

Prediction of Determination of Rice Farming Production using the Naïve Bayes Method

Diana Laily Fithri¹, Noor Latifah²
{Diana.laily@umk.ac.id¹, Noor.latifah@umk.ac.id²}

Faculty of Engineering, Universitas Muria Kudus, Gondang Manis Bae Kudus, Kudus, 59324, Indonesia¹²

Abstract. Agricultural crops of rice are the main products of agricultural production in Indonesia, many of the farmers do not have enough time to maintain their agricultural products. So that many of these farmers have poor crop production. The criteria used in determining the yield of rice crops are the sex of a farmer, level of education, type of work, position, monthly income, and age level of the farmer. But after the prediction using the Naïve Bayes method, the most dominant criterion in predicting the yield of rice crops is the level of education and occupation at work. Because of the higher job position, it is not necessarily a good agricultural product.

Keywords: yield, crop, rice, turnover, criteria, naïve bayes.

1 Introduction

Rice plants are the result of Indonesian agricultural commodities which must be preserved. Many of the farmers have extensive agricultural land, but are not used properly by farmers. So that for food needs in Indonesia, many governments import rice from abroad so that food needs in Indonesia remain fulfilled. Actually, this can be overcome by mapping farmers who have large agricultural land. So that farmers can use their agricultural land well. By comparing farmland owned by farmers with agricultural products can help farmers to increase income in selling agricultural produce. Rice is a food crop in the form of clump grass that is widely grown in Indonesia. Based on the cultivation system, rice is divided into two types, namely dry rice (gogo) which is planted in dry land / fields and paddy fields planted in fields that are always flooded [1]. One method used in overcoming these problems is by using the Naïve Bayes method. The criteria used in determining the yield of rice crops are the sex of a farmer, level of education, type of work, occupation, monthly income, and age level of the farmer. Many of the farmers who have a high level of education, but can not use their knowledge because of the high level of employment can also affect the results of agricultural production.

In research on rice production to minimize the problem of food insecurity in Kubu Raya district, a solution is needed to predict rice production in Kubu Raya District. This system was built to predict rice production in the January - April period, May - August period, and the September - December period with interval regression methods with fuzzy neural.

The Naïve Bayes method is a method used in determining the most dominant criteria for measurement, in Naïve Bayes a comparison is used between the first probability (h1) with the second probability (h2), which can later be taken a conclusion in determining the main criteria to be selected in the increase agricultural production.

ICCSET 2018, October 25-26, Kudus, Indonesia
Copyright © 2018 EAI
DOI 10.4108/eai.24-10-2018.2280524

Tantiny's research, Budi Susanto and Widi Hapsari stated that Naive Bayesian can predict the possibilities of member classes, such as the possibility of placing new samples in particular classes. Naive Bayesian is based on the Bayesian theorem which always shows fast and precise performance even when applied to very large databases [2].

2 Methods

2.1 Concept of Data Mining

Data mining is the activity of finding interesting patterns from large amounts of data, data can be stored in a database, data warehouse, or other information storage[3]. Data mining is related to other fields of science, such as database systems, data warehousing, statistics, machine learning, information retrieval, and high-level computing. In addition, data mining is supported by other sciences such as neural networks, pattern recognition, spatial data analysis, image databases, signal processing. Data mining is defined as the process of finding patterns in data. This process is automatic or often semi-automatic. The pattern found must be meaningful and the pattern provides benefits, usually economic benefits. Data needed in large numbers [2].

In the calculation of a Naive Bayes represented by h_1 and the second object is represented by h_2 . because what will appear is probabilistic, h is a variable that must be described probabilistically. Furthermore, the probabilities are a priori, $(P h_1)$ and $(P h_2)$ each see the probability of the emergence of objects 1 and object 2. Although the probability of the appearance of the two objects is not known with certainty but at least it can be estimated from the available data. Suppose N is the total number of two objects, then N_1 and N_2 each represent the number of objects 1 and object 2,

$$P(h_1) \approx \frac{N_1}{N} \text{ dan } P(h_2) \approx \frac{N_2}{N}. \quad (1)$$

If $P(h_1)$ is much larger than $P(h_2)$ then it is logical to predict that the most frequent occurrence is object 1. But if $P(h_1) = P(h_2)$, then the probability of predictions will be 50-50

The probability-intensive function, $p(x | h_j)$ is often referred to as the likelihood function of j and h for x . from the additional information in the form of likelihood, $P(x | h_j)$, can be obtained

$$\text{posterior } P(h_j + x) = \frac{p(\frac{x}{h_j})p(h_j)}{P(x)} \text{probability} \quad (2)$$

2.2 Data Collection Method

The method used in data collection through several stages, including:

- a. Observation Stage is the stage that is carried out by observing the object to be observed, including the data of farmers and the criteria used in determining the prediction of agricultural yields of rice.
- b. The interview stage is a stage that is carried out by conducting direct interviews with farmers which include interviews on agricultural returns and turnover which includes

several criteria including the influence of gender, level of education, occupation, position, and age level of a farmer.

- c. The Literature Stage is the stage that is carried out by reading some literature from research journals that are in accordance with the prediction system for determining agricultural production using the Naïve Bayes method[5].

3 Research Result

Data mining processing by using sample data obtained from several farmers of 33 farmers who have answers to several different criteria. From the farmers' answers then data mining processing is carried out which consists of several stages, namely cleaning (data cleaning), data integration (merging of several data) and data normalization which will produce an answer from the most dominant criteria in determining the prediction of rice crop . The criteria used in determining the prediction of agricultural production are in table 1.

Table 1. Criteria for determining agricultural production results.

Criteria's	Values	Values Conversion
Gender	Male	1
	Female	2
Level of education	SMU	1
	S1	2
	S2	3
Type of work	entrepreneur	1
	Private	2
	PNS	3
	Others	4
Position	Top management	1
	Middle management	2
	Low Management	3
Income Per Month	<5 Million	1
	5 Million - 10 Million	2
	15 Million - 20 Million	3
	>20 Million	4
Age Group	< 18 year	1
	18 year - 40 year	2
	>40 year	3

After choosing from several criteria for data mining, then by selecting data samples from several farmers for testing to determine agricultural yield. Farmer sample data for mapping agricultural production results can be seen in table 2.

Table 2. Sample data predicting the determination of the yield of rice plants.

No.	Gender	Level of education	Type of work	Position	Income Per Month	Age Group
1	2	2	2	3	1	2
2	2	2	2	3	1	2
3	2	2	2	2	2	3
4	1	2	2	2	2	3
5	1	2	2	3	2	2
6	1	2	2	3	1	2
7	1	2	2	3	1	2
8	1	2	2	2	2	3
9	2	3	1	1	4	2
10	1	2	2	3	2	2
11	1	2	2	3	1	2
12	2	2	1	1	3	3
13	2	2	2	3	1	2
14	2	2	2	3	1	2
15	1	2	2	2	2	3
16	1	2	2	3	2	2
17	1	3	2	3	1	2
18	1	2	2	3	1	2
19	1	2	2	3	1	2
20	1	2	2	3	1	3
21	2	3	2	3	1	2
22	1	3	1	1	3	3
23	2	2	2	3	2	2
24	1	2	2	3	1	3
25	2	2	2	3	1	2
26	2	3	1	3	1	2
27	2	3	1	1	4	3
28	1	2	2	3	2	2
29	2	2	2	3	2	3
30	2	2	2	3	1	2
31	1	2	2	2	2	2
32	2	3	2	3	1	2
33	2	2	2	3	1	3

From the results obtained by the farmer, each criterion gets different values from each farmer. Which can later be continued into the data mining process stage.

Stages in making data mining:

1. Cleaning data is the cleaning process of data that is not used in the data mining process. In the process of cleaning the data for sample data used, the previous data cleaning process has been carried out. So that in the data cleaning process there is no data that is empty or not used[4].
2. Data integration is data integration must be done carefully because it must integrate unique entities so that they will not get distorted results. The criteria used in mapping the production results are gender, level of education, type of work, position, income per month, age group.
3. Data normalization to determine the productivity of agricultural products

- a. Calculation for the results of determining the productivity of plants with Good Results criteria
- b. Calculation for the results of determining the productivity of plants with poor criteria.

Based on the formula $P(h_1) \approx \frac{N_1}{N}$ and $P(h_2) \approx \frac{N_2}{N}$ in the equation (1)

a. $P(h_1) \approx \frac{N_1}{N}$ then, $P(h_1) = 8/33 = 0.24$

b. $P(h_2) \approx \frac{N_2}{N}$ then $P(h_2) = 25/33 = 0.76$

c. To determine the predictive value of each variable, then use equation (10) to generate the probability value that is $P(a_i|v_j) = \frac{n_c + mp}{n + m}$

d. Description: * n_c = The number of chance opportunities that will be predicted

* m = Determine arbitrary values starting at 0.1 - 1

* p = The specified value is 3

* n = The total amount that will be predicted

4. Calculation of values to determine predictions of good results

$$P(a_i|v_j) = \frac{n_c + mp}{n + m}$$

To determine the gender prediction value, then:

$$\begin{aligned} P(\text{Male I good result}) &= (16 + (3 * 0.5)) / (33 + 3) \\ &= 0.5 \\ P(\text{female I good result}) &= (17 + (3 * 0.5)) / (33 + 3) \\ &= 0.5 \end{aligned}$$

To determine the predicted level of education level, then:

$$\begin{aligned} P(\text{high school I good results}) &= (0 + (3 * 0.5)) / (33 + 3) \\ &= 0 \\ P(\text{S1 bagus Good results}) &= (27 + (3 * 0.5)) / (33 + 3) \\ &= 0.8 \\ P(\text{S1 bagus Good results}) &= (27 + (3 * 0.5)) / (33 + 3) \\ &= 0.8 \\ P(\text{S2 bagus Good results}) &= (6 + (3 * 0.5)) / (33 + 3) \\ &= 0.2 \end{aligned}$$

To determine the predicted value of the Job Type, then:

$$\begin{aligned} P(\text{Entrepreneur I Good results}) &= (5 + (3 * 0.5)) / (33 + 3) \\ &= 0.2 \\ P(\text{PNS Bagus Good Results}) &= (28 + (3 * 0.5)) / (33 + 3) \\ &= 0.8 \\ P(\text{Private Bagus Good Results}) &= (0 + (3 * 0.5)) / (33 + 3) \\ &= 0 \end{aligned}$$

To determine the value of Job prediction, then:

$$\begin{aligned}
P(\text{Top Management bagus Good results}) &= (6 + (3 * 0.5)) / (33 + 3) \\
&= 0.2 \\
P(\text{Middle Management Bagus Good Results}) &= (5 + (3 * 0.5)) / (33 + 3) \\
&= 0.2 \\
P(\text{Low Management Bagus Good Results}) &= (22 + (3 * 0.5)) / (33 + 3) \\
&= 0.7 \\
\text{To determine the predicted value of income per month, then:} \\
P(<5 \text{ million I Good results}) &= (18 + (3 * 0.5)) / (33 + 3) \\
&= 0.5 \\
P(5-10 \text{ million I Good Results}) &= (10 + (3 * 0.5)) / (33 + 3) \\
&= 0.3 \\
P(15-20 \text{ million I Good Results}) &= (2 + (3 * 0.5)) / (33 + 3) \\
&= 0.1 \\
P(> 20 \text{ million Good Results}) &= (22 + (3 * 0.5)) / (33 + 3) \\
&= 0.1 \\
\text{To determine the predicted value of the age group, then:} \\
P(<18 \text{ years I Good results}) &= (0 + (3 * 0.5)) / (33 + 3) \\
&= 0 \\
P(18-40 \text{ years old Good Results}) &= (22 + (3 * 0.5)) / (33 + 3) \\
&= 0.7 \\
P(> 40 \text{ years Good Results}) &= (11 + (3 * 0.5)) / (33 + 3) \\
&= 0.3
\end{aligned}$$

So, the predictive value for the example above is $0.5 * 0.8 * 0.8 * 0.7 * 0.5 * 0.7 = 0.0784$.

Problems that usually occur in agriculture are agricultural products that are not increasing, this is due to the patterns of life and problems that exist in the farmer. Predictions to improve agricultural yields use several criteria, namely gender, position, occupation, age, and income. And the most dominant criterion in predicting agricultural results is age and position in the job. The results of these predictions can be used to increase income and welfare turnover for farmers.

4 Conclusion

Rice plants are a staple of agricultural production in Indonesia, many of the farmers do not have enough time to maintain their agricultural products. So that many of these farmers have poor crop yields. The Naïve Bayes method is used to predict rice plants by using several criteria

used in determining the yield of rice production is the sex of a farmer, level of education, type of work, position, monthly income and the age level of farmers. But after the prediction is done using the Naïve Bayes method, the most dominant criteria in predicting the yield of rice crops is the level of education and occupation at work.

References

- [1] K. Asih, F. A. Setyaningsih, and D. M. Midyanti, "Aplikasi Prediksi Produksi Padi Menggunakan Regresi Interval Dengan Neural Fuzzy Di Kabupaten Kubu Raya," *Coding, Sistem Komput. Untan*, vol. 05, pp. 108–118, 2017.
- [2] M. Ramaswami and R. Bhaskaran, "A CHAID Based Performance Prediction Model in Educational Data Mining," *IJCSI Internatuonal J. Comput. Sci*, vol. 7, pp. 10–18, 2010.
- [3] S. Conference, "DESIGN AND IMPLEMENTATION OF," in in SEAMS-GMU, 2007, pp. 1–8.
- [4] R. A. Santoso, D. Syauqy, M. Hannats, and H. Ichsan, "Pengembangan Sistem Prediksi Hama Wereng Berdasarkan Data Cuaca Sensor Dan Cuaca Online Menggunakan Metode Naive Bayes," vol. 2, no. 10, pp. 4002–4010, 2018.
- [5] M. Syarief et al., "Penerapan Metode Naïve Bayes Classifier Untuk Deteksi Penyakit pada tanaman jagung," *J. Ilm. NERO*, vol. 3, no. 1, pp. 61–68, 2017.