

**Comportamento da nota de corte do Prouni
e a sua relação com a política de acesso
ao ensino superior:
um estudo com Mineração de Dados**

**Prouni approval score behavior and its relation
with the policy of access to higher education:
a study with Data Mining**

**Comportamiento del puntaje de aprobación del Prouni
y su relación con la política de acceso
a la enseñanza superior:
un estudio con Minería de Datos**

Rodrigo Oliveira Zacarias

Instituto Federal Fluminense (IF Fluminense), Campos dos Goytacazes/RJ – Brasil

Leila de Matos Abreu

Instituto Federal Fluminense (IF Fluminense), Campos dos Goytacazes/RJ – Brasil

Henrique Rego Monteiro da Hora

Instituto Federal Fluminense (IF Fluminense), Campos dos Goytacazes/RJ – Brasil

Aline Pires Vieira de Vasconcelos

Instituto Federal Fluminense (IF Fluminense), Campos dos Goytacazes/RJ – Brasil

Resumo

Este artigo tem objetivo de analisar os fatores que influenciam na variação da nota de corte das bolsas ofertadas para cursos de graduação no Prouni e sua relação com a política de acesso ao ensino superior. Foram obtidos 41.437 registros de cursos com bolsas ofertadas no 1º semestre de 2018 na plataforma Brasil.io. Foi aplicado o algoritmo de classificação PART para a mineração dos dados, que gera regras por meio da indução de conhecimento. Os resultados indicaram que, nos menores intervalos, predominaram os cursos ofertados na modalidade a distância, de grau tecnológico e com mensalidades menores que R\$ 300,00 enquanto nos demais intervalos estavam presentes os cursos presenciais, de grau bacharelado e com mensalidade superior a R\$ 1.000,00. Isso auxilia o estudo de aspectos relacionados à inclusão social, pluralidade, destinação de recursos públicos, políticas públicas, entre outras dimensões relacionadas ao programa.

Palavras-chave: Mineração de dados educacionais, Prouni, Algoritmo PART, Política pública, Recursos públicos

Abstract

This article aims to analyze the factors that influence the variation in the scholarship approval score offered for the undergraduate courses in Prouni and its relation with the Higher Education access policy. There were 41,437 records of courses with scholarships offered in the first half of 2018 on the Brasil.io platform. We applied the classification algorithm PART for the data mining, which generates rules through the knowledge induction. The results indicated that in shortest periods there were a predomination of the distance modality technological degree courses, and the ones with monthly payments lower than R\$ 300,00, while in other intervals the baccalaureate degree presence courses predominated, and the ones with monthly payment higher than R\$ 1,000.00. This helps the study of aspects related to social inclusion, plurality, public resource allocations, public policies, among other dimensions related to the program.

Keywords: Educational Data Mining, Prouni, Algorithm PART, Public policy, Public resources

Resumen

Este artículo pretende analizar los factores que influyen en la variación del puntaje de aprobación de las becas ofrecidas para cursos de graduación en el PROUNI y su relación con la política de acceso a la enseñanza superior. Se obtuvieron 41.437 registros de cursos con becas ofrecidas en el 1er semestre de 2018 en la plataforma Brasil.io. Se aplicó el algoritmo de clasificación PART para la minería de los datos, que genera reglas por medio de la inducción de conocimiento. Los resultados indicaron que, en los menores intervalos, predominaron los cursos ofrecidos en la modalidad a distancia, de grado tecnológico y con mensualidades menores que R\$ 300,00 mientras en los demás intervalos estaban presentes los cursos presenciales, de grado bachillerato y con una mensualidad superior a R\$ 1.000.00. Esto ayuda el estudio de aspectos relacionados a la inclusión social, pluralidad, asignación de recursos públicos, políticas públicas, entre otras dimensiones relacionadas con el programa.

Palabras clave: Minería de datos educacionales, Prouni, Algoritmo PART, Política pública, Recursos públicos

1. Introdução

Nos últimos anos, houve avanços consideráveis na expansão e na regionalização de oportunidades de acesso ao ensino superior no Brasil, tanto em instituições públicas quanto particulares. Em relação aos estabelecimentos privados de ensino, podem-se destacar as seguintes políticas públicas afirmativas: o Financiamento Estudantil (Fies) e o Programa Universidade para Todos (Prouni) (CASALI; MATTOS, 2015).

O Prouni é um programa criado pelo Governo Federal, por meio da Medida Provisória nº 213 de 10 de setembro de 2004, e institucionalizado pela Lei nº 11.096, em 13 de janeiro de 2005. Essa política tem a finalidade de

conceder bolsas de estudos integrais e parciais, em cursos de graduação e sequenciais de formação específica, em instituições de ensino superior privadas. O programa é direcionado aos estudantes egressos do ensino médio da rede pública ou particular na condição de bolsistas integrais, com renda familiar *per capita* máxima de três salários mínimos. Os candidatos são selecionados conforme seus respectivos desempenhos pelas notas obtidas no Exame Nacional do Ensino Médio (Enem). A partir dessas notas, é determinada a nota de corte das bolsas ofertadas pelo programa, sendo a nota mínima necessária para a aprovação (BRASIL, 2019).

Segundo Moraes e Telles (2018), o Prouni é um marco histórico para o ensino superior no Brasil, pois permitiu a implantação de uma política pública que favorecesse estudantes oriundos de famílias cuja renda *per capita* não exceda um salário mínimo e meio, ou seja, de grupos populares. Nesse sentido, o programa tem sido objeto de estudo em diversos campos da educação, sociologia e direito, abordando aspectos relacionados à inclusão social, pluralidade, destinação de recursos públicos, políticas públicas sociais, entre outros.

Casali e Mattos (2015) salientam a existência de diversos debates de cunho ideológico sobre o Prouni, considerando quatro pontos críticos: políticas públicas, política de inclusão social, isenção de impostos e percepção e impactos segundo os beneficiários. As discussões se concentram nos impactos das práticas de inclusão social e meritocracia, assim como a respeito da constitucionalidade da isenção das instituições de ensino superior privadas que disponibilizam as bolsas para os alunos do programa.

Como política social de inclusão, Moreira (2017) avalia o Prouni como estratégia de promoção de democratização de acesso ao nível superior de ensino e os perfis dos alunos contemplados, concluindo por meio dos dados socioeducacionais coletados no Exame Nacional de Desempenho dos Estudantes (Enade). Nesse estudo, verifica-se que o programa cumpre o objetivo de inserir estudantes de baixa renda no ensino superior, porém, ao mesmo tempo, o impacto do Prouni é inversamente proporcional à massificação do curso. Por exemplo, há promoção de maior inclusão em cursos de medicina do que de pedagogia, embora a pluralidade de perfis de alunos pagantes seja maior no primeiro curso.

Esse fenômeno impacta diretamente a procura das vagas e a existência de vagas ociosas em diversos cursos espalhados pelo país. Isso desencadeia uma discussão acerca dos recursos públicos destinados ao programa, bem como à política de isenção de impostos para as instituições participantes. Costa e Ferreira (2017) apontam que, em média, 30% de todas as vagas ofertadas estão ociosas e que o custo é R\$ 281,78/mês por bolsista. Ainda destacam que as isenções concedidas impactam significativamente a arrecadação do governo, que a ociosidade pode afetar a otimização de investimentos e que a discrepância na distribuição dos bens educacionais entre alunos contemplados e pagantes induz os alunos de camadas mais populares para os cursos menos procurados e/ou com baixa qualidade.

Esses tópicos de discussão incidem na criação de medidas que melhorem a eficiência do Prouni como política de acesso ao ensino superior. A utilização de métodos e/ou ferramentas computacionais, como Data Mining ou Mineração de Dados (MD), podem auxiliar na obtenção de informações valiosas para esse processo.

A MD é o processo que consiste na preparação e análise de dados; identificação de anomalias, padrões e correlações em grandes conjuntos de dados, para prever resultados ou detectar possíveis comportamentos, uma excelente ferramenta para tomadas de decisões em algum processo. A MD é uma das etapas do processo de Descoberta de Conhecimento em Base de Dados, ou KDD – *Knowledge Discovery in Databases*. Ele é composto pelas etapas de seleção, pré-processamento, transformação, mineração de dados e interpretação/avaliação (AHLEMEYER-STUBBE; COLEMAN, 2014; FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). As tarefas de MD mais conhecidas e utilizadas são: agrupamento, classificação, associação e regressão (HORA et al., 2018).

Uma área em constante desenvolvimento para exploração de dados oriundos de ambientes educacionais é a Mineração de Dados Educacionais (MDE), em inglês *Educational Data Mining* (EDM) (BAKER; ISOTANI; CARVALHO, 2011). A MDE aplica diversos métodos estatísticos, aprendizado de máquina e algoritmos de MD, para análise de diferentes tipos de dados educacionais, permitindo a extração de novos conhecimentos e proporcionando

subsídios para a execução de processos educacionais mais eficazes (ROMERO; VENTURA, 2010).

Nesse contexto, o objetivo deste trabalho é investigar os fatores que influenciam na variação de notas de cortes das bolsas ofertadas pelas instituições de ensino superior privadas, por meio do Prouni, a partir dos dados presentes na base do programa, por meio da aplicação da técnica de classificação da MD. Esta pesquisa visa a contribuir para a extração de informações que auxiliem no aumento da eficiência do Prouni enquanto política pública afirmativa para acesso ao nível superior de ensino.

Na primeira seção deste artigo, descreve-se o contexto do Prouni, as principais temáticas de discussão na área e a possibilidade de aplicação de MDE para complementá-la. Na segunda seção, são apresentados trabalhos correlatos envolvendo a MDE. A terceira apresenta os procedimentos metodológicos e, na seguinte, a análise e a discussão sobre os resultados obtidos. Na última seção, descrevem-se as considerações acerca do trabalho.

2. Trabalhos relacionados

Para Baker et al. (2011), a MDE contribui de forma significativa para as diversas atividades do ambiente educacional. No campo de ensino, auxilia a investigação de fatores que promovem a melhoria da aprendizagem do aluno, bem como os motivos que levam à evasão de estudantes. Na gestão escolar, a MDE pode identificar possibilidades para implementação de ações institucionais mais eficazes e para otimização de investimentos de recursos financeiros em ações pedagógicas.

Nesse contexto de MDE aplicada a bases governamentais, alguns autores relatam a obtenção de resultados satisfatórios. Gualhano et al. (2018) apresentam a MD dentro do contexto educacional, com objetivo de extrair conhecimento da avaliação dos programas de pós-graduação *stricto sensu* para identificar os principais quesitos que impactam a nota final do programa. Para os autores, o conhecimento obtido auxiliará o planejamento estratégico e a vantagem competitiva da instituição. Como resultado, foi identificado que a produção intelectual nesses programas é um dos itens mais importantes para manutenção do seu credenciamento.

Maschio et al. (2018) trazem um panorama sobre as técnicas de MD mais utilizadas em bases de dados educacionais no contexto brasileiro. Grande parte dos trabalhos aplica o aprendizado de máquina como a técnica principal para alcance do objetivo que, em sua maioria, são voltados para verificação de desempenho. Os autores ainda apresentam que a indução de conhecimento, método utilizado neste artigo, não é aplicada em larga escala. Ainda se destaca que grande parte dos trabalhos são voltados para problemas relacionados ao ensino médio e ensino superior, o que evidencia a relevância da temática em estudo, porém não apresentam trabalhos específicos com dados minerados oriundos de programas de fomento educacional, como o Prouni.

3. Metodologia

A partir dos conceitos apresentados por Gil (2017), esta pesquisa é classificada como aplicada em sua natureza, abordando aspectos quantitativos, com base nas ferramentas estatísticas e dados numéricos utilizados, e exploratórios, pelo intuito da geração de resultados ainda desconhecidos. A estratégia de investigação foi organizada e estruturada em quatro etapas sequenciais, conforme apresentado na Figura 1.

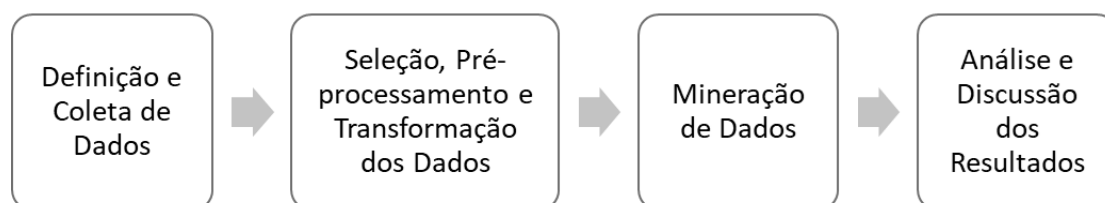


Figura 1. Etapas da estratégia de pesquisa. Fonte: Elaboração Própria.

3.1 Definição e Coleta de Dados

A coleta de dados foi realizada na plataforma Brasil.io, um repositório de dados públicos, em fase beta, referente a diversas temáticas do governo brasileiro, a fim de atender à Lei de Acesso à Informação - Lei 12.527/2011 (BRASIL, 2011). Para o objetivo deste trabalho, foi selecionada a base de dados intitulada “Cursos e notas de corte do Prouni 2018” (BRASIL, 2018). Por meio de um arquivo de extensão CVS (Valores Separados por Vírgula, no inglês), foram obtidos 41.437 registros, subdivididos em 16 atributos, relativos aos cursos de

graduação de instituições particulares com bolsas ofertadas no programa, durante o 1º semestre de 2018, em todo o território nacional.

O Quadro 1 apresenta uma breve descrição do conteúdo de cada atributo que subdivide cada registro da base de dados coletada.

Quadro 1. Atributos dos registros da base de dados coletada e suas respectivas descrições de conteúdo

Atributo	Conteúdo
UF	Unidades Federativas do Brasil
Cidade	Cidade onde se encontram as IES privadas
Universidade	Nome da IES privada
Nome do Campus	Nome do campus da IES privada
Curso	Nome do curso de graduação
Grau	Grau do curso de graduação (Licenciatura, Bacharelado ou Tecnologia)
Turno	Turno de funcionamento do curso
Mensalidade	Valor de custo do curso
Bolsa Integral Ampla	Número de bolsas disponíveis nesta modalidade
Bolsa Integral Cota	Número de bolsas disponíveis nesta modalidade
Bolsa Parcial Ampla	Número de bolsas disponíveis nesta modalidade
Bolsa Parcial Cota	Número de bolsas disponíveis nesta modalidade
Nota de Corte da Bolsa Integral Ampla	Valor de nota de corte nesta modalidade
Nota de Corte da Bolsa Integral Cota	Valor de nota de corte nesta modalidade
Nota de Corte da Bolsa Parcial Ampla	Valor de nota de corte nesta modalidade
Nota de Corte da Bolsa Parcial Cota	Valor de nota de corte nesta modalidade

Fonte: Elaboração Própria.

3.2 Seleção, pré-processamento e transformação dos dados

Nesta etapa, os dados coletados são preparados para o processo de mineração. Os mesmos podem ser generalizados e agregados, ou até mesmo novos atributos podem ser criados ou excluídos para melhor compreensão do domínio do problema. Um processo comum é a discretização ou transformação dos dados em categorias ou intervalos quantitativos, para melhor interpretação do domínio do problema (COSTA; BERNARDINI; VITERBO FILHO, 2014).

No fim dessa etapa, a base de dados passou a ter uma nova configuração de atributos. O Quadro 2 apresenta os novos atributos, juntamente com seus

respectivos domínios. Todos os atributos relativos a “Bolsas” e “Notas de Corte” possuem o mesmo domínio, por isso, aparecem uma única vez no quadro, mas representam respectivamente a quatro desse tipo na base de dados, conforme a modalidades de bolsas disponíveis no Prouni. Assim, a base transformada apresenta 13 atributos por registro.

Quadro 2. Atributos de descrição e seus respectivos domínios da base de dados processada – ano de 2018

Atributo	Domínio do Atributo
Região	Norte; Nordeste; Centro-Oeste; Sudeste e Sul
Classificação	Negócios, administração e direito; Agricultura, silvicultura, pesca e veterinária; Engenharia, produção e construção; Tecnologias da informação e comunicação; Artes e humanidades; Ciências sociais, jornalismo e informação; Ciências naturais matemática e estatística; Saúde e bem-estar; Educação; e Serviços
Grau	Bacharelado; Tecnológico; e Licenciatura
Turno	Curso a Distância; Noturno; Vespertino; Matutino; e Integral
Mensalidade	Menor que 300; De 300 a 699; De 700 a 999; e Maior que 1000
Bolsas	Nenhuma bolsa; 1 bolsa; De 2 a 5 bolsas; e Mais de 5 bolsas
Notas de corte	Nenhuma bolsa; De 500 a 599 pontos; De 600 a 699 pontos; Menor que 500 pontos; e A partir de 700 pontos

Fonte: Elaboração Própria.

Os atributos “UF”, “Cidade”, “Universidade” e “Nome do Campus” foram transformados em “Região do Brasil”, com agrupação de cada estado em sua respectiva região, objetivando uma visão espacial macro das características que influenciam a nota de corte.

O atributo “Nome do Curso” acarretava numa grande sobrecarga de informações no momento de mineração, gerando resultados que contribuíam de forma superficial para o objeto de estudo. Por isso, optou-se por substituir esse atributo por sua classificação no Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), com base nas áreas apresentadas e de acordo com as especificidades apresentadas no Manual do CINE BRASIL (2018).

Para o atributo “Mensalidade” e os demais relativos a “Bolsas” e “Notas de Corte”, foi realizada a transformação discretizada, criando intervalos de valores para esses dados. Assim, o número real que representava o dado de cada um desses atributos nos registros foi substituído pelo intervalo ao qual está

inserido. Com isso, foram obtidos resultados de fácil legibilidade e compreensão, facilitando o alcance do objetivo do trabalho.

3.3. Mineração de Dados

Para avaliar os conhecimentos gerados com o processamento da base de estudo, a MD foi realizada utilizando aprendizado supervisionado para problemas de classificação. Foi aplicado o algoritmo de classificação PART, que oferece como saída um conjunto de regras de fácil interpretação por seres humanos, por meio do *software* WEKA (*Waikato Environment for Knowledge Analysis*).

O algoritmo PART foi desenvolvido com base no algoritmo J48. Por meio da técnica dividir-para-conquistar, é gerada uma lista de decisão, que origina uma árvore de decisão C4.5 a cada iteração. A partir dela, as regras são induzidas e refinadas. Para cada regra gerada, há uma estimativa de sua cobertura sobre as instâncias da base de dados. Esse processo é realizado inúmeras vezes, até que todas as instâncias sejam abrangidas. As regras com mais instâncias cobertas são apresentadas ao usuário, de forma legível e de fácil compreensão, e as demais descartadas (FRANK; WITTEN, 1998).

Além das características mencionadas, outro fator que contribuiu para a escolha do PART foi o fato de, durante os testes, esse algoritmo ter sido o que proporcionou uma maior porcentagem de instâncias corretamente classificadas na base de dados selecionada, em relação aos demais algoritmos de classificação.

No momento de mineração, foram realizadas quatro formas de processamento dos dados, tendo como desfecho, ou o resultado esperado, sempre um dos atributos de nota de corte. A Figura 2 apresenta um exemplo de um desses processamentos realizados no WEKA, em que o desfecho escolhido é a Nota de Corte de Bolsa Integral em Ampla Concorrência, uma das modalidades de assistência financeira para uma vaga ofertada no Prouni.

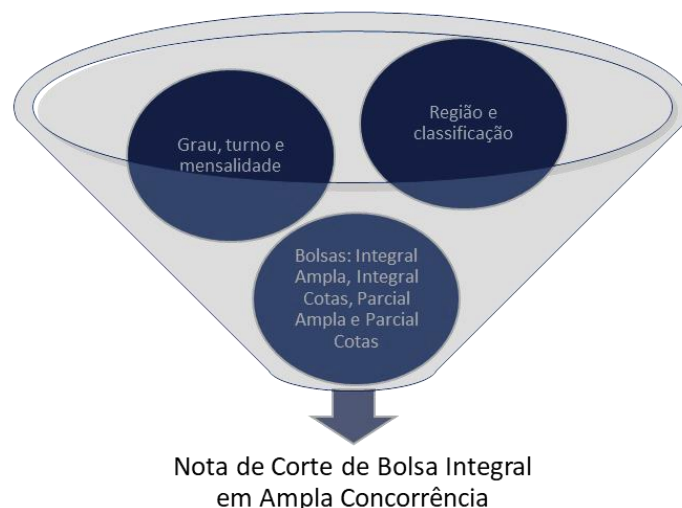


Figura 2. Exemplo de um dos processamentos realizados no WEKA.
Fonte: Elaboração Própria.

Os demais atributos de nota de corte eram excluídos temporariamente da base durante cada processo, pois se optou por analisar o fenômeno da variação de forma isolada, uma vez que as notas de corte das demais modalidades não estavam disponíveis para o estudante no momento de escolha do curso. Nesse sentido, isso não influenciou em sua decisão, ao contrário das demais nove características, representadas pelas bolinhas na figura. O algoritmo PART foi aplicado, utilizando as configurações e os parâmetros padrões do WEKA, tendo como configurações de teste o *cross-validation* (validação cruzada) com 10 *folds* (partições de dados), de forma a avaliar a capacidade de generalização do modelo.

Em cada processamento, foram selecionadas as regras de instâncias classificadas corretamente, que apresentavam pelo menos 100 ocorrências e precisão acima de 70%. As regras foram organizadas pelos intervalos do domínio de nota de corte.

Por fim, ao término desse refinamento, as regras restantes foram submetidas a uma análise, a fim de verificar quais valores dos atributos eram mais frequentes. Por exemplo, se grande parte das regras apresentaram o atributo turno com valor “Curso a distância”, pode-se inferir que esse valor/fator teve grande influência na escolha do candidato naquele intervalo de nota de corte. Assim, foram evidenciados todos os valores de atributos que tiveram

influência em cada intervalo de nota de corte da modalidade de bolsa selecionada como desfecho.

Na última etapa, após todo o processamento de dados realizado em todas as modalidades de bolsa, os resultados foram consolidados. Com isso, foram feitas a análise e a discussão sobre as informações obtidas, os trabalhos correlatos levantados e possíveis contribuições na área de MDE.

4. Análise e discussão dos resultados

Nesta seção, são apresentados os resultados obtidos por meio da MD, realizada na base “Cursos e notas de corte do Prouni 2018”. Ela contém um total de 41.437 registros de cursos de graduação, capturados pela plataforma Brasil.io, em 08 de fevereiro de 2018. Estatisticamente, as Figuras 3 e 4 apresentam uma visão macro das características dos cursos.

No 1º semestre de 2018, 40,3% dos cursos de graduação cadastrados com bolsas no Prouni localizavam-se na região sudeste do país, seguida pela região sul, com 22,1%. A região nordeste contabilizava 17,2% dos cursos, enquanto a centro-oeste, 10,1%, e a norte, 7,2%. Quanto ao grau, destacam-se os Bacharelados (46,2%), seguidos pelos Tecnológicos (32,5%) e, por último, as Licenciaturas (21,2%). A maior parte dos cursos está classificada na área de “Negócios, administração e direito”, seguidos pelas de “Saúde e bem-estar” e “Engenharia, produção e construção”, que somam pouco mais de 25.000 cursos cadastrados.

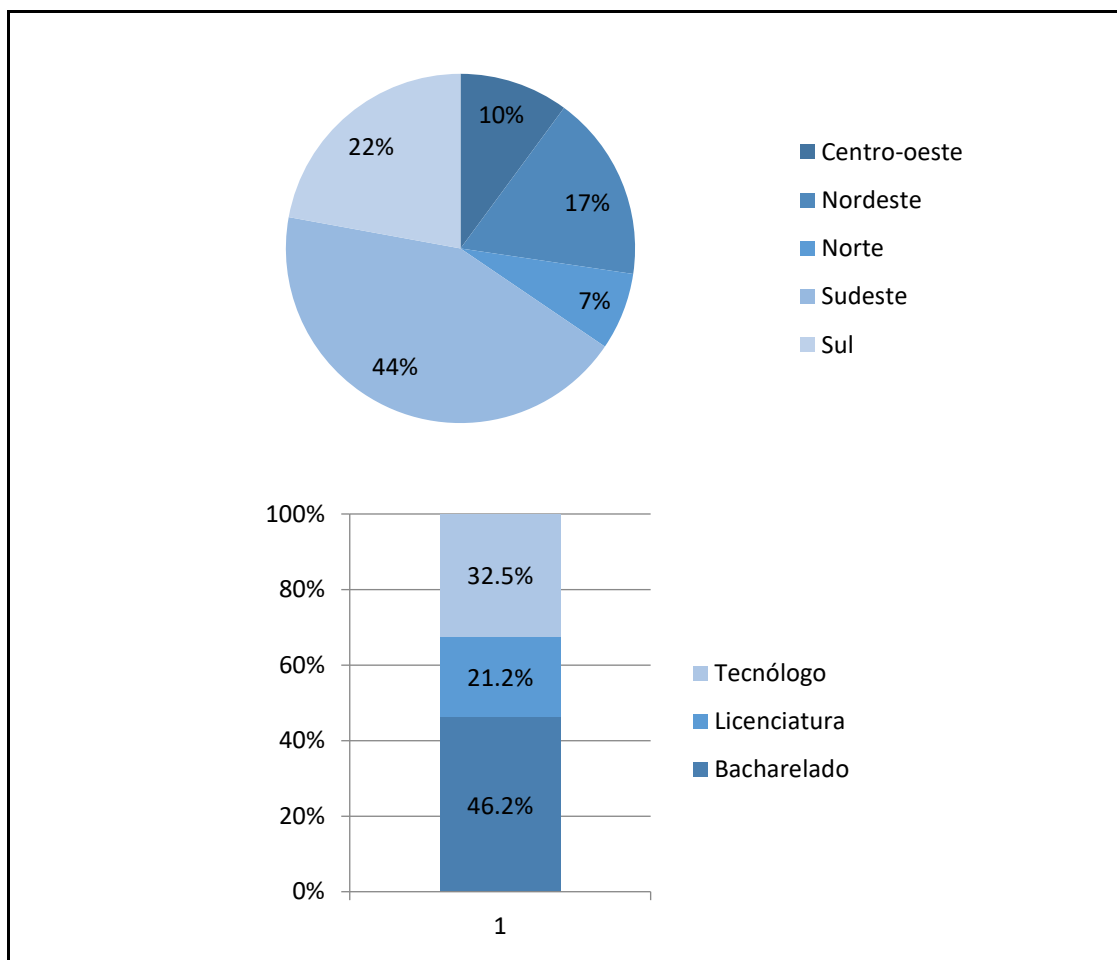


Figura 3. Dados estatísticos referentes a regiões e graus dos cursos.
Fonte: Elaboração Própria.

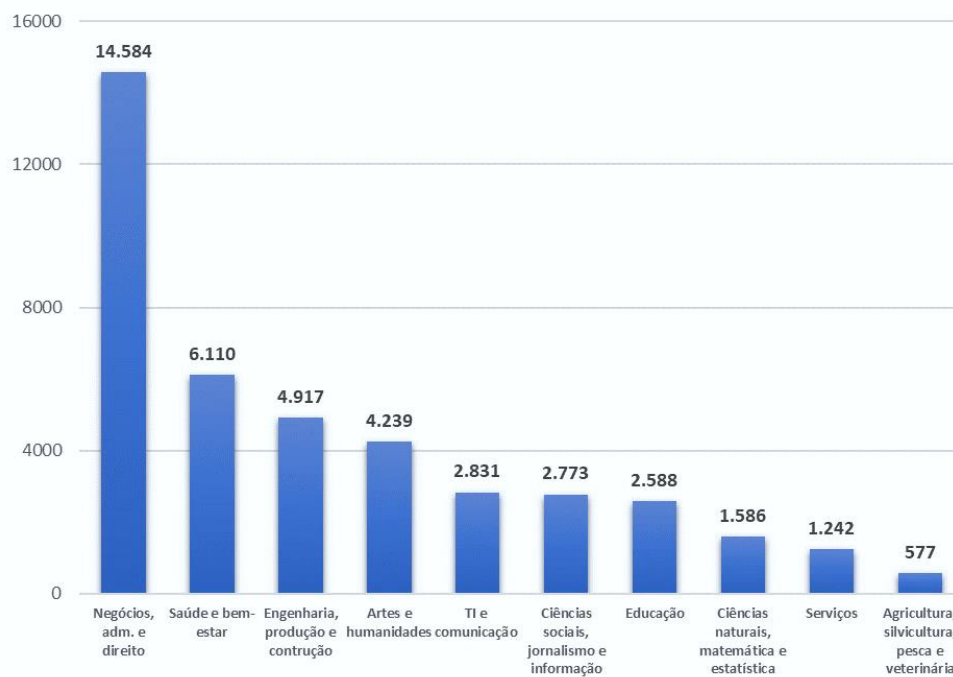


Figura 4. Dados estatísticos referentes às áreas de classificação do Inep para os cursos.
Fonte: Elaboração Própria.

Conforme mencionado anteriormente, foi aplicado o algoritmo de classificação PART para a mineração dos dados, utilizando as configurações e os parâmetros padrões do WEKA, tendo como configurações de teste o *Cross-validation* com 10 *folds*. Foram realizados 04 (quatro) processamentos com os seguintes desfechos: (1) Nota Bolsa Integral Ampla Concorrência, (2) Nota Bolsa Integral Cotas, (3) Nota Bolsa Parcial Ampla Concorrência e (4) Nota Bolsa Parcial Cotas.

A Tabela 1 apresenta as porcentagens de Acurácia e *F-Measure* obtidas pelos algoritmos em cada um dos desfechos. Segundo Baeza-Yates e Ribeiro-Neto (2013), a Acurácia representa a quantidade de instância de entrada que foram classificadas corretamente, conforme a associação entre as instâncias e as classes pré-estabelecidas. Já o *F-Measure* representa uma média harmônica entre precisão e cobertura.

Tabela 1. Valores de Acurácia e F-Measure obtidos pelo algoritmo PART em cada desfecho.

Desfecho	Quantidade de Regras	Acurácia	<i>F-Measure</i>
(1)	798	75,9%	74,1%
(2)	583	87,1%	86,7%
(3)	638	84,8%	84,4%
(4)	353	95,5%	95,3%

Fonte: Elaboração Própria.

Os altos valores de porcentagens alcançados nos parâmetros apresentados indicam que o nível de confiabilidade dos resultados obtidos é satisfatório, o que permite a realização de inferências que condizem com a realidade do objeto de estudo.

Na Tabela 2, são apresentadas as regras selecionadas em cada um dos desfechos apresentados (1ª coluna), com base nos critérios mínimos de, pelo menos, 100 ocorrências e precisão acima de 70%. A coluna “Regra” apresenta um conjunto de características, com base nos domínios de atributos do Quadro 2, que se repetem para aquele desfecho em um determinado intervalo de nota de corte. A coluna “Resultado” mostra o intervalo da regra com a quantidade de ocorrências de instâncias classificadas (primeiro número) e quantas foram

classificadas incorretamente (segundo número) pelo algoritmo PART. Com base nesses números, é calculada a precisão de regra apresentada na última coluna.

Tabela 2. Conjunto de regras selecionadas por desfecho.

Desfecho	Regra	Resultado	Precisão
(1)	Bolsa Integral Cotas = 1 bolsa E Mensalidade = De 300 a 699 E Grau = Tecnológico E Bolsa Integral Ampla = De 2 a 5 bolsas	De 500 a 599 pontos (555.0/103.0)	81,4%
	Bolsa Integral Cotas = 1 bolsa E Turno = Curso a Distância E Região = Nordeste E Grau = Tecnológico E Bolsa Integral Ampla = Nenhuma bolsa	De 500 a 599 pontos (448.0/108.0)	75,9%
	Bolsa Integral Ampla = 1 bolsa E Turno = Matutino E Mensalidade = Maior que 1000 E Bolsa Parcial Cotas = Nenhuma bolsa	De 600 a 699 pontos (395.0/55.0)	86,1%
	Bolsa Integral Ampla = 1 bolsa E Turno = Noturno E Grau = Bacharelado E Mensalidade = Maior que 1000 E Bolsa Integral Cotas = De 2 a 5 bolsas	De 600 a 699 pontos (285.0/32.0)	88,8%
	Bolsa Integral Ampla = De 2 a 5 bolsas E Turno = Noturno E Mensalidade = Maior que 1000 E Grau = Bacharelado E Região = Sudeste E Bolsa Parcial Ampla = Nenhuma bolsa	De 600 a 699 pontos (220.0/19.0)	91,4%
(2)	Mensalidade = Menor que 300 E Bolsa Integral Ampla = De 2 a 5 bolsas E Bolsa Integral Cotas = De 2 a 5 bolsas	Menor que 500 pontos (100.0/19.0)	81%
	Mensalidade = De 300 a 699 E Bolsa Integral Ampla = 1 bolsa E Bolsa Integral Cotas = De 2 a 5 bolsas E Região = Norte	De 500 a 599 pontos (221.0/28.0)	84,6%
	Mensalidade = Menor que 300 E Bolsa Parcial Cotas = Nenhuma bolsa E Região = Nordeste E Turno = Curso a Distância	De 500 a 599 pontos (519.0/47.0)	90,9%
	Mensalidade = De 300 a 699 E Bolsa Integral Ampla = 1 bolsa E Região = Nordeste E Bolsa Integral Cotas = De 2 a 5 bolsas E Grau = Tecnológico	De 500 a 599 pontos (177.0/23.0)	87,0%
	Mensalidade = De 300 a 699 E Bolsa Integral Ampla = De 2 a 5 bolsas E Turno = Matutino E Grau = Bacharelado	De 500 a 599 pontos (266.0/63.0)	76,3%
	Mensalidade = Maior que 1000 E Bolsa Integral Cotas = De 2 a 5 bolsas E Turno = Noturno E Região = Nordeste	De 600 a 699 pontos (186.0/25.0)	86,6%
	Bolsa Integral Ampla = 1 bolsa E Bolsa Integral Cotas = 1 bolsa E Turno = Noturno E Mensalidade = Maior que 1000 E Região = Sudeste E Grau = Bacharelado	De 600 a 699 pontos (188.0/41.0)	78,2%
(3)	Turno = Curso a Distância E Região = Nordeste E Bolsa Parcial Cotas = 1 bolsa	De 500 a 599 pontos (690.0/178.0)	74,2%
(4)	Bolsa Parcial Ampla = 1 bolsa E Bolsa Parcial Cotas = 1 bolsa E Turno = Curso a Distância E Região = Sudeste E Bolsa Integral Cotas = Nenhuma bolsa	Menor que 500 pontos (247.0/73.0)	70,4%

Desfecho	Regra	Resultado	Precisão
	Turno = Curso a Distância E Bolsa Parcial Ampla = 1 bolsa E Região = Sudeste E Bolsa Integral Cotas = Nenhuma bolsa E Grau = Tecnológico E Classificação = Negócios, administração e direito	Menor que 500 pontos (215.0/31.0)	85,6%

Fonte: Elaboração Própria.

Foram selecionadas 16 regras: 05 para o Desfecho (1); 07 para o Desfecho (2); 01 para o Desfecho (3); e 03 para o Desfecho (4). O Quadro 3 apresenta as características mais frequentes, em cada intervalo de nota de corte, considerando os resultados de todos os desfechos no processo de MD.

Os tópicos descritos na coluna “Tendência” permitem a inferência de tendências e possíveis padrões presentes naquele intervalo, devido à frequência de aparição das características nas regras selecionadas do respectivo intervalo nos desfechos. O intervalo “A partir de 700 pontos” não apresentou o mínimo de ocorrências delimitado para a seleção de regras.

Quadro 3. Características mais frequentes por intervalo de nota de corte.

Intervalo	Tendência
Menor que 500 pontos	<ul style="list-style-type: none"> • Mensalidade menor que R\$ 300,00 • Curso de modalidade a distância • De 2 a 5 bolsas integrais para cotas
De 500 a 599 pontos	<ul style="list-style-type: none"> • Mensalidade de R\$ 300,00 a R\$ 699,00 • Curso de modalidade a distância • Grau Tecnológico • Região Nordeste • De 2 a 5 bolsas integrais na ampla
De 600 a 699 pontos	<ul style="list-style-type: none"> • Mensalidade maior que R\$ 1.000,00 • Turno Noturno • Grau Bacharelado • Região Sudeste • A partir de 2 bolsas integrais na ampla ou cotas
A partir de 700 pontos	*As regras geradas não apresentaram o mínimo de ocorrências necessárias para seleção.

Fonte: Elaboração Própria.

Em um panorama geral, é possível perceber que, em todos os desfechos, os cursos de modalidade a distância são os mais presentes entre os intervalos das menores notas de corte, assim como os cursos de grau tecnológico e os com mensalidade menor que R\$ 300,00. Em contrapartida, os cursos de grau bacharelado e com mensalidade superior a R\$ 1.000,00 aparecem mais nos intervalos das maiores notas de corte.

É possível destacar também que os cursos de graus bacharelado e tecnológico estão estatisticamente mais presentes na base de dados, o que contribui para a presença desses dados em grande parte das regras geradas. Apesar de a maioria dos cursos constar nas regiões sul e sudeste estatisticamente, muitas regras classificaram os cursos da região nordeste, o que implica na existência de um padrão nos aspectos dos cursos ofertados nessa região.

No contexto técnico, este trabalho realiza a indução de regras de classificação por meio do algoritmo PART, conforme apresentado por Frank e Witten (1998). Segundo Maschio et al. (2018), a técnica de indução de regras ainda não é amplamente aplicada na área de MDE, em que é preferencialmente utilizado o aprendizado de máquina. Neste trabalho, a técnica de indução alcançou os resultados mais satisfatórios para o alcance do objetivo durante as fases de testes, o que evidencia a boa aplicabilidade para estudos nesse campo.

Os dados analisados referentes ao Prouni, obtidos na plataforma Brasil.io, comprovam que o programa está presente em todas as regiões brasileiras, conforme a Figura 3. Isso atesta um dos seus principais objetivos, ou seja, a atuação como política de acesso ao ensino superior para jovens de família de baixa renda, até um salário e meio *per capita* (MORAES; TELLES, 2018). No entanto, ainda é possível perceber a concentração de mais de 50% das ofertas de bolsas para as regiões sul e sudeste. Nesse sentido, a chance de escolha ainda não é acessível a todos, pois muitos estudantes têm que considerar suas limitações socioeconômicas e regionais, que podem ser impeditivas para se candidatar a uma bolsa.

Esse aspecto corrobora os tópicos apresentados por Casali e Mattos (2015), sobre os impactos do programa nas práticas de inclusão social e meritocracia. Isso provoca a reflexão sobre até que ponto o Prouni transcende ser objeto de ajuda de custo financeiro para uma fonte de oportunidades de inserção no ensino superior, de forma completa para todos os candidatos em todas as suas dimensões político-sociais.

O fato de a frequência de cursos de modalidade a distância, de grau tecnológico e mensalidade até R\$ 699,00 estar em intervalos de nota de corte menores que 600 pontos, em contrapartida aos cursos presenciais, de grau bacharelado e com mensalidades mais caras estarem em intervalos de notas de

corde superiores, está em consonância com o estudo sobre a massificação de cursos de graduação apresentado por Moreira (2017).

Isso indica que é maior a disputa pelos cursos mais caros, muitas vezes, provenientes de áreas tradicionais como engenharias, medicina e direito. Há promoção da pluralidade de perfis socioeconômicos de estudantes nesses cursos, prevalecendo a democracia pelo mérito, como um aspecto positivo, uma vez que alguns não teriam condições financeiras para arcar com as despesas de tais cursos. Em outras palavras, existe ampliação do acesso de estudantes oriundos de famílias de baixa renda aos cursos mais caros e mais prestigiados e pleiteados. No entanto, o mesmo efeito positivo não acontece nas demais áreas como licenciaturas e/ou pedagogia, que passam a apresentar baixa procura, ociosidade e um perfil mais padronizado de estudantes.

A partir disso, o percentual de vagas ociosas levanta questionamentos em relação à eficiência e ao impacto econômico do Prouni no orçamento federal. Costa e Ferreira (2017) salientam as isenções concedidas para as instituições de ensino participantes do programa, pois, mesmo com as vagas ociosas, as isenções continuam garantidas. Isso faz com que valor médio investido de R\$ 281,78/mês por bolsista deixe de ser realocado para outros cursos, áreas ou programas educacionais. Além disso, a diferença na distribuição dos bens educacionais entre alunos contemplados e pagantes induz os candidatos de famílias de baixa renda para os cursos menos procurados e/ou com baixa qualidade.

O Quadro 3 indica que cursos de grau tecnológico e de modalidade a distância com bolsas integrais, ampla concorrência ou cotista, se encontram em menores intervalos que 600 pontos, o que pode caracterizar a baixa procura pelas vagas ofertadas. Corroborando o estudo de Gualhano et al. (2018), esses dados podem auxiliar o planejamento estratégico das instituições de ensino superior e o aumento sua vantagem competitiva. Nesse caso, um olhar crítico sobre as informações levantadas sobre o Prouni pode ajudar essas instituições a alocar ofertas de bolsas no programa, bem como a identificação do perfil dos cursos mais atrativos para suas clientelas, o que impactaria o direcionamento de seus investimentos. No caso do Governo Federal, as informações poderiam ser utilizadas como referência para a disponibilização de recursos para determinadas regiões do país para o fomento do referido programa.

5. Considerações finais

Esta pesquisa se propôs a analisar o comportamento das notas de corte do Prouni, a partir de variáveis contidas em sua base de dados e sua relação com a política de acesso ao Ensino Superior. A aplicação do algoritmo PART para a geração de regras de classificação, por meio da indução de conhecimento, retornou resultados satisfatórios, com porcentagens de instâncias classificadas corretamente acima dos 70%, o que expressa a confiabilidade das informações geradas.

Os resultados apresentados permitiram a inferência de tendências e possíveis padrões nos quatro desfechos propostos para o processo de mineração de dados. Nos intervalos “Menor que 500 pontos” e “De 500 a 599”, percebe-se grande presença de cursos de modalidade a distância, de grau tecnológico, bem como os que apresentam as menores mensalidades, demonstrando uma menor concorrência para as bolsas ofertadas para eles.

O intervalo “De 600 a 699 pontos” apresenta os cursos presenciais, de grau bacharelado e com mensalidade a partir de R\$ 1.000,00 em sua maioria. As regras do intervalo “A partir de 700 pontos” não se adequaram aos requisitos mínimos de seleção. Dessa forma, obteve-se um panorama do comportamento das notas de corte, que permitiu ampliar a discussão sobre aspectos de política de acesso ao ensino superior, com destaque para os pontos de políticas públicas e de inclusão social, isenção de impostos e investimentos de recursos públicos.

Tendo em vista a carência de estudos específicos envolvendo técnicas de mineração e bases de dados do Prouni, este trabalho contribui, apresentando a viabilidade de extração de informações estratégicas, a partir da própria base dos programas governamentais de fomento educacional. Para trabalhos futuros, sugere-se um estudo aprofundado das informações geradas, com objetivo de traçar um planejamento estratégico para a alocação eficiente dos recursos destinados ao programa.

Referências bibliográficas

- AHLEMEYER-STUBBE, A.; COLEMAN, S. *A practical guide to Data Mining for business and industry*. Chichester, UK: John Wiley & Sons, Ltd, 2014.
- BAEZA-YATES, R.; RIBEIRO-NETO, B. *Recuperação de informação: conceitos e tecnologia das máquinas de busca*. 2ªed. Porto Alegre: Bookman, 2013.
- BAKER, R.; ISOTANI, S.; CARVALHO, A. Mineração de Dados Educacionais: oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, v.19, n.02, p. 03, 31 ago 2011.
- BRASIL. *Lei nº 12.527, de 18 de novembro de 2011*. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>. Acesso em: 14 dez. 2018.
- BRASIL. *Cursos e notas de corte do Prouni 2018 - Datasets - Brasil.IO*. Disponível em: <<https://brasil.io/dataset/cursos-prouni/cursos>>. Acesso em: 14 dez. 2018.
- BRASIL, Ministério da Educação. *Prouni - O Programa*. Disponível em: <<http://prouniportal.mec.gov.br/o-programa>>. Acesso em: 4 abr. 2019.
- CASALI, A. M. D.; MATTOS, M. J. V. M. DE. Análise de estudos e pesquisas sobre o sentido social do programa Universidade para Todos (Prouni). *Ensaio: Avaliação e Políticas Públicas em Educação*, v.23, n.88, p.681–716, set 2015.
- COSTA, D. D. DA; FERREIRA, N. DE B. O Prouni na educação superior brasileira: indicadores de acesso e permanência. *Avaliação: Revista da Avaliação da Educação Superior*, v.22, n.1, p.141–163, abr 2017.
- COSTA, J. D. J.; BERNARDINI, F. C.; VITERBO FILHO, J. A mineração de dados e a qualidade de conhecimentos extraídos dos boletins de ocorrência das rodovias federais brasileiras. *AtoZ: novas práticas em informação e conhecimento*, v.3, n.2, p.139, 31 dez 2014.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, v.17, n.3, p.37–54, mar 1996.
- FRANK, E.; WITTEN, I. H. *Generating accurate rule sets without global optimization*. Hamilton, New Zeland: University of Waikato, 1998.
- GIL, A. C. *Como elaborar projetos de pesquisa*. 6ªed. São Paulo: Atlas, 2017.
- GUALHANO, M. A.; SALLES, S. A. F.; DA HORA, H. R. M. Mineração de dados das fichas da avaliação quadrienal da Capes dos programas da área interdisciplinar: engenharia, tecnologia e gestão. *Revista Meta: Avaliação*, v.10, n. 29, p.417, 15 ago 2018.
- HORA, G. S. et al. Avaliação de ferramentas de mineração de dados: uma abordagem com o modelo TAM. *Interfaces Científicas - Exatas e Tecnológicas*, v.2, n.3, p.109, 13 mar 2018.
- INEP. *Manual para classificação dos cursos de graduação e sequenciais - Cine Brasil 2018*. 2018. Disponível em: <http://download.inep.gov.br/educacao_superior/censo_superior/apresentacao/2018/Manual_Preliminar_para_a_Classificacao_dos_Cursos_Cine_Brasil_2018.pdf>. Acesso em: 2 dez. 2018

MASCHIO, P. et al. Um panorama acerca da mineração de dados educacionais no Brasil. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (BRAZILIAN SYMPOSIUM ON COMPUTERS IN EDUCATION), 29., Fortaleza, 2018. Disponível em: <<http://br-ie.org/pub/index.php/sbie/article/view/8194>>. Acesso em: 28 nov. 2018

MORAES, M. M. DE; TELLES, J. C. DA S. Prouni: uma política pública sob a ótica da educação popular. *Revista Científica - Faculdade de Balsas*, v.9, n.1, p.59-66, 2018.

MOREIRA, C. R. B. S. *Um olhar sobre o muro: avaliação do programa universidade para todos (Prouni)*. 2017. Tese (Doutorado em Educação) - Universidade Federal do Paraná, Curitiba, 2017.

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