



Análise da mudança de processo seletivo para ingresso no ensino superior usando ferramentas de mineração de dados

Analysing the changes in the selection process for admission to higher education using data mining tools

J. F. O. Gaya*; S. F. Pereira Jr; P. F. Schiavo; E. N. Borges; S. S. C. Botelho

Centro de Ciências Computacionais, Universidade Federal do Rio Grande, 96203-900, Rio Grande-RS, Brasil

**joelfelipe94@gmail.com*

(Recebido em 12 de maio de 2017; aceito em 22 de maio de 2017)

Mineração de Dados é um processo de extração de informação implícita, previamente desconhecida e potencialmente útil de bases de dados. O resultado da extração é conhecimento que pode ser analisado para planejamentos futuros e para maior entendimento de um processo. Neste artigo, foi analisada uma base de dados de estudantes do curso de Engenharia de Computação da Universidade Federal do Rio Grande (FURG), onde houve uma mudança na forma de avaliação para o ingresso dos alunos no ensino superior. Com isso, observou-se que o impacto desta mudança foi negativo no desempenho acadêmico dos estudantes. Também durante a pesquisa buscamos identificar quais modelos de dados demonstram se o aluno conclui ou não o curso. Pôde-se observar que a idade, as notas de ingresso e o número de repetições nas disciplinas são fatores preponderantes, para que aluno obtenha a graduação.

Palavras-chave: mineração de dados, classificação, regressão

Data mining is a process of extracting implicit, previously unknown and potentially useful information from databases. The result of the extraction is knowledge that can be analyzed for future planning and for a better understanding of a process. In this article, a database of students of the Computer Engineering course of the Universidade Federal do Rio Grande (FURG) was analyzed, where there was a change in the admission exam to college. With this, it was observed that the impact of this change was negative in the academic performance of the students. Also during the research we try to identify which data models demonstrate whether the student completes the course or not. It could be observed that the age, the entrance grades and the number of repetitions in the disciplines are preponderant factors, so that the student obtains the graduation.

Keywords: data mining, classification, regression

1. INTRODUÇÃO

Por muitos anos, cada uma das universidades públicas brasileiras possuía seu próprio método de processo seletivo. Esta realidade mudou recentemente com a adesão das universidades ao Exame Nacional do Ensino Médio (ENEM). O ENEM é uma avaliação realizada pelo Ministério da Educação (MEC) desde o ano de 1998. E tem como objetivo avaliar o conhecimento dos estudantes do ensino médio.

A partir de 2009, o ENEM passou a ser utilizado nos processos seletivos da grande maioria das instituições públicas do país. Na primeira edição desse modelo de seleção, o ENEM registrou cerca de 4,5 milhões de inscritos. Já em 2013, foram registradas mais de 7 milhões de inscrições. Assim como o número de inscrições, o número de vagas ofertadas em instituições públicas de ensino superior também cresceu de maneira significativa no país durante esse período.

Na Universidade Federal do Rio Grande (FURG), onde o estudo relatado neste artigo foi conduzido, a adesão ao exame foi feita no ano de 2011. Porém, como mencionado anteriormente esta não foi a única mudança que ocorreu na época, ela veio associada a uma série de mudanças como o aumento da oferta de vagas disponíveis na região, assim como no resto do país. Permitindo que estudantes de qualquer parte do país concorressem a uma vaga nas

universidades, desde que ela aderisse ao SISU, sem a necessidade do estudante se deslocar até a universidade para prestar vestibular.

Apesar de esta mudança ter afetado diversas universidades em todo o país, ainda não há muito estudo sobre o impacto no desempenho dos acadêmicos. Portanto, o presente trabalho busca verificar o impacto desta mudança em alguns aspectos do desempenho acadêmico.

Diversos estudos em aplicações e métodos de mineração de dados foram realizados em diversas áreas, tendo como objetivo extrair conhecimento de bases de dados. No entanto, mesmo após anos de pesquisa na área, ainda não há um sistema de mineração genérico capaz de extrair conhecimento de uma base qualquer. Então, é esperado que a tarefa de projetar um sistema de mineração que realize tal tarefa usando métodos existentes de mineração, para uma base ainda não explorada seja um desafio científico [1].

Anteriormente, foram propostas algumas abordagens para previsão de desempenho acadêmico [2, 3], que trazem ideias bastante úteis para o presente trabalho.

Kabakchieva (2013) [4] realizou um estudo em bases de dados de estudantes aplicando classificação Bayesiana para prever o desempenho de um indivíduo, baseado nos dados do ano anterior. Eles evidenciam, neste estudo, que não apenas o esforço do estudante, mas também influência familiar, hábitos do estudante e escolaridade dos pais influenciam dramaticamente na performance do mesmo. Vale ressaltar que este trabalho relata a importância dos hábitos do estudante em seu desempenho, mas esta informação ainda não está presente na base da universidade em questão. Porém, se de fato essa informação é relevante para o desempenho do acadêmico seria interessante que a instituição passasse então a armazená-la.

Outro estudo abordou a previsão do desempenho de estudantes de engenharia de universidades indianas, a partir de suas notas do ano anterior ou exame do primeiro ano [5]. Eles testaram três algoritmos de classificação diferentes: ID3 [6], C4.5 [7] e CART [8]. Dentre estes algoritmos, selecionaram o C4.5 como o melhor preditor. Os pesquisadores implementaram o resultado do estudo de previsão alertando os alunos que tinham chances de serem reprovados e obtiveram uma melhora no desempenho dos mesmos.

Utilizando mineração de dados, pesquisadores conseguiram prever em uma base de estudantes indianos o desempenho dos alunos com base em suas notas anteriores [9]. Foi usado o algoritmo de classificação C4.5. Os autores concluíram que estudantes e professores podem melhorar seu desempenho com base em sua previsão.

Este trabalho tem como objetivo principal avaliar o impacto da mudança de método de ingresso na aprovação e nas notas finais dos alunos. O estudo foi estendido para avaliar se os auxílios solicitados pelos estudantes e atendidos pela Universidade Federal do Rio Grande contribuem de fato no desempenho acadêmico do estudante. Além disso, buscamos identificar modelos de dados para prever se o aluno conclui ou não o curso.

O restante deste artigo é organizado da seguinte forma. A Seção 2 é feita uma descrição breve da base de dados e de como os dados foram pré-processados. Também é apresentada a metodologia adotada. A Seção 3 aborda os resultados obtidos a partir dos algoritmos empregados para cada tarefa de mineração realizada. E finalmente, a Seção 4 resume as contribuições deste artigo e aponta a direção que será tomada nos trabalhos futuros.

2. MATERIAL E MÉTODOS

A base de dados fornecida pela FURG compreende alunos apenas do curso de Engenharia de Computação, de 1994, ano em que o curso teve início, até o final do ano de 2015. A Figura 1 ilustra o modelo da base de dados utilizada neste trabalho. Nelas, as informações que seriam interessantes minerar se encontram dispersas em diversas tabelas e, além disso, existe uma série de dados faltantes e mal cadastrados. As consultas então tiveram que agrupar os dados, eliminar registros ruidosos e transformar a informação, para que essa pudesse ser minerada.

A partir da base fornecida, foram realizadas consultas SQL (*Structured Query Language*), que geraram as relações passíveis de serem mineradas. Ao fim desta transformação, identificamos dois casos que podemos trabalhar neste artigo.

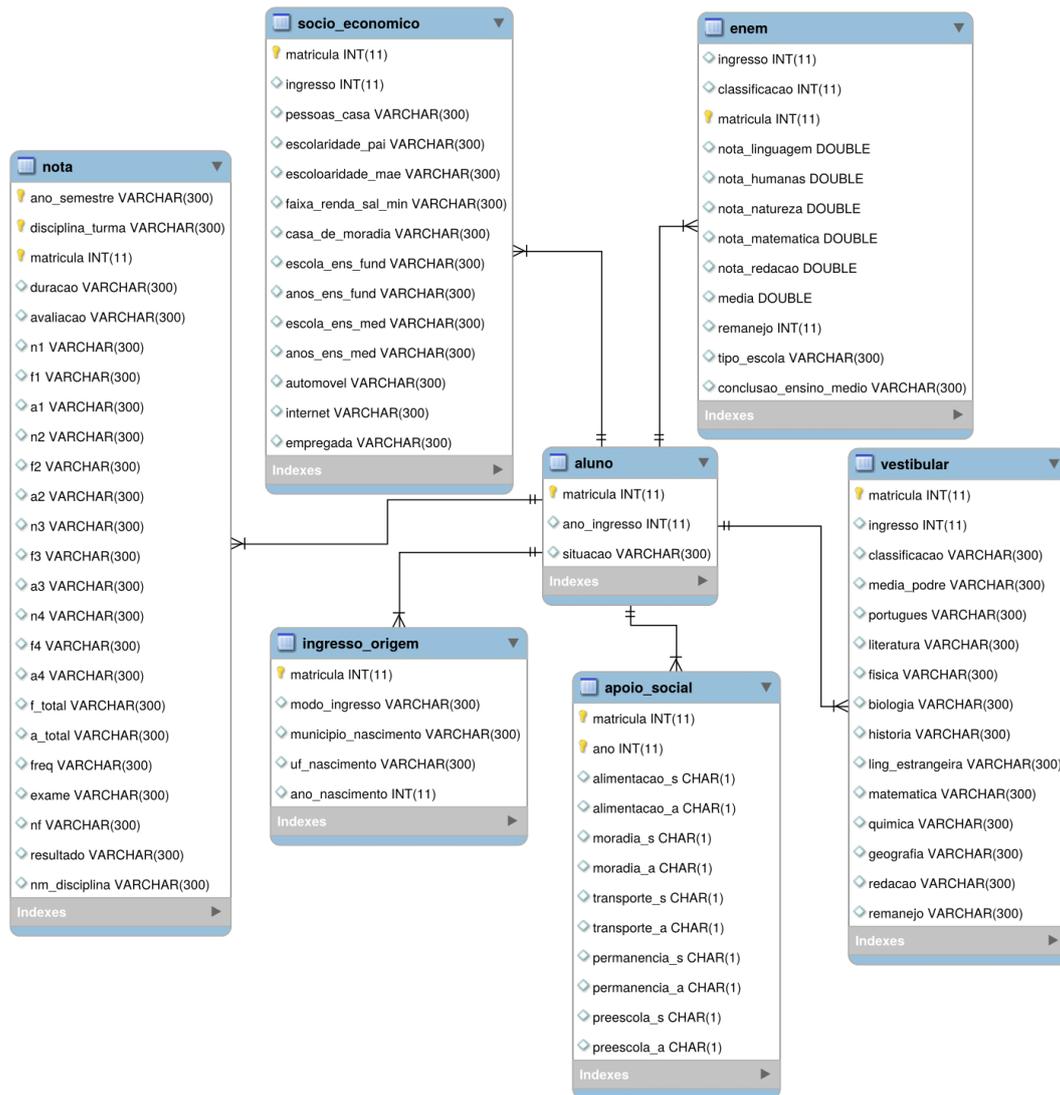


Figura 1: Esquema lógico das tabelas usadas no presente trabalho.

O primeiro caso diz respeito ao impacto dos atributos no desempenho dos alunos. Os dados necessários para modelar este impacto são obtidos a partir de duas tabelas. Uma delas, utilizada para as técnicas de classificação e regressão, a fim de verificar o impacto direto de cada atributo no desempenho. A outra tabela foi utilizada para a associação, onde se procura alguma inconsistência que possa vir a prejudicar os alunos de baixa renda e ou oriundos de outras regiões.

A tabela usada na classificação e na regressão, após as etapas de transformação e pré-processamento, contém os atributos idade de ingresso, nota obtida no exame admissional, nome da disciplina cursada, modo de ingresso, resultado e nota final. Onde, o modo de ingresso pode ser Enem, vestibular ou outros e resultado é o atributo que contém a informação se o aluno aprovou ou reprovou na disciplina. Cada instância representa o desempenho de um determinado aluno em uma disciplina.

Para a tarefa de regressão não é interessante que a tabela contenha o atributo resultado, já que o resultado se dá em função da nota final. Então este atributo foi retirado para a execução desta tarefa. Da mesma forma para o atributo nota final na tarefa de classificação, e por isto foi retirado da tabela.

A medida de desempenho dos alunos foi feita por disciplina, devido ao fato da base em questão não conter turmas formadas que ingressaram através do novo exame (SISU - ENEM) e de mudanças no currículo que mudaram o período do curso no qual as disciplinas eram

ministradas. Portanto, essa base não permite uma análise que envolva o número de formados ou desistências.

A tabela usada para associação reúne informações provenientes de tabelas que continham dados dos benefícios concedidos aos alunos e com questionários socioeconômicos respondidos no momento do ingresso. O pré-processamento dos atributos dos benefícios para associação foi feito de uma forma bastante peculiar: se o benefício foi pedido e foi concedido o atributo recebe o valor *sim*, se o benefício foi pedido e não obtido, este recebe *não*, e caso ele não tenha sido pedido o atributo não possui valor. Esta abordagem foi adotada para que benefícios que não fossem solicitados não formassem regras.

O segundo caso de estudo diz respeito a modelagem para predição do impacto na conclusão do curso. No pré-processamento deste caso foram filtradas as tabelas de modo a obter os atributos: número de matrícula, idade de ingresso, notas do ingresso (valores obtidos nas provas do vestibular correlacionados com as disciplinas utilizadas no ENEM, conforme Tabela 1), quantidade de vezes que o aluno cursou as disciplinas da primeira série Cálculo 1, Física 1, Química e Computação, quantidade de vezes que cursou as disciplinas da segunda série Cálculo 2 e Física 2, e da terceira série Método Numérico e Teoria da Computação. Ainda foram adicionadas as informações se o aluno estava *formado*, *cursando* ou *não formado*, este último inclui todos os alunos que não concluíram o curso por qualquer motivo.

Tabela 1: Representação das notas de ingresso.

Disciplinas	Processo Seletivo	
	Enem	Vestibular
Linguagem	Nota Linguagem	(Português + Literatura + Língua Estrangeira) / 3
Humanas	Nota Humanas	(História + Geografia) / 2
Natureza	Nota Natureza	(Física + Biologia + Química) / 3
Matemática	Nota Matemática	Matemática

Após a consulta finalizada, foram identificadas duas situações que precisaram de ajustes manuais. A primeira ocorreu quando se percebeu que alguns alunos formados não possuíam a informação de quantidade de vezes que cursaram algumas matérias. Foi considerado que se a informação de *formado* estiver correta, os discentes poderiam ter feito equivalência das disciplinas em outro curso ou instituição, por exemplo. Dessa forma, considera-se que todos os alunos que concluíram o curso fizeram pelo menos uma vez a cada disciplina consultada.

A segunda situação foi que, alguns alunos não possuíam a nota de ingresso. Esses cadastros foram excluídos da seleção, por se tratarem de poucos registros.

Por fim, restaram 820 instâncias com o perfil homogêneo, onde 305 registros representavam estudantes já formados, 322 não formados e 193 cursando.

A fim de realizar os experimentos, foi usado como ferramenta de mineração de dados o Waikato Environment for Knowledge Analysis (WEKA) [10]. WEKA é uma ferramenta de mineração de dados escrita em Java pela universidade de Waikato e é excelente para tarefas como classificação, regressão, associação e agrupamento, tarefas essas bastante úteis para a condução do presente trabalho.

Foram utilizadas técnica de classificação, regressão e associação. Todas essas dispõem de diversos algoritmos no WEKA. Para cada uma destas técnicas será apresentado um algoritmo com duas parametrizações distintas. Os algoritmos usados foram escolhidos com base nos trabalhos relacionados e também com base em testes preliminares.

Vale ressaltar, entretanto, que a pesquisa apresentada neste artigo não modela o desempenho de um acadêmico, mas procura entender quais são as variáveis que levam um aluno a obter um determinado desempenho. Mais especificamente, procuramos saber se a mudança na forma de ingresso tem afetado significativamente o desempenho dos acadêmicos. Porém, é de fundamental importância obter modelos dos dados que tenham boa acurácia, pois isso indica que o modelo representa de forma fiel os dados que compõem o seu treinamento. Por essas razões, este trabalho dá preferência a modelos de fácil interpretação.

3. RESULTADOS E DISCUSSÃO

Analisando se houve impacto na aprovação dos alunos após a alteração do método do processo seletivo foi utilizada a tarefa de classificação para a predição do atributo alvo resultado (aprovação ou reprovação). Foi utilizado o classificador J48 [11].

O experimento foi conduzido usando validação cruzada com 10 partições. Os resultados obtidos através desta análise podem ser observados na Tabela 2 e na Tabela 3. Ambos os resultados parecem representar bem o conjunto de dados usado no treinamento e com isso pode-se passar à análise do modelo gerado e a partir dela tirar conclusões sobre os dados. Na parametrização do teste foi utilizada validação cruzada com 10 partições.

Tabela 2: Medidas de avaliação obtidas do resultado do classificador J48, através da validação cruzada, para divisões não necessariamente binárias.

Medida de avaliação	Valor
Instâncias corretamente classificadas	80.17%
Instâncias incorretamente classificadas	19.82 %
Kappa	0.244
Erro absoluto médio	0.2647
Raiz do erro quadrático médio	0.3723
Erro relativo absoluto	80.98%
Raiz do erro quadrático relativo	92.08%

Tabela 3: Medidas de avaliação obtidas do resultado do classificador J48, através da validação cruzada, para divisões binárias.

Medida de avaliação	Valor
Instâncias corretamente classificadas	80.08%
Instâncias incorretamente classificadas	17.92%
Kappa	0.3491
Erro absoluto médio	0.252
Raiz do erro quadrático médio	0.3636
Erro relativo absoluto	77.08%
Raiz do erro quadrático relativo	89.92%

A Figura 2 mostra o modelo para a disciplina Fundamentos de Química, porém o resto do modelo se comporta de forma análoga. Levando em consideração o modelo gerado, pode-se observar de maneira clara três fenômenos. Primeiro, o impacto negativo da idade de ingresso nas notas, o que pode ser explicada pelo fato de alunos mais novos terem mais disponibilidade de tempo para as tarefas do curso. Em segundo lugar, o impacto positivo da nota obtida no ingresso, seja ele por vestibular ou SISU, fenômeno esse não estranho visto que alunos com bom desempenho em avaliações mantenham o bom desempenho. E finalmente, o impacto negativo da mudança de avaliação na aprovação dos alunos.

Feita a análise de impacto na quantidade de aprovações, passamos ao impacto nas notas finais obtidas pelos alunos em cada disciplina. Para isso iremos utilizar o método de regressão linear para os mesmos atributos utilizados na análise anterior, conforme descrito na Seção 2. A Equação 1 apresenta o modelo aprendido pelo algoritmo, onde Ganho é um fator aditivo definido de acordo com base na disciplina. Como esperado, o modelo gerado por esse método confirma os três fenômenos mencionados anteriormente, mas não adiciona novas conclusões.

$$\text{NotaFinal} = 0.0009 * \text{MediaIngresso} - 0.196 * \text{IdadeIngresso} + 0.5511 * [\text{ModoIngresso} = \text{Vestibular}] + 4.0567 + \text{Ganho} \quad (1)$$

De forma semelhante à tarefa anterior, analisamos os resultados obtidos usando validação cruzada com 10 partições. A análise dos resultados obtidos através deste algoritmo pode ser observada na Tabela 4 e na Tabela 5. Podemos observar que o uso ou não de um método de seleção de atributos representa, neste caso, um ganho marginal.

Tabela 4: Medidas de avaliação obtidas do resultado da regressão linear, através da validação cruzada, usando o método M5 de seleção de atributos.

Medida de Avaliação	Valor
Coefficiente de correlação	0.512
Erro absoluto médio	2.0188
Raiz do erro quadrático médio	2.67
Erro relativo absoluto	81.53%
Raiz do erro quadrático relativo	85.90%

Tabela 5: Medidas de avaliação obtidas do resultado da regressão linear, através da validação cruzada, sem o uso de métodos de seleção de atributos.

Medida de Avaliação	Valor
Coefficiente de correlação	0.5119
Erro absoluto médio	2.018
Raiz do erro quadrático médio	2.6702
Erro relativo absoluto	81.50%
Raiz do erro quadrático relativo	85.91%

O impacto negativo nas notas e no número de aprovações faz pensar sobre o que leva estes alunos a terem um desempenho inferior ao dos alunos que ingressaram pelo antigo vestibular. Uma das possibilidades é que com o aumento dos alunos oriundos de outras regiões existe uma demanda maior por auxílio que pode não estar sendo atendida.

Para verificar a hipótese levantada, foi usada uma tabela que possui informações socioeconômicas do aluno e dos pedidos de auxílio. Através dessas informações, procura-se identificar padrões que levem a universidade a acolher de forma mais eficaz os alunos. Neste trabalho optou-se por usar associação para encontrar tal padrão.

O algoritmo escolhido para a tarefa de associação foi o Apriori [12]. As regras obtidas e sua confiança são mostradas nas Tabelas 6 e 7. Nelas não observamos a existência de nenhuma anomalia aparente nos dados o que não é um resultado conclusivo. Porém, apesar de não conclusivo é um indicativo de que a causa da queda de desempenho dos alunos que ingressaram pelo vestibular em relação aos que entraram pelo SISU-ENEM não foi por falta de auxílio da universidade.

No segundo estudo de caso, para encontrar um modelo que auxilie a prever se o aluno conclui ou não o curso, retorna-se à tarefa de classificação, utilizando o algoritmo J48, com a validação cruzada com 10 partições. Os resultados obtidos através desta análise podem ser observados na Tabela 8 e na Tabela 9.

Tabela 6: Regras obtidas pelo algoritmo Apriori, para suporte mínimo 0.05 e confiança 0,8.

Regra	Predicado	Confiança
1	Alimentação=Sim, Moradia=Sim, Esc. Ens. Med.=Pública → Resultado=Aprov	0.83
2	Alimentação=Sim, Moradia=Sim, Esc. Ens. Med.=Pública Internet=Sim → Resultado=Aprov	0.83
3	Alimentação=Sim, Moradia=Sim, Esc. Ens. Fund.=Pública Esc. Ens. Med.=Pública → Resultado=Aprov	0.82
4	Alimentação=Sim, Moradia=Sim, → Resultado=Aprov	0.81
5	Alimentação=Sim, Moradia=Sim, Internet=Sim → Resultado=Aprov	0.8

Tabela 7: Regras obtidas pelo algoritmo Apriori, para suporte mínimo 0,013 e confiança 0,9.

Regra	Predicado	Confiança
1	Transporte=Sim, Casa de Moradia=Própria, Esc. Ens. Fund.=Pública, Internet=Não → Resultado=Rep	1
2	Permanência=Não, Casa de Moradia=Própria, Esc. Ens. Fund.=Pública, Internet=Não → Resultado=Rep	1
3	Transporte=Sim, Permanência=Não, Casa de Moradia=Própria, Esc. Ens. Fund.=Pública, Internet=Não → Resultado=Rep	1
4	Alimentacao=Sim, Transporte=Sim, Casa de Moradia=Própria, Esc. Ens. Fund.=Pública, Internet=Não → Resultado=Rep	1
5	Transporte=Sim, Casa de Moradia=Alugada, Esc. Ens. Fund.=Pública, Internet=Sim → Resultado=Rep	0.96
6	Moradia=Sim, Casa de Moradia=Alugada, Internet=Sim → Resultado=Rep	0.93
7	Casa de Moradia=Própria, Esc. Ens. Fund.=Pública, Internet=Não → Resultado=Rep	0.91

Em ambos os resultados, o conjunto de dados usado no treinamento atingiu um alto coeficiente de acertos, mostrando que a idade de ingresso, as médias das provas de ingresso e o número de repetições nas disciplinas possuem grande relevância na predição sobre a conclusão do curso. Pode-se considerar o modelo muito especializado, mas a proposta é encontrar quais os atributos podem revelar a predição.

Tabela 8: Medidas de avaliação obtidas do resultado do classificador J48, através da validação cruzada, para divisões não necessariamente binárias.

Medida de avaliação	Valor
Instâncias corretamente classificadas	95.8533 %
Instâncias incorretamente classificadas	4.1467 %
Kappa	0.9172
Erro absoluto médio	0.0706
Raiz do erro quadrático médio	0.1983
Erro relativo absoluto	14.1254 %
Raiz do erro quadrático relativo	39.6707 %

Tabela 9: Medidas de avaliação obtidas do resultado do classificador J48, através da validação cruzada, para divisões binárias.

Medida de avaliação	Valor
Instâncias corretamente classificadas	96.3317 %
Instâncias incorretamente classificadas	3.6683%
Kappa	0.9267
Erro absoluto médio	0.0661
Raiz do erro quadrático médio	0.1854
Erro relativo absoluto	13.2237 %
Raiz do erro quadrático relativo	37.1002 %

Analisando o modelo percebe-se também que o número de repetições de um aluno na disciplina Cálculo I é semelhante a disciplina de Física I. Ambas as disciplinas são ministradas no primeiro ano do curso. Alguns alunos que obtiveram notas muito altas no ingresso, independente se pelo Vestibular ou ENEM, não concluíram a Engenharia de Computação na FURG. Acreditamos que estas pessoas tenham se deslocado para outras universidades ou cursos. Como já havíamos abordado a situação de origem do aluno é um fator importante para a continuidade e desempenho no curso. Seria necessária uma avaliação específica destes casos.

Em um dos ensaios, onde os dados somente para a disciplina Cálculo I foram usados, o modelo confirma a referência anterior que a idade de ingresso é um fator de grande influência. A Figura 3 apresenta o modelo treinado. Mesmo errando consideravelmente (vide Tabela 10), o modelo é significativo.

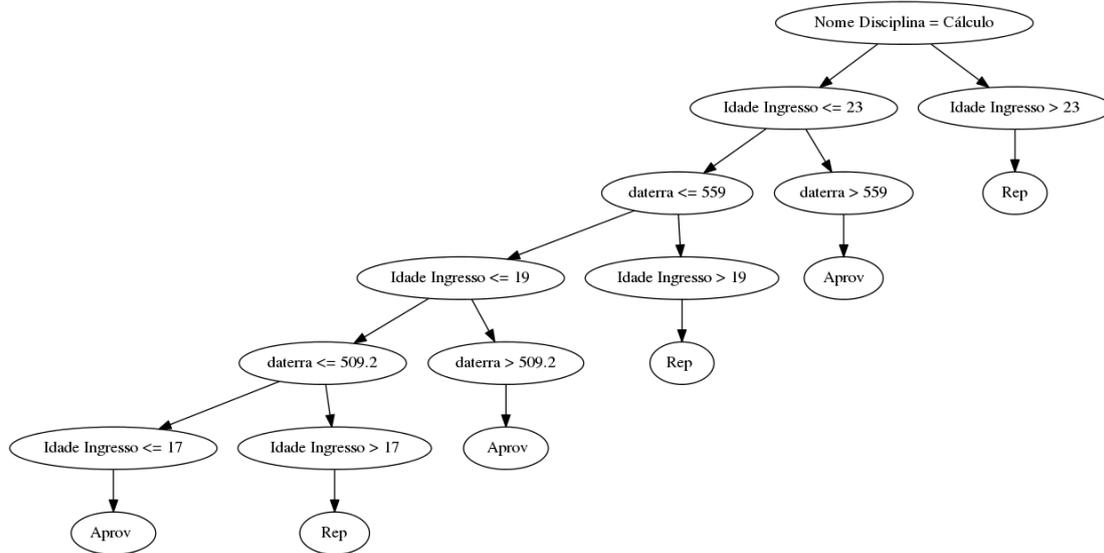


Figura 3: Resultado Árvore Cálculo I.

Tabela 10: Medidas de avaliação obtidas do resultado para ser aprovado em Cálculo I, através da validação cruzada.

Medida de Avaliação	Valor
Instâncias corretamente classificadas	69.2214 %
Instâncias incorretamente classificadas	30.7786 %
Kappa	0.2666
Erro absoluto médio	0.4094
Raiz do erro quadrático médio	0.4626
Erro relativo absoluto	87.4535 %
Raiz do erro quadrático relativo	95.6273 %

4. CONCLUSÃO

Este trabalho mostrou que a mudança no método de ingresso afetou de forma significativa o desempenho dos alunos nesta base de dados. Além disso, o impacto no desempenho dos alunos foi negativo, ou seja, segundo os modelos obtidos alunos que ingressaram pelo SISU-ENEM, tem desempenho pior que os alunos que entraram através do vestibular. Esse é um dado bastante preocupante, visto que essa foi uma mudança ocorreu em escala nacional. Os resultados aqui obtidos não são conclusivos para todo país, mas indicam a necessidade de um estudo em maior escala. Um estudo dessa magnitude se justifica pelo investimento que foi feito por parte do governo federal tanto no próprio exame quanto no número de vagas.

O impacto negativo da mudança de avaliação para SISU-ENEM no desempenho dos acadêmicos pode ter diversas causas. Uma delas é que muitos estudantes chegam à Universidade oriundos de outras regiões, o que leva a pensar se existe algum aspecto socioeconômico que justifique essa queda. Uma das abordagens para lidar com isto são auxílios fornecidos pela universidade como moradia para alunos com necessidades de habitação.

Além disso, identificamos que o número de repetições em disciplinas chaves tem relevância na formação do aluno no curso. Usando os dados de desempenho no ingresso e o número de repetições dos alunos nas disciplinas anteriores, os professores podem ajudar a concentrar esforço do aluno em áreas com problemas potenciais no referido curso. Os educadores também

podem usar esta informação para orientar a sua implementação e avaliação das mudanças curriculares.

Neste trabalho, foi relatado o uso de técnicas de análise para construir modelos preditivos. Embora muitos dos modelos gerados não possuam poder preditivo suficiente para ser útil, os modelos mais fortes e outras observações a partir da análise fornecem indicações úteis sobre as relações entre as disciplinas e o histórico de ingresso do aluno.

Para avaliar o impacto da situação socioeconômica dos alunos e os auxílios recebidos, foi realizada uma tarefa de associação em dados socioeconômicos, dados de auxílios fornecidos pela universidade e os resultados. O resultado desta tarefa, no entanto, não apontou nenhuma regra que indicasse uma falta de suporte, por parte da universidade, ao aluno impactando em seu resultado. E portanto, apesar de não ser conclusivo, esse resultado leva a pensar que de fato pode existir outra causa para essa queda de desempenho.

Portanto, considerando os resultados obtidos nesse trabalho, não foi possível inferir com confiança suficiente a causa exata para o impacto negativo da mudança de método de ingresso, visto que essa mudança não se deu isoladamente, pois no mesmo período houve um aumento significativo de vagas em toda rede de ensino superior pública federal, e qualquer conclusão que tente apontar causas para este fenômeno deve levar isso em consideração.

Contudo, o resultado obtido com essa base de dados é um indício de que o mesmo fenômeno pode estar acontecendo em muitos cursos da FURG ou em muitas instituições em todo o país. E caso seja verificado que este não é um fenômeno isolado, devem ser tomadas as providências cabíveis.

5. REFERÊNCIAS BIBLIOGRÁFICAS

1. Deshpande SP, Thakare VM. Data mining system and applications: a review. *International Journal of Distributed and Parallel Systems* 2010 Sep; 1(1): 32-44.
2. Khan B, Khiyal MSH, Khattak MD. Final grade prediction of secondary school student using decision tree. *International Journal of Computer Applications* 2015; 115(21): 32-36.
3. Pandey UK, Pal S. Data mining: a prediction of performer or underperformer using classification [Internet]. New York: Computing Research Repository; 2011 [cited 2017 May 11]. Available from: <https://arxiv.org/abs/1104.4163v1>.
4. Kabakchieva D. Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies* 2013; 13(1): 61-72, doi: 10.2478/cait-2013-0006.
5. Yadav SK, Pal S. Data mining: A prediction for performance improvement of engineering students using classification. [Internet]. New York: Computing Research Repository; 2012 [cited 2017 May 11]. Available from: <https://arxiv.org/abs/1203.3832>.
6. Quinlan JR. Induction of decision trees. *Machine learning* 1986; 1(1): 81-106, doi: 10.1023/A:1022643204877
7. Quinlan JR. C4.5: programs for machine learning. Elsevier; 2014. Chapter 3, Constructing decision trees; p. 17-25.
8. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. Belmont: Wadsworth International Group; 1984.
9. Prasad GNR, Babu AV. Mining previous marks data to predict students performance in their final year examinations. *International Journal of Engineering Research and Technology* 2013 Sep; 2(2).
10. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The weka data mining software: an update. *ACM SIGKDD explorations newsletter* 2009; 11(1):10-18, doi: 10.1145/1656274.1656278.
11. Bhargava N, Sharma G, Bhargava R, Mathuria M. Decision tree analysis on J48 algorithm for data mining. *International Journal of Advanced Research in Computer Science and Software Engineering* 2013; 3(6).
12. Agrawal R, Srikant R. Fast algorithms for mining association rules. *Proceedings of the International Conference in Very Large Data Bases*; 1994. p. 487-499.