PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DALAM MENENTUKAN PELAJARAN YANG DIMINATI SISWA DI SDN 016 UJUNGBATU

Aurora Attila Naufal[1], Riri Okra[2], Sarwo Derta[3], Hari Antoni Musril [4]

[1,2,3,4]*Universitas Islam Negeri Sjech M.Djamil Djambek Bukittinggi*
E-mail : [1]auro124an@gmail.com, [2]ririokra@uinbukittinggi.ac.id,
[3]sarwoderta@uinbukittinggi.ac.id, [4]hariantonimusril@uinbukittinggi.ac.id

ABSTRAK

Penentuan minat belajar siswa di sekolah dasar selama ini masih dilakukan secara subjektif melalui tanya jawab dan tes tertulis, sehingga hasilnya kurang akurat dan tidak berbasis data. Penelitian ini bertujuan menerapkan algoritma Support Vector Machine (SVM) untuk mengklasifikasikan pelajaran yang paling diminati siswa kelas VI SDN 016 Ujungbatu berdasarkan data akademik. Data yang digunakan meliputi nilai UTS, nilai harian, nilai UAS, kehadiran, dan partisipasi siswa dari 9 mata pelajaran (18 fitur). Dari 60 data siswa, 40 data digunakan sebagai data latih dan 20 data sebagai data uji. Model SVM multi-kelas diimplementasikan dengan pendekatan One-vs-All menggunakan fungsi fitcecoc pada software MATLAB. Hasil pengujian menunjukkan akurasi model pada data latih sebesar 100%, sedangkan pada data uji diperoleh akurasi sebesar 33,33% dengan presisi makro 15,90%, recall makro 23,30%, dan F1-score 18,70%. Mata pelajaran Ilmu Pengetahuan Alam (IPA) teridentifikasi sebagai pelajaran yang paling diminati siswa. Keterbatasan akurasi pada data uji terutama disebabkan oleh ketidakseimbangan distribusi kelas dan jumlah data yang relatif kecil. Penelitian ini memberikan fondasi bagi pengembangan sistem rekomendasi pelajaran berbasis data di sekolah dasar.

Kata Kunci : *Support Vector Machine, Minat Belajar, Klasifikasi, Data Mining, Sekolah Dasar*

ABSTRACT

Determining student learning interests in elementary schools has traditionally been carried out subjectively through question-and-answer sessions and written tests, resulting in inaccurate and non-data-driven outcomes. This study aims to apply the Support Vector Machine (SVM) algorithm to classify the most preferred subjects of sixth-grade students at SDN 016 Ujungbatu based on academic data. The data used includes mid-term scores, daily grades, final exam scores, attendance, and student participation across 9 subjects (18 features). From 60 student records, 40 were used for training and 20 for testing. The multi-class SVM model was implemented using a One-vs-All approach via the fitcecoc function in MATLAB software. The results showed a training accuracy of 100%, while the testing accuracy was 33.33% with a macro precision of 15.90%, macro recall of 23.30%, and F1-score of 18.70%. Natural Science (IPA) was identified as the most preferred subject. The limited testing accuracy is primarily attributed to class distribution imbalance and relatively small dataset size. This study provides a foundation for developing data-driven subject recommendation systems in elementary schools.

Keywords : *Support Vector Machine, Learning Interest, Classification, Data Mining, Elementary School*

PENDAHULUAN

Pendidikan dasar memegang peran krusial dalam membentuk fondasi pengetahuan dan karakter peserta didik. Pada jenjang ini, pengenalan mata pelajaran yang sesuai dengan minat dan bakat siswa menjadi faktor penting dalam meningkatkan motivasi belajar dan pencapaian akademis. Namun, dalam praktiknya, identifikasi minat siswa masih sering dilakukan secara konvensional dan subjektif, sehingga hasilnya kurang dapat diandalkan.

Minat belajar merupakan daya tarik intrinsik yang mendorong individu untuk mencari dan menggali pengetahuan. Siswa yang memiliki minat tinggi pada suatu mata pelajaran cenderung lebih terlibat dalam pembelajaran, lebih antusias menyelesaikan tugas, dan lebih tekun dalam mengeksplorasi materi [1]. Sebaliknya, rendahnya minat belajar sering ditandai dengan perilaku pasif, mudah bosan, dan kurang termotivasi dalam proses pembelajaran [2].

Karakteristik siswa sekolah dasar yang masih berada pada tahap perkembangan emosi dan sosial menyebabkan minat belajar siswa cenderung belum stabil dan mudah berubah, serta sangat dipengaruhi oleh lingkungan sekolah dan teman sebaya. Kondisi ini tidak jarang berdampak pada kenyamanan belajar siswa lain, termasuk adanya perilaku bullying yang dapat menurunkan rasa percaya diri dan minat terhadap mata pelajaran tertentu.

Berdasarkan hasil wawancara dengan guru di SDN 016 Ujungbatu, diketahui bahwa penentuan minat siswa selama ini dilakukan melalui tanya jawab dan tes tertulis. Cara ini dinilai masih subjektif karena siswa sering memilih mata pelajaran berdasarkan pengaruh teman atau sekadar ikut-ikutan, bukan berdasarkan minat yang sesungguhnya. Akibatnya, potensi siswa tidak tergali secara optimal, dan proses pembelajaran belum berjalan sesuai kebutuhan masing-masing siswa.

Kemajuan teknologi informasi membuka peluang untuk mengatasi permasalahan tersebut melalui pendekatan berbasis data. Salah satu teknik yang dapat digunakan adalah algoritma machine learning, khususnya Support Vector Machine (SVM). SVM merupakan metode supervised learning yang efektif untuk masalah klasifikasi, bekerja dengan mencari hyperplane optimal yang memisahkan data ke dalam kelas-kelas berbeda dengan margin terbesar [3]. Algoritma ini memiliki keunggulan dalam menangani data berdimensi tinggi, kemampuan generalisasi yang baik, dan ketahanan terhadap overfitting [4].

Penelitian terkait menunjukkan bahwa SVM mampu memberikan akurasi tinggi dalam berbagai konteks pendidikan. Sulistilawati (2024) menerapkan SVM untuk memprediksi minat dan nilai belajar siswa berdasarkan data historis akademik. Bachtiar (2020) menggunakan SVM untuk membantu mahasiswa menentukan peminatan berdasarkan prestasi dua tahun pertama dengan hasil memuaskan. Penelitian lain di SMK Adiluhur berhasil memprediksi kelulusan siswa dengan akurasi 95,06% [5]. Namun, penerapan SVM di tingkat sekolah dasar untuk klasifikasi minat pelajaran masih sangat terbatas.

Penelitian ini bertujuan menerapkan algoritma SVM untuk mengklasifikasikan mata pelajaran yang paling diminati siswa kelas VI SDN 016 Ujungbatu berdasarkan data akademik yang mencakup nilai ujian, nilai harian, kehadiran, dan partisipasi. Implementasi dilakukan menggunakan software MATLAB dengan Statistics and Machine Learning Toolbox, dan evaluasi model menggunakan confusion matrix serta metrik akurasi, presisi, recall, dan F1-score.

METODE

Jenis dan Sumber Data

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan teknik data mining berbasis supervised learning. Data diperoleh dari SDN 016 Ujungbatu, Kecamatan Ujungbatu, Kabupaten Rokan Hulu, Provinsi Riau, pada bulan Juni–Juli 2025. Subjek penelitian adalah siswa kelas VI sebanyak 60 siswa. Atribut data yang digunakan mencakup nilai Ujian Tengah Semester (UTS) dan nilai harian (NH) dari 9 mata pelajaran: PAI, PKN, Bahasa Indonesia, IPA, IPS, Matematika, SBDP, PJOK, dan Bahasa Melayu Riau (BMR), sehingga menghasilkan 18 fitur numerik per siswa. Label target adalah mata pelajaran yang paling diminati dari masing-masing siswa.

Tahapan Penelitian

Tahapan penelitian meliputi: (1) identifikasi masalah, (2) studi literatur, (3) pengumpulan data, (4) pra-pemrosesan data, (5) penentuan atribut, (6) penerapan algoritma SVM, dan (7) evaluasi model menggunakan confusion matrix.

Pra-Pemrosesan Data

Pra-pemrosesan data meliputi tiga tahap. Pertama, pembersihan data (data cleaning) untuk menghapus nilai kosong dan atribut tidak relevan seperti nomor urut dan nama siswa[6]. Kedua, transformasi data untuk memastikan seluruh nilai dalam format numerik sesuai kebutuhan algoritma SVM. Ketiga, normalisasi fitur agar seluruh atribut berada dalam skala yang seragam. Data kemudian dibagi menjadi data latih (training) sebanyak 40 data (70%) dan data uji (testing) sebanyak 20 data (30%).

Algoritma Support Vector Machine (SVM)

SVM merupakan metode klasifikasi yang bekerja berdasarkan prinsip Structural Risk Minimization (SRM), dengan tujuan menemukan hyperplane optimal yang memisahkan data dari kelas berbeda dengan margin maksimum [7]. Untuk masalah klasifikasi biner, SVM mencari fungsi keputusan:

$$f(x) = w \cdot x + b$$

di mana w adalah vektor bobot dan b adalah bias. Data yang paling dekat dengan hyperplane disebut support vector. Karena terdapat 9 kelas mata pelajaran, digunakan pendekatan One-vs-All (OvA) melalui fungsi fitcecoc pada MATLAB dengan kernel linear[8].

Implementasi pada MATLAB R2024b

Implementasi dilakukan secara bertahap pada MATLAB R2024b menggunakan Statistics and Machine Learning Toolbox. Kode program disusun dalam satu skrip yang mencakup tahap pembacaan data, pra-pemrosesan, pembagian data, pelatihan model, prediksi, dan evaluasi.

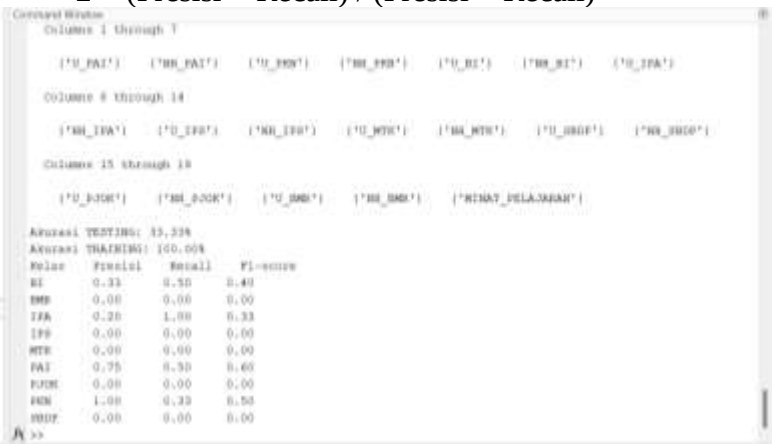


Gambar 1. Implementasi MATLAB R2024b

Evaluasi Model

Evaluasi model dilakukan menggunakan confusion matrix yang menghasilkan nilai True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Dari confusion matrix dihitung empat metrik evaluasi[9]:

Akurasi = (TP + TN) / (TP + TN + FP + FN)
Presisi = TP / (TP + FP)
Recall = TP / (TP + FN)
F1-Score = 2 × (Presisi × Recall) / (Presisi + Recall)



Gambar 2. Hasil Convusion Matrix MATLAB R2024b

HASIL PENELITIAN DAN PEMBAHASAN

Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari 60 record siswa kelas VI SDN 016 Ujungbatu dengan 18 fitur numerik yang mencakup nilai UTS dan nilai harian dari 9 mata pelajaran. Label kelas terdiri dari 9 kategori minat pelajaran. Setelah melalui tahap pra-pemrosesan, seluruh data siap digunakan dalam format numerik tanpa nilai kosong.

Distribusi kelas pada dataset menunjukkan ketidakseimbangan yang signifikan. PAI merupakan kelas dengan jumlah data tertinggi (10 siswa; 16,67%), diikuti IPA (14 siswa; 23,33%) dan PKN (5 siswa; 8,33%). Sebaliknya, beberapa kelas lain seperti MTK dan BMR memiliki jumlah data yang lebih sedikit — MTK hanya 4 siswa (6,67%) dan SBDP yang paling sedikit dengan 2 siswa (3,33%). Ketidakseimbangan kelas ini menjadi salah satu faktor yang memengaruhi performa model dalam penelitian ini.

Tabel 1. Distribusi Jumlah Siswa per Kategori Minat Pelajaran

	Mata Pelajaran	Jumlah Siswa	Persentase
1	BI	7	11,67%
2	BMR	6	10,00%
3	IPA	14	23,33%
4	IPS	8	13,33%
5	MTK	4	6,67%
6	PAI	10	16,67%
7	PJOK	4	6,67%

8	PKN	5	8,33%
9	SBDP	2	3,33%
	Total	60	100%

Hasil Perhitungan Manual (Sampel 4 Data)

Untuk memberikan gambaran konseptual algoritma SVM, dilakukan perhitungan manual menggunakan 4 sampel data dengan 3 kelas (MTK, IPA, PAI). Dengan pendekatan OvA, dibentuk 3 model biner. Sebagai ilustrasi, Model 1 (MTK vs lainnya) menghasilkan vektor bobot w = [−10, 0, 13] dan bias b = −229,5, sehingga fungsi klasifikasinya:

$$f(x) = -10x_1 + 0x_2 + 13x_3 - 229{,}5$$

Hasil prediksi dari 4 data sampel menunjukkan 3 data berhasil diklasifikasikan dengan benar (akurasi 75%). Satu kesalahan terjadi pada kelas PAI yang diprediksi sebagai IPA. Tabel 2 menyajikan confusion matrix dari sampel perhitungan manual.

Tabel 2. Confusion Matrix Perhitungan Manual (4 Sampel Data)

Aktual \ Prediksi	MTK	IPA	PAI
MTK	1	0	0
IPA	0	2	0
PAI	0	1	0

Hasil Implementasi MATLAB (60 Data)

Model SVM multi-kelas diimplementasikan menggunakan perangkat lunak MATLAB R2024b dengan memanfaatkan seluruh 60 data siswa. Data dibagi menjadi 40 data latih dan 20 data uji dengan perbandingan 70:30[10]. Fungsi fitcecoc digunakan untuk membangun model OvA secara otomatis.

Verifikasi nama kolom menunjukkan bahwa dataset terdiri dari 19 kolom (18 fitur numerik + 1 label kelas). Setelah skrip dijalankan, Command Window menampilkan hasil akurasi sebagai berikut:

- Akurasi training: 100,00%
- Akurasi testing: 33,33%

Selain akurasi, dilakukan pula perhitungan metrik evaluasi lainnya menggunakan data uji. Tabel 3 merangkum seluruh metrik evaluasi model SVM pada penelitian ini.

Tabel 3. Ringkasan Metrik Evaluasi Model SVM

Metrik Evaluasi	Nilai
Akurasi Data Latih	100,00%
Akurasi Data Uji	33,33%
Presisi (Makro)	15,90%
Recall (Makro)	23,30%
F1-Score (Makro)	18,70%

Analisis dan Pembahasan

Berdasarkan tabulasi output MATLAB, IPA teridentifikasi sebagai pelajaran yang paling diminati dengan 14 siswa (23,33%), diikuti PAI dengan 10 siswa (16,67%), BI dengan 7 siswa (11,67%), dan IPS dengan 8 siswa (13,33%). Mata pelajaran dengan peminat paling sedikit adalah SBDP (2 siswa, 3,33%), PJOK, dan MTK (masing-masing 4 siswa, 6,67%). Temuan ini konsisten dengan penelitian sebelumnya yang menyebutkan bahwa IPA dan Matematika merupakan mata pelajaran favorit di sekolah dasar ketika disajikan dengan metode pembelajaran yang menarik [8].

Model SVM mencapai akurasi 100% pada data latih, menunjukkan bahwa model berhasil mempelajari pola dari data pelatihan secara sempurna. Namun, akurasi data uji hanya 33,33%, mengindikasikan terjadinya overfitting. Kondisi ini terutama disebabkan oleh dua faktor: (1) ketidakseimbangan distribusi kelas — kelas IPA dan PAI memiliki jauh lebih banyak data dibanding SBDP (2 data) sehingga model cenderung bias; dan (2) jumlah data yang terbatas (60 record) untuk 9 kelas multi-label mengakibatkan setiap kelas hanya memiliki sedikit sampel pada data uji.

Nilai F1-score makro sebesar 18,70% menunjukkan bahwa model kesulitan mengklasifikasikan kelas-kelas minoritas secara merata. Hal ini merupakan keterbatasan umum SVM standar pada dataset tidak seimbang. Perbandingan dengan penelitian terdahulu yang mencapai akurasi lebih tinggi (94–98%) dapat dijelaskan oleh perbedaan jumlah kelas: penelitian tersebut umumnya menangani klasifikasi biner atau tiga kelas dengan lebih banyak data per kelas, sedangkan penelitian ini mengklasifikasikan 9 kelas sekaligus.

Meskipun akurasi uji belum optimal, penelitian ini berhasil membuktikan bahwa SVM dapat mengidentifikasi pola minat dari data akademik siswa SD, dan IPA terbukti menjadi pelajaran yang paling diminati di SDN 016 Ujungbatu. Informasi ini dapat dimanfaatkan oleh guru dan pihak sekolah untuk merancang program pembelajaran yang lebih sesuai dengan minat siswa.

KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Support Vector Machine (SVM) multi-kelas dengan pendekatan One-vs-All pada MATLAB R2024b untuk mengklasifikasikan mata pelajaran yang paling diminati siswa kelas VI SDN 016 Ujungbatu. Menggunakan 18 fitur dari nilai UTS dan nilai harian 9 mata pelajaran dengan pembagian 70:30, model mengidentifikasi IPA sebagai pelajaran yang paling banyak diminati (14 siswa, 23,33%), diikuti PAI (10 siswa) dan IPS (8 siswa).

Model mencapai akurasi 100% pada data latih dan 33,33% pada data uji, dengan presisi makro 15,90%, recall 23,30%, dan F1-score 18,70%. Kesenjangan akurasi mengindikasikan overfitting akibat ketidakseimbangan kelas dan jumlah data terbatas. Untuk penelitian lanjutan disarankan: (1) menambah jumlah data, (2) menerapkan teknik penyeimbangan kelas seperti SMOTE, (3) membandingkan dengan algoritma lain seperti Random Forest atau K-Nearest Neighbor, (4) mengeksplorasi kernel RBF, serta (5) mengembangkan antarmuka berbasis web atau mobile.

UCAPAN TERIMAKASIH

Terima kasih penulis ucapkan yang setulus-tulusnya untuk semangat yang diberikan oleh orang tua, dosen pembimbing yang telah berjasa dalam pembuatan skripsi ini.

DAFTAR PUSTAKA

[1] S. Perancangan, J. Komputer, M. Cisco, and P. Tracer, “Jurnal Ilmu Komputer Revolusioner,” vol. 8, no. 1, pp. 1–10, 2024.

[2] R. Salfina, H. A. Musril, R. Okra, and S. Derta, “Perancangan Game Edukasi untuk Mata Pelajaran Bahasa Jepang Kelas X Menggunakan Aplikasi Educandy di SMKN 4

Payakumbuh," *J. Inf. Syst. Educ. Dev.*, vol. 1, no. 1, pp. 1–7, 2023, [Online]. Available: http://journal.mwsfoundation.or.id/index.php/jised/article/view/1

[3] Y. G. Afnila, K. Khairuddin, H. A. Musril, and R. Okra, "Pengaruh Media Pembelajaran Interaktif Terhadap Hasil Belajar Siswa Dalam Pembelajaran Bimbingan Teknologi Informasi dan Komunikasi Kelas X di SMAN 1 Tigo Nagari," *Indones. Res. J. Educ.*, vol. 3, no. 1, pp. 544–551, 2022, doi: 10.31004/irje.v3i1.321.

[4] D. I. Pushpita Anna Octaviani, Yuciana Wilandari, "Penerapan Metode SVM Pada Data Akreditasi Sekolah Dasar Di Kabupaten Magelang," *J. Gaussian*, vol. 3, no. 8, pp. 811–820, 2014.

[5] E. Haryatmi and S. Pramita Hervianti, "Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan MahasiswaTepat Waktu," vol. 1, no. 10, pp. 386–392, 2021.

[6] N. Leli, F. Rakhmawati, and R. Widyasari, "Penerapan Metode Support Vector Machine (Svm) Untuk Klasifikasi Uang Kuliah Tunggal Di Universitas Islam Negeri Sumatera Utara," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 4, no. 2, pp. 903–915, 2023, doi: 10.46306/lb.v4i2.354.

[7] M. Rafly and I. Veritawati, "Penggunaan Algoritma Support Vector Machine Untuk Penentuan Rekomendasi Penerima Beasiswa Di SMA Negeri 8 Kota Bogor," vol. 9, no. 1, pp. 102–113, 2025.

[8] B. Pada and S. M. A. Ulun, "Penerapan metode forecasting dengan algoritma support vector manchine untuk memprediksi penerimaan peserta didik baru pada sma ulun nuha," vol. 7, pp. 931–937, 2024, doi: 10.37600/tekinkom.v7i2.1525.

[9] A. Pratama, R. C. Wihandika, and D. E. Ratnawati, "Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa," vol. 2, no. 4, pp. 1704–1708, 2018.

[10] I. O. Muraina, "Ideal Dataset Splitting Ratios in Machine Learning Algorithms: General Concerns for Data Scientists and Data Analysts," *7th Int. MARDIN ARTUKLU Sci. Res. Conf.*, no. February, pp. 496–504, 2022, [Online]. Available: https://www.researchgate.net/publication/358284895

.