

DATA-DRIVEN APPROACH TO RUBBER YIELD FORECASTING: A CASE STUDY

H.R.N.B. Meththananda*, J. Senarathne and S. Kodithuwakku

Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka.

**bhagyameththananda97@gmail.com*

Rubber cultivation plays a vital role in Sri Lanka's economy, with yield influenced by various climatic and agronomic factors. This study investigated rubber yield dynamics using five years of monthly data from 2019 to 2023, comprising 6,007 observations collected from a large commercial rubber processing industry. Key variables included tappable trees, stand per hectare, tapping days, number of trees per tapper, rainfall, and tapping techniques. To capture these complex relationships, machine learning (ML) models, Lasso Regression, K -Nearest Neighbours (KNN), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) regressor were employed. An ensemble model using a voting regressor was constructed by combining the outputs of all four models to enhance predictive robustness. Additionally, a Long Short-Term Memory (LSTM) network was utilised for long-term forecasting, effectively identifying temporal dependencies and seasonal patterns. The dataset was shuffled and divided into 70% for training, 20% for testing, and 10% for validation. XGBoost achieved the best short-term performance, with a coefficient of determination (R^2) of 0.93 and the lowest mean absolute error (MAE) of 489.17 on the test set, outperforming other short-term forecasting models and effectively capturing complex, non-linear relationships. For long-term forecasting, the LSTM model delivered exceptional precision with minimal errors, achieving an MAE of 0.18 on the test set, enabling accurate yield predictions critical for planning tapping activities and managing seasonal variations. The study reveals that tappable trees, stand per hectare, and tapping days are the most significant yield determinants, with cross-correlation analysis indicating that prior-month rainfall has a 0.002 higher correlation with yield than current-month rainfall due to the disruptive impact of heavy rain on tapping activities. These findings highlight the importance of integrating agronomic practices and climatic factors into predictive models, offering critical insights to enhance productivity and sustainability in Sri Lanka's rubber industry.

Keywords: Climatic factors, k -nearest neighbours, Long short-term memory, Random forest, Voting regressor