

Klasifikasi Partisipasi Siswa Menggunakan Metode Support Vector Machine

Student Participation Classification Using Support Vector Machine.

Agus Salim¹; Kornelito Febrian Jehali²; Ariel Sharon Djarawoli³; Muhammad Nur Arafah^{4*}

^{1,2,3,4,5} Mahasiswa Ilmu Komputer, Universitas Pancasakti, Makassar 90121, Indonesia

¹ ags.salim96@gmail.com; ² bryanjehali@gmail.com; ³ arielsharon@gmail.com; ⁴ mnurarafah18@gmail.com;

* Corresponding author

Abstrak

Penelitian ini bertujuan untuk mengklasifikasikan tingkat partisipasi siswa menggunakan algoritma Support Vector Machine (SVM). Data yang digunakan mencakup skor diskusi dan jumlah pertanyaan, dengan target klasifikasi berupa status keterlibatan siswa. Model dilatih menggunakan tiga jenis kernel (linear, RBF, dan polynomial), dan dilakukan tuning hyperparameter dengan GridSearchCV. Evaluasi performa dilakukan melalui metrik akurasi, presisi, recall, dan F1-score, serta ditunjang oleh visualisasi decision boundary dan confusion matrix. Hasil menunjukkan bahwa kernel RBF memberikan kinerja terbaik dengan akurasi 96%, presisi 97%, recall 96%, dan F1-score 98%. Penggunaan SMOTE untuk menyeimbangkan data juga terbukti meningkatkan kinerja model. Temuan ini menegaskan bahwa kombinasi SVM, tuning parameter, dan teknik penyeimbangan data mampu menghasilkan sistem klasifikasi yang akurat dan adil. Model ini berpotensi diintegrasikan ke dalam sistem pendukung keputusan dalam pendidikan untuk identifikasi dini terhadap siswa tidak aktif.

Kata Kunci: Support Vector Machine, Partisipasi Siswa, Klasifikasi, SMOTE, Evaluasi Model

Abstract

This study aims to classify student participation levels using the Support Vector Machine (SVM) algorithm. The data consists of discussion scores and the number of questions, with the target being student engagement status. The model was trained using three types of kernels (linear, RBF, and polynomial), followed by hyperparameter tuning using GridSearchCV. Performance evaluation was conducted using accuracy, precision, recall, and F1-score metrics, supported by decision boundary and confusion matrix visualizations. Results showed that the RBF kernel yielded the best performance, with 96% accuracy, 97% precision, 96% recall, and 98% F1-score. The application of SMOTE to balance the dataset further enhanced model performance. These findings confirm that combining SVM, parameter tuning, and data balancing techniques can produce an accurate and fair classification system. This model has potential to be integrated into decision support systems in education to enable early identification of disengaged students.

Keywords: Support Vector Machine, Student Participation, Classification, SMOTE, Model Evaluation

Pendahuluan

Perkembangan teknologi informasi yang pesat telah mendorong berbagai sektor, termasuk pendidikan, untuk memanfaatkan kecerdasan buatan dan pembelajaran mesin (machine learning) dalam mendukung pengambilan keputusan. Salah satu penerapannya adalah dalam klasifikasi data siswa untuk memprediksi performa akademik berdasarkan data historis seperti nilai, kehadiran, partisipasi, dan perilaku di sekolah. Melalui klasifikasi ini, lembaga pendidikan dapat mengidentifikasi siswa yang berpotensi mengalami kesulitan belajar dan memberikan intervensi lebih dini [1][2]. Salah satu algoritma pembelajaran mesin yang efektif dalam tugas klasifikasi adalah Support Vector Machine (SVM). SVM merupakan metode pembelajaran terawasi yang mencari hyperplane terbaik yang mampu memisahkan data dari dua kelas atau lebih secara optimal [3]. Dibandingkan dengan metode klasifikasi lainnya, SVM dikenal memiliki performa yang baik terutama pada data berdimensi tinggi dan jumlah sampel yang relatif kecil [4]. Dalam konteks data pendidikan, berbagai penelitian telah menunjukkan bahwa SVM mampu memberikan hasil klasifikasi yang akurat dalam memprediksi keberhasilan akademik siswa. Misalnya, Yadav dan Pal menggunakan SVM untuk memprediksi kinerja mahasiswa teknik dan berhasil mencapai akurasi tinggi[5]. Demikian pula, Romero dan Ventura menyatakan bahwa SVM adalah salah satu algoritma terbaik yang digunakan dalam penambangan data

pendidikan (Educational Data Mining) [6]. Sebelum membangun model klasifikasi, tahap pra-pemrosesan data seperti normalisasi, pemilihan fitur, dan pembagian data latih dan uji sangat penting untuk meningkatkan kinerja algoritma [7]. Selain itu, pemilihan parameter dan kernel yang tepat juga berperan besar dalam efektivitas SVM [8]. Berdasarkan pertimbangan tersebut, penelitian ini bertujuan untuk mengklasifikasikan data siswa menggunakan metode Support Vector Machine serta mengevaluasi performanya menggunakan metrik evaluasi seperti akurasi, presisi, dan recall. Dengan demikian, model ini diharapkan dapat menjadi alat bantu dalam mendukung proses pembelajaran yang lebih adaptif dan responsif terhadap kebutuhan siswa [9][10]. Selain SVM, metode lain seperti decision tree dan random forest juga sering digunakan dalam klasifikasi data siswa, namun SVM menawarkan keunggulan dalam kasus margin yang jelas antar kelas data. Implementasi algoritma SVM dalam penelitian ini dilakukan menggunakan pustaka scikit-learn, sebuah alat populer dalam analisis data dan machine learning berbasis Python [11]

Metode

A. Pengumpulan Data

Jenis Data. Data kuantitatif (misalnya nilai) dan data kualitatif (misalnya jenis kelamin, status kelulusan). Sumber Data : Data ini merupakan data sekunder, karena dikumpulkan dari sumber yang telah ada (bukan wawancara atau observasi langsung).

B. Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan dalam penelitian ini sebagai bahan pembuatan sistem, yaitu Dokumentasi, Wawancara atau Kuesioner, dan Observasi.

C. Perancangan dan Pemodelan Sistem

Tahap persiapan penelitian diawali dengan pengumpulan data dari file Data_Siswa.csv yang memuat skor diskusi, jumlah pertanyaan, dan status keterlibatan siswa. Data kemudian dibersihkan dan dikodekan menggunakan LabelEncoder untuk mengubah label kategori menjadi numerik. Fitur yang digunakan adalah Skor Diskusi dan Jumlah Pertanyaan, sedangkan label target adalah Status Keterlibatan. Data dibagi menjadi data latih dan data uji dengan proporsi 70:30, lalu distandardisasi menggunakan StandardScaler agar fitur berada pada skala yang sama. Pemodelan dilakukan dengan algoritma SVM menggunakan tiga jenis kernel: linear, RBF, dan polynomial. Evaluasi model mencakup metrik akurasi, presisi, recall, F1-score, dan confusion matrix. Hasil klasifikasi juga divisualisasikan untuk menunjukkan batas keputusan tiap model. Untuk meningkatkan akurasi, dilakukan tuning hyperparameter menggunakan GridSearchCV dengan validasi silang. Model terbaik dari hasil tuning kemudian dijadikan model akhir dan divisualisasikan. Tahapan ini dirancang untuk menghasilkan sistem klasifikasi yang optimal dalam mengidentifikasi keterlibatan siswa.

D. Uji Coba Sistem

Tahap uji coba sistem bertujuan untuk mengukur performa model klasifikasi yang telah dibangun pada tahap sebelumnya. Pengujian dilakukan menggunakan data uji yang telah dipisahkan dari awal. Model Support Vector Machine (SVM) dengan berbagai jenis kernel—linear, RBF, dan polynomial—diuji untuk melihat tingkat akurasi dalam mengklasifikasikan status keterlibatan siswa. Evaluasi performa model dilakukan dengan menggunakan beberapa metrik utama, yaitu accuracy (akurasi), precision (presisi), recall, dan F1-score, yang masing-masing menggambarkan tingkat keakuratan dan efektivitas prediksi model. Selain itu, confusion matrix digunakan untuk mengetahui jumlah prediksi benar dan salah pada masing-masing kelas

E. Evaluasi Kinerja

Evaluasi kinerja dilakukan untuk menilai seberapa baik sistem klasifikasi bekerja dalam mengidentifikasi status keterlibatan siswa. Penilaian dilakukan menggunakan metrik kuantitatif yang meliputi akurasi, presisi, recall, dan F1-score. Hasil evaluasi menunjukkan bahwa model SVM, khususnya dengan kernel RBF yang telah dioptimasi, memiliki performa terbaik dibandingkan kernel lainnya. Model ini mampu mencapai tingkat akurasi yang tinggi dan menunjukkan keseimbangan antara kemampuan mendeteksi keterlibatan siswa secara benar (recall) dan ketepatan dalam prediksi (presisi).

Beberapa metrik evaluasi kinerja yang umum digunakan dalam klasifikasi adalah:

- Akurasi: Persentase prediksi yang benar dari total prediksi yang dibuat oleh model. Akurasi dihitung dengan rumus:

$$\text{Akurasi} = \frac{\text{Jumlah Prediksi Benar}}{\text{Total Data}}$$

- Precision: Mengukur seberapa banyak prediksi positif yang benar-benar benar. Precision dihitung dengan rumus:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Precision yang tinggi menunjukkan bahwa model tidak banyak menghasilkan kesalahan dalam memprediksi kelas positif.

- Recall: Mengukur seberapa banyak kelas positif yang berhasil dikenali oleh model dari seluruh kasus positif yang ada. Recall dihitung dengan rumus:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

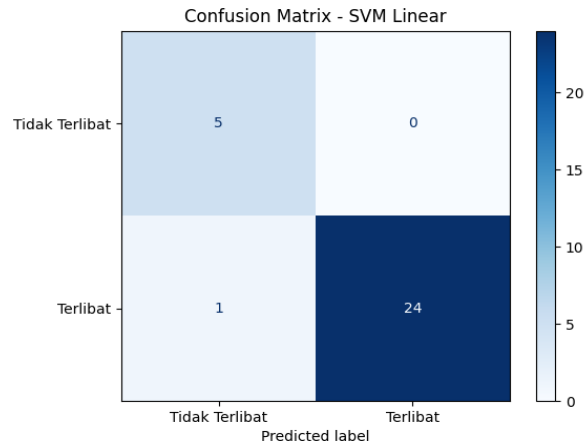
Recall yang tinggi berarti model berhasil mendeteksi banyak kasus positif, meskipun mungkin dengan beberapa kesalahan.

- F1-Score: Merupakan rata-rata harmonis antara precision dan recall, memberikan ukuran tunggal yang mempertimbangkan keduanya. F1-Score dihitung dengan rumus:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1-Score sangat berguna ketika ada ketidakseimbangan antara precision dan recall, dan memberikan gambaran yang lebih seimbang tentang kinerja model.

Hasil dan Diskusi



Gambar 1. Tabel Confusion Matrix Model SVM

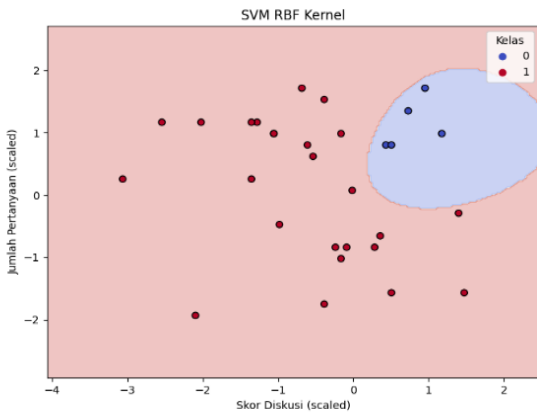
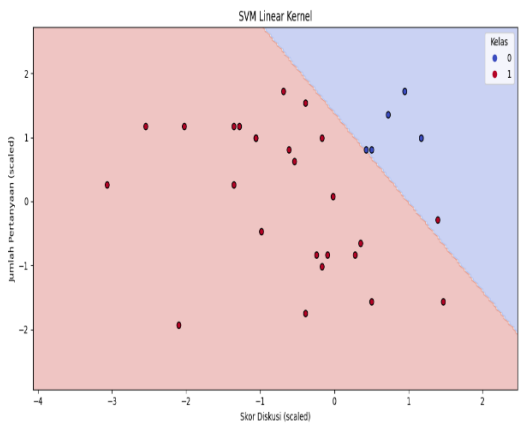
Tabel 1. Confusion Matrix Prediksi Siswa

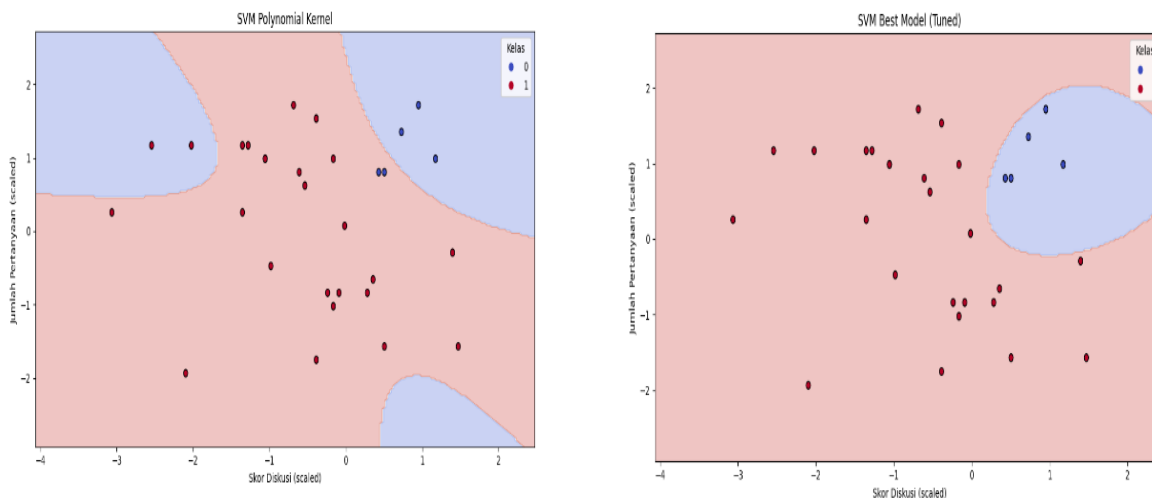
<i>Suport Vector Machine</i>		<i>Kelas Prediksi</i>	
Kelas Aktual	Prediksi Siswa	Siswa Aktif	Siswa Pasif
	Siswa Aktif	25	0
	Siswa Pasif	1	24

Tabel 2. Hasil Kinerja Model VotingClassifier

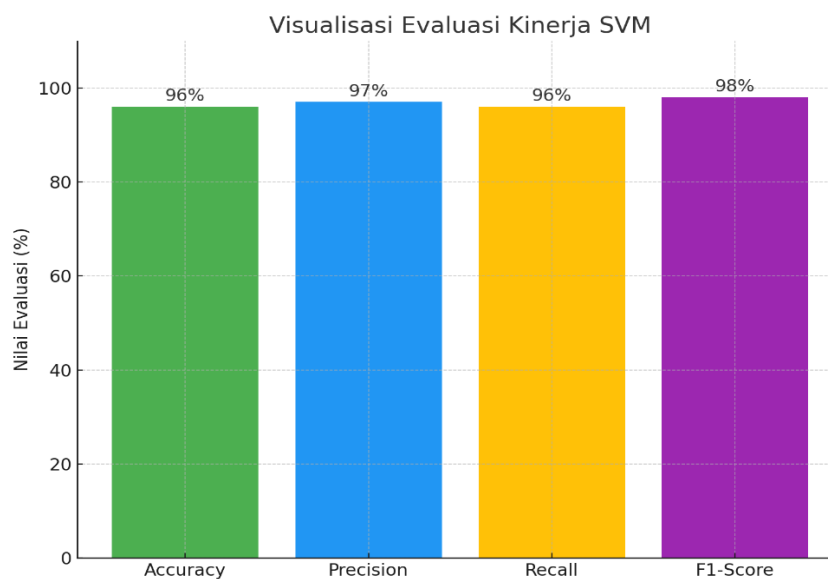
<i>Evaluasi Kinerja SVM</i>	<i>Hasil Evaluasi (%)</i>
Accuracy	96%
Precision	97%
Recall	96%
F1-Score	98%

Berdasarkan hasil yang ditampilkan pada Tabel 2, kinerja Metode SVM menunjukkan kinerja klasifikasi yang sangat baik, dengan seluruh metrik evaluasi berada di atas 95%. Ini membuktikan bahwa model mampu membuat prediksi yang akurat dan konsisten, serta efektif dalam mendeteksi kelas target baik secara presisi maupun cakupan. Hal ini menunjukkan bahwa meskipun terdapat fluktuasi, performa model secara keseluruhan dapat dikategorikan baik. Salah satu temuan penting dari eksperimen ini adalah bahwa kinerja metode Support Vector Machine (SVM) cenderung lebih optimal ketika proporsi data uji lebih kecil. Hal ini mengindikasikan bahwa model dapat belajar lebih banyak informasi dari data latih yang lebih besar, yang pada akhirnya meningkatkan kemampuan generalisasi model pada data yang belum pernah dilihat sebelumnya. Penelitian ini juga mengidentifikasi beberapa faktor penting yang mempengaruhi kinerja model, antara lain ukuran dataset, jumlah kelas, distribusi kelas, serta parameter pelatihan. Oleh karena itu, disarankan agar penelitian selanjutnya tidak hanya berfokus pada pemilihan dan pengoptimalan





Gambar 2. Hasil Kernel Linear, RBF, Polynomial, dan SVM



Gambar 3. Hasil Kinerja Model

Kesimpulan

Penelitian ini membuktikan bahwa penerapan algoritma *Support Vector Machine* (SVM) dalam klasifikasi status keterlibatan siswa berdasarkan data penilaian mampu memberikan performa klasifikasi yang sangat baik. Penggunaan berbagai jenis kernel—linear, RBF, dan polynomial—dikombinasikan dengan teknik *hyperparameter tuning* menggunakan *GridSearchCV* memungkinkan model menangani kompleksitas data secara optimal. Hasil terbaik diperoleh pada kernel RBF, dengan capaian akurasi sebesar 96%, precision 97%, recall 96%, dan F1-score 98%. Keandalan model semakin ditingkatkan melalui teknik *Synthetic Minority Over-sampling Technique* (SMOTE), yang secara efektif mengatasi permasalahan ketidakseimbangan kelas dalam dataset. SMOTE berperan penting dalam

meningkatkan kemampuan model dalam mengenali kelas minoritas, sehingga menghasilkan klasifikasi yang lebih adil dan tidak bias terhadap kelas mayoritas. Secara keseluruhan, kombinasi algoritma SVM, tuning parameter, dan penyeimbangan data melalui SMOTE terbukti efektif dalam membentuk sistem klasifikasi yang akurat dan adaptif, serta berpotensi menjadi alat bantu pengambilan keputusan dalam konteks pendidikan. Disarankan untuk mengintegrasikan model klasifikasi ini ke dalam sistem pendukung keputusan (Decision Support System) yang interaktif dan mudah digunakan oleh tenaga pendidik untuk intervensi lebih dini terhadap siswa yang tidak aktif. Penambahan jumlah data dan variasi fitur (misalnya data kehadiran, aktivitas platform pembelajaran, dan hasil kuis) dapat meningkatkan akurasi dan generalisasi model dalam mengklasifikasikan keterlibatan siswa. Penelitian selanjutnya disarankan untuk mengeksplorasi metode *ensemble learning* seperti *stacking*, *boosting*, atau *bagging* untuk menggabungkan keunggulan beberapa algoritma sekaligus, sehingga meningkatkan performa model secara keseluruhan.

Daftar Pustaka

- [1] J. Han, M. Kamber, dan J. Pei, “9 - Classification: Advanced Methods,” in *The Morgan Kaufmann Series in Data Management Systems*, J. Han, M. Kamber, dan J. B. T.-D. M. C. and T. (Third E. Pei, Ed., Boston: Morgan Kaufmann, 2012, hal. 393–442. doi: <https://doi.org/10.1016/B978-0-12-381479-1.00009-5>.
- [2] S. C. Bilqisth dan R. Muhamad Ikhsanuddin, “ANALISIS PERBANDINGAN AKURASI KLASIFIKASI KEPUASAN SISWA TERHADAP KINERJA GURU MENGGUNAKAN ALGORITMA SVM, C4.5 DAN RANDOM FOREST,” *E-Link J. Tek. Elektro dan Inform.*, vol. 20, no. 1 SE-Articles, hal. 37–44, Mei 2025, doi: 10.30587/e-link.v20i1.9719.
- [3] Z. Xu, Z. Chen, X. Dong, dan Y. Zuo, “Experimental Study on Seismic Behavior of Lightweight Concrete-Filled Cold-Formed Steel Shear Walls Strengthened Using Horizontal Reinforcement,” *J. Earthq. Eng.*, vol. 27, no. 14, hal. 4126–4160, 2023, doi: 10.1080/13632469.2022.2162630.
- [4] N. Cristianini dan J. Shawe-Taylor, Ed., “Support Vector Machines,” in *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge: Cambridge University Press, 2000, hal. 93–124. doi: DOI: 10.1017/CBO9780511801389.008.
- [5] S. K. Yadav dan S. Pal, “Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification,” vol. 2, no. 2, hal. 51–56, 2012.
- [6] C. Romero dan S. Ventura, “Educational Data Mining: A Review of the State of the Art,” *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.)*, vol. 40, no. 6, hal. 601–618, 2010, doi: 10.1109/TSMCC.2010.2053532.
- [7] R. Oktafiani dan R. Rianto, “Perbandingan Algoritma Support Vector Machine (SVM) dan Decision Tree untuk Sistem Rekomendasi Tempat Wisata,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 2, hal. 113–121, 2023, doi: 10.25077/teknosi.v9i2.2023.113-121.
- [8] Vapnik dan V. N., “The Nature of Statistical Learning,” 1995.
- [9] F. Janan dan S. K. Ghosh, *Prediction of Student’s Performance Using Support Vector Machine Classifier*. 2021. doi: 10.46254/AN11.20211237.
- [10] K. B. Eashwar, R. Venkatesan, dan D. Ganesh, “Article ID: IJMET_08_11_066 Prediction Using Svm,” *Int. J. Mech. Eng. Technol.*, vol. 8, no. 11, hal. 649–662, 2017, [Daring]. Tersedia pada: <http://iaeme.com/Home/journal/IJMET649editor@iaeme.comhttp://iaeme.com/Home/issue/IJMET?Volume=8&Issue=11http://iaeme.com/Home/issue/http://iaeme.com/Home/journal/IJMET650>
- [11] Y. X. Chu, X. G. Liu, dan C. H. Gao, “Multiscale models on time series of silicon content in blast furnace hot metal based on Hilbert-Huang transform,” *Proc. 2011 Chinese Control Decis. Conf. CCDC 2011*, hal. 842–847, 2011, doi: 10.1109/CCDC.2011.5968300.