

Comparação de *Random Survival Forest* e modelo de Cox com relação a performance de previsão: Um estudo de caso

Tiago Almeida de OLIVEIRA ¹, Pedro Augusto Ferreira da SILVA ², Hiago José Andrade de Albuquerque MARTINS ³, Lucas Cardoso PEREIRA ⁴, Alisson de Lima BRITO ⁵, Ednário Barbosa de MENDONÇA ⁶

Resumo: *A análise da sobrevivência é atualmente uma das ferramentas estatísticas que mais crescem no meio acadêmico. Na análise de sobrevivência existe uma sólida teoria de modelos de regressão que pode ser usada para modelar dados com observações incompletas chamadas censuras, a maioria desses modelos são paramétricos e existe também o modelo de riscos proporcionais de Cox. Eles não tem a melhor capacidade preditiva, neste sentido, os modelos Machine Learning em conjunto com o Random Forest em análise de sobrevivência (RSF) são uma alternativa crescente para o uso em predição. Foram ajustados 4 diferentes configurações de coariáveis no RSF, partindo de um modelo saturado com presença de interação até um modelo parcimonioso baseado em critérios próprios a área de Machine Learning para escolha de variáveis. Os modelos foram comparados contra o modelo de Cox via critérios C-index e Brier Score Index - IBS. O melhor modelo ajustado para predição foi o modelo completo com todas as covariáveis sob modelagem de Random Survival Forest.*

Palavras-chave: Análise de Sobrevivência, Riscos Proporcionais, Machine Learning.

¹Departamento de Estatística - UEPB. e-mail: *email*

²UEPB - Universidade Estadual da Paraíba. e-mail: *pedro1_20@hotmail.com*

³UEPB - Universidade Estadual da Paraíba. e-mail: *hiago1803@gmail.com*

⁴UEPB - Universidade Estadual da Paraíba. e-mail: *lacascp@hotmail.com*

⁵Pós graduação em Estatística e Experimentação Agropecuária - UFLA. e-mail: *alissonjs95@gmail.com*

⁶Departamento de Estatística - UEPB. e-mail: *eddybarbosa92@gmail.com*