No. 341 | Received: 2020-09-21 | Revised: 2020-10-13 | Accepted: 2020-10-15

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
http://tip.ppj.unp.ac.id

Jurnal Teknologi Informasi dan Pendidikan
Volume 13, No. 2, September 2020
P. 50 – 54 | https://doi.org/10.24036/tip.v13i2

SVM ALGORITHM FOR PREDICTING RICE YIELDS

Agus Setia Budi[1], Purnomo Hadi Susilo[2], Nur Nafi' iyah[3*]
[1,2,3]Program Studi Teknik Informatika, Fakultas Teknik, Universitas Islam Indonesia, Indonesia
Corresponding Author: mynaff26@gmail.com

ABSTRACT

Agriculture in Indonesia is highly dependent on reservoir, irrigation water sources and rain. Some agricultural land in Indonesia is rain-fed. Plants in Indonesia rely on water from rain and irrigation. Weather conditions greatly affect the number of farmers' harvest. Farmers often experience crop failures due to changing weather. From data from the Central Statistics Agency, it is stated that the amount of rice harvested in 2019 decreased by 7.76% compared to 2018. In order to avoid rice imports and rice food shortages, something is needed that can help the government in making policies. One of them is creating a rice yield prediction system. This research process consists of collecting data via the web. The data shows the variables of province, year, land area, production. The total number of data is 170 rows, with a division of 130 lines for training, and 40 for testing. Furthermore, the data is processed and processed and normalized. The results of data processing are then trained and predicted with a linear SVM kernel. The results of SVM prediction with original data without normalization of MAPE 6635.53%, and RMSE 1094810.74. The results of SVM prediction with normalized data first, the MAPE value is 9427.714%, and RMSE is 0.017.

Keywords: *SVM, prediction of rice yield, normalization.*

PENDAHULUAN

Paddy land area in 2019 is around 10.68 million hectares and has decreased by 700.05 thousand hectares or 6.15% when compared to 2018. Production of rice harvest in 2019 is 54.60 million tons and decreased by around 4.60 million tons or 7.76% when compared to 2018. The data come from the Indonesian Central Bureau of Statistics.

Hartato's research states that by making a prediction system for Biopharmaca harvest area in Indonesia, it can help the government in making policies so that there is no decrease in supply [1].

Agriculture in Indonesia is very diverse and several agricultural sources are the main or staple food source, for example rice. From data from the Indonesian Central Statistics Agency, it was stated that in 2019 there was a decline in rice harvest production of 7.76%. So that rice imports are one way out to meet food needs [2]. The productivity or yield of rice yields decreases due to changes in climate, or weather, and changes in nutrients in the soil. So that farmers must anticipate by planning a planting schedule that is adjusted to the current climate and weather conditions [3]. Another way to produce rice yields is by making planting models legowo (30x25xlarikan), legowo (30x25x12,5), legowo (30x20xlarikan), and legowo (30x20x10). In this way it can increase the yield of upland rice varieties Situ Patenggang [4].

Rice is a staple food source for several residents in Indonesia. Processed rice will produce rice and contain lots of carbohydrates. Because rice is a source of carbohydrates and is the main source of consumption by Malang residents, a rice yield forecasting system using the SVR (Support Vector Regression) method was developed. However, the MAPE results showed 10.133% [5]. In addition to using the SVR method, to estimate crop yields can use multiple linear regression algorithms. The result of prediction of plant productivity using multiple linear regression with standard error is 206.3075 [6] [9]. The KNN method can also be used to predict rice yields, with an absolute error value of 80876.83 [7] [10].

Indonesia is an agricultural country whose largest agricultural product is rice. Rice is a staple or the main source of food for the Indonesian population. So it is hoped that the yield or rice production in Indonesia cannot decrease. In order to avoid a decrease in rice yields, it is necessary to make an estimate of the yield of rice crops. The

method that can be used to estimate rice yields is the Support Vector Machine [5] [8]. The purpose of this study is to predict rice yields with SVM in order to assist the government in determining future policies.

This research will discuss:

1. How to implement the SVM algorithm to predict rice yields based on provincial, year and land area variables.
2. How to implement the SVM algorithm to predict rice yields with normalized data first with Equation 1.

METHOD

Preprocessing Data

The data used in this study is data on crop production in Indonesia. Data taken from the website: https://www.pertanian.go.id/home/?show=page&act=view&id=61. The data is still in pdf form and each variable is still separate. The variables of this data are province, year, land area, production, and productivity. Land area is the area of land planted with rice in hectares. Production is the yield of rice harvest per year in tonnes. Productivity is the yield divided by the area of land in units (tonnes/hectare). The data used are 34 provinces in Indonesia from 2014-2018. Total data is 170 rows, and is divided into 130 lines for training, and 40 lines for testing.

The data used were also normalized by Equation 1 [1].

$$v\_new = \frac{v - v\_\min}{v\_maks - v\_\min} * (v\_new\_maks - v\_new\_\min) \quad (1)$$

Proposed Method

Prediction is determining a future value, the data is in the form of continuous data, prediction or estimation is prediction [11]. Predicting must do pattern recognition and must study how that data can be modeled. Predicting the SVM algorithm must carry out a training process whose goal is to find the optimal position of the hyperplane [12] [13] [14].

Figure 1: 130 rows of dataset consisting of input variables in the form of province, year, land area and output variables are production or crop yields. Training with the SVM algorithm, the training results are in the form of line equations consisting of weights and biases. And as many as 40 lines of test data were tested to produce predictive data, then calculated the MAPE [5] and RMSE [9] values.

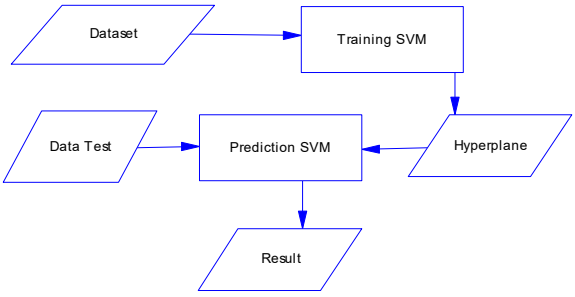

Figure 1. SVM Block Diagram

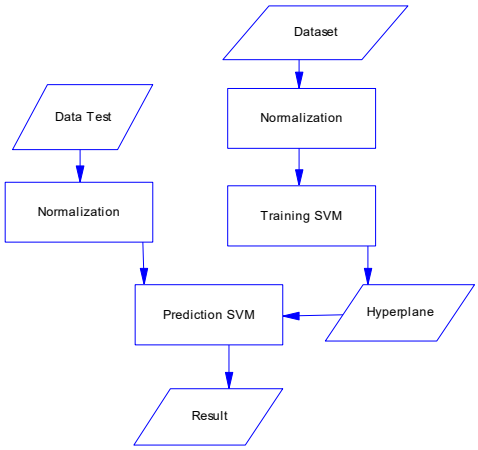

Figure 2. SVM Block Diagram with Normalized Data

Equations 2 and 3 are sequentially used to calculate the MAPE and RMSE error values.

$$MAPE = \sum_{i=1}^{n} \left| \frac{t_i - y'_i}{y'_i} \right| x100\% \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum (t_i - y'_i)^2} \quad (3)$$

Note: t is the target or real data, while y 'is the predicted data, and n is the total test data.

Figure 2: The input data is normalized first with Equation 1. Then the data from normalization results are carried out by SVM training and test data are tested, the results are in the form of predictive data.

The data in Table 1 is input data with variables of province, year, land area, and crop yield. The data in Table 2 are data that have been normalized.

The data in Table 1 is that 34 provinces in Indonesia are symbolized by numbers, 1 means Aceh province, 2 is North Sumatra province. Then the year is data from 2014 to 2018. Land area is the area of land planted with rice in hectares. And production is the yield per year in tonnes. The data in Table 2 is data that has been normalized using Equation 1. Each variable looks for a maximum and minimum value, then a new value is calculated with a range of [0 1].

Jurnal Teknologi Informasi dan Pendidikan
Volume 13, No. 2, September 2020
http://tip.ppj.unp.ac.id

The results data from the website include the number, name of the province and year. The data is then pre-processed, changing the name of the province to a number, for example Aceh province to 1, North Sumatra province 2, and so on. Initial processing results data obtained column province number, year area and crop yields. Total data is 170 rows, and divided into 130 lines of training data and 40 lines of testing data. The next steps to carry out normalization, for example, the normalization process are:

Column area, the maximum area is 2285232 Ha, and the minimum area is 186 Ha, normalized to [0 1]. New data minimum 0 and maximum old data 1. The result of data normalization is 376137 Ha.

$$data\_new = \frac{376137 - 186}{2285232 - 186} * (1 - 0) = 0.1645$$

Table 1. Dataset

Province	Year	Land Area (Ha)	Production (ton)
1	2014	376137	1820062
1	2015	461060	2331046
1	2016	429486	2205056
1	2017	470351	2494613
1	2018	463485	2516221
2	2014	717318	3631039
2	2015	781769	4044829

Table 2. Data Normalization

Province	Year	Land Area (Ha)	Production (ton)
0.00	0.00	0.16	0.13
0.00	0.25	0.20	0.17
0.00	0.50	0.19	0.16
0.00	0.75	0.21	0.18
0.00	1.00	0.20	0.18
0.03	0.00	0.31	0.27
0.03	0.25	0.34	0.30

RESULTS AND DISCUSSION

1. Yield Prediction using SVM

In predicting crop yields using the SVM algorithm, the MAPE value is 6635.53%, and the RMSE value is 1094810.74. The prediction process using the SVM algorithm uses a linear kernel. And the results of the training are the weights and biases used to predict the test data. Weights and bias in Table 3.

Table 3. Training Result Weight and Bias

Weight	Bias
-31.67	-1291356.6
-2.42	
6.69	

The prediction results from the SVM algorithm are in Table 4. The actual data is the test data that will be tested, while the prediction data is the data from the calculation of the equation from the SVM training results.

Table 4. Test Prediction Results

Actual Data	Prediction Results
5426097	5657146
5471806	5683930
5727081	6252905
6055404	6652682
6196737	6477784
657617	-358267
660720	-358457
695329	-139553
711401	-151056
716156	-152851
314704	-877969
331220	-898178
344869	-854510
350193	-780893
350256	-778930
449621	-666296
461844	-672189
548536	-485294
667100	-355442
751531	-170172
102761	-1152631
117791	-1155856
99088	-1153525
104716	-1125136
132852	-1087500
72074	-1155545
75265	-1153902
82213	-1128322
84037	-1113520
101054	-1096010
27665	-1251275
30219	-1249312
27840	-1254690
29516	-1252479
27736	-1252575
196015	-993117
181769	-1020795
233599	-959642
257888	-946030
288335	-891356

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390

2. Normalized Data Predicted by the SVM Algorithm

Table 5. Training Result Weight and Bias Data Normalization

Weight	Bias
-0.00243	-0.0172011
0.00293	
0.97071	

Table 6. Training Prediction Results Data Normalization

Actual Data	Prediction Results
0.40	0.423
0.40	0.425
0.42	0.462
0.44	0.488
0.45	0.478
0.05	0.040
0.05	0.041
0.05	0.056
0.05	0.056
0.05	0.056
0.02	0.007
0.02	0.007
0.03	0.010
0.03	0.016
0.03	0.017
0.03	0.021
0.03	0.021
0.04	0.034
0.05	0.043
0.06	0.055
0.01	-0.010
0.01	-0.010
0.01	-0.009
0.01	-0.006
0.01	-0.003
0.01	-0.011
0.01	-0.010
0.01	-0.007
0.01	-0.006
0.01	-0.004
0.00	-0.017
0.00	-0.016
0.00	-0.015
0.00	-0.015
0.00	-0.014
0.01	0.000
0.01	-0.001
0.02	0.003
0.02	0.005
0.02	0.009

Prediction using the SVM algorithm where the data is normalized first, the result is the MAPE value is 9427.714, and the RMSE value is 0.017. The results of the SVM training are shown in Table 5. The training results are in the form of weight and bias. And the test results of the test data are shown in Table 6. The SVM training process uses a linear kernel. The process of calculating MAPE and RMSE is based on Table 6, namely the real data value minus the predicted value.

Table 7 describes the results of the MAPE and RMSE accuracy of rice yield predictions using the SVM algorithm, the SVM regression algorithm, and the linear regression algorithm. The SVM column uses the fitrsvm function, and the SVM regression algorithm uses the fitrlinear function, and the linear regression algorithm uses the regress function. Table 8 describes the results of the MAPE and RMSE accuracy from normalized rice data.

Table 7. Accuracy Results

Type	SVM	SVM Regresi	Regresi Linear
MAPE	6635.53	3008.31	4024
RMSE	1094810.74	1319401.1081	104449563.2

Table 8. Accuracy Results Data Normalization

Type	SVM	SVM Regresi	Regresi Linear
MAPE	9427.714	10764	1576
RMSE	0.017	0.0979	0.0996

CONCLUSION

The SVM algorithm in predicting rice yields in 34 provinces in Indonesia produces MAPE values of 6635.53 and RMSE 1094810.74. This study also compares the results of the SVM regression algorithm on data that is not normalized with MAPE values of 3008.31, RMSE 1319401.1. The results of the linear regression algorithm are MAPE value 4024, RMSE 104449563.2. Furthermore, the SVM algorithm is also used to predict rice yields before the data is normalized with MAPE values of 9427.714 and RMSE 0.017. While the MAPE and RMSE values with the SVM regression algorithm and linear regression on the normalized data, MAPE and RMSE, respectively, were 10764, 0.0979 and 1576, 0.0996.

SUGGESTION

Predict rice yields with other algorithms, or with the SVM algorithm with kernels other than linear.

BIBLIOGRAPHY

[1] E. Hartanto, D. Sitorus and A. Wanto, "Analisis Jaringan Saraf Tiruan untuk Prediksi Luas Panen Biofarmaka di Indonesia," *SEMANTIK,* vol. 4, no. 1, pp. 49-56, 2018.

[2] F. Maspiyanti and M. I. Fanany, "Klasifikasi Fase Pertumbuhan Padi Berdasarkan Citra Hiperspektral dengan Modifikasi Logika Fuzzy," *Jurnal Pengindraan Jauh,* vol. 10, no. 1, pp. 41-48, 2013.

[3] H. Herminingsih, "Hubungan Adaptasi Petani Terhadap Perubahan Iklim dengan Produktivitas Tembakau pada Lahan Sawah dan Tegalan di Kabupaten Jember," *JSEP,* vol. 7, no. 2, pp. 31-44, 2014.

[4] S. Putra, "Pengaruh Jarak Tanam Terhadap Peningkatan Hasil Padi Gogo Varietas Situ Patenggang," *Agrin,* vol. 15, no. 1, pp. 54-63, 2011.

[5] D. A. Mardhika, B. D. Setiawan and R. C. Wihandika, "Penerapan Algoritma Support Vector Regression Pada Peramalan Hasil Panen Padi Studi Kasus Kabupaten Malang," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* vol. 3, no. 10, pp. 9402-9412, 2019.

[6] E. I. A. Warih and Y. Rahayu, "PENERAPAN DATA MINING UNTUK MENENTUKAN ESTIMASI PRODUKTIVITAS TANAMAN TEBU DENGAN MENGGUNAKAN ALGORITMA LINIER REGRESI BERGANDA DI KABUPATEN REMBANG," Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang.

[7] W. T. Panjaitan, E. Utami and H. Al-Fatta, "PREDIKSI PANEN PADI MENGGUNAKAN ALGORITMA K-NEAREST NEIGBOUR," in *SNATIF Universitas Muara Kudus*, Kudus, 2018.

[8] Fatkhuroji, S. Santosa and R. A. Pramunendar, "PREDIKSI HARGA KEDELAI LOKAL DAN KEDELAI IMPOR DENGAN METODE SUPPORT VECTOR MACHINE BERBASIS FORWARD SELECTION," *Jurnal Teknologi Informasi,* vol. 15, no. 1, pp. 61-77, 2019.

[9] A. Mayasari and Y. Rahayu, "PENERAPAN ALGORITMA LINIER REGRESSION UNTUK MENENTUKAN ESTIMASI LUAS LAHAN PANEN TANAMAN JAGUNG TERHADAP CURAH HUJAN DAN AREA TAMBAH TANAM DI KABUPATEN REMBANG," Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang.

[10] R. Nuari, A. Apriliyani, Juwari and Kusrini, "IMPLEMENTASI METODE K-NEAREST NEIGHBOR (KNN) UNTUK MEMPREDIKSI VARIETAS PADI YANG COCOK UNTUK LAHAN PERTANIAN," *Jurnal INFORMA Politeknik Indonusa Surakarta,* vol. 4, no. 2, pp. 28-34, 2018.

[11] D. Rohmayani, "ANALYSIS OF STUDENT TUITION FEEPAY DELAY PREDICTION USING NAIVE BAYES ALGORITHM WITH PARTICLE SWARM OPTIMATION OPTIMAZATION (CASE STUDY : POLITEKNIK TEDC BANDUNG)," *Jurnal Teknologi Informasi dan Pendidikan,* vol. 13, no. 2, pp. 1-8, 2020.

[12] E. Prasetyo, Data Mining Konsep dan Aplikasi Menggunakan Matlab, Yogyakarta: Andi, 2012.

[13] B. Santosa, Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis Teori dan Aplikasi, Yogyakarta: Graha Ilmu, 2007.

[14] B. Santosa, Data Mining Terapan dengan Matlab, Yogyakarta: Graha Ilmu, 2007.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390