INTERNATIONAL JOURNAL OF
INNOVATIVE RESEARCH AND KNOWLEDGE
ISSN-2213-1356 www.ijirk.com

PREDICTIVE ANALYSIS OF RICE PRODUCTION IN RELATION TO CLIMATIC PARAMETERS: A SUPERVISED MACHINE LEARNING APPROACH

Myelinda A. Baldelovar, MIT
Surigao del Sur State University Tagbina Campus
Tagbina, Surigao del Sur

Abstract
Climate change will adversely affect the agricultural production. The changes in climatic parameters such as rainfall, temperature, pressure and humidity would implicate to food security worldwide. In this study, using the five year data collected from Philippine Atmospheric Geophysical and Astronomical Services Administration (PAGASA) and Philippines Statistics Authority (PSA), the relationship between the production of irrigated rice, rainfed rice and climatic parameters such as maximum temperature, minimum temperature, rainfall, humidity and pressure was analyzed. Using Rapidminer Studio version 9.1.0, a multiple linear regression analysis was applied to the data. Results revealed that for rainfed rice production, maximum temperature, minimum temperature and rainfall are significant. Maximum temperature has a negative relationship for both irrigated and irrigated rice production. A positive relationship for both irrigated and irrigated rice production with minimum temperature was noted. It was also revealed that rainfall has a positive significant relationship with rainfed rice production.

Keywords: *climatic parameters, multiple linear regression, predictive analysis, rice production, supervised machine learning*

1. Introduction

Agriculture is an economic activity that is highly dependent upon weather and climate in order to produce the food and fiber necessary to sustain human life. High productivity growth in the agricultural sector has been a key driver of structural transformation promoting long-term economic growth. According to Gornall et al

(2010), agriculture is strongly influenced by weather and climate. While farmers are often flexible in dealing with weather and year-to-year variability, there is nevertheless a high degree of adaptation to the local climate in the form of established infrastructure, local farming practice and individual experience. Climate change can therefore be expected to impact on agriculture, potentially threatening established aspects of farming systems but also providing opportunities for improvements.

Climate change affects agriculture in a number of ways; including changes in average temperatures; rainfall and climate extremes with an important impact on soil erosion (i.e. floods, drought, etc): changes in pests and diseases, changes in atmospheric carbon dioxide, changes in the nutritional quality of some foods, changes in growing season, and changes in sea level (WorldBank, 2008). In the study of Hoffmann (2013), crop yields show a strong correlation with temperature change and with the duration of heat or cold waves, and differ based on plant maturity stages during extreme weather events. Modified precipitation patterns will enhance water scarcity and associated drought stress for crops and alter irrigation water supplies. They also reduce the predictability for farmers' planning (OECD, 2014). Tirado et al (2010) found that in an indirect way, a change in temperature and moisture levels may lead to a change in the absorption rate of fertilizers and other minerals, which determine yield output. In short, the rise in temperature along with the reduction in rainfall reduces agricultural productivity if both are beyond the threshold that is suitable for rice production. According to Ignaciuk and Mason-D'Croz (2014), climate change currently decreases the yield of maize, rice, wheat, potatoes and vegetables and continues to reduce seriously by 2050 globally.

Past studies showed that much have been done in studying the relationship between climatic parameters and production of rice using different statistical techniques but few studies were conducted applying data mining analysis using multiple linear regression technique in rainfed and irrigated rice ecosystem.

This study aims to determine the climatic parameters that are significant in the production of rice (rainfed and irrigated) in order to provide a learning agent that can aid in the farmers in making decisions to make farming more efficient and profitable through technology. The research focuses only on rice for CARAGA region in the Philippines.

2. Literature Survey

2.1 Application of machine learning in crops production

Machine learning provides many effective algorithms which can identify the input and output relationship in crop selection and yield prediction. Techniques like Artificial neural networks, K-nearest neighbors and Decision Trees have carved a niche for themselves in the context of crop selection which is based on various factors. Crop selection based on the effect of natural calamities like famines has been done by Washington et al (2011) based on machine learning. The use of artificial neural networks to choose the crops based on soil and climate has been shown by researchers Snehal et. al (2014).

2.2 Machine learning techniques used in predicting crop production in relation to climate change

Dahikar and Rode (2014) used the artificial neural network to predict the crop by using the soil parameters such as types of soil, pH, nitrogen, phosphate, potassium, organic carbon, calcium, magnesium, sulphur, manganese, copper, iron, depth and climate parameters such as temperature, rainfall, humidity. In conducting the experiment, Cotton, Sugarcane, Jawar, Bajara, Soyabean, Corn, Wheat, Rice and Groundnut crops were used. Rossanna et al (2013) concluded that climate related variables were not the main determinants of corn yield, rather yield was greatly affected by planting practices, particularly by the application right amount of fertilization. Priya and Suresh (2009) developed a model for forecasting the yield of the sugarcane in Coimbatore district by using the fortnightly weather variable such as average daily maximum and minimum temperature, relative humidity in the morning and evening and total fortnightly rainfall and the yield data.

Their forecast model was able to explain 87% of variation in the sugarcane yield and they conclude that the sugarcane yield can be forecasted using the regression technique successfully two months before harvest.

Rosa et al (1981) developed a regression models and techniques to predict the response variable that is yield and the explanatory variables such as weather, soil properties. Kaspar (2003) found that parametric regression model is used in many of the yield forecasting method. According to Shibayama (1991), the commonly used models are linear regression models. Wilcox (2001) found the use of polynomial regression models.

Yiqun Gu et al (1994) formulated a belief network to assess the effect of climate change on potato production was formulated. They have shown a belief network combining the uncertainty of future climate change, considering the variability of current weather parameters such as temperature, radiation, rainfall and the knowledge about potato development. They thought that their network give support for policy makers in agriculture. They test their model by using synthetic weather scenarios and then the results are compared with the conventional mathematical model and conclude that the efficiency is more for the belief network.

Hong-Ying et al (2012) described the new concept of crop yield under average climate conditions and used the time series techniques on the past yield data to set up a forecasting model. They conclude that the moving average model is regarded as the potential yield forecasting model. The strong point is it needed a relatively small amount of data.

3. Methodology

3.1 Data gathering Procedure

The data utilized in this research are secondary data which consist of climate parameters, rice production. Rice production (irrigated and rainfed) data were collected from Philippine Statistics Authority (PSA) and Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) for climatic data. Prior to the collection of data, the researcher seeks an approval from higher authorities in every department through a letter.

3.2 Data Selection

The climatic parameters considered for analysis are average of the minimum rainfall, average of the maximum rainfall, pressure, temperature and humidity. Rice production data include crops' type/ecosystem (rainfed or irrigated), geolocation, year, period and production in metric ton.

3.3 Data Preprocessing

Data preprocessing is an important step in data mining since raw data are usually susceptible to missing values, noisy data, incomplete data, inconsistent data and outlier data. This is done to enhance data efficiency. Data preprocessing deals with data preparation and transformation of the dataset and seeks at the same time to make knowledge discovery more efficient. Preprocessing include several techniques like cleaning, integration, transformation, and reduction. It aims to reduce the data size, find the relation between data, normalize data, remove outliers and extract features for data.

3.4 Data Analysis

Multiple Linear Regression analysis was utilized using Rapidminer data mining application program as a tool for determining relevant knowledge. MLR was used to determine the relationship between climatic parameters and rice and corn production. It is a technique used for numerical prediction. It is a statistical measure that attempts to determine the strength of the relationship between one dependent variable and a series of other changing variables known as independent variables.

4. Results and Discussions

Tables presented below are the results after performing linear regression analysis using Rapidminer Studio.

Table 1. MLR Analysis for Irrigated Rice Production

Attribute	Coefficient	t-Stat	p-Value
Humidity	8107.3	1.005	0.329
Max	-47204.1	-2.394	**0.028**
Min	42.8	3.245	**0.005**
Period	140146.7	0.745	0.466
Pressure	7666.8	0.603	0.554
Rainfall	-214.3	-1.988	0.063

As shown in Table 1, maximum temperature and minimum temperature are significant in the production of rice that utilized the irrigation system. The negative coefficient of maximum temperature implies that there is a significant negative relationship between maximum temperature and irrigated rice production which means that it can cause a significant reduction on the production of irrigated rice. On the other hand, coefficient of minimum temperature implies a positive effect in the irrigated rice production.

Table 2. MLR Analysis for Rainfed Rice Production

Attribute	Coefficient	t-Stat	p-Value
Max	-36949.28	-2.420	**0.027**
Min	127176.99	3.996	**0.001**
Humidity	6293.05	1.186	0.252
Rainfall	-173.75	-2.331	**0.032**

Table 2 above revealed that for rainfed rice production, maximum temperature, minimum temperature, and rainfall are significant. This means that these climatic parameters are important factors in the production of rainfed rice. The negative coefficient of maximum temperature indicates a negative relationship between rainfed rice production and maximum temperature. Conversely, rainfall and minimum temperature coefficients signify positive relationship between rainfed rice production and rainfall and minimum temperature.

5. Conclusion

Results of the study revealed that for irrigated rice production, maximum temperature and minimum temperature are significant. For rainfed rice production, maximum temperature, minimum temperature and rainfall are significant. Maximum temperature has a negative relationship for both irrigated and irrigated rice production. A positive relationship for both irrigated and irrigated rice production with minimum temperature was noted. It was also revealed that rainfall has a positive significant relationship with rainfed rice production.

References

[1] Atkinson, R. C., and Geiser, S. 2009. Reflections on a century of college admissions tests. Educational Researcher 38, 9, 665-676.

[2] Dahikar MSS, Rode SV. Agricultural crop yield prediction using artificial neural network approach. International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE). 2014 Jan; 2(1):1–4.

[3] Horie T, Yajima M, Nakagawa H. Yield Forecasting. Agricultural Systems. 1992; 40(1):211–36.

[4] Priya SRK, Suresh KK. A study on pre-harvest forecast of sugarcane yield using climatic variables, Statistics and Applications. 2009; 8(2):1–8.

[5] Rossana MC, Leon D, Rex EL, Jalao J. A Prediction Model Framework for Crop Yield Prediction. Asia Pacific Industrial Engineering and Management Society Conference Proceedings Cebu, Phillipines. 2013. p. 185.

[6] Rosa DLD, Cardona F, Almorza J. Crop Yield Prediction Based on Properties of Soils in Sellilla, Spain, Geoderma. 1981; 25(3-4):267–74.

[7] 33. Kaspar TC. Relationship between six years of corn yields and terrain attributes. Precision Agriculture. 2003; 4(1):87– 101.

[8] 34. Shibayama M. Estimating grain yield of maturing rice canopies using high spectral resolution reflectance measurements. Remote Sensing of Environment. 1991; 36(1):45–53.

[9] 35. Wilcox A. Factors Affecting the Yield of Winter Cereals in Crop Margins. Journal of Agricultural Science. 2000; 135(4):335–46.

[10] 36. House CC. Forecasting Corn Yields: A comparison Study using Missouri Data, Statistical Research Division, United States Department of Agriculture. 1979; 17(16):3189–200.

[11] Yiqun Gu Y, James W, McNicol M. An Application of Belief Networks to Future Crop Production. IEEE Conference on Artificial Intelligence for Applications, San Antonia, TX. 1994. p. 305–9.

[12] Hong-Ying L, Yan-Lin H, Yong-Juan Y, Hui-Ming Z. Crop yield forecasted model based on time series techniques. Journal of Northeast Agricultural University (English Edition). 2012; 19(1):73–7.