# FORECASTING RICE YIELDS VIA SUPPORT VECTOR MACHINES AND LINEAR REGRESSION MODELS WITH CLIMATE AS PREDICTOR

Rizza G. Baltazar, Graduate School, Angeles University Foundation, Philippines, rgbaltazar@psau.edu.ph
Rosanna A. Esquivel, DIT, College of Computer Studies, Angeles University Foundation
Angeles City, Philippines, esquivel.rosanna@auf.edu.ph

## ABSTRACT-

**Changes in the climate in the next decades will have an impact on rice yield crops. It is necessary to forecast rice yield in terms of climatological factors since these factors may have positive or negative effects on rice production. A temperature rise will allow more rice production to occur in other regions of the Philippines or growing more than two rice yield harvests per year. According to researches, most of the over-all impacts of climate change is likely to be negative. Rice farming requires water. Rainless days can significantly reduce rice yields. Crop management, use of technology and government support will improve varieties of rice that are ready for the climate change. However, before the implementation of any crop management or intervention to improve rice production, validated forecasting models should be the basis. The climatological factors used as predictors in this study are pressure, temperature, relative humidity, rainfall, wind, cloudness, and sunshine from the Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA). The rice yield reports came from the Department of Agriculture (DA). The predictors and the output datasets used in this study are from Region III and inclusive of the years 2009 to 2017 as training data set. This study focused on the development of rice yield forecasting models using Support Vector Machines (SVM) and Linear Regression methods. The models were validated for accuracy using Error Rates using the 2018 actual data as test data set.**

**Keywords-** Support Vector Models, Linear Regression, Forecasting, Rice Yield, Climatologic factors

## I. INTRODUCTION

Rice is a primary food source for more than three billion people in the world. It is an important crop yield in most Asian countries like the Philippines. In rice consuming or producing countries, the major issues in rice demand and its production are population growth and climate change. Forecasting of rice yield is crucial, especially in regions with climatological uncertainties. It enables government to come up with strategic contingency plans for the redistribution of rice during times of shortage [1].

Rice yield modeling is an important tool in modern agricultural research [2]. Climatological factors can have effects on rice growth, development and yield. The relationship between climatological factors and rice production is an important issue since the world's rice production resources are already under pressure due to rapidly increasing population. Climatological factors can also affect land use patterns. Good understanding of the changes in climatological factors and changes on the growth and development of rice crops are essential. The impact of climatological factors on rice production is of particular importance because it is a food source especially in in Asia [3]. It was estimated that about 60% increase in rice yield would be likely needed by 2020 due to population growth. The rice consumption in the global market is projected to reach 873 million tons by 2030 [4]. Thus, immediate attention to climatological factors in rice yield is urgent as it poses threat to food supplies and security.

Reports and researches showed that variations in rice yields are typically related to climatological factors and weather patterns [5]. Rice yield is significantly influenced by climatological factors such as rainfall, temperature, sunshine and relative humidity. The impact of rainfall variations in rice yield had been emphasized [6]. Increase in temperature had also been found to reduce rice yields and quality. Increased temperature shortens the life cycle of rice crops resulting in shorter grain filling time and thus producing smaller and lighter grains leading to lesser rice yields and poorer quality. Higher temperature decreases rice yield because it increases rice plant respiration rate and decrease the net photosynthesis. As a consequence, higher temperature lowers biomass production and yield [7].

Sunshine determines solar radiation that affects photosynthesis, carbon uptake, heat balance of agricultural land. It then affects temperatures of soil and air which had been identified as major factors in regulating rice crop development. Any significant change in sunshine will be of major impact in climate change and rice crop production. Studies of [8] have concluded significant influence of sunshine hours on rice yields.

Existing studies showed that rainfall, temperature and relative humidity are some of the factors that affect rice yield. Jeong, et.al. analyzed the patterns in climatologic variables and rice yield and observed a generally reducing trend in rice yield [9]. Most of the researches for increased rice production are focused on genetic modification and soil management. There is however a need to supplement such existing researches to provide information on climatological factors that may help in higher rice yield [10]. Historical data analysis between rice yield and climatological factors will provide such information. Cause-effect relationships are expected to exhibit non-linearity. This will require machine learning techniques. Moreover, works similar in the said field and in rice yield in particular are currently limited in the Philippines and thus this research is geared in this direction.

Developing forecasting models had encountered its bottlenecks. With the prevalent use of non-linear and linear statistical theory, particularly in the field of machine learning, breakthroughs had been made since the 1990s [11]. The use of artificial intelligence has undergone various developments and applications through the use of computer iterative algorithms such as Support vector machine (SVM) and linear regression (LR). Through the utilization of good sparsity, ability to fit small samples and global optimization, SVM and LR have outperformed other statistical models [12]. The non-linearity of SVM had made it a choice among researchers to be applied in agricultural production such as remote sensing, moisture prediction, plant and disease warning.

Various models and methods have been used in attempt to understand the impact of climatologic factors in rice yield. The aim of this study is to apply support vector machines and linear regression models in forecasting rice yields with climatologic factors as predictors.

### A. Statement of the Problem

The study aimed to provide a validated set of forecasting rice yields models with climatologic factors using support vector machine and linear regression.

Specifically, this study aimed to answer the following:

1. What is the demographic profile of the locale in terms of climatological parameters and rice yield?
2. What data cleaning protocols were used to detect and remove corrupt or inaccurate records from the data sets?
3. How to apply support vector machine model in developing rice yield forecasting model?
4. How to use linear regression in forecasting rice yields?
5. How to evaluate the accuracy of the developed rice yield forecasting models?

### B. Scope and Limitations of the Study

The study is focused on providing a validated set of models in forecasting rice yields through climatologic factors. The first paradigm in this study described Region III in terms of climatological parameters and rice yields for the past ten (10) years. The climatologic reports from PAGASA situated in Clark Air Base Zone on factors such as pressure, temperature, relative humidity, rainfall, wind, cloudiness and sunshine in Region 3 were considered in this study. The said reports detailed the input parameters in this study and these are the climatologic factors from 2008 to 2018. The rice yield reports from 2009 to 2018 came from the Regional Field Office III of the Department of Agriculture. Region III will be described in terms of the climatologic factors and rice yield reports from 2009 to 2018. Using Albeit data preprocessing for data cleaning, the researchers conducted collecting the data, cleaning the data, analyzing the data, and publishing the results to the locale's representatives. SMOreg implements the support vector machine for regression was generated through Weka 3.8 to determine the initial weight for each of the climatologic predictors which served as inputs. In adjusting the gradient weights, Microsoft Excel's Generalized Reduced Gradient (GRG) Nonlinear in Solver was used. In the SVM model, the model had an adjusted weight of all those parameters under study. In the linear regression model, the same data sets used in SVM was used to generate the model using Weka. The study is limited to the data for Region III pertaining to the recently mentioned irrigated rice yield and climatologic factors.

### C. Conceptual Framework

The researcher used the framework below in conducting the study. It was anchored on The Knowledge Discovery Databases (KDD) model.

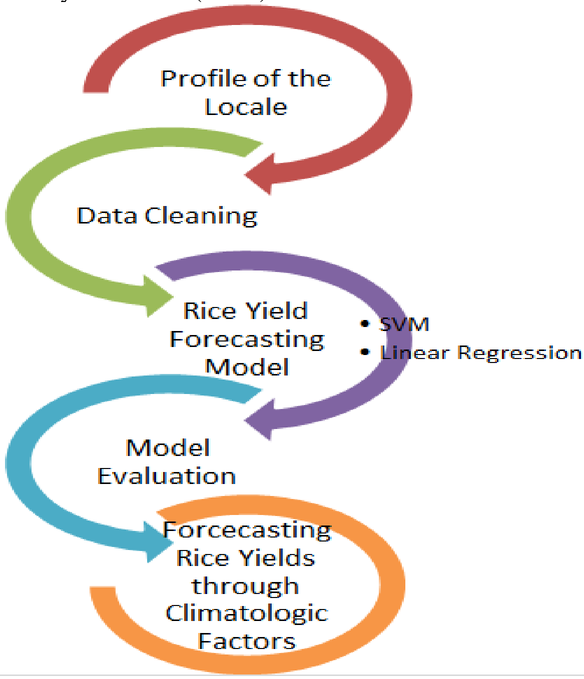



Figure 1. Conceptual Framework

As depicted in figure 1, the researcher presented all the gathered demographic profile of Region III in terms of irrigated rice yield and the climatologic factors. Then, data cleaning tasks were performed to prepare the data sets for computations. Using Weka, initial weights were determined and the SVM was constructed. Then, another model was built using linear regression. Mean absolute percentage error (MAPE) and Mean Absolute Deviation (MAD) were used to compute for the most accurate forecast. The models were compared and results of the study were discussed and presented.

## II. METHODOLOGY

The data sets from DA and PAGASA were used to describe Region III through graphs. Changes in climactic factors over the past 10 years were noted as well as the irrigated rice yield variations were presented. The the data sets were subjected to data cleaning tasks in prepartion for computations and analysis. Using Weka 3.8, the SMOreg and linear regression model were determined. Running the data sets in Weka was necessary to determine the initial weight for computation the SVM implementation.

Using the inputs namely: pressure, temperature, relative humidity, rainfall, wind, cloudness and sunshine and the weights yielded in Weka, Class 1 for SVM was computed.

Considering that every class calculation in SVM is similar, the delta weight (Δw) between the input variables and the weights were assumed as $w_1$ to $w_n$.

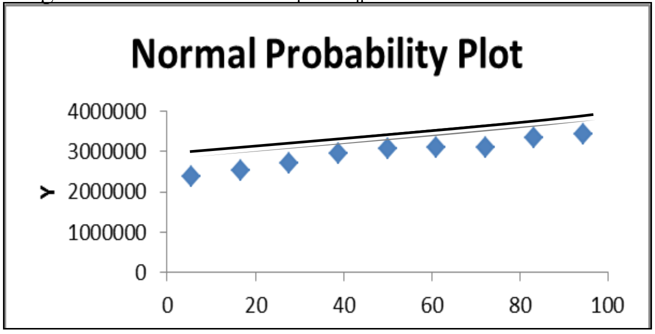



Figure 2. Rice Regression Line of Probability Plot

The SVM's vector [13] can be thought of as a directed line segment as shown in Figure 2. A vector has a displacement from one point P (head) to another point Q (tail). A vector starts from the origin called a position vector. Another important property that will be relevant in our discussion is thatThen, multiply vector w and b by a constant (scalar multiplication) in the normal vector in the original direction.

To separate the plane: w*x+b= 0. The decision rule for the model will be: $Y_i$=sign(w*$x_i$+b). Use this and compute the direction of the normal vector by dividing it by its norm and then multiply it by delta to gives us a vector of length delta and pointing in the direction of the norm. Then, subtract this value from A to arrive at point B on our hyperplane. Note that the researcher tried to to maximize function since it is only a function of alphas. This is the so-called Dual formulation of the optimization problem ans resolved in this study through the MS Excel Solver using the Max values of the gradient function. The Max value is computed using:

$$\text{Max } L(\alpha)=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i x_j \cdot \text{ where } \alpha_i \geq 0 \ \& \ \sum_{i=1}^{N}\alpha_i y_i = 0$$

The linear regression model [14] is the relationship between two different variables, called the dependent and independent variables. When the relative relationship between the two variables is calculated, a linear regression model to forecast the rice yield can be formulated. y = bx + a is the formula for a linear regression. The "y" is the value to forecast, the "b" is the slope of the regression line, the "x" is the value of the independent value, and the "a" represents the y-intercept. The regression equation simply describes the relationship between the dependent variable (y) and the independent variable (x).

# III. RESULTS

The following section presents the results of analyzing the historical data on climatologic factors and rice yield towards the development of various forecasting models using ANN.

## A. *Demographic Profile*

Data on the reports from PAGASA on climatologic factors from 2008 to 2018 particularly pressure, temperature, relative humidity, rainfall, wind, cloudiness and sunshine in Region 3 are presented in the next section.

The irrigated rice yield reports from 2009 to 2018 from DA were also used to describe Region III.

Table 1 shows the historical data sets of the climatologic factors used in this study. The recorded daily climatologic factors were summed up to get the monthly and then the yearly average values of each of the climatologic factors. X0 to X6 represents pressure, temperature, relative humidity, rainfall, wind, cloudiness and sunshine respectively for the 11-year periods of 2008 to 2018.

TABLE 1. REGION III DEMOGRAPHICS

| Year | X0 | X1 | X2 | X3 | X4 | X5 | X6 |
|---|---|---|---|---|---|---|---|
| 2008 | 993.7 | 31.04 | 74.54 | 155.55 | 2.66 | 6.75 | 5.83 |
| 2009 | 993.27 | 30.89 | 74.92 | 239.74 | 2.65 | 6.79 | 5.53 |
| 2010 | 994.22 | 32.12 | 71.31 | 133.58 | 2.65 | 6.51 | 6.64 |
| 2011 | 993.02 | 30.83 | 74.58 | 203.33 | 2.39 | 6.81 | 5.42 |
| 2012 | 992.68 | 31.35 | 74.50 | 219.78 | 2.48 | 6.56 | 5.83 |
| 2013 | 993.08 | 31.44 | 75.29 | 182.05 | 2.35 | 6.60 | 5.91 |
| 2014 | 993.53 | 31.23 | 74.30 | 140.90 | 2.41 | 6.42 | 6.33 |
| 2015 | 994.37 | 31.45 | 74.67 | 181.70 | 2.48 | 6.23 | 7.01 |
| 2016 | 993.89 | 32.04 | 78.39 | 157.56 | 2.38 | 6.30 | 6.33 |
| 2017 | 993.45 | 32.03 | 73.16 | 144.23 | 2.44 | 6.41 | 6.19 |
| 2018 | 993.16 | 31.77 | 73.41 | 229.74 | 2.37 | 6.23 | 6.29 |

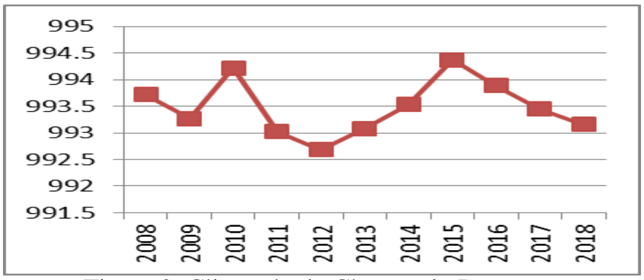



Figure 3. Climatologic Changes in Pressure

Figure 3 depicts the changes in pressure in Region III as recorded by PAGASA. High air pressure produces clear sky, dry and stable weather. In a low pressure zone, wind is circulated inwards and upwards rapidly. As a result, air rises and cools; clouds and precipitate are formed. Low air pressure produces unstable weather conditions like rain or storms ???. The observed air pressure in Region III shows sudden upsurges in the year 2010 and after (5) years in 2015. The lowest was observed in the year 2012.

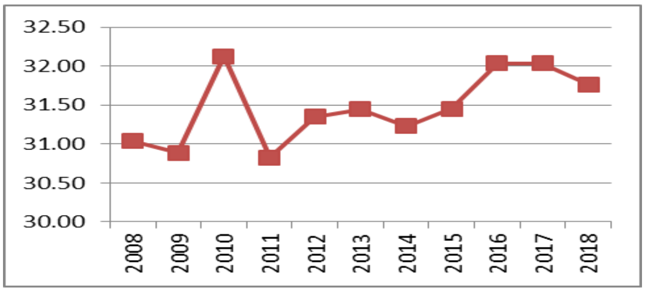



Figure 4. Climatologic Changes in Temperature

Figure 4 depicts the changes in temperature over the past 11 years. Temperature changes could make it too hot to grow rice crops, and droughts caused by such change could reduce the amount of water available for irrigation. It is also likely to cause stronger storms and more floods. Higher temperatures could help some kinds of weeds and pests to spread to new areas ???. The chart shows an upsurge in temperature in 2010. It was also noted that the temperature had not gone lower than 31 degree celcius since 2014.

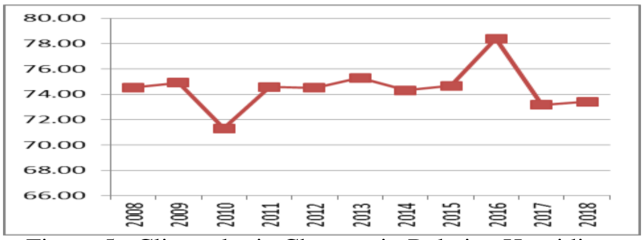



Figure 5. Climatologic Changes in Relative Humidity

Figure 5 graphically shows the changes in relative humidity of Region III from 2008 to 2018. When water vapor stays the same while the temperature decreases, relative humidity increases. When water vapor remains the same and the temperature increases, the relative humidity decreases. The reason behind this is when that colder air does not require as much moisture to become saturated as warmer air ???. The relative humidity in Region III is relatively the same except for the lowest in 2010 and highest in 2016.

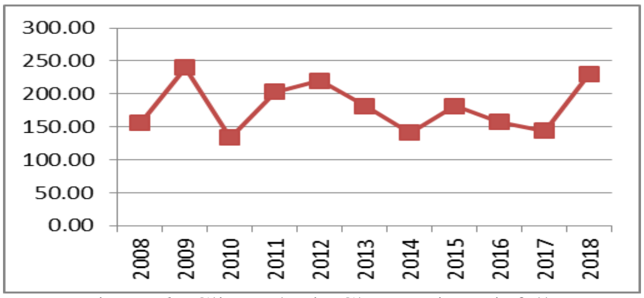



Figure 6. Climatologic Changes in Rainfall

Figure 6 shows that changes in rainfall in Region III as recorded by PAGASA. With warmer climate, heavy rainfall will increase with fewer but more intense events. This then will lead to longer dry climate and a higher flood risks ???. Rainfall was lowest in 2010 and recurrred to be low in 2014 and 2017. Rainfall will affect how irrigation should be planned for rice crops.

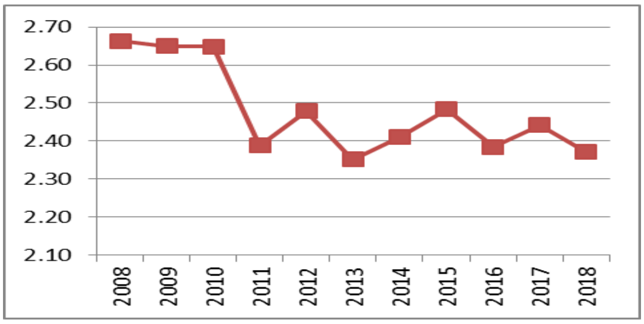



Figure 7. Climatologic Changes in Wind

Figure 7 depicts the wind changes over the past 11 years. Wind power in most studies have the lowest environmental impacts except in the field of generating electricity. Wind power reduces carbon emissions, saves gallons of water a year and reduces pollution [???]. Wind power in Region had been consistently high from 2008 to 2010. Its sudden decrease in 2011 was never as high again. Its value had not been consistent ever since 2011.

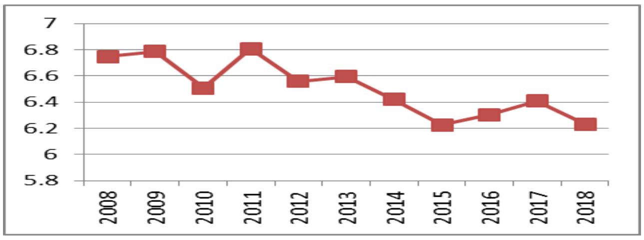



Figure 8. Climatologic Changes in Cloudness

Values in Cloudness changes is depicted in Figure 8. Sunlight is reflected by the clouds to cool the Earth. Clouds , however, also traps heat to warm the Earth [???]. The highest cloudness recorded was in 2011 and from then on, it had shown a decrease. The highest cloudness after 2011 was in 2013 and slightly lower was on 2017.

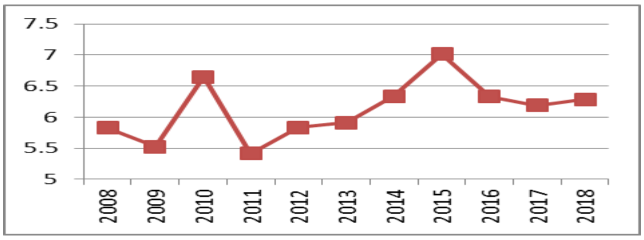



Figure 9. Climatologic Changes in Sunshines

Figure 9 shows the daily average of number of hours of sunshines from 2008 to 2018. Changes in climate depend on changes in the net sunlight reaching the Earth [???]. The number of hours of sunshine was high in 2010 and was even higher after five (5) years in 2015. Since the lowest sunshine in 2011, the value was never again low from 2012 until 2018.

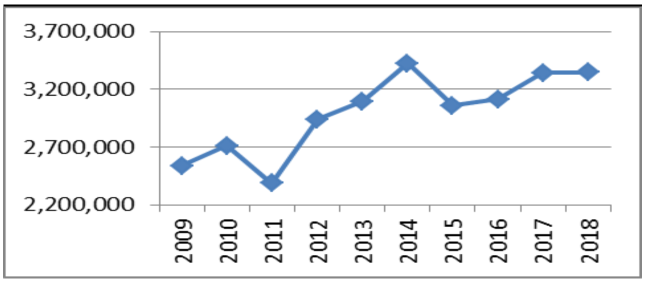



Figure 10. Rice Yield for the Past 10 Years

Figure 10 shows the yearly rice yields in metric tons from 2009 to 2018 from the records of DA. The rice yields had increase continually since 2011 with its peak on 2014. It was however noted the the production had only gradually changed from 2015 to 2018. With the increasing population [???], it was worth mentioning that more rice should had been yielded. The rice yield increase rate is not coping with the population increase rate. The 2015 census results showed that the population increased by over 8 million people or 1.89% compared to the 2010 results. The rice yield only increase by 1.007% from 2017 to 2018.

### B. Data Cleaning

Data cleaning started upon the collection of the needed reports from PAGASA and DA. Reports of the daily records of the seven (7) climactic factors was taken into account while the yearly rice yield was also included.

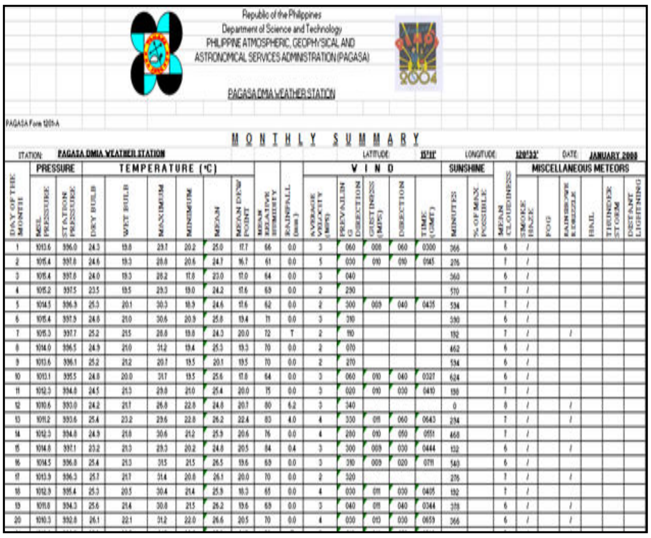


Figure 11. Data from PAGASA

Figure 11 depicts how the data of climatologic factors from PAGASA looks like in MS Excel. The data from 2008 to 2018 are saved on separate files every month and then collated yearly according to the predictors. The predictors are the factors under study and these are the climatologic facotrs particulary pressure, temperature, relative humidity, rainfall, wind, cloudness and sunshine. PAGASA collects the data from the various weather centers located within Pampanga. PAGASA in Clark consolidate the said climatologic reports and may be requested for research purposes such as this undertaking.

The data cleaning tasks used by the researcher was conducted to create a dataset that is uniform with other related datasets depicting the various climatologic factors and rice yield.

The MS Excel formatting on the given data sets affect the structure of records. This is a problem that when left unhandled because it would create errors in reading the data and applying computations when using either MS Excel.

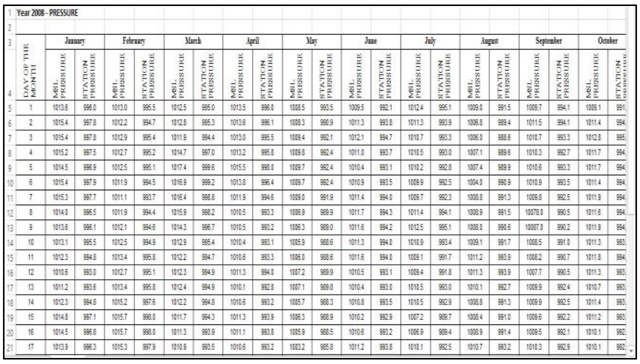


Figure 12. Unformatted Data

Figure 12 shows the unformatted data from 2008 to 2018 which were saved in comma-separated values (.CSV) file formats. These actions were necessary for the data set in MS Excel to be read in Weka 3.8 without a glitch. There are items in the PAGASA reports where some worksheets in MS Excel have with fields that are beyond the scope of this study. These fields were needed to be removed as part of data restructuring. Going back to Table 1, it indicates the augmented data from PAGASA from 2008 to 2018. The data from DA included rice yields from 2009 to 2018. The data sets from these years would be used as historical data for the computations of the forecasting models through ANN. Numbers stored as Text were converted into numeric formats. This activity was accomplished through the formatting tool of MS Excel. This task was necessary to ensure that all the computations will not encounter NaN or Not-a-Number error.

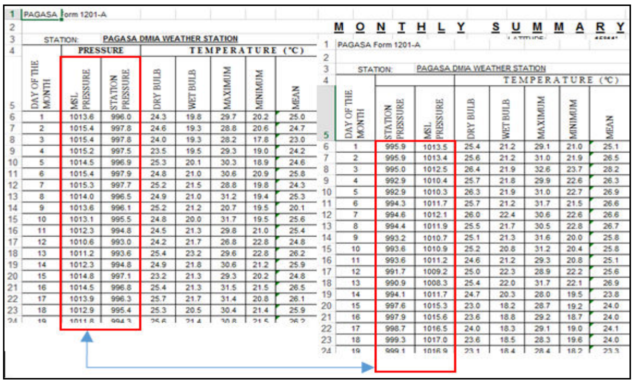


Figure 13. Unformatted Data

Reports from PAGASA from 2008 to 2018 were merged together with the rice yield from DA with data only from 2009 to 2018 as shown in Figure 13. All the data sets merged were using 2009 to 2018 only to match reports both from PAGASA and DA in terms of available data for the years included in this study. Mistake in inputting data from columns (due to difference in value) had also been corrected as part of data cleaning.

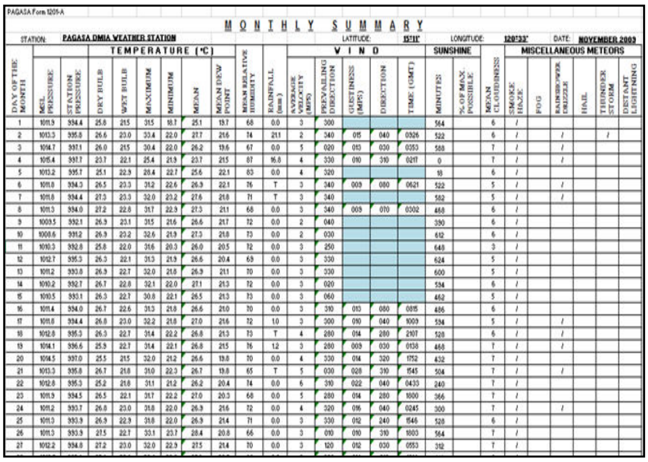


Figure 14. Data with Blank Cells

Another task of data cleaning which is removing blank or empty cells had to be treated also by using MS Excel as shown in Figure 14. These rows with empty cells needed to be removed and not set to 0 because it will affect the computations of forecasting once applied to the entire data set.

### C. *Support Vector Machine*

The cleaned data that summarized the climatologic factors and rice yield from 2009 to 2017 were used as training data sets in Weka 3.8. Data set on rice yield for 2018 was used as validation or test data set. The training data set saved as .CSV file was run using the SMOreg in Weka 3.8.

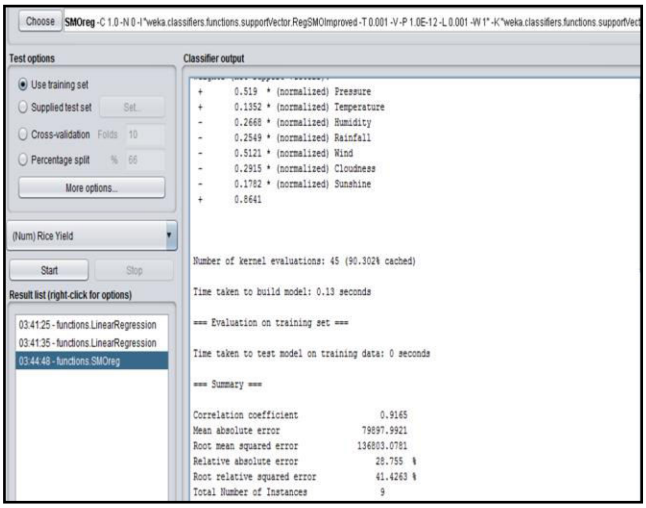



Figure 15. Sets of Weights Generated Using Weka

As shown in Figure 15, Weka 3.8 provided sets of weights and threshold or bias values used in the computation of the SVM class and adjusted weights.

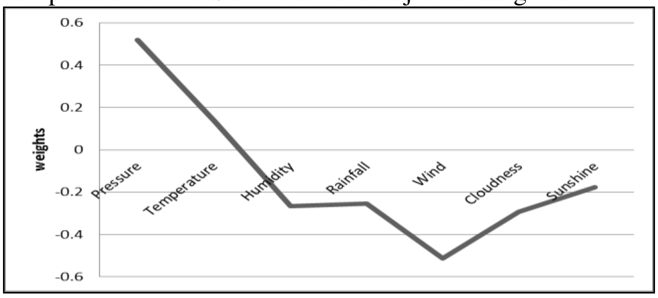
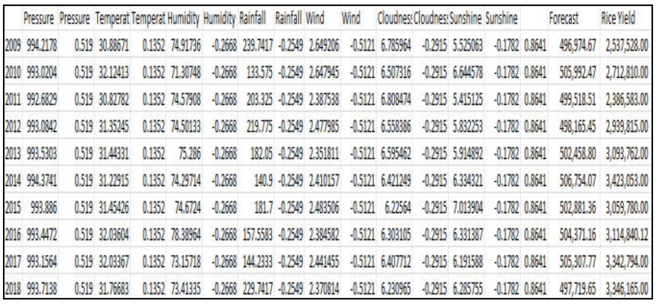



Figure 16. Initial Weights

Figure 16 depicts the initial weights assigned by Weka as plotted in a chart using MS Excel. From the initial weights, Weka assumed that the given training data set should allocate more weight to pressure and temperature with respect to the rice yields.



| | Pressure | Pressure | Temperat | Temperat | Humidity | Humidity | Rainfall | Rainfall | Wind | Wind | Cloudnes | Cloudnes | Sunshine | Sunshine | | Forecast | Rice Yield |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2009 | 994.2178 | 0.519 | 30.88671 | 0.1352 | 74.91736 | -0.2668 | 239.7417 | -0.2549 | 2.649206 | -0.5121 | 6.785964 | -0.2915 | 5.525063 | -0.1782 | 0.8641 | 496,974.67 | 2,537,528.00 |
| 2010 | 993.0204 | 0.519 | 32.12413 | 0.1352 | 71.30748 | -0.2668 | 133.575 | -0.2549 | 2.647945 | -0.5121 | 6.507316 | -0.2915 | 6.644578 | -0.1782 | 0.8641 | 505,992.47 | 2,712,810.00 |
| 2011 | 992.6829 | 0.519 | 30.82782 | 0.1352 | 74.57908 | -0.2668 | 203.325 | -0.2549 | 2.387538 | -0.5121 | 6.808474 | -0.2915 | 5.415125 | -0.1782 | 0.8641 | 499,518.51 | 2,386,583.00 |
| 2012 | 993.0842 | 0.519 | 31.35245 | 0.1352 | 74.50133 | -0.2668 | 219.775 | -0.2549 | 2.477985 | -0.5121 | 6.558386 | -0.2915 | 5.832253 | -0.1782 | 0.8641 | 498,165.45 | 2,939,815.00 |
| 2013 | 993.5303 | 0.519 | 31.44331 | 0.1352 | 75.286 | -0.2668 | 182.05 | -0.2549 | 2.351811 | -0.5121 | 6.595462 | -0.2915 | 5.914892 | -0.1782 | 0.8641 | 502,458.80 | 3,093,762.00 |
| 2014 | 994.3741 | 0.519 | 31.22915 | 0.1352 | 74.29714 | -0.2668 | 140.9 | -0.2549 | 2.410157 | -0.5121 | 6.421249 | -0.2915 | 6.334321 | -0.1782 | 0.8641 | 506,754.07 | 3,423,053.00 |
| 2015 | 993.886 | 0.519 | 31.45426 | 0.1352 | 74.6724 | -0.2668 | 181.7 | -0.2549 | 2.483506 | -0.5121 | 6.22564 | -0.2915 | 7.013904 | -0.1782 | 0.8641 | 502,881.36 | 3,059,780.00 |
| 2016 | 993.4472 | 0.519 | 32.03604 | 0.1352 | 78.38964 | -0.2668 | 157.5583 | -0.2549 | 2.384582 | -0.5121 | 6.303105 | -0.2915 | 6.331387 | -0.1782 | 0.8641 | 504,371.16 | 3,114,840.12 |
| 2017 | 993.1564 | 0.519 | 32.03367 | 0.1352 | 73.15718 | -0.2668 | 144.2333 | -0.2549 | 2.441455 | -0.5121 | 6.407712 | -0.2915 | 6.191588 | -0.1782 | 0.8641 | 505,307.77 | 3,342,794.00 |
| 2018 | 993.7138 | 0.519 | 31.76683 | 0.1352 | 73.41335 | -0.2668 | 229.7417 | -0.2549 | 2.370814 | -0.5121 | 6.230965 | -0.2915 | 6.285755 | -0.1782 | 0.8641 | 497,719.65 | 3,346,165.00 |

Figure 17. SVM Computation Using Excel

As shown in Figure 17, to compute for the class activation of a particular year, $Y_1 = W_0 * X_0 + W_1 * X_1 + \ldots W_n * X_n + b = f$. The process is repeated for all the inputs and weights for all the years under study.

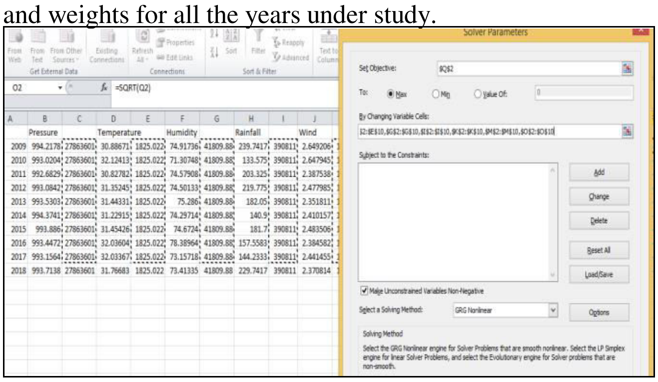


Figure 18. Adjusted Weight Computation Using Solver

As shown in Figure 18, gradient is a mathematical way of propagating the total loss due to error back into the model to know how much is the error of the linkage of each the inputs and subsequently updating the weights to minimize the error by giving the parameters with higher error rates lower weights and vice versa.

### D. *Linear Regression*

For the 2nd model that uses linear regression, the same pre-processed data that summarized the climatologic factors and rice yield from 2009 to 2017 were used as training data sets in Weka 3.8. Data set on rice yield for 2018 was used as validation or test data set. The training data set saved as .CSV file was run using the Linear Regression in Weka 3.8.

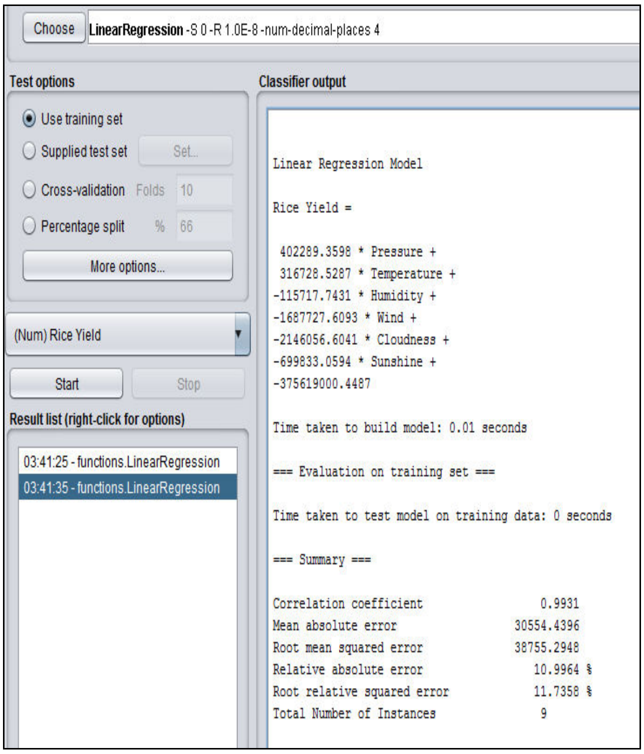



Figure 19. Sets of Weights Generated Using Weka

As shown in Figure 19, Weka 3.8 provided sets of slopes (b) for every input (x) and y-intercept value used in the computation of the Linear Regression rice yield model.

| | Pressure | | Temperature | | Humidity | | Rainfall | | Wind | | Cloudness | | Sunshine | | | Forecast | Rice Yield |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2009 | 994.2178 | 402289.4 | 30.88671 | 316728.5 | 74.91736 | -115718 | 239.7417 | 0 | 2.649206 | -1687728 | 6.785964 | -2146057 | 5.525063 | -699833 | -375619000.4 | 2,556,834.74 | 2,537,528.00 |
| 2010 | 993.0204 | 402289.4 | 32.12413 | 316728.5 | 71.30748 | -115718 | 133.575 | 0 | 2.647945 | -1687728 | 6.507316 | -2146057 | 6.644578 | -699833 | -375619000.4 | 2,701,468.96 | 2,712,810.00 |
| 2011 | 992.6829 | 402289.4 | 30.82782 | 316728.5 | 74.57908 | -115718 | 203.325 | 0 | 2.387538 | -1687728 | 6.808474 | -2146057 | 5.415125 | -699833 | -375619000.4 | 2,430,111.99 | 2,386,583.00 |
| 2012 | 993.0842 | 402289.4 | 31.35245 | 316728.5 | 74.50133 | -115718 | 219.775 | 0 | 2.477985 | -1687728 | 6.558386 | -2146057 | 5.832253 | -699833 | -375619000.4 | 2,858,866.33 | 2,939,815.00 |
| 2013 | 993.5303 | 402289.4 | 31.44331 | 316728.5 | 75.286 | -115718 | 182.05 | 0 | 2.351811 | -1687728 | 6.595462 | -2146057 | 5.914892 | -699833 | -375619000.4 | 3,051,848.28 | 3,093,762.00 |
| 2014 | 994.3741 | 402289.4 | 31.22915 | 316728.5 | 74.29714 | -115718 | 140.9 | 0 | 2.410157 | -1687728 | 6.421249 | -2146057 | 6.334321 | -699833 | -375619000.4 | 3,419,761.57 | 3,423,053.00 |
| 2015 | 993.886 | 402289.4 | 31.45426 | 316728.5 | 74.6724 | -115718 | 181.7 | 0 | 2.483506 | -1687728 | 6.22564 | -2146057 | 7.013904 | -699833 | -375619000.4 | 3,071,700.69 | 3,059,780.00 |
| 2016 | 993.4472 | 402289.4 | 32.03604 | 316728.5 | 78.38964 | -115718 | 157.5583 | 0 | 2.384582 | -1687728 | 6.303105 | -2146057 | 6.331387 | -699833 | -375619000.4 | 3,127,619.73 | 3,114,840.12 |
| 2017 | 993.1564 | 402289.4 | 32.03367 | 316728.5 | 73.15718 | -115718 | 144.2333 | 0 | 2.441455 | -1687728 | 6.407712 | -2146057 | 6.191588 | -699833 | -375619000.4 | 3,392,753.10 | 3,342,794.00 |
| 2018 | 993.7138 | 402289.4 | 31.76683 | 316728.5 | 73.41335 | -115718 | 229.7417 | 0 | 2.370814 | -1687728 | 6.230965 | -2146057 | 6.285755 | -699833 | -375619000.4 | 3,935,455.78 | 3,346,165.00 |

Figure 20. Linear Regression Computation Using Excel

As shown in Figure 20, to compute for the forecast of a particular year, Y = bx+a. $b_0 * X_0 + b_1 * X_1 + \ldots b_n * X_n + a$. The process is repeated for all the inputs and slopes and the y-intercept (a) for all the years under study.

### E. *Evaluation of Accuracy*

Table 2 shows the MAD and MAPE results of the SVM and Linear Regression rice yield models developed in this study. MAD is computed as (∑(actual-forecast))/n while MAPE is computed as ∑((actual-forecast)/actual*100)/n.

Using MAD, Linear Regression has a lower value indicating a lower error. Using MAPE, Linear Regression also has a lower value of percentage error. In both evalutaion criteria, Linear Regression is more suited to be the forecasting model for rice yield using climatolic factors as predictors.

TABLE 2. EVALUATION SUMMARY OF ANN MODELS

| Criteria | SVM | Linear Regression |
|---|---|---|
| MAD | 2,288,704.64 | 30,554.44 |
| MAPE | 79.76 | 1.06 |

## IV. CONCLUSIONS AND RECOMMENDATIONS

Changes in pressure in Region III as recorded by PAGASA showed sudden upsurges in the year 2010 and 2015 with lowest in 2012. Temperature showed an upsurge in 2010 and it had not gone lower than 31 degree celcius since 2014. The relative humidity in Region III had been relatively the same except for the lowest in 2010 and highest in 2016. Rainfall was lowest in 2010 and recurrred to be low in 2014 and 2017. Wind power had been consistently high from 2008 to 2010 and its sudden decrease in 2011 was never as high again. The highest cloudness recorded was in 2011 and from then on, it had shown a decrease. The number of hours of sunshine was high in 2010 and was even higher in 2015 and it had never been again low from 2012 until 2018. The rice yield increase rate is not coping with the population increase rate. The population census results showed increased by 1.89% compared rice yield increase of only by 1.007%. Data cleaning commenced right after the collection of the needed reports from PAGASA and DA. Reports of the daily records of the seven (7) climactic factors was taken into account as well as the yearly rice yield. The data cleaning tasks were conducted to create a dataset that is uniform with other related datasets depicting the various climatologic factors and rice yield. These tasks were also used to handle errors in reading the data and applying computations when using either MS Excel and Weka. There are items in the PAGASA and DA reports where some fields are beyond the scope of this study. These fields were needed to be removed as part of data restructuring. Numbers stored as Text and null values were also corrected. In the SVM model with initial weights generated through Weka 3.8, the seven (7) climatologic factors were set as inputs while the rice yield was set as output. The Linear Regression model implicated that rice yield can be forecasted using the formula of y = bx+a. It was able to generate forecast for the next year with an accuracy higher than SVM. Using MAPE and MAD, it is concluded that Linear Regression produced more accurate forecast that SVM.

It is recommended that future endeavors also considers the type of rice in forecasting. With the changing climatological factors, resilience will be the toll that will give to the survival of roce crops. With the knowledge on future rice yields, it is recommended to use such in decision-making such the amount of imported rice or exported rice. This will pave way to rice sufficiency in the entire country.

[1] Wright, H. (2016). Using ICT to Strengthen Agricultural Extension Systems for Plant Health. Journal of Agricultural & Food Information, vol. 17 (1)., pp. 23-26, DOI: 10.1080/10496505.2015.1120214.

[2] Ghosh, K. Singh, A. Mohanty, U., Acharya, N., Pal, R., Singhe, K. and Pasupalak, S. (2015). Development of a Rice Yield Prediction System Over Bhubaneswar, India: Combination of Extended Range Forecast and CERES-Rice Model. Meteorological Applications Meteorol. Appl. 22: 525–533 (2015) DOI: 10.1002/met.1483.

[3] Intergovernmental Panel on Climate Change (IPCC). 2014. Annual Assessment Report: Climate Change. Cambridge, United Kingdom: Cambridge University Press.

[4] Social Weather Stations. (2018). First Quarter 2018 Social Weather Survey: Hunger among families falls to 9.9%; Moderate Hunger 8.6%, Severe Hunger 1.3%. Retrieved from https://www.sws.org.ph/swsmain/artcldisppage/?artcsyscode=ART-20180430001037 on May 28, 2018.

[5] Piñol, E. (2018). Philippine Agriculture: Today and Tomorrow. Retrieved from http://oldweb.da.gov.ph/The_Philippine_Agriculture_Today_and_the_Future.pdf on May 28, 2018.

[6] Rahman, M., Kang, S., Nagabhatla, N. & Macnee, R. (2017). Impacts of Temperature and Rainfall Variation on Rice Productivity in Major Ecosystems of Bangladesh. Agriculture & Food Securityvolume 6, Article number: 1.

[7] Swansea University. (2019, March 25). Clouds have moderated warming triggered by climate change: A new study has revealed how clouds are modifying the warming created by human-caused climate change in some parts of the world. ScienceDaily. Retrieved August 27, 2019 from sciencedaily.com/releases/2019/03/19025120401.htm

[8] Goode, P.R., & Dziembowski, W.A. (2003). Sunshine , Earthshine and Climate Change I . Origin of , and Limits on Solar Variability.

[9] Jeong JH, Resop JP, Mueller ND, Fleisher DH, Yun K, Butler EE, et al. (2016) Random Forests for Global and Regional Crop Yield Predictions. PLoS ONE 11(6): e0156571. DOI:10.1371/journal.pone.0156571.

[10] Priyanka, T., Soni, P., and Malathy, C. (2018). Agricultural Crop Yield Prediction Using Artificial Intelligence and Satellite Imagery. Eurasian Journal of Analytical Chemistry, vol. 13, pp. 6-12.

[11] Gandhi, N., and Armstrong, L. (2016). A review of the application of data mining techniques for decision making in agriculture. In

2016 2nd International Conference on Contemporary Computing and Informatics, DOI: 10.1109/IC3I.2016.7917925.

[12] Ying-xue, S., Huan, X. & Li-jiao, Y. (2017). Support Vector Machine-Based Copen Crop Model: Case of Rice Production in China. Saudi Journal of Biological Sciences 2017, 24, 537-547.

[13] Gandhi, N., Petkar, O., Armstrong, L. & Cowan, E. (2016). Rice Crop Yield Prediction in India using Support Vector Machines. 13th International Joint Conference on Computer Science and Software Engineering (JCSSE).

[14] Dhakal, C. (2018). Single Regression to Forecast Rice Production: A Case of time, Period and Harvest Area. Journal of Institute of Science and Technology, 18(2): 132-137.