

PAPER • OPEN ACCESS

Support Vector Regression Algorithm Modeling to Predict the Availability of Foodstuff in Indonesia to Face the Demographic Bonus

To cite this article: Sari Devia Agustina *et al* 2018 *J. Phys.: Conf. Ser.* **1028** 012240

View the [article online](#) for updates and enhancements.

Related content

- [Elementary school students visual spatial comprehension based on van Hiele Theory: the case in Madiun, East Java, Indonesia](#)
S Andini, L Fitriana and Budiyono
- [Geo-electrical and geological strikes of the Mount Lamongan geothermal area, East Java, Indonesia – preliminary results](#)
L R Nugraheni, S W Niasari and M Nukman
- [The Strategic Industry of East Java Economy](#)
Arman and Maulidian

Recent citations

- [Mathematical and geophysical methods for searching anomalies of the Radon signal related to earthquakes](#)
F Ambrosino *et al*



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Support Vector Regression Algorithm Modeling to Predict the Availability of Foodstuff in Indonesia to Face the Demographic Bonus

Sari Devia Agustina^{1,4*}, Mustakim^{2,4}, Okfalisa³, Celsa Bella^{1,4}, Muhammad Anang Ramadhan^{1,5}

¹Departemen of Information System, UIN Sultan Syarif Kasim Riau, Pekanbaru 28293, Indonesia

²Laboratory of Data Mining Departemen of Information System, UIN Sultan Syarif Kasim Riau, Pekanbaru 28293, Indonesia

³Departemen of Informatic Engineering, UIN Sultan Syarif Kasim Riau, Pekanbaru 28293, Indonesia

⁴Research Group of Puzzle Research Data Technology, Faculty of Science and Technology, UIN Sultan Syarif Kasim Riau, Pekanbaru 28293, Indonesia

*sari.devia.agustina@student.uin-suska.ac.id

Abstract. The demographic bonus becomes a valuable phenomenon for Indonesian. One of the positive effects of this phenomenon is the increase of productive age proclaimed which will be the future of Indonesian economy. The agricultural sector plays an important role of the overall national economy which is indicated by an increase from year to year. However, the level of nutritional adequacy declined by a few percent each year due to an increase in the number of people who are not balanced by increased demand for food. In this case the government is expected to determine the policy priorities related to Demographic Bonus issues by predicting the future. Computing and data mining technologies play an important role in prediction cases by drawing conclusions based on regression lines. The technique is called Support Vector Regression, which is able to handle some cases of statistical data. Three determinant attributes used in this research are (1) Harvest Area; (2) Number of Harvest Production; and (3) Food Productivity, become the main reference for 714 data from 1998-2015 in 34 Provinces in Indonesia containing 7 types of crops. Three distribution data experiments conducted using K-Fold Cross Validation have the highest accuracy on Fold-1 with correlation coefficient value (R) of 92% with the smallest error value at fold-1 with MSE value of 14%. Predicted results show a decline in the number of food production in almost every province in Indonesia. From the experimental results, it is known that the biggest contributor of food products is Java Island, especially in East Java. Almost every kind of palawija plant, East Java plays an important role in the production of food needed in Indonesia.

1. Introduction

The demographic bonus becomes a valuable phenomenon for Indonesian nation. One of the positive effects of this phenomenon is the increasing of productive age which is proclaimed to become the stakeholder of Indonesia's economy. With a good preparation and the right investment, demographic



bonuses can change the future of Indonesia to be more prosperous and advanced. Nevertheless, according to data from Indonesian Central Bureau of Statistics (BPS) in 2010, Indonesia's population growth rate is still showing worrying sign because it reaches 1.49% or about 4.5 million. The Association of Indonesian Internet Service Providers (APJII) estimates that by 2019 the number of people will increase [1]. High population growth will also increase people's need for food. Therefore, one of the factors that should be the focus of development towards demographic bonus Indonesia is in terms of food security which is part of the agricultural sector [2].

The agricultural sector plays an important role for the overall national economy which can be demonstrated by the large number of people or workers who live or work in agricultural sector or from national products derived from agricultural sector [3]. One of agriculture sectors which become the main priority in National Medium-Term Development Plan (RPJMN) 2015-2019 is Food Security which focuses on increasing food availability, stabilizing food distribution, accelerating food diversification, and supervising fresh food safety. On the other hand, the development of food security is implemented in an effort to increase economic growth and reduce poverty as a manifestation of social, cultural and economic development as part of overall development [4].

The availability of food and employment is a factor which determines the welfare of Indonesian people. Sufficient food availability is influenced by the increase of employment, but the fact is that the number of agricultural workers has decreased [5]. Indonesia currently has 90 million farmers or about 45% of the population feeding the entire population (about 230 million people). It can be seen that the most of staple food comes from cereals consisting of rice, corn and flour and the largest staple food of the people is rice [6].

When calculated from the level of nutritional adequacy, the level of food security in Indonesia during the period 2009-2013 has decreased. In 2013 the number of population in the category of food-security amounted to 116 million people or 47.09% of the population [7]. With the increasing number of population every year which is not matched by the increase of food demand, and the decrease of food security in Indonesia is also accompanied by the decline of labor in agriculture sector. This is a serious problem which is to be the focus of government. Especially in the face of demographic bonus in the year 2020-2030, at that time the population of Indonesia will reach 305 million. This means that the number of people who need food will be more or in other words, the national food needs will increase. The fulfillment of these food needs is absolutely supported by the agricultural sector [8].

Increasing the number of residents must be balanced with adequate food availability. Therefore, a strategic plan to realize the sufficient food availability is by doing the planning. The government is expected to make policy priorities related to food availability to face demographic bonus by making decisions both in the current and future periods in the form of predictions.

The development of computational technology has generated many models of computational intelligence which can be applied to settle various cases such as prediction. Various research has been done by applying the Support Vector Regression (SVR) method to solve prediction cases with good accuracy compared to other methods. SVR is a method that can overcome the overfitting, resulting in an excellent performance [9], giving conclusions about the advantages and accuracy results [10] and can be applied to various cases with continuous data [11]. Previous research using SVR conducted by Mustakim, Agus Buono and Irman Hermandi in 2015 shows the accuracy in predicting the productivity of oil palm in Riau Province [12] [22].

The main objective of this research is to apply SVR to predict the amount of Indonesian food security to face the demographic bonus for the future. With some predictor parameters as well as using data of 10 years of food security.

2. Method and Materials

The research data is obtained from the official website of Central Bureau of Statistics (BPS) www.bps.go.id [5]. Based on the data of food security in all provinces of Indonesia from 1998-2015 with 714 records. Food security itself is a condition of the food fulfillment for the household which is reflected from the availability of adequate food, both quantity and quality, safe, equitable and

affordable (physical or economic). The types of food crops are (1) Rice, (2) Corn, (3) Cassava, (4) Sweet Potato, (5) Peanuts, (6) Green Beans and (7) Soybean. These types of food data includes of Harvested Area data, Total Production and Productivity on each type of food crop which is also part of the attribute.

The data processing stage is divided into three stages: normalization for each target variable, dividing the training data and testing data from the whole data and the last stage is predicting the pattern using SVR method. After normalization, the next step is to divide the data into two parts: training data set and testing data set. From the experiments conducted, the optimum result is obtained as training data and data testing that will be used to predict the food security using RBF kernel.

Some steps in the prediction experiments using SVR in this case are shown in Figure 1.

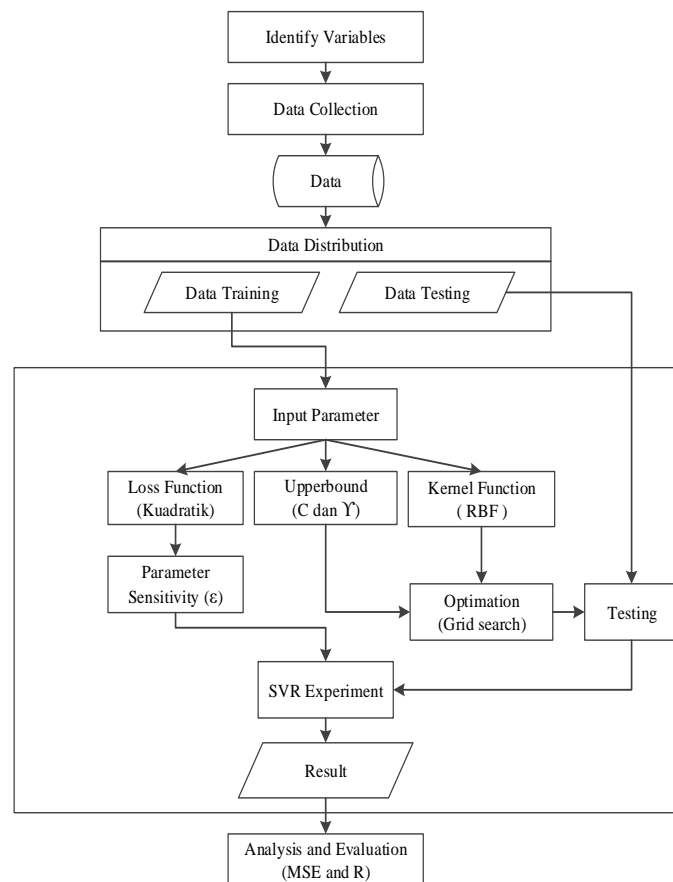


Figure 1. Research Methodology

2.1. Demography Bonus

The demographic bonus is the future in which the burden of dependence rate between the productive age population and the unproductive age population has decreased to a level below 50. In other words every working age population bears a little of the unproductive age population. To get the demographic bonus, the quality of human resources must be maximally enhanced through education, health services and the provision of employment. Population projection data shows that Demographic Bonus in Indonesia is estimated to occur in the period of 2020 to 2030 [13].

2.2. Food Security

The availability of adequate and distribution of food evenly is one of the pillars of the realization of sustained and sustainable food security (UU No. 7 of 1996 on food, and explanation of PP No. 68 of 2002). The objective of development in the field of food is to strengthen the food security which is

characterized by the increase of food availability and increasing diversification of food consumption. In the session of FAO Food Security Committee in 1991 and the Law of Republic Indonesia Number 71 1996 on food, explained that the food security is a condition of food fulfillment for the household which is reflected from the availability of adequate food, both quantity and quality, safe, equitable and affordable (physical or economic) [14].

2.3. Support Vector Regression (SVR)

SVR is an application of Support Vector Machine (SVM) for regression cases. SVR is a method which can overcome the overfitting, so it will produce a good performance [15]. Smola and Scholkopf (2003) described the SVR by assuming a condition which has a set of λ training data, (x_j, y_j) with $j = 1, 2, \dots, \lambda$ with input $x = \{x_1, x_2, x_3\} \subseteq \mathbb{R}^N$ and the corresponding output $y = \{y_1, \dots, y_\lambda\} \subseteq \mathbb{R}^N$. With SVR, we will find a function of $f(x)$ having the largest deviation ε of the actual target for all training data. Then by using SVR, when the value of ε is equal to 0, perfect regression will be obtained. Based on the data, the SVR wants to find a function of $f(x)$ that can approximate the output to an actual target, with a ε -tolerance error, and minimal complexity. The regression function of $f(x)$ can be expressed by the following formula [15]:

$$f(x) = w^T \varphi(x) + b \quad (1)$$

Where $\varphi(x)$ denotes a point in the higher-dimensional feature space, the mapping results of input vector x in input space which has lower dimension. The coefficients of w and b are estimated by minimizing the risk function defined in the equation:

$$\min \frac{1}{2} \|w\|^2 + C \frac{1}{\tau} \sum_{i=1}^{\tau} L_{\varepsilon}(y_i, f(x_i)) \quad (2)$$

Which fulfill:

$$\begin{aligned} y_i - w\varphi(x_i) - b &\leq \varepsilon \\ w\varphi(x_i) - y_i + b &\leq \varepsilon, i = 1, 2, \dots, \tau \end{aligned} \quad (3)$$

Where,

$$L_{\varepsilon}(y_i, f(x_i)) = \begin{cases} |y_i - f(x_i)| - \varepsilon & |y_i - f(x_i)| \geq \varepsilon \\ 0 & \text{for others} \end{cases} \quad (4)$$

2.4. Min-Max Normalization

Data normalization is part of data transformation, which is the technique to convert data into a more easily understood [16]. The goal of data normalization is to obtain the same weight from all attributes of data and doesn't vary or the result of weighting has no dominant attribute or considered more important than the others [17].

2.5. K-Fold Cross Validation

K-fold cross validation is performed to divide the data into training set and testing set [18]. The point of this type of validation is to divide the data randomly into desired subset of k . K-fold cross validation repeats k -times to split a set randomly into the most free set of section [19].

2.6. Grid Search

The Grid search method is one of the simplest methods to overcome the optimization problem [20]. This method involves of composing a suitable grid in a dimensional space, evaluating the objective function of the entire grid point, and finding the grid point corresponding to the objective function which has the optimum value [21].

3. Result and Discussions

SVR requires parameters which match the kernel in training. To get the optimal kernel, optimization using grid search is done in training stage. There are 2 parameters which are optimized by grid search, they are parameter C and parameter γ . Parameter C is penalty value against SVR model error, while parameter γ is parameter used as input for kernel functions to be used. The best RBF kernel simulation results in SVM are shown in Table I below.

Table 1. RBF kernel simulation in Fold 1

| γ | C | | 0.5 | 1 | 2 | 4 |
|----------|---|-----|--------|--------|--------|--------|
| | | | 1 | 2 | 4 | 8 |
| 0.5 | 1 | MSE | 0.0020 | 0.0023 | 0.0034 | 0.0039 |
| | | R | 0.9251 | 0.8994 | 0.6512 | 0.501 |
| 0.25 | 2 | MSE | 0.0017 | 0.0020 | 0.0026 | 0.0022 |
| | | R | 0.9261 | 0.9210 | 0.8499 | 0.8716 |
| 0.125 | 4 | MSE | 0.0014 | 0.0015 | 0.0015 | 0.0016 |
| | | R | 0.7887 | 0.7861 | 0.7674 | 0.7532 |
| 0.0625 | 8 | MSE | 0.0026 | 0.0026 | 0.0026 | 0.0024 |
| | | R | 0.5348 | 0.5329 | 0.5258 | 0.5783 |

From each experiment, two best simulations were obtained through the largest correlation coefficient (R) and the smallest Mean Squared Error (MSE) error, using the formula;

$$MSE = \frac{1}{m} \sum_{i=1}^m (\hat{y}_i - y_i)^2 \quad (5)$$

and

$$R = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2} \sqrt{\sum(y - \bar{y})^2}} \quad (6)$$

The simulation results will be divided into two experiments, among them (1) Experiment 1, is the predicted result with the best simulation result through the smallest MSE value. (2) Experiment 2, is the result of prediction with the best simulation result through the biggest R value:

Table 2. Result of modeling on Fold 1

| ALT | ACTUAL | EXPERIMENT 1 | EXPERIMENT 2 |
|-------|---------|--------------|--------------|
| P-1 | 0.00314 | -0.00032 | -0.00015 |
| P-2 | 0.02098 | -0.00033 | -0.00004 |
| P-3 | 0.00435 | -0.00042 | -0.00029 |
| P-4 | 0.00169 | -0.00008 | -0.00001 |
| P-5 | 0.00097 | 0.00000 | 0.00003 |
| P-6 | 0.00307 | -0.00016 | -0.00007 |
| P-7 | 0.00245 | -0.00014 | -0.00005 |
| P-8 | 0.03343 | 0.00164 | 0.00105 |
| P-9 | 0.00398 | 0.00006 | 0.00006 |
| P-10 | 0.0093 | 0.00006 | 0.00006 |
| ... | ... | ... | ... |
| P-714 | 0.00001 | 0.00006 | 0.00006 |

Table 2 describes the result obtained from SVR modeling using fold-1. From the modeling we get the result of RBF kernel with MSE of 0.0014 R value of 0.7887 in experiment 1 and MSE 0.0017 R

value of 0.9261 in experiment 2. The predicted result graph of fold-1 in each experiment can be seen in Figures 2 and 3.

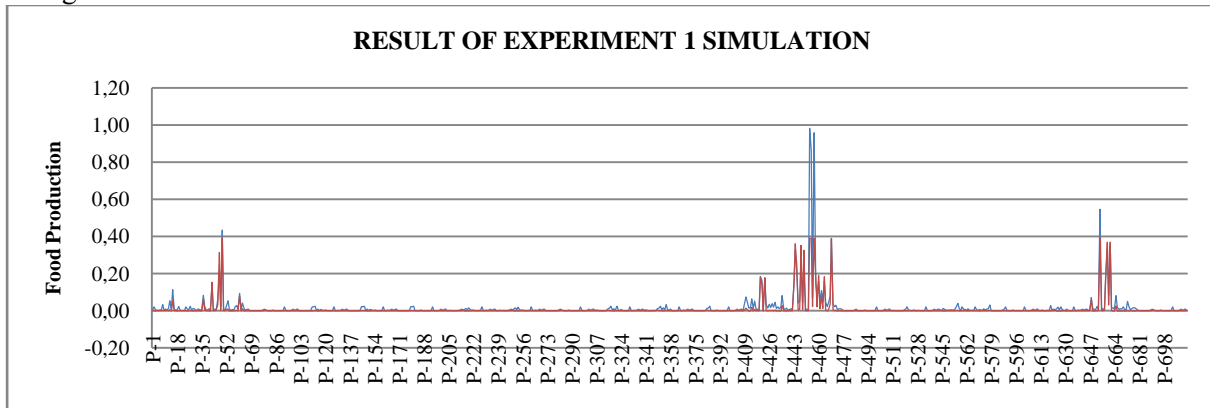


Figure 2. Predicted graph of first modeling results

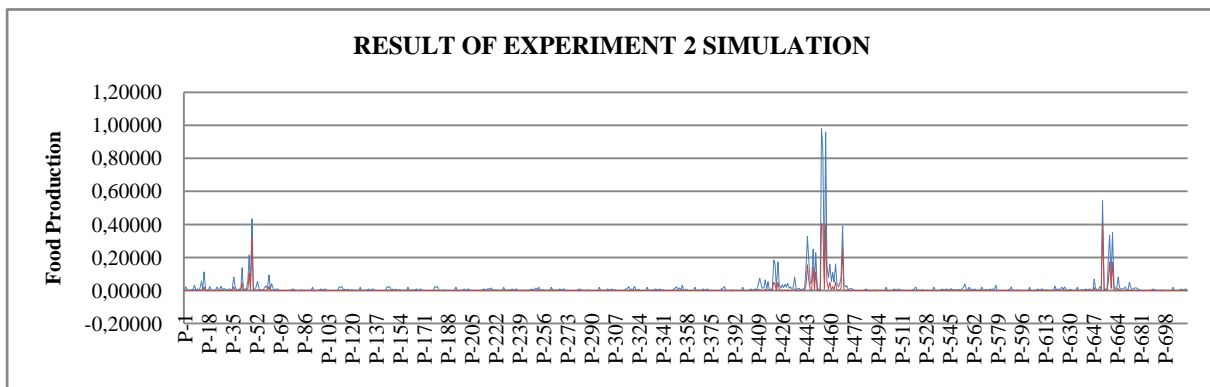


Figure 3. Predicted graph of second modeling results

From figure 2 and 3 above can be known what is related to predict the next production, it obtained the best model on SVR kernel that the RBF with correlation coefficient (R) equal to 0,9261 and mean squared error (MSE) equal to 0,0014 is found in fold 1. Predicted results by comparing actual (pre-predicted) with predictions on each food requirement were obtained by focusing on three provinces contributed to the highest production of foodstuffs that can be seen in Table 3.

Table 3. Food demand predictions based on the largest contributors on production

| No | Food | Province with the Most Contribution of Production | The prediction results are based on the coefficient of R and MSE | | | | | |
|----|-------------|---|--|------------|----------|---------|------------|----------|
| | | | MSE | | | R | | |
| | | | Actual | Prediction | Result | Actual | Prediction | Result |
| 1 | Rice | West Java | 0.98209 | 0.38951 | Decrease | 0.98209 | 0.40198 | Decrease |
| | | Central Java | 0.8612 | 0.38951 | Decrease | 0.8612 | 0.40118 | Decrease |
| | | East Java | 0.95734 | 0.38951 | Decrease | 0.95734 | 0.40198 | Decrease |
| 2 | Corn | East Java | 0.43478 | 0.38916 | Decrease | 0.43478 | 0.31396 | Decrease |
| | | Central Java | 0.21595 | 0.31417 | Increase | 0.21595 | 0.10455 | Decrease |
| | | South Sulawesi | 0.09282 | 0.07976 | Decrease | 0.09282 | 0.02201 | Decrease |
| 3 | Green beans | Central Java | 0.00784 | -0.00061 | Decrease | 0.00784 | -0.00052 | Decrease |
| | | East Java | 0.0075 | -0.0004 | Decrease | 0.0075 | -0.00022 | Decrease |
| | | South Selatan | 0.00312 | -0.00004 | Decrease | 0.00312 | -0.00009 | Decrease |
| 4 | Peanuts | East Java | 0.01864 | -0.00053 | Decrease | 0.01864 | -0.00022 | Decrease |

| No | Food | Province with the Most Contribution of Production | The prediction results are based on the coefficient of R and MSE | | | | | |
|----|--------------|---|--|------------|----------|---------|------------|----------|
| | | | MSE | | | R | | |
| | | | Actual | Prediction | Result | Actual | Prediction | Result |
| 5 | Soybeans | Central Java | 0.01483 | -0.00093 | Decrease | 0.01483 | -0.00029 | Decrease |
| | | DI Yogyakarta | 0.00574 | -0.00036 | Decrease | 0.00574 | -0.00021 | Decrease |
| | | East Java | 0.03283 | 0.00078 | Decrease | 0.03283 | 0.00024 | Decrease |
| | | Central Java | 0.01472 | -0.00125 | Decrease | 0.01472 | -0.00037 | Decrease |
| | | West Nusa Tenggara | 0.00813 | -0.00034 | Decrease | 0.00813 | -0.00021 | Decrease |
| 6 | Sweet potato | West Java | 0.0388 | 0.00301 | Decrease | 0.0388 | 0.00138 | Decrease |
| | | Papua | 0.0302 | 0.00168 | Decrease | 0.0302 | 0.00066 | Decrease |
| | | East Java | 0.02 | 0.00000 | Decrease | 0.02 | -0.00011 | Decrease |
| 7 | Cassava | Lampung | 0.54583 | 0.38951 | Decrease | 0.54583 | 0.39621 | Decrease |
| | | Central Java | 0.33278 | 0.36837 | Increase | 0.33278 | 0.17196 | Decrease |
| | | East Java | 0.35021 | 0.36865 | Increase | 0.35021 | 0.16762 | Decrease |

From Table 3 above, it is known that the biggest contributor of food products is Java Island, especially in East Java. Almost every type of food which exists, East Java plays an important role in the production of food needed in Indonesia. Almost every province of the largest food supply contributor in Indonesia experience the decreased in production. This decrease can be caused by several factors, such as the number of workers [4].

Therefore, with the decrease of food prediction results every year, and to face demographic bonus in the next few years, it is hoped that the government will prioritize the increase of productive labor force in agriculture sector. This is due to the decline in the number of workers of productive age, in 2010 to 2014 there was a decline in labor by 1.49%. The largest decrease in labor growth was in the group of youth, between the ages of 15 to 29 with an average reduction of 3.41% per year [4]. So to face the future demographic bonus the government should divert the productive workforce into agricultural sector. Because it corresponds to the introduction at the beginning of this study, the pertinent sector holds an important role in political, economic, social and cultural stability.

4. Conclusion

From the results of analysis and exposure in this study, obtained some conclusions which are the best model obtained for prediction results using RBF kernel on SVR found on the fold-1 with the correlation coefficient (R) by 0,9261 are mean squared error of 0,0014. East Java, belong to the province that plays an important role in the production of food needed in Indonesia because almost every type of food, East Java is included in the provinces which produce the largest amount of food production. Overall predictions indicate a decrease in the amount of production in almost every province in Indonesia, even the provinces which contribute as the largest contributor of production in Indonesia also experienced a decrease in production followed by a decrease in harvested area.

Acknowledgments

A biggest thanks to Faculty of Science and Technology UIN Sultan Syarif Kasim Riau on the financial support for this research, the facilities and mental support from the leaders. And also thanks to Puzzle Reseach Data Technology (Predatech) Team Faculty of Science and Technology UIN Sultan Syarif Kasim Riau for their feedbacks, corrections and their assistance in implementing these activities so that research can be done well.

References

- [1] Association of Indonesian Internet Service Providers 2017 *Internet Survey APJII 2016* [Online] Available <https://apjii.or.id/content/read/39/264/Survei-Internet-APJII-2016>, accessed July 5, 2017 In Indonesian.

- [2] Wibisono and Kunto 2016 *Demographic Bonus as an Opportunity to Create Rice Self-Sufficiency Through Tidal Land 09 June 2016* [Online] Available <http://indonesiana.tempo.co>, accessed July 10, 2017 In Indonesian
- [3] Tunjung 2010 Efficiency Analysis of Soybean Raw Material Supply Management at Kecap Company Pt Lombok Gandaria Food Industry Palanglara Karanganyar [skripsi] (Faculty of Agriculture Sebelas Maret University of Surakarta) In Indonesian
- [4] Ministry of Agriculture 2016 Annual Report of Food Security Agency 2015 Ministry of Agriculture In Indonesian
- [5] Central Bureau of Statistics 2016 Population Census [Online] Available <https://www.bps.go.id>, accessed July 4, 2017 In Indonesian
- [6] Prabowo R 2010 Government Policy in Achieving Food Resilience in Indonesia *Journal of Agricultural Sciences* **6**(2) 62-73
- [7] Abidin M Z 2015 Impact of Rice Import Policy and Food Security in Social Welfare Perspective *Sosio Informa* **1**(3) 213-230
- [8] Asmuhammad 2016 Demographic Bonus: The Dilemma of Agriculture Sector 21 September 2016 [Online] Available <http://kompasiana.com>, accessed July 11, 2017
- [9] Piantari E 2011 Feature Selection Hiperspektral Data for Predicted Rice Productivity with Genetic Algorithm Support Vector Regression [Essay] Bogor (ID): Bogor Agricultural University
- [10] Ibrahim N and Wibowo A 2014 Support Vector Regression with Missing Data Treatment Based Variables Selection for Water Level Prediction of Galas River in Kelantan Malaysia *Wseas Transactions on Mathematics* E-ISSN: 2224-2880 **13**(1)
- [11] Paz J F 2010 A Support Vector Regression Approach to Predict Carbon Dioxide Exchange A P de Leon F de Carvalho (Eds.): *Distrib. Computing & Artif. Intell AISC* **79** 157–164.
- [12] Mustakim, Agus Buono, and Irman Hermadi 2015 Support Vector Regression for Predicted Palm Oil Productivity in Riau Province *Journal of Science, Technology and Industry* **12**(2) 179 - 188
- [13] National Population and Family Planning Agency 2016 Demography bonus [Online] Available <http://www.bkkbn.go.id>, accessed July 10, 2017 In Indonesian
- [14] Anwar S, Mukson and Rahmadi D 2017 Analysis of Potential and Food Availability in Relation with Food Security in Central Java *Journal of Research and Development of Universitas Muhammadiyah Semarang* 5-17
- [15] Smola A and Schölkopf B 2003 A Tutorial on Support Vector Regression Neuro COLT, Technical Report NC-TR-98-030 (Royal Holloway College, University of London, UK)
- [16] Patel V R and Mehta R G 2011 Impact of Outlier Removal and Normalization Approach in Modified K-Means Clustering Algorithm *IJCSI International Journal of Computer Science Issues* **8**(5)
- [17] Jain Y K and Bhandare S K 2011 Min Max Normalization Based Data Perturbation Method for Privacy Protection *International Journal of Computer & communication Technology* **2**(8)
- [18] Agmalaro M A and Buono A 2011 Statistical Downscaling GCM Data Modeling Using Support Vector Regression to Predict Monthly Indramayu Rainfall [Thesis] (Bogor (ID): Bogor Agricultural University)
- [19] Gunawan H 2012 Hyperspectral Band Selection Using Recursive Feature Elimination to Predict Rice Production with Support Vector Regression [Essay] (Bogor (ID): Bogor Agricultural University)
- [20] Rao S S 2009 *Engineering Optimization: Theory and Practice* (John Wiley and Sons, Newyork)
- [21] Hidayat R 2013 Prediction System of Nutritional Status of Toddlers by Using Support Vector Regression [Essay] (Bogor (ID): Bogor Agricultural University)
- [22] Mustakim, Agus Buono, and Irman Hermadi 2016 Performance Comparison Between Support Vector Regression And Artificial Neural Network for Prediction of Oil Palm Production *Jurnal Ilmu Komputer dan Informasi* **9**(1) 1 - 8