

UNIVERSIDADE DE CAXIAS DO SUL  
ÁREA DO CONHECIMENTO DE CIÊNCIAS EXATAS E ENGENHARIAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA  
MESTRADO PROFISSIONAL EM ENGENHARIA MECÂNICA

JONAS MIGUEL DALL AGNOL

INTELIGÊNCIA ARTIFICIAL APLICADA AO RECONHECIMENTO DE PADRÕES  
DA EVASÃO DE ESTUDANTES DA ÁREA DE CIÊNCIAS EXATAS EM UMA  
INSTITUIÇÃO DE ENSINO SUPERIOR

CAXIAS DO SUL

2019

JONAS MIGUEL DALL AGNOL

INTELIGÊNCIA ARTIFICIAL APLICADA AO RECONHECIMENTO DE PADRÕES  
DA EVASÃO DE ESTUDANTES DA ÁREA DE CIÊNCIAS EXATAS EM UMA  
INSTITUIÇÃO DE ENSINO SUPERIOR

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Mecânica da Universidade de Caxias do Sul, visando à obtenção do título de Mestre em Engenharia Mecânica.

Área de concentração: Processos de Fabricação e Produção de Componentes, Ferramentas e Matrizes.

Orientador: Prof. Dr. Leandro Luís Corso

Coorientador: Prof. Dr. Marcos Alexandre Luciano

CAXIAS DO SUL

2019

Dados Internacionais de Catalogação na Publicação (CIP)  
Universidade de Caxias do Sul  
Sistema de Bibliotecas UCS - Processamento Técnico

D144i Dall Agnol, Jonas Miguel

Inteligência artificial aplicada ao reconhecimento de padrões da evasão de estudantes da área de ciências exatas em uma instituição de ensino superior / Jonas Miguel Dall Agnol. – 2019.

73 f. : il. ; 30 cm

Dissertação (Mestrado) - Universidade de Caxias do Sul, Programa de Pós-Graduação em Engenharia Mecânica, 2019.

Orientação: Leandro Luís Corso.

Coorientação: Marcos Alexandre Luciano.

1. Redes neurais (Computação). 2. Inteligência artificial. 3. Evasão universitária. I. Corso, Leandro Luís, orient. II. Luciano, Marcos Alexandre, coorient. III. Título.

CDU 2. ed.: 004.032.26

Catalogação na fonte elaborada pela(o) bibliotecária(o)  
Paula Fernanda Fedatto Leal - CRB 10/2291

***“INTELIGÊNCIA ARTIFICIAL APLICADA AO RECONHECIMENTO DE  
PADRÕES DA EVASÃO DE ESTUDANTES DA ÁREA DE CIÊNCIAS EXATAS EM  
UMA INSTITUIÇÃO DE ENSINO SUPERIOR”***

**Jonas Miguel Dall Agnol**

Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Mecânica da Universidade de Caxias do Sul, como parte dos requisitos necessários para a obtenção do título de Mestre em Engenharia Mecânica.

Área de Concentração: Projeto e Fabricação.

Caxias do Sul, 30 de agosto de 2019.

Banca Examinadora:

Prof. Dr. Leandro Luís Corso

Orientador

Universidade de Caxias do Sul

Prof. Dr. Marcos Alexandre Luciano

Coorientador

Universidade de Caxias do Sul

Prof. Dr. Alexandre Fassini Michels

Universidade de Caxias do Sul

Prof. Dr. Alexandre Vieceli

Universidade de Caxias do Sul

Prof. Dr. Guilherme Holsbach Costa

Universidade de Caxias do Sul

À Deus, por ter me dado forças  
para chegar até aqui.

## AGRADECIMENTOS

A Deus, por ter me dado condições, desafios e aprendizados necessários para chegar até este momento.

Aos meus pais, Ivo Dall Agnol e Joaneta Dall Agnol, que me ensinaram que a ética, o empenho e o trabalho sério fazem e farão sempre a diferença na vida.

A minha companheira, Viviane Vidal, por estar ao meu lado durante este período, incentivando e ajudando sempre que possível.

Ao meu orientador e irmão de caminhada, Leandro Corso, por ser exemplo de caráter e empenho e por ter feito a diferença sempre que precisei.

Ao meu coorientador, Marcos Luciano, pelo apoio e auxílio em todas as etapas deste estudo e por ser exemplo de coragem e retidão.

Aos professores da banca Alexandre Michels, Alexandre Vieceli e Guilherme Costa por terem aceitado e contribuído com este trabalho.

Aos demais familiares e amigos, pela compreensão nas inúmeras ausências que se fizeram necessárias durante este período e pelas palavras de apoio em todos os momentos.

A todos os professores e funcionários da Universidade de Caxias do Sul, em especial aos prof. César Augusto Bernardi, Marcelo Rossato e Nilda Stecanela, respectivamente Diretor Administrativo e Financeiro, Pró-Reitor Acadêmico e Pró-Reitora de Pós-Graduação e Pesquisa em 2016, que acreditaram em meu potencial e possibilitaram a transformação do prêmio Vocação para o Sucesso obtido no mesmo ano junto a Câmara de Indústria e Comércio de Caxias do Sul em desconto para cursar o Programa.

“Você precisa fazer aquilo que pensa  
que não é capaz de fazer”.

**Eleanor Roosevelt**

## RESUMO

Depois de um forte crescimento nas últimas duas décadas o mercado do ensino superior passa por um período de estagnação, sendo necessária, neste contexto, a retenção dos acadêmicos que já fazem parte de sua base de estudantes. Este trabalho contribui neste cenário por meio da identificação dos acadêmicos da Área de Ciências Exatas e Engenharias de uma Instituição de Ensino Superior da Serra Gaúcha que possuem maior propensão à evasão dos estudos, utilizando sua base de dados e a Inteligência Artificial, com o modelo Redes Neurais Artificiais (RNAs). Primeiramente foram mapeados aspectos financeiros, acadêmicos e sociais disponibilizados de 2006 a 2018, seguindo pela etapa de tratamento de dados, na qual apenas as interrupções causadas por troca de currículo foram descartadas. Os dados foram então processados em uma RNA de topologia *perceptron*, de tal maneira a obter acurácia de 93% na identificação do padrão de tendência à evasão dos estudos (conjunto de validação). Também foram realizados processamentos por tipo de graduação, que apresentam resultados igualmente satisfatórios em termos de reconhecimento de padrões. Soma-se a esta avaliação o comportamento adequado da curva ROC (*Receiver Operating Characteristic*) e os baixos índices de correlação entre fatores e matriz-alvo que comprovam, desta maneira, que a técnica RNAs de topologia *backpropagation* pode ser uma importante aliada nas tomadas de decisão de instituições no que tange a gestão da retenção discente.

**Palavras-chave:** Redes Neurais Artificiais, *backpropagation*, evasão.

## ABSTRACT

After a strong growth in the last two decades the market of the higher education goes through a period of stagnation, being necessary, in this context, the retention of the academics that already are part of its base of students. This work contributes in this scenario through the identification of the academics of the Area of Exact Sciences and Engineering of a Higher Education Institution of the Serra Gaúcha who are more likely to avoid studies using their database and Artificial Intelligence, with the Neural Networks model Artificial (RNAs). Firstly, financial, academic and social aspects were mapped from 2006 to 2018, followed by the data processing stage, where only the interruptions caused by curriculum change were discarded. The data were then processed in a perceptron topology RNA, in order to obtain a 93% accuracy in the identification of the tendency pattern to evasion of the studies (validation set). Also performed were by type of graduation, which have similarly satisfactory results in terms of pattern recognition. It is added to this evaluation the appropriate behavior of the Receiver Operating Characteristic (ROC) curve and the low correlation coefficients between factors and target matrix, which confirm that the RNA technique of backpropagation topology can be an important ally in the decisions of institutions regarding the management of student retention.

**Keywords:** Artificial Neural Networks, *backpropagation*, evasion.

## LISTA DE FIGURAS

Figura 1 - Modelo não linear de um neurônio .....	23
Figura 2 - Rede alimentada adiante com camada única .....	25
Figura 3 - Rede alimentada diretamente com múltiplas camadas .....	25
Figura 4 - Rede recorrente com neurônios ocultos.....	26
Figura 5 - Curva ROC .....	33
Figura 6 - Etapas de trabalho propostas .....	36
Figura 7 - Decisão sobre a utilização dos dados.....	38
Figura 8 - Definição da quantidade de neurônios – processamento completo .....	41
Figura 9 - Configuração de implementação da RNA .....	42
Figura 10 - Definição da quantidade de neurônios - bacharelado .....	43
Figura 11 - Definição da quantidade de neurônios - licenciatura.....	44
Figura 12 - Definição da quantidade de neurônios - tecnologia.....	44
Figura 13 - Número de iterações x função de custo .....	50
Figura 14 - Número de iterações x gradiente e checagens de validação .....	51
Figura 15 - Desempenho RNAs Tecnologia e Licenciatura.....	52
Figura 16 - Gradientes e Checagens de Validação - processamento por tipo de graduação ....	53
Figura 17 - Matriz de Confusão para o Conjunto de Treinamento.....	54
Figura 18 - Matriz de Confusão para o Conjunto de Validação .....	55
Figura 19 - Matriz de Confusão para o Conjunto de Teste .....	56
Figura 20 - Matriz de Confusão Total .....	56
Figura 21 - Curvas ROC para os conjuntos de dados.....	57
Figura 22 - Matrizes de Confusão - Subconjuntos de treinamento .....	60
Figura 23 - Matrizes de Confusão - Subconjuntos de validação .....	60
Figura 24 - Matrizes de Confusão - Subconjuntos de teste .....	61
Figura 25 - Matrizes de Confusão - Todos os subconjuntos .....	61
Figura 26 - Curvas ROC - Todos os subconjuntos.....	62
Figura 27 - Coeficiente de Correlação - todos os subconjuntos .....	63
Figura 28 - Coeficientes de Correlação com Matriz Alvo - todos os subconjuntos.....	64

## LISTA DE QUADROS

Quadro 1 - Síntese dos tipos de aprendizado.....	27
Quadro 2 - Matriz de confusão para duas classes.....	32
Quadro 3 - Cursos analisados .....	35
Quadro 4 - Tipos e parâmetros de dados .....	37
Quadro 5 - Tratamento de parâmetros .....	39
Quadro 6 - Amostras por subconjunto.....	45
Quadro 7 - Configuração das RNAs por tipo de graduação .....	46
Quadro 8 - Comparativo de dados após o tratamento .....	48
Quadro 9 - Parâmetros de processamento .....	49
Quadro 10 - Parâmetros de processamento - por tipo de graduação .....	52
Quadro 11 - Coeficientes de Correlação.....	58

## SUMÁRIO

1	INTRODUÇÃO .....	13
1.1	JUSTIFICATIVA .....	14
1.2	OBJETIVOS DA DISSERTAÇÃO .....	16
1.2.1	Objetivo Geral .....	16
1.2.2	Objetivos Específicos .....	16
1.3	ABORDAGEM METODOLÓGICA.....	16
1.4	ESTRUTURA DA DISSERTAÇÃO.....	17
2	REVISÃO BIBLIOGRÁFICA .....	19
2.1	EVASÃO .....	19
2.2	REDES NEURAIS ARTIFICIAIS.....	22
2.2.1	Modelo de um Neurônio.....	23
2.2.2	Arquitetura de Rede.....	24
2.2.3	Aprendizado .....	27
2.2.4	Treinamento por Correção de Erros .....	28
2.2.5	Métricas de avaliação de RNAs.....	31
3	MÉTODOS.....	35
3.1	DEFINIÇÃO E COLETA DE DADOS .....	36
3.2	TRATAMENTO DOS DADOS .....	38
3.3	DEFINIÇÃO, IMPLEMENTAÇÃO E PROCESSAMENTO DO MODELO .....	39
3.3.1	Processamento completo .....	40
3.3.2	Processamento por tipo de graduação .....	43
3.4	ANÁLISE DOS RESULTADOS .....	46
3.4.1	Matriz de Confusão .....	46
3.4.2	Curva ROC .....	47
3.4.3	Coefficiente de correlação .....	47
4	APLICAÇÃO DOS MÉTODOS .....	48
4.1	DEFINIÇÃO, COLETA E TRATAMENTO DOS DADOS .....	48
4.2	IMPLEMENTAÇÃO E PROCESSAMENTO DOS MODELOS .....	49
4.2.1	Processamento completo .....	49
4.2.2	Processamento por tipo de graduação .....	52

4.3	ANÁLISE DOS RESULTADOS .....	53
4.3.1	Processamento completo .....	54
4.3.2	Processamento por tipo de graduação .....	59
5	CONCLUSÕES .....	65
5.1	CONSIDERAÇÕES FINAIS.....	65
5.2	SUGESTÕES DE TRABALHOS FUTUROS .....	66
	REFERÊNCIAS BIBLIOGRÁFICAS .....	67

## 1 INTRODUÇÃO

Conforme a Sinopse Estatística do Ensino Superior 2016 (2017), o número de alunos matriculados no ensino superior no Brasil foi de 8,05 milhões em 2016, enquanto em 2015 este número foi de 8,03 milhões, ou seja, crescimento de 0,2%. No período de 2006 a 20015, no entanto, o crescimento médio de matrículas no ensino superior do país foi de 6% e, extraindo-se apenas as instituições de caráter privado (responsáveis por 75% do sistema), houve recuo de matrículas de 6,07 milhões em 2015 para 6,05 milhões em 2016.

Analisando, além das novas matrículas e total de matriculados no ensino superior, os concluintes desta modalidade de ensino têm-se um novo parâmetro cabível de atenção: o percentual elevado de evasão, ou seja, alunos que, por diferentes motivos, desistem de seus cursos. De acordo com o Mapa do Ensino Superior 2016 (2016), a taxa de evasão nos cursos presenciais no país em 2014 foi de 25,4%, sendo que nas instituições públicas esta taxa foi de 18,3% e nas instituições privadas, de 27,9%. Tratando-se apenas dos cursos de engenharia, a evasão média no mesmo ano foi de 23,2%, sendo essa de 26,2% em instituições privadas e 14,7% nas instituições públicas.

Comparando-se a quantidade de concluintes na área de engenharia a cada 10.000 habitantes em relação a média dos países membros do OCDE (Organização para a Cooperação e Desenvolvimento Econômico), no ano de 2016 verifica-se que no Brasil este índice é de 6,1, enquanto que na OCDE o mesmo é de 9,1, ou seja, 33% maior no grupo citado. Faz-se necessário pontuar que índices como estabilidade populacional, crescimento do PIB (Produto Interno Bruto) e diminuição da desigualdade social interferem negativamente nas taxas brasileiras, explica Amaral (2016). Comparando-se, dentre os oito agrupamentos estudados, as áreas de engenharia, produção e construção e ciências, matemática e computação foram as com maior diferença entre os índices brasileiro e da OCDE.

Analisar os fatores que fazem com que os grupos das áreas de exatas e engenharias possuam menores índices de concluintes se torna necessário para o entendimento das causas e a proposição de alternativas para o cenário apresentado. Neste contexto, a inteligência artificial e suas variadas formas de aplicação emergem como ferramentas confiáveis e adequadas para os processos de análise e tomada de decisão. Tontini e Walter (2011) exemplificam tal possibilidade em pesquisa realizada com estudantes, por intermédio de questionários estruturados tratados com Redes Neurais Artificiais (RNA), onde se traçou, por intermédio de *clusters*, o perfil do aluno com tendência à evasão tendo como resultado o

acompanhamento dos que foram identificados neste grupo, reduzindo a mesma em 18% na instituição analisada.

Kim e Bayler (2016), em outra pesquisa, avaliaram a eficácia dos agentes de tutoria inteligente e verificaram que os sistemas que se valeram de “mentores”, solidários e conhecedores do assunto, obtiveram melhores resultados de aprendizagem dos que utilizavam modelos “motivadores” e “expert”, sendo o último focado exclusivamente na transmissão de conhecimento. Já Çebi e Karal (2017) aplicaram o Método Analítico Hierárquico (AHP) para definir um modelo imparcial para avaliação de trabalhos de alunos, sendo este um método amplamente utilizado para tomada de decisões baseadas em critérios.

Neste estudo a inteligência artificial aplicada à educação superior será abordada no âmbito da evasão discente, fazendo uso do modelo de reconhecimento de padrões baseado em RNAs que, por meio de informações acadêmicas, sociais e financeiras, propõe identificar os alunos que possuem maior tendência à evasão.

A utilização de ferramentas de inteligência artificial, como as que permitem realizar o reconhecimento de padrões em diferentes contextos, possibilita a mudança de paradigma na tomada de decisões: deixa-se o método empírico e utiliza-se o método analítico, que leva em consideração a conjuntura única e particular de cada aluno da Instituição estudada.

## 1.1 JUSTIFICATIVA

Levando-se em consideração a inédita redução de estudantes em 25 anos no ensino superior privado e a crise vivenciada no país no ano de 2015, com redução do PIB de 3,6%, e em 2016, quando esta queda foi de 3,8% (Instituto Brasileiro de Geografia e Estatística, 2016), faz-se necessário aprimorar a gestão universitária nos âmbitos acadêmico e administrativo-financeiro. Conforme apontam Rejas e Ponce (2015) a gestão de universidades necessita ser participativa, onde cada ator possui as competências necessárias para o gerenciamento sistemático dos resultados obtidos. Schmidt (2002) complementa esta visão ao afirmar que para dirigir uma universidade é necessário um equilíbrio entre a utopia de identificar horizontes amplos que antecipem o futuro e a realidade para que sejam identificadas e superadas suas mazelas, através de experiência, planejamento claro e trabalho coletivo.

A gestão é um dos núcleos vitais da universidade. É ela que busca e emprega recursos, propicia os necessários estímulos à qualificação docente, favorece a

ampliação das oportunidades educacionais, estimula a qualidade do ensino e apoia o desenvolvimento da pesquisa e da extensão. (SCHMIDT, 2002, p. 78)

Em seu estudo, Fernandes e Flores (2012) concluem que as instituições de ensino superior demonstram crescente interesse em implantar melhorias em suas práticas de gestão, de modo especial no que tange a questões da avaliação da qualidade dos docentes e do ensino. Neste sentido, a utilização de estudos cujos objetivos sejam visualizar o fenômeno da evasão, sistematizar ações de avaliação institucional e desenvolvimento de docentes, bem como prever os estudantes com tendência a deixarem seus cursos podem se constituir em ferramentas de apoio à tomada de decisão dos gestores universitários.

Lobo (2012) sintetiza a importância do estudo da evasão ao afirmar que “todos perdem com isso: além do próprio aluno, a instituição de ensino e todos os que nela trabalham, quem o financia e a sociedade como um todo”. Silva Filho et al. (2007) apresentam um conceito similar: as perdas de estudantes que iniciaram mas não terminaram seus cursos são desperdícios sociais, acadêmicos e econômicos.

Em termos de predição na área de gestão universitária, destacam-se alguns trabalhos: Superby et al. (2006) mapearam elementos culturais, sociais e psicológicos de estudantes do ensino médio para prever os que teriam maior probabilidade de evasão no ensino superior. Já Dekker et al. (2009) utilizaram dados relativos a notas, cursos e currículos do ensino médio para identificar os estudantes do departamento de engenharia elétrica com maior chance de evadirem no início do ensino superior. No caso de Bayer et al. (2012) foram utilizados dados derivados de comportamento social dos alunos coletados em fóruns e grupos de discussão para aprimorar a acurácia da previsão de evasão.

No Brasil pode-se destacar o trabalho de Campello e Lins (2008), que usaram a técnica de *clustering* para identificar e agrupar estudantes com maior tendência a evasão a partir das notas obtidas nas disciplinas, bem como a quantidade de disciplinas cursadas ao longo do tempo. Outro estudo que deve ser ressaltado é o de Manhães et al. (2011), que usaram técnicas de mineração de dados e tiveram por objetivo prever os alunos com maior probabilidade de evasão nos cursos de engenharia a partir das notas e do resultado final (aprovação ou reprovação) nas disciplinas cursadas no primeiro semestre dos cursos.

Este trabalho utiliza Inteligência Artificial por intermédio de algoritmos de RNAs como ferramenta de apoio à gestão acadêmica, no sentido em que possibilita, em um curto espaço de tempo, a visualização, compreensão e análise de grande quantidade de dados – cerca de 179 mil, oferecendo assim embasamento para tomadas de decisão que podem ser

adequadas à realidade vivida pelas Instituições de Ensino Superior no que tange evasão de alunos.

## 1.2 OBJETIVOS DA DISSERTAÇÃO

A seguir são definidos o objetivo geral e os objetivos específicos, que estabelecem a ação que conduz ao tratamento do tema abordado no problema de pesquisa e as ações realizadas de forma a estabelecer relação com a temática trabalhada.

### 1.2.1 Objetivo Geral

O objetivo geral desta dissertação é propor um método para identificar os estudantes do grupo de risco à evasão da Área do Conhecimento de Ciências Exatas e Engenharia de uma instituição de ensino superior comunitária fazendo uso da técnica computacional de Redes Neurais Artificiais.

### 1.2.2 Objetivos Específicos

Os objetivos específicos determinados para que o objetivo geral seja atingido consistem nos seguintes itens:

- a) explorar, por meio da pesquisa bibliográfica, o estudo da técnica de Redes Neurais Artificiais aplicados à evasão no ensino superior;
- b) efetuar a coleta de dados dos alunos evadidos e não evadidos, de 2006 a 2018;
- c) organizar e tratar os dados coletados para o formato necessário para o processamento, definindo *inputs* e *outputs*;
- d) configurar e aplicar o algoritmo de Redes Neurais Artificiais para o reconhecimento de padrão da evasão;
- e) validar o modelo e discutir os resultados.

## 1.3 ABORDAGEM METODOLÓGICA

Esta pesquisa tem, enquanto forma de abordagem, o método quantitativo uma vez que, para Marques e Melo (2017) ela está focada na objetividade e no fato de que a

compreensão acerca da realidade apenas pode ser obtida com a análise de dados brutos coletados com ferramentas neutras e padronizadas. Fonseca (2012) corrobora desta visão ao concluir que este tipo de pesquisa procura estabelecer uma regra, um princípio e a uniformidade de determinado fenômeno, levando em consideração o que é comum à maioria dos casos, sendo inadequada para as exceções.

No que se refere aos objetivos, pode-se afirmar que o caráter desta pesquisa é explicativo pois ela objetiva a identificação dos fatores que determinam ou contribuem para a ocorrência de fatores, segundo Gil (2002). Marconi e Lakatos (2011) complementam esta visão ao afirmarem que ela registra fatos, analisa-os, interpreta-os e identifica suas causas visando ampliar generalizações, estruturar e definir modelos teóricos, relacionar hipóteses e gerar hipótese a partir da dedução lógica.

Quanto aos procedimentos, a pesquisa em questão é considerada do tipo *ex-post-facto*, uma vez que o objetivo é a investigação das possíveis relações de causa e efeito entre um fato identificado pelo pesquisador e um fenômeno que ocorre posteriormente, sendo utilizada quando há impossibilidade de aplicação da pesquisa experimental (GERHARDT, SILVEIRA, 2009; FONSECA, 2012).

#### 1.4 ESTRUTURA DA DISSERTAÇÃO

O presente estudo de dissertação encontra-se estruturado em cinco capítulos, sendo o primeiro a contextualização do mesmo, no que tange a importância do estudo da evasão no ensino superior, apresenta as justificativas e motivações para o desenvolvimento do trabalho, bem como o objetivo geral e objetivos específicos, a metodologia utilizada e a estruturação dos capítulos.

O capítulo dois apresenta a revisão da literatura com os principais conceitos, definições e métodos abordados por outros autores e que foram adaptados para utilização nesta dissertação. Optou-se, em especial no que tange aos conceitos de RNAs, pela ênfase aos aspectos conceituais do que aos formais dos modelos e das técnicas, porém, em equilíbrio com tópicos atualizados relacionados ao tema.

O capítulo três aborda os métodos utilizados para desenvolver os algoritmos de reconhecimento de padrões, contextualizando o local e a motivação das escolhas, detalhando as etapas necessárias para definição das variáveis, aplicação das técnicas, teste e, por fim, a validação do modelo.

No capítulo quatro abordam-se as parametrizações realizadas, visualizam-se as configurações e simulações realizadas e discutem-se os resultados obtidos no *software* escolhido para implementação das RNAs.

Por fim, o capítulo cinco apresenta as conclusões do estudo, suas limitações e dificuldades encontradas, bem como suas contribuições e as sugestões para trabalhos futuros que possam aprimorar a pesquisa realizada.

## 2 REVISÃO BIBLIOGRÁFICA

Este capítulo compreende a revisão bibliográfica iniciando pelo conceito de evasão seguido pela apresentação das RNAs, fornecendo assim o aparato bibliográfico necessário para o desenvolvimento da proposta desta pesquisa.

### 2.1 EVASÃO

Define-se evasão como fuga ou escapada (BUENO, 1996). Quando levado ao âmbito da educação, ou especificadamente à educação superior, este alcança uma delimitação mais estreita: estudantes que abandonaram, trancaram, desligaram-se ou transferiram-se para outra instituição de ensino, de acordo com Rodriguez (2011). Para Lobo (2012), a evasão no ensino superior deve ser tratada levando-se em consideração a natureza do desligamento: evasão do curso, evasão da Instituição de Ensino Superior (IES) e evasão do sistema.

Ainda conforme Lobo (2012), a evasão do curso pode ser entendida como aquela em que o aluno troca de curso mas permanece na mesma instituição, troca de instituição ou como aquela em que ele evade sem motivo aparente. No caso da evasão da instituição, é a que o aluno troca de instituição, mas se mantém no ensino superior. Já na evasão do sistema, o aluno deixa o ensino superior, independente da instituição de ensino.

Para fins de delimitação, este trabalho enquadra-se como “evasão de curso”, considera como evadido o aluno que saiu do curso, mantendo-se ou não na instituição de ensino, bem como se mantendo ou não no ensino superior. No que se refere a diferenciação entre trancamento e evasão, este estudo apoia-se na definição de Gaioso (2005) que considera a evasão a interrupção do ciclo de ensino e em Baggi e Lopes (2011), que consideram a evasão como fuga de alunos. Isto deve-se ainda a imprevisibilidade de retorno de alunos que trancam suas matrículas, fator este apontado por Lobo (2012) como dificultador para o cálculo de evasão onde há esta diferenciação.

No que tange as causas da evasão, uma ampla gama de estudos visa este tópico, como as pesquisas de Spady (1970), Tinto (1975), Gaioso (2005) e Silva Filho (2007). Spady (1970) ancorou seus estudos nas teorias de Durkheim sobre suicídio. Essa teoria sugere que a propensão ao suicídio é aumentada quando o indivíduo não está integrado a algum grupo social. A partir deste entendimento Spady (1970) propôs que o indivíduo que compartilha valores com o grupo social é menos propenso a abandonar o curso. O momento de ingresso em um curso superior também é analisado por Almeida et al. (2002), corroborando com a

percepção de que esta etapa da vida do estudante o coloca em um estado potencial de angústia, uma vez que deve responder a questionamentos pessoais de cunho vocacional, ético, acadêmico e social.

O Modelo de Integração desenvolvido por Tinto (1975) avançou sobre os estudos de Spady (1970) propondo que a integração, a adaptação social e acadêmica do estudante na instituição determinará a sua decisão de permanecer ou não no curso. O autor então centra seu raciocínio em dois vetores: no indivíduo, considerando seu contexto pessoal e cultural, e na instituição, por meio de sua oferta cultural, acadêmica e de integração social.

Veloso (2001) conduziu um estudo empírico sobre a evasão nos cursos de graduação da Universidade Federal de Mato Grosso, buscando explicar a evasão através da percepção da IES. A pesquisa foi realizada utilizando entrevistas e análise de dados. O autor concluiu que a evasão naquele contexto era um fenômeno institucional, reflexo da ausência de uma política de permanência do aluno no curso de sua opção. Veloso (2001) aponta ainda uma série de ações possíveis para a redução da evasão, como o desenvolvimento de programas de intercâmbio junto às escolas e o desenvolvimento de ações de acompanhamento e integração do aluno à vida universitária. O autor ainda sugere que a evasão é uma ferramenta de exclusão social, que pode ser agravada em diferentes cenários ou grupos sociais.

Para estudar este contexto, Cardoso (2008) pesquisou, através de análise qualitativa, os efeitos das políticas de cotas na Universidade de Brasília (UnB), tendo como vetores principais o rendimento e evasão dos alunos cotistas. O estudo apontou que os cotistas evadem em percentual menor do que outros alunos. Por outro lado, os alunos que conciliavam seu estudo ao trabalho demonstravam indecisão sobre a escolha de seu curso, e por este motivo tinham índices de evasão maiores do que alunos cotistas que não trabalhavam.

Oliveira (2015), utilizando como ferramenta de pesquisa o Questionário de Vivências Acadêmicas – QVA-r, estudou o nível de adaptação, rendimento e permanência de graduandos de engenharia de uma universidade federal paranaense. Suas conclusões demonstram que questões relacionadas com as dimensões vocacionais e econômicas foram áreas de maior vulnerabilidade dos acadêmicos. Para o autor, tais áreas necessitariam de maior atenção institucional por parte da IES. Este entendimento também é encontrado em estudo anterior de Silva Filho et al. (2007), apontando que a falta de recursos financeiros para a conclusão dos estudos, bem como uma incompatibilidade entre a expectativa e a realidade acadêmica são fatores fundamentais para a evasão.

Pesquisa conduzida por Almeida et al. (2002) apontaram que cerca de 30% dos alunos que frequentam o ensino superior estudam em cursos que não correspondem às suas primeiras

opções vocacionais. Esse fato pode explicar, ao menos em parte, o crescente número de jovens universitários insatisfeitos e desestimulados com seus estudos, especialmente em disciplinas da área de exatas (OLIVEIRA, 2015).

Há a necessidade de se levar em consideração ainda, em instituições privadas, o valor das mensalidades – especialmente em cursos considerados “caros”, como na área de ciências exatas e engenharia. O aspecto financeiro, porém, não é visto como um fator determinante para a evasão por alguns autores como Tinto (1975), Silva Filho et al. (2007), Cardoso (2008) e Lobo (2012) pois, conforme verificado, há a percepção de que aspectos ligados a qualidade do ensino, engajamento com a comunidade acadêmica e curso, bem como percepções pessoais, tendem a embasar a decisão final.

Em contraponto a isto, para Tontini e Walter (2011) a dificuldade financeira é um dos principais fatores apontados pelos estudantes como justificativa para evasão. Conforme Cabrera et al. (1992), se faz importante pontuar que nem todos estudantes tem condições financeiras adequadas, seja por apoio governamental, empresarial ou familiar – o que acaba por impactar na continuidade do curso. Dowd e Coury (2006) corroboram desta visão e argumentam que os estudantes com melhores condições financeiras são mais propensos a se manterem no ensino superior, uma vez que não necessitam arcar com despesas como moradia e alimentação, por exemplo. Em cursos onde o mercado de trabalho absorve alunos durante a graduação, como nos cursos de ciências exatas e engenharia, há o desafio que o tempo investido no trabalho impõe a participação em atividades acadêmicas, o que impacta na percepção do mesmo acerca do seu curso.

O tempo investido no trabalho como forma de sobrevivência impõe, em vários casos, limites acadêmicos, como na participação em encontros organizados no interior ou fora da universidade, nos trabalhos coletivos com os colegas, nas festas organizadas pela turma, entre outras circunstâncias. Há uma luta constante entre o que gostariam de fazer e o que é possível fazer, materializada em uma gama variada de situações: carga horária de trabalho, tempo insuficiente para dar conta das solicitações do curso e outras, de ordem social e cultural, condicionadas pelos baixos recursos financeiros (ZAGO, 2006, p. 235).

Seja o motivo pessoal, financeiro ou acadêmico, faz-se importante ressaltar que o momento vivido traz à tona a necessidade de gestão de todos os atores envolvidos na

problemática. Santos (2001, p. 6) confirma a afirmação ao citar que “durante muito tempo as universidades não tiveram muita preocupação com a gestão dos seus custos, pois a receita sempre maior permitia que esses custos fossem absorvidos facilmente pelos aumentos nas mensalidades”. Contudo, na opinião de Pereira (2003), com exceção de alguns cursos como medicina e odontologia, muitos cursos não ocupam a totalidade das vagas oferecidas, ou seja, o aluno não mais concorre a uma vaga pelo vestibular e sim a Instituição que precisa concorrer no processo de escolha do aluno.

## 2.2 REDES NEURAIS ARTIFICIAIS

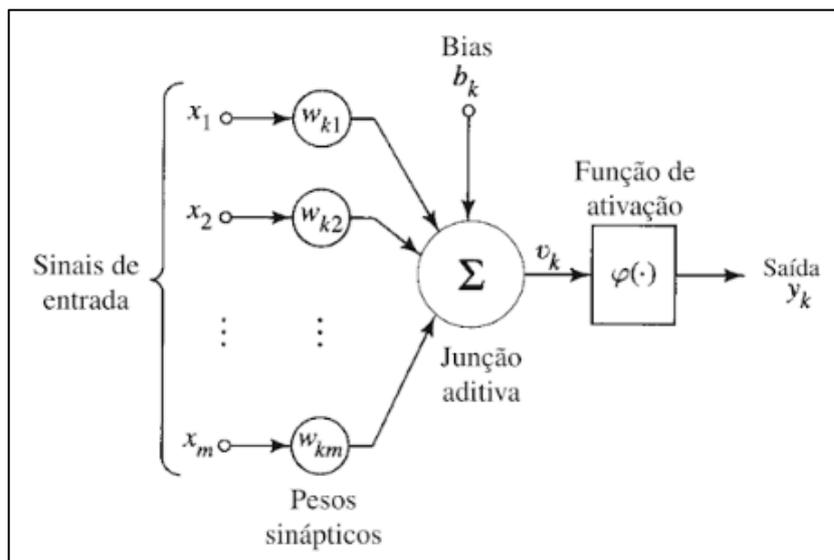
De acordo com Haykin (2001) e Mackay (2005), uma rede neural tem inspiração no cérebro humano, sendo um sistema inteligente com capacidade de aprendizado e reconhecimento de padrões nos dados apresentados em suas entradas, bem como a generalização deste conhecimento para a análise de novos dados. O primeiro estudo com RNAs foi publicado em 1943, de acordo com Dartora (2003), com o título *A Logical Calculus of the Ideas Immanent in Nervous Activity*, que propunha a modelagem de um neurônio artificial como uma unidade de processamento simples. Desde então diferentes estudos vêm sendo apresentados na área, dada a robustez, a habilidade de aprender com exemplos e a velocidade de processamento desta técnica.

Von Zuben (2003) cita alguns dos principais tipos de problemas que podem ser resolvidos atualmente utilizando-se RNAs: aproximação de funções, identificação de sistemas dinâmicos, emulação de atividades humanas, predição de séries temporais, agrupamento de dados, Caixeiro Viajante e Roteamento de veículos, Memória Endereçável por Conteúdo e classificação de padrões. Esta última aplicação, de acordo com Moraes (2010), tem recebido atenção especial devido as ocorrências humanas, como linguagem, entendimento de imagens, modo de falar, entre outros, tomarem forma de padrões. O reconhecimento de padrões é, conforme o autor, uma tarefa complexa onde busca-se avaliar as situações em termos de padrões das circunstâncias que as constituem, descobrir relações existentes no meio para melhor entendê-las, possibilitando adaptação.

### 2.2.1 Modelo de um Neurônio

Conforme Haykin (2001), “um neurônio é uma unidade de processamento de informação que é fundamental para a operação de uma rede neural”. A Figura 1 apresenta o modelo de um neurônio.

Figura 1 - Modelo não linear de um neurônio



Fonte: adaptado de Haykin (2001, p. 36).

Na figura supracitada pode ser verificado o conjunto de sinapses caracterizadas, cada uma, por um peso próprio. Um sinal  $x_j$  na entrada da sinapse  $j$  conectada ao neurônio  $k$  é multiplicado pelo peso sináptico  $w_{kj}$ , onde  $k$  se refere ao neurônio em questão e  $j$  se refere ao terminal de entrada da sinapse a qual o peso se refere. Podem ser visualizados ainda um somador ponderado pelas sinapses do neurônio denominado combinador linear e uma função de ativação (restritiva). Há, finalmente, um *bias* aplicado externamente, representado por  $b_k$  cujo objetivo é aumentar ou diminuir a entrada da função de ativação. Em termos matemáticos, de acordo com Haykin (2001), um neurônio  $k$  pode ser descrito de acordo com as Equações 1 e 2.

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (1)$$

$$y_k = \varphi(u_k + b_k) \quad (2)$$

Nas equações  $x_1, x_2, \dots, x_m$  são os sinais de entrada;  $w_{k1}, w_{k2}, \dots, w_{km}$  são os pesos sinápticos do neurônio  $k$ ;  $u_k$  é a saída do combinador linear devido ao sinais de entrada;  $b_k$  é o bias;  $\varphi(\cdot)$  é a função de ativação e  $y_k$  é o sinal de saída do neurônio.

Quanto à função de ativação, a mesma define a saída de um neurônio em relação a um campo local induzido ( $v$ ), podendo ser dos tipos informados a seguir:

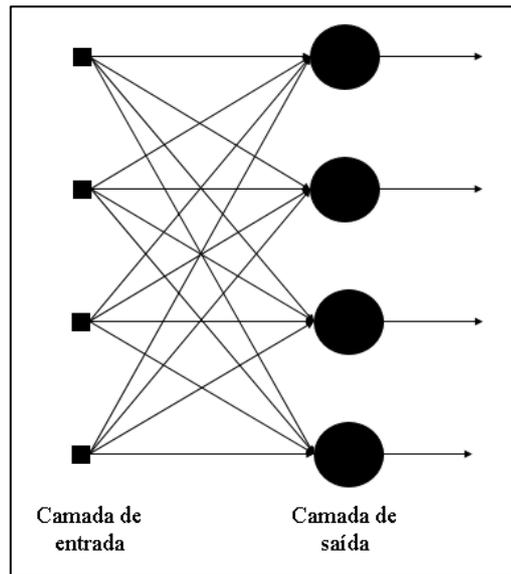
- a) Função de Limiar: de acordo com Haykin (2001), também conhecida como *função de Heaviside*, a saída de um neurônio assume o valor 1 se o campo local induzido daquele neurônio é não negativo, e 0 caso contrário.
- b) Função Linear por Partes: esta forma de função de ativação pode ser interpretada como uma aproximação de um amplificador não linear, conforme Haykin (2001)
- c) Função Sigmóide: é a função de ativação mais utilizada na construção de Redes Neurais Artificiais, possui o formato de “s”, é estritamente crescente e exibe um balanceamento adequado entre comportamento linear e não linear, explica Haykin (2001). Quando deseja-se valores de saída diferentes de 0 ou 1, a função tangente hiperbólica é utilizada, pois é sua forma escalonada. Um outro tipo de função sigmoide é a função Softmax, que é utilizada para problemas de classificação na última camada e garante que as saídas formem uma distribuição de probabilidade válida, de tal maneira que nenhuma saída possua valor 0 e soma de todas seja 1, de acordo com Mikolov et al. (2011).

### 2.2.2 Arquitetura de Rede

Conforme indicam Souza (2008) e Haykin (2001), a arquitetura da rede é o modo de conexão entre os diferentes neurônios que formam a RNA, onde o encadeamento destes neurônios forma as chamadas “camadas”, sendo que a maneira pela qual estes neurônios estão estruturados está intimamente ligada com o algoritmo de aprendizado utilizado para treinar a rede. Em geral, são identificadas três classes de arquitetura diferentes:

- a) Redes Alimentadas Adiante com Camada Única: este tipo de rede é estritamente do tipo alimentada adiante ou acíclica, uma vez que há uma camada de entrada de nós de fonte que se projeta sobre uma camada de saída de nós computacionais, mas não ocorre no sentido contrário, como demonstrado na Figura 2.

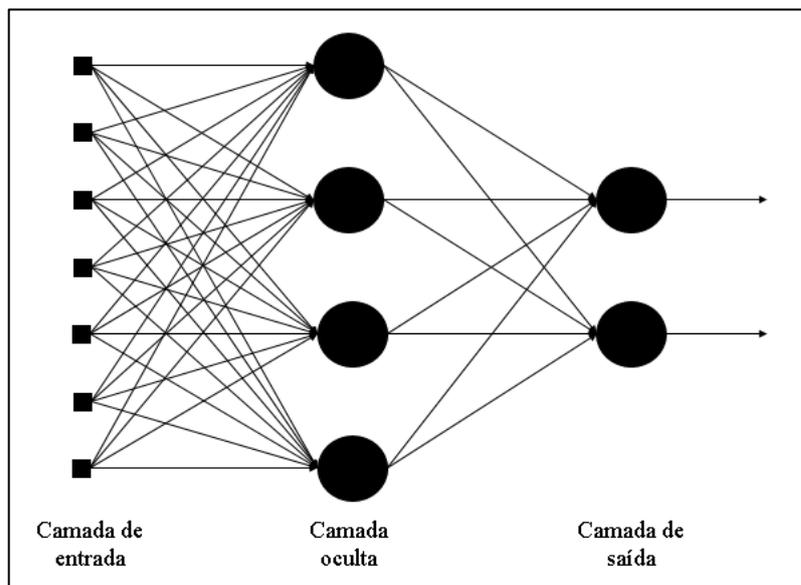
Figura 2 - Rede alimentada adiante com camada única



Fonte: o autor (2018)

- b) Redes Alimentadas Diretamente com Múltiplas Camadas: a distinção desta rede se dá pela presença de uma ou mais camadas ocultas, cujos nós são chamados de neurônios ocultos, e tem por função intervir entre a entrada externa e a saída da rede, conforme indica a Figura 3. Kayri (2015) demonstrou que este tipo de rede se mostra adequada em pesquisas relacionadas à predição de sucesso de alunos no ensino superior.

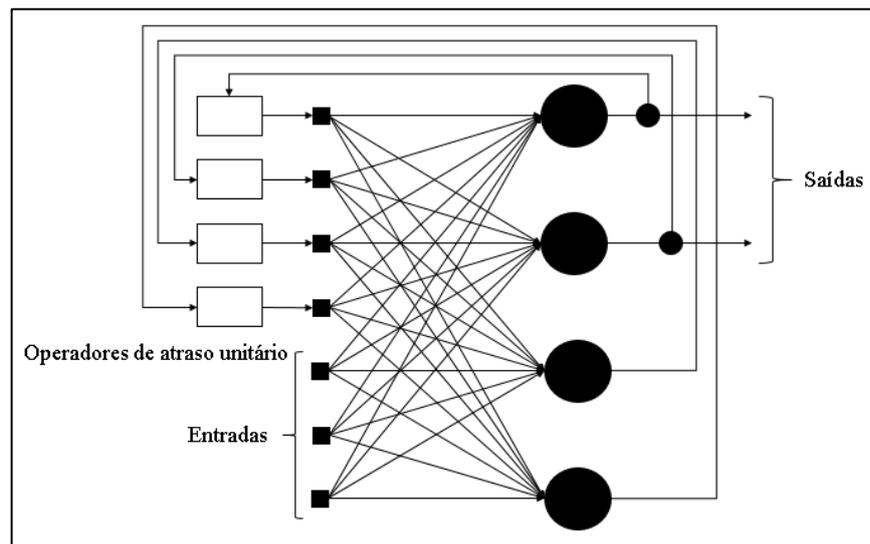
Figura 3 - Rede alimentada diretamente com múltiplas camadas



Fonte: o autor (2018)

- c) Redes Recorrentes: distinguem-se das anteriores por possuir ao menos um laço de realimentação, que possui a capacidade de impactar profundamente a aprendizagem da rede e seu desempenho. Conforme aponta Pereira (2017) este tipo de rede é fortemente utilizado para predição, especialmente nos que possuem como foco as séries temporais, como em índices financeiros. Tal característica de aplicação tem, entre outras motivações, a forte dependência com as últimas entradas, tendo como efeito prático a memória de curto prazo na rede.

Figura 4 - Rede recorrente com neurônios ocultos



Fonte: o autor (2018)

Estudos relacionados à área educacional citam a dificuldade em definir a arquitetura ideal e as definem, quando explicado, a partir de seu funcionamento e suas aplicações (Karamouzis e Vrettos, 2008; Oladokun et. al, 2008; Martinho et. al, 2013; Musso et. al, 2013). Em termos práticos, para problemas de classificação, as redes alimentadas diretamente são usadas comumente para buscar a resolução dos problemas.

Kayri (2015) analisou em seu estudo o desempenho de predição de sucesso de alunos por meio da rede Multilayer Perceptron – MLP (algoritmo *backpropagation*) e da rede RBF (*Radial Basis Function*), que são duas das arquiteturas mais comuns deste tipo de rede, e concluiu que a primeira, para estudos ligados à área educacional, possui melhor desempenho devido a não linearidade dos dados, além do menor tempo para processamento. A amostragem, de 1271 universitários, foi feita por meio de questionário com oito questões e,

como resultado, dentre os parâmetros avaliados, apenas o erro quadrático médio e a raiz do erro quadrático médio foram inferiores na rede RBF. O erro relativo, a soma dos quadrados dos erros e o erro absoluto médio comprovaram que a rede MLP é mais adequada para o atingimento do objetivo.

### 2.2.3 Aprendizado

De acordo com Rezende (2003), as RNAs têm como característica o aprendizado por meio de exemplos, sendo o algoritmo de aprendizado o responsável pela adaptação dos parâmetros da rede para que ocorra a convergência para uma solução. Podem ser considerados três tipos de aprendizados, resumidos no Quadro 1.

Quadro 1 - Síntese dos tipos de aprendizado

Tipo	Característica	Aplicações	Benefícios
Supervisionado	Existência de supervisor externo à rede que monitora as resposta para cada vetor de entrada	Aproximação de funções, modelagem de sistemas e classificação de dados	Há transferência de conhecimento do supervisor para a rede neural da forma mais completa possível.
Não supervisionado	Não há saídas desejadas para as entradas e não há supervisor externo	Categorização de dados	Melhoramento do custo computacional e redução do tempo de treinamento
Por Reforço	Aplica reforço ou penalidade associado à última ação da rede. Não há supervisor externo	Sistemas observados por sensores que não possuem limites definidos (ou erros de medição)	O sistema interage com o ambiente e desenvolve a habilidade de aprender com base apenas na sua experiência

Fonte: o autor (2018)

Entende-se que uma RNA aprendeu adequadamente um problema quando a mesma é capaz de generalizá-lo, ou seja, quando o mapeamento de entrada-saída computado pela rede for correto, ou aproximadamente correto, de acordo com Haykin (2011). Em termos de processo, o aprendizado deve ser entendido como um “ajuste de curva” e, neste caso, deve ser levado em consideração o cuidado para que a rede, quando exposta a uma grande quantidade de exemplos, não memorize os dados do treinamento (efeito conhecido como *overfitting*), tornando-o insipiente para novos dados. Neste caso, o objetivo do treinamento deve ser aproximar as funções geradoras dos dados e não simplesmente minimizar o erro do conjunto de treinamento, de acordo com Rezende (2003).

Para que a possibilidade de ocorrência do *overfitting* seja minimizada, Rezende (2003) sugere que seja aplicado o método construtivo ou o algoritmo de poda. O primeiro, de acordo com Castro (1999) e Rezende (2003), consiste na adição gradual de neurônios ou camadas até que uma solução adequada seja encontrada ou um critério de parada envolvendo o erro de treinamento e de generalização seja atingido. O segundo método consiste no processo inverso, conforme Haykin (2001): começar com uma rede *perceptron* de múltiplas camadas grande, com desempenho adequado, e então podá-la por meio da eliminação ou redução de certos pesos sinápticos.

#### 2.2.4 Treinamento por Correção de Erros

Rezende (2003) cita que o primeiro modelo de RNA que conceituou o aprendizado foi o Perceptron Simples, em 1958, de apenas uma camada, porém, sabendo-se desde já que para a resolução de problemas de maior complexidade haveria necessidade de mais camadas. Na década de 80 foi descrito o modelo *backpropagation*, cuja estrutura está baseada na retropropagação dos erros para realização dos ajustes de pesos das camadas intermediárias.

Haykin (2001) pontua que o perceptron é a forma mais simples de uma rede neural usada para a classificação de padrões linearmente separáveis, sendo que quando as entradas não são linearmente separáveis, o algoritmo de convergência do perceptron apresenta problemas pois as fronteiras de decisão entre as diferentes classes podem oscilar continuamente.

No que se refere a aprendizagem por retropropagação de erro, esta consiste em dois passos, conforme explica Haykin (2001): um para frente (propagação, com pesos sinápticos fixos) e um para trás (retropropagação, com pesos sinápticos ajustados de acordo com a regra de correção de erro). Este tipo de rede é capaz aprender tarefas complexas extraindo

progressivamente as características mais significativas dos vetores de entrada, ao mesmo tempo que exhibe alto grau de conectividade, sendo este proporcionado pelas sinapses da rede.

Para que a RNA possa ser treinada, oferecendo o resultado esperado, se faz necessário que alguns parâmetros sejam definidos, de tal maneira que podem interferir diretamente na qualidade do mesmo, tais quais os relacionados a seguir.

#### 2.2.4.1 Métodos de Amostragem

De acordo com Haykin (2001), visto que o processo de aprendizagem pode ser entendido como uma escolha de parametrização da rede para um conjunto de dados, a ferramenta estatística conhecida como validação cruzada serve como princípio orientador para buscar a validação do modelo. Neste contexto o que ocorre é a divisão do conjunto de treinamento em subconjuntos, sendo um para o próprio treinamento, um subconjunto de estimação para selecionar o modelo e um subconjunto de validação, usado para testar ou validar o modelo. Rezende (2003) complementa este entendimento informando que existem diferentes métodos para estimar uma medida verdadeira, conforme descrito a seguir.

- a) *Houldout*: este estimador divide os exemplos em uma porcentagem fixa de exemplos para treinamento e teste. Valores típicos são 2/3 para treinamento e 1/3 para teste.
- b) *Stratified Cross-Validation*: estimador que gera partições exclusivas considerando a proporção de exemplos em cada uma das classes.
- c) *Leave-One-Out*: é um estimador computacionalmente dispendioso e frequentemente utilizado em amostras pequenas onde, em uma amostra  $n$ , uma hipótese é induzida utilizando  $(n - 1)$  exemplos e então a hipótese é testada no único exemplo remanescente.
- d) *Bootstrap*: estimador cuja ideia básica é repetir o processo de classificação diversas vezes, estimando então valores como erro e bias, a partir dos experimentos replicados, sendo que cada experimento é conduzido com base em um novo conjunto de treinamento obtido por amostragem com reposição do conjunto original de exemplos.

#### 2.2.4.2 Algoritmo de treinamento

Parte relevante das tarefas das RNA referem-se a problemas de otimização, sendo o Algoritmo do Gradiente Descendente um dos mais usados para cumprir esta função, constituindo-se de uma ferramenta para otimizar funções complexas dentro de um programa de computador, objetivando encontrar seu valor mínimo. De acordo com Rezende (2003), como normalmente os pesos são inicializados de forma aleatória, o desempenho de treinamento depende das condições iniciais e o valor da taxa de aprendizado influencia também a velocidade de convergência.

Para muitas funções pode haver mais de um valor mínimo, podendo ser este o mínimo local ou o mínimo global. A tarefa dos algoritmos do gradiente descente é então, minimizar a possibilidade de parada em um mínimo local, forçando a convergência em um mínimo global. Rezende (2003) explica que uma alternativa para que isto seja realizado é a adição de um termo de *momentum*, que é o responsável pelo acúmulo do histórico de ajustes anteriores, resultando em termo residual quando o ajuste é nulo por causa do gradiente, fazendo com que o treinamento permaneça em torno do mínimo local até a convergência.

Rezende (2003) cita, no entanto, que outras variações do backpropagation surgiram na literatura, entre os anos de 1980 e 1990, como o *Quick Propagation*, *Resilient Propagation*, Levenberg-Marquadt e treinamento por modos deslizantes, entre outros, que resultaram em soluções mais eficientes para os problemas de treinamento.

Atualmente, no *software* MathWorks Inc. *Matlab R2017b*®, usado pela comunidade científica, há variados modelos de algoritmos baseados no gradiente que possuem diferentes características de atualização de pesos e bias, porém, levando em consideração o estudo realizado por Baptista et al. (2013), que comparou o desempenho destes algoritmos para problemas de classificação, a seguir pode ser visualizada a descrição dos que mais se destacaram neste:

- a) *trainoss (One-step secant backpropagation)* - atualiza os valores de acordo com o método secante de uma etapa. A retropropagação é usada para calcular os derivativos de desempenho em relação às variáveis de peso e polarização e uma função de busca de linha é usada para localizar o ponto mínimo. Este algoritmo requer mais espaço de armazenamento e capacidade de processamento por iteração do que os algoritmos de gradiente conjugado;
- b) *trainlm (Levenberg-Marquardt backpropagation)* - atualiza os valores de peso e polarização de acordo com a otimização Levenberg-Marquardt, sendo

frequentemente o algoritmo mais rápido disponível, embora exija mais memória que do que os outros algoritmos;

- c) *trainscg (Scaled conjugate gradient backpropagation)* - atualiza os valores de ponderação e polarização de acordo com o método do gradiente conjugado escalonado. O mesmo não realiza pesquisa de linha em cada iteração, o que garante menor necessidade de armazenamento e processamento que os demais algoritmos.

#### 2.2.4.3 Função de custo

O treinamento e a qualidade da predição realizada necessitam ser computados e avaliados por uma função, sendo esta conhecida como função de custo. De acordo com Bezerra (2016), dentre as funções de custo mais utilizadas estão a entropia cruzada (*cross-entropy*) e a soma dos erros quadrados, sendo que a escolha adequada para utilização deve levar em consideração os dados do estudo e seu objetivo.

Para dados contínuos, conforme Bezerra (2016), é de uso popular a soma dos quadrados dos erros, porém, quando os dados são binários indica-se a utilização da entropia cruzada. Em seu estudo teórico e prático, Golik et al. (2013) concluíram que a entropia cruzada possui a capacidade de encontrar um ótimo local melhor do que a soma do quadrado dos erros, visto que o último ficou preso em um pior local, onde o gradiente ficou nulo e não houve mais ajustes de erro. O autor cita ainda que os pesos foram inicializados aleatoriamente e, caso não fossem, seria possível um melhor resultado da soma dos quadrados dos erros, porém, a inicialização aleatória é a maneira mais comum de iniciar o treinamento em grande parte dos estudos.

#### 2.2.5. Métricas de avaliação de RNAs

Para que se possa avaliar a qualidade do processamento da rede, algumas técnicas são sugeridas, de tal forma que podem variar conforme a topologia de rede utilizada e o objetivo para o qual a mesma foi desenvolvida. No caso das Redes Neurais Artificiais aplicadas ao reconhecimento de padrões, são sugeridas as seguintes ferramentas:

### 2.2.5.1 Matriz de Confusão

De acordo com Rezende (2003), a matriz de confusão possibilita uma medida efetiva para os modelos de classificação, uma vez que demonstra o número de classificações corretas ante as classificações preditas para cada classe. Para problemas com duas classes, tais como o do objeto deste estudo, as classes são rotuladas comumente como “+” (positiva) e “-” (negativa). Neste caso, os dois erros possíveis são chamados de falso positivo ( $Fp$ ) e falso negativo ( $Fn$ ), enquanto os acertos são denominados de verdadeiro positivo ( $Tp$ ) e verdadeiro negativo ( $Tn$ ), conforme ilustrado pelo Quadro 2. A medida de desempenho é chamada de acurácia ( $(Tp + Tn) / (Tp + Fn + Fp + Tn)$ ).

Quadro 2 - Matriz de confusão para duas classes

<b>Classe</b>	<b>Predita C+</b>	<b>Predita C-</b>
<b>Verdadeira C+</b>	Verdadeiros positivos ( $Tp$ )	Falsos negativos ( $Fn$ )
<b>Verdadeira C-</b>	Falsos positivos ( $Fp$ )	Verdadeiros negativos ( $Tn$ )

Fonte: adaptado de Rezende (2013)

A rede, quando treinada de forma adequada, possui percentuais menores nas células definidas como falso negativo ( $Fn$ ) e falso positivo ( $Fp$ ). O inverso ocorre a partir de uma alta taxa de acertos nas células correspondentes ao verdadeiro negativo ( $Tn$ ) e verdadeiro positivo ( $Tp$ ) pois, neste caso, considera-se que a rede demonstrou a capacidade de prever adequadamente os conjuntos a partir do treinamento realizado.

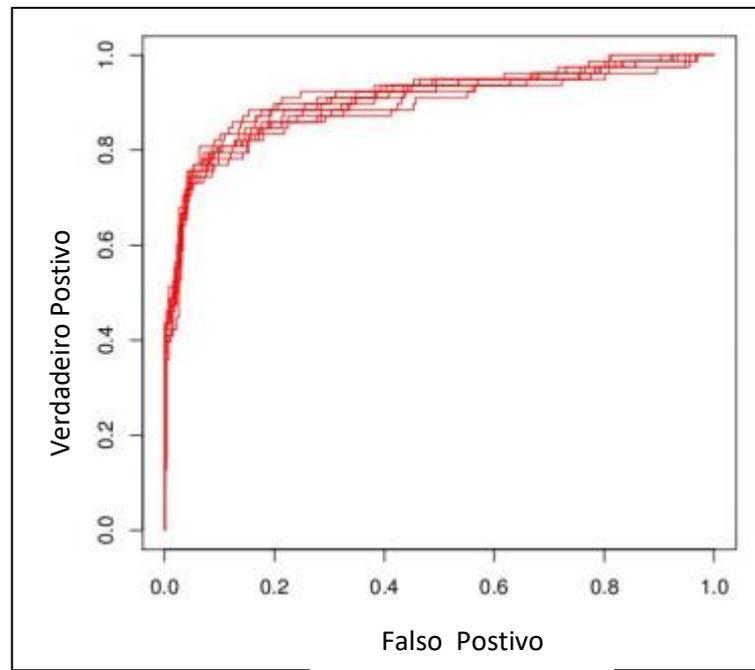
### 2.2.5.2 Curvas ROC

A curva ROC, conforme indicado por Prati et. al (2008), é um método gráfico para avaliação, organização e seleção de sistemas de diagnóstico e predição, provendo uma avaliação diferenciada e complementar ao método que avalia o treinamento por apenas uma única medida. Para traçar a curva faz-se necessário calcular a sensibilidade ( $Tp / (Tp + Fn)$ ), definida como a probabilidade do teste apresentar resultado positivo em um indivíduo que interrompeu seus estudos e, também, a especificidade ( $Tn / (Tn + Fp)$ ), que é a probabilidade do teste apresentar resultado negativo em um aluno que não interrompeu seu ciclo acadêmico.

De acordo com a Figura 5, a sensibilidade é plotada no eixo Y e no eixo X, 1 - a especificidade, sendo ambos em valores decimais, de tal maneira que para cada ponto de corte

utilizado no teste são calculadas a especificidade e a sensibilidade e inserido um ponto no gráfico, formando assim a curva ROC.

Figura 5 - Curva ROC



Fonte: adaptado de Prati et. al (2008)

A área sob a curva descreve o desempenho global do teste, uma vez que leva em consideração todos os valores de sensibilidade e especificidade para cada valor da variável do teste. Sendo assim, quanto mais a curva se aproximar do canto superior esquerdo, maior a capacidade do teste em distinguir os alunos que interromperam ou não seus estudos.

### 2.2.5.3 Coeficiente de correlação

De acordo com Larson e Farber (2010), uma maneira de medir o tipo e a força de uma correlação entre duas variáveis é calcular o coeficiente de correlação. A amplitude do coeficiente é de -1 para 1, sendo que, se a relação é forte, o valor está próximo de +1 (correlação positiva) ou de -1 (correlação negativa) e se não há relação, está próximo a 0. Conforme indica Field (2009), quando as variáveis não são paramétricas deve ser calculado o coeficiente de correlação de Spearman.

A utilização da correlação em conjunto com a aplicação das Redes Neurais Artificiais pode ser encontrada em diferente áreas, com o intuito principal de escolher as variáveis de entrada com maior correlação em relação à variável de saída. Soares et al. (2015) aplicaram o

coeficiente de correlação desta maneira em seu estudo sobre predição da produtividade do milho e Nascimento et al. (2017), na área de saúde pública, buscaram identificar os fatores de influência do ecossistema correlacionando-os com os óbitos por dengue, seguindo então de um processamento com Redes Neurais Artificiais.

Braga et al. (2000), no entanto, chamam a atenção para o fato de que não se pode julgar quais das variáveis de entrada foram mais importantes no processo de aprendizagem pois não existe correlação linear entre as mesmas. Esta visão é respaldada por Nagaoka et al. (2005), que afirmam que a inexistência de qualquer correlação pode ser resultado dos pesos atribuídos pela rede, visto que estes são definidos com o objetivo de minimizar a função de custo de todos os padrões de treinamento.

### 3 MÉTODOS

O ambiente escolhido para aplicação do estudo é uma Universidade Comunitária do Estado do Rio Grande Sul (RS) com mais de 80 cursos de graduação, além de programas de pós-graduação *Lato Sensu* e *Stricto Sensu*. O público-alvo são os acadêmicos dos cursos de graduação da área do conhecimento de Ciências Exatas e Engenharia devido ao interesse na análise global e local da área tecnológica da instituição. O Quadro 3 ilustra os cursos contemplados na análise entre os anos de 2006 e 2018.

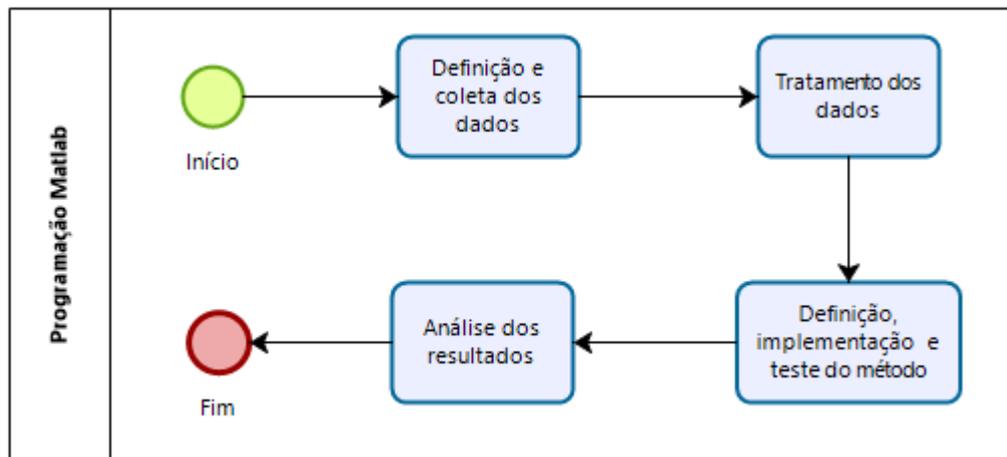
Quadro 3 - Cursos analisados

Num.	Curso
1	Análise e Desenvolvimento de Sistemas
2	Automação Industrial
3	Eletrônica Industrial
4	Polímeros
5	Sistemas de Qualidade
6	Computação
7	Física
8	Matemática
9	Matemática EaD
10	Química – Licenciatura
11	Ciência da Computação
12	Engenharia Ambiental
13	Engenharia Automotiva
14	Engenharia Civil
15	Engenharia de Alimentos
16	Engenharia de Computação
17	Engenharia de Controle e Automação
18	Engenharia de Materiais
19	Engenharia de Produção
20	Engenharia Elétrica
21	Engenharia Mecânica
22	Engenharia Química
23	Química – Bacharelado
24	Sistemas de Informação
25	Tecnologias Digitais

Fonte: o autor (2018)

Para o alcance do objetivo proposto neste estudo foram desenvolvidas as etapas ilustradas na Figura 6, sendo estas a definição e coleta dos dados; tratamento dos dados; definição, implementação e teste do método e análise dos resultados.

Figura 6 - Etapas de trabalho propostas



Fonte: o autor (2018)

O estudo se inicia com a definição dos dados utilizados para a classificação, seguido da coleta dos mesmos em *software* de gestão institucional, exportados para uso no *software* Microsoft Excel®. O tratamento de dados consiste em verificar possíveis inconsistências, bem como proceder com a conversão dos dados para facilitar o processamento da rede. Para a definição do modelo de RNA utilizado foram levados em consideração a topologia da rede e os parâmetros de algoritmo de treinamento. Na fase de implementação do método, ou treinamento da rede, é buscado o ponto ótimo de parada do processamento, de tal forma que o erro seja minimizado e a capacidade de generalização maximizada de acordo com os critérios escolhidos. Por fim tem-se a análise de resultados, que permite visualizar e comparar os resultados obtidos com os dados reais a fim de validar o processo de treinamento do algoritmo. As etapas detalhadas encontram-se descritas a seguir.

### 3.1 DEFINIÇÃO E COLETA DE DADOS

A definição dos dados requer uma análise do problema para evitar ambiguidades e erros, porém, requer também o correto entendimento sobre quais dados são suficientemente necessários para que o restante do processo ocorra da forma esperada.

A natureza do estudo pressupõe a utilização de fatores acadêmicos, sociais e financeiros para o reconhecimento do padrão de tendência à evasão, ou seja, visto que na literatura os parâmetros são diferentes para cada um dos estudos, optou-se, por meio empírico e a partir das bases consultadas, buscar os atributos considerados importantes para este processo.

Conforme indicado pelas referências deste trabalho, neste estudo foram contemplados parâmetros acadêmicos, sociais e financeiros, levando em consideração a disponibilização e a possibilidade de acesso de dados da Instituição, que podem ser verificados no Quadro 4.

Quadro 4 - Tipos e parâmetros de dados

Tipo	Parâmetro
Acadêmico	<ul style="list-style-type: none"> <li>- Tipo de curso (bacharelado, licenciatura ou tecnologia)</li> <li>- Curso</li> <li>- Currículo</li> <li>- Ano de ingresso na Instituição</li> <li>- Média harmônica do vestibular</li> <li>- Ano/semestre da última matrícula</li> </ul>
Social	<ul style="list-style-type: none"> <li>- Sexo</li> <li>- Ano de nascimento</li> <li>- Escola (pública ou privada)</li> <li>- Cidade de residência</li> </ul>
Financeiro	<ul style="list-style-type: none"> <li>- Possui financiamento estudantil</li> <li>- Possui débito com a Instituição</li> <li>- Faz parte do Programa Universidade para Todos (ProUni)</li> </ul>

Fonte: o autor (2018)

Outros parâmetros poderiam ter sido utilizados, como quantidade de solicitações de cancelamentos, suspensões e trocas de formas de pagamento na Central de Atendimento e quantidade de demandas na Ouvidoria da Instituição, porém, considera-se que o objetivo de se capturar diferentes ângulos do problema foi atingido com os dados obtidos.

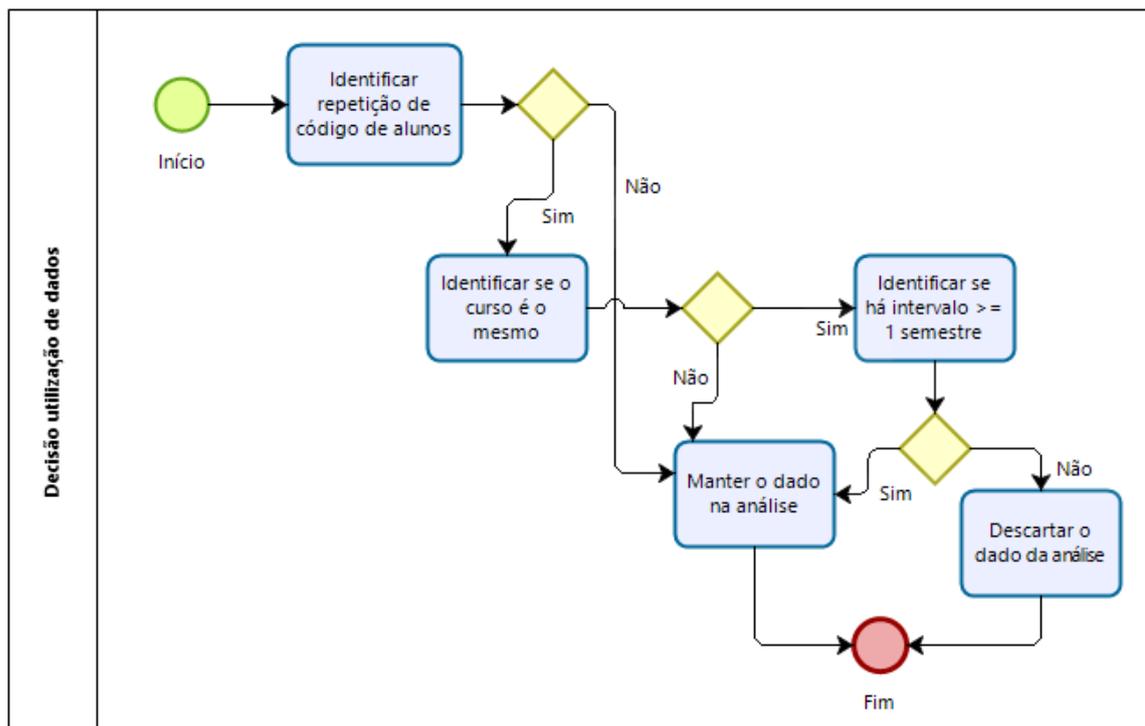
### 3.2 TRATAMENTO DOS DADOS

O tratamento de dados consiste em verificar os dados coletados e realizar as modificações necessárias (escalonamentos, normalizações, conversões, entre outros) para que estes estejam apropriados ao processamento na RNA.

Uma vez que os dados fornecidos correspondiam a novas matrículas por currículo, organizadas por semestre, houve a necessidade de identificar se havia recorrências de código do acadêmico (representando o mesmo estudante) causadas por troca de currículo no mesmo curso. Quando estas foram localizadas os dados relativos à primeira ocorrência de código do acadêmico foram atualizados com a informação sobre o último semestre cursado e os dados repetidos, excluídos da análise.

O tratamento em questão foi motivado pela necessidade de considerar que as trocas de currículo, sem intervalo, não representassem a evasão de acadêmico, visto que esta informação, pelo seu caráter interpretativo, não foi obtida diretamente pelo sistema da Instituição. A Figura 7 ilustra o processo de decisão a respeito da utilização dos dados originalmente obtido, que gerou uma matriz de 13 colunas (parâmetros) e 12785 linhas (matrículas) para utilização em planilha eletrônica.

Figura 7 - Decisão sobre a utilização dos dados



Fonte: o autor (2018)

Pode-se verificar que a última etapa da análise constante na Figura 7 consiste em identificar se o aluno apresenta interrupções em seu período acadêmico. Os dados obtidos não permitem identificar se os alunos que apresentaram interrupções realizaram evasão do curso, da Instituição ou do sistema, sendo que, para isto, seria necessário um acompanhamento e cadastramento do motivo que levou o aluno a evadir. Desta forma, optou-se por definir que a falta de matrícula em um semestre representa a evasão do curso, métrica possível de acompanhamento a partir dos dados disponibilizados pela Instituição, e que não possui tempo específico para que receba esta denominação.

Um segundo tratamento de dados necessitou ser realizado para a o processamento na RNA, visto que a entrada dos mesmos necessita estar adequada e padronizada no *software*, conforme indicado no Quadro 5.

Quadro 5 - Tratamento de parâmetros

Parâmetro	Tratamento
- Sexo - Escola - Cidade de residência - Débito com a Instituição - Possui financiamento - Faz parte do ProUni	Conversão em números binários
- Tipo de curso - Curso	Criação de números sequenciais
- Currículo	Remoção de caracteres alfabéticos

Fonte: o autor (2018)

### 3.3 DEFINIÇÃO, IMPLEMENTAÇÃO E PROCESSAMENTO DO MODELO

A escolha do modelo de RNA deu-se a partir da consulta à literatura referida neste estudo, sendo composta especialmente de artigos relacionados às RNAs para reconhecimento de padrões diversos, que apontaram para a rede neural de topologia *backpropagation* como a mais adequada para a obtenção do resultado esperado neste estudo. Após a definição constatou-se que esta topologia é a utilizada na *toolbox* (interface gráfica que possui

diferentes arranjos de RNAs) para reconhecimento de padrões do *software* escolhido para a aplicação, corroborando assim os estudos que sugeriam esta tipologia.

A rede neural foi então criada com o auxílio da *toolbox* do *Matlab R2017b*®, sendo que a definição da utilização desta *toolbox* deveu-se a praticidade da implementação com a possibilidade posterior do ajuste de parâmetros por meio da exportação do código, que é largamente utilizado pela comunidade científica. Em termos de Redes Neurais Artificiais, o *software* possui diferentes opções de *toolbox*, que podem ser entendidas como bibliotecas de funções capazes de resolver classes particulares de problemas, de acordo com Demuth et al. (2010).

Quanto a implementação, a organização dos dados foi dada por duas matrizes (entrada e alvo). A matriz de entrada possui 13 colunas e a matriz alvo uma, com números binários para marcar a interrupção dos estudos, quando ocorrida. Para o processamento da RNA na *toolbox* escolhida fez-se necessária a especificação de métodos e parâmetros, conforme descrito a seguir.

### 3.3.1 Processamento completo

O processamento completo levou em consideração o escopo total de dados, incluindo todos os tipos de graduação e cursos obtidos, de acordo com as etapas detalhadas a seguir.

#### 3.3.1.1 Determinação dos parâmetros de treinamento

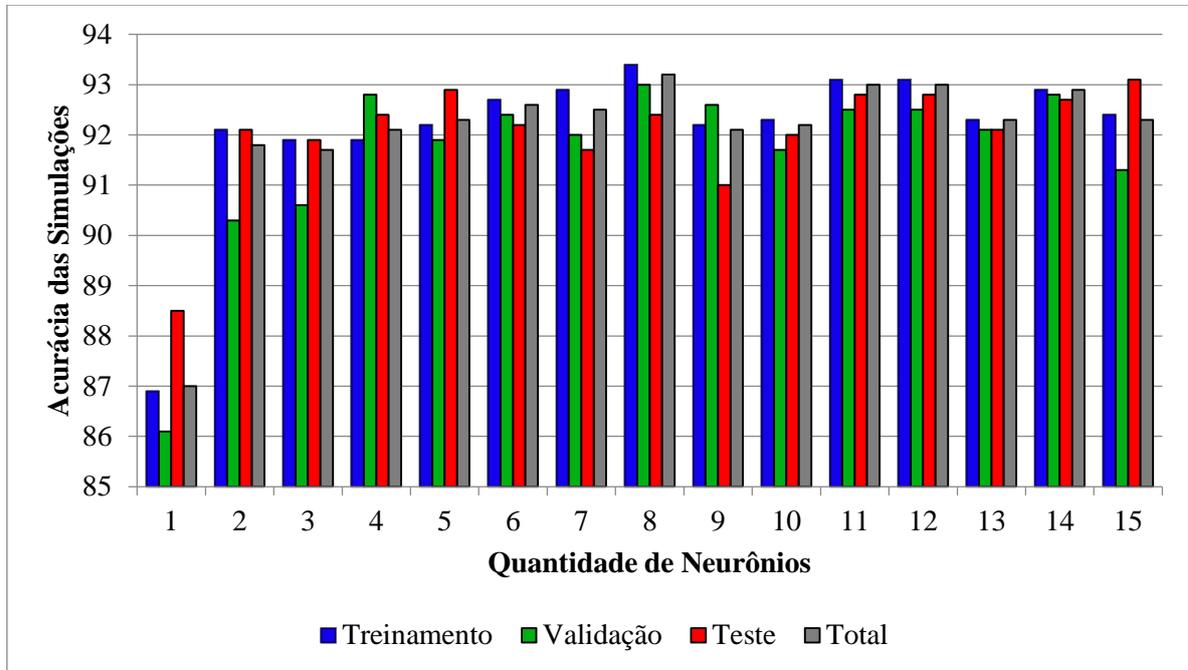
Além do carregamento dos dados de entrada e alvo no *software*, foi realizada também a determinação dos parâmetros de treinamento da rede neural artificial, sendo o primeiro a definição do número de neurônios na camada oculta.

Em relação à quantidade ideal de neurônios, visto sua dependência com o problema que se deseja tratar, optou-se por seguir a técnica dos algoritmos construtivos, de acordo com o item 2.2.3, e adicionar neurônios gradualmente, avaliando, por meio do treinamento, a capacidade de generalização atingida.

A Figura 8 ilustra o processo de determinação da quantidade de neurônios na camada oculta, iniciando o mesmo com 01 neurônio e avaliando sua capacidade de generalização em cada um dos subconjuntos de dados. Verifica-se que, a partir do processamento com 02 neurônios a acurácia encontrada em todas as simulações variou de 90% a 93%, com os

melhores resultados em 02 dos 03 conjuntos e no somatório dos mesmos no processamento com 08 neurônios.

Figura 8 - Definição da quantidade de neurônios – processamento completo



Fonte: o autor (2018)

O teste para definição do número de neurônios encerrou-se com a análise da rede com 15 neurônios uma vez que a máxima capacidade de generalização foi encontrada no processamento com 08 neurônios, sendo testado em mais 07 oportunidades sem que se obtivesse melhor resultado.

### 3.3.1.2 Definição do método de amostragem

Conforme citado, o *software* escolhido para a aplicação predefine o método *holdout* como método de amostragem, possibilitando a escolha de percentual de amostras para cada um dos conjuntos.

O tamanho dos conjuntos de validação e teste pode ser definido no *software* de 5% a 35%, cabendo o restante ao conjunto de treinamento, ou seja, no mínimo 30%. O conjunto de treinamento foi definido em 70% do total de amostras disponíveis, ou seja, 8949 amostras. Os 30% restantes foram subdivididos igualmente entre o conjunto de validação e o conjunto de teste, totalizando 1918 amostras para cada um dos grupos. Levando em consideração que a

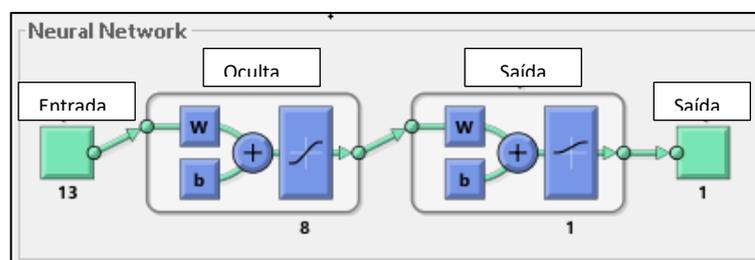
rede apresenta grande quantidade de dados e teve adequado comportamento de treinamento para os conjuntos definidos, o método foi mantido para o estudo em questão.

### 3.3.1.3 Análise dos parâmetros predefinidos

Devido a utilização da *toolbox* para criação e processamento da RNA, parte dos parâmetros foram preestabelecidos pelo *software*. No entanto, visto que ao final do processamento, por meio da exportação do algoritmo, se fazem possíveis a análise e substituição destes, os parâmetros de processamento foram definidos conforme a seguir.

- a) Número de camadas ocultas: com predefinição de uma camada, optou-se por manter a configuração devido aos algoritmos construtivos. Neste contexto, iniciou-se com o mínimo de camadas e avaliou-se o resultado do processamento, considerado adequado para o estudo em questão, tal como proposto no item 2.2.3.
- b) Funções de ativação: para este tipo de problema o *software* sugere a utilização de duas funções de ativação: a função de ativação tangente hiperbólica para a camada oculta e a função Softmax para a camada de saída, que foram mantidas por encontrar embasamento na literatura, de acordo com o item 2.2.1. A Figura 9 ilustra graficamente a RNA implementada, onde há 13 fatores de entrada, 01 camada oculta (com 08 neurônios como exemplo), uma camada de saída com 01 neurônio e uma saída com apenas 01 fator.

Figura 9 - Configuração de implementação da RNA



Fonte: o autor (2018).

- c) Treinamento: o algoritmo SCG (*Scaled Conjugate Gradient*) é predefinido pelo *software* e foi mantido devido ao estabelecido no item 2.2.4.2.

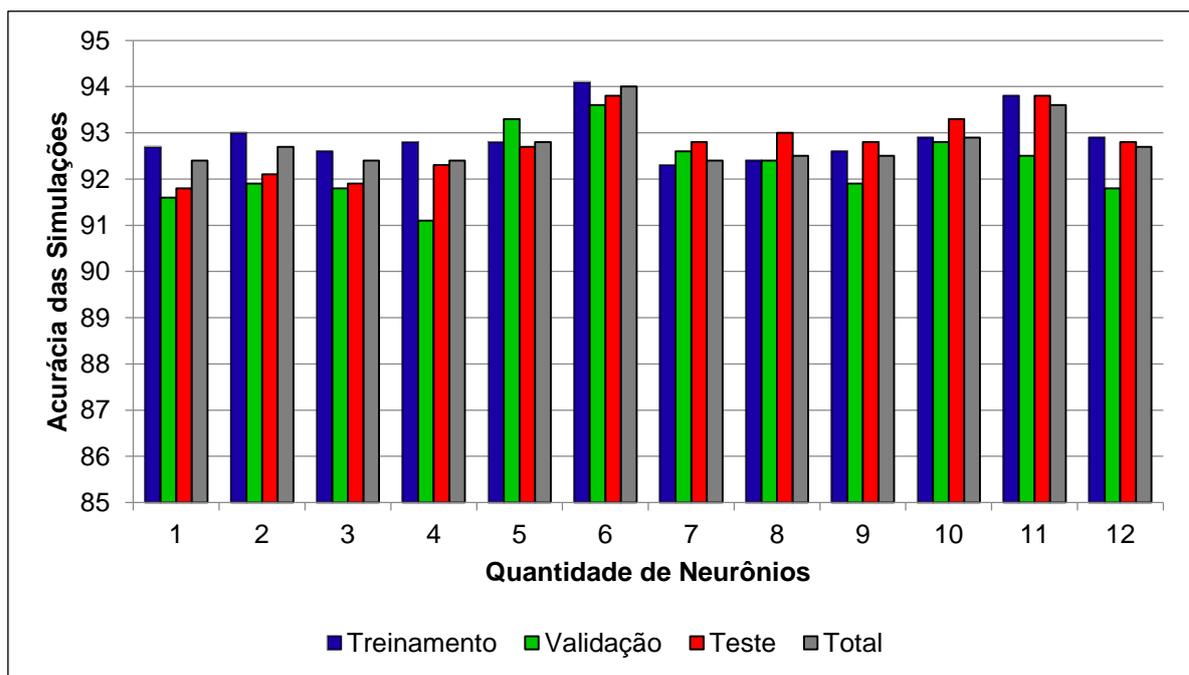
### 3.3.2 Processamento por tipo de graduação

Visto que os ajustes de pesos das RNAs são feitos levando em consideração os dados a elas fornecidos, faz-se necessário redefinir os parâmetros de treinamento para encontrar a quantidade adequada de neurônios para os novos conjuntos.

#### 3.3.2.1 Determinação dos parâmetros de treinamento

Tal qual o processamento completo, para o processamento por tipo de graduação optou-se por manter a técnica dos algoritmos construtivos e acrescentar neurônios gradativamente com o intuito de encontrar o ponto ótimo.

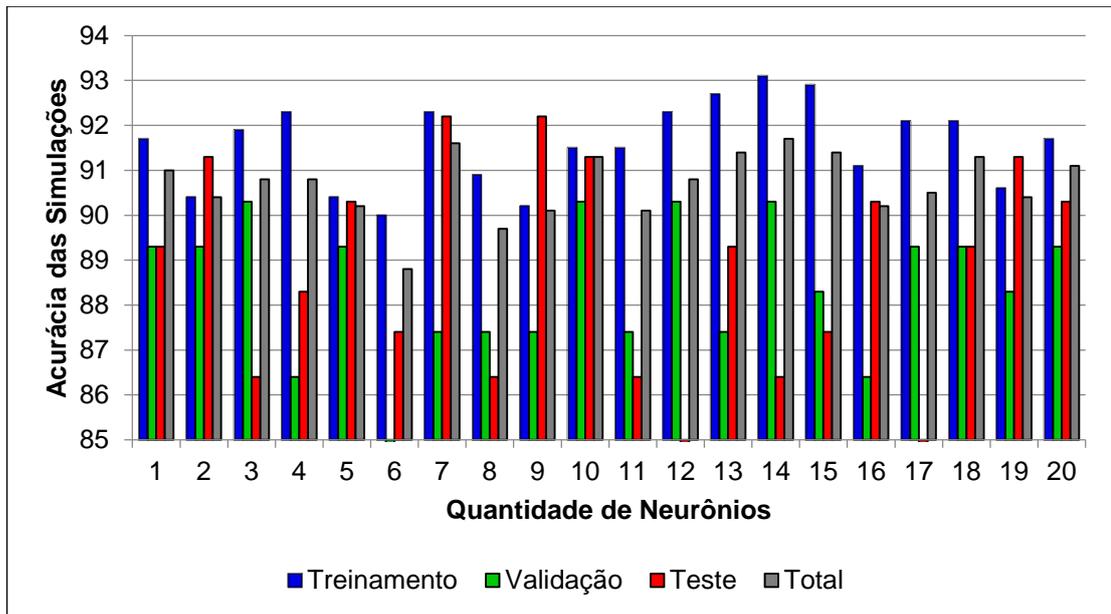
Figura 10 - Definição da quantidade de neurônios - bacharelado



Fonte: o autor (2019)

Pode ser verificado na Figura 10 que, com 6 neurônios, houve o melhor desempenho de todos os subconjuntos, sendo que o subconjunto de validação obteve acurácia de 93,6%. Utilizando o critério de duplicar o número de processamentos a partir do ponto ótimo encontrado, os mesmos foram interrompidos após o processamento com 12 neurônios.

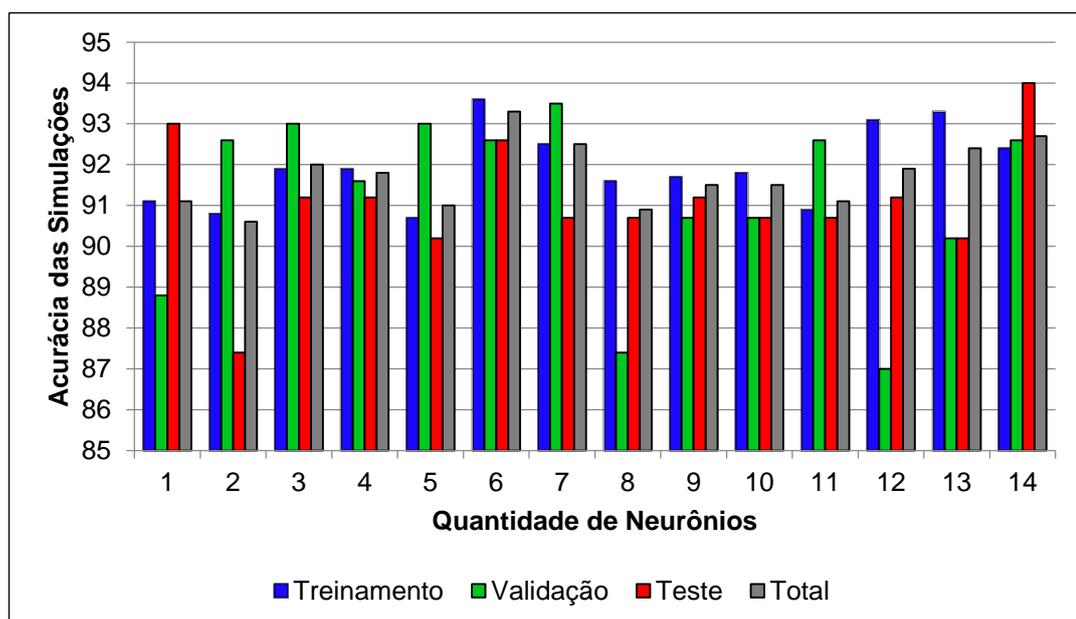
Figura 11 - Definição da quantidade de neurônios - licenciatura



Fonte: o autor (2019)

Diferentemente do processamento para definição do número de neurônios para os cursos bacharelado onde a acurácia para cada número de neurônios apresentou constância, para os cursos de licenciatura verificou-se uma relevante diferença entre os subconjuntos, conforme mostra a Figura 11 sendo que a melhor acurácia das simulações foi obtida com 10 neurônios, levando em consideração a acurácia de 90,3% para o subconjunto de validação.

Figura 12 - Definição da quantidade de neurônios - tecnologia



Fonte: o autor (2019)

Utilizando o mesmo critério para interrupção dos testes, verifica-se que na Figura 12, com 7 neurônios, a RNA que processou apenas os cursos de tecnologia obteve a melhor acurácia do subconjunto de validação: 93,5%.

### 3.3.2.2 Definição do método de amostragem

Mantendo o entendimento do processamento completo dos dados, optou-se por seguir com o método *holdout* para os processamentos por tipo de graduação seguindo, da mesma forma, a escolha percentual de 70% para o subconjunto de treinamento, 15% para o subconjunto de validação e 15% para o subconjunto de teste. O Quadro 6 ilustra a quantidade de amostras utilizada por subconjunto para cada um dos tipos de graduação:

Quadro 6 - Amostras por subconjunto

Graduação	Treinamento	Validação	Teste	Total
Bacharelado	7464	1600	1600	10664
Licenciatura	481	103	103	687
Tecnologia	1004	215	215	1434

Fonte: o autor (2019)

Pode-se perceber que os cursos de bacharelado representam a maioria das amostras utilizadas no processamento completo – cerca de 83%, sendo este o processamento por tipo de graduação que apresentou maior estabilidade de acurácia entre os subconjuntos, o que pode estar relacionado ao número de amostras disponibilizadas para a aprendizagem do algoritmo. Cabe ressaltar, no entanto, que não é possível aumentar a quantidade de amostras nos cursos de licenciatura e tecnologia pois a base utilizada compreende todos os dados disponíveis na Instituição.

### 3.3.2.3 Análise dos parâmetros predefinidos

Os parâmetros predefinidos pelo software foram revisados para o processamento por tipo de graduação. O Quadro 7 apresenta o resumo dos parâmetros de cada RNA por tipo de graduação.

Quadro 7 - Configuração das RNAs por tipo de graduação

Graduação	Entradas	Camadas Ocultas	Neurônios	1º Função de ativação	Camada de Saída	2º Função de ativação	Saídas
Bacharelado	13	1	6	Tang. Hiperb	1	Softmax	1
Licenciatura	13	1	10	Tang. Hiperb	1	Softmax	1
Tecnologia	13	1	7	Tang. Hiperb	1	Softmax	1

Fonte: o autor (2019)

Pode ser verificado que se optou por manter o mesmo número de camadas ocultas para os três tipos de graduação, as funções de ativação e o algoritmo de treinamento. Assim sendo o número de neurônios foi o único parâmetro diferente do processamento completo.

### 3.4 ANÁLISE DOS RESULTADOS

A análise de resultados de uma RNA utilizada para reconhecimento de padrões compreende principalmente a matriz de confusão e a curva ROC (*Receiver Operating Characteristic*), visto que representam gráfica e estatisticamente a capacidade de reconhecimento do treinamento efetuado. Há ainda a possibilidade de visualizar a correlação entre as variáveis da matriz de entrada com a matriz alvo, a fim de identificar se algum conjunto de dados possui capacidade de induzir de maneira indesejada a matriz alvo.

#### 3.4.1 Matriz de Confusão

O *software* escolhido para a implementação da RNA permite a visualização de quatro matrizes de confusão após o treinamento da rede, sendo uma para cada conjunto de dados e uma com o somatório dos dados processados.

A importância de avaliar os conjuntos separadamente consiste em identificar a acurácia de processamento de cada um deles e, neste sentido, o conjunto de validação recebe especial atenção pois é, por meio dele, que se pode verificar a existência de *overfitting* de processamento, uma vez que os dados do conjunto de validação não passaram pela etapa de treinamento da rede.

### 3.4.2. Curva ROC

Da mesma maneira que o *software* escolhido para aplicação da rede neural permite a visualização das matrizes de confusão para cada um dos conjuntos de dados, o mesmo ocorre para as curvas ROC. O objetivo, tal qual o anterior, é identificar a consistência e a similaridade das curvas nos conjuntos de dados, de tal maneira que, caso um dos mesmos se apresente destoante em relação aos demais, o processamento deva ser verificado.

No caso da análise individual espera-se que as curvas se apresentem com inclinações próximas ao canto superior esquerdo e distantes da reta diagonal, representando assim alta capacidade de identificar os padrões desejados no estudo.

### 3.4.3. Coeficiente de correlação

Neste estudo a identificação das correlações entre cada uma das variáveis e a matriz alvo busca identificar se algum dos fatores da matriz de entrada interfere fortemente sobre a matriz alvo. A partir desta análise pode-se identificar tendências que podem afetar o desempenho da rede, positiva ou negativamente. Os cálculos para definição dos coeficientes de correlação foram realizados pelo método de *Spearman*, conforme indicado em 2.2.5.3, devido tratarem-se de variáveis não-paramétricas, utilizando o *software* IBM SPSS Statistics®.

## 4 APLICAÇÃO DOS MÉTODOS

Neste capítulo são apresentados os resultados dos métodos inseridos no Capítulo 3, utilizando a fundamentação teórica estudada no Capítulo 2, bem como as considerações e conclusões pertinentes ao estudo realizado.

### 4.1 DEFINIÇÃO, COLETA E TRATAMENTO DOS DADOS

A etapa de definição, coleta e tratamento de dados resultou em uma menor quantidade de informações em relação ao que foi obtido inicialmente. O conjunto total de dados obtido compreendeu 16 parâmetros e 16057 matrículas, que resultaram, após o tratamento dos dados, em um total de 13 parâmetros e 12075 matrículas. Os parâmetros removidos foram o cadastro, o percentual cursado e a cidade de residência, devido não ser possível identificar as diferenças necessárias por meio deles.

Em relação ao número de matrículas, foram removidas 3982 entradas devido repetição do cadastro sem a mudança de situação, ou seja, estas matrículas possuíam apenas troca de currículo sem interrupção de pelo menos um semestre. O Quadro 8 apresenta o número de matrículas divididas por tipo de graduação, antes e após o tratamento.

Quadro 8 - Comparativo de dados após o tratamento

<b>Tipo de Graduação</b>	<b>Antes do Tratamento</b>	<b>Após o Tratamento</b>	<b>Saldo</b>
Bacharelado	13719	10664	- 3055
Licenciatura	791	687	- 104
Tecnologia	1547	1434	- 113

Fonte: o autor (2019)

É possível identificar que a redução da quantidade de dados se deu principalmente nos cursos de bacharelado (cerca de 20%), porém, mesmo após o tratamento, este tipo de graduação permaneceu com a maioria da base de dados utilizada para o processamento.

## 4.2 IMPLEMENTAÇÃO E PROCESSAMENTO DOS MODELOS

A implementação e processamento dos modelos foi subdividida em duas etapas: o processamento completo e o processamento por tipo de graduação, os quais objetivam identificar as características obtidas em cada uma RNAs.

### 4.2.1 Processamento completo

Após a definição dos parâmetros de processamento conforme indicado no Capítulo 3, a RNA completa foi treinada para visualização dos resultados na *toolbox* que, de acordo com as definições dadas, teve as características apresentadas no Quadro 9.

Quadro 9 - Parâmetros de processamento

<b>Parâmetro</b>	<b>Valor</b>
Número de iterações	101
Tempo	0:00:01
Desempenho	0,162
Gradiente	0,00572
Checagens de validação	6

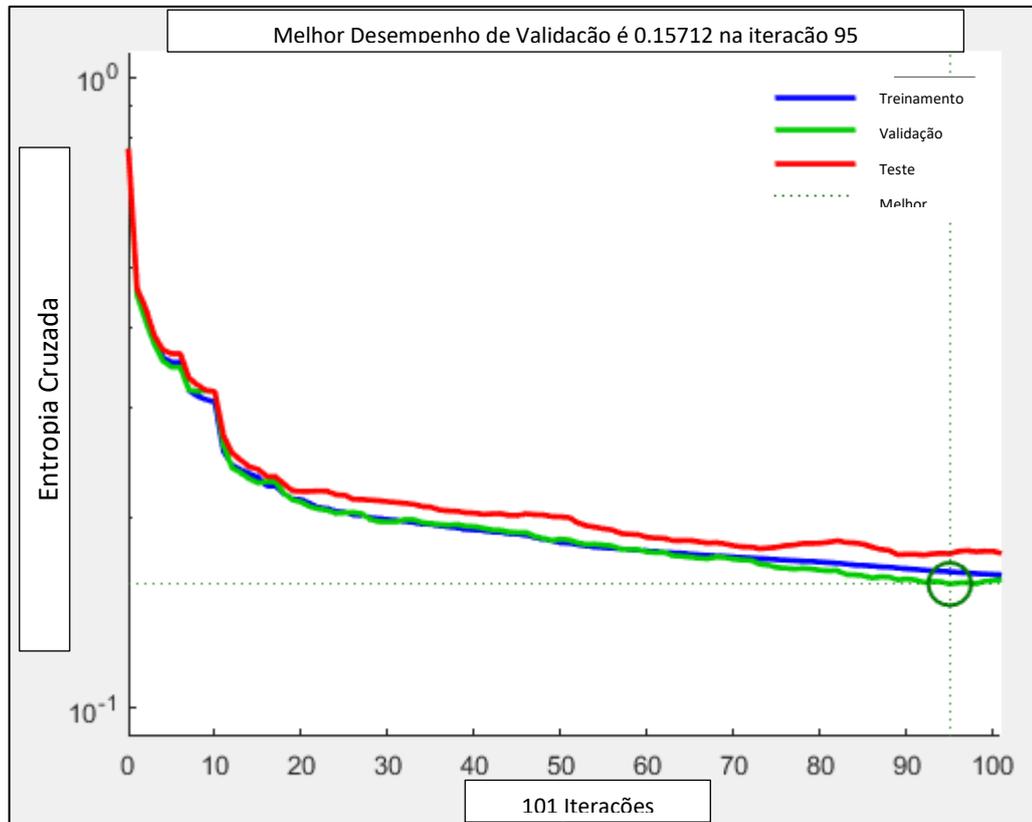
Fonte: o autor (2018)

Em relação aos parâmetros observados no treinamento, pode-se verificar que o critério de parada acionado pelo *software* foi o atingimento do número de checagens de validação (6), o que garantiu a inexistência de *overfitting* no processamento. Quanto ao número de iterações, o treinamento foi encerrado na iteração 101 após 01 segundo de treinamento. O desempenho atingido, para o conjunto de treinamento, foi de 0,162, sendo que a meta determinada pelo algoritmo era 0. Quanto ao gradiente, o valor predefinido como limite de parada adotando este critério era  $1 \times 10^{-6}$  e o treinamento atingiu um gradiente de 0,00572.

Apesar de os valores dos parâmetros, de forma tabelada, explicitarem o comportamento do treinamento, a visualização do mesmo graficamente permite extrair outras informações relevantes a respeito do processamento para cada um dos subconjuntos. A

Figura 13 ilustra o decaimento do custo em função do número de iterações de treinamento.

Figura 13 - Número de iterações x função de custo



Fonte: o autor (2018)

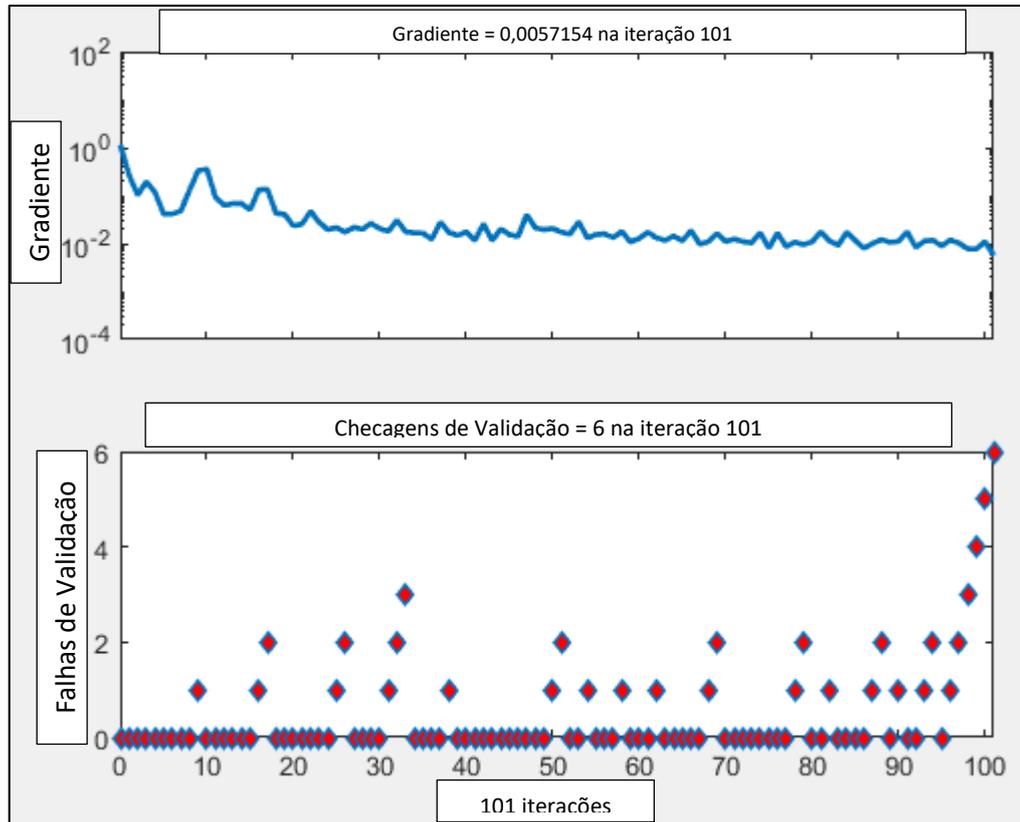
Além do número de iterações e do custo associado, fica evidenciado que o treinamento foi encerrado na iteração 101 devido ao mínimo custo do conjunto de validação ter sido encontrado na iteração 95, sendo testado por mais 6 iterações até a iteração 101.

Os conjuntos de treinamento e teste, à iteração 101, apresentaram maior custo, porém, caso o treinamento tivesse continuado o conjunto de treinamento possivelmente teria seu valor reduzido abaixo do valor de validação, visto o comportamento da curva, porém, para evitar o *overfitting* a curva que necessita ser analisada é a do conjunto de validação.

Como informação adicional pode ser verificado ainda que o custo do treinamento caiu fortemente nas primeiras 20 iterações, demonstrando que o algoritmo de treinamento cumpriu com o esperado e convergiu às regiões de mínimo logo após o início do processo.

De maneira complementar ao citado anteriormente pode se visualizado com mais detalhes na Figura 14 a descida do gradiente e as checagens de validação ocorridas ao longo de todo o processo de treinamento.

Figura 14 - Número de iterações x gradiente e checagens de validação



Fonte: o autor (2018)

A análise da descida do gradiente em função do número de iterações corrobora o argumento de que o algoritmo cumpriu o esperado, uma vez que a partir da iteração 20 até a iteração 101 o valor do gradiente foi praticamente estável, atingindo valores de mínimo local seguidas vezes sem ficar preso aos mesmos.

Já as checagens de validação demonstram que valores mínimos do conjunto de validação foram encontrados antes da iteração 10 e se repetiram em variados momentos até a iteração 101, porém, não superando as seis validações consecutivas necessárias para que o treinamento fosse encerrado.

#### 4.2.2 Processamento por tipo de graduação

O processamento por tipo de graduação foi realizado da mesma forma que o processamento completo, porém, devido à particularidade de cada RNA, os mesmos obtiveram diferentes resultados finais, conforme ilustrado no Quadro 10.

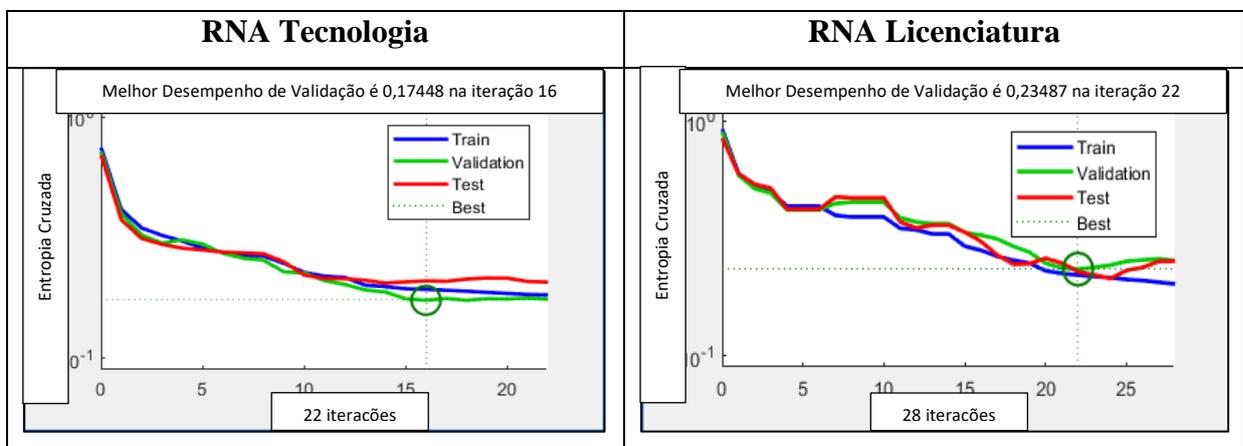
Quadro 10 - Parâmetros de processamento - por tipo de graduação

Graduação	Iterações	Desempenho	Gradiente	Checagens de Validação
Bacharelado	104	0,14612	0,00869	6
Licenciatura	22	0,23487	0,0675	
Tecnologia	16	0,17448	0,0246	

Fonte: o autor (2019)

Nos resultados apresentados verifica-se que a RNA que processou os cursos de bacharelado apresentou o melhor desempenho e gradiente e a maior quantidade de iterações. Este fato, no entanto, não representa um problema visto que o tempo de processamento foi reduzido. Já as RNAs dos cursos de licenciatura e tecnologia obtiveram menores números de iterações, porém, também menor desempenho e maior gradiente, ou seja, um resultado inferior a RNA dos cursos de bacharelado. A Figura 15 ilustra o desempenho das RNAs que processaram os cursos de tecnologia e licenciatura.

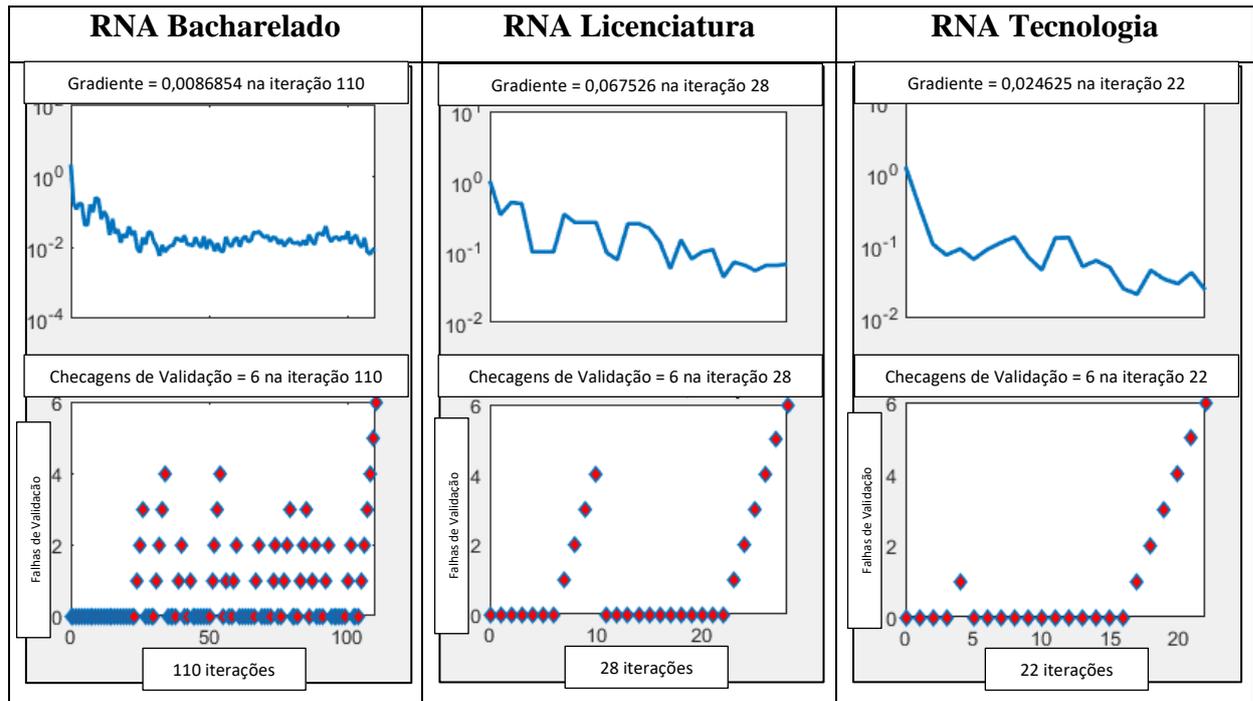
Figura 15 - Desempenho RNAs Tecnologia e Licenciatura



Fonte: o autor (2019)

O desempenho das RNAs apresentadas demonstra um decaimento diferente entre ambas, tendo o processamento dos cursos de tecnologia um decaimento mais uniforme sem apresentação de picos. Esta visão pode ser complementada por meio da Figura 16, que apresenta os gradientes e as checagens de validação das RNAs de cada tipo de graduação.

Figura 16 - Gradientes e Checagens de Validação - processamento por tipo de graduação



Fonte: o autor (2019)

Pode-se perceber, por meio da análise da Figura 16 que o processamento das RNAs dos cursos de bacharelado e tecnologia apresentaram queda acentuada do gradiente no início do aprendizado e um equilíbrio ao fim do aprendizado, enquanto que, na RNA de licenciatura, este comportamento é menos intenso, sendo a descida do gradiente irregular e com picos. Em termos de checagens de validação fica evidente que na RNA de bacharelado houve mais tentativas de encerrar o treinamento, porém, deve ser levado em consideração o número de iterações que a RNA foi treinada, de 4 a 5 vezes mais do que as RNAs de licenciatura e tecnologia.

#### 4.3 ANÁLISE DOS RESULTADOS

A análise dos resultados foi dividida em três partes para que possa ser verificada a capacidade de reconhecimento de padrões da rede neural, sua representatividade para a

Instituição e a importância da utilização da ferramenta para obtenção dos resultados propostos, sendo ainda dividida em processamento completo e processamento por tipo de graduação.

#### 4.3.1 Processamento completo

A análise dos resultados no processamento completo tem o objetivo de identificar a capacidade de generalização por meio de um único processamento, que não diferencia o tipo de graduação e compreende a maior quantidade de dados possível.

##### 4.3.1.1 Matriz de Confusão

O *software* utilizado possibilita a visualização da matriz de confusão de cada um dos conjuntos de dados e, por fim, uma única matriz com o somatório dos conjuntos. A matriz de confusão do conjunto de treinamento, conforme mostrado na Figura 17, obteve acurácia de 93,4% (célula azul). Além disto, podem ser verificados ainda a quantidade de amostras e o percentual de acertos (células verdes) e a quantidade de amostras e o percentual de erros (células vermelhas). As células de cor cinza representam os percentuais de acerto (fonte verde) e de erro (fonte vermelha) por linhas e colunas.

Figura 17 - Matriz de Confusão para o Conjunto de Treinamento

saída	0	7587 84,8%	384 4,3%	95,2% 4,8%
	1	209 2,3%	769 8,6%	78,6% 21,4%
		97,3% 2,7%	66,7% 33,3%	93,4% 6,6%
		0	1	target

Fonte: o autor (2018)

O conjunto de treinamento fornece métrica importante para avaliar o desempenho da RNA pois é o conjunto com a maior quantidade de dados (70%), porém, para que a avaliação seja consistente é necessário levar em consideração também a acurácia obtida nos conjuntos de validação e de teste.

Para o conjunto de validação, conforme demonstrado na Figura 18, a acurácia do teste foi de 93,0%, apenas 0,4% inferior à acurácia do conjunto de treinamento. O resultado do conjunto de validação é similar ao do conjunto de treinamento e comprova que a rede possui alta capacidade de identificar o padrão desejado.

Figura 18 - Matriz de Confusão para o Conjunto de Validação

saída	0	1633 85,1%	90 4,7%	94,8% 5,2%
	1	45 2,3%	150 7,8%	76,9% 23,1%
		97,3% 2,7%	62,5% 37,5%	93,0% 7,0%
		0	1	target

Fonte: o autor (2018)

Quanto ao conjunto de teste, conforme Figura 19 a seguir, que possui a mesma quantidade de dados do conjunto de validação, a acurácia foi de 92,4%, 0,6% menor do que do conjunto de validação e 1,0% menor do que a acurácia do conjunto de treinamento.

Figura 19 - Matriz de Confusão para o Conjunto de Teste

saída	0	1600 83,4%	91 4,7%	94,6% 5,4%
	1	54 2,8%	173 9,0%	76,2% 23,8%
		96,7% 3,3%	65,5% 34,5%	92,4% 7,6%
		0	1	target

Fonte: o autor (2018)

Por fim, para extração de uma métrica única sobre capacidade de identificação de padrões da rede neural criada pode ser visualizada a matriz de confusão total, que é um somatório dos acertos e erros dos conjuntos de treinamento, validação e teste, com o respectivo percentual em relação ao total, conforme Figura 20.

Figura 20 - Matriz de Confusão Total

saída	0	10820 84,6%	565 4,4%	95,0% 5,0%
	1	308 2,4%	1092 8,5%	78,0% 22,0%
		97,2% 2,8%	65,9% 34,1%	93,2% 6,8%
		0	1	target

Fonte: o autor (2018)

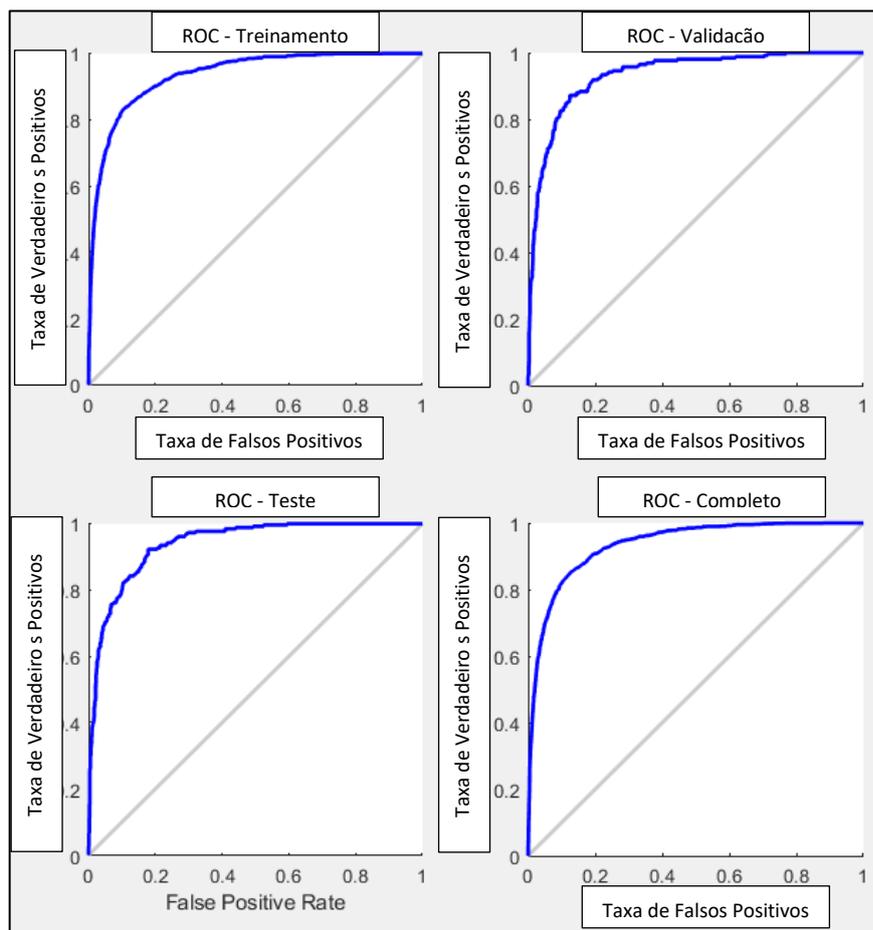
A matriz de confusão total obtida para a rede neural artificial, levando em consideração o somatório de todos os conjuntos, obteve acurácia de 93,2%, ou seja, o algoritmo utilizado acertou a classificação 11912 amostras do total de 12785 matrículas,

resumindo assim que a rede neural em questão, em sua totalidade, tem elevada taxa de assertividade na identificação do padrão desejado.

#### 4.3.1.2 Curva ROC

Tal qual a matriz de confusão é apresentada para cada um dos conjuntos de dados e também para o total dos dados, o mesmo ocorre para a curva ROC. Devido a similaridade das curvas e, levando em consideração que as matrizes de confusão apresentadas comprovam que o algoritmo apresenta comportamento semelhante para todos os conjuntos de dados, as mesmas são apresentadas juntas na Figura 21.

Figura 21 - Curvas ROC para os conjuntos de dados



Fonte: o autor (2018)

De acordo com as curvas ROC (linhas azuis na Figura 21), é possível identificar que a rede neural possui alto poder discriminante, ou seja, a capacidades do teste em distinguir

indivíduos com tendência à evasão dos estudos visto que as mesmas situam-se próximas ao canto esquerdo superior.

Pode ser verificado que, por meio do cálculo da sensibilidade e da especificidade, o teste em questão possui alta taxa de sensibilidade (95%) e moderada taxa de especificidade (79%). Isto significa que a aplicação desta rede resultará na identificação de parte relevante dos acadêmicos com tendência à evasão, porém, devido a moderada especificidade pode considerar acadêmicos que não possuem propensão à evasão como grupo de risco para realização de ações de retenção.

Considerando a temática do trabalho tem-se como adequada a rede com alta sensibilidade por considerar que, caso algum acadêmico seja erroneamente inscrito no grupo de propensos à evasão, este não sofre danos – apenas receberia maior atenção e cuidado institucional, o que não é entendido como prejudicial ao mesmo.

Apesar de o custo do erro não ser mapeado para este estudo, visto ser necessário ligá-lo a possíveis ações de retenção, toma-se como pressuposto que este é irrelevante se comparado à interrupção de, no mínimo, um semestre, que foi o critério adotado para mapeá-lo no grupo de risco de propensão à evasão. Sendo assim, pelo viés da Instituição, se mostra adequado um teste com maior sensibilidade, mesmo que oferecendo menor especificidade.

#### 4.3.1.3 Coeficiente de correlação

Os coeficientes de correlação de Spearman foram calculados individualmente para cada uma das variáveis de entrada em relação a matriz alvo, com significância 0,01, e estão mostradas no Quadro 11.

Quadro 11 - Coeficientes de Correlação

<b>Identificação</b>	<b>Variável</b>	<b>Coeficiente de Correlação</b>
1	Curso	+ 0,038
2	Currículo	- 0,085
3	Sexo	- 0,100
4	Ano de nascimento	- 0,102
5	Tipo de escola (pública ou privada)	- 0,140
6	Ano de ingresso na Instituição	- 0,194
7	Média harmônica do Vestibular	+ 0,068

8	Ano/semestre da última matrícula	+ 0,194
9	Cidade de residência	- 0,009
10	Possui débito com a Instituição	- 0,098
11	Faz parte do Programa Universidade para Todos (Pro Uni)	+ 0,074
12	Possui financiamento estudantil	+ 0,060
13	Tipo de curso	- 0,021

Fonte: o autor (2018)

Pode-se perceber, por meio dos coeficientes de correlação, que, isoladamente, nenhuma variável possui relação média ou forte com a matriz alvo, uma vez que todos os coeficientes ficaram determinados de  $-0,2$  a  $+0,2$ . Caso uma das variáveis apresentasse forte correlação com a matriz alvo, em detrimento das outras, esta poderia influenciar diretamente no resultado da rede neural, comprometendo a decisão sobre sua utilização. Desta maneira conclui-se que não se pode identificar os grupos com tendência à evasão do ciclo de estudos analisando apenas a correlação dos variados fatores compostos no estudo, cabendo à técnica das Redes Neurais Artificiais o processamento destes para entrega do resultado esperado.

#### 4.3.2 Processamento por tipo de graduação

A análise dos resultados por tipo de graduação tem o objetivo de comparar os desempenhos obtidos por cada uma das RNAs processadas, buscando identificar qual das mesmas obteve resultados mais satisfatórios e se, individualmente, alguma apresenta resultado superior a RNA do processamento completo.

##### 4.3.2.1 Matriz de Confusão

A comparação de desempenho por meio das matrizes de confusão é fundamental para identificar a capacidade de generalização das RNAs utilizadas no estudo. A Figura 22 ilustra as matrizes de confusão dos subconjuntos de treinamento por tipo de graduação.

Figura 22 - Matrizes de Confusão - Subconjuntos de treinamento

		Bacharelado	Licenciatura	Tecnologia
saída	0	6359 85,2%	266 3,6%	96,0%
	1	172 2,3%	667 8,9%	79,5%
		97,4%	71,5%	94,1%
		0	1	
		target		

		Bacharelado	Licenciatura	Tecnologia
saída	0	349 72,6%	23 4,8%	93,8%
	1	18 3,7%	91 18,9%	83,5%
		95,1%	79,8%	91,5%
		0	1	
		target		

		Bacharelado	Licenciatura	Tecnologia
saída	0	880 87,6%	59 5,9%	93,7%
	1	14 1,4%	51 5,1%	78,5%
		98,4%	46,4%	92,7%
		0	1	
		target		

Fonte: o autor (2019)

Percebe-se que todos os subconjuntos apresentaram resultados satisfatórios, porém, o subconjunto do bacharelado obteve a melhor capacidade de generalização - 94,1%. Constatase ainda que este foi o subconjunto com o maior número de amostras, o que auxilia o aprendizado da RNA. Para o subconjunto de validação, a Figura 23 ilustra o comparativo.

Figura 23 - Matrizes de Confusão - Subconjuntos de validação

		Bacharelado	Licenciatura	Tecnologia
saída	0	1357 84,6%	63 3,9%	95,6%
	1	40 2,5%	140 8,8%	77,8%
		97,1%	69,0%	93,6%
		0	1	
		target		

		Bacharelado	Licenciatura	Tecnologia
saída	0	72 69,9%	6 5,8%	92,3%
	1	4 3,9%	21 20,4%	84,0%
		94,7%	77,8%	90,3%
		0	1	
		target		

		Bacharelado	Licenciatura	Tecnologia
saída	0	191 88,8%	10 4,7%	95,0%
	1	4 1,9%	10 4,7%	71,4%
		97,9%	50,0%	93,5%
		0	1	
		target		

Fonte: o autor (2019)

Nas matrizes de confusão dos subconjuntos de validação a RNA dos cursos de bacharelado obteve novamente o melhor resultado, porém, com resultado muito próximo da RNA dos cursos de tecnologia, apesar da diferença entre o tamanho de amostras. Constatase então que, apesar de poucos dados, o subconjunto dos cursos de tecnologia possui um

processamento seguro e capaz de identificar o padrão em novos dados. A Figura 24 apresenta as matrizes de confusão dos subconjuntos de teste.

Figura 24 - Matrizes de Confusão - Subconjuntos de teste

		Bacharelado			Licenciatura			Tecnologia		
saída	0	1359 84,9%	60 3,8%	95,8% 4,2%	69 67,0%	2 1,9%	97,2% 2,8%	188 87,4%	14 6,5%	93,1% 6,9%
	1	39 2,4%	142 8,9%	78,5% 21,5%	7 6,8%	25 24,3%	78,1% 21,9%	6 2,8%	7 3,3%	53,8% 46,2%
		97,2% 2,8%	70,3% 29,7%	93,8% 6,2%	90,8% 9,2%	92,6% 7,4%	91,3% 8,7%	96,9% 3,1%	33,3% 66,7%	90,7% 9,3%
		0	1		0	1		0	1	
		target			target			target		

Fonte: o autor (2019)

Os resultados dos subconjuntos de teste comprovam que a RNA dos cursos de bacharelado possui a melhor capacidade de generalização (93,8%), uma vez que apresenta valores substancialmente maiores do que os outros subconjuntos. A Figura 25 apresenta as matrizes de confusão para o total de dados de cada tipo de graduação.

Figura 25 - Matrizes de Confusão - Todos os subconjuntos

		Bacharelado			Licenciatura			Tecnologia		
saída	0	9075 85,1%	389 3,6%	95,9% 4,1%	490 71,3%	31 4,5%	94,0% 6,0%	1259 87,8%	83 5,8%	93,8% 6,2%
	1	251 2,4%	949 8,9%	79,1% 20,9%	29 4,2%	137 19,9%	82,5% 17,5%	24 1,7%	68 4,7%	73,9% 26,1%
		97,3% 2,7%	70,9% 29,1%	94,0% 6,0%	94,4% 5,6%	81,5% 18,5%	91,3% 8,7%	98,1% 1,9%	45,0% 55,0%	92,5% 7,5%
		0	1		0	1		0	1	
		target			target			target		

Fonte: o autor (2019)

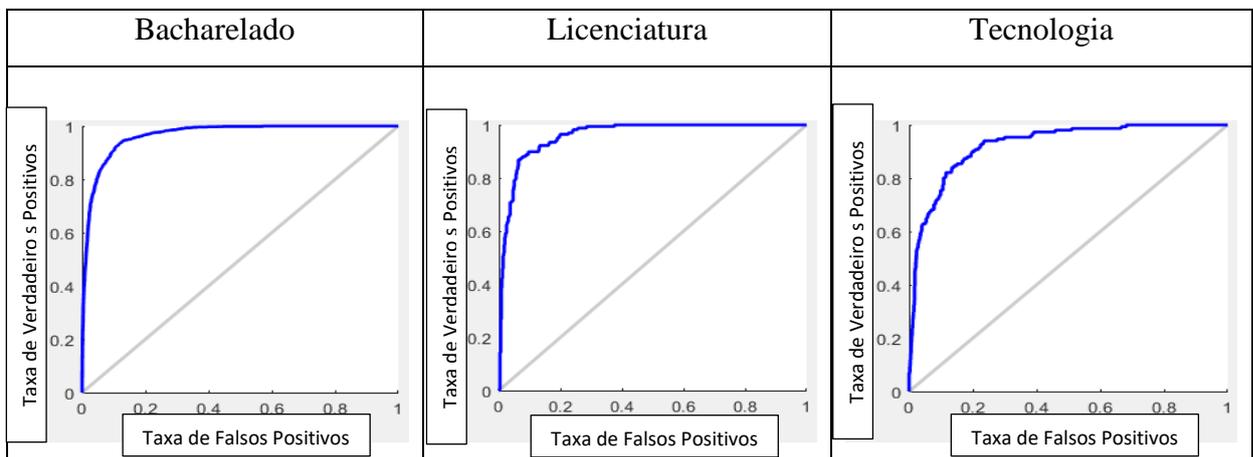
Comprova-se, a partir da Figura 25, que a RNA dos cursos de bacharelado obteve a melhor capacidade de generalização em comparação aos demais subconjuntos – 94,0% de acurácia contra 91,3% para o subconjunto dos cursos de licenciatura e 92,5% para o

subconjunto dos cursos de tecnologia. Apesar da diferença, considera-se que todos os subconjuntos obtiveram resultados satisfatórios nos processamentos realizados.

#### 4.3.2.2 Curva ROC

Para a análise das curvas ROC, visto que sua representatividade se dá graficamente por meio do posicionamento da curva em relação ao vértice superior esquerdo, optou-se por apresentar apenas o comparativo da totalidade dos dados dos subconjuntos de bacharelado, licenciatura e tecnologia. As curvas de desempenho dos processamentos são representadas pelas linhas azuis, conforme demonstrado na Figura 26.

Figura 26 - Curvas ROC - Todos os subconjuntos



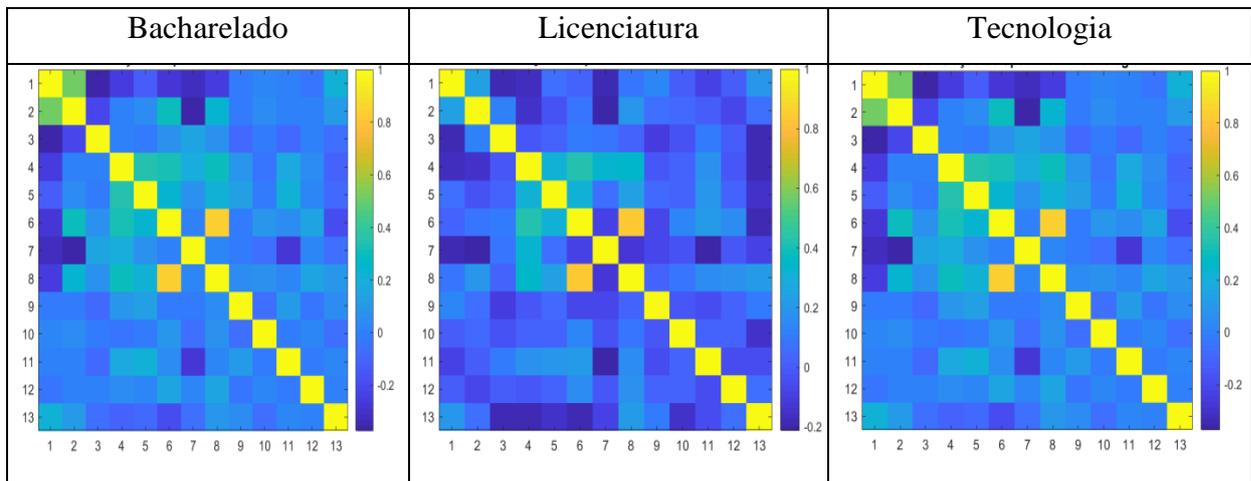
Fonte: o autor (2019)

De acordo com a Figura 26 todos os subconjuntos apresentaram capacidade adequada de generalização, com destaque para a RNA dos cursos de bacharelado, que apresenta a maior área abaixo da curva. No entanto, as RNAs dos cursos de licenciatura e tecnologia, apesar de áreas menores, também apresentaram resultados satisfatórios.

#### 4.3.2.3 Coeficiente de correlação

A análise dos coeficientes de correlação por tipo de graduação tem por objetivo verificar, dentro de cada subconjunto, possíveis correlações fortes entre as próprias variáveis, bem como as correlações entre as variáveis e a matriz alvo. A Figura 27 ilustra a correlação entre cada uma das variáveis de entrada em todos os subconjuntos.

Figura 27 - Coeficiente de Correlação - todos os subconjuntos



Fonte: o autor (2019)

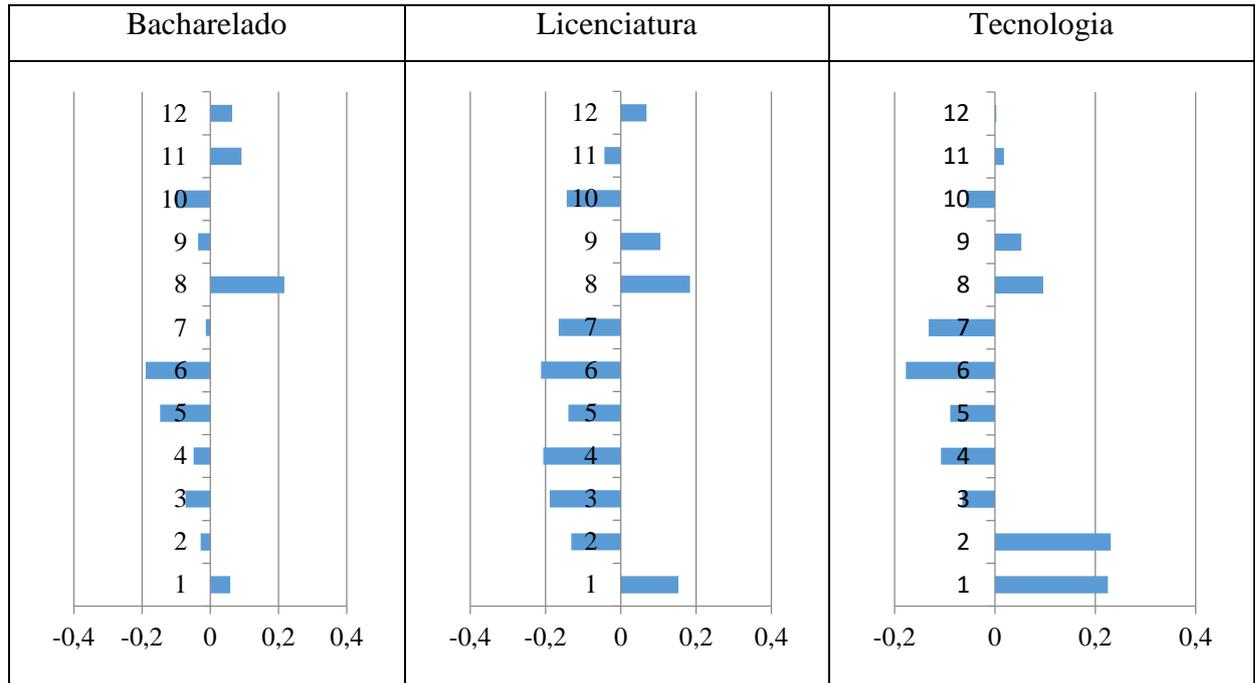
A Figura 27 apresenta a correlação entre cada um dos fatores de entrada, identificados de 01 a 13 conforme o Quadro 11. A força da correlação entre as mesmas é representada por cores, sendo as mais fracas as azuis escuras (-0,2) e as mais fortes as amarelas (+1,0). As correlações intermediárias podem ser verificadas pelas cores laranja, verde e azul claro e a escala lateral, presente em todos os subconjuntos, identifica e orienta sobre a força das correlações.

A análise da figura em questão permite identificar que a diagonal principal de todos os subconjuntos é amarela, visto representar o ponto em que a correlação se dá entre as mesmas variáveis. As demais correlações entre as variáveis, em todos os subconjuntos, são, em sua maioria, fracas (até 0,2).

Percebe-se, porém, que as variáveis 6 (ano de ingresso) e 8 (semestre da última matrícula) possuem forte correlação em todos os subconjuntos. Este comportamento, no entanto, já era esperado pela natureza do estudo, onde estas variáveis tem ligação direta.

A Figura 28 ilustra o coeficiente de correlação entre as variáveis de entrada e a matriz alvo. Para a construção desta análise foi removida a variável correspondente ao tipo de curso, visto ser este ser o parâmetro para definição dos subconjuntos. A identificação das variáveis refere-se a identificação apresentada no Quadro 11.

Figura 28 - Coeficientes de Correlação com Matriz Alvo - todos os subconjuntos



Fonte: o autor (2019)

Verifica-se que as correlações entre as variáveis de entrada e a matriz alvo variam, em todos os subconjuntos, em sua maioria, de  $-0,2$  a  $+0,2$ , ou seja, são consideradas fracas, apesar de se mostrarem superiores aos valores encontrados dos coeficientes do processamento completo. Esta análise permite concluir que nenhuma variável, isoladamente, influenciou o resultado do processamento da RNA e que o elevado resultado de acurácia de previsão se deu mediante a utilização desta ferramenta.

## 5 CONCLUSÕES

Este capítulo foi dividido em duas partes: as considerações finais e as sugestões de trabalhos futuros, conforme apresentado a seguir.

### 5.1 CONSIDERAÇÕES FINAIS

O presente trabalho demonstrou atingir os objetivos propostos inicialmente, visto que apresentou alta taxa de acurácia de acertos de alunos com propensão à evasão dos estudos, tanto na RNA de processamento completo quanto nas RNAs que processaram separadamente os cursos por tipo de graduação. Além disso, por meio dos cálculos dos coeficientes de correlação, demonstrou ser uma ferramenta essencial para este tipo de análise pois permite identificar se algum dos dados interfere diretamente no processamento da RNA.

A realização do estudo foi dificultada devido a definição das variáveis, realizada de maneira empírica visto que não existe, na literatura ou em estudos relacionados, uma conclusão a respeito do processo de escolha, bem como a necessidade de tratar repetidamente os dados por inconsistências encontradas na base dos mesmos. Já a aplicação da rede neural no *software Matlab R2017b*® foi facilitada devido à intuitiva interface de treinamento e a possibilidade de exportação do algoritmo.

Deve ser ponderado ainda que a abordagem sobre acadêmicos que porventura enquadrem-se erroneamente no grupo de propensos a evasão não expõe prejuízos aos mesmos ou ao Instituição, visto que as medidas adotadas tendem a ser de maior proximidade e acolhimento em relação a suas demandas.

Enquanto contribuições, além do estudo em si, deseja-se evidenciar a necessidade de que seja criada uma cultura preventiva de resolução de problemas, uma vez que se concede tempo para que gestores acadêmicos e administrativos possam criar, implementar e monitorar medidas de maneira mais efetiva, tendo maior probabilidade de resultados positivos. Além disso, deseja-se frisar que, em muitos casos, as ferramentas e o conhecimento para resolução de problemas existentes estão mais próximos do que se pressupunha – cabe então conhecer os recursos existentes e criar uma cultura de incentivo e valorização dos agentes que se propõe a aplicar estas metodologias.

## 5.2 SUGESTÕES DE TRABALHOS FUTUROS

A metodologia proposta neste estudo demonstrou uma alta capacidade de identificação do padrão desejado, porém, este pode ser aprimorado e maximizado. Algumas sugestões podem ser citadas para trabalhos futuros, como por exemplo:

- a) Incluir outros fatores relacionados ao curso, tais quais as percepções a respeito da qualidade, atualização e organização do mesmo, bem como a relação entre teoria e prática.
- b) Abranger aspectos pertencentes à estrutura física da Instituição, como qualidade dos laboratórios, conservação das salas de aula, limpeza de banheiros e atualização dos equipamentos presentes nas salas de aula.
- c) Englobar percepções pessoais a respeito da qualidade do atendimento, sendo ele do curso, da Área do Conhecimento, do coordenador e da Instituição, como um todo.
- d) Adicionar elementos relativos ao grau de motivação pessoal, nível de relacionamento e importância do relacionamento com colegas e amigos.
- e) Incorporar nos aspectos econômicos o trabalho do acadêmico, sua relação com o curso, a quantidade de horas disponibilizada para tal e o nível de compromisso assumido com este.
- f) Conter os dados relativos ao protocolo acadêmico e Ouvidoria da Instituição, como a quantidade de solicitações de cancelamentos de disciplinas, trocas de forma de pagamento e reclamações realizadas.

## REFERÊNCIAS BIBLIOGRÁFICAS

AMARAL, N. C. A educação superior brasileira: dilemas, desafios e comparações com os países da OCDE e do BRICS. **Revista Brasileira da Educação**, v. 21, n. 66, p. 717-736, 2016.

ALMEIDA, L. S.; SOARES, A. P. C.; FERREIRA, J. A. G. Questionário de Vivências Acadêmicas (QVA-r): avaliação do ajustamento de estudantes universitários. **Avaliação Psicológica**, v. 1, n. 2, p. 81-93, 2002.

BAGGI, C. A. S.; LOPES, D. A. Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica. **Avaliação**, v. 16, n. 2, p. 355-374, 2011.

BAYER, J.; BYDZOVSKA, H.; GERYK, J.; OBSIVAC, T.; POPELINSKY, L. Predicting drop-out from social behaviour of students. In: **Anais 5th International Conference on Educational Data Mining – EDM 2012**, Chania, 2012.

BARRETO, J. B. **Introdução às Redes Neurais Artificiais**. Universidade Federal de Santa Catarina, 2002.

BAPTISTA, F.; RODRIGUES, S.; MORGADO-DIAS, F. Performance comparison of ANN training algorithms for classification. **IEEE 8th International Symposium on Intelligent Signal Processing**, Funchal, 2013, p. 115-120.

BEZERRA, E. **Introdução à Aprendizagem Profunda**. Salvador: SBC, 2016.

BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LUDEMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: LTC, 2000.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. Instituto Brasileiro de Geografia e Estatística. PIB. Disponível em <[http://www.ibge.gov.br/home/estatistica/pesquisas/anos\\_anteriores\\_2016.shtm](http://www.ibge.gov.br/home/estatistica/pesquisas/anos_anteriores_2016.shtm)> Acesso em: mar. 2018.

BUENO, F. S. Minidicionário da língua portuguesa. São Paulo: FTD, 1996. 703 p.

CABRERA, A. F.; NORA, A.; CASTAÑEDA, M. B. The role of finances in the persistence process: a structural model. **Research in Higher Education**, v. 33, n. 5, p. 571-593, 1992.

CAMPELLO, A. V. C., LINS, L. N. Metodologia de análise e tratamento da evasão e retenção em cursos de graduação em instituições federais de ensino superior. In: **Anais XXVIII Encontro Nacional de Engenharia de Produção**, Rio de Janeiro, 2008.

CARDOSO, C. B. **Efeitos da política de cotas na Universidade de Brasília**: uma análise do rendimento e da evasão. 2008. 134 f. Dissertação (Mestrado em Educação) – Faculdade de Educação, Universidade de Brasília, Brasília, 2008.

CASTRO, L. N.; IYODA, E. M.; PINHEIRO, E.; VON ZUBEN, F. Redes neurais construtivas: uma abordagem comparativa. **Proceedings of the IV Brazilian Conference on Neural Networks**, São José dos Campos, 1999, p. 102-107.

ÇEBI, A.; KARAL, H. An application of fuzzy analytic hierarchy process (FAHP) for evaluating students Project. **Education Research and Reviews**, 2017, v. 12, p. 120-132.

DARTORA, G. A. **Redes neurais artificiais**: aplicação na tipificação de carcaça de suínos. 2003. 105 f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Santa Catarina, Florianópolis, 2003.

DEKKER, G.; PECHENIZKIY, M.; VLEESHOUWERS, J. Predicting students drop out: a case study. In: **Anais Proceeding of the International Conference on Educational Data Mining**, Córdoba, 2009, p. 41-50.

DEMUTH, H.; BEALE, M.; HAGAN, M. **Neural network toolbox – User's guide**, version 6.0.4. The Mathworks Inc., 2010, 901 p.

DOWD, A. C.; COURY, T. The effect of loans on the persistence and attainment of community college students. **Research in Higher Education**, v. 47, n. 1, p. 33-62, 2006.

FERNANDES, S.; FLORES, M. A. A docência no contexto da avaliação do desempenho no ensino superior: reflexões no âmbito de um estudo em curso. **Revista Iberoamericana de Evaluación Educativa**. v. 5, n. 2, p. 82-98, 2012.

FIELD, A. **Descobrimdo a estatística usando o SPSS**. 2. ed. Porto Alegre: Artmed. 2009.

FONSECA, R. C. V. **Metodologia do trabalho científico** – 1. ed. Curitiba: IESBE Brasil, 2012. 90 p.

GERHARDT, T. E.; SILVEIRA, D. T. **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009, 120 p.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GOLIK, P.; DOETSCH, P.; NEY, H. Cross-entropy vs. squared error training: a theoretical and experimental comparison. In: **Proceedings of the 14th Annual Conference of the International Speech Communication Association**, Lyon, 2013, p. 1756-1760.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Sinopse Estatística da Educação Superior 2016**. Brasília: Inep, 2017. Disponível em: < <http://inep.gov.br/web/guest/sinopses-estatisticas-da-educacao-superior>>. Acesso em: 18 03. 2018.

GAIOSO, N. P. L. **O fenômeno da evasão escolar na educação superior no Brasil**. 2005. 75 f. Dissertação (Mestrado em Educação) – Programa de Pós-Graduação em Educação da Universidade Católica de Brasília, Brasília, 2005.

HAYKIN, S. **Redes Neurais: princípios e práticas**. Porto Alegre: Bookman, 2001, 893 p.

KAYRI, M. An Intelligent Approach to Educational Data: Performance Comparison of the Multilayer Perceptron and the Radial Basis Function Artificial Neural Networks. **Educational Sciences: Theory and Practice**, 2015, v. 15, n. 15.

KARAMOUZIS, S. T.; VRETTOS, A. An Artificial Neural Network for Predicting Student Graduation Outcomes. **Proceedings of the World Congress on Engineering and Computer Science 2008**, 2008.

KIM, Y. BAYLOR, A. L. Research-Based Design of Pedagogical Agent Roles: a Review, Progress and Recommendations. **International Journal of Artificial Intelligence and Education**, 2016, v. 26, p. 160-169.

LARSON, R. FARBER, B. **Estatística Aplicada**. 4. ed. São Paulo: Pearson Prentice Hall, 2010. 658 p.

LOBO, M. B. C. M. Panorama de evasão no ensino superior brasileiro: aspectos gerais das causas e soluções. **Associação Brasileira de Mantenedoras de Ensino Superior. Cadernos**, n. 25, 2012.

MACEDO, C. **Evasão estudantil nos cursos de matemática, química e física da Universidade Federal Fluminense**: uma silenciosa problemática. 2012. 106 f. Dissertação (Mestrado em Serviço Social) – Departamento de Serviço Social, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2012.

MACKAY, D. J. C. **Information Theory, Inference and Learning Algorithms**. Cambridge: Cambridge University Press, 2005.

MARCONI, M. A.; LAKATOS, E. M. **Metodologia científica**. 6. ed. São Paulo: Atlas, 2011.

MARQUES, K. A.; MELO, A. F. F. Abordagens metodológicas no campo da pesquisa científica. In: **Anais do Simpósio de Metodologias Ativas**: inovações para o ensino e aprendizagem na educação básica e superior, São Paulo, 2017.

MARTINHO, V. R. C.; NUNES, C.; MINUSSI, C. R. Predição do grupo de risco de evasão discente em cursos superiores presenciais utilizando uma rede neural ARTMAP-Fuzzy. In: **Anais do Congresso Brasileiro de Educação em Engenharia – COBENGE 2013**, Gramado, 2013, v. 1, p. 1-6.

MANHÃES, L. M. B.; CRUZ, S. M. S.; COSTA, R. J. M.; ZAVALETA, J. ZIMBRÃO, G. Identificação dos fatores que influenciam a evasão em cursos de graduação através de sistemas baseados em mineração de dados: uma abordagem quantitativa. In: **Anais do VIII Simpósio Brasileiro de Sistemas de Informação**, São Paulo, 2012.

MIKOLOV, T.; KOMBRINK, S.; BURGET, L.; CERNOCKY, J.; KHUNDANPUR S.; Extensionsof recurrente neural network language model. **IEEE International Conference on Acoustics**, Praga, 2011.

MORAIS, E. C. **Reconhecimento de padrões e redes neurais artificiais em predição de estruturas secundárias de proteínas**. 2010. 135 f. Tese (Doutorado em Engenharia de Sistemas e Computação) – Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010.

MUSSO, M. F.; KYNDT, E.; CASCALLAR, E. C.; DOCHY, F. Predicting General Academic Performance and Identifying the Differential Contribution of Participating Variables Using Artificial Neural Networks. **Frontline Learning Research**, 2013, v. 1, n. 1.

NAGAOKA, M. E.; SILVA, I. N.; LANCAS, K. P. Um método para estimar densidade de solo utilizando redes neurais artificiais. **Energia na Agricultura**, 2005, v. 20, p. 53-68.

NASCIMENTO, E. O.; COSTA, R. G.; OLIVEIRA, L. N. Redes neurais artificiais aplicadas na correlação entre óbitos de dengue, automedicação e fatores abióticos em Goiânia-Goiás. **Scientia Plena**, 2017, v. 12, n. 3.

OLADOKUN, V. O.; ADEBANJO, A. T.; CHARLES-OWABA, O. E. Predictind Students Academic Performance using Artificial Neural Network: A Case Study of na Engineering Course. **The Pacific Journal of Science and Technology**, 2008, v. 9, n. 1.

PEREIRA, F. C. B. **Determinantes da evasão de alunos e os custos ocultos para as instituições de ensino superior**: uma aplicação na Universidade do Extremo Sul Catarinense. 2003. 173 f. Tese (Doutorado em Engenharia de Produção) – Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina, Florianópolis, 2003.

PRATI, R. C.; BATISTA, G. E. A. P. A.; MONARD, M. C. Curvas ROC para avaliação de classificadores. **IEEE Latin America Transactions**, v. 6, n. 2, p. 215-222, 2008.

REJAS, L. P.; PONCE, E. R. An exploratory study of the relation between academic management and quality in universities. **Interciencia**, v. 40, 2015

REZENDE, S. O. Sistemas Inteligentes: fundamentos e aplicações. Barueri: Manole, 2003, 525 p.

RIBEIRO, M. A. O projeto profissional familiar como determinante da evasão universitária: um estudo preliminar. **Revista Brasileira de Orientação Profissional**, p.55-70, 2005.

RODRIGUEZ, A. **Fatores de permanência e evasão de estudantes do ensino superior privado brasileiro – um estudo de caso**. Caderno de Administração, v. 5, n. 1, Pontifícia Universidade Católica de São Paulo / PUC SP, São Paulo, 2011.

RUSSELL, S.; NORVIG, P. **Inteligência artificial**. 3. ed. Rio de Janeiro: Elsevier, 2013, 1016 p.

SANTOS, A. P.; SANTOS, S. I. F.; RISSOLI, V. R. V. A predição da evasão de estudantes de graduação como recurso de apoio fornecido por um assistente inteligente. In: Simpósio de Excelência em Gestão e Tecnologia, 2013, Resende. **Anais do Simpósio de Excelência em Gestão e Tecnologia**, 2013, v. 10.

SANTOS, E. M. **Modelo conceitual de sistema de custos por atividades para as universidades**: um estudo de caso na Universidade do Vale do Itajaí. 2001. 159 f. Dissertação (Mestrado em Engenharia de Produção e Sistemas) – Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina, Florianópolis, 2001.

SCHMIDT, L. M. Gestão Universitária: uma relação pedagógico-administrativa. **Olhar de professor**, v. 5, n. 1, p. 77-80, 2002.

SILVA FILHO, R. L. L., MOTEJUNAS, P. R., HIPOLITO, O., LOBO, M. A evasão no ensino superior brasileiro. **Cadernos de pesquisa**, São Paulo, v.37, n. 132, p. 641-659, 2007.

SINDICATO DAS MANTENEDORAS DO ENSINO SUPERIOR. **Mapa do Ensino Superior 2016**. São Paulo: Semesp, 2016.

SOARES, F. C.; ROBAINA, A. D.; PEITER, M. X.; RUSSI, J. L. Predição da produtividade da cultura do milho utilizando rede neural artificial. **Ciência Rural**, 2015, v. 45, p. 1987-1993.

SOUZA, R. C. T. Previsão de séries temporais utilizando Rede Neural treinada por filtro de Kalman e evolução diferencial. 2009. 98 f. Dissertação (Mestrado em Engenharia de Produção e Sistemas) – Pontifícia Universidade Católica do Paraná, Curitiba, 2008.

SPADY, W. Dropouts from higher education: an interdisciplinary review and synthesis. **Interchange**, v. 1, p. 64-85, 1970.

SUPERBY, J.; VANDAMME, J. P.; MESKENS, N. Determination of factors influencing the achievement of the first-year university students using data mining methods. In **Proc. Of the Workshop on Educational Data Mining at ITS'06**, p. 37-44, 2006.

TINTO, V. Dropout from higher education: a theoretical synthesis of recent research. **Journal of Educational Research**, v. 45, p. 89-125, 1975.

TONTINI, G., WALTER, S. A. Podemos identificar a propensão e reduzir a evasão de alunos? Ações estratégicas e resultados táticos para instituições de ensino superior. **Colóquio Internacional sobre Gestão Universitária na América do Sul**, 11, Florianópolis, p. 1-18, 2011.

VELOSO, T. C. M. A. Evasão nos cursos de graduação da Universidade Federal de Mato Grosso, campus universitário de Cuiabá – um processo de exclusão. In: **Anais da Reunião Anual da Associação Nacional de Pós-Graduação e Pesquisa em Educação**, 24., Caxambu, 2001.

VON ZUBEN, F. J. Uma caricatura Funcional de Redes Neurais Artificiais. In: **Anais do Congresso Brasileiro de Redes Neurais**, 6., v. 1, n. 2, p. 66-76, 2003.

ZAGO, N. Do acesso à permanência no ensino superior: percursos de estudantes universitários de camadas populares. **Revista Brasileira de Educação**, Universidade Federal de Santa Catarina, v. 11, n. 32, p. 226-237, 2006.