

UFRRJ
INSTITUTO DE CIÊNCIAS EXATAS
CURSO DE PÓS-GRADUAÇÃO EM MODELAGEM
MATEMÁTICA E COMPUTACIONAL

DISSERTAÇÃO

Aplicação de Redes Neurais RBF e MLP na Análise de Evasão
Discente do Curso de Sistemas de Informação da UFRRJ

Kelly Harumi Fausta Kawase

2015



**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM
MATEMÁTICA E COMPUTACIONAL**

**APLICAÇÃO DE REDES NEURAIS RBF E MLP NA ANÁLISE DE
EVASÃO DISCENTE DO CURSO DE SISTEMAS DE INFORMAÇÃO
DA UFRRJ**

KELLY HARUMI FAUSTA KAWASE

Sob a Orientação do Professor
Robson Mariano da Silva

Dissertação submetida como requisito parcial para obtenção do grau de **Mestre em Ciências**, no Curso de Pós-Graduação em Modelagem Matemática e Computacional, Área de Concentração em Inteligência Computacional e Otimização.

Seropédica, RJ
Outubro de 2015

006.32

K22a

T

Kawase, Kelly Harumi Fausta, 1983-

Aplicação de redes neurais RBF e MLP na análise de evasão discente do curso de sistemas de informação da UFRRJ / Kelly Harumi Fausta Kawase. – 2015.

70 f.: il.

Orientador: Robson Mariano da Silva.

Dissertação (mestrado) – Universidade Federal Rural do Rio de Janeiro, Curso de Pós-Graduação em Modelagem Matemática e Computacional.

Bibliografia: p.46-49.

1. Redes neurais (Computação) – Teses. 2. Evasão universitária – Análise - Teses. 3. Universidade Federal Rural do Rio de Janeiro. Curso de Sistemas de Informação – Estudantes - Avaliação - Teses. 4. Universidade Federal Rural do Rio de Janeiro. Curso de Sistemas de Informação – Processamento de dados – Teses. I. Silva, Robson Mariano da, 1963- II. Universidade Federal Rural do Rio de Janeiro. Curso de Pós-Graduação em Modelagem Matemática e Computacional. III. Título.

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE CIÊNCIAS EXATAS
CURSO DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E
COMPUTACIONAL

KELLY HARUMI FAUSTA KAWASE

Dissertação submetida como requisito parcial para obtenção do grau de Mestre em Ciências,
no Curso de Pós-graduação em Modelagem Matemática e Computacional, área de
concentração em Inteligência Computacional e Otimização.

Dissertação aprovada em 05/10/2015.

Robson Mariano da Silva. Prof. Dr. UFRRJ
(Orientador)

Lígia Cristina Ferreira Machado. Prof. Dr. UFRRJ

Reinaldo Bellini Gonçalves. Prof. Dr. PUC - RJ

Dedicatória

Dedico este trabalho primeiramente a Deus, por iluminar meus pensamentos e por me socorrer nas horas difíceis. Aos meus pais e irmã pelo exemplo, incentivo e carinho. Aos meus amigos e cunhado pelo apoio nos momentos tristes e alegres. E ao meu querido sobrinho Gabriel que sempre alegra o meu coração.

Agradecimentos

Á Deus, por ter me dado forças para enfrentar mais esta jornada.

Aos meus pais Mauricio Massao Kawase e Geni Fausta da Silva Kawase pela ajuda e o apoio;

A minha irmã Kátia Yuri Fausta Kawase pelo companheirismo, dedicação e incentivo.

Ao Prof. Robson Mariano da Silva que aceitou ser meu orientador, me ajudando e apoiando desde o início, com palavras de incentivo e motivação, mesmo diante das adversidades.

A UFRRJ e ao seu Programa de Pós-Graduação em Modelagem Matemática e Computacional por ter possibilitado realizar este mestrado.

Aos secretários Geneilson Barcellos e Janaina Gama, por toda atenção, dedicação e carinho que disponibilizaram ao longo dessa jornada.

Ao colega Eluã Ramos Coutinho por ter me ajudado no início da minha dissertação.

Ao Coordenador Rafael Bernardo Teixeira por ter cedido os dados do curso de Sistemas de Informação da UFRRJ.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa concedida.

E por fim, a todos os meus amigos, colegas e conhecidos que me ajudaram.

RESUMO

KAWASE, Kelly Harumi Fausta. **Aplicação de redes neurais RBF e MLP na análise de evasão discente do curso de Sistemas de Informação da UFRRJ**. 2015. 70p. Dissertação (Mestrado em Modelagem Matemática e Computacional). Instituto de Ciências Exatas, Departamento de Matemática, Universidade Federal Rural do Rio de Janeiro, Seropédica, 2015.

A evasão escolar é um problema complexo, por reunir diversos fatores que influenciam na decisão do aluno em permanecer ou não no curso. Identificar as características envolvidas neste processo possibilita traçar o perfil dos alunos propensos à evasão para poder tomar medidas eficazes. Segundo a literatura, os cursos na área de Tecnologia da Informação vêm apresentando altos índices de evasão. Não obstante disto, o curso de Sistemas de Informação da UFRRJ vem apresentando índices preocupantes. Dentre os diversos trabalhos encontrados na literatura para prever a evasão no Ensino Superior do Brasil, a utilização de técnicas computacionais vem despertando interesse dos pesquisadores, precisamente Redes Neurais Artificiais ARTMAP-Fuzzy, Funções de Base Radial (RBF) e Redes Bayesianas, que vem mostrando bons resultados na predição de evasão discente em Instituições de Ensino Superior. O objetivo dessa dissertação é propor um sistema inteligente baseado em redes neurais de Perceptron de Múltiplas Camadas (MLP) e Funções de Base Radial (RBF) na análise de evasão discente do curso de Bacharel em Sistemas de Informação da UFRRJ. O conjunto de dados utilizados corresponde a 148 amostras de alunos matriculados no curso. Para tal estudo foi considerado a evasão até o 3º período, onde foram utilizadas as seguintes variáveis: gênero, origem escolar, desempenho nas disciplinas específicas do 1º período, distância, tempo de ingresso entre o Ensino Médio e o Ensino Superior, desempenho no ENEM (Exame Nacional do Ensino Médio), estado civil e situação com relação à evasão. Foram realizadas 100 simulações, sendo selecionadas as 25 melhores. Os resultados obtidos pelas redes RBF e MLP mostram ser promissores na análise de evasão discente do curso. Sendo que a rede RBF apresentou melhor desempenho (acurácia = 91,03%) do que o modelo MLP, com diferença significativa ($p\text{-valor} < 0,05$) entre os resultados dos modelos.

Palavras-chave: *Evasão, Sistemas de Informação, Redes Neurais Artificiais.*

ABSTRACT

KAWASE, Kelly Harumi Fausta. **Application of neural networks RBF and MLP in student evasion analysis of the Information Systems Course of UFRRJ**. 2015. 70p. Dissertation (Masters in Computational and Mathematical Modeling). Instituto de Ciências Exatas, Departamento de Matemática, Universidade Federal Rural do Rio de Janeiro, Seropédica, 2015.

The school dropout is a complex problem, which gathering various factors that influence the decision of the student to stay or not in the course. The identification of the characteristics involved in this process turned it possible to analyse the profile of the students prone to evasion, and thereby to take effective action. According to the literature, courses in Information Technology area have shown high dropout rates. However, the course of Information Systems of UFRRJ has shown alarming rates. Among the many studies in the literature to predict dropout in higher education in Brazil, the use of computational techniques has aroused the interest of researchers, precisely Artificial Neural Networks ARTMAP-Fuzzy, Radial Basis Function (RBF) and Bayesian networks, which has been showing good results in predicting student dropout in higher education institutions. The objective of this dissertation is to propose an intelligent system based on Multilayer Perceptron (MLP) and Radial Basis Function (RBF) in student evasion analysis of the course of Bachelor of Information Systems UFRRJ. The used data set corresponds to 148 samples of students enrolled in the course. For such a study was considered evasion to the 3rd period, where the following variables were used: gender, school origin, performance in the specific subjects of the 1st period, distance, time between high school and higher education, performance in ENEM (National Secondary Education Examination), civil status and situation regarding evasion. It were performed 100 simulations, being selected the top 25. The results obtained by the RBF and MLP networks prove to be promising in the course Student dropout analysis. Since the RBF network performed better (accuracy = 91.03%) than MLP model, with a significant difference ($p\text{-value} < 0.05$) between the results of the models.

Keywords: *Evasion, Information Systems, Artificial Neural Networks.*

LISTA DE FIGURAS

Figura 1: Gráfico do índice estimado da evasão (%) nas instituições públicas e privadas do Brasil.....	6
Figura 2: Neurônio Biológico. Fonte: adaptado de MACHADO, 2005.	10
Figura 3: Modelo do Neurônio Artificial. Fonte: adaptado de BRAGA; CARVALHO; LUDEMIR, 2012.	11
Figura 4: Rede MLP com 2 camadas ocultas.	13
Figura 5: Modelo de neurônio não linear. Fonte: Haykin (2001).	14
Figura 6: Representação gráfica das funções de transferência: (a) Função Degrau, (b) Função Linear, (1) Função Logística, (2) Função Tangente Hiperbólica.	15
Figura 7: Fluxo do sinal do neurônio oculto j. Fonte: Adaptado de Haykin (2001).	16
Figura 8: Arquitetura da rede RBF.	21
Figura 9: Causas da evasão no curso de Sistemas de Informação da UFRRJ.....	23
Figura 10: Fluxograma do modelo de evasão do curso de SI da UFRRJ.....	24
Figura 11: Criação das 100 simulações aleatórias.....	27
Figura 12: Aspecto da Rede RBF, adaptado do MATLAB (MATHWORKS, 2012).	28
Figura 13: Aspecto da Rede MLP, adaptado do MATLAB (MATHWORKS, 2012).	29
Figura 14: Evasão relativa, até o 3º período, das turmas de 2010 a 2014 do curso de SI da UFRRJ.	32
Figura 15: Percentual de evadidos e não evadidos de acordo com a categoria da variável Tempo de ingresso.	33
Figura 16: Percentual de evadidos e não evadidos de acordo com a categoria da variável origem escolar.	34
Figura 17: Percentual de evadidos e não evadidos de acordo com a categoria da variável nota em matemática no ENEM.....	34
Figura 18: Percentual de evadidos e não evadidos de acordo com a categoria da variável distância.	35
Figura 19: Percentual de evadidos e não evadidos de acordo com a categoria da variável melhor desempenho no ENEM.....	35
Figura 20: Percentual de evadidos e não evadidos de acordo com a categoria da variável gênero.	36
Figura 21: Percentual de evadidos de acordo com as sete categorias com maior índice de evasão (causas externas).	37
Figura 22: Percentual dos gêneros de acordo com as áreas de melhor desempenho no ENEM.	37
Figura 23: Percentual de evadidos e não evadidos da variável aprovação em cálculo.	38
Figura 24: Percentual de evadidos e não evadidos da variável média no conjunto de disciplinas específicas do 1º período	39
Figura 25: As nove categorias com maior índice de evasão no curso de Sistemas de Informação da UFRRJ.	40
Figura 26: Percentual de evadidos em relação aos atributos de maior índice de evasão das variáveis média das disciplinas específicas e a distância.	42

LISTA DE TABELAS

Tabela 1: Evasão média (%) em alguns países da OECD – Organização para Cooperação e Desenvolvimento Econômico – 2005. Fonte: Lobo (2007) apud OCDE (2006).	4
Tabela 2: Número de ingressantes, matriculados, concluintes e índice de evasão nos cursos superiores do Brasil de 2003 a 2013.....	5
Tabela 3: Funções de base radial mais comumente utilizadas.....	20
Tabela 4: Algoritmos utilizados no aprendizado de redes RBF.....	22
Tabela 5: Categorias das variáveis envolvidas no processo de evasão do curso de SI da UFRRJ.	26
Tabela 6: Parâmetros utilizados no treinamento da rede RBF.	28
Tabela 7: Parâmetros utilizados no treinamento da rede MLP.	28
Tabela 8: Evasão por turma e período no curso de SI da UFRRJ.	31
Tabela 9: Percentual de evadidos e não evadidos com relação às causas da evasão externas a instituição/curso de Sistemas de Informação da UFRRJ.....	33
Tabela 10: Percentual das características (categorias) que os alunos do curso de Sistemas de Informação da UFRRJ mais possuíam ao ingressar no curso.	36
Tabela 11: Percentual de evadidos e não evadidos com relação às causas da evasão externas à instituição/curso de SI da UFRRJ.....	38
Tabela 12: Percentual das categorias que os alunos do curso de SI da UFRRJ tiveram no 1º período do curso.	39
Tabela 13: Resultados das 25 simulações com relação à média e o desvio padrão das medidas de desempenho.	40
Tabela 14: Acurácia das 25 simulações da rede RBF quando deixado uma variável de fora.	41
Tabela 15: Sensibilidade das 25 simulações da rede RBF quando deixado uma variável de fora.....	41
Tabela 16: Especificidade das 25 simulações da rede RBF quando deixado uma variável de fora.....	42

SUMÁRIO

CAPÍTULO I	1
1. INTRODUÇÃO	1
1.1. OBJETIVOS	2
1.1.1. Objetivo geral	2
1.1.2. Objetivos específicos	2
1.1.3. Estrutura do trabalho	2
CAPÍTULO II	3
2. REVISÃO DE LITERATURA	3
2.1. O ENSINO SUPERIOR NO BRASIL	3
2.1.1. Evasão no ensino superior	4
2.1.2. Definições e classificação quanto ao tipo de evasão discente	7
2.1.3. Métodos para precaver a evasão no ensino superior	8
2.1.4. Métodos de predição da evasão no ensino superior	9
2.2. REDES NEURAIS	10
2.2.1. Tipos de arquitetura e formas de aprendizagem das RNAs	12
2.2.2. Redes Perceptrons de Múltiplas Camadas (MLP)	13
2.2.3. Algoritmo de retropropagação (<i>Backpropagation</i>)	15
2.2.5. Redes de função de base radial (RBF)	18
CAPÍTULO III	23
3. MATERIAIS E MÉTODOS	23
3.1. BASE DE DADOS	23
3.2. MODELO PROPOSTO PARA ANÁLISE DA EVASÃO DO CURSO DE SI DA UFRRJ	24
3.2.1 Primeira fase	25
3.2.2 Segunda fase	26
3.2.3 Terceira fase	30
3.2.4 Quarta fase	30
CAPÍTULO IV	31
4. RESULTADOS E DISCUSSÃO	31
4.1. ANÁLISE EXPLORATÓRIA DA EVASÃO NO CURSO DE SI DA UFRRJ.	31
4.2. ESTUDO DAS VARIÁVEIS ENVOLVIDAS NO PROCESSO DE EVASÃO DO CURSO DE SI DA UFRRJ	32
4.2.1 Causas externas à instituição/curso de Sistemas de Informação da UFRRJ	32
4.2.2 Causas internas a instituição/curso de Sistemas de Informação da UFRRJ	38
4.2.3 Causas da evasão com relação a todos os atributos	39
4.3. COMPARAÇÃO NO DESEMPENHO DAS REDES MLP E RBF	40
4.4. VERIFICAÇÃO DO IMPACTO DAS VARIÁVEIS NO DESEMPENHO DA REDE DE MELHOR PERFORMANCE	40
CAPÍTULO V	44
5. CONCLUSÕES E TRABALHOS FUTUROS	44
5.1. CONCLUSÕES	44
5.2. TRABALHOS FUTUROS	45

6. REFERÊNCIAS BIBLIOGRÁFICAS	46
APÊNDICES	50
APÊNDICE A – Resultados das 25 melhores simulações da rede RBF.....	50
APÊNDICE B – Resultados das 25 melhores simulações da rede MLP.....	51
APÊNDICE C – Resultados das 25 simulações da rede RBF sem a variável Tempo de Ingresso (E.M.x E.S)	52
APÊNDICE D – Resultados das 25 simulações da rede RBF sem a variável Estado Civil	53
APÊNDICE E – Resultados das 25 simulações da rede RBF sem a variável Origem Escolar	54
APÊNDICE F – Resultados das 25 simulações da rede RBF sem a variável Nota em Matemática no ENEM.....	55
APÊNDICE G – Resultados das 25 simulações da rede RBF sem a variável Distância.	56
APÊNDICE H – Resultados das 25 simulações da rede RBF sem a variável Área de melhor desempenho no ENEM.....	57
APÊNDICE I – Resultados das 25 simulações da rede RBF sem a variável Média das Disciplinas Específicas.	58
APÊNDICE J – Resultados das 25 simulações da rede RBF sem a variável Aprovação em Cálculo.	59

CAPÍTULO I

1. INTRODUÇÃO

A evasão nas instituições de Ensino Superior é um problema mundial. Que afeta o resultado dos sistemas educacionais, gerando prejuízos significativos para o avanço das nações. Os discentes que iniciam seus estudos e não concluem são desperdícios sociais, econômicos e acadêmicos. A busca de suas causas vem sendo objeto de estudos de pesquisadores e órgãos governamentais em diversos países, incluindo o Brasil.

Devido ao crescimento populacional e a expansão da economia, o governo federal vem investindo na abertura de novos cursos e vagas a fim de incluir mais jovens de 18 a 24 anos nas instituições de ensino superior e poder fortalecer o capital intelectual. Esta inclusão depende também de uma maior participação das classes de renda baixa, que necessitam de maiores iniciativas do governo, como programas de cotas, bolsas de estudos integrais ou parciais – Programa Universidade para Todos (PROUNI) e financiamento estudantil (FIES). Aliado a isto, as instituições cada vez mais vem aderindo ao Sistema de Seleção Unificada (Sisu), que vem viabilizando o aumento da mobilidade acadêmica, descentralizando o processo seletivo e possibilitando a democratização das oportunidades.

Entretanto, ao passo que o governo vem buscando meios para atrair estes alunos para o ensino superior, principalmente os de classes de baixa renda, ele também vem fazendo um grande esforço para que estes discentes continuem nas instituições e concluam seus cursos. E à medida que os cursos exijam maior base em matemática aumenta-se a chance do discente evadir, como é o caso dos cursos do setor de Tecnologia da Informação e Comunicação (TIC).

O setor de Tecnologia de Informação e Comunicação vem contribuindo para o aumento da competitividade do Brasil no exterior, tendo um papel fundamental para o desenvolvimento social e econômico do país. Segundo a empresa de consultoria *International Data Corporation* - IDC Brasil (2015), o mercado de TIC deverá movimentar neste ano R\$ 165,6 bilhões, além de continuar crescendo acima do Produto Interno Bruto-PIB.

Paralelamente a este crescimento, haverá uma necessidade de profissionais nesta área, e que de acordo com o estudo feito pela Associação Brasileira de Empresas de Tecnologia da Informação e Comunicação (BRASSCOM) não será suprida nos estados do Rio de Janeiro, São Paulo, Distrito Federal, Paraná e Rio Grande do Sul, devido à necessidade do mercado ser superior ao número de discentes que as instituições destes estados podem formar. Esta carência se justifica pelo alto índice de evasão nos cursos superiores do Brasil nesta área, que é de aproximadamente 87%. Além de muitos deles chegarem à graduação com deficiência em matemática, que com o tempo diminui o interesse do discente pela área (BRASSCOM, 2015).

De acordo com a literatura, a evasão é um objeto complexo, pois, reúne diversas variáveis e muitas inter-relações entre elas que resultam no processo de decisão do aluno em permanecer ou não no curso.

Uma proposta que vem despertando interesse dos pesquisadores é a utilização de técnicas computacionais, precisamente Redes Neurais Artificiais ARTMAP-Fuzzy, Funções de Base Radial – RBF, Perceptron de Múltiplas Camadas – MLP e Redes Bayesianas, que vem mostrando bons resultados na predição de evasão discente em Instituições de Ensino Superior. Esta predição possibilita identificar os alunos propensos à evasão, facilitando a tomada de decisão de modo a identificar com antecedência o perfil destes alunos e, assim, desenvolver estratégias e ações eficientes de sorte a mantê-los no curso.

1.1. Objetivos

1.1.1. Objetivo geral

- Desenvolver um sistema inteligente utilizando Redes Neurais Artificiais Perceptron de Múltiplas Camadas - MLP e Redes de Função de Base Radial - RBF na análise de evasão discente do curso de Sistemas de Informação da UFRRJ.

1.1.2. Objetivos específicos

- Analisar a evasão do curso de Sistemas de Informação por período e turma;
- Realizar um estudo das variáveis envolvidas no processo de evasão do curso de Sistemas de Informação da UFRRJ;
- Comparar o desempenho das redes neurais MLP e RBF;
- Verificar o impacto das variáveis no desempenho da rede de melhor performance.

1.1.3. Estrutura do trabalho

A apresentação deste trabalho está organizada em quatro capítulos, além deste. O Capítulo 2 encontra-se um resgate dos conceitos que foram relevantes para a execução desta pesquisa, através da revisão de literatura. Nesta é abordada os seguintes tópicos: o ensino superior no Brasil e as redes neurais.

No tópico do ensino superior no Brasil inicia-se com um breve histórico e segue-se até o estado atual desta modalidade de ensino – a problemática da evasão discente. Abordando as definições, a classificação e os métodos para precaver e prever esse problema nas Instituições de Ensino Superior (IES). Além disso, é abordado também sobre as Redes Neurais (RNs) de Perceptron de Múltiplas Camadas (MLP) e Função de Base Radial (RBF), os tipos de arquitetura e as formas de aprendizagem, bem como o algoritmo de retropropagação (*Backpropagation*).

O Capítulo 3 consiste nos materiais e métodos, com os procedimentos metodológicos adotados para a elaboração desta pesquisa: a base de dados, o modelo proposto para a análise da evasão do curso de Sistemas de Informação (SI) da UFRRJ, bem como as suas fases.

É apresentado no Capítulo 4 os resultados e discussão, sendo realizado através da análise da evasão do curso de SI, bem como o estudo das variáveis envolvidas neste processo, além da comparação dos modelos de redes neurais e, por conseguinte, a verificação do impacto das variáveis no desempenho da rede de melhor performance.

O Capítulo 5 tem-se as conclusões e as sugestões para os trabalhos futuros.

E por fim, têm-se as referências deste trabalho e os apêndices.

CAPÍTULO II

2. REVISÃO DE LITERATURA

2.1. O Ensino Superior no Brasil

A universidade é considerada o elemento central do ensino superior e as demais instituições tentam posicionar-se a ela, podendo então, ser estudada como referência. Mesmo que no caso do Brasil ela tenha sido mais uma aspiração do que uma instituição concreta durante muito tempo (MENDONÇA, 2000).

De acordo com Cunha (2003) e Durham (2003), o ensino superior no Brasil iniciou-se de forma tardia. Pois, diferentemente da Espanha, que no século XVI criou universidades em suas colônias, Portugal proibia que estas fossem criadas no Brasil, sendo isto, uma forma de impedir movimentos independentistas. Em seu lugar, eram oferecidas algumas bolsas para alguns filhos de colonos estudarem na Universidade de Coimbra ou Évora, em Portugal. Mesmo a iniciativa dos jesuítas de estabelecer um seminário ou colégios na colônia foi prejudicada com a expulsão da Companhia de Jesus no final do século XVIII.

Somente em 1808, com a ameaça de invasão das tropas napoleônicas a Portugal, a sede do reino transferiu-se para o Brasil. E assim começa a história dos cursos superiores propriamente ditos. Entretanto, em vez de universidades foram criadas cátedras isoladas para a formação de profissionais que atendessem as necessidades do estado. A partir destas surgiram as escolas, as academias e as faculdades. E a primeira instituição que assumiu duradouramente o *status* de universidade e que serviu de modelo para as demais, foi a Universidade do Rio de Janeiro, atual Universidade Federal do Rio de Janeiro - UFRJ, que inicialmente resultou da união de três faculdades – de medicina, engenharia e direito, autorizada a sua criação em 1915 e criada em 1920 (CUNHA, 2003).

Entretanto, mesmo com a reunião dessas instituições em universidade elas continuaram a funcionar de forma isolada, não tendo qualquer alteração nos seus currículos e nem nas práticas desenvolvidas no seu interior. Também não havia articulação entre elas, exceto quando envolvia a disputa pelo poder. (MENDONÇA, 2000).

Isto passa a acontecer em 1934 com a criação da Universidade de São Paulo – USP, que além de reunir diversas faculdades, tinha como diferencial a criação da Faculdade de Educação e a Faculdade de Filosofia, Ciências e Letras, esta última pensada como o “coração da universidade”, que iria promover a integração entre as diferentes áreas do saber. (CUNHA, 2003).

“A partir do final da década de 1940 e ao longo da década de 1950 vão ocorrer as federalizações, estendendo-se pelas décadas de 1960 e 1970 o processo de criação das universidades federais [...]” (SAVIANI, 2011, p.7).

Segundo Sobral (2015) e Martins (2007), a grande expansão das matrículas no ensino superior começa a ocorrer a partir da década de 1970, devido à concentração urbana das massas e a exigência de uma melhor formação da mão-de-obra dos trabalhadores. Entretanto, contraditoriamente a isto, enquanto as matrículas crescem, poucos são os alunos que concluem seus cursos; sendo a evasão discente a grande responsável por este fenômeno.

Atualmente, em reconhecimento ao papel estratégico das universidades federais para o desenvolvimento econômico e social, o Governo Federal desenvolveu o Reuni- Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais. O Reuni, teve o seu Programa de Expansão Fase I das universidades federais, iniciado em 2003, que

possibilitou adotar uma série de medidas no intuito de retomar o crescimento do ensino superior público, além de buscar a ampliação do acesso e a permanência na educação superior. Sendo o combate à evasão uma de suas metas.

2.1.1. Evasão no ensino superior

A evasão no Ensino Superior é um problema global, comum às instituições de ensino superior no mundo contemporâneo, devido a isto, vem sendo objeto de estudos e análises em diversos países, incluindo o Brasil. As perdas destes discentes que iniciam e não concluem seus cursos são desperdícios sociais, acadêmicos e econômicos. No setor público há um descompasso entre os diversos recursos públicos investidos sem o devido retorno e no setor privado uma considerável perda de receitas (MEC, 1996; LOBO, et al., 2007).

Segundo pesquisa feita por Hipólito (2015), em 2009 as instituições privadas tiveram um gasto em torno de R\$ 9 bilhões com a evasão. E a universidade pública gasta aproximadamente 15 mil reais por aluno ao ano, podendo ser maior quando envolve outros custos como alimentação e transporte.

Na tabela 1, mostra que os altos índices de evasão no Ensino Superior são problemas vivenciados em várias instituições internacionais, inclusive em países europeus e nos Estados Unidos da América (USA).

Tabela 1: Evasão média (%) em alguns países da OECD – Organização para Cooperação e Desenvolvimento Econômico – 2005. Fonte: Lobo (2007) apud OCDE (2006).

Países	Índice de evasão média
Japão	7%
Turquia	12%
UK	17%
Coréia	22%
Alemanha	30%
México	31%
USA	34%
França	41%
Itália	58%
Suécia	52%

A avaliação do ensino superior no Brasil é realizada através de dados coletados pelo INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira pelo Censo da Educação superior. Por meio destes dados é possível verificar informações detalhadas sobre a situação deste ensino e poder estimar a evasão.

Apesar de existirem várias formas para realizar o cálculo da evasão, dependendo da sua finalidade e os critérios adotados pelo pesquisador, o importante é verificar como anda esta tendência. Foi utilizada a equação (1), adotada nos trabalhos de Silva Filho (2014) e Martinho (2014), onde é calculado o índice anual de evasão.

$$\text{Evasão (\%)} = \frac{(M(n-1) - C(n-1)) - (M(n) - I(n))}{(M(n-1) - C(n-1))} \times 100 \quad (1)$$

Evasão (%): índice da evasão em determinado ano (n);

n: ano em que se deseja estimar o índice da evasão;

M(n) = matrículas em determinado ano (n);
M (n-1) = matrículas do ano anterior a n;
C (n-1) = concluintes do ano anterior a n;
I (n) = novos ingressantes em determinado ano (n).

Considerando os dados descritos na tabela 2, ainda que exista um crescimento nas matrículas e uma menor evasão em 2013, em relação ao ano anterior; houve uma queda no número de concluintes. Ao longo de uma década, isto foi à primeira vez que aconteceu. Além do número de ingressantes ter diminuído em relação ao ano anterior.

Tabela 2: Número de ingressantes, matriculados, concluintes e índice de evasão nos cursos superiores do Brasil de 2003 a 2013.

Ano	Categoria Administrativa	Matrícula	Ingressos	Concluintes	Evadidos(%)
2003	Total	3.936.933	1.554.664	532.228	-
	Pública	1.176.174	325.405	173.141	-
	Privada	2.760.759	1.229.259	359.087	-
2004	Total	4.223.344	1.646.414	633.363	24,31
	Pública	1.214.317	364.647	209.008	15,29
	Privada	3.009.027	1.281.767	424.355	28,08
2005	Total	4.567.798	1.805.102	730.484	23,04
	Pública	1.246.704	362.217	203.689	12,02
	Privada	3.321.094	1.442.885	526.795	27,33
2006	Total	4.883.852	1.965.314	762.633	23,94
	Pública	1.251.365	368.394	195.231	15,34
	Privada	3.632.487	1.596.920	567.402	27,15
2007	Total	5.250.147	2.138.241	786.611	24,49
	Pública	1.335.177	416.178	197.040	12,98
	Privada	3.914.970	1.722.063	589.571	28,46
2008	Total	5.808.017	2.336.899	870.386	22,23
	Pública	1.552.953	538.474	195.933	10,86
	Privada	4.255.064	1.798.425	674.453	26,12
2009	Total	5.954.021	2.065.082	959.197	21,24
	Pública	1.523.864	422.320	206.877	18,83
	Privada	4.430.157	1.642.762	752.320	22,15
2010	Total	6.379.299	2.182.229	973.839	15,97
	Pública	1.643.298	475.884	190.597	11,36
	Privada	4.736.001	1.706.345	783.242	17,62
2011	Total	6.739.689	2.346.695	1.016.713	18,73
	Pública	1.773.315	490.680	218.365	11,71
	Privada	4.966.374	1.856.015	798.348	21,31
2012	Total	7.037.688	2.747.089	1.050.413	25,03
	Pública	1.897.376	547.897	237.546	13,21
	Privada	5.140.312	2.199.192	812.867	29,44
2013	Total	7.305.977	2.742.950	991.010	23,79
	Pública	1.932.527	531.846	229.278	15,61
	Privada	5.373.450	2.211.104	761.732	26,92

Fonte: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2015).

A partir de 1990 com a abertura de novos cursos e vagas para o ensino superior e com a criação do sistema de cotas, do Programa Universidade para Todos – PROUNI e o Fundo de Financiamento Estudantil – FIES, além de mais facilidades com bolsas e a redução do valor das matrículas nas instituições privadas; estas passam a responder por uma demanda reprimida – a inclusão de classes de rendas mais baixas. Havendo uma tendência no crescimento destas classes. Entretanto, muitos destes alunos vêm totalmente despreparados, tendo dificuldades de aprendizagem e sem maturidade social. E à medida que os cursos exijam maior base em matemática, eles evadem (MARTINHO, 2014; CAPELATO, 2011).

Uma hipótese para o crescimento da evasão a partir de 2010 (Figura 1) pode estar relacionado a uma maior inclusão destas classes de rendas mais baixas e a criação do SISU – Sistema de Seleção Unificada.

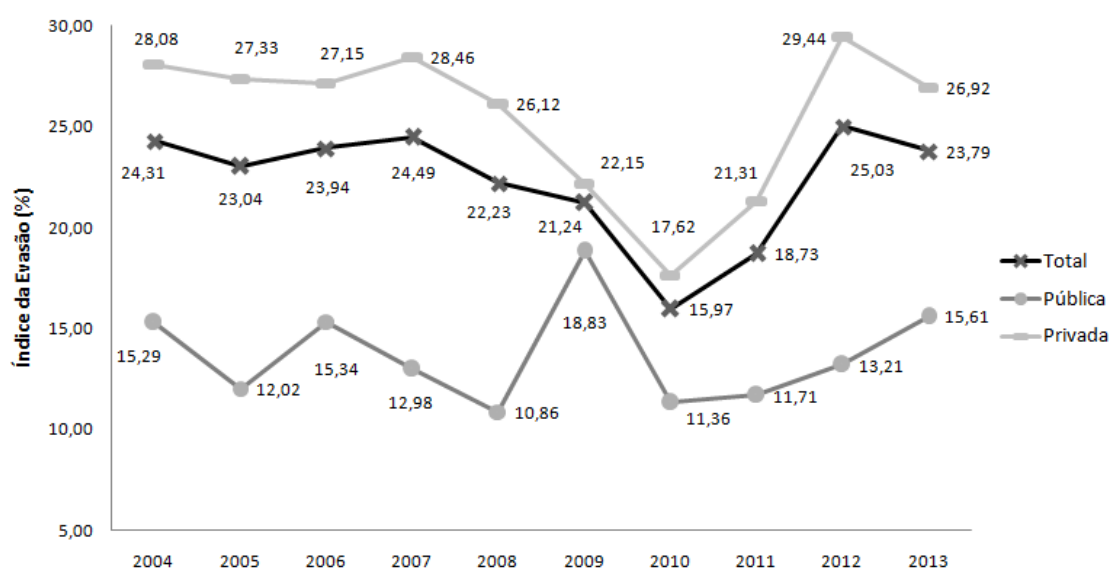


Figura 1: Gráfico do índice estimado da evasão (%) nas instituições públicas e privadas do Brasil

Em 2009 o Ministério da Educação (MEC) apresentou as Instituições Federais de Ensino Superior – IFES uma nova proposta de ENEM – Exame Nacional de Ensino Médio, que seria uma unificação do processo de seleção das vagas das IFES, sendo uma alternativa a descentralização do processo seletivo e possibilitando a democratização das oportunidades de acesso as vagas federais. Sendo instituído o Sistema de Seleção Unificada (SISU), no qual os alunos através deste sistema informatizado são selecionados para estas vagas nos cursos de graduação disponibilizadas pelas IFES.

Se por um lado o aluno passa a ter mais oportunidades de acesso, por outro, a evasão dispara no seu primeiro ano de implantação. Segundo Associação Nacional dos Dirigentes das Instituições Federais de Ensino Superior - ANDIFES (2014) e Ceretta (2013), no primeiro ano de graduação após a implantação do SISU a evasão aumentou, apresentando uma taxa duas vezes maior que a evasão geral em algumas instituições.

De acordo com Sena (2014), o aluno acaba entrando no curso não por vocação, mas pela nota de corte, ingressando muitas vezes em um curso que não é nem a sua primeira ou segunda opção. Dessa forma, tem a oportunidade de escolher dois cursos totalmente diferentes.

No entanto, segundo o estudo realizado pela Comissão Especial para Estudos sobre Evasão nas Universidades Públicas Brasileiras – MEC (1996), a análise global do sistema ou das instituições em algumas situações pode sustar a evasão, não permitindo uma política que

contemple a diversidade de casos. Assim podem ocorrer efeitos incoerentes com a real situação dos cursos, em lugar de minimizar problemas podem intervir apenas em cursos que não apresentam níveis altos de evasão.

Conforme pesquisa realizada por Lobo *et al* (2007), o estado do Rio de Janeiro apresentou a maior taxa do Brasil, enquanto as áreas de Serviços e de Ciências, Matemática e Computação tiveram as mais altas taxas de evasão anual média em 2005.

Por isso, é interessante que os pesquisadores delimitem o seu trabalho, definindo prioridades e identificando qual o tipo de evasão, de acordo com o real problema encontrado.

2.1.2. Definições e classificação quanto ao tipo de evasão discente

As definições de evasão discente vêm sendo discutidas ao longo do tempo com diferentes conceituações sobre evasão no Ensino Superior do Brasil:

Bueno (1993) faz uma distinção entre evasão e “exclusão”, a primeira é uma postura do aluno que decide desligar-se por sua própria vontade, enquanto a segunda, a responsabilidade passa a ser da universidade e tudo o que a cerca, por não aproveitar e nem direcionar este discente a permanecer.

Na pesquisa realizada por Ristoff (1995), é abordada a diferença entre evasão e “mobilidade”, sendo que no primeiro caso corresponde ao abandono nos estudos e no segundo a migração para outro curso. Esta última parcela não é um desperdício, nem fracasso do aluno ou da instituição, mas um investimento na tentativa de buscar o sucesso ou a felicidade.

Já em estudos feitos pela Comissão Especial do MEC, a evasão dos cursos de graduação é “[...] a saída definitiva do aluno de seu curso de origem, sem concluí-lo” (MEC, 1996, p.15).

Segundo Biazus (2003) citado por Costa (1991), nos cursos de graduação da UFGRS, a evasão é definida como a saída do aluno, temporária ou definitiva, da universidade ou do curso, exceto pela diplomação. Ainda, de acordo com o autor, a evasão é caracterizada em três fases:

- definitiva: quando o aluno se afasta da instituição, seja porque abandonou, desistiu ou transferiu-se para outra universidade.
- temporária: leva-se em conta qualquer trancamento, de um até dez semestres.
- de curso: seria a transferência interna.

Devido à dificuldade de padronizar tudo que está relacionado à evasão, já que não é unânime o seu conceito, é interessante estar atento às distinções de base e a sua classificação quanto ao tipo de evasão que se pretende trabalhar. Isto além de evitar o risco de distorcer a realidade com simplificações ou generalizações inadequadas, permite identificar adequadamente o objeto a ser estudado (MEC, 1996).

De acordo com Lobo (2012), existem quatro níveis de evasão discente: a evasão do sistema, a evasão da instituição de ensino, a evasão do curso e a evasão do aluno.

Na evasão do sistema o discente abandona o sistema de ensino superior, de forma definitiva ou temporária, isto dentro do sistema estudado. Podendo variar de abrangência, como exemplo, indo do sistema federal de ensino ao sistema de ensino superior brasileiro como um todo. Para efeito de cálculo da evasão, não são levados em consideração os alunos transferidos internamente e nem externamente, desde que seja dentro do sistema estudado. No caso do estudante que não solicita transferência e se submete a um novo processo de seleção é considerado abandono, por não conter uma informação detalhada sobre o ocorrido (MEC, 1996; LOBO, 2012).

“O perfil real da evasão do sistema de ensino só poderia ser traçado se fossem cruzados os dados, por aluno, tanto intra quanto inter-universidades (MEC, 1996, p.21)”.

A evasão da instituição de ensino caracteriza-se quando o aluno deixa a instituição de ensino ao qual está matriculado. Não leva em consideração para o cálculo da evasão os alunos que mudaram de curso internamente. Pode-se medir a evasão da instituição ou de um conjunto de instituições, através de cálculos de dados gerais agregados.

A evasão do curso ocorre quando o aluno deixa o curso por qualquer motivo. Muitas IES não consideram como evasão propriamente dita quando o aluno muda de curso dentro da instituição ou quando a vaga deixada por um discente for preenchida por outro. Porém, quando o aluno sai de um curso, tem-se uma perda (evasão), que precisa ser analisada.

Medir a Evasão não se trata só de verificar um “saldo de caixa”, ou seja, quantos alunos entraram menos quanto saíram, mas quem entrou e quem saiu e por quais razões, para que seja possível evitar outras perdas pelos mesmos motivos com ações que gerem mudanças e essas só acontecem se entendemos, claramente, o que está ocorrendo (LOBO, 2012, p.8).

A evasão do aluno é a origem, que gera todos os outros tipos de evasão. É utilizado para o seu cálculo o que se chama de acompanhamento da coorte.

“[...] Acompanhamento da coorte, que trabalha com o dado individualizado do próprio aluno, permitindo medir com exatidão a evasão até do indivíduo, que vai além da vacância da vaga preenchida por outro aluno transferido (SILVA FILHO; LOBO, 2014, p.2)”.

2.1.3. Métodos para precaver a evasão no ensino superior

De acordo com a literatura, a análise do perfil dos alunos evadidos vem sendo proposta na maioria das pesquisas científicas relacionadas à evasão no Ensino Superior. E todos são unânimes em afirmar que a evasão resulta da união de diversos fatores que determinam na decisão do aluno em permanecer ou não no curso (MARTINS, 2007).

Na pesquisa realizada por Biazus (2003), foi estudado e elaborado um instrumento para análise das causas da evasão discente nos Cursos de Ciências Contábeis da Universidade Federal de Santa Maria (UFSM) e da Universidade Federal de Santa Catarina (UFSC). O modelo desenvolvido por ele possui duas categorias definidas da causa da evasão: interna e externa. A categoria interna possui os seguintes indicadores: atitude comportamental, motivos institucionais e requisitos didático-pedagógicos. Na externa os indicadores são vocação pessoal, componentes conjunturais, características individuais e componente sócio-político-econômico.

No estudo feito por Gaioso (2006), são identificadas as principais causas da evasão nos cursos de Direito, Engenharia Civil e Medicina nas Instituições de Educação Superior Brasileira. Algumas das causas da evasão relatadas são: falta de orientação profissional e desconhecimento da metodologia do curso, deficiência da educação básica, busca de herança profissional, mudança de endereço, problemas financeiros, horário de trabalho incompatível com o do estudo, concorrência entre as IES privadas, imaturidade, reprovação sucessiva, Falta de perspectivas de trabalho e ausência de laços afetivos na universidade.

Slhessarenko *et al* (2013), realizou uma pesquisa documental no curso de Bacharelado em Sistema de Informação do Instituição Federal de Educação em Santa Catarina (IFSC), no período de 2010 a 2011, e identificou que as principais causas da evasão neste curso foram: a mudança de interesse, opção de vida ou indecisão profissional, a didática não eficiente dos professores, expectativas não atendidas em relação ao curso, a carga horária semanal de trabalho, os aspectos inadequados das salas de aula ao ensino, a falta de orientação adequada por parte da coordenação do curso e as dificuldades de acompanhamento do curso. Sendo que

a mudança de interesse, opção de vida e/ou indecisão profissional foi a causa que obteve o maior percentual.

2.1.4. Métodos de predição da evasão no ensino superior

Nos últimos anos, as técnicas computacionais vêm sendo utilizadas para a previsão de evasão discente no Ensino Superior. Tanto a análise do perfil dos alunos como também a utilização destas ferramentas possibilita a instituição ou a coordenação do curso a criar melhores ações pedagógicas no combate à evasão.

Entretanto, estas análises realizadas sobre o perfil dos alunos, geralmente apoiadas na aplicação de questionários, podem ter possível viés na seleção dos respondentes, já que é muito difícil estabelecer contato com estes alunos. Além da descrição do perfil do evadido não ser contrastada com os que permanecem na instituição, não sendo possível uma visão comparativa entre estes dois grupos, o que será de pouca serventia para a consequente estratégia para conter esta ocorrência (DA SILVA, 2012). A previsão da evasão discente mostra-se eficiente para conter esta ocorrência, pois, possibilita identificar estes dois grupos, verificando se o aluno estará propenso ou não à evasão.

No trabalho realizado por Karamouzis e Vrettos (2008), as Redes Neurais MLP foram utilizadas para prever a taxa de aprovação nos dois primeiros anos e se o aluno iria continuar na faculdade. Os autores utilizaram doze parâmetros de entrada: cor, gênero, intenção de se matricular na faculdade, idade, escola de origem, necessidade de atendimento especial, apoio/reforço acadêmico, CEP, endereço, idade na conclusão do ensino médio, dedicação na faculdade e se é um estudante sucedido. A rede foi configurada com três camadas e utilizou-se o algoritmo *backpropagation*. A taxa média de acerto obtida foi de 72% com o próprio conjunto de treinamento e 68% com o conjunto de teste.

Tontini e Walter (2014) realizaram um estudo que envolvia um estudo de um método de identificação do risco de evasão de alunos de graduação para o semestre seguinte, analisando o primeiro e o segundo semestre de 2009. Utilizou-se a Rede RBF e os seguintes parâmetros de entrada: qualidade do curso, organização do curso, atualização do curso, relação entre teoria e prática, aplicação profissional do aprendizado, competência dos professores, atendimento do curso, atendimento da coordenação, atendimento do centro, atendimento da praça de alimentação, qualidade dos laboratórios, conservação das salas de aula, conservação da infraestrutura, limpeza e conservação dos banheiros, equipamentos em salas de aula, aprendizado com o curso, desenvolvimento pessoal, obtenção de boas notas e/ou conceitos, necessidade de reforço de aulas, oportunidades profissionais, perspectivas de melhoria de vida, integração com a turma, escolha do curso, identificação com o curso, importância profissional do curso, comprometimento com o curso, tempo disponível para o estudo, disponibilidade para frequentar as aulas, suficiência de renda pessoal ou familiar, necessidade bolsa de estudo, estabilidade pessoal e familiar, saúde pessoal, motivação para vida e persistência pessoal nos objetivos. Através da identificação dos alunos em risco de evasão, os alunos foram contatados e acompanhados por seus coordenadores de curso, o que possibilitou uma redução de 18% no índice de evasão na Fundação Universidade Regional de Blumenau (FURB).

Em Lameira e Silva (2014) encontramos a aplicação de Redes Bayesianas com a finalidade de identificar os fatores que causam a evasão no curso de Sistemas de Informação da UFRRJ. O conjunto de dados constituiu-se de 70% (n=105) dos alunos com matrícula no curso de Sistemas de Informação, e destes 105 alunos, 90% foi usado para treino e o restante para teste. As variáveis de entrada da rede foram: gênero (masculino ou feminino), origem escolar (pública ou privada), reprovação (sim ou não), distância (perto ou longe), para o parâmetro perto foi adotado uma distância inferior a 40 km da universidade e como saída a

situação de evasão (sim ou não). Foi identificado a não correlação da origem escolar como fator de evasão. A sensibilidade no conjunto de validação foi de 93,3%.

Martinho (2014) desenvolveu um sistema utilizando rede neural ARTMAP-Fuzzy, para classificação dos alunos propensos ou não à evasão escolar no Curso Superior de Tecnologia em Automação e Controle do Instituto Federal de Mato Grosso - IFMT. As variáveis utilizadas como entradas na rede foram: sexo, faixa etária, etnia, estado civil, números de pessoas/residência, renda familiar, possui computador, escolaridade dos pais, escola de origem, autoavaliação, procedência, distância escola – residência, meio de transporte, situação com relação à trabalho, turno de estudo, alunos na sala. Obteve um índice de acerto do grupo evasivo entre 95% e 100% e acurácia global média em torno de 95%.

2.2. Redes Neurais

Uma das Técnicas de Inteligência Computacional são as Redes Neurais Artificiais (RNAs), usualmente conhecidas como redes neurais (RN) e referidas nas diversas literaturas também como neurocomputadores, redes conexionistas, processadores paralelamente distribuídos, etc. É inspirado em um modelo neural biológico, que é altamente complexo, não-linear e paralelo. Com isso tem a capacidade de aprender e extrair informações que não são explícitas nos exemplos, através de experiências, utilizando um processamento paralelo e distribuído simples (neurônios artificiais) que calculam funções matemáticas normalmente não lineares, podendo assim fazer generalizações das informações aprendidas e resolver problemas complexos e intratáveis (BRAGA; CARVALHO; LUDEMIR, 2012; HAYKIN, 2001).

Como descrito anterior o neurônio artificial é baseado em um neurônio biológico (Figura 2), que possui três partes principais: corpo celular, dendritos e o axônio. Os dendritos tem a função de receber informações e conduzir até o corpo celular (soma) onde é processada a informação. A soma destes sinais quando alta gera um impulso que passa pelo axônio e é transmitido pela terminação axônica e o dendrito de outro neurônio. Esse ponto de contato é chamado de sinapse. A operação em paralelo de todos os neurônios possibilita a realização de funções complexas. (BRAGA; CARVALHO; LUDEMIR, 2012; MACHADO, 2005).

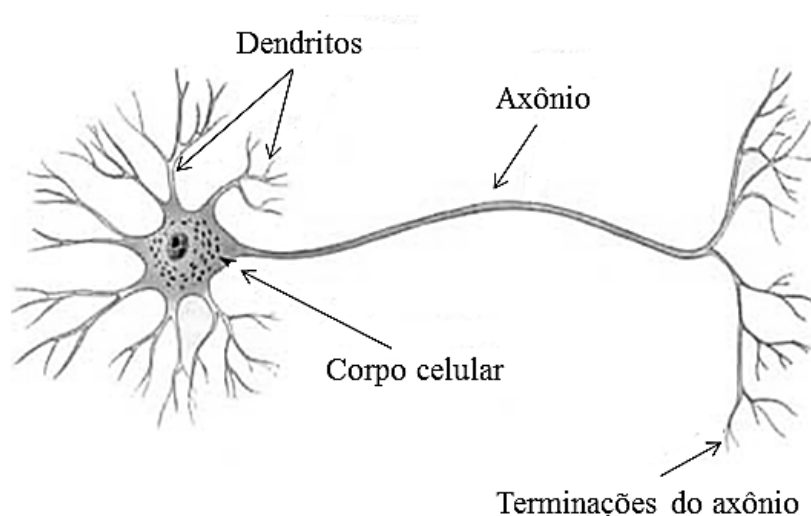


Figura 2: Neurônio Biológico. Fonte: adaptado de MACHADO, 2005.

O primeiro modelo neural artificial – MCP foi criado por Warren McCulloch e Walter Pitts em 1943, em seu trabalho *A Logical Calculus of the Ideas Immanent in Nervous Activity*, onde simula o comportamento de um neurônio biológico.

Os terminais de entrada $x_1, x_2, \dots, x_m$ seriam os dendritos e a saída y o axônio. As sinapses são as entradas com os pesos acoplados $w_k x_k$. Os pesos que determinam o grau para o disparo da conexão (Figura 3).

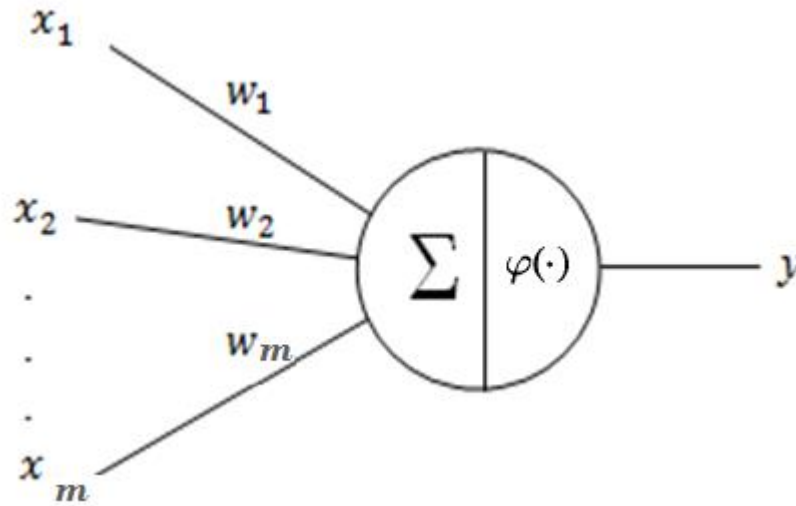


Figura 3: Modelo do Neurônio Artificial. Fonte: adaptado de BRAGA; CARVALHO; LUDEMIR, 2012.

O corpo celular é o somatório, que é a soma ponderada das entradas dos pesos (Equação 2):

$$\sum_{k=1}^m w_k x_k \quad (2)$$

E a função de transferência $\varphi(.)$ define o valor de saída do neurônio, dependendo da comparação feita ao seu limiar θ (threshold), gerando saídas 0 ou 1 (Equação 3).

$$\varphi(.) = \begin{cases} 1 & \text{se } \sum_{k=1}^m x_k w_k \geq \theta \\ 0 & \text{se } \sum_{k=1}^m x_k w_k < \theta \end{cases} \quad (3)$$

Entretanto, este e o modelo *perceptron* desenvolvido por Rosembat enfrentavam uma grande limitação, resolviam somente problemas linearmente separáveis, sendo que os mais relevantes são não lineares (GOLDSCHMIDT, 2010).

2.2.1. Tipos de arquitetura e formas de aprendizagem das RNAs

Neurônios individuais possuem uma limitada capacidade computacional, já um conjunto de neurônios artificiais conectados em forma de uma rede neural tem a capacidade de resolver problemas complexos (BRAGA; CARVALHO; LUDEMIR, 2012).

Os principais tipos de arquitetura de redes são: *feedforward* de uma única camada, *feedforward* de múltiplas camadas e recorrentes. Na primeira, suas sinapses estão ligadas somente aos neurônios da camada seguinte, sendo acíclica. A segunda rede, além de possuir camadas de entrada e saída, possui também camada intermediária, sendo alimentada para frente. E a última rede possui pelo menos um ciclo de realimentação, podendo ter sinapses com neurônios de uma mesma camada, tanto lateral quanto anterior. (MARTINHO, 2014).

Com relação à forma de aprendizagem da rede os algoritmos também podem ser classificados. Podendo ser supervisionado, quando existe um supervisor e já se sabe os padrões de entrada e a saída desejada para o treinamento, e o não supervisionado quando não tem um supervisor e só estão disponíveis os padrões de entrada. Mas como existem regularidades nos dados isso possibilita o aprendizado.

Na aprendizagem supervisionada o treinamento por correção de erros é o mais conhecido. Este treinamento ocorre a cada interação, sendo o erro, verificado através da diferença entre o valor desejável do neurônio k ($y_{dk}(t)$) e o valor predito ($y_{pk}(t)$), no tempo t (Equação 4).

$$e_k(t) = y_{dk}(t) - y_{pk}(t) \quad (4)$$

E a cada passo t o sinal de erro pode ser calculado para cada neurônio k de saída. Sendo a função de custo redefinida como o somatório dos k sinais de erro, como é verificado na equação 5.

$$E(t) = \frac{1}{2} \sum_{k=1}^m e_k^2(t) \quad (5)$$

Isto possibilitará que a próxima interação ($t+1$) tenha um erro menor, através de ajustes calculados e aplicados aos pesos de j (Equação 6). No qual, $w_{kj}(t)$ é o valor do peso no tempo t , $\Delta w_{kj}(t)$ o valor do ajuste a ser aplicado ao peso e $w_{kj}(t+1)$ é o próximo valor assumido, interação seguinte (BRAGA; CARVALHO; LUDEMIR, 2012; GUARNIERI, 2006).

$$w_{kj}(t+1) = w_{kj}(t) + \Delta w_{kj}(t) \quad (6)$$

Isto resultada na minimização da regra do delta no qual o valor do ajuste $\Delta w_{kj}(t)$ é ocasionado pela multiplicação da taxa de aprendizagem η pelo erro $e(t)$ e pela entrada da sinapse x_k (Equação 7) (BRAGA; CARVALHO; LUDEMIR, 2012; GUARNIERI, 2006).

$$\Delta w_{kj}(t) = \eta e_k(t) x_k(t) \quad (7)$$

2.2.2. Redes Perceptrons de Múltiplas Camadas (MLP)

Através de Donald Hebb, Frank Rosenblatt, Hopfield e outros pesquisadores, foi possível criar novas formas de aprendizagem e treinamento, possibilitando a criação de uma rede neural que pudesse resolver problemas complexos, gerando qualquer saída - a rede de Perceptrons de Múltiplas Camadas – MLP (NIED, 2007). Esta rede é uma alternativa eficiente na resolução de problemas complexos, pois além de ser um aproximador universal, possibilita configurar a sua rede (figura 4) com três camadas: de entrada, oculta e de saída. Sendo que a camada oculta pode ter uma ou mais camadas intermediárias.

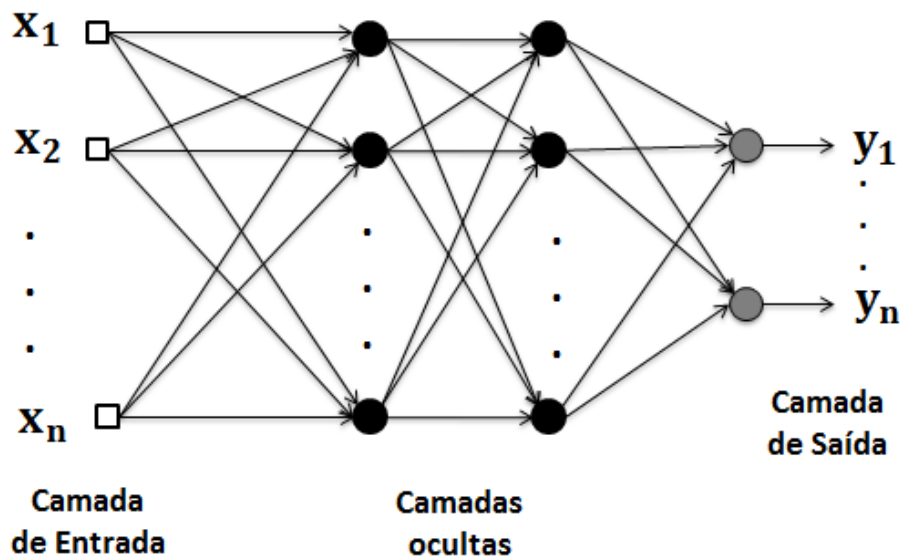


Figura 4: Rede MLP com 2 camadas ocultas.

Na rede MLP (Figura 5) é incluído um elemento externo b_k (bias), denominado viés, (ruído aleatório). Este tem o efeito de diminuir ou aumentar a entrada líquida da função de ativação. Em termos matemáticos, o neurônio k é descrito pelas equações 8 e 9:

$$u_k = \sum_{j=1}^n w_{kj} x_j \quad (8)$$

$$y_k = \varphi(u_k + b_k) \quad (9)$$

Onde y_i é o sinal de saída do neurônio, $\varphi(\cdot)$ é a função de ativação e u_i é a saída da combinação linear (somatório das entradas ponderadas dos pesos) (HAYKIN, 2001).

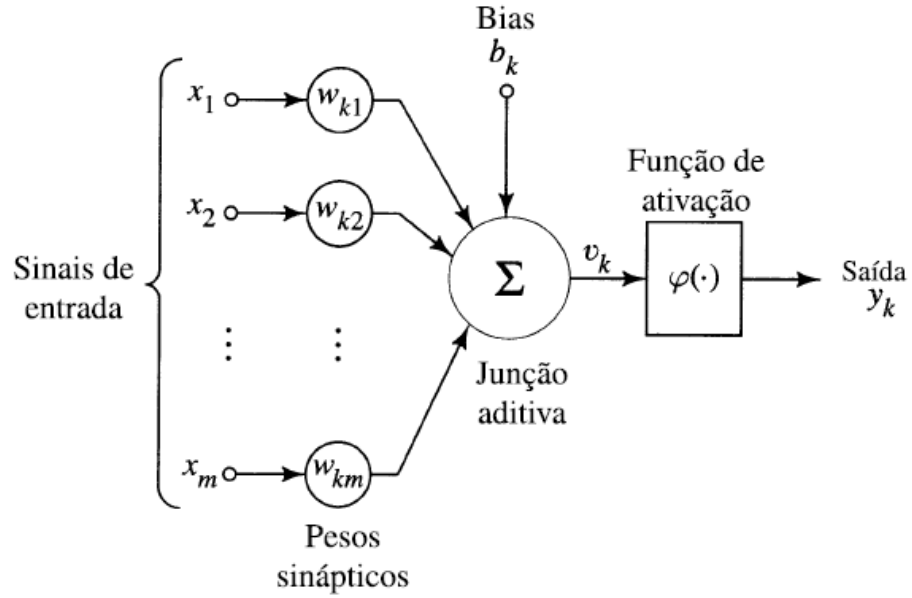


Figura 5: Modelo de neurônio não linear. Fonte: Haykin (2001).

As funções mais comumente utilizadas são: função limiar (degrau), linear, logística e tangente hiperbólica (FIORIN, et al; 2011) representadas pelas equações (10, 11, 12, 13) e pelo comportamento esboçados na figura (6).

a) Função de ativação degrau:

$$\varphi(u_k) = \begin{cases} 1 & \text{se } u_k \geq 0 \\ 0 & \text{se } u_k < 0 \end{cases} \quad (10)$$

b). Função Linear :

$$\varphi(u_k) = u_k \quad (11)$$

c) Funções Sigmóides:

c. 1) Função Logística:

$$\varphi(u_k) = \frac{1}{(1 + \exp(-uk))} \quad (12)$$

c. 2) Função Tangente Hiperbólica:

$$\varphi(u_k) = \tanh\left(\frac{u_k}{2}\right) = \frac{1 - \exp(-uk)}{1 + \exp(-uk)} \quad (13)$$

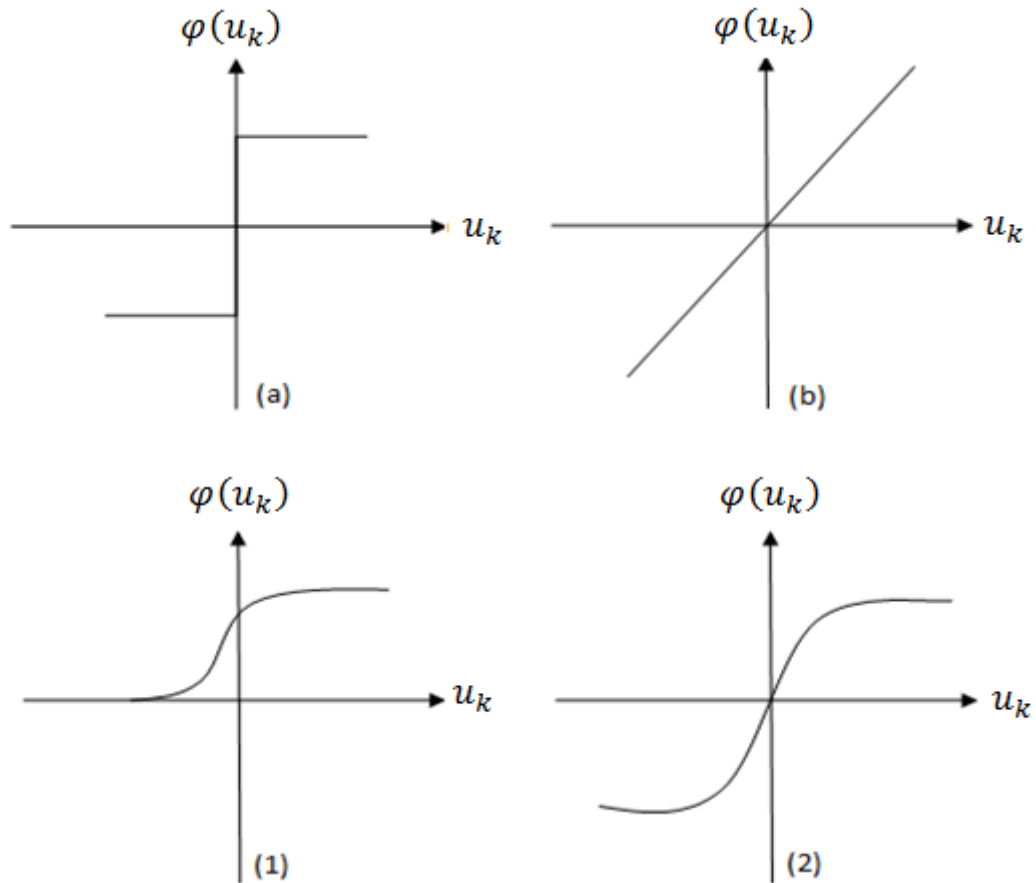


Figura 6: Representação gráfica das funções de transferência: (a) Função Degrau, (b) Função Linear, (1) Função Logística, (2) Função Tangente Hiperbólica.

2.2.3. Algoritmo de retropropagação (*Backpropagation*)

Existem muitos tipos de algoritmos utilizados na rede MLP, sendo o *Back-propagation* o com maior aplicação. Este consiste em duas fases: *forward* e *backward*. Simplificando as fases, na primeira, as informações são apresentadas a rede e são propagadas da camada de entrada, passando pelas intermediárias até chegar à camada de saída, não havendo alteração nos pesos. E na segunda, ocorre o inverso, caso o erro esteja acima do desejado ele retropropaga da sua saída até a origem ajustando os pesos sinápticos a fim de diminuir o erro.

Para que isto ocorra, inicialmente é verificado o erro de saída do neurônio j (camada oculta), na interação t , que é a diferença entre o erro desejado $d_j(t)$ e o erro predito $y_j(t)$ (Equação 14).

$$e_j(t) = d_j(t) - y_j(t) \quad (14)$$

O fluxo do sinal do neurônio j é representado pela figura 7, no qual ele é alimentado por vários sinais funcionais vindos dos neurônios da camada a sua esquerda.

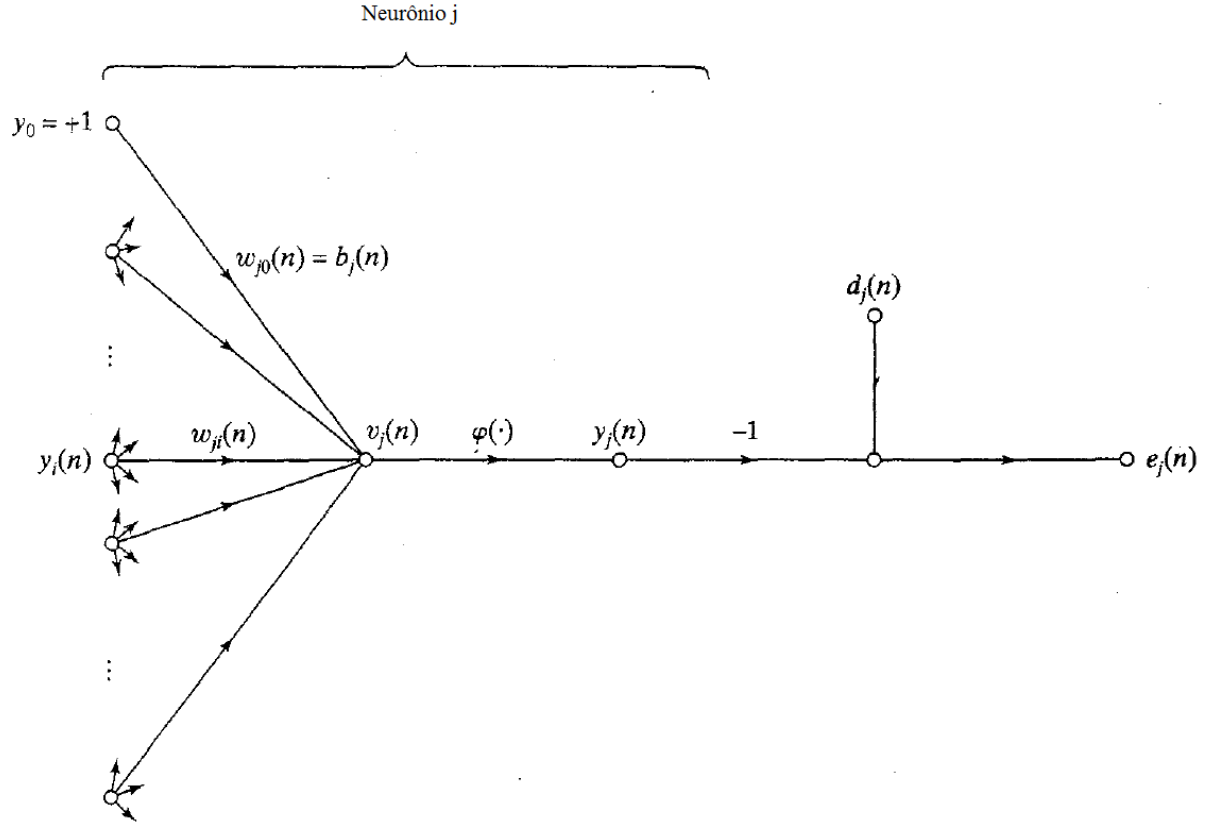


Figura 7: Fluxo do sinal do neurônio oculto j. Fonte: Adaptado de Haykin (2001).

O processo de aprendizagem pela minimização do erro ocorre aplicando-se o método do gradiente descendente na função de energia total instantânea do erro, que é obtida pelo somatório dos termos $\frac{1}{2}e_j^2(t)$ e C representa o conjunto de todos os neurônios da camada de saída da rede (Equação 15).

$$E(t) = \frac{1}{2} \sum_{j \in C} e_j^2(t) \quad (15)$$

O campo local induzido $u_j(t)$ e o sinal funcional que aparece na saída do neurônio $y_j(t)$, na iteração t , são descritos pelas equações 16 e 17:

$$u_j(t) = \sum_{i=0}^m w_{ji}(t) y_i(t) \quad (16)$$

$$y_j(t) = \varphi_j(u_j(t)) \quad (17)$$

Onde:

$y_i(t)$ é o sinal de entrada no neurônio j ,

$w_{ji}(t)$ é o peso sináptico que conecta o neurônio i ao j ,

$y_j(t)$ é a saída no neurônio j ,

$\varphi_j(\cdot)$ é a função de ativação aplicada ao neurônio.

Aplicando-se a regra da cadeia (Equação 18) e diferenciando ambos os lados das equações 15, 14, 17 e 16 em relação a $e_j(t)$, $y_j(t)$, $u_j(t)$ e $w_{ji}(t)$ respectivamente, é produzido a equação dos ajustes dos pesos (Equação 19), no qual $e_j(t)$ é o erro de saída da rede, $\varphi_j'(u_j(t))$ é a derivada da função de ativação e $y_i(t)$ é o sinal de entrada do neurônio j.

$$\frac{\partial E(t)}{\partial w_{ji}(t)} = \frac{\partial E(t)}{\partial e_j(t)} \frac{\partial e_j(t)}{\partial y_j(t)} \frac{\partial y_j(t)}{\partial u_j(t)} \frac{\partial u_j(t)}{\partial w_{ji}(t)} \quad (18)$$

$$\frac{\partial E(t)}{\partial w_{ji}(t)} = -e_j(t) \varphi_j'(u_j(t)) y_i(t) \quad (19)$$

Segundo Haykin (2001), a correção ou ajuste $\Delta w_{ji}(t)$ aplicado ao peso $w_{ji}(t)$ é determinado pela regra do delta (Equação 20), sendo η a taxa de aprendizagem. O sinal negativo indica a descida do gradiente, determinado a direção de busca no espaço de pesos de forma a reduzir o valor da função custo $E(t)$. E aplicando-se a equação 19 em 20, a equação pode ser reescrita como (Equação 21):

$$\Delta w_{ji}(t) = -\eta \frac{\partial E(t)}{\partial w_{ji}(t)} \quad (20)$$

$$\Delta w_{ji}(t) = \eta \delta_j(t) y_i(t) \quad (21)$$

O gradiente local $\delta_j(t)$ indica as modificações que serão necessárias nos pesos sinápticos e é definido pelo produto do erro $e_j(t)$ do neurônio correspondente pela derivada da função de ativação associada $\varphi_j'(u_j(t))$, como é verificado na equação 22:

$$\begin{aligned} \delta_j(t) &= -\frac{\partial E(t)}{\partial u_j(t)} \\ &= \frac{\partial E(t)}{\partial e_j(t)} \frac{\partial e_j(t)}{\partial y_j(t)} \frac{\partial y_j(t)}{\partial u_j(t)} \\ &= e_j(t) \varphi_j'(u_j(t)) \end{aligned} \quad (22)$$

Um fator fundamental envolvido no cálculo da correção do peso $\Delta w_{ji}(t)$ é o sinal do erro de saída $e_j(t)$, como é verificado nas equações 21 e 22. Pode-se verificar este ajuste em duas situações distintas, de acordo com a localização do neurônio j, podendo este estar na camada de saída ou na camada intermediária. Para a primeira situação é calculado o gradiente local de maneira direta, sendo comparado com uma saída desejada. No segundo caso, devido ao neurônio não estar acessível diretamente por não ter uma saída desejada específica, os neurônios da camada intermediária acabam compartilhando a responsabilidade pelo erro cometido na saída da rede. Devido a isto, os erros de todos os neurônios conectados diretamente ao neurônio oculto são definidos recursivamente (HAYKIN, 2001). Nesta situação pode-se redefinir o gradiente local para o neurônio oculto como (Equação 23):

$$\begin{aligned}\delta_j(t) &= - \frac{\partial e_j(t)}{\partial y_j(t)} \frac{\partial y_j(t)}{\partial u_j(t)} \\ &= \frac{\partial E(t)}{\partial y_j(t)} \varphi_j'(u_j(t))\end{aligned}\quad (23)$$

Em sequência, para calcular a derivada parcial $\partial E(t)/\partial y_j(t)$ (Equação 24) deve-se calcular o sinal do $E(t)$, mas utilizando o neurônio k da camada de saída em vez do neurônio da camada intermediária (neurônio j).

$$E(t) = \frac{1}{2} \sum_{k \in C} e_k^2(t) \quad (24)$$

Diferenciando a equação 25 em relação a $y_j(t)$ e em seguida utilizando-se a regra da cadeia na derivada parcial $\partial e_k(t)/\partial y_j(t)$ e as devidas deduções e substituições, sendo melhor detalhado nas páginas 187 a 201 no livro do Haykin(2001), é obtido a fórmula de retropropagação para o gradiente local (Equação 25).

$$\delta_j(t) = \varphi_j'(u_j(t)) \sum_k \delta_k(t) w_{kj}(t) \quad (25)$$

Este gradiente local da camada oculta $\delta_k(t)$ é o produto da derivada da função de ativação aplicada aos pesos dos neurônios e aos sinais de entrada da camada oculta, $\varphi_j'(u_j(t))$, pela soma ponderada dos gradientes $\delta_k(t)$, que dependem dos erros $e_k(t)$ dos neurônios que estão diretamente conectados ao neurônio j e à sua direita, pelos pesos associados nestas conexões.

Resumidamente, a atualização dos pesos na camada oculta é definida pela regra do delta (Equação 21) e o seu gradiente dependerá da localização do seu neurônio, se ele é um nó de saída ou se é um nó oculto. No primeiro caso o gradiente local é igual a equação 22 e para o segundo, é igual a equação 25.

Entretanto, este algoritmo possui uma deficiência com redes pequenas, pois requerer taxas pequenas de aprendizagem estável, tornando-se muito lento para esse tipo de rede (ZOUKAS, 2009).

O *Levenberg-Marquardt* é uma variação do algoritmo *Back-Propagation*, entretanto é baseado na aproximação do método de Newton. Ele é muito utilizado para redes pequenas e tendo um tempo de convergência menor quando comparado com outros métodos. Demanda de grande quantidade de memória, o que torna ruim a sua utilização para redes muito grandes (BRAGA; CARVALHO; LUDEMIR, 2012).

2.2.5. Redes de função de base radial (RBF)

As redes neurais artificiais do tipo RBF são uma poderosa ferramenta de geração de mapeamento não linear multivariado, tendo como base a utilização de funções de base radial.

As funções de base radial (RBFs) são geralmente funções não-lineares cujo valor cresce ou decresce monotonicamente à medida que a distância do centro da função de base radial aumenta. Sua aplicação no contexto de aproximação de funções tem sua origem na teoria da interpolação multivariada (POGGIO, 1990), sendo o problema de interpolação formulado da seguinte forma:

Dado um conjunto de N pontos diferentes $\{\underline{x}_i \in \Re^m; i = 1, 2, \dots, N\}$, onde m representa a dimensão do espaço de entrada e um conjunto correspondente de N números reais $\{d_i \in \Re; i = 1, 2, \dots, N\}$, encontre uma função $F: \Re^m \rightarrow \Re$ que satisfaça a condição de interpolação.

$$F(\underline{x}_i) = d_i, \quad i = 1, 2, \dots, N \quad (26)$$

A técnica das RBFs, (POWELL, 1990), consiste em escolher uma função de aproximação F que pode ser representada por:

$$F(x) = \sum_{i=1}^n w_i \varphi(\|\underline{x} - \underline{x}_i\|) \quad (27)$$

onde $\{\varphi(\|\underline{x} - \underline{x}_i\|); i = 1, 2, \dots, N\}$ é um conjunto de N funções arbitrárias, geralmente não lineares conhecidas como função de base radial, e $\|\cdot\|$ a representa uma norma (distância) em geral euclidiana. Os pontos de dados conhecidos $\underline{x}_i \in \Re^m$, $i=1,2,\dots,N$ são tomados como centros das funções de base radial, já que as normas são tomadas em relação a eles.

Com os parâmetros das RBFs definidos, basta determinar os coeficientes w_i , onde inserindo as condições de interpolação obtém-se o seguinte conjunto de equações lineares (Equação 28):

$$\underline{w} = \begin{bmatrix} \varphi_{11} & \varphi_{12} & \dots & \varphi_{1p} \\ \varphi_{21} & \varphi_{22} & \dots & \varphi_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{p1} & \varphi_{p2} & \dots & \varphi_{pp} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_p \end{bmatrix} \quad (28)$$

$\Phi \qquad \underline{w} \qquad \underline{d}$

onde $\varphi_{ij} = \varphi_i(\underline{x}_j)$, Φ é denominada matriz de interpolação e $\underline{w}$ e $\underline{d}$ correspondem, respectivamente, ao vetor de pesos lineares e ao vetor de respostas desejadas.

Assumindo que Φ seja não-singular, o vetor de pesos lineares $\underline{w}$ pode ser obtido através da seguinte equação 29:

$$\underline{w} = \Phi^{-1} \underline{d} \quad (29)$$

Michelli (1986) mostrou a existência de um conjunto de funções que são apropriadas para a interpolação por resultarem em um conjunto de equações lineares para as variáveis w_i nas quais existe uma única solução. A tabela 3 mostra exemplos destas funções que são mais comumente utilizadas.

Tabela 3: Funções de base radial mais comumente utilizadas.

Função	Equação
Gaussiana	$\varphi(\underline{x}) = \exp\left(-\frac{\ \underline{x}-c_i\ ^2}{2\sigma_i^2}\right)$
Multiquadrada Inversa	$\varphi(\underline{x}) = \frac{1}{\sqrt{\ \underline{x}-c_i\ ^2 + \sigma^2}}$
Função multiquadrática	$\varphi(\underline{x}) = \sqrt{\ \underline{x}-c_i\ ^2 + \sigma^2}$
Lâmina spline fina	$\varphi(\underline{x}) = \left(\frac{\ \underline{x}-c_i\ ^2}{\sigma^2}\right) \log\left(\frac{\ \underline{x}-c_i\ ^2}{\sigma^2}\right)$

O parâmetro c_i representa o centro da função de base radial e σ_i a sua largura. O parâmetro σ_i pode ser pensado como um fator de escala para a distância $\|\underline{x} - c_i\|^2$. No caso da função gaussiana, por exemplo, o valor de $\varphi_i(\underline{x})$ decresce mais rapidamente quando $\sigma_i \rightarrow 0$.

Originalmente, as redes RBF utilizavam tantas funções de base radial quantos fossem os padrões do conjunto de dados, de sorte a obter uma exatidão da aproximação. Essa abordagem, além de requerer um esforço computacional elevado, gerava uma super, gerava uma super especialização da rede (BISHOP, 1997).

Broomhead e Lowe (1988) propuseram que nem todos os vetores de entrada tivessem uma função de base radial associada, ou seja, o número de vetores pertencentes ao conjunto de treinamento da rede neural (N) é maior que o número de funções de base radial utilizadas na rede RBF (K). Assim, como, $N > K$, o sistema de equações lineares cujas variáveis são os pesos w_i , terá mais equações do que variáveis.

A abordagem proposta por Broomhead e Lowe (1988) resultou em uma significativa redução do custo computacional e no aumento da capacidade de generalização das redes RBF, fato este que possibilitou a aplicação das redes RBF em áreas, tais como: modelagem de sistemas, processamento digital de sinais e predição de séries temporais, sendo sua arquitetura representada na figura 8.

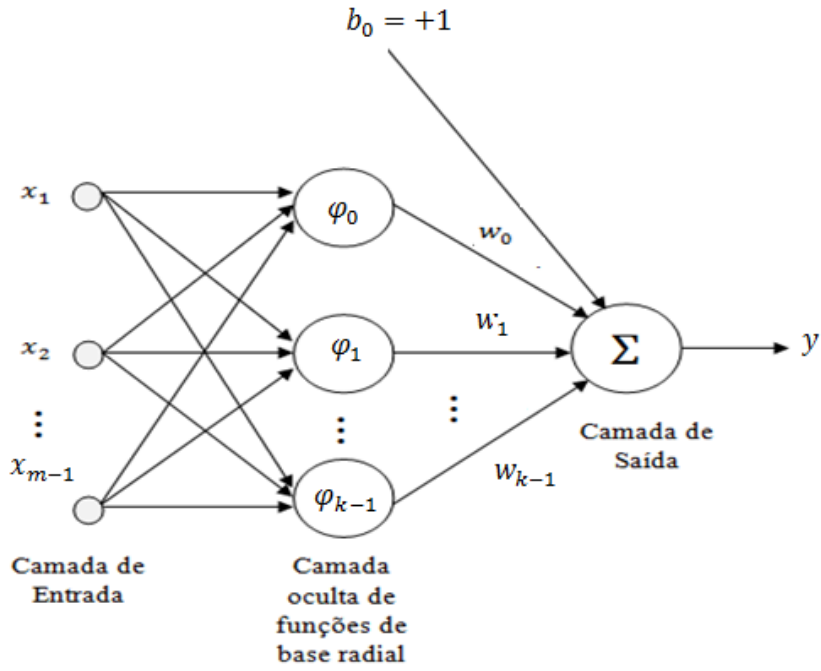


Figura 8: Arquitetura da rede RBF.

Esta rede é constituída de uma camada de nós entrada, a qual é apresentada o vetor de entrada $\underline{x}(n) \in \Re^m$; uma camada intermediária de neurônios não-lineares, onde cada neurônio calcula a distância entre o vetor de entrada e o centro da função de base radial associada (camada escondida). A camada de saída é constituída por um único neurônio linear, que é a soma das saídas de cada função de base radial, ponderadas pelos respectivos pesos sinápticos w_k , de modo que a combinação linear é expressa pela equação 30:

$$y = \sum_{k=0}^{K-1} w_k \varphi_k(\underline{x}, t_k, \sigma_k^2) + b_0 \quad (30)$$

Onde o termo $(\underline{x}, t_k, \sigma_k^2)$ é a k-ésima função de base radial; $\underline{x}(n) \in \Re^m$ representa o vetor de entrada $\underline{x}$ no instante n; $t_k(n) \in \Re^m$ é o vetor centro da k-ésima função de base radial; k é o número de funções de base radial, $\sigma_k^2(n) \in \Re$ é a variância associada a cada uma das funções no instante n; w_k são os pesos sinápticos e b_0 é o termo constante de polarização ou bias.

O processo de aprendizagem em uma rede RBF consiste em determinar os parâmetros livres, que são as variâncias, os centros e os pesos sinápticos. Este processo constituído de dois estágios, o primeiro utilizando técnicas não-supervisionadas para determinação das funções de base radial na camada oculta, usando para tal apenas os dados de entrada e segundo, utilizando métodos lineares supervisionados para determinar os pesos na segunda camada, com rápida convergência (BISHOP,1997).

Em Haykin (2001), encontramos vários algoritmos que podem ser utilizados para a adaptação dos parâmetros livres (σ_k^2, t_k, w_k) das redes neurais RBF. A tabela 4 apresenta alguns dos algoritmos utilizados para o ajuste dos parâmetros livres.

Tabela 4: Algoritmos utilizados no aprendizado de redes RBF.

Centros das RBF	Pesos Sinápticos	Variâncias dos centros
Conhecimento prévio e inferência a partir do conjunto de vetores de treinamento	Gradiente estocástico (LMS), supervisionado: usa $\underline{e}(n) = \underline{d}(n) - \underline{y}(n)$	Por conhecimento prévio e inferência a partir do conjunto de vetores de treinamento
Aglomerção pelo algoritmo “ <i>k-média</i> ”. Não supervisionado	Pseudo Inversa por decomposição em valores singulares: $\underline{w}(n) = \Phi^{-1}(n) \underline{d}(n)$	Gradiente estocástico (LMS), Supervisionado: usa $\underline{e}(n) = \underline{d}(n) - \underline{y}(n)$

Os processos de aprendizado utilizados em redes RBF são geralmente mais fáceis do que os das redes de perceptrons de multicamadas (MLP), em virtude dos processos de aprendizagem para os parâmetros centros, variâncias e pesos sinápticos poderem ser encadeados sequencialmente, possibilitando assim, a otimização do aprendizado das redes neurais RBF (MULGREW, 1996).

CAPÍTULO III

3. MATERIAIS E MÉTODOS

3.1. Base de Dados

Esta pesquisa foi realizada com os dados do curso de Bacharel em Sistemas de Informação (SI) da Universidade Federal Rural do Rio de Janeiro – UFRRJ, fornecidos pela coordenação do curso (2015). A base possui 153 dados, relativos às turmas de 2010 a 2014.

As variáveis foram selecionadas de acordo com as possíveis causas da evasão no curso. Estas foram divididas em duas categorias: causas internas a instituição/curso e externas a ela (Figura 9).

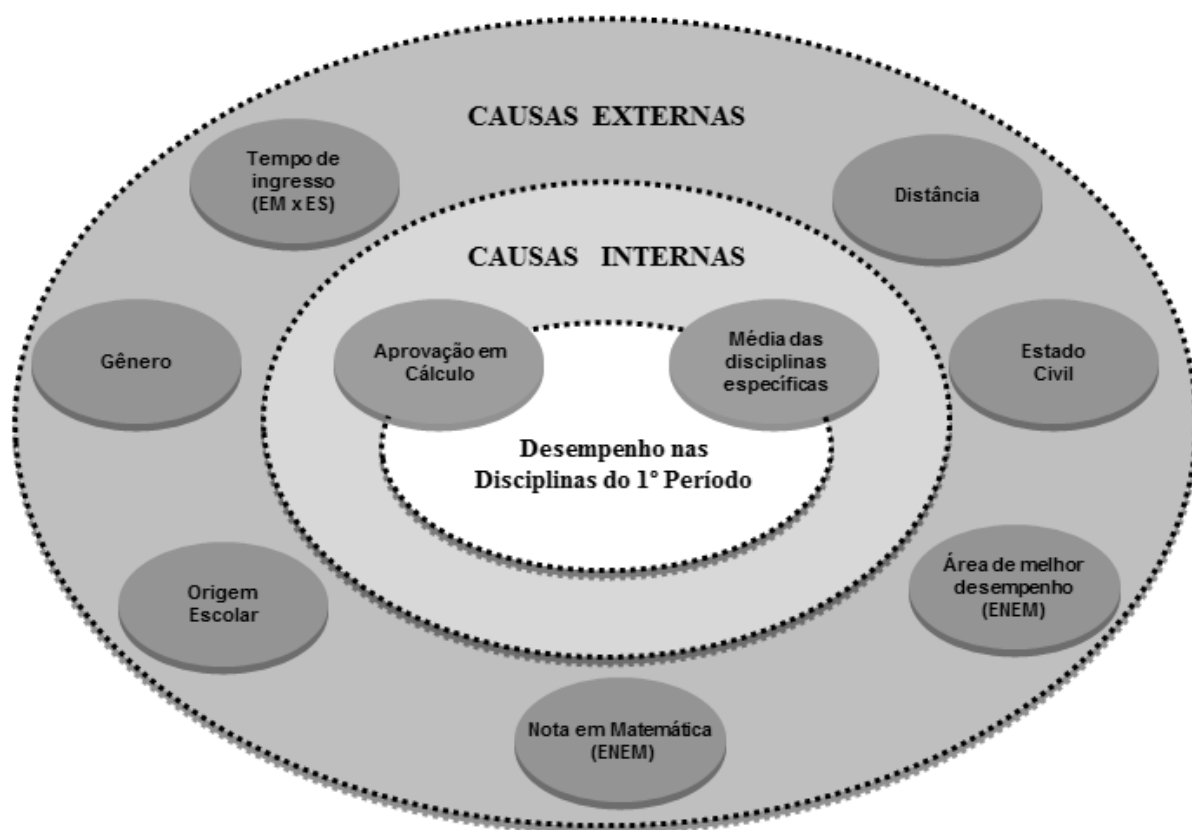


Figura 9: Causas da evasão no curso de Sistemas de Informação da UFRRJ.

São causas externas a instituição/curso as características que os alunos apresentavam ao ingressar no curso. As seis variáveis destas causas foram:

- tempo de ingresso: tempo entre o Ensino Médio – E.M e o Ensino Superior – E.S;

O aluno que terminou o Ensino Médio no final de um ano e ingressou no ano seguinte foi considerado como ‘0 ano’ e após isto 1 ano ou mais.

- origem escolar: a modalidade de ensino da escola de origem do aluno;

- nota em matemática: obtido pelo estudante no ENEM, na área de Matemática e suas Tecnologias;
- distância: a distância entre a localidade anterior ao ingresso do aluno e a UFRRJ;
- estado civil (ingressar no curso);
- área de melhor desempenho: área em que o aluno obteve o melhor desempenho no ENEM, sendo que estas áreas são: matemática e suas tecnologias, ciências da natureza, ciências humanas e linguagens e suas tecnologias;
- gênero.

As causas internas a instituição/curso levam em consideração o desempenho do aluno no 1º período do curso. As duas variáveis destas causas foram:

- aprovação em cálculo (aprovação na disciplina de Cálculo I no 1º período);

O aluno foi considerado não aprovado na disciplina de cálculo quando ocorresse trancamento ou outro motivo que caracterizasse reprovação.

- média das disciplinas específicas: média das notas das disciplinas de Linguagem de Programação I, Introdução a Sistemas Digitais e Introdução a Sistemas de Informação, cursada no 1º período do curso, de acordo com a atual grade curricular do curso (2015).

3.2. Modelo Proposto para Análise da Evasão do Curso de SI da UFRRJ

O modelo de evasão discente está dividido em quatro fases (Figura 10).

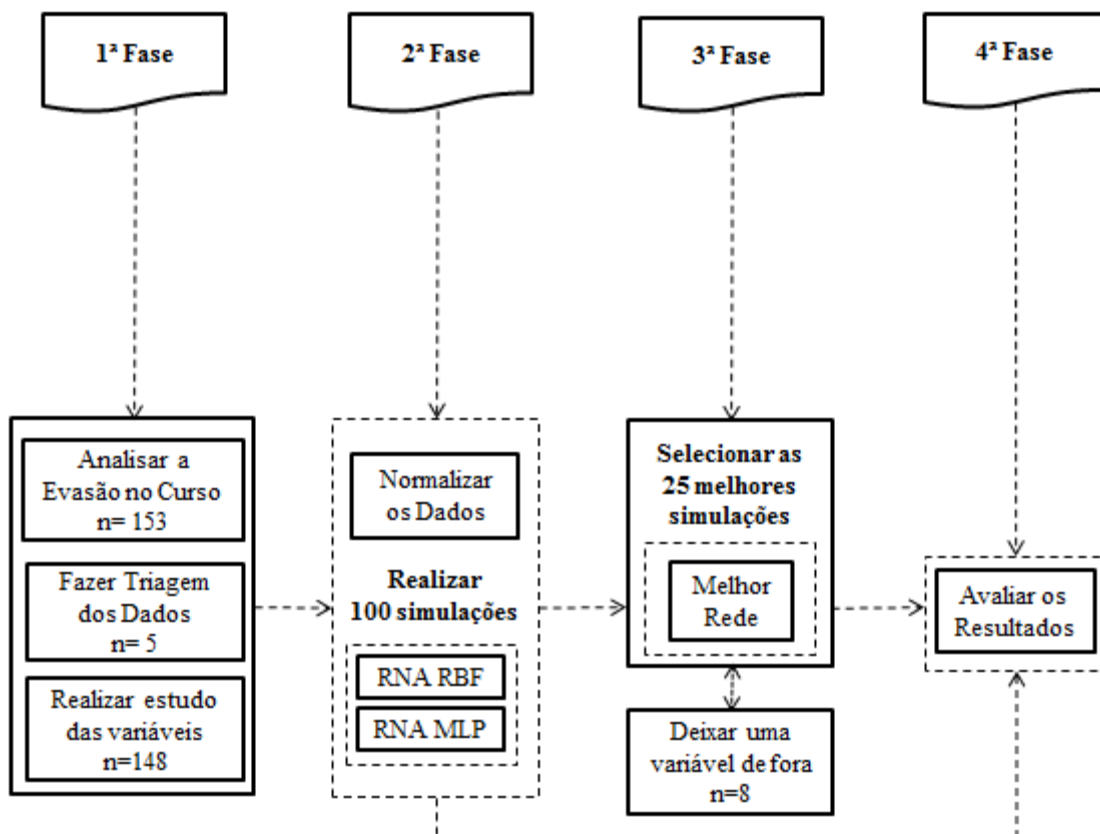


Figura 10: Fluxograma do modelo de evasão do curso de SI da UFRRJ.

3.2.1 Primeira fase

A primeira fase corresponde ao pré-processamento e foi dividida em três etapas: analisar a evasão no curso, fazer triagem e realizar estudo das variáveis.

A análise da evasão foi realizada até o 3º período, este corte permite termos uma amostra com dados referentes as turmas de 2010 a 2014, visto que os alunos ingressantes em 2014 cursaram somente até este semestre. Em seguida foi realizada uma análise das variáveis envolvidas no processo de evasão do curso;

Foram considerados evadidos todos os alunos que se desligaram (D) do curso definitivamente, seja por cancelamento, abandono ou transferência interna ou externa (não houve falecimento). Sendo considerado abandono em determinado período o momento exato em que o aluno largou o curso e não considerando o trancamento automático. Pois, pode acontecer do discente, por exemplo, abandonar o curso no 1º período, ocorrer o trancamento automático no 2º e o sistema informa que ele foi evadido no 3º período, o que não aconteceu.

De acordo com a deliberação nº 58 de 05/09/2007 da UFRRJ, o aluno da graduação que não efetuar a sua matrícula em determinado período terá a sua matrícula trancada automaticamente e se no período subsequente caso não faça a manutenção de matrícula ou o destrancamento (continuar com o curso) será desligado.

Para o cálculo da Evasão por turma ou por período foi utilizada a fórmula de Rodrigues (2013) com adaptações; trocando-se somente o termo I(n)- ingressantes no ano ou semestre por M(n) – matriculados em determinado ano ou período (Equação 31).

$$E(n) = \frac{D(n) * 100}{M(n)} \quad (31)$$

n: ano ou período que se deseja estimar a evasão;

E(n) : taxa de evasão;

D(n): número de desligados;

M(n): número de matriculados.

Na etapa de triagem foram verificadas as inconsistências dos dados. Nela foi realizado o corte dos dados faltantes, sendo retirados 5 dados.

Para a realização do estudo das variáveis envolvidas no processo de evasão foram utilizadas 148 amostras, sendo categorizadas de acordo com a tabela 5.

Tabela 5: Categorias das variáveis envolvidas no processo de evasão do curso de SI da UFRRJ.

Variáveis	Categorias
Tempo de ingresso (E.M x E.S)	1 ano ≠ 1 ano (0 ou >1 ano)
Origem Escolar	Pública Privada
Nota em matemática (ENEM)	≥700 <700
Distância	>40km ≤40km
Estado civil	Solteiro Casado
Área de melhor desempenho (ENEM)	Linguagens ou Humanas Matemática Ciências da Natureza
Aprovação em cálculo	Aprovado Não Aprovado
Média das disciplinas específicas	<6 entre 6 e 9 >9
Gênero	Masculino Feminino

3.2.2 Segunda fase

A segunda fase consistiu em realizar a normalização dos dados, a realização das 100 simulações aleatórias e a comparação entre as redes MLP e RBF.

Na etapa de normalização dos dados estes foram categorizados e normalizados para que fosse possível a sua utilização nas redes neurais.

Os dados que não possuíam inconsistências foram codificados e em seguida normalizados na escala de 0 a 1, utilizando a Equação 32; para possibilitar a sua utilização no desenvolvimento das redes neurais MLP e RBF.

$$X(n) = \frac{X(n) - X_{\min}}{X_{\max} - X_{\min}} \quad (32)$$

n: posição da variável a ser normalizada;

X(n): variável a ser normalizada;

X_{min}: o menor valor entre as variáveis;

X_{máx}: o maior valor entre as variáveis.

Para obter as 100 simulações aleatórias, nas redes neurais MLP e RBF, foram realizadas as seguintes etapas 100 vezes (Figura 11):

- separação em dois grupos: não evadidos e evadidos;
- permutação aleatória destes grupos pelo software MATLAB (MATHWORKS, 2012);
- seleção de 80% para treino dos evadidos (n=30) e dos não evadidos (n=89);
- seleção de 20% para teste dos evadidos (n=7) e dos não evadidos (n=22).

Esta divisão possibilitou obter um conjunto proporcionalmente equivalente entre os evadidos e não evadidos, além de 100 grupos aleatórios.

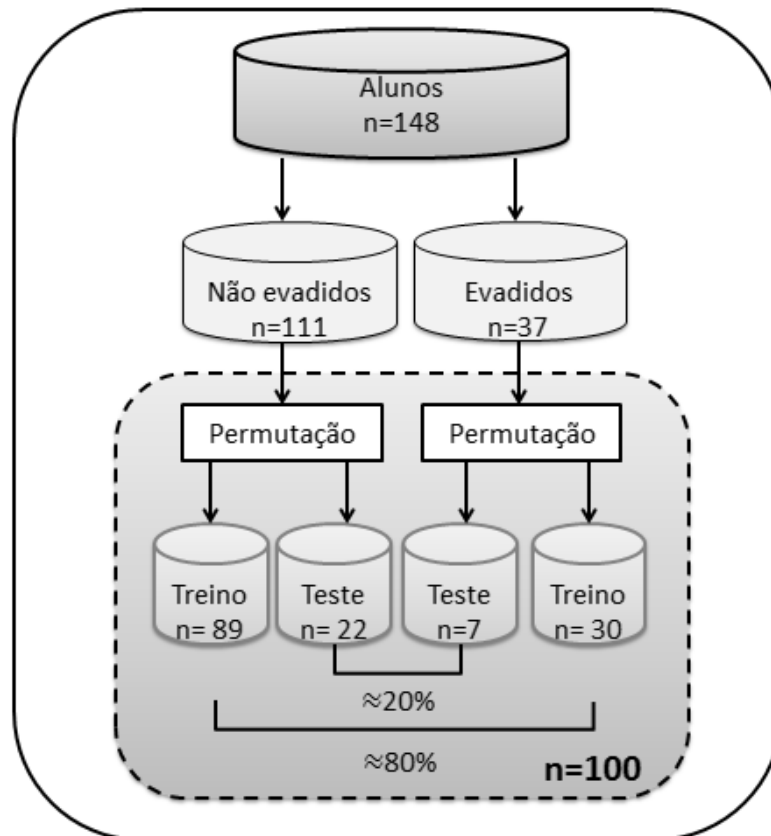


Figura 11: Criação das 100 simulações aleatórias.

Inicialmente foram utilizadas 9 entradas e em seguida 8, quando testadas todas as variáveis no modelo ou oito quando foi tirada a variável gênero. Estas variáveis são as causas da evasão descritas no modelo anterior. E foi usada uma saída para as duas redes, que no caso é a situação do aluno podendo ser evadido ou não, até o terceiro período.

Foi descrito no trabalho a melhor configuração para as duas redes após realizados diversos testes e configurações.

Para a implementação da rede RBF foi utilizada a função *newrb* do software MATLAB (MATHWORKS, 2012), tendo seus parâmetros descritos na tabela 6.

Tabela 6: Parâmetros utilizados no treinamento da rede RBF.

Parâmetros	Valor
Épocas/Máximo de neurônios	200
Espalhamento	0,5
Erro de treinamento desejável	0,01
Nº de neurônios na camada intermediária	5

A função *newrb* segundo Lima e Maia (2009), gera uma rede RBF eficiente, acrescentando neurônios até que atinja o erro especificado. O seu treinamento funciona de forma híbrida, utilizando a função gaussiana na camada escondida e a função linear na camada de saída (Figura 12).

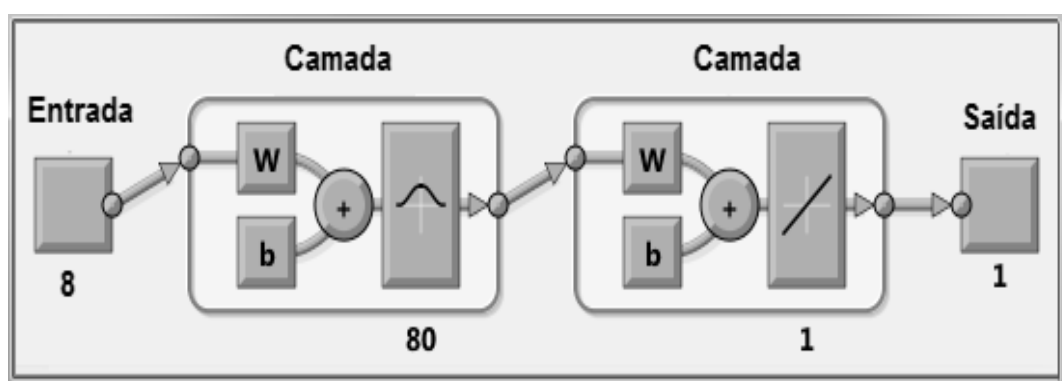


Figura 12: Aspecto da Rede RBF, adaptado do MATLAB (MATHWORKS, 2012).

“As funções de ativação gaussiana e as sigmóides são usadas quase que exclusivamente em classificadores. Por serem não lineares, possuem uma forte capacidade de separação de grupos” (DIEGUEZ, 2010, p.14).

Na implementação da rede MLP foi utilizada a função *newff* do software MATLAB (MATHWORKS, 2012), tendo os seus parâmetros descritos na tabela 7.

Tabela 7: Parâmetros utilizados no treinamento da rede MLP.

Parâmetros	Valor
Nº de épocas	200
Função de ativação	Sigmoide
Algoritmo de treinamento	<i>Levenberg-Marquardt</i>
Nº de neurônios na camada intermediária	5
Erro de treinamento desejável	0,01

A função *newff* utiliza uma função linear na camada de saída. Como é verificado na figura 9. Nesta é mostrado o aspecto que a rede possui quando utilizado 8 entradas, cinco neurônios na camada intermediária e um neurônio na camada de saída (Figura 13).

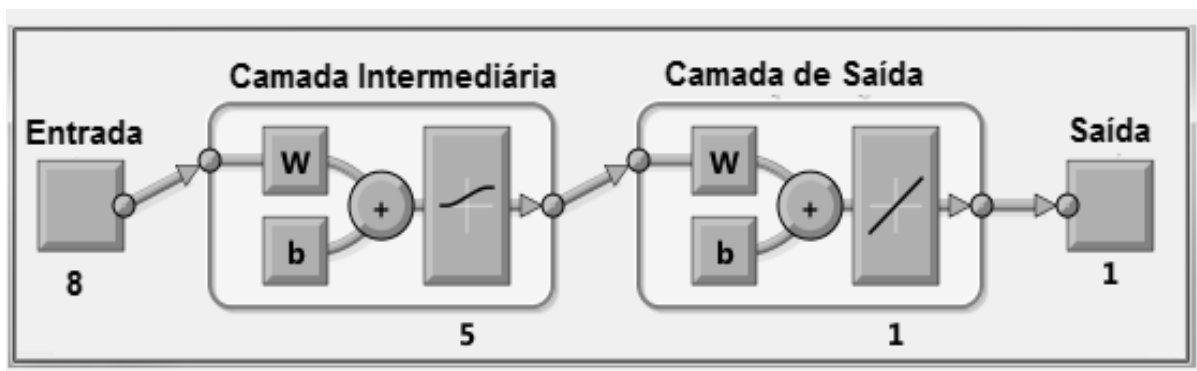


Figura 13: Aspecto da Rede MLP, adaptado do MATLAB (MATHWORKS, 2012).

Para avaliar o desempenho das redes MLP e RBF, foi utilizada a matriz de confusão. Este classificador é um método comumente utilizado por vários autores, tanto para avaliar estes dois tipos de redes como também para mostrar a eficiência da predição da evasão discente (MANHÃES et al, 2011; DOS SANTOS; DE ALBUQUERQUE, SOARES, 2014; DA COSTA; CAZELLA; RIGO, 2014).

Esta matriz é dividida em duas dimensões: valores verdadeiros e preditos. O cruzamento entre estas duas dimensões gera quatro classes: verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN). Outras medidas podem ser derivadas desta matriz, como a especificidade, sensibilidade e acurácia (MONARD; BARANAUSKAS, 2003).

Neste trabalho as classes: verdadeiro positivo e/ou verdadeiro negativo seriam os acertos da rede, ou seja, quando a rede identifica que o aluno foi evadido ou não e realmente foi isto que ocorreu. Para as outras duas classes: falso positivo e/ou falso negativo seriam os erros da rede, ou seja, quando a rede prediz que o aluno foi evadido ou não e isto não foi o que aconteceu.

Os três parâmetros derivados da matriz de confusão utilizados foram: especificidade, sensibilidade e acurácia.

A especificidade (E) foi utilizada para medir a capacidade da rede em identificar os alunos que não estão propensos à evasão (Equação 21).

$$E = \frac{VN}{(VN + FP)} \quad (21)$$

A sensibilidade (S) foi avaliada pela equação 22 e possibilitou medir a eficiência da rede em apontar os estudantes em risco de evasão.

$$S = \frac{VP}{(VP + FN)} \quad (22)$$

A acurácia (A), que é o acerto global da rede (equação 23), foi aplicada para detectar os discentes não evadidos e evadidos, ou seja, a soma do verdadeiro negativo e do verdadeiro positivo dividindo-se pelo total da amostra (n).

$$A = \frac{VN + VP}{n} \quad (23)$$

3.2.3 Terceira fase

Foi calculado a média e o desvio padrão das 25 melhores simulações da rede MLP e RBF para verificar qual a rede de melhor desempenho. Em seguida, foram utilizadas as 25 melhores interações (deixando uma variável de fora) para verificar o impacto de cada variável no desempenho da melhor rede. Isto possibilitou verificar a variação do modelo quando cada uma destas variáveis era retirada, além de averiguar a importância destas para o modelo.

3.2.4 Quarta fase

Nesta fase foram avaliados os resultados das redes MLP e RBF com relação as 25 simulações. Além da avaliação dos resultados quando retirado (deixado de fora) as variáveis envolvidas no processo de evasão. O que possibilitou verificar quais variáveis influenciaram significativamente para os alunos permanecerem ou não no curso e também a variável de maior impacto na rede.

CAPÍTULO IV

4. RESULTADOS E DISCUSSÃO

4.1. Análise exploratória da evasão no curso de SI da UFRRJ.

Através da análise da evasão por períodos verificou-se que a média da evasão no período avaliado de 2010 a 2014 diminui ao longo dos semestres, ocorrendo os maiores índices nos três primeiros períodos. De acordo com a tabela 8, verifica-se um maior percentual deste parâmetro no 1º período (11,26%) decrescendo até o valor de 6,01% no terceiro.

Tabela 8: Evasão por turma e período no curso de SI da UFRRJ.

Turmas	1º Período			2º Período			3º Período		
	M	D	Ev(%)	M	D	Ev(%)	M	D	Ev(%)
2010	30	7	23,33	23	3	13,04	20	0	0,00
2011	32	3	9,38	29	4	13,79	25	4	16,00
2012	32	2	6,25	30	3	10,00	27	2	7,41
2013	32	2	6,25	30	0	0,00	30	2	6,67
2014	27	3	11,11	24	4	16,67	20	0	0,00
Média	-	-	11,26	-	-	10,70	-	-	6,01

*M: alunos matriculados; D: alunos desligados; Ev(%):percentual de alunos evadidos.

Foi verificado que não houve diferença significativa quando comparado as proporções entre o percentual de evadidos x número de matriculados entre as turmas, em nível de 5% de significância, nos três períodos.

Um fator preocupante ocorreu na análise feita da evasão relativa por turma até o terceiro período. Quando analisado esta evasão, verificou-se que o índice de evadidos vem diminuindo de 2011 a 2013 e volta a crescer em 2014 (Figura 14). Aliado a isto, a única turma que teve vagas ociosas foi de 2014.

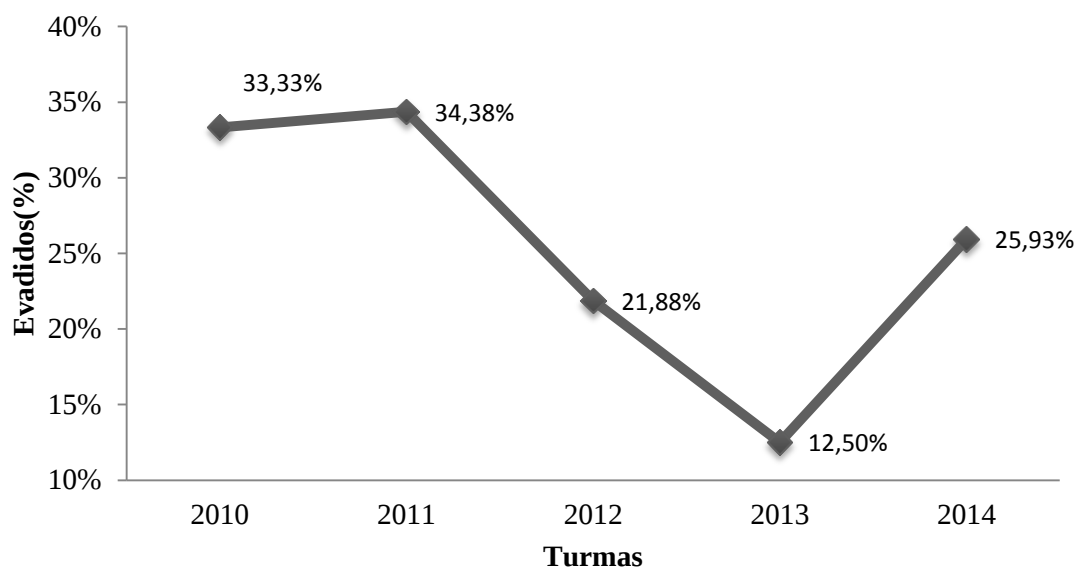


Figura 14: Evasão relativa, até o 3º período, das turmas de 2010 a 2014 do curso de SI da UFRRJ.

Apesar de não haver evadidos no terceiro período para a turma de 2014, deve-se ter uma preocupação com relação às próximas turmas, com relação ao percentual de evadidos até o terceiro período, caso este crescimento torne-se uma tendência.

4.2. Estudo das Variáveis Envolvidas no Processo de Evasão do curso de SI da UFRRJ

As variáveis foram analisadas de acordo com as suas causas, podendo ser externas a instituição/curso de Sistemas de Informação da UFRRJ ou externas a ela.

4.2.1 Causas externas à instituição/curso de Sistemas de Informação da UFRRJ

Através da tabela 9 foi possível realizar a análise das causas externas a instituição/curso de Sistemas de Informação da UFRRJ. Nela consta as variáveis envolvidas e suas categorias, de acordo com o percentual de evadidos ou não até o terceiro período. Contém as seguintes variáveis: tempo de ingresso (E.M x E.S), origem escolar, nota em matemática (ENEM); distância; estado civil, área de melhor desempenho (ENEM) e gênero.

Tabela 9: Percentual de evadidos e não evadidos com relação às causas da evasão externas a instituição/curso de Sistemas de Informação da UFRRJ.

Variável	Categoria	Ev (n)	Ev(%)	Não Ev(n)	Não Ev(%)	Total (n)
Tempo de ingresso	1 ano	2	5,4	35	94,6	37
	≠ 1 ano	35	31,5	76	68,5	111
Origem Escolar	Privada	8	20,5	31	79,5	39
	Pública	29	26,6	80	73,4	109
Nota em Matemática	≥700	6	15,4	33	84,6	39
	<700	31	28,4	78	71,6	109
Distância	≤40	3	13,0	20	87,0	23
	>40km	34	27,2	91	72,8	125
Estado civil	Casado	0	0,0	4	100,0	4
	Solteiro	37	25,7	107	74,3	144
Área de melhor Desempenho	Ling./ Hum.	22	45,8	26	54,2	48
	Matem.	15	16,9	74	83,1	89
	Ciências	0	0,0	11	100,0	11
Gênero	Feminino	16	38,1	26	61,9	42
	Masculino	21	19,8	85	80,2	106

*Ev (n): n° de alunos evadidos; Ev(%): percentual de alunos evadidos; Não Ev(n): n° de alunos não evadidos; Ev(%):percentual de alunos não evadidos.

Em relação ao tempo de ingresso do aluno no ensino superior foi verificada uma menor evasão para a categoria de 1 ano (Figura 15). Este resultado pode estar relacionado com o amadurecimento deste com relação à escolha do curso ou mesmo por já ter passado por um curso que não era a sua 1ª opção e estando atualmente no que pretendia.

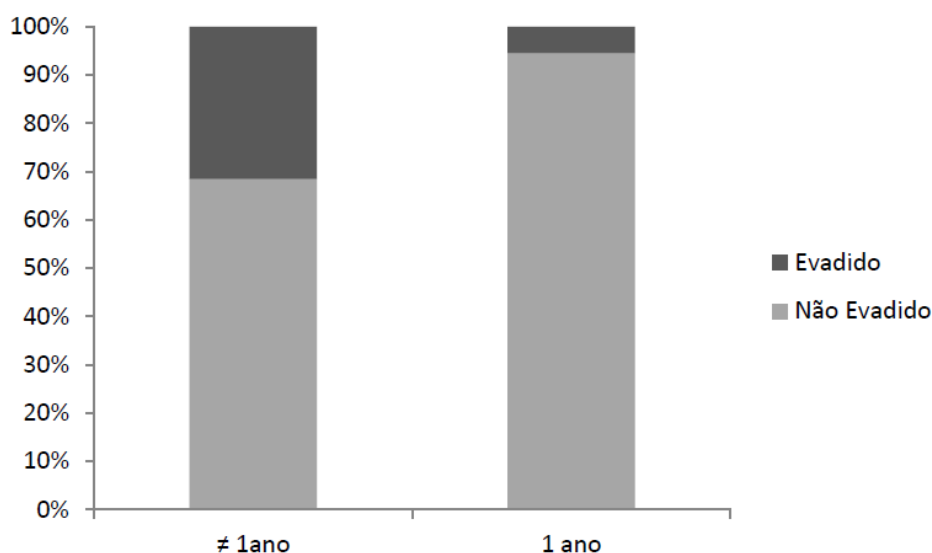


Figura 15: Percentual de evadidos e não evadidos de acordo com a categoria da variável Tempo de ingresso.

A origem escolar nesta dissertação está dividida em duas categorias: pública (federal ou estadual) e privada. Essa variável pode muitas das vezes influenciar tanto no desempenho como também ao aluno a se manter na universidade.

Os alunos da rede pública tendem a evadir mais do que os da particular (Figura 16).

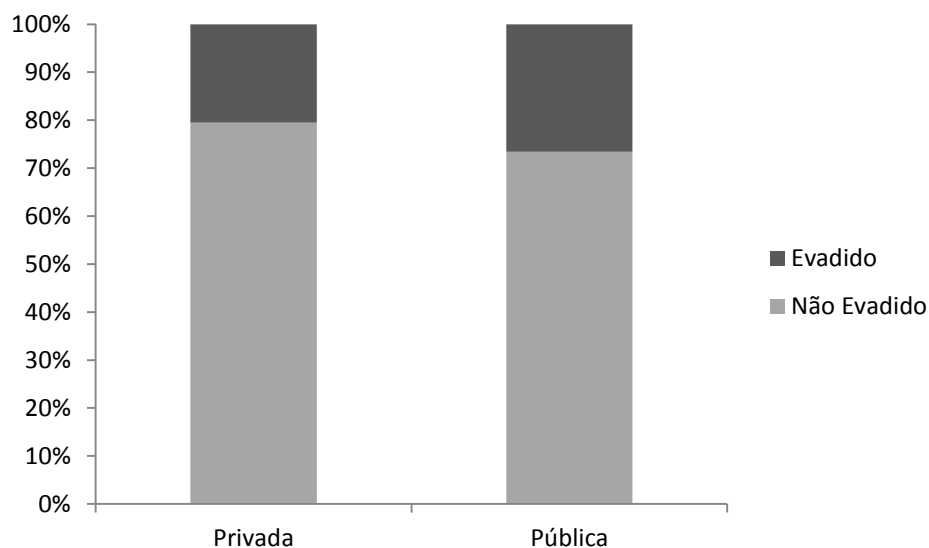


Figura 16: Percentual de evadidos e não evadidos de acordo com a categoria da variável origem escolar.

Os alunos que tiveram a nota superior a 700 em matemática e suas tecnologias no ENEM têm menos chances de serem evadidos, conforme (Figura 17).

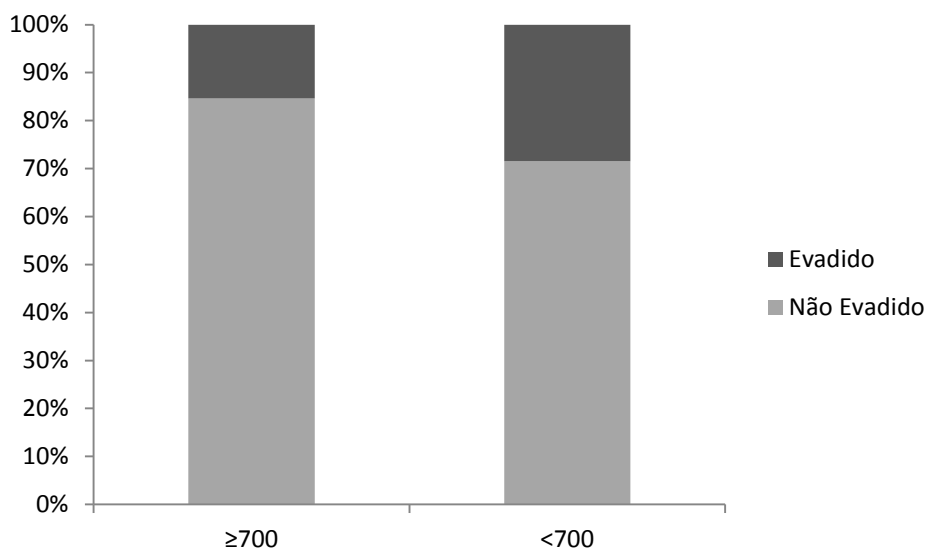


Figura 17: Percentual de evadidos e não evadidos de acordo com a categoria da variável nota em matemática no ENEM.

Em relação à variável distância, foi verificado que os alunos que residiam a mais de 40km da UFRRJ, antes do ingresso, evadiram mais (Figura 18).

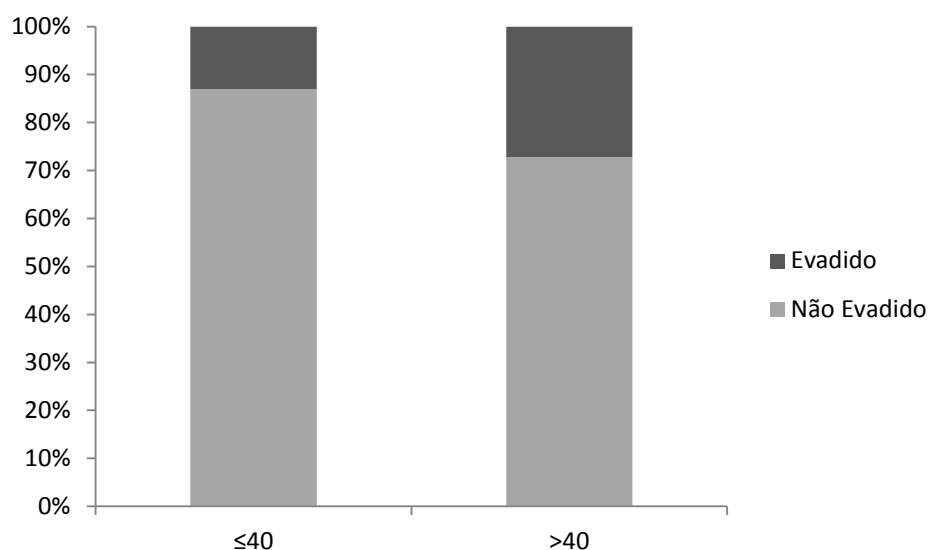


Figura 18: Percentual de evadidos e não evadidos de acordo com a categoria da variável distância.

Na variável estado civil, foi observado que os alunos que ingressaram na situação de casado não evadiram, fato este que não foi verificado para os alunos com estado civil solteiro.

Para a variável área de melhor desempenho no ENEM, que leva em consideração as quatro áreas: matemática e suas tecnologias, ciências da natureza, ciências humanas e linguagens e suas tecnologias, foi observado que a evasão ocorre com maior frequência com os alunos que apresentaram o melhor desempenho nas áreas de Linguagens ou Humanas. Não houve evadidos para os discentes que obtiveram o melhor desempenho em ciências da natureza (Figura 19).

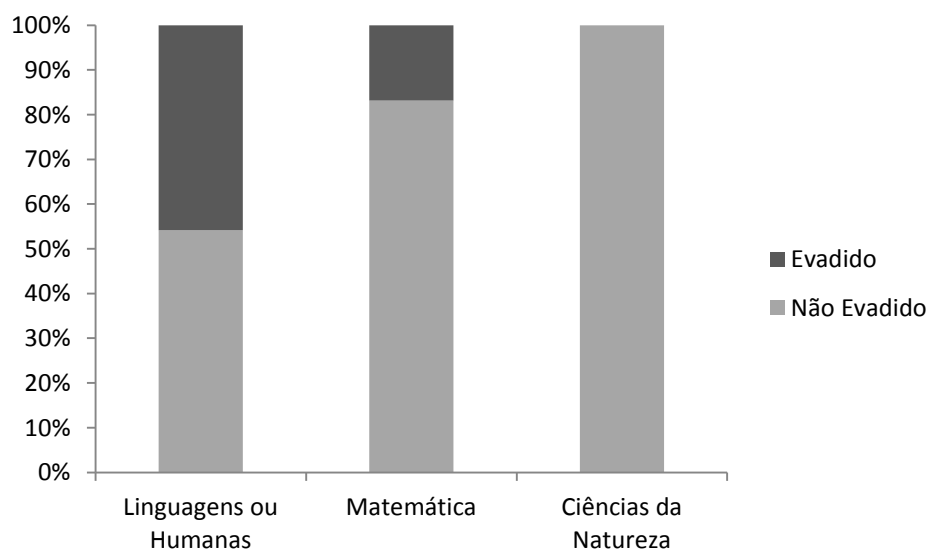


Figura 19: Percentual de evadidos e não evadidos de acordo com a categoria da variável melhor desempenho no ENEM.

Com relação ao gênero, como era de se esperar, as mulheres evadem mais que os homens (Figura 20).

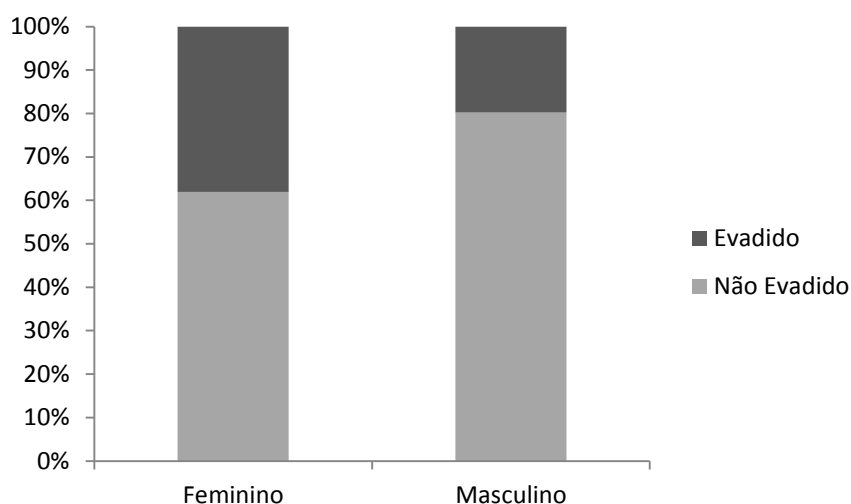


Figura 20: Percentual de evadidos e não evadidos de acordo com a categoria da variável gênero.

Ao analisar as características (categorias) que os alunos do curso de Sistemas de Informação da UFRRJ mais apresentavam ao ingressar no curso verificou-se também que a maioria destes possuía também atributos com maior índice de evasão, exceto o gênero masculino e o melhor desempenho na área de matemática (ENEM), que são atributos com um percentual inferior de evadidos (Tabela 10).

Tabela 10: Percentual das características (categorias) que os alunos do curso de Sistemas de Informação da UFRRJ mais possuíam ao ingressar no curso.

Variável	Categoria	Total(n)	Total(%)
Tempo de ingresso	≠1ano	111	75,0
Origem Escolar	Pública	109	73,6
Nota em matemática no ENEM	<700	109	73,6
Distância	>40km	125	84,5
Estado civil	Solteiro	144	97,3
Área de melhor desempenho no ENEM	Matemática*	89	60,1
Gênero	Masculino*	106	71,6

*categorias com menor índice de evasão.

Paralelamente a isto, os dois atributos que possuíram o maior índice de evasão são os do gênero feminino e a área de melhor desempenho em Linguagens ou Humanas (Figura 21).

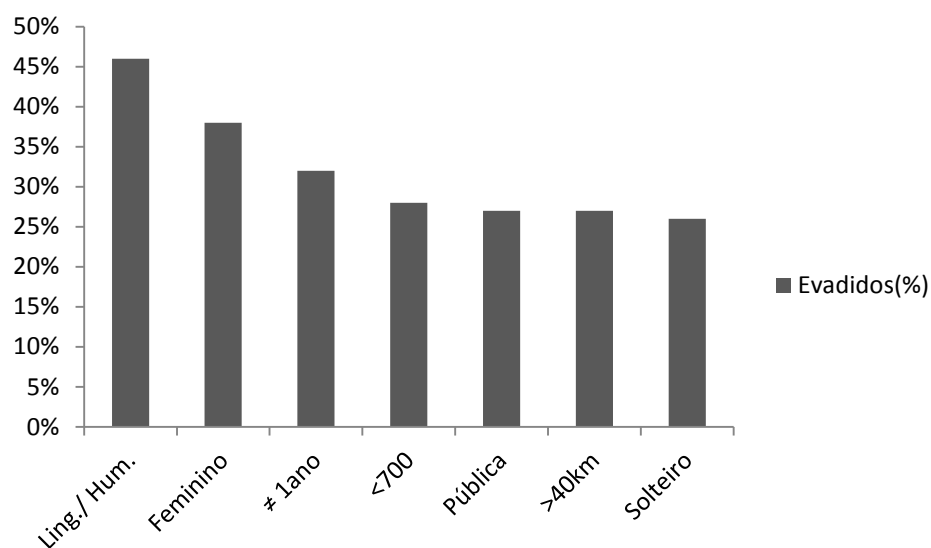


Figura 21: Percentual de evadidos de acordo com as sete categorias com maior índice de evasão (causas externas).

De acordo com a literatura a maioria dos homens preferem os cursos ligados à área de exatas, enquanto as mulheres as outras áreas. Observando o gráfico da figura 22, pode realmente haver uma relação entre os gêneros e as áreas. Mesmo que este curso seja da área de exatas, as mulheres possuíram melhor desempenho em outras áreas (53%) no ENEM, enquanto os homens tiveram melhor desempenho na área de matemática (65%).

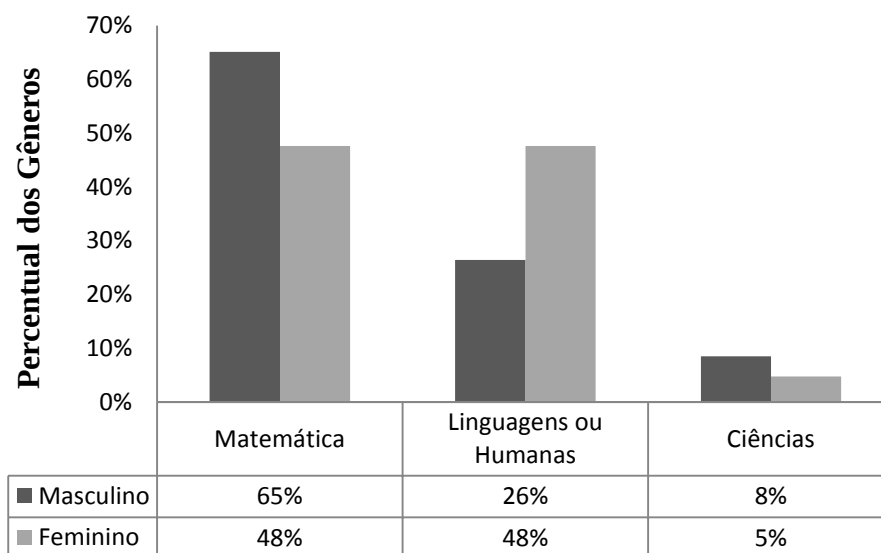


Figura 22: Percentual dos gêneros de acordo com as áreas de melhor desempenho no ENEM.

4.2.2 Causas internas a instituição/curso de Sistemas de Informação da UFRRJ

Na tabela 11 consta o percentual de evadidos e não evadidos com relação às causas internas a instituição/curso de SI da UFRRJ. Contém as seguintes variáveis: aprovação em cálculo e média das disciplinas do 1º período do curso.

Tabela 11: Percentual de evadidos e não evadidos com relação às causas da evasão externas à instituição/curso de SI da UFRRJ.

Variável	Categoria	Ev (n)	Ev(%)	Não Ev(n)	Não Ev(%)	Total (n)
Aprovação em Cálculo	Aprovado	1	5,0	19	95,0	20
	Não aprovado	36	28,1	92	71,9	128
Média do conjunto das disciplinas específicas	<6	34	49,3	35	50,7	69
	entre 6 e 9	3	4,3	67	95,7	70
	>9	0	0,0	9	100,0	9

*Ev (n): nº de alunos evadidos; Ev(%): percentual de alunos evadidos; Não Ev(n): nº de alunos não evadidos; Ev(%):percentual de alunos não evadidos.

A evasão, como seria de se esperar, ocorre de forma mais acentuada para alunos que não foram aprovados em cálculo (Figura 23).

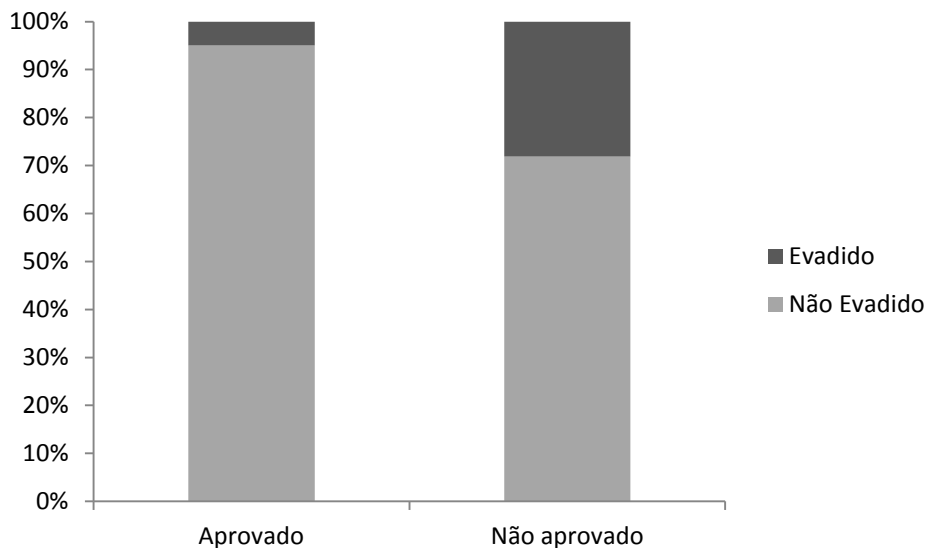


Figura 23: Percentual de evadidos e não evadidos da variável aprovação em cálculo.

A evasão ocorre com maior frequência entre os alunos que tiveram a média no conjunto de disciplinas específicas menor que 6,0. Nenhum aluno que obteve média acima de 9,0 evadiu (Figura 24).

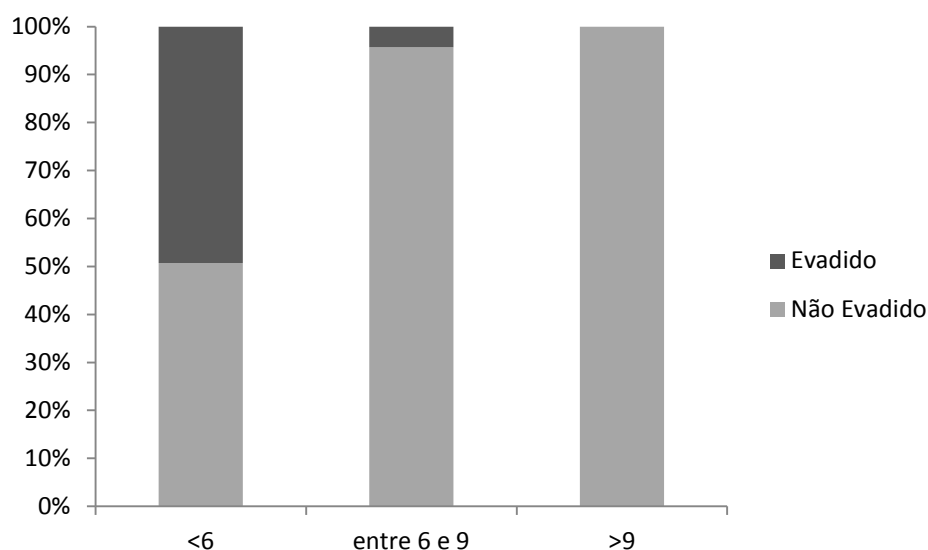


Figura 24: Percentual de evadidos e não evadidos da variável média no conjunto de disciplinas específicas do 1º período

Analisando os atributos que os alunos mais possuíam com relação as causas internas verificou-se que a maioria destes não foi aprovado na disciplina de Cálculo I e tiveram uma média entre 6 e 9 no conjunto de disciplinas específicas cursadas no 1º período do curso de Sistemas de Informação (Tabela 12).

Tabela 12: Percentual das categorias que os alunos do curso de SI da UFRRJ tiveram no 1º período do curso.

Variável	Categoria	Total (n)	Total (%)
Aprovação em Cálculo	Não aprovado	128	86,5
Média no conjunto de disciplinas específicas	entre 6 e 9	70	47,3

Cálculo I é a disciplina que possui o maior índice de reprovação dentre as matérias ofertadas no 1º período do curso de Sistemas de Informação.

Apesar da maioria dos alunos estarem com média entre 6 e 9 (conjunto de disciplinas específicas) os alunos que obtiveram desempenho inferior à 6 tiveram quase o mesmo quantitativo ($\approx 47\%$).

4.2.3 Causas da evasão com relação a todos os atributos

Dentre os nove atributos que possuem o maior índice de evasão, a média nas disciplinas específicas foi à característica que obteve o maior índice de evasão. Outro atributo que também teve um grande indicador foi a dos alunos que tiveram melhor empenho nas áreas de Linguagens ou Humanas no ENEM (Figura 25). De todos os alunos que evadiram 97% deles possuíam uma das duas características.

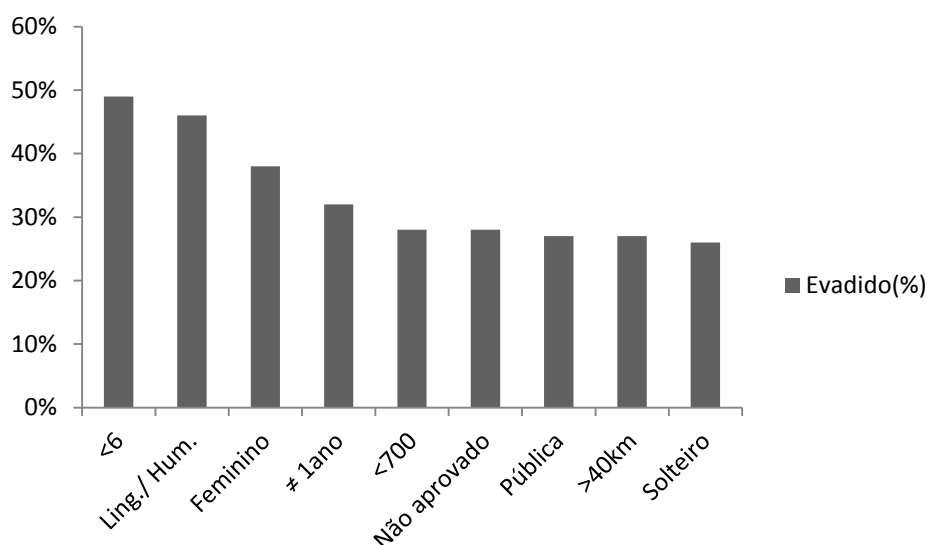


Figura 25: As nove categorias com maior índice de evasão no curso de Sistemas de Informação da UFRRJ.

Os alunos que tiveram todos os atributos com maior índice de evasão (causas externas) tiraram média inferior a 9 (conjunto de disciplinas específicas do 1º período). Demonstrando que pode haver uma relação de influência das causas externas no desempenho dos alunos do curso de Sistemas de informação da UFRRJ.

4.3. Comparação no Desempenho das Redes MLP e RBF

Ao executar as redes MLP e RBF verificou-se que a variável gênero piorou os modelos e como a finalidade desta análise é também buscar um melhor empenho das redes, optou-se por não incluí-la nesta etapa.

Após realizar o número de 100 interações, verificando as 25 melhores simulações de cada (APÊNDICES A e B), a RBF apresentou um melhor comportamento na média com relação às medidas de sensibilidade, especificidade e acurácia (Tabela 13). Além de apresentar o nível de significância menor que 5% para as três medidas. Isto demonstra estatisticamente que realmente a rede RBF foi mais eficiente.

Tabela 13: Resultados das 25 simulações com relação à média e o desvio padrão das medidas de desempenho.

	Sensibilidade (%) Média ± Desvio Padrão	Especificidade (%) Média ± Desvio Padrão	Acurácia (%) Média ± Desvio Padrão
RBF	91,09 ± 4,64*	90,86 ± 7,00*	91,03 ± 2,63*
MLP	83,64 ± 11,13*	85,14 ± 9,65*	84,00 ± 6,74*

* $p\text{-value} < 0,05$

4.4. Verificação do impacto das variáveis no desempenho da rede de melhor performance

Através das 25 melhores sequências aleatórias realizadas na rede RBF, que apresentou melhor desempenho, foi possível testar qual variável apresentou maior influência na execução dessa rede.

Nesta etapa, foi utilizado um método que deixasse uma variável de fora (*leave one out*), para observar o comportamento da rede sem ela. Este procedimento possibilitou verificar qual variável apresentou o maior impacto na rede. Sendo descrito nos Apêndices C, D, E, F, G, H, I e J as 25 simulações para cada variável retirada.

Na realização deste método foi verificado que a retirada de qualquer uma das variáveis diminuiu o desempenho da rede com relação à acurácia, mostrando que todas foram fundamentais para este modelo (Tabela 14).

Tabela 14: O percentual do índice de acurácia das 25 simulações da rede RBF quando deixado uma variável de fora.

Variável retirada	Média ± Desvio Padrão
Média do conjunto das disciplinas específicas	77,52 ± 5,98*
Tempo de ingresso (Ensino Médio x Ensino Superior)	87,59 ± 4,34*
Origem Escolar	87,72 ± 4,78*
Distância	88,97 ± 4,22*
Área de melhor desempenho (ENEM)	89,38 ± 4,77
Aprovação em Cálculo	89,66 ± 3,86
Nota em matemática (ENEM)	90,07 ± 3,35
Estado Civil	90,34 ± 3,15
Todas as variáveis	91,03 ± 2,63*

* $p\text{-value} < 0,05$

As quatro primeiras variáveis descritas na tabela 14: média das disciplinas específicas, tempo de ingresso, origem escolar e distância tiveram o nível de significância menor que 5% quando comparado com a utilização de todas as variáveis. Sendo que a média das disciplinas específicas foi a que mais influenciou, diminuindo este parâmetro de medida em 13,51%.

Com relação à sensibilidade, somente as variáveis referentes a média das disciplinas específicas e distância tiveram $p\text{-value} < 0,05$ quando comparado com a inclusão de todas as variáveis. O que significa que no conjunto de variáveis estas duas são as que mais influenciaram os alunos a evadirem, já que a tabela 15 está relacionada à sensibilidade.

Tabela 15: O percentual do índice de sensibilidade das 25 simulações da rede RBF quando deixado uma variável de fora.

Variável retirada	Média ± Desvio Padrão
Média do conjunto das disciplinas específicas	84,55±8,70*
Distância	87,45±6,05*
Aprovação em Cálculo	89,82±5,45
Estado Civil	90,18±5,36
Tempo de ingresso (Ensino Médio x Ensino Superior)	90,55±6,28
Nota em matemática (ENEM)	91,09±4,64
Área de melhor desempenho (ENEM)	92,73±5,08
Origem Escolar	94,36±4,60
Todas as variáveis	90,86 ± 7,00*

* $p\text{-value} < 0,05$

Dos alunos que foram evadidos 83,8% possuíam as duas características com maior índice de evasão das duas variáveis, sendo que todos eles possuíam uma dessas características (Figura 26).

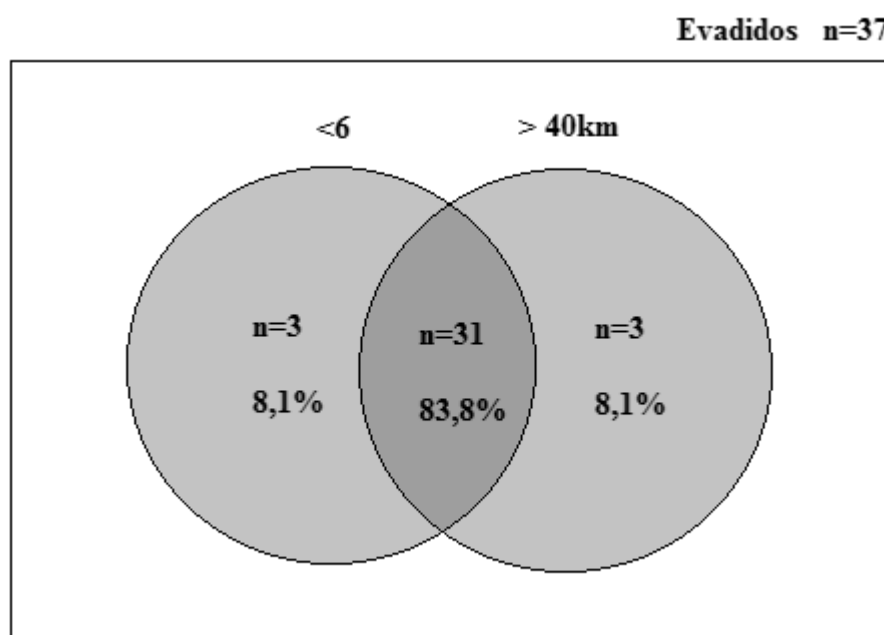


Figura 26: Percentual de evadidos em relação aos atributos de maior índice de evasão das variáveis média das disciplinas específicas e a distância.

Analisando a especificidade (Tabela 16), verificou-se que as variáveis referentes à média nas disciplinas específicas, origem escolar, tempo de ingresso e área de melhor desempenho (ENEM), influenciaram significativamente ($p\text{-valor} < 0,05$) na permanência dos discentes no curso.

Tabela 16: O percentual do índice de especificidade das 25 simulações da rede RBF quando deixado uma variável de fora.

Variável retirada	Média $\pm$ Desvio Padrão
Média do conjunto das disciplinas específicas	55,43 $\pm$ 13,25*
Origem Escolar	66,86 $\pm$ 12,86*
Tempo entre o E.M e o E.S	78,29 $\pm$ 20,67*
Área de melhor desempenho no ENEM	78,86 $\pm$ 13,75*
Nota em matemática no ENEM	86,86 $\pm$ 12,32
Aprovação em Cálculo	89,14 $\pm$ 11,87
Estado Civil	90,86 $\pm$ 7,00
Distância	93,71 $\pm$ 7,24
Todas as variáveis	90,86 $\pm$ 7,00*

* $p\text{-value} < 0,05$

Através desse resultado foi realizada uma análise nos alunos que ingressaram no curso de Sistemas de Informação após um ano do término do ensino médio e tiveram melhor desempenho na área de matemática no ENEM e verificou-se que nenhum dos discentes que possuía estes dois atributos evadiram. O que mostra a grande influência destas características na permanência dos alunos no curso de Sistemas de Informação.

Independente da origem escolar, os alunos que possuíram os seguintes atributos: média das disciplinas específicas igual ou superior a 6, tempo de ingresso igual a 1 e melhor desempenho na área de matemática não evadiram.

Analisando a origem escolar, todos os alunos oriundos de rede privada que permaneceram no curso tiveram desempenho igual ou superior a 6 na média das disciplinas específicas, o que não ocorreu com todos os alunos de origem pública. Alguns destes foram evadidos e tiveram em comum um melhor desempenho na área de humanas (ENEM) e tempo de ingresso diferente de 1.

Apesar das variáveis: aprovação em cálculo, estado civil e nota em matemática (ENEM) terem ajudado as duas redes a apresentarem um melhor desempenho, elas não obtiveram influência significativa na rede RBF para os três parâmetros de medida (sensibilidade, especificidade e acurácia).

Analisando a grade curricular (2015) verificou-se que a disciplina de Cálculo I também não prende outras disciplinas específicas do curso nos três primeiros períodos, sendo somente pré-requisito da disciplina de Cálculo II.

CAPÍTULO V

5. CONCLUSÕES E TRABALHOS FUTUROS

5.1. Conclusões

Nesta dissertação foi comparado o desempenho de um sistema computacional baseado em redes neurais MLP e RBF na análise de evasão discentes no curso de Sistemas de Informação da UFRRJ, utilizando uma base de dados composta basicamente por informações fornecidas pela coordenação do curso.

A análise dos resultados e a avaliação de desempenho demonstrou que os modelos neurais propostos são robustos, eficientes e com significativo grau de confiabilidade, alcançando índices médios de acurácia em torno de 90,0% e 84,0%, respectivamente. Entretanto, o desempenho das redes RBF apresentou um comportamento superior à MLP nas 25 melhores simulações (p-valor $<0,05$) para a média da acurácia, especificidade e sensibilidade, obtendo 91,03% de acurácia.

Com relação as turmas do curso de Sistemas de Informação, verificou-se que houve um crescimento da evasão da turma de 2014 em relação a de 2013, fato este gera uma preocupação, caso isso venha a se tornar uma tendência.

No estudo foi verificado que 97,0% dos alunos que evadiram tiveram melhor desempenho na área de Linguagens ou Humanas (ENEM) ou obtiveram a média das disciplinas específicas inferior a 6,0.

Apesar das variáveis - aprovação em cálculo, nota em matemática no ENEM e estado civil - terem contribuído para um melhor comportamento das redes, estas não influenciaram significativamente na classificação no que tange a análise da evasão.

A variável que apresentou maior impacto na performance do modelo neural proposto foi a média das disciplinas específicas, influenciando tanto para os alunos permanecerem ou não no curso. Esse resultado pode demonstrar que o conjunto de disciplinas específicas no 1º período, e não somente a disciplina de cálculo, tem grande influência nesta decisão.

Observa-se que não se identificou, na literatura pesquisada, nenhuma proposta de sistema de análise de evasão com as características e peculiaridade do modelo desenvolvido nesta dissertação.

O desempenho obtido pelo discente no primeiro período é de suma importância na decisão de permanecer no curso, fato este que permite concluir que esta variável possuiu um impacto superior a qualquer outra variável utilizada pelo modelo.

Frente aos resultados obtidos, acredita-se que a utilização de redes neurais pode ser estendida e recomendada para outros cursos de graduação onde se observam consideráveis níveis de evasão.

Uma limitação do modelo proposto é a dificuldade na obtenção e compilação dos dados para a elaboração da base de dados. O desenvolvimento de um sistema de aquisição de dados é de suma importância no caso específico da UFRRJ.

5.2. Trabalhos futuros

Apesar das redes MLP e RBF terem apresentado bons resultados na análise da evasão discente do curso de Sistemas de Informação da UFRRJ, ainda há muitas melhorias e abordagens que podem ser realizadas. A utilização de modelos de redes neurais do tipo ArtMapFuzzy, possibilita incorporar o conceito de elasticidade e plasticidade da Teoria da Ressonância Adaptativa associada a teoria da lógica nebulosa (Fuzzy) e o aprendizado estatístico das redes Bayesianas.

Devido ao alto índice de evasão nas áreas de Tecnologia da Informação é importante que ocorram maiores análises em outros cursos destas áreas a fim de diminuir esta problemática e possibilitar maiores comparações entre estes cursos.

É notável a importância de uma maior participação da comunidade acadêmica na elaboração de pesquisas relacionadas a este tema, já que isto não ocorre somente em âmbito nacional, mas internacionalmente. Comparar esta problemática em diversos cursos e instituições possibilita respostas e soluções mais eficazes.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ANDIFES. **A graduação do Sisu**. Disponível em: <http://www.andifes.org.br/?p=24734>. Acesso em: 04 abr. 2014.
- BLAZUS Cleber Augusto. **Sistema de fatores que influenciam o aluno a evadir-se dos cursos de graduação na UFSM e na UFSC: Um estudo no curso de Ciências Contábeis**. 2003. 190 f. Tese (Doutorado em Engenharia da Produção). Universidade Federal de Santa Catarina (UFSC), 2003.
- BISHOP, C, M. **Neural Networks for Pattern Recognition**. Oxford University Press, 1997.
- BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LUDEMIR, T. B. **Redes Neurais Artificiais - Teoria e Aplicações**. Ed. 2, p. 225, 2012.
- BRASSCOM. **Procuram-se profissionais de TI**. Disponível em: <http://www.brasscom.org.br/brasscom/Portugues/detNoticia.php?codArea=2&codCategoria=26&codNoticia=400>. Acesso em: 01jul 2015.
- BROOMHEAD, D, S., LOWE, D., **"Multivariable functional interpolation and adaptive networks"**, *Complex Systems*. v.2, pp. 321-355, 1988
- BUENO; J.L.O. **A evasão de alunos**. Pandéia, FFCLRP – USP, Ribeirão Preto, 1993.
- CAPELATO, R. **O que os números revelam sobre o perfil da evasão no ensino superior brasileiro**. Palestra. Seminário ABMES: Evasão e retenção: problemas e soluções. Brasília: ABMES, 2011.
- CERETTA, Raul. **Estudo apresenta dados sobre a evasão nas universidades federais**. UFSM, 2013. Disponível em: <http://site.ufsm.br/noticias/exibir/8800>. Acesso em: 04 abr 2014.
- CUNHA, L. A. (2003). **Ensino superior e universidade no Brasil**. Em E. M. T. Lopes, L. M. Faria Filho & C. G. Veiga (Orgs.), *500 anos de educação no Brasil* (3ª ed.) (pp. 151-204). Belo Horizonte: Autêntica.
- DA COSTA, S.; CAZELLA, S; RIGO, S. J. **Minerando dados sobre o desempenho de alunos de cursos de educação permanente em modalidade EAD: Um estudo de caso sobre evasão escolar na UNA-SUS**. RENOTE, 2014. v. 12, n. 2.
- DA SILVA, G. P. **Análise de evasão no ensino superior: uma proposta de diagnóstico e seus determinantes**. Avaliação: Revista da Avaliação da Educação Superior, Campinas, v. 18, n. 2, p. 311-333, 2013.
- DIEGUEZ, A. A. **Reconhecimento de Caracteres de Placa Veicular Usando Redes Neurais**. 2010. 69 f. Trabalho de conclusão de curso. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014.

DOS SANTOS, R. N.; DE ALBURQUEQUE, C.; SOARES, E. D. **Uma Abordagem Genérica de Identificação Precoce de Estudantes com Risco de Evasão em um AVA utilizando Técnicas de Mineração de Dados.** TISE. 2014, vol.10,794-799.

DURHAM, E.R. **O ensino superior no Brasil: público e privado.** São Paulo: NUPES/USP, 2003.

FIORIN, D.V. et al. **Aplicações de redes neurais e previsões de disponibilidade de recursos energéticos solares.** *Rev. Bras. Ensino Fis.* [online]. 2011, vol.33, n.1, pp. 01-20. ISSN 1806-1117. <http://dx.doi.org/10.1590/S1806-11172011000100009>.

GAIOSO, N. P. L. **Evasão discente na educação superior: a perspectiva dos dirigentes e dos alunos.** 2005. Dissertação (Mestrado em Educação) - Faculdade de Educação, Universidade Católica de Brasília, Brasília, DF, 2005.

GOLDSCHMIDT, R. R. **Uma Introdução à Inteligência Computacional: fundamentos, ferramentas e aplicações.** 1ª. ed. Rio de Janeiro - RJ: IST-Rio. 142p.

GUARNIERI, R. A. **Emprego de Redes Neurais Artificiais e Regressão Linear Múltipla no Refinamento das Previsões de Radiação Solar do Modelo ETA.** 171 p. Dissertação (Programa de Pós-graduação em Meteorologia). Instituto Nacional de Pesquisas Espaciais- INPE, 2006. São José dos Campos-SP.

HAYKIN, S. **Redes Neurais Princípios e Prática**, Ed. 2, p. 902, 2001.

HIPÓLITO, O. **País perde R\$ 9 bilhões com evasão no ensino superior, diz pesquisador.** UFCG. Disponível em: http://www.ufcg.edu.br/prt_ufcg/assessoria_imprensa/mostra_noticia.php?codigo=11662. Acesso em: 04 abr 2014.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS – INEP. Sinopses Estatísticas da Educação Superior – graduação. Disponível em: <http://www.inep.gov.br/>. Acesso em: 05 ago. 2015.

IDC BRASIL. **IDC Brasil aponta que mercado de TIC deve crescer 5% e movimentar US\$ 165,6 bilhões no país em 2015.** São Paulo: IDC BRASIL, 2015. Disponível em: <http://br.idclatin.com/releases/news.aspx?id=1779>. Acesso em: 01 abr 2015.

KARAMOUZIS, S. T.; VRETTOS, A. **An artificial neural network for predicting student graduation outcomes.** In: WORLD CONGRESS ON ENGINEERING AND COMPUTER SCIENCE – WCECS, 2008, San Francisco. *Proceedings of the WCECS*. San Francisco: Newswood Limited, 2008. p. 991-994. Disponível em: http://www.iaeng.org/publication/WCECS2008/WCECS2008_pp991-994.pdf. Acesso em: 12 nov. 2014.

LAMEIRA, R; DA SILVA, R.M. **Aplicação de Rede Bayesiana na identificação de fatores que causam a evasão no curso de Sistemas de Informação da UFRRJ.** In: Anais da I Escola Regional de Sistemas de Informação do Rio de Janeiro, p. 13-16. Niterói: SBC, 2014.

LIMA, R.A.G; MAIA, N.A. **PREDIÇÃO DE TRÁFEGO EM DOMÍNIO MPLS UTILIZANDO REDES NEURAIS ARTIFICIAIS DO TIPO RBF**.In: Anais do IX Congresso Brasileiro de Redes Neurais. Ouro Preto: SBRN, 2009.

LOBO, R. L., et al.**A evasão no ensino superior brasileiro**. Cadernos de Pesquisa 37.132 (2007): 641-659.

LOBO, M.B.C.M. **Panorama da Evasão no Ensino Superior Brasileiro: aspectos gerais das causas e soluções**. Cadernos 25. ABMES. Brasília, 2012. Evasão no Ensino Superior Brasileiro. 82p.

MACHADO, F. W. **Modelagem Chuva-Vazão Mensal Utilizando Redes Neurais Artificiais**. 174 p. Dissertação (Programa de Pós-graduação em Engenharia de Recursos Hídricos e Ambiental- Mestre em Engenharia Hidrológica).Universidade Federal do Paraná, 2005. Curitiba-PR.

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

MARTINHO, V. R.C. **Sistema inteligente para a predição de grupo de risco de evasão discente**. 2014. 145 f. Tese (doutorado) - Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade de Engenharia de Ilha Solteira, 2014.

MARTINS, C.B.N. **Evasão de alunos nos cursos de graduação em uma instituição de ensino superior**.2007. 116 f. Dissertação (mestrado). Fundação Pedro Leopoldo, 2007.

MCCULLOCH, W.S.;PITTS,W. **A Logical Calculus of the Ideas Immanent in Nervous Activity**.BulletinofMathematicalBiophysics, 1943.

MEC. Relatório da Comissão Especial para Estudos sobre Evasão nas Universidades Públicas Brasileiras. **Diplomação, Retenção e evasão nos cursos de graduação em IES públicas**. Brasília, 1996.

MENDONÇA, A. W. P.C. **A universidade no Brasil**. *Rev. Bras. Educ.* [online]. 2000, n.14, pp. 131-150.ISSN 1413-2478.

MICHELLI, C. **Interpolation of scatterd data: distance matrices and conditionally positive definite functions**. ConstructiveApproximation, v.2, 11-22, 1986.

MONARD, M. C; BARANAUSKAS, J. A. **Conceitos sobre aprendizado de máquina**. Sistemas Inteligentes-Fundamentos e Aplicações, v. 1, p. 1, 2003.

MULGREW, B. **Applying Radial Basis Functions**. IEEE Signal Processing Magazine.pp. 50-65, 1996.

NIED, Ademir. **Treinamento de redes neurais artificiais baseado em sistemas de estrutura variável com taxa de aprendizado adaptativa**.2007.128 p. Tese (doutorado) - Universidade Federal de Minas Gerais, 2007.

POWELL.M.J.D. **The Theory of Radial Basis Function Approximation in 1990**. Numerical Analysis, University of Cambridge (1990).

POGGIO.T; GIROSI.F.**Regularization algorithms for learning that are equivalent to multilayer networks**. Science, 247 (1990), pp. 978–982.

RISTOFF, D. **Evasão: Exclusão ou Mobilidade**. Santa Catarina, UFSC, 1995 (MIMEO).

RODRIGUES, F.S.; DANTE, A.C. **Estudo sobre a evasão no curso de Ciência da Computação da UFRGS**. UFRGS, 2013.

SAVIANI, D. **A expansão do ensino superior no Brasil: mudanças e continuidades**. Poiesis Pedagógica 8.2 (2011): 4-17.

SENA, M.J. **Evasão explode após adesão ao Sisu**. UFRPE. Disponível em:https://www.ufpe.br/agencia/clipping/index.php?option=com_content&view=article&id=15574:evasao-explode-apos-adesao-ao-sisu&catid=34&Itemid=122. Acesso em: 04 abr 2014.

SLHESSARENKO, M. et al. **A evasão na educação superior para o curso de bacharelado em sistema de informação**. Revista GUAL, Florianópolis, v. 7, n. 1, p. 128-147, jan. 2014.

SILVA FILHO, R.L.L.; LOBO, M.B.C.M. **Esclarecimentos metodológicos sobre os cálculos de evasão**. Disponível em:<http://institutolobo.org.br/imagens/pdf/artigos/art_078.pdf>. Acesso em: 09 nov 2014.

SOBRAL, F.J.M. **Os Desafios do Ensino Superior nos Institutos Federais**. Disponível em: <http://www.ifc-riodosul.edu.br/secao/capacitacao/artigos/Palestra%20Rio%20do%20Sul%20II.pdf>, Acessado em 10 de junho de 2015.

TONTINI, Gerson; WALTER, Silvana Anita. **Pode-se identificar a propensão e reduzir a evasão de alunos? Ações estratégicas e resultados táticos para instituições de ensino superior**. Avaliação (Campinas), Sorocaba, v. 19, n. 01, março 2014 . Disponível em:<http://educa.fcc.org.br/scielo.php?script=sci_arttext&pid=S141440772014000100005&lng=es&nrm=iso>. Acesso em: 04 dez. 2014.

ZOUCAS, F.A.M. **Aplicação de Redes Neurais para a Previsão de Séries Temporais em Segmentos Industriais**. 170 p. Dissertação (Programa de Mestrado de Engenharia Mecânica). Centro Universitário da FEI, 2009. São Bernardo do Campo-SP.

APÊNDICES

APÊNDICE A – Resultados das 25 melhores simulações da rede RBF.

A tabela abaixo apresenta os resultados das 25 melhores simulações da rede RBF.

Simulação	Especificidade (%)	Sensibilidade (%)	Acurácia (%)
1	95,45	100,00	96,55
2	90,91	100,00	93,10
3	90,91	100,00	93,10
4	95,45	85,71	93,10
5	95,45	85,71	93,10
6	95,45	85,71	93,10
7	95,45	85,71	93,10
8	95,45	85,71	93,10
9	95,45	85,71	93,10
10	95,45	85,71	93,10
11	95,45	85,71	93,10
12	95,45	85,71	93,10
13	86,36	100,00	89,66
14	86,36	100,00	89,66
15	86,36	100,00	89,66
16	90,91	85,71	89,66
17	90,91	85,71	89,66
18	90,91	85,71	89,66
19	90,91	85,71	89,66
20	90,91	85,71	89,66
21	90,91	85,71	89,66
22	90,91	85,71	89,66
23	81,82	100,00	86,21
24	81,82	100,00	86,21
25	81,82	100,00	86,21

APÊNDICE B – Resultados das 25 melhores simulações da rede MLP.

A tabela abaixo apresenta os resultados das 25 melhores simulações da rede MLP.

Simulação	Especificidade (%)	Sensibilidade (%)	Acurácia (%)
1	86,36	100,00	89,66
2	72,73	100,00	79,31
3	72,73	100,00	79,31
4	68,18	100,00	75,86
5	50,00	100,00	62,07
6	90,91	85,71	89,66
7	86,36	85,71	86,21
8	86,36	85,71	86,21
9	86,36	85,71	86,21
10	86,36	85,71	86,21
11	86,36	85,71	86,21
12	86,36	85,71	86,21
13	86,36	85,71	86,21
14	81,82	85,71	82,76
15	81,82	85,71	82,76
16	81,82	85,71	82,76
17	77,27	85,71	79,31
18	72,73	85,71	75,86
19	72,73	85,71	75,86
20	100,00	71,43	93,10
21	95,45	71,43	89,66
22	95,45	71,43	89,66
23	95,45	71,43	89,66
24	95,45	71,43	89,66
25	95,45	71,43	89,66

APÊNDICE C – Resultados das 25 simulações da rede RBF sem a variável Tempo de Ingresso (E.M.x E.S)

A tabela abaixo apresenta os resultados das 25 simulações da rede RBF sem a variável Tempo de Ingresso (E.M.x E.S)

Simulação	Especificidade (%)	Sensibilidade (%)	Acurácia (%)
1	95,45	42,86	82,76
2	90,91	85,71	89,66
3	95,45	100,00	96,55
4	95,45	71,43	89,66
5	95,45	85,71	93,10
6	95,45	57,14	86,21
7	95,45	85,71	93,10
8	90,91	100,00	93,10
9	100,00	42,86	86,21
10	90,91	85,71	89,66
11	100,00	42,86	86,21
12	95,45	85,71	93,10
13	86,36	100,00	89,66
14	86,36	71,43	82,76
15	77,27	100,00	82,76
16	90,91	85,71	89,66
17	81,82	85,71	82,76
18	90,91	85,71	89,66
19	90,91	42,86	79,31
20	86,36	100,00	89,66
21	86,36	85,71	86,21
22	95,45	57,14	86,21
23	90,91	57,14	82,76
24	81,82	100,00	86,21
25	77,27	100,00	82,76

APÊNDICE D – Resultados das 25 simulações da rede RBF sem a variável Estado Civil

A tabela abaixo apresenta os resultados das 25 simulações da rede RBF sem a variável Estado Civil.

Simulação	Especificidade (%)	Sensibilidade (%)	Acurácia (%)
1	95,45	100,00	96,55
2	86,36	100,00	89,66
3	86,36	100,00	89,66
4	95,45	85,71	93,10
5	95,45	85,71	93,10
6	95,45	85,71	93,10
7	95,45	85,71	93,10
8	95,45	85,71	93,10
9	95,45	85,71	93,10
10	95,45	85,71	93,10
11	95,45	85,71	93,10
12	95,45	85,71	93,10
13	81,82	100,00	86,21
14	86,36	100,00	89,66
15	86,36	100,00	89,66
16	90,91	85,71	89,66
17	86,36	85,71	86,21
18	90,91	85,71	89,66
19	90,91	85,71	89,66
20	86,36	85,71	86,21
21	90,91	85,71	89,66
22	90,91	85,71	89,66
23	81,82	100,00	86,21
24	86,36	100,00	89,66
25	77,27	100,00	82,76

APÊNDICE E – Resultados das 25 simulações da rede RBF sem a variável Origem Escolar

A tabela abaixo apresenta os resultados das 25 simulações da rede RBF sem a variável Origem Escolar.

Simulação	Especificidade (%)	Sensibilidade (%)	Acurácia (%)
1	95,45	71,43	89,66
2	90,91	85,71	89,66
3	95,45	85,71	93,10
4	100,00	57,14	89,66
5	95,45	85,71	93,10
6	95,45	71,43	89,66
7	100,00	85,71	96,55
8	95,45	71,43	89,66
9	100,00	42,86	86,21
10	100,00	57,14	89,66
11	100,00	57,14	89,66
12	100,00	71,43	93,10
13	95,45	71,43	89,66
14	86,36	71,43	82,76
15	90,91	57,14	82,76
16	90,91	71,43	86,21
17	95,45	57,14	86,21
18	100,00	85,71	96,55
19	90,91	42,86	79,31
20	90,91	71,43	86,21
21	95,45	57,14	86,21
22	90,91	57,14	82,76
23	86,36	57,14	79,31
24	86,36	71,43	82,76
25	90,91	57,14	82,76

APÊNDICE F – Resultados das 25 simulações da rede RBF sem a variável Nota em Matemática no ENEM

A tabela abaixo apresenta os resultados das 25 simulações da rede RBF sem a variável Nota em Matemática no ENEM

Simulação	Especificidade (%)	Sensibilidade (%)	Acurácia (%)
1	95,45	71,43	89,66
2	90,91	85,71	89,66
3	90,91	100,00	93,10
4	95,45	71,43	89,66
5	95,45	71,43	89,66
6	95,45	71,43	89,66
7	95,45	100,00	96,55
8	95,45	71,43	89,66
9	95,45	100,00	96,55
10	95,45	71,43	89,66
11	95,45	100,00	96,55
12	95,45	71,43	89,66
13	86,36	100,00	89,66
14	86,36	85,71	86,21
15	86,36	100,00	89,66
16	90,91	85,71	89,66
17	90,91	85,71	89,66
18	90,91	100,00	93,10
19	90,91	100,00	93,10
20	90,91	85,71	89,66
21	90,91	85,71	89,66
22	90,91	71,43	86,21
23	81,82	100,00	86,21
24	81,82	100,00	86,21
25	81,82	85,71	82,76

APÊNDICE G – Resultados das 25 simulações da rede RBF sem a variável Distância.

A tabela abaixo apresenta os resultados das 25 simulações da rede RBF sem a variável Distância.

Simulação	Especificidade (%)	Sensibilidade (%)	Acurácia (%)
1	95,45	100,00	96,55
2	86,36	100,00	89,66
3	81,82	100,00	86,21
4	95,45	85,71	93,10
5	90,91	100,00	93,10
6	95,45	85,71	93,10
7	95,45	100,00	96,55
8	90,91	85,71	89,66
9	90,91	85,71	89,66
10	95,45	85,71	93,10
11	95,45	100,00	96,55
12	90,91	85,71	89,66
13	81,82	100,00	86,21
14	81,82	100,00	86,21
15	81,82	100,00	86,21
16	81,82	85,71	82,76
17	90,91	85,71	89,66
18	86,36	100,00	89,66
19	86,36	85,71	86,21
20	86,36	85,71	86,21
21	81,82	100,00	86,21
22	86,36	85,71	86,21
23	81,82	100,00	86,21
24	77,27	100,00	82,76
25	77,27	100,00	82,76

APÊNDICE H – Resultados das 25 simulações da rede RBF sem a variável Área de melhor desempenho no ENEM.

A tabela abaixo apresenta os resultados das 25 simulações da rede RBF sem a variável Área de melhor desempenho no ENEM.

Simulação	Especificidade (%)	Sensibilidade (%)	Acurácia (%)
1	100,00	71,43	93,10
2	90,91	85,71	89,66
3	95,45	85,71	93,10
4	95,45	71,43	89,66
5	95,45	57,14	86,21
6	100,00	71,43	93,10
7	90,91	85,71	89,66
8	95,45	71,43	89,66
9	95,45	100,00	96,55
10	95,45	85,71	93,10
11	95,45	85,71	93,10
12	95,45	71,43	89,66
13	81,82	100,00	86,21
14	95,45	85,71	93,10
15	90,91	85,71	89,66
16	81,82	71,43	79,31
17	90,91	71,43	86,21
18	90,91	85,71	89,66
19	100,00	85,71	96,55
20	90,91	42,86	79,31
21	90,91	71,43	86,21
22	95,45	85,71	93,10
23	90,91	100,00	93,10
24	90,91	57,14	82,76
25	81,82	85,71	82,76

APÊNDICE I – Resultados das 25 simulações da rede RBF sem a variável Média das Disciplinas Específicas.

A tabela abaixo apresenta os resultados das 25 simulações da rede RBF sem a variável Média das Disciplinas Específicas.

Simulação	Especificidade (%)	Sensibilidade (%)	Acurácia (%)
1	90,91	42,86	79,31
2	72,73	71,43	72,41
3	86,36	71,43	82,76
4	90,91	42,86	79,31
5	90,91	42,86	79,31
6	86,36	42,86	75,86
7	90,91	71,43	86,21
8	86,36	57,14	79,31
9	90,91	42,86	79,31
10	90,91	42,86	79,31
11	77,27	42,86	68,97
12	86,36	57,14	79,31
13	90,91	71,43	86,21
14	77,27	57,14	72,41
15	77,27	57,14	72,41
16	90,91	71,43	86,21
17	72,73	71,43	72,41
18	90,91	57,14	82,76
19	77,27	42,86	68,97
20	81,82	71,43	79,31
21	86,36	57,14	79,31
22	59,09	71,43	62,07
23	77,27	57,14	72,41
24	95,45	28,57	79,31
25	95,45	42,86	82,76

APÊNDICE J – Resultados das 25 simulações da rede RBF sem a variável Aprovação em Cálculo.

A tabela abaixo apresenta os resultados das 25 simulações da rede RBF sem a variável Aprovação em Cálculo.

Simulação	Especificidade (%)	Sensibilidade (%)	Acurácia (%)
1	95,45	100,00	96,55
2	90,91	100,00	93,10
3	86,36	100,00	89,66
4	95,45	85,71	93,10
5	95,45	85,71	93,10
6	95,45	85,71	93,10
7	95,45	85,71	93,10
8	95,45	85,71	93,10
9	90,91	85,71	89,66
10	90,91	85,71	89,66
11	95,45	85,71	93,10
12	95,45	85,71	93,10
13	81,82	100,00	86,21
14	86,36	100,00	89,66
15	81,82	100,00	86,21
16	90,91	85,71	89,66
17	90,91	85,71	89,66
18	90,91	85,71	89,66
19	90,91	42,86	79,31
20	86,36	85,71	86,21
21	90,91	85,71	89,66
22	90,91	85,71	89,66
23	81,82	100,00	86,21
24	81,82	100,00	86,21
25	77,27	100,00	82,76