

## Aprendizado de máquina supervisionado para predição de dados clínicos para a busca do melhor modelo

Mateus Padilha da Mota<sup>1</sup>

Randhall Bruce Kreismann<sup>2</sup>

**Resumo:** Atualmente estabelecido como uma preocupação de saúde mundial, o traumatismo cranioencefálico (TCE) apresenta alto impacto nos sistemas de saúde pública, causando uma grande perda de saúde e cuidados de alto custo. O estado nutricional é comprometido após um TCE grave, ocorrendo alto risco de desnutrição, evidenciando a necessidade de monitoramento da adequação nutricional e de marcadores antropométricos. Ao mesmo tempo, o uso de aprendizado de máquina (ML) tem crescido, seu principal objetivo é utilizar dados para identificar padrões e oferecer alternativas que contribuam para a tomada de decisões. Para a área da saúde esse tipo de tecnologia contribui para diagnósticos mais precisos, automação de processos e para a redução do custo e do tempo de diagnóstico. O objetivo desse trabalho foi avaliar o poder preditivo prognóstico de parâmetros de adequação nutricional em pacientes que sofreram traumatismo cranioencefálico grave, usando o aprendizado de máquina. Foi realizado um estudo retrospectivo com 71 pacientes (61 homens e 10 mulheres) em um hospital de referência na região metropolitana de Porto Alegre. Foi avaliada a adequação nutricional (média da primeira semana de internação), índice de massa corporal, perímetro de braço e perímetro de panturrilha. A análise dos dados incluiu testes de Shapiro-Wilk, teste t para amostras independentes. A análise de ML incluiu a construção de um algoritmo para identificar o potencial preditivo das diferentes variáveis para o desfecho, em sete fases: análise exploratória, processamento de dados, análise de features, divisão dos dados, treinamento, validação e otimização de hiperparâmetros. Iniciou-se realizando a análise exploratória dos dados para resumir suas características, posteriormente foi realizado o processamento de dados, a análise de features, a divisão de dados sendo 80% para treino e 20% para teste, foi realizado o treinamento com diferentes algoritmos, e avaliados pelo f1-score, os 2 melhores algoritmos foram comparados pela correlação de Matthews e foram hipertunados com GridSearchCV, RandomSearchCV e otimização Bayesiana (Reg.Plataforma Brasil: 3.912.623). Os pacientes tinham a idade média de 43.68+17.6 anos e foram divididos pelo desfecho (Sobrevida (S): n = 49 e Óbito (O): n = 22). Diferenças significativas foram encontradas para todos os parâmetros nutricionais avaliados ( $p < 0.005$ ).

---

<sup>1</sup> Discente do Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas do Centro Universitário Cesuca. E-mail: mateuspmota2010@gmail.com

<sup>2</sup> Docente do Curso de Nutrição do Centro Universitário Cesuca. Doutor em Bioquímica. E-mail: randhall.carteri@cesuca.edu.br

O algoritmo com melhor predição para o conjunto de variáveis foi o Extra tree com RandomSearch que apresentou um f1-score de 80.87% e acurácia de 37.50%. Por meio do coeficiente de correlação ponto-bisserial e com o SHAP notou-se que as variáveis que mais influenciam na predição são quilocaloria, adequação e proteína que devido os valores da correlação serem altos ( $>0.30$ ) são boas preditoras.

**Palavras-chave:** Machine learning; Data science; Nutrição