

**Instituto Federal de Educação, Ciência e Tecnologia Fluminense**  
**Programa de Pós-graduação em Sistemas Aplicados à Engenharia e Gestão**

**DISSERTAÇÃO**

**ANÁLISE DO DESEMPENHO DOS CURSOS DE LICENCIATURA DA  
REDE FEDERAL NO CENÁRIO NACIONAL POR MEIO DO CONCEITO  
PRELIMINAR DE CURSO UTILIZANDO MINERAÇÃO DE DADOS**

**MARIANA ABREU GUALHANO**

**2019**

**Instituto Federal de Educação, Ciência e Tecnologia Fluminense**  
**Programa de Pós-graduação em Sistemas Aplicados à Engenharia e Gestão**

ANÁLISE DO DESEMPENHO DOS CURSOS DE LICENCIATURA DA  
REDE FEDERAL NO CENÁRIO NACIONAL POR MEIO DO CONCEITO  
PRELIMINAR DE CURSO UTILIZANDO MINERAÇÃO DE DADOS

**MARIANA ABREU GUALHANO**

*Sob orientação da professora*

**Aline Pires Vieira de Vasconcelos**

Dissertação submetida como requisito parcial para obtenção do grau de **Mestre** no Programa de Pós-graduação em Sistemas Aplicados à Engenharia e Gestão, Área de Concentração em Sistemas Computacionais.

Campos dos Goytacazes, RJ

Fevereiro de 2019

Biblioteca Anton Dakitsch  
CIP - Catalogação na Publicação

G911a Gualhano, Mariana Abreu  
Análise do desempenho dos cursos de licenciatura da Rede Federal no cenário nacional por meio do Conceito Preliminar de Curso utilizando Mineração de Dados. / Mariana Abreu Gualhano - 2019.  
79 f.: il. color.

Orientadora: Aline Pires Vieira de Vasconcelos

Dissertação (mestrado) -- Instituto Federal de Educação, Ciência e Tecnologia Fluminense, Campus Campos Centro, Curso de Mestrado Profissional em Sistemas Aplicados à Engenharia e Gestão, Campos dos Goytacazes, RJ, 2019.  
Referências: f. 65 a 70.

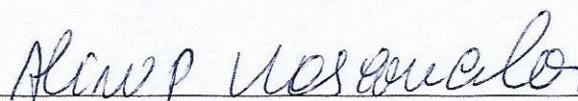
1. Avaliação. 2. Instituto Federal. 3. CPC. 4. Mineração de Dados. 5. Licenciatura. I. Vasconcelos, Aline Pires Vieira de , orient. II. Título.

**Instituto Federal de Educação, Ciência e Tecnologia Fluminense**  
**Programa de Pós-graduação em Sistemas Aplicados à Engenharia e Gestão**

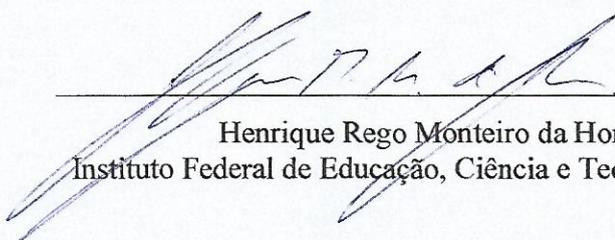
**MARIANA ABREU GUALHANO**

Dissertação submetida como requisito parcial para obtenção do grau de **Mestre** no Programa de Pós-graduação em Sistemas Aplicados à Engenharia e Gestão, Área de Concentração em Sistemas Computacionais.

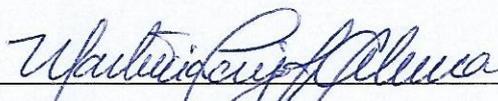
APRESENTADO EM 05/02/2019.



Aline Pires Vieira de Vasconcelos, D. Sc.  
Instituto Federal de Educação, Ciência e Tecnologia Fluminense  
(Orientadora)



Henrique Rego Monteiro da Hora, D. Sc.  
Instituto Federal de Educação, Ciência e Tecnologia Fluminense



Marlúcia Cereja de Alencar  
Instituto Federal de Educação, Ciência e Tecnologia Fluminense

## AGRADECIMENTO

Agradeço a Deus e Nossa Senhora, por estarem presentes em todos os momentos da minha vida e por me sustentarem o tempo todo durante esta caminhada.

Aos meus pais, Márcia e Toninho, meu infinito agradecimento, por nunca medirem esforços para minha educação, por todo suporte e toda base que sempre me proporcionaram: sem vocês, com certeza nunca teria conseguido alcançar este objetivo.

As minhas irmãs, Gabriela e Fernanda, pelo apoio em todos os momentos, pelas dicas, seja na hora de saber como menciono uma lei no texto como em que lugar usar uma vírgula, vocês são essenciais para minha vida e foram para esse momento.

Ao meu marido Roberto, que sempre esteve ao meu lado, ouvindo minhas reclamações e sempre me motivando a continuar, por todo o suporte, amor, carinho e companheirismo, e por sempre acreditar no meu potencial, isso tudo foi primordial para cumprimento desta etapa. Te amo!

Aos meus primos “falsiones”, Thays e Léo, por além de sempre estarem prontos a ajudar, me ofereceram casa, comida e roupa lavada aqui em Campos, sempre serei grata a vocês por todo apoio.

A minha orientadora, Aline Vasconcelos, por toda a paciência e empenho com que sempre me orientou neste trabalho. Muito obrigada por ter me corrigido quando necessário sem nunca me desmotivar, meus sinceros agradecimentos por isso.

Aos meus amigos de trabalho do IFF Guarus, que sempre estiveram dispostos a contribuírem para o desenvolvimento deste trabalho, independente da área em que atuam. Vocês são demais e fazem parte dessa conquista.

Aos meus amigos da vida e do Mestrado e a todos que direta e indiretamente fizeram parte dessa caminhada. Muito obrigada!

## RESUMO

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 busca utilizar técnicas de Mineração de Dados para extrair conhecimento das avaliações do CPC do ano de 2014 e identificar os principais critérios e resultados da avaliação dos cursos de graduação em Licenciatura, fazendo uma análise e comparação dos dados da Rede Federal com as Universidades Federais, Estaduais e Privadas 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 (*Knowledge Discovery in Databases*), 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. Espera-se que as informações extraídas a partir do presente trabalho possam ser úteis e subsidiar à gestão educacional, além de serem utilizadas para o aprimoramento dos cursos de graduação.

**Palavras-chave:** Avaliação; Instituto Federal; CPC; Mineração de Dados; Licenciatura.

## ABSTRACT

The evaluation of the quality of undergraduate courses in Brazil is done through the Preliminary Concept of Courses (CPC), which is an indicator of quality that evaluates these courses. The present seeks to use Data Mining techniques to extract knowledge of the CPC evaluations of the year 2014 and to identify the main criteria and results of the evaluation of the undergraduate courses in Bachelor's degree, making an analysis and comparison of the data from the Federal Network with Federal, State and Private Universities from all over Brazil. For this, the database was made available on the portal of the Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), and the KDD process steps were used, focusing on Data Mining to extract knowledge from the database. The results allowed the identification of the evaluation criteria with the greatest impact for the evaluation of the CPC of the Bachelor's degree courses. It is hoped that the information extracted from the present work can be useful and subsidize to the educational management, besides being used for the improvement of the undergraduate courses.

**Keywords:** Evaluation; Federal Institute; CPC; Data mining; Teacher training program.

## LISTA DE ABREVIACES

CPC	Conceito Preliminar de Curso
CONIF	Conselho Nacional das Instituies da Rede Federal
ENADE	Exame Nacional de Desempenho dos Estudantes
ENEM	Exame Nacional do Ensino Mdio
IDD	Indicador de Diferena entre Desempenho Observados e Esperados
IES	Instituio de Ensino Superior
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Ansio Teixeira
KDD	<i>Knowledge-Discovery in Databases</i>
MDE	Minerao de Dados Educacionais
OAF	Oportunidades de Ampliao da Formao Acadmica e Profissional
ODP	Organizao Didtico-Pedaggica
RT	Regime de Trabalho
SAEB	Sistema de Avaliao da Educao Bsica
TI	Tecnologia da Informao

## LISTA DE FIGURAS

Figura 1 – Etapas do KDD .....	10
Figura 2 – Exemplo de Árvore de Decisão.....	13
Figura 3 – Expansão da Rede Federal .....	15
Figura 4 – Palavras Chaves e <i>Tesauros</i> .....	19
Figura 5 – Diagrama de Venn relacionado aos trabalhos da base <i>Scopus</i> . .....	20
Figura 6 – Evolução das Publicações para os termos ABCD, ABC e ACD. ....	21
Figura 7 – Evolução das Publicações para os termos BCD e ABD.....	22
Figura 8 – Quantidade de Periódicos e/ou Conferências por área. ....	23
Figura 9 – Quantidade de Periódicos e/ou Conferências na área ABD. ....	24
Figura 10 – Método para calcular o <i>CitScore</i> . ....	24
Figura 11 – Palavras Chaves e <i>Tesauros</i> .....	28
Figura 12 – Diagrama de Venn relacionado aos trabalhos da base <i>Scielo</i> . ....	29
Figura 13 – Evolução das Publicações para os termos ABC, AC e BC. ....	30
Figura 14 – Evolução das Publicações para os termos AB.....	30
Figura 15 – Quantidade de Periódicos e/ou Conferências por área. ....	31
Figura 16 – Metodologia do Trabalho. ....	37
Figura 17 – Composição da Base de Dados do estudo proposto.....	41
Figura 18 – Tela Inicial do <i>Software Weka</i> .....	42
Figura 19 – Árvore de Decisão da Rede Federal. ....	44
Figura 20 – Árvore de Decisão Universidades Estaduais (a). ....	46
Figura 21 – Árvore de Decisão Universidades Estaduais (b). ....	47
Figura 22 – Árvore de Decisão Universidades Estaduais (c). ....	48
Figura 23 – Árvore de Decisão Universidades Estaduais (d). ....	49
Figura 24 – Árvore de Decisão Universidades Estaduais (e). ....	49
Figura 25 – Árvore de Decisão Universidades Federais (a). ....	51
Figura 26 – Árvore de Decisão Universidades Federais (b). ....	51
Figura 27 – Árvore de Decisão Universidades Federais (c). ....	52
Figura 28 – Árvore de Decisão Universidades Federais (d). ....	52
Figura 29 – Árvore de Decisão Universidades Privadas (a). ....	54
Figura 30 – Árvore de Decisão Universidades Privadas (b). ....	54
Figura 31 – Árvore de Decisão Universidades Privadas (c). ....	55

Figura 32 – Árvore de Decisão Universidades Privadas (d). .....	56
Figura 33 – Árvore de Decisão Universidades Estaduais (f).....	72
Figura 34 – Árvore de Decisão Universidades Federais (e). .....	72
Figura 35 – Árvore de Decisão Universidades Federais (f). .....	73
Figura 36 – Árvore de Decisão Universidades Federais (g).....	73
Figura 37 – Árvore de Decisão Universidades Federais (h).....	74
Figura 38 – Árvore de Decisão Universidades Federais (i).....	74
Figura 39 – Árvore de Decisão Universidades Federais (j).....	75
Figura 40 – Árvore de Decisão Universidades Privadas (e).....	75
Figura 41 – Árvore de Decisão Universidades Privadas (f). .....	76
Figura 42 – Árvore de Decisão Universidades Privadas (g). .....	76
Figura 43 – Árvore de Decisão Universidades Privadas (h). .....	77
Figura 44 – Árvore de Decisão Universidades Privadas (i). .....	77
Figura 45 – Árvore de Decisão Universidades Privadas (j). .....	78
Figura 46 – Árvore de Decisão Universidades Privadas (k). .....	78
Figura 47 – Árvore de Decisão Universidades Privadas (l). .....	79
Figura 48 – Árvore de Decisão Universidades Privadas (m). .....	79

## LISTA DE QUADROS

Quadro 1 – Descrição da Pesquisa realizada no <i>Scopus</i> .....	20
Quadro 2 – Descrição da Pesquisa realizada no <i>Scielo</i> .....	28
Quadro 3 – Dados retirados da base de dados do INEP.....	38
Quadro 4 – Critérios considerados nas Avaliações Nota 2 e Nota 5 .....	58
Quadro 5 – Dados do INEP sobre o curso da Rede Federal Nota 5. ....	59
Quadro 6 – Síntese dos Critérios relacionados às IES.....	60

## LISTA DE TABELAS

Tabela 1 – Critérios para Cálculo do CPC.....	7
Tabela 2 – Critérios de Avaliação do CPC.....	8
Tabela 3 – Modalidade de Ensino/Peso dos Cursos .....	17
Tabela 4 – Quantidade de Artigos por Periódicos/Anais de Conferência.....	22
Tabela 5 – Quantidade de Artigos por Periódicos das áreas analisadas. ....	23
Tabela 6 – Análise dos Periódicos/Anais de Conferência .....	25
Tabela 7 – Artigos mais citados.....	26
Tabela 8 – Quantidade de Artigos por Periódicos das áreas analisadas. ....	31
Tabela 9 – Artigos mais citados.....	32
Tabela 10 – Quantidade de vezes os critérios estão presentes nas árvores.....	57
Tabela 11 – Critérios de maior impacto presentes nas árvores.....	58

# SUMÁRIO

1. INTRODUÇÃO.....	1
1.1. Contextualização.....	1
1.2. Objetivos.....	3
1.2.1. Objetivo Geral.....	3
1.2.2. Objetivos Específicos.....	3
1.3. Justificativa.....	4
1.4. Estrutura do Trabalho.....	5
2. REFERENCIAL TEÓRICO.....	6
2.1. Conceito Preliminar de Cursos.....	6
2.1.1. Desempenho dos Estudantes.....	8
2.1.2. Corpo Docente.....	9
2.1.3. Percepção Discente sobre as Condições do Processo Formativo.....	9
2.2. Knowledge Discovery in Databases (KDD).....	9
2.3. Mineração de Dados.....	11
2.3.1. Árvore de Decisão.....	12
2.3.2. Algoritmo C4.5.....	13
2.3.3. Mineração de Dados Educacionais (MDE).....	14
2.4. Rede Federal.....	15
3. ESTUDO BIBLIOMÉTRICO.....	18
3.1. Pesquisa na Base <i>Scopus</i> .....	18
3.1.1. Identificação da Evolução Cronológica das Áreas.....	21
3.1.2. Identificação dos Periódicos com maior Número de Publicações.....	22
3.1.3. Identificação dos Trabalhos mais Citados na Área.....	26
3.2. Pesquisa na Base <i>Scielo</i> .....	27
3.2.1. Identificação da Evolução Cronológica das Áreas.....	29

3.2.2.	Identificação dos Periódicos com maior Número de Publicações .....	30
3.2.3.	Identificação dos Trabalhos mais citados na Área .....	32
3.3.	Trabalhos Relacionados .....	33
3.3.1.	Trabalhos que utilizam a Mineração de Dados na área Educacional .....	33
3.3.2.	Trabalhos que avaliam a Qualidade do Ensino .....	35
3.3.3.	Contribuição do Presente Trabalho .....	36
4.	MATERIAIS E MÉTODOS.....	37
4.1.1.	Coleta de Dados .....	37
4.1.2.	Tratamento dos Dados.....	40
4.1.3.	Mineração dos Dados .....	42
4.1.4.	Análise e Discussão dos Resultados.....	43
5.	RESULTADOS .....	44
5.1.	Avaliação da Rede Federal .....	44
5.2.	Avaliação das Universidades Estaduais.....	46
5.3.	Avaliação das Universidades Federais.....	50
5.4.	Avaliação das Universidades Privadas .....	53
5.5.	Discussão dos Resultados .....	57
6.	CONCLUSÕES .....	62
6.1.	Contribuições .....	63
6.2.	Trabalhos Futuros .....	63
	REFERÊNCIAS BIBLIOGRÁFICAS .....	65
	APÊNDICE I – MATRIZES DE CONFUSÃO .....	71
	APÊNDICE II – ÁRVORES DE DECISÃO .....	72

# **1. INTRODUÇÃO**

## **1.1. Contextualização**

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 se destacam 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, oferecendo cursos de qualificação, ensino médio integrado, cursos superiores de tecnologia e Licenciaturas.

A Lei de criação dos Institutos Federais, 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 nesta 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 nos Institutos Federais, 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 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 (MEC, 2015a). 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 a avaliação do CPC das IES está disponível no *site* do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP)<sup>1</sup>.

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 *Educational data mining (EDM)*, conceito apresentado nos trabalhos de Baker *et al.* (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 (2015) afirma que grande parte dos dados pertinentes à educação estão relacionados, direta e indiretamente, a 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 através de uma combinação de diferentes técnicas de mineração

---

<sup>1</sup> <http://inep.gov.br/conceito-preliminar-de-curso-cpc->

em associação com esses dados, visando otimizar a prática e a gestão educacional. Têm-se como exemplo de dados que podem ser extraídos para contribuir com o conhecimento na gestão educacional: perfil de aluno, nota dos alunos do Enade, número de mestres e doutores, entre outros.

Em consonância com os aspectos relacionados a importância da avaliação da qualidade dos cursos da IES concomitantemente com a relevância dos cursos de Licenciatura, principalmente nos Institutos Federais, 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.

## **1.2. Objetivos**

### **1.2.1. Objetivo Geral**

Nesse contexto, este trabalho tem o objetivo de extrair conhecimento da avaliação de qualidade dos cursos de Licenciatura ofertados na Rede Federal de todo o Brasil, com foco nos Institutos Federais, a partir do CPC do ano de 2014 (avaliação mais atual referente ao Ano II), comparando com as Universidades Federais, Estaduais e Privadas, utilizando técnicas de mineração de dados.

### **1.2.2. Objetivos Específicos**

Para atingir o objetivo geral, foram estabelecidos os seguintes objetivos específicos:

- Levantar Dados relacionados à Avaliação de Qualidade das IES por meio do CPC de 2014 na base do INEP;
- Possibilitar a identificação de perfis específicos para os cursos de Licenciatura das IES;
- Verificar os perfis relacionados com as avaliações mais significativas: Nota 2 (próximo ao descredenciamento do curso) e Nota 5 (curso de excelência);
- Traçar o perfil da Rede Federal em comparação com as outras Universidades;
- Analisar os conhecimentos obtidos com a aplicação das tarefas de mineração de dados;

- Gerar subsídios para auxiliar a gestão dos cursos de graduação para a melhoria do seu desempenho.

### **1.3. Justificativa**

É incontestável a importância de se aferir a qualidade dos cursos superiores no Brasil, afirma Sousa (2017), pois a educação superior é um dos domínios de grande importância no processo de desenvolvimento do país, em seus mais variados setores e dimensões.

Na avaliação da qualidade dos cursos das IES, são disponibilizados grandes repositórios de dados que podem propiciar informações valiosas para os profissionais da área de educação. Cretton (2016) salienta a possibilidade de serem encontrados padrões relevantes e novos conhecimentos, que podem ser aplicados no aprimoramento das IES do Brasil.

O curso de Licenciatura é o espaço legal e institucional para a formação do professor da educação básica no Brasil, e tem como objetivo capacitar o futuro docente em termos de conhecimentos, habilidades e atitudes necessárias para o ensino, apontam Oliveira e Oliveira (2016). Mediante o exposto, é notável que os cursos de Licenciatura têm um papel fundamental na sociedade como um todo e isso é corroborado dentro dos Institutos Federais e expresso em forma da Lei nº 11.892, a qual garante o mínimo de 20% (vinte por cento) de suas vagas para atender aos cursos de Licenciatura (BRASIL, 2008).

Além disso, a Matriz do Conselho Nacional das Instituições da Rede Federal de Educação Profissional, Científica e Tecnológica (CONIF) que trata de uma Matriz Orçamentária de Custeio e Investimento para as Instituições Federais da Rede de Ensino Profissional e Tecnológico, definiu na edição de 2018 o aluno de Licenciatura com um peso considerável (2.5), frente aos alunos de outras modalidades de ensino (FORPLAN, 2018).

Por fim, a partir da pesquisa bibliográfica, observou-se diversos trabalhos relacionados inseridos nesse contexto. Os trabalhos de Ferreira (2015), Cretton e Gomes (2016) e Fonseca e Namem (2016) avaliam o perfil dos alunos e cursos utilizando a técnica de mineração de dados com base no Censo Escolar da Educação Básica, no Sistema de Avaliação da Educação Básica (Saeb) e no Enade, respectivamente. Já os trabalhos de Bittencourt *et al.* (2010), Lacerda *et al.* (2015), Gonçalves *et al.* (2017) abarcam a avaliação realizada através do CPC nas Instituições de Ensino Superior. Porém, nenhum dos trabalhos encontrados faz uma avaliação com base no CPC utilizando a técnica de mineração de dados como o presente trabalho.

#### **1.4. Estrutura do Trabalho**

Partindo desta Introdução, o presente trabalho está estruturado da seguinte forma: na Seção 2 é apresentada uma contextualização dos temas base deste trabalho, sendo eles: Avaliação de Qualidade dos cursos de Graduação utilizando o CPC, Descoberta em bases de conhecimento (KDD) e Mineração de Dados; em seguida, na Seção 3 é apresentado um estudo bibliográfico acerca do tema da presente pesquisa, e são levantados os trabalhos relacionados existentes na literatura; na Seção 4 é definida e explicada a metodologia utilizada no trabalho; na Seção 5 são explicitados os resultados gerados pelo trabalho e é apresentada uma discussão dos resultados obtidos; por fim, na Seção 6 são apresentadas as conclusões e os trabalhos futuros propostos neste trabalho.

## 2. REFERENCIAL TEÓRICO

### 2.1. Conceito Preliminar de Cursos

O Conceito Preliminar de Curso (CPC) é um indicador de qualidade que tem como sua principal função a avaliação dos cursos de graduação de todo o Brasil. Seu cálculo é baseado na avaliação de desempenho dos estudantes, no valor agregado pelo processo formativo e em insumos referentes às condições de oferta (corpo docente, infraestrutura e recursos didáticos-pedagógicos). Sua divulgação ocorre sempre no ano seguinte ao da realização do Exame Nacional de Desempenho de Estudantes (Enade) (MEC, 2015a).

O Enade é um exame que tem como objetivo avaliar o desempenho dos estudantes com relação aos conteúdos programáticos previstos nas diretrizes curriculares dos cursos de graduação, o desenvolvimento de competências e habilidades necessárias ao aprofundamento da formação geral e profissional, como também, o nível de atualização dos estudantes com relação a realidade brasileira e mundial (MEC, 2015b).

De acordo com o INEP (2017), a composição e o cálculo do CPC são realizados através de oito componentes, os quais são agrupados em três dimensões, objetivando a avaliação da qualidade dos cursos de graduação:

1. **Desempenho dos Estudantes:** medido a partir das notas dos estudantes concluintes no Enade e dos valores do Indicador da Diferença entre os Desempenhos Observado e Esperado (IDD);
2. **Corpo Docente:** baseado em informações obtidas a partir do Censo da Educação Superior, que é um instrumento de pesquisa que reúne informações sobre as IES, seus cursos de graduação presencial ou a distância, cursos sequenciais, vagas oferecidas, inscrições, matrículas, ingressantes e concluintes e informações sobre docentes nas diferentes formas de organização acadêmica e categoria administrativa vinculados aos cursos avaliados (MEC, 2015c);
3. **Percepção Discente sobre as Condições do Processo de Formação:** é obtida por intermédio do levantamento de informações relativas à organização didático-pedagógica, à infraestrutura e instalações físicas e as oportunidades de ampliação da formação acadêmica e profissional, com base nas respostas obtidas com a aplicação do questionário Socioeconômico do Enade. Este questionário tem como objetivo subsidiar a construção do perfil socioeconômico do estudante e obter uma apreciação quanto ao seu processo formativo, o concluinte habilitado a participar do Enade tem acesso a informação sobre seu local de prova somente após o preenchimento do

Questionário através do acesso ao Sistema de Cadastro do Estudante<sup>2</sup> em um período estabelecido pelo INEP.

As três dimensões supracitadas abrangem itens específicos para o cálculo do CPC. 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 a ser avaliado.

**Tabela 1** – Critérios para Cálculo do CPC.

<b>Dimensão</b>	<b>Componentes</b>	<b>Peso</b>	
Desempenho dos Estudantes	Nota dos Concluintes no Enade (NC)	20%	55%
Valor Agregado	Nota do Indicador de Diferença entre os Desempenhos Observados e Esperados (NIDD)	35%	
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: (INEP, 2015)

A composição e a forma de cálculo do CPC de 2014 (base de dados utilizada no presente trabalho), com os respectivos pesos dos componentes, são apresentadas na Fórmula 1. Destaca-se que todos os componentes têm suas notas padronizadas para uma faixa de 0 a 5, antes da utilização da Fórmula 1. A padronização ocorre de acordo com a Nota Técnica Daes/Inep nº 58/2015 (INEP, 2015).

$$CPC = 0,2NC + 0,35NIDD + 0,075NM + 0,15ND + 0,075NR + 0,075NF + 0,05NF + 0,025NA \quad (1)$$

<sup>2</sup> <http://Enade.inep.gov.br/Enade/#!/index>

**Tabela 2** – Critérios de Avaliação do CPC.

CPC (Valor Contínuo)	CPC (Faixa)
$0 \leq \text{CPC} < 0,945$	1
$0,945 \leq \text{CPC} < 1,945$	2
$1,945 \leq \text{CPC} < 2,945$	3
$2,945 \leq \text{CPC} < 3,945$	4
$3,945 \leq \text{CPC} \leq 5$	5

Fonte: (INEP, 2015)

A nota final do CPC é expressa em uma escala de 1 a 5 (baseado na conversão da nota contínua para a faixa, apresentado na Tabela 2) e trata-se de um indicador prévio dos cursos de graduação do país. Para consolidar esses valores e assegurar que os cursos representem efetivamente o que se espera em termos de qualidade e excelência, comissões de avaliadores realizam visitas *in loco* para corroborar ou alterar o conceito obtido preliminarmente. Os Cursos que obtiverem notas entre 1 e 2 automaticamente estão sujeitos a visita dos avaliadores do Inep para averiguação *in loco* das condições de ensino ofertadas. Os cursos com conceito igual ou superior a 3 podem optar por não receber a visita dos avaliadores e, assim, transformar o CPC em conceito permanente (MEC, 2015a).

Cada critério exposto na Tabela 1, é apresentado com maior detalhe nas subseções que seguem.

#### 2.1.1. Desempenho dos Estudantes

- **Nota dos Concluintes no Enade (ENADE):** a Nota dos Concluintes no Enade de cada curso de graduação referente ao ano do cálculo do CPC;
- **Nota do Indicador da Diferença entre os Desempenhos Observado e Esperado (IDD):** busca aferir aquilo que diz respeito especificamente ao valor agregado pelo curso no desenvolvimento dos estudantes concluintes. A partir disso, a determinação do desempenho dos concluintes de cursos de graduação estão relacionados, principalmente, a dois fatores: características de desenvolvimento do estudante concluinte ao ingressar na Educação Superior e qualidade das condições do processo formativo oferecido pelo curso. Os cálculos realizados são baseados no desempenho observado do estudante concluinte em comparação a estimativa da parte do desempenho do estudante concluinte, decorrente de suas características quando

ingressante no curso. Esses dados referentes a características dos ingressantes são obtidos através dos dados do Exame Nacional do Ensino Médio (ENEM), que a partir do exame de 2009, tornou o preenchimento dessas informações obrigatórias para inscrição no exame.

### 2.1.2. Corpo Docente

- **Nota de Proporção de Mestres:** o cálculo realizado é baseado na razão entre o número de docentes do curso de graduação com titulação igual ou superior a mestre e o número total de docentes vinculados ao curso de graduação avaliado;
- **Nota de Proporção de Doutores:** o cálculo realizado é baseado na razão entre o número de docentes do curso de graduação com título de doutor e o número total de docentes vinculados ao curso de graduação avaliado;
- **Nota de Regime de Trabalho:** o cálculo realizado é baseado na razão entre o número de docentes do curso de graduação com regime de trabalho parcial ou integral e o número total de docentes vinculados ao curso de graduação avaliado.

### 2.1.3. Percepção Discente sobre as Condições do Processo Formativo

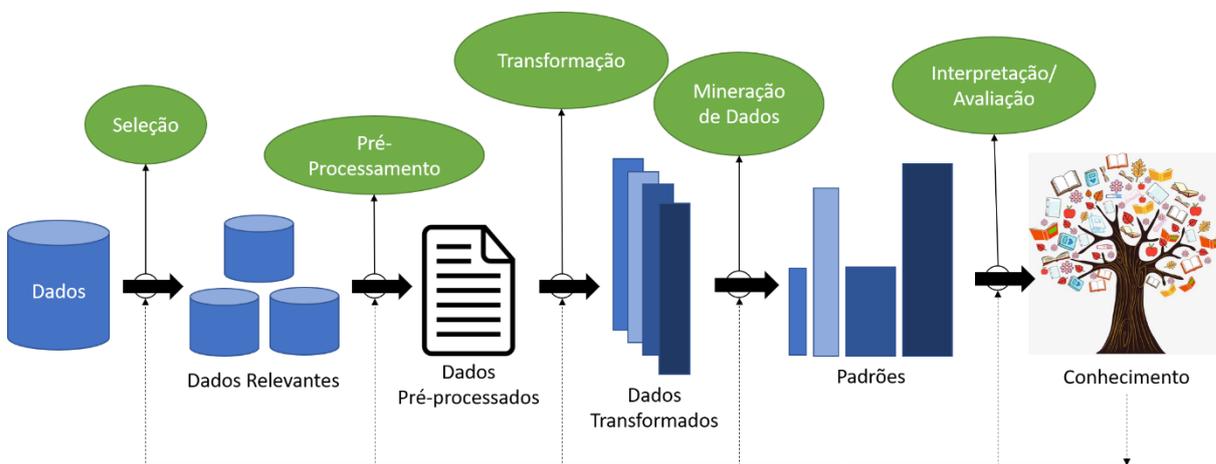
- **Nota referente à Organização Didático-Pedagógica:** Esta nota é baseada nas respostas de itens relativos à Organização Didático-Pedagógica, que estão presentes no Questionário Socioeconômico para cada estudante do curso de graduação;
- **Nota referente à Infraestrutura e Instalações Físicas:** Esta nota é baseada nas respostas de itens relativos à Infraestrutura e Instalações Físicas, que estão presentes no Questionário Socioeconômico para cada estudante do curso de graduação;
- **Nota referente às oportunidades de ampliação da formação acadêmica e profissional:** Esta nota é baseada nas respostas de itens relativos a oportunidades de ampliação da formação acadêmica e profissional, que estão presentes no Questionário Socioeconômico para cada estudante do curso de graduação, esse critério é baseado nas oportunidades dos estudantes participarem de programas, projetos ou atividades de extensão universitária, iniciação científica, entre outros.

## 2.2. Knowledge Discovery in Databases (KDD)

É importante utilizar uma metodologia e métodos que auxiliem na concepção de informações de um grande volume de dados armazenados em diversas áreas, inclusive no

âmbito educacional, afirma Cordeiro *et al.* (2017). Os autores apontam a mineração de dados como uma das principais técnicas utilizadas para este fim, mas enfatizam que é necessária a realização do tratamento prévio dos dados, utilizando o processo de descoberta de conhecimento em base de dados, que tem a mineração de dados como uma de suas etapas principais.

O termo “descoberta de conhecimento em base de dados” ficou conhecido como KDD devido a abreviação da tradução para o inglês *Knowledge Discovery in Databases*. Foi definido, segundo Fayyad *et al.* (1996), como um processo não trivial, que possui várias etapas de identificação de diferentes padrões ainda não conhecidos, que sejam potencialmente úteis, corretos e compreensíveis. As etapas e a construção deste conhecimento podem ser observadas na Figura 1.



**Figura 1** – Etapas do KDD

Fonte: Elaboração própria; adaptado de Fayyad *et al.* (1996)

Como descrito por Fayyad *et al.* (1996) e corroborado por Colak *et al.* (2015), o processo KDD possui cinco etapas, que estão apresentadas na Figura 1:

1. **Seleção de Dados:** Identificar e selecionar os dados que irão auxiliar na extração do conhecimento desejado por meio da análise dos bancos de dados existentes.
2. **Pré-processamento dos Dados:** Remover redundâncias, valores discrepantes e dados inconsistentes. As incorreções devem ser resolvidas utilizando técnicas como remoção, codificação ou padronização, com o objetivo de garantir a qualidade relacionada à integridade dos dados.

3. **Transformação e Redução de Dados:** Transformar os dados em estruturas padronizadas que sejam compatíveis com o algoritmo e ferramenta utilizada na etapa de mineração de dados.
4. **Mineração de Dados:** Processo em que os dados são analisados e aplica-se métodos a fim de extrair conhecimento da base analisada.
5. **Interpretação/Avaliação dos Padrões:** Identificar, interpretar e avaliar os padrões gerados a partir da etapa anterior, de acordo com o critério definido pelo usuário. A partir disso, tornar as informações obtidas disponíveis aos interessados de uma forma compreensível.

### 2.3. Mineração de Dados

A mineração de dados compõe a quarta etapa do processo de KDD, que ocorre após a seleção, pré-processamento e transformação dos dados, ou seja, com os dados já selecionados e padronizados, de acordo com o algoritmo e ferramenta escolhida, a etapa de mineração de dados pode ser iniciada. Esta etapa é responsável pela extração dos padrões e conhecimentos, independentemente da quantidade de dados e de sua complexidade, afirma Cretton (2016).

Segundo Zaki e Meira Jr (2014), a Mineração de Dados (DM – *Data Mining*) é um processo capaz de descobrir novos padrões, além de modelos descritivos, compreensíveis e preditivos de dados massivos. Destacando-se assim como uma das principais ferramentas na função de análise de um vasto número de registros, utilizando algoritmos que permitem a extração de conhecimento desses dados. Assim, os autores salientam que é um campo interdisciplinar que relaciona conceitos de áreas aliadas, por exemplo: sistemas de banco de dados, estatísticas, aprendizado de máquina e reconhecimento de padrões.

A DM é definida por Wagner *et al.* (2016) como uma técnica na qual é extraído conhecimento de uma grande base de dados e auxilia na tomada de decisão. Os autores enfatizam que esta é uma técnica aplicada a diferentes áreas, como em pesquisas, empresas, indústrias, etc: objetiva a análise de tendências e conseqüentemente a melhoria dos processos.

Como caracterizado por MacCue (2007), no processo de mineração de dados, há diversas técnicas que devem ser testadas e ajustadas de acordo com a necessidade e a base de dados utilizada. As técnicas mais utilizadas na mineração de dados são: Classificação, Regressão, Clusterização e Associação, cada uma utilizando métodos específicos. Estes métodos são classificados de acordo com as tarefas que realizam, corrobora Han *et al.* (2012), que descrevem e definem cada uma das técnicas e métodos em seu trabalho.

**Técnica de Classificação:** É utilizada com o objetivo de prever valores de variáveis do tipo decisivas, em que o primeiro passo é a escolha de um atributo central a partir da base selecionada, também conceituado como desfecho. Um exemplo utilizado por Han *et al.* (2012) é a criação de um modelo/padrão que classifica os clientes de um banco como especiais ou de risco. Nesse cenário, por exemplo, o modelo é criado para classificar a qual categoria um certo cliente do banco pertence: especial ou de risco (possíveis desfechos).

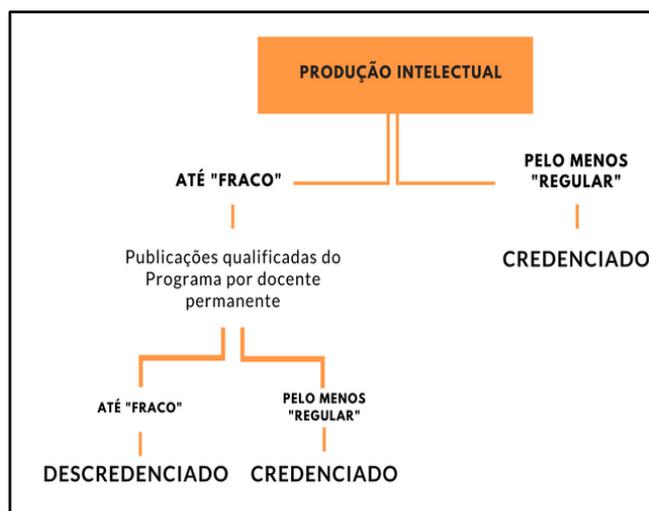
Após a criação do modelo de classificação, é necessário estimar a precisão do classificador. Há diversas formas de avaliar a precisão do classificador, uma das mais utilizadas é a matriz de confusão, que fornece uma visão geral do resultado da classificação e através dela é possível obter a porcentagem dos atributos classificados de forma correta e incorreta, também conhecido como taxa de acerto (CAZZOLATO e VIEIRA, 2011).

### 2.3.1. Árvore de Decisão

Segundo Mantas e Abellán (2014), a árvore de decisão é um método de classificação de dados com estrutura simples, fácil interpretação para seus resultados e de grande utilidade para as mais diversas problemáticas de diferentes áreas do conhecimento (HARRISON *et al.* 2018; SANKARI E MANIMEGALAI, 2017).

O método de classificação por Árvore de Decisão é explicado por Cretton e Gomes (2016): o atributo de maior significância encontra-se na raiz da árvore e a partir dele, geram-se ramificações, que representam a relevância desta ligação. Assim, os autores afirmam que esta estrutura tem a capacidade de representar, de forma intuitiva, padrões simples e complexos, que possibilitam a obtenção de informações importantes.

Um exemplo da árvore de decisão pode ser observado no Figura 2, em que o desfecho é (Credenciado ou Não Credenciado), ou seja, toda a árvore de decisão irá convergir apenas para dois resultados, o do Credenciamento de um curso ou Descredenciamento do mesmo. Além disso, o atributo de maior significância é Produção Intelectual e a ramificação gerada é referente ao atributo de Publicações qualificadas do Programa por docente permanente.



**Figura 2** – Exemplo de Árvore de Decisão

Fonte: Gualhano *et. al.* (2018).

O presente trabalho utiliza a técnica de Classificação e o método de Árvore de Decisão.

### 2.3.2. Algoritmo C4.5

O ID3 (*Iterative Dichotomiser*) é um algoritmo desenvolvido por Quilan (1987) para geração de árvore de decisão e, posteriormente, otimizado para uma versão chamada de C4.5, que possui sua implementação de código aberto em Java, conhecida como Algoritmo J48. Este algoritmo tem como principal função obter modelos de árvores de decisão partindo do topo para base, aplicando o método de divisão e conquista, assim, ele funciona dividindo o problema em vários problemas menores, o que conseqüentemente acarreta em menor complexidade e em aumento da eficiência da árvore de decisão (CRETTON, 2016).

Cordeiro *et al.* (2017) analisaram uma grande gama de trabalhos relacionados a Mineração de Dados Educacionais e pontuou o método de árvore de decisão como um dos mais utilizados na literatura. De encontro a esta afirmação, vários trabalhos na literatura relacionado a mineração de dados utilizam a árvore de decisão, apontando o algoritmo J48 como o de maior eficiência, quando comparado a outros.

O trabalho de Baby (2018) utilizou para a previsão do desempenho pedagógico em instituições de ensino superior três algoritmos de classificação: Árvores de decisão (C4.5), *Sequential Minimal Optimization* (SMO) e redes neurais artificiais, também conhecida como Perceptron de Múltiplas Camadas (MLP). E o algoritmo J48 obteve uma precisão de 94,37% comparado aos outros. Corroborando isso, no trabalho de Ramaphosa *et al.* (2018) é apresentado a comparação de quatro classificadores (Naïve Bayes, BayesNet, JRip e J48) e

apresentado o algoritmo de classificação de melhor desempenho entre eles. Com base nos resultados obtidos neste trabalho, os autores afirmam que o algoritmo J48 superou os outros algoritmos com 99,13% de precisão de previsão.

Além disso, é possível utilizar parâmetros para treinamento da rede, como por exemplo o *cross-validation*, que é explicado por (Cordeiro, 2017).

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. (pág 45).

### 2.3.3. Mineração de Dados Educacionais (MDE)

Uma das áreas em que a Mineração de Dados é aplicada é a área Educacional, também conhecida como Mineração de Dados Educacionais. De acordo com Cordeiro *et al.* (2017), a área da MDE tem uma importante característica e tem se tornado uma área atrativa, pois consegue converter dados sem reais significados em informações de grande importância na pesquisa e prática educacional.

De acordo com Kamal e Ahuja (2019), a mineração de dados educacionais (MDE) é utilizada para converter os dados disponíveis em ambientes educacionais em informações úteis. Auxiliando na construção de soluções para diferentes questões que surgem em contextos educacionais, como previsão de desempenho de estudantes acadêmicos, criação de novos cursos, método ou modo de ensino.

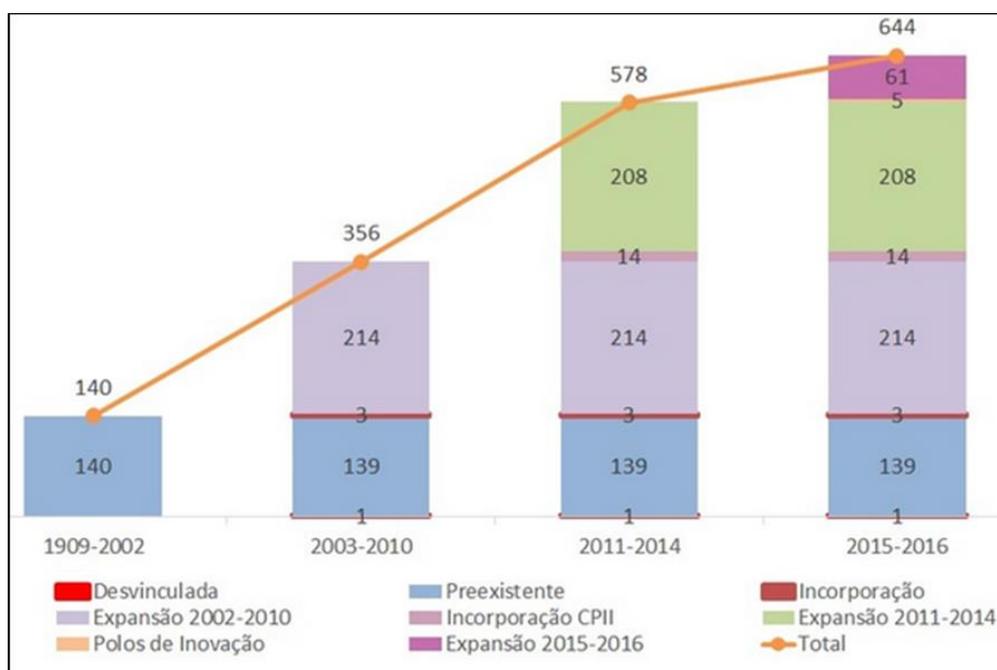
A MDE utiliza ciências como o aprendizado de máquina e a estatística para auxiliar na construção de conhecimentos, além de aplicar algoritmos de mineração de dados a diferentes tipos de dados relacionados ao ensino e assim, estabelecer bases para um processo de avaliação/aprendizagem mais eficaz, confirmam Romero e Ventura (2010).

A diferença entre a Mineração de Dados e a Mineração de Dados Educacionais é apontada tanto por Romero e Ventura (2010) quanto por Rigo *et al.* (2012). Os autores afirmam que a principal diferença está na natureza dos dados, pois os dados utilizados na mineração de dados no contexto educacional são mais diversificados, precisando assim de ajustes para sua mineração. A MDE trabalha, por exemplo, com diferentes grupos que analisam informações educacionais de diferentes ângulos, de acordo com seus objetivos. A partir disso, o conhecimento descoberto por algoritmos de MDE pode ser utilizado tanto para auxiliar os professores a compreender os processos de aprendizado de seus alunos e refletir sobre os métodos de ensino, como também para apoiar as reflexões do próprio aluno sobre a

sua situação escolar. Assim, na MDE utilizam-se alguns métodos, por exemplo, correlação que não fazem parte da mineração de dados.

## 2.4. Rede Federal

Por meio da Lei nº 11.892, de 29 de dezembro de 2008, ficou instituída a Rede Federal de Educação Profissional, Científica e Tecnológica por meio da criação dos Institutos Federais de Educação, Ciência e Tecnologia (IFs). Nesta Lei, no artigo 7º, que trata dos objetivos dos IFs, fica estabelecido como finalidade a oferta da educação profissional técnica de nível médio, cursos de formação inicial e continuada, pesquisa e extensão e no que se refere ao nível de educação superior, ficou constituído que devem ser oferecidos tanto cursos superiores relacionados à tecnologia e engenharia, quanto cursos de Licenciatura voltados para a formação de professores para a educação básica e profissional, além de cursos de pós-graduação (BRASIL, 2008).



**Figura 3** – Expansão da Rede Federal

Fonte: (BRASIL, 2016)

De 1909 a 2002, foram construídas 140 escolas técnicas no país. Entre 2003 e 2016, o Ministério da Educação concretizou a construção de mais de 500 novas unidades referentes ao plano de expansão da educação profissional.

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, totalizando 644 campi em funcionamento.

Como pode ser observado na Figura 3, os Institutos Federais estão em constante crescimento e atendendo a uma parcela significativa da população brasileira, afirma Moura (2018). São 38 Institutos Federais presentes em todos os estados, oferecendo cursos de qualificação, ensino médio integrado, cursos superiores de tecnologia e Licenciaturas.

O artigo 8º da Lei nº 11.892 dispõe que no desenvolvimento da sua ação acadêmica, o Instituto Federal, em cada exercício, deverá garantir o mínimo de 50% (cinquenta por cento) de suas vagas para atender a educação profissional técnica de nível médio, prioritariamente, na forma de cursos integrados para os concluintes do ensino fundamental e para o público da educação de jovens e adultos e o mínimo de 20% (vinte por cento) de suas vagas para atender aos cursos de Licenciatura, bem como programas especiais de formação pedagógica, com vistas na formação de professores para a educação básica, sobretudo nas áreas de ciências e matemática e para a educação profissional (BRASIL, 2008).

A Lei nº 9394, em seu artigo 62º aponta que a formação de docentes para atuar na educação básica far-se-á em nível superior, em curso de licenciatura plena, admitida, como formação mínima para o exercício do magistério na educação infantil e nos cinco primeiros anos do ensino fundamental, a oferecida em nível médio, na modalidade normal (BRASIL, 1996).

Os cursos de Licenciatura têm suas diretrizes definidas pela Resolução CNE/CP nº 02/2015, a qual apresenta princípios, fundamentos, dinâmica formativa e procedimentos a serem observados nas políticas, na gestão e nos programas e cursos de formação, bem como nos processos de avaliação e de regulação das instituições de educação que as ofertam (MEC, 2015d).

Gatti (2015) salienta a necessidade de ponderar o conceito da qualidade da educação no que se refere às Licenciaturas, pois o contexto social e as condições formativas devem ser considerados como proposições na formação de professores. Além disso, Gatti (2016) ressalta a importância da discussão da formação, das condições de trabalho e carreira dos professores, visto que é o docente que exercerá o papel central nos processos educativos formais, deste modo deve-se compreender e discutir a qualidade educacional de um país, ou de uma região.

A Matriz do Conselho Nacional das Instituições da Rede Federal de Educação Profissional, Científica e Tecnológica (CONIF) é uma matriz Orçamentária de Custeio e Investimento para as Instituições Federais da Rede de Ensino Profissional e Tecnológico que é divulgada anualmente. Ela é utilizada para distribuição dos recursos destinados aos Institutos Federais de Educação, Ciência e Tecnologia, por parte da Secretaria de Educação Profissional e Tecnológica (SETEC), para o exercício do ano vigente (FORPLAN, 2018).

A Matriz do CONIF analisa um conjunto de dados acadêmicos, que são informados pelos institutos através do Sistema Nacional de Informações da Educação Profissional e Tecnológica (SISTEC). E a partir desses dados e de outros parâmetros calculam variáveis, por exemplo, o peso dos cursos de cada modalidade, que trabalha com variação de 0,5 pontos entre eles. Menor peso igual a 1,0 e maior peso igual a 2,5. A Tabela 3 apresenta a modalidade de ensino e o peso referente a ela, divulgado na Matriz CONIF de 2018.

**Tabela 3 – Modalidade de Ensino/Peso dos Cursos**

<b>Modalidade</b>	<b>Peso</b>
Curso de Formação Inicial Continuada (FIC)	1,0
Ensino Básico	1,0
Ensino Fundamental I	1,0
Ensino Fundamental II	1,5
Ensino Médio	1,5
Cursos Técnicos	1,5
Cursos Educação de Jovens e Adultos (Proeja)	2,5
Tecnologia e Bacharelado	1,0 – 2,5
Licenciaturas	2,5
Pós-Graduação <i>Lato Sensu</i>	1,0 – 2,5
Pós-Graduação <i>Stricto Sensu</i>	2,5

Fonte: (FORPLAN, 2018).

A partir disso, fica evidente a importância dos cursos de Licenciatura no cenário dos Institutos Federais, importância observada na porcentagem de vagas destinadas aos cursos de Licenciatura pela Lei nº 11.892, que tem uma proporção maior comparada a outras modalidades individualmente, além do valor do peso do curso apresentado na Matriz CONIF do ano de 2018, em que se tem a totalidade dos cursos de Licenciatura com o peso fixo de 2,5 (maior peso possível).

### 3. ESTUDO BIBLIOMÉTRICO

Conforme Chueke e Amatucci (2015), os estudos bibliométricos auxiliam na sistematização de pesquisas realizadas no mesmo âmbito do saber e auxiliam na busca por problemas a serem investigados em pesquisa futuras, o que corrobora o fato de que o conhecimento científico é desenvolvido de forma gradual.

Com a finalidade de identificar os trabalhos na área de avaliação educacional com foco na Avaliação de cursos de Graduação do Ensino Superior, assim como, apresentar uma visão da produção científica nacional e internacional dentro deste contexto, foi realizado um estudo bibliométrico.

Para o estudo bibliométrico referente a esta pesquisa é utilizado como base a metodologia proposta por Chueke e Amatucci (2015). São realizados os 5 passos descritos a seguir:

1. Elaboração do Protocolo de Pesquisa;
2. Identificação dos estudos mais relevantes no campo;
3. Avaliação da qualidade dos Estudos levantados;
4. Sintetização dos Dados Coletados;
5. Análise dos Resultados Obtidos.

Seguindo essa metodologia, foi elaborado um estudo bibliométrico na base *Scopus* em um contexto internacional, com objetivo de apresentar a visão da produção científica através de trabalhos relacionados a Avaliação de Cursos de Graduação e a utilização de técnicas de mineração de dados. Além disso, com o intuito de examinar a produção no cenário nacional, foi realizada uma pesquisa na base de dados *Scielo*, utilizando o termo CPC, que é um indicador exclusivo do Brasil.

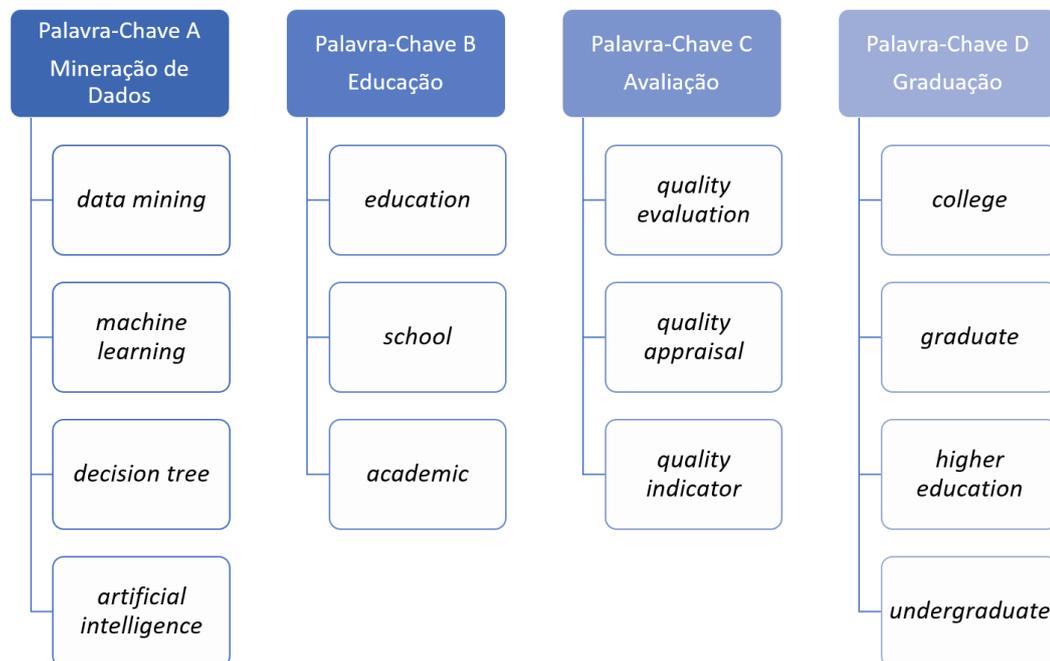
#### 3.1. Pesquisa na Base *Scopus*

Na primeira etapa é realizada a elaboração do Protocolo da Pesquisa, na qual são definidos os seguintes parâmetros:

- **Amostra:** Artigos indexados na base de dados SCOPUS (base confiável e com grande representatividade e abrangência na área acadêmica), utilizando com filtro: artigo de periódico e artigo de conferência publicados de 2005 até 2017.

- **Definição de Palavras-chave, *tesauros* e termos correspondentes:** *data mining, education, evaluation, college education*, além disso, foi realizada uma consulta no site Thesaurus<sup>3</sup> para a definição dos *tesauros* (Figura 4). Vale ressaltar que o termo Conceito Preliminar de Curso (CPC) foi suprimido, pois esse não foi encontrado em nenhum trabalho na base *Scopus*, assim, qualquer interseção não retornava nenhum artigo.

Além disso, realizou-se a Identificação dos Estudos mais relevantes através do seguinte recorte: Periódicos e os autores com pelo menos duas publicações, posteriormente, são identificados os autores e os trabalhos com mais citações, e são considerados os cinco trabalhos mais citados e avaliados através dos seguintes critérios, respectivamente: título, os autores, o ano de publicação e o número de citações, e em seguida sintetizados os dados. Na quinta e última etapa, é realizado um estudo cronológico dos trabalhos produzidos e é apresentada a quantidade de publicações por ano em formato gráfico.



**Figura 4** – Palavras Chaves e *Tesauros*

Fonte: Elaboração própria

<sup>3</sup> <https://www.thesaurus.com/>

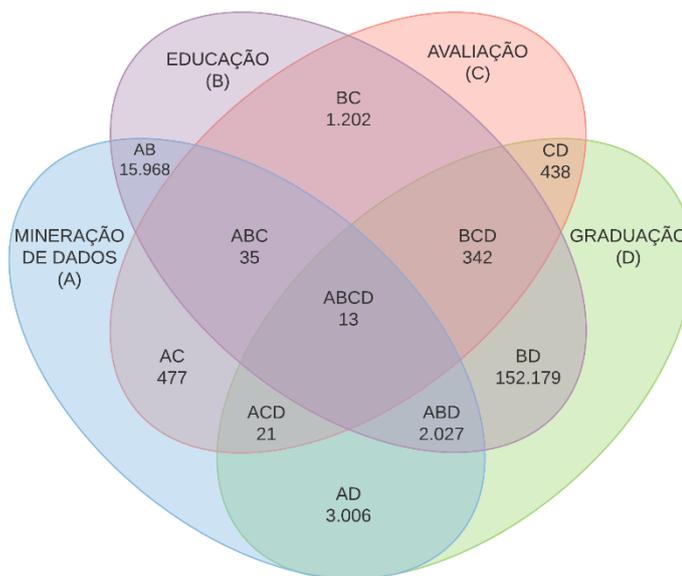
A pesquisa realizada na base *Scopus* está apresentada no Quadro 1, utilizando os padrões da ferramenta de busca da base supracitada.

**Quadro 1** – Descrição da Pesquisa realizada no *Scopus*

<b>Tesouro A</b>	TITLE-ABS-KEY ("data mining" OR "machine* Learning" OR "Decision* Tree*" OR "artificial* intelligence*")
<b>Tesouro B</b>	AND TITLE-ABS-KEY ( education* OR school* OR academic* )
<b>Tesouro C</b>	AND TITLE-ABS-KEY ("quality evaluation*" OR "quality appraisal*" OR "quality* indicator*")
<b>Tesouro D</b>	AND TITLE-ABS-KEY ("graduat*" OR "college*" OR "undergraduate*" OR "higher* education*")
<b>Corte Tipo</b>	AND ( LIMIT-TO ( DOCTYPE , "cp " ) OR LIMIT-TO ( DOCTYPE , " ar " ) )
<b>Corte Temporal</b>	PUBYEAR > 2004 AND PUBYEAR < 2018

Fonte: Elaboração Própria

A Figura 5 apresenta o diagrama de Venn relacionando todas as intersecções presentes entre os termos, em que é apresentado a quantidade de trabalhos encontrados na base de conhecimento *Scopus* de cada Palavra-chave.



**Figura 5** – Diagrama de Venn relacionado aos trabalhos da base *Scopus*.

Fonte: Elaboração própria

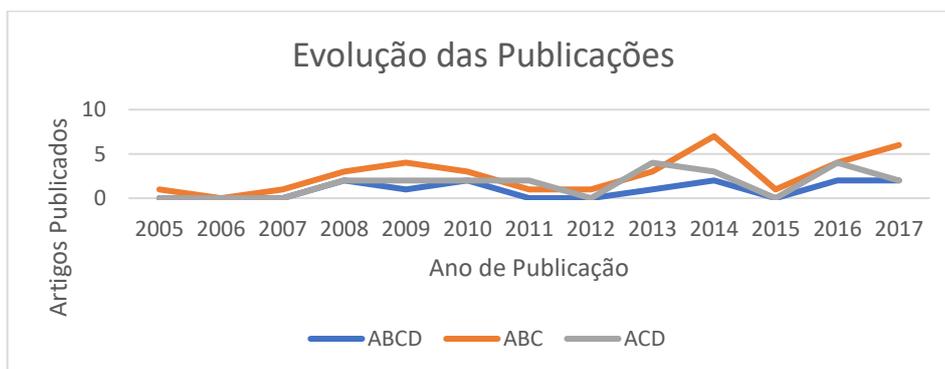
Através da Figura 5, fica evidente que a interseção entre todos os termos supracitados apresenta uma pequena quantidade de trabalhos, além disso, como mencionado nesta Seção, o termo CPC (Conceito Preliminar de Curso) foi retirado da pesquisa, visto que trata-se de um termo não encontrado em diversas pesquisas realizadas na base *Scopus* e, conseqüentemente, sua interseção com qualquer outro termo, retornava a zero trabalhos na área. Este fato, concomitante a pequena gama de trabalho na interseção ABCD, demonstra que esta é ainda uma área pouco explorada na literatura.

Tanto na intersecção ACD (mineração de dados, avaliação e graduação), quanto na ABC (mineração de dados, educação e avaliação) também foi encontrado um pequeno número de trabalhos comparado às outras intersecções, por exemplo, nas buscas de duas palavras-chaves.

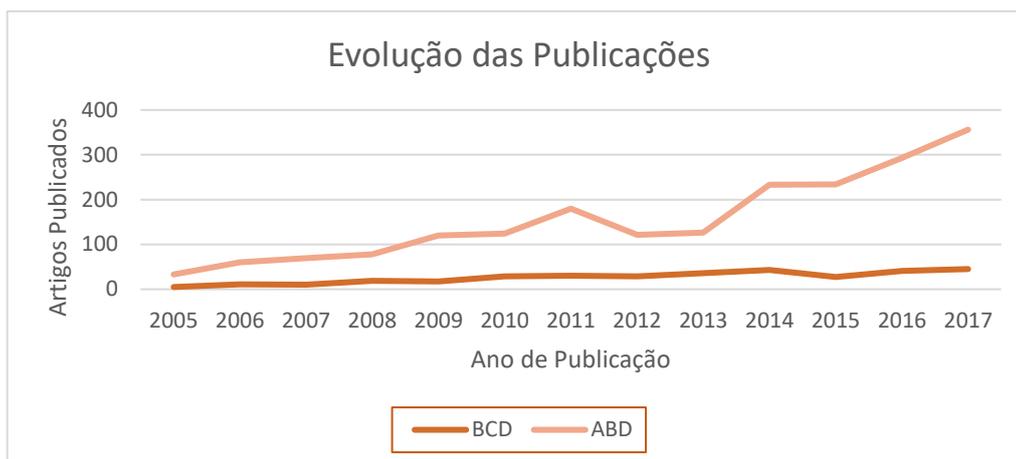
A partir disso, as etapas posteriores foram realizadas utilizando as intersecções ABCD (mineração de dados, educação, avaliação e graduação), ABC (mineração de dados, educação e avaliação), ACD (mineração de dados, avaliação e educação), ABD (mineração de dados, educação e graduação), e BCD (educação, avaliação e graduação), por serem de maior relevância para o presente trabalho. É importante evidenciar que os trabalhos que continham a palavra “Graduação” foram verificados e eram de fato a respeito de cursos de graduação, visto que a tradução para o termo apresenta polimorfismo.

### 3.1.1. Identificação da Evolução Cronológica das Áreas

As Figuras 6 e 7 apresentam a evolução das áreas pesquisadas no que se refere ao número de publicações de acordo com o tempo. A quantidade de trabalhos encontrada na intersecção ABD (mineração de dados, educação e graduação) e BCD (educação, avaliação e graduação) é discrepante uma das outras, por isso são apresentadas de forma independente na Figura 7.



**Figura 6** – Evolução das Publicações para os termos ABCD, ABC e ACD.  
Fonte: Elaboração própria



**Figura 7** – Evolução das Publicações para os termos BCD e ABD.

Fonte: Elaboração própria

### 3.1.2. Identificação dos Periódicos com maior Número de Publicações

Nesta seção, apresenta-se os periódicos com mais publicações em cada uma das pesquisas realizadas. A Tabela 4 apresenta a totalidade de periódicos e anais de conferências encontrados para as intersecções consideradas para a pesquisa.

**Tabela 4** – Quantidade de Artigos por Periódicos/Anais de Conferência.

Termos	Quantidade de Artigos por Periódicos	Quantidade de Anais de Conferência
ABCD	2	11
ABC	8	27
ABD	605	1.422
ACD	15	6
BCD	214	128

Fonte: Elaboração própria

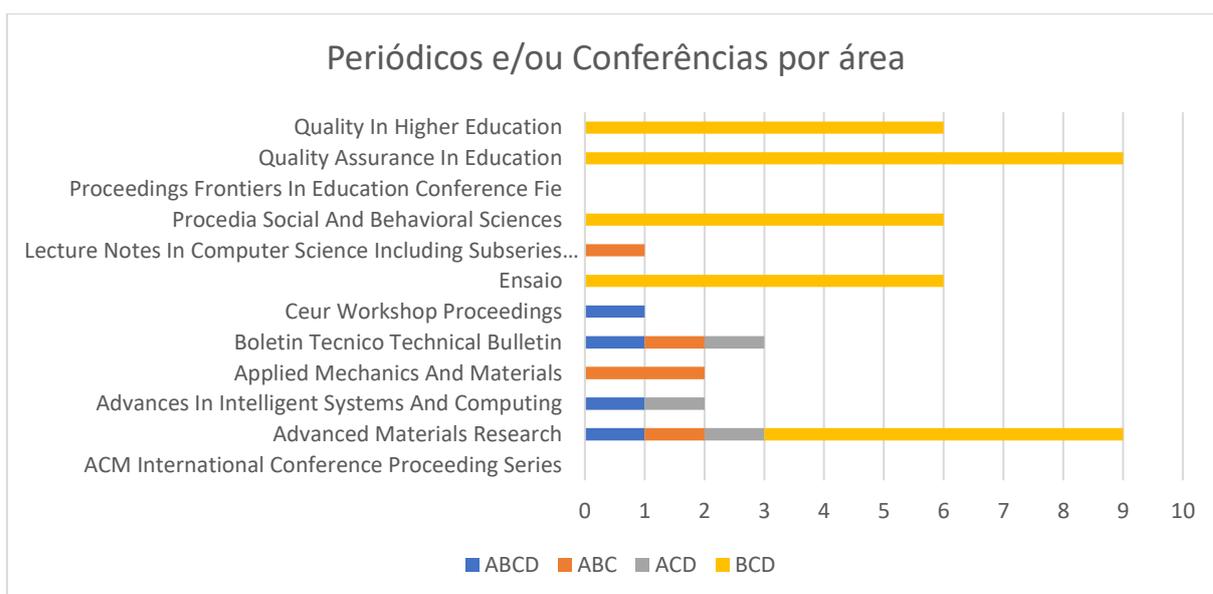
A Tabela 5 apresenta a relação dos principais periódicos com no mínimo duas publicações nas áreas utilizadas como base para pesquisa. Assim, pode-se observar a presença de 12 canais de comunicação científica em que se destaca o *Advanced Materials Research*, que apresentou publicações em quatro das cinco intersecções pesquisadas e o periódico *Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics* devido ao grande número de publicações nas áreas ABD, que é o tema central pesquisado.

**Tabela 5** – Quantidade de Artigos por Periódicos das áreas analisadas.

Periódico	ABCD	ABC	ABD	ACD	BCD
ACM International Conference Proceeding Series	-	-	44	-	-
Advanced Materials Research	1	1	-	1	6
Advances In Intelligent Systems And Computing	1		34	1	-
Applied Mechanics And Materials	-	2	-	-	-
Boletin Tecnico Technical Bulletin	1	1	-	1	-
Ceur Workshop Proceedings	1	-	33	-	-
Ensaio	-	-	-	-	6
Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics	-	1	82	-	-
Procedia Social And Behavioral Sciences	-	-	-	-	6
Proceedings Frontiers In Education Conference Fie	-	-	44	-	-
Quality Assurance In Education / ISSN:0968-4883	-	-	-	-	9
Quality In Higher Education	-	-	-	-	6

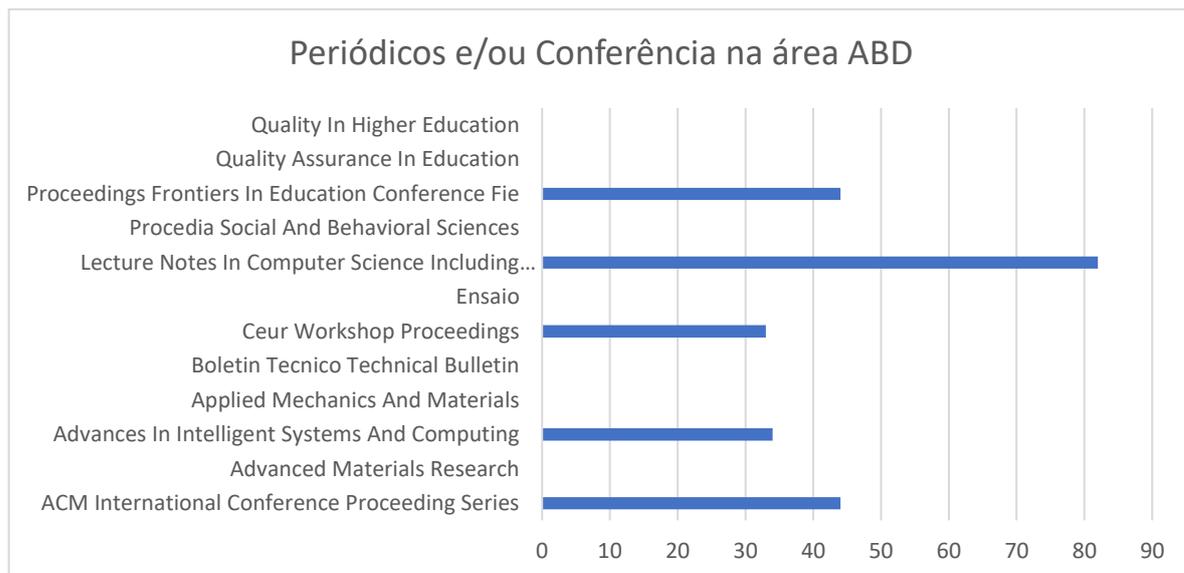
Fonte: Elaboração própria

As Figuras 8 e 9 apresentam os principais periódicos e/ou conferências que tiveram maior número de publicação dentro da respectiva área de acordo com termos pesquisados. Visto que a quantidade de publicações na área dos termos ABD se destaca das outras, foi necessário dividir essa análise em dois gráficos, para melhor visualização dos dados.



**Figura 8** – Quantidade de Periódicos e/ou Conferências por área.

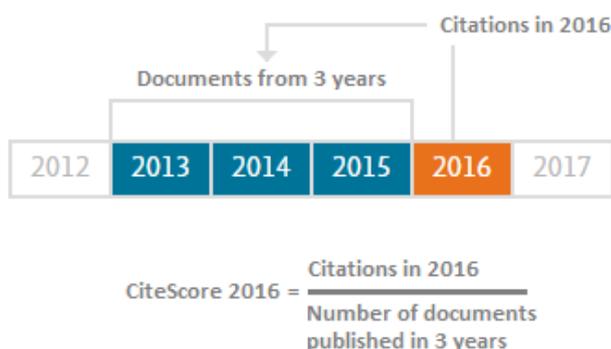
Fonte: Elaboração própria



**Figura 9** – Quantidade de Periódicos e/ou Conferências na área ABD.

Fonte: Elaboração própria

O *Scopus* possui uma ferramenta chamada *CiteScore*, que tem como objetivo medir o impacto das citações dos canais de comunicação, principalmente os periódicos. O cálculo é realizado através da média de citações recebidas por documento. *CiteScore* é o número de citações recebidas por uma revista em um ano para documentos publicados nos três anos anteriores, dividido pelo número de documentos indexados na *Scopus*, publicados nesses mesmos três anos. O cálculo é apresentado na Figura 10.



**Figura 10** – Método para calcular o *CiteScore*.

Fonte: *Scopus*

A fim de avaliar o impacto dos periódicos mais publicados nas áreas pesquisadas, foi utilizada a ferramenta supracitada para corroborar os veículos de comunicação científica que devem ser monitorados com maior frequência para pesquisas e futuras publicações.

As métricas utilizadas pelo *CiteScore* são:

- **Citações no ano:** Quantidade de citações recebidas por um periódico em um ano;
- **Documentos:** Documentos publicados nos três anos anteriores;
- **CiteScore:** Fórmula Apresentada na Figura 10.
- **CiteScore Campo de Estudo:** Indica a posição relativa de um periódico em seu campo de estudo. Um Percentual *CiteScore* de 98% significa que o periódico está no top 2% do seu campo de assunto.

**Tabela 6 – Análise dos Periódicos/Anais de Conferência**

<b>Periódico</b>	<b>Citações em 2017</b>	<b>Documentos 2014-2016</b>	<b>CiteScore</b>	<b>CiteScore Campo de Estudo (%)</b>
ACM International Conference Proceeding Series	3.814	6.788	0.56	24% Reconhecimento de Padrões
Advanced Materials Research	N/A	N/A	N/A	N/A
Advances in Intelligent Systems and Computing	5.040	12.471	0.4	19% Ciência da Computação
Applied Mechanics and Materials	N/A	N/A	N/A	N/A
Boletim Tecnico Technical Bulletin	0	15	0.0	12% Ciência dos Materiais
Ceur Workshop Proceedings	3.677	11.687	0.31	12% Ciência da Computação
Ensaio	35	133	0.26	19% Educação
Lecture Notes in Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics	59.443	66.063	0.9	53% Ciência da Computação
Procedia Social and Behavioral Sciences	N/A	N/A	N/A	N/A
Proceedings Frontiers in Education Conference Fie	594	1.315	0.45	34% Educação
Quality Assurance in Education	123	81	1.52	79% Educação
Quality in Higher Education	56	66	0.85	57% Educação

Fonte: Elaboração própria

De acordo com a Tabela 6, fica evidente que os dois periódicos que apresentam melhores resultados são o *Quality In Higher Education* e o *Quality Assurance In Education*, ambos atuantes do campo de estudo Educação, área de estudo de fundamental importância para o presente trabalho.

### 3.1.3. Identificação dos Trabalhos mais Citados na Área

Os trabalhos mais citados entre os autores sobre a área pesquisada estão apresentados na Tabela 6, em que está mapeado os cinco artigos mais citados de cada um dos conceitos, com exceção das áreas ACD, que retornou apenas quatro artigos. Esses trabalhos são de fundamental importância e devem servir como base para futuras pesquisas.

**Tabela 7 – Artigos mais citados**

<b>Termos</b>	<b>Nome do Artigo</b>	<b>Autores</b>	<b>Qtd. de Citações</b>	<b>Ano</b>
<b>ABCD</b>				
	Developing an intelligent diagnosis and assessment e-learning tool for introductory programming	Huang, C.-J., Chen, C.-H., Luo, Y.-C., Chen, H.-X., Chuang, Y.-T.	12	2008
	Apply data mining to students' choosing teachers under complete credit hour	Wu, F.	4	2010
	A study of teaching evaluation in adult higher education based on decision tree	Deng, J., Hu, J., Chi, H., Wu, J.	2	2010
	Classroom teaching quality evaluation based on neuro-fuzzy ID3 algorithm	Jin, H., Yao, H.	2	2008
	Evaluating the quality of teaching in higher education institutes using clustering approach and set pair analysis	Imtiyaz, M., Singhal, A	1	2016
<b>ABC</b>				
	Using machine learning to support quality judgments	Custard, M., Sumner, T.	22	2005
	Automatically characterizing resource quality for educational digital libraries	Bethard, S., Wetzler, P., Butcher, K., Martin, J.H., Sumner, T.	21	2009
	Characterizing and predicting the multifaceted nature of quality in educational web resources	Wetzler, P., Bethard, S., Leary, H., (...), Martin, J.H., Sumner, T.	9	2013
	An evaluation model of polytechnic teaching quality based on ID3 decision making tree	Jiang, W.X., Zhong, Y.Z., Liang, H.	2	2014
Classroom teaching quality evaluation based on neuro-fuzzy ID3 algorithm	Jin, H., Yao, H.	2	2008	
<b>ABD</b>				
	"flipping" the classroom to explore active learning in a large undergraduate course	Zappe, S., Leicht, R., Messner, J., Litzinger, T., Lee, H.W.	100	2009
	A pilot study to investigate the effect of a simulation strategy on the clinical decision making of midwifery students	Cioffi, J., Purcal, N., Arundell, F.	76	2005
	Using learning analytics to assess students' behavior in open-ended programming tasks	Blikstein, P.	70	2011
	Blended' education and the transformation of teachers: A long-term case study in postgraduate UK higher education	Motteram, G.	68	2006
	Data mining for providing a personalized learning path in creativity: An application of decision trees	Lin, C.F., Yeh, Y.-C., Hung, Y.H., Chang, R.I.	67	2013

Fonte: Elaboração própria

Tabela 7 – Artigos mais citados

Termos	Nome do Artigo	Autores	Qtd. de Citações	Ano
<b>ACD</b>				
	Application of data mining technology in the information technology of college english teaching	Zhimei, Z.	2	2013
	Evaluating the quality of teaching in higher education institutes using clustering approach and set pair analysis	Imtiyaz, M., Singhal, A.	1	2016
	Application and study of ordinal decision tree in the teaching quality evaluation	Ma, H.-Y., Chen, J.-K., Yang, N., Wang, L.-L.	1	2013
	Data mining in teaching quality analysis: A case study in college english teaching	Wu, Y., Huang, H.	1	2009
<b>BCD</b>				
	A systematic review of peer teaching and learning in clinical education	Secomb, J.	152	2008
	Service quality in postgraduate education	Angell, R.J., Heffernan, T.W., Megicks, P.	78	2008
	Value co-creation and university teaching quality: Consequences for the European Higher Education Area (EHEA)	Díaz-Méndez, M., Gummesson, E.	32	2012
	Democratizing information creation from health care data for quality improvement, research, and education- The montefiore medical center experience	Bellin, E., Fletcher, D.D., Geberer, N., Islam, S., Srivastava, N.	31	2010
	Reflections on using journals in higher education: A focus group discussion with faculty	O'Connell, T., Dymont, J.	26	2006

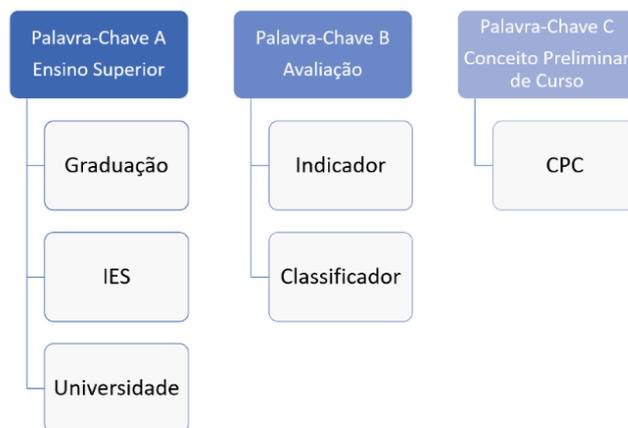
Fonte: Elaboração própria

### 3.2. Pesquisa na Base *Scielo*

Na primeira etapa é realizada a elaboração do Protocolo da Pesquisa, na qual são definidos os seguintes parâmetros:

- **Amostra:** Artigos indexados na base de dados *Scielo* (base confiável e com grande representatividade e abrangência na área acadêmica, principalmente no Brasil), utilizando com filtro: artigo de periódico e artigo de conferência publicados de 2005 até 2018 e Coleções no Brasil.
- **Definição de Palavras-chave, *tesauros* e termos correspondentes:** ensino/educação, indicadores de qualidade e IES, ademais, foi realizada uma consulta no site Thesaurus<sup>4</sup> para a definição dos *tesauros* relacionados aos termos. A Figura 11 apresenta as Palavras-chave utilizadas.

<sup>4</sup> <https://www.thesaurus.com/>



**Figura 11** – Palavras Chaves e *Tesouros*

Fonte: Elaboração própria

Foi realizada a Identificação dos Estudos mais relevantes através do seguinte recorte: Periódicos e os autores com pelo menos duas publicações. Posteriormente, são identificados os autores e os trabalhos com mais citações. E são considerados os cinco trabalhos mais citados e avaliados através dos seguintes critérios, respectivamente: título, os autores, o ano de publicação e o número de citações e, em seguida, os dados são sintetizados. Na última etapa é realizado um estudo cronológico dos trabalhos produzidos e apresentada a quantidade de publicações por ano em formato gráfico.

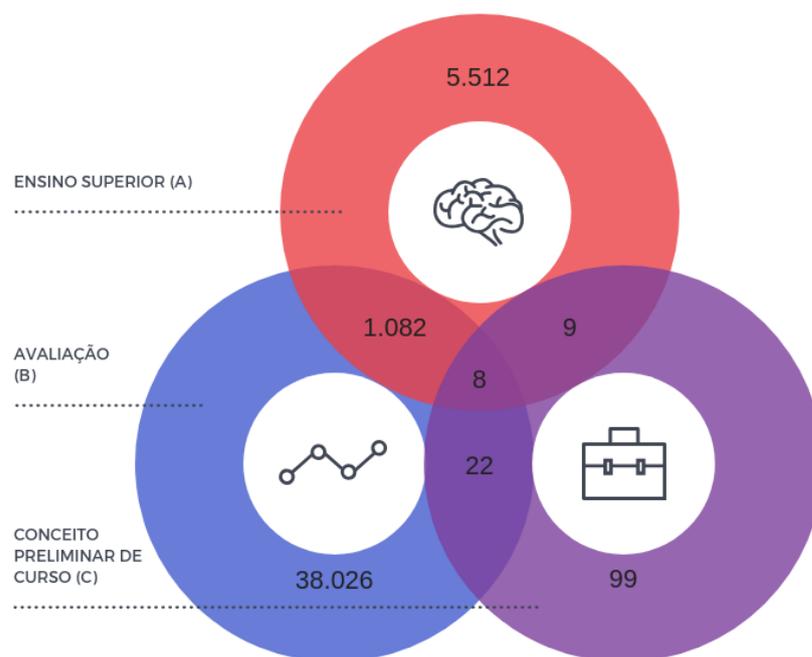
A pesquisa realizada no *Scielo* está apresentada no Quadro 2, utilizando os padrões da ferramenta de busca da base supracitada.

Quadro 2 – Descrição da Pesquisa realizada no *Scielo*

<b><i>Tesouro A</i></b>	("Ensino Superior" OR graduação OR "IES")
<b><i>Tesouro B</i></b>	(Avaliação OR Indicador OR Classificador)
<b><i>Tesouro C</i></b>	("CPC" OR "Conceito Preliminar de Curso")
<b>Corte Tipo</b>	"research-article"
<b>Corte Tempo</b>	2005 até 2018
<b>Corte Coleções</b>	Brasil

Fonte: Elaboração Própria

A Figura 12 apresenta o diagrama de Venn relacionando todas as intersecções presentes entre os termos, que é apresentada a quantidade de trabalhos encontrados na base de conhecimento *Scielo* de cada Palavra-chave.



**Figura 12** – Diagrama de Venn relacionado aos trabalhos da base *Scielo*.

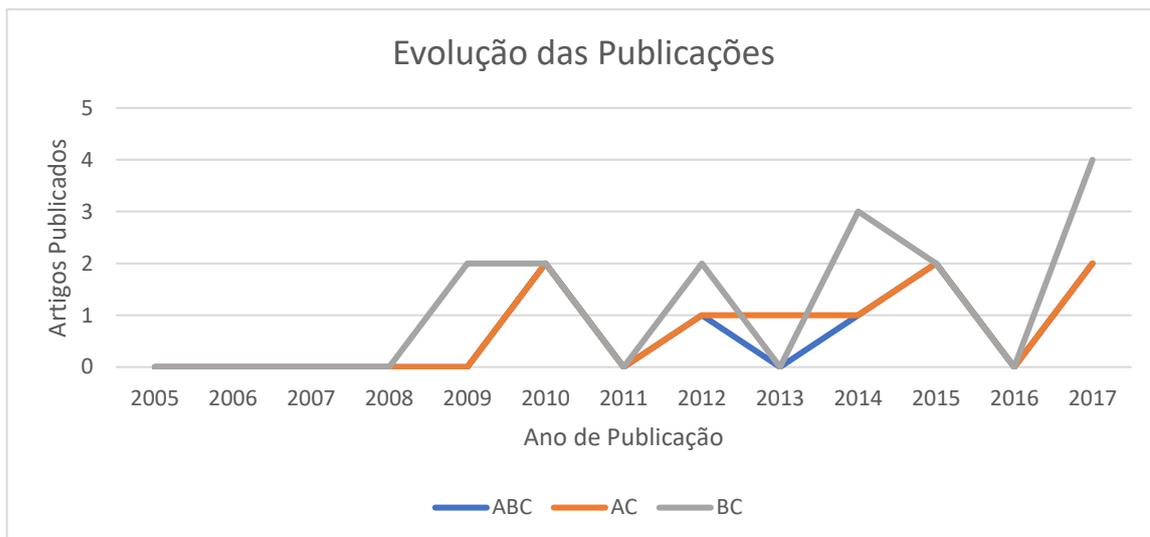
Fonte: Elaboração própria

Através da Figura 12, fica evidente que a interseção de todos os termos supracitados apresenta uma pequena quantidade de trabalhos, além disso, toda interseção que tem presente a Palavra-chave Conceito Preliminar de Curso, que é o objeto de estudo do presente trabalho, exibe um quantitativo baixo de trabalhos, o que demonstra que esta ainda é uma área pouco explorada na literatura.

A partir disso, as etapas posteriores foram realizadas utilizando as interseções ABC (ensino superior, avaliação e CPC), AB (ensino superior e avaliação), AC (ensino superior e CPC) e BC (avaliação e CPC), por serem de maior relevância para o presente trabalho.

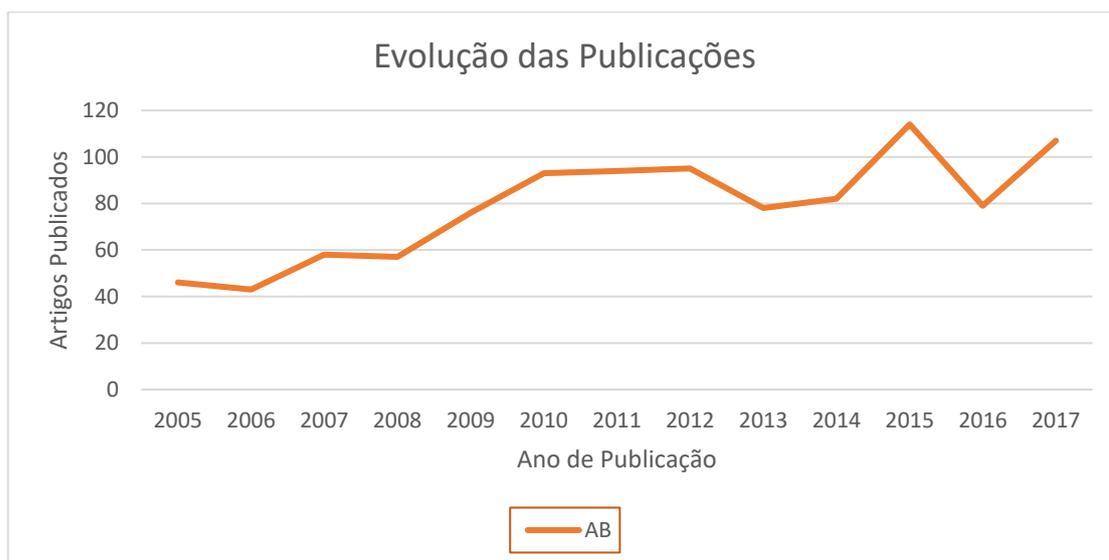
### 3.2.1. Identificação da Evolução Cronológica das Áreas

As Figuras 13 e 14 apresentam a evolução das áreas pesquisadas no que se refere ao número de publicações de acordo com o tempo. A quantidade de trabalhos encontrada na interseção AB (ensino superior e educação) é discrepante das outras, por isso é apresentada na Figura 14 de forma independente.



**Figura 13** – Evolução das Publicações para os termos ABC, AC e BC.

Fonte: Elaboração própria



**Figura 14** – Evolução das Publicações para os termos AB.

Fonte: Elaboração própria

### 3.2.2. Identificação dos Periódicos com maior Número de Publicações

Nesta seção são apresentados os periódicos com mais publicações em cada uma das pesquisas realizadas. A Tabela 8 apresenta a relação dos principais periódicos com no mínimo duas publicações nas áreas utilizadas como base para pesquisa.

**Tabela 8** – Quantidade de Artigos por Periódicos das áreas analisadas.

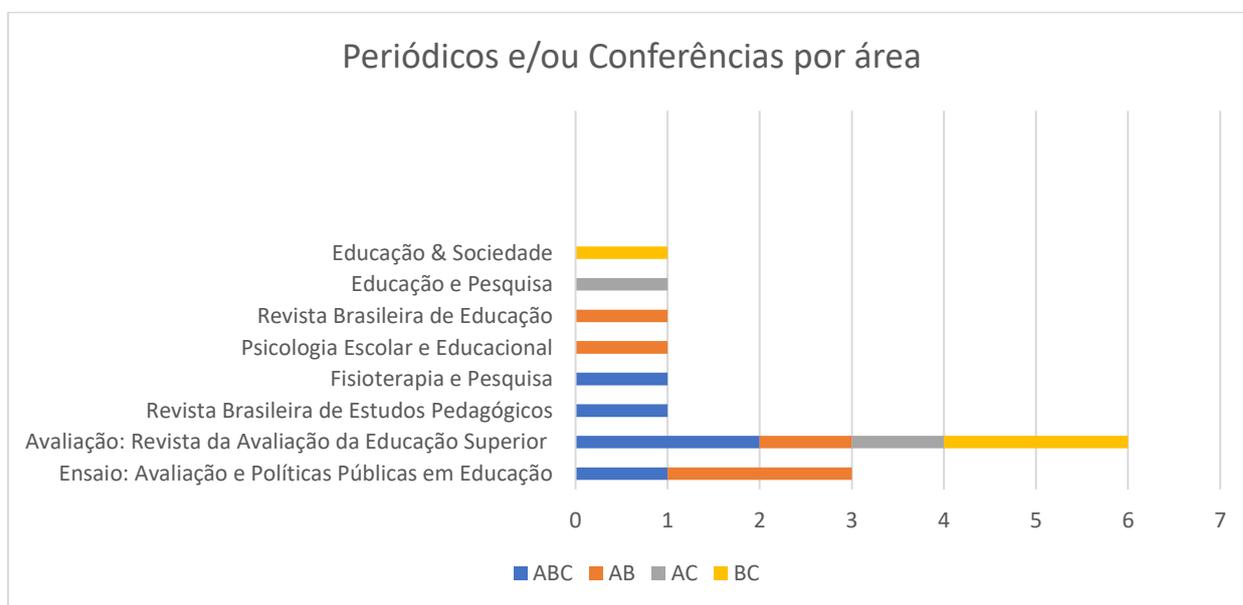
<b>Periódico</b>	<b>ABC</b>	<b>AB</b>	<b>AC</b>	<b>BC</b>
Avaliação: Revista da Avaliação da Educação Superior	3	196	3	1
Ensaio: Avaliação e Políticas Públicas em Educação	1	98	1	6
Educação e Pesquisa	1	12	1	2
Revista Brasileira de Educação	-	16	-	-
Educar em Revista	-	15	-	-
Educação & Sociedade	-	14	-	1

Fonte: Elaboração própria

Através da Tabela 8, pode-se observar a presença de seis canais de comunicação científica, no qual se destacam os periódicos:

1. Avaliação: Revista da Avaliação da Educação Superior;
2. Ensaio: Avaliação e Políticas Públicas em Educação;
3. Educação e Pesquisa.

A Figura 15 apresenta os principais periódicos e/ou conferências que tiveram maior número de publicação dentro da respectiva área de acordo com termos pesquisados.



**Figura 15** – Quantidade de Periódicos e/ou Conferências por área.

Fonte: Elaboração própria

### 3.2.3. Identificação dos Trabalhos mais citados na Área

Os trabalhos mais citados entre os autores sobre a área pesquisada estão apresentados na Tabela 9, na qual está mapeado os cinco artigos mais citados de cada um dos conceitos, com exceção das áreas AC, que retornou apenas dois artigos.

**Tabela 9** – Artigos mais citados

<b>Termos</b>	<b>Nome do Artigo</b>	<b>Autores</b>	<b>Citações</b>	<b>Ano</b>
<b>ABC</b>				
	Análise do Exame Nacional de Desempenho dos Estudantes após o primeiro ciclo avaliativo das áreas de agrárias, saúde e serviço social do Estado do Piauí	Gurgel, Carmesina Ribeiro.	4	2010
	As atuais políticas públicas de avaliação para a educação superior e os impactos na configuração do trabalho docente	Ribeiro, Elisa Antonia	2	2012
	Mudanças nos pesos do CPC e seu impacto nos resultados de avaliação em universidades federais e privadas	Bittencourt, Hélio Radke; Viali, Lorí; Rodrigues, Alziro Cesar de Moraes; Casartelli, Alam de Oliveira.	2	2010
	Relações entre indicadores de qualidade de ensino e desempenho de estudantes dos cursos de Pedagogia do Brasil no Exame Nacional de Desempenho dos Estudantes	Lacerda, Leo Lynce Valle de; Ferri, Cássia.	-	2015
	Avaliação dos cursos de Fisioterapia nos anos de 2004 a 2013	Gonçalves, Rogério Fabiano; Sandes, Aline Araújo Gomes; Nascimento, Isadora Yasmim Monteiro; Amaral, Auxiliadora Renê.	-	2017
<b>AB</b>				
	O processo de adaptação ao ensino superior e o rendimento acadêmico	Chauí, Marilena.Cunha, Simone Miguez; Carrilho, Denise Madruga.	20	2005
	Docência na universidade, cultura e avaliação institucional: saberes silenciados em questão	Cunha, Maria Isabel da.	13	2006
	Do Provão ao ENADE: uma análise comparativa dos exames nacionais utilizados no Ensino Superior Brasileiro	Verhine, Robert Evan; Dantas, Lys Maria Vinhaes; Soares, José Francisco.	12	2006
	Avaliação das universidades brasileiras: as possibilidades de avaliar e as dificuldades de ser avaliado	Ribeiro, Jorge Luiz Lordêlo de Sales.	3	2011
	Emprego de uma abordagem multicritério para classificação do desempenho de instituições de ensino superior	Freitas, André Luís Policani; Rodrigues, Sidilene Gonçalves; Costa, Helder Gomes.	7	2009

Fonte: Elaboração própria

Tabela 9 – Artigos mais citados

Termos	Nome do Artigo	Autores	Citações	Ano
<b>AC</b>				
	O desempenho das universidades brasileiras na perspectiva do Índice Geral de Cursos (IGC)	Hoffmann, Celina; Zanini, Roselaine Ruviano; Corrêa, Ângela Cristina; Siluk, Julio Cezar Mairesse; Schuch Júnior, Vitor Francisco; Ávila, Lucas Veiga.	-	2014
	A avaliação da educação superior nas modalidades presencial e a distância: análises com base no Conceito Preliminar de Cursos (CPC)	Oliveira, Édison Trombeta de; Piconez, Stela Conceição Bertholo.	-	2017
<b>CD</b>				
	Políticas de avaliação da educação superior brasileira: Provão, SINAES, IDD, CPC, IGC e... outros índices	Polidori, Marlis Morosini.	9	2009
	De exames, rankings e mídia	Barreyro, Gladys Beatriz.	7	2008
	Avaliação da educação superior no segundo governo Lula: "provão II" ou a reedição de velhas práticas?	Rothen, José Carlos; Barreyro, Gladys Beatriz.	4	2011

Fonte: Elaboração própria

Esses trabalhos são de fundamental importância e devem servir como base para futuras pesquisas.

### 3.3. Trabalhos Relacionados

De acordo com as pesquisas realizadas, os trabalhos mais relevantes vão ser descritos a seguir. Vale ressaltar que serão apresentados alguns trabalhos provenientes do Estudo Bibliométrico (Subseções 3.1.3 e 3.2.3), como também de uma Pesquisa Bibliográfica mais ampla, considerando além das bases científicas *Scopus e Scielo*, o Portal da Capes, pois esse apresentou trabalhos importantes e convergentes com o presente trabalho.

#### 3.3.1. Trabalhos que utilizam a Mineração de Dados na área Educacional

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.

No que se refere à utilização da técnica de mineração de dados com foco na área educacional, através do perfil de discentes e docentes, tem-se os seguintes trabalhos que se originaram das pesquisas supracitadas:

A pesquisa de Eashwar *et al.* (2017) buscou analisar estudantes de pós-graduação, com o objetivo de identificar os indivíduos que estão mais propensos a desistir de seus cursos, além daqueles que apresentam uma baixa produtividade. Desta forma, é possível dar maior atenção aos mesmos.

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 sobre o 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.

Undavia *et al.* (2016) por intermédio da aplicação de Árvores de Decisão, desenvolveram um modelo para auxiliar estudantes de pós-graduação, no sentido de melhor direcionar seus esforços para uma área que esteja de acordo com seus interesses. Tal medida foi realizada a partir da coleta de dados por questionários aplicados para alunos de 13 universidades que indagavam sobre as características pessoais e suas respectivas graduações, a fim de realizar previsões sobre os cursos de pós-graduação a ser escolhido.

Ferreira (2015) também investigou fatores relacionados a evasão de estudantes, porém tendo como estudo de caso o ensino fundamental. Sua pesquisa apresentou um modelo capaz de prever 96,17% dos casos de evasão escolar, que foram encontradas evidências sobre algumas características dos alunos relacionadas com a não conclusão do Ensino Fundamental, por exemplo, sua etnia.

Também sobre desempenho de alunos, o trabalho de França e Amaral (2013) relata a problemática de que alunos possuem velocidades diferentes de aprendizado devido às suas

necessidades individuais, o que gera dificuldades para cursos de ensino presencial dada a complexidade de personalização de uma sala de aula. Logo, é necessária uma atenção individualizada para estes alunos com dificuldades, o que é o objetivo da pesquisa, em que os autores utilizaram o algoritmo de *clusterização kmeans* para agrupar os alunos de acordo com seus desempenhos em matérias do curso de programação. A criação destes grupos permite que os educadores forneçam atenções diferenciadas para cada um.

Baker *et al.* (2011) apresentaram uma revisão da literatura acerca dos métodos e aplicações que influenciam a pesquisa e a prática da educação em diversos países. A pesquisa permeia o escopo de mineração de dados educacionais, verificando assim, as aplicações internacionais e os desafios para a consolidação da área no Brasil. O trabalho abordou o potencial impacto da Mineração de Dados Educacionais (EDM) na melhora da qualidade dos cursos com foco em cursos de Ensino a Distância (EAD).

Da mesma forma, a pesquisa de Manhães *et al.* (2011) utilizou a mineração de dados para identificar alunos de graduação propensos a evasão de seus cursos. Seus resultados demonstraram que utilizando apenas as notas do primeiro semestre, é possível realizar essa identificação para 80% dos estudantes, atestando assim, a situação final no curso do aluno, ou seja, se permanece matriculado ou não.

Wu (2010), em seu trabalho, utiliza a mineração de dados com enfoque no direcionamento dos alunos na escolha de matérias eletivas no curso de administração de uma Universidade da China. Através de questionários aplicados aos estudantes a respeito do perfil ideal de professores, utilizando o método *clusterização kmeans*, os alunos foram categorizados em quatro grupos, cada grupo era referente a um perfil de docente e disciplina específica.

### 3.3.2. Trabalhos que avaliam a Qualidade do Ensino

No que se refere a avaliação da qualidade do ensino superior e dos seus respectivos cursos, têm-se os seguintes trabalhos:

Gonçalves *et al.* (2017) realizaram um estudo sobre o desempenho dos cursos de fisioterapia no Brasil através 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 resultado que as IES públicas apresentaram melhor desempenho do que as instituições privadas.

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.

Deng *et al.* (2010) utilizam em seu trabalho o algoritmo ID3 com o objetivo de extrair conhecimento acerca da avaliação da qualidade do ensino superior, utilizando como ferramenta a árvore de decisão. Para isso, os autores inicialmente analisaram as deficiências do algoritmo e propuseram algumas melhorias. Posteriormente, validaram o modelo com um estudo de caso, assim foi possível constatar a viabilidade do método proposto que vem a auxiliar em aspectos da melhoria da qualidade de ensino.

O trabalho de Bittencourt *et al.* (2010) faz uma análise acerca do efeito das ponderações utilizadas no CPC e estudam o impacto que eventuais mudanças teriam sobre os conceitos de cursos oferecidos por universidades privadas e federais brasileiras, fazendo uma comparação entre elas.

### 3.3.3. Contribuição do Presente Trabalho

Assim, pode-se observar que se encontram na literatura científica outras pesquisas relacionadas a esta, inseridas no contexto de avaliação de cursos de graduação, além de estudos os quais utilizam técnicas da mineração de dados na área educacional.

Duas abordagens principais foram encontradas, a primeira refere-se ao perfil de alunos e cursos extraídos através da técnica de mineração de dados utilizando bases de dados diversas, como Enade, Saeb, entre outras; e a segunda mostra a análise e avaliações realizadas através do CPC e de alguns outros métodos para aferição da qualidade do ensino e dos cursos. Com isso, destaca-se que não foi encontrado nenhum trabalho que obtivesse as duas abordagens concomitantemente, em que é a proposta do presente trabalho.

## 4. MATERIAIS E MÉTODOS

Com o objetivo de garantir a cientificidade do trabalho, a pesquisa foi classificada dos seguintes pontos de vista: na perspectiva de sua natureza, é considerado como uma pesquisa aplicada, pois está relacionado a resolução de problemas específicos, com o objetivo de obter soluções práticas, além disso, como foi realizada uma análise dos trabalhos publicados na área de mineração de dados em bases relacionadas a educação, além da atribuição de significados aos dados, configura-se uma pesquisa com abordagem qualitativa, corroboram (Silva e Menezes, 2005); a pesquisa é caracterizada como descritiva, pois envolve o uso de técnicas padronizadas de coletas de dados, e descreve características específicas de uma população, estabelecendo relações entre as variáveis, e o fato de utilizar bases de dados reconhecidas cientificamente, utilizando material já publicado, torna a pesquisa classificada no que tange a procedimentos técnicos como pesquisa bibliográfica, afirma (Gil, 2009).

A estratégia de pesquisa foi organizada em quatro etapas macro e sequenciais, conforme ilustrado na Figura 16. O processo do trabalho é baseado nas etapas do KDD, seguindo a proposta de Colak *et al.* (2015).

**Figura 16** – Metodologia do Trabalho.



Fonte: Elaboração própria

A presente pesquisa é realizada na base de dados referente a avaliação da qualidade dos cursos de graduação das IES através do CPC, a partir da plataforma digital do INEP. Tendo como principal foco os cursos de Licenciatura da Rede Federal, estabelecendo um paralelo com as demais IES.

### 4.1.1. Coleta de Dados

**Seleção dos Dados:** Inicialmente é realizada a obtenção e 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.

São considerados como população todos os cursos de Licenciatura da Rede Federal, Universidades Estaduais, Federais e Particulares que obtiveram avaliação no CPC de 2014. O

ano de 2014 foi selecionado, pois refere-se a última avaliação do CPC realizada para os cursos pertencentes ao Ano II (Ciências Exatas, Licenciaturas e áreas afins), com dados disponíveis. A próxima divulgação da avaliação desses cursos está prevista para o final do ano de 2018.

O Quadro 3 apresenta os atributos presentes na base de dados que foi utilizada neste trabalho, no qual é apresentado cada atributo, sua definição e um exemplo. Vale ressaltar, que dentre os dados provenientes do site do INEP, estão inseridos os critérios para cálculo da nota do Conceito Preliminar de Curso. Os demais dados estão relacionados a informações técnicas que possibilitam a identificação dos cursos de graduação.

**Quadro 3** – Dados retirados da base de dados do INEP.

<b>Atributos</b>	<b>Definição</b>	<b>Exemplo</b>
Ano	Ano referente a 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

Fonte: Elaboração Própria

**Quadro 3** – Dados retirados da base de dados do INEP (Continuação).

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 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 - Faixa	4

Fonte: Elaboração Própria

#### 4.1.2. Tratamento dos Dados

Observou-se que os dados extraídos provenientes do Quadro 3 tinham excesso de informações e por isso, foi necessário realizar o Tratamento dos mesmos, com o objetivo de trabalhar apenas com os dados pertinentes ao presente trabalho. Seguindo as etapas do KDD, iniciou-se a atividade de Pré-processamento dos Dados e, posteriormente, a Transformação dos mesmos.

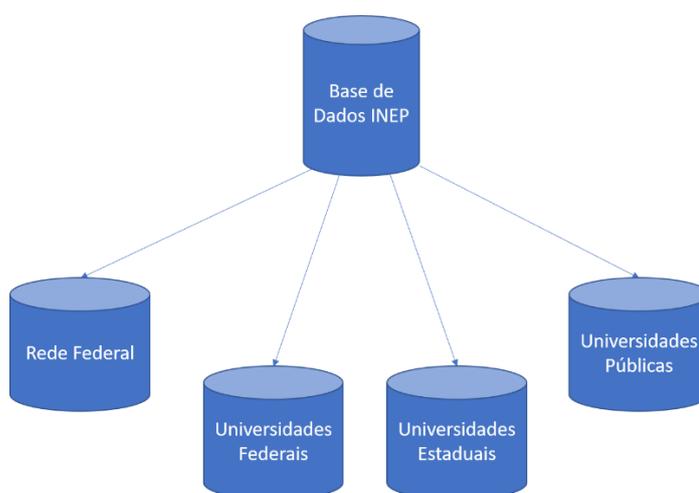
**Pré-processamento dos Dados:** Nesta etapa são realizados o tratamento e a higienização na base de dados supracitada.

Para isso, foi realizado o seguinte procedimento: Iniciou-se por um filtro de todos os cursos de Licenciatura, utilizando o texto Licenciatura na Área de Enquadramento, depois foi feito um filtro por tipo de Instituição (Rede Federal, Universidade Federal, Universidade Estadual e Universidade Privada). A partir disso, foi gerada uma base de dados para cada tipo de IES com os seguintes filtros:

- **Rede Federal:** Foi realizado um filtro com as palavras “Instituto Federal de Educação, Ciência e Tecnologia” E “Centro Federal de Educação Tecnológica” em Organização Acadêmica e um filtro com “UTFPR” em Sigla da IES. Totalizando 290 cursos de Licenciatura da Rede Federal nesta base de dados;
- **Universidade Federal:** Foi realizado um filtro com a palavra “Universidade” em Organização Acadêmica, e a palavra “Federal” em Categoria Administrativa, além disso, foi realizada a exclusão da sigla “UTFPR” da busca, pelo fato dessa Universidade fazer parte da Rede Federal. Totalizando 886 cursos de Licenciatura das Universidades Federais nesta base de dados;
- **Universidade Estadual:** Foi realizado um filtro com a palavra “Universidade” em Organização Acadêmica, e a palavra “Estadual” em Categoria Administrativa. Totalizando 967 cursos de Licenciatura das Universidades Estaduais nesta base de dados;
- **Universidade Privada:** Foi realizado um filtro com a palavra “Universidade” em Organização Acadêmica, e a palavra “Privada” em Categoria Administrativa. Totalizando 779 cursos de Licenciatura das Universidades Privadas nesta base de dados.

A partir disso, tem-se a base de dados principal dividida em outras quatro bases, separando os dados nas categorias: Rede Federal, Universidades Federais, Universidades Estaduais e Universidades Privadas. A Figura 17 apresenta as bases formadas após a divisão.

Nesta etapa do Tratamento da base de dados, cada base de dados resultante contém todos os atributos apresentados no Quadro 3, 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).



**Figura 17** – Composição da Base de Dados do estudo proposto.

Fonte: Elaboração própria

Para a **Higienização dos Dados**, foram realizados os seguintes procedimentos: de todos os atributos apresentados no Quadro 3, foram extraídos nesta 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 os critérios supracitados são os mesmos para todas as IES.

**Transformação dos Dados:** 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<sup>5</sup>, 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 a Mineração de Dados Educacionais, os autores corroboram que a ferramenta WEKA foi a que mais apresentou suporte para a realização do processo de mineração.

A partir da escolha da ferramenta, as quatro bases de dados foram moldadas para que houvesse compatibilidade da utilização das mesmas no *software* WEKA, que tem sua tela inicial apresentada na Figura 18.



**Figura 18** – Tela Inicial do *Software* Weka

Fonte: Elaboração própria

#### 4.1.3. Mineração dos Dados

Na etapa de mineração de dados é utilizado o método de classificação J48, explicado na Seção 2.3. 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 nos trabalhos de Cordeiro *et al.* (2017), Baby (2018) e Ramaphosa *et al.* (2018), que corroboram o método como um dos mais utilizados na literatura no que se refere à mineração de dados educacionais, além de ser o mais preciso quando comparado a outros métodos de classificação, em trabalhos que apresentam propostas convergentes com o presente trabalho. Para a execução do método, como parâmetro de treinamento da rede, foi selecionada a opção *cross-validation*, também conceituada na Seção 2.3.2.

---

<sup>5</sup> <https://www.cs.waikato.ac.nz/ml/weka/>

#### 4.1.4. Análise e Discussão dos Resultados

Na última etapa, as árvores de decisão geradas foram confrontadas com os procedimentos de atribuição de notas do próprio INEP, e com trabalhos encontrados na literatura, gerando discussão acerca dos resultados encontrados.

## 5. 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 – Subseção 2.3) de cada avaliação são disponibilizadas no Apêndice I deste trabalho.

### 5.1. Avaliação da Rede Federal

A Figura 19 ilustra os resultados para o modelo de árvore de decisão calculado tendo como desfecho a Nota Final do CPC dos cursos de Licenciatura da Rede Federal de todo o Brasil. O total de cursos avaliados foi de 290. A taxa de acerto (Subseção 2.3) obtida da base de dados para o método de classificação J48 e parâmetro de treinamento *cross-validation* (Subseção 2.3.2) foi de 80,29%.

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

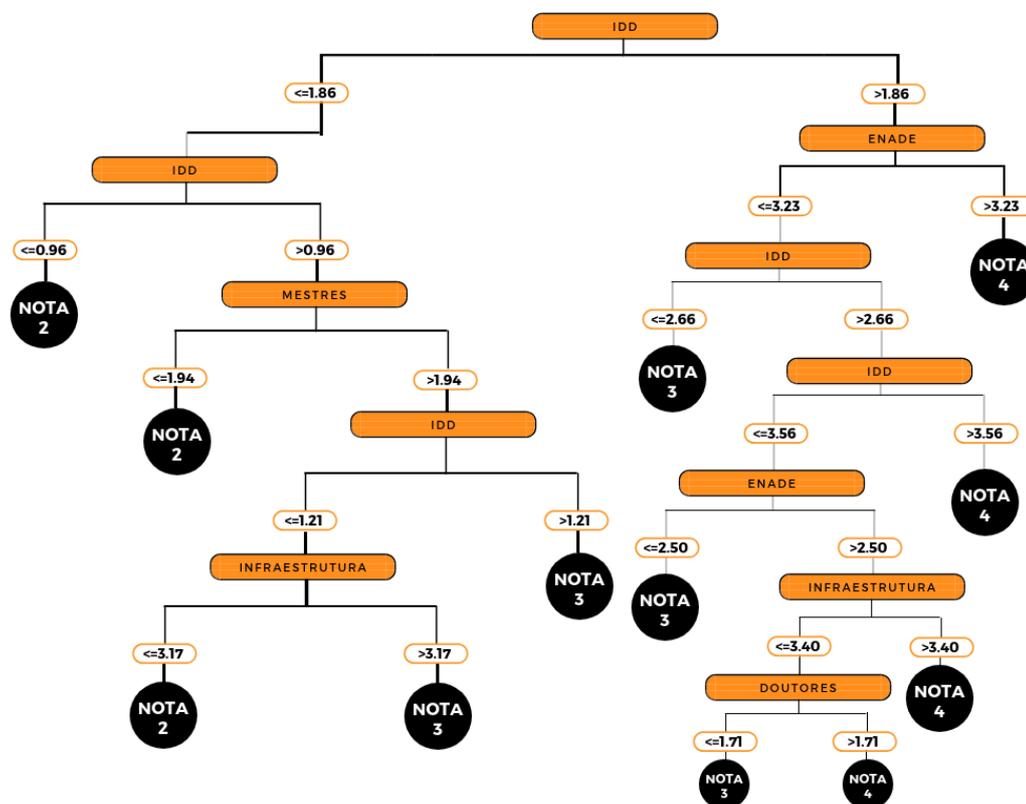


Figura 19 – Árvore de Decisão da Rede Federal.

Fonte: Elaboração própria

A partir da árvore gerada, constataram-se os seguintes perfis para os cursos de Licenciatura dos Institutos Federais, considerando como desfecho a Nota de Avaliação do CPC. A árvore apresentada na Figura 18 será descrita tendo como referência de leitura da esquerda para a direita.

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 nota do IDD apresentar valores entre 0.96 e 1.21, a nota referente a 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 a 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 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 a 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 a avaliação 5 no CPC, visto que apenas um curso (Licenciatura em Química na Universidade Tecnológica Federal do Paraná - UTFPR) foi avaliado com este conceito em 2014.

Através 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).

## 5.2. Avaliação das Universidades Estaduais

As Figuras 20, 21, 22, 23 e 24 estão particionadas representando uma única árvore de decisão, devido a delimitação de espaço e para melhor compreensão da mesma. Elas ilustram os resultados para o modelo de árvore de decisão calculado tendo como desfecho a Nota do CPC dos cursos de Licenciatura das Universidades Estaduais de todo o Brasil. O total de cursos avaliados foi de 967. A taxa de acerto obtida da base de dados para o método de classificação J48 e parâmetro de treinamento *cross-validation* foi de 82,68%.

A partir disso, é possível analisar os principais padrões referentes a avaliação final do CPC dos cursos de Licenciatura nas Universidades Estaduais, como também, obter conhecimento importante relativo a estes padrões. Vale ressaltar que no presente trabalho serão descritos e analisados apenas a Avaliação Nota 2 (curso próximo ao descredenciamento), e a Avaliação Nota 5 (curso com excelência na avaliação), as demais Figuras que compõem a árvore encontram-se no Apêndice II.

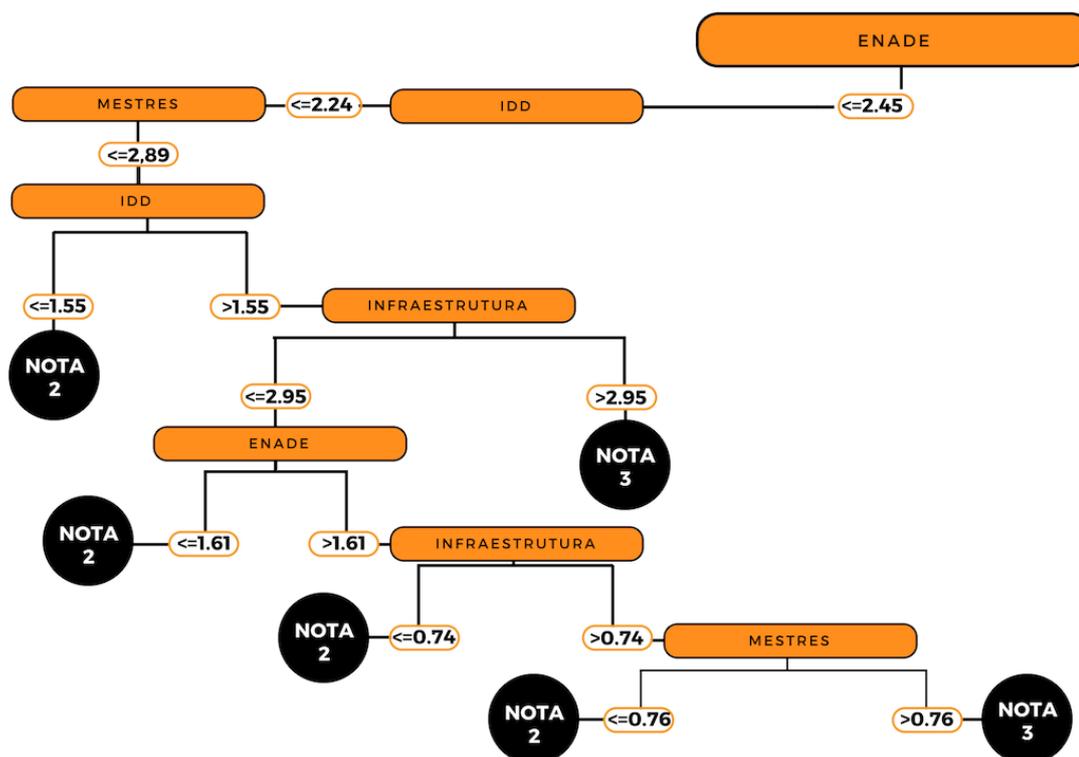
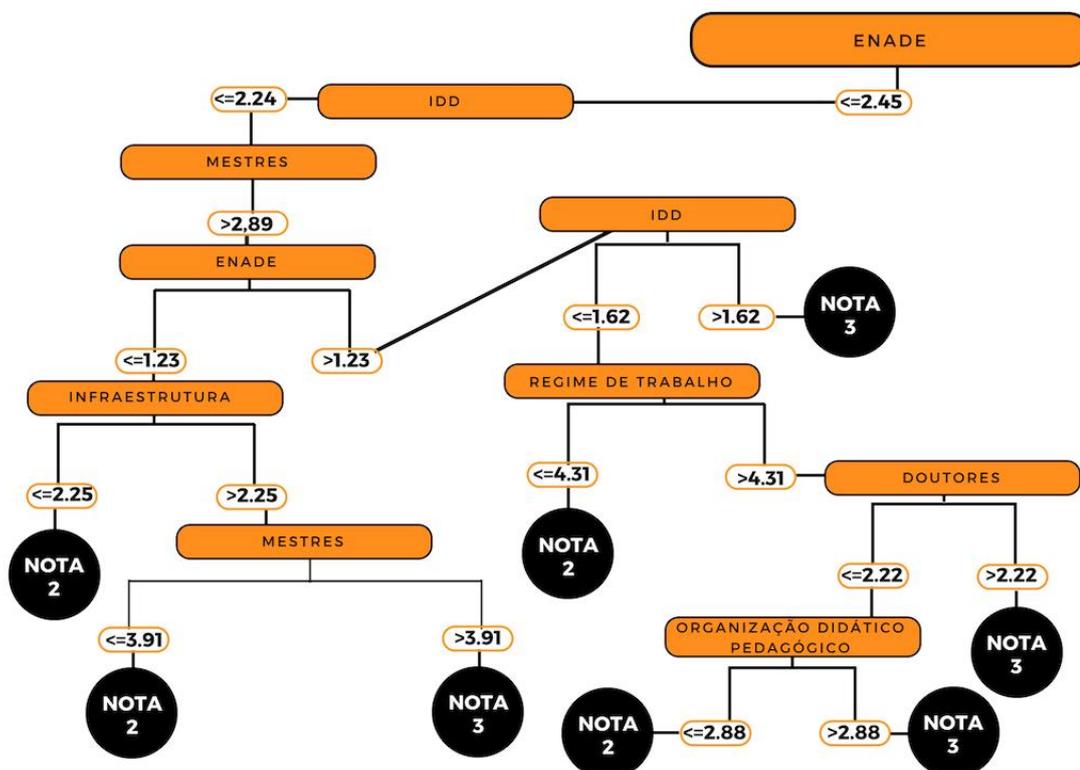


Figura 20 – Árvore de Decisão Universidades Estaduais (a).

Fonte: Elaboração própria

Como apresentado na Figura 20, a avaliação com **Nota 2** pode ser obtida nos seguintes casos: se a nota do Enade for menor ou igual a 2.45 combinado com a nota referente a Mestres menor ou igual a 2.89 e a nota do IDD menor ou igual a 1.55; se a nota do IDD no intervalo de 1.55 e 2.24, a nota de Mestres for menor ou igual a 2.89, a nota da Infraestrutura menor ou igual a 2.95 e a nota do Enade for menor que 1.61; se a nota do Enade estiver entre 1.61 e 2.45, a nota do IDD no intervalo de 1.55 e 2.24, a nota de Mestres for menor ou igual a 2.89 e a nota da Infraestrutura menor ou igual a 0.74; se a nota do Enade estiver entre 1.61 e 2.45, a nota do IDD no intervalo de 1.55 e 2.24, a nota da Infraestrutura entre 0.74 e 2.95 e a nota de Mestres menor ou igual a 0.76.



**Figura 21** – Árvore de Decisão Universidades Estaduais (b).

Fonte: Elaboração própria

Como apresentado na Figura 21, a avaliação com **Nota 2** pode ser obtida nos seguintes casos: se a nota do IDD for menor que 2.24, a nota do Enade inferior ou igual a 1.23 e a nota relativa a Infraestrutura menor ou igual a 2.25; se a nota do IDD for menor que 2.24, a nota do Enade inferior ou igual a 1.23, a nota relativa a Infraestrutura maior a 2.25 e a nota referente a Mestres entre 2.89 e 3.91; se a nota do Enade estiver no intervalo de 1.23 e 2.45, a nota

relativa a Mestres for superior a 2.89, a nota do IDD menor ou igual a 1.62 e a nota relacionado ao Regime de Trabalho (RT) inferior a 4.31; se a nota do Enade estiver no intervalo de 1.23 e 2.45, a nota relativa a Mestres for superior a 2.89, a nota do IDD menor ou igual a 1.62, a nota de RT maior que 4.31, a nota relativa a Doutores menor ou igual a 2.22 e a nota relacionada a ODP menor ou igual a 2.88.

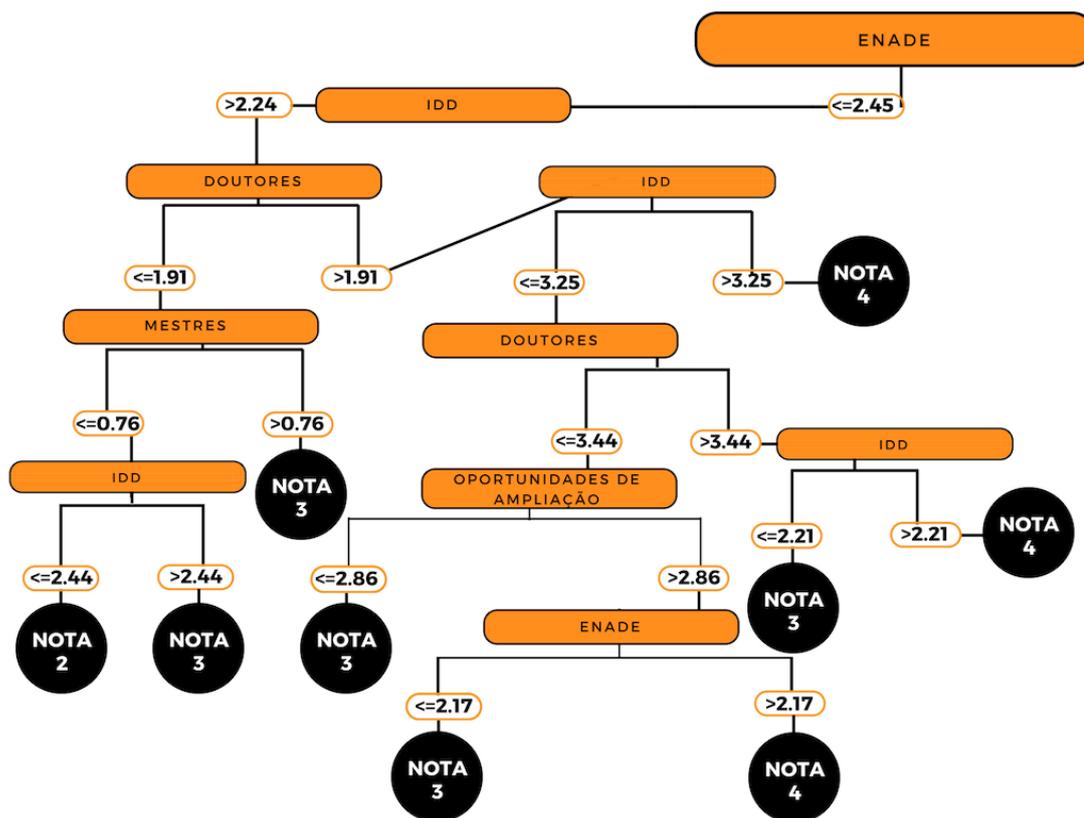


Figura 22 – Árvore de Decisão Universidades Estaduais (c).

Fonte: Elaboração própria

Como apresentado na Figura 22 a avaliação com **Nota 2** pode ser obtida no seguinte caso: a nota do Enade menor ou igual a 2.45, a nota relativa a Doutores menor ou igual a 1.91, a nota referente a Mestres menor ou igual a 0.76 e a nota do IDD no intervalo de 2.24 e 2.44.

A Figura 23 demonstra que a avaliação com **Nota 2** pode ser obtida no seguinte caso: Se a nota do Enade estiver no intervalo de 2.45 e 3.54, a nota relativa a Doutores inferior ou igual a 2.35, a nota do IDD menor ou igual a 1.69 e a nota referente a Infraestrutura menor ou igual a 1.57.

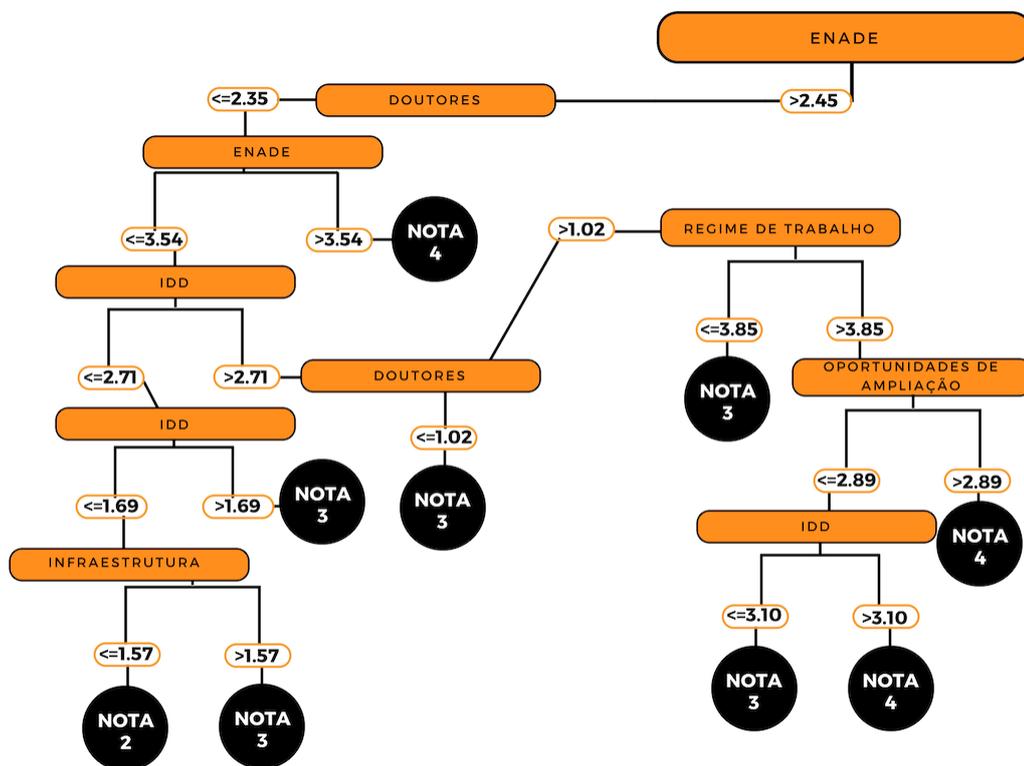


Figura 23 – Árvore de Decisão Universidades Estaduais (d).

Fonte: Elaboração própria

. A **Nota 5** no CPC é obtida para os cursos que apresentam apenas o seguinte perfil, como demonstrado na Figura 24: A nota relativa a Doutores superior a 2.35, a nota do IDD maior que 3.06 e a nota do Enade maior que 4.18.

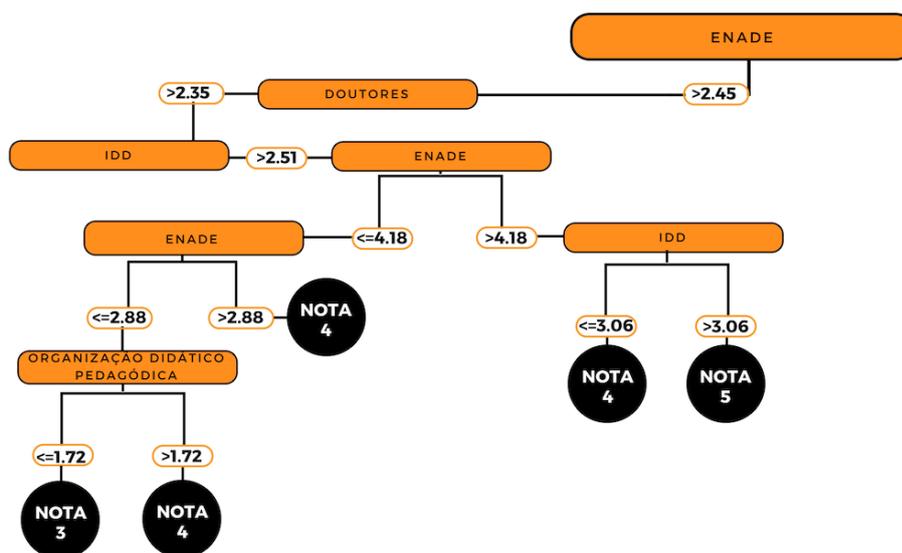


Figura 24 – Árvore de Decisão Universidades Estaduais (e).

Fonte: Elaboração própria

Através da análise da árvore de decisão apresentada, é possível inferir que os critérios que mais impactam tanto na Avaliação Nota 5, quanto na Avaliação Nota 2 são 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. A partir disso, é possível concluir que um bom desempenho nestes critérios é primordial para os cursos avaliados. No que tange a avaliação do curso com Nota 2 (próximo ao descredenciamento), dois critérios, Infraestrutura e Nota de Mestres, se destacam e podem auxiliar os cursos a evoluírem para pelo menos a Nota 3, visto que são critérios que não dependem da atuação do aluno e sim da Universidade. Já na proposição de obter uma avaliação de excelência, além dos critérios nota do IDD e nota do Enade, é imprescindível uma boa atuação na Nota referente a Doutores no que tange as Universidades Estaduais.

### 5.3. Avaliação das Universidades Federais

As Figuras 25, 26, 27, 28 estão particionadas representando uma única árvore de decisão, devido a delimitação de espaço e para melhor compreensão da mesma. Elas ilustram os resultados para o modelo de árvore de decisão calculado tendo como desfecho a Nota do CPC dos cursos de Licenciatura das Universidades Federais de todo o Brasil.

O total de cursos avaliados foi de 886. A taxa de acerto obtida da base de dados para o método de classificação J48 e parâmetro de treinamento *cross-validation* foi de 85%. A partir disso, é possível analisar os principais padrões referentes à avaliação final do CPC dos cursos de Licenciatura nas Universidades Federais, como também, obter conhecimento importante relativo a estes padrões. Vale ressaltar que no presente trabalho serão descritos e analisados apenas a Avaliação Nota 2 (curso próximo ao descredenciamento), e a Avaliação Nota 5 (curso com excelência na avaliação). As demais Figuras que compõem a árvore encontram-se no Apêndice II.

A Figura 25 demonstra que a avaliação com **Nota 2** pode ser obtida nos seguintes casos: Se a nota do Enade for menor ou igual a 1.49 combinado com a nota referente ao IDD menor ou igual a 0.86; Se a nota do Enade for menor ou igual a 1.49, a nota referente ao IDD superior a 0.86, mas a nota relacionada a Organização Didático-pedagógica (ODP) inferior ou igual a 0.65; Se a nota do Enade for menor ou igual a 1.49, a nota do IDD estiver no intervalo de 0.86 e 2.46, a nota do ODP superior a 0.65, mas a nota referente a Mestres for inferior a 2.46.

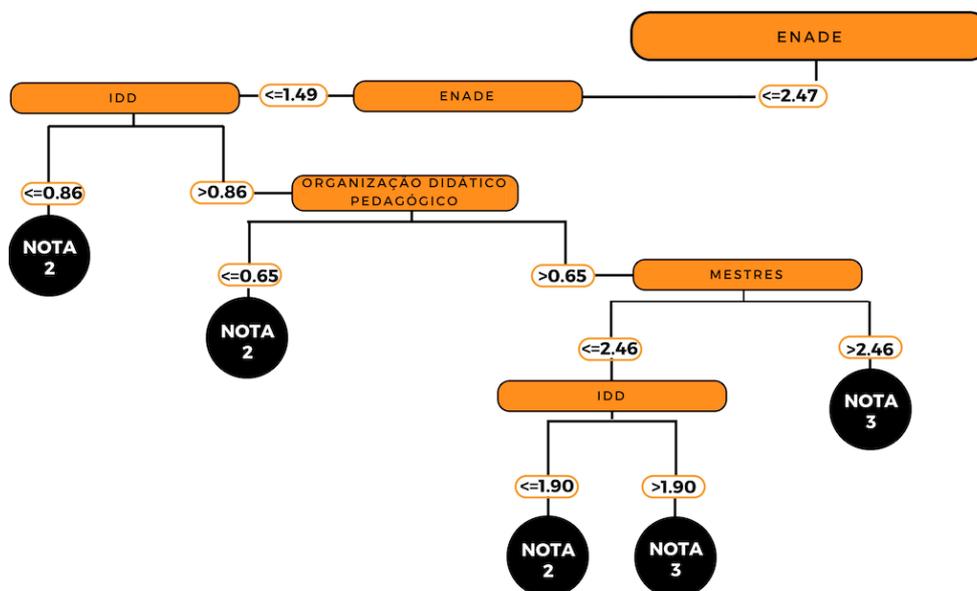


Figura 25 – Árvore de Decisão Universidades Federais (a).

Fonte: Elaboração própria

A Figura 26 demonstra que a avaliação com **Nota 2** pode ser obtida no seguinte caso: Se a nota do Enade for superior a 2.47, mas a nota do IDD for menor ou igual a 1.04 e a nota que se refere a Doutores menor ou igual a 2.53.

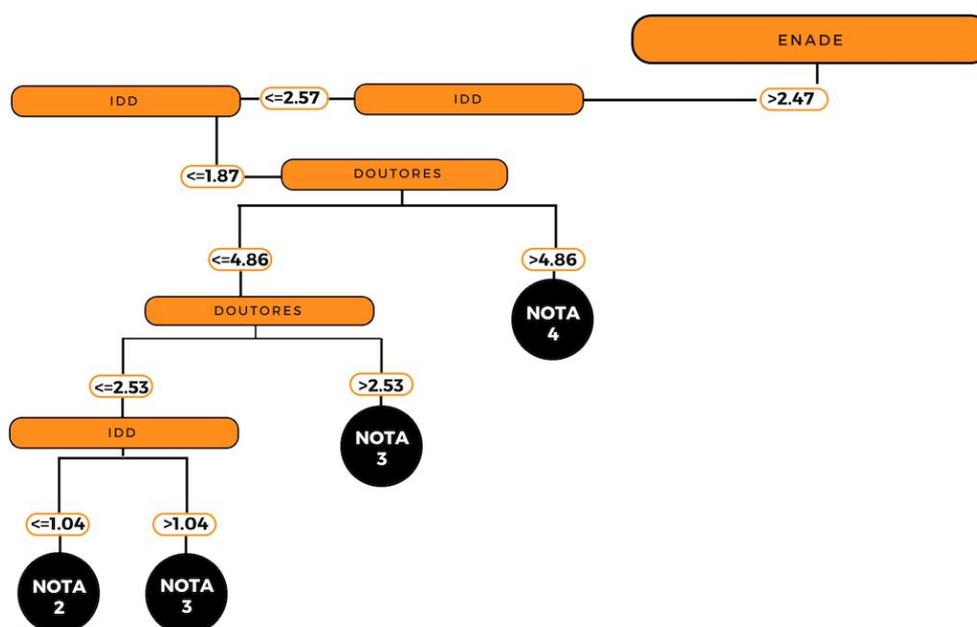


Figura 26 – Árvore de Decisão Universidades Federais (b).

Fonte: Elaboração própria

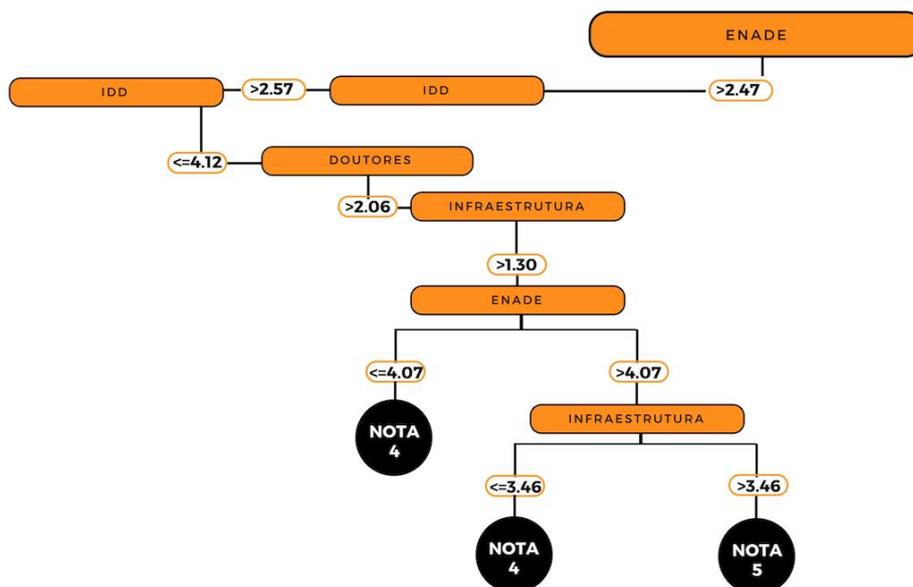


Figura 27 – Árvore de Decisão Universidades Federais (c).

Fonte: Elaboração própria

Como apresentado na Figura 27, a avaliação com **Nota 5** pode ser obtida no seguinte caso: se a nota do Enade for maior que 4.07, a nota do IDD estiver entre 2.57 e 4.12, a nota referente a Doutores ser superior a 2.06 e a nota da Infraestrutura maior que 3.46.

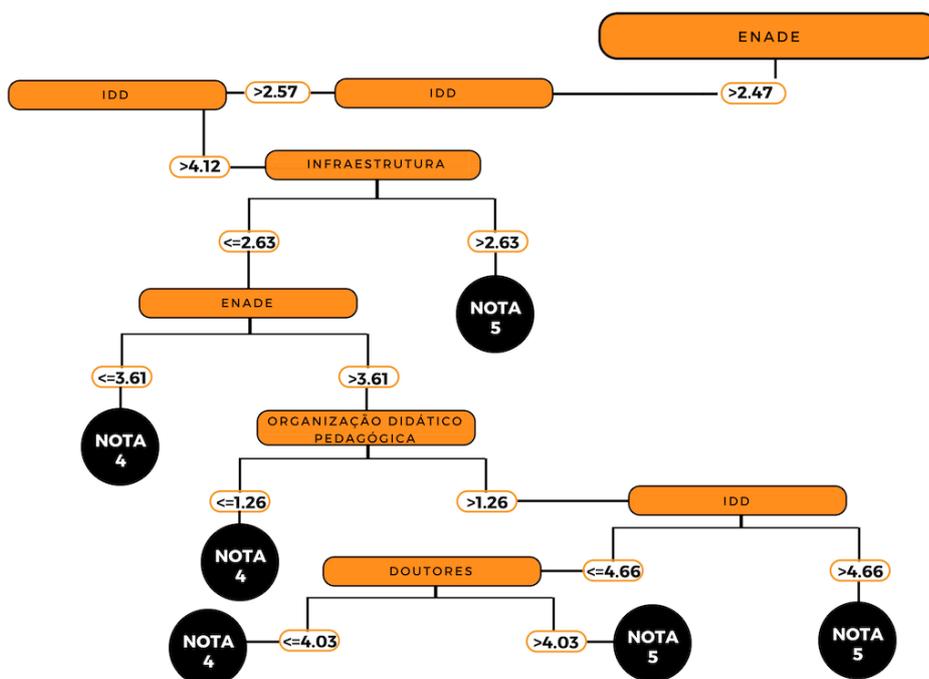


Figura 28 – Árvore de Decisão Universidades Federais (d).

Fonte: Elaboração própria

Como apresentado na Figura 28, a avaliação com **Nota 5** pode ser obtida no seguintes casos: se a nota do Enade for maior que 2.47, a nota referente ao IDD superior a 4.12 e a nota da Infraestrutura maior que 2.63; se a nota do Enade for maior que 3.61, a nota da ODP superior a 1.26 e a nota referente do IDD superior a 4.66; se a nota do Enade for maior que 3.61, o IDD estiver no intervalo de 4.12 e 4.66, a nota da ODP superior a 1.26 e a nota referente a Doutores superior a 4.03.

Através da análise da árvore de decisão apresentada, é possível inferir que os critérios que mais impactam tanto na Avaliação Nota 5, quanto na Avaliação Nota 2 são as notas do Enade e do IDD, respectivamente. A nota da Organização Didático-pedagógica (ODP) também foi um critério determinante nas duas avaliações.

No que se refere a Avaliação Nota 2 foram observados dois critérios importantes (Nota de Mestres e Nota de Doutores) que se tiverem melhores desempenhos, podem contribuir para a evolução da nota do CPC dos cursos, evoluindo de Nota 2 para Nota 3. É importante ressaltar que o número de Doutores também impacta na Nota de Mestres, assim os critérios estão relacionados e merecem atenção no momento de estruturação do curso.

Na Avaliação Nota 5, a nota referente a Doutores e a Infraestrutura são critérios de extrema importância para alcançar a excelência na avaliação, além de serem critérios que não estão ligados diretamente aos alunos.

#### **5.4. Avaliação das Universidades Privadas**

As Figuras 29, 30, 31, 32, 33 estão particionadas representando uma única árvore de decisão, devido a delimitação de espaço e para melhor compreensão da mesma. Elas ilustram os resultados para o modelo de árvore de decisão calculado tendo como desfecho a Nota Final do CPC dos cursos de Licenciatura das Universidades Privadas de todo o Brasil.

O total de cursos avaliados foi de 779. A taxa de acerto obtida da base de dados para o método de classificação J48 e parâmetro de treinamento *cross-validation* foi de 81,03%. A partir disso, é possível analisar os principais padrões referentes à avaliação final do CPC dos cursos de Licenciatura nas Universidades Privadas, como também, obter conhecimento importante relativo a estes padrões. Vale ressaltar que no presente trabalho serão descritos e analisados apenas a Avaliação Nota 2 (curso próximo ao descredenciamento), e a Avaliação Nota 5 (curso com excelência na avaliação).

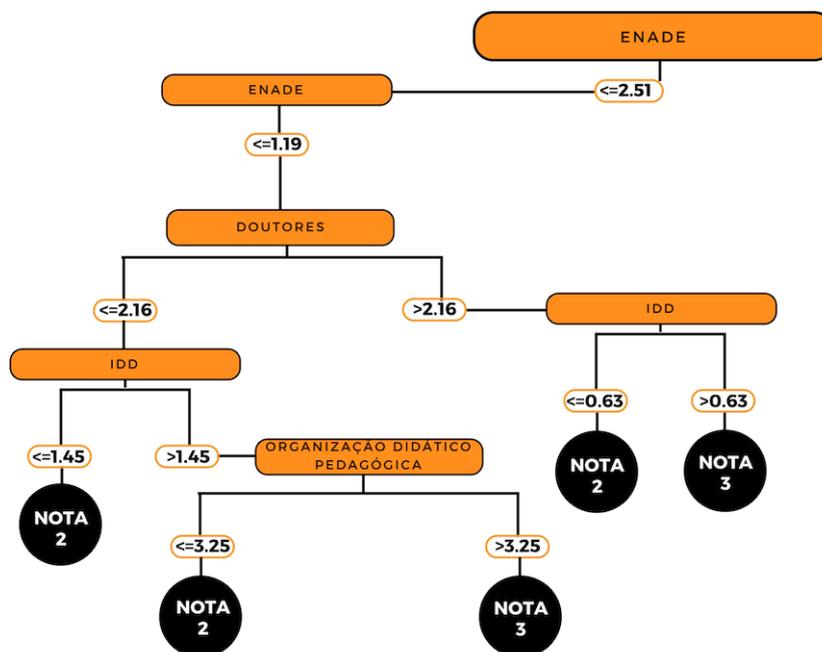


Figura 29 – Árvore de Decisão Universidades Privadas (a).

Fonte: Elaboração própria

Como apresentado na Figura 29 a avaliação com **Nota 2** pode ser obtida nos seguintes casos: Nota do Enade menor ou igual a 1.19, nota referente a Doutores menor que 2.16 e nota do IDD menor ou igual a 1.45; Nota do Enade menor ou igual a 1.19, nota referente a Doutores menor que 2.16 e a nota do ODP menor que 3.25; Nota do Enade menor ou igual a 1.19, nota referente a Doutores superior a 2.16, porém a Nota no IDD menor ou igual a 0.63.

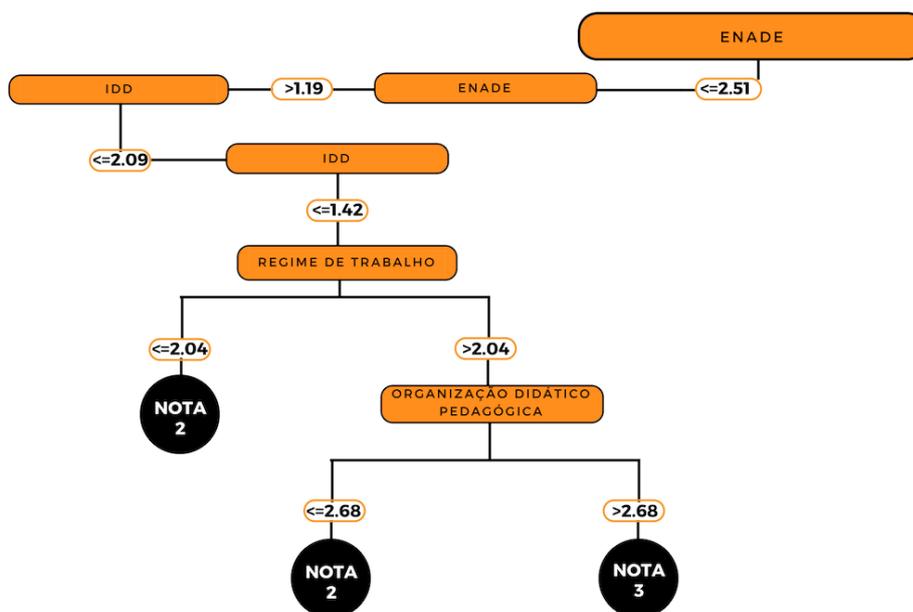
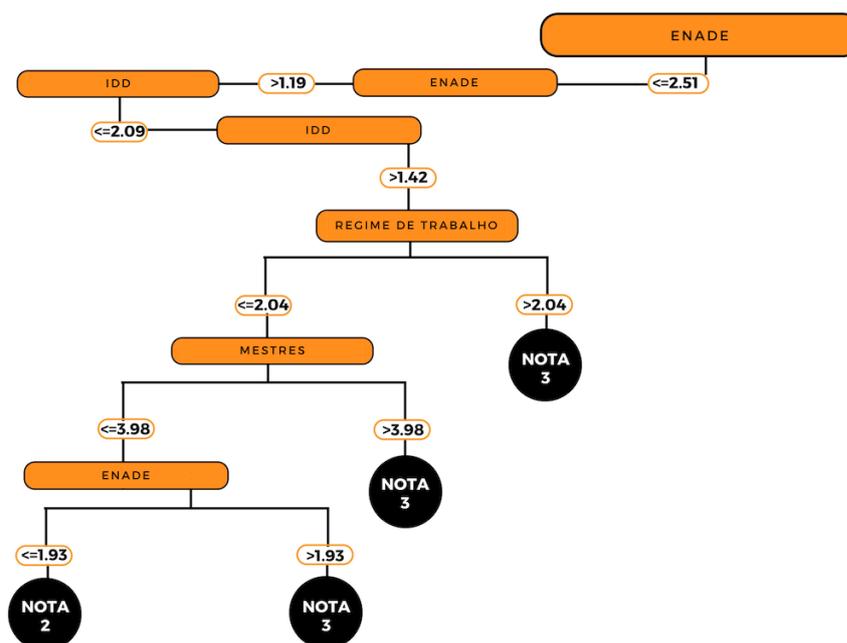


Figura 30 – Árvore de Decisão Universidades Privadas (b).

Fonte: Elaboração própria

Como apresentado na Figura 30 a avaliação com **Nota 2** pode ser obtida nos seguintes casos: se a nota do Enade estiver entre 1.19 e 2.51, a nota do IDD não atingir pelo menos 1.42 e a Nota referente ao Regime de Trabalho for menor ou igual a 2.04; se a nota do Enade estiver entre 1.19 e 2.51, a nota do IDD não atingir pelo menos 1.42, a nota relativa ao Regime de Trabalho for superior a 2.04, porém a nota referente ao ODP for menor ou igual a 2.68.



**Figura 31** – Árvore de Decisão Universidades Privadas (c).

Fonte: Elaboração própria

Como apresentado na Figura 31 a avaliação com **Nota 2** pode ser obtida no seguinte caso: se a nota do Enade for menor ou igual a 1.93, a nota do IDD no intervalo de 1.42 e 2.09, a nota referente ao Regime de Trabalho menor ou igual a 2.04 e a Nota de Mestres não atingir um valor de pelo menos 3.98.

A Figura 32 demonstra que a avaliação com **Nota 5** pode ser obtida nos seguinte caso: se a nota referente ao Enade for superior a 2.51, a Nota de Mestres superior a 4, a nota referente ao IDD maior que 4,51 e a nota de Infraestrutura maior que 3.58; se a nota referente ao Enade for superior a 2.51, a Nota de Mestres superior a 4.0, a nota estiver no intervalo de 3.62 e 4.51 e a nota de Infraestrutura maior que 4.13 e a nota relativa a Doutores maior que 3.23; Nota do Enade entre 4.06 e 4.29, a nota referente ao IDD no intervalo de 3.62 e 4.51 e a

nota referente a Doutores inferior ou igual a 3.23, é necessário que a Nota relativa a Infraestrutura seja maior que 4.12.

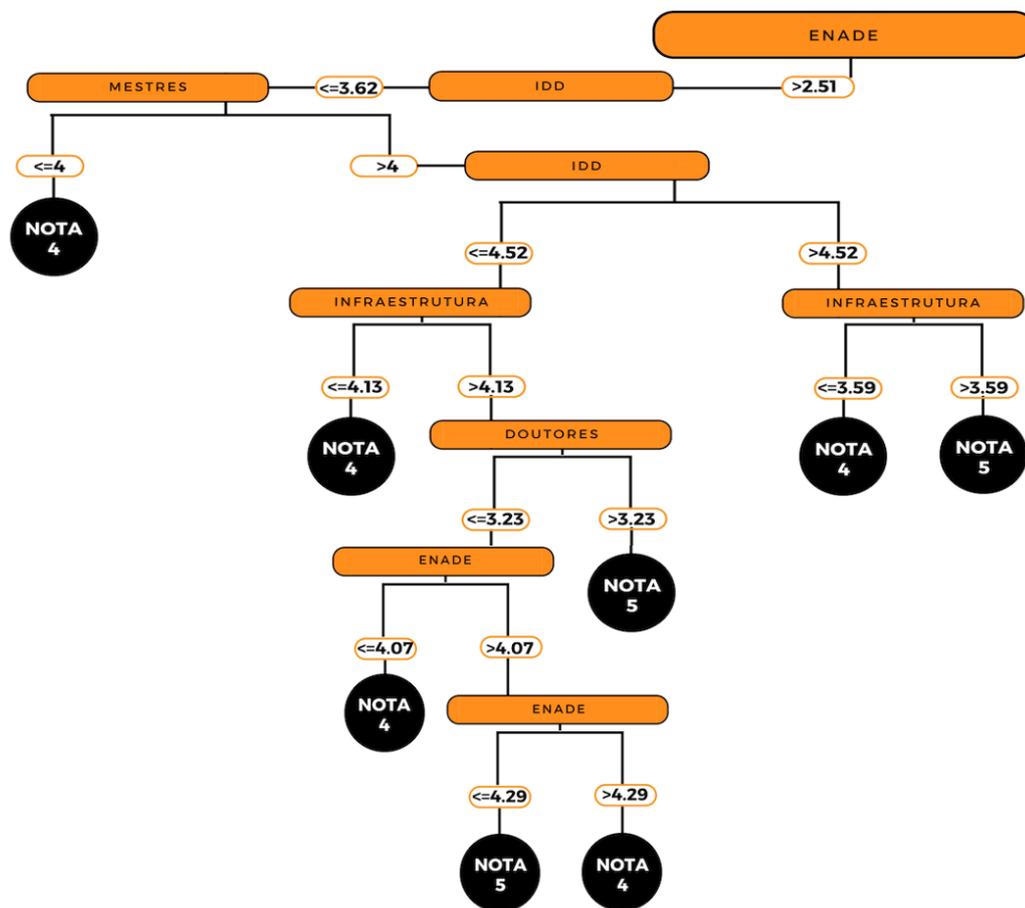


Figura 32 – Árvore de Decisão Universidades Privadas (d).

Fonte: Elaboração própria

Através da análise da árvore de decisão apresentada, é possível inferir que as Universidades privadas têm em seus perfis um número maior de critérios que impactam tanto na Avaliação Nota 5, quanto na Avaliação Nota 2, sendo os principais: Nota do Enade e IDD. Contudo apenas a obtenção de bons índices nos critérios supracitados não garante aos cursos das Universidades Privadas uma boa avaliação.

No que tange a Nota 2, além da Nota do Enade e do IDD, a Organização Didático-pedagógica e o Regime de Trabalho são critérios importantes a serem melhorados para uma possível evolução de Nota. Já para Nota 5, os critérios Nota relativa de mestres e a doutores combinado com a Infraestrutura merecem destaque no panorama das Universidades Privadas.

## 5.5. Discussão dos Resultados

Na Tabela 10, 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 10** – Quantidade de vezes os critérios estão presentes nas árvores.

<i>Crítérios</i>	<b>Rede Federal</b>	<b>Universidades Estaduais</b>	<b>Universidades Federais</b>	<b>Universidades Privadas</b>
<i>ENADE</i>	2	8	6	10
<i>Indicador de Diferença entre Desempenhos Observados e Esperados (IDD)</i>	5	13	12	11
<i>Mestres</i>	1	4	2	2
<i>Doutores</i>	1	6	11	10
<i>Regime de Trabalho</i>	-	3	-	6
<i>Organização Didático-pedagógica (ODP)</i>	-	2	4	4
<i>Infraestrutura e Instalações Físicas</i>	2	4	4	4
<i>Oportunidades de Ampliação da Formação Acadêmica e Profissional (OAF)</i>	-	2	1	5

A partir da Tabela 10, 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 quatro árvores geradas.

A partir disso, pode-se inferir que são os critérios determinantes para o desfecho “Avaliação do CPC”. Este 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 possuem o maior peso no cálculo do CPC, corroborando a extração de conhecimento das árvores geradas.

Assim sendo, na Tabela 11 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 em todas as quatro árvores geradas, confirmando a importância dos mesmos.

Tabela 11 – Critérios de maior impacto presentes nas árvores.

<i>Critérios</i>	<b>Rede Federal</b>	<b>Universidades Estaduais</b>	<b>Universidades Federais</b>	<b>Universidades Privadas</b>
<i>ENADE</i>	2º	1º	1º / 2º	1º / 2º
<i>Indicador de Diferença entre Desempenhos Observados e Esperados (IDD)</i>	1º	2º	2º	2º
<i>Nota de Doutores</i>	-	2º	-	-

**Quadro 4** – Critérios considerados nas Avaliações Nota 2 e Nota 5

IES	Critérios Considerados	
	Nota 2	Nota 5
Rede Federal	<ul style="list-style-type: none"> <li>• IDD</li> <li>• Mestres</li> <li>• Infraestrutura</li> </ul>	-
Universidades Estaduais	<ul style="list-style-type: none"> <li>• Enade</li> <li>• IDD</li> <li>• Mestres</li> <li>• Doutores</li> <li>• Infraestrutura</li> <li>• ODP</li> <li>• Regime de Trabalho</li> </ul>	<ul style="list-style-type: none"> <li>• Enade</li> <li>• IDD</li> <li>• Doutores</li> </ul>
Universidades Federais	<ul style="list-style-type: none"> <li>• Enade</li> <li>• IDD</li> <li>• Mestre</li> <li>• Doutores</li> <li>• ODP</li> </ul>	<ul style="list-style-type: none"> <li>• Enade</li> <li>• IDD</li> <li>• Doutores</li> <li>• Infraestrutura</li> <li>• ODP</li> </ul>
Universidades Privadas	<ul style="list-style-type: none"> <li>• Enade</li> <li>• IDD</li> <li>• Mestre</li> <li>• Doutores</li> <li>• ODP</li> <li>• Regime de Trabalho</li> </ul>	<ul style="list-style-type: none"> <li>• Enade</li> <li>• IDD</li> <li>• Mestres</li> <li>• Doutores</li> <li>• Infraestrutura</li> </ul>

Fonte: Elaboração Própria

O Quadro 4 apresenta um panorama dos critérios considerados nas Avaliações dos extremos (Nota 2 e Nota 5). A partir disso, pode-se concluir que os critérios Mestres e Doutores apareceram em três das quatro árvores geradas para a Avaliação Nota 2, e são critérios importantes que podem impulsionar a nota do CPC dos cursos de graduação, fazendo com que os mesmos consigam desempenho melhores e alcancem Nota 3. Isso é corroborado com o trabalho de Lacerda *et al.* (2015), que através de procedimentos estatísticos demonstram que existem relações significativas dos percentuais de doutores e de mestres com o Enade e o IDD.

No que se refere à Avaliação Nota 5, o critério Doutores apresentou maior significância pois está presente em todas as árvores geradas que tiveram perfis para esta nota, da mesma forma em que o trabalho de Lacerda *et al.* (2015) confirmam que o indicador percentual de doutores foi aquele que apresentou as maiores e mais significativas correlações para todas as combinações avaliadas.

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 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. O curso foi avaliado conforme demonstrado no Quadro 5.

**Quadro 5** – Dados do INEP sobre o curso da Rede Federal Nota 5.

<b>Dados do CPC</b>
<ul style="list-style-type: none"> <li>• Nota Contínua do Enade: 4,75</li> <li>• Nota - Organização Didático-Pedagógica: 3,08</li> <li>• Nota - Infraestrutura e Instalações Físicas: 3,19</li> <li>• Nota - Oportunidades de Ampliação da Formação: 3,34</li> <li>• Nota – IDD: 4,74</li> <li>• Nota – Mestres: 4,72</li> <li>• Nota – Doutores: 3,51</li> <li>• Nota – Regime de Trabalho: 5,0</li> </ul>

Fonte: Elaboração Própria

**Quadro 6 – Síntese dos Critérios relacionados às IES.**

<b>Critérios</b>	<b>Rede Federal</b>	<b>Universidades Estaduais</b>	<b>Universidades Federais</b>	<b>Universidades Privadas</b>
ENADE	<b>Critério determinante:</b> mantém a avaliação do CPC entre 3 e 4.	Critério de maior impacto na avaliação do CPC.	Critério de maior impacto na avaliação do CPC.	Critério de maior impacto na avaliação do CPC.
Indicador de Diferença entre Desempenhos Observados e Esperados (IDD)	Critério de maior impacto na avaliação do CPC.	<b>Critério determinante:</b> mantém a avaliação do CPC entre 4 e 5.	<b>Critério determinante:</b> mantém a avaliação do CPC entre 4 e 5.	<b>Critério determinante:</b> mantém a avaliação do CPC entre 4 e 5.
Nota de Mestres	<b>Critério Significativo:</b> para obter avaliação (Nota 2).	<b>Critério Significativo:</b> para obter avaliação (Nota 2).	<b>Critério Suporte:</b> Se bem avaliado consegue evoluir a Nota do CPC.	<b>Critério Suporte:</b> Se bem avaliado consegue evoluir a Nota do CPC.
Nota de Doutores	<b>Critério determinante:</b> mantém a avaliação do CPC entre 3 e 4.	<b>Critério Significativo:</b> para obter avaliação de Excelência (Nota 5).	<b>Critério determinante:</b> mantém a avaliação do CPC entre 4 e 5.	<b>Critério Significativo:</b> para obter avaliação de Excelência (Nota 5).
Regime de Trabalho	Não apresentou impacto significativo na avaliação.	<b>Critério Suporte:</b> Se bem avaliado consegue evoluir a Nota do CPC.	Não apresentou impacto significativo na avaliação.	<b>Critério Suporte:</b> Se bem avaliado consegue evoluir a Nota do CPC.
Organização Didático-pedagógica (ODP)	Não apresentou impacto significativo na avaliação.	Não apresentou impacto significativo na avaliação.	<b>Critério Suporte:</b> Se bem avaliado consegue evoluir a Nota do CPC.	<b>Critério Suporte:</b> Se bem avaliado consegue evoluir a Nota do CPC.
Infraestrutura e Instalações Físicas	<b>Critério Suporte:</b> Se bem avaliado consegue evoluir a Nota do CPC.	<b>Critério Significativo:</b> para obter avaliação (Nota 2).	<b>Critério Significativo:</b> para obter avaliação de Excelência (Nota 5).	<b>Critério Significativo:</b> para obter avaliação de Excelência (Nota 5).
Oportunidades de Ampliação da Formação Acadêmica e Profissional (OAF)	Não apresentou impacto significativo na avaliação.	Não apresentou impacto significativo na avaliação	Não apresentou impacto significativo na avaliação	Não apresentou impacto significativo na avaliação.

Fonte: Elaboração Própria

O Quadro 6 apresenta uma síntese dos resultados obtidos, relacionando os critérios considerados na avaliação do CPC com as Instituições de Ensino Superior (IES). Pode-se observar os seguintes critérios:

- **Critério de maior Impacto:** critério que impacta diretamente em qualquer perfil dos cursos analisados, ou seja, qualquer avaliação do CPC dependerá inicialmente deste critério. Um exemplo pode ser observado relacionando a Rede Federal com o IDD, nota-se que em qualquer perfil dos cursos de Licenciatura, a avaliação do CPC dependerá prioritariamente de qual foi a Nota resultante do IDD.

- Critério Significativo: critério que impacta diretamente avaliações específicas do CPC, como por exemplo, Nota 2 ou Nota 5. Um exemplo é observado relacionando às Universidades Estaduais com a Nota de Doutores, ou seja, para que seja possível conseguir uma avaliação de excelência (Nota 5), o critério Nota de Doutores precisa receber uma boa avaliação (maior que 2.3).
- Critério Determinante: critério que impacta diretamente nos perfis relacionados às melhores avaliações possíveis do CPC. Um exemplo é observado relacionando às Universidades Privadas com o IDD, ou seja, se este critério for bem avaliado (maior que 4), a avaliação do CPC sempre ficará entre 4 e 5.
- Critério Suporte: critério 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. Um exemplo pode ser observado relacionando as Universidades Federais e a ODP, se o critério possuir uma avaliação maior que 0.7, consegue evoluir a Nota de CPC de 2 para 3.

Vale ressaltar que alguns critérios não apresentaram impacto significativo para determinadas Instituições de Ensino, destacando-se o critério “Oportunidades de Ampliação da Formação Acadêmica e Profissional” que não apresentou impacto em nenhuma das IES analisadas.

## 6. CONCLUSÕES

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, 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 a nota obtida na avaliação do CPC.

Como resultados de maior importância, pode ser destacada a predominância dos critérios Indicador da Diferença entre os Desempenhos Observados e Esperados (IDD) e a Nota do Enade para todos os cursos das IES analisados. Ressalta-se que os critérios considerados no presente trabalho são os mesmos para todas as instituições, porém estes apresentaram impactos diferentes para cada tipo de IES.

Para a Rede Federal, a nota do Indicador da Diferença entre os Desempenhos Observados e Esperados (IDD) e a Nota do Enade foram os critérios que mais influenciaram na avaliação do CPC, ambos aspectos relacionados diretamente com o aluno. Os critérios Infraestrutura e Nota de Doutores apresentaram-se como alicerces fundamentais para um bom desempenho na avaliação.

Nas Universidades Estaduais, os critérios que mais impactaram na Avaliação foram a nota do Indicador da Diferença entre os Desempenhos Observados e Esperados (IDD) e a Nota do Enade. No que se refere à avaliação Nota 2, Infraestrutura e Nota de Mestres são critérios que se bem avaliados podem auxiliar os cursos a evoluírem a Nota do CPC, visto que são critérios que dependem apenas da atuação da Universidade. Já na avaliação nota 5, além dos critérios nota do IDD e nota do Enade, é necessária uma boa atuação na Nota de Doutores.

Nas Universidades Federais, além da nota do IDD e do Enade figurarem os critérios de maior importância, a nota da Organização Didático-pedagógica (ODP) também foi um critério determinante nas duas avaliações. Além disso, no que tange a avaliação Nota 2, os critérios Nota de Mestres e Nota de Doutores podem contribuir para a evolução da nota do CPC dos cursos. A nota de Infraestrutura e Doutores são critérios que não dependem da atuação do discente e são imprescindíveis para alcançarem a avaliação de Excelência (Nota 5).

Nas Universidades Privadas, observa-se um número maior de critérios que impactam na Avaliação, assim, a obtenção de bons índices apenas na nota do IDD e do Enade não são garantia de uma boa avaliação no CPC. Na avaliação Nota 2, além da Nota do Enade e do IDD, a Organização Didático-pedagógica (ODP) e o Regime de Trabalho são critérios importantes. Os critérios Nota de mestres e Doutores combinados com a Infraestrutura merecem destaque para obtenção de uma nota de Excelência.

Baseado nestes resultados, espera-se que a partir dos padrões e conhecimentos extraídos e apresentados, seja possível auxiliar os gestores nas suas tomadas de decisão, 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 2014.

### **6.1. Contribuições**

Acredita-se que a principal contribuição dessa pesquisa é explicitação dos perfis que descrevem os cursos de Licenciatura das IES analisadas e conseqüentemente o subsídio de informações para os gestores dos cursos de graduação. Uma outra contribuição está relacionada a 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 esta uma modalidade de ensino pouco explorada na mineração de dados educacionais.

### **6.2. Trabalhos Futuros**

Como trabalhos futuros propõe-se a utilização de outras bases de conhecimento relacionadas a avaliação do Ensino Superior, o que resultaria em um aumento da amostra da pesquisa. Além disso, pode-se realizar outras análises relacionadas com os dados provenientes da pesquisa, por exemplo, a relação do perfil do aluno com a nota obtida no CPC. Outra análise importante que pode ser realizada é a respeito da nota do Enade em relação a nota do CPC, realizando um estudo mais aprofundado da relação entre os dois.

Vale ressaltar que o estudo foi baseado na Avaliação do CPC do ano de 2014, pois quando iniciada a pesquisa, foi a mais atual encontrada na plataforma do INEP para os cursos de Licenciatura. Sugere-se que o método seja replicado com a utilização das próximas avaliações, previstas para o final do ano de 2018.

Uma outra aplicação seria a seleção de outros atributos além dos determinantes para a nota do CPC, previamente a fase de mineração de dados, com o intuito de verificar outras

abordagens e analisar se houve melhoria nas taxas de acerto obtidas. Ademais, com o mesmo intuito, poderá ser investigada a utilização de outros métodos de classificação.

## REFERÊNCIAS BIBLIOGRÁFICAS

AHUJA, R. *et al.* Analysis of Educational Data Mining. *In: YADAV, N. et al.* (Eds.). **Harmony Search and Nature Inspired Optimization Algorithms**. Singapore: Springer Singapore, 2019. v. 741p. 897–907.

BABY, A. M. Pedagogue Performance Assessment (PPA) using Data mining Techniques. **IOP Conference Series: Materials Science and Engineering**, v. 396, p. 012024, 29 ago. 2018.

BAKER, R.; ISOTANI, S.; CARVALHO, A. Mineração de Dados Educacionais: Oportunidades para o Brasil. **Revista Brasileira de Informática na Educação**, v. 19, n. 02, 31 ago. 2011.

BITTENCOURT, H. R. *et al.* Mudanças nos pesos do CPC e seu impacto nos resultados de avaliação em universidades federais e privadas. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 15, n. 3, p. 147–166, 2010.

BRASIL. Lei de Diretrizes e Bases da Educação Nacional. Lei número 9394, 20 de dezembro de 1996.

\_\_\_\_\_. 11.892. Ministério da Educação. Escassez de Professores no Ensino Médio: Propostas estruturais e Emergenciais. 2007.

\_\_\_\_\_. 11.892. 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. 2008.

\_\_\_\_\_. **Expansão da Rede Federal**. Disponível em: <<http://redefederal.mec.gov.br/expansao-da-rede-federal>>. Acesso em: 10 maio. 2017.

CAZZOLATO, M. T.; VIEIRA, M. T. P. Inclusão de um novo algoritmo de classificação na ferramenta Kira. p. 5, 2011.

CHUEKE, G. V.; AMATUCCI, M. O que é bibliometria? Uma introdução ao Fórum. **Internext**, v. 10, n. 2, p. 1, 9 set. 2015.

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.

CORDEIRO, R. G. **Identificação do comportamento dos estudantes evadidos de cursos técnicos utilizando técnicas de mineração de dados**. Campos dos Goytacazes: Instituto Federal de Educação, Ciência e Tecnologia Fluminense, 2017.

CORDEIRO, R. G.; MUSSA, M. DE S.; HORA, H. R. M. DA. Modelo para análise do perfil dos alunos evadidos em curso técnico utilizando técnicas de mineração de dados. 2017.

CRETTON, N. N. **Mineração de Dados aplicado na Base do Enade com enfoque na criação de Perfis dos Estudantes que prestaram o Exame utilizando o algoritmo J48**. Campos dos Goytacazes: Universidade Cândido Mendes, 2016.

CRETTON, N. N.; GOMES, G. R. Aplicação de Técnicas de Mineração de Dados na base de dados do ENADE com enfoque nos cursos de Medicina. **Acta Biomédica Brasiliensia**, v. 7, n. 1, p. 74, 20 jun. 2016.

SOUSA, J. V. Internacionalização da Educação Superior como indicador do Sinaes: de qual qualidade estamos falando? **Educação**, v. 40, n. 3, p. 343, 31 dez. 2017.

DENG, J. *et al.* **A Study of Teaching Evaluation in Adult Higher Education Based on Decision Tree** 2010 Second International Conference on Information Technology and Computer Science. **Anais... In: 2010 2ND INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY AND COMPUTER SCIENCE (ITCS 2010)**. Kiev, Ukraine: IEEE, jul. 2010 Disponível em: <<http://ieeexplore.ieee.org/document/5557350/>>. Acesso em: 3 dez. 2018

EASHWAR, K. B.; VENKATESAN, R.; GANESH, D. Student performance prediction using SVM. **International Journal of Mechanical Engineering and Technology.**, v. 8, n. 17, p. 649–662, 2017.

FERREIRA, G. **Investigação acerca dos fatores determinantes para a conclusão do Ensino Fundamental utilizando Mineração de Dados Educacionais no Censo Escolar da Educação Básica do INEP 2014**. *In: II WORKSHOP DE MINERAÇÃO DE DADOS EDUCACIONAIS*. Maceió, Alagoas, Brasil: 2015. Disponível em: <<http://br-ie.org/pub/index.php/wcbie/article/view/6211>>. Acesso em: 30 ago. 2018

FONSECA, S. O. DA; NAMEN, A. A. Mineração em Bases de Dados do INEP: Uma Análise Exploratória para Nortear melhorias no Sistema Educacional Brasileiro. **Educação em Revista**, v. 32, n. 1, p. 133–157, mar. 2016.

FORPLAN. **Metodologia da Matriz Orçamentária da Rede de Ensino Profissional e Tecnológico de 2018**, 2018. Disponível em: <<https://www2.ifal.edu.br/ifal/reitoria/pro-reitorias/administracao/orcamento-e-contabilidade/orcamento/comissao-fordap-de-gestao-orcamentaria/material-apresentado-nas-visitas-aos-campi-1o-semester-2018/metodologia-da-matriz-conif-2018.pdf/view>>

FRANÇA, R. S. DE; AMARAL, H. J. C. DO. Mineração de Dados na Identificação de Grupos de Estudantes com Dificuldades de Aprendizagem no Ensino de Programação. **RENOTE - Revista Novas Tecnologias na Educação**, v. 11, n. 1, 2013.

FRANK, E. *et al.* Weka-A Machine Learning Workbench for Data Mining. *In: MAIMON, O.; ROKACH, L. (Eds.). Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer US, 2009. p. 1269–1277.

GATTI, B. Formação de professores: licenciaturas, currículos e políticas. **movimento-revista de educação**, n. 2, 31 ago. 2015.

GATTI, B. A. Formação de professores: condições e problemas atuais. **Revista Internacional de Formação de Professores**, v. 1, n. 2, p. 161-171–171, 23 maio 2016.

GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2009.

GONÇALVES, R. F. *et al.* Avaliação dos cursos de Fisioterapia nos anos de 2004 a 2013. **Fisioterapia e Pesquisa**, v. 24, n. 4, p. 392–398, dez. 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.

HAN, J.; KAMBER, M.; PEI, J. **Data mining: concepts and techniques**. Burlington: Elsevier Science, 2012.

HARRISON, P. A. *et al.* Selecting methods for ecosystem service assessment: A decision tree approach. **Ecosystem Services**, v. 29, p. 481–498, fev. 2018.

INEP. **Nota Técnica Daes/Inep nº 58/2015 - Cálculo do Conceito Preliminar de Curso 2014**, 2015. Disponível em: <[http://download.inep.gov.br/educacao\\_superior/enade/notas\\_tecnicas/2014/nota\\_tecnica\\_daes\\_n582015\\_calculo\\_do\\_cpc2014.pdf](http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2014/nota_tecnica_daes_n582015_calculo_do_cpc2014.pdf)>

\_\_\_\_\_. **NOTA TÉCNICA Nº 3/2017/CGCQES/DAES**, 2017. Disponível em: <[http://download.inep.gov.br/educacao\\_superior/enade/notas\\_tecnicas/2015/nota\\_tecnica\\_daes\\_n32017\\_calculo\\_do\\_cpc2015.pdf](http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2015/nota_tecnica_daes_n32017_calculo_do_cpc2015.pdf)>

KAMAL, P.; AHUJA, S. Academic Performance Prediction Using Data Mining Techniques: Identification of Influential Factors Effecting the Academic Performance in Undergrad Professional Course. *In: YADAV, N. et al. (Eds.). Harmony Search and Nature Inspired Optimization Algorithms*. Singapore: Springer Singapore, 2019. v. 741p. 835–843.

LACERDA, L. L. V. DE *et al.* 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**, v. 96, n. 242, p. 129–145, abr. 2015.

MACCUE, C. **Data mining and predictive analysis: intelligence gathering and crime analysis**. Amsterdam: Elsevier/Butterworth-Heinemann, 2007.

MACHADO, R. D. *et al.* Estudo Bibliométrico em Mineração de Dados e Evasão Escolar. 2015.

MANHÃES, L. M. B. *et al.* Previsão de Estudantes com Risco de Evasão Utilizando Técnicas de Mineração de Dados. **Anais do XXII SBIE - XVII WIE**, p. 150–159, 2011.

MANTAS, C. J.; ABELLÁN, J. Credal-C4.5: Decision tree based on imprecise probabilities to classify noisy data. **Expert Systems with Applications**, v. 41, n. 10, p. 4625–4637, ago. 2014.

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 Disponível em: <<http://br-ie.org/pub/index.php/sbie/article/view/8194>>. Acesso em: 4 nov. 2018.

MEC. **Conceito Preliminar de Curso (CPC)**, 2015a. Disponível em: <<http://portal.inep.gov.br/conceito-preliminar-de-curso-cpc->>

\_\_\_\_. **ENADE**, 2015b. Disponível em: <<http://portal.inep.gov.br/enade?>>

\_\_\_\_. **Censo da Educação Superior**, 2015c.

\_\_\_\_. **Resolução CNE/CP nº 02/15**, 2015d.

MOURA, A. G. DE. **Uma Proposta de Melhoria de Processos para os Institutos Federais**. Campos dos Goytacazes: Instituto Federal de Educação, Ciência e Tecnologia Fluminense, 2018.

OLIVEIRA, B. M.; OLIVEIRA, M. R. N. S. Licenciatura nos Institutos Federais: Aspectos para Discussão. **Revista Brasileira da Educação Profissional e Tecnológica**, v. 1, n. 10, p. 22, 15 jun. 2016.

PEREIRA, C. W. DOS S.; MORORÓ, L. P. A Expansão dos Institutos Federais e dos Cursos de Licenciaturas no Brasil. v. 12, n. 1, 2017.

QUINLAN, J. R. Simplifying decision trees. **International Journal of Man-Machine Studies**, v. 27, n. 3, p. 221–234, set. 1987.

RAMAPHOSA, K. I. M.; ZUVA, T.; KWUIMI, R. **Educational Data Mining to Improve Learner Performance in Gauteng Primary Schools**. 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD). **Anais**. In: INTERNATIONAL CONFERENCE ON ADVANCES IN BIG DATA, COMPUTING AND DATA COMMUNICATION SYSTEMS (ICABCD). Durban, South Africa: IEEE, ago. 2018. Disponível em: <<https://ieeexplore.ieee.org/document/8465478/>>. Acesso em: 9 nov. 2018.

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, 2012.

\_\_\_\_\_. 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.

ROIGER, R. J. **Data mining: a tutorial-based primer**. Boca Raton: Taylor & Francis, CRC Press, 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.

SANKARI, E. S.; MANIMEGALAI, D. Predicting membrane protein types using various decision tree classifiers based on various modes of general PseAAC for imbalanced datasets. **Journal of Theoretical Biology**, v. 435, p. 208–217, dez. 2017.

SILVA, E. L.; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. 4. ed. Florianópolis: UFSC, 2005.

SUKHIJA, K.; JINDAL, M.; AGGARWAL, N. **The recent state of educational data mining: A survey and future visions** 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE). **Anais**. In: 2015 IEEE 3RD INTERNATIONAL CONFERENCE ON MOOCs, INNOVATION AND TECHNOLOGY IN EDUCATION (MITE). Amritsar, India: IEEE, out. 2015 Disponível em: <<http://ieeexplore.ieee.org/document/7375344/>>. Acesso em: 3 set. 2018.

UNDAVIA, J. N.; DOLIA, P.; PATEL, A. Customized Prediction Model to Predict Post-Graduation Course for Graduating Students Using Decision Tree Classifier. **Indian Journal of Science and Technology**, v. 9, n. 12, 30 mar. 2016.

USAMA FAYYAD; GREGORY PIATETSKY-SHAPIRO; PADHRAIC SMYTH. From Data Mining to Knowledge Discovery in Databases. v. 17, n. 3, p. 37, 1996.

WAGNER, M. M.; RIBEIRO, V.; ZABADAL, J. Mineração de Dados Meteorológicos Empregando Dados de Temperatura: o caso de uma cidade gaúcha. **Interfaces Científicas - Exatas e Tecnológicas**, v. 2, n. 2, p. 57, 11 out. 2016.

WU, F. **Apply Data Mining to Students' Choosing Teachers Under Complete Credit Hour** 2010 Second International Workshop on Education Technology and Computer Science. **Anais...** *In:* 2010 SECOND INTERNATIONAL WORKSHOP ON EDUCATION TECHNOLOGY AND COMPUTER SCIENCE. Wuhan, China: IEEE, 2010 Disponível em: <<http://ieeexplore.ieee.org/document/5458759/>>. Acesso em: 3 dez. 2018.

ZAKI, M. J.; MEIRA JR, W. **Data Mining and Analysis: Fundamental Concepts and Algorithms**. 1. ed. [s.l.] Cambridge University Press, 2014.

## APÊNDICE I – MATRIZES DE CONFUSÃO

### Nota Final (*cross-validation*)

Classificado como:	Institutos Federais					Universidades Estaduais					Universidades Federais					Universidades Privadas				
	N1	N2	N3	N4	N5	N1	N2	N3	N4	N5	N1	N2	N3	N4	N5	N1	N2	N3	N4	N5
<b>Nota 1</b>	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>Nota 2</b>	0	10	5	0	0	0	60	27	0	0	0	14	13	0	0	0	25	15	0	0
<b>Nota 3</b>	0	5	79	8	0	0	22	350	27	0	0	8	316	38	0	0	7	329	42	0
<b>Nota 4</b>	0	0	7	21	0	0	0	27	119	3	0	0	37	318	10	0	0	50	209	11
<b>Nota 5</b>	0	0	0	1	0	0	0	0	4	1	0	0	0	11	15	0	0	0	10	14

## APÊNDICE II – ÁRVORES DE DECISÃO

### *Universidades Estaduais*

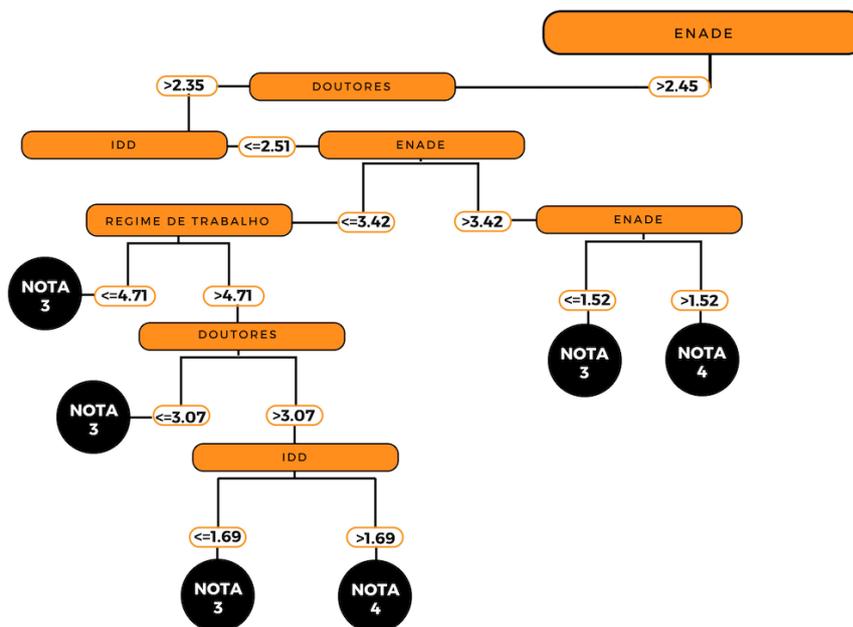


Figura 33 – Árvore de Decisão Universidades Estaduais (f).

Fonte: Elaboração própria

### *Universidades Federais*

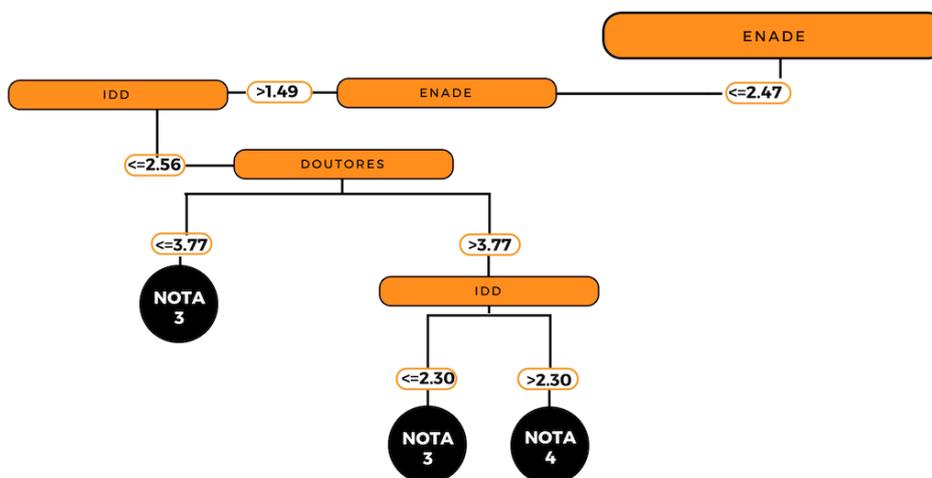


Figura 34 – Árvore de Decisão Universidades Federais (e).

Fonte: Elaboração própria

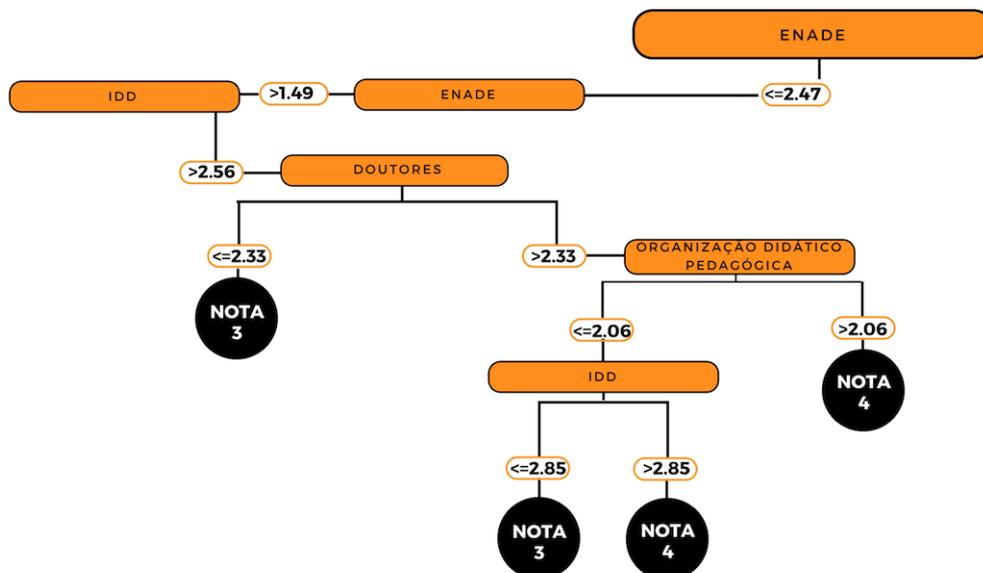


Figura 35 – Árvore de Decisão Universidades Federais (f).

Fonte: Elaboração própria

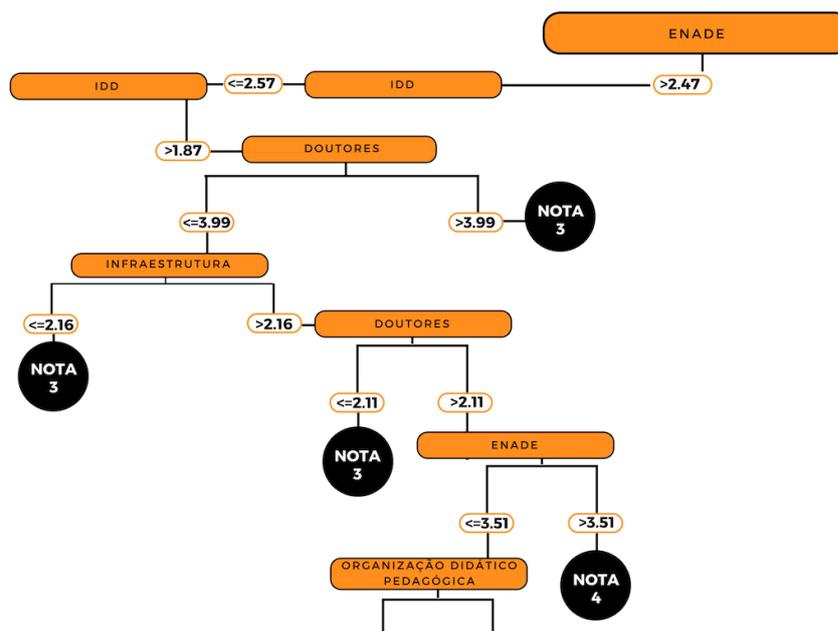


Figura 36 – Árvore de Decisão Universidades Federais (g).

Fonte: Elaboração própria

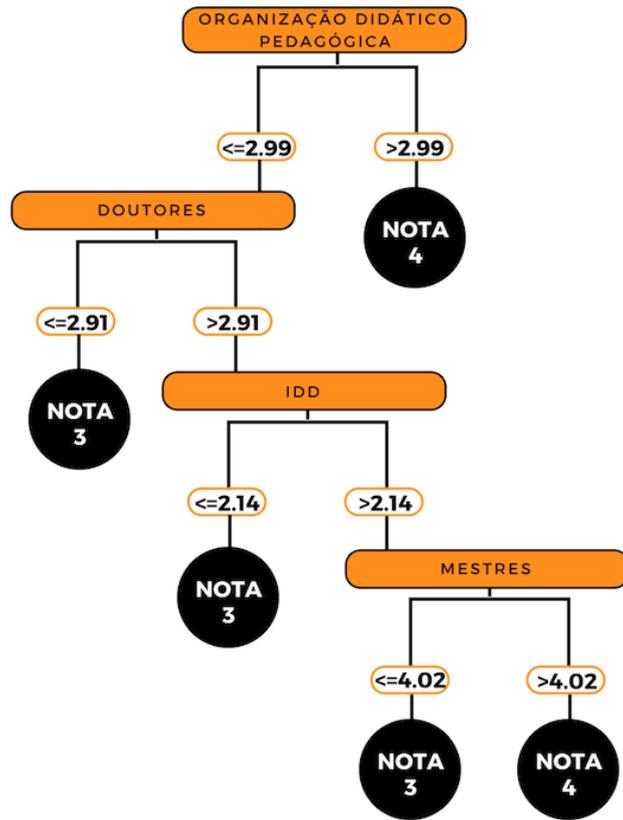


Figura 37 – Árvore de Decisão Universidades Federais (h).

Fonte: Elaboração própria

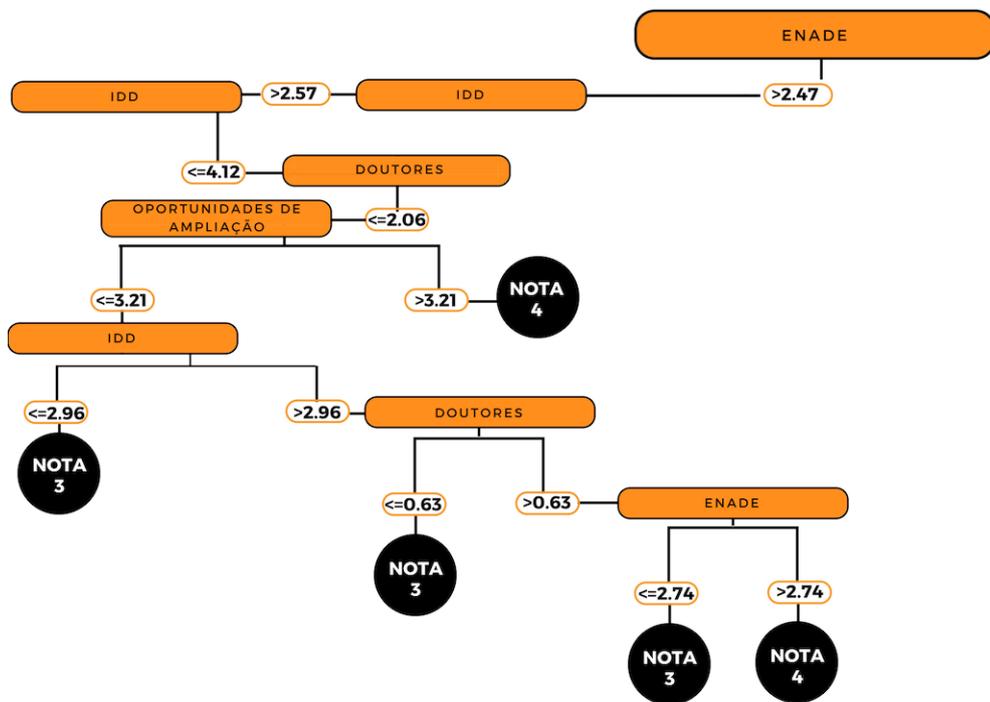


Figura 38 – Árvore de Decisão Universidades Federais (i).

Fonte: Elaboração própria

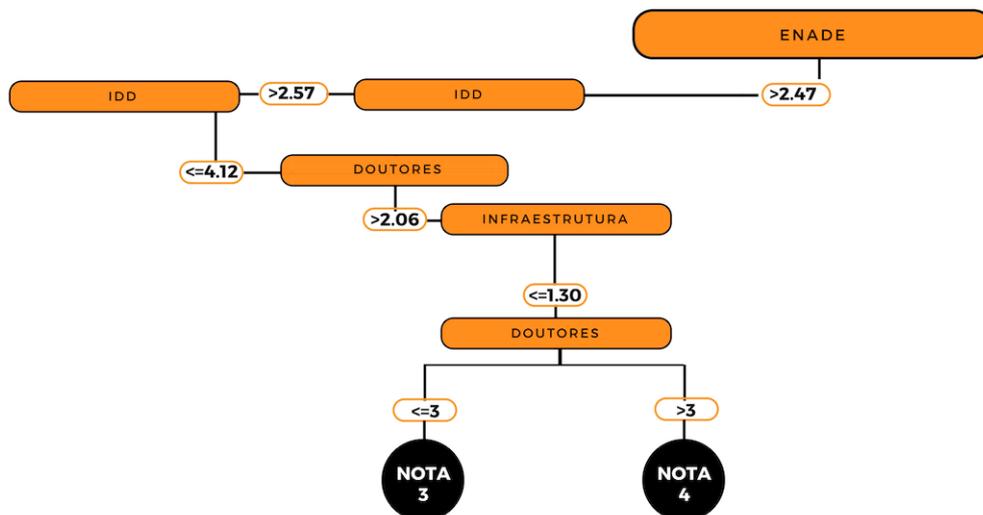


Figura 39 – Árvore de Decisão Universidades Federais (j).

Fonte: Elaboração própria

### Universidades Privadas

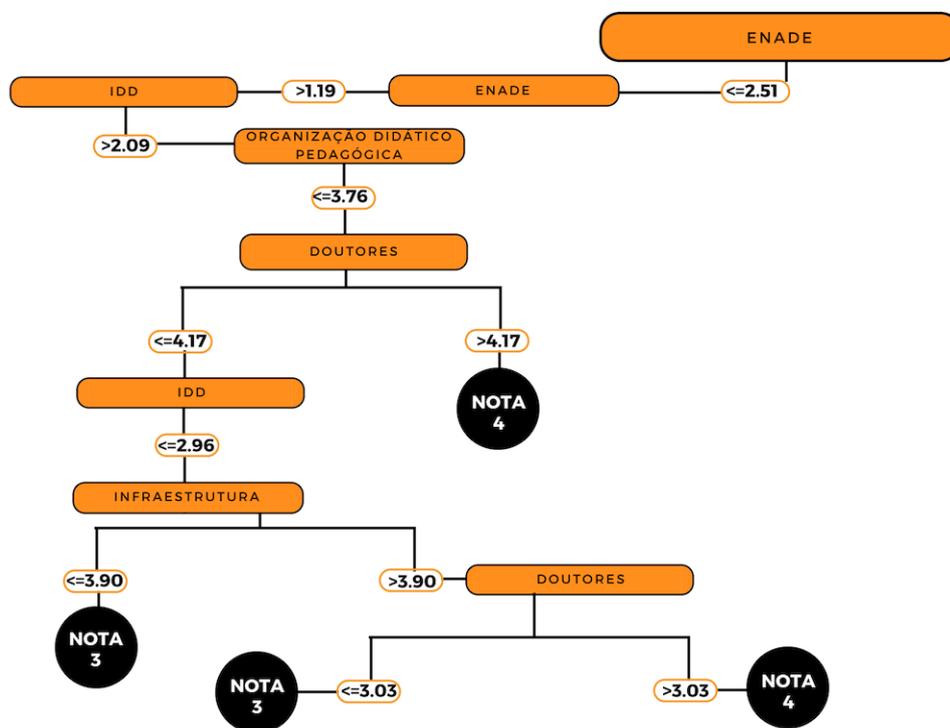


Figura 40 – Árvore de Decisão Universidades Privadas (e).

Fonte: Elaboração própria

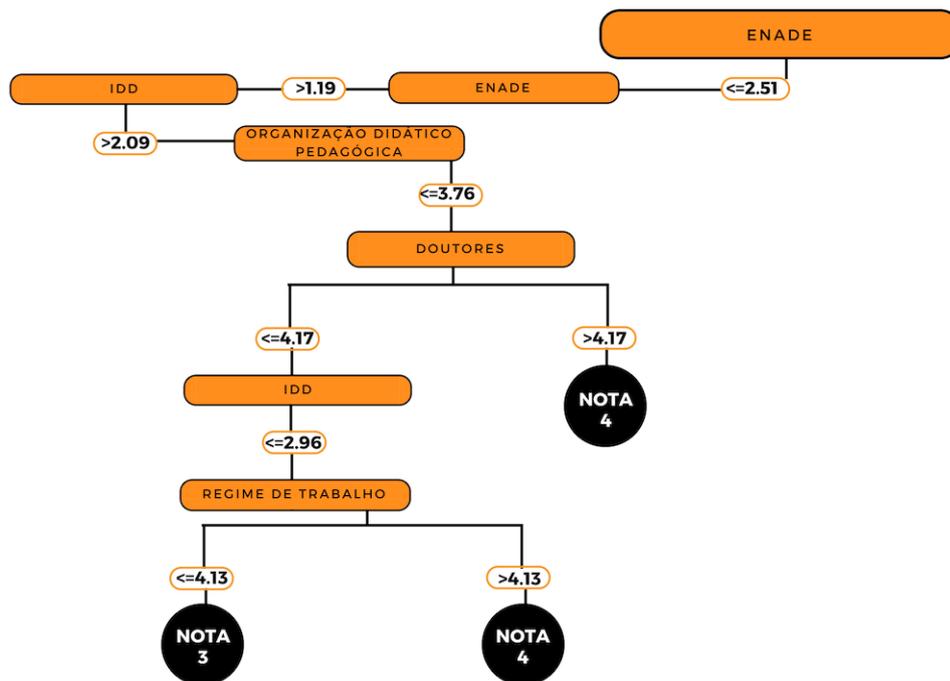


Figura 41 – Árvore de Decisão Universidades Privadas (f).

Fonte: Elaboração própria

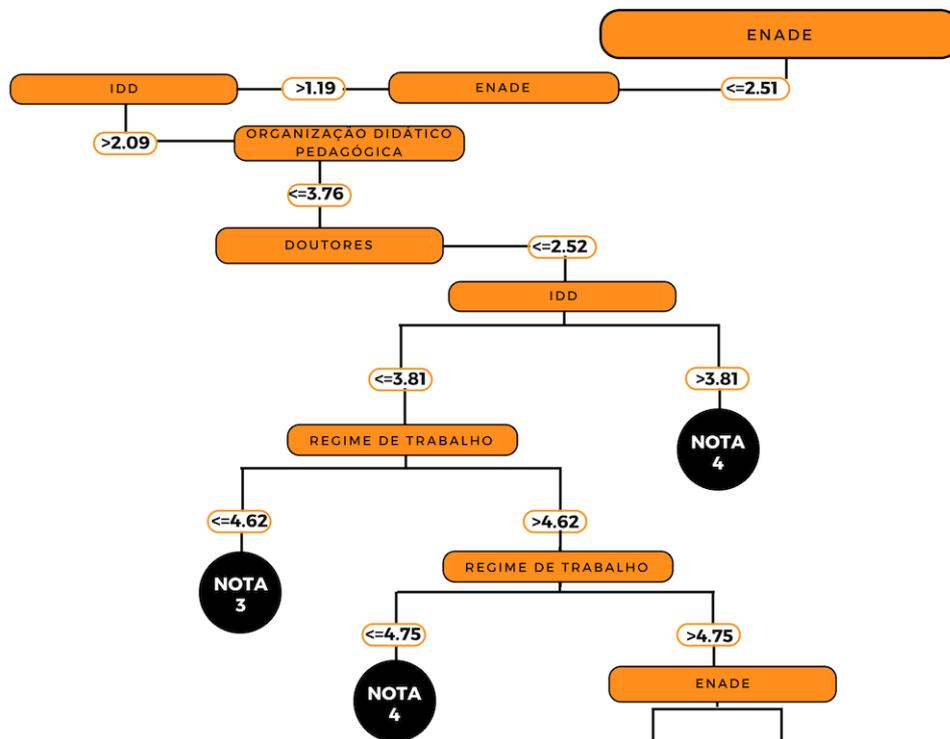


Figura 42 – Árvore de Decisão Universidades Privadas (g).

Fonte: Elaboração própria

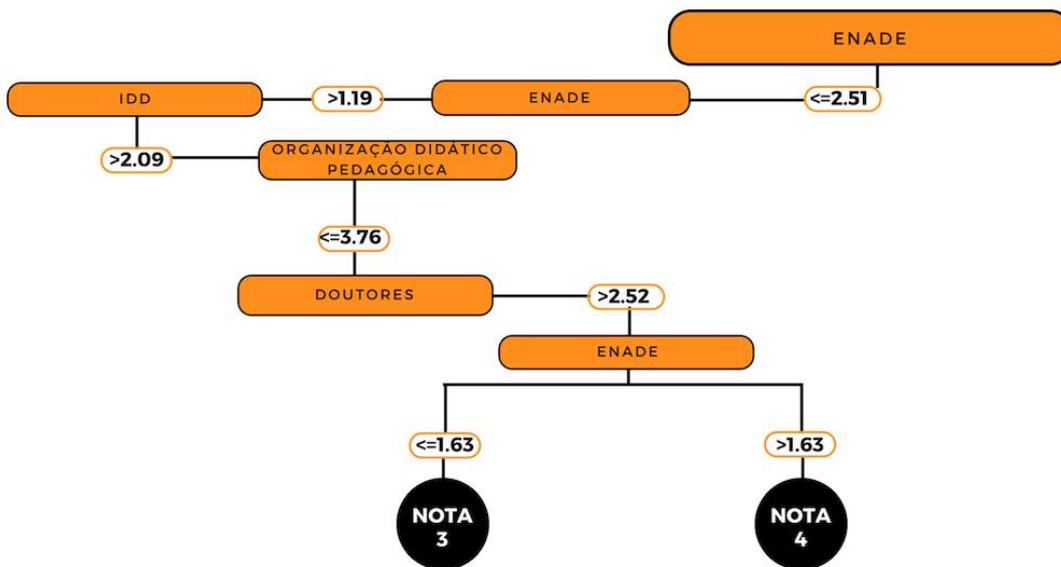


Figura 43 – Árvore de Decisão Universidades Privadas (h).

Fonte: Elaboração própria

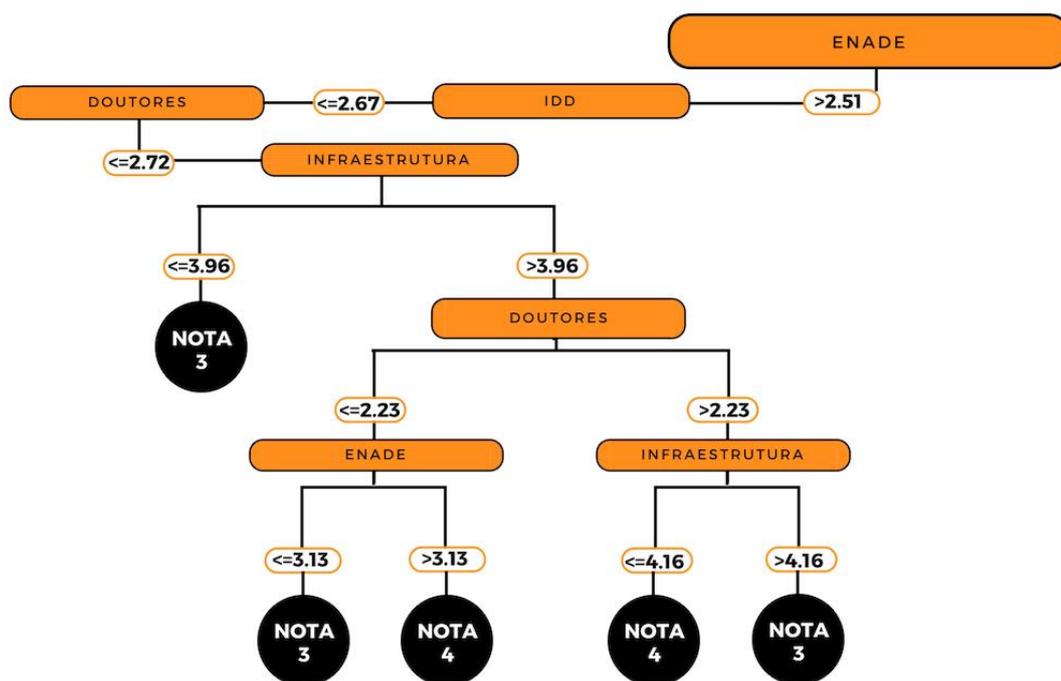


Figura 44 – Árvore de Decisão Universidades Privadas (i).

Fonte: Elaboração própria

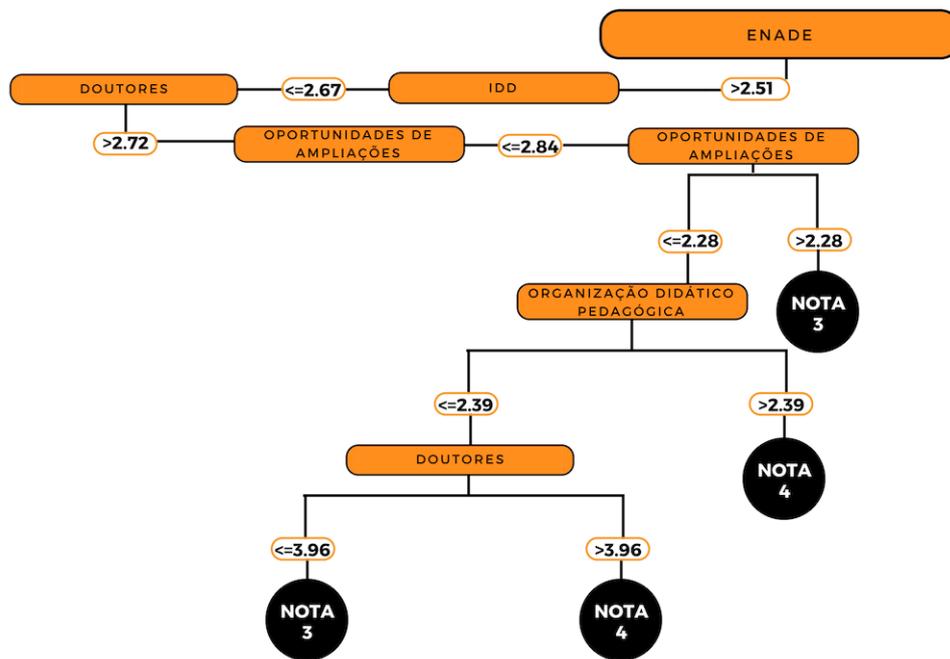


Figura 45 – Árvore de Decisão Universidades Privadas (j).

Fonte: Elaboração própria

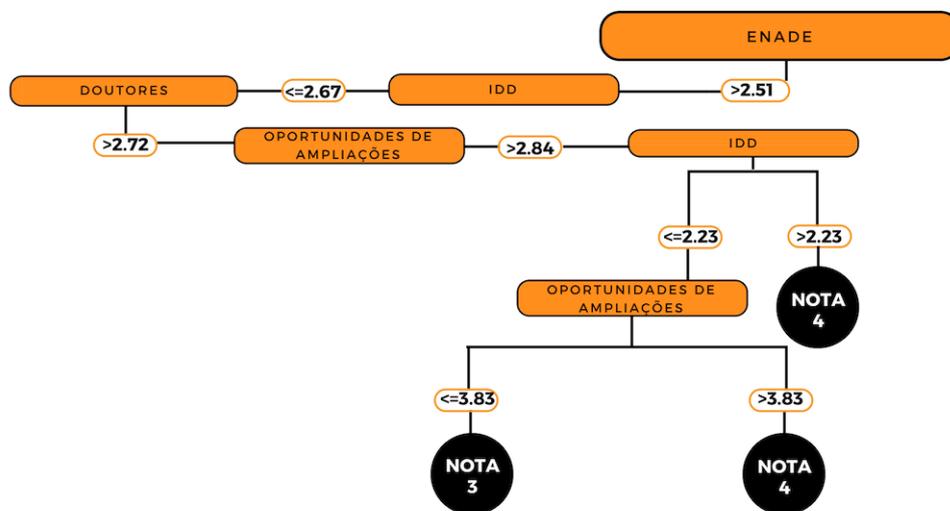


Figura 46 – Árvore de Decisão Universidades Privadas (k).

Fonte: Elaboração própria

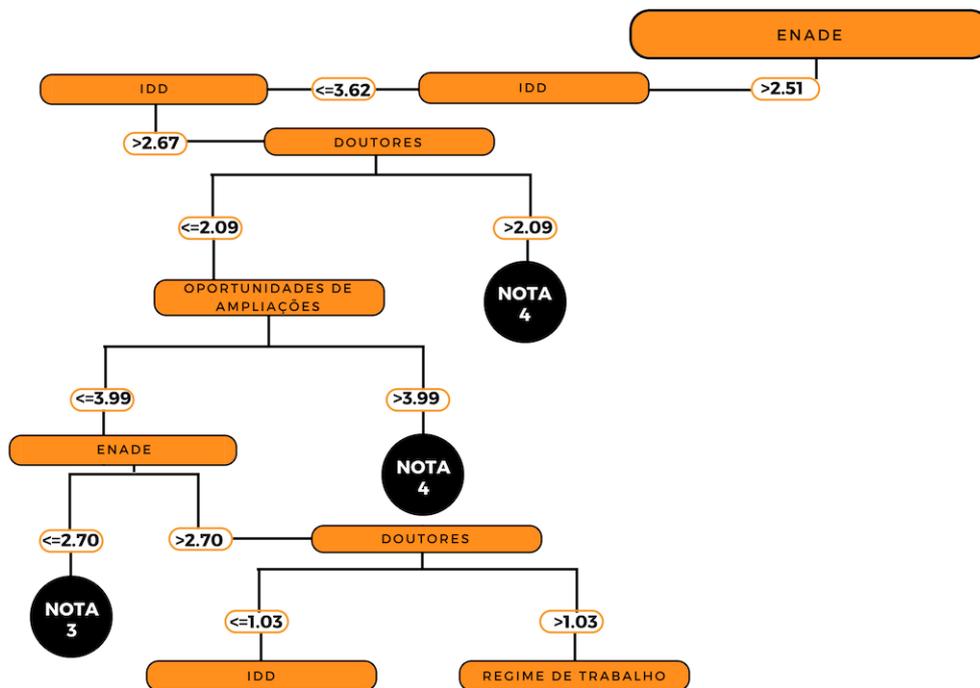


Figura 47 – Árvore de Decisão Universidades Privadas (l).

Fonte: Elaboração própria

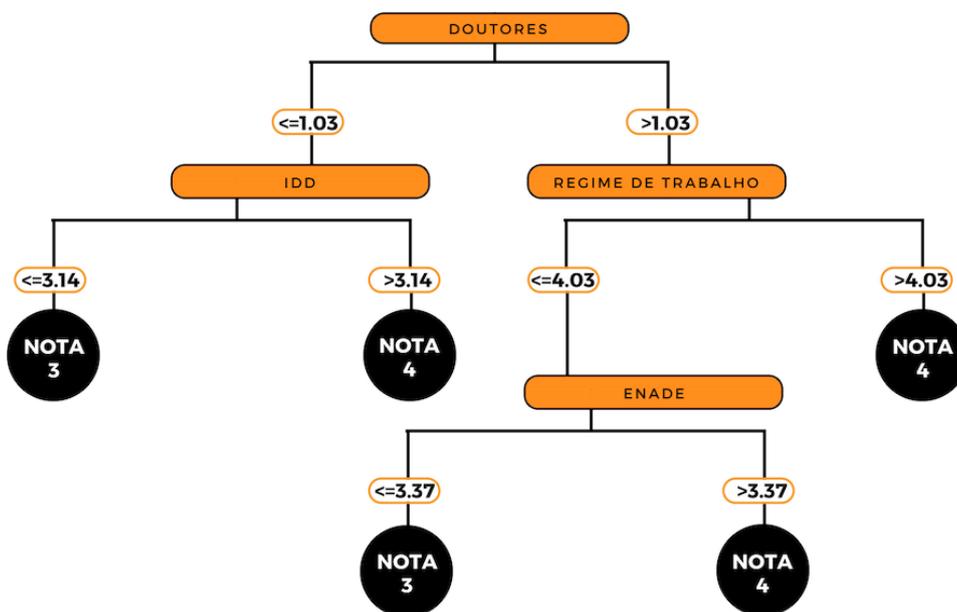


Figura 48 – Árvore de Decisão Universidades Privadas (m).

Fonte: Elaboração própria