

1ª Etapa: Compreensão do domínio. Esta etapa foca no conhecimento dos objetivos do projeto e, então, converte-se esse conhecimento em uma definição de problema de mineração de dados e um plano de projeto preliminar com a intenção de alcançar os objetivos. 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 do projeto, nesse caso, é encontrar os fatores (atributos) responsáveis pela melhoria do desempenho escolar, visando auxiliar no processo decisório dos gestores educacionais para adotar medidas de melhoria do IDEB. Este estudo fará a análise dos dados educacionais de 80 escolas públicas municipais de Maceió dos anos iniciais do ensino fundamental. 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 (**Tabela 1**).

Tabela 1. Metas para a educação básica de Alagoas.

	IDEB Observado					Metas					
	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)

2ª Etapa: Entendimento dos dados. Os dados podem vir de diversas fontes e ter diversos formatos. Esta é a etapa de familiarização com os dados do problema e identificação da qualidade dos mesmos, bem como pela construção das primeiras percepções e indagações sobre o que os dados podem mostrar (PÁDUA E SOUSA, 2018). As bases de dados usadas nesse estudo foram disponibilizadas abertamente pelo portal do INEP. Segundo o INEP (2016), o questionário do aluno dos anos iniciais do ensino fundamental consiste de 58 itens, distribuídos em 6 (seis) categorias conforme o **Quadro 1**.

Quadro 1. Estrutura básica do questionário do aluno.

Categoria	Descrição
Caracterização sociodemográfica	Sexo, cor, raça e idade
Informações socioeconômicas	Capital econômico
Capital social	Convívio, formação e atitude dos pais ou responsáveis na educação do aluno
Capital cultural	Hábitos de leitura e gestão do tempo
Trajetória escolar	Tempo de permanência na escola, reprovação e abandono
Atitudes em relação a estudos específicos	Atitudes do aluno e do professor em relação ao estudo de português e matemática, somados a um item sobre uso da biblioteca ou sala de leitura

Fonte: Adaptado de INEP (2016)

3ª Etapa: Preparação dos dados. Esta etapa 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. 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, remoção de dados duplicados. Coletou-se os dados educacionais das 80 escolas selecionadas no portal do INEP. Tanto para a coleta, quanto para o pré-processamento dos dados utilizou-se o MS Excel 2003. Consultou-se um especialista em gestão educacional para selecionar as questões mais relevantes da prova Brasil. Segue a **Tabela 2** com os atributos escolhidos e sua representação na ferramenta WEKA.

Tabela 2. Questões e seus correspondentes atributos na ferramenta WEKA.

Questões da Prova Brasil	Atributos (WEKA)
Sexo	Masculino, Feminino
Qual a sua idade?	8 anos, 9 anos, 10 anos...
Na sua casa possui computador?	Possui_Computador
Até que série sua mãe ou a mulher responsável por você estudou?	Escolaridade_Mae
Até que série seu pai ou o homem responsável por você estudou?	Escolaridade_Pai
Seus pais ou responsáveis incentivam você a estudar?	Incentiva_a_Estudar
Seus pais ou responsáveis incentivam você a fazer o dever de casa?	Incentiva_Deber_Casa
Seus pais ou responsáveis incentivam você a ler?	Incentiva_a_Ler
Seus pais ou responsáveis incentivam você a ir à escola e não faltar às aulas?	Incentiva_ir_escola
Você lê livros em geral?	Ler_Livro
Você costuma frequentar bibliotecas?	Freq_Biblioteca
Desde a primeira série em que tipo de escola você estudou?	Escola_Estudou
Você já foi reprovado?	Reprovacao
Você já abandonou a escola durante o período de aulas e ficou fora da escola o resto do ano?	Evasao
Você faz o dever de casa de língua portuguesa?	Dever_Portugues
O professor corrige o dever de casa de língua portuguesa?	Corrige_Portugues
Você faz o dever de casa de matemática?	Dever_Matematica
O professor corrige o dever de casa de matemática?	Corrige_Matematica
Você utiliza a biblioteca ou sala de leitura da sua escola?	Biblioteca_Salaleitura
Censo	Infraestrutura

Fonte: Dos autores.

Nesta etapa alguns atributos passaram por um processo de discretização, ou seja, foram transformados em outros valores que correspondem à mesma informação. Essa estratégia é usada para descobrir mais informações sobre os dados, uma vez que a variável contínua não permite isso. Como exemplo, pode-se citar a *feature* idade que corresponde uma faixa de idade para verificar a distorção idade-série. Ao se identificar essa distorção cria-se um novo campo na tabela de dados, isso permite verificar o impacto da distorção idade-série no IDEB.

4ª Etapa: Modelagem. Nesta etapa, várias técnicas de modelagem são selecionadas e aplicadas. Tipicamente, existem diversas técnicas para o mesmo tipo de problema de mineração. No entanto, há algumas que dependem do objetivo desejado. A escolha do pacote WEKA para realização do trabalho deu-se em virtude de sua facilidade de uso, bem como da disponibilização de diversos algoritmos que implementam técnicas de *data mining*. Nesta etapa buscou-se priorizar duas técnicas de utilização dos algoritmos, a saber: a técnica de mineração por regressão linear simples e a técnica de mineração por árvore de decisão (J48).

Para a realização dos testes foi usado o algoritmo de Regressão Linear Simples (função *SimpleLinearRegression* da ferramenta WEKA), o qual utilizou apenas o atributo classe IDEB e o atributo no qual está sendo realizada sua correlação, removendo no WEKA todos os atributos e mantendo apenas estes dois. Realizou-se este

teste com cada atributo individualmente e obteve-se os modelos de regressão e valores de correlação observado na **Tabela 3**.

Tabela 3. Modelo de regressão e coeficiente dos 20 atributos mais relevantes.

Atributo	Coefficiente de correlação
Incentiva dever de casa	0.5477
Média de incentivo	0.5102
Idade de 14 anos	0.4473
Escola que estudou	0.4301
Evasão	0.4043
Incentiva ir à escola	0.3776
Corrige o dever de matemática	0.3467
Escolaridade da mãe	0.3355
Incentiva a leitura	0.3307
Média da escolaridade dos pais	0.3263
Idade de 13 anos	0.2912
Incentiva a estudar	0.2739
Sexo masculino	0.2703
Sexo feminino	0.2703
Média dever corrigido	0.2559
Escolaridade do pai	0.2418
Idade de 10 anos	0.2346
Infraestrutura	0.1870
Idade de 8 anos	0.1753
Frequência a Biblioteca	0.1705

Fonte: Dos autores.

Nessa análise foram considerados todos os atributos e realizado a sua influência em relação ao IDEB. Com esses resultados, pode-se analisar a influência de cada atributo na construção do IDEB, visando avaliar uma possível dependência de Y (atributo classe IDEB) em X (demais atributos), os valores do coeficiente de correlação apresentam o quão boa é a correlação. No modelo de regressão é observado se a correlação é positiva ou negativa, ou seja, se o fator contribui positivamente ou negativamente no IDEB.

A **Figura 2** mostra a representação gráfica da correlação entre os atributos “X: MEDIA_INCENTIVO” (linha horizontal) e “Y: IDEB_2011” (linha vertical). Nela é possível perceber a correlação entre esses dois atributos. É possível observar os 80 pontos que representam cada uma das instâncias (escolas), em sua grande maioria localizadas na parte superior direita, evidenciando o quanto o incentivo aos estudos ajuda na melhoria do IDEB.

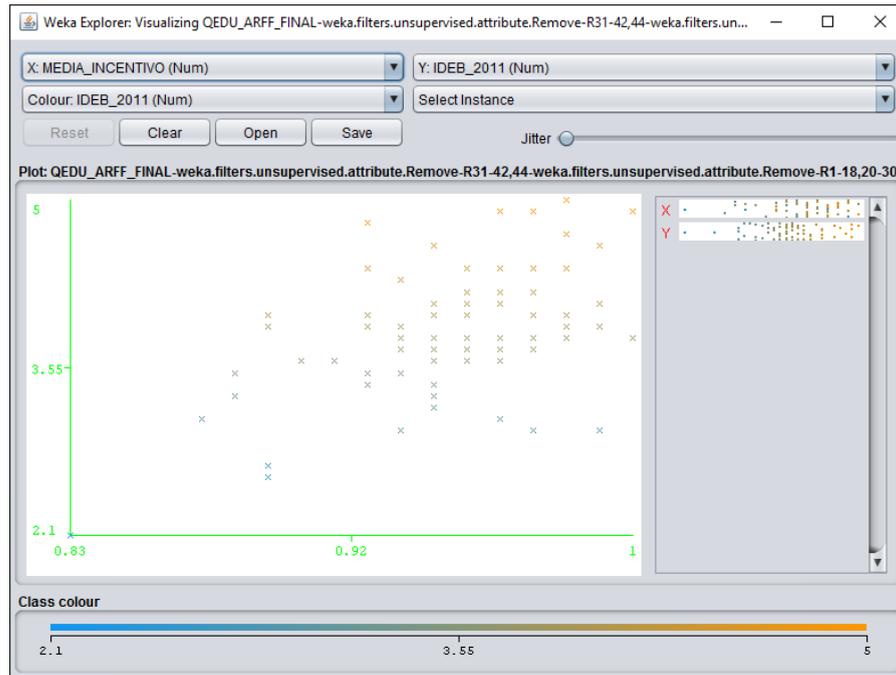


Figura 2. Visualização do algoritmo de regressão linear no WEKA.

Fonte: Dos autores.

Para a realização dos testes com o algoritmo de classificação J48, utilizou-se a base de dados com todos os atributos, porém com uma pequena alteração no atributo classe, que foi necessário ser categorizado, esta etapa de discretização se faz necessária para se trabalhar com o algoritmo J48. É importante ressaltar que os algoritmos de classificação requerem que os dados estejam na forma de atributos categorizados. Mantiveram-se as configurações padrões do algoritmo J48 na ferramenta WEKA. Com esse resultado obteve-se uma árvore que pode ser visualizada graficamente pela **Figura 3**.

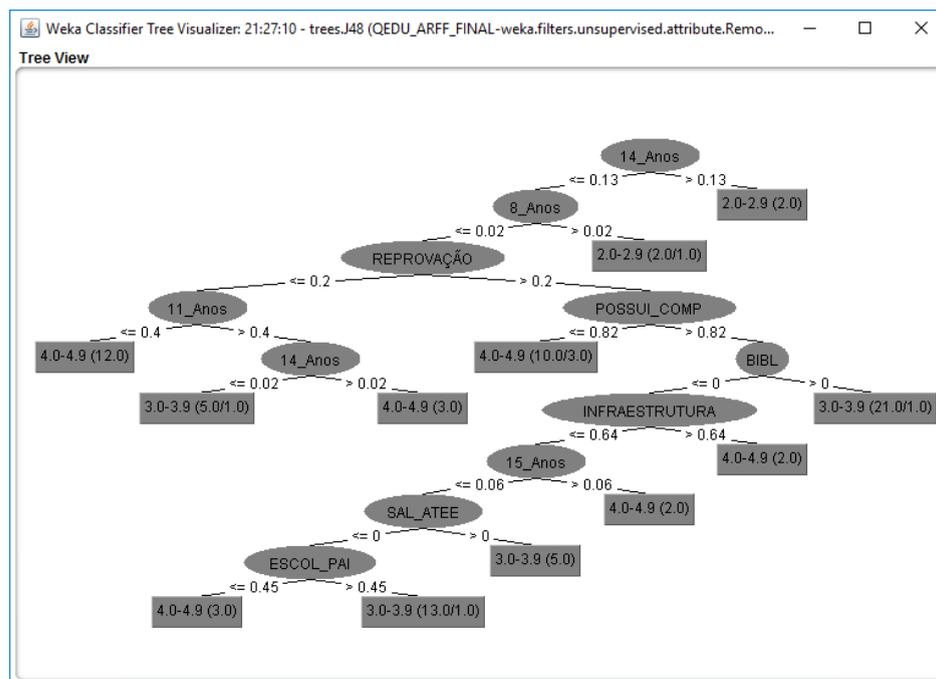


Figura 3. Visualização da árvore de decisão J48 do WEKA.

Fonte: Dos autores.

Analisando a **Figura 3**, percebeu-se que infraestrutura não tem muita relevância no IDEB. Notou-se, ainda, através da análise da árvore de decisão, que a melhoria do IDEB está diretamente ligada as ações pedagógicas dos professores em sala de aula e da gestão escolar.

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 (PÁDUA E SOUSA, 2018). Esta é a etapa no qual os conhecimentos encontrados são interpretados e utilizados em processo decisório. Com base nas análises efetuadas, pode-se apontar para as conclusões do **Quadro 2**.

Quadro 2. Dimensões e atributos relevantes para o IDEB segundo os algoritmos.

Dimensão	Atributo	Reflexão sobre o IDEB
Discente	Sexo	Percebe-se que os discentes do sexo feminino têm um destaque maior na melhoria do IDEB.
	Distorção entre idade-série	Os alunos que estão na faixa correta com relação à idade e série (9 a 12 anos) contribuem positivamente no IDEB. Todavia, os alunos que se encontram na faixa errada com relação à idade e série (8 anos ou de 13 anos ou mais) contribuem de forma negativa no IDEB.
	Trajetória escolar (evasão)	Alunos com altos índices de evasão contribuem negativamente no IDEB.
	Informações socioeconômicas	Observa-se que alunos que estudaram anteriormente em escolas particulares contribuem de maneira expressiva na melhoria do IDEB.
	Escolaridade dos pais	Percebe-se que a escolaridade dos pais exerce influência positiva no IDEB. Todavia, apesar dos pais terem uma escolaridade maior que as mães, ficou constatado que a mãe possui maior influência no aprendizado do aluno.
	Incentivo aos estudos	Quanto à análise dos incentivos aos estudos, ficou claro que é o fator que mais pesa para obtenção de um melhor IDEB.
Docente	Compromisso do docente	Observa-se que a cobrança maior por parte dos professores nas disciplinas de língua portuguesa e matemática contribui positivamente no IDEB, sendo que a disciplina de matemática tem um peso ainda maior.
	Formação do docente	Nota-se que a formação adequada do docente interfere de forma positiva no IDEB. Como formação adequada considera-se docente com curso superior de licenciatura na disciplina lecionada.
Gestor	Estilo de gestão	Observa-se que o estilo de gestão tem impactado no IDEB. Um gestor que atua como um líder democrático e que realiza uma gestão baseado em resultados contribui significativamente para a melhoria do IDEB da sua unidade escolar.
Infraestrutura	Infraestrutura em geral (biblioteca, sala de leitura, etc.)	Ficou constatado que leituras de livros em geral, frequência à biblioteca e a existência de biblioteca, sala de leitura e outras questões de infraestrutura em geral nas escolas, não exercem influência considerável no IDEB. Também ficou constatado que à posse de computadores não causa grande interferência no IDEB.

Fonte: Dos autores.

6ª Etapa: Distribuição. A criação do modelo não é o fim do projeto. Em alguns casos, o conhecimento adquirido precisará ser organizado e apresentado de uma forma que

possa ser usado (PÁDUA E SOUSA, 2018). Conforme dito antes, o resultado do presente trabalho pode ser aplicado tanto em outros novos projetos, que porventura podem aparecer como nos objetivos inicialmente traçados. São os gestores educacionais os responsáveis pelo projeto que decidem se e de que maneira o conhecimento será utilizado. Esse estudo de caso serviu para perceber o quanto cada atributo influencia na determinação da classe IDEB, influenciando, assim na melhoria dos índices nas escolas públicas do município de Maceió.

Com essa metodologia, pode-se obter informações importantes no âmbito educacional, que possa ser útil no processo decisório de gestores educacionais, visando à melhoria da gestão escolar.

4. Conclusão

Os resultados advindos do processamento dos dados na ferramenta WEKA proporciona uma espécie de provocação, pautando discussões interessantes na área pedagógica. Por exemplo, ao se considerar que foi constatado na análise dos dados tratados pelo WEKA que leituras de livros em geral, frequência à biblioteca e a existência de biblioteca, sala de leitura e outras questões de infraestrutura em geral nas escolas, não têm influência considerável no IDEB, suscita-se minimamente uma discussão entre o ter e o usar. Uma pergunta pode ser feita à equipe pedagógica (junto às escolas): como as unidades escolares estão utilizando esses equipamentos? A resposta pode ser surpreendente, a de que o uso pode não ser o mais adequado. Uma outra pergunta, então, poderia melhor qualificar o dado a ser processado. Por exemplo, ao invés de se trabalhar com ter livros ou biblioteca disponíveis, obter informações da quantidade de livros que um estudante lê de fato. Pode-se descobrir distorções importantes no fazer pedagógicos desses ambientes, e que passam despercebidos pela gestão (da secretaria e da unidade escolar).

Utilizou-se técnicas de regressão linear simples e de árvore de decisão (J48) para identificar os fatores que influenciavam o IDEB das escolas da Rede Municipal de Maceió. Buscou-se demonstrar que a MDE pode auxiliar os gestores a traçar ações que possam reduzir a taxa de evasão e reprovação escolar, por exemplo, melhorando com isso o IDEB na sua unidade escolar.

Os resultados obtidos com a aplicação da MDE poderão servir como indicativos para a direção das escolas públicas municipais no tocante ao auxílio ao processo decisório, principalmente no que se refere a investimentos, visando à melhoria no padrão do aprendizado, como consequência o aumento do IDEB.

Verifica-se que a ferramenta WEKA pode ser utilizada na MDE e contribuir no tratamento da informação a partir de atributos preestabelecidos na comunicação ou socialização dos conhecimentos oriundos dos questionários do INEP, ou de outro banco de dados disponível.

Neste trabalho junto às escolas da Rede Municipal de Maceió no tocante à análise de dados IDEB da educação básica, constatou-se uma lacuna epistemológica e que em futuros estudos poderia melhorar a capacidade e o domínio da ferramenta WEKA, podendo contribuir na melhoria do tratamento dos dados educacionais das escolas públicas. Como trabalhos futuros, pretende-se aplicar outros algoritmos de mineração de dados, expandir o escopo do estudo para outros níveis de escolaridade e para outros cenários como o estadual e nacional.

5. Referências

- BEZERRA, C.; SCHOLZ, R.; ADEODATO, P.; PONTES, T.; SILVA, I. (2016). **Evasão Escolar: Aplicando Mineração de Dados para Identificar Variáveis Relevantes**. V Congresso Brasileiro de Informática na Educação (CBIE 2016). Anais do XXVII Simpósio Brasileiro de Informática na Educação (SBIE 2016).
- CHAPMAN, P. *et al.* (2000). **CRISP-DM 1.0 step-by-step data mine guide**. CRISP-DM Consortium.
- COELHO, V. C.; COSTA, J. P. C. L.; SOUSA, D. C. R.; CANEDO, E. D.; SILVA, D. G.; SOUSA JÚNIOR R. T. (2015). **Mineração de dados educacionais para identificação de barreiras na utilização da educação a distância**. Escola Nacional de Administração Pública. ENAP. Ministério do Planejamento, Orçamento e Gestão, Brasília -DF.
- FILHO, R.L.C.S.; ADEODATO,P.J.L. (2017). **Solução de Mineração de Dados para Avaliação do Ensino Médio dos Institutos Federais a partir do Censo Escolar e do Enem**. 15ª Conferências Ibero-Americanas WWW/Internet e computação aplicada 2017. Vilamoura, Algarve, Portugal.
- IDEB. (2016). **Índice de Desenvolvimento da Educação Básica**. Disponível em: <http://ideb.inep.gov.br/>>. Acesso em: 10 de setembro de 2016.
- INEP/MEC. (2007). **Indicadores da Qualidade da Educação**. São Paulo: ação educativa.
- INEP. (2016). **Prova Brasil**. Sistema de Avaliação da Educação Básica (Saeb). Disponível em: <http://provabrasil.inep.gov.br/>>. Acessado em: 10 de setembro de 2016.
- MANHÃES, L. M. B. (2015). **Predição do desempenho acadêmico de graduandos utilizando mineração de dados educacionais**. Tese (Doutorado em Engenharia de Sistemas e Computação. Universidade Federal do Rio de Janeiro. Rio de Janeiro.
- MÁRQUEZ-VERA, C.; Morales, C. R.; Soto, S. V. (2013). **Predicting School Failure and Dropout by Using Data Mining Techniques**. IEEE Journal of Latin American Learning Technologies. Vol. 8, no. 1, February.
- MEC. IDEB. (2018). **Índice de Desenvolvimento da Educação Básica**. Disponível em: <http://ideb.inep.gov.br/>>. Acessado em: 10 de setembro de 2018.
- NASCIMENTO, R. L. S.; Cruz Junior, G. G; Fagundes, R. A. A. (2018). **Mineração de Dados Educacionais: Um estudo sobre indicadores da educação em bases de dados do INEP**. Revista RENOTE. Novas Tecnologias na Educação. CINTED. UFRGS.
- PÁDUA, A. F. L. O.; SOUSA, F. A. (2018). **Metodologia CRISP-DM: Potencialidades na descoberta do conhecimento em dados educacionais**. XVI Congresso Internacional de Tecnologia na Educação. 2018. Disponível em:<<http://www.pe.senac.br/congresso/anais/2018/SENAC/pdf/poster>>. Acessado em: 25 de março de 2019.
- PAIVA, R.; Bittencourt, I. I.; Pacheco, H.; Da Silva, A. P.; Jacques, P.; Isotani, S. (2012). **Mineração de dados e a gestão inteligente da aprendizagem: desafios e direcionamentos**. Instituto de Computação. Universidade Federal de Alagoas (UFAL).