Journal of Intelligent System and Telecommunications

Journal homepage: https://journal.unesa.ac.id/index.php/jistel/index

Predicting Regional Sanitation Conditions Using Support Vector Machines and Neural Networks

Rosalia Indah Trisanti[1], Agam Nizar Dwi Nur Fahmi[2]*

[1]Department of Electrical Engineering, Faculty of Sains and Engineering, Universitas Bangka Belitung, Bangka, Indonesia

[2]Department of Electrical Engineering, Faculty Engineering, Universitas Muhammadiyah Jember, Jember, Indonesia

*Correspondence: E-mail: [1]trisantirosalia@gmail.com, [2]agamnizardwi10@gmail.com

ARTICLE INFO

Article History:
Submitted/Received 10 December 2024
First Revised 20 December 2024
Accepted 20 December 2024
First Available Online 21 December 2024
Publication Date 21 December 2024

Keyword:
Sanitation, Neural Network, SVM, dengue fever, Artificial Intelligence.

ABSTRACT

Sanitation conditions are very important for an area. This is what makes an area comfortable and free from various diseases, one of which is dengue fever. This is because sanitation conditions are one of the important factors in reducing the number of dengue fever patients. This study aims to predict the level of sanitation conditions in an area using Neural Network and Support Vector Machine (SVM). This study is part of a study to determine dengue fever in the Bangka Belitung area of Indonesia. The parameters used are a learning rate of 0.1, momentum 0.1 and learning is carried out for 500 epochs. The results show that using the Neural Network method, an accuracy of 95% was obtained and the SVM method obtained an accuracy of 96%. Based on the results of the system, it can be concluded that the system has been able to run well. The contribution of this study is that it can be a Decision Support System in determining the sanitation level of an area, and related parties can immediately take follow-up action.

1. INTRODUCTION

Dengue fever, caused by mosquito bites where the mosquito's saliva contains the dengue virus. Dengue fever does not only occur in Indonesia, but has become a global community problem [1]. Cases of fever are increasingly worrying and the number is increasing [2]. In Indonesia in 2022, according to the Ministry of Health of the Republic of Indonesia, cases of dengue fever reached 131,265 cases, around 40% of sufferers were children aged 0 to 14 years and around 73% of deaths due to dengue fever were children. Cases of dengue fever in Bangka Regency have increased in recent years. Based on information from the Bangka Regency Health Office, cases of dengue fever in the 2020 period were 119 cases, then increased in the 2021 period to 149 cases, then jumped to 290 cases in the 2022 period [4]. To deal with the increase in dengue fever cases, the Bangka Regency Health Office has initiated various actions, short-term and long-term actions. In short-term actions, adult mosquito vector interventions such as fumigation or fogging are carried out, while for long-term actions, mosquito larvae control is carried out in the form of the One House One Mosquito Net Movement (G1R1J) for school children [5]. Although various actions have been taken, dengue fever cases are still increasing.

Currently, the classification of dengue fever cases in the Regency is carried out manually, namely by examining historical data from dengue fever cases that have occurred. A very important thing is the sanitation conditions of an area. Based on this, a decision support system is needed to classify the sanitation conditions of an area quickly and efficiently. In this case, it can utilize technological developments by implementing a more sophisticated classification model. This is also in accordance with the national strategy program for dengue control from 2021 to 2026 [6], which is the development of research, studies, and innovations which are the basis for policies that can be used as a guideline to introduce more sophisticated and up-to-date forecasting methods.

The research that has been done is a study on the prediction of dengue fever. The KNN method has good performance in predicting the period of hospitalization days for dengue fever patients, with an accuracy value of 65.67% [7]. Clustering of dengue fever patients devided into 2 clusters. The Medoids method in clustering is better than the K-Means method, because the K-Medoids method can overcome the problem of outliers in the data used [8]. The K-Means method is effective for predicting dengue fever, 60% of this model is effective in predicting dengue fever cases with positive results [9]. The SVM method has better performance in predicting dengue fever cases, where the accuracy value using the SVM method is higher than the Decision Tree method, which is 99% [10]. CBR has good performance with an accuracy value of 80%, this shows that CBR can identify vital signs and blood profiles for the treatment of dengue fever at the Hospital in Negombo Sri Lanka [11]. SVM has better performance in predicting dengue fever compared to the KNN method, with an accuracy value of 100% [12]. The KNN model is able to predict dengue fever cases, where the MAE and RMSE values obtained in the KNN model with K = 3 are 16,637 and 29,145 [13]. SVM has good performance in predicting dengue fever outbreaks in Selangor, Malaysia, with an accuracy value of 70% [14]. The ANN model has better performance than Decision Tree in Predicting Dengue Fever Outbreaks Based on Meteorological Data, with an accuracy value of 68.85% [15]. The SVM method has good performance in predicting dengue fever, where the cross validation, K-Fold validation, Precision and F1-Score values are higher than other methods, namely 98.5%, 97.5%, 98.2% and 98% respectively [16].

The above research discusses more about predicting dengue fever cases in various places, but there is still very limited research on the sanitation conditions of an area. Research discussing sanitation using K-Means shows a system accuracy of 0.8645, average sensibility

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

is 0.8775, average specificity 0.9231, and average precision 0.9176 [17]. The difference with this research is that the SVM and Artificial Intelligence methods are used, where previous research shows that using this method produces high accuracy.

The purpose of this study is to predict the potential sanitation conditions of an area using Support Vector Machines and Artificial Neural Networks. The areas that are the objects of research are 8 Bangka Regencies, Bangka Belitung Province, Indonesia.

The contribution of this research is to provide considerations or Decision Support Systems to related parties to immediately take action before cases occur. The hope is that if the sanitation of an area is good, it can at least reduce the potential for the spread of dengue fever. The originality of this research is that this measurement or prediction is usually done manually, but by using Artificial Intelligence, the prediction becomes more accurate.

2. METHODS

2.1. Research stage

The research stage begins with first observing the area that is the object of research and conducting interviews with related parties. Furthermore, data collection is carried out, learning processes are carried out, testing is carried out, and results are evaluated. The learning process is carried out using Support Vector Machine and Neural Network. The input factors of the system are septic tank management, the habit of washing hands with soap, household food and drinking water management, household waste disposal management, and household liquid waste management. While the output is the level of regional sanitation conditions. The level of regional sanitation consists of levels 1 (low), 2 (good), and 3 (very good).

2.2. Material

The data used was obtained from the Bureau of Statistics regarding the number of culverts and rivers, population, and altitude of the area), for data on sanitation conditions and data on the number of dengue fever patients obtained from the Bangka Regency Health Office, Bangka Belitung Province. This research was conducted from 2018 to 2022, where the data was divided into 2, namely learning data and testing data.

2.3 Support Vector Machine

The SVM algorithm works by mapping data to a feature space so that data points can be categorized, even when the data cannot be separated linearly. If a separator between categories is successfully found, the data can be transformed in such a way that the separator can be described as a hyperplane. Then, the characteristics of the new data can be used to predict which group the new record should be in. SVM can categorize linear and non-linear data. In this study, it is divided into 3 clusters, namely the sanitation level of this area consisting of level 1 (low), 2 (good), and 3 (very good).

2.3 Neural Network

The algorithm of the Neural Network is as follows: Enter the vector X into the input layer; calculate the output of each neuron in the hidden layer and output layer; calculate the activation function in the output layer; update the weights so that the output of the ANN is consistent with the class label in the training data.

2.4 System evaluation

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

Evaluation process is important because it measures the performance of the proposed model and ensures how well the model is used in producing the right classification. Model performance usually uses measures such as accuracy, precision, recall and F1-Score. Precision describes the amount of data with positive categories that are correctly classified. Recall shows what percentage of positive category data is correctly classified by the system. F1-Score shows the balance between recall and precision. In this study, the matrix used to evaluate model performance is the Confusion Matrix which can be used to calculate the accuracy, precision, recall and F1-Score values [18].

$$Accuracy = \frac{TP + TN}{n} x100 \quad ...(1)$$

$$Precision = \frac{TP}{TP + FP} x100 \quad ...(2)$$

$$Recall = \frac{TP}{TP + FN} x100 \quad ...(3)$$

$$F1\ Score = 2x \frac{Precision\ x\ Recall}{Precision + Recall} \quad ...(4)$$

Whereas:
TP : True Possitive
TN : True Negative
FP : False Possitive
FN : False Negative

3. RESULTS AND DISCUSSION

The research has been conducted, the first stage used Neural Network and the second stage used SVM. In the first study using Neural Network, and the Confusion matrix was obtained in **Table 1**. **Table 1** explains that level 1 can be classified correctly 193 data, level 2 there are 189 data that are classified correctly, and level 3 there are 189.

Table 1. Confusion Matrix using Neural Network

	Level 1	Level 2	Level 3
Level 1	193	7	0
Level 2	3	189	8
Level 3	0	11	189

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

The system performance is described in **Table 2**, where an accuracy of 0.95 is obtained, with a precision value of 0.95, a recall value of 0.94, and an F1 score of 1. This fact shows that the system has been able to run well.

Table 2. System performance using Neural Network

Performance	Value
Accuracy	0.95
Precision M1	0.97
Precision M2	0.95
Precision M3	0.95
Precision	0.95
Recall M1	0.95
Recall M2	0.91
Recall M3	0.96
Recall	0.94
F1 score	1.00

Receiver operating characteristic (ROC) on the results of the testing process using Neural Network is depicted as in **Figure 1**. Based on this graph, calculations are carried out to calculate the Area Under The Curve (AUC). The AUC calculation obtained a value of 0.95, based on the AUC criteria if the value is between 0.9 and 1 then it is in the excellent criteria. So we can conclude that the system is in the excellent criteria.

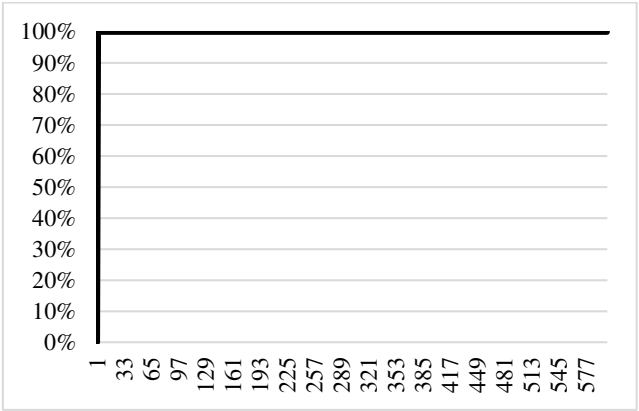

Figure 1. ROC using Neural Network

The results of the Confusion Matrix when applying Support Vector Machine are explained in **Table 3**. When this method is applied, it can be seen that level 1 can correctly classify 194 data, and this is 1 data higher when compared to using Neural Network. Likewise at level 2, the data that can be correctly classified is 190 data. While at the third level that can be correctly classified is 192. This shows (in this case) that SVM produces higher accuracy than using Neural Network.

Table 3. Confusion Matrix menggunakan Support Vector Machine (SVM)

	Level 1	Level 2	Level 3
Level 1	194	6	0
Level 2	2	190	8
Level 3	0	8	192

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

The system performance is described in **Table 4**, where the accuracy is 0.96, with a precision value of 0.96, a recall value of 0.95, and an F1 score of 1. This fact shows that the system using SVM has also been able to run well. Based on **Table 4**, it also shows that the system performance (precision and recall values) tends to be higher than using Neural Network. Both systems get an F1 score of 1, because the precision and recall values have almost the same value.

Table 4. System performance using Support Vector Machine (SVM)

Performance	Value
Accuracy	0.96
Precision M1	0.97
Precision M2	0.95
Precision M3	0.96
Precision	0.96
Recall M1	0.96
Recall M2	0.93
Recall M3	0.96
Recall	0.95
F1 score	1.00

Receiver operating characteristic (ROC) on the results of the testing process using SVM is depicted as in **Figure 2**. Based on this graph, calculations are carried out to calculate the Area Under The Curve (AUC). The AUC calculation obtained a value of 0.96, so we can conclude that the system is in the excellent criteria.

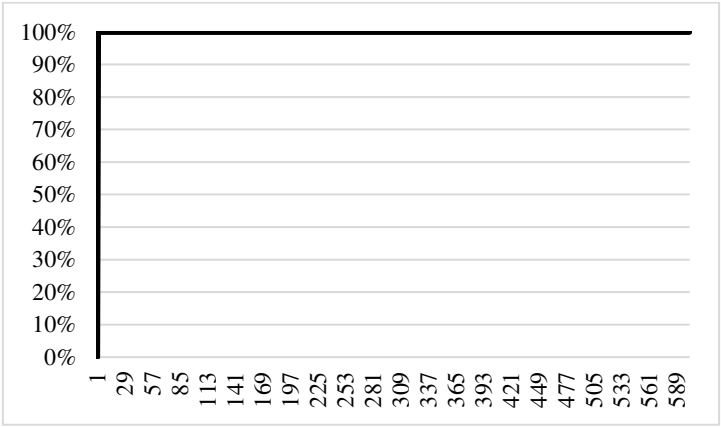

Figure 2. ROC using SVM

Efforts to prevent disease by removing or controlling various environmental components that are responsible for the chain of disease transmission are also called environmental sanitation [19]. An environment with good sanitation can prevent the spread of disease, while an environment with poor sanitation can contribute to providing a habitat for the breeding of Ae. Aegypti mosquitoes. The population of Ae. Aegypti mosquitoes is often associated with poor sanitation, as well as inadequate waste management (WHO, 2009). Thus, to improve environmental sanitation, the STBM program can be carried out, which is a government program to improve a clean environment. Community-Based Total Sanitation or STBM is a policy made by the government to encourage behavioral changes towards a clean and healthy

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

life by involving active participation from individuals and the community [20]. Some inputs from sanitation conditions are septic tank management, the habit of washing hands with soap, household food and drinking water management, household waste disposal management, and household liquid waste management.

Sungailiat and Pemali areas are included in cluster 3, which is a high level of sanitation. This monitoring is very important. This is evident from the Belinyu, Riau Silip, and Mendo Barat areas from January to September being included in cluster 1 (having a low level of sanitation), but its results followed up by anticipation action, from October to December they changed to being included in cluster 3, which is having a high level of sanitation. Another fact is that the Merawang and Puding Besar areas from January to September were included in cluster 2 (having a moderate level of sanitation), and from October to December they were included in cluster 3, which was having a high level of sanitation. This also happened in the Bakam area from January to July it was included in cluster 2, which was having a moderate level of sanitation and from August to December it was included in cluster 3, which was having a high level of sanitation. Then the results of the sanitation level clustering in Bangka Regency in 2022 showed that all areas were included in cluster 3, which was having a high level of sanitation.

Research on SVM has been widely used by previous researchers [21] where it was carried out in the prediction of total soluble solids of industrial tomatoes. Other applications are used to predict heterogeneous machine parameters in industry 4.0 [22]. Based on the results above, good results were obtained. Meanwhile, the implementation of Neural Network in sanitation is to predict water pollution [23], the results of this study can predict water pollution so that it can be anticipated immediately. In addition, it can model and simulate the environment [24].

Research on sanitation using machine learning has been studied by previous researchers, the indicators used are such as washing hands before meals and drinking water from natural water sources, and sanitation facilities. The results of this study showed an accuracy of 70% [25]. The overall system, in this case, shows that using SVM produces higher accuracy than Neural Network. But both still show good performance, when viewed from the performance of accuracy, precision, recall, and F1 score. When viewed from the AUC value, both methods have obtained excellent results, because the AUC value of the system (using both Neural Network and SVM) is still in the range between 0.9 and 1.

4. CONCLUSION

In the performance evaluation, it was found that the proposed model in classifying the level of sanitation has good performance. The results of the performance evaluation of the proposed model (using Neural Network) have values of accuracy, precision, recall and F1-Score obtained respectively of 0.95, 0.954, 0.94 and 1. While the application of SVM the results of the calculation of the values of accuracy, precision, recall and F1-Score are respectively 0.96, 0.96, 0.95 and 1. Based on this, the proposed model can be relied on in the level of regional sanitation based on septic tank management, the habit of washing hands with soap, household food and drinking water management, household waste disposal management, and household liquid waste management.

Future research is proposed to explore the causes/determinants of sanitation in an area, and not only dengue fever. Then re-learning is carried out to obtain a more robust system. Other methods can also be used, with the hope of obtaining a higher accuracy value.

5. ACKNOWLEDGMENT

DOI: http://dx.doi.org/10.xxxxx/jistel.vXiX
e- ISSN xxxx-xxxx

We would like to thank the Statistics Bureau and the Health Service of Bangka Regency, Bangka Belitung Province, Indonesia, for providing data for this study.

6. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. The authors confirmed that the paper was free of plagiarism.

7. AUTHORS' CONSTRIBUTION/ROLE

Rosalia Indah Trisanti1: Conceptualization, Methodology, Writing Original Draft, Investigation; Agam Nizar Dwi Nur Fahmi: Formal Analysis, Writing Original Draft, Supervision.

8. REFERENCES

[1] M. R. Pereira, "Global rise in dengue transmission—2023," Am. J. Transplant., vol. 24, no. 3, pp. 318–319, Mar. 2024, doi: 10.1016/j.ajt.2024.01.003.

[2] N. Haider, M. N. Hasan, J. Onyango, and M. Asaduzzaman, "Global landmark: 2023 marks the worst year for dengue cases with millions infected and thousands of deaths reported," IJID Reg., vol. 13, p. 100459, 2024, doi: https://doi.org/10.1016/j.ijregi.2024.100459.

[3] Rokom. (2023). Atasi Dengue, Kemenkes Kembangkan Dua Teknologi ini. https://sehatnegeriku.kemkes.go.id/baca/umum/202 3025/3642353/

[4] Kasmono. (2023). Dinkes Kabupaten Bangka catat penyakit DBD sebanyak 290 kasus. Antaranews.Com. https://babel.antara news.com/berita/328068/dinkes-kabupaten-bangka-catat penyakit-dbd-sebanyak-290-kasus

[5] Dinkes Kabupaten Bangka. (2022). Rapat Koordinasi Lintas Program Penanggulangan KLB DBD di Kabupaten Bangka. https://dinkes.bangka.go.id/berita/rapat-koordinasi lintas -program-penanggulangan-klb-dbd-di-kabupaten bangka

[6] Samad, I., Handito, A., Sugiarto, A., Setiani, E., Gunawan, D., & Silalahi F. S. M., et al. (2022). Membuka Lembaran Baru : Laporan Tahunan 2022 Demam Berdarah Dengue

[7] Wulandari, D. A. P., Permana, K. A. B., & Sudarma, M. (2018). Prediction of Days in Hospital Dengue Fever Patients using K-Nearest Neighbor. International Journal of Engineering and Emerging Technology, 3, No.1, 23–25

[8] Alibuhtto, M. C., & Mahat, N. I. (2018). Applications of Partition Based Algorithms for Clustering Dengue Patients in Sri Lanka. International Journal of Advance Scientific Research and Engineering Trends, 3(7), 27–31.

[9] Yogapriya, P., & Geetha, P. (2019). K-means Clustering Algorithm for Dengue Disease Detection using Tanagra Tool. International Journal of Computer Sciences and Engineering, 7(2), 765–768. https://doi.org/10.26438/ijcse/v7i2.765768.

[10] Sanjudevi, R., & Savitha, D. (2019). Dengue Fever Prediction Using Classification Techniques. International Research Journal of Engineering and Technology (IRJET), 06(02), 558–563.

[11] Briyatis, S. H. U, Premaratne, D. G., & Silva, H. D. (2019). A Novel Method for Dengue Management Based on Vital Signs and Blood Profile. International Journal of Engineering and Advanced Technology (IJEAT), 8(6S3), 154–159. https://doi.org/10.35940/ijeat.F1025.0986S319

[12] Ganthimathi, M., Thangamani, M., Mallika, C., Balaji, V. P. (2020). Prediction of Dengue Fever Using Intelligent Classifier. International Journal of Emerging Trends in Engineering Research, 8, No.4, 1338–1341. https://doi.org/10.30534/ijeter/2020/65842020

[13] Parikh, K., Makadia, T., & Patel, H. (2020). Feature Extraction and Prediction of Dengue Outbreaks. Interntional Journal of Scientific Research in Computer Science, Engineering and Information Technology, 6(5), 216–222. https://doi.org/10.32628/CSEIT206544

[14] Salim, N. A. M., Wah, Y. B., Reeves, C., Smith, M., Yaacob, W. F. W., Mudin, R. N., Dapari, R., Sapri, N. N. F. F., & Haque, U. (2021). Prediction of Dengue Outbreak in Selangor Malaysia Using Machine Learning Techniques. Scientific Reports, 11(1), 1–9. https://doi.org/10.1038/s41598-020-79193-2

[15] Krishnan, N. F. M., Zukarnain, Z. A., Ahmad, A., & Jamaludin, M. (2022). Predicting Dengue Outbreak based on Meteorological Data Using Artificial Neural Network and Decision Tree Models. International Journal on Informatics Visualization, 6(3), 597–603. https://doi.org/10.30630/joiv. 6.2.788

[16] Thirugnanam, A., & Hussain, F. B. J. (2023). Exploring Machine Learning Algorithms for the Prediction of Dengue: A Comprehensive Review. International Information and Engineering Technology Association, 37(5), 1281–1290. https://doi.org/10.18280/ria.370521.

[17] R. I. Trisanti, G. P. A. Buditjahjanto, and L. Anifah, "Application of K-Means Algorithm for Clustering Regions Based on Sanitation Level: A Case Study in Bangka Regency, Indonesia," in 2023 Sixth International Conference on Vocational Education and Electrical Engineering (ICVEE), 2023, pp. 144–149. doi: 10.1109/ICVEE59738.2023.10348267.

[18] Bahri, S., Wibowo, A., Wajhillah, R., & Suhada, S. (2020). Data Mining; Algoritma Klasifikasi & Penerapannya dalam Aplikasi. Graha Ilmu.

[19] Ehlers, V. M., & Steel, E. W. (1937). Municipal and rural sanitation (2nd ed.). New York: McGraw-Hill Book Company.

[20] Peraturan Menteri Kesehatan Republik Indonesia, No.3 Tahun 2014 Tentang Sanitasi Total Berbasis Masyarakat, 2014

[21] C.-N. Kasimatis et al., "Implementation of a decision support system for prediction of the total soluble solids of industrial tomato using machine learning models," Comput. Electron. Agric., vol. 193, p. 106688, 2022, doi: https://doi.org/10.1016/j.compag.2022.106688.

[22] L. Romeo, J. Loncarski, M. Paolanti, G. Bocchini, A. Mancini, and E. Frontoni, "Machine learning-based design support system for the prediction of heterogeneous machine parameters in industry 4.0," Expert Syst. Appl., vol. 140, p. 112869, 2020, doi: https://doi.org/10.1016/j.eswa.2019.112869.

[23] J. Pal and D. Chakrabarty, "Effects of input/output parameters on artificial neural

network model efficiency for breakthrough contaminant prediction," Water Supply, vol. 21, no. 7, pp. 3614–3628, 2021, doi: 10.2166/ws.2021.125.

[24] I. Ciuca and A. Ware, "A Neural Network Environment for Modelling and Simulation," IFAC Proc. Vol., vol. 30, no. 19, pp. 535–541, 1997, doi: https://doi.org/10.1016/S1474-6670(17)42354-8.

[25] A. Taweesan, T. Koottatep, T. Kanabkaew, and C. Polprasert, "Application of machine learning in sanitation management prediction: Approaches for achieving sustainable development goals," Environ. Sustain. Indic., vol. 22, p. 100374, 2024, doi: https://doi.org/10.1016/j.indic.2024.100374.