Forecasting of Arabica Coffee Production in Bali Province Using Support Vector Regression

Ni Made Ratna Putri Udiani
Department of Information Technology
Udayana University
Jimbaran, Indonesia

I Ketut Gede Darma Putra
Department of Information Technology
Udayana University
Jimbaran, Indonesia

Gusti Made Arya Sasmita
Department of Information Technology
Udayana University
Jimbaran, Indonesia

Abstract: Coffee is one from 40 the leading commodities of national commodities and one of superior commodities in the province of Bali. Bali Province's records shows an increasing growth from Arabica coffee plantations in the period covering 8,205 hectares in 2009 increased to 13,155 hectares in 2014 (Bali Central Bureau of Statistics, 2015). Based on the amount of Arabica coffee production that continues to increase, we need a study to find out the achievement of results and policies that will be carried out in order to increase the results of Arabica coffee production. Research on forecasting Arabica coffee production in Bali using Support Vector Regression and several kernels. Pearson Universal Kernel and RBF kernel. Forecasting in the future three years from 2019-2021 has increased. The test MAPE results using the Universal Person Kernel is 5.14% and the RBF kernel is 7.68%.

Keywords: Forecasting, SVR, PUK, RBF, Arabica Coffee

1. INTRODUCTION

Indonesia is an Agricultural Country, this can be seen from the large land area used for agriculture. Based on the existing land area in Indonesia, around 74.68% is used for agricultural land. Various research results, concluded that the biggest contribution in reducing the number of poverty is the growth of the agricultural sector. Whereas the contribution of the agricultural sector in reducing the number of poverty reached 66 percent, with about 74 percent in rural areas and 55 percent in urban areas [1].

The plantation sector has a significant potential role in the natural resource [2]. Coffee is one of the leading commodities from 40 national commodities and one of the most superior commodities in the Bali Province. Based on the Bali Province's statistics shows an increasing growth of Arabica coffee plantation's field in the period covering 8,205 hectares in 2009 increased to 13,155 hectares in 2014 (Bali Central Bureau of Statistics, 2015). With the increase in plantation area, the amount of Arabica coffee production also increased in 2009 which amounted to 3,135,750 tons and increased to 4,183,924 tons in 2014 [3].

Based on the amount of Arabica coffee production that continues to increase, research is needed to obtain results and policies that will be carried out in order to increase the results of Arabica coffee production. Research can be done by doing forecasting to find out the amount of coffee production in the future.

Several studies on Coffee Production forecasting have been carried out previously, namely research on forecasting for robusta coffee production demand at PT. XYZ. From the results of forecasting, the robusta coffee demand pattern is obtained which tends to be constant and based on forecasting calculations with the 3 Center Moving Average method, the result of forecasting the number of requests for robusta coffee is 8754MT [4]. Other studies regarding rainfall forecasting use Support Vector Regression. The results of the research evaluation conducted on the data show that the projected technique performs higher than the conventional framework in terms of accuracy. and the time process runs. The proposed approach produces a maximum prediction of 99.92% [5].

Other studies regarding forecasting the availability of food use Support Vector Regression shows three defining attributes used in this study are (1) Harvest Area; (2) Total Harvest Production; and (3) Food

From the results of experiments, it is known that the biggest contributor to food products is Java Island, especially in East Java. Almost every type of crops, East Java plays an important role in the production of food needed in Indonesia [6].

This study shows a forecasting of the amount of Arabica coffee production using a different method from the literature study described earlier in order to produce a forecast that is close to the actual data. This study uses the Support Vector Regression (SVR) method.

2. METHODOLOGY

The methodology used on this research will be explained by the sequence of research conducted. They are data collection, data normalization, forecasting by sharing training data and test data, forecasting results, data visualization, forecasting accuracy. A general chart of the research can be seen in Figure 1.
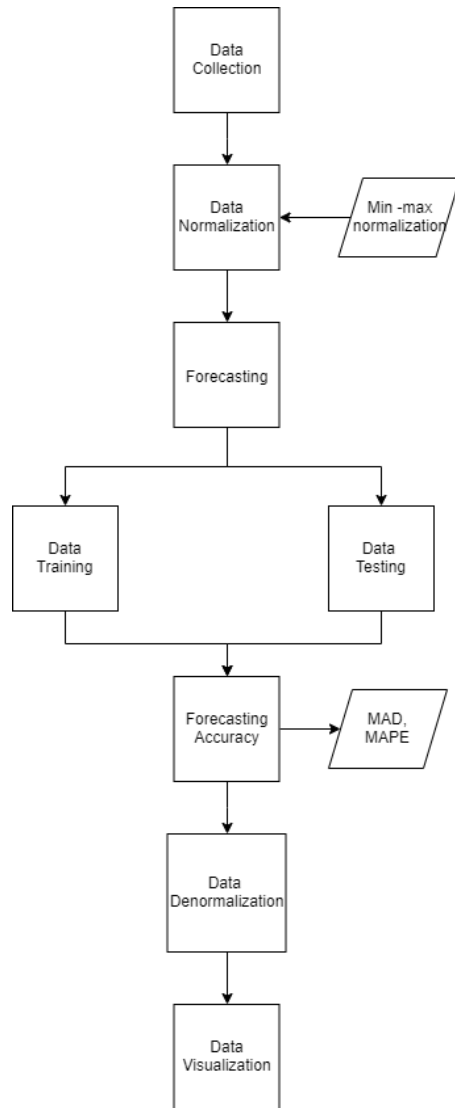
Figure 1. General Chart

Figure 1. is a general chart of research. The study began with data collection, data normalization, forecasting (test data and training data), forecasting results, data denormalization, and data visualization..

2.1 Data Collection

The Data used on this study obtained from the Bali Central Statistics Agency which can be accessed via the web address https://bali.bps.go.id/. The research data used were from 2000 to 2018 for 19 years.

2.2 Data Normalization

Normalization of data is part of data transformation, which is a technique to convert data into values that are more easily understood [7]. Normalization is very important in handling real-time data, because real-time data has different ranges and units. Normalization is used to scale real-time data in the range of 0 to 1. The normalization process helps make numerical calculations more precise and improves the accuracy of forecast results. The formula for data normalization is in Figure 2.

$$v' = \frac{v - \min_A}{\max_A - \min_A}(new\_\max_A - new\_\min_A) + new\_\min_A$$

Figure 2. Normalization Formula

V is the actual data, minA is the lowest actual data, maxA is the highest actual data, new_maxA is the highest data scale that is 1, and new_minA is the lowest data scale that is 0.

2.3 Forecasting

The forecasting will use the Weka v3.9.3. The data used in the forecasting is the amount of Arabica coffee production from 19 years (2000-2018). Forecasting is done by dividing the data into two parts, they are training data and test data. Training data used in 2000-2012. Test data used in 2013-2018. Data predicted in 2019-2021 are measured for accuracy using MAPE percentages.

This study uses the Support Vector Regression method. The Support Vector Regression method in this study uses two kernels, namely Pearson Universal Kernel, and RBF Kernel

2.4 Forecasting Accuracy

The accuracy of forecasting is shown by the percentage of MAPE. Mean Absolute Percentage Error (MAPE) is the average absolute percent error, where the absolute deviation value for each period is divided by the actual value for that period. The formula for calculating MAPE is as follows.

$$MAPE = \frac{\sum ((deviation\ absolute/actual\ data)x100)}{n}$$

MAPE (Mean Absolute Percentage Error) is the easiest step to interpret. For example, the MAPE 2% result is a clear statement that does not depend on problems such as the size of the input data. According to Lewis, C.D. (1982), the interpretation of MAPE is divided into 4 categories [8], as shown in Table 1.

Table 1. MAPE Intepretation

MAPE Value	Interpretation
< 10%	Very Good
10 - 20%	Good
20 - 50%	Enough
>50%	Bad

Table 1 is an interpretation of MAPE. The MAPE category consists of four divisions: very good, good, enough, and bad.

2.5 Data Denormalization

Forecasting results at WEKA are data with a scale of 0-1. Data must be normalized so that it can be compared with actual research data. The equation used in min-max denormalization is as follows.

Denormalization = ($normalizedValue * ($max-$min) + $min)

Where $normalizedValue is the normalized value, $max is the highest actual data, and $min is the lowest actual data.

2.6 Data Visualization

Data visualization is important to do in presenting data. Data visualization is useful to facilitate understanding of data, as well as being communicative and more interesting. This research visualizes data in graphical form with the help of Microsoft Excel.

3. THEORY AND CONCEPTS

Concepts and theories contain reference material that is used according to the topic of research conducted. The topic of the research was forecasting Arabica coffee production. References that are loaded are forecasting, SVR, and the kernel.

3.1 Forecasting

Forecasting is an important issue that covers many fields including business and industry, government, economics, environmental science, medicine, social sciences, politics, and finance. Forecasting problems are often classified as short term, medium term, and long term. Short-term forecasting problems involve predictive events for only a few time periods (days, weeks, months) in the future. The medium-term forecast extends from one to two years into the future, and the problem of long-term forecasting can extend beyond the next few years. Short and medium term forecasting is needed for activities that range from operations management to budgeting and selection of new research and development projects. Long-term forecasting impacts on issues such as strategic planning. Short and medium term forecasting is usually based on the identification, modeling and extrapolation of patterns found in historical data. This historical data usually shows inertia and does not change dramatically very quickly [9].

3.2 Support Vector Regression

Support Vector Regression (SVR) is the improvement from time series forecasting and forecasting in various domains, including business and management science [10]. Support Vector Regression is a Support Vector Machine method for regression cases. The basic problem of regression is to determine a function that can accurately predict future values. This can be done by forming the dividing plane with the smallest size while minimizing the amount of distance between data points to the separating plane [11].

Support Vector Regression has parameter C. Parameter C is the penalty value for SVR model errors.

3.3 Kernel

The kernel method is a class of algorithms that developed in the 90s in the field of machine learning. The best-known example of the kernel method is Support Vector Machines (SVMs), which are good for classification problems. The kernels used in this study are Pearson Universal Kernel and RBF [12].

3.3.1 Pearson Universal Kernel

Pearson Universal Kernel is famous in the field of spectroscopy. The Pearson Universal Kernel function was taken as an alternative kernel function in this study. Pearson Universal Kernel functions as a kind of universal kernel that can replace (by selecting the appropriate parameter settings) kernel functions that are commonly applied, namely the Linear, Polynomial, Gaussian and Sigmoid kernels. The function of the Pearson Universal Kernel is as follows.

$$K(x_i, x_j) = 1/[1 + (2\sqrt{\|x_i - x_j\|^2}\sqrt{2^{(1/\omega)} - 1}/\sigma)^2]^\omega$$

xi and xj are two vectors. The Pearson VII kernel function will point to the symmetrical matrix with the diagonal and all other entries ranging between 0 and 1 for each random pair (xi and xj). PUK has the resistance and strength in mapping the same or even stronger compared to standard kernel functions,

which results in the same or better generalization performance than SVM [13].

3.3.2 RBF

The RBF kernel is a kernel that can generally be used for all types of data. This kernel uses the RBF kernel function to get inside x and y using the equation.

$$K(x, y) = \exp(-gamma*(x-y)^2)$$

where x, y are data values in two different feature spaces and σ (sigma) is a free parameter in the RBF kernel that determines kernel weight. The parameter σ needs to be adjusted to provide more accurate classification results. The default value σ is 1. The use of the gamma parameter can be used on the RBF function whose value $\gamma = \frac{1}{\sigma^2}$ [14].

4. RESULT AND DISCCUSION

Forecasting the case of the amount of Arabica coffee production with several kernels. Arabica coffee production data in Bali is 19 years. Forecasting results are checked for accuracy and compared with actual data using graphs.

4.1 Pearson Universal Kernel

Forecasting uses Arabica coffee production data in Bali using the Support Vector Regression and Pearson Universal Kernel methods. Forecasting Arabica coffee production in Bali is done by finding the best parameters to produce the lowest percentage of MAPE. The results of testing the parameters used are in Table 2.

Table 2. C Testing

C	Error Percentage (MAPE%)
1	5.14%
10	15.30%
100	15.30%

Table 2 is the result of a comparison of several parameter tests. The parameters used in the Support Vector Regression and Pearson Universal Kernel methods are C. Magnitude C affects the test results. C is smaller, resulting in a better percentage of MAPE. The best percentage of MAPE produced was 5.14%, with C 1.0. The PUK kernel has parameters namely sigma and omega. The best sigma and omega test results are in Table 3.

Table 3. PUK Parameter Testing Result

Sigma	Omega	Error Percentage (MAPE%)
0.1	0.1	5.14%
1.0	1.0	13.69%
2.0	2.0	17.21%
3.0	3.0	20.97%

Table 3 is the result of a comparison of several parameter tests. The parameters used in the Support Vector Regression and Pearson Universal Kernel methods are sigma and omega. The amount of sigma and omega affects the test results. Sigma and omega are smaller, resulting in a better percentage of MAPE. The best percentage of MAPE produced was 5.14%, with sigma 0.1 and omega 0.1. Forecasting results on Arabica coffee production are in Table 4.

Tabel 4. Arabica Coffee Forecasting Result

Year	Actual	Training	Testing	MAD	MAPE
2000	6796				
2001	5644	5639.946			
2002	3768	3774.945			
2003	4413	4406.493			
2004	3696	3700.428			
2005	3279	3280.905			
2006	2679	2682.294			
2007	3296	3304.372			
2008	3136	3140.927			
2009	3475	3481.403			
2010	3485	3485.109			
2011	3123	3127.341			
2012	4200	4195.291			
2013	4215		3897.632	317.368	7.53%
2014	3804		3928.51	124.5095	3.27%
2015	4154		3944.566	209.4342	5.04%
2016	4052		3967.209	84.7907	2.09%
2017	3473		3733.364	260.3637	7.50%
2018	4217		3988.206	228.794	5.43%
2019			4003.851		
2020			4093.19		
2021			4141.77		
Average				**204.21**	**5.14%**

Table 4 is the result of forecasting Arabica coffee production data in Bali with Support Vector Regression and Pearson Universal Kernel. The results of the performance of Arabica coffee production using Support Vector Regression and RBF kernels are "very good" because they are below 10%, according to Lewis, C.D. (1982) [8]. Visualization of forecast results is illustrated using a line graph that can be seen in Figure 3.
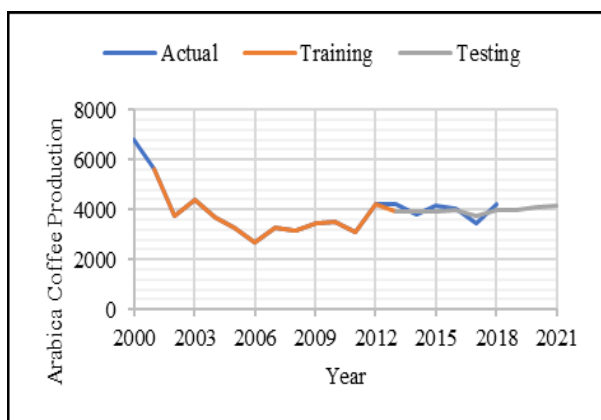

Figure 3. Forecasting Result With SVR and PUK

Figure 3 is a graph of forecasting results on the amount of Arabica coffee production using Support Vector Regression and Pearson Universal Kernel Forecasting results consisting of three years (2019-2021) which have increased. Forecasting using Support Vector Regression and Pearson Universal Kernel 1 is influenced by the magnitude of the parameters C, sigma and omega which can be seen in Table 2 and Table 3

4.2 RBF Kernel

Forecasting uses Arabica coffee production data in Bali using the Support Vector Regression and RBF Kernel methods. Forecasting Arabica coffee production in Bali is done by finding the best parameters to produce the lowest percentage of MAPE. The results of testing the parameters used are in Table 5.

Table 5. C Testing

C	Error Percentage (MAPE%)
1	7.68%
10	33.25%
100	52.52%

Table 5 is the result of a comparison of several parameter tests. The parameters used in the Support Vector Regression and RBF Kernel methods are C. Magnitude C affects the test results. C is smaller, resulting in a better percentage of MAPE. The best percentage of MAPE produced was 7.68%, with C 1.0. The PUK kernel has parameters namely sigma and omega. The best sigma and omega test results are in Table 5. The RBF kernel has the parameters gamma. The best gamma test results are in Table 5.

Table 6. PUK Parameter Testing Result

Gamma	Error Percentage (MAPE%)
0.1	7.68%
1.0	16.59%
2.0	16.82%
3.0	16.29%

Table 6 is the result of comparison testing of several parameters. The parameters used in the Support Vector Regression and RBF Kernel methods are gamma. The amount of gamma affects the test results. Smaller gamma, results in a better percentage of MAPE. The best percentage of MAPE produced was 7.68%, with gamma 0.1. Forecasting results on Arabica coffee production are in Table 7.

Table 7. Arabica Coffee Forecasting Result

Year	Actual	Training	Testing	MAD	MAPE
2000	6796				
2001	5644	4355.442			
2002	3768	4068.488			
2003	4413	3665.022			
2004	3696	3700.016			
2005	3279	3522.985			
2006	2679	3000.126			
2007	3296	3300.255			
2008	3136	3339.367			
2009	3475	3298.609			
2010	3485	3325.369			
2011	3123	3446.821			
2012	4200	4117.48			
2013	4215		3915.747	317.368	7.53%
2014	3804		3536.571	124.5095	3.27%
2015	4154		3953.212	209.4342	5.04%
2016	4052		3648.965	84.7907	2.09%
2017	3473		3666.257	260.3637	7.50%
2018	4217		3726.777	228.794	5.43%
2019			4060.665		
2020			4077.957		
2021			4138.477		
Total				**308.9976**	**7.68%**

Table 7 is the result of forecasting Arabica coffee production data in Bali with Support Vector Regression and RBF Kernel. The results of the performance of Arabica coffee production using Support Vector Regression and RBF kernels are "very good" because they are below 10%, according to Lewis, C.D. (1982) [8]. Visualization of the results of the forecasting is illustrated using a line graph that can be seen in Figure 4.
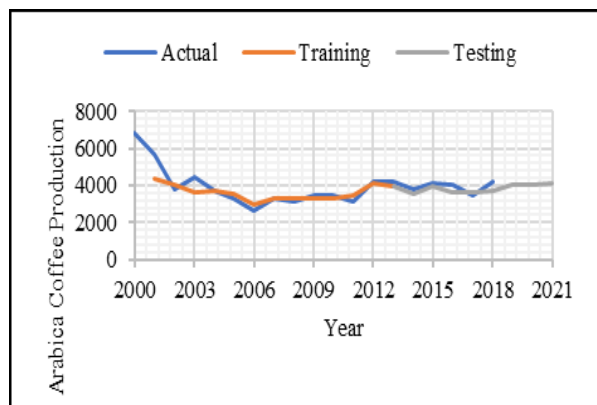

Figure 4. SVR and PUK Forecasting Result

Figure 4 is a graph of forecasting results on the amount of Arabica coffee production using Support Vector Regression and RBF Kernel. Forecasting results consisted of three years (2019-2021) which experienced an increase. Forecasting using Support Vector Regression and RBF Kernel is influenced by the magnitude of the parameters C, sigma and gamma which can be seen in Table 5.

5. CONCLUSION

Coffee is one of the leading commodities of 40 national commodities in the province of Bali. Statistics on the Province of Bali show an increase in the area of Arabica coffee plantations in the period covering 8,205 hectares in 2009 increased to 13,155 hectares in 2014 (Bali Central Bureau of Statistics, 2015). With the increase in plantation area, the amount of Arabica coffee production also increased in 2009 which amounted to 3,135,750 tons and increased to 4,183,924 tons in 2014 [3].

Based on the amount of Arabica coffee production that continues to increase, we need a study to find out the achievement of results and policies that will be carried out in order to increase the results of Arabica coffee production. Research can be done by doing forecasting to find out the amount of coffee production in the future.

Research on forecasting Arabica coffee production in Bali using Support Vector Regression and several kernels. Pearson Universal Kernel and RBF kernel. Forecasting in the future three years from 2019-2021 has increased. The test results using the Universal Person Kernel is 5.14% and the RBF kernel is 7.68%.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] E. D. Martauli, "Analysis of Coffee Production in Indonesia Analisis Produksi Kopi Diindonesia," vol. 01, no. 02, p. 2, 2018.

[2] R. R. Novanda *et al.*, "A Comparison of Various Forecasting Techniques for Coffee Prices," *J. Phys. Conf. Ser.*, vol. 1114, no. 1, 2018.

[3] I. G. B. Udayana, "Marketing Strategies Arabica Coffee with Information Technology in Kintamani District Bangli," *Int. Res. J. Eng. IT Sci. Res.*, vol. 3, no. 3, pp. 93–102, 2017.

[4] S. R. P. N. Hidayatika, "Usulan Penggunaan Metode Forecasting untuk Permintaan Kopi Robusta Pada PT. XYZ," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2013.

[5] N. Hasan, N. C. Nath, and R. I. Rasel, "A Support Vector Regression Model for Forecasting Rainfall," *2nd Int. Conf. Electr. Inf. Commun. Technol. EICT 2015*, no. September 2016, pp. 554–559, 2016.

[6] S. D. Agustina, Mustakim, Okfalisa, C. Bella, and M. A. Ramadhan, "Support Vector Regression Algorithm Modeling to Predict the Availability of Foodstuff in Indonesia to Face the Demographic Bonus," *J. Phys. Conf. Ser.*, vol. 1028, no. 1, 2018.

[7] R. Rismala, "Prediksi Harga Saham menggunakan Support Vector Regression dan Firefly Algorithm," vol. 2, no. 2, p. 15, 2015.

[8] J. J. Montaño Moreno, A. Palmer Pol, A. Sesé Abad, and B. Cajal Blasco, "Using the R-MAPE Index as a Resistant Measure of Forecast Accuracy," *Psicothema*, Vol. 25, No. 4, Pp. 500–506, 2013.

[9] D. C. Montgomery, C. L. Jennings, And M. Kulahci, *Introduction to Time Series Analysis and Forecasting*. 2008.

[10] S. F. Crone, J. Guajardo, and R. Weber, "A study on the ability of Support Vector Regression and Neural Networks to forecast basic time series patterns," *IFIP Int. Fed. Inf. Process.*, vol. 217, pp. 149–158, 2006.

[11] T. B. Trafalis and H. Ince, "Support Vector Machine for Regression Applications to Financial Forecasting," *Inst. Electr. Electron. Eng.*, 2000.

[12] G. Rubio, H. Pomares, L. J. Herrera, and I. Rojas,

"Kernel Methods Applied to Time Series Forecasting," *Comput. Ambient Intell.*, pp. 782–789, 2007.

[13] K. A. A. Abakar and C. Yu, "Performance of SVM based on PUK kernel in comparison to SVM based on RBF kernel in prediction of yarn tenacity," *Indian J. Fibre Text. Res.*, vol. 39, no. 1, pp. 55–59, 2014.

[14] R. Indraswari and A. Arifin, "RBF Kernel Optimization Method with Particle Swarm Optimization on SVM using The Analysis of Input Data's Movement," vol. 1, no. December, pp. 36–42, 2017.