Questionnaire Based Hospital Patient Satisfaction Level Classification With Support Vector Machine

Khoirunsyah Dalimunthe[1], Hartono[2], B. Herawan Hayadi[3]

[1,2,3]*Universitas Potensi Utama*

Jl. KL. Yos Sudarso Km. 6,5 No. 3-A, Tanjung Mulia, Medan

[1]choyrun2@gmail.com
[2]hartonoibbi@gmail.com
[3]b.herawan.hayadi@gmail.com

Abstract—**The utilization of machine learning in various questionnaire-based classifications, especially using the Support Vector Machine (SVM) algorithm, has piqued our interest in conducting research on hospital patient satisfaction levels through a survey. Using nine questions as features and measuring the patients' willingness to recommend RS Haji Medan to others, we built three classification models with Polynomial, RBF, and Sigmoid kernel functions. Out of the 86 responses we received, our t-test validation test revealed that all the questions we asked are valid for use in the classification process. The results show that the Polynomial model produced the highest accuracy (90.5%), precision (91.8%), and recall (90.5%) when compared to the RBF and Sigmoid models. Furthermore, the generated model exhibits stable performance, with an average difference of less than 7% between the training and testing performance. This stability suggests promising resistance to overfitting and underfitting.**

***Keywords*—Questionnaire, Classification, SVM, Polynomial, RBF, Sigmoid**

I. Introduction

Nowadays, hospitals have become a business model aimed at gaining profit, which follows four standard concepts: preventive medicine, participatory medicine, predictive medicine, and personalized medicine [1]. By utilizing these four concepts, each hospital competes to attract patients with various services to satisfy their needs [2]. Given the competitiveness of hospital services, it is inevitable that patient satisfaction is an important factor for them to return to use those services or even recommend the hospital to others [3]. In this study, we analyze the level of patient satisfaction using a classification model based on the Support Vector Machine (SVM) algorithm. We sourced our data from RS Haji Medan patients and their family members by providing them with a questionnaire about their experiences while using the hospital's services.

Questionnaire-based classification is not a new field in machine learning, especially when using the SVM algorithm. Some research studies we have found already demonstrate the feasibility of using this type of dataset, as seen in soccer player talent identification [4], piano lesson teaching satisfaction [5], engineering student choice of learning media [6], evaluation of fruit waste recycling behavior [7], and determining the factors related to hallux valgus [8]. From these research studies, we conclude that validating the questionnaire responses is crucial before using them in the classification process. That's why, in this study, we employ the t-test validation method to assess the validity of all the answers received.

In this study, we compared three kernel functions, namely Polynomial, Radial Basis Function (RBF), and Sigmoid. These three functions are frequently compared in terms of their performance within SVM models, such as in the prediction of Covid-19 cases based on X-ray images, where the Polynomial kernel produced a model with a higher accuracy rate (86%) compared to the other two models [9]. In [10], the RBF kernel function outperformed the other two kernels with the highest accuracy rate of 99.26% when implemented in the DDoS Attack classification model. In the research on consumer price index prediction, the RBF kernel was once again chosen as the best model because it produced the smallest MAPE value, which was 1.8242 [11]. Another study on online spam reviews classification once again demonstrated the superiority of the RBF kernel over Polynomial and Sigmoid kernels, achieving the highest accuracy rate of 89.02% [12]. The comparison of Polynomial, RBF, and Sigmoid in [13] shows that the Sigmoid kernel achieved the best performance with an accuracy of 80.49% when classifying online sentiment. From these research studies, we observed that each kernel function produced the best model against the others in different classification problems. When comparing the three kernel functions, we used a standardized kernel equation with the same parameter values for each kernel.

When splitting the data into training and testing sets, it's important to choose the best sampling ratio, such as the common 70:30 split, where the dataset is divided into 70% training data and 30% testing data [14]. The drug abuse detection research successfully implemented a 70:30 ratio in the SVM algorithm, achieving an accuracy rate of 83.3% [15]. In the research on multi-label text classification using the SVM algorithm, the 70:30 ratio was again employed and successfully resulted in a model with an accuracy rate of 81% [16]. In [17], the dataset of participants in the Jogja International Scout Camp 2020 was split into a 70:30 ratio, which was then used for selecting potential participants to represent the North Sumatra region, resulting in an SVM model achieving an accuracy score of 95%. The 70:30 sampling ratio used in sentiment analysis of the Ruang Guru application with the SVM algorithm yielded an accuracy result of 84% [18]. In the phishing website detection research using the SVM algorithm, the 70:30 ratio was employed and managed to obtain an accuracy value of 89.84% [19]. The results of these five research studies show that the 70:30 sampling ratio produced high accuracy values, reaching above 80%. In this research, we modified the ratio from 70:30 to 75:25 to determine whether the model's performance could still achieve an accuracy above 80%.

When splitting the data into training and testing sets, it's important to choose the best sampling ratio, such as the common 70:30 split, where the dataset is divided into 70% training data and 30% testing data. The drug abuse detection research successfully implemented a 70:30 ratio in the SVM algorithm, achieving an accuracy rate of 83.3% [15]. In the research on multi-label text classification using the SVM algorithm, the 70:30 ratio was again employed and successfully resulted in a model with an accuracy rate of 81% [16]. In [17], the dataset of participants in the Jogja International Scout Camp 2020 was split into a 70:30 ratio, which was then used for selecting potential participants to represent the North Sumatra region, resulting in an SVM model achieving an accuracy score of 95%. The 70:30 sampling ratio used in sentiment analysis of the Ruang Guru application with the SVM algorithm yielded an accuracy result of 84% [18]. In the phishing website detection research using the SVM algorithm, the 70:30 ratio was employed and managed to obtain an accuracy value of 89.84% [19]. The results of these five research studies show that the 70:30 sampling ratio produced high accuracy values, reaching above 80%. In this research, we modified the ratio from 70:30 to 75:25 to determine whether the model's performance could still achieve an accuracy above 80%.

By utilizing a dataset directly obtained from the questionnaire responses of patients and their families at RS Haji Medan, this research implements the SVM algorithm with three different kernel functions (Polynomial, RBF, Sigmoid) to classify the level of patient satisfaction with the hospital, using a 75:25 sampling ratio. The target class of respondents' willingness to recommend RS Haji Medan to others is used as the target variable in this study, which is evaluated using accuracy, precision, and recall values based on the confusion matrix results generated by each model.

II. Research Methods

A. Data Collection

This research used primary data collected directly through questionnaires distributed to 100 patients/families of patients at RS Haji Medan. Out of the 100 questionnaires distributed, only 86 were returned, so the data used in this study amounted to 86. We provided 9 questions about the services received by patients at the hospital, which were then used as features in the classification. For the target class, we asked a question about the level of willingness of patients to recommend RS Haji Medan to others. Table 1 shows the questions provided in the questionnaire.

Based on the questionnaire responses obtained, we conducted tabulation for each answer option, and the results are as seen in Table 2.

B. Data Validation

Before using the questionnaire results as a dataset in the classification process, we conducted validation of the respondents' answers using a t-test. This research uses a significance level of 0.05 (95% confidence) with degrees of freedom equal to n-2. By using the values in the t-test statistical table, the t-table value for a 95% confidence level and 84 degrees of freedom is 1.99. We use this value to compare the calculated t-value from the questionnaire responses to assess the validity of the answers. Table 3 shows the comparison results between the calculated t-statistic and the t-table values based on the questionnaire responses in this study.

Based on the values indicated in Table 3, it is evident that all the questionnaire responses have t-statistic values greater than the t-table value (1.99). This demonstrates that all the questionnaire responses are valid and can be used in the subsequent classification process.

C. Data Validation

Before using the questionnaire results as a dataset in the classification process, we conducted validation of the respondents' answers using a t-test. This research uses a significance level of 0.05 (95% confidence) with degrees of freedom equal to n-2. By using the values in the t-test statistical table, the t-table value for a 95% confidence level and 84 degrees of freedom is 1.99. We use this value to compare the calculated t-value from the questionnaire responses to assess the validity of the answers. Table 3 shows the comparison results between the calculated t-statistic and the t-table values based on the questionnaire responses in this study.

TABLE I
QUESTIONNAIRE QUESTIONS

Questions	Answer Options	Values
Overall, how satisfied are you with your experience at this hospital?	Very Satisfied Satisfied Neutral Dissatisfied	4 3 2 1
To what extent do you feel attended to by the medical staff at this hospital?	Very Attentive Attentive Neutral Inattentive	4 3 2 1
To what extent do you feel medical information is conveyed clearly and easily understood by the medical staff at this hospital?	Very Clearly Clearly Neutral Unclear	4 3 2 1
To what extent do you feel your waiting time in the waiting area was in line with the estimated time given?	Very Much in Line In Line Neutral Not in Line	4 3 2 1
To what extent do you feel the facilities at this hospital are sufficient and comfortable?	Very Sufficient and Comfortable Sufficient and Comfortable Neutral Insufficient and Uncomfortable	4 3 2 1
To what extent do you feel the administrative procedures at this hospital are efficient and time-saving?	Very Efficient and Time-saving Efficient and Time-saving Neutral Inefficient and Time-consuming	4 3 2 1
To what extent do you feel the hospital staff is friendly and courteous?	Very Friendly and Courteous Friendly and Courteous Neutral Unfriendly and Discourteous	4 3 2 1
To what extent do you feel involved in the decision-making process regarding your treatment at this hospital?	Very Involved Involved Neutral Not Involved	4 3 2 1
To what extent do you feel you were provided with information about post-hospitalization treatment and care at this hospital?	Very Informed Informed Neutral Uninformed	4 3 2 1
To what extent are you willing to recommend this hospital to others based on your experience?	Very Willing Willing Neutral Unwilling	4 3 2 1

TABLE II
DISTRIBUTION OF QUESTIONNAIRE RESPONSES

Question Number	Number of Answer per Value			
	4	**3**	**2**	**1**
1	35	32	13	6
2	27	39	13	7
3	29	33	19	5
4	23	43	14	6
5	30	33	17	6
6	26	36	18	6
7	36	28	16	6
8	31	40	8	7
9	20	41	21	4
10	49	24	13	0

TABLE III
DATA VALIDATION RESULT

Question Number	t-value	Validation
1	4,88298	Valid
2	3,892552	Valid
3	3,356974	Valid
4	5,167423	Valid
5	4,456581	Valid
6	5,754879	Valid
7	4,447735	Valid
8	3,783667	Valid
9	3,146756	Valid
10	4,985699	Valid

D. Support Vector Machine

We implemented the SVM (Support Vector Machine) algorithm in the patient satisfaction classification process at RS Haji Medan, using three different kernel function variations. The first kernel function we used is Polynomial, which transforms data into a higher-dimensional space based on the given degree (d) [20]. In the Polynomial kernel function, we use the scaling factor (g) = 0.41, the constant value (c) = 0.1, and the degree (d) = 2, into the following equation (1) [21]:

$$K(x, y) = (g * x.y + c)^d \quad (1)$$

Where:
K(x, y) : The kernel function value between data points x and y.
x and y : The data points being compared.
d : The degree of the polynomial.
g : The scaling factor value.
c : The constant term value.

The second kernel function used in this study is the Radial Basis Function (RBF), which transforms data into a higher-dimensional space based on the value of gamma (g) that serves as the decision boundary [22]. In the RBF kernel function, we use the gamma (g) = 0.41 into the following equation (2) [23]:

$$K(x, y) = exp(-g * \|x - y\|^2) \quad (2)$$

Where:
K(x, y) : The kernel function value between data points x and y.
x and y : The data points being compared.
g : The gamma value.

The last kernel function used in this study is the Sigmoid, which transforms data into a higher-dimensional space using similarity scaling (g) and decision boundary control (c) values [24]. In the Sigmoid kernel function, we use the similarity scaling (g) = 0.41 and decision boundary control (c) = 0.1 into the following equation (3) [25]:

$$K(x, y) = tanh(g * x.y + c) \quad (3)$$

Where:
K(x, y) : The kernel function value between data points x and y.
x and y : The data points being compared.
g : The similarity scaling value.
c : The decision boundary control value.

E. Confusion Matrix

This research employs a 75:25 sampling ratio, where 75% of the data is used as training data, and 25% is used as testing data in the classification process. The classification results, in the form of predictions for each category of the target class (Satisfied, Neutral, and Dissatisfied), are presented in the form of a confusion matrix. The true positive, false positive, true negative, and false negative values from the resulting confusion matrix are then calculated to obtain accuracy, precision, and recall values using equations (4) through (6) as follows [26]:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

Where:
TP : Represents the cases where the model correctly predicted a positive class (predicts a response as "Satisfied" when it is indeed a "Satisfied" response).
FP : Represents the cases where the model incorrectly predicted a positive class (predicts a response as "Satisfied" when it is actually "Neutral" or "Dissatisfied").
TN : Represents the cases where the model correctly predicted a negative class (predicts a response as "Neutral" or "Dissatisfied" when it is indeed "Neutral" or "Dissatisfied").
FN : Represents cases where the model incorrectly predicted a negative class (predicts a response as "Neutral" or "Dissatisfied" when it is actually "Satisfied").

III. RESULT

A. Classification Result for Polynomial Model

We obtained the confusion matrix for the Polynomial model shown in Figure 1 using the following values: g = 0.41, c = 0.1, and d = 2. Figure 1 (a) depicts the confusion matrix from the training process, while Figure 1 (b) depicts the confusion matrix from the testing process.

	Neutral	Satisfied	Dissatisfied
Neutral	17	2	0
Satisfied	0	37	0
Dissatisfied	0	0	9

(a) Training Result

	Neutral	Satisfied	Dissatisfied
Neutral	3	2	0
Satisfied	0	12	0
Dissatisfied	0	0	4

(b) Testing Result

Fig1. Polynomial Model Confusion Matrix

In the training process, the Polynomial model correctly predicted 17 out of 19 Neutral class targets, 37 out of 37 Satisfied class targets, and 9 out of 9 Dissatisfied class targets. In the testing process, the Polynomial model correctly predicted 3 out of 5 Neutral class targets, 12 out of 12 Satisfied class targets, and 4 out of 4 Dissatisfied class targets.

B. Classification Result for RBF Model

We obtained the confusion matrix for the RBF model shown in Figure 2 using the value of g = 0.41. Figure 2 (a) depicts the confusion matrix from the training process, while Figure 2 (b) depicts the confusion matrix from the testing process.

	Neutral	Satisfied	Dissatisfied
Neutral	19	0	0
Satisfied	0	37	0
Dissatisfied	1	2	6

(a) Training Result

	Neutral	Satisfied	Dissatisfied
Neutral	3	2	0
Satisfied	0	12	0
Dissatisfied	0	4	0

(b) Testing Result

Fig 2. RBF Model Confusion Matrix

In the training process, the RBF model correctly predicted 19 out of 19 Neutral class targets, 37 out of 37 Satisfied class targets, and 6 out of 9 Dissatisfied class targets. In the testing process, the RBF model correctly predicted 3 out of 5 Neutral class targets, 12 out of 12 Satisfied class targets, and 0 out of 4 Dissatisfied class targets.

C. Classification Result for Sigmoid Model

We obtained the confusion matrix for the Sigmoid model shown in Figure 3 using the values of g = 0.41, and c = 0.1. Figure 3 (a) depicts the confusion matrix from the training process, while Figure 3 (b) depicts the confusion matrix from the testing process.

	Neutral	Satisfied	Dissatisfied
Neutral	19	0	0
Satisfied	0	37	0
Dissatisfied	9	0	0

(a) Training Result

	Neutral	Satisfied	Dissatisfied
Neutral	4	1	0
Satisfied	0	12	0
Dissatisfied	4	0	0

(b) Testing Result

Fig 3. Sigmoid Model Confusion Matrix

In the training process, the Sigmoid model correctly predicted 19 out of 19 Neutral class targets, 37 out of 37 Satisfied class targets, and 0 out of 9 Dissatisfied class targets. In the testing process, the Sigmoid model correctly predicted 4 out of 5 Neutral class targets, 12 out of 12 Satisfied class targets, and 0 out of 4 Dissatisfied class targets.

IV. DISCUSSION

We calculate the accuracy, precision, and recall values of each model using equations (1) to (3), which are tabulated in Table 4.

TABLE IV
MODEL PERFORMANCE

Model	Training			Testing		
	Acc	Prec	Rec	Acc	Prec	Rec
Polynomial	96.9	97.1	96.9	90.5	91.8	90.5
RBF	95.4	95.6	95.4	71.4	61.9	71.4
Sigmoid	86.2	76.8	86.2	76.2	64.7	76.2

According to the training performance results presented in Table 4, the Polynomial model achieved the highest values for accuracy (96.9%), precision (97.1%), and recall (96.9%), outperforming the RBF and Sigmoid models. Similarly, in the testing results, the Polynomial model also outperformed the others, achieving an accuracy of 90.5%, a precision of 91.8%, and a recall of 90.5%.

We also found that the Sigmoid model performed the worst in terms of performance, both in the training and testing results. In terms of training performance, this model generated the lowest values for accuracy (86.2%), precision (76.8%), and recall (86.2%). In the testing performance, it only managed to achieve an accuracy of 76.2%, a precision of 64.7%, and a recall of 76.2%.

After obtaining the performance metrics, we analyze and interpret the results to gain insights into the stability of the models. We calculate the performance metric differences between the training and testing values for each model, as shown in Table 5.

TABLE V
PERFORMANCE ANALYSIS

Model	|Δ(Training, Testing)|		
	Accuracy	**Precision**	**Recall**
Polynomial	6,4	5,3	6,4
RBF	24	33,7	24
Sigmoid	10	12,1	10

From Table 5, the Polynomial model produced the lowest training-testing performance difference values, with Δ accuracy = 6.4%, Δ precision = 5.3%, and Δ recall = 6.4%. Considering that all these values are below 7%, it can be concluded that the Polynomial model is the most stable compared to the RBF and Sigmoid models. This stability suggests that the Polynomial model is less likely to encounter overfitting or underfitting scenarios when new data is added in the future.

The results in Table 5 also show that the RBF model generates the highest training-testing performance difference values, with Δ accuracy = 24%, Δ precision = 33.7%, and Δ recall = 24%. All these values exceed 10%, indicating that this model is the most unstable. Therefore, the RBF model is the most likely to encounter overfitting or underfitting scenarios (should new data be added) among the three models, and it should not be considered as an alternative solution for the hospital patient satisfaction level classification problem.

After analyzing the performance of each model, we found that the Polynomial model outperforms the others in terms of performance metrics and stability. Therefore, we recommend using this model configuration to address the research problem.

V. CONCLUSIONS

Based on the questionnaire responses, we have concluded that patient satisfaction with hospital services is influenced by several factors, such as their experiences in the hospital, the attention provided by the medical staff, the clarity of medical information conveyed, service waiting time estimation, sufficient facilities, service procedure efficiency, hospital staff friendliness, involvement in treatment decision-making, and information regarding inpatient treatment and care. All of these factors have been shown to be valid for assessing patient responses in terms of recommending the hospital to others, as indicated by the t-test results. Since this study distributed questionnaires manually to patients/family members, not all the questionnaires that were distributed were returned (86 out of 100 questionnaires). Therefore, we recommend using more efficient methods, such as digital or online media. We also suggest increasing the sample size and/or the number of questions provided to cover a broader scope, with the expectation of obtaining more accurate results. From the various model variations using the Polynomial, RBF, and Sigmoid kernel functions, the classification using a 75:25 sampling ratio demonstrates that the Polynomial kernel function outperforms the other two kernels. With testing data accuracy of 90.5%, precision of 91.8%, and recall of 90.5%, this model exhibits excellent performance in classifying the questionnaire dataset. It is also worth noting that, based on the training-testing performance ratio, this model shows stable accuracy, precision, and recall values, with less than a 7% difference. With an approximate 6.4% difference in accuracy, a 5.3% difference in precision, and a 6.4% difference in recall, this model exhibits promising stability against overfitting and underfitting. Overall, this study presents an alternative solution for classifying a questionnaire-based dataset using machine learning algorithms, and we sincerely hope it will serve as a valuable reference for future studies in the same field.

REFERENCES

[1] I. Ilin, O. Iliyaschenko, and A. Konradi, "Business model for Smart Hospital health organization," *SHS Web Conf.*, vol. 44, p. 00041, 2018, doi: 10.1051/shsconf/20184400041.

[2] M. W. Lestari, A. Surjoputro, and A. Sriatmi, "Sustainable competitive advantage strategy in hospital industry to response policy change in Indonesia: A narrative review," *Public Heal. Indones.*, vol. 7, no. 3, pp. 95–100, 2021, doi: 10.36685/phi.v7i3.421.

[3] Y. Soumokil, M. Syafar, and A. Yusuf, "Analisis Kepuasan Pasien Di Rumah Sakit Umum Daerah Piru," *J. Ilm. Kesehat. Sandi Husada*, vol. 10, no. 2, pp. 543–551, 2021, doi: 10.35816/jiskh.v10i2.645.

[4] S. Jauhiainen, S. Aÿrämö, H. Forsman, and J. P. Kauppi, "Talent identification in soccer using a one-class support vector machine," *Int. J. Comput. Sci. Sport*, vol. 18, no. 3, pp. 125–136, 2019, doi: 10.2478/ijcss-2019-0021.

[5] T. Li, "Application of slack variable-optimized SVM on piano teaching reform," *J. Phys. Conf. Ser.*, vol. 1941, no. 1, p. 012084, Jun. 2021, doi: 10.1088/1742-6596/1941/1/012084.

[6] A. Larasati, A. Miftahul Hajji, and A. N. Handayani, "Preferences Analysis of Engineering Students on Choosing Learning Media using Support Vector Machine (SVM) Model," in *Proceedings of the 2nd International Conference on Vocational Education and Training (ICOVET 2018)*, Paris, France: Atlantis Press, 2019, pp. 57–59. doi: 10.2991/icovet-18.2019.15.

[7] J. Farjami, S. Dehyouri, and M. Mohamadi, "Evaluation of waste recycling of fruits based on Support Vector Machine (SVM)," *Cogent Environ. Sci.*, vol. 6, no. 1, p. 1712146, Jan. 2020, doi: 10.1080/23311843.2020.1712146.

[8] H. Nakao *et al.*, "Determination of individual factors associated with hallux valgus using SVM-RFE," *BMC Musculoskelet. Disord.*, vol. 24, no. 1, pp. 1–7, 2023, doi: 10.1186/s12891-023-06303-2.

[9] D. N. Avianty, P. I. G. P. S. Wijaya, and F. Bimantoro, "The Comparison of SVM and ANN Classifier for COVID-19 Prediction," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 13, no. 2, p. 128, 2022, doi: 10.24843/lkjiti.2022.v13.i02.p06.

[10] R. I. Perwira and H. Prapcoyo, "Software Defined Network : The Comparison of SVM kernel on DDoS Detection," *RSF Conf. Ser. Eng. Technol.*, vol. 1, no. 1, pp. 281–290, 2021, doi: 10.31098/cset.v1i1.413.

[11] M. F. Rohmah, I. K. G. D. Putra, R. S. Hartati, and L. Ardiantoro, "Comparison Four Kernels of SVR to Predict Consumer Price Index," *J. Phys. Conf. Ser.*, vol. 1737, no. 1, 2021, doi: 10.1088/1742-6596/1737/1/012018.

[12] K. N and R. V, "Ensemble of Multiple Kernel SVM Classifiers for Detection of Online Spam Reviews," *SSRN Electron. J.*, no. Icicnis, pp. 479–488, 2021, doi: 10.2139/ssrn.3769761.

[13] Istiadi and A. Y. Rahman, "Optimisasi Parameter Support Vector Machine Berbasis Algoritma Genetika Pada Klasifikasi Teks," *Conf. Innov. Appl. Sci. Technol.*, no. Desember 2020, pp. 481–488, 2020, [Online]. Available: http://publishing-widyagama.ac.id/ejournal-v2/index.php/ciastech/article/view/1904

[14] Q. H. Nguyen *et al.*, "Influence of data splitting on performance of machine learning models in prediction of shear strength of soil," *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/4832864.

[15] R. Dasmasela, B. P. Tomasouw, and Z. A. Leleury, “PENERAPAN METODE SUPPORT VECTOR MACHINE (SVM) UNTUK MENDETEKSI PENYALAHGUNAAN NARKOBA Application of Support Vector Machine (SVM) Method to Detect Drug Abuse,” *Param. J. Mat. Stat. dan Ter.*, vol. 1, no. 2, pp. 111–122, 2022.
[16] L. Efrizoni, S. Defit, M. Tajuddin, and A. Anggrawan, “Komparasi Ekstraksi Fitur dalam Klasifikasi Teks Multilabel Menggunakan Algoritma Machine Learning,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 653–666, 2022, doi: 10.30812/matrik.v21i3.1851.
[17] M. Azhari, Z. Situmorang, and R. Rosnelly, “Comparison of Classification Accuracy, Recall, and Precision on C4.5, Random Forest, SVM, and Naive Bayes Algorithms,” *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 2, p. 640, Apr. 2021, doi: 10.30865/mib.v5i2.2937.
[18] I. Nurul Hassanah, S. Faisal, A. Mutoi Siregar, U. Buana Perjuangan Karawang Jl HSRonggo Waluyo, T. Timur, and J. Barat, “Perbandingan Algoritma Support Vector Machine Dengan Decision Tree Pada Aplikasi Ruang Guru,” *Kumpul. J. Ilmu Komput.*, vol. 10, no. 1, pp. 39–50, 2023.
[19] A. Sreenidhi, B. Shruti, A. Divya, and N. Subhashini, “Detecting Phishing Websites Using Machine Learning,” *Smart Innov. Syst. Technol.*, vol. 286, pp. 353–362, 2022, doi: 10.1007/978-981-16-9873-6_32.
[20] N. W. E. Rosiana Dewi, I. G. Aris Gunadi, and G. Indrawan, “Detection of Class Regularity with Support Vector Machine methods,” *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 11, no. 1, p. 20, 2020, doi: 10.24843/lkjiti.2020.v11.i01.p03.
[21] D. Pardede, Wanayumini, and R. Rosnelly, “A Combination Of Support Vector Machine And Inception-V3 In Face-Based Gender Classification,” *Int. Conf. Inf. Sci. Technol. Innov.*, vol. 2, no. 1, pp. 34–39, Mar. 2023, doi: 10.35842/icostec.v2i1.30.
[22] A. F. Indriani and M. A. Muslim, “SVM Optimization Based on PSO and AdaBoost to Increasing Accuracy of CKD Diagnosis,” *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 10, no. 2, p. 119, 2019, doi: 10.24843/lkjiti.2019.v10.i02.p06.
[23] M. Handayani, R. Rosnelly, and Hartono, “Classification of Basurek Batik Using Pre-Trained VGG- 16 and Support Vector Machine,” in *2nd International Conference on Information Science and Technology Innovatin (ICoSTEC)*, 2023, pp. 3–7.
[24] I. S. Al-Mejibli, J. K. Alwan, and D. H. Abd, “The effect of gamma value on support vector machine performance with different kernels,” *Int. J. Electr. Comput. Eng.*, vol. 10, no. 5, pp. 5497–5506, 2020, doi: 10.11591/IJECE.V10I5.PP5497-5506.
[25] J. T. Samudra, R. Rosnelly, and Z. Situmorang, “Comparative Analysis of SVM and Perceptron Algorithms in Classification of Work Programs,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 22, no. 2, pp. 285–298, Mar. 2023, doi: 10.30812/matrik.v22i2.2479.
[26] N. A. Sinaga, R. Ramadani, K. Dalimunthe, M. S. A. A. Lubis, and R. Rosnelly, “Comparison of Decision Tree, KNN, and SVM Methods for Determining Majors in Vocational Schools,” *J. Sist. Komput. dan Inform.*, vol. 3, no. 2, p. 94, Dec. 2021, doi: 10.30865/json.v3i2.3598.