

Investigação acerca dos fatores determinantes para a conclusão do Ensino Fundamental utilizando Mineração de Dados Educacionais no Censo Escolar da Educação Básica do INEP 2014

Gisele da S. Ferreira

Programa de Pós-Graduação em Informática na Educação
Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre, RS – Brasil

`gisele.ferreira@ufrgs.br`

Abstract. *This article focuses on the identification of factors related to the completion of elementary school using data mining techniques applied to Microdata of Education Basic School Census of 2014. INEP The advantage of this approach is the use of public data that provide information about students, classes, schools and teachers from all over Brazil. The experiments carried out from decision trees were, at first, analyzed with full information and, secondly, the information is divided into: Data of the students, the classes and data from schools. The experiments performed with all the information provided aggregate average accuracies of 96.17%.*

Resumo. *Este artigo tem como foco a identificação de fatores relacionados à conclusão do Ensino Fundamental utilizando técnicas de mineração de dados aplicadas aos Microdados do Censo Escolar da Educação Básica do INEP de 2014. A vantagem desta abordagem é a utilização de dados públicos que trazem informações sobre estudantes, turmas, escolas e docentes de todo o Brasil. Os experimentos realizados a partir de árvores de decisão foram, num primeiro momento, analisados com sua totalidade de informações e, num segundo momento, as informações foram divididas entre: dados dos alunos, dados das turmas e dados das escolas. Os experimentos realizados com todas as informações agregadas forneceram precisões médias de 96,17%.*

1. Introdução

Nas últimas décadas foram produzidos inúmeros trabalhos teóricos e empíricos acerca dos fatores que contribuem para o atraso, abandono, permanência e interrupção dos estudos. Trabalho, gravidez, insegurança, violência, problemas emocionais, afetivos e familiares como falta de incentivo familiar, desinteresse dos responsáveis, desestruturação familiar, desamparo, doenças como depressão ou inimizades com colegas, falta de investimento na estrutura das instituições de ensino para prover as necessidades básicas dos estudantes, como boas instalações de salas de aula, manutenção das bibliotecas, laboratórios com computadores, laboratórios de ciências, quadras esportivas, cozinhas, mesas, cadeiras, banheiros e pinturas, desvalorização dos professores que resultam em despreparo ou falta de docentes, localização das escolas, etc. são fatores que podem contribuir para desincentivar os estudos. É importante que se procure entender como e o quanto cada fator exerce influência sobre o estudo para que

se possa pensar em políticas públicas mais adequadas ao combate da evasão escolar e melhora da educação.

Existe uma preocupação crescente com políticas afirmativas de incentivo ao estudo por parte do governo e sociedade, ao mesmo tempo em que faltam estudos sobre que fatores influenciam o acesso à educação por parte da população e como cada fator contribui positiva ou negativamente para que o incentivo ao estudo seja garantido à população em geral. A maior parte dos estudos sobre Mineração de Dados Educacionais prioriza a descrição das técnicas e poucos estudos procuram detectar as causas da evasão e repetência escolar, enquanto esses estudos são muito importantes por apresentarem grandes efeitos financeiros e de impacto em toda a cadeia produtiva do país [RIGO *et al* 2014].

O Instituto de Pesquisas Aplicadas (IPEA), a Fundação João Pinheiro (FJP) e o Programa das Nações Unidas Para o Desenvolvimento (PNUD) lançaram, em 2014, o Atlas do Desenvolvimento Humano nas Regiões Metropolitanas Brasileiras, que apresentou o Índice de Desenvolvimento Municipal (IDHM) calculado para permitir uma análise da realidade socioeconômica intrametropolitana e entre as 16 regiões metropolitanas pesquisadas no estudo [PNUD, IPEA e FJP 2014]. O IDHM é obtido da média aritmética de outros três subíndices: IDH-Educação, IDH-Longevidade e IDH-Renda. O IDHM de todas as 16 regiões metropolitanas apresentaram uma melhora significativa e foi verificada também uma redução das disparidades entre as 16 RMs estudadas [MARTINS e FERREIRA 2014]. O IDHM da Região Metropolitana de Porto Alegre (RMPA) figurava na quarta posição dentre as 16 regiões metropolitanas analisadas em 2000, caindo para a nona posição em 2010. Essa queda de cinco posições foi ocasionada pelo desempenho do IDH-Educação, que sofreu aumento muito inferior à média das demais regiões metropolitanas brasileiras analisadas na década, caindo da 8ª posição em 2000 para a 15ª em 2010. Enquanto o IDH-Renda, que é calculado a partir da variável “renda *per capita*”, ganhou uma posição no ranking metropolitano, passando da 5ª posição para a 4ª, o IDH-Longevidade, que é calculado a partir da variável “esperança de vida ao nascer”, da RMPA perdeu apenas uma posição, de 1º em 2000 para 2º em 2010.

É possível identificar exatamente quais variáveis foram responsáveis pela queda do IDH-Educação decompondo sua fórmula e analisando todas as variáveis que a compõem separadamente. O IDH-Educação é obtido da média ponderada das variáveis “frequência escolar”, com peso 2/3, e “escolaridade”, com peso 1/3. A escolaridade, medida a partir do percentual de jovens de 18 anos ou mais que possuem ensino fundamental completo, perdeu apenas duas posições no período 2000-2010, passando do 10º lugar para o 12º. Já a frequência escolar perdeu 7 posições no mesmo período, passando do 7º lugar para o 14º dentre as regiões metropolitanas brasileiras do estudo. A variável “frequência escolar” é obtida da média aritmética de outras quatro variáveis: 1) percentual de crianças de 5 a 6 anos que frequentam escola (que figurava a última posição dentre as 16 regiões metropolitanas em 2000 e também em 2010); 2) percentual de jovens de 11 a 13 anos frequentando os anos finais do ensino fundamental ou com ensino fundamental completo (tanto em 2000 quanto em 2010, a RMPA ficou na 2ª posição); 3) percentual de jovens de 15 a 17 anos com ensino fundamental completo (que estava em 4º lugar em 2000 e caiu para 10º em 2010) e; 4) percentual de jovens de 18 a 20 anos com ensino médio completo (que estava em 3º lugar em 2000 e caiu para 12º em 2010, sendo a principal causa da queda de 5 posições da RMPA).

Este trabalho pretende estudar quais características individuais, familiares e locais contribuem para que os alunos possuam ou não o Ensino Fundamental concluído, focando nos alunos de 16 anos das escolas da cidade de Porto Alegre, a partir de técnicas de mineração de dados educacionais (MDE) utilizando os Microdados do Censo Escolar da Educação Básica do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) de 2014. O Censo Escolar da Educação Básica do INEP é produzido e divulgado anualmente com dados educacionais com informações coletadas tanto sobre a infra-estrutura da escola, como sobre o pessoal docente, matrículas, jornada escolar, rendimento e movimento escolar, por nível, etapa e modalidade de ensino, dentre outros, com população alvo: escola, aluno e professor da Educação Básica, referentes à Educação Infantil (Creche e Pré-Escola), Ensino Fundamental, Ensino Médio e Educação Profissional de nível técnico, nas modalidades Regular, Educação de Jovens e Adultos e Educação Especial, vinculados à rede pública (federal, estadual e municipal) e à rede privada.

A hipótese adotada neste trabalho é de que as características dos alunos, das turmas escolares e das escolas influenciam fortemente a conclusão do Ensino Fundamental, exercendo influência sobre o atraso e a evasão escolar. Para testar-se esta hipótese será utilizada uma série de variáveis relacionadas à escola (tais como situação e local de funcionamento, caracterização da infraestrutura -abastecimento de água, energia elétrica, etc.-, disponibilidade de equipamentos didático/pedagógicos, existência de dependências físicas, fornecimento de alimentação escolar, oferta de atendimento educacional especializado e disponibilidade de materiais didáticos de atendimento à diversidade sociocultural), à turma (tais como horário de funcionamento, tipo de atendimento, existência de atividade complementar, atendimento educacional especializado e tipo de disciplinas) e aos alunos (tais como idade, sexo, cor ou raça, uso de transporte escolar, necessidade educacional especial e rendimento escolar), disponíveis nos Microdados do Censo Escolar da Educação Básica do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP).

2. Evasão e repetência escolar

Apesar de que nas últimas décadas tenham sido produzidos inúmeros trabalhos teóricos e empíricos acerca dos efeitos de diversas variáveis sobre os alunos no quesito estudo, existem poucos estudos que se propõem em quantificar as influências, e os poucos pesquisadores que utilizaram métodos estatísticos a fim de quantificar os efeitos das variáveis, optaram por focar em um número reduzido de fatores. Este trabalho pretende, além de reunir os fatores analisados em diferentes estudos, analisar seus efeitos e explorar efeitos talvez ainda não identificados em pesquisas anteriores.

A educação não está ao pleno alcance de todos os cidadãos e tanto a escola quanto a família pouco tem feito pela criança que evade [QUEIROZ 2001], mas recentemente um estudo apontou que no período compreendido entre 1984 e 1997 foi constatado que as características familiares vieram perdendo parte de sua importância para o avanço escolar, mostrando um possível sinal de uma universalização do ensino [LEON e MENEZES-FILHO 2001].

O acesso ao Ensino Fundamental está universalizado no Brasil, mas não sua conclusão, as taxas de repetência e evasão estão subindo no Ensino Médio e a expansão do Ensino Médio estacionou [KLEIN 2006]. O acesso ao ensino fundamental está

praticamente assegurado a todas as crianças e jovens em nosso país, no entanto, a abertura da escola não veio acompanhada de transformações capazes de alterar suas formas de organização e funcionamento, de assegurar condições de trabalho, salário e formação aos professores, comprometendo sua mudança qualitativa [ALMEIDA 2002]. A evasão escolar é um problema crônico no Brasil e tem sido assimilada e tolerada por escolas e sistemas de ensino, chegando ao ponto de admitirem um número mais elevado de matrículas do que o adequado, por contar com o abandono dos estudos por parte de alunos ao longo do ano letivo [DIGIÁCOMO 2011]. Segundo COSTA RIBEIRO [1991 *apud* BORUCHOVITCH 1999] o problema mais sério no ensino do país não é a evasão escolar e sim o alto índice de repetência, pois ainda que o tempo em que muitos alunos permanecem na escola devesse ser suficiente para a conclusão do Ensino Fundamental, devido à repetência muitos desses alunos acabam completando apenas duas ou três séries escolares. O Estado não tem efetivado políticas públicas eficazes que possam garantir melhores condições de trabalho para os docentes que nela atuam [BATISTA 2009].

Estudo anterior que entrevistou alunas que interromperam seus estudos no Programa de Educação de Jovens e Adultos da Universidade Federal de Minas Gerais (UFMG) concluiu que os principais motivos que as levavam a interromperem seus estudos foram externos ao Programa: trabalho, situação econômica, cansaço, relações familiares, etc. [OLIVEIRA 2008]. Outro estudo posterior da mesma autora constatou que a oferta de uma infraestrutura para receber os filhos das estudantes, além de ser um importante incentivo à permanência das mulheres que não tem com quem deixar o(s) filho(s) para estudar, pode ser um fator determinante para a permanência/interrupção dos estudos dessas mulheres [OLIVEIRA 2011].

A partir dos dados obtidos do suplemento de educação da Pesquisa Nacional por Amostra de Domicílios de 2006 sobre as motivações para não estudar de alunos de 10 anos a 17 anos de idade que estavam fora da escola, os resultados apontaram que 67,7% deles alegaram como principal motivo a falta de demanda por educação e apenas 10,9% alegaram deficiências de oferta de estudo, os demais 21,7% não frequentavam a escola por outros motivos [NERI 2009].

Um estudo de caso que pesquisou a evasão escolar em uma comunidade de pequenos produtores rurais imigrantes do norte da Alemanha para o estado do Espírito Santo constatou que para grande parte dos pais dessa comunidade a escola ensina tudo aquilo que não lhes interessa e no futuro ainda leva seus filhos para a cidade, fazendo uma equivalência entre “sair da escola” e “continuar na roça” [BAHIA 2001].

Ceratti (2008) aplicou um questionário para alunos do Centro Estadual de Educação Básica para Jovens e Adultos de Nova Londrina no estado do Paraná a fim de estudar causas e consequências da evasão escolar. O questionário consistia em questões de múltipla escolha com espaço reservado para uma justificativa após cada resposta. Por trás das justificativas apresentadas nas questões, foram identificadas algumas das razões da evasão escolar. As justificativas encontravam-se repletas de explicações e desculpas, apontando medo do fracasso e do esforço, necessidade de preservação da auto-estima, dificuldade de enfrentar dificuldades, dentre outros fatores que revelaram a crença dos alunos de que as razões da evasão escolar estavam neles próprios. Barbeti (2007) apontou para o perigo dos discursos sobre a incapacidade do aluno e sua consequente evasão e fracasso escolar, que podem produzir marcas que prevalecem para além da

escola, passando para suas relações familiares, modificando negativamente as percepções dos pais sobre os filhos e consequentemente o autoconceito dos filhos.

Os resultados de um estudo que pesquisou as causas da evasão escolar no ensino médio regular da rede pública do município de Maracanaú, no Ceará, aplicando um questionário aos alunos, professores e à direção de uma Escola Estadual, concluiu que tanto a direção, quanto os docentes e discentes apontam como causas para evasão escolar: a falta de incentivo da família e da escola, a necessidade de trabalhar, o excesso de conteúdos escolares e as amizades erradas [SOUSA *et al* 2011].

Um estudo sobre evasão e repetência escolar questionou adolescentes evadidos de escolas de Ribeirão Preto, em São Paulo, e apontou a necessidade do trabalho para auxiliar no orçamento familiar e a desmotivação pelo ensino como os fatores que mais contribuíram para o abandono escolar e repetência [CANO *et al* 2006].

A fonoaudióloga Cláudia Alaminos (2005) deparou-se com experiências de exclusão social vividas por alguns de seus pacientes portadores de deficiências. A questão transformou-se em um tema de estudo: se a evasão escolar dos alunos pobres não poderia estar associada ao preconceito contra suas condições econômicas. Do estudo concluiu-se que os principais agentes envolvidos no processo da educação dos alunos, que são os educadores e a família, muitas vezes não enxergam a necessidade de permanência de jovens com menos condições de aprender do que os demais na escola. Assim, foi constatado que um baixo nível socioeconômico da família do estudante pode gerar um preconceito impeditivo para que o direito à educação seja desfrutado por esse aluno.

Recente estudo [TEIXEIRA 2011] avaliou dados de criminalidade e educação nos estados brasileiros entre 2001 e 2005 e constatou que a elevação na taxa de evasão escolar contribui com o aumento da taxa de homicídios um ano mais tarde. As análises concluíram que provavelmente quando os jovens evadidos defrontam-se com problemas como o desemprego, influência negativa de gangues e baixos salários, são incentivados a conduzirem-se para o mundo do crime, ocorrendo uma consequente elevação nas taxas de criminalidade.

3. Mineração de Dados Educacionais

Para analisar a conclusão do Ensino Fundamental será aplicado o processo de Descoberta de Conhecimento em Base de Dados (DCDB), que procura identificar novos padrões válidos, úteis e compreensíveis de comportamento, onde a Mineração de Dados é uma etapa da DCDB que procura extrair conhecimento de grandes volumes de dados para apoiar a análise de perfis de alunos para geração de informações e alertas que possam apoiar atividades preventivas pelos professores, embasando e agilizando as tomadas de decisões que visam contribuir com a redução dos problemas educacionais brasileiros; a técnica também pode ser utilizada integrando múltiplas bases de dados nos estudos [COSTA *et al* 2015]. As técnicas de mineração de dados que serão utilizadas no presente trabalho são a Regra de Classificação J48 e o filtro *CfsSubsetEval*, ambos disponíveis no *software* Weka. As Regras de Classificação têm por objetivo encontrar relacionamentos entre atributos e uma classe, gerando regras que podem posteriormente ser utilizadas para fazer previsões [KAMPFF 2009]. A mineração de dados tem sido muito utilizada em estudos interdisciplinares, envolvendo, principalmente, as áreas de Estatística, Inteligência Artificial e Banco de Dados (Figura 1) [LIN & CERCONE

1997 apud KAMPPFF 2009]. O *software* que será utilizado para a aplicação das ferramentas de mineração de dados propostas será o Weka 3.6 (*software* livre de mineração de dados).

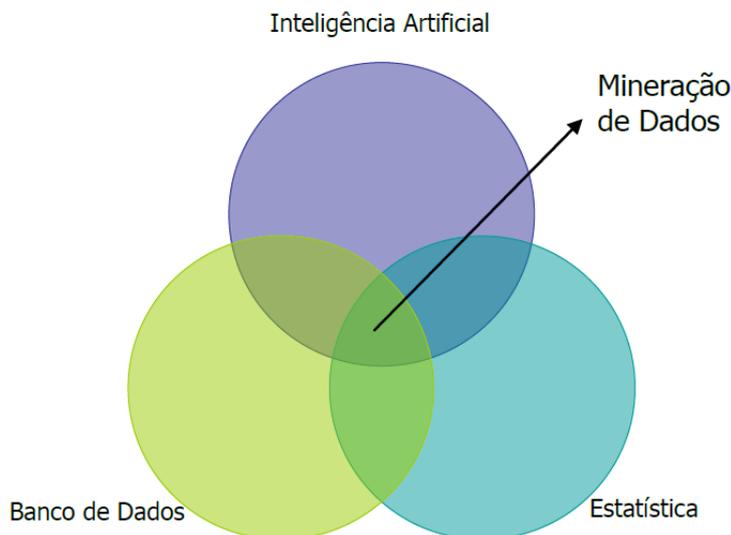


Figura 1. Mineração de Dados: intersecção de várias áreas [KAMPPFF 2009]

As árvores de decisão são construídas do topo para a base e no presente trabalho foi escolhido o algoritmo J48 para avaliar a significância de cada atributo individualmente na construção dos nós da árvore. Cada algoritmo classificador foi executado 10 vezes para os conjuntos de treinamento e de teste, através do método *10-fold cross validation*, e seu desempenho final foi obtido a partir da média dessas execuções.

4. Pré-Processamento, Experimentos Realizados e Análise dos Resultados

Para analisar os perfis dos alunos com e sem o Ensino Fundamental Completo, foram utilizados os Microdados do Censo Escolar da Educação Básica do INEP de 2014. As bases de dados são disponibilizadas na internet em formato .csv. Foi utilizado o software SPSS para agregar a base com informações dos dados das matrículas dos alunos às demais. A variável “PK_COD_ENTIDADE” (Código da Escola) foi utilizada como variável identificadora na agregação da base de dados com informações das escolas à base de dados com as informações das matrículas dos alunos. A variável “PK_COD_TURMA” (Código da Turma) foi utilizada como variável identificadora na agregação da base de dados com informações sobre as turmas escolares à base de dados de matrículas. A base de dados com as informações das matrículas escolares é composta de 85 variáveis, a base dos dados das escolas é composta de 143 variáveis e a base de dados das turmas escolares por 79 variáveis.

Após a agregação das três bases de dados em uma única, foi criada a variável de interesse que identifica se o aluno possui ou não o Ensino Fundamental Completo. Houve um descarte das informações sobre as escolas de fora da cidade de Porto Alegre e das informações dos alunos não nascidos em 1998 (que são os que completaram 16 anos em 2014). O banco de dados final totalizou 21.114 alunos de escolas de Porto Alegre nascidos em 1998.

Foram retiradas do banco de dados todas as variáveis que permitissem a identificação da série que o aluno cursa, as variáveis compostas por códigos (de matrículas, de turmas, de escolas, país de origem do aluno, etc.), variáveis repetidas e variáveis muito específicas, como datas e horários. As variáveis que formaram o banco de dados exportado do SPSS em .csv e importado no Weka totalizaram 176.

O algoritmo de classificação por árvore de decisão J48 foi aplicado à base de dados dividindo-a em dez conjuntos e sendo executado dez vezes pelo método da validação cruzada (10 *fold cross-validation*), que é o padrão nas configurações do ambiente Weka. A acurácia média (*Precision*) obtida nas rodadas foi de 98,58% e a abrangência (*Recall*) foi de 98,6%, ambas medidas atingiram valores muito satisfatórios, conforme a Figura 2. Analisando-se a Matriz de Confusão observa-se que 19.289 alunos foram identificados corretamente pelo algoritmo J48 (11.357 como possuidores do Ensino Fundamental completo e 7.932 como alunos sem o Ensino Fundamental completo) e 276 alunos foram incorretamente classificados.

O algoritmo J48 gerou uma árvore de decisão com 82 nodos. Diante de árvores com muitos nodos, surge a necessidade da utilização de filtros, que permitem a identificação dos atributos mais relevantes. Após a identificação desses atributos, o algoritmo de classificação pode ser rodado novamente somente com esses atributos, gerando árvores de decisão com uma quantidade menor de nodos, o que possibilita a geração de regras mais simples.

Foi aplicado o filtro *CfsSubsetEval* na tentativa de identificar quais atributos têm mais impacto na variável preditiva final (possui ou não o Ensino Fundamental completo). Os filtros têm a característica de avaliar os atributos independentemente do algoritmo de aprendizagem. O algoritmo do filtro foi executado 10 vezes (método 10-*fold cross-validation*) e em todas as execuções identificou os seguintes atributos como os mais influentes na conclusão ou não do Ensino Fundamental: sexo do aluno, dependência administrativa da escola (Federal, Estadual, Municipal ou Privada), destinação do lixo, existência de laboratório de ciências na escola, existência de auditório na escola, disponibilização de internet banda larga, se a escola oferece atendimento educacional especializado, se a turma tem aulas de inglês, se a turma tem aulas de espanhol, se a turma tem outras disciplinas (exceto as obrigatórias).

Foi executado novamente o Algoritmo de Classificação por Árvore de Decisão J48, desta vez considerando somente os dez atributos identificados pelo filtro como mais influentes sobre a conclusão do Ensino Fundamental. A acurácia média obtida foi de 96,17% e a abrangência foi de 96,2% (Figura 2), ambas as medidas sofreram leve redução quando comparadas com a execução anterior que considerou os 176 atributos, porém a árvore de decisão foi gerada com apenas 29 nodos, o que permitiu uma melhor visualização das regras obtidas. Diante da árvore de decisão gerada pelo algoritmo de classificação J48 após a aplicação do filtro, pudemos concluir que os recursos de internet banda larga, laboratório de ciências, auditório na escola e ensino privado apresentaram forte ligação com maiores chances de o aluno apresentar o Ensino Fundamental concluído. Já a ausência desses recursos apresentou ligação com os alunos que não possuem o Ensino Fundamental concluído.

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      19289      98.5893 %
Incorrectly Classified Instances    276        1.4107 %
Kappa statistic                    0.9709
Mean absolute error                0.0231
Root mean squared error            0.1112
Relative absolute error            4.7522 %
Root relative squared error        22.5631 %
Total Number of Instances         19565
Ignored Class Unknown Instances    1549

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  Class
          0.977   0.008   0.989     0.977  N
          0.992   0.023   0.984     0.992  S
Weighted Avg.  0.986   0.017   0.986     0.986

=== Confusion Matrix ===
  a  b  <-- classified as  Number of Leaves : 82
7932 185 | a = N
 91 11357 | b = S          Size of the tree : 163

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      18816      96.1717 %
Incorrectly Classified Instances    749        3.8283 %
Kappa statistic                    0.9208
Mean absolute error                0.0656
Root mean squared error            0.1817
Relative absolute error            13.5145 %
Root relative squared error        36.8738 %
Total Number of Instances         19565
Ignored Class Unknown Instances    1549

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  Class
          0.938   0.021   0.969     0.938  N
          0.979   0.062   0.957     0.979  S
Weighted Avg.  0.962   0.045   0.962     0.962

=== Confusion Matrix ===
  a  b  <-- classified as  Number of Leaves : 29
7612 505 | a = N
 244 11204 | b = S          Size of the tree : 57

```

Figura 2. Resultado obtido utilizando 176 atributos (à esq.) e 10 atributos (à dir.)

As variáveis do banco de dados foram divididas em três grupos: informações do aluno, da escola e da turma. Foi aplicado o filtro *CfsSubsetEval* utilizando o método de teste *cross-validation*, assumindo o valor de 10 *folds* para cada um dos grupos e depois executado o Algoritmo J48 (também utilizando o *10-fold cross validation*) nos três grupos separadamente considerando apenas as variáveis apontadas como as mais influentes na conclusão do Ensino Fundamental dos alunos.

A árvore de decisão resultante da análise dos atributos dos alunos (apontados após a aplicação do filtro) apresentou 3 nodos e associou o fato dos alunos apresentarem necessidades especiais com a não conclusão do Ensino Fundamental e o aluno de cor ou raça branca com a conclusão do Ensino Fundamental, com acurácia média de 63,11%. A árvore de decisão resultante da análise dos atributos das escolas (somente os identificados pelo filtro como mais significativos) apresentou 14 nodos e associou o fato das escolas não possuírem laboratório de ciências, auditório e internet banda larga com a não conclusão do Ensino Fundamental e a dependência administrativa da escola ser privada com a conclusão do Ensino Fundamental, com acurácia média de 81,56%. A árvore de decisão resultante da análise dos atributos das turmas (após aplicação do filtro) apresentou 8 nodos e associou o fato das turmas usufruírem de aulas de inglês, espanhol, artes e outras disciplinas não obrigatórias com a conclusão do Ensino Fundamental, com acurácia média de 91,64%.

5. Considerações Finais

Ao trabalhar-se com um banco de dados rico em variáveis é interessante efetuar a divisão dos atributos em classes para agregar mais informações às conclusões do estudo. Se for observada somente a precisão dos resultados, pode-se considerar satisfatória uma árvore de decisão com acurácia média de 96,17%, mas ao ir-se além nas análises e considerar a perda de acurácia como válida diante do ganho de informação, torna-se possível desfrutar de um ganho em conclusões.

A acurácia média obtida da árvore de decisão resultante da aplicação do filtro sobre os 176 atributos da base de dados foi de 96,17%, gerando uma árvore com 29 nodos e associando recursos de internet banda larga, laboratório de ciências, auditório

na escola e ensino privado com maiores chances de o aluno apresentar o Ensino Fundamental concluído. Ao dividir-se o banco de dados em tipos de atributos, adicionamos às conclusões mais informações relevantes à pesquisa: foram encontradas evidências de que necessidades especiais dos alunos estejam ligadas com a não conclusão do Ensino Fundamental e de que a cor/raça branca, aulas de inglês, espanhol, artes e outras disciplinas não obrigatórias estejam associadas à conclusão do Ensino Fundamental.

Diante dos resultados encontrados, também podemos concluir que é interessante a aplicação de filtros, mesmo diante de uma diminuição da acurácia dos resultados, pois a vantagem proporcionada pela obtenção de regras mais simples possibilitam a melhor tomada de decisões e a visualização com mais clareza dos atributos que mais exercem influência sobre a variável de interesse.

Para trabalhos futuros fica a sugestão do estudo de viabilização da utilização das demais variáveis do Censo Escolar não utilizadas nesse trabalho. A análise dessas variáveis pode ser feita em conjunto com especialistas da área de educação que possam auxiliar em agregações, que possibilitem seu uso a fim de agregar informação relevante ao estudo.

Referências

- Alaminos, Cláudia. (2005) Evasão escolar na adolescência: necessidade ou ideologia. Proceedings of the 1th Simpósio Internacional do Adolescente.
- Almeida, Maria Isabel de. (2002) Ações organizacionais e pedagógicas dos sistemas de ensino: políticas de inclusão.
- Bahia, Joana. (2001) A "lei da vida": confirmação, evasão escolar e reinvenção da identidade entre os pomeranos. Educação e Pesquisa, v. 27, n. 1, p. 69-82.
- Barbeti, Renato de Souza. (2007) A evasão escolar e seus significados para alunos, professores e família. 5ª Mostra Acadêmica UNIMEP. Universidade Metodista de Piracicaba.
- Batista, Santos Dias; Souza, Alesxsandra Matos; Oliveira, Júlia Mara da Silva. (2009) A evasão escolar no ensino médio: um estudo de caso. Revista Profissão Docente, UNIUBE. Uberaba/MG.
- Boruchovitch, Evely. (1999) Estratégias de aprendizagem e desempenho escolar: considerações para a prática educacional.
- Brasil. PNAD 2007. Instituto Brasileiro de Geografia e Estatística. Pesquisa nacional por amostra de domicílios: aspectos complementares da Educação de Jovens e Adultos e Educação Profissional.
- Cano, Maria Aparecida Tedeschi; Ferriani, Maria das Graças, C.; Mendonça, Mônica Luis. (2006) Repetência e evasão escolar de adolescentes em Ribeirão Preto-SP: uma primeira abordagem. Revista Eletrônica de Enfermagem, v. 1, n. 1.
- Ceratti, Márcia Rodrigues Neves. (2008) Evasão escolar, causas e consequências. Curitiba/PR.

- Costa, Susane Santos da; Cazella, Silvio; Rigo, Sandro José. (2015) Minerando dados sobre o desempenho de alunos de cursos de educação permanente em modalidade EAD: Um estudo de caso sobre evasão escolar na UNA-SUS. *RENOTE*, v. 12, n. 2.
- Digiácomo, Murillo José. (2011) Evasão escolar: não basta comunicar e as mãos lavar. *V. 23, n. 11, p. 2012*.
- Kampff, Adriana Justin Cerveira. (2009) Mineração de dados educacionais para geração de alertas em ambientes virtuais de aprendizagem como apoio à prática docente. Tese (Doutorado em Informática na Educação)-Universidade Federal do Rio Grande do Sul-UFRGS, Porto Alegre.
- Klein, Ruben. (2006) Como está a educação no Brasil? O que fazer? Ensaio: avaliação de políticas públicas na educação, Rio de Janeiro, v.14, n.51, p. 139-172.
- Leon, Fernanda Leite Lopez de; Menezes-Filho, Naércio Aquino. (2001) Reprovação, avanço e evasão escolar no Brasil. *Cultura*, v. 1998, n. 1999.
- Martins, C. M. dos R.; Ferreira, G. da S. (2014) Índice de Desenvolvimento Humano (IDH) na Região Metropolitana de Porto Alegre (RMPA) — 2000-10. Carta de conjuntura Fundação de Economia e Estatística, ANO 23, Nº 12, PG. 7. Porto Alegre.
- Neri, Marcelo *et al.* (2009) Motivos da evasão escolar. Brasília: Fundação Getulio Vargas,
- Oliveira, Paula Cristina Silva de; Eiterer, Carmem Lúcia. (2008) Evasão escolar de alunos trabalhadores na EJA. Faculdade de educação/UFMG.
- Oliveira, Paula Cristina Silva de; (2011) Alfabetizando/as na EJA: as razões da permanência nos estudos. Diss. Dissertação (Mestrado em Educação)-Programa de Pós-Graduação em Educação, Faculdade de Educação, Universidade Federal de Minas Gerais, Belo Horizonte, MG.
- Pnud; Ipea; Fjp. (2014) Atlas do Desenvolvimento Humano nas Regiões Metropolitanas Brasileiras. Brasília.
- Queiroz, Lucileide Domingos. (2001) Um estudo sobre a evasão escolar: para se pensar na inclusão escolar. Associação Nacional de Pós-Graduação e Pesquisa em Educação - ANPAD.
- Rigo, S. J.; Cambruzzi, W.; Barbosa, J. L. V.; Cazella, Sílvio. (2014) Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios. *Revista Brasileira de Informática na Educação*, v. 22, p.132-146.
- Sousa, Antonia de Abreu *et al.* (2011) Evasão escolar no ensino médio: velhos ou novos dilemas? *Vértices*, v. 13, n. 1, p. 25-36.
- Teixeira, E. C. (2011) Dois ensaios acerca da relação entre criminalidade e educação. Tese (Doutorado em Economia Aplicada)-Escola Superior de Agricultura “Luiz de Queiroz” – ESALQ, Piracicaba, SP.