTE, O SEU SALÁRIO	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro ()	Concordo Parcialmente ()	Concordo Totalmente ()

BRUTO? (COM ADICIONAIS, SE HOVER).					
11. ALÉM DA ATIVIDADE COMO DIRETOR NESTA ESCOLA, VOCÊ EXERCE OUTRA ATIVIDADE QUE CONTRIBUI PARA SUA RENDA PESSOAL?	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro ()	Concordo Parcialmente ()	Concordo Totalmente ()
12. CONSIDERANDO TODAS AS ATIVIDADES QUE VOCÊ EXERCE (DENTRO E FORA DA ÁREA DE EDUCAÇÃO), QUAL É, APROXIMADAMENTE, O SEU SALÁRIO BRUTO? (COM ADICIONAIS, SE HOVER).	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro ()	Concordo Parcialmente ()	Concordo Totalmente ()
13. QUAL É A SUA CARGA HORÁRIA DE TRABALHO SEMANAL, COMO DIRETOR, NESTA ESCOLA?	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro ()	Concordo Parcialmente ()	Concordo Totalmente ()
14. VOCÊ ASSUMIU A DIREÇÃO DESTA ESCOLA POR MEIO DE:	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro ()	Concordo Parcialmente ()	Concordo Totalmente ()
15. POR QUANTO TEMPO VOCÊ TRABALHOU COMO PROFESSOR(A) ANTES DE SE TORNAR DIRETOR(A)?	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro 0	Concordo Parcialmente ()	Concordo Totalmente ()
NESTA ESCOLA, O ENSINO RELIGIOSO: 109. É de presença obrigatória.	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro 0	Concordo Parcialmente ()	Concordo Totalmente ()
110. Segue uma religião específica.	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro 0	Concordo Parcialmente ()	Concordo Totalmente ()
111. Nesta escola há atividades alternativas para os estudantes que não queiram participar das aulas de ensino religioso.	Discordo Totalmente ()	Discordo Parcialmente ()	Neutro 0	Concordo Parcialmente ()	Concordo Totalmente ()

Classifique **por ordem de prioridade da mais importante para a menos importante** as 10 (dez) questões dentre as selecionadas na coluna anterior

Caro diretor (a), gostaria de acrescentar alguma outra questão que não fora contemplada por este questionário, e que julga ser relevante?

Questão (os) inserida (s):

Muito obrigado por sua participação nesta pesquisa!

APÊNDICE C – Script Escrito em Python 3 utilizado para gerar nuvens de palavras das entrevistas dos diretores

```

import nltk
import
gensimimport nltk
from nltk.probability import FreqDist import
pandas as pd
from wordcloud import WordCloud
import matplotlib.pyplot as plt
nltk.download('stopwords') texto = ""

f = open ('texto.txt', 'r')
for line in f:
    texto += line
f.close()

stop_words = nltk.corpus.stopwords.words("portuguese")

clean_text = []
for i in texto.splitlines(): clean_text += [i.strip()]
    for i in
gensim.utils.simple_preprocess(i) if i not in stop_words]
texto_sem_stopwords = ""
for a in clean_text:
    texto_sem_stopwords = texto_sem_stopwords + " " + a
    # remover palavras específicas
stop_words_new = ['segue', 'cidade', 'viu', 'escola', 'tudo', 'grave', 'dá', 'menos',
    'seção',
    'identificar', 'civil', 'gente', 'porque', 'então', 'né', 'aqui',
    'tá', 'aí', 'fazer',
    'sim', 'vai', 'questão', 'ser', 'acho', 'pode',
    'sempre', 'vezes', 'muita', 'disso', 'existe', 'toda', 'melhor',
    'vem', 'ainda', 'disse', 'ter', 'outro', 'onde', 'dia', 'pouco',
    'quer', 'dessa', 'ver', 'muitas', 'dentro', 'consegue',
    'sabe', 'faz',
    'junto', 'graças', 'deus', 'assim', 'ano', 'deve',
    'grande', 'podem',
    'parte', 'fica', 'certo', 'tras', 'nesse', 'chegar', 'momento',
    'grande',
    'desse', 'vou', 'podem', 'somente', 'meio',
    'possa', 'feito', 'tipo', 'aberta',
    'sexo', 'têm', 'falando', 'nesse', 'semana',
    'vou', 'ano', 'causa', 'diz',
    'primeiro', 'assim', 'bom', 'processa',
    'resolver', 'quanto', 'dizer', 'posso',
    'cozinha', 'influencia', 'bem', 'nada', 'traz',
    'manhã', 'ajudar', 'melhorar', 'limpeza',
    'dificuldade', 'lado', 'todo', 'falta', 'poder',
    'lar', 'agora', 'coisa', 'lá']

clean_text_new = []
for i in texto_sem_stopwords.splitlines():
    clean_text_new += [i.strip() for i in gensim.utils.simple_preprocess(i) if i not
in stop_words_new]

```

```

texto_sem_stopwords_new = ""
for a in clean_text_new:
    texto_sem_stopwords_new = texto_sem_stopwords_new + " " + a

texto_sem_stopwords_new
wordcloud = WordCloud(max_font_size=100,width = 1520, height =
535).generate(texto_sem_stopwords_new) plt.figure(figsize=(16,9))
plt.imshow(wordcloud)
plt.axis("off")
plt.show()

all_words = texto_sem_stopwords_new.lower().split() freq =

nlk.FreqDist(all_words)

df = pd.DataFrame(data=freq.items())

# ordena por quantidade
df = df.sort_values(1, ascending=False)
#exibe df.head(55).reset_index()
freq.plot(30, cumulative=False, title="Frequência das 30 palavras mais repetidas") #
gráfico com as 30 palavras mais comuns
freq.most_common(30) # lista com as 50 palavras que mais se repetem
# lista os primeiros itens
df.head()
# lista todos os últimos df.tail()
# salva em um diretório formato txt separado por espaço df.to_csv(r'c:\temp\pandas.txt',
header=None, index=None, sep=' ', mode='a')
#texto.txt

```