Weather Analysis to Predict Rice Cultivation Time Using Multiple Linear Regression to Escalate Farmer’s Exchange Rate

Luminto
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
luminto_luhur@hotmail.com

Harlili
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
harlili@informatika.org

***Abstract*—Agriculture is one of primary sectors of the national economy and is receiving more attention from government annually in order to increase productions and boost national economy. Agriculture, especially rice cultivation, has been challenged with various issues for the past decades such as extreme weather (global warming) which could result in crop failure. From the weather aspect, this paper aims to build weather analysis program to predict rice cultivation time in hope to escalate Farmer’s Exchange Rate (FER) . Farmer’s Exchange Rate is an proxy indicator to determine how prosperous farmers from certain regions are. Weather analysis is conducted by retrieving weather data from National Weather Forecast and Farmer’s Exchange Rate data from National Statistics Authority for the past 1 year and using the obtained data to build a regression model using Multiple Linear Regreesion (MLR) to determine the correlation between weather and FER. The variables are “Average Temperature”, “Average Humidity”, “Rainfall”, and “Solar Radiation”. The resulted model is then projected using line chart. Based on evaluation the proposed analysis from 2 different regions tested gives overall Root Mean Square Error (RMSE) between 0.39 – 1.34.**

Keywords—weather analysis, Farmer’s Exchange Rate, Multiple Linear Regression

I. Introduction

Weather is the state of the atmosphere, to the degree that it is hot or cold, wet or dry, calm or stormy, clear or cloudy. Weather refers to day-by-day temperature and precipitation activity, whereas climate is the term for the averaging of atmospheric conditions over longer periods of time [1]. Climate change and agriculture are interrelated processes, both of which take place on a global scale. Climate change affects agriculture in a number of ways, including through changes in average temperatures, rainfall, and heat waves. It will likely negatively affect crop production in low latitude countries and increase the risk of food insecurity for some vulnerable groups, such as the poor.

One of the highest worldwide production agricultural commodity is rice. Rice is the seed of the grass species *Orzya Sativa* (Asian rice) or *Oryza glaberrima* (African rice). Rice is the most widely consumed food for a large part of the world’s human population, especially in Asia. For rice-growing environments, there are biotic and abiotic factors. Biotic factors are living factors such as pests and diseases, meanwhile abiotic factors are non-living factors i.e. weather. This paper only focuses on abiotic factors especially weather that impact rice-growing environments. The weather abiotic factors are temperature, humidity, rainfall, and solar radiation.

Multiple linear regression is a useful approach to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. There are two main advantage of using multiple linear regression model to analyze data. First, multiple linear regression has the ability to determine the relative influence of one or more predictor variables to the criterion value. Second, it has the ability to identify outliers or anomalies.

In this paper, we propose a model of rice cultivation time prediction with weather data from National Weather Forecast using 4 variables, Average Temperature, Average Humidity, Rainfall and Solar Radiation. These variables are abiotic factors related to weather in rice-growing environments. We implement this into a software named Cultivation Time Predictor. Our work will predict the best cultivation time for farmers to plant rice for each season in 2 seasons regions, and provide information on the predicted Farmer’s Exchange Rate in comparison with the actual value, and the error value between them. Farmer’s Exchange Rate is an proxy indicator to determine how prosperous farmers from certain regions are. The Farmer’s Exchange Rate data will be retrieved from National Statistics Authority.

This paper is organized as follows; in section 2, related work on this research. Our proposed method will be explain thoroughly in section 3. Section 4 will focus on how we conducted our experiments and section 5 will conclude our experiments analysis. Finally, section 6 will wrap up our work described in this paper as we will conclude everything up in this section.

II. Related Work

V. Sellam and E. Poovammal [2] conducted experiment to predict crop yield using regression analysis on agriculture in India. They use linear regression by computing the linear regression model, compute the residual values and obtain the R^2.

978-1-5386-3001-3/17/$31.00 ©2017 IEEE

They use variables such as Annual Rainfall, Area under Cultivation (Million Hectare), Food Price Index (FPI) to determine the relationship between them and yield of rice crop. They also use rice production data in India from year 2000-2011 provided by Department of Statistics and Agriculture, National Informatic Centre. The result clearly states that Annual Rainfall, Area under Cultivation, and Food Price Index have an average of 70% influence in the crop yield.

Anjela Diana Corraya and Sonia Corraya [3] conducted experiment on predicting yield of rice and price in Bangladesh. They use Weighted Linear Regression Based method to do the prediction. The system takes data from past years as inputs and uses Weighted Linear Regression model to produce two graphs representing both old and predicted price and yield for any particular crop. Experimental results show that the method is quite successful in predicting price and yield considering inflation factor, with the performance accuracy for predicting price and yield value are 78.75% and 83.55% respectively. This system is believed to help farmers know about the future prediction beforehand so they can prepare for any kind of unwanted disaster. The proposed method can be applied not only for rice, but also for another agricultural commodity in order to predict their yield and price.

Niketa Gandhi, Owaiz Petkar, Leisa J. Armstrong, and Amiya Kumar Tripathy [4] condcuted experiment on predicting rice crop yield in India using Support Vector Machines (SVM). All the dataset used in this research were sourced from the openly accessible records of the Indian Government. This was sourced for the years 1998-2002 for the Kharif season of rice production. The parameters selected for the present study were precipitation, minimum, average and maximum temperature, reference crop evapotranspiration, area, and production. The algorithm achieved the accuracy of 78.76%, sensitivity of 68.17%, and specificity of 83.97%. They achieved a score of F1 0.69, in which was computed to measure the test's accuracy. Meanwhile, error results of the classifier are mean absolute error of 0.23, root mean squared error of 0.39, relative absolute error of 67.38%, and root relative squared error of 82.51%. The experimental results showed that the other classifiers such as Naïve Bayes, BayesNet, and Multilayer Perceptron performed better by achieving the highest accuracy, sensitivity, and specificity compared to SMO classifier.

Saisunee Jabjone and Chatchai Jiamrum [5] also conducting experiment on predicting rice yield in Phimai District of Thailand using Artificial Neural Network. They designed the artificial neural networks by using six meteorological factors : rainfall, evapotranspiration, temperature, humidity, water distribution, and wind speed. The monthly dataset during 2002 to 2007 of Phimai district was used to train the model and used to predict the rice yield in 2008 to 2012. The result showed that the BP neural computing technique could be employed successfully in ANN. Empirical rice yield predicting model produced consistently higher R^2 (0.99) and lower RMSE value (9.94) than linear regression-based yield models.

III. PROPOSED METHOD

Multiple Linear Regression is a statistical process for estimating the relationship among multiple independent variables with dependent variables. The multiple linear regression equation then will be

$$yi=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4 \qquad (1)$$

Where y_i is the dependent variable (Farmer's Exchange Rate), β_0 is the constant, β_1 is the coefficient for variable X_1 (Avg. Temperature), β_2 coefficient for X_2 (Avg. Humidity), β_3 coefficient for variable X_3 (Rainfall), and β_4 coefficient for variable X_4 (Solar Radiation) [4][5][6].

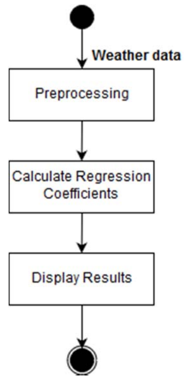

Figure 1. Proposed Process Flow

The proposed process flow can be seen at Figure 1. We obtained our weather data from http://dataonline.bmkg.go.id/, and Farmer's Exchange Rate monthly data for each regions from http://www.bps.go.id/index.php/publikasi. The regions used in this experiment are Kabupaten Majalengka, Jawa Barat, Indonesia and Kabupaten Badung, Bali, Indonesia. Both of this region have weather data available from National Weather Forecast. Obtained weather data from 2016 year is stored in CSV (comma separated files) type file. Meanwhile, Farmer's Exchange Rate data is stored in database. After the weather data file is loaded, it will go through preprocessing, in which all data will be loaded and parsed into its own variables, and calculate the average of each variables per month. The next step is to separate data training and data testing. Separation of data training and testing is based on season. The tested region has two season; the dry season (April to October), and the monsoon season (November to March). We use data from Kabupaten Majalengka, Jawa Barat for the first and second iteration and Kabupaten Badung, Bali for the third and forth iteration. We use the dry season data to train the model and monsoon season data to test the model for odd iteration, and vice-versa for the even iteration. Afterwards, for each iteration, it will calculate regression coefficients using multiple linear regression, and after the coefficients are estimated, it will predict the Farmer's Exchange Rate using testing data. Finally, it will be displayed as the predicted vs. actual Farmer's Exchange Rate chart on the testing season.

To test the accuracy, we used the actual Farmer's Exchange Rate actual records that have already been stored in our database. Using the predicted results earlier, each month of the testing season is compared to the actual result to count the accuracy using Root Mean Square Error (RMSE). The RMSE equation will be

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (2)$$

IV. EXPERIMENT

Experiment is conducted to get the lowest RMSE value possible on a certain region. To get this purpose, we used 2016/2017 weather data as our training and testing data for our experiment. It is to determine whether the training data on a certain region has any effect on the accuracy result. We choose districts with highest crop productivity in their respective province. We conducted 4 experiments, which shown below on Table I.

Table I. Experiment Data

Num.	Training Data	Testing Data
1	Dry season of 2016/2017 at Majalengka District, Jawa Barat, Indonesia	Monsoon season of 2016/2017 at Majalengka District, Jawa Barat, Indonesia
2.	Monsoon season of 2016/2017 at Majalengka District, Jawa Barat, Indonesia	Dry season of 2016/2017 at Majalengka District, Jawa Barat, Indonesia
3	Dry season of 2016/2017 at Badung District, Bali, Indonesia	Monsoon season of 2016/2017 at Badung District, Bali, Indonesia
4	Monsoon season of 2016/2017 at Badung District, Bali, Indonesia	Dry season of 2016/2017 at Badung District, Bali, Indonesia

Each of the experiments produced a different set of coeffcients and RMSE value. The experiment result can be seen at Table II.

Table II. Experiment Result

Num.	Variables	Coeff.	Month with highest actual FER	Month with highest predicted FER
1	Constant	37.61	November	November
	AT	2.23		
	AH	-		
	RF	-		
	SR	0.19		
2	Constant	-124.5	April	April
	AT	5.67		
	AH	0.81		
	RF	-		
	SR	-0.01		
3	Constant	87.23	October	October
	AT	0.191		
	AH	-		
	RF	0.029		
	SR	0.467		
4	Constant	26.1	May	April
	AT	1.68		
	AH	0.29		
	RF	-0.08		
	SR	-		

AT is short for Averge Temperature, AH for Average Humidity, RF for Rainfall and SR for Solar Radiation.

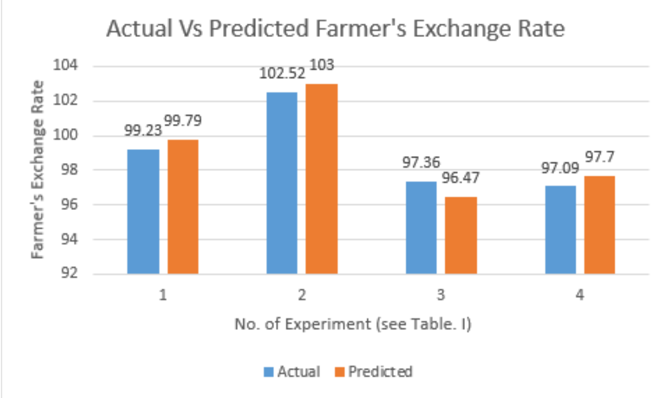

Figure 2. Actual Vs Predicted Farmer's Exchange Rate Comparison

Experiment 4 has a difference in peak time of Farmer's Exchange Rate, in which the actual peak time was in May, while the predicted peak time results in April. The other experiments give the same peak time as the actual time; November for experiment 1, April for experiment 2, and October for experiment 3. All of the experiments result in slight difference value between actual and the predicted rate as you can seen in Figure 2. Only experiment 2 results in surplus FER, while the other 3 experiments, gives deficit FER.

Rice cultivation predicted time is determined by substracting rice harvest time (3 months) from month with highest predicted Farmer's Exchange Rate because we would like to suggest farmer to cultive rice at certain time will give the best result 3 months later. On the other hand, the root mean square error is determined by rooting the mean square of the difference between the actual and the predicted Farmer's Exchange Rate. The experiment's root mean square error and the cultivation time table can be seen at Table III.

Table III. Cultivation Time

Num.	District	Season	Highest FER	Cultivation Time	RMSE
1.	Majalengka	Monsoon	November	August	1.17
2.	Majalengka	Dry	April	January	1.15
3.	Badung	Monsoon	October	July	1.03
4.	Badung	Dry	April	January	1.34

V. RESULT ANALYSIS

For experiment no. 2 and no. 4, they produce the same cultivation time of dry season, which is January. It means that both Kabupaten Majalengka and Kabupaten Badung will produce the highest Farmer's Exchange Rate on April and farmers are highly suggested to start farming on January. Meanwhile, for the monsoon season, Kabupaten Majalengka produce peak FER value on November and Kabupaten Badung on October. Therefore, farmers at Kabupaten Majalengka will be advised to start farming on August, and July for Kabupaten Badung. All of these result in very low RMSE value.

As can be seen from Table II, the coefficient that produces the highest FER is always variable AT. It is marked as the significant variable, the factor that determines how well the rice-growing environment are. The variable AH only affects the prediction if the training data is monsoon season data. On the other hand, the variable SR only affects the predicition if the training data is dry season data. While the variable RF has slight affection toward the prediction simply because the imbalanced and invalid data in RF.

However, when we experimented using all of the variables for the prediction, the results are not as good as using just some of the variables. They tend to give a bigger RMSE value and produce an inconsistency rate value. They seem to produce relatively bigger error compare to the other parameter combinations.

VI. CONCLUSION

In this research, we have built a multiple linear regression model to predict rice cultivation time that results the highest Farmer's Exchange Rate of season 2016/2017 at 2 seasons regions. We also conclude that Average Temperature and Solar Radiation are the significant variables, but the prediction cannot be done using only these two variables. Prediction can be done on certain regions by simply testing out all of the variables combinations that result in low RMSE value. Multiple linear regression is a suitable method for predicting problem that based on multiple dependent variables and it is easy to be implemented and has a faster performance compared to machine learning techniques.

For the near future works, we could develop this into further uses and more region can be tested. A prediction that not only focus on rice but also another agricultural commodity such as sugarcane and maize, or it could be building a more highly accurate crop calendar for certain region that have high productivity on certain commodity.

ACKNOWLEDGMENT

We would like to thank Dr. techn Wikan Danar Sunindiyo and Dicky Prima Satya, S.T., M.T. for the feedbacks and improvements towards our research development. We would also like to thank our colleagues that gives us motivation to do this research.

REFERENCES

[1] G. James, D. Witten, T. Hastie, and R. Tibshirani, "An Introduction to Statistical Learning with Applications in R," Springer New York Heidelberg Dordrecht London, 2013, pp. 71-75.

[2] V. Sellam, E. Poovammal, "Prediction of Crop Yield using Regression Analysis," in *Indian Journal of Science and Technology*, Vol. 9. Kattankulathur: SRM University, Oct. 2016, pp. 1-5.

[3] Anjela Diana Corraya, Sonia Corraya, "Regression based Price and Yield Prediction of Agricultural Crop," in *International Journal of Computer Applications (0975-8887)*, Vol. 152, No. 5, Bangladesh, Oct. 2016.

[4] N. Gandhi, O. Petkar, Leisa J. Armstrong, Amiya Kumar Tripathy, "Rice Crop Yield Prediction in India using Support Vector Machines," in *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Jul. 2016.

[5] S. Jabjone, C. Jiamrum, "Artificial Neural Networks for Predicting the Rice Yield in Phimai District of Thailand," in *International Journal of Electrical Energy*, Vol. 1, No. 3, Nakhon Ratchasima, Sept. 2013.

[6] Willy H. Verheye, "Soils, Plant Growth and Crop Production", Eolls Publishers Co.Ltd, vol 2, 2010. Transl. J. Magn. Japan, vol. 2, pp. 740-741, Aug. 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].