

FUNDAMENTOS E APLICAÇÃO DE REDES NEURAIS ARTIFICIAIS À SAÚDE E HUMANIDADES: TUTORIAL – PARTE 1

BRASIL, Roxana Macedo¹; BRITO, Diogo de Freitas^{2,3}; BARRETO, Ana Cristina Lopes y Glória²; JUNIOR, Homero da Silva Nahum^{2,4}

405

Resumo

O objetivo foi prever o comportamento de universitários quanto à matrícula. O modelo utilizou um banco de dados com 326 registros, mas realizou a predição sobre outro conjunto de dados com 64 pessoas. A codificação foi realizada em R 4.4.3. As redes neurais desenvolvidas foram *feedforward*, RNN e LSTM, com aplicação do *Backpropagation* e função de ativação *Softmax*. Em todas as métricas, a primeira rede apresentou os piores desempenhos para o banco de dados de desenvolvimento. Entretanto, no conjunto de dados para predição, os resultados foram equivalentes, porém com distintos padrões identificados de influência das características sobre a matrícula. A conclusão foi pelo alcance do objetivo, dado que foi possível prever os resultados da variável de saída.

Palavras-chave: Heurística. Processo estocástico. Administração. Economia. Negócio.

Abstract

The aim was to predict the enrolment behaviour of university students. The model used a database with 326 records, but performed the prediction on another dataset with 64 people. Coding was carried out in R 4.4.3. The neural networks developed were feedforward, RNN and LSTM, with the application of Backpropagation and the Softmax activation function. In all the metrics, the first network performed worst for the development database. However, in the prediction dataset, the results were equivalent, but with different patterns identified for the influence of characteristics on enrolment. The conclusion was that the objective had been achieved, since it was possible to predict the results of the output variable.

Keywords: Heuristics. Stochastic process. Management. Economics. Business.

Introdução

Detectar arritmias cardíacas (Maggioni e Rebonatto, 2024), prever a incidência de câncer pela dicotomia sexual (Souza *et al.*, 2025), identificar o estadiamento de câncer de pele (Esteva *et al.*, 2017) e identificar doença cutâneas (Alencar, Silva e Mendes, 2021), assim como controlar internamente Organizações financeiras (Gonçalves e Reinaldi, 2024), analisar o perfil do Microempreendedor Individual (Curty *et al.*, 2024) e identificar o

¹ Docente Ph.D. em Educação Física;

² Docentes do Curso de Educação Física do Centro Universitário Celso Lisboa;

³ Docente do Curso de Gestão Desportiva e do Lazer do Centro Universitário Celso Lisboa;

⁴ Docente da Escola de Saúde da Universidade Cândido Mendes.

sentimento de clientes externos de Organização varejista (Carneiro *et al.*, 2025) seriam algumas das possibilidades de aplicação de Redes Neurais Artificiais (RNA).

As RNAs seriam modelos matemáticos que emitiriam o processamento de informações realizado pelo sistema nervoso central, portanto apresentariam nós interligados que realizariam aquele processo e organizados estariam nas seguintes camadas: 1) entrada, por onde a rede seria alimentada de dados, logo cada nó representaria uma variável ou característica; 2) ocultas, posicionamento intermediário, local em que ocorreriam as transformações de toda ordem, cálculos, identificações de padrões e previsões, então a quantidade dessas camadas deteria relação direta com a complexidade da rede; 3) saída, fornecedora do resultado das camadas anteriores (Haykin, 2000). A dedução imediata seria que as conexões entre os nós (relações entre as variáveis) seriam determinantes aos resultados. Por causa disso, as RNAs seriam treinadas, permitindo o ajuste de pesos entre aquelas conexões (pesos sinápticos), visando atenuar erros e potencializar a precisão de previsões das funções de ativação (I) (Braga, Ludermir e Carvalho, 2007).

$$y = f\left(\sum_{i=1}^n w_i x_i + b\right) \quad (I)$$

Onde: x_i = entrada i -ésima; w_i = peso associado à entrada i -ésima; b = bias (ou viés); f = função de ativação.

Valeria destacar que o bias seria um valor de ajuste adicionado em cada nó, visando melhorar a previsão pelo deslocamento da função de ativação, dotando-a de flexibilidade, dessa forma a dependência da camada de entrada seria reduzida e a possibilidade de identificação de padrões complexos no conjunto de dados se elevaria (Dreyfus, 2005). Objetivamente, a função de ativação poderia fornecer saída significativa mesmo na existência de entrada nula. O valor do bias não seria aleatório, mas estimado durante o treinamento, paralelamente aos pesos sinápticos, segundo determinada função de perda codificada no algoritmo (Rojas, 1993). Com relação à função de ativação, diversas seriam as possibilidades (Quadro I), a seleção dependeria do problema (por exemplo: classificação ou regressão), da velocidade e eficiência do treinamento, e arquitetura da RNA. Além disso,

o custo computacional poderia influenciar a escolha da função aplicada (Quadro II), dado que essencialmente, as funções de avaliação introduziriam não linearidade.

Quadro I: Exemplificação de Possibilidades de Função de Avaliação.

| Função | Utilidade e Característica |
|-------------------------------------|--|
| Linear | Regressão, pois não haveria limite à saída. |
| Sigmoide | Classificação binária, saídas entre 0 e 1 (probabilística). Não adequada ao treinamento de redes profundas em razão do gradiente que desaparece (<i>vanishing gradient</i>). |
| Tangente Hiperbólica (Tanh) | Dados com valores negativos e positivos. saídas entre -1 e 1. Em menor intensidade que a anterior, teria o problema do gradiente que desaparece. |
| <i>Rectified Linear Unit</i> (ReLU) | Visão computacional, redes profundas. Parcialmente, resolveria o problema do gradiente que desaparece, mas poderia fornecer saídas sempre nulas (neurônios "mortos"). |
| Leaky ReLU | Melhoria da ReLU. Não haveria neurônios mortos e permitiria valores negativos. |
| Softmax | Classificação com múltiplas categorias. O somatório das saídas seria igual à unidade (100,00%). |
| Swish | Combinação de várias funções, melhorando o desempenho de redes complexas. |

Fonte: Os Autores (2025).

Quadro II: Exemplificação de Aplicação da Função de Avaliação, e = 2,718 e com w = 0,5, x = 2 e b = 1.

| Função [saída] | Matemática | Exemplo |
|-----------------------|--|---|
| Sigmoide [0,00; 1,00] | $f(x) = \frac{1}{1 + e^{-(wx+b)}}$ | $f(x) = \frac{1}{1 + e^{-(0,5 \cdot 2 + 1)}} \approx 0,88$ |
| Tanh [-1,00; 1,00] | $f(x) = \frac{e^{(wx+b)} - e^{-(wx+b)}}{e^{(wx+b)} + e^{-(wx+b)}}$ | $f(x) = \frac{e^{(0,5 \cdot 2 + 1)} - e^{-(0,5 \cdot 2 + 1)}}{e^{(0,5 \cdot 2 + 1)} + e^{-(0,5 \cdot 2 + 1)}} \approx 0,96$ |

Fonte: Os Autores (2025).

Em termos computacionais, as variáveis seriam processadas de forma distribuída (as tarefas seriam divididas pelos nós) e paralela (várias tarefas executadas simultaneamente), então na existência de *M* camadas (redes profundas) ocorreria a composição de funções (II), tendo cada camada *m* atribuição específica (III) pela aplicação da função de ativação (σ).

$$f(x) = f^{(M)} \left(f^{(M-1)} \left(f^{(M-2)} \left(\dots f^{(1)}(x) \right) \right) \right) \quad (II)$$

$$f^{(m)}(x^{(m-1)}) = \sigma^{(m)}(w^{(m)}x^{(m-1)} + b^m) \quad (III)$$

Pelo exposto, possível seria deduzir que a RNA exigiria grande volume de dados à realização do treinamento, como regra prática, comumente se utilizaria 10 vezes a quantidade de parâmetros (pesos). Em problemas de classificação, necessário seria a realização do balanceamento das classes, quando haveria desproporção entre quantidade de exemplos, induzido o modelo ao viés, então necessário seria providenciar o ajuste (Krizhevsky, Sutskever e Hinton, 2012).

Todavia, o pré-processamento, também, seria requisito, especialmente pela exigência de codificação de variáveis categóricas e normalização das quantitativas. No primeiro caso, se houver múltiplas categorias únicas (por exemplo: em variáveis como cidade e profissão), poderia ser empregado o *Target Encoding*, especialmente, substituindo as categorias pelas respectivas probabilidades condicionais (Micci-Barreca, 2001). Possível, também, seria optar por transformar n categorias em n variáveis binárias, processo conhecido como *One-hot Encoding* (Harris e Harris, 2012).

A normalização das variáveis quantitativas deveria ser realizada, porque RNAs seriam sensíveis às escalas de medição (Goodfellow, Bengio e Courville, 2018), então poderia ser aplicado o método Min-Max (IV) ou Z-score (V), tornando o valor observado (x) ajustado (x') em determinada escala comum, isso eliminaria discrepâncias em virtude de unidades e magnitudes distintas. Idealmente, utilizada seria a normalização por lotes (*Batch normalization*), que consistiria em aplicar os parâmetros aprendíveis para adicionar flexibilidade de reescalar (γ) e deslocar (β), $y = \gamma x' + \beta$, possibilitando o *aprendizado* de valores ótimos de escala e deslocamento durante o treinamento (Ioffe e Szegedy, 2015).

$$x' = \frac{x - \min(X)}{\max(X) - \min(X)} \quad (IV)$$

Onde: X = conjunto de valores de determinada variável.

$$x' = \frac{x - \mu}{\sigma} \quad (V)$$

Onde: μ = a média populacional da variável X ; σ : o desvio padrão populacional da variável X .

Em outras situações, poderia haver relações complexas entre variáveis categóricas (por exemplo: na variável disciplina, Matemática e Física se relacionariam; como na variável compra, Pão e Manteiga), então o modelo deveria identifica-las, transformando as categorias (Matemática e Física, por exemplo) em vetores, isso seria o processo de *Embedding* (Sousa, Santiago e Dias, 2019; Sousa, 2016). Ao treinamento de redes profundas o emprego de Unidades de Processamento Gráfico (*Graphics Processing Units* – GPU) seria necessário, dada a existência de milhares de núcleos de cálculos simultâneos (Gurgel e Formiga, 2013).

O funcionamento da RNA dependeria das seguintes características: 1) número de camadas, simples (um ou poucas) ou profunda (muitas camadas ocultas); 2) conexões entre os neurônios, diretas (rede *feedforward*) ou cíclicas (rede recorrente); 3) funções de ativação; 4) tipo de aprendizado, supervisionado, não supervisionado ou reforçado; 5) hiperparâmetros, quantidade de neurônios por camada e taxa de aprendizado, por exemplo. Tal conjunto determinaria o desenho e a estrutura da RNA, ou seja, a arquitetura dela. Em outras palavras, a maneira como neurônios e camadas foram organizados, portanto o fluxo de dados entre entrada, ocultas e saídas, obedecendo a forma de aprendizado (Goodfellow, Bengio e Courville, 2018; Bishop, 2006; Hochreiter, 1991). As arquiteturas mais comumente utilizadas seriam (Fleck *et al.*, 2016; Fiorin *et al.*, 2011; Wueges e Borba, 2010; Ferneda, 2006):

- *Feedforward Neural Network* (FFNN): a forma mais simples, portanto os dados teriam fluxo unidirecional (entrada → saída). Indicada à classificação e regressão. A simplicidade a tornaria eficiente, porém a exporia à *maldição da dimensionalidade*, quando a quantidade de variáveis fosse elevada, o espaço de busca se tornaria demasiadamente amplo, o que poderia impossibilitar encontrar a solução ótima, caso a quantidade de dados não fosse adequadamente elevada. Outras consequências seriam a não captura de padrões de complexos (dependências temporais ou sequenciais nos dados), e as demandas computacional e temporal de treino requisitariam amplo volume de recursos;
- Rede Neural Recorrente (RNN): processaria dados sequenciais (por exemplo séries temporais e linguagem natural), porque as conexões ocorreriam em ciclos, possibilitando que informações de estados anteriores influenciassem o estado atual. Contudo, como os ajustes dos pesos ocorreria pelos gradientes

das funções de erro, e esses tenderiam a ser reduzidos à medida que retornariam às camadas iniciais (*vanishing gradient* ou gradiente que desaparece) no treino, poderia ocorrer o comprometimento do aprendizado tornando-o ínfimo (camadas iniciais teriam dificuldade de ajuste nos pesos), muito lento ou ineficaz, ou dificultoso para padrões complexos em cadeias longas ou dados intensamente interdependentes;

- *Long Short-Term Memory* (LSTM): adequada à modelagem de dependências de longo prazo em dados sequenciais (por exemplo, tradução automática, análise de sentimentos e reconhecimento de fala). Capaz seria de, seletivamente, armazenar ou esquecer informações pela introdução de células de memória, então, superaria o gradiente que desaparece.

Em inteligência artificial, especialmente, nas aplicações envolvendo aprendizado de máquina, o ponto pacífico seria a divisão do conjunto de dados em grupos de treino, validação e teste. O primeiro seria utilizado para ajustar os parâmetros do modelo, ele o ensinaria a reconhecer padrões. A importância estaria em evitar o ajustamento excessivo aos dados de treinamento (*memorização*), efeito denominado *overfitting* (Asogwa *et al.*, 2023; Xu e Goodacre, 2018).

O objetivo da validação estaria no ajuste dos hiperparâmetros (por exemplo, taxa de aprendizado e número de neurônios em uma camada), permitindo avaliar o desempenho do modelo em dados distintos daqueles do treino, garantindo a generalização. Essa entendida como a capacidade do modelo de funcionar adequadamente diante de novos dados, logo sem validação a robustez das previsões tenderia a ser perdida (Lumumba *et al.*, 2024; Arlot e Celisse, 2010).

O grupo de teste seria empregado para simular o comportamento do modelo diante de dados reais, diferentes do treino, evitando a possibilidade de ajuste enviesado de hiperparâmetros, o que comprometeria a capacidade de generalização e impediria a otimização (Li *et al.*, 2022; Khan *et al.*, 2021).

Dentre os algoritmos de otimização (Haykin, 2000), o *Backpropagation* seria de aprendizado supervisionado, ajustando os pesos das conexões entre neurônios para minimizar o erro entre as saídas observada e predita, o que equivaleria a afirmar que estimaria o erro na camada de saída (VI), propagaria à oculta (VII), e atualizaria o peso (VIII). Semanticamente seria estimar o gradiente da função de erro em relação aos pesos

da rede, subsequentemente, atualizar atualizá-los na direção oposta ao gradiente (Oliveira e Melo Júnior, 2024).

$$\delta^{(M)} = h_{\theta}(x) - y \quad (VI)$$

Onde: $\delta^{(M)}$: erro da camada de saída; $h_{\theta}(x)$: saída predita; y : saída observada.

$$\delta^{(m)} = (w^{(m+1)})^T \delta^{(m+1)} \odot g'(z^{(m)}) \quad (VII)$$

Onde: $w^{(m+1)}$: peso da próxima camada; $g'(z^{(m)})$: derivada da função de ativação da camada m .

$$w^{(m)} := w^{(m)} - \alpha \Delta w^{(m)} \quad (VIII)$$

Onde: $\Delta w^{(m)}$: gradiente do erro em relação aos pesos.

A alternativa, mais comumente, empregada seria o *Gradient Descente*, cujo método de otimização seria minimizar a função de custo ($j(\theta)$), ajustando os parâmetros (θ) da função iterativamente, na direção do gradiente negativo (Almeida, 2023). O processamento dos dados poderia ser realizado de forma distinta, gerando variações como o *Stochastic Gradient Descent – SGD* (Souza, 2019) e o *Mini-Batch Gradient Descent* (Mamann, 2023).

$$\theta := \theta - \alpha \frac{\partial j(\theta)}{\partial \theta} \quad (VIII)$$

Onde: α : taxa de aprendizado; $\frac{\partial j(\theta)}{\partial \theta}$: gradiente da função de custo em relação a θ .

A comparação de desempenhos entre redes exigiria a estimativa de métricas como (James *et al.*, 2013; Bishop, 2006): 1) acurácia, quantidade de predições correta em relação ao total predito (IX); 2) precisão, quantidade de predições positivas corretas em relação ao total de positivos preditos (X); 3) sensibilidade (*recall*), proporção de predições positivas corretas (XI); e 4) *F1-Score*, média harmônica entre as duas últimas, equilibrando as

estimativas (XII). Então, alicerçado no exposto, o presente estudo objetivou prever o comportamento de universitários quanto à matrícula.

$$Acurácia = \frac{VPositivo + VNegativo}{VPositivo + VNegativo + FPositivo + FNegativo} \quad (IX)$$

$$Precisão = \frac{VPositivo}{VPositivo + FPositivo} \quad (X)$$

$$Sensibilidade = \frac{VPositivo}{VPositivo + FNegativo} \quad (XI)$$

$$F1 - Score = 2 \cdot \frac{Precisão \cdot Sensibilidade}{Precisão + Sensibilidade} \quad (XII)$$

Onde: *VPositivo*: verdadeiro positivo; *FPositivo*: falso positivo; *VNegativo*: verdadeiro negativo; *FNegativo*: falso negativo.

Metodologia

Utilizou um banco de dados com 326 registros de graduandos em Educação Física em instituição localizada em Pernambuco. Desse total, 212 (65,03%) correspondiam ao sexo feminino. As variáveis disponíveis eram: Sexo, 0 para feminino e 1 para masculino; Idade, entre 18 e 68 anos, Coeficiente de Rendimento Acumulado (CR), entre 2,30 e 9,90 pontos; Frequência com valores de 18,00% a 100,00%; Pagamento, codificado como 0 para atraso, 1 para em dia, e 2 para antecipado; participação Extracurricular, recebendo 0 para nunca, 1 para às vezes, 2 para a maioria das vezes e 3 para sempre, e Matrícula, classificada como 0 para abandono, 1 para trancamento, 2 para transferência para outra instituição, 3 para transferência de outra instituição, e 4 para rematrícula. Esse conjunto de dados foi acumulado entre os anos de 2020 e 2023 (BD2023), e utilizado para modelar a rede neural. Outro banco contendo 64 registros com similar estrutura, mas correspondendo ao primeiro semestre de 2024 foi empregado para a predição (BD2024).

A codificação foi realizada em R 4.4.3, empregando o RStudio 2022.07.1+554 "Spotted Wakerobin" for Windows. O conjunto de pacotes utilizados incluiu neuralnet 1.44.2 (*Training Neural Network*), keras 2.15.0 (*R Interface to 'Keras'*), tidyverse 2.0.0 (*Easily Install and Load the 'Tidyverse'*), caret 7.0-1 (*Classification and Regression Training*), pROC

1.18.5 (*Display and Analyze ROC Curves*) e moments 0.14.1 (*Moments, Cumulants, Skewness, Kurtosis and Related Tests*), respectivamente para redes *feedforward*, RNN/LSTM, gerenciamento de dados, pré-processamento, métricas e inferências.

Inicialmente, o banco de dados foi dividido nos grupos de Treino, Validação e Teste nas respectivas proporções de 70,00%, 15,00% e 15,00%, tal realização objetivou evitar a contaminação do conjunto de treino por informações do grupo de teste, o que originaria viés (Lumumba *et al.*, 2024; Asogwa *et al.*, 2023; Li *et al.*, 2022). Entretanto, à variável Pagamento foi imposto o processo *One-hot Encoding* (Harris e Harris, 2012). Em sequência, avaliada foi a similaridade das respectivas distribuições de frequências, garantindo que aqueles grupos não guardavam distinções, para tanto utilizou-se o teste Qui-quadrado para variáveis categóricas (Siegel e Castellan Jr, 2006) e Anova para contínuas (Triola, 2017), sempre com $\alpha = 0,05$. Essas submetidas à normalização pelo Z-score (Goodfellow, Bengio e Courville, 2018).

Os modelos de RNAs desenvolvidos foram *feedforward*, RNN e LSTM (Fleck *et al.*, 2016; Fiorin *et al.*, 2011), treinados pelo algoritmo *backpropagation* (Oliveira e Melo Júnior, 2024), porque o fenômeno era simples. A função de ativação selecionada foi a *softmax*, convergente à classificação multiclases (Silva e Amado, 2023; Silva, 2021). A função de perda à otimização do modelo foi a Entropia Cruzada (Silva *et al.*, 2021; Figueiredo, Oliveira e Souto, 2020). O desempenho dos modelos foi avaliado pelas métricas: acurácia, precisão, sensibilidade e F1-Score (Goodfellow, Bengio e Courville, 2018; James *et al.*, 2013; Bishop, 2006), em BD2023 e BD2024.

Resultados e Discussão

A divisão do banco de dados (BD2023) em Treino, Validação e Teste não produziu distinção em qualquer distribuição de frequência das variáveis contínuas (Tabela 1) ou categóricas (Tabela 2). Portanto, o risco de viés foi reduzido. Submetidos às métricas de desempenho, o *Feedforward* (Tabela 3) se apresentou inferiormente aos demais modelos em todos os grupos.

Tal quadro convergiu às expectativas, pois aquela rede não capturaria dependências temporais e as interações entre variáveis seriam modeladas de forma estática (Goodfellow, Bengio e Courville, 2018; Haykin, 2000). Contudo, identificou-se relações não-lineares no Rendimento, dado que haveria um platô no comportamento de matrículas para $CR > 8,50$, com melhora marginal acima desse ponto. O risco de perda existiria para Idade $< 25,00$ e

Idade > 60,00, portanto essa variável apresentava influência parabólica. Muito embora, os fatores de risco dominantes fossem CR e Frequência com Pesos Relativos, respectivamente, de 0,38 e 0,29, a antípoda seria o Pagamento antecipado (fator protetor), peso Relativo = -0,21, ou seja, o universitário que o realizasse teria maior probabilidade de se manter na instituição.

Tabela 1: Comparação das Distribuições das Variáveis Contínuas nos Grupos Treino, Validação e Teste. Anova, GL(Grupos) = 2,00 e GL(Resíduos) = 323,00.

| Variável | Domínio | Σ Quadrados | Média Quadrados | Estatística Teste (F) | Valor-p |
|------------|----------|--------------------|-----------------|-----------------------|---------|
| Idade | Grupo | 1,70 | 0,84 | 0,84 | 0,43 |
| | Resíduos | 323,30 | 1,00 | | |
| CR | Grupo | 3,10 | 1,56 | 1,57 | 0,21 |
| | Resíduos | 321,90 | 1,00 | | |
| Frequência | Grupo | 2,30 | 1,13 | 1,13 | 0,32 |
| | Resíduos | 322,70 | 1,00 | | |

Fonte: Os Autores (2025).

Tabela 2: Comparação das Distribuições das Variáveis Categóricas nos Grupos Treino, Validação e Teste. Teste Qui-quadrado, GL = 2,00.

| Variável | Estatística Teste (χ) | Valor-p |
|------------------|------------------------------|---------|
| Sexo | 1,03 | 0,60 |
| Pagamento | | |
| 0 | 0,60 | 0,74 |
| 1 | 3,84 | 0,15 |
| 2 | 3,83 | 0,15 |

Fonte: Os Autores (2025).

Tabela 3: Desempenhos dos Modelos no Banco de Dados de Implementação.

| Rede | Grupo | Acurácia | Precisão | Sensibilidade | F1-Score |
|-------------|-----------|----------|----------|---------------|----------|
| Feedforward | Treino | 0,78 | 0,71 | 0,55 | 0,62 |
| | Validação | 0,53 | 0,30 | 0,16 | 0,21 |
| | Teste | 0,56 | 0,23 | 0,21 | 0,22 |
| RNN | Treino | 0,85 | 0,83 | 0,84 | 0,84 |
| | Validação | 0,82 | 0,80 | 0,81 | 0,81 |
| | Teste | 0,81 | 0,80 | 0,80 | 0,80 |
| LSTM | Treino | 0,88 | 0,86 | 0,85 | 0,86 |
| | Validação | 0,84 | 0,83 | 0,83 | 0,83 |
| | Teste | 0,84 | 0,82 | 0,82 | 0,82 |

Fonte: Os Autores (2025).

A análise da RNN permitiu identificar que houve captura de padrões temporais, que pese ter conquistado Sensibilidade excessiva aos *outliers* temporais e, conforme esperado, sofreu efeito *vanishing gradient* em sequências com mais de 12 passos. O conjunto de

sequências de risco identificadas relevaram que a variável Pagamento com comportamento atraso → em dia → atraso elevaria o risco de perda de matrícula em 2,10 vezes. Mas, esse seria, aproximadamente, 78,00% maior que o padrão quando o cliente interno combinasse CR decrescente e Frequência oscilante. Mereceriam destaque as duas janelas temporais críticas, a primeira das quais permitiu afirmar que o risco seria reduzido em 34,00%, quando a participação em atividades Extracurriculares fosse consistente, a partir de três meses. A segunda, capturou o pico de risco (Sensibilidade = 0,85) como consequência de três ciclos consecutivos com CR < 5,00 pontos.

A LSTM identificou padrões complexos revelados nas sequências de efeito blindagem (Sensibilidade = 0,89) com consequência de histórico de, pelo menos, dois com Frequência > 0,80, e Pagamento = Antecipado → Extracurricular = Sempre → CR > 7,50 pontos seria a combinação com 51,00% de risco reduzido. A memória de longo prazo da rede descortinou o efeito recuperação, quando o Pagamento fosse mantido em dia por 12 meses teria por consequência o risco residual 22,00% menor, e que mesmo se houvesse melhora no desempenho, o CR < 6,00 pontos por seis ciclos sequenciais manteria o risco elevado. Com relação às interações não lineares, identificadas foram que 68,00% do risco seriam modulados pela relação entre as variáveis Extracurricular e Frequência, e que o CR corrente teria o triplo do peso do histórico da variável no comportamento sobre a matrícula.

Limitando as observações aos aspectos técnicos revelados na Tabela 3, percebeu-se que a LSTM apresentou métricas melhores que a RNN, sobretudo em Sensibilidade e F1-Score, devido à captura de dependências de longo prazo. A ocorrência de *overfitting* se apresenta quando a diferença entre acurácias de Treino e Validação se tornaria superior a 0,10, o que não se manifestou para RNN (0,85 – 0,82 = 0,03) ou LSTM (0,88 – 0,84 = 0,04), logo nesses casos a generalização foi adequada. O modelo RNN priorizou a captura de casos positivos, não impedindo a ocorrência de algum falso positivo, dado que a Sensibilidade (0,81) > Precisão (0,80) no grupo Validação. Finalmente, a validade do modelo se constatou pela proximidade entre as acurácias no Teste (0,84) e Validação (0,83) do modelo LSTM.

As considerações anteriores seriam válidas aos resultados do BD2024 (Tabela 4), destacando a melhora de desempenho do *Feedforward*, possivelmente, em razão do menor volume de dados (Braga, Ludermir e Carvalho, 2007; Dreyfus, 2005), $n_{BD2023} = 236$ e $n_{BD2024} = 64$. Apesar das distintas estimativas quantitativas, os modelos identificaram que CR < 5,00 pontos, Pagamento em atraso e Frequência < 50,00% seriam fatores de risco à

Matrícula. A concordância, também, se deu no reconhecimento da participação Extracurricular como fator de proteção (Tabela 5). Fundamental seria a compressão de que a absorção do padrão temporal tenderia a favorecer a compreensão do fenômenos, como consequência as estimativas foram mais elevadas, especialmente na LSTM pela captura de padrão complexos.

Tabela 4: Desempenhos dos Modelos no Banco de Dados de Predição.

| Grupo | Acurácia | Precisão | Sensibilidade | F1-Score |
|--------------------|-----------------|-----------------|----------------------|-----------------|
| <i>Feedforward</i> | 0,82 | 0,78 | 0,85 | 0,81 |
| RNN | 0,84 | 0,81 | 0,83 | 0,82 |
| LSTM | 0,86 | 0,83 | 0,87 | 0,85 |

Fonte: Os Autores (2025).

Tabela 5: Comparação dos Fatores de Risco no Banco de Dados de Predição. Padrões temporais (*).

| Fator | Feedforward | RNN* | LSTM* |
|---------------------|--------------------|-------------|--------------|
| CR < 5,00 pontos | 38,00% | +72,00% | +68,00% |
| Pagamento = Atraso | 25,00% | +41,00% | +53,00% |
| Frequência < 50,00% | 19,00% | 28,00% | +31,00% |
| Extracurricular | -12,00% | -34,00% | -51,00% |

Fonte: Os Autores (2025).

Em síntese, os resultados permitiram discriminar as aplicações por modelo, então o *feedforward* foi suficiente em BD2023 nos casos em que a média do CR > 6,50 pontos e o desvio padrão da Frequência não superasse 15,00%. Porém, para considerações temporais complexas, o LSTM seria mais adequado, se o período de análise fosse, minimamente, de seis ciclos e houvesse mais do que 100 registros. De forma geral, à tomada de decisão e às intervenções baseadas em risco, o monitoramento do universitário deveria ocorrer sempre que o valor predito fosse 0,70 no LSTM ou 0,60 no RNN, classificados, nessa ordem, como Crítico ou Alerta.

Considerações Finais

Tendo por objetivo prever o comportamento de universitários quanto à matrícula, três modelos de RNA foram estimados. Todos identificaram os fatores de risco e o fator de proteção, os quais influenciariam a renovação, transferência de e para outra instituição, de ensino superior, trancamento e abandono. Por conseguinte, concluiu-se que o objetivo foi alcançado, tendo cada modelo especificidade de aplicação.

Aos estudos futuros poderia ser interessante investigar a outros métodos de normalização com a finalidade de identificar aquele que melhor se adequa à situação posta. Assim como, o emprego de outros algoritmos de treinamento poderia fornecer métricas distintas às arquiteturas de rede similares. A inclusão de outras variáveis como quantidade de reprovações, avaliação sobre a instituição e autopercepção poderiam permitir o refinamento dos modelos.

Referências

- AGGARWAL, CC. **Neural networking and deep learning**. New York (USA): Springer, 2023.
- ALENCAR, VC; SILVA, LER; MENDES, MS. **Deteção de doenças de pele utilizando redes neurais convolucionais – CNN**. Curitiba (PR): CRV, 2021.
- ALMEIDA, APCR. **Avaliação de métodos de otimização em redes neurais convolucionais para identificação de dígitos**. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) - Universidade Tecnológica Federal do Paraná, Santa Helena, 2023.
- ARLOT, S; CELISSE, A. A survey of cross-validation procedures for model selection. **Statistics Surveys**, v. 4, p. 40–79, 2010.
- ASOGWA, SC *et al.* Smart city PM2. 5 air pollution modeling techniques: train-test data split versus k-fold cross validation techniques. **Journal of Inventive Engineering and Technology (JIET)**, v. 3, n. 1, p. 30-44, 2023.
- BISHOP, CM. **Pattern recognition and machine learning**. New York (USA): Springer, 2006.
- BRAGA, AP; LUDERMIR, TB; CARVALHO, ACPLF. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: LTC, 2007.
- CARNEIRO, MS *et al.* Rede Neural Convolucional aplicada na análise de sentimentos em comentários de clientes de empresa do ramo varejista. **Navus-Revista de Gestão e Tecnologia**, v. 16, p. 1-24, 2025.
- CURTY, LMA *et al.* Microempreendedor Individual: análise de perfil baseada no planejamento inicial do negócio. **Revista de Empreendedorismo e Gestão de Micro e Pequenas Empresas**, v. 9, n. 03, p. 01-26, 2024.
- DREYFUS, G. **Neural Networks: methodology and applications**. Berlin (Deutschland): Springer, 2005.
- ESTEVA, A *et al.* Dermatologist-level classification of skin cancer with deep neural networks. **Nature**, v. 542, p. 115-118, 2017.
- FENEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, v. 35, n. 1, p. 25-30, 2006.
- FIGUEIREDO, AF; OLIVEIRA, HS; SOUTO, EJ Pereira. Redes neurais densas para classificação de estresse. **Journal of Health Informatics**, v. 12, p. 202-208, 2020.
- FIORIN, DV *et al.* Aplicações de redes neurais e previsões de disponibilidade de recursos energéticos solares. **Revista Brasileira de Ensino de Física**, v. 33, n. 1, a. 1309, 2011.

FLECK, L *et al.* Redes neurais artificiais: princípios básicos. **Revista Eletrônica Científica Inovação e Tecnologia**, v. 1, n. 13, p. 47-57, 2016.

GONÇALVES, LB; REINALDI, MAA. Impacto da inteligência artificial no controle interno de Instituições Financeiras. **Revista Contabilidade & Inovação**, v. 3, n. 1, p. 49-59, 2024.

GOODFELLOW, I; BENGIO, Y; COURVILLE, A. **Deep Learning**. Das umfassende Handbuch: Grundlagen, aktuelle Verfahren und Algorithmen, neue Forschungsansätze. Frechen (Deutschland): MITP Verlags GmbH, 2018.

GURGEL, S; FORMIGA, AA. Parallel implementation of feedforward neural networks on gpus. In Intelligent Systems (BRACIS), 2013 **Brazilian Conference on**, Oct 2013, pp. 143–149.

HARRIS, D; HARRIS, S. **Digital design and computer architecture**. California (USA): Morgan Kaufmann. 2012.

HAYKIN, S. **Redes neurais: princípios e práticas**. Porto Alegre (RS): Bookman, 2000.

HOCHREITER, S. **Untersuchungen zu dynamischen neuronalen Netzen**. Diplomarbeit im Fach Informatik. Institut für Informatik Technische. Universität München. München (Deutschland).1991.

IOFFE, S; SZEGEDY, C. Batch normalization: accelerating deep network training by reducing internal covariate shift. **Proceedings of the 32nd International Conference on Machine Learning (ICML)**. PML, v. 37, p. 448-456, 2015.

JAMES, G *et al.* **An introduction to statistical learning: with applications in R**. New York (USA): Springer, 2013.

KHAN, MU *et al.* Detecting wake lock leaks in android apps using machine learning. **IEEE Access**, v. 9, p. 125753-125767, 2021.

KRIZHEVSKY, A; SUTSKEVER, I; HINTON, GE. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F *et al.* (Eds.) **Advances in Neural Information Processing Systems**, Vol. 25, Curran Associates, Inc., Red Hook, NY, 2012, p. 1097-1105.

LI, ZJ *et al.* Detection of nut–bolt loss in steel bridges using deep learning techniques. **Sustainability**, v. 14, n. 17, a. 10837, 2022.

LUMUMBA, VW *et al.* Comparative Analysis of Cross-Validation Techniques: LOOCV, K-Folds Cross-Validation, and Repeated K-Folds Cross-Validation in Machine Learning Models. **American Journal of Theoretical and Applied Statistics**, v. 13, p. 127-137, 2024.

MAGGIONI, AF; REBONATTO, MT. Detecção de arritmias cardíacas: abordagem da derivação I com redes neurais híbridas. **Revista Interfaces: Saúde, Humanas e Tecnologia**, v. 12, n. 4, p. 4585-4600, 2024.

MAMANN, LVS. **Aprendizado offline e online de redes neurais no contexto de casas inteligentes e de computação em névoa**. Dissertação (Mestrado em Engenharia Elétrica e Informática Industrial) - Universidade Tecnológica Federal do Paraná. Curitiba (PR), 2023.

MICCI-BARRECA, D. A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems. **ACM SIGKDD Explorations Newsletter**, v. 3, n. 1, p. 27-32, 2001.

OLIVEIRA, MA; MELO JUNIOR, FAP. **Backpropagation e redes neurais**. Rio de Janeiro: Ciência Moderna, 2024.

ROJAS, R. **Theorie der neuronalen Netze**. Berlin (Deutschland): Springer-Verlag, 1993.

SIEGEL, S; CASTELLAN JR, NJ. **Estatística não-paramétrica para ciências do comportamento**. Porto Alegre (RS): Artmed, 2006.

SILVA, DCG. Predição de estresse em ovelhas prenhas e lactantes com o uso de redes neurais artificiais. **Ensaio e Ciência: Ciências Biológicas, Agrárias e da Saúde**, v. 25, n. 2, p. 160-165, 2021.

SILVA, JAM; AMADO, FG. Conceptron: Framework para desenvolvimento de redes neurais artificiais. **Revista Eletrônica Científica Inovação e Tecnologia**, v. 14, n. 35, p. 32-44, 2023.

SILVA, JVS *et al.* Combining deep learning model compression techniques. **IEEE Latin America Transactions**, v. 20, n. 3, p. 458-464, 2021.

SOUSA, S. **Estudo de modelos de word embedding**. Trabalho de Conclusão de Curso (Graduação) - Universidade Tecnológica Federal do Paraná. Medianeira (PR), 2016.

SOUSA, SJ; SANTIAGO, MO; DIAS, TMR. Uma estratégia para identificação de gênero em repositórios de dados abertos utilizando um modelo de rede neural artificial. **Ciência da Informação**, v. 48, n. 3, p. 432-433, 2019.

SOUZA, CFC *et al.* Previsão da incidência de câncer por gênero em belo horizonte: inovação no uso de redes neurais artificiais para planejamento de saúde pública. **ARACÊ**, v. 7, n. 3, p. 15184-15201, 2025.

SOUZA, HA. **Treinamento de redes neurais com arquitetura *multilayer perceptron* em FPGA**. Trabalho de Conclusão de Curso (Bacharelado em Engenharia Elétrica) – Departamento de Engenharia Elétrica e Eletrônica. Universidade Federal de Santa Catarina. Florianópolis (SC), 2019.

TRIOLA, MF. **Introdução à estatística**. Rio de Janeiro: LTC, 2017.

WUEGES, AFE; BORBA, JA. Redes neurais, lógica nebulosa e algoritmos genéticos: aplicações e possibilidades em finanças e contabilidade. **Revista de Gestão da Tecnologia e Sistemas de Informação**, v. 7, n. 1, p. 163-182, 2010.

XU, Y; GOODACRE, R. On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. **Journal of analysis and testing**, v. 2, n. 3, p. 249-262, 2018.