RICE YIELD PREDICTION IN THE NORTHERN AND EASTERN PROVINCES OF SRI LANKA USING WEATHER DATA

Kirushika Jeyachandran*

Department of Computer Science, Open University of Sri Lanka.

INTRODUCTION

The agricultural sector is the backbone of the national economy of Sri Lanka, as it consists 21.7% of the total national exports. Rice is the major agricultural crop, and occupies 34% of the total cultivated area (RRDI, 2021). The country has been entirely dependent on rice as a primary food source for several centuries. Diverse factors influence the production of rice in Sri Lanka and those facts should be investigated to predict and manage future demands.

Sri Lankan farmers face several hazards due to biotic and abiotic stresses, and among them, unpredicted climate change is crucial. Every year, a large amount of rice production is affected due to floods, high temperature, wind speed and several other weather elements. High Temperatures make rice flowers sterile and therefore grain is not produced (Makoto Endo, 2009). Rice cannot survive if submerged under water for long periods of time due to flooding. In this background, rice production and climate changes should be investigated together to produce a robust rice prediction model in Sri Lanka.

Over the past few years, several studies have been conducted to predict rice production and the weather elements in Sri Lanka and other several countries. An early study investigated the factors which affected paddy yield in the Anuradhapura and Ratnapura districts (Leaker, 1984).A weather forecasting model is proposed for dry and wet zones of Sri Lanka(Upul Sonnadara, 2002). A data mining and machine learning based model is proposed to predict the weather elements in India(Aishwarya Dhore, 2017). The impact of climate change in the Sri Lankan agricultural sector was also investigated (ROBERT MENDELSOHN, 2005). These studies either predict or investigate the weather elements and rice production separately. From our knowledge, no model was proposed to investigate the climate change and rice production together in the Sri Lankan context.

This study aims to investigate the trend of rice yield production in the Northern and Eastern provinces of Sri Lanka, by analysing the corresponding regional climatic data. In addition, this study is focused on identifying the weather elements, which affect the paddy growth in Sri Lanka. This study was conducted in the Jaffna, Vavuniya, Mannar, Batticaloa and Trincomalee districts, based on the collected yield data during the *Yala* and *Maha* seasons. The raw weather data of the past eleven years was collected from the Department of Meteorology. Regional rice production details were collected from the Department of Census and Statistics.

OBJECTIVES

The objective of this study is to investigate the trend of rice yield production in Sri Lanka by analyzing the regional climatic data. In addition, this study focuses on identifying the weather elements, which are affecting the paddy growth in Sri Lanka.

METHODOLOGY

Data Collection and Pre-processing

Figure 1 Study areas

Monthly raw climatic data were collected from the Department of Meteorology, Sri Lanka. Among the seven districts of Northern and Eastern provinces, based on the data availability, five districts were selected - namely Jaffna, Vavuniya, Mannar, Batticaloa, and Trincomalee. Figure 1 shows the location map of this study area. In these districts, climate data was collected for the past eleven years, from 2009 to 2019. Based on the availability, three weather elements were collected and used for this study: Temperature, Humidity and Rainfall.

Northern and Eastern provinces of Sri Lanka experience two rainfall seasons and farmers cultivate rice during these seasons. These seasons are known as *Maha* and *Yala*. *Maha* season starts from October and ends in January. *Yala* season starts from April and ends in July (RRDI, 2021). Details of the months for sowing paddy were collected from the Provincial Department of Agriculture of Northern and Eastern Province.

The climatic data of the *Maha* and *Yala* period were used in the present study. Since the paddy crop takes 106 to 110 days to harvest, four months were considered the period of paddy growth. Average values of weather elements were obtained and used to predict the paddy production in the selected region. Figure 2 illustrates the *Maha* and *Yala* seasons' weather data and the corresponding rice yield data of the Vavuniya district.

	RF	RH	Temp	Yield
2009	953.5	NaN	26.86	24457
2010	1181.0	NaN	26.32	49171
2011	842.7	NaN	26.95	58888
2012	1397.2	NaN	26.37	62512
2013	698.0	NaN	26.64	17515
2014	1236.4	NaN	26.05	76363
2015	936.9	NaN	26.67	71889
2016	607.3	80.35	25.70	22083
2017	509.0	82.90	26.60	35071
2018	840.3	83.80	26.30	96567
2019	944.1	85.50	28.70	77646

	RF	RH	Temp	Yield
2009	159.8	68.81	29.93	789
2010	201.8	70.18	29.92	3685
2011	200.0	70.92	29.55	16781
2012	132.4	67.07	29.83	2692
2013	212.5	72.69	29.24	23930
2014	413.1	67.58	29.85	5455
2015	472.8	72.25	29.49	24978
2016	547.5	75.10	29.50	4405
2017	355.7	70.60	30.10	1940
2018	373.0	74.60	29.20	1557
2019	152.7	64.90	30.90	11730

Figure 2 Climate and corresponding rice yield data of Vavuniya district Left: Maha season, Right: Yala season.

Implementation

Feature Scaling

The weather and rice yield data samples collected were measured in different ranges and scales. Rainfall measured in ‘mm’ was in the range of 0-1000. The temperature was measured in Celsius and in between 25°C and 35°C. Average relative humidity lies between 50 and 100. Rice yield were varying from 250 to 100,000 MT as gross harvest for the districts. Therefore, the standardization technique is used to scale the weather elements and corresponding rice yield values.

Prediction

We have used the linear regression to predict the rice yield values based on the weather elements. Linear regression is used to model the linear relationship between a dependent variable and one or more explanatory variable (independent variable). The dependent variable is termed as the predicted and the independent variables are called predictors. In this study, Multiple Linear Regression is used to model the linear relationship between a dependent variable (rice yield) and independent variables or predictors (rainfall, humidity, and temperature). The following liner regression equation (D.Ramesh, 2015) is used in this study:

$$Y = \lambda_1 X_1 + \lambda_2 X_2 + \lambda_3 X_3 + \lambda_0 + \epsilon \qquad (1)$$

Where, X_1, X_2, and X_3 are the rainfall, relative humidity, and temperature, respectively. Y is the predicted paddy yield value. Further, λ_1, λ_2, λ_3, are the corresponding regression coefficient, λ_0 is a constant, and ϵ is the error term.

The proposed liner regression model was trained with the collected data samples. The dataset was divided into training and testing sets with the ratio of 70% and 30 %, respectively. The effectiveness of the proposed model is measured using Root Mean Squared Error (RMSE). It is the square root of Mean Squared Error (MSE). MSE is measured as,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - Y_i)^2 \qquad (2)$$

Where n is the number of samples and Y_i and y_i are the predicted and observed values. In addition to the RMSE, we have used R-squared scale (R^2) to measure the fraction of predicted values. It is measured as,

$$R^2 = 1 - \frac{SSE}{TSS} \qquad (3)$$

Where SSE and TSS are the sum of squared error and the total sum of squares.

RESULTS AND DISCUSSION

Seventy-four weather element data and corresponding yield data have been used to train the prediction model. The remaining 24 data sample pairs are used for testing. Training and testing data were obtained from both *Yala* and *Maha* seasons with equal proportions. Figure 3 compares the predicted values with the actual values of the rice yield data.

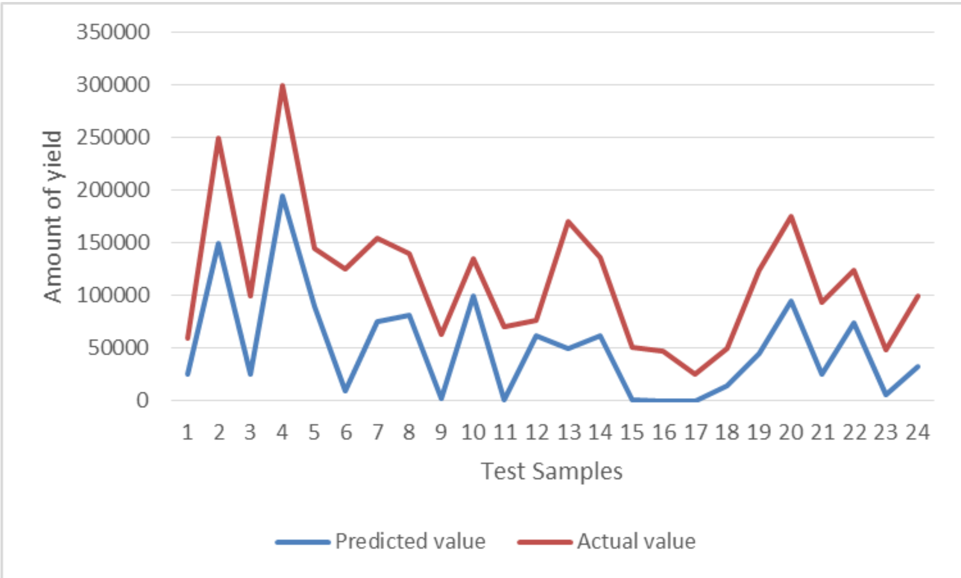

Figure 3 Comparison between the predicted and actual rice yield values

As shown in Figure 3, the proposed linear regression model predicts the corresponding rice yield values for 24 test samples. The proposed model showed 0.86 RMSE based on the prediction of rice yield values. Further, the proposed approach showed the 0.71 R^2 value based on the fraction of predicted values.

In addition to the yield prediction, the correlation between the individual weather elements and the corresponding yield value is also analysed. Figure 4 illustrates the correlation between the rice yield and rainfall.

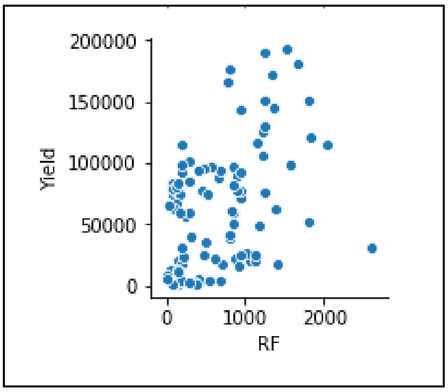

Figure 4: Rice Yield Vs Rain fall correlation.

From the plot in Figure 4, each dot represents a season. Each point's vertical position indicates the yield for the season and horizontal position indicates rainfall prevailed for the particular season. According to the plot, whenever there is an increase in rainfall, yield also shows increment. In observing the plot, it can be concluded that a tight positive correlation exits between rainfall and rice yield.

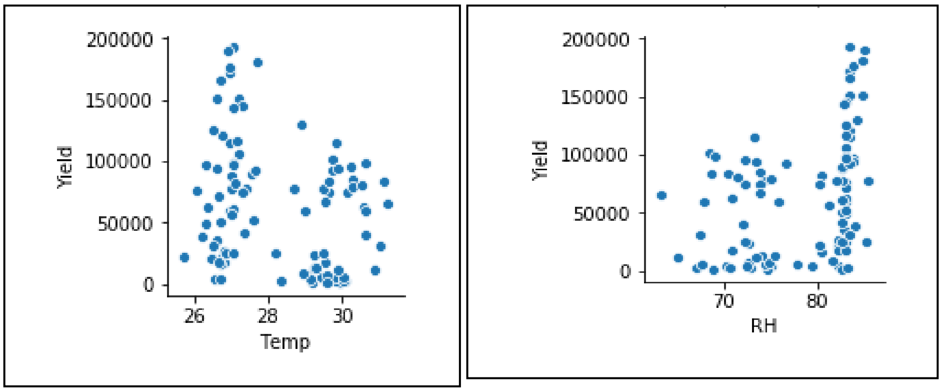

Figure 5: Left: Temperature Vs Rice yield, Right: Relative humidity Vs Rice yield

Figure 5 shows the relationship between the rice yield, temperature, and relative humidity. Yield shows peak amount when temperature lies between 26°C and 28°C. If temperature increases beyond that range, yield shows gradual decrement. Whenever the relative humidity is above 80, yield reaches optimum amount. As shown in these graphics, rainfall positively influences the paddy yield while temperature negatively influences.

CONCLUSION

This study used a model to predict the rice yield using the weather elements in Northern and Eastern provinces of Sri Lanka. Weather and yield data were collected and pre-processed. Linear regression is used to predict the yield values and RMSE and R^2metrices are used to measure the performance of the model. Based on the experimental results, proposed model able to predict the rice yield values with 0.86 RMSE. In addition, the results showed that the correlation between paddy growth and weather elements: rainfall, relative humidity, and temperature are +0.46, +0.35, and -0.25, respectively. On this basis, rainfall positively influences the paddy yield while temperature negatively influences. This prediction study might be beneficial for farmers as it predicts the future demand of rice based on the weather changes and identifies the correlation between paddy growth and weather elements in the local regions of Sri Lanka.

REFERENCES

Aishwarya Dhore, A. B. B. S. M. W., 2017. Weather prediction using the data mining Techniques. *International Research Journal of Engineering and Technology,* 4(5).

D.Ramesh, B., 2015. *ANALYSIS OF CROP YIELD PREDICTION USING DATA MINING TECHNIQUES.* India, IJRET.

Leaker, A., 1984. *An Investigation of the Factors Affecting Paddy yields.* Colombo, Sri Lanka, National Science Council.

Makoto Endo, T. T. K. H. S. K., 2009. *High Temperature Cause Male Sterility in Rice Plants with Transcriptional Alterations During Pollen Development.* s.l., PubMed.

ROBERT MENDELSOHN, M. M. S.-N. N. S., 2005. *Climate change and agriculture in Sri Lanka:a Ricardian valuation.* United Kingdom, Environment and Development Economics.

RRDI, 2021. *Rice Research & Development Institute.* [Online]
Available at:
https://doa.gov.lk/rrdi/index.php?option=com_sppagebuilder&view=page&id=42&lang=en
[Accessed 10 04 2021].

Upul Sonnadara, R. J., 2002. Forecasting the Occurrence of Rainfall in Selected Weather Stations in the Wet and Dry Zones of Sri Lanka. *Sri Lankan Journal of Physics,* 3(1), pp. 39-52.