



**Evento:** XXXIV Seminário de Iniciação Científica

## **APLICAÇÃO DE MODELOS MATEMÁTICOS NA PREDIÇÃO DO TEMPO DE VIDA DE BATERIAS QUE ALIMENTAM DISPOSITIVOS MÓVEIS<sup>1</sup>**

**Leonardo Ojczenasz Schmidt<sup>2</sup>, Airam Teresa Zago Romcy Sausen<sup>3</sup>, Julia Dammann<sup>4</sup>, Jean Schäffel Moreira<sup>5</sup>**

<sup>1</sup> Projeto de pesquisa desenvolvido na Unijuí; com financiamento da agência CNPq.

<sup>2</sup> Estudante do curso Ciência da Computação, bolsista do Programa Institucional de Bolsas de Iniciação Científica do Conselho Nacional de Desenvolvimento Científico e Tecnológico - PIBIC/CNPq.

<sup>3</sup> Professor orientador do projeto Aplicação de Modelos Matemáticos na Predição do Tempo de Vida de Baterias que Alimentam Dispositivos Móveis.

<sup>4</sup> Doutoranda em Modelagem Matemática e Computacional da UNIJUÍ.

<sup>5</sup> Mestrando em Modelagem Matemática e Computacional da UNIJUÍ.

### **INTRODUÇÃO**

Esse trabalho apresenta os resultados obtidos na terceira etapa de uma bolsa de Iniciação Científica (PIBIC/CNPq), vinculada à pesquisa sobre aplicação de modelos matemáticos na predição do tempo de vida de baterias de Lítio-Íon Polímero (Li-Po) que alimentam dispositivos móveis. O projeto é desenvolvido em apoio a uma tese de doutorado e utiliza como base um conjunto de dados de 850 ciclos de carga e descarga extraídos de baterias por meio de uma plataforma de testes do Grupo de Automação Industrial e Controle (GAIC).

Na etapa anterior (2024/2025), foi implementado a técnica SMOTER (Synthetic Minority Oversampling Technique for Regression), uma adaptação do SMOTE (CHAWLA et al., 2002) para problemas de regressão proposta por Torgo et al. (2013), com o objetivo de ampliar artificialmente o conjunto de dados de treinamento e avaliar se essa ampliação contribuiria para reduzir o erro de predição da rede neural. Os resultados daquela etapa revelaram uma ambiguidade: o modelo com SMOTER se saiu melhor ao prever um modelo teórico de referência, mas pior ao prever os dados experimentais reais, levantando dúvidas sobre a validade metodológica da comparação.

Ao reanalisar o protocolo experimental, identificou-se que a divisão automática realizada pelo MATLAB entre os conjuntos de treino, validação e teste variava a cada execução, tornando as condições experimentais incompatíveis entre si. Além disso, ao



concatenar dados sintéticos antes da divisão, havia o risco de que amostras artificiais fossem incluídas no conjunto de teste, comprometendo a avaliação da capacidade de generalização do modelo em dados reais. Diante dessas limitações, esta etapa teve como objetivo estruturar um protocolo experimental mais rigoroso, reproduzível e estatisticamente fundamentado, capaz de avaliar de forma consistente o impacto do SMOTER no desempenho da rede neural.

## METODOLOGIA

A arquitetura da rede neural foi mantida com duas camadas ocultas de 10 neurônios cada, algoritmo de treinamento Levenberg-Marquardt (trainlm) e função de ativação logsig, garantindo coerência com as etapas anteriores do projeto, com os dados organizados nas variáveis de entrada (2x850) e saída (1x850) no MATLAB.

A primeira correção metodológica consistiu em substituir a divisão automática aleatória do MATLAB por uma divisão manual com índices fixos, salvos em `indices_reais.mat`, de modo que todos os experimentos passaram a ser avaliados sobre exatamente o mesmo conjunto de testes, composto exclusivamente por dados reais. No experimento com SMOTER, apenas o conjunto de treino foi ampliado com dados sintéticos, enquanto a validação e teste permaneceram inalterados.

Para lidar com a variabilidade decorrente da inicialização aleatória dos pesos da rede neural, foi implementado um protocolo com 30 seeds (sementes) distintas, onde cada seed gera uma inicialização diferente da rede e, para cada uma delas, obteve-se um MSE (Mean Squared Error - Erro Quadrático Médio) de teste tanto para o baseline (modelo treinado apenas com dados reais) quanto para o SMOTER, ambos avaliados sobre o mesmo conjunto de teste real. Esses pares diretamente comparáveis foram então submetidos ao teste t pareado, que verifica se a média das diferenças entre os métodos é estatisticamente diferente de zero, complementado pelo coeficiente d de Cohen para amostras pareadas, que quantifica a magnitude prática dessa diferença.

O script em Python responsável pela geração das amostras sintéticas foi ajustado para controle explícito da taxa de oversampling (sobreamostragem), possibilitando testar sistematicamente taxas entre 10% e 50% em incrementos de 5%, cada uma avaliada com 30 seeds. Por fim, foi conduzida uma análise de curva de aprendizado, treinando o modelo com



frações crescentes do conjunto de treino - 20%, 40%, 60%, 80% e 100% - com 30 seeds por fração e a taxa ótima identificada.

## RESULTADOS E DISCUSSÃO

Na comparação inicial com seed fixa ( $\text{rng} = 10$ ), o baseline (referência sem dados sintéticos) apresentou MSE de teste de 38,87 e generalization gap (diferença entre erro de treino e erro de teste) de 16,28, enquanto o SMOTER apresentou MSE de 36,49 e gap de 12,27, indicando redução tanto no erro quanto na diferença entre desempenhos em treino e teste.

Com o protocolo de 30 seeds e taxa de oversampling de 20%, o baseline apresentou MSE médio de 41,02 e SMOTER de 38,93. O teste t pareado resultou em  $p\text{-value} = 0,000411$  e intervalo de confiança de 95% igual  $[1,0167; 3,1553]$ , o que significa que a diferença média real entre os MSEs do baseline e do SMOTER está estimada entre aproximadamente entre 1,02 e 3,16 pontos, sempre favorável ao SMOTER. Como o intervalo não inclui o valor zero e o  $p\text{-value}$  é inferior a 0,05, a redução é estatisticamente significativa. O coeficiente d de Cohen resultou em 0,73, caracterizando um efeito moderado a alto. O SMOTER apresentou MSE menor que o baseline em 23 das 30 seeds, reforçando que a melhoria é consistente.

Na análise das taxas de oversampling, a taxa de 10% não produziu ganho significativo ( $p\text{-value} = 0,537$ ), indicando que uma ampliação muito pequena da base não é suficiente para impactar o modelo de forma consistente. A partir de 15%, as reduções passaram a ser estatisticamente significativas em todas as taxas testadas, sendo que o menor MSE médio foi obtido com 40% de oversampling ( $\text{MSE} = 38,64$ ,  $p\text{-value} = 0,000018$ ), confirmando esse valor como a taxa ótima para o conjunto de dados estudado.

A análise de curva de aprendizado, conduzida com a taxa ótima de 40%, revelou o resultado mais expressivo do trabalho, apresentado na Tabela 1.

Tabela 5 — Curva de aprendizado: MSE médio de teste por fração do conjunto de treino.

| Fração do treino(%) | MSE baseline | MSE SMOTER 40% |
|---------------------|--------------|----------------|
| 20                  | 3316,41      | 110,08         |
| 40                  | 1444,95      | 192,77         |



|     |        |        |
|-----|--------|--------|
| 60  | 539,45 | 118,93 |
| 80  | 192,19 | 176,79 |
| 100 | 41,02  | 38,64  |

Fonte: elaborado pelo autor (2026).

Com apenas 20% do conjunto de treino, o baseline atingiu MSE médio de 3316,41, enquanto o SMOTER manteve 110,08 - uma diferença de aproximadamente 30 vezes. Conforme o volume de dados reais aumenta, as duas curvas convergem, chegando a valores muito próximos com 100% do treino (41,02 vs 38,64). Esse comportamento sugere que o SMOTER atua como forma de compensar a escassez de dados, com impacto naturalmente diminuído à medida que mais dados reais são incluídos no treinamento - que é exatamente o cenário deste projeto, dado que o conjunto de dados é limitado pelos ciclos coletados na plataforma do GAIC. Esse resultado também esclarece a ambiguidade do ano anterior: com o protocolo corrigido, o ganho do SMOTER mostrou-se consistente e estatisticamente robusto.

## CONSIDERAÇÕES FINAIS

Os resultados obtidos nesta etapa indicam que a técnica SMOTER pode contribuir para a redução do erro de predição da rede neural, desde que aplicada com um protocolo experimental rigoroso e calibrada adequadamente. A principal contribuição desta etapa foi a correção metodológica ao fixar os índices de divisão dos dados e garantir que o conjunto de teste fosse exclusivamente composto por dados reais, foi possível estabelecer uma comparação válida entre baseline e SMOTER, algo que não havia sido garantido na etapa anterior.

A taxa ótima identificada foi de 40% de oversampling, com  $p\text{-value} = 0,000018$  e  $d$  de Cohen = 0,73. A análise de curva de aprendizado revelou que o benefício do SMOTER é especialmente pronunciado em cenários de escassez de dados, com reduções de até 96,7% no MSE quando apenas 20% dos dados de treino estão disponíveis.

**Palavras-chave:** Redes Neurais Artificiais. Baterias Li-Po. SMOTER. Aprendizado de Máquina. Modelagem Matemática.



## AGRADECIMENTOS

Ao CNPq, pelo financiamento desta pesquisa por meio da bolsa PIBIC.

## REFERÊNCIAS BIBLIOGRÁFICAS

CHAWLA, N. V. et al. Smote: synthetic minority over-sampling technique. **Journal of artificial intelligence research**, v. 16, p. 321-357, 2002.

DAMMANN, Julia. **Modelagem Matemática do Tempo de Vida de Baterias Através da Proposição de um Modelo Híbrido Personalizado que Considera o Número de Ciclos de Descarga**. 2022. Dissertação (Mestrado) - Universidade Regional do Noroeste do Estado do Rio Grande do Sul - UNIJUÍ, Ijuí. 2022.

MATLAB Documentation. Disponível em: <<https://www.mathworks.com/help/index.html>>. Acesso em: dez. 2025.

TORGO, L. et al. Smote for regression. In: SPRINGER. **European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases**. 2013. p. 348-363.