



Educação em Revista

ISSN: 0102-4698

ISSN: 1982-6621

Faculdade de Educação da Universidade Federal de Minas Gerais

Fonseca, Stella Oggioni da; Namen, Anderson Amendoeira
MINERAÇÃO EM BASES DE DADOS DO INEP: UMA ANÁLISE EXPLORATÓRIA
PARA NORTEAR MELHORIAS NO SISTEMA EDUCACIONAL BRASILEIRO
Educação em Revista, vol. 32, núm. 1, 2016, Janeiro-Março, pp. 133-157
Faculdade de Educação da Universidade Federal de Minas Gerais

DOI: 10.1590/0102-4698140742

Disponível em: <http://www.redalyc.org/articulo.oa?id=399362312007>

- Como citar este artigo
- Número completo
- Mais informações do artigo
- Site da revista em redalyc.org

redalyc.org

Sistema de Informação Científica Redalyc

Rede de Revistas Científicas da América Latina e do Caribe, Espanha e Portugal

Sem fins lucrativos acadêmica projeto, desenvolvido no âmbito da iniciativa
acesso aberto

MINERAÇÃO EM BASES DE DADOS DO INEP: UMA ANÁLISE EXPLORATÓRIA PARA NORTEAR MELHORIAS NO SISTEMA EDUCACIONAL BRASILEIRO

Stella Oggioni da Fonseca*
Universidade do Estado do Rio de Janeiro (UERJ)

Anderson Amendoeira Namen**
Universidade do Estado do Rio de Janeiro (UERJ)

RESUMO: O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) possui bases de dados provenientes de levantamentos estatísticos e avaliativos em diversos níveis e modalidades de ensino, incluindo a Educação Básica, cuja avaliação é efetuada pelo Sistema de Avaliação da Educação Básica (Saeb). Com o intuito de analisar o Ensino Fundamental público brasileiro, o Saeb é composto de uma avaliação denominada Prova Brasil, que, por intermédio da aplicação de testes e questionários aos alunos, professores e diretores, torna possível a extração de informações relevantes para identificação de caminhos visando à melhoria da qualidade do ensino. Mediante o desafio de explorar esta base, constituída de um grande volume de dados, este estudo aplica o processo de Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in Databases* – KDD) com o intuito de identificar fatores que relacionam o perfil de professores que lecionam Matemática com a proficiência obtida por seus alunos. São apresentados os passos deste processo no contexto desta aplicação, explicitando, principalmente, a Mineração de Dados (*Data Mining*). Por fim são tecidos comentários sobre os padrões descobertos.

Palavras-chave: Educação Básica. Aprendizagem de Matemática. Mineração de Dados.

<http://dx.doi.org/10.1590/0102-4698140742>

* Mestre em Modelagem Computacional e Doutoranda do Programa de Pós-Graduação em Modelagem Computacional pela Universidade do Estado do Rio de Janeiro – Instituto Politécnico. E-mail: sfonseca@iprj.uerj.br

** Doutor em Engenharia de Sistemas e Computação pela COPPE/UFRJ. Professor do Programa de Pós-Graduação em Modelagem Computacional do Instituto Politécnico da Universidade do Estado do Rio de Janeiro e da Universidade Veiga de Almeida. Email: aanamen@iprj.uerj.br

DATA MINING ON INEP DATABASES: AN INITIAL ANALYSIS AIMING TO IMPROVE BRAZILIAN EDUCATIONAL SYSTEM

ABSTRACT: Anísio Teixeira National Institute for Education Research and Studies (INEP) provides databases from the evaluations of various levels and modalities of education. The Elementary Education Assessment System, called SAEB, provides mechanisms for evaluation of elementary education. One of these mechanisms is Prova Brasil, which includes the application of tests and questionnaires to students, teachers and school principals. Due to the large data volume involved, information extraction and its analysis is not a simple task. The general process of converting raw data into useful information is called Knowledge Discovery in Databases (KDD). It is applied here in order to identify factors that relate profiles and their influence - positively and negatively - on students' Mathematics learning. KDD steps are presented, with focus on the Data Mining stage. Some discovered patterns are analysed and discussed in this paper's conclusion.

Keywords: Elementary Education. Mathematics Learning. Data Mining.

INTRODUÇÃO

O século XXI se caracteriza por um cenário em que as atividades humanas se organizam em torno da geração, recuperação e utilização do conhecimento. Esse novo contexto, proveniente da revolução tecnológica e do processo de globalização das relações econômicas e culturais, fez com que a educação fosse revisitada, em relação às suas finalidades e aos seus meios, com vistas a garantir a formação de pessoas capazes de enfrentar os múltiplos desafios do futuro (MARTUCCI, 2000). Dentro dessa perspectiva, há um empenho das organizações governamentais, em diversos âmbitos, em efetuar ações que possibilitem a melhoria do ensino em todos os níveis, desde a alfabetização até os níveis mais altos de especialização. Sistemas de avaliação visando à elaboração de diagnósticos para a melhoria do ensino público e à melhoria do gerenciamento dos recursos disponíveis passaram a ocupar papel de destaque na agenda política educacional (BAUER, 2012).

No Brasil, a entidade responsável por promover estudos, pesquisas e avaliações sobre o sistema educacional é o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP. Vinculado ao Ministério da Educação, esse instituto objetiva subsidiar a formulação e implementação de políticas públicas e produzir informações confiáveis aos gestores, educadores e público em geral (BRASIL, 2011). O INEP é

atuante em múltiplos níveis e modalidades de ensino, compreendendo, por exemplo, avaliações a respeito da Educação Básica, exames aplicados aos concluintes do Ensino Médio e levantamento de informações acerca das instituições de Ensino Superior.

As avaliações realizadas pelo INEP, além de mensurarem o aprendizado dos conteúdos propostos, procuram avaliar também os vários fatores que possam afetar a qualidade do ensino dos discentes. Dentre esses fatores, destaca-se o nível socioeconômico e cultural no qual os alunos se enquadram, sendo considerado um aspecto essencial para compreender o desempenho educacional (ALVES; SOARES, 2009). Informações acerca de professores, diretores e escolas também são coletadas, por intermédio de questionários, permitindo o registro de dados como formação profissional, práticas pedagógicas e gerenciais e perfil socioeconômico e cultural dos profissionais envolvidos com a educação.

A aplicação de testes e questionários gera um grande volume de dados e estes, por sua vez, são armazenados em diferentes arquivos que podem ser obtidos no *site* do referido instituto (via *download*). Desse modo, as bases de dados disponibilizadas constituem um importante acervo, contendo uma grande quantidade de informações para gestores, pesquisadores, educadores e a comunidade em geral (RIGOTTI; CERQUEIRA, 2004).

O estudo realizado por Araújo e Luzio (2005) enfatiza a importância dessas bases de dados, afirmando que “tem-se a convicção de que os dados produzidos a partir da avaliação educacional podem subsidiar, de forma efetiva, ações em prol de melhorias na qualidade do aprendizado e das oportunidades educacionais oferecidas à sociedade brasileira” (ARAÚJO; LUZIO, 2005, p. 9).

Soares (2004, p.18), por sua vez, afirma que “a escola dos alunos de pior desempenho é pior que a dos alunos de melhor desempenho”, considerando e englobando no termo “escola” não somente questões como recursos pedagógicos disponíveis, infraestrutura e segurança, mas também o corpo de professores e funcionários que a constitui. O autor evidencia desse modo que o aprendizado está diretamente relacionado com o meio em que o aluno está inserido.

É notório, portanto, que há um conjunto de informações embutido nos dados coletados pelo INEP que possibilita a compreensão dos fatores que afetam a qualidade do aprendizado. Contudo, a indagação a ser feita é se, de fato, esses dados estão norteando melhorias no processo ensino-aprendizagem. Questiona-se se esse rico banco de

dados tem sido utilizado pela comunidade educacional para a extração de informações correlacionadas com o desempenho dos estudantes e se tem sido feita uma análise profunda desses dados.

Obviamente, para a formulação de ações governamentais e um mapeamento da situação do sistema educacional, a resposta a essas questões deveria ser positiva. No entanto, segundo Horta Neto (2010), apesar de os gestores apoiarem e reconhecerem a importância das avaliações, colocando-as em posição de destaque, elas ainda não influenciam como deveriam os processos e medidas adotados pelas secretarias de educação. Tais dificuldades estariam relacionadas com a divulgação dos resultados, limitada a relatórios que apontam somente dados numéricos, sem a disponibilização de informações complementares acerca dos problemas educacionais apontados pela avaliação e de possíveis soluções para superá-los. Cerdeira, Almeida e Costa (2014) analisam os efeitos adversos das avaliações externas quando percebidas essencialmente como um ranqueamento escolar, o que acarretaria na aplicação de excessivos testes focando somente em melhor desempenho (e não na melhoria do processo de ensino-aprendizagem), na submissão de testes sem avaliar o trabalho cotidiano escolar e no estreitamento das práticas curriculares, o que reduziria a autonomia docente e escolar. Paula e Moreira (2014) corroboram essa visão, afirmando que professores e escolas assumem o papel de objetos e não de sujeitos do processo de avaliação já que, de modo superficial e leviano, simplesmente são responsabilizados pelos resultados alcançados pelos estudantes.

Por outro lado, conforme discutido por Machado (2012), avaliações censitárias, como a Prova Brasil, foco deste estudo, fornecem dados que podem nortear melhorias quanto à qualidade do ensino. Segundo a autora:

As avaliações externas, especialmente a Prova Brasil, [...], fornecem dados que, se apropriados de forma consistente, podem revigorar os contornos da escola pública que realiza a sua função social na sociedade democrática de garantir o ensino-aprendizagem para todos os seus alunos. Porém, a avaliação da escola e a reflexão sobre a sua realidade não podem se esgotar nelas, que podem ser tomadas como ponto de partida para a trajetória da escola rumo à sua avaliação institucional, que não pode prescindir de uma autoavaliação. (MACHADO, 2012, p.79).

Ademais, as informações referentes a avaliações censitárias subsidiaram diversas pesquisas na área de educação. Gewehr (2010), por exemplo, enfatiza a importância das bases de dados do INEP e seu potencial de modificar as práticas com que os gestores e

professores conduzem o cotidiano escolar. Em seu trabalho, que foca nas escolas públicas do Paraná, concluiu-se que um alto percentual de docentes reviu suas práticas após a implementação da Prova Brasil. Souza e Oliveira (2010), por sua vez, apresentam alguns movimentos promissores que buscam, em alguns estados da federação, articular os resultados avaliativos em larga escala a processos de autoavaliação, com a perspectiva de focalizar não apenas a escola, mas todas as instâncias do sistema no processo avaliativo.

Diante dos fatos expostos, fica evidente a exigência de que o estudo aprofundado dos dados educacionais se torne uma vertente cada vez mais crescente. E para o cumprimento de tal exigência há a necessidade do emprego de tecnologias que possibilitem analisar uma vasta quantidade de dados e, por conseguinte, extrair informações e conhecimento relevante.

Nesse contexto, este estudo primeiramente explana sobre a possibilidade de realização de análises profundas dos dados educacionais por meio de metodologias avançadas baseadas em inferência estatística. Em seguida, são apresentadas as etapas que constituem o processo utilizado neste trabalho, denominado Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in Databases* – KDD), que permite a extração de conhecimento embutido nos dados.

Posteriormente, apresenta-se a aplicação desse processo em bases de dados do INEP. Mais especificamente, aplica-se o processo KDD em dados provenientes da avaliação da Educação Básica, objetivando avaliar algumas questões referentes aos docentes que podem influenciar, tanto de forma positiva quanto negativa, o desempenho dos alunos, em particular na disciplina de matemática. Por fim, são apresentadas algumas conclusões preliminares que podem auxiliar educadores e gestores públicos na busca de ações visando à melhoria na qualidade do ensino.

ANÁLISE DOS DADOS

Conforme mencionado, nos últimos anos houve um acúmulo de dados oriundos do empenho de organizações governamentais em efetuar análises do sistema educacional. Dentro dessa perspectiva, o desafio é fazer uso efetivo desses dados por meio do emprego de diferentes metodologias e ferramentas.

Esse desafio já vem sendo discutido por diversos pesquisadores. Como exemplo, o relato do organizador da Primeira Conferência Internacional em Educação, Economia e Sociedade (*Paris International Conference on Education, Economy and Society*), afirmando que o uso geral e sistemático de metodologias avançadas para análise de dados é certamente a “nova fronteira” para a pesquisa educacional (TCHIBOZO, 2009). Tchibozo menciona ainda o pouco uso de estatística inferencial nos trabalhos de pesquisa de ensino e enfatiza que a aplicação dessas metodologias tornaria os resultados das pesquisas mais convincentes; possibilitaria o aumento da divulgação em toda a comunidade de investigação; facilitaria a comparação com outros estudos; e, por fim, possibilitaria um monitoramento dos métodos de pesquisa, reforçando o acúmulo de conhecimento, fundamental para a evolução do processo científico (TCHIBOZO, 2009).

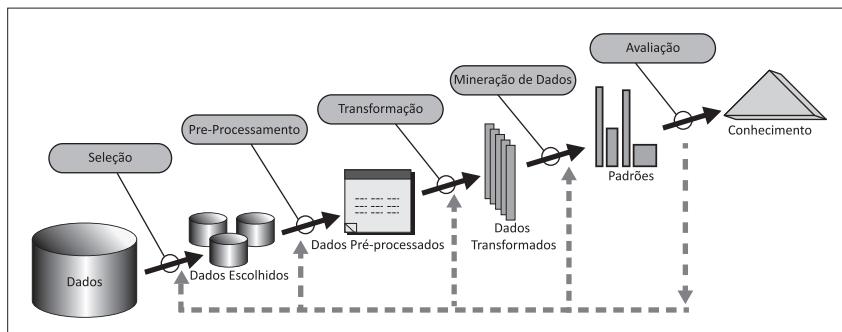
Baseando-se nessas perspectivas, alguns trabalhos que analisam dados educacionais já vêm utilizando uma metodologia promissora denominada mineração de dados (*Data Mining*), que realiza a junção entre a Estatística e a Inteligência Computacional (ver LOPES et al., 2007; NAMEN; BORGES; SADALA, 2013; FONSECA; NAMEN, 2013; KAMPFF; REATEGUI; LIMA, 2008; GOMES; LEVY; LACHTERMACHER, 2004).

Segundo Tan, Steinbach e Kumar (2009), mineração de dados é uma tecnologia que emprega métodos de análise capazes de processar um vasto conjunto de dados, com o objetivo de extrair informações relevantes. Fayyad, Piatetsky-Shapiro e Smyth (1996) afirmam que mineração de dados consiste na realização da análise dos dados e na aplicação de algoritmos que, sob certas limitações computacionais, produzem um conjunto de padrões relacionados aos dados. Declararam ainda que para extrair conhecimento dos dados é necessário seguir um processo denominado KDD e que a mineração constitui um dos passos desse processo. Uma descrição sucinta das etapas do KDD será apresentada a seguir.

PROCESSO KDD

Conforme descrito por Fayyad, Piatetsky-Shapiro e Smyth (1996), Descoberta de Conhecimento em Bancos de Dados consiste em um processo não trivial que almeja a identificação de padrões válidos, novos, potencialmente úteis e comprehensíveis, que estão embutidos nos dados. As cinco etapas que constituem este processo são apresentadas na Figura 1.

FIGURA 1 – Processo KDD



Fonte: FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996.

Inicialmente, por meio de um entendimento bem definido do domínio da aplicação, é necessário selecionar as bases de dados, bem como os dados, que serão usados no processo de descoberta de conhecimento. Em seguida, é efetuada a limpeza e o pré-processamento, uma vez que, frequentemente, os dados são encontrados com inúmeras inconsistências. Essas tarefas são fundamentais, pois o objetivo é eliminar incongruências, de modo que não influenciem o resultado dos algoritmos de mineração que serão aplicados. Posteriormente, realiza-se a transformação que consiste em reduzir ou projetar tais dados.

Essas três etapas iniciais podem ser agrupadas, originando uma grande fase nesse processo, conhecida como preparação dos dados. A partir da preparação apropriada dos dados inicia-se a etapa de mineração que, conforme já mencionado, consiste em escolher técnicas e algoritmos que possibilitem a extração de padrões.

Finalmente, efetua-se a etapa de avaliação que compreende na interpretação dos padrões minerados. Após uma análise minuciosa, usa-se o conhecimento diretamente, incorporando-o a sistemas de apoio a decisões, ou simplesmente se documenta esse conhecimento, expondo-o às partes interessadas. Portanto, essa fase é importante, pois assegura que apenas resultados úteis e válidos sejam utilizados.

BASES USADAS E PREPARAÇÃO DOS DADOS

Conforme anteriormente mencionado, esse estudo se propôs a analisar as bases de dados do INEP referentes à Educação Básica. Tais dados são oriundos de um conjunto de avaliações em larga escala que compõem o Sistema de Avaliação da Educação Básica – Saeb, responsável

por realizar um diagnóstico do sistema educacional brasileiro visando identificar fatores relacionados ao desempenho dos discentes.

A primeira aplicação do Saeb ocorreu em 1990 e ao longo dos anos vem sendo aprimorada por inúmeras modificações (HORTA NETO, 2007). Em 2005, o Saeb foi reestruturado passando a ser composto por duas avaliações: a Avaliação Nacional da Educação Básica (Aneb) e a Avaliação Nacional do Rendimento Escolar (Anresc). Em 2013, acrescentou-se a Avaliação Nacional da Alfabetização (Ana), que avalia de forma censitária os alunos do 3º ano do Ensino Fundamental das escolas públicas.

Seguindo as características anteriores do Saeb, a Aneb realiza a avaliação por amostragem dos alunos matriculados nos 5º e 9º anos do Ensino Fundamental e 3º ano do Ensino Médio, presentes nas redes públicas e privadas, expondo seus resultados em nível nacional, por região e por unidade da federação. Já a Anresc, avalia alunos de Ensino Fundamental dos 5º e 9º anos da rede pública, diferenciando-se principalmente pelo fato de apresentar também os resultados por escola e município (BRASIL, 2011).

A Anresc, por ser uma avaliação mais extensa e detalhada, se tornou conhecida como Prova Brasil. Esse tipo de avaliação, conforme discutido por Bonamino e Souza (2012), possui um grande potencial, já que repassa os resultados para cada escola, permitindo assim um maior acompanhamento dos pais e da sociedade.

A Prova Brasil é composta por aplicações de testes de Língua Portuguesa (com foco em leitura) e Matemática (ênfase em resolução de problemas), bem como questionários aos alunos, diretores e professores, além de englobar questionários preenchidos pelos aplicadores da prova sobre as condições de infraestrutura das escolas. Tal avaliação é realizada a cada dois anos, sendo que em 2013, resultado mais recente disponibilizado no *site* do INEP, ocorreu sua quinta edição. Para acesso a mais detalhes, todos os questionários estão disponíveis no endereço www.inep.gov.br.

Cabe ressaltar que a avaliação Prova Brasil apresenta um desenho seccional, não longitudinal, ou seja, os testes são aplicados a um grupo de discentes de uma determinada série e esses alunos são diferentes a cada ano de avaliação. Limitações relacionadas a essa característica são discutidas mais adiante, nas conclusões do trabalho.

Tomando como base o banco de dados correspondente à Prova Brasil 2011, o presente trabalho deu enfoque aos dados relacionados ao questionário dos professores que lecionam Matemática para alunos do 9º ano do Ensino Fundamental, juntamente com os dados relacionados à proficiência dos seus alunos. Considerou-se ainda

nesta pesquisa, apesar de essa avaliação ser em âmbito nacional, a seleção de docentes do estado do Rio de Janeiro. Desse modo, os resultados obtidos para essa seleção não necessariamente retratam a realidade de outras regiões do país, uma vez que são influenciados pelas características presentes na rede pública do estado do Rio.

O trabalho objetivou identificar o perfil dos professores que pudesse influenciar o processo de ensino-aprendizagem de Matemática dos seus alunos. O agente professor foi selecionado, pois este estudo compartilha o exposto por Nicolella, Kassouf e Belluzzo Junior (2014): “Um elemento importante entre os fatores que determinam a educação de um indivíduo é o professor. Ele é responsável pela facilitação da transmissão do conhecimento e por colaborar na formação pessoal” (NICOLELLA; KASSOUF; BELLUZZO JUNIOR, 2014, p. 157).

Assim, por meio do questionário respondido pelos professores, que aborda questões como formação profissional, salário obtido, expectativas inerentes aos alunos, relacionamento entre os funcionários, práticas pedagógicas, perfil socioeconômico e cultural, buscou-se analisar correlações entre esses aspectos e o desempenho dos estudantes.

SELEÇÃO DOS DADOS

Em conformidade com o objetivo exposto, que é a extração de padrões que relacionam o perfil dos professores com o resultado obtido por seus alunos nas provas de Matemática, foram selecionados os dados referentes às respostas do questionário do professor e à proficiência obtida por seus alunos. Assim, as bases utilizadas, dispostas em diferentes arquivos disponibilizados pelo INEP, foram:

- TS_RESULTADO_ALUNO: arquivo composto das informações com os resultados dos alunos nos testes de Língua Portuguesa e Matemática;
- TS_QUEST_PROFESSOR: arquivo contendo os dados das respostas ao questionário aplicado ao professor de cada disciplina.

O arquivo TS_RESULTADO_ALUNO contém 5.201.730 registros correspondentes, em âmbito nacional, aos alunos do 5º e 9º anos do Ensino Fundamental, com 22 campos que compreendem diversos atributos identificadores e a proficiência nas disciplinas de Matemática e Língua Portuguesa. Já o arquivo TS_QUEST_PROFESSOR é composto por 304.412 registros que correspondem aos professores de todo o país que lecionam para as disciplinas e

séries acima mencionadas, contendo 161 atributos, incluindo 152 referentes às respostas das perguntas do questionário e os restantes, que possibilitam a identificação de cada professor.

Para executar as tarefas que transformam os dados em um formato apropriado, esses arquivos foram importados para tabelas a partir do uso de um sistema gerenciador de banco de dados, sendo usado o *software PostgreSQL*, sistema de código aberto, disponibilizado gratuitamente e amplamente utilizado por desenvolvedores de sistemas (mais detalhes em POSTGRESQL, 2007).

Portanto, como o objetivo é realizar um estudo acerca dos professores que lecionam Matemática para alunos do 9º ano do Ensino Fundamental, foram selecionados na tabela TS_QUEST_PROFESSOR somente os registros de professores que lecionam essa disciplina na série mencionada. Do mesmo modo, foram selecionados apenas os registros de alunos correspondentes da tabela TS_RESULTADO_ALUNO.

Outra tarefa realizada foi a extração, em ambas as tabelas, dos professores e alunos pertencentes ao estado do Rio de Janeiro que, como dito anteriormente, é a região selecionada para este estudo.

É importante mencionar que todas essas seleções foram realizadas com base nos valores dos atributos que constituem as tabelas. Tais atributos armazenam as informações necessárias para se efetuarem essas identificações. Mais detalhes sobre o processo de seleção de dados podem ser encontrados em Fonseca (2014).

Após realizar essas seleções, restaram 153446 registros de alunos na tabela TS_RESULTADO_ALUNO e 3043 registros de professores na tabela TS_QUEST_PROFESSOR.

PRÉ-PROCESSAMENTO E TRANSFORMAÇÃO DOS DADOS

Após a etapa de seleção, algumas tarefas foram realizadas a fim de garantir a qualidade dos dados e, consequentemente, resultados mais confiáveis na etapa posterior de mineração. Essas tarefas demandaram um grande tempo do processo KDD e, por limitações de espaço, são apresentadas de forma sucinta no presente trabalho, sendo detalhadas em Fonseca (2014).

A principal tarefa realizada nesta etapa foi a criação de um atributo que armazenasse, para cada professor, informação que quantificasse se a sua turma obteve (ou não) êxito no teste de Matemática. Assim, seria possível relacionar as variáveis presentes no questionário do professor ao desempenho dos alunos. A informação escolhida para quantificar o desempenho dos estudantes foi o cálculo do percentual de alunos, de

cada professor, que se encontrava acima da média geral de Matemática. Assim, computou-se para cada docente o total de alunos para os quais ele(a) lecionava, bem como quantos de seus alunos obtiveram um resultado superior ao da média geral. Por intermédio desses totais foi possível calcular o percentual de alunos acima da média.

De posse dessa informação foi possível separar os professores em dois conjuntos ou, equivalentemente, em duas classes. Essa distinção em classes foi realizada de duas formas: uma visando à análise da descoberta de fatores que pudessem influenciar de forma positiva e a outra a descoberta de fatores que pudessem influenciar negativamente o desempenho dos alunos.

Portanto, para a análise da influência positiva, os professores que apresentaram um percentual de até 65% dos alunos acima da média geral foram inseridos na classe nomeada “Até 65%”. Os restantes, ou seja, os professores com percentual acima de 65%, foram incluídos na classe intitulada “Maior que 65%”. Acredita-se que a análise do perfil de professores cujos percentuais de alunos acima da média fossem superiores a 65% (ou seja, classificados como “Maior que 65%”) pode levantar alguns indicadores quanto as causas que influenciam positivamente o processo de ensino-aprendizagem dos estudantes.

De modo análogo, para a análise da influência negativa, os professores foram divididos em outras duas classes: “Até 35%”, se possuíssem um percentual de alunos acima da média inferior ou igual a 35%, e “Maior que 35%” caso tivessem um percentual superior a 35%. Assim, a análise dos perfis dos professores da classe “Até 35%” possibilita identificar fatores que acarretariam em um desempenho negativo dos estudantes.

Cabe salientar que a escolha das duas classes foi feita de modo que os percentuais de professores a elas relacionados fossem semelhantes. Desse modo, aproximadamente 26,5% do total de professores pertencia à classe intitulada “Maior que 65%” (influência positiva), com percentual semelhante de professores pertencentes à classe “Até 35%” (influência negativa). As turmas do primeiro grupo, por sua vez, tinham uma média 12,23% maior do que a média geral, enquanto as do último, uma média 12,44% menor, contabilizando uma diferença aproximada de 25% entre as médias dos alunos atendidos por professores do primeiro e do segundo grupos. Apesar das possíveis limitações da escolha realizada, os dois grupos representam diferenças significativas nos resultados de desempenho dos estudantes. Além disso, um percentual expressivo da base de dados original foi mantido na análise (aproximadamente 53%).

Conforme apresentado em detalhes em Fonseca (2014), a tabela TS_QUEST_PROFESSOR teve seus dados transformados, de modo que fosse possível identificar as classes às quais os professores pertenciam, possibilitando que, na mineração de dados, tarefa apresentada na próxima seção, fizesse-se uso desses dados para identificação de padrões relevantes.

MINERAÇÃO DOS DADOS

Após o processo de preparação dos dados, avançou-se para a etapa de mineração de dados. Conforme mencionado, essa etapa compreende a escolha de algoritmos que efetuam a extração de padrões válidos, novos, potencialmente úteis e compreensíveis, embutidos nos dados.

No caso do presente trabalho, baseando-se nas alternativas que os professores assinalaram como resposta às questões, pretendeu-se identificar a qual classe os docentes se enquadram. Em outras palavras, buscou-se identificar quais os atributos, com seus respectivos valores, que permitem classificá-los como “Maior que 65%” (influência positiva) ou “Até 35%” (influência negativa).

Com a finalidade de efetuar essa classificação, foi aplicado o algoritmo de mineração de dados denominado *Naïve Bayes*. Esse algoritmo classifica um registro em uma determinada classe, baseando-se na probabilidade de esse registro pertencer a essa classe. Como característica principal, essa metodologia assume que os atributos não são correlacionados, isto é, o valor de um atributo não influencia os valores dos outros atributos (TAN; STEINBACH; KUMAR, 2009).

Inicialmente, considera-se que se deseja classificar um registro de professor, com base nas respostas dadas ao questionário, para uma classe. Em termos matemáticos, tem-se um registro XX , descrito por um conjunto de atributos X_1, X_2, \dots, X_d ; X_1, X_2, \dots, X_d , e deseja-se classificá-lo para uma classe C_1, C_2, \dots, C_m ; C_1, C_2, \dots, C_m . Essa classificação é feita computando-se a probabilidade de $C_j|C_j$ ocorrer, dado que XX ocorra, $P(C_j|X) P(C_j|X)$, para cada classe $C_j, j = 1, 2, \dots, m$; $C_j, j = 1, 2, \dots, m$. O registro então será rotulado para a classe que maximizar essa probabilidade.

O algoritmo *Naïve Bayes* calcula a probabilidade $P(C_j|X)$ $P(C_j|X)$ por meio do Teorema de Bayes (HAN; KAMBER; PEI, 2012):

$$P(C_j|X) = \frac{P(C_j)P(X|C_j)}{P(X)}.$$

Contudo, como o algoritmo assume que os atributos não são correlacionados, a probabilidade $P(X|C_j)P(X|C_j)$ é computada para cada $X_i|X_i$ independentemente dos outros atributos. Logo, essa expressão é simplificada por:

$$P(C_j|X) = P(C_j) \prod_{i=1}^d P(X_i|C_j) = P(C_j) \times (P(X_1|C_j) \times P(X_2|C_j) \times \dots \times P(X_d|C_j)).$$

Portanto, primeiramente deve-se calcular a probabilidade para cada atributo $(P(X_i|C_j))(P(X_i|C_j))$ e, posteriormente, multiplicá-los. O resultado obtido é, então, multiplicado por $P(C_j)P(C_j)$ a fim de obter $P(C_j|X)P(C_j|X)$. Finalmente, como dito anteriormente, basta verificar qual a classe $C_j|C_j$ que possui maior probabilidade dado XX . Relembrando que as classes $C_j|C_j$, aqui mencionadas, referem-se às classes anteriormente citadas: “Até 65%” ou “Maior que 65%”, esta última relacionada à análise da influência positiva no processo de ensino-aprendizagem; e “Maior que 35%” ou “Até 35%”, sendo a última relacionada à análise da influência negativa.

Neste trabalho foi utilizada uma implementação do algoritmo *Naïve Bayes* disponibilizada no *software* Weka (ver HALL et al., 2009), que é uma ferramenta gratuita composta de uma coleção de algoritmos de mineração de dados e uma série de funcionalidades que auxiliam na etapa de pré-processamento (WITTEN; FRANK; HALL, 2011).

RESULTADOS E AVALIAÇÃO

Primeiramente foi executado o algoritmo *Naïve Bayes*, com o objetivo de identificar quais valores de atributos favorecem um professor ser classificado como da classe “Maior que 65%”, isto é, fatores que podem influenciar positivamente o desempenho dos discentes. Mais detalhes relacionados à preparação do ambiente computacional, definição dos parâmetros utilizados e configuração da ferramenta Weka para essa tarefa podem ser vistos em Fonseca (2014).

Conforme já discutido, *Naïve Bayes* apresenta como resultado as probabilidades de um determinado atributo ocorrer dado que uma

classe ocorra $(P(X_i|C_j))(P(X_i|C_j))$. Portanto, o modelo gerado por *Naïve Bayes* é dado por um conjunto de probabilidades correspondente a cada atributo, com seu respectivo valor. Desse modo, apresentam-se na Tabela 1 as probabilidades relacionadas às respostas a questões que implicaram na classificação do professor como “Maior que 65%”, ou seja, que indicaram que mais que 65% de seus alunos obtiveram desempenho acima da média.

TABELA 1

Resultado de mineração de dados gerado pelo algoritmo *Naïve Bayes* (influência positiva)

Resultado da etapa de Mineração de Dados (Influência Positiva)		
Atributos: Perguntas do Questionário do Professor	Valores	Probabilidade (%)
Questão 20) Nesta escola, qual a sua carga horária semanal?	Até 24h-aula	63,71
	De 25 a 30h-aula	9,52
	Mais de 30h-aula	24,79
	Desconhecido	1,98
Questão 23) Qual é a sua situação trabalhista nesta escola?	Estatutário	83,09
	CLT	5,09
	Sem contrato ou Temporário	7,39
	Desconhecido	4,43
Questão 76) Ocorreu na escola: inexistência de professores para algumas disciplinas ou séries?	Não	47,62
	Sim, não grave	31,86
	Sim, e grave	18,88
	Desconhecido	1,64
Questão 81) Ocorreu na escola: alto índice de faltas por parte dos alunos?	Não	64,70
	Sim, não grave	27,59
	Sim, e grave	6,40
	Desconhecido	1,31
Questão 113) Você conhece os resultados do sistema de avaliação da Educação Básica – Saeb: do seu município?	Sim	74,67
	Não	24,18
	Desconhecido	1,15
Questão 121) Quanto dos conteúdos previstos você conseguiu desenvolver com os alunos desta turma, neste ano?	Menos de 40%	1,32
	Entre 40% e 80%	30,21
	Mais de 80%	66,83
	Desconhecido	1,64
Questão 124) Quantos dos alunos desta turma você acha que: concluirão o Ensino Médio?	Poucos	2,63
	Próximo da metade	25,94
	Quase todos	53,04
	Desconhecido	18,39
Questão 125) Quantos dos alunos desta turma você acha que: entrarão para a universidade?	Poucos	25,78
	Próximo da metade	44,17
	Quase todos	11,33
	Desconhecido	18,72

Fonte: FONSECA, 2014

É importante, neste ponto, relembrar que as informações extraídas referem-se aos docentes que lecionam Matemática para os alunos do 9º do Ensino Fundamental no Estado do Rio de Janeiro, não necessariamente representando, por completo, a realidade de outras regiões do país.

Note que na Tabela 1 os valores dos atributos que possuem uma maior probabilidade de ocorrer, dado que a classe “Maior que 65%” ocorra, estão em destaque (células preenchidas com cor cinza). Assim, os valores desses atributos favorecem para que um registro seja classificado como da classe “Maior que 65%”. Além disso, é importante ressaltar que os valores “Desconhecido” nas perguntas desta tabela referem-se às respostas em branco ou nulas.

Ao se analisarem os atributos com seus respectivos valores, podem-se observar alguns resultados significativos. Nota-se que o desenvolvimento de um alto percentual do conteúdo originalmente previsto, a frequência regular dos alunos na escola e a inexistência de problemas relacionados à falta de professores favorecem um desempenho positivo dos discentes. Além disso, aparecem como fatores favoráveis a situação trabalhista estável, bem como a carga horária semanal não elevada dos docentes.

Os resultados obtidos reforçam outros estudos efetuados, como o de Carvalho et al. (2012), que analisou escolas da Região Metropolitana do Rio de Janeiro a partir dos dados dos questionários da Prova Brasil 2009, objetivando estudar fatores intraescolares que favoreciam o desempenho dos alunos do 5º ano do Ensino Fundamental em Matemática. Os autores concluíram que a estabilidade do vínculo profissional do corpo docente é relevante para um desempenho positivo. Além disso, avaliou-se que quando não havia carência de professores para lecionar algumas disciplinas ou altos índices de faltas de professores e alunos, o desempenho dos estudantes era superior.

Como pode ser visto ainda, as perguntas de número 124 e 125 (que abordam a expectativa que o professor deposita em relação à formação futura de seus educandos) com os valores “Quase todos” ou “Próximo da metade” configuram-se como importantes fatores para classificar um registro como pertencente à classe “Maior que 65%”. Isso indica que a crença dos professores em relação à turma ter, futuramente, uma boa formação educacional afeta positivamente o desempenho dos alunos.

É importante mencionar que a detecção desse último padrão corrobora pesquisas realizadas desde a década de 60. O primeiro estudo que expõe esse resultado foi realizado por Rosenthal e Jacobson (1966) e destacou que professores que têm uma visão positiva dos alunos tendem a estimulá-los e estes, por conseguinte, obtêm

melhor desempenho. Por outro lado, professores que não acreditam no potencial de seus alunos adotam posturas que comprometem negativamente os resultados dos educandos. Em pesquisa mais recente, Barbosa e Randall (2004) reforçam essa questão, demonstrando que há uma correspondência muito forte entre expectativas positivas dos professores e melhores resultados dos alunos. A discussão sobre esse aspecto é feita também em Namen, Borges e Sadala (2013), que obtiveram conclusão semelhante ao analisarem dados da Prova Brasil 2007 relacionados a professores que lecionam Língua Portuguesa para alunos do 5º ano do Ensino Fundamental.

Outra questão a ser ressaltada é a necessidade de repasse dos resultados das avaliações a respeito do ensino básico para os docentes, pois verificou-se que ao conhecer os resultados obtidos nas avaliações do Saeb, os professores tendem a ter alunos com uma melhor proficiência em Matemática (questão 113).

Uma possível explicação para esta última relação citada consiste no argumento de que o professor, ao saber que seus alunos estão sendo avaliados, modifica sua forma de ensinar para que eles obtenham êxito nas próximas avaliações. De acordo com Libâneo (2013), a partir das avaliações e seus respectivos diagnósticos, os professores são levados a repensar suas práticas de ensino, visando à melhoria do desempenho de seus pupilos nas futuras avaliações.

Após observar os fatores que afetam positivamente o desempenho dos discentes, foi executado novamente o algoritmo *Naïve Bayes*, a fim de identificar quais valores de atributos favoreciam a classificação do professor na classe “Até 35%”. Conforme mencionado anteriormente, como o percentual de alunos acima da média é muito pequeno, infere-se que alguns fatores relacionados a estes professores possam influenciar negativamente o desempenho dos estudantes.

As probabilidades de cada atributo ocorrer, com seus respectivos valores, dado que a classe “Até 35%” ocorra são apresentadas na Tabela 2. Novamente, os valores que obtiveram uma maior probabilidade foram destacados (células preenchidas com cor cinza).

Ao se analisar a Tabela 2 podem-se destacar alguns fatores que influenciam negativamente a proficiência dos alunos, como o baixo salário recebido pelo educador. Observa-se que, neste caso, a probabilidade da menor faixa “Até R\$ 1.090,00” (equivalente a 2 salários mínimos em 2011) é próxima à probabilidade do valor mediano “De R\$ 1.090,01 a R\$ 2.180,00”, e a soma dos percentuais das duas menores faixas salariais corresponde a mais de cinquenta por cento das situações.

TABELA 2Resultado de mineração de dados gerado pelo algoritmo *Naïve Bayes* (influência negativa)

Resultado da etapa de Mineração de Dados (Influência Positiva)		
Atributos: Perguntas do Questionário do Professor	Valores	Probabilidade (%)
Questão 5) Há quantos anos você obteve o nível de escolaridade?	2 ou menos	4,92
	De 3 a 14	60,16
	Mais de 15	32,95
	Desconhecido	1,97
Questão 14) Nesta escola, qual, aproximadamente, o seu salário bruto?	Até 1090,00	26,56
	De 1090,01 a 2180,00	29,18
	Mais de 2180,00	21,64
	Desconhecido	22,62
Questão 81) Ocorreu na escola: alto índice de faltas por parte dos alunos?	Não	41,31
	Sim, não grave	36,89
	Sim, e grave	20,00
	Desconhecido	1,80
Questão 113) Você conhece os resultados do sistema de avaliação da Educação Básica – Saeb: do seu município?	Sim	56,82
	Não	41,54
	Desconhecido	1,64
	Menos de 40%	2,30
Questão 121) Quanto dos conteúdos previstos você conseguiu desenvolver com os alunos desta turma, neste ano?	Entre 40% e 80%	47,38
	Mais de 80%	48,52
	Desconhecido	1,80
	Poucos	41,14
Questão 125) Quantos dos alunos desta turma você acha que: entrarão para a universidade?	Próximo da metade	27,38
	Quase todos	4,43
	Desconhecido	27,05
	Sim, a maioria	86,23
Questão 126) Os alunos desta turma têm livros didáticos?	Metade da turma	3,93
	Não	8,20
	Desconhecido	1,64
	Frequente	93,12
Questão 149) Indique a frequência com a qual você desenvolve as seguintes práticas pedagógicas com seus alunos nesta turma: experimentar diferentes modos de resolver um problema ou de efetuar um cálculo.	Menos frequente	2,62
	Nunca	3,28
	Desconhecido	0,98

Fonte: FONSECA, 2014

Nesse contexto, referente à desvalorização da profissão de educador, Barbosa (2011) considera que o salário do professor brasileiro é considerado baixo e esse fato acarretaria em dificuldades para se manter os bons professores, bem como atrair novos profissionais para lecionar. O autor declara ainda que, como alternativa para contornar a pouca renda resultante dos baixos salários, os professores intensificam a jornada de trabalho o que,

consequentemente, compromete a realização de outras atividades desenvolvidas pelo docente. Assim, tais fatores não permitiriam que o docente pudesse investir no seu aprimoramento profissional o que poderia implicar em pior desempenho dos seus alunos.

Se por um lado a questão salarial parece apontar para o senso comum, os resultados referentes à questão 126 são, de certo modo, surpreendentes, pois indicam que mesmo nas situações em que a maioria dos alunos possuía livros didáticos, a probabilidade de obtenção de maus resultados nas provas foi alta. Pode-se concluir, num primeiro momento, que esse fator não representa tanta importância, em comparação a outros aspectos, em relação à melhoria do desempenho do aluno. Acredita-se que um estudo mais aprofundado em relação ao ponto deva ser realizado por educadores e gestores educacionais.

Analizando a Tabela 2, nota-se também que assim como a crença dos professores na boa formação educacional futura dos seus alunos contribuía para o bom desempenho dos discentes, a pequena expectativa do professor quanto ao número de alunos que iniciarão o nível superior reflete negativamente no desempenho dos estudantes. Portanto, a situação inversa apresentada por Rosenthal e Jacobson (1966) também foi verificada.

A situação inversa acima percebida não ocorre no caso da questão 113, que versa sobre o conhecimento dos resultados do Saeb. Se a Tabela 1 indica que o conhecimento dos resultados do Saeb (74,67%) favorece positivamente para o bom desempenho dos alunos, a Tabela 2 mostra que esse fator não exerce tanta influência em relação ao desempenho ruim dos alunos (56,82% conhecem e 41,54% desconhecem). Poderia se explicar esse ponto, afirmando que conhecer os resultados do Saeb é uma questão diretamente relacionada aos professores que obtêm bons resultados, mas que há também um bom percentual de professores que conhecem esses resultados e que têm a grande maioria de seus alunos abaixo da média. Em outras palavras, para obtenção de bons resultados em Matemática é fundamental conhecer os resultados do Saeb mas, por outro lado, conhecer estes resultados não garante necessariamente o sucesso.

As células destacadas em cinza não devem ser analisadas isoladamente nos casos das questões 81 e 121. Quanto à primeira, pode-se perceber que a maior probabilidade (56,89%) de ocorrência de maus resultados refere-se aos casos relacionados ao absenteísmo dos alunos (respostas “Sim, e grave” e “Sim, não grave”). Quanto à ultima, a maioria das respostas, que engloba exatamente as respostas não destacadas, indica que menos de 80% dos conteúdos previstos foram desenvolvidos em sala de aula.

Finalmente, as respostas relacionadas às questões 5 e 149 que se referem, respectivamente, a aspectos como tempo de formação e práticas pedagógicas do professor, apresentam resultados que devem ser estudados com mais profundidade para identificação das reais explicações para o mau desempenho dos alunos.

CONCLUSÕES

Conforme mencionado, o INEP organiza e disponibiliza uma ampla base de dados relacionada à educação para a realização de estudos e diagnósticos. Avaliando essas bases pode-se concluir que, apesar de serem importante fonte de informação, muitas vezes não são exploradas em todo o seu potencial, devido às dificuldades inerentes ao grande volume de dados envolvido, que está além da capacidade humana de realizar sua interpretação. Diante desse contexto, o desafio aqui apresentado foi fazer uso efetivo desses dados, analisando-os por meio de metodologias que propiciassem a extração de informações que oferecessem rotas para a definição de ações voltadas à melhoria dos resultados do processo de ensino-aprendizagem de Matemática na Educação Básica. A alternativa apresentada foi a Mineração de Dados, uma tecnologia que se alicerça em conceitos estatísticos e de inteligência computacional.

Cabe ressaltar que o foco do presente trabalho se restringiu às bases de dados do INEP, apesar de existirem bases de dados estaduais e municipais repletas de informações valiosas que também poderiam subsidiar o processo de análise e tomada de decisão. Nesse sentido, alguns estados como Ceará e São Paulo podem ser citados como exemplos em relação à utilização efetiva de suas bases de dados com vistas à elaboração de políticas educacionais visando à melhoria no processo de ensino-aprendizagem (ver SOUSA; OLIVEIRA, 2010).

Ademais, poderiam ser abordadas mais dimensões do problema, não somente o professor como o caso aqui exposto, mas também dados referentes aos questionários respondidos pelos alunos, diretores e sobre a escola. Isso possibilitaria uma avaliação mais profunda da influência das condições socioeconômicas dos alunos, bem como aspectos relacionados à eficácia das escolas e o efeito escola. Alguns trabalhos como os de Paul e Barbosa (2008), Costa e Guedes (2009), Costa (2010), Barbosa e Sant'Anna (2010) e Alves e Passador (2011) abordam bem essas questões e sua relevância. Tal análise possibilitaria a extração de padrões e de relações que poderiam enriquecer os resultados aqui obtidos, e que estão sujeitos a algumas limitações, pelo fato de restringirem-se somente à identificação de correlações entre os perfis dos professores e o desempenho de seus alunos.

Conforme mencionado anteriormente, a Prova Brasil não é uma avaliação longitudinal, ou seja, não permite coletar informações a respeito de um mesmo aluno ao longo do tempo, característica que gera algumas limitações quanto à identificação de fatores que influenciam o processo de aprendizagem. Analisando esse aspecto, Bonamino e Oliveira (2013) argumentam que o aprendizado do aluno é um processo que ocorre ao longo dos anos, ou seja, o desempenho do aluno no 9º ano do Ensino Fundamental retrata não apenas o que ele aprendeu no 9º ano, mas também nos anos letivos anteriores. Além disso, como as variáveis do professor são observadas apenas no ano da avaliação (no presente caso, o 9º ano), há falta de sintonia temporal entre a medida de desempenho e as medidas de contexto aferidas por esse tipo de avaliação seccional. Essa limitação é amplamente discutida em outros trabalhos (FRANCO; BROOKE; ALVES, 2008; LEE, 2010).

É importante salientar que há algumas restrições relacionadas aos mecanismos de aferição utilizados na Prova Brasil e aos indicadores que podem ser gerados por intermédio das ferramentas aqui apresentadas, e que são analisadas mais detalhadamente no trabalho de Namen, Borges e Sadala (2013). Assim como Piaget (2011, p. 71) considera que o exame escolar “orienta ele todo o trabalho do aluno para o resultado artificial, o bom êxito nas provas finais, no lugar de apelar para as suas reais atividades e sua personalidade”, André (2009), abordando especificamente o Saeb e a Prova Brasil, cita a armadilha de a aprendizagem se reduzir ao ensino de conteúdos e ao que for passível de mensuração, deixando de fora outros aspectos extremamente ricos, como, por exemplo, o desenvolvimento de valores, de atitudes, de convivência social, de solidariedade, de práticas culturais. Esse aspecto também é abordado por Ravitch (2011), que levanta a necessidade de formar jovens com bom embasamento em história, ciência, geografia, matemática, leitura, e não apenas formar gerações que aprendam a responder testes de múltipla escolha.

Apesar de todas essas limitações, acredita-se que as informações coletadas por intermédio da Prova Brasil podem fornecer perspectivas importantes sobre alunos, professores, diretores e escolas. Segundo Alves e Passador (2011), a comparação de desempenho das redes administradas por diferentes esferas de governo em diferentes regiões, desde que não utilizadas para a geração de um simples esquema de competição, pode auxiliar a elaboração de políticas públicas na área de educação.

Por intermédio dos dados oriundos da Prova Brasil, que avalia o Ensino Fundamental regular das redes públicas, esta pesquisa teve como propósito aplicar um algoritmo de mineração de dados com

a finalidade de relacionar a aprendizagem de Matemática e o perfil de professores do Ensino Fundamental. Mais especificamente, foram abordados os fatores referentes aos educadores do Estado do Rio de Janeiro que influenciam positivamente e negativamente o desempenho dos alunos do 9º ano do ensino básico nas provas de Matemática.

Para extrair essas informações foi aplicado o processo de descoberta de conhecimento em bancos de dados. Primeiramente, efetuou-se a etapa de preparação dos dados que, apesar de despender um grande tempo do processo, é essencial para garantir a qualidade dos dados e, consequentemente, a extração de padrões mais confiáveis. Posteriormente, aplicou-se o algoritmo de mineração *Naïve Bayes*, que permite a classificação de registros baseando-se na probabilidade de pertencerem a uma determinada classe. É importante ressaltar que, mesmo assumindo que não há correlação entre os atributos, esse algoritmo tem se mostrado eficaz em diversas aplicações, apresentando um bom desempenho computacional (HAN; KAMBER; PEI, 2012).

A etapa de mineração possibilitou observar alguns fatores que são favoráveis ao aprendizado de matemática, como alto percentual de conteúdo previsto desenvolvido; assiduidade dos discentes na escola; baixos índices de absenteísmo dos professores; a expectativa, por parte do docente, de que muitos dos seus alunos terão uma boa formação educacional; e a necessidade de divulgação, para o corpo docente, dos resultados das avaliações a respeito do ensino básico.

Alguns aspectos que tendem a influenciar negativamente o desempenho dos estudantes também puderam ser notados, como a desvalorização salarial da profissão de educador; os altos índices de absenteísmo dos alunos; e a crença do professor de que poucos alunos entrarão em uma universidade. Questões que, a princípio, podem se situar distantes do senso comum, como a pouca relevância dos livros didáticos em relação à atenuação do mau desempenho dos alunos, foram também observadas.

Os fatores descobertos apresentados reforçam e/ou acrescentam questões que vêm sendo discutidas por educadores e pesquisadores. No entanto, este trabalho não objetivou realizar uma análise profunda desses resultados, cabendo futuramente uma consulta aos especialistas na área da educação.

É importante ressaltar que foi aplicado apenas um algoritmo de mineração devido às limitações de espaço inerentes a um artigo. No entanto, outras metodologias de mineração de dados podem ser aplicadas, como árvores de decisão e classificadores baseados em regras. Outra possibilidade a ser explorada consiste em uma análise

temporal a respeito dos dados da Prova Brasil, uma vez que no *site* do INEP são disponibilizados os dados das edições anteriores da avaliação. Assim, seria possível comparar os fatores que vêm afetando o desempenho dos discentes ao longo dos anos.

Diante desse campo abrangente de aplicações, este estudo se propôs a utilizar uma metodologia capaz de realizar uma análise dos dados da avaliação da Educação Básica. Espera-se que possa servir como estímulo para o uso mais intenso de métodos de mineração de dados em bases de dados do INEP, com o intuito de mapear a situação de diversos níveis e modalidades de ensino e, consequentemente, nortear melhorias no sistema educacional brasileiro.

AGRADECIMENTOS

O presente trabalho foi realizado com o apoio financeiro da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro - FAPERJ.

REFERÊNCIAS

ALVES, T.; PASSADOR, C.S. *Educação Pública no Brasil*: condições de oferta, nível socioeconômico dos alunos e avaliação. São Paulo: Annablume; Brasília: Capes, Inep, 2011.

ALVES, M. T. G.; SOARES, J. F. Medidas de nível socioeconômico em pesquisas sociais: uma aplicação aos dados de uma pesquisa educacional. *Opinião Pública*, Campinas, v. 15, n. 1, p. 1-30, jun. 2009.

ANDRÉ, M. A complexa relação entre pesquisas e políticas públicas no campo da formação de professores. *Educação*, Porto Alegre, v. 32, n. 3, p. 270-276, set./dez. 2009.

ARAÚJO, C. H.; LUZIO, N. *Avaliação da Educação Básica*: em busca da qualidade e eqüidade no Brasil. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2005.

BARBOSA, A. *Os salários dos professores brasileiros*: implicações para o trabalho docente. 2011. 208 f. Tese (Doutorado em Educação Escolar) – Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara, 2011.

BARBOSA, M. L. O.; RANDALL, L. Desigualdades sociais e a formação de expectativas familiares e de professores. *Caderno CRH* (UFBA .Impresso), Salvador, v. 17, n. 41, p. 299-308, mai./ago. 2004.

BARBOSA, M. L. O.; SANT'ANNA, M. J. G. O espaço urbano na escola: efeitos sobre a distribuição da qualidade. *Sociologia da Educação*, Rio de Janeiro, v. 1, p. 40-70, 2010.

BAUER, A. É possível relacionar avaliação discente e formação de professores? A experiência de São Paulo. *Educ. rev.*, Belo Horizonte, v. 28, n. 2, p. 61-82, jun. 2012.

BONAMINO, A.; OLIVEIRA, L. H. G. Estudos longitudinais e pesquisa na educação básica. *Linhas Críticas*, Brasília, v. 19, n. 38, p. 33-50, jan./abr. 2013.

BONAMINO, A.; SOUSA, S. Z. Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola. *Educação e Pesquisa*, São Paulo, v. 38, n. 2, p. 373-388, abr./jun. 2012.

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *PDE/PROVA BRASIL Plano de Desenvolvimento da Educação 2011*. Brasília, 2011. Disponível em: <http://portal.mec.gov.br/dm/documents/prova%20brasil_matriz2.pdf>. Acesso em: 10 ago. 2014.

CARVALHO, C. P. de; WALHELM, A. P. S.; ALVES, F.; KOSLINSKI, M. Gestão e desempenho escolar: um estudo nas redes municipais da Região Metropolitana do Rio de Janeiro a partir dos resultados da Prova Brasil 2009. In: CONGRESSO IBERO AMERICANO DE POLÍTICA E ADMINISTRAÇÃO DA EDUCAÇÃO. 3., 2012, Zaragoza. *Cadernos ANPAE*. Timbaúba: Biblioteca ANPAE - Cadernos ANPAE, 2012.

CERDEIRA, D. G.; ALMEIDA, A. B; COSTA, M. Indicadores e avaliação educacional: percepções e reações a políticas de responsabilização. *Estudos em Avaliação Educacional* (Impresso), São Paulo, v. 25, n. 57, p. 198-225, jan./abr. 2014.

COSTA, M. Famílias e acesso diferenciado a escolas públicas prestigiadas: um estudo de caso. *Educ. rev.*, v. 26, n. 2, p. 227-248, ago. 2010.

COSTA, M.; GUEDES, R. Expectativas de futuro como efeito-escola: explorando possibilidades. *São Paulo em Perspectiva* (Impresso), São Paulo, v. 23, n.1, p. 101-114, jan./jun. 2009.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI Magazine*, American Association for Artificial Intelligence, Califórnia, USA, v. 17, n. 3, p. 37-54, 1996.

FONSECA, S. O. *Utilização de modelos de classificação para mineração de dados relacionados à aprendizagem de matemática e ao perfil de professores do ensino fundamental*. 2014. 121 f. Dissertação (Mestrado em Modelagem Computacional) – Instituto Politécnico, Universidade do Estado do Rio de Janeiro, Nova Friburgo, 2014.

FONSECA, S. O; NAMEN, A. A. Um modelo de classificação para a descoberta de conhecimento relacionado a professores de matemática do ensino básico. In: ENCONTRO DE MODELAGEM COMPUTACIONAL, 16. 2013, Ilhéus, Bahia. ENCONTRO DE CIÊNCIA E TECNOLOGIA DE MATERIAIS, 4., 2013, Ilhéus, Bahia. ENCONTRO REGIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, 3., 2013, Ilhéus, Bahia. *Anais...* Rio de Janeiro: Rede Sírius, UERJ, 2013, v.1.

FRANCO, C.; BROOKE, N.; ALVES, F. Estudo longitudinal sobre qualidade e equidade no ensino fundamental brasileiro: GERES 2005. *Ensaio: avaliação e políticas públicas em educação*, Rio de Janeiro, v. 16, n.61, p. 625-638, out./dez. 2008.

GEWEHR, G. G. *Avaliação da educação básica: políticas e práticas no contexto de escolas públicas municipais*. 2010. 115 f. Dissertação (Mestrado em Educação) – Programa de Pós-graduação em Educação, Centro de Teologia e Ciências Humanas, Pontifícia Universidade Católica do Paraná, Curitiba, 2010.

GOMES, J. C.; LEVY, A.; LACHTERMACHER, G. Segmentação do censo educacional 2000 utilizando técnicas de mineração de dados: o impacto da pesquisa operacional nas novas tendências multidisciplinares. In: SIMPÓSIO BRASILEIRO DE PESQUISAS OPERACIONAIS, 36. 2004, São João del-Rei, Minas Gerais. *Anais...* São João del-Rei, MG, Brasil, nov. 2004, p. 820-831.

HAN, J.; KAMBER, M.; PEI, J. *Data Mining: concepts and techniques*. 3. ed. Waltham, USA: Morgan Kaufmann Publishers, 2012.

HORTA NETO, J. L. Um olhar retrospectivo sobre a avaliação externa no Brasil: das primeiras medições em educação até o SAEB de 2005. *Revista Iberoamericana de Educación*, Madrid, Espanha, n. 42/5, abr. 2007.

_____. Os desafios da utilização dos resultados de avaliações nacionais para o desenvolvimento de políticas educacionais por um governo subnacional no Brasil. *Revista Iberoamericana de Educación*, Madrid, Espanha, n. 53, p. 65-82, 2010.

KAMPFF, A. J. C.; REATEGUI, E. B.; LIMA, J. V. Mineração de dados educacionais para a construção de alertas em ambientes virtuais de aprendizagem como apoio à prática docente. *Novas Tecnologias na Educação*, Rio Grande do Sul, Brasil, v. 6, n. 2, dez. 2008.

LEE, V. E. A necessidade dos dados longitudinais na identificação do efeito-escola. *R. bras. Est. Pedag.*, Brasília, v. 91, n. 229, p. 471-480, set./dez. 2010.

LIBÂNEO, J. C. *Didática*. 2. Ed. São Paulo: Cortez Editora, 2013.

LOPES, C. B. et al. Identificação das características associadas com a aprovação de candidatos de escolas públicas e privadas: Vestibular-2004 UFMG. *Educ. Rev.*, Belo Horizonte, n. 46, p. 167-194, dez. 2007.

MACHADO, C. Avaliação externa e gestão escolar: reflexões sobre usos dos resultados. *Revista @ambienteeducação*, v. 5, n. 1 p.70-82, jan./jun. 2012.

MARTUCCI, E.M. Informação para educação: os novos cenários para o ensino fundamental. *Informação & Sociedade: Estudos*, João Pessoa, v. 10, n. 2, p. 13-36, 2000.

NAMEN, A. A.; BORGES, S. X. de A.; SADALA, M. da G. S. Indicadores de qualidade do ensino fundamental: o uso das tecnologias de mineração de dados e de visões multidimensionais para apoio à análise e definição de políticas públicas. *Revista Brasileira Estudos Pedagógicos*, Brasília, v. 94, n. 238, p. 677-700, set./dez. 2013.

NICOLELLA, A. C.; KASSOUF, A. L.; BELUZZO JUNIOR, W. Programas de qualificação do corpo docente e sua relação com o desempenho dos estudantes. In: FERNANDES, R. et. al. (Org.). *Políticas públicas educacionais e desempenho escolar dos alunos da rede pública de ensino*. Ribeirão Preto, SP: FUNPEC-Editora, 2014. p. 157-195.

PAULA, H. de F.; MOREIRA, A. F. Atividade, ação mediada e avaliação escolar. *Educ. Rev.*, Belo Horizonte, v. 30, n. 1, p. 17-36, mar. 2014.

PAUL, J.; BARBOSA, M. L. O. Qualidade docente e eficácia escolar. *Tempo Social: Revista de Sociologia da USP*, São Paulo, v. 20, n. 1, p. 119-133, 2008.

PIAGET, J. *Para onde vai a educação?* 20. ed. Rio de Janeiro: José Olympio, 2011.

POSTGRESQL. *Documentação do PostgreSQL 8.0.0*. Rio de Janeiro, Brasil, 2007. 1310 p.

RAVITCH, D. *Vida e morte do grande sistema escolar americano: como os testes padronizados e o modelo de mercado ameaçam a educação*. Porto Alegre: Sulina, 2011.

RIGOTTI, J.I.R.; CERQUEIRA, C. A. As bases de dados do INEP e os indicadores educacionais: conceitos e aplicações. In: RIOS-NETO, E. L. G.; RIANI, J. de L. R. (Org.). *Introdução à demografia da educação*. Campinas: Associação Brasileira de Estudos Populacionais, p. 73-88, 2004.

ROSENTHAL, R.; JACOBSON, L. Teachers' expectancies: determinants of pupils' IQ gains. *Psychological Report*, Missoula, USA, v. 19, p. 115-118, 1966.

SOARES, J. F. Qualidade e equidade na Educação Básica brasileira: a evidência do Saeb-2001. *Arquivos Analíticos de Políticas Educativas*, Tempe, USA, v. 12, n. 38, 2004.

SOUZA, S. Z.; OLIVEIRA, R. P. de. Sistemas estaduais de avaliação: uso dos resultados, implicações e tendências. *Cadernos de Pesquisa*, São Paulo, v. 40, n. 141, p 793-822, 2010.

TAN, P.; STEINBACH, M.; KUMAR, V. *Introdução ao data mining*: mineração de dados. Rio de Janeiro: Ciência Moderna, 2009.

TCHIBOZO, G. Applications in data analysis for educational research. *Policy Futures in Education*, Paris, França, v. 7, n. 4, p. 364-367, 2009.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. *The WEKA Data Mining Software: An Update*; SIGKDD Explorations, New York, USA, v. 11, n. 1, p. 10-18, jun. 2009.

WITTEN, I. H.; FRANK, E.; HALL, M. A. *Data Mining: practical machine learning tools and techniques*. 3. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011.

Recebido: 29/09/2014

Aprovado: 16/04/2015

Contato:

Stella Oggioni da Fonseca
Universidade do Estado do Rio de Janeiro
Instituto Politécnico
Rua Bonfim, 25, Vila Amélia, Nova Friburgo
Rio de Janeiro | RJ | Brasil
CEP 28.625-570