

PREDICTING LOAN ELIGIBILITY WITH SUPPORT VECTOR MACHINE: A MACHINE LEARNING APPROACH

Rajunaidi¹, Herman Yuliansyah^{2*}, Sunardi³, Murinto²

¹Magister Informatika, Universitas Ahmad Dahlan

²Informatika, Universitas Ahmad Dahlan

³Teknik Elektro, Universitas Ahmad Dahlan

email: *herman.yuliansyah@tif.uad.ac.id

Abstract: Non-performing loans remain one of the main challenges faced by cooperatives, particularly when the loan eligibility assessment process is still conducted manually. This traditional approach tends to be time consuming, subjective, and prone to inaccurate decisions. This study aims to develop a predictive model for borrower eligibility using the Support Vector Machine (SVM) algorithm as a more efficient and objective machine learning-based solution. A total of 1,000 loan history records were processed using RapidMiner software, taking into account variables such as salary, years of employment, loan amount, monthly installment, employment status, monthly expenses, number of dependents, housing status, age, and collateral value. The model's performance was evaluated using a confusion matrix and classification metrics including accuracy, precision, recall, and kappa. The results indicate that the SVM model achieved an accuracy of 90.05%, precision of 90.13%, recall of 90.05%, and f1 score of 90.08%, reflecting a strong performance in classifying borrower eligibility. The application of this method makes a significant contribution to the development of data driven decision support systems within cooperative environments. This finding expands the scientific understanding in the field of microfinance and supports the implementation of artificial intelligence technologies in making decisions that are more precise, rapid, and accurate.

Keywords: cooperative; eligibility prediction; machine learning; non-performing loan; SVM

Abstrak: Kredit macet merupakan salah satu permasalahan utama yang dihadapi koperasi, terutama ketika proses penilaian kelayakan peminjam masih dilakukan secara manual. Pendekatan ini cenderung lambat, subjektif, dan berisiko menghasilkan keputusan yang kurang akurat. Penelitian ini bertujuan untuk membangun model prediksi kelayakan peminjam menggunakan algoritma Support Vector Machine (SVM) sebagai solusi berbasis machine learning yang lebih efisien dan objektif. Sebanyak 1.000 data riwayat pinjaman diolah menggunakan tools RapidMiner dengan mempertimbangkan variabel: gaji, lama bekerja, besar pinjaman, angsuran per bulan, status pegawai, pengeluaran bulanan, jumlah tanggungan, status rumah, umur, dan nilai jaminan. Evaluasi model dilakukan menggunakan confusion matrix dan metrik klasifikasi seperti akurasi, presisi, recall, dan kappa. Hasil menunjukkan bahwa model SVM mencapai akurasi 90,05%, presisi 90,13%, recall 90,05%, dan f1 score 90,08%, yang mencerminkan performa model yang sangat baik dalam mengklasifikasikan kelayakan peminjam. Penerapan metode ini memberikan kontribusi penting dalam pengembangan sistem pendukung keputusan berbasis data di lingkungan koperasi. Temuan ini memperluas wawasan keilmuan di bidang keuangan mikro dan mendukung penerapan teknologi kecerdasan buatan dalam pengambilan keputusan yang lebih tepat, cepat, dan akurat.

Kata Kunci: koperasi; kredit macet; machine learning; prediksi kelayakan; SVM



INTRODUCTION

The increasing societal demands, such as educational expenses and business capital needs, have driven a rise in loan applications to cooperative institutions. Non-performing loans (NPL) have become the most significant issue within cooperatives[1]. Bangnano Cooperative, headquartered in Jakarta with a branch office in Yogyakarta, is a cooperative focused on improving the welfare of its members. However, the loan eligibility evaluation process at Bangnano Cooperative is still conducted manually, posing risks of non-performing loans and slowing down decision-making due to a lack of accuracy in the evaluation process[2]. This situation indicates an urgent need for a system that operates efficiently, objectively, and in a measurable manner.

Therefore, a more systematic, efficient, and technology-based approach is necessary[3]. One relevant approach is the application of machine learning methods to predict the eligibility of prospective loan recipients[4]. This technology is capable of processing large volumes of data and addressing complex problems that are difficult to solve using traditional methods [5][6]. The integration of machine learning into the loan evaluation process is expected to improve classification accuracy and speed, reduce human error, and accelerate more accurate decision-making [7][8].

The implementation of machine learning methods is seen as a promising solution to enhance the quality of eligibility predictions through the analysis of historical data and borrower behavior patterns. Previous studies have generally categorized loan eligibility into two classes: eligible and ineligible[9][10]. As an improvement, this study proposes three decision classes: Approved, Considera-

tion, Rejected. This aims to provide a more nuanced and practical evaluation outcome in assessing borrower eligibility.

In prior research, various machine learning algorithms were used with varying degrees of accuracy: SVM and Naïve Bayes for cooperative loans achieved an average accuracy of 90% [1]. Credit card application prediction using the C4.5 algorithm reached 83.33% [11]. Motorcycle loan eligibility with Naïve Bayes achieved 100% [12]. credit approval using K-Nearest Neighbors (KNN) achieved 98% [13]. Predictive vehicle maintenance using Naïve Bayes and Decision Tree resulted in 33% and 75% accuracy respectively [14]. Classification of Indonesia Smart Card recipients using SVM achieved 93% [15]. Credit approval prediction using SVM reached 85% [5]. Insurance customer prediction using SVM attained 88% [16]. Car loan eligibility using the SMART method achieved 65% [10]. Flood disaster prediction using DNIN achieved 94% [3]. organizational data clustering with three decision classes [17]. Negative content prediction on social media using SVM reached 97% accuracy [18]. Credit eligibility prediction using Random Forest achieved 83% accuracy [19]. Newton's Law of Universal Gravitation has been applied in various fields, including as a benchmark method for cold-start prediction, achieving an accuracy of 99.94% [20]. In machine translation, the Bilingual Evaluation Understudy (BLEU) metric has been utilized, with the highest average score of 0.601 on BLEU-1 and the lowest average score of 0.368 on BLEU-4 [21]. In the education sector, the C4.5 algorithm has been employed to predict student graduation outcomes, demonstrating an accuracy of 90% [22]. Furthermore, in link prediction tasks, methods such as KNN and Naive Bayes have been tested, with Ran-

dom Forest outperforming these methods across six datasets[23].

The purpose of this research is to develop a predictive model using the SVM method to assess loan feasibility and identify customers at risk of loan default. The purpose will evaluate the performance of the SVM model based on metrics such as training time, testing time, accuracy, precision, recall, and F1 score. The main objective is to improve the accuracy and efficiency of credit risk assessments and support data driven decision making in loan management.

METODE

This study utilizes 1,000 data records from members of the Bangnano Cooperative. Eligibility prediction is performed using the Support Vector Machine (SVM) algorithm with the RapidMiner tool. The stages of the research process are illustrated in Figure 1.

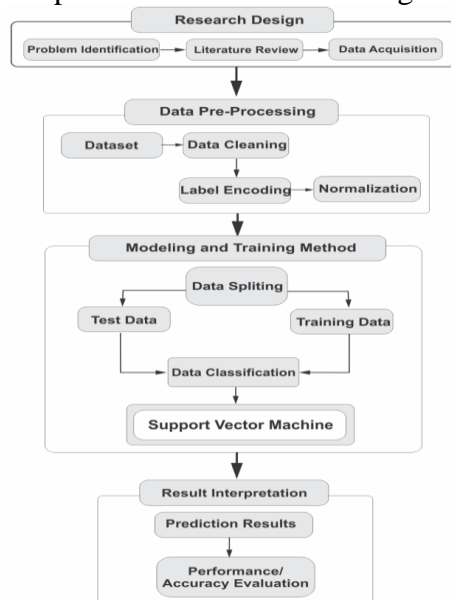


Figure 1. Research Flow and Methodology

This research method begins with a design phase that includes problem identification, literature review, and data

acquisition. Problem identification focuses on the classification of loan eligibility using machine learning algorithms. The literature review provides a theoretical foundation for classification methods, especially SVM, and insights from relevant studies. Data acquisition involves collecting a dataset with information on loan applicants, as shown in Table 1. The dataset contains 12 attributes: name, income, length of employment, loan amount, monthly installment, employment status, monthly expenses, number of dependents, housing status, age, collateral value, and eligibility. The eligibility attribute is the label, categorized into three classes: Approved, Consideration, and Rejected.

Once the data is collected, a pre-processing phase is carried out. This includes data cleaning to address duplication, missing values, or inconsistent formats, and label encoding to convert categorical data into numerical form so it can be processed by classification algorithms. Normalization is then applied to ensure that all numerical features fall within a uniform range, typically between 0 and 1.

After pre-processing, the dataset is split into two subsets through a data splitting process: training data and testing data, with a ratio of 80:20. The training data is used to build the classification model, while the testing data is used to evaluate the model's performance on unseen data. The classification algorithm used in this study is SVM, which is known for its ability to optimally classify data by maximizing the margin between classes[20]. The model is trained using the training dataset and then tested on the testing dataset to generate predictions for the eligibility label. The prediction results are then analyzed in the interpretation

Table 1. Cooperative Loan Applicant Data

No	Income	Length of Employment	Loan Amount	Monthly installment	Employment Status	Monthly Expenses	Number of Dependents	Housing Status	Age	Collateral Value	Eligibility
1.	10760000	14	15000000	1075000	Permanent	3800000	3	Owned	47	40300000	Approved
2.	8250000	7	15000000	1075000	Permanent	1300000	5	Owned	63	8600000	Consideration
3.	7950000	15	7000000	501666	Permanent	4900000	3	Owned	50	19100000	Consideration
....
1000	6850000	12	7000000	501666	Permanent	4800000	1	Rented	44	17800000	Rejected

Stage, which includes evaluating the model outcomes and assessing its performance using accuracy, fl score, precision, and recall metrics. This evaluation is essential to determine the model's effectiveness in performing the classification task comprehensively. The entire methodological process concludes with drawing conclusions based on the findings and model evaluation results.

This study applies the SVM algorithm with a Radial Basis Function (RBF) kernel as the classification method. The main objective of SVM is to construct an optimal hyperplane that maximally separates data classes in order to accurately predict the eligibility of loan recipients.

The basic formula of SVM can be expressed as shown in Equation (1):

$$f(x) = \text{sign}(w \cdot x + b) \quad (1)$$

Description:

x represents the input feature vector, w is the weight vector, b denotes the bias, and $f(x)$ indicates the predicted class output.

For data that cannot be separated linearly, a kernel function is used to map the data into a higher-dimensional space. In this study, the Radial Basis Function (RBF) kernel is applied, which is defined by Equation (2)[24]:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (2)$$

Description:

$K(x_i, x_j)$ represents the kernel value between data points x_i and x_j , γ (gamma)

is the kernel parameter that controls the model's complexity, and $\|x_i - x_j\|^2$ denotes the squared Euclidean distance between the two data points.

The parameter settings for SVM in this study follow the configuration provided by the LibSVM library, using the C-Support Vector Classification (C-SVC) type, which is suitable for multi-class classification tasks. The RBF kernel is chosen due to its effectiveness in handling non-linear data by mapping it into a higher-dimensional space. The gamma parameter is set to 0.1, which controls the influence of a single training data point. Lower gamma values result in a wider influence range. The regularization parameter C is set to 1.0, indicating a balance between maximizing the margin and minimizing classification errors.

The evaluation is conducted using a Confusion Matrix, with the formulas expressed in Equations (3)(4)(5) and (6) [24][25]:

$$\text{Accuracy} = (TP + TN) / (TP + FP + FN + TN) \quad (3)$$

$$\text{Precision} = (TP) / (TP + FP) \quad (4)$$

$$\text{Recall} = (TP) / (TP + FN) \quad (5)$$

$$F1 \text{ Score} = F1 = (2 \times TP) / (2 \times TP + FP + FN) \quad (6)$$

Description:

TP (True Positive): The number of correctly predicted positive instances.

TN (True Negative): The number of cor-

rectly predicted negative instances.

FP (False Positive): The number of negative instances incorrectly predicted as positive.

FN (False Negative): The number of positive instances incorrectly predicted as negative.

RESULT AND DISCUSSION

This research uses data from 1,000 cooperative members with 12 parameters. Loan eligibility predictions are made using the SVM algorithm implemented in RapidMiner. The results of the classification analysis performed with SVM are presented, showing the actual and predicted values for each category, along with their respective accuracy percentages, as shown in Figure 2.

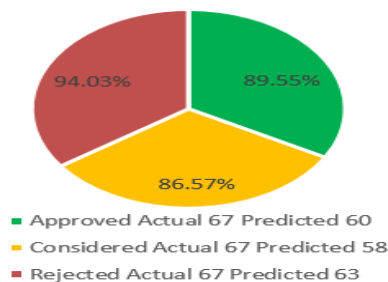


Figure 2. Actual and Predicted Accuracy

Based on Figure 2, the analysis of the SVM model shows accurate predictions for the three categories with varying accuracy levels. In the "Approved" category, the model successfully predicted 60 out of 67 data points, achieving an accuracy of 89.55%. In the "Considered" category, the model predicted 58 out of 67 data points with an accuracy of 86.57%. Meanwhile, in the "Rejected" category, the model achieved the highest accuracy of 94.03%, with 63 out of 67 data points correctly predicted. Overall, the SVM model demonstrated a total accuracy of 90.05%, proving its effectiveness in clas-

sifying data into the designated categories. The validation results for the loan eligibility prediction in Bangnano Cooperative are presented in Table 2.

Table 2. SVM Model Validation Results

Categories	Value	Approved	Consideration	Rejected
Accuracy	90,05%	89,55%	86,57%	94,03%
Precision	90,13%	92,31%	84,06%	94,03%
Recall	90,05%	89,55%	86,57%	94,03%
F1 Score	90,08%	90,09%	85,29%	94,03%

Based on Table 2, the SVM model's performance evaluation includes accuracy, confusion matrix, precision, recall, and F1-Score. The model achieved 90.05% accuracy, indicating strong alignment with actual labels. Precision was 90.13%, reflecting accurate positive predictions with low false positives, while Recall was 90.05%, showing effective detection of actual positive cases. The F1-Score of 90.08% confirms the model's balanced and stable performance.

These metrics demonstrate that the model is effective in both key aspects: the accuracy of positive predictions and the comprehensiveness of detection, making it suitable for applications where both types of errors have significant consequences. A detailed evaluation of the Confusion Matrix, showing the number of correct and incorrect predictions for each class, is presented in Figure 3.

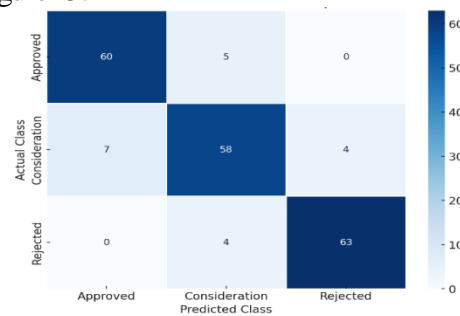


Figure 3. Confusion Matrix

The evaluation in Figure 3 presents the Confusion Matrix, highlighting correct and incorrect predictions for each

class: Approved, Considered, and Rejected. The model correctly classified 60 Approved instances, with 5 misclassified as Considered. For Considered, 58 were correct, but 7 were misclassified as Approved and 4 as Rejected. The Rejected class was perfectly classified with all 63 correct predictions. Overall, the SVM algorithm demonstrated effective and reliable performance in classifying loan eligibility, achieving near-optimal results in all categories.

CONCLUSION

This study was conducted with the objective of designing a predictive model to assess loan eligibility for recipients at Koperasi Bangnano using the SVM algorithm, which classifies data into three categories: Approved, Consideration, Rejected. The testing results showed that the model achieved an accuracy of 90.05%, indicating that SVM performs optimally in classifying eligibility based on the available variables. The implementation of this method provides a significant contribution to the development of data driven decision support systems within cooperative environments. These findings expand scientific insight in the field of microfinance and support the application of artificial intelligence technology in enabling more accurate, faster, and well-informed decision-making.

Future research is recommended to include additional variables such as credit history and type of employment to enhance classification accuracy. The SVM algorithm could also be further developed into a decision support application for cooperatives. Moreover, a comparison with other algorithms such as Random Forest or Naïve Bayes should be conducted to identify a more optimal

predictive model.

ACKNOWLEDGMENTS

This research was supported by the Direktorat Penelitian dan Pengabdian kepada Masyarakat (DPPM) Kementerian Pendidikan Tinggi, Sains dan Teknologi under Penelitian Tesis Magister (Master's Thesis Research) with grant number:126/C3/DT.05.00/PL/2025; and thanks to Universitas Ahmad Dahlan for their support.

BIBLIOGRAPHY

- [1] S. Riyadi, M. M. Siregar, K. fadhli F. Margolang, and K. Andriani, "Analysis of SVM and Naive Bayes Algorithm In Classification of Nad Loans in save and Loan Cooperatives," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 8, no. 3, pp. 261–270, Aug. 2022, doi: 10.33330/jurteks.v8i3.1483.
- [2] N. K. H. Aripuspa, N. M. Estiyanti, and E. M. Dharma, "Model Sistem Pendukung Keputusan Kelayakan Pemberian Kredit Pada BUMDes," *Progresif J. Ilm. Komput.*, vol. 18, no. 2, p. 231, Jul. 2022, doi: 10.35889/progresif.v18i2.925.
- [3] S. Sandiwarno, "Penerapan Machine Learning Untuk Prediksi Bencana Banjir," *J. Sist. Inf. Bisnis*, vol. 14, no. 1, pp. 62–76, Jan. 2024, doi: 10.21456/vol14iss1pp62-76.
- [4] A. Prawira, D. Arisandi, and T. Sutrisno, "Penerapan Algoritma Naive Bayes dan Multiple Linear Regression Untuk Prediksi Status dan Plafon Kredit (Studi Kasus:

- Bank ABC),” *J. Educ.*, vol. 5, no. 1, pp. 1075–1087, Dec. 2022, doi: 10.31004/joe.v5i1.720.
- [5] I. M. D. P. Asana and N. P. D. T. Yanti, “Sistem Klasifikasi Pengajuan Kredit Dengan Metode Support Vector Machine (SVM) I Made Dwi Putra Asana,” *J. Sist. Cerdas*, vol. 6, no. 2, pp. 123–133, 2023.
- [6] H. Yuliansyah, S. Sulistyawati, T. W. Sukesu, S. A. Mulasari, and W. N. S. Wan Ali, “Artificial intelligence in malnutrition research: a bibliometric analysis,” *Bull. Soc. Informatics Theory Appl.*, vol. 7, no. 1, pp. 32–42, Jul. 2023, doi: 10.31763/businta.v7i1.605.
- [7] A. Bagja, K. Kusri, and M. R. Arief, “Komparasi Algoritma Naïve Bayes Dan SVM Untuk Klasifikasi Kelayakan Pemberian Pinjaman,” *Infotek J. Inform. dan Teknol.*, vol. 6, no. 2, pp. 513–523, Jul. 2023, doi: 10.29408/jit.v6i2.20059.
- [8] E. Alaros, M. Marjani, D. A. Shafiq, and D. Asirvatham, “Predicting Consumption Intention of Consumer Relationship Management Users Using Deep Learning Techniques: A Review,” *Indones. J. Sci. Technol.*, vol. 8, no. 2, pp. 307–328, Dec. 2022, doi: 10.17509/ijost.v8i2.55814.
- [9] N. Windy Mardiyyah, N. Rahaningsih, and I. Ali, “Penerapan Data Mining Menggunakan Algoritma K-Nearest Neighbor Pada Prediksi Pemberian Kredit Di Sektor Finansial,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1491–1499, Apr. 2024, doi: 10.36040/jati.v8i2.9010.
- [10] D. W. Sinaga, A. Z. Syah, and E. Saputra, “Implementasi Simple Multi Attribute Rating Technique untuk Menentukan Kelayakan Pengajuan Kredit Mobil,” *Edumatic J. Pendidik. Inform.*, vol. 8, no. 1, pp. 203–211, Jun. 2024, doi: 10.29408/edumatic.v8i1.25687.
- [11] M. R. Wirasena and J. Warman, “Penerapan Algoritma C4.5 Untuk Prediksi Kelayakan Pengajuan Kartu Kredit Visa Bagi Nasabah,” *TeknoIS J. Ilm. Teknol. Inf. dan Sains*, vol. 14, no. 2, pp. 296–302, Jul. 2024, doi: 10.36350/jbs.v14i2.266.
- [12] R. Halila Nasution and S. Humaira, “Implementasi Algoritma Naïve Bayes Untuk Memprediksi Kelayakan Pemberian Kredit Sepeda Motor,” *J. Sci. Soc. Res.*, vol. 4307, no. 2, pp. 546–554, 2024, doi: <https://doi.org/10.54314/jssr.v7i2.1872>.
- [13] R. Wahyusari, D. P. Putro, P. S. Informatika, S. Tinggi, and T. Ronggolawe, “Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Pemberian Kredit Application of the K-Nearest Neighbor Algorithm for Determining Credit,” pp. 1–5, 2024.
- [14] Islamia Nuraini, Arnowo Hari Wibowo, Muhammad Asrol, Taufik, and Dyah Lang Wilis, “Analisis Pemeliharaan Prediktif Kendaraan Operasional Menggunakan Algoritma Naïve Bayes Dan Decision Tree,” *J. Publicuho*, vol. 7, no. 2, pp. 591–603, May 2024, doi: 10.35817/publicuho.v7i2.412.
- [15] A. Nata and S. Suparmadi, “Analisis Sistem Pendukung Keputusan Dengan Model Klasifikasi Berbasis Machine Learning Dalam Penentuan Penerima Program Indonesia Pintar,” *J. Sci. Soc. Res.*, vol. 5, no. 3, p. 697, Oct. 2022, doi: 10.54314/jssr.v5i3.1041.

- [16] R. Pratama, M. I. Herdiansyah, D. Syamsuar, and A. Syazili, "Prediksi Customer Retention Perusahaan Asuransi Menggunakan Machine Learning," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, pp. 96–104, Mar. 2023, doi: 10.32736/sisfokom.v12i1.1507.
- [17] Y. Mulyana, A. Fadlil, and I. Riadi, "Classification of Organizational Data Using the K-Means Algorithm," *Komputasi J. Ilm. Ilmu Komput. dan Mat.*, vol. 19, no. 2, pp. 123–137, Jul. 2022, doi: 10.33751/komputasi.v19i2.5676.
- [18] H. Syahputra and A. Wibowo, "Comparison of Support Vector Machine (SVM) and Random Forest Algorithm for Detection of Negative Content on Websites," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 1, pp. 165–173, Mar. 2023, doi: 10.26555/jiteki.v9i1.25861.
- [19] B. Prasojo and E. Haryatmi, "Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest," *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 2, pp. 79–89, Sep. 2021, doi: 10.25077/TEKNOSI.v7i2.2021.79-89.
- [20] H. Yuliansyah, Z. A. Othman, and A. A. Bakar, "A new link prediction method to alleviate the cold-start problem based on extending common neighbor and degree centrality," *Phys. A Stat. Mech. its Appl.*, vol. 616, p. 128546, Apr. 2023, doi: 10.1016/j.physa.2023.128546.
- [21] B. O. S. Miranda, H. Yuliansyah, and M. K. Biddinika, "Machine Translation Indonesian Bengkulu Malay Using Neural Machine Translation-LSTM," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 18, no. 3, Jul. 2024, doi: 10.22146/ijccs.98384.
- [22] H. Yuliansyah, R. A. P. Imaniati, A. Wirasto, and M. Wibowo, "Predicting Students Graduate on Time Using C4.5 Algorithm," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 7, no. 1, p. 67, Apr. 2021, doi: 10.20473/jisebi.7.1.67-73.
- [23] S. A. Koni'ah and H. Yuliansyah, "Classification Algorithm for Link Prediction Based on Generated Features of Local Similarity-Based Method," *SISTEMASI*, vol. 11, no. 2, p. 317, May 2022, doi: 10.32520/stmsi.v11i2.1641.
- [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, p. 5497, Oct. 2020, doi: 10.11591/ijece.v10i5.pp5497-5506.
- [25] S. Fide, S. Suparti, and S. Sudarno, "Analisis Sentimen Ulasan Aplikasi Tiktok Di Google Play Menggunakan Metode Support Vector Machine (Svm) Dan Asosiasi," *J. Gaussian*, vol. 10, no. 3, pp. 346–358, Dec. 2021, doi: 10.14710/j.gauss.v10i3.32786.