# MACHINE LEARNING TECHNIQUES TO PREDICT COTTON YIELD USING SATELLITE REMOTE SENSING

**Francielle Morelli-Ferreira**
**Getulio de F. Seben Junior**
**Philip Lanza**
**L. S. Shiratsuchi**
**LSU AgCenter**
**Baton Rouge, LA**
**Nayane Jaqueline C. Maia**
**Elizabeth H. Kazama**
**Danilo Tedesco**
**Glauco de S. Rolim**
**Rouverson P. Silva**
**Unesp, Jaboticabal, SP, Brazil**
**Ziany Neiva Brandao**
**Embrapa Cotton, Brasilia, Brazil**

## Abstract

Cotton yield prediction models using machine learning and remote sensing data as input is incipient, specially conducting on-farm experimentation. Thus, commercial fields equipped with technologies in Mato Grosso, Brazil, were monitored by satellite images to predict cotton yield using supervised learning techniques. The objective of this study was to identify how early, which vegetation indices, and what machine learning algorithm is optimal to predict cotton yield at farm level, and for that, we have gone through the following steps: 1) We observed the cotton yield in 3 fields (983 acres) and eight vegetation indices (VI) were calculated on five dates during the cotton growing cycle. 2) Scenarios were created to facilitate the analysis and interpretation of results: Scenario 1: All Data (8 indices on 5 dates = 40 inputs) and Scenario 2: the best variable selected by Stepwise (1 input). 3) In the search for the best algorithm, hyperparameter adjustments, calibrations, and tests using machine learning were performed to predict cotton yield and performances were evaluated. Scenario 1 (All Data) presented the best metrics in all fields of study, and the Multilayer Perceptron (MLP) and Random Forest (RF) algorithms showed the best performances with adjusted R2 of 47% and RMSE of only 0.24 t $ha^{-1}$, however, in this scenario all inputs that were generated throughout the growing season (approx. 180 days) are needed, so we optimized the prediction and tested only the best VI in each field, and the Simple Ratio (SR) plus the K-Nearest Neighbor (KNN) algorithm, predicts with 0.26 and 0.28 t $ha^{-1}$ of RMSE and 5.20% MAPE, anticipating the cotton yield with low error by ±143 days, and with important aspect of requiring less computational demand in the generation of the prediction when compared to MLP and RF, for example, enabling its use as a technique that helps farmers to predict cotton yield, resulting in time savings for planning, whether in marketing or in crop management strategies. Authors acknowledge the projects USDA (LAB94427) and Louisiana Cotton Board and Cotton INC for the PhD sandwich of Francielle Morelli Ferreira financial support.