

Desenvolvimento e análise de modelos de estatística multivariada baseados em Aprendizado de Máquina: predição de partos pré-termo na coorte de nascimento BRISA em Ribeirão Preto

Gabriel Bazo¹, Ricardo Vêncio², Marco Antonio Barbieri¹

Introdução: O nascimento pré-termo (PT) é um dos principais fatores que levam a morte e causa, a longo prazo, significativa perda de capacidade, tanto física quanto psicológica entre os sobreviventes.

Objetivos: Desenvolver uma ferramenta computacional baseada em Aprendizado de Máquina capaz de prever o nascimento PT em uma coorte de nascimentos brasileira.

Métodos: Trata-se de um estudo transversal, baseado na coorte BRISA, que coletou informações sobre todos os nascimentos em 8 hospitais de Ribeirão Preto, Brasil em 2010. Foram aplicados questionários sobre a mãe e o recém-nascido. Nesta análise, apenas as informações maternas anteriores ao nascimento foram consideradas. Foram utilizadas características sociodemográficas, estilo de vida, características da gestação e acompanhamento pré-natal. Uma regressão não-paramétrica foi utilizada para determinar as variáveis independentes e a variável dependente foi criada a partir da categorização da idade gestacional (<37s PT, ≥37s termo). Seis algoritmos de predição foram testados, *Random Forest* (RF), *Support vector machines* (SVM), *Naive Bayes* (NB), *Decision Tree* (DT), *Linear discriminant analysis* (LDA) e *K-nearest neighbors* (KNN).

Resultados: 7699 gestantes participaram e 7651 atenderam todos os critérios de inclusão. A idade média foi de 27 anos ($\pm 6,25$), 44,22% eram primíparas. A taxa total de PT foi de 12,7%. As variáveis com maior poder de predição para o parto PT, foram: ocorrência de um parto anterior PT, ameaça de parto PT na gestação atual, quantidade de consultas do pré-natal no segundo e terceiro trimestre da gestação, realização e tipo de serviço do pré-natal, número de fetos, hospitalização durante a gestação, ocorrência de doenças durante a gestação corrente, hipertensão gestacional, anemia e quantidade de filhos vivos. RF apresentou a maior acurácia (acc=0,913), maior sensibilidade (s=0,81) e maior área sob a curva (auc=0,842). KNN apresentou a maior especificidade (e=0,972). SVM obteve o maior valor positivo preditivo (vpp=0,649). NB, por sua vez, obteve o maior valor preditivo negativo (vpn=0,984). A simulação computacional de um pré-natal, nas normas do Ministério da Saúde, em mães nas quais o modelo de predição (SVM) indicou a ocorrência de um possível parto PT, demonstrou uma redução de 3,9% no número de nascimentos PT (0,5 na porcentagem). Assim sendo, em âmbito nacional, no ano de 2017 seria possível uma redução de cerca de 12,7 mil partos PT, uma economia média de 119 milhões de reais, para ambos sistemas, público e privado.

Conclusão: Fatores como hospitalização durante a gravidez, hipertensão, tipo de serviço e características do pré-natal foram determinantes para a predição do parto PT. O estudo demonstrou ainda que, foi possível desenvolver modelos de predição baseados em Aprendizado de Máquina para partos PT, e sugere-se que esses modelos, adotados em nível nacional por meio de políticas públicas de saúde possam reduzir o número de partos PT, impactando o custo da saúde pública no país, e muito além disso, a qualidade de vida da mãe e seu recém-nascido.

Palavras-chave: parto pré-termo, saúde perinatal, políticas de saúde, modelos de predição, aprendizado de máquina.

1 – Faculdade de Medicina de Ribeirão Preto – FMRP-USP

2 – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto - FFCLRP-USP