

Leidiane Teixeira dos Reis

Redes Neurais Artificiais na Identificação da Propensão à Evasão nas Licenciaturas do IFNMG-Salinas

Salinas - Minas Gerais

26 de agosto de 2021

Leidiane Teixeira dos Reis

Redes Neurais Artificiais na Identificação da Propensão à Evasão nas Licenciaturas do IFNMG-Salinas

Trabalho de conclusão de curso apresentado
ao Curso de Bacharelado em Sistemas de
Informação do Instituto Federal do Norte
de Minas Gerais, como exigência para
obtenção do grau de Bacharel em Sistemas
de Informação.

Instituto Federal do Norte de Minas Gerais - Campus Salinas

Bacharelado em Sistemas de Informação

Orientador: Alison Zille Lopes

Salinas - Minas Gerais

26 de agosto de 2021

Leidiane Teixeira dos Reis Redes Neurais Artificiais na Identificação da Propensão à Evasão nas Licenciaturas do IFNMG-Salinas/ Leidiane Teixeira dos Reis. – Salinas - Minas Gerais, 26 de agosto de 2021- 54 p. : il. (algumas color.) ; 30 cm.

Orientador: Alison Zille Lopes

– Instituto Federal do Norte de Minas Gerais - Campus Salinas

Bacharelado em Sistemas de Informação , 26 de agosto de 2021.

1. Redes Neurais Artificiais. 2. Multilayer Perceptron. 3. Evasão 4. Ensino Superior. 5. Classificação I. Alison Zille Lopes. II. Instituto Federal do Norte de Minas Gerais - Campus Salinas. III. Bacharelado em Sistemas de Informação. IV. Título



Ministério da Educação
Secretaria de Educação Profissional e Tecnológica
Instituto Federal do Norte de Minas Gerais

ATA DE REUNIÃO

ATA DE DEFESA DE TRABALHO DE CONCLUSÃO DE CURSO

ATA DE DEFESA DE TRABALHO DE CONCLUSÃO DE CURSO da aluna **Leidiane Teixeira dos Reis** - registro de matrícula de número 142382. Às **13** horas do dia **17** do mês de **agosto** de **2021**, reuniu-se de modo virtual, pela ferramenta Google Meet, a Banca Avaliadora do Trabalho de Conclusão de Curso para julgar o trabalho intitulado "**Redes Neurais Artificiais na Análise de Evasão nas Licenciaturas do IFNMG-Salinas**". O **Professor Alison Zille Lopes, orientador da aluna** e Presidente da Banca, abriu a sessão apresentando os demais membros da Banca, os Professores **Arthur Faria Porto** e **Patrícia de Oliveira e Lucas**. Dando continuidade aos trabalhos, o **Presidente** informou aos presentes que, de acordo com o Regulamento de Trabalho de Conclusão de Curso do Bacharelado em Sistemas de Informação no seu Art. 32, será considerado APROVADO na qualificação o discente que obtiver nota igual ou superior a 60. Em seguida, deu início a apresentação da aluna. Ao final da apresentação, seguiu-se a arguição da aluna pelos membros da Banca. Logo após o término da arguição a Banca Avaliadora se reuniu, sem a presença da aluna e do público, para avaliação do trabalho, constituída de deliberação individual e expedição do resultado final de acordo com o Regulamento de Trabalho de Conclusão de Curso do Bacharelado em Sistemas de Informação no seu Art. 31. As deliberações individuais de cada membro da banca, bem como o resultado final estão no **ANEXO 0916577**. Tendo como base as deliberações dos membros da Banca Avaliadora, o Trabalho de Conclusão de Curso foi **APROVADO**. O resultado final de **APROVAÇÃO** foi comunicado publicamente à aluna pelo Presidente da Banca. Nada mais havendo a tratar o **Presidente** encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros da Banca Avaliadora. Salinas, 17 de agosto de 2021.

ATENÇÃO: INSERIR COMO ANEXO, NA ORDEM DE NUMERAÇÃO, O PDF COM AS PLANILHAS USADAS NA AVALIAÇÃO.



Documento assinado eletronicamente por **Alison Zille Lopes, Professor(a) do Ensino Básico, Técnico e Tecnológico**, em 18/08/2021, às 12:25, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Patrícia de Oliveira e Lucas, Professor(a) do Ensino Básico, Técnico e Tecnológico**, em 19/08/2021, às 12:02, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Arthur Faria Porto, Professor(a) do Ensino Básico, Técnico e Tecnológico**, em 19/08/2021, às 12:10, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).

A autenticidade deste documento pode ser conferida no site

[http://sei.ifnmg.edu.br/sei/controlador_externo.php?](http://sei.ifnmg.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0)

[acao=documento_conferir&id_orgao_acesso_externo=0](#), informando o código verificador **0915304** e o código CRC **6EC0E876**.

Dedico este trabalho à minha família e a todos os meus professores.

Agradecimentos

Agradeço a Deus, por ter me dado forças para continuar persistindo mesmo nos momentos difíceis.

A minha família pelo apoio nesta jornada, em especial a minha mãe Cleusa por sempre acreditar em mim e nunca permitir que eu desistisse e a minha irmã Luana pelo apoio e participação no projeto.

Ao meu professor e orientador Alison Zille Lopes por todo suporte, apoio e conhecimento transmito que serviram de inspiração para o desenvolvimento deste trabalho.

A minha amiga e colega de turma Janina Barbosa de Aguiar pelo apoio e ajuda nesta jornada que trilhamos juntas.

A todos os professores do curso de Bacharelado em Sistemas de Informação do IFNMG-Salinas por todo conhecimento transmito.

Ao Instituto Federal do Norte de Minas Gerais - Campus Salinas (IFNMG-Salinas) pelo suporte dado para o desenvolvimento deste trabalho com o fornecimento dos dados.

“O insucesso é apenas uma oportunidade para recomeçar de novo com mais inteligência.”

Henry Ford

Resumo

A evasão escolar é um fenômeno complexo que atinge todos os níveis de ensino, gerando prejuízos acadêmicos, sociais e econômicos. Ao longo dos anos, estudos do tema vêm se proliferando, fomentados pela busca por compreender e combater esse fenômeno. Nesse cenário, as redes neurais artificiais vêm se destacando com a produção de classificadores da propensão à evasão. Dessa forma, o presente trabalho objetivou, utilizando redes Multilayer Perceptron (MLP), produzir um classificador que traduza dados socioeconômicos na situação discente, evasão ou permanência, para estudantes dos cursos de licenciatura do IFNMG-Salinas. Assim, uma base de dados formada por 256 amostras, cedidas pelo IFNMG-Salinas, foi utilizada no treinamento (192) e na avaliação (64) de diferentes redes MLP. O melhor classificador MLP, obteve resultados promissores, apresentando acurácia geral de 67,19% e identificando cerca de 72% dos discentes evadidos. Logo, redes MLP são potenciais aliados no combate à evasão e, ao utilizar dados socioeconômicos, ajudariam na identificação de alunos ingressantes propensos à evasão.

Palavras-chave: Redes Neurais Artificiais, Multilayer Perceptron, Evasão, Ensino Superior, Classificação.

Abstract

School dropouts are a complex phenomenon that affects all levels of education, causing academic, social, and economic losses. Over the years, studies on the subject have been proliferating due to the search to understand and fight back this phenomenon. In this scenario, artificial neural networks have been standing out by producing classifiers to predict the inclination to dropout. Thus, the present work objective was to adjust a Multilayer Perceptron (MLP) neural network classifier that translates socioeconomic data into the student situation, dropout or permanence, based on data from teaching degree courses at IFNMG-Salinas. A dataset formed by 256 samples, provided by IFNMG-Salinas, was used in training (192) and evaluation (64) of different MLP networks. The best MLP classifier achieved promising results, an overall accuracy of 67.19% and a 72% success rate in dropout identification. So, MLP networks are potential allies in the fight against dropouts and, by using socioeconomic data, they would help to identify newcomer students prone to dropout.

Keywords: Artificial Neural Networks, School Dropout, Higher Education, Multilayer Perceptron, Classification.

Lista de Ilustrações

Figura 1 – Neurônio Artificial	27
Figura 2 – Funções de ativação: (a) Sigmóide, (b) Tangente Hiperbólica (c) ReLU	28
Figura 3 – Arquiteturas de RNAs: (a) Camada única (b) Múltiplas camadas (c) Recorrente	28
Figura 4 – Rede MLP com duas camadas escondidas	29
Figura 5 – Validação <i>k-fold</i> com 10 partições	30
Figura 6 – Arquitetura do modelo selecionado utilizando validação <i>k-fold</i>	37
Figura 7 – Histórico de perda logarítmica do treinamento do modelo selecionado .	38

Lista de Tabelas

Tabela 1 – Codificação 1-de-n para um atributo com três valores.	34
Tabela 2 – Topologias de rede selecionadas para avaliação.	35
Tabela 3 – Matriz de confusão para a rede MLP segundo o conjunto de treinamento	38
Tabela 4 – Matriz de confusão para a rede MLP segundo o conjunto de teste . . .	39
Tabela 5 – Modelos Avaliados	53

Lista de Abreviaturas e Siglas

IFNMG	Instituto Federal do Norte de Minas Gerais
RNA	Redes Neurais Artificiais
MLP	Multilayer Perceptron
IES	Instituição de Ensino Superior
RBF	Radial Basis Function
OECD	Organization for Economic Cooperation and Development
UFRRJ	Universidade Federal Rural do Rio de Janeiro
ENEM	Exame Nacional do Ensino Médio
RELU	Rectified Linear Unit

Sumário

1	Introdução	23
1.1	Objetivos	24
2	Revisão de Literatura	25
2.1	Ensino Superior	25
2.2	Evasão no Ensino Superior	25
2.3	Redes Neurais Artificiais	27
2.3.1	Neurônio Artificial	27
2.3.2	Arquiteturas	28
2.3.3	Aprendizagem	28
2.3.4	Multilayer Perceptron	29
2.3.5	Validação Cruzada	30
3	Materiais e Métodos	31
3.1	Base de dados	31
3.2	Ajuste e Avaliação da Rede Neural Artificial	33
3.2.1	Pré-Processamento	34
3.2.2	Topologias Avaliadas	34
3.2.3	Validação Cruzada e Treinamento	35
3.2.4	Métricas de Avaliação	36
4	Resultados e Discussão	37
5	Conclusão	41
	Referências	43
	Anexo A Questionário Socioeconômico	47
	Apêndice A Modelos Avaliados	53

Capítulo 1

Introdução

O ensino superior é uma importante ferramenta de transformação social, visto que tem a capacidade de difundir conhecimento e transformar positivamente a comunidade ao seu redor, aumentando o desenvolvimento econômico, a capacidade crítica das pessoas e a chance de obter melhores ganhos salariais [Fábio, 2016, Martins and Oliveira, 2017, Fiúza, 2019]. Contudo, o ensino superior enfrenta alguns desafios, e o maior deles é manter os alunos ingressantes nas instituições de ensino superior (IES) até a conclusão do curso, ou seja, conter a evasão escolar. No Brasil, esse fenômeno gera um desperdício de recursos na ordem dos 15 bilhões de reais anuais, desperdício de vagas, entre outros [Lobo, 2019].

A evasão no ensino superior caracteriza-se como um grande desafio para as IES devido a sua capacidade de causar diversos prejuízos econômicos, acadêmicos e sociais. Evasão significa a saída definitiva do aluno do seu curso de origem, sem concluí-lo, e pode ser causada por diferentes fatores, e esses são comumente divididos em fatores internos e externos às IES [Brasil, 1996, Santiago, 2015]. No ano de 2018, a taxa média anual de evasão no ensino superior atingiu 30,1% na rede privada e 18,5% na rede pública de ensino [Bocchini, 2018].

A evasão é um fenômeno preocupante e persistente, dessa forma, uma grande variedade de métodos vem sendo aplicados na análise desse fenômeno. Nos últimos anos, técnicas de inteligência artificial têm se popularizado entre pesquisadores que buscam compreender e combater a evasão. Dentre essas técnicas, as redes neurais artificiais (RNAs), paradigmas computacionais de processamento de informação inspirados no funcionamento do cérebro humano, vêm se destacando pelo seu potencial adaptativo a diferentes cenários. Kawase [2015], por exemplo, obteve resultados promissores ao aplicar RNAs *Multilayer Perceptron* (MLP) e *Radial Basis Function* (RBF) na análise da evasão discente do curso de Sistemas de Informação da Universidade Federal Rural do Rio de Janeiro. Soares and Barbosa [2018] produziram um estudo comparativo da capacidade de predição da evasão por redes MLP, RBF e ARTMAP-Fuzzy, sendo que a primeira obteve o melhor desempenho frente ao estudo de caso utilizado. Melo [2019] propôs um classificador MLP para auxiliar

na redução da evasão na Universidade Federal do Triângulo Mineiro, conseguindo identificar a tendência à evasão com até um mês de antecedência.

Dessa forma, o presente trabalho, visando contribuir com um futuro instrumento que identifique alunos propensos a evadir, aplicou redes neurais artificiais para classificar a situação do discente, permanência ou evasão, frente aos seus dados socioeconômicos. A arquitetura de RNAs selecionada para esse fim foi a *Multilayer Perceptron*, visto que esta é amplamente utilizada na solução de problemas que envolvem classificação de padrões [Silva et al., 2010], produzindo um classificador ajustado a partir de dados de discentes dos cursos de licenciatura do Instituto Federal do Norte de Minas Gerais (IFNMG) - Campus Salinas.

1.1 Objetivos

O objetivo geral deste trabalho foi produzir um classificador, utilizando Redes Neurais Artificiais *Multilayer Perceptron* (MLP), da propensão à evasão ou à permanência de discentes nos cursos de licenciatura do IFNMG-Salinas perante seus dados socioeconômicos.

Como objetivos específicos têm-se:

- Criar a base de dados de evasão.
- Avaliar a capacidade preditiva do modelo ajustado.
- Desenvolver um modelo que sirva de base para uma futura ferramenta de identificação de alunos propensos a evadir.

Capítulo 2

Revisão de Literatura

2.1 Ensino Superior

O ensino superior é uma importante ferramenta de transformação social devido a sua capacidade de impulsionar o desenvolvimento tecnológico e aumentar a capacidade crítica da população [Martins and Oliveira, 2017]. Ter um curso de ensino superior pode trazer, além da vantagem de adquirir conhecimento, a chance de obter melhores ganhos salariais. Segundo dados da *Organization for Economic Cooperation and Development*, nos 36 países que compõem a organização, em média os adultos (de 24 a 64 anos) com ensino superior têm uma vantagem de ganhos de cerca de 55% em relação a adultos sem ensino superior [OECD, 2018].

A educação superior no Brasil tem se expandido ao longo dos anos, o número de matrículas (graduação e sequencial) continua crescendo, atingindo a marca de 8,45 milhões de alunos em 2018. Entre os anos de 2008 e 2018, as matrículas aumentaram 44,6% [Brasil, 2019].

O crescimento do número de matrículas no ensino superior traz grandes benefícios para a sociedade e, ao mesmo tempo, alguns desafios e preocupações para as Instituições de Ensino Superior (IES), sendo uma das principais manter os alunos ingressantes na instituição até a conclusão do curso.

2.2 Evasão no Ensino Superior

A evasão no ensino superior é um fenômeno mundial que afeta os sistemas educacionais e constitui-se como um grande desafio para todas as Instituições de Ensino Superior, sejam elas públicas ou privadas [Santiago, 2015]. Em 2018, por exemplo, a taxa de evasão dos cursos de ensino superior no Brasil atingiu 30,1% na rede privada e 18,5% na rede pública [Bocchini, 2018].

Evasão é um termo que não possui uma definição única, podendo ser: evasão do curso, o estudante desliga-se do curso superior sem concluí-lo; da instituição, o estudante se desvincula da instituição; e do sistema de ensino, quando ocorre o abandono dos estudos, seja numa situação temporária ou definitiva [Brasil, 1996, Chaves, 2016]. A definição mais utilizada é a dada pela Comissão Especial de Estudos sobre Evasão nas Universidades Públicas Brasileiras, a qual consiste na saída definitiva do aluno do seu curso de origem sem concluí-lo [Brasil, 1996].

Entender os motivos pelos quais os alunos evadem é um grande desafio, uma vez que existem diferentes fatores relacionados a esse fenômeno. Esses são comumente agrupados em fatores internos e externos às IES. Os fatores internos estão ligados ao curso, abordando as questões de infraestrutura, corpo docente e assistência socioeducacional. Já os fatores externos são relacionados aos alunos e abordam aspectos como vocação, questões socioeconômicas e problemas de ordem pessoal [Santiago, 2015, Feitosa, 2016]. Os fatores causadores da evasão não são sempre os mesmos, variando de uma IES para outra, inclusive na forma que devem ser abordados. No trabalho desenvolvido por Feitosa [2016], por exemplo, os motivos que levaram os alunos a evadir foram a localização do campus, os horários das disciplinas, a conciliação com o trabalho e a desmotivação. Já Garcia et al. [2017] identificaram a falta de vocação como principal motivação, enquanto Rocha [2015] observou a dificuldade de conciliar trabalho e estudo, as questões financeiras, as questões familiares e a dificuldade de aprendizagem.

A evasão é um problema que gera consequências sérias para IES e alunos brasileiros. Nas IES, ocorre o desperdício de recursos, na ordem de 15 milhões de reais anuais, tempo e vagas. Os alunos, em contrapartida, são assolados com problemas de autoestima, dificuldade de inclusão no mercado de trabalho e substanciais perdas salariais futuras. Assim, a sociedade perde importantes contribuições para o seu desenvolvimento [Nagai and Cardoso, 2017, Gargantini, 2019, Lobo, 2019].

Dessa forma, a evasão tem sido objeto de estudo de muitos autores ao longo dos anos. Souza [2016], por exemplo, investigou a relação entre reprovações em disciplinas obrigatórias e os altos índices de evasão do curso de Licenciatura em Matemática (Noturno) da Universidade de Brasília. Monteiro [2018] caracterizou a evasão do Curso de Graduação em Saúde Coletiva da Faculdade de Ceilândia da Universidade de Brasília através do levantamento do perfil dos evadidos, identificando semestres com maior evasão e os motivos que levaram à desistência do curso. Neto et al. [2018] propuseram um modelo de análise de evasão dos cursos superiores do Instituto Federal da Bahia. Ferreira [2019] analisou os fatores que influenciaram os alunos a abandonar o curso de agronomia da Universidade Federal do Cariri.

2.3 Redes Neurais Artificiais

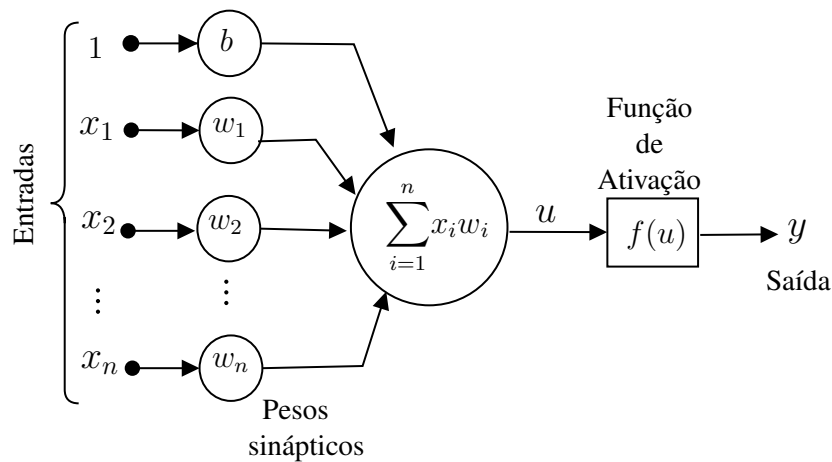
Inspiradas na forma como o cérebro humano processa informações e, assim, realizando processamento paralelo e distribuído a partir de unidades simples (neurônios artificiais), as redes neurais artificiais são técnicas de inteligência computacional capazes de aprender e extrair informações através de experiências [Haykin, 2001, Kawase, 2015].

2.3.1 Neurônio Artificial

Um neurônio artificial é um modelo matemático baseado no funcionamento de um neurônio biológico. O primeiro modelo foi criado por McCulloch e Pitts em 1943 e foi aperfeiçoado ao longo dos anos por outros autores. Esse modelo é considerado o marco inicial para o desenvolvimento das RNAs e continua sendo referência até os dias atuais [Kawase, 2015, Soares and Barbosa, 2018, Melo, 2019].

O funcionamento de um neurônio artificial (Figura 1) pode ser descrito da seguinte forma: sinais são apresentados às entradas (x_1, x_2, \dots, x_n); cada sinal é ponderado por um peso (w_1, w_2, \dots, w_n); a soma ponderada dos sinais de entrada produz o nível de atividade (u); e a função de ativação $f(u)$ é aplicada com o objetivo de limitar a saída do modelo. O bias (b), ponderação de um sinal fixo igual a um, também é usado na computação do nível de atividade e serve ao propósito de aumentar ou diminuir a influência dos sinais de entrada [Negnevitsky, 2005, Furtado, 2019].

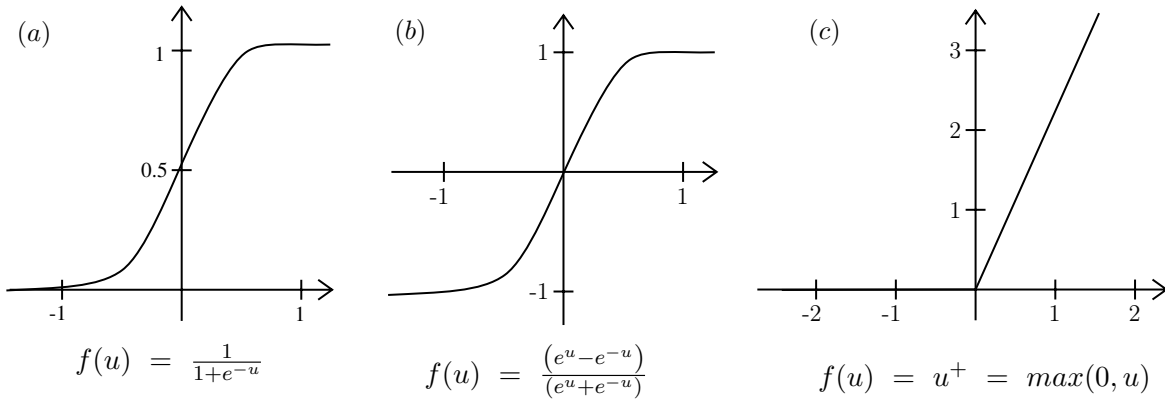
Figura 1 – Neurônio Artificial



Fonte: Adaptado de Negnevitsky [2005].

Esse modelo se apresenta constante para quase todas as redes neurais, variando somente a função de ativação, a qual limita a amplitude do sinal de saída do neurônio. Dentre as funções de ativação mais comumente utilizadas estão a Sigmóide, a Tangente Hiperbólica e a Unidade Linear Retificada (Rectified Linear Unit - ReLU) [Araújo, 2019, Furtado, 2019], apresentadas na Figura 2.

Figura 2 – Funções de ativação: (a) Sigmóide, (b) Tangente Hiperbólica (c) ReLU

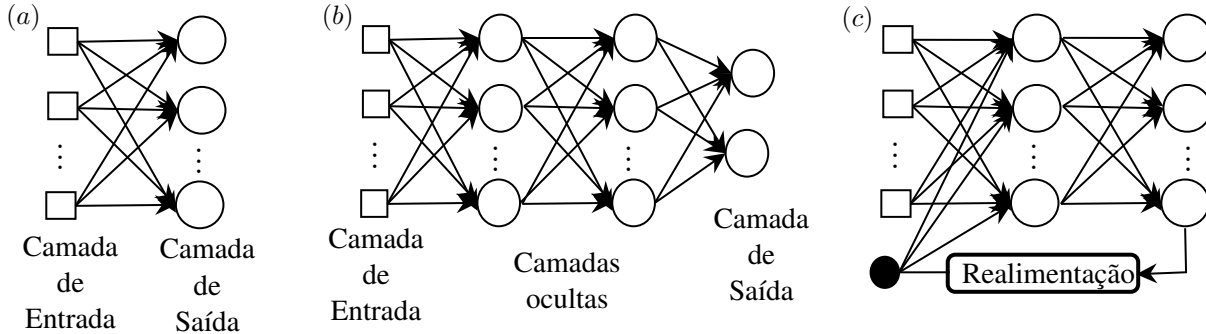


Fonte: Adaptado de Araújo [2019].

2.3.2 Arquiteturas

As arquiteturas de RNAs podem ser classificadas em: redes *feedforward* ou diretas, nas quais não existem ciclos de realimentação, e redes recorrentes, redes que possuem pelo menos um ciclo de realimentação (Figura 3).

Figura 3 – Arquiteturas de RNAs: (a) Camada única (b) Múltiplas camadas (c) Recorrente



Fonte: Adaptado de Soares and Barbosa [2018].

As redes *feedforward* são subclassificadas em redes: *feedforward* de camada única, a camada de entrada se comunica diretamente com a camada de saída, e *feedforward* de múltiplas camadas, apresentando uma ou mais camadas ocultas, situadas entre as camadas de entrada e saída, responsáveis por grande parte do processamento (extração de características) efetuado pela rede [Haykin, 2001].

2.3.3 Aprendizagem

As redes neurais artificiais têm como principal característica a capacidade de aprender a partir de seu ambiente e com isso melhorar o seu desempenho. Esse aprendizado é feito por meio de um processo iterativo de ajustes de pesos, ou seja, o treinamento. O

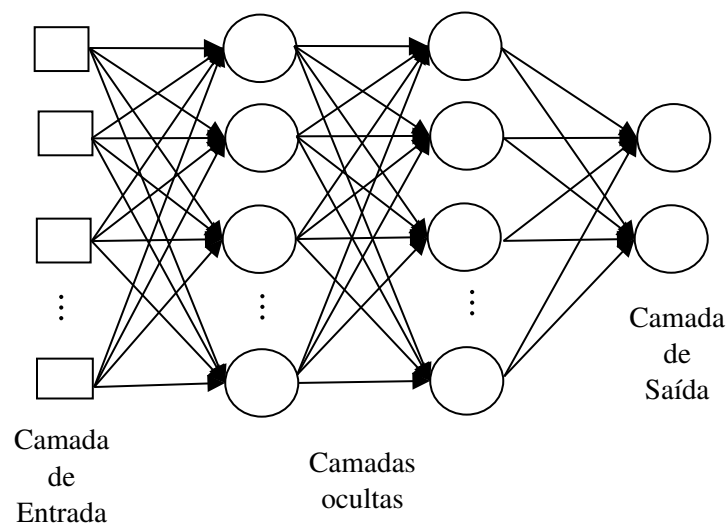
aprendizado ocorre quando a rede neural atinge uma solução que possa ser generalizada para uma classe de problemas [Haykin, 2001, Gorski, 2019].

O processo de aprendizagem é realizado através de algoritmos de treinamento, os quais podem ser organizados em aprendizado supervisionado, não-supervisionado e por reforço. No aprendizado supervisionado existe um professor (supervisor) que fornece à rede a resposta desejada para cada padrão de entrada, assim, o treinamento é realizado a partir de exemplos. Esses padrões conhecidos (exemplos) não existem no aprendizado não-supervisionado, portanto a própria rede deverá ser capaz de extrair as características relevantes dos impulsos (entradas). Já o aprendizado por Reforço é muitas vezes tratado como um caso especial de aprendizado supervisionado, pois, apesar de respostas desejadas não serem fornecidas, a rede conhece a qualidade da resposta produzida ao receber o reforço (comportamento desejado) ou a punição (comportamento indesejado) [Furtado, 2019].

2.3.4 Multilayer Perceptron

A *Multilayer Perceptron* (MLP) é uma rede neural *feedforward* de múltiplas camadas (Figura 4), que se destaca no reconhecimento de padrões. Aplicadas a problemas que envolvem classificação e regressão não linear, as MLPs são as RNAs mais populares, sendo isso devido, principalmente, à facilidade de modelagem e treinamento. A aprendizagem dessas redes é realizada através de treinamento supervisionado, geralmente, através do algoritmo *backpropagation* [Sousa, 2000, Soares, 2010]

Figura 4 – Rede MLP com duas camadas escondidas



Fonte: Adaptado de Soares and Barbosa [2018].

O *backpropagation* consiste em duas fases, *forward* (para frente) e a *backward* (para trás). Na fase *forward*, os sinais de entrada são propagados através da rede, da entrada

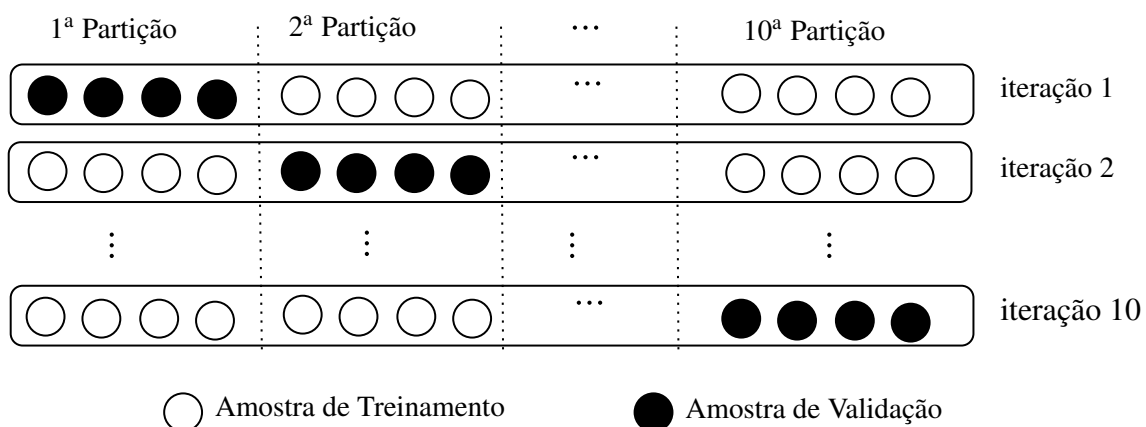
para a saída, produzindo a saída estimada. Na fase *backward*, núcleo da aprendizagem da rede, os pesos sinápticos são ajustados à medida que um sinal de erro é retropropagado através da rede. Esse sinal de erro tem origem na diferença entre a saída desejada e a estimada pela rede. Ao executar ambas as fases por múltiplas iterações, o *backpropagation* reduz essa diferença, aproximando a saída estimada pela rede da saída desejada [Haykin, 2001, Soares and Barbosa, 2018].

2.3.5 Validação Cruzada

A validação cruzada é uma das técnicas estatísticas mais utilizadas para seleção das melhores topologias para redes neurais artificiais. Os modelos são confrontados perante sua capacidade de generalização ao serem avaliados com base em um conjunto de dados diferente do usado no ajuste de parâmetros [Silva et al., 2010, Fontana, 2020].

Existem diferentes técnicas de validação cruzada, *hold-out*, *leave-one-out*, *k-fold*, entre outras, sendo que a *k-fold* se destaca como a mais empregada na avaliação de modelos [Soares and Barbosa, 2018]. No método de validação *k-fold* (Figura 5), o conjunto de dados é dividido em k subconjuntos (partições), sendo um subconjunto utilizado na validação e os $k-1$ subconjuntos restantes empregados no treinamento. A validação cruzada é repetida k vezes, de forma que cada um dos k subconjuntos sejam utilizados uma única vez como conjunto de validação do modelo. O resultado final é a média dos valores de validação, segundo a métrica adotada na avaliação, obtidos em cada iteração [Scheiber et al., 2017, Fontana, 2020].

Figura 5 – Validação *k-fold* com 10 partições



Fonte: Adaptado de Silva et al. [2010].

Capítulo 3

Materiais e Métodos

Neste capítulo serão descritos os materiais e métodos utilizados no ajuste de uma rede neural artificial para classificação da situação de alunos, permanência ou evasão, frente a seus dados socioeconômicos. Assim, o presente trabalho pode ser organizado em duas etapas: formação da base de dados e ajuste e avaliação da rede neural artificial.

3.1 Base de dados

A base de dados, disponível em <https://github.com/LeidianeTeixeira/TCC.git>, foi formada a partir da tabulação de arquivos físicos de questionários socioeconômicos (ANEXO A) em planilha eletrônica. A coleta foi realizada na secretaria de registros acadêmicos do IFNMG - Campus Salinas, mediante autorização da direção de ensino, desde que fosse preservado o anonimato dos discentes.

Foram coletadas 256 amostras de dados relativas às turmas de 2017 e 2018 dos cursos de licenciatura em Ciências Biológicas, Física, Matemática e Química. Dessas 256 amostras, 135 eram de discentes evadidos e 121 de discentes regularmente matriculados (não evadidos).

Cada amostra é caracterizada por 22 atributos categóricos (respostas do questionário e situação do discente), sendo estes:

- Atividade Remunerada: questiona se o discente exerce alguma atividade remunerada ("não", "sim, em tempo parcial", "sim, em tempo integral" e "sim, mas se trata de um trabalho eventual").
- Renda Mensal: renda mensal do discente em salários mínimos ("Não tenho nenhuma renda mensal", "menos de 0,5 salário mínimo", "0,5 até 1 salário mínimo", "1 até 1,5 salário mínimo", "1,5 até 2,5 salários mínimos", "2,5 até 3 salários mínimos" e "mais do que 3 salários mínimos").

- Número de membros da família: número de membros da família que residem com o discente ("uma a três pessoas", "quatro a seis pessoas", "sete a nove pessoas", "mais de dez pessoas" e "não tenho família").
- Renda mensal familiar: renda mensal per capita da família do discente ("minha família não tem renda mensal", "menos de 0,5 salário mínimo", "de 0,5 até 1 salário mínimo", "de 1 até 1,5 salário mínimo", "de 1,5 até 2,5 salários mínimos", "de 2,5 até 3 salários mínimos" e "mais do que 3 salários mínimos").
- Situação do pai no trabalho: situação do pai em relação ao trabalho ("trabalha regularmente", "está desempregado", "vive de renda", "é falecido e não deixou pensão" e "outra situação").
- Situação da mãe no trabalho: situação da mãe em relação ao trabalho ("trabalha regularmente", "está desempregada", "vive de renda", "é falecida e não deixou pensão" e "outra situação").
- Como concluiu o ensino fundamental: tipo de escola em que o discente concluiu o ensino fundamental ("pública municipal", "pública estadual", "pública federal", "particular", "parte em escola pública e parte em escola particular" e "supletivo ou telecurso").
- Como concluiu o ensino médio: tipo de escola em que o discente concluiu o ensino médio ("pública municipal", "pública estadual", "pública federal", "particular", "parte em escola pública e parte em escola particular", "supletivo ou telecurso" e "apenas conclui o ensino fundamental").
- Tempo de conclusão do ensino médio: há quanto tempo o discente concluiu o ensino médio ("menos de um ano", "entre um e dois anos", "entre dois e três anos", "entre três e quatro anos", "entre quatro e cinco anos", "há mais de cinco anos", "apenas conclui o ensino fundamental" e "ainda estou cursando o ensino médio").
- Cor/Raça: cor ou raça do discente ("branco (a)", "negro(a)", "pardo(a) ou mulato(a)", "indígena" e "amarelo(a) de origem asiática").
- Hábito de ler jornais ou revistas: frequência com a qual o discente lê jornais ou revistas ("diariamente", "semanalmente", "ocasionalmente" e "não").
- Livros lidos por ano: quantidade de livros não escolares lidos no ano ("nenhum", "01 e 02 livros", "03 e 05 livros" e "mais de 05 livros").
- Atividade em que ocupa mais tempo: atividade com a qual ocupa mais tempo ("televisão", "teatro", "cinema", "música", "dança", "artesanato", "leitura", "internet" e "nenhuma dessas atividades").

- Meio de se manter informado: forma pela qual o discente se mantém informado sobre os acontecimentos atuais ("jornal escrito", "jornal falado (TV)", "jornal falado (Rádio)", "revistas", "através de pessoas", "internet" e "nenhum desses").
- Faixa etária: faixa de idade do discente ("até 14 anos", "de 15 a 17 anos", "de 18 a 19 anos", "de 20 a 24 anos", "de 25 a 29 anos", "de 30 a 39 anos", "de 40 a 49 anos" e "a partir de 50 anos").
- Procedência : local no qual o discente reside ("zona rural", "urbana" e "comunidade remanescente quilombola").
- Repetiu ensino fundamental: quantidade de vezes que o discente repetiu durante o ensino fundamental ("01 vez", "02 vezes", "03 vezes", "mais de 03 vezes" e "nenhuma vez").
- Repetiu ensino médio: quantidade de vezes que o discente repetiu durante o ensino médio ("01 vez", "02 vezes", "03 vezes", "mais de 03 vezes", "apenas concluí o ensino fundamental" e "nenhuma vez").
- Grau de escolaridade do pai: nível de escolaridade do pai do discente ("analfabeto", "ensino fundamental incompleto", "ensino fundamental completo", "ensino médio incompleto", "ensino médio completo", "superior incompleto", "superior completo" e "pós-graduado").
- Grau de escolaridade da mãe: nível de escolaridade da mãe do discente ("analfabeta", "ensino fundamental incompleto", "ensino fundamental completo", "ensino médio incompleto", "ensino médio completo", "superior incompleto", "superior completo" e "pós-graduada").
- Deficiência: tipo de deficiência apresentada pelo discente ("visual", "auditiva", "motora/física", "intelectual" e "não apresento nenhum tipo de deficiência").
- evadido: situação do discente perante a instituição de ensino ("sim" e "não").

3.2 Ajuste e Avaliação da Rede Neural Artificial

Para a solução do problema proposto optou-se pela utilização de uma arquitetura de rede multicamadas, a *Multilayer Perceptron*. Esta foi escolhida por ser uma técnica de fácil modelagem e alto poder computacional muito utilizada no reconhecimento de padrões [Soares and Barbosa, 2018, Melo, 2019].

A implementação foi realizada em um notebook do *Google Colaboratory*, o qual permite a fácil escrita e execução de códigos em *Python*. Esse notebook, disponível no

mesmo repositório da base de dados, fez uso do pacote de aprendizagem de máquina *Scikit-learn* no ajuste e avaliação dos modelos.

3.2.1 Pré-Processamento

A *Multilayer Perceptron* não consegue processar diretamente dados categóricos, dessa forma, no pré-processamento os dados foram convertidos para numéricos [Melo, 2019]. Para isso, utilizou-se a codificação 1-de-n, na qual é criado um atributo para cada valor categórico. O atributo que corresponde ao valor é preenchido com 1 e os demais com 0 [Silva, 2018]. Um exemplo da codificação 1-de-n é mostrado na tabela 1.

Tabela 1 – Codificação 1-de-n para um atributo com três valores.

Valor categorizado (Procedência)	Zona Rural	Zona Urbana	Comunidade Quilombola
Zona Rural	1	0	0
Zona Urbana	0	1	0
Comunidade Quilombola	0	0	1

Fonte: Adaptado de Silva [2018].

Após o pré-processamento, os 22 atributos originais (21 entradas e 1 saída) foram transformados em 119 atributos (117 entradas e 2 saídas).

3.2.2 Topologias Avaliadas

A etapa mais importante no projeto de uma rede *Multilayer Perceptron* é a seleção da topologia, a qual está organizada em: camada de entrada, camadas ocultas e camada de saída. O número de entradas e o número de neurônios na camada de saída dependem diretamente do problema ao qual é aplicada, assim, neste trabalho são, respectivamente, 117 e 2. Entretanto, a escolha do número de camadas ocultas e de neurônios em cada uma delas é uma questão de projeto e impacta na capacidade de aprendizado da rede. Uma única camada oculta é suficiente para resolver a maioria dos problemas de classificação de padrões [Silva et al., 2010], sendo, assim, empregada em todas as topologias avaliadas. Quanto ao número de neurônios a ser empregado nessa única camada oculta, um bom ponto de partida é a média aritmética entre entradas e saídas da rede [Silva, 2019], logo, partiu-se de 59 neurônios (média entre 117 entradas e 2 saídas) para o presente trabalho.

A partir do número inicial de neurônios na camada oculta da rede, foram realizados testes preliminares para decidir entre aumentar ou diminuir os neurônios nas demais topologias avaliadas. Nesses testes, ao treinar a rede com 59 neurônios na camada oculta, observou-se uma acurácia próxima de um (100% de acerto na classificação) para o conjunto de treinamento, a qual era pouco impactada pela redução do número de neurônios. Dessa forma, uma vez que menos neurônios implicam em modelos menos complexos, optou-se

por reduzir o número de neurônios na camada oculta. Além disso, segundo [Furtado \[2019\]](#), é uma boa prática reduzir o número de neurônios se, com base no número utilizado como ponto de partida, a rede foi capaz de aprender. Assim, as topologias avaliadas são apresentadas pela Tabela 2, as quais foram definidas ao ir dividindo por dois o número de neurônios, partindo de 59 e indo até um único neurônio na camada oculta.

Tabela 2 – Topologias de rede selecionadas para avaliação.

Número de Entradas	Número de neurônios na camada oculta	Número de Saídas
117	59	2
117	29	2
117	14	2
117	7	2
117	3	2
117	1	2

3.2.3 Validação Cruzada e Treinamento

Inicialmente, a base de dados foi dividida em dois conjuntos: treinamento e teste. A divisão foi realizada aleatoriamente, sendo 75 %, 192 amostras, reservado ao conjunto de treinamento e os 25 % restantes, 64 amostras, ao conjunto de teste. A divisão foi baseada na recomendação de empregar entre 60% e 90% dos dados disponíveis para o treinamento [[Silva et al., 2010](#)], sendo o restante destinado ao teste. O conjunto de treinamento foi usado na seleção e ajuste do modelo, enquanto o conjunto de teste serviu para verificar a capacidade de generalização do modelo.

Na seleção do modelo, o conjunto de treinamento foi empregado a um processo conhecido como validação cruzada, o qual serviu para ajustar diferentes modelos e comparar suas respectivas capacidades de generalização. Durante esse processo, foram avaliados 60 modelos (Apêndice A), obtidos ao gerar 5 instâncias para cada combinação possível formada a partir de duas funções de ativação, Sigmóide e ReLU, e 6 topologias (Tabela 2). Cada instância apresenta pesos iniciais diferentes, os quais influenciam no treinamento, produzindo modelos distintos. As funções de ativação Sigmóide e ReLU foram selecionadas por serem as mais comumente empregadas a problemas de classificação [[Godoy, 2018](#), [Araújo, 2019](#)]. Os neurônios na camada de saída, em todas as topologias, usam a função de ativação Sigmóide e, após a aplicação da mesma, arredondam a saída, produzindo 0 ou 1.

Na validação cruzada, a capacidade de generalização dos modelos foi mensurada a partir da acurácia média de validação, a qual é obtida ao aplicar o método *k-fold*. O *k-fold* foi empregado com *k* igual a dez, logo, o conjunto de treinamento foi dividido em dez partições. Assim, cada modelo foi ajustado dez vezes, sempre reservando diferentes partições para a validação. Para cada partição dessas, não empregada no ajuste do modelo,

foi calculada a acurácia. Dessa maneira, são dez valores de acurácia por modelo, obtendo a acurácia média de validação. O modelo com maior acurácia média é, então, o produto final da validação cruzada, sendo esse ajustado usando o conjunto de treinamento em sua totalidade.

O algoritmo de treinamento utilizado, tanto na validação cruzada quanto no ajuste do modelo selecionado, foi o *backpropagation*, o qual empregou o gradiente descendente estocástico na otimização dos pesos. Além disso, dentre os principais parâmetros de configuração do algoritmo, destacam-se: aprendizagem adaptativa, taxa de aprendizagem inicial de 0,1, coeficiente de momentum de 0,9, número máximo de iterações igual a 5000 épocas e a parada do treinamento caso não haja queda na perda logarítmica por 10 épocas consecutivas. Os valores de taxa de aprendizagem e coeficiente de momentum utilizados foram definidos com base nas recomendações da literatura, na qual recomenda-se a utilização de um valor entre 0 e 1 para taxa de aprendizagem e valores entre 0 e 0,9 para o coeficiente de momentum [Silva et al., 2010]

3.2.4 Métricas de Avaliação

A avaliação do desempenho da rede ajustada foi efetuada através da acurácia e da matriz de confusão, as quais são as principais e mais comumente utilizadas métricas de avaliação de modelos de classificação [Kawase, 2015, Soares and Barbosa, 2018]. A acurácia é calculada através da divisão do números de predições corretas pela quantidade total de predições [Araújo, 2019]. A matriz de confusão cruza as classes reais com as classes previstas pela rede [Kawase, 2015, Soares and Barbosa, 2018], permitindo visualizar os acertos e erros da rede, assim, complementando a acurácia.

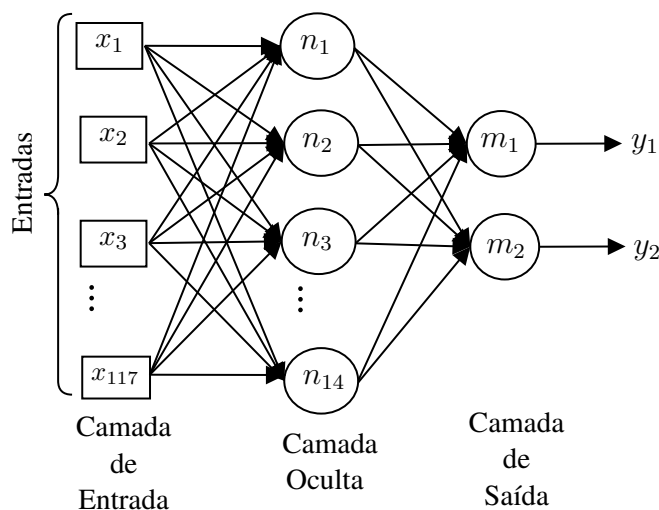
Capítulo 4

Resultados e Discussão

No presente capítulo serão apresentados e discutidos os resultados obtidos a partir da modelagem e avaliação do classificador MLP da situação discente, evasão ou permanência, perante dados socioeconômicos, para os cursos de licenciatura do IFNMG-Salinas.

Durante a validação cruzada, os valores de acurácia média de validação, calculados através do k-fold para os 60 modelos, variaram entre $42,42\% \pm 17,86$ e $61,95\% \pm 8,24$. Assim, o modelo com $61,95\%$ de acurácia média de validação foi selecionado, o qual é composto por 117 entradas, 14 neurônios na camada oculta e 2 neurônios de saída (Figura 6), sendo empregada a função de ativação Relu nos neurônios da camada oculta.

Figura 6 – Arquitetura do modelo selecionado utilizando validação *k-fold*



Fonte: Adaptado de Araújo [2019].

O treinamento do modelo selecionado, finalizado com melhor perda logarítmica de 0,06, tem sua evolução ilustrada pela Figura 7. Após o término do treinamento, essa rede MLP foi capaz de classificar corretamente as 192 amostras do conjunto de treinamento, logo, 100% de acurácia.

Figura 7 – Histórico de perda logarítmica do treinamento do modelo selecionado



Fonte: A autora [2021]

A Tabela 3, matriz de confusão, torna mais claro os efeitos da aprendizagem da rede perante os dados empregados no treinamento. O modelo classificou corretamente 89 amostras da classe ‘sim’ (evadiu) e 103 amostras da classe ‘não’ (não evadiu).

Tabela 3 – Matriz de confusão para a rede MLP segundo o conjunto de treinamento

	Valor Predito			Taxa (%)
	Evadiu	Sim	Não	
Valor Real	Sim	89	0	100%
	Não	0	103	100%
Taxa Média (%)				100%

Fonte: A autora [2021]

Ao apresentar o conjunto de teste ao modelo, observou-se acurácia de 67,19%, classificando corretamente 43 dentre 64 amostras. A Tabela 4 detalha os acertos e erros de classificação da rede, bem como a acurácia, para ambas as classes. A identificação de evadidos (classe ‘sim’) obteve acurácia de aproximadamente 72% (18 em 25 amostras), enquanto não evadidos (classe ‘não’) obteve acurácia próxima de 64% (25 em 39 amostras).

O modelo desenvolvido apresentou resultados promissores, sinalizando que cerca de 72% dos alunos com tendência à evasão poderiam ser identificados ao ingressar nos cursos. Isso seria de grande valia para as licenciaturas do IFNMG-Salinas, uma vez que apresentam um alto índice de evasão, cerca de 47% perante os dados coletados. Entretanto, a rede seria um instrumento de apoio, não único, uma vez que nem todos os discentes evadidos foram identificados (cerca de 28% do conjunto de teste).

Tabela 4 – Matriz de confusão para a rede MLP segundo o conjunto de teste

	Valor Predito			Taxa (%)
	Evadiu	Sim	Não	
Valor Real	Sim	18	7	71,2%
	Não	14	25	63,2%
Taxa Média (%)				67,2%

Fonte: A autora [2021]

Na literatura são encontrados alguns trabalhos semelhantes ao desenvolvido, contudo uma comparação com os mesmos deve ser realizada com cautela, uma vez que envolvem atributos, cursos e instituições distintas. Kawase [2015], por exemplo, alcançou resultados superiores ao aplicar redes RBF e MLP, obtendo, respectivamente, 91% e 84% de acurácia, na classificação da situação discente no curso de Sistemas de informação da UFRRJ. Melo [2019], por outro lado, obteve resultados próximos ao deste trabalho, no qual a rede MLP ajustada, com base em dados extraídos do sistema acadêmico da Universidade Federal do Triângulo Mineiro, registrou uma acurácia de 70,5% na classificação entre evadidos e não evadidos. Ambos os autores, em contraste com o presente trabalho, consideraram atributos de desempenho acadêmico, os quais podem ter tido grande influência na capacidade preditiva dos classificadores, embora retardem a identificação de discentes predispostos à evasão. Além disso, Kawase [2015] ajustou seus modelos para um único curso, o que, provavelmente, também repercutiu nos resultados alcançados.

Capítulo 5

Conclusão

O classificador da situação discente desenvolvido neste trabalho, utilizando Redes Neurais Artificiais *Multilayer Perceptron*, é um instrumento promissor no combate à evasão nas licenciaturas do IFNMG-Salinas. Esse classificador, como parte de dados socioeconômicos coletados no ingresso dos discentes, auxiliaria na identificação prematura de alunos com possível tendência a evadir, permitindo uma intervenção mais cedo por parte da instituição.

Apesar do potencial como um instrumento de suporte pedagógico, o classificador proposto apresenta limitações e deve ser empregado em conjunto com outras medidas de combate à evasão. Essas limitações podem ser reduzidas ao aumentar a acurácia da classificação, o que pode ser alcançado ao explorar novas técnicas e ao ampliar o número de atributos considerados. A usabilidade do classificador, também, precisa ser melhorada, integrando-o ao sistema acadêmico. Assim, sugere-se como trabalhos futuros:

- Avaliar e comparar diferentes técnicas de aprendizagem de máquina, também empregadas para classificação, na predição da situação discente;
- Buscar maior acurácia ao ampliar a quantidade de atributos empregados na construção do classificador. Assim, mantendo o foco na identificação da tendência à evasão no ingresso, os candidatos seriam atributos relacionados ao desempenho escolar no ensino médio, bem como aqueles associados ao perfil psicológico do aluno;
- Estender a abrangência do modelo a outros cursos superiores do IFNMG-Salinas. O curso, provavelmente, se tornaria um novo atributo do modelo;
- Integrar o modelo ao sistema acadêmico, englobando: ampliar o registro acadêmico com os atributos utilizados e prover o acesso remoto pela direção de ensino, setor pedagógico e secretarias.

Referências

- A. M. Araújo. Classificação e detecção de pessoas em ambientes não controlados utilizando redes neurais convolucionais. Monografia (bacharel em ciência da computação), Universidade Federal da Paraíba, João Pessoa, 2019.
- B. Bocchini. Pesquisa mostra evasão de 30% em cursos superiores privados: na rede pública, percentual dos que deixaram os estudos chega a 18,5%. <https://bit.ly/3bI0F5s>, 2018. Acesso em 24 mar. 2020.
- Brasil. Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas. relatório da comissão especial de estudos sobre evasão nas universidades públicas brasileiras. <https://bit.ly/3exVhDF>, 1996. Acesso em 10 abr. 2020.
- Brasil. Censo da educação superior 2018 : notas estatísticas. Technical report, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), 2019.
- V. S. Chaves. Evasão nos cursos de graduação em física, matemática e química da ufrn. Dissertação (mestrado), Universidade Federal do Rio Grande do Norte, Natal, 2016.
- J. M. Feitosa. Análise de evasão no ensino superior: uma proposta de diagnóstico para o campus de laranjeiras. Dissertação (mestrado), Universidade Federal do Rio Grande do Norte, São Cristóvão, 2016.
- A. K. M. S. Ferreira. Evasão no ensino superior: Uma análise no curso de agronomia da universidade federal do cariri. Dissertação (mestrado), Universidade Federal da Paraíba, João Pessoa, 2019.
- S. Fiúza. O impacto social da educação através das instituições de ensino superior. <http://bit.ly/3rIpBiL>, 2019. Acesso em 19 ago. 2020.
- E. Fontana. Introdução aos algoritmos de aprendizagem supervisionada. <https://bit.ly/3qIHe2f>, 2020. Acesso em 17 abr. 2021.
- M. I. V. Furtado. *Redes Neurais Artificiais: Uma Abordagem Para Sala de Aula*. Atena Editora, Ponta Grossa, 2019. ISBN 9788572473262. URL <https://bit.ly/30zGdgJ>.

- A. C. Fábio. Qual o impacto das universidades no crescimento econômico. <http://bit.ly/3eB2iDE>, 2016. Acesso em 19 ago. 2020.
- L. M. L. S. Garcia, D. F. Lara, F. Antunes, C. C. de Miranda, and C. A. Peres. Análise da evasão no ensino superior e suas motivações: Um estudo de caso em um curso de sistemas de informação. In *Anais do XIII Simpósio Brasileiro de Sistemas de Informação*, pages 527–534, Lavras, 2017.
- T. Gargantini. Evasão nas universidades: a frustração de trilhões de dólares. <http://bit.ly/3eyKRDY>, 2019. Acesso em 10 nov. 2020.
- D. Godoy. Hiper-parâmetros em ação! parte i — funções de ativação. <https://bit.ly/3jGDr30>, 2018. Acesso em 14 abr. 2021.
- J. R. Gorski. Redes neurais. <http://bit.ly/3lc5iHK>, 2019. Acesso em 12 ago. 2020.
- S. Haykin. *Redes Neurais Princípios e Práticas*. Bookman, Porto Alegre, 2 edition, 2001. ISBN 978-85-7307-718-6. Tradução de: Paulo Martins Engel.
- K. H. F. Kawase. Aplicação de redes neurais rbf e mlp na análise de evasão discente do curso de sistemas de informação da ufrj. Dissertação (mestrado), Universidade Federal Rural do Rio de Janeiro, Seropédica, 2015.
- R. Lobo. O acolhimento do estudante e a evasão. <https://bit.ly/3rHEulk>, 2019. Acesso em 11 ago. 2020.
- B. V. Martins and S. R. Oliveira. Expansão e diversificação do ensino superior, impactos no mercado de trabalho e inserção profissional no brasil: reflexões iniciais e proposta de agenda de pesquisa. *Revista de Gestão do Unilasalle*, 6(2):53–70, 7 2017.
- A. L. Melo. Uso da técnica de mineração de dados como uma ferramenta de gestão da evasão no ensino superior. Dissertação (mestrado), Universidade Federal do Triângulo Mineiro, Uberaba, 2019.
- D. N. Monteiro. Perfil discente e razões de evasão no ensino superior: o caso da graduação em saúde coletiva da faculdade de ceilândia da universidade de brásilia no período de 2013 a 2017. Trabalho de conclusão de curso (bacharelado em saúde coletiva), Universidade de Brasília, Brasília, 2018.
- N. P. Nagai and A. L. J. Cardoso. A evasão universitária: uma análise além dos números. *Revista Estudo e Debate*, 24(1):193–215, 3 2017.
- M. Negnevitsky. *Artificial Intelligence A Guide to Intelligent Systems*. Pearson Education, Harlow, 2 edition, 2005. ISBN 0-321-20466-2.

- A. C. A. Neto, M. A. da Silva Modesto, C. R. B. Souza, R. S. Argolo, M. T. T. Andrade, J. Araújo, J. Rios, and T. K. G. Carneiro. Uma proposta de análise da evasão no ensino superior do ifba. *Revista Ensino em Foco*, 2 2018.
- OECD. Education at a glance: Oecd indicators. <http://bit.ly/3viDHJ0>, 2018. Acesso em 16 mar. 2020.
- C. S. Rocha. Por que eles abandonam? evasão e bolsistas prouni dos cursos de licenciaturas. Tese (doutorado), Universidade Federal do Rio Grande do Norte, São Leopoldo, 2015.
- E. F. B. Santiago. Evasão no ensino superior: estudo na universidade estadual de montes claros. Dissertação(mestrado), Faculdade Novos Horizontes, Belo Horizonte, 2015.
- J. N. C. Scheiber, A. L. Beskow, J. C. T. Muller, E. O. B. Nara, J. I. da Silva, and J. W. Reuter. Técnicas de validação de dados para sistemas inteligentes: uma abordagem do software sdbayes. In *XVII Colóquio Internacional de Gestão Universitária*, Mar del Plata - Argentina, 2017.
- C. Q. Silva. Redes neurais aplicadas no reconhecimento de símbolos matemáticos manuscritos online. Tcc (bacharel em matemática aplicada), Universidade de São Paulo, São Paulo, 2019.
- D. A. Silva. Aplicação de técnicas de pré-processamento e agrupamento na base de dados de benefícios previdenciários do ministério público do trabalho. Trabalho de conclusão de curso (bacharel em sistemas de informação), Universidade Federal de Uberlândia, Uberlândia, 2018.
- I. N. Silva, D. H. Spatti, and R. A. Flauzino. *Redes Neurais Artificiais para engenharia e ciências aplicadas*. Artliber Editora, São Paulo, 2010. ISBN 8588098539.
- G. F. Soares and J. P. N. F. Barbosa. Comparação entre as redes neurais artificiais mlp, rbf e artmap-fuzzy na predição da evasão escolar. Trabalho de conclusão de curso (graduação), Instituto Federal de Educação, Ciência e Tecnologia do Pará, Belém, 2018.
- J. D. R. Soares. Modelagem matemática na previsão de colheita de bananeira: regressão linear múltipla x redes neurais artificiais. Tese (doutorado), Universidade Federal de Lavras, Lavras, 2010.
- H. C. Sousa. Um framework para a criação e simulação de redes neurais artificiais utilizando component object model. Dissertação (mestrado), Universidade de São Paulo, São Carlos, 2000.
- L. F. D. Souza. Evasão do curso de licenciatura em matemática (noturno) da universidade de Brasília. Trabalho de conclusão de curso (licenciatura em ciência da computação), Universidade de Brasília, Brasília, 2016.

Anexo A

Questionário Socioeconômico

1 - Você exerce alguma atividade remunerada?

- ☐ Não.
- ☐ Sim, em tempo parcial (até vinte horas semanais).
- ☐ Sim, em tempo integral (mais de trinta horas semanais).
- ☐ Sim, mas se trata de um trabalho eventual.

2- Qual é a sua renda mensal ?

- ☐ Não tenho nenhuma renda mensal.
- ☐ Menos de 0,5 salário mínimo.
- ☐ De 0,5 até 1 salário mínimo.
- ☐ De 1 até 1,5 salário mínimo.
- ☐ De 1,5 até 2,5 salários mínimos.
- ☐ De 2,5 até 3 salários mínimos.
- ☐ Mais do que 3 salários mínimos.

3 - Qual é o número de membros da sua família?

- ☐ De uma a três pessoas.
- ☐ De quatro a seis pessoas.
- ☐ De sete a nove pessoas.
- ☐ Mais de dez pessoas.
- ☐ Não tenho família.

4 - Renda mensal da família

- ☐ Minha família não tem renda mensal.
- ☐ Menos de 0,5 salário mínimo.
- ☐ De 0,5 até 1 salário mínimo.
- ☐ De 1 até 1,5 salário mínimo.
- ☐ De 1,5 até 2,5 salários mínimos.

- ☐ De 2,5 até 3 salários mínimos.
- ☐ Mais do que 3 salários mínimos.

5 - Qual das seguintes alternativas melhor expressa a atual situação de seu pai no trabalho?

- ☐ Trabalha regularmente.
- ☐ Está desempregado.
- ☐ Vive de renda.
- ☐ É falecido e não deixou pensão.
- ☐ Outra situação.

6 - Qual das seguintes alternativas melhor expressa a atual situação de sua mãe no trabalho?

- ☐ Trabalha regularmente.
- ☐ Está desempregada.
- ☐ Vive de renda.
- ☐ É falecida e não deixou pensão.
- ☐ Outra situação.

7 - Como você fez seus estudos de Ensino Fundamental ou equivalente?

- ☐ Escola pública municipal.
- ☐ Escola pública estadual.
- ☐ Escola pública federal.
- ☐ Escola particular.
- ☐ Parte em escola pública e parte em escola particular.
- ☐ Supletivo ou telecurso.

8 - Como você fez seus estudos de Ensino Médio ou equivalente?

- ☐ Escola pública municipal.
- ☐ Escola pública estadual.
- ☐ Escola pública federal.
- ☐ Escola particular.
- ☐ Parte em escola pública e parte em escola particular.
- ☐ Supletivo ou telecurso.
- ☐ Apenas conclui o ensino fundamental.

9 - Há quanto tempo você concluiu o Ensino Médio ?

- ☐ Menos de um ano.
- ☐ Entre um e dois anos (incompletos).
- ☐ Entre dois e três anos (incompletos).

- ☐ Entre três e quatro anos (incompletos).
- ☐ Entre quatro e cinco anos (incompletos).
- ☐ Há mais de cinco anos.
- ☐ Apenas conclui o ensino fundamental.
- ☐ Ainda estou cursando o Ensino Médio.

10 - Você se considera:

- ☐ Branco(a).
- ☐ Negro(a).
- ☐ Indígena.
- ☐ Pardo(a), Mulato(a).
- ☐ Amarelo(a) de origem asiática.

11 - Você tem o hábito de ler jornais ou revistas?

- ☐ Diariamente.
- ☐ Semanalmente.
- ☐ Ocasionalmente.
- ☐ Não.

12 - Excetuando os livros escolares, quantos livros você lê por ano?

- ☐ Nenhum.
- ☐ 01 e 02 livros.
- ☐ 03 e 05 livros.
- ☐ Mais de 05 livros.

13 - Com quais das atividades citadas abaixo você ocupa mais tempo?

- ☐ Televisão.
- ☐ Teatro.
- ☐ Cinema.
- ☐ Música.
- ☐ Dança.
- ☐ Artesanato.
- ☐ Leitura.
- ☐ Internet.
- ☐ Nenhuma dessas atividades.

14 - Qual é o meio que você mais utiliza para se manter informado sobre os acontecimentos atuais?

- ☐ Jornal escrito.

- ☐ () Jornal falado (TV).
- ☐ () Jornal falado (Rádio).
- ☐ () Revistas.
- ☐ () Através de pessoas.
- ☐ () Internet.
- ☐ () Nenhum desses.

15 - Indique sua Faixa Etária:

- ☐ () Até 14 anos.
- ☐ () De 15 a 17 anos.
- ☐ () De 18 a 19 anos.
- ☐ () De 20 a 24 anos.
- ☐ () De 25 a 29 anos.
- ☐ () De 30 a 39 anos.
- ☐ () De 40 a 49 anos.
- ☐ () A partir de 50 anos.

16 - Qual sua procedência?

- ☐ () Zona rural.
- ☐ () Zona urbana.

17 - Se você repetiu alguma série no Ensino Fundamental, informe o número de vezes:

- ☐ () 01 vez.
- ☐ () 02 vezes.
- ☐ () 03 vezes.
- ☐ () Mais de 03 vezes.
- ☐ () Nenhuma vez.

18 - Se você repetiu alguma série no Ensino Médio, informe o número de vezes:

- ☐ () 01 vez.
- ☐ () 02 vezes.
- ☐ () 03 vezes.
- ☐ () Mais de 03 vezes.
- ☐ () Apenas concluí o ensino fundamental.
- ☐ () Nenhuma vez.

19 - O grau de escolaridade do seu pai é:

- ☐ () Analfabeto.
- ☐ () Ensino fundamental incompleto.

- ☐ Ensino fundamental completo.
- ☐ Ensino médio incompleto.
- ☐ Ensino médio completo.
- ☐ Superior incompleto.
- ☐ Superior completo.
- ☐ Pós-Graduado.

20 - O grau de escolaridade da sua mãe é:

- ☐ Analfabeta.
- ☐ Ensino fundamental incompleto.
- ☐ Ensino fundamental completo.
- ☐ Ensino médio incompleto.
- ☐ Ensino médio completo.
- ☐ Superior incompleto.
- ☐ Superior completo.
- ☐ Pós-Graduada.

21 - Você apresenta algum tipo de deficiência?

- ☐ Visual.
- ☐ Auditiva.
- ☐ Motora/Física.
- ☐ Intelectual.
- ☐ Não apresento nenhum tipo de deficiência.

Apêndice A

Modelos Avaliados

Tabela 5 – Modelos Avaliados

Nº de neurônios	Função de ativação	Instâncias	Acurácia média de validação (%) \pm Desvio Padrão
59	Logistic	1	58,31 \pm 7,07
		2	58,28 \pm 5,98
		3	58,34 \pm 6,52
		4	57,76 \pm 8,80
		5	60,42 \pm 9,11
	Relu	1	56,21 \pm 8,33
		2	58,31 \pm 5,82
		3	55,21 \pm 8,48
		4	55,18 \pm 6,82
		5	56,78 \pm 8,67
29	Logistic	1	58,86 \pm 5,29
		2	56,76 \pm 8,80
		3	55,71 \pm 5,45
		4	57,81 \pm 8,56
		5	55,71 \pm 6,39
	Relu	1	56,23 \pm 4,23
		2	55,68 \pm 8,41
		3	57,78 \pm 8,03
		4	57,81 \pm 8,56
		5	55,13 \pm 9,15
	Logistic	1	57,31 \pm 5,88
		2	57,73 \pm 8,26
		3	57,81 \pm 7,52
		4	54,65 \pm 6,37

	Relu	5	$57,73 \pm 9,18$
		1	$56,18 \pm 6,64$
		2	$58,31 \pm 7,07$
		3	$61,94 \pm 8,24$
		4	$59,31 \pm 9,43$
		5	$55,15 \pm 6,53$
7	Logistic	1	$57,21 \pm 5,63$
		2	$58,81 \pm 6,59$
		3	$58,28 \pm 5,50$
		4	$57,23 \pm 5,96$
		5	$58,78 \pm 6,69$
	Relu	1	$57,15 \pm 8,51$
		2	$56,18 \pm 6,21$
		3	$58,31 \pm 6,95$
		4	$55,57 \pm 14,64$
		5	$58,73 \pm 9,58$
3	Logistic	1	$57,23 \pm 6,40$
		2	$42,42 \pm 17,85$
		3	$57,23 \pm 7,95$
		4	$56,23 \pm 6,75$
		5	$42,81 \pm 17,42$
	Relu	1	$53,60 \pm 8,39$
		2	$53,13 \pm 9,14$
		3	$55,13 \pm 9,09$
		4	$53,63 \pm 11,87$
		5	$54,68 \pm 9,86$
1	Logistic	1	$53,65 \pm 9,25$
		2	$53,65 \pm 9,25$
		3	$53,65 \pm 9,25$
		4	$54,65 \pm 7,57$
		5	$56,26 \pm 8,40$
	Relu	1	$50,42 \pm 10,03$
		2	$60,34 \pm 9,92$
		3	$53,65 \pm 9,25$
		4	$52,52 \pm 10,00$
		5	$59,28 \pm 11,21$