I.J. Modern Education and Computer Science, **2021, 3, 23-33**
Published Online June 2021 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2021.03.03

Rainfall Forecasting to Recommend Crops Varieties Using Moving Average and Naive Bayes Methods

Muhammad Resa Arif Yudianto
Master Program of Informatic Engineering, Universitas Amikom Yogyakarta, Indonesia
Email: muhammadresa0203@students.amikom.ac.id

Tinuk Agustin
Master Program of Informatic Engineering, Universitas Amikom Yogyakarta, Indonesia
Email: tinuk.agustin@students.amikom.ac.id

Ronaldus Morgan James
Master Program of Informatic Engineering, Universitas Amikom Yogyakarta, Indonesia
Email: ronaldus.1188@students.amikom.ac.id

Firstyani Imannisa Rahma
Master Program of Informatic Engineering, Universitas Amikom Yogyakarta, Indonesia
Email: firstyani.rahma@students.amikom.ac.id

Arham Rahim
Master Program of Informatic Engineering, Universitas Amikom Yogyakarta, Indonesia
Email: arham.rahim@students.amikom.ac.id

Ema Utami
Magister of Informatics Engineering, Universitas Amikom Yogyakarta, Indonesia
Email: ema.u@amikom.ac.id

Received: 11 March 2020; Revised: 06 April 2020; Accepted: 08 May 2020; Published: 08 June 2021

Abstract: Indonesia has been known as an agrarian country because of its fertile soil and is very suitable for agricultural land, including rice. Yogyakarta is one of the most significant granary regions in Indonesia, especially in the Sleman region. However, one of the main challenges in rice planting in recent years is the erratic rainfall patterns caused by climate anomalies due to the El Nino and La Nina phenomena. As a result of this phenomenon, farmers have difficulty determining planting time and harvest time and planting other plants. Therefore, we make rainfall predictions to recommend planting varieties with Moving Average and Naive Bayes Methods in Sleman District. The results showed that moving averages well use in predicting rainfall. From these results, we can estimate that in 2020 rice production will below. That can saw from the calculation of the probability of naive Bayes on rice plants being low at 0.999 and 0.923. So that the recommended intercrops planted in 2020 are corn and peanuts. We also find that rainfall prediction with Moving Average using data from several previous years in the same month is more accurate than using data from four past months or periods.

Index Terms: Rainfall forecasting, Naive Bayes, Moving Average, Crops, Prediction.

1. Introduction

Indonesia located at an active volcano meeting (Ring of Fire) that makes lands of Indonesia is very fertile and very suitable for agriculture. Indonesia is also on the Equator so that Indonesia is in a warm tropical region with year-round sunshine and has two seasons, namely the dry and rainy seasons. The tropical countries usually have extraordinarily abundant natural resources. No wonder Indonesia is known as an agricultural country, and the third-largest rice producer in the world [1].

Indonesia has eight rice granaries. This rice granary is the cornerstone of rice production. One of the rice barns in Indonesia located in the province of Yogyakarta. Sleman is one of the districts, which covers only 18% of the province of Yogyakarta, but the region is becoming a significant granary of rice in the area. The success of agricultural products (food availability) support by factors such as land, seeds, fertilizer, and irrigation, all of which can be regulated.

However, climate conditions are not negotiable because they cannot refuse, but we can adjust to knowing the weather forecast that will come. The planting schedule in dryland depends on the condition of monthly rainfall in the concerned area. Currently, farmers can no longer rely on custom cropping patterns and have always been a habit hereditary. Lately, natural phenomena have shown an increasingly essential and unpredictable role through the emergence of El Nino and La Nina climate anomalies. The fact that El Nino and La Nina make weather predictions is not like before [2]. The consequences of weather prediction errors can lead to the risk of crop failure and cause substantial losses in the agriculture sector. Rainfall and water availability in the soil are critical factors when meeting water needs for plants or crops. The rainfall forecast is an essential feature in the study of agriculture science and technology. The study of the weather needs to do in predicting rainfall in each area, based on these conditions.

There is a discussion about the behavior of rainfall in previous studies that helps Indian farmers create policy and decrease crop damage due to irregular rainfall patterns [3]. Likewise, in Indonesia, because uncertain weather conditions affect water availability and cause a decrease in crop quality, crop failure, and crop damage increase. As a result of the decline in the quality and quantity of the crop, agricultural production declines and causes losses in the farming sector, negatively impacting the welfare of farmers. In this study, the authors wanted to predict the weather to determine when the right time for planting rice and other crops interlude plants and recommend the appropriate varieties of crops for farmers to maximize yields and reduce losses due to changes in the weather conditions are uncertain.

2. Related Studies

A study of the temporal-spatial dynamics of meteorological variables in the context of climate change to examine changes in rainfall and temperature in northern Ethiopia, resulting in recommendations for strategies designed in the agricultural sector, must take into account the nature of decreased and erratic rainfall [4]. Some research [5,6,7] provides strong evidence that changes in temperature and rainfall can lead to additional severe obstacles to agriculture in Africa [8].

Another research explained that The ideal planting period in Bangladesh divided into three periods of each year [3]. In this research, the cropping patterns from all periods are always the same (rice-rice-rice) because the country tends to have high rainfall. Nevertheless, this precipitation is not evenly distributed in each region. What distinguishes between these periods is not a type of crop for rice, but only the variety of farmed rice. For example, there is a crop calendar (aus-aman-boro) adjusted to the high rainfall in each planting period.

Moreover, there is a paper describing the resource management characteristics in agricultural land and the impact of season conditions, which is utilized by farmers in maintaining the farm's sustainability at Imogiri district, Bantul Regency of Yogyakarta [9]. The research described the features of resource management in the agricultural field in the Imogiri district, depending on the type of land, the pattern of crop planting, and the dominant types of products. Furthermore, season conditions can improve agricultural land resource management's sustainability up to 86.2% in the irrigated rice fields, 92.7% in the rainfed rice field, and 88.6% in the dry land.

Forecasting precipitation can do by using a statistical model that is ARMA (Auto Regressive Moving Average), M.A. (Moving Average), A.R. (Auto-Regressive), Multiple Regression, and ARIMA (Auto-Regressive Integrated Moving Average) [10]. We will use the moving average to predict the rainfall. Generally, the moving average method use to statistical prediction in business and engineering purposes. Some applications of moving average for forecasting are predicting stock price [11], counting electricity demands [12], and foreseeing total delivery goods [13].

The classification method can use the Naive Bayes method. Previous studies analyzed wheat production using Naïve Bayes conducted by Kaur and Kalsi to produce future trends from a collection of historical information [14]. This study concludes that naïve Bayes' performance is better than SVM and KNN with an accuracy of Naive Bayes 95.34%, SVM 74.48%, and KNN 91.53%. Furthermore, Tsangaratos and Ilia's research compared logistic regression and naïve Bayes in the assessment of landslide vulnerability [15]. Of the 116 data is divided into 58 Data landslides and 58 area non-landslide conclude the performance and accuracy of naïve Bayes slightly higher with a value of 87.50% of the logistic regression with an amount of 82.61%. For prediction cases, Hou and Yang also used Naïve Bayes to predict Conotoxin Superfamilies [16]. This study concluded that from 305 data conotoxins, the accuracy value was 84.92%.

3. Proposed Methods

The research begins by identifying the problem by conducting a literature study to describe the issues that occur in the agricultural sector in Sleman District. The stages in this research flow shown in Fig. 1.

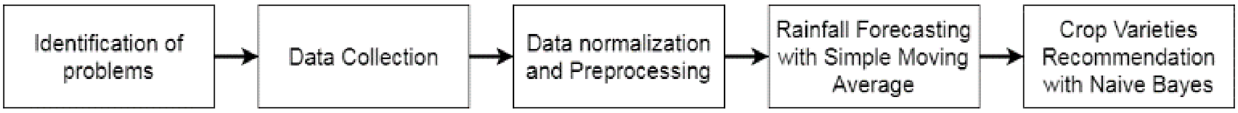

Fig. 1. Research Flow

A. Data Collection

The dataset used in this study consisted of several types of data, including crop harvest and rainfall data. Data on plant commodities obtain from the primary sources, namely from the Sleman Agriculture Service, which are presented in the form of tables from the last five years, starting from the amount of production from various plant varieties, pests that attack, and pest control data from 2014-2018. At the same time, rainfall data obtained from secondary sources, namely from the BPS (Statistics Indonesia) government site in Sleman from 2005-2018.

The data obtained is still in the form of raw and unstructured data, so a preprocessing process is needed to clean and normalize the data so that the data is feasible and valid for further processing. After the data is declared normal, the next stage is the process of rainfall forecasting carried out on rainfall data in previous years. The results of rainfall forecasting use for the classification of plant varieties that are appropriate for plants in certain months based on several parameters / dependent variables.

B. Normalization

Normalization is one of the preprocess texts that functions to process raw and unstructured data from the data mining process, so it expects that the words or text resulting from the normalization process can extract into features that affect their respective classes [17]. The process of data normalization in this study is carried out by grouping data based on the appropriate attributes and will use as dependent variables. The next step is cleaning the irrelevant data and not affecting the variables that affect the process of determining plant varieties. The process of data normalization is different from the normalization of data in a relational database.

C. Data Pre-processing

Data preprocessing is an important step that must be done before the classification or forecasting process in time series data to improve performance accuracy. This process involves reducing or eliminating incomplete words and noise from unstructured data into more complete, neat, and clean data from noise [18,19]. This preprocessing phase aims to minimize errors caused by bias from each dataset and get the results of feature extraction that is not widespread [19].

The preprocessing stage in this study with a variety of treatments for a variety of varied data, starting from the process of cleaning incomplete data, cleaning noise, and completing incomplete data. One example of a preprocessing method by completing incomplete data on rainfall data shown in Table 1. If a value data is missing or incomplete, the default value, 0, can be given in mathematical operation performed [20].

Assumed the column has a value of 0 but given the amount "-" indicating the absence of rainfall in the month and year by changing it to the number 0 like the techniques in previous studies [20].

Table 1. Example Pre-Processing Rainfall Data

MONTH	2005	2006		MONTH	2005	2006
January	198	268		January	198	268
February	271	131		February	271	131
March	76	277		March	76	277
April	100	132		April	100	132
May	0	123		May	0	123
June	121	-		June	121	0
July	42	-		July	42	0
August	-	-		August	0	0
September	10	-		Septembe r	10	0
October	51	-		October	51	0
November	57	26		Novembe r	57	26
December	417	182		December	417	182

D. Forecasting with Moving Average

Moving Average is often defined as a mean of some parts of the entire sample [21]. In time series context, moving average is a method for finding an average of certain items that move from series to another by eliminating top numbers of the previous averaged group and adding the next in the following group [22]. Moving Average is also for smoothing the time series where the number of variations in the data reduced [23].

Advantages of Moving Average are 1) The moving average removes short-term fluctuations; 2) decrease the impact of extreme values in time series; 3) The method is adjustable for every need [22]. There are many types of moving average methods, but we will use Simple Moving Average, which is often used to estimate the current level of a time series and make a prediction for the future [24].

In this paper, we will use two approaches to rainfall forecasting. In the first approach, we will calculate a prediction using actual data from previous periods and divide them by the number of periods, as explained in (1)

$$F_{(t+1)} = \frac{A_{(1)}+A_{(2)}+A_{(3)}+\cdots+A_{(t)}}{n} \quad (1)$$

Where F(t+1) is a prediction result for the later period, A(1) to A(t) are actual data from the previous period, and n is the number of prior periods.

On the other hand, we also forecast data in a month next year with data from the same month of several previous years, as explained in (2)

$$F_{(Month\ t+1)} = \frac{A_{(Month\ 1)}+A_{(Month\ 2)}+\cdots+A_{(Month\ t)}}{n} \quad (2)$$

Where F(Month t+1) is a prediction result for a month next year, A(Month 1) to A(Month n) are actual data from the same month in previous years, and n is a total of prior years.

To determine the best approach for rainfall forecasting, we must know how far the result values with the real one. We will use Mean Square Error to estimate the mean of the squared difference between prediction results and actual data [25] as explained in (3)

$$MSE = \frac{\sum(A_t-F_t)^2}{n} \quad (3)$$

Where F(t) is the forecasting result in period t, A(t) is real data from period t, and n is a total of periods.

E. Classification with Naive Bayes

The Naive Bayes is a classification algorithm based on probability, which includes a strong independence assumption [26]. This model will calculate the probabilities of class members with a feature only and the previous likelihood [27]. It has shown as a successful classifier in many fields [28].

The strong points of Naive Bayesian Classifier are: 1) can guess predictor variables that do not depend on their effects in the classification; 2) to accept any high number either constant or categorical variables; 3) can cut the high-dimensional to the one-dimensional basic density estimation; 4) They can train and classify faster; 5) They insensitive to a variable that not useful [29]. Unlike neural networks, the Naive Bayesian Classifier does not require settings with many free parameters, and its results, which returns as probabilities, can be used in many jobs [30]. To calculate the numerical probability of data. It is necessary to find the mean and standard deviation of each parameter. The equation used to calculate the value – average arithmetic (mean) can see in (4)

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (4)$$

Information :

µ: calculated average (mean)
xi: the i-th sample value
n: number of samples
The equation for calculating the standard deviation can see in (5)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n-1}} \quad (5)$$

Information :

σ: standard deviation
xi: i-th x value
µ: calculated average (mean)
n: number of samples

After getting the mean value (mean) and the standard deviation, then classify with continuous data used using the Gaussian Density as explained in (6)

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma^2{}_{ij}}} \tag{6}$$

Information :

P: Opportunity
Xi: Attribute to i
xi: Attribute value to i
Y: Class sought
yi: Subclass Y is sought
µ: mean, represents the average of all attributes
σ: Standard deviation, denotes variants of all attributes.

4. Experiments and Results

A. Validity Test Both Methods of Choosing an SMA (Simple Moving Average)

We will use the Sleman district at Sleman regency rainfall dataset that contains rainfall data from 2005 to 2018 to predict the level of rainfall each month in 2019 and 2020. We will test two approaches, as explained in part 2 of this paper, before calculating the rainfall prediction.

1. First Approach

In this first approach, we performed the forecast with data from several previous periods, and we calculated the prediction with equation (1). We used data in February 2013 - June 2013 and February 2014 - June 2014. We chose those data because there is no value of 0, which means it is not raining, for more accurate results. The following calculation results show in Table 2.

Table 2. Calculation Results of the First Approach

Period + 1	A_i				$\sum_{i=t-n+1}^{t} A_i$	F_{t+1}
February 2013	Oct-2012	Nov-2012	Dec-2012	Jan-2013	1490	372.5
	118	531	378	463		
March 2013	Nov-2012	Dec-2012	Jan-2013	Feb-2013	1820	455
	531	378	463	448		
April 2013	Dec-2012	Jan-2013	Feb-2013	Mar-2013	1603	400.75
	378	463	448	314		
May 2013	Jan-2013	Feb-2013	Mar-2013	Apr-2013	1545	386.25
	463	448	314	320		
June 2013	Feb-2013	Mar-2013	Apr-2013	May-2013	1275	318.75
	448	314	320	193		
February 2014	Oct-2013	Nov-2013	Dec-2013	Jan-2014	1423	355.75
	248	294	390	491		
March 2014	Nov-2013	Dec-2013	Jan-2014	Feb-2014	1587	396.75
	294	390	491	412		
April 2014	Dec-2013	Jan-2014	Feb-2014	Mar-2014	1597	399.25
	390	491	412	304		
May 2014	Jan-2014	Feb-2014	Mar-2014	Apr-2014	1492	373
	491	412	304	285		
June 2014	Feb-2014	Mar-2014	Apr-2014	May-2014	1238	309.5
	412	304	285	237		

2. Second Approach

In the second approach, we predict the rainfall level for a month in the year with equation (2). We applied data from several previous years in the same month as used in the first approach. The following calculation results are shown in Table 3.

Table 3. Calculation Results of the Second Approach

Period + 1	A_i				$\sum_{i=t-n+1}^{t} A_i$	F_{t+1}
Feb-2013	Feb-2009	Feb-2010	Feb-2011	Feb-2012	1465	366.25
	238	323	358	546		
March 2013	Mar-2009	Mar-2010	Mar-2011	Mar-2012	1255	313.75
	270	351	278	356		
April 2013	Apr-2009	Apr-2010	Apr-2011	Apr-2012	1243	310.75
	462	119	377	285		
May 2013	May-2009	May-2010	May-2011	May-2012	981	245.25
	200	430	193	158		
June 2013	Jun-2009	Jun-2010	Jun-2011	Jun-2012	191	47.75
	26	162	0	3		
Feb-2014	Feb-2010	Feb-2011	Feb-2012	Feb-2013	2111	527.75
	455	654	539	463		
March 2014	Mar-2010	Mar-2011	Mar-2012	Mar-2013	1675	418.75
	323	358	546	448		
April 2014	Apr-2010	Apr-2011	Apr-2012	Apr-2013	1299	324.75
	351	278	356	314		
May 2014	Mei-2010	Mei-2011	Mei-2012	Mei-2013	1101	275.25
	119	377	285	320		
June 2014	Jun-2010	Jun-2011	Jun-2012	June-2013	974	243.5
	430	193	158	193		

From the calculation of the two approaches above, a graph can be made as follows in Fig. 2 :

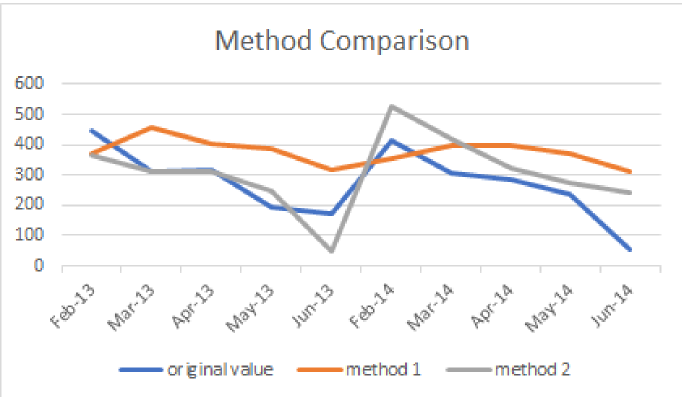

Fig. 2. Graphic Pattern Similarity

From the Fig.2. , it can be seen that the second approach has a pattern that is more similar to the original data compared to the first approach. Therefore, we chose the second approach to calculate rainfall predictions in 2019 and 2020.

B. *Rainfall Prediction with Moving Average Method*

The following is an example calculation to predict rainfall in January 2019 as shown in equation (7) and get the results of the January rainfall calculation of 316.5 mm

$$F_{(Jan\ 2019)} = \frac{A_{(Jan\ 2005)}+A_{(Jan\ 2006)}+\cdots+A_{(Jan\ 2018)}}{(2018-2005)+1}$$

$$F_{(Jan\ 2019)} = \frac{324+432+\cdots+525}{(2018-2005)+1}$$

$$F_{(Jan\ 2019)} = \frac{4431}{14}$$

$$F_{(Jan\ 2019)} = 316{,}5$$

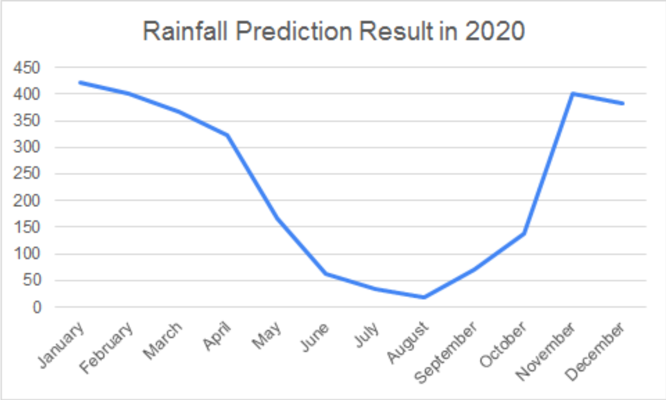

Fig. 3. The graph in 2020

C. *Crops Varieties Recommendation with Naive Bayes*

The Naive Bayes algorithm is used to classify the yields of rice, tomatoes, chilies, peanuts, corn, and cassava to determine which crops will be recommended as a crop in the next planting period. The recommendations are obtained from the district data, harvested area, rainfall, pests, pest control, and production results from previous years. Before carrying out the classification process to recommend intermittent crops in 2020, the first Naive Bayes was tested by comparing the Naive Bayes classification of rice plants with actual data in 2018.

1. *Finding Value Mean*

To get the mean value calculated using equation (4) on each type of label from the data. For example, looking for the mean values of the harvest area attribute for high and low tags:

$$\mu_{high} = \frac{3911+3225+3006+3236+3408+3185+749+792+992}{9} = \frac{22504}{9} = 2550044$$

$$\mu_{low} = \frac{2785+3044+104}{3} = \frac{6873}{3} = 2291$$

From the above calculation, the mean value of the harvested area attribute is obtained for the high class = 2500.44 and the low class = 2291. This process is carried out for all attributes of the training data.

Finding Value Standard Deviation

After getting the mean value, the next process is to find the standard deviation using equation (5) on each type of label from the data. For example, looking for a standard deviation of the harvest area attribute for high and low labels:

$$\sigma^2_{high} = \frac{\begin{array}{c}(3911-2500.44)^2+(3225-2500.44)^2+(3006-2500.44)^2\\+(3236-2500.44)^2+(3408-2500.44)^2+(3185-2500.44)^2+\\(749-2500.44)^2+(792-2500.44)^2+(992-2500.44)^2\end{array}}{9-1}$$

$$= \frac{\begin{array}{c}(1989666.98)+(524980.75)+(255586.42)+\\(541041.98)+(823657.09)+(468616.31)+\\(3067557.64)+(2918782.41)+(2275404.64)\end{array}}{8}$$

$$= \frac{128652994.22}{8} = 1608161.78$$

$$\sigma^2_{high} = \sqrt{1608161.78} = 1268.13$$

$$\sigma^2_{low} = \frac{(2785-2291)^2+(3044-2291)^2+(1044-2291)^2}{3-1}$$

$$= \frac{(244036)+(567009)+(1555009)}{2} = 1183027$$

$$\sigma^2_{low} = \sqrt{1183027} = 1087.67$$

From the above calculation, the standard deviation value of the harvested area attribute is obtained for the high class = 2500.44 and the low class = 2291. This process is carried out for all characteristics of the training data.

2. Naive Bayes Calculation

After demanding mean and standard deviation value for continuous data, the next process is naive Bayes calculation with a Gaussian density, as mentioned in equation (6). For example below, we will count the harvested area attribute for high class and low class:

$$
\begin{aligned}
HarvestProd.Area\ (High|2946.89) &= \frac{1}{\sqrt{2\pi(1268.133186)}}\ e^{\frac{-(2946.89-2500.444444)^2}{2(1268.133186)^2}} \\
&= \frac{1}{\sqrt{2x3.14(1268.133186)}}\ 2.7183^{-0.061969398} \\
&= \frac{1}{3177.932656}\ 0.93991126 \\
&= 0.000295762
\end{aligned}
$$

$$
\begin{aligned}
HarvestProd.Area\ (Low|2946.89) &= \frac{1}{\sqrt{2\pi(1087.670446)}}\ e^{\frac{-(2946.89-2291)^2}{2(1087.6704446)^2}} \\
&= \frac{1}{\sqrt{2x3.14(1087.670446)}}\ 2.7183^{-181818205} \\
&= \frac{1}{2725.694326}\ 0.833751885 \\
&= 0.000305886
\end{aligned}
$$

3. Looking for Likelihood Value

From the calculation above, the next step is to calculate the likelihood value as explained in the equation below:

$$P(X|High) = \left(\frac{3}{9}\right) \times 0.000295762 \times 0.000175041 \times 0.004905898 \times 0.00476015 \times 0.009260825 \times 0.01357245$$

$$P(X|Low) = \left(\frac{1}{9}\right) \times 0.000305886 \times 0.000102939 \times 0.003911756 \times 0.003911756 \times 0.00494205 \times 0.00749228 \times 0.01357245$$

4. Probability value normalization

From the likelihood value above, we will do the normalization of probability value. The result of the process is explained at the equation below:

$$
\begin{aligned}
P(X|High) &= \frac{5.0632 \times 10^{-17}}{(5.0632 \times 10^{-17} + 3.14716 \times 10^{-16})} \\
&= 0.138635614
\end{aligned}
$$

$$
\begin{aligned}
P(X|Low) &= \frac{3.14716 \times 10^{-16}}{(5.0632 \times 10^{-17} + 3.14716 \times 10^{-16})} \\
&= 0.861364386
\end{aligned}
$$

5. Finding of maximum probability value

Before we classify rice crops' production results, we must look at the final probability value that closes to 1 or equal to 1. As explained in the normalization process above, the HIGH class has a probability value 0.138635614 and the probability value of the LOW class is 0.86136438, we can conclude that the production of rice corps in 2018 is LOW.

tahun	kecamatan	luas_panen	produksi	produksi_divide_lahan	STATUS	
2014	Turi	749	4548	6.0720961281708945	High	
2014	Sleman	3236	19902	6.150185414091471	High	
2014	Gamping	3911	23674	6.053183329071849	High	
2015	Turi	792	4998	6.3106060606060606	High	
2015	Sleman	3408	21902	6.426643192488263	High	
2015	Gamping	3225	21274	6.596589147286822	High	
2016	Turi	992	6058	6.106854838709677	High	Data Train
2016	Sleman	3185	19902	6.248665620094192	High	
2016	Gamping	3006	18290	6.084497671324018	High	
2017	Turi	1044	5831	5.585249042145594	Low	
2017	Sleman	3044	17843	5.861695137976347	Low	
2017	Gamping	2785	16398	5.887971274685817	Low	
2018	Gamping	2561.56	14097.7	5.503575911514799	Low	
2018	Turi	804.82	4231.25	5.2573866970176555	Low	Data Test
2018	Sleman	2946.89	15743	5.342228768230145	Low	

Fig. 4. Training data and Testing data

Table 4. Classification from Each Region For 2020

Plant	Status	Probability Value	MAX Probability Value	Result
Gamping	High	0.518391677	0.518391677	High
	Low	0.481608323		
Sleman	High	0.138635614	0.861364386	Low
	Low	0.861364386		
Turi	High	0.370654592	0.629345408	Low
	Low	0.629345408		

To test the accuracy of the Naive Bayes test is carried out on test data consisting of data from 3 districts, Gamping, Sleman, and Turi, as explained in Table 4. In the process of testing the reliability of the algorithm in predicting/classifying alternating plants used as much as 3 test data that has a label. The results of testing 3 test data show that 2 data, namely the Sleman and Turi regions, are classified correctly. Still, the Gamping area cannot be classified correctly, where the original data of the Gamping area is low but highly classified.

From Fig. 4 above, we found a similarity when compared with the results of the Naive Bayes classification. Therefore, we did the process for the classification of crop production in 2020.

Table 5. Crop Production Classification Result For 2020

Plant	Status	Probability Value	MAX Probability Value	Result
Rice	High	0.394546519	1	Low
	Low	0.605453481		
Corn	High	0.999901685	0.999901685	High
	Low	9.83146E-05		
Tomato	High	3.4671E-275	1	Low
	Low	1		
Chili	High	0.001243174	0.998756826	Low
	Low	0.998756826		
Peanuts	High	0.923361151	0.923361151	High
	Low	0.076638849		
Cassava	High	0.368670906	0.631329094	Low
	Low	0.631329094		

Table 5 above shows the results of the classification using the Naive Bayes method on rice, corn, tomatoes, chilies, peanuts, and cassava with rainfall forecasting data in 2020. We can predict that rice production is low in 2020, and the productivity of corn and peanut is high. So, we recommend planting corn or peanuts during the month.

5. Conclusion and Future Works

The previous research [9] explains when they are looking for time to plant rice. The farmers still rely on season conditions. This study also tells about crop rotation with alternative plants, such as palawija. Even, it does not clearly explain what kind of crops to be planted in particular either situation or planting time. Otherwise, Our research describes obtaining planting time using the Naive Bayes algorithm of several variables that affect plant growth and yield. One of these variables is the prediction of rainfall occurring each year in the future. In this study, researchers also specifically recommend palawija, or secondary crop, interlude plants that are good for planting.

Our research finds that rainfall forecasting with Moving Average with data from several previous years in the same month that applied in the second approach is more accurate than the first approach that uses data from four past months. For future works, we will use annual data instead of data from each harvest season and adding more data from the previous periods.

References

[1] Data on the Amount of Production, Export and Import are quoted from the FAO (Food and Agriculture Organization) Publication of the United Nations, Rice Market Monitor, Volume XVIII Issue No. 2, July 2015.

[2] Prasetyo, Y., Nabilah, F. 2017. "Pattern Analysis of El Nino and La Nina Phenomenon Based on Sea Surface Temperature (SST) and Rainfall Intensity using Oceanic Nino Index (ONI) in West Java Area". Conference Series Earth and Environmental Science 98(1):012041

[3] Geetha,G., and Selvaraj,R.S. 2011. "Prediction of monthly rainfall in Chennai using Back Propagation Neural Network model". Int. J. of Eng. Sci. and Technology, vol. 3, no. 1, pp. 211-213

[4] Darji, M.P., Dabhi,V.K., Prajapati, H.B., "Rainfall Forecasting Using Neural Network: A Survey". Conference: Computer Engineering and Applications (ICACEA), International Conference on Advances in, At Ghaziabad, India, 10.1109/ICACEA.2015.7164782
[5] Asfawa, A. Simaneb, B. Hassenc, A. Bantiderd, A. 2017. "Variability and time series trend analysis of rainfall and temperature in north central Ethiopia: A case study in Woleka sub-basin". Weather and Climate Extremes. Volume 19. Elsevier
[6] Seo, S. N., Mendelsohn, R., Dinar, A., Hassan, R., & Kurukulasuriya, P. (2009). "A ricardian analysis of the distribution of climate change impacts on agriculture across agro-ecological zones in Africa". Environmental and Resource Economics, 43(3), 313–332.
[7] Nelson, G. C., Rosegrant, M. W., Palazzo, A., Gray, I., Ingersoll, C., Robertson, R., et al. (2010). Food security, farming, and climate change to 2050: Scenarios, Results, Policy Options (p 155). Washington, DC: IFPRI.
[8] Benin, S., Wood, S., & Nin-Pratt, A. (2016). "Introduction In Agricultural productivity in Africa: Trends, patterns, and determinants". In S. Benin (Ed.), Chapter 1. (pp. 1–23). Washington, DC: International Food Policy Research Institute (IFPRI)
[9] Khotimah, N. (2019). Pranata mangsa and the sustainability of agricultural land resources management in Imogiri sub-district of Bantul regency. IOP Conference Series: Earth and Environmental Science, 338, 1–8.
[10] Lana, M.A., Vasconcelos, A.C.F., Gornott, C. et al. "Is dry soil planting an adaptation strategy for maize cultivation in semi-arid Tanzania?". Food Sec. 10, 897–910 (2018) doi:10.1007/s12571-017-0742-7
[11] Hari, Y., & Dewi, L. P. (2018). "Forecasting system approach for stock trading with relative strength index and moving average indicator". Journal of Telecommunication, Electronic and Computer Engineering, 10(2–3), 25–29.
[12] Kartikasari, M. D., & Prayogi, A. R. (2018). "Demand forecasting of electricity in Indonesia with limited historical data". Journal of Physics: Conference Series, 974(1).
[13] Maricar, M. A., Widiadnyana, P., & Arta Wijaya, I. W. (2017). "Analysis of Data Mining for Forecasting Total Goods Delivery with Moving Average Method". International Journal of Engineering and Emerging Technology, 2(1), 7.
[14] S. Kaur and S. Kalsi. 2019. "Analysis of Wheat Production using Naïve Bayes Classifier". Int. J. Comput. Appl., vol. 178, no. 14, pp. 38–41.
[15] P. Tsangaratos and I. Ilia. 2016. "Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size". Catena, vol. 145, pp. 164–179.
[16] H. Huo and L. Yang. 2018. "Prediction of conotoxin superfamilies by the Naive Bayes classifier". Proc. - 2017 10th Int. Congr. Image Signal Process. Biomed. Eng. Informatics, CISP-BMEI 2017, vol. 2018-January, pp. 1–5.
[17] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," Appl. Soft Comput. J., no. xxxx, p. 105524, 2019.
[18] Geoff Dougherty. 2012. "Pattern Recognition and Classification: an Introduction". Springer Science & Business Media.
[19] Salvador García, Julián Luengo, Francisco Herrera. 2015. “Data Preprocessing in Data Mining”. Springer.
[20] S. A. Alasadi and W. S. Bhaya. 2017. "Review of data preprocessing techniques in data mining". J. Eng. Appl. Sci., vol. 12, no. 16, pp 4102–4107.
[21] Evans, M. K. 2001. "Practical Business Forecasting". Wiley-Blackwell.
[22] Molugaram, K., Rao, G. S., Shah, A., & Davergave, N. 2017. "Statistical Techniques for Transportation Engineering". First Edition. Butterworth-Heinemann.
[23] Storm, K. 2019. "Industrial Process Plant Construction Estimating and Man-Hour Analysis". First Edition. Gulf Professional Publishing.
[24] Johnston, F. R., Boyland, J. E., Meadows, M., & Shale, E. 1999. "Some Properties of A Simple Moving Average When Applied to Forecasting A Time Series". Journal of the Operational Research Society, 50(12), 1267–1271.
[25] Lehmann, E. L.; Casella, George. 1998. "Theory of Point Estimation". Second Edition. New York: Springer.
[26] Zhao, Y., & Cen, Y. 2013. "Data Mining Applications with R". Academic Press.
[27] Grimnes, S., & Martinsen, Ø. G. 2015. "Bioimpedance and Bioelectricity Basics". Third Edition. Academic Press.
[28] Mao, W., & Wang, F.-Y. 2012. "New Advances in Intelligence and Security Informatics". Academic Press.
[29] Nisbet, R., Miner, G., & Yale, K. 2018. "Handbook of Statistical Analysis and Data Mining Applications". Second Edition. Academic Press.
[30] Stern, M., Beck, J., & Woolf, B. P. 1999. "Naive Bayes Classifiers for User Modeling". Center for Knowledge Communication, Computer Science Department, University of Massachusetts.

Authors' Profiles

Mr. Muhammad R.A. Yudianto has degrees Bachelor of Information System from Amikom University Yogyakarta. He had been working for Logique Digital Indonesia for 6 months as a software engineer. He is currently a master of Informatic Engineering at the Department of Computer Science, Amikom University Yogyakarta.

Mrs. Tinuk Agustin has a Bachelor of Informatic Engineering from Amikom University Yogyakarta. She is currently a master of Informatic Engineer at the Department of Computer Science, Amikom University Yogyakarta.

Mr. Ronaldus M. James has a Bachelor of Information System from Amikom University Yogyakarta. He is currently a master of Informatic Engineering at the Department of Computer Science, Amikom University Yogyakarta.

Mrs. Firstyani I. Rahma got a Bachelor of Education degree in Informatics Engineering Education in Yogyakarta State University in 2018. Currently, she is studying for a Master of Informatic Engineering in Amikom University Yogyakarta.

Mr. Arham Rahim has a Bachelor of Informatic Engineering from STMIK AKBA Makassar. He is currently a master of Informatic Engineering at the Department of Computer Science, Amikom University Yogyakarta.

Prof. Ema Utami received the S.Si, M.Kom and Doctoral degrees in Computer Science from Gadjah Mada University, Yogyakarta, Indonesia in 1997, 2002 and 2010 respectively. Since 1998 she has been a lecturer in Universitas Amikom Yogyakarta, Indonesia. Her areas of interest are Natural Language Processing, Computer Algorithms, and Database Programming.

How to cite this paper: Muhammad Resa Arif Yudianto, Tinuk Agustin, Ronaldus Morgan James, Firstyani Imannisa Rahma, Arham Rahim, Ema Utami, " Rainfall Forecasting to Recommend Crops Varieties Using Moving Average and Naive Bayes Methods", International Journal of Modern Education and Computer Science(IJMECS), Vol.13, No.3, pp. 23-33, 2021.DOI: 10.5815/ijmecs.2021.03.03