University of Peradeniya

Machine Learning for rice yield prediction based on weather data

Aminda Amarasinghe[1] Ishini Sangarasekara[1] Damayanthi Herath[1] Nuwan de Silva[2] Mojith Ariyaratne[2] Ruwanga Amarasinghe[3] Janaka Alawatugoda[4]

[1]Department of Computer Engineering, Faculty of Engineering, University of Peradeniya. [2]Department of Crop Science, Faculty of Agriculture, University of Peradeniya.
[3]School of Technology, Sri Lankan Technological Campus. [4]Institute for Integrated and Intelligent Systems, Griffith University

Abstract – Rice is an important crop in Sri Lanka, primarily grown during the two distinct seasons of the year (*Yala* and *Maha*) in Sri Lanka. Accurate and timely rice yield prediction is crucial for food security in the country. This study focuses on rice yield prediction using less number of variables: Rainfall, Maximum temperature, Minimum temperature and Radiation. The data from Kurunegala district in the *Yala* and *Maha* seasons collected from 1982 to 2013 were used to analyze the crop yield. Data preprocessing included outliers and missing values handling and normalization. The Machine Learning models considered are Linear Regression (LR), Support Vector Regression (SVR), K-Nearest Neighbour (KNN) regression and Random Forest (RF) regression. The performance of these models was evaluated using three metrics: Root Mean Squared Error (RMSE), Relative Root Mean Squared Error (RRMSE), and Mean Absolute error (MAE). Random Forest regression models are the best models for both seasons according to results.

Introduction:

- Climate changes and yield uncertainties pose challenges in rice cultivation.
- Research focuses on employing machine learning to predict rice yields based on weather variables.
- Aims to support farmers' decisions and enhance agricultural outcomes.

Data:

- Daily weather data (radiation, maximum and minimum temperature, and rainfall) of Kurunegala district:
  collected from the Natural Resources Management Centre (NRMC), Department of Agriculture, Peradeniya.
- Total Rice Yield amount of two seasons:
  collected from the Department of Census and Statistics

Methodology:

After data preprocessing, and feature selection, K-fold Cross-validation was used to develop machine-learning models. Interquartile Range method was used for outlier handling under data preprocessing. Feature selection was done considering feature importances. For both Yala and Maha season, min. temperature is the only feature selected for further processing based on feature selection.

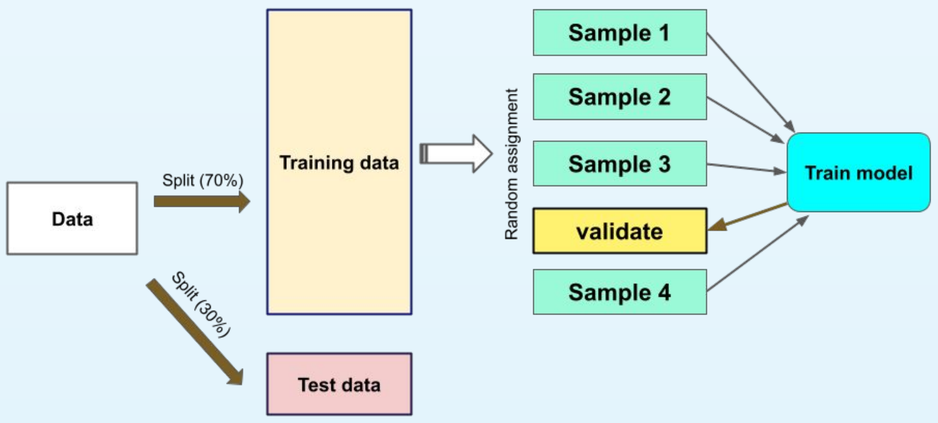

Figure 1: Machine learning workflow with Cross validation

***Yala* season**

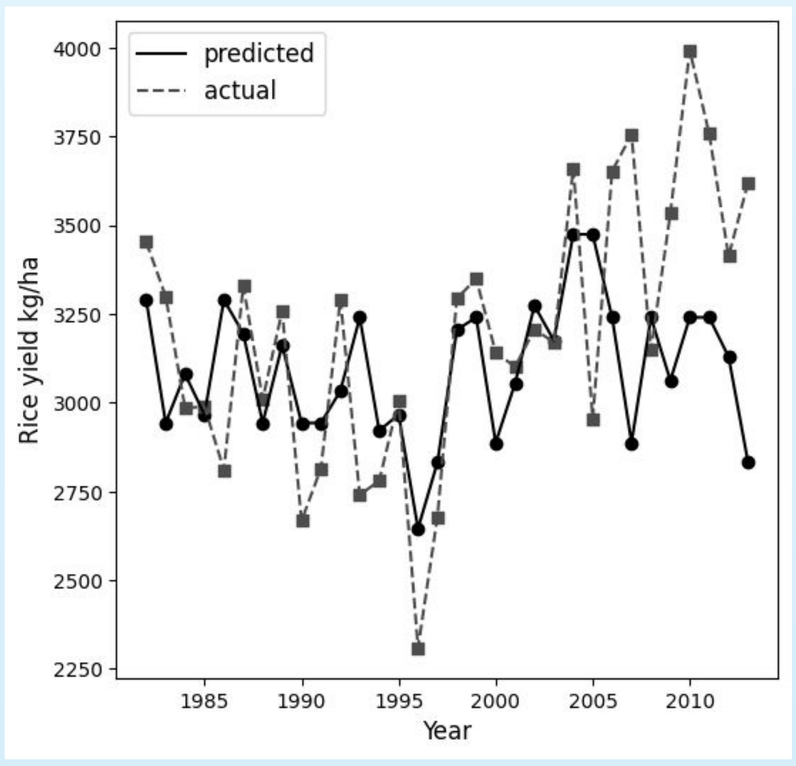

Figure 2: Actual and predicted rice yield using Random Forest model for *Yala* season

***Maha* season**

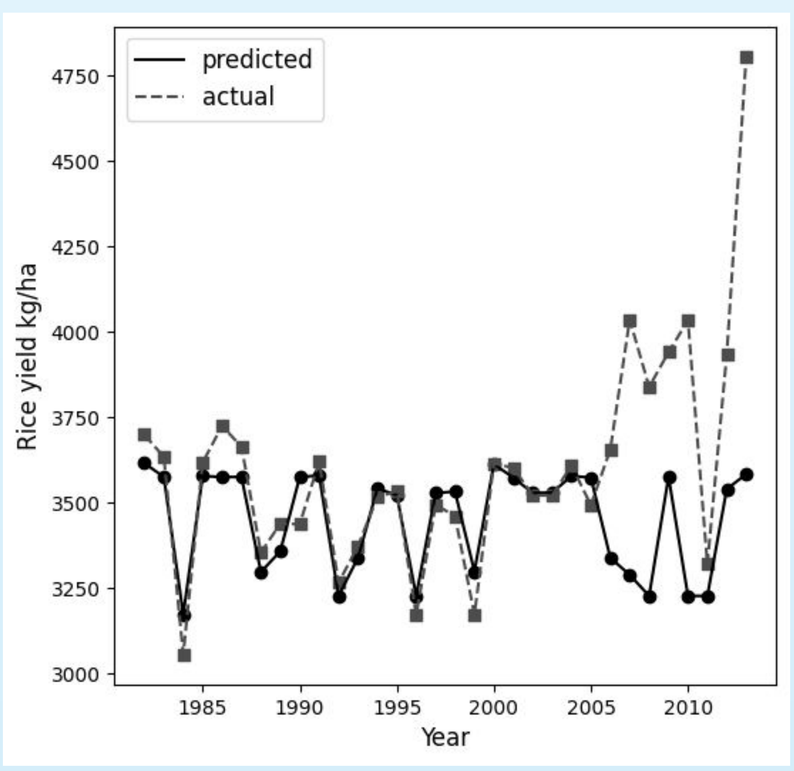

Figure 3: Actual and predicted rice yield using Random Forest model for *Maha* season

Conclusion:

- ML models show better performance for the Maha season compared to the Yala season.
- Machine Learning can be used for prediction of rice yield using weather parameters.

References:

1. Piyal Ekanayake, Windhya Rankothge, Rukmal Weliwatta, Jeevani W. Jayasinghe, "Machine Learning Modelling of the Relationship between Weather and Paddy Yield in Sri Lanka", Journal of Mathematics, vol. 2021, Article ID 9941899, 14 pages, 2021.
2. Lasini Wickramasinghe, Rukmal Weliwatta, Piyal Ekanayake, Jeevani Jayasinghe, "Modeling the Relationship between Rice Yield and Climate Variables Using Statistical and Machine Learning Techniques", Journal of Mathematics, vol. 2021, Article ID 6646126, 9 pages, 2021. https://doi.org/10.1155/2021/6646126

Contact details
Name : Damayanthi Herath
Tel. No.: +94 77 966 7468
Email : damayanthiherath@eng.pdn.ac.lk

Multidisciplinary AI Research Centre (MARC)
University Research Council
University of Peradeniya
Peradeniya, 20400, Sri Lanka