

## **Relação do perfil do aluno com o Conceito Preliminar dos Cursos da Rede Federal na Região Sudeste utilizando Mineração de Dados**

Mariana Abreu Gualhano<sup>1\*</sup>; Roberto Coutinho Medeiros Junior<sup>2</sup>; Yago Pessanha Corrêa<sup>3</sup>  
<sup>1</sup>Instituto Federal Fluminense; <sup>2</sup>Instituto Federal Fluminense; <sup>3</sup>Instituto Federal Fluminense  
\*mariana.gualhano@ifff.edu.br

### **Resumo**

A avaliação da qualidade dos cursos de graduação no Brasil é feita por meio do Conceito Preliminar de Cursos (CPC). Este artigo busca utilizar técnicas de Mineração de Dados para extrair conhecimento das avaliações do CPC do ano de 2017 com o intuito de relacionar o perfil do aluno e o conceito obtido pelos cursos da Rede Federal na região Sudeste. Para isso, utilizou-se a base de dados disponibilizada no portal do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, e foram utilizadas as etapas do processo KDD (*Knowledge Discovery in Databases*) para extração de conhecimento da base. Os resultados foram satisfatórios e mostraram que a modalidade de ensino é um critério de grande impacto para o desfecho analisado, além disso, a renda familiar obteve maior preeminência dentre os itens avaliados. Espera-se que as informações extraídas a partir do presente trabalho possam ser úteis e subsidiar à gestão educacional.

**Palavras-chave:** Rede Federal, CPC, Mineração de Dados.

### **1. 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<sup>[1]</sup>.

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 a 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<sup>[2]</sup>. 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.

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. Contudo, não há utilidade em grandes bases de dados sem a utilização de ferramentas que auxiliem na análise e interpretação desses dados<sup>[3]</sup>.

Em consonância com os aspectos relacionados a importância da avaliação da qualidade dos cursos da IES concomitantemente com a relevância dos cursos de graduação e seus alunos, 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.

## 2. Metodologia

A estratégia 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<sup>[4]</sup>.



**Figura 1.** Metodologia do Trabalho.

Inicialmente é realizada a obtenção e seleção dos dados. Neste trabalho os dados são originários de duas bases distintas:

- Base de dados que o INEP disponibiliza referente à avaliação do CPC dos cursos de graduação e podem ser acessados publicamente por qualquer usuário.
- Base de dados que o INEP disponibiliza referente aos relatórios dos cursos de graduação, no presente trabalho foi utilizado a parte que traduz o perfil dos estudantes de um determinado curso avaliado pelo Enade.

São considerados como amostra todos os cursos de graduação da Rede Federal, que obtiveram avaliação no CPC de 2017 da Região Sudeste. O ano de 2017 foi selecionado, pois refere-se a última avaliação do CPC realizada para os cursos pertencentes ao Ano II (Ciências Exatas, Licenciaturas e áreas afins).

Seguindo as etapas do KDD, iniciou-se a atividade de Tratamento dos Dados, realizados através do Pré-processamento dos Dados e, posteriormente, a Transformação dos mesmos. Para isso, foi realizado o seguinte procedimento: Iniciou-se por um filtro por tipo de Instituição (Rede Federal) e posteriormente pela Região do Curso (Sudeste). Nesta etapa da pesquisa, 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 moldar as bases de dados de modo compatível com a mesma.

A ferramenta escolhida para aplicar as técnicas de Mineração de Dados foi o software Waikato Environment for Knowledge Analysis – WEKA. De acordo com a literatura<sup>[5]</sup> a ferramenta WEKA foi a que mais apresentou suporte para a realização do processo de mineração. A partir da escolha da ferramenta, a base de dados foi moldada para que houvesse compatibilidade da utilização da mesma no *software*. Na etapa de mineração de dados é utilizado o método de classificação J48. Sendo esse, um método de árvore de decisão implementado com base no algoritmo C4.5 e disponível para utilização na ferramenta WEKA. O método foi escolhido fundamentado no trabalho<sup>[6]</sup> que corrobora o método como um dos mais utilizados na literatura no que se refere à mineração de dados educacionais. Para a execução do método, como parâmetro de treinamento da rede, foi selecionada a opção *cross-validation*.

Na última etapa, a árvore de decisão gerada foi confrontada com os procedimentos de atribuição de notas do próprio INEP, e com trabalhos encontrados na literatura, gerando discussão acerca dos resultados encontrados.

### 3. Resultados e Discussão

Nesta seção, são apresentados os resultados obtidos após a execução da etapa de mineração de dados da metodologia apresentada.

#### 3.1 Avaliação da Rede Federal na Região Sudeste

A Figura 2 ilustra os resultados para o modelo de árvore de decisão calculado tendo como desfecho a Nota do CPC dos cursos de Licenciatura da Região Sudeste do Brasil. O total de cursos avaliados foi de 84. A taxa de acerto obtida da base de dados para o método de classificação J48 e parâmetro de treinamento *cross-validation* foi de 80,60%. A partir disso, é possível analisar os principais perfis de estudantes referentes à avaliação final do CPC dos cursos de Licenciatura na Rede Federal na região Sudeste, como também, obter conhecimento importante relativo a estes padrões.

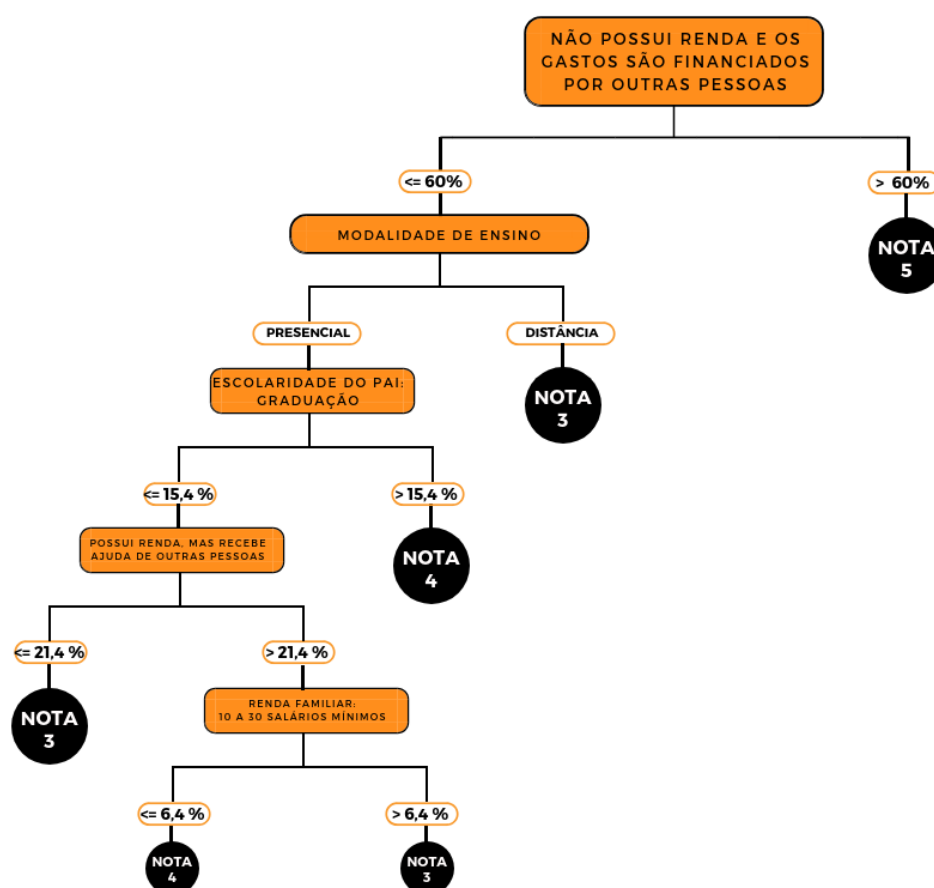


Figura 2. Árvore de Decisão da Região Sudeste.

Como pode-se observar na Figura 2, tem-se o seguinte perfil dos estudantes para os cursos com avaliação:

Nota 3: se menos de 60% dos estudantes não possuem renda e os gastos são financiados por outra pessoa, a modalidade de ensino é a distância; se menos de 60% dos estudantes não possuem renda e os gastos são financiados por outra pessoa, a modalidade de ensino é presencial, a escolaridade do pai de menos de 15,4% dos estudantes for graduação e menos de 21,4% dos estudantes possuem renda, mas recebem ajuda de outras pessoas para os gastos; se

menos de 60% dos estudantes não possuem renda e os gastos são financiados por outra pessoa, a modalidade de ensino é presencial, a escolaridade do pai de menos de 15,4% dos estudantes for graduação e mais de 21,4% dos estudantes possuem renda, mas recebem ajuda de outras pessoas para os gastos e mais de 6,4% possuem renda familiar de 10 a 30 salários mínimos.

Nota 4: se menos de 60% dos estudantes não possuem renda e os gastos são financiados por outra pessoa, a modalidade de ensino é presencial e a escolaridade do pai de mais de 15,4% dos estudantes for graduação; se menos de 60% dos estudantes não possuem renda e os gastos são financiados por outra pessoa, a modalidade de ensino é presencial, a escolaridade do pai de menos de 15,4% dos estudantes for graduação, mais de 21,4% dos estudantes possuem renda, mas recebem ajuda de outras pessoas para os gastos e menos de 6,4% possuem renda familiar de 10 a 30 salários mínimos.

Nota 5: se mais de 60% dos estudantes não possuem renda e os gastos são financiados por outra pessoa.

A partir da análise da Figura 2, nota-se que a modalidade de ensino (Presencial ou Distância) possui um impacto considerável na avaliação do CPC, visto que se a modalidade de ensino for à Distância, infere-se ao curso nota 3, apresentando assim a necessidade de um estudo mais aprofundado frente à este tipo de oferta.

Além disso, o item com maior preeminência nas árvores de decisão foi a Renda Familiar, de forma clara, observa-se que um curso no qual maior parte dos alunos são financiados por outra pessoa, obtém-se conceito máximo no CPC (Nota 5), visto que os discentes tendem a se preocupar apenas com o estudo. Corroborando isso, a árvore nos níveis mais baixos, apresentam dados referentes a Renda, porém as avaliações dos cursos estão entre 3 e 4.

#### 4. Conclusão

A partir deste contexto, pode-se concluir que através da extração de conhecimento da avaliação do CPC de 2017, foi possível identificar os critérios com poder discricionário e impacto na avaliação. As taxas de acertos obtidas ultrapassaram 80%, assim sendo, os resultados são considerados satisfatórios pois permitiram traçar perfis sólidos dos cursos de graduação e dos alunos em relação a nota obtida na avaliação do CPC.

Baseado nestes resultados, 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 as medidas a serem tomadas e melhoria dos projetos de ensino para aprimorar os cursos avaliados no CPC do ano de 2017.

#### Referências

- [1] BRASIL. 11.892. Lei nº 11.892, de 29 de dezembro de 2008. Rede Federal cria os Institutos Federais de Educação, Ciência e Tecnologia, e dá outras providências. 2008.
- [2] MEC. **Conceito Preliminar de Curso (CPC)**, 2015. Disponível em: <<http://portal.inep.gov.br/conceito-preliminar-de-curso-cpc>>
- [3] RIGO, S. J. 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**, v. 22, n. 01, p. 132, 18 maio 2014.
- [4] COLAK, C.; KARAMAN, E.; TURTAY, M. G. Application of knowledge discovery process on the prediction of stroke. **Computer Methods and Programs in Biomedicine**, v. 119, n. 3, p. 181–185, maio 2015.
- [5] MASCHIO, P. et al. **Um Panorama acerca da Mineração de Dados Educacionais no Brasil**. In: XXIX SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (BRAZILIAN SYMPOSIUM ON COMPUTERS IN EDUCATION). Fortaleza, Ceará, Brasil: 28 out. 2018.
- [6] CORDEIRO, R. G.; MUSSA, M. DE S.; DA HORA, H. R. M. Modelo para análise do perfil dos alunos evadidos em curso técnico utilizando técnicas de mineração de dados. 2017.