Forecasting the Number of Unemployment in Bali Province using the Support Vector Machine Method

Imelda Alvionita Tarigan

Department of Information Technology, Faculty of Engineering,
Udayana University, Bukit Jimbaran, Bali, Indonesia

ABSTRACT

Unemployment has an impact on economic development in Indonesia. Bali is one of the provinces in Indonesia which has had a high unemployment rate in the last 13 years. Forecasting the number of unemployed in Bali Province is needed so that government policies can more optimally handle unemployment. This study aims to forecast the number of unemployed in the next five years. The method used is the Support Vector Machine because it is capable of forecasting a certain time series or time series. The data used are unemployment data from 2007 to 2019. The results of the analysis in this study show that the best SVM kernel type for forecasting the number of unemployed is radial. This type of kernel is used because it shows the smallest error value, namely MSE 0.007022, MAE 0.071292, and MAPE 23.24%. Forecasting results in the coming year an increase in the number of unemployed people from 2020 to 2024.

KEYWORDS: Unemployment, Time Series, SVM

INTRODUCTION:

Unemployment for someone who is already classified in the active workforce is looking for work at a certain wage level, but cannot get the job he wants. Unemployment becomes a problem in the economy because of unemployment, the productivity and income of the community will decrease so that it can cause poverty and other social problems [1]. Unemployment can be caused by various things such as work or difficulty in getting the opportunity to get a job. Developing countries are faced with the problem of unemployment. The government must take steps to reduce the unemployment rate in Indonesia. Bali is one of the provinces in Indonesia which has shown a high number of unemployed people over the past 13 years [2]. There is an increase and decrease each year. In order for the steps taken by the government to be in synergy with the actual situation, it is necessary to forecast the number of unemployed for the coming years. The level of attractiveness or reduction in the number of unemployed can be predicted using time series forecasting or forecasting methods using analysis of the relationship pattern between the variables to be calculated and the time variable[3]such as the Support Vector Machine method. Support Vector Machine is a learning system that uses a hypothetical space in the form of linear functions in a high-dimensional feature and uses a learning algorithm based on optimization theory. The SVM method has 3 model approaches or so-called kernels, namely radial, linear, and polynomial kernels [4]. This study predicts the number of unemployed in Bali Province using the Support Vector Machine method using data on the number of unemployed from 2007 to 2019 obtained from the Central Bureau of Statistics Bali.

A. Research Method

Forecasting the number of unemployed uses a case study in Bali Province. Forecasting is done using the Support Vector Machine method. There are several stages of research for forecasting can be seen in Figure 1.
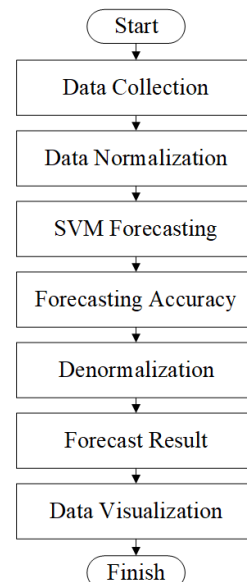

Figure 1Research Stage

1. Data Collection

Data collection on the number of unemployed in Bali Province from 2007 to 2019 was obtained from the Central Bureau of Statistics Bali. The data obtained is an annual documentation, in order to obtain 13 data. The unemployment rate data is used as the basis for forecasting.

2. Data Normalization

Data normalization aims to facilitate data processing so that it is able to produce training outputs that are in accordance with the activation function used.

$$\frac{0.8\,(x_0 - x_{min})}{x_{max} - x_{min}} + 0.1 \qquad (1)$$

The normalization stage was carried out on 13 data on the number of unemployed. The results of data normalization on the number of unemployed can be seen in Table 1.

Table 1 Results of Data Normalization on the Number of Unemployment

Year	Unemployment Number	X'
2007	77577	0.900
2008	69548	0.751
2009	66470	0.694
2010	68791	0.737
2011	65716	0.680
2012	48230	0.355
2013	41820	0.236
2014	44126	0.279
2015	47210	0.336
2016	46484	0.323
2017	36143	0.131
2018	34485	0.100
2019	37551	0.157

Table 1 is the results of the data normalization on the number of unemployed. The maximum value in the data is 77577, namely in 2007. The minimum value in the data is 34485, namely in 2018. The data on the number of unemployed is processed using Equation 1. The value of normalization results can be seen in Table 1 with information X '.

3. Support Vector Machine Forecast

Forecasting by building an SVM model using the R Studio application by adding the e1071 package.

4. Forecasting Accuracy

Forecasting accuracy uses MSE, MAE, and MAPE to determine the error value from the results of the formation of the SVM model which is processed by adding the package metrics from R Studio.

5. Denormalization

Denormalization which aims to change the predicted data back to the beginning in order to maintain stability and prevent data confusion.

$$X_i = \frac{(y - 0.1)\,(x_{max} - x_{min})}{0.8} + x_{min} \quad (2)$$

6. Forecast Result

The results of forecasting the number of unemployed in the next five years, namely 2020 to 2024.

7. Data Visualization

Data visualization is carried out to display forecasting results via graphs to facilitate data analysis.

B. Literature Study

Previous research conducted by Muhammad Fajar was to forecast the unemployment rate using a multilayer perceptron neuron network using Central Bureau of Statistics Data. The results obtained by the multilayer perceptron neuron network method are very good for forecasting with a MAPE value of 1.4465%. With the unemployment rate has increased from February to August [5].

Research conducted by Zainal Arifin, Junita Herliani, and Hamdani, namely forecasting unemployment using the double exponential smoothing method in East Kalimantan Province using East Kalimantan BPS data. The results obtained by an alpha value of 0.6 have the best results with the smallest MAPE value of 6.89%. With the prediction that the number of unemployed will decrease in 2018[6].

Research conducted by EnsiwiMunarsi is forecasting the number of unemployed people in South Sumatra Province using the ARIMA method using data from 2002 to 2016. The results obtained by the best arima model are 2,1,1. With the prediction results show a decrease in the number of unemployed in South Sumatra Province compared to the previous time [7].

C. Result

Forecasting is carried out on the number of unemployed in Bali Province using the Support Vector Machine method with different kernel functions. The kernel function is applied to the number of unemployed data to get the best forecasting results. The results of forecasting using 3 kernel functions can be seen in Table 2.

Table 2 Forecasting Accuracy Results

Model	Kernel	Hasil		
		MSE	MAE	MAPE
Model 1	Radial	0.007022	0.071292	23.24%
Model 2	Linear	0.009604	0.079980	28.84%
Model 3	Polynomial	0.024858	0.118461	40.72%

Table 2 shows the results of the accuracy of forecasting the number of unemployed in Bali Province. The kernel functions used are radial, linear, and polynomial. The accuracy results show that the best model is model 1. Model 1 shows that the radial kernel function has the lowest error value than other kernel functions. The MSE values obtained were 0.007022, MAE 0.071292, and MAPE 23.24%.

Table 3Forecasting Results

Year	Unemployment Number	Forecasting	Error
2007	77577	70130.516	7446.484
2008	69548	70290.189	-742.189
2009	66470	67952.301	-1482.301
2010	68791	63312.678	5478.322
2011	65716	57418.734	8297.266
2012	48230	51854.863	-3624.863
2013	41820	47880.304	-6060.304
2014	44126	45608.301	-1482.301
2015	47210	44055.654	3154.346
2016	46484	42162.265	4321.735
2017	36143	39851.592	-3708.592
2018	34485	38106.945	-3621.945
2019	37551	38052.019	-501.019
2020		39985.664	

Table 3 is a comparison of the actual data and the results of forecasting the number of unemployed. The best SVM kernel function to use is Model 1, namely the radial kernel function. Forecasting is carried out from 2007 to 2020. Forecasting is carried out for the next one year, namely 2020. The forecast results show that the number of unemployed is predicted to be 39,985 people. In 2019 the number of unemployed is predicted to be 38,052, the 2019 prediction has an error value of 501,019 from the actual data.
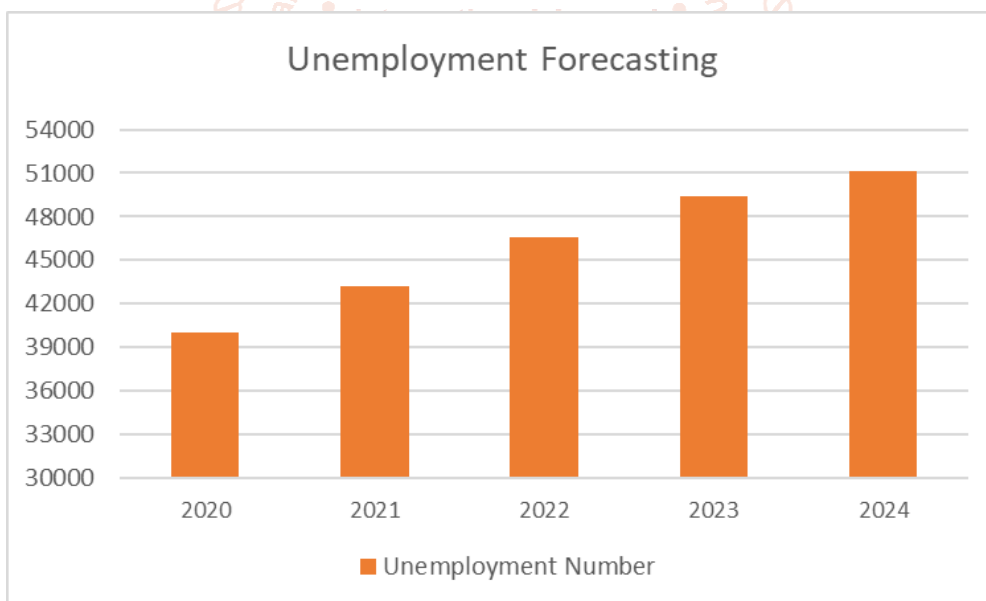

Figure 2Unemployment Forecasting Results

Figure 2 shows the long-term results of forecasting the number of unemployed in Bali Province. Forecasting results show the number of unemployed people in 2020 will increase to 39,986 people. In 2021 the number of unemployed will increase to 43,210 people. The number of unemployed is predicted to increase again in 2022 to 46,614. The years 2023 and 2024 are predicted to increase to 49,354 and 51,144 people, respectively.

D. Conclusion

Forecasting the number of unemployed in Bali Province uses 3 kernel functions with different results. Forecasting accuracy is done using MSE, MAE, and MAPE. The best SVM model is Model 1 using the radial kernel function because it has the lowest error value, namely MSE 0.007022, MAE 0.071292, and MAPE 23.24%. Forecasting results show the number of unemployed will increase in 2020 by 39,985 people. The number of unemployed people in 2021 to 2024 has increased with the highest number in 2024 of 51,144 people. Further research can be carried out by developing forecasts using a larger dataset size, to obtain better accuracy. Other unemployment factors can be added to make a complete and accurate research. The use of other forecasting methods can be done to get results with better accuracy values.

References

[1] S. Sukirno, *Pengantar Ekonomi Makroekonomi*. Jakarta: PT Raja Grafindo Persada, 2004.

[2] BPS Provinsi Bali, "Banyaknya Pengangguran Provinsi Bali Menurut Kabupaten/Kota 2007-2019," 2020. [Online]. Available: https://bali.bps.go.id/indicator/6/285/1/banyaknya-pengangguran-provinsi-bali-menurut-kabupaten-kota.html.

[3] D. W. Hanke, J. E. Wichers, *Business Forecasting Eight Edition*. New Jersey: Pearson Prentice Hall, 2005.

[4] V. Kecman, "Support Vector Machines – An Introduction," no. May 2005, pp. 1–47, 2005.

[5] M. Fajar, "Peramalan Tingkat Pengangguran Terbuka Menggunakan Multilayer Perceptrons NeuralL Network," no. April, 2018.

[6] Z. Arifin, "Peramalan Pengangguran Menggunakan Metode Double Exponential Smoothing Di Provinsi Kalimantan Timur," *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 1, pp. 24–29, 2019.

[7] E. Munarsih, "Peramalan Jumlah Pengangguran di Provinsi Sumatera Selatan dengan Metode Autoregressive Integreted Moving Average (ARIMA)," *J. Penelit. Sains*, vol. 19, no. 1, p. 168439, 2017.