

## Mineração de Dados Educacionais nos Resultados do ENEM de 2015

Augusto Simon<sup>1</sup>, Sílvio César Cazella<sup>2,3</sup>

<sup>1</sup> Programa de Pós-Graduação em Educação (PPGEDU) – Universidade Federal do Rio Grande do Sul (UFRGS) – Porto Alegre – RS – Brasil

<sup>2</sup> Programa de Pós-Graduação em Informática na Educação (PPGIE) – Universidade Federal do Rio Grande do Sul (UFRGS) – Porto Alegre – RS – Brasil

<sup>3</sup> Programa de Pós-Graduação em Ensino na Saúde (PPGENSAU) – Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA) – Porto Alegre – RS – Brasil.

augusto.simon@ufrgs.br, silvioc@ufcspa.edu.br

**Abstract.** *The objective of this study is to generate a predictive model of the average performance indicator in nature sciences and its technologies of high school students from public data referring to the Brazilian National High School Examination (ENEM) of 2015. Such a proposal is justified in an alarming scenario of low performance of primary school students in Brazil, according to the PISA (Program for International Student Assessment) 2015 data. The data mining technique chosen is the decision tree and the algorithm used is j48, through Waikato Environment for Knowledge Analysis (WEKA).*

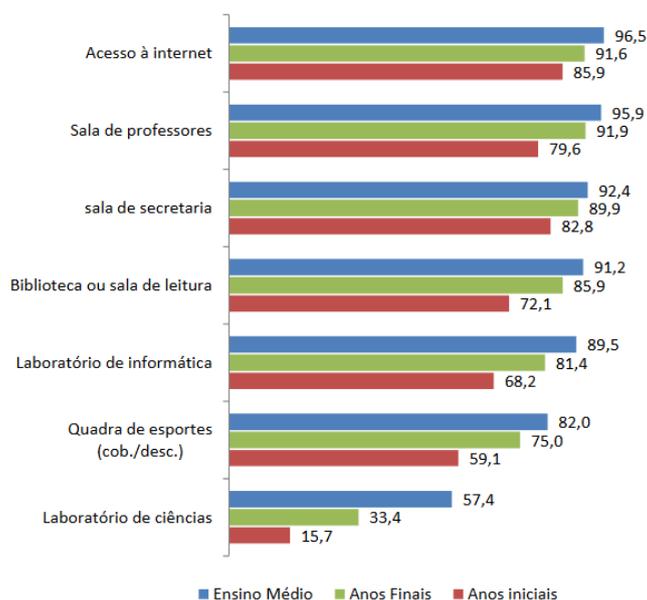
**Resumo.** *O objetivo deste estudo consiste em gerar um modelo preditivo do indicador de desempenho médio em ciências da natureza e suas tecnologias dos alunos de escolas do ensino médio a partir dos dados públicos referentes ao Exame Nacional do Ensino Médio (ENEM) de 2015. Tal proposta justifica-se em um cenário alarmante de baixo desempenho dos alunos do ensino básico no Brasil, de acordo com os dados de 2015 do PISA (Programme for International Student Assessment). A técnica de mineração de dados escolhida é a árvore de decisão e o algoritmo utilizado é o j48, através do software Waikato Environment for Knowledge Analysis (WEKA).*

### 1. Introdução

A análise dos indicadores educacionais brasileiros dos últimos anos não permite sensação de conforto, muito menos oferece motivos para comemorações. Segundo os dados de 2015 do PISA (Programme for International Student Assessment), o programa internacional de avaliação de estudantes, desenvolvido e coordenado pela OCDE (Organização para Cooperação e Desenvolvimento Econômico), o desempenho dos estudantes do ensino básico no Brasil, nas três áreas avaliadas pelo programa, ciências, leitura e matemática, está abaixo da média dos estudantes em países membros da organização [OECD 2016]. Dos 70 países avaliados pelo programa, considerando os resultados para a área de ciências, o Brasil posicionou-se na 63<sup>a</sup> colocação. Reforçando esse mal-estar, o relatório indica que, em ciências, o valor médio de desempenho dos

estudantes brasileiros tem se mantido estável, sem apresentar mudanças estatisticamente significativas, desde o início da avaliação na área, em 2006.

Estudos como a edição de 2016 do Censo Escolar da Educação Básica [INEP 2017], realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP)<sup>1</sup>, proporcionam reflexões sobre a infraestrutura das instituições de ensino básico no território nacional em relação aos baixos índices apresentados pelo PISA de 2015. No contexto do ensino de ciências, o censo identificou que o percentual de estudantes brasileiros que possuem acesso a laboratórios de ciências no ensino básico é expressivamente inferior à estrutura prevista no PNE (Plano Nacional de Educação). Conforme observado na Figura 1, entre os recursos avaliados pelo censo, os laboratórios de ciências possuem os mais baixos índices de presença nas escolas.



**Figura 1 - Percentual de alunos dos anos iniciais, dos anos finais e do ensino médio por recurso disponível na escola. Fonte: INEP (2017)**

Os resultados apresentados pelo PISA de 2015 e os problemas de infraestrutura evidenciados pelo Censo Escolar da Educação Básica de 2016 demandam séria atenção. Investigações relacionando o perfil socioeconômico e de infraestrutura das escolas de ensino básico no Brasil associadas ao desempenho dos seus alunos em exames como o PISA atualmente podem contar com técnicas presentes na área de Mineração de Dados Educacionais (MDE).

Este estudo possui o objetivo de gerar um modelo preditivo do indicador de desempenho médio em ciências da natureza e suas tecnologias dos alunos de escolas do ensino médio a partir dos dados públicos referentes ao Exame Nacional do Ensino Médio (ENEM) de 2015. Tais dados são fornecidos pelo INEP, agrupados por escolas, como subsídio a estratégias que visem a melhoria da qualidade da educação<sup>2</sup>.

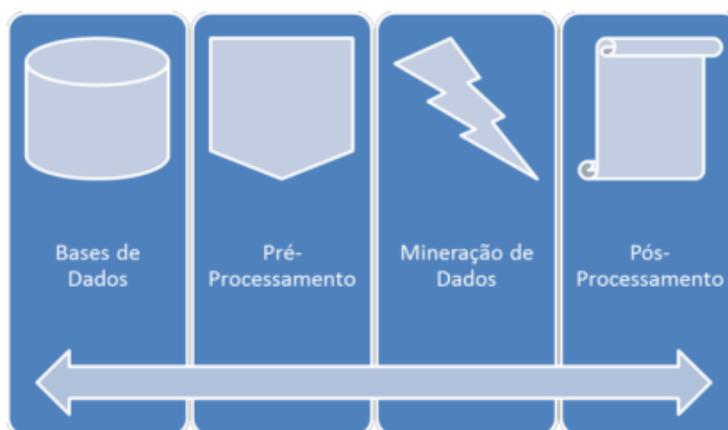
<sup>1</sup> <http://portal.inep.gov.br>. Acesso em 16/08/2017

<sup>2</sup> <http://portal.inep.gov.br/web/guest/enem-por-escola>. Acesso em 18/08/2017.

Iniciando a partir da introdução ao tema de pesquisa, este estudo é apresentado da seguinte maneira: a seção 2 trata da MDE em geral; na seção 3 são discutidos estudos que apresentem aplicações da MDE relacionadas ao ENEM; na seção 4 são explicitados os materiais e métodos utilizados, com a descrição da etapa de pré-processamento de dados para a mineração; na seção 5 é descrita a aplicação da técnica de mineração de dados escolhida, a árvore de decisão, e do algoritmo utilizado, o j48, através do software WEKA; e na seção 6 são discutidos os resultados obtidos, as conclusões deste estudo e suas possibilidades para trabalhos futuros.

## 2. Mineração de Dados Educacionais

A MDE consiste na aplicação da Mineração de Dados (MD) aplicada ao contexto educacional. A MD é a etapa principal do processo de Descoberta de Conhecimento em Bases de Dados (DCDB, sigla em português para o termo em inglês KDD, *Knowledge Discovery in Database*) [Silva e Silva 2014]. Os autores abordam a DCDB através da Figura 2, que apresenta o processo em suas quatro etapas: A base de dados, o pré-processamento, a MD e o pós-processamento.



**Figura 2 – Processo de Descoberta de Conhecimento em Banco de Dados.**  
Fonte: [Silva e Silva 2014]

Silva e Silva (2014) explicam que, no DCDB, podem ser utilizadas uma ou mais bases de dados, disponibilizadas de forma estruturada ou não estruturada. A integração das bases utilizadas ocorre na etapa de pré-processamento, bem como a correção de valores ausentes, ruidosos, inconsistentes e redundantes. Na etapa da MD, segundo Silva e Silva, é extraído algum tipo de conhecimento a partir da execução de algum algoritmo da tarefa utilizada. Essas tarefas podem ser: modelo preditivo, análise de agrupamento e regras de associação. A análise dos resultados obtidos é realizada na quarta e última etapa, o pós-processamento.

Segundo Silva e Silva (2014), variados estudos são encontrados na literatura acadêmica utilizando técnicas da MD no contexto educacional. É citada, por exemplo: a aplicação de árvores de decisão para o fornecimento de informações em apoio a gestores de instituições de ensino na avaliação do desempenho dos seus alunos; e a clusterização de dados na identificação de fatores relacionados à satisfação de alunos em instituições de ensino, também em apoio à tomada de decisão de seus gestores.

### 3. Trabalhos Relacionados

A proposta de utilização da MDE no contexto dos dados abertos sobre o ENEM não é exclusiva deste estudo e tampouco recente na literatura acadêmica nacional. Esta seção discutirá projetos relacionando a MDE e o ENEM em áreas de avaliação como: processos seletivos no ensino superior; agrupamentos de regiões para tomada de decisões no contexto administrativo e educacional; fatores socioeconômicos que diminuem o desempenho dos alunos no exame; e matérias em que os alunos tiveram mais dificuldade no exame.

Martinago (2005) investiga a base de dados sobre os vestibulandos da UFPR de dezembro de 2003. O autor utiliza o software WEKA, a técnica de mineração do tipo árvore de decisão e os algoritmos de classificação J48 e J48.PART a fim de traçar o perfil do vestibulando da Universidade Federal do Paraná, campus Curitiba. Em sua base de dados constam informações coletadas no questionário socioeducacional preenchido pelos candidatos no momento da inscrição, os dados do cadastro geral dos candidatos contendo o registro das notas obtidas pelo candidato nas provas e na redação, a opção pelo ENEM, a nota do ENEM, a média das notas do candidato e o resultado do vestibular para o aluno. Em relação à nota do Exame Nacional do Ensino Médio, O autor obteve: resultados gerais, como estudantes que obtiveram notas entre 4,208 pontos e 10 (máxima) obtiveram sucesso no vestibular para os onze cursos mais concorridos; e resultados específicos, como a aprovação dos candidatos que obtiveram nota no ENEM entre 4,208 a 8,969 pontos e prestaram vestibular para o curso de Comunicação Social.

Furtado (2014) realiza o agrupamento dos municípios do estado do Rio de Janeiro a partir das notas de matemática do ENEM de 2011 utilizando um algoritmo de agrupamento espacial, o SKATER, com a técnica de mineração do tipo árvore geradora mínima. O algoritmo permite a descoberta e o agrupamento de municípios cujos resultados de seus estudantes foi semelhante, o que pode auxiliar na tomada de decisões no contexto administrativo e educacional. A prova de matemática foi uma escolha arbitrária do autor, que utilizou os resultados do exame como estudo de caso para sua proposta. Furtado apresenta uma nova abordagem para a utilização do algoritmo SKATER, que considera as medidas estatísticas dos seus atributos. O autor conclui que a sua proposta resulta em agrupamentos de melhor qualidade em relação à abordagem tradicional, contribuindo assim com estudos para os quais ela se aplique, como os realizados no contexto administrativo e educacional.

Silva, Morino e Sato (2014) utilizam dados dos questionários socioeconômicos preenchidos por alunos participantes da edição de 2010 do ENEM, oriundos das capitais da região Sudeste do Brasil. Essas informações, especificamente as respostas de quatro perguntas, são relacionadas com os resultados de desempenho desses alunos. Os autores adotam a técnica de mineração baseada em regras de associação, utilizando diferentes parametrizações do algoritmo A Priori. Em suas conclusões, são identificados fatores que diminuem o desempenho dos alunos, como a renda familiar baixa, a escolaridade dos pais de nível primário e a quantidade alta de pessoas que moram com o estudante.

Braga e Drummond (2017) investigam a utilização da linguagem R como ferramenta para mineração de dados abertos governamentais. Como estudo de caso empregam os dados do ENEM de 2013. As autoras utilizam técnicas de mineração por agrupamento, extração de regras através de árvore de decisão e mineração de texto.

Após a demonstração do emprego de diferentes técnicas de mineração através da linguagem R, as conclusões do estudo limitam-se a ressaltar termos mais frequentes apresentados pelas provas de 2013 do ENEM e as matérias em que os alunos tiveram mais dificuldade: química, física e biologia.

Stearns et al. (2017) utilizam dados socioeconômicos do ENEM de para comparar a capacidade de generalização de dois métodos de agrupamento por árvore de decisão para a nota dos estudantes no exame. Foram utilizados os resultados das notas de matemática devida a sua alta variância. Os modelos utilizados foram os métodos de regressão baseados em boosting de árvore de decisão, o AdaBoost e o Gradient Boosting. Os melhores resultados foram obtidos com o Gradient Boosting. Os autores concluem que os indicadores socioeconômicos podem explicar um viés na pontuação dos alunos, sendo possível prever a nota do aluno com valores de métricas (MAE, do inglês *Mean Absolute Error*) adequadas.

A proposta apresentada neste artigo se difere das discutidas nesta seção. Busca-se, neste estudo, relacionar fatores socioeconômicos das escolas do ensino médio de todo o território nacional com o indicador desempenho médio em ciências da natureza e suas tecnologias dos seus alunos. Para tanto, será construído um modelo preditivo utilizando a técnica de mineração de árvore de decisão, através do algoritmo j48, com software WEKA.

#### 4. Materiais e Métodos

Os dados fornecidos pelo INEP apresentam valores médios de proficiência dos alunos agrupados por escolas. É disponibilizado um conjunto de dados para cada área avaliada, sendo elas: ciências da natureza e suas tecnologias; ciências humanas e suas tecnologias; linguagens, códigos e suas tecnologias; matemática e suas tecnologias; e redação. As escolas presentes nos conjuntos disponibilizados são aquelas em que pelo menos 10 dez alunos estiveram em fase de conclusão do ensino médio regular e participaram do ENEM de 2015 e que possuíam pelo menos 50% de alunos participantes do exame. Como delimitação do estudo foi definida a utilização dos resultados para a área de ciências da natureza e suas tecnologias. Tal escolha foi motivada pelos baixos índices de presença de laboratórios de ciências nas escolas evidenciados pelo Censo Escolar da Educação Básica de 2016 [INEP 2017].

O software utilizado para a MDE foi o Waikato Environment for Knowledge Analysis (WEKA), em sua versão 3.8.1. O WEKA consiste em um projeto de software livre iniciado em 1992 com o objetivo de prover uma coleção compreensível de algoritmos de aprendizado de máquina e ferramentas de processamento de dados para pesquisadores e praticantes [HALL et al. 2009]. Segundo os autores, atualmente o WEKA é aceito largamente tanto na academia e em empresas, com uma comunidade de usuários ativa, contando com mais de um milhão e quatrocentos mil downloads desde abril do ano de 2000. Sua implementação é realizada em linguagem Java, o que também contribui para sua manutenção e modificação.

Para a análise dos dados do ENEM por escola, optou-se pela técnica de mineração de árvore de decisões, através do algoritmo j48. Segundo Bhargava et al. (2013), a árvore de decisão é um sistema de suporte a decisão baseado em grafos de decisão em formato de árvore. É utilizado para aprender funções de classificação que

indicam o valor de uma variável dependente através de dados de variáveis independentes fornecidos. Para os autores, árvores de decisão consistem nas abordagens mais poderosas em mineração de dados e descoberta de conhecimento. Tais ferramentas, dessa forma, são altamente efetivas em muitas áreas, como a mineração de texto, extração de informação, aprendizado de máquina e reconhecimento de padrões. Entre os algoritmos de árvores de decisão, o utilizado por Bhargava et al. (2013) é o j48, também evidenciada por Peña-Ayala (2014) como bastante popular em aplicações educacionais.

### Pré-Processamento do Dados

Cada tabela fornecida pelo INEP dos dados do ENEM por escola, para cada área avaliada, apresenta 25 colunas, representando uma informação, ou variável, cada, divididas em três categorias: características da escola, indicadores e desempenho. A Tabela 1 apresenta cada uma dessas 25 variáveis e, conforme indicado na coluna Id, as nove selecionadas para serem utilizadas na mineração de dados a ser realizada.

**Tabela 1 – Variáveis disponibilizadas pelo ENEM por escola. Fonte: o autor.**

<b>Categoria</b>	<b>Id</b>	<b>Coluna</b>	<b>Opções de Resposta</b>
<b>Características Escola</b>		Código Da Entidade	String
		Nome Da Entidade	String
		Sigla Da UF	String
		Nome Município	String
	<b>1</b>	<b>Dependência Administrativa</b>	<b>Privada</b>
			<b>Municipal</b>
			<b>Estadual</b>
			<b>Federal</b>
	<b>2</b>	<b>Localizacao</b>	<b>Urbana</b>
			<b>Rural</b>
<b>3</b>	<b>Número De Alunos No Censo</b>	<b>Valor Numérico</b>	
	Porte Da Escola	De 1 A 30 Alunos	
		De 31 A 60 Alunos	
		De 61 A 90 Alunos	
		Maior Que 90 Alunos	
	Número De Participantes No Enem Por Escola	Valor Numérico	
<b>4</b>	<b>Taxa De Participação</b>	<b>Valor Percentual</b>	
	Número De Participantes Com Necessidades Especiais	Valor Numérico	
<b>Indicadores</b>	<b>5</b>	<b>Indicador De Permanência Na Escola</b>	<b>Menos De 20%</b>
			<b>De 20% A 40%</b>
			<b>De 40% A 60%</b>
			<b>De 60% A 80%</b>
			<b>80% Ou Mais</b>
	<b>6</b>	<b>Indicador De Nível Socioeconômico</b>	<b>Muito Baixo</b>
			<b>Baixo</b>
			<b>Médio Baixo</b>
			<b>Médio</b>
			<b>Médio Alto</b>
		<b>Alto</b>	
		<b>Muito Alto</b>	
	Faixa - Indicador De Formação Docente	Maior Que 50%	
		Até 50%	
<b>7</b>	<b>Indicador De Formação Docente</b>	<b>Valor Percentual</b>	

<b>Desempenho Ciências Da Natureza E Suas Tecnologias</b>	<b>8</b>	<b>Taxa De Aprovação</b>	<b>Valor Percentual</b>
		Taxa De Reprovação	Valor Percentual
		Taxa De Abandono	Valor Percentual
		Média Dos 30 Melhores Alunos Da Escola	Valor Numérico
	<b>9</b>	<b>Média Escola</b>	<b>Valor Numérico</b>
		Percentual Alunos Nível 1	Valor Percentual
		Percentual Alunos Nível 2	Valor Percentual
		Percentual Alunos Nível 3	Valor Percentual
	Percentual Alunos Nível 4	Valor Percentual	
	Percentual Alunos Nível 5	Valor Percentual	

A variável Média Escola indica o desempenho médio dos alunos da escola na área de ciências da natureza e suas tecnologias. O INEP fornece esse dado como um valor numérico, cujo valor máximo é inferior a 750 pontos para a área. Para a tarefa de mineração de dados, entretanto, por se tratar do valor de referência a ser utilizado, foi necessário convertê-lo em categorias limitadas. Na Tabela 2 são exibidas as categorias divididas em apenas quatro valores possíveis, utilizando a escala prevista pelo INEP (2016).

**Tabela 1 – Categorias para os valores de Média Escola. Fonte: o autor.**

Coluna	Opções de Resposta
<b>Média Escola</b>	I (menor de 450)
	II (de 450 a 549,99)
	III (de 550 a 649,99)
	IV (de 650 a 749,99)

Os dados do ENEM por escola, fornecidos pelo site do INEP, são disponibilizados através de uma planilha no formato .xlsx. Tal formato não é suportado pelo software WEKA, a ser utilizado neste estudo, sendo convertido para um formato de dados separados por vírgula, o .csv. Para a análise foi salvo então um arquivo com as nove colunas selecionadas e 15599 linhas, com a identificação das colunas e as escolas cujos resultados são informados.

## 5. Mineração de Dados

Através do software WEKA, a análise por árvore de decisão utilizando o algoritmo j48 foi executada utilizando-se a opção cross-validation, com o valor para fold igual a dez e com a variável dependente Média Escola. A árvore obtida apresentou como variável independente principal para classificação a Tipo Escola, que se divide entre os quatro valores: privada, federal, estadual e municipal. A variável hierarquicamente seguinte é a Nível socioeconômico. A Figura 3 e a Figura 4 apresentam, respectivamente, os dois primeiros níveis hierárquicos da árvore e a árvore completa, ressaltando-se as divisões entre os tipos de escolas.



Figura 3 – Principais variáveis da árvore de decisão. Fonte: o autor.

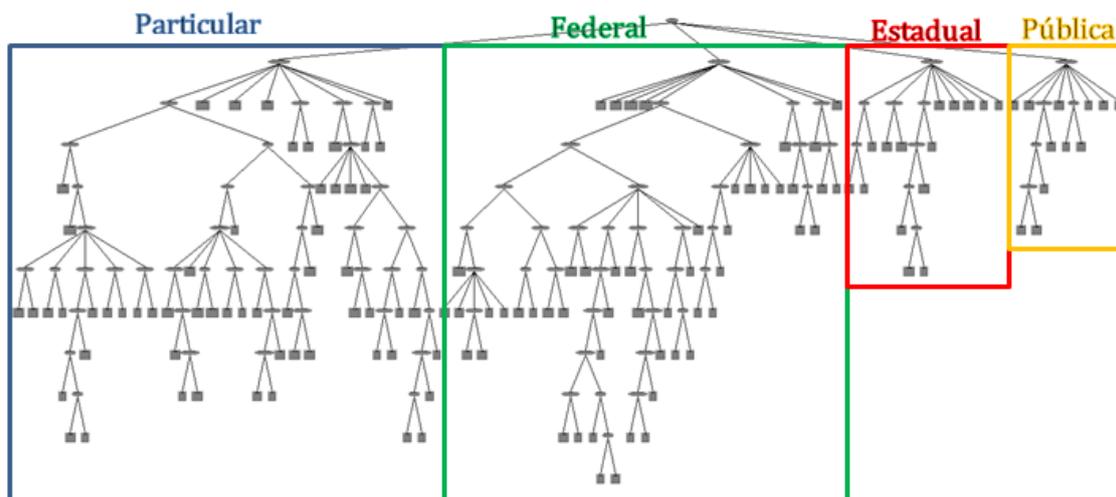


Figura 4 – Árvore de decisão dividida por tipo de escola. Fonte: o autor.

A árvore de decisão foi capaz de identificar corretamente 77,02% instâncias das 15998 inseridas. A Tabela 3 apresenta os valores de acurácia para cada classe da variável média escola. Os valores de precisão apresentam resultados interessantes para as classes I a III. Entretanto a classe IV, para médias de escolas entre 650 e 749,99 pontos, apresenta valor consideravelmente inferior aos das outras classes. A Tabela 4, que apresenta matriz de classificação para cada classe, indica que, na classe IV, de nove ocorrências, apenas uma foi corretamente classificada.

Tabela 3 – Acurácia detalhada por classe. Fonte: o autor.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,491	0,038	0,647	0,491	0,558	0,511	0,912	0,58	III
0,897	0,493	0,798	0,897	0,845	0,447	0,786	0,853	II
0,508	0,05	0,703	0,508	0,59	0,523	0,874	0,623	I
0,015	0,001	0,111	0,015	0,027	0,04	0,815	0,043	IV
0,77	0,351	0,758	0,77	0,758	0,467	0,819	0,772	Weighted Avg.

Tabela 4 – Matriz de classificação. Fonte: o autor.

a	b	c	D	classified as
953	982	2	3	a = III
464	9574	626	5	b = II
7	1431	1485	0	c = I
49	16	0	1	d = IV

## 6. Conclusões

Os resultados obtidos com a árvore de decisão utilizando-se o algoritmo j48 se mostraram promissores para três das quatro classes de valores da variável dependente Média Escola. Os ramos com maior variação hierárquica são justamente aqueles de maior diversidade de valores para a variável Média Escola. Segundo o algoritmo J48, as variáveis independentes mais importantes hierarquicamente para a decisão sobre a classe da variável independente, respectivamente e em ordem decrescente, são: no primeiro nível o Tipo Escola; no segundo nível o Nível Socioeconômico; no terceiro

nível em diante as variáveis alternaram de posição em função dos valores dos níveis superiores.

Sobre o desempenho médio dos alunos em ciências da natureza e suas tecnologias, destaca-se que os valores dos grupos III e IV, acima de 550 pontos, ocorrem nas escolas: privadas, apenas no nível socioeconômico muito alto; federais, nos níveis muito alto, alto e médio alto; estaduais, apenas no nível muito alto; e nas municipais, apenas no nível médio alto.

A continuação deste estudo prevê o enriquecimento dos dados utilizados com as informações de infraestrutura das escolas do ensino básico brasileiro. Através dos dados abertos do Censo Escolar da Educação Básica [INEP 2017], disponibilizados pelo INEP, a presença de laboratórios de ciências nas escolas poderá ser analisada como elemento adicional na nova árvore de decisão a ser desenvolvida.

## 6. Referências

- BHARGAVA, N. et al. (2013) Decision Tree Analysis on J48 Algorithm for Data Mining. *International Journal of Advanced Research in Computer Science and Software Engineering*, v. 3, n. 6, p. 1114–1119.
- BRAGA, L. C. C.; DRUMMOND, I. N. (2017) Uma Abordagem de Mineração Descritiva Aplicada a Dados Abertos Governamentais Empregando a Ferramenta R. *Anais do Computer on the Beach*, p. 51–60.
- FURTADO, V. M. (2014) Agrupamento de Conjuntos de Instâncias: Uma Aplicação ao ENEM. Dissertação. Programa de Pós-graduação em Engenharia de Sistemas e Computação. Universidade Federal do Rio de Janeiro.
- HALL, M. et al. The WEKA data mining software. (2009) *ACM SIGKDD Explorations Newsletter*, v. 11, n. 1, p. 10.
- INEP. (2016) DIVULGAÇÃO ENEM 2015 POR ESCOLA. Brasília - DF. Disponível em:  
<[http://download.inep.gov.br/educacao\\_basica/enem/enem\\_por\\_escola/2015/apresentacao\\_enem\\_por\\_escola\\_2015.pdf](http://download.inep.gov.br/educacao_basica/enem/enem_por_escola/2015/apresentacao_enem_por_escola_2015.pdf)>. Acesso em: 14 ago. 2017.
- INEP. (2017) CENSO ESCOLAR 2016 - Notas Estatísticas. Brasília – DF. Disponível em:  
<[http://download.inep.gov.br/educacao\\_basica/censo\\_escolar/apresentacao/2017/apresentacao\\_censo\\_escolar\\_da\\_educacao\\_basica\\_2016.pdf](http://download.inep.gov.br/educacao_basica/censo_escolar/apresentacao/2017/apresentacao_censo_escolar_da_educacao_basica_2016.pdf)>. Acesso em: 31 mar. 2017.
- MARTINHAGO, S. (2005) Descoberta de Conhecimento sobre o Processo Seletivo da UFPR. Dissertação. Programa de Pós-Graduação em Métodos Numéricos em Engenharia. Universidade Federal do Paraná.
- OECD. (2016) PISA 2015 - Resultados Principais Brasil. Disponível em:  
<<http://www.oecd.org/pisa/PISA-2015-Brazil-PRT.pdf>>. Acesso em: 31 mar. 2017.
- PEÑA-AYALA, A. (2014) Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, v. 41, n. 4, p. 1432–1462.

SILVA, L. A.; MORINO, A. H.; SATO, T. M. C. (2014) Prática de Mineração de Dados no Exame Nacional do Ensino Médio. Anais dos Workshops do Congresso Brasileiro de Informática na Educação, v. 3, n. 1, p. 651–660.

SILVA, L. A.; SILVA, L. (2014) Fundamentos de Mineração de Dados Educacionais. Anais dos Workshops do Congresso Brasileiro de Informática na Educação, v. 3, n. 1, p. 568 - 581.

STEARNS, B. et al. (2014) Prevendo Desempenho dos Candidatos do ENEM Através de Dados Socioeconômicos. Anais do Congresso da Sociedade Brasileira de Computação. p. 2522-2530.