Published online on the journal's webpage: **http://jurnal.iaii.or.id**

Forecasting Cases of Dengue Hemorrhagic Fever Using the Backpropagation, Gaussians and Support-Vector Machine Methods

I Made Yudha Arya Dala[1], I Ketut Gede Darma Putra[2], Putu Wira Buana [3]
[1, 2, 3] Department of Information Technology, Engineering Faculty, Udayana University
[1]yudhaaryadala@student.unud.ac.id, [2]ikgdarmaputra@unud.ac.id, [3]wbhuana@gmail.com

Abstract

Dengue disease has been known to the people of Indonesia since 1779. The Aedes mosquito has two types, namely Aedes aegypti and Aedes albopictus. Aedes aegypti is a mosquito that carries the dengue virus. The dengue fever cases in Bali province tend to increase from year to year, especially when approaching the rainy season. The government's preventive action is needed to tackle the spread of the dengue virus and casualties. Data mining attempts to extract known knowledge or use historical data to find regularity patterns and relationships in a set of data. In this study, data mining predicts the number of dengue cases in Bali's province. The prediction uses several database variables to predict future variables' values, which are not currently known. The process of estimating predictive values based on patterns in a data set. This forecasting aims to assist the government in predicting dengue fever cases in the coming period to prepare appropriate prevention efforts. Forecasting dengue fever cases are carried out using three methods: backpropagation, gaussians, and support-vector machine. The amount of data used was 528 sample data, from 2008 to 2018. The results obtained are that the backpropagation method is better at predicting dengue fever cases with a MAPE error rate of 0.025. Simultaneously, the gaussian method has a MAPE error rate of 0.035, and support-vector machine has a MAPE error rate of 0.060.

Keywords: Dengue Fever, Data Mining, Backpropagation, Gaussians, Support-Vector Machine, MAPE.

1. Introduction

Dengue virus infection (DENV) is a global health threat, affecting at least 3.6 billion people living in more than 125 countries in tropical and subtropical regions[1]. Dengue disease has been known to the people of Indonesia since 1779. This disease is a precursor to dengue fever (DHF). Dengue hemorrhagic fever (DHF) or dengue hemorrhagic fever (DHF) attacked Indonesians for the first time in Jakarta in 1968[2]. The Aedes mosquito spreads the dengue virus. The Aedes mosquito has two types, namely Aedes aegypti and Aedes albopictus. Aedes aegypti is a mosquito that carries the dengue virus; this mosquito lives indoors. Aedes albopictus is a type of mosquito that carries the dengue virus, but its concentration is small compared to Aedes aegypti. The Aedes aegypti mosquitoes' habitat is indoors, while the Aedes albopictus is in the bush. Ordinary people, it is not easy to detect early diagnosis of DHF. Due to the typical early symptoms of DHF, such as bleeding on the skin or other signs of bleeding, it sometimes occurs only at the end of the disease period [3].

Data mining attempts to extract knowledge that includes collecting or using historical data to find regularity patterns and relationships in a data set [4]. Artificial Neural Networks are potent data modeling capable of capturing and representing complex Input-Output relationships, besides a microscopic approach where information is presented by neuronal excitation patterns [5]. Artificial neural networks (ANN), as a part of Artificial Intelligence in computer science, are widely used to solve forecasting or prediction problems, especially those based on time-series data [6]. Rapid Miner is a data mining tool that provides data mining and machine learning procedures, including data preprocessing, visualization, modeling, and evaluation[7]. This research aims to know which method is best for predicting Dengue Hemorrhagic Fever cases based on the level of measurement accuracy and forecasting of Dengue Hemorrhagic Fever Cases in Bali Province in the next phase.

In this study, data mining predicts the number of dengue cases in Bali's province. The prediction uses several database variables to predict future variables' unknown

values. The process of estimating predictive values is based on patterns in a data set[8].

This research use backpropagation, gaussian, and support-vector machine methods. Backpropagation is a learning method whose architecture has multi-layer layers where learning is repeated to create a system that is resistant to damage and consistently works well, and is often used to solve complex problems [9, 10]. Gaussian Processes are one of the most widely used stochastic processes to model-dependent power that is observed over time, gaussian process collects random variables where these variables' value has a finite value[11]. Support-vector machine is one of the most effective machine learning algorithms. From a practical and theoretical perspective, support-vector machine problems are solved by solving Lagrangian equations that are a dual form of support-vector machine through quadratic programming[12].

Previous research that used the backpropagation methodology was the study of "Implementation of the Backpropagation Algorithm to Predict Student Graduation." The data used were 318 samples of data, 70% of which used as training data, and the other 30% will be used as testing data [13]. Poonpong Suksawang researched gaussian method forecasting in a journal entitled "Electricity Consumption Forecasting in Thailand using Hybrid Model SARIMA and Gaussian Process with Combine Kernel Function Technique." The data used is data on electricity consumption for 11 years, from 2005 to 2015[14]. Other research conducted at PT. The Java Bali Gresik Power Plant using the support-vector machine method. The data used were taken from the history of water usage for seven months, while the test data used were 49, which were taken randomly from all data[15].

Approaches with backpropagation, gaussians, and support-vector machine that use learning and training methods can be applied to identify data patterns from the prediction system for dengue hemorrhagic fever (DHF) using data from the Bali Provincial Health Office and data mining tools that be used in this research is rapidminer.

2. Research Method

The data mining method or knowledge discovery in database (KDD) is a term used to describe knowledge discovery in databases. Data Mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various large databases[4]. Figure 1 is an overview of this research.
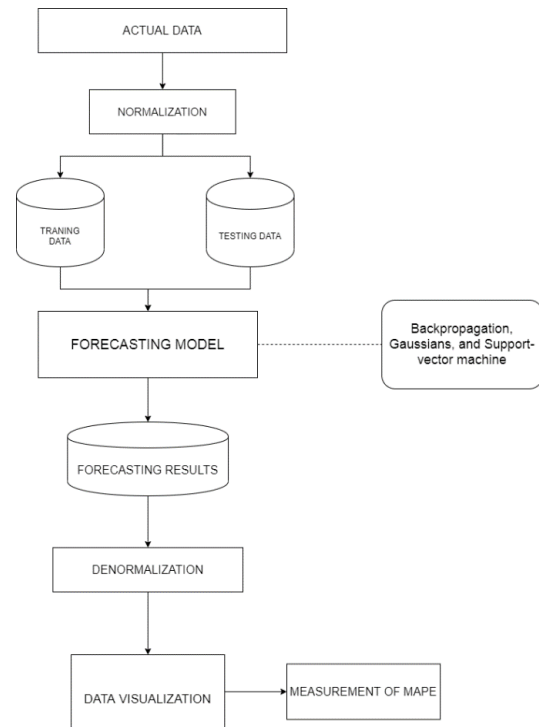

Figure 1. Research Stage

2.1 Actual Data

The data used for forecasting is dengue fever case data in Bali province in data from 2008 to 2018. Data collection was taken from the Bali provincial health office. There are 4 data attributes used, namely data on the number of dengue fever sufferers, the number of deaths, the number of male victims, and female victims. The data will be divided into 2, namely data training and data testing.

Table 1. Dengue Fever Case Data

Year	Number of cases	Number of deaths	Number of male deaths	Number of female deaths
2008	4361	9	5	4
2009	4385	9	3	6
2010	4774	13	8	5
2011	2996	7	5	2
2012	2650	3	2	1
2013	7077	6	3	3
2014	8629	17	9	8
2015	10704	28	14	14
2016	19427	56	35	21
2017	17287	43	22	21
2018	16415	37	14	23

2.2 Normalization

Normalization data is a technique that functions to eliminate data redundancies. Normalization process helps make more precise numerical calculations and improves the accuracy of forecast results. The equation used in the min-max normalization is as follows[16]

$$v' = \frac{(v - min_A)}{max_A - min_A} \qquad (1)$$

Which v' is normalization data, v is actual data, min_A is lowest data value and max_A is highest data value.

Normalization is carried out on each attribute, namely the number of sufferers, the total number of deaths, the number of male deaths, and female deaths. Normalization results of data on the number of dengue fever cases can be seen in Table 2.

Table 2 Normalization Results

Year	Number of cases	Number of deaths	Number of male deaths	Number of female deaths
2008	0,182	0,191	0,173	0,209
2009	0,183	0,191	0,124	0,282
2010	0,201	0,251	0,245	0,245
2011	0,116	0,160	0,173	0,136
2012	0,100	0,100	0,100	0,100
2013	0,311	0,145	0,124	0,173
2014	0,385	0,311	0,270	0,355
2015	0,484	0,477	0,391	0,573
2016	0,900	0,900	0,900	0,827
2017	0,798	0,704	0,585	0,827
2018	0,756	0,613	0,391	0,900

2.3 Training Data and Testing data

Dengue case data is divided into two datasets, namely training data and testing data. A division ratio of 50-50 60-40, 70-30, and 80-20 of 528 sample data. Training data is pre-existing data based on facts that have occurred, and testing data is data used for the accuracy of the data mining model that has been created. The technique used is split validation.

2.4 Forecasting Model

Forecasting models are created using the Rapidminer 9.2 tools. This study uses a forecasting model using three methods: the backpropagation method, gaussian, and support-vector machine. Figure 2 is a picture of the forecasting model made in Rapidminer 9.2.

The data retrieve operator functions as a place to accommodate training data and test data. The windowing operator functions to increase accuracy in forecasting data in the form of time series. In contrast, the set role operator functions to determine the attributes processed by the method used. The model operator functions to carry out the forecasting process using the backpropagation, gaussian, and support-vector machine methods. The operator applies model functions to store the data mining process that has been carried out on training data to apply the process to data testing. Operator Performance Operators results and performance functions to display data mining results and the level of accuracy of the methods used. Excel Write operator functions to convert data mining results from Rapidminer into Excel.
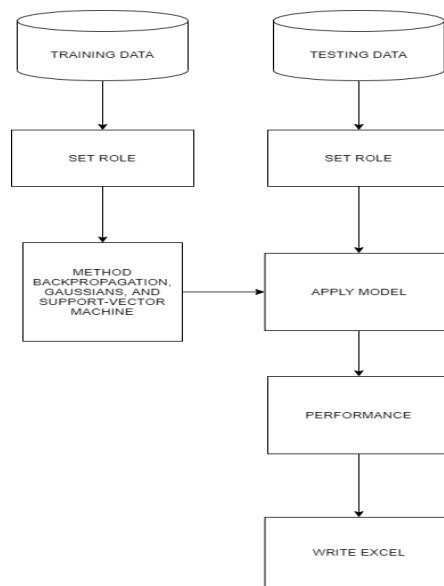

Figure 2. Forecasting Model

2.5 Forecasting Results

Forecasting results that have been obtained, of course, will vary depending on the prediction model used. Forecasting results containing forecasts from the test data will be visualization using a line graph. Prediction results will display the forecasting of the Number of cases, male deaths, female deaths, and the total number of deaths.

2.6 Denormalization

Data from forecasting results are reprocessed with data denormalization. The denormalization process serves to return the data value to its original form. The forecast results are converted back into values like the initial data with Equation[17].

$$v = (v' * (max_A - min_A)) + min_A \qquad (2)$$

2.7 Visualization

Data visualization is done to make it easier to understand the data that will be used in forecasting. The visualization used in this research is in the form of a graphic created using Microsoft Excel tools.

2.8 Measurement of MAD and MAPE

This study's two measurements to calculate the forecasting results' accuracy using the backpropagation method, gaussians, and support-vector machine are as follows.

2.8.1 *Mean Absolute Deviation* (MAD)

Mean Absolute Deviation is the average absolute error over a certain period regardless of whether the forecast

result is greater or less than the reality. The MAD equation is as follows[18].

$$MAD = \frac{\sum(X_i - F_i)}{n} \quad (3)$$

2.8.2 *Mean Absolute Percentage Error* (MAPE)

Mean Absolute Percentage Error is a measure of relative error. MAPE states the percentage of error forecasting results against actual requests during a specific period, which will provide information on the percentage of too high or too low errors. The MAPE equation is as follows[18].

$$MAPE = \frac{\sum \frac{|X_i - F_i|}{X_i}}{n} \times 100\% = \frac{\sum \frac{|e_i|}{X_i}}{n} \times 100\% \quad (4)$$

3. Results and Discussion

They were making a forecast model for dengue fever case data in Bali Province on the Rapidminer 9.2 application. The backpropagation, gaussians, and support-vector machine forecasting methods are used to predict the attributes in dengue fever cases in Bali province, namely the number of dengue cases, the number of women who died, the number of men who died, and the total number of deaths.
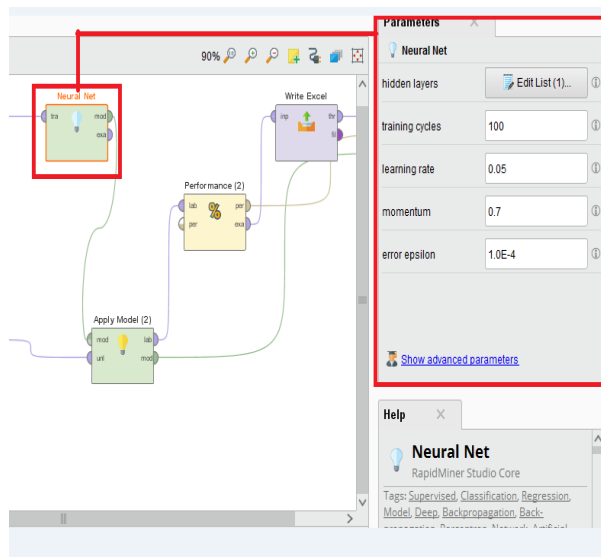

Figure 3 Architecture of the Backpropagation Forecasting Model

Figure 3 shows the forecasting model using the backpropagation method found in the modeling operator view. The determination of hidden layers, training cycles, and learning rates from the backpropagation method is carried out in the Parameters view. The results of forecasting the backpropagation method can be seen in Figure 4.

Figure 4 shows the consequence of forecasting in the Rapidminer application using the backpropagation method. The expected value above is still in normalized form, and it will be entered into Microsoft Excel to be

normalized to make data and MAPE measurements easier to interpret.

Row No.	jumlah pend...	prediction(ju...	tahun
1	0.798	0.786	2017
2	0.756	0.759	2018
3	0.651	0.651	2019
4	?	0.673	2020

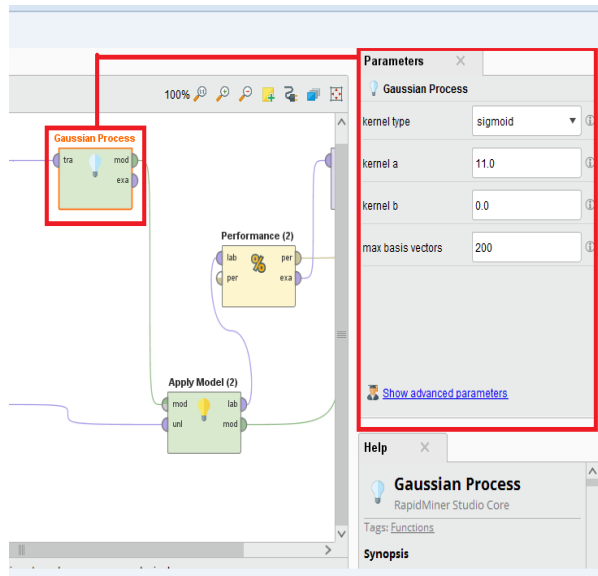Figure 4 Forecasting Results of the Backpropagation Method

Figure 5 Gaussian Forecasting Model Architecture

Figure 5 shows the forecasting model using the Gaussian method in the modeling operator view. The determination of the kernel type, kernel A and kernel B parameters from the Gaussian method can be done in the Parameters view. The results of forecasting the Gaussian method can be seen in Figure 6.

jumlah pend...	prediction(ju...	jumlah pend...
0.756	0.750	0.798
0.739	0.739	0.756
0.709	0.734	0.739

Figure 6 Forecasting Results of the Gaussian Method

Figure 4 shows the consequence of forecasting in the Rapidminer application using the gaussian method. The expected value above is still in normalized form, and it will be entered into Microsoft Excel to be normalized to make data and MAPE measurements easier to interpret.

Figure 7 shows the forecasting model using the Support-vector machine method in the modeling operator view. Determination of the L Pos and L Neg parameters from the Support-vector machine method can be done in the

Parameters view. The results of forecasting the support-vector machine method can be seen in Figure 8.
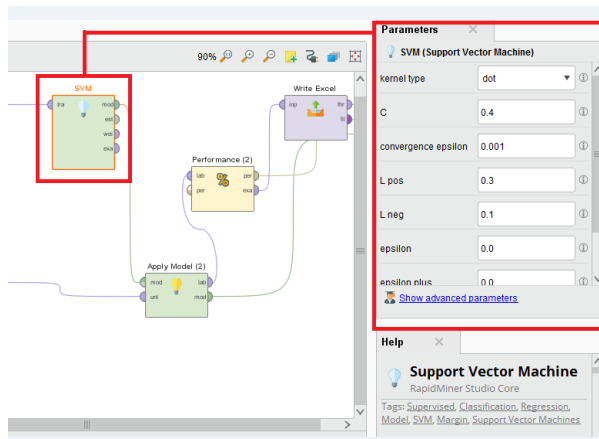

Figure 7 Support-vector machine forecasting model architecture

Figure 8 Forecasting Results of the Support-vector machine method

Figure 4 shows the consequence of forecasting in the Rapidminer application using the Support-vector machine method. The expected value above is still in normalized form, and it will be entered into Microsoft Excel to be normalized to make data and MAPE measurements easier to interpret.

Forecasting results and each attribute's best architecture in the backpropagation, gaussians, and support-vector machine methods are as follows.

Tabel 3. The best architecture on the backpropagation method

No	Attribute Name	Epoch	MAPE	Prediction	
				2019	2020
1	Number of cases	100	0,013	12915,118	13289,213
2	Number of deaths	650	0,022	29,397	29,079
3	Number of male deaths	800	0,042	18,539	18,535
4	Number of female deaths	900	0,018	14,750	12

Table 3 is the best architecture for each parameter of the backpropagation method and the results of forecasting the next two years for each attribute. As predicted, the number of dengue fever cases has decreased from 2018, which previously amounted to 16415 to 12915,118 and 13289,213, respectively. The total number of deaths from the forecast results decreased from 2018, which previously amounted to 37, respectively, to 29,397 and

29,079. The number of male deaths resulting from the forecast increased from 2018, which previously amounted to 14, respectively, to 18,539 and 18,535. The number of women who died due to the forecast decreased from 2018, which previously amounted to 23 to 14,750 and 12.

Tabel 4. The best architecture on the gaussian method

No	Attribute Name	Kernel	MAPE	Prediction	
				2019	2020
1	Number of cases	6,0	0,009	15805,447	15406,820
2	Number of deaths	8,0	0,017	34,826	33,550
3	Number of male deaths	4,0	0,069	11,080	9,434
4	Number of female deaths	4,0	0,045	23,156	23,272

Table 4 is the best architecture for each parameter of the Gaussian method and the results of forecasting the next two years for each attribute. The type of kernel used is sigmoid. As predicted, the number of dengue fever cases has decreased from 2018, which previously amounted to 16415 to 15805,447 and 15406,820, respectively. The total number of deaths from the forecast results decreased from 2018, which previously amounted to 37, respectively, to 34,826 and 33,550. The number of male deaths resulting from the forecast decreased from 2018, which previously amounted to 14 to 11,080 and 9,434, respectively. The number of female deaths from the forecast is fixed from 2018, which previously amounted to 23, respectively, to 23,156 and 23,272.

Tabel 5. The best architecture in the support-vector machine method

No	Attribute Name	L pos	L neg	MAPE	Prediction	
					2019	2020
1	Number of cases	0,2	0,8	0,015	15427,168	15053,472
2	Number of deaths	0,4	0,4	0,016	33,264	32,488
3	Number of male deaths	0,2	0,3	0,084	12,538	11,331
4	Number of female deaths	0,3	0,3	0,047	24,869	25,033

Table 5 is the best architecture for each parameter of the support-vector machine method and the forecast results for the next two years for each attribute. As predicted, the number of dengue fever cases has decreased from 2018, which previously totaled 16415 to 15427,168 and 15053,472, respectively. The total number of deaths from the forecast results decreased from 2018, which previously amounted to 37, respectively, to 33,264 and 32,488. The number of male deaths resulting from the forecast decreased from 2018, which previously

amounted to 14 to 12,538 and 11,331, respectively. The number of women who died due to the forecast increased from 2018, which previously amounted to 23 to 24,869 and 25,033.
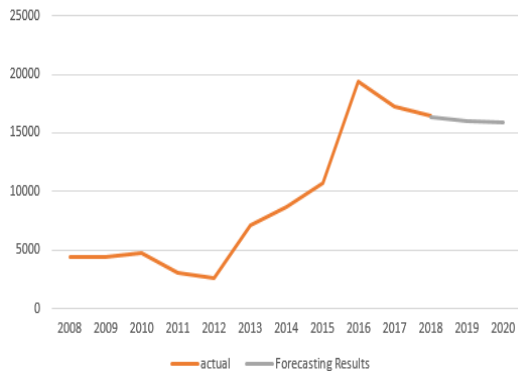

Figure 3. The best forecast graph for the next period is the number of DHF cases

Figure 3 results from forecasting the future dengue fever cases with two years from 2019 to 2020. The prediction results decreased from 2018, which previously amounted to 16415 to 15805,447 and 15053,472, respectively.
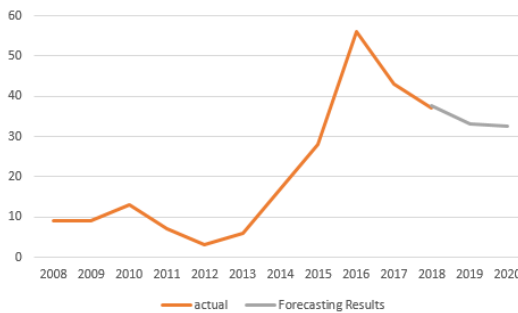

Figure 4. Forecasting graph for the coming period the number of dengue fever deaths

Figure 4 results from forecasting the future period of the number of deaths over the next two years from 2019 to 2020. The prediction results decreased from 2018, which previously amounted to 37, respectively to 33,264 and 32,488.
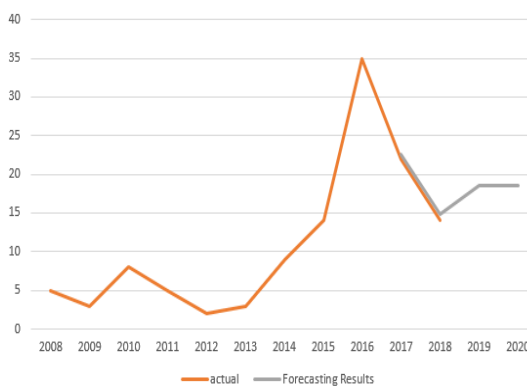

Figure 5. Forecasting graph in the coming period the number of DHF male deaths

Figure 5 results from forecasting the future period of the number of male deaths with two years from 2019 to 2020. The prediction results decreased from 2018. The forecast results increased from 2018, which previously amounted to 14 consecutively to 18,539 and 18,535.
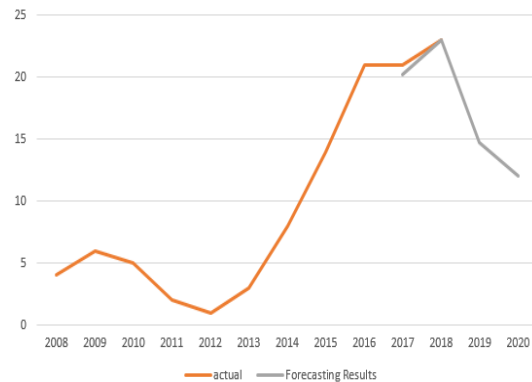

Figure 5. Forecasting graph for the coming period the number of DHF women died

Figure 5 is the result of forecasting the future period of the number of women's deaths with a span of 2 years from 2019 to 2020. The prediction results decreased from 2018. The forecasting results decreased from 2018, which previously totaled 23 to 14,750 and 12.
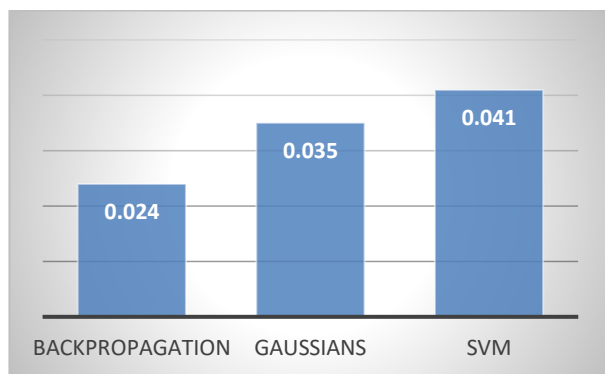

Figure 6. MAPE Comparison

Figure 6 is a bar chart of the final forecasting results from dengue fever data. This value is obtained from the average results for each attribute based on the method used. Backpropagation is the best method with an error rate of 0.024, while forecasting using the Gaussian method has an error rate of 0.035. Moreover, support-vector machine has an error rate of 0.041.

4. Conclusion

The backpropagation method is a better method than the Gaussian method and the Support-Vector machine in predicting cases of dengue fever in Bali Province based on the analysis carried out because the error rate value of the backpropagation method is smaller, namely 0.024, while forecasting using the Gaussian method has an error rate of 0.035, and the support-vector machine error rate of 0.041. The advantage of this backpropagation

method is that it has several units in one or more hidden screens because the backpropagation method is the best forecasting method for this study. The development of dengue fever case forecasting results can be done by adding additional methods, the forecasting method used should be more diverse. This aims to get better forecasting results.

References

[1] N. K. Iin, "Keterkaitan Antara Kondisi Lingkungan Dan Perilaku Masyarakat Terhadap Keberadaan Vektor Demam Berdarah Dengue (DBD)," *Journal of Borneo Holistic Health,* vol. 3, 2020.

[2] R. wowor, " Pengaruh Kesehatan Lingkungan terhadap Perubahan Epidemiologi Demam Berdarah di Indonesi," vol. 5, 2017.

[3] T. NurAkmalinaMatJusoh, "Performance Evaluation of Commercial Dengue Diagnostic Tests for Early Detection of Dengue in Clinical Samples," *Journal of Tropical Medicine* 2017.

[4] Y. W. Wahdi, "Implementasi Data Mining Untuk Analisa Tingkat Pelanggaran Lalu Lintas Dengan Algoritma Association Rule " vol. 1, 2018.

[5] N. E.-K. Mohamed, "Artificial Neural Networks in Data Mining " *Journal of Computer Engineering* 2016.

[6] R. Aulia, "Penerapan Metode Backpropagation Untuk Memprediksi Jumlah Kunjungan Wisatawan Berdasarkan Tingkat Hunian Hotel," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi),* vol. 4, 2018.

[7] A. Kori, "Comparative Study of Data Classifiers Using Rapidminer," 2017.

[8] A. Muqtadir, "The Implementation of Grey Forecasting Model for Forecast Result's Food Crop Agricultural " *Scientific Journal of Informatics* 2016.

[9] A. Wanto, "Analysis of Artificial Neural Network Backpropagation Using Conjugate Gradient Fletcher Reeves In The Predicting Process " *International Conference on Information and Communication Technology* 2018.

[10] W. Saputra, "Analysis Resilient Algorithm on Artificial Neural Network Backpropagation " *International Conference on Information and Communication Technology,* 2017.

[11] J. Zhao, "Gaussian Process Regression for Prediction of Sulfate Content in Lakes of China " *Journal of Engineering and Technological Sciences,* vol. 51, pp. 198-215, 2019.

[12] R. R. Fiska, "Penerapan Teknik Data Mining dengan Metode Support Vector Machine (SVM) untuk Memprediksi Siswa yang Berpeluang Drop Out " *Jaringan Sistem Informasi Robotik* 2017.

[13] M. D. Yalidhan, "Implementasi Algoritma Backpropagation Untuk Memprediksi Kelulusan Mahasiswa," *Kumpulan jurnaL Ilmu Komputer,* vol. 5, 2018.

[14] P. Suksawang, "Electricity Consumption Forecasting in Thailand using Hybrid Model SARIMA and Gaussian Process with Combine Kernel Function Technique," *International Journal of Energy Economics and Policy* 2018.

[15] N. N. Putriwijaya, "Peramalan Jumlah Pemakaian Air di PT. Pembangkitan Jawa Bali Unit Pembangkit Gresik Menggunakan Support Vector Regression," 2018.

[16] D. A. Nasution, "Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-Nn," *Journal of Computer Engineering System and Science,* vol. 4, pp. 78-82, 2019.

[17] N. M. Ashar1, "Penerapan Metode Extreme Learning Machine (ELM) Untuk Memprediksi Jumlah Produksi Pipa Yang Layak (Studi Kasus Pada PT. KHI Pipe Industries)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Kompute,* vol. 2, pp. 4621-4628 2018.

[18] U. Khair1, "Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error," *International Conference on Information and Communication Technology* 2017.