

Análise dos cursos de licenciatura da Rede Federal utilizando mineração de dados

Mariana Abreu Gualhano¹

ORCID: 0000-0002-8103-7125

Aline Pires Vieira de Vasconcelos²

ORCID: 0000-0002-2683-595X

Resumo

A extração do conhecimento, também conhecida como processo KDD (*Knowledge Discovery in Databases*), é um conjunto de técnicas (Seleção, Pré-processamento, Tratamento, Mineração e Interpretação de Dados) que tem como objetivo analisar e extrair padrões e informações potencialmente úteis de grandes bases de dados. A avaliação da qualidade dos cursos de graduação no Brasil é feita por meio do Conceito Preliminar de Cursos (CPC), que é um indicador de qualidade que avalia esses cursos. Portanto, a presente pesquisa insere-se no discutido contexto, buscando utilizar técnicas de mineração de dados para extrair conhecimento das avaliações do CPC dos anos de 2014 e 2017 e identificar os principais critérios e resultados da avaliação dos cursos de graduação em Licenciatura, fazendo uma análise dos dados da Rede Federal de todo o Brasil. Para isso, utilizou-se a base de dados disponibilizada no portal do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) e foram utilizadas as etapas do processo KDD, com foco na mineração de dados para extração de conhecimento da base. Os resultados permitiram realizar a identificação dos critérios de avaliação com maior impacto para a avaliação do CPC dos cursos de Licenciatura, além de uma comparação entre a avaliação de 2014 e 2017. Espera-se que as informações extraídas a partir do presente trabalho possam ser úteis e subsidiar a gestão educacional, além de serem utilizadas para o aprimoramento dos cursos de graduação.

Palavras-chave

Avaliação – Rede Federal – CPC – Mineração de dados – Licenciatura.

1- Instituto Federal de Educação, Ciência e Tecnologia Fluminense (IFF), Macaé – RJ, Brasil. Contato: mariana_gualhano@hotmail.com.

2- Instituto Federal de Educação, Ciência e Tecnologia Fluminense (IFF), Campos dos Goytacazes – RJ, Brasil. Contato: alineprofcefet@gmail.com



<https://doi.org/10.1590/S1678-4634202046219576>

This content is licensed under a Creative Commons attribution-type BY-NC.

Analysis of teaching certification programs in federal schools using data mining

Abstract

Knowledge extraction, also known as the KDD (Knowledge Discovery in Databases) process, is a set of techniques (Selection, Pre-processing, Treatment, Data Mining and Data Interpretation) that aims to analyze and extract potentially useful patterns and information from large databases. The evaluation of the quality of undergraduate courses in Brazil is done through the Preliminary Concept of Courses (PCC), which is a quality index that evaluates these courses. Therefore, this research is part of the context presented here, seeking to use data mining techniques to extract knowledge from the PCC evaluations of the years 2014 and 2017 and to identify the main criteria and results of the evaluation of Teaching Certification programs, making an analysis of data from Federal Schools all over Brazil. In order to do that, the database available on the website of INEP (National Institute for Educational Studies and Research) and the KDD process steps were used, focusing on Data Mining to extract knowledge from the database. The results enabled the identification of the evaluation criteria with the greatest impact on the evaluation of the PCC of Teaching Certification programs, in addition to a comparison between the evaluation of 2014 and 2017. The information extracted from the present work will hopefully be useful and may subsidize educational management, in addition to being used to improve undergraduate courses.

Keywords

Educational evaluation – Federal Schools – PCC – Data mining – Teaching Certification programs.

Introdução

Segundo a Lei 11.892 de 29 de dezembro de 2008, a Rede Federal de Educação Profissional, Científica e Tecnológica é composta por Institutos Federais, Centros Federais de Educação Tecnológica, Escolas Técnicas Vinculadas às Universidades Federais, Universidade Tecnológica Federal do Paraná e Colégio Pedro II (BRASIL, 2008).

No contexto da Rede Federal destacam-se os Institutos Federais, que estão em constante crescimento e atendendo uma parcela significativa da população brasileira, afirma Moura (2018). São 38 Institutos Federais presentes em todos estados, que oferecem cursos de qualificação, ensino médio integrado, cursos superiores de tecnologia e Licenciaturas.

A Lei de criação da Rede Federal, no seu artigo 8º, reserva um mínimo de 20% de sua oferta para cursos de Licenciatura. A ideia de integrar a formação de professores aos Institutos tem como uma de suas bases a necessidade desses profissionais para atuação no Ensino Médio, decorrente de um estudo realizado por conselheiros da Câmara de Educação

Básica do Conselho Nacional de Educação (CNE/CEB/2007), que apresentou o problema de defasagem do número de professores nessa modalidade de ensino (BRASIL, 2007).

Mororó e Pereira (2017) salientam que a partir do crescimento dos Institutos Federais, também houve a interiorização dos cursos das Licenciaturas, o que é uma particularidade da política dessas instituições de ensino técnico federal e da própria política de formação de professores no país, agregando assim, valor e oportunidades a localidades com diferentes especificidades e que antes não ofertavam esses cursos.

Ademais, a Matriz do Conselho Nacional das Instituições da Rede Federal de Educação Profissional, Científica e Tecnológica (CONIF), que tem como função custear e investir nas Instituições Federais, na sua edição de 2018, consolidou um aumento do peso orçamentário para os alunos de Licenciatura, quando comparados a alunos de outras modalidades de ensino (FORPLAN, 2018).

Nesse contexto, fica evidente a importância dos cursos de Licenciatura para a formação de professores na atuação em diversos níveis de ensino, cursos esses que são ofertados na Rede Federal, em Universidades públicas e privadas e que são avaliados periodicamente por meio de órgãos governamentais.

As avaliações da qualidade de ensino de graduação têm como principal objetivo apresentar dados referentes a critérios pré-estabelecidos pelos órgãos competentes e a partir da avaliação desses critérios, produzir informações relacionadas à eficiência das instituições, por exemplo, do Ensino Superior (IES). Assim, essas avaliações são de fundamental importância no que tange ao controle de qualidade das instituições de ensino.

O Conceito Preliminar de Curso (CPC) é um indicador de qualidade resultante da avaliação dos cursos de graduação de todo o Brasil, contendo quatro pilares fundamentais: desempenho de estudantes, corpo docente, infraestrutura e recursos didático-pedagógicos.

O CPC tem os cursos avaliados segundo as áreas a ele vinculadas, em um ciclo avaliativo com referência nos resultados trienais de desempenho de estudantes: Ano I - Saúde, Ciências Agrárias e áreas afins; Ano II - Ciências Exatas, Licenciaturas e áreas afins; Ano III - Ciências Sociais Aplicadas, Ciências Humanas e áreas afins (BRASIL, 2015). O ano representa o ciclo de avaliação dos cursos que ocorre a cada três anos, iniciando com os cursos do Ano I, em seguida os cursos do Ano II e, por fim, os cursos do Ano III. A base de dados que contém os dados referentes à avaliação do CPC das IES está disponível no *site* do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP)³.

Devido ao crescente avanço tecnológico, grandes volumes de dados são gerados pela constante utilização da Tecnologia da Informação (TI) em todos os âmbitos, inclusive na educação, como Rigo *et al.* (2014) afirmam. No Brasil, têm-se dados disponíveis de avaliações da qualidade de ensino do Ensino Fundamental até a Pós-graduação, perpassando pelo Ensino Médio e Graduação.

No entanto, não há utilidade em grandes bases de dados sem a utilização de ferramentas que auxiliem na análise e interpretação desses, corroboram Machado *et al.* (2015). Dentre as ferramentas de análise de uma grande massa de dados, destaca-se a técnica de mineração de dados, que permite extrair conhecimento de um vasto número de registros (ROIGER, 2017), em especial em dados relacionados à educação, conhecida como

3- Informações disponíveis em: <<http://inep.gov.br/conceito-preliminar-de-curso-cpc>>

Educational data mining (EDM), conceito apresentado nos trabalhos de Baker, Isotani e Carvalho (2011) e Ahuja *et al.* (2019).

A mineração de dados é uma das etapas do KDD, que vem a ser *Knowledge Discovery in Databases* (Descoberta em bases de conhecimento), em que envolve também as seguintes etapas: seleção, pré-processamento, transformação e interpretação dos dados.

A partir disso, Sukhija, Jindal e Aggarwal (2015) afirmam que grande parte dos dados pertinentes à educação estão relacionados, direta e indiretamente, à qualidade de ensino e ao comportamento de aprendizado dos estudantes e educadores. O autor salienta que a EDM tem como objetivo principal desenvolver uma compreensão dos dados presentes nesses repositórios e isso é realizado por meio de uma combinação de diferentes técnicas de mineração em associação com esses dados, visando a otimizar a prática e a gestão educacional. São exemplos de dados que podem ser extraídos para contribuir com o conhecimento na gestão educacional: perfil de aluno, nota dos alunos do Exame Nacional de Desempenho dos Estudantes (Enade), número de mestres e doutores, entre outros.

Em consonância com os aspectos relacionados à importância da avaliação da qualidade dos cursos da IES concomitantemente com a relevância dos cursos de Licenciatura, principalmente na Rede Federal, espera-se que o presente trabalho possa contribuir para os gestores por meio das informações extraídas da base de dados, auxiliando, então, tanto na tomada de decisão quanto nas medidas de qualidade e estruturação dos cursos de graduação.

Mineração de dados e educação

De acordo com Cordeiro (2017), existe um crescimento na utilização de sistemas relacionados à Tecnologia da Informação (TI) na área da educação, o que gera um grande volume de dados armazenados, que muitas vezes não produzem nenhum conhecimento. Assim, surge a necessidade da associação desses dados com ferramentas e metodologias que são capazes de fazer uma análise e interpretação desses e, a partir disso, gerar informações que auxiliem na redução de problemas no processo de ensino-aprendizagem, gestão acadêmica, entre outros, com o intuito de melhorar o processo educacional.

Gonçalves *et al.* (2017) realizaram um estudo acerca do desempenho dos cursos de fisioterapia no Brasil por meio da relação das pontuações do Enade com o Conceito Preliminar de Curso (CPC), verificando a qualidade dos cursos entre as instituições de ensino superior públicas e privadas. No resultado do trabalho, os autores chegaram ao seguinte resultado: as IES públicas apresentaram melhor desempenho do que as instituições privadas.

O trabalho de Cretton e Gomes (2016) aplica técnicas de mineração de dados para identificar perfis de estudantes do curso de Medicina, utilizando a base de dados do Enade, além de extrair conhecimento a respeito do nível de dificuldade do componente específico da prova. A partir da mineração de dados, pode-se identificar a influência da categoria e dos tipos das instituições de ensino superior na criação dos perfis, sendo diretamente ligadas ao nível de desempenho dos estudantes e sua opinião acerca do nível do exame.

Fonseca e Namem (2016) analisaram informações relacionadas ao ensino fundamental público brasileiro. De maneira mais específica, foi relacionado o perfil dos professores de Matemática com a proficiência obtida por seus alunos. Os registros foram

extraídos da base de dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Segundo os autores, foi possível verificar fatores positivos e negativos que impactam no desempenho dos estudantes.

Lacerda *et al.* (2015) investigam as relações entre os indicadores presentes no Conceito Preliminar de Curso (CPC) dos cursos de Pedagogia no Brasil, com base nos resultados do Exame Nacional de Desempenho dos Estudantes (Enade) para essa área. Sete indicadores foram analisados: percentual de doutores, percentual de mestres, percentual de docentes em regime integral ou parcial, infraestrutura, organização didático-pedagógica, Indicador de Diferença entre os Desempenhos Observado e Esperado (IDD) e a nota no Enade. Os resultados apresentados, segundo as combinações de organização acadêmica e categoria administrativa das IES brasileiras, mostraram relações significativas dos percentuais de doutores e de mestres com o Enade e o IDD, sendo mais fortes nas universidades estaduais e nas faculdades integradas privadas.

Metodologia da pesquisa

A metodologia de pesquisa foi organizada em quatro etapas macro e sequenciais, conforme ilustrado na Figura 1. O processo do trabalho é baseado nas etapas do KDD, seguindo a proposta de Colak, Karaman e Turtay (2015).

Figura 1 – Metodologia do trabalho



Fonte: Elaboração própria

Na primeira etapa da pesquisa, foi realizada uma pesquisa bibliográfica, nas bases científicas *Scopus*[®] e *ScienceDirect*[®], relacionada com aspectos fundamentais de mineração de dados, de avaliações de cursos de graduação, o CPC e seus respectivos tesouros, de modo a conferir o arcabouço bibliográfico para tanto fundamentar a investigação científica quanto para comparar os resultados alcançados com já relatados pela comunidade acadêmica na quarta etapa.

Na segunda etapa, foi realizada a coleta de dados, a qual é composta por duas fases importantes: a obtenção e a seleção dos dados. Neste trabalho, os dados são originários da base de dados que o INEP disponibiliza referente à avaliação do CPC dos cursos de graduação e podem ser acessados publicamente por qualquer usuário na própria plataforma do INEP (inep.gov.br/conceito-preliminar-de-curso-cpc).

São considerados como população todos os cursos de Licenciatura da Rede Federal que obtiveram avaliação no CPC de 2014 e/ou no CPC de 2017, recuperando as fichas de avaliação de todos os cursos de graduação listados e as informações organizadas em tabela, na qual cada linha representa um curso e cada coluna, um item de avaliação. Os anos de 2014 e 2017 foram selecionados, pois referem-se às últimas avaliações do CPC

realizadas para os cursos pertencentes ao Ano II (Ciências Exatas, Licenciaturas e áreas afins), com dados disponíveis.

O Quadro 1 apresenta de forma clara e coesa os atributos presentes na base de dados do INEP, que foram utilizados nesta etapa do presente trabalho. Dentre os dados provenientes estão inseridos os critérios para cálculo da nota do CPC. Os demais dados estão relacionados a informações técnicas dos cursos de graduação.

A partir da realização da seleção de dados na base supracitada, foram gerados dois arquivos: um com os dados dos cursos de Licenciatura da Rede Federal avaliados em 2014 (290 registros) e outro com os dados dos cursos de Licenciatura da Rede Federal avaliados em 2017 (338 registros).

Quadro 1 – Dados retirados da base de dados do INEP

Atributos	Definição	Exemplo
Ano	Ano referente à avaliação realizada.	2014
Código de Área	Código específico para cada curso de graduação.	702 – Matemática (Licenciatura)
Área de Enquadramento	Nome do Curso de Graduação.	Matemática (Licenciatura)
Código/Nome e Sigla da IES	- Código que identifica uma IES específica; - Nome da IES; - Sigla da IES.	- 1120 - Instituto Federal de Educação, Ciência e Tecnologia Fluminense - IF Fluminense
Categoria Administrativa	As IES podem ser classificadas em: Federal, Municipal, Estadual ou Privada.	Federal
Organização Acadêmica	As IES podem ser: Faculdade, Centro Universitário, Instituto Federal de Educação, Ciência e Tecnologia, Universidade ou Centro Federal de Educação e Tecnologia.	Instituto Federal de Educação, Ciência e Tecnologia
Código do Município, Município, UF e Região do Curso	Informações referentes a localidade dos cursos.	Código do Município: 3301009 Município: Campos dos Goytacazes UF: RJ Região: Sudeste
Número de Cursos da Unidade	Quantidade total de cursos existentes na unidade específica.	1
Concluintes inscritos	Número de Alunos inscritos no exame do Enade	27
Concluintes Participantes	Número de Alunos que efetivamente realizaram o exame do Enade	26
Nota Bruta – FG	Desempenho médio de seus concluintes na Formação Geral (FG) – Base para cálculo da nota do Enade.	55,11
Nota Bruta – CE	Desempenho médio de seus concluintes no Componente Específico (CE) – Base para cálculo da nota do Enade.	31,79
Nota Bruta Geral	Nota bruta Geral do curso de graduação específico e da área de avaliação correspondente – Base para cálculo da nota do Enade.	37,63
Nota Contínua do Enade	Nota Contínua do Enade.	2,79

Nota - Organização Didático-Pedagógica	O conceito está explicado na subseção 2.1.3 com maior detalhamento.	Bruta: 5,55 Padronizada: 3,87
Nota - Infraestrutura e Instalações Físicas	O conceito está explicado na subseção 2.1.3 com maior detalhamento.	Bruta: 5,37 Padronizada: 3,94
Nota - Oportunidades de Ampliação da Formação	O conceito está explicado na subseção 2.1.3 com maior detalhamento.	Bruta: 4,90 Padronizada: 3,20
Concluintes Participantes com Nota no Enem (Percentual)	Número Total e Percentual dos concluintes que participaram do Enade, e possuem Nota no Enem.	Total: 11 Percentual: 42,3 %
Nota – IDD	O conceito está explicado na subseção 2.1.1 com maior detalhamento.	Bruta: 1,70 Padronizada: 2,84
Número de Docentes	Quantidade total de docentes que estão vinculados ao curso de graduação.	22
Número de Matrículas	Quantidade total de matrículas realizadas no curso de graduação.	86
Nota – Mestres	O conceito está explicado na subseção 2.1.2 com maior detalhamento.	Bruta: 0,86 Padronizada: 4,31
Nota – Doutores	O conceito está explicado na subseção 2.1.2 com maior detalhamento	Bruta: 0,18 Padronizada: 0,90
Nota – Regime de Trabalho	O conceito está explicado na subseção 2.1.2 com maior detalhamento	Bruta: 1,00 Padronizada: 5,00
CPC – Contínuo	Nota do Conceito Preliminar de Curso	2,96
CPC – Faixa	Nota do Conceito Preliminar de Curso	4

Fonte: Elaboração própria.

Observou-se que os dados extraídos provenientes do Quadro 1 tinham excesso de informações, por isso foi necessário realizar o tratamento, com o objetivo de trabalhar apenas com os dados pertinentes ao presente trabalho.

Assim, foi realizada uma limpeza na base, por meio dos seguintes procedimentos: de todos os atributos apresentados no Quadro 1, foram extraídos nessa etapa apenas os critérios utilizados efetivamente no cálculo do CPC: Nota dos concluintes no Enade, Nota do Indicador de Diferença entre os Desempenhos Observados e Esperados, Nota de Proporção de Mestres, Nota de Proporção de Doutores, Nota de Regime de Trabalho, Nota Referente à Organização Didático-pedagógica, Nota Referente à Infraestrutura e Instalações Físicas, Nota Referente às Oportunidades de Ampliação da Formação Acadêmica e Profissional e CPC - Faixa. Os demais atributos foram excluídos das bases de dados geradas. Vale ressaltar que entre esses dados está o dado classificador do presente trabalho (desfecho), que é aquele que descreve a nota final do CPC do curso de graduação avaliado, podendo variar de 1 a 5 (CPC – Faixa).

Na tabela 1 é possível observar os quesitos utilizados pelo INEP para orientar as avaliações de todas as áreas, bem como as ponderações de cada critério avaliado (BRASIL; INEP, 2015).

Tabela 1 – Critérios para cálculo do CPC

Dimensão	Componentes	Peso	
Desempenho dos Estudantes	Nota dos Concluintes no Enade (NC)	20%	
Valor Agregado	Nota do Indicador de Diferença entre os Desempenhos Observados e Esperados (NIDD)	35%	55%
Corpo Docente	Nota de Proporção de Mestres (NM)	7,5 %	30 %
	Nota de Proporção de Doutores (ND)	15,0 %	
	Nota de Regime de Trabalho (NR)	7,5 %	
Percepção Discente sobre as Condições do Processo Formativo	Nota Referente à Organização Didático-pedagógica (NO)	7,5 %	15 %
	Nota Referente à Infraestrutura e Instalações Físicas (NF)	5,0 %	
	Nota Referente às Oportunidades de Ampliação da Formação Acadêmica e Profissional (NA)	2,5 %	

Fonte: (BRASIL; INEP, 2015).

Os dados foram preparados para serem submetidos ao processo de mineração de dados. Para transformação dos dados, foi necessária a escolha da ferramenta utilizada, pois se faz necessário formatar as bases de dados de modo compatível com ela.

Após a coleta e o tratamento desses dados, esses foram processados e minerados utilizando o *software Waikato Environment for Knowledge Analysis – WEKA*⁴, versão 3.8.3 (FRANK *et al.*, 2009). De acordo com Maschio *et al.* (2018), que realizou um Mapeamento Sistemático da Literatura no que se refere à mineração de dados educacionais, os autores corroboram que a ferramenta WEKA foi a que mais apresentou suporte para a realização do processo de mineração.

Para a geração dos resultados tendo como desfecho a Nota Final do CPC (2014/2017), o método J48 foi executado no WEKA, utilizando como parâmetro de treinamento da rede o método *crossvalidation*, que assim é explicado por Cordeiro (2017, p. 45):

Nesse método, os dados são divididos em k conjuntos mutuamente exclusivos. A validação é realizada k vezes, de forma que a cada validação um conjunto diferente é utilizado como teste e os outros k-1 conjuntos são utilizados para treinamento. O resultado final é obtido a partir da média de todas as k validações.

4- Disponível em: <https://www.cs.waikato.ac.nz/ml/weka/>

Assim, foi possível, por meio de um vasto número de dados referentes às avaliações, obter informações importantes para auxiliar na análise e interpretação desses resultados, corroborando o objetivo da mineração de dados.

Na quarta e última etapa, as árvores de decisão, geradas por meio do processo de mineração de dados realizado pelo *software* WEKA, foram confrontadas com os procedimentos de atribuição de notas do próprio INEP e com trabalhos encontrados na literatura, gerando discussão acerca dos resultados encontrados.

Resultados

Nesta seção, são apresentados os resultados obtidos após a execução da etapa de mineração de dados da metodologia apresentada. As matrizes de confusão (utilizadas para, entre outros, calcular a taxa de acerto) de cada avaliação podem ser observadas na tabela 2.

Tabela 2 – Matrizes de confusão – Nota final (*cross-validation*)

Classificado como:	Rede Federal (2014)					Rede Federal (2017)				
	N1	N2	N3	N4	N5	N1	N2	N3	N4	N5
Nota 1	0	1	0	0	0	0	0	0	0	0
Nota 2	0	10	5	0	0	0	7	6	0	0
Nota 3	0	5	79	8	0	0	11	180	23	0
Nota 4	0	0	7	21	0	0	0	21	77	4
Nota 5	0	0	0	1	0	0	0	0	3	0

Fonte: Elaboração própria.

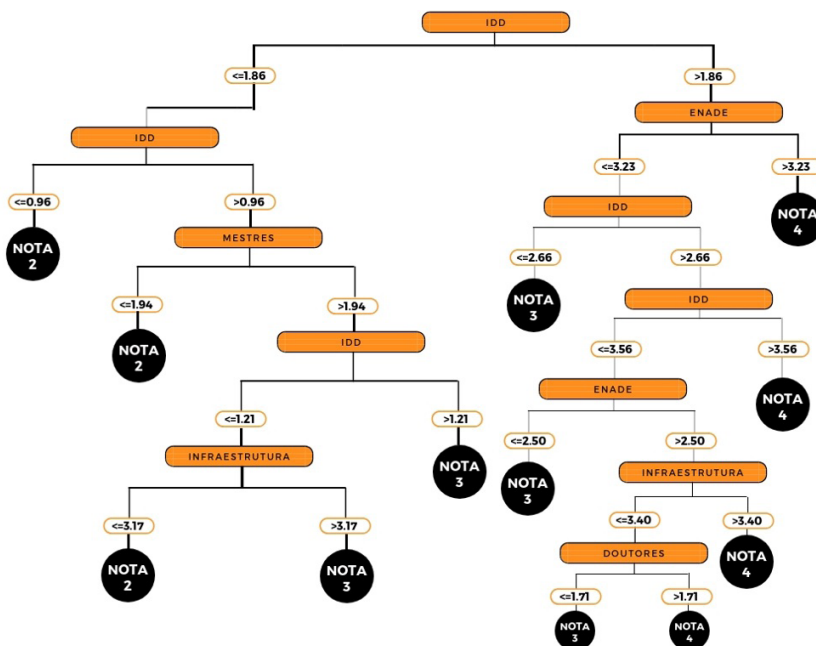
Avaliação da Rede Federal (2014)

A Figura 2 ilustra os resultados para o modelo de árvore de decisão calculado tendo como desfecho a Nota Final do CPC de 2014 dos cursos de Licenciatura da Rede Federal de todo o Brasil. O total de cursos avaliados foi de 290. A taxa de acerto obtida da base de dados para o método de classificação J48 e parâmetro de treinamento *cross-validation* foi de 80,29%.

A partir disso, é possível analisar os principais padrões referentes à avaliação do CPC de 2014 dos cursos de Licenciatura na Rede Federal, como também, obter conhecimento importante relativo a estes padrões.

A partir da árvore gerada, constataram-se os seguintes perfis para os cursos de Licenciatura da Rede Federal, considerando como desfecho a Nota de Avaliação do CPC de 2014. A árvore apresentada na Figura 2 será descrita tendo como referência de leitura da esquerda para a direita.

Figura 2 – Árvore de decisão da Rede Federal



Fonte: Elaboração própria.

A avaliação com nota 2 pode ser obtida nos seguintes casos: se a nota do Indicador da Diferença entre os Desempenhos Observados e Esperados (IDD) for menor ou igual a 0,96; se a nota do IDD estiver entre 0,96 e 1,86 e a nota de Mestres não for superior a 1,94; se o IDD estiver entre 0,96 e 1,21 e a infraestrutura não atingir pelo menos uma nota superior a 3,17.

A avaliação com nota 3 pode ser obtida nos seguintes casos: se a nota do IDD apresentar valores entre 0,96 e 1,21, a nota referente à infraestrutura precisa ser pelo menos 3,17; se a nota do IDD estiver entre 1,21 e 1,86, a nota de Mestres precisa ser maior que 1,94; se a nota do IDD estiver no intervalo de 1,86 e 2,66 e a nota do Enade não atingir pelo menos 3,23; se a nota do IDD estiver entre 2,66 e 3,56 e a nota do Enade não atingir 2,50; se a nota do IDD estiver entre 2,66 e 3,56, a nota referente à infraestrutura não atingir 3,40 e a Nota de Doutores for inferior a 1,71.

A avaliação com nota 4 pode ser obtida nos seguintes casos: se a nota do IDD for superior a 1,86 e a nota do Enade for maior que 3,23; se a nota do Enade for igual ou inferior a 3,23, a nota do IDD precisa ser superior a 3,56; se a nota do IDD estiver entre 2,66 e 3,56, e a nota do Enade no intervalo de 2,50 e 3,23, é necessário que a nota referente à infraestrutura atinja pelo menos 3,40, senão é necessário que a nota de Doutores seja superior a 1,71.

É importante ressaltar que a árvore gerada não apresentou nenhum perfil referente à avaliação 5 no CPC, visto que apenas um curso (Licenciatura em Química na Universidade Tecnológica Federal do Paraná - UTFPR) foi avaliado com esse conceito em 2014.

Por meio da análise da árvore de decisão apresentada, é possível inferir que os critérios que mais impactaram em uma boa avaliação no CPC dos cursos da Rede Federal

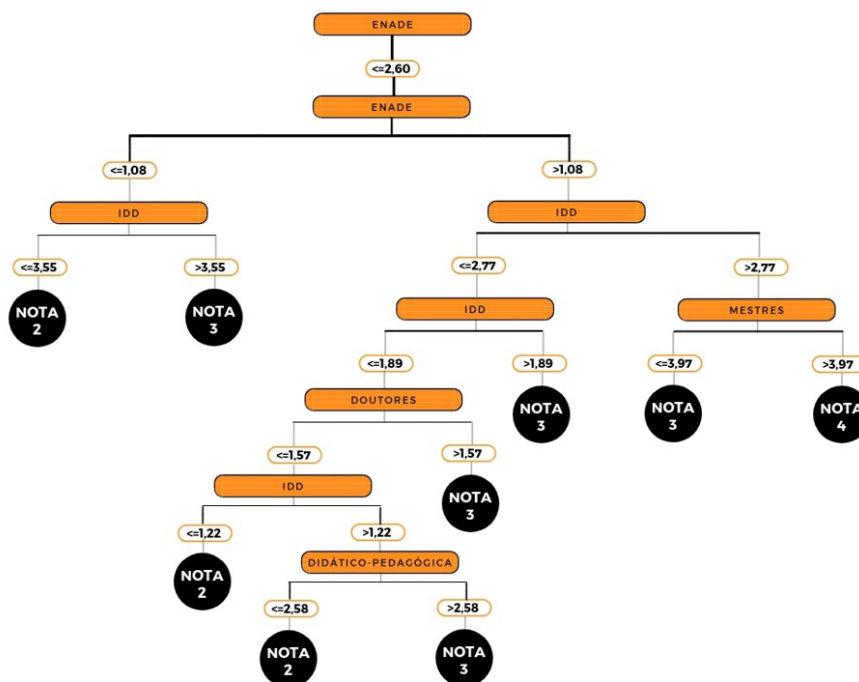
foram: a nota do Indicador da Diferença entre os Desempenhos Observados e Esperados (IDD) e a nota do Enade, ambos aspectos relacionados diretamente com o discente. Porém, vale ressaltar, que os critérios infraestrutura e nota de doutores (critérios independentes dos alunos) são alicerces fundamentais para um bom desempenho na avaliação, pois tendo esses dois critérios bem avaliados, é possível manter a boa avaliação do curso (nota 4).

Avaliação da Rede Federal (2017)

As Figuras 3 e 4 estão particionadas representando uma única árvore de decisão, devido à delimitação de espaço e para melhor compreensão dela. Elas ilustram os resultados para o modelo de árvore de decisão calculado tendo como desfecho a nota final do CPC de 2017 dos cursos de Licenciatura da Rede Federal de todo o Brasil. O total de cursos avaliados foi de 338. A taxa de acerto obtida da base de dados para o método de classificação J48 e parâmetro de treinamento (*cross-validation*) foi de 80,51%.

A partir disso, é possível analisar os principais padrões referentes à avaliação do CPC de 2017 dos cursos de Licenciatura na Rede Federal, como também, obter conhecimento importante relativo a esses padrões. Vale ressaltar que para a avaliação de 2017 serão descritas apenas a avaliação nota 2 (curso próximo ao descredenciamento) e a avaliação nota 5 (curso com excelência na avaliação).

Figura 3 – Árvore de decisão da Rede Federal



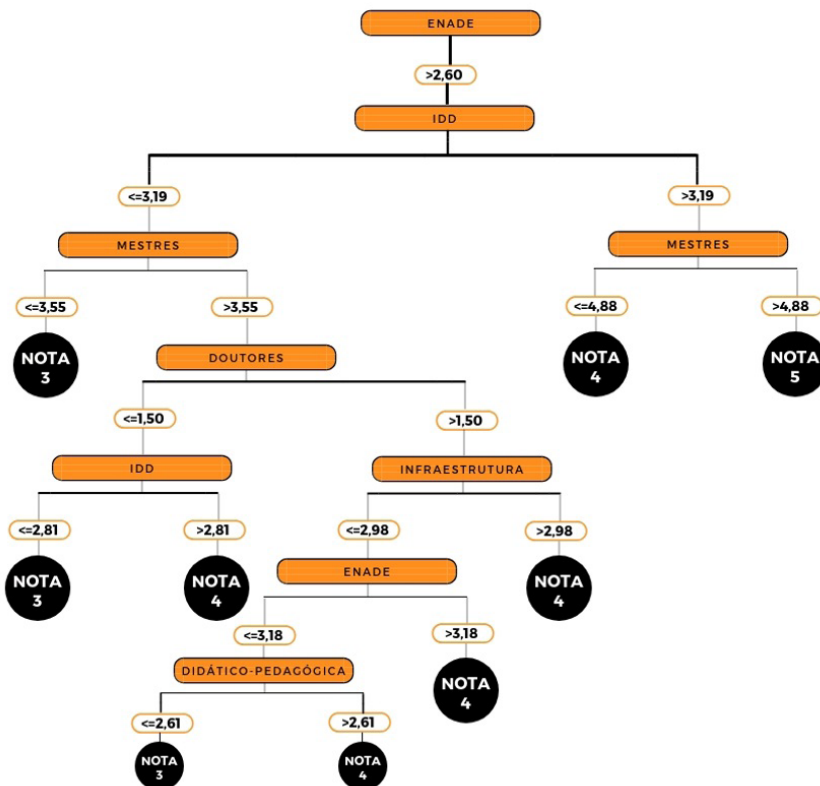
Fonte: Elaboração própria.

A avaliação com nota 2 pode ser obtida nos seguintes casos, como apresentado na Figura 3: se a nota do Enade for menor ou igual a 1,08 e a nota do IDD for inferior a 3,55; se a nota do Enade estiver entre 1,08 e 2,60, a nota de Doutores for inferior ou igual a 1,57 e a nota do IDD menor ou igual a 1,22; se a nota do Enade estiver entre 1,08 e 2,60, a nota de Doutores for inferior ou igual a 1,57, a nota do IDD estiver entre 1,22 e 2,77 e a nota referente à Organização Didático-Pedagógica for menor ou igual a 2,58.

A avaliação com nota 3 pode ser obtida nos seguintes casos, como apresentado na Figura 3: se a nota do Enade for menor ou igual a 1,08 e a nota do IDD for superior a 3,55; se a nota do Enade estiver entre 1,08 e 2,60, a nota do IDD estiver entre 1,22 e 1,89, e a nota da Organização Didático-Pedagógica superior a 2,58; se a nota do Enade estiver entre 1,08 e 2,60, a nota do IDD inferior ou igual a 1,89 e a nota de Doutores for superior a 1,57; se a nota do Enade estiver entre 1,08 e 2,60, a nota do IDD estiver entre 1,89 e 2,77; se a nota do Enade estiver entre 1,08 e 2,60, a nota do IDD for superior a 2,77 e a nota de Mestres menor ou igual a 3,97.

A avaliação com nota 4 pode ser obtida nos seguintes caso, como apresentado na Figura 3: se a nota do Enade estiver entre 1,08 e 2,60, a nota do IDD for superior a 2,77 e a nota de Mestres superior a 3,97.

Figura 4 – Árvore de decisão da Rede Federal



Fonte: Elaboração própria.

A avaliação com nota 3 pode ser obtida nos seguintes casos, como apresentado na Figura 4: se a nota do Enade for superior a 2,60, a nota do IDD menor ou igual a 3,19 e a nota de Mestres inferior ou igual a 3,55; se a nota do Enade for superior a 2,60, a nota de Mestres superior a 3,55, a nota de Doutores menor ou igual a 1,50 e a nota do IDD menor que 2,81; se a nota do Enade estiver entre 2,60 e 3,18, a nota do IDD menor ou igual a 3,19, a nota de Mestres superior a 3,55, a nota de Doutores superior a 1,50, a nota de Infraestrutura menor ou igual a 2,98 e a nota da Organização Didático-Pedagógica inferior ou igual a 2,61.

A avaliação com nota 4 pode ser obtida nos seguintes casos, como apresentado na Figura 4: se a nota do Enade for superior a 2,60, a nota de Mestres superior a 3,55, a nota de Doutores menor ou igual a 1,50 e a nota do IDD estiver entre 2,81 e 3,19; se a nota do Enade estiver entre 2,60 e 3,18, a nota do IDD menor ou igual a 3,19, a nota de Mestres superior a 3,55, a nota de Doutores superior a 1,50, a nota de Infraestrutura menor ou igual a 2,98 e a nota da Organização Didático-Pedagógica superior a 2,61; se a nota do Enade for maior que 3,18, a nota do IDD menor ou igual a 3,19, a nota de Mestres superior a 3,55, a nota de Doutores superior a 1,50 e a nota de Infraestrutura menor ou igual a 2,98; se a nota do Enade for maior que 2,60, a nota do IDD menor ou igual a 3,19, a nota de Mestres superior a 3,55, a nota de Doutores superior a 1,50 e a nota de Infraestrutura superior a 2,98; se a nota do Enade for maior que 2,60, a nota do IDD for maior que 3,19, a nota de Mestres for menor ou igual a 4,88.

A avaliação com nota 5 pode ser obtida no seguinte caso, como apresentado na Figura 4: se a nota do Enade for superior a 2,60, a nota do IDD maior que 3,19 e a nota de Mestres for superior a 4,88.

Discussão dos resultados

Na tabela 3, apresenta-se a quantidade de vezes que os critérios (quesitos e/ou itens) figuraram em cada uma das árvores de decisão geradas a partir dos dados coletados.

Tabela 3 – Quantidade de vezes os critérios estão presentes nas árvores

Critérios	Rede Federal (2014)	Rede Federal (2017)
ENADE	2	3
Indicador de Diferença entre Desempenhos Observados e Esperados (IDD)	5	6
Mestres	1	3
Doutores	1	2
Regime de Trabalho	-	-
Organização Didático-pedagógica (ODP)	-	2
Infraestrutura e Instalações físicas	2	1
Oportunidades de Ampliação da Formação Acadêmica e Profissional (OAF)	-	-

Fonte: Elaboração própria.

A partir da tabela 3, nota-se que os itens com maior preeminência nas árvores de decisão são o IDD e o Enade, sendo que figuram como os quesitos com maior dependência dos discentes, estando no topo da árvore de decisão nas duas árvores geradas.

A partir disso, pode-se inferir que são os critérios determinantes para o desfecho Avaliação do CPC. Esse resultado é aderente à proposta do cálculo do CPC, que define como ponderações de cada critério 35% e 20% para o IDD e Enade, respectivamente. Ou seja, esses critérios têm o maior peso no cálculo do CPC, corroborando a extração de conhecimento das árvores geradas.

Assim sendo, na tabela 4 apresentam-se os critérios com maior impacto em cada uma das árvores de decisão gerada a partir dos dados coletados, sendo representados de forma ordinal (1º e 2º). É possível verificar a predominância dos itens Enade e IDD nas duas árvores geradas, o que confirma a importância deles para a avaliação.

Tabela 4 – Critérios de maior impacto presentes nas árvores

Critérios	Rede Federal 2014	Rede Federal 2017
Enade	2º	1º
Indicador de Diferença entre Desempenhos Observados e Esperados (IDD)	1º	2º

Fonte: Elaboração própria.

Como a Rede Federal é composta em sua maioria por instituições de ensino relativamente novas comparados às Universidades, além de ofertarem outras modalidades de ensino, o número de curso existentes é consideravelmente menor que nas outras IES. Assim, a árvore gerada de 2014 apresentou um número reduzido de folhas devido a essas características e não apresentou nenhum perfil para Nota 5, pois na avaliação de 2014 apenas 1 curso da Rede Federal obteve essa nota, que foi a Licenciatura em Química da UTFPR (Universidade Tecnológica Federal do Paraná).

Quadro 2 – Critérios considerados nas avaliações Nota 2 e Nota 4/5

IES	Critérios Considerados	
	Nota 2	Nota4/Nota 5
Rede Federal 2014	<ul style="list-style-type: none"> • IDD • Mestres • Infraestrutura 	<ul style="list-style-type: none"> • Enade • IDD • Doutores
Rede Federal 2017	<ul style="list-style-type: none"> • Enade • IDD • Doutores 	<ul style="list-style-type: none"> • Enade • IDD • Mestres

Fonte: Elaboração própria.

O quadro 2 apresenta um panorama dos critérios considerados nas Avaliações dos extremos (Nota 2 e Nota 4/5). A partir disso, pode-se concluir que os critérios Mestres e Doutores apareceram nas duas árvores geradas, podendo impulsionar a nota do CPC dos cursos de graduação, fazendo com que os mesmos consigam desempenho melhores. Isso é corroborado com o trabalho de Lacerda *et al.* (2015), que por meio de procedimentos estatísticos demonstram que existem relações significativas dos percentuais de doutores e de mestres com o Enade e o IDD.

O método de classificação por árvore de decisão é explicado por Cretton e Gomes (2016): o critério de maior significância encontra-se na raiz da árvore e a partir dele, geram-se ramificações, que representam a relevância dessa ligação. No processo da geração das árvores de decisão, as ramificações (folhas) são geradas baseadas na utilidade do atributo para a classificação. Assim, ocorre um determinado ganho de informação a cada atributo verificado, ou seja, o atributo escolhido como raiz da árvore é aquele que possui o maior ganho de informação.

A partir disso, é possível relacionar a posição do nó nas árvores de decisão geradas com a importância do critério, porém, vale ressaltar que além disso, é importante verificar e analisar as especificidades de cada árvore, tendo assim as seguintes características no presente trabalho:

- Critério de maior impacto (raiz da árvore): que impacta diretamente em qualquer perfil dos cursos analisados, ou seja, qualquer avaliação do CPC dependerá inicialmente dele.
- Critério significativo (ramificações): que impacta diretamente avaliações específicas do CPC, como por exemplo, Nota 2 ou Nota 5.
- Critério determinante (ramificações): que impacta diretamente nos perfis relacionados às melhores avaliações possíveis do CPC.
- Critério suporte (ramificações): que tem como objetivo a evolução da Nota do CPC, ou seja, este critério permeia duas avaliações e de acordo com a nota recebida pelo critério, consegue evoluir a nota final do CPC.

O quadro 3 apresenta uma síntese dos resultados obtidos, relacionando os critérios considerados na avaliação do CPC com as Instituições de Ensino Superior (IES). Vale ressaltar que alguns critérios não apresentaram impacto significativo para a Rede Federal, destacando-se os critérios: oportunidades de ampliação da formação acadêmica e profissional e regime de trabalho, ou seja, não apareceram em nenhum momento nas árvores geradas.

Quadro 3 – Síntese dos critérios relacionados às IES.

Crítérios	Rede Federal - 2014	Rede Federal - 2017
Enade	Crítério determinante: mantém a avaliação do CPC entre 3 e 4.	Crítério de maior impacto na avaliação do CPC.
Indicador de Diferença entre Desempenhos Observados e Esperados (IDD)	Crítério de maior impacto na avaliação do CPC.	Crítério Suporte: Se bem avaliado consegue evoluir a Nota do CPC.
Nota de Mestres	Crítério Significativo: para obter avaliação (Nota 2).	Crítério Significativo: para obter avaliação (Nota 5).
Nota de Doutores	Crítério determinante: mantém a avaliação do CPC entre 3 e 4.	Crítério Suporte: Se bem avaliado consegue evoluir a Nota do CPC.
Regime de Trabalho	Não apresentou impacto significativo na avaliação.	Não apresentou impacto significativo na avaliação.
Organização Didático-pedagógica (ODP)	Não apresentou impacto significativo na avaliação.	Crítério Suporte: Se bem avaliado consegue evoluir a Nota do CPC.
Infraestrutura e Instalações Físicas	Crítério Suporte: Se bem avaliado consegue evoluir a Nota do CPC.	Crítério determinante: mantém a avaliação do CPC entre 3 e 4.
Oportunidades de Ampliação da Formação Acadêmica e Profissional (OAF)	Não apresentou impacto significativo na avaliação.	Não apresentou impacto significativo na avaliação

Fonte: Elaboração própria.

Por meio de uma análise comparativa das árvores de decisão de 2014 e 2017 referente à avaliação CPC da Rede Federal, pode-se observar:

- O critério principal, que está no topo da árvore e relacionado com o maior impacto na avaliação do CPC, em 2014 era a Nota do IDD, já em 2017 passou a ser a Nota do Enade;
- Como a Rede Federal apresentou um número maior de cursos avaliados em 2017, foi possível determinar um perfil para avaliação nota 5 (Nota do Enade superior a 2,60, Nota do IDD maior que 3,19 e Nota de Mestres superior a 4,88);
- O critério organização didático-pedagógica não impactou em nenhum dos perfis dos cursos na avaliação de 2014, já na avaliação de 2017 foi considerado um critério suporte, ou seja, se bem avaliado consegue evoluir a nota do CPC;
- O critério Mestre passou a ter uma grande preeminência na árvore de 2017, desempenhando um papel fundamental na obtenção de avaliações Nota 4 e Nota 5.

Conclusão

Este trabalho não teve a ambição de propor e apresentar medidas para a evolução na nota do CPC dos cursos das IES, mas sim determinar os critérios os quais gestores dos cursos de graduação, em especial dos cursos de Licenciatura, devem observar. Isso foi realizado a partir da extração de conhecimento da avaliação do CPC de 2014 e 2017, identificando os critérios com poder discricionário na avaliação e gerando as árvores de decisão de modo a subsidiar os gestores dos cursos de graduação.

As taxas de acertos obtidas ultrapassaram 80%, assim sendo, os resultados são considerados satisfatórios pois permitiram traçar perfis sólidos dos cursos de Licenciatura em relação à nota obtida na avaliação do CPC.

Espera-se que a partir dos padrões e conhecimentos extraídos e apresentados, seja possível auxiliar os gestores nas suas tomadas de decisões, no que se refere às medidas a serem tomadas e melhoria dos projetos de ensino para aprimorar os cursos avaliados no CPC.

Acredita-se que a principal contribuição desta pesquisa é explicitação dos perfis que descrevem os cursos de Licenciatura das IES analisadas e, consequentemente, o subsídio de informações para os gestores dos cursos de graduação. Outra contribuição está relacionada à utilização de técnicas de mineração de dados no contexto educacional, área que foi definida como emergente, e a identificação dos critérios que descrevem o perfil dos cursos de graduação em Licenciatura, sendo essa uma modalidade de ensino pouco explorada na mineração de dados educacionais.

Referências

AHUJA, Ravinder *et al.* Analysis of educational data mining. In: YADAV, Neha *et al.* (ed.). **Advances in intelligent systems and computing**. v. 741. Singapore: Springer, 2019. p. 897-907.

BAKER, Ryan; ISOTANI, Seiji; CARVALHO, Adriana. Mineração de dados educacionais: oportunidades para o Brasil. **Revista Brasileira de Informática na Educação**, Porto Alegre, v. 19, n. 02, p. 3-13, 31 ago. 2011. Disponível em: <http://dx.doi.org/10.5753/rbie.2011.19.02.03>. Acesso em: 12 out. 2020.

BRASIL. **Lei nº 11.892**, de 29 de dezembro de 2008. Institui a Rede Federal de Educação Profissional, Científica e Tecnológica, cria os Institutos Federais de Educação, Ciência e Tecnologia, e dá outras providências. Brasília, DF: [s. n.], 2008. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/lei/11892.htm. Acesso em: 10 dez. 2018.

BRASIL. Conselho Nacional de Educação. Câmara de Educação Básica. **Escassez de professores no ensino médio: propostas estruturais e emergenciais**. Brasília, DF: MEC, 2007. Disponível em: <http://portal.mec.gov.br/cne/arquivos/pdf/escassez1.pdf>. Acesso em: 10 dez. 2018.

BRASIL; INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Nota técnica DAES/ INEP Nº 58/2015**: cálculo do conceito preliminar de curso 2014. Brasília, DF: INEP, 2015. Disponível em: http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2014/nota_tecnicadaes_n582015_calculo_do_cpc2014.pdf. Acesso em: 15 nov. 2018.

BRASIL. Ministério da Educação e Cultura. **Conceito Preliminar de Curso (CPC)**. Brasília, DF: MEC, 2015. Disponível em: <http://portal.inep.gov.br/web/guest/educacao-superior/indicadores-de-qualidade>. Acesso em: 18 dez. 2018.

COLAK, Cemil; KARAMAN, Esra; TURTAY, M. Gokhan. Application of knowledge discovery process on the prediction of stroke. **Computer Methods and Programs in Biomedicine**, Dublin, v. 119, n. 3, p. 181-185, May 2015. <http://dx.doi.org/10.1016/j.cmpb.2015.03.002>. Acesso em: 12 out. 2020.

CORDEIRO, Renata Gomes. **Identificação do comportamento dos estudantes evadidos de cursos técnicos utilizando técnicas de mineração de dados**. 2017. 91 f. Dissertação (Mestrado) – Instituto Federal Fluminense, Campos dos Goytacazes, 2017.

CRETTON, Nícollas Nogueira; GOMES, Georgia Rodrigues. Aplicação de técnicas de mineração de dados na base de dados do ENADE com enfoque nos cursos de medicina. **Acta Biomédica Brasiliensia**, Itaperuna, v. 7, n. 1, p. 74-89, 20 jun. 2016. <http://dx.doi.org/10.18571/acbm.100>. Acesso em 14 out. 2020.

FONSECA, Stella Oggioni da; NAMEN, Anderson Amendoeira. Mineração em bases de dados do INEP: uma análise exploratória para nortear melhorias no sistema educacional brasileiro. **Educação em Revista**, Belo Horizonte, v. 32, n. 1, p. 133-157, mar. 2016. <http://dx.doi.org/10.1590/0102-4698140742>. Acesso em 14 out. 2020.

FORPLAN. Fórum de Planejamento e Administração. **Metodologia da matriz orçamentária da rede de ensino profissional e tecnológico de 2018**. Brasília, DF: Conif, 2018. Disponível em: <https://www.ifpb.edu.br/joaopessoa/institucional/planejamento-institucional/documentos-insitucionais/metodologia-da-matriz-conif-2018.pdf>. Acesso em: 10 nov. 2018.

FRANK, Eibe *et al.* Weka-A machine learning workbench for data mining. *In*: MAIMON, Oded; ROKACH, Lior (org.). **Data mining and knowledge discovery handbook**. Boston: Springer, 2009. p. 1269-1277.

GONÇALVES, Rogério Fabiano *et al.* Avaliação dos cursos de fisioterapia nos anos de 2004 a 2013. **Fisioterapia e Pesquisa**, São Paulo, v. 24, n. 4, p. 392-398, dez. 2017. <http://dx.doi.org/10.1590/1809-2950/17167124042017>. Acesso em 14 out. 2020.

LACERDA, Leo Lynce Valle de; FERRI, Cássia. Relações entre indicadores de qualidade de ensino e desempenho de estudantes dos cursos de pedagogia do Brasil no Exame Nacional de Desempenho dos Estudantes. **Revista Brasileira de Estudos Pedagógicos**, Brasília, DF, v. 96, n. 242, p. 129-145, abr. 2015. <http://dx.doi.org/10.1590/s2176-6681/330512908>. Acesso em 14 out. 2020.

MACHADO, Roger Douglas *et al.* Estudo bibliométrico em mineração de dados e evasão escolar. *In*: CONGRESSO NACIONAL DE EXCELÊNCIA EM GESTÃO, 11., 2015, Rio de Janeiro. **Anais...** Rio de Janeiro: CNEG, 2015. p. 1-21.

MASCHIO, Pedro *et al.* Um panorama acerca da mineração de dados educacionais no Brasil. *In*: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (SBIE), 29., 2018, Fortaleza. **Anais do SBIE 2018**. Porto Alegre: Brazilian Computer Society, 2018. p. 1936-1940.

MORORÓ, Leila Pio; PEREIRA, Cláudio Wilson dos Santos. A expansão dos institutos federais e dos cursos de licenciaturas no Brasil. *In: XII COLÓQUIO NACIONAL DO MUSEU PEDAGÓGICO*, 12., 2017, Vitória da Conquista. **Anais dos...** v. 12. Vitória da Conquista: UESB, 2017. p. 468-472.

MOURA, Amanda Gomes de. **Uma proposta de melhoria de processos para os institutos federais**. 2018. 144 f. Dissertação (Mestrado) – Instituto Federal Fluminense, Campos dos Goytacazes, 2018.

RIGO, Sandro José *et al.* Aplicações de mineração de dados educacionais e learning analytics com foco na evasão escolar: oportunidades e desafios. **Revista Brasileira de Informática na Educação**, Porto Alegre, v. 22, n. 01, p. 132-146, 18 maio 2014. <http://dx.doi.org/10.5753/rbie.2014.22.01.132>. Acesso em: 14 out. 2020.

ROIGER, Richard J. **Data mining: a tutorial-based primer**. Boca Raton: Crc Press, 2017. 487 p.

SUKHIJA, Karan; JINDAL, Manish; AGGARWAL, Naveen. The recent state of educational data mining: a survey and future visions. *In: INTERNATIONAL CONFERENCE ON MOOCS, INNOVATION AND TECHNOLOGY IN EDUCATION (MITE)*, 3., 2015, Amritsar. **Proceedings - IEEE...** Amritsar: IEEE, 2015. p. 354-359.

Recebido em: 06.02.2019

Revisado em: 01.10.2019

Aprovado em: 12.11.2019

Mariana Abreu Gualhano é mestre em engenharia de sistemas aplicados à engenharia e gestão. Atua como professora dos cursos de eletrônica, automação, eletromecânica e engenharia de controle e automação no Instituto Federal Fluminense (IFF), tendo como principais temas na sua área de pesquisa: arduino, *Android* e mineração de dados.

Aline Pires Vieira de Vasconcelos é doutora em engenharia de sistemas e computação pelo Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, da Universidade Federal do Rio de Janeiro (COPPE/UFRJ). É professora titular na área de informática do Instituto Federal Fluminense (IFF) e integrante do colegiado do Mestrado em Sistemas Aplicados à Engenharia e Gestão.