

KLASIFIKASI PRESTASI AKADEMIK PESERTA DIDIK DENGAN METODE *MACHINE LEARNING* DI SMP X

Fandi Chriswantoro Putro¹, Ahmad Chusyairi^{2*}, Cian Ramadhona Hassolthine³

Universitas Siber Asia, Jakarta Selatan, Jakarta, Indonesia¹

Email: fni.chriz27@gmail.com

Universitas Siber Asia, Jakarta Selatan, Jakarta, Indonesia²

Email*: ahmadchusyairi@lecturer.unsia.ac.id

Universitas Siber Asia, Jakarta Selatan, Jakarta, Indonesia³

Email: cianhassolthine@lecturer.unsia.ac.id

(*) *Corresponding Author*

ABSTRAK

Machine learning (ML) salah satu bidang ilmu yang berfokus pada perancangan dan pengembangan model algoritma untuk menciptakan perilaku berdasarkan data yang tersedia. Prestasi akademik adalah metrik yang digunakan untuk penilaian lembaga pendidikan berkualitas. Dengan menggunakan data akademik peserta didik SMP X dan algoritma dari *machine learning* klasifikasi seperti *Random Forest*, *Naïve Bayes*, *k-Nearest Neighbours* (k-NN), dan *Support Vector Machine* (SVM), sehingga penelitian ini dapat melakukan klasifikasi prestasi akademik peserta didik di SMP X dengan optimal dilihat dari perbandingan tingkat akurasi yang paling baik diantara algoritma klasifikasi. Keakuratan suatu algoritma adalah ukuran seberapa tepat algoritma tersebut mengklasifikasikan suatu sampel. Hasil evaluasi yang dibandingkan berupa akurasi validasi dan standar deviasi. Perbandingan dilakukan untuk menentukan algoritma terbaik berdasarkan akurasi dan stabilitas. Hasil penelitian menunjukkan algoritma SVM memiliki akurasi validasi tertinggi dengan nilai 0.987410 yang menunjukkan kinerja terbaik dalam memprediksi kelas dan nilai deviasi standar terendah yaitu 0.005132 yang menunjukkan kinerja yang lebih stabil dan konsisten, dibandingkan dengan algoritma lainnya. Hal ini mengindikasikan bahwa SVM unggul dalam memprediksi kelas yang benar dengan kinerja stabil. Berdasarkan dari hasil dan analisa disimpulkan bahwa algoritma SVM yang terpilih digunakan untuk mengembangkan model klasifikasi prestasi akademik peserta didik dalam bentuk program *python* yang masih sederhana namun memiliki akurasi yang tinggi, stabil dan konsisten. Program ini telah menjadi alat bantu bagi pihak SMP X dalam mengidentifikasi prestasi akademik peserta didik dan sebagai bahan pelaporan hasil belajar peserta didik kepada orang tua.

Kata kunci: Klasifikasi, *machine learning*, peserta didik, prestasi akademik

ABSTRACT

Machine learning (ML) is a field of science that focuses on designing and developing algorithmic models to create behavior based on available data. Academic achievement is a metric used for the assessment of quality educational institutions. By using academic data of students in SMP X and machine learning classification algorithms such as *Random Forest*, *Naïve Bayes*, *k-Nearest Neighbors* (k-NN), and *Support Vector Machine* (SVM), so that this research can classify the academic achievement of students in SMP X optimally seen from the comparison of the best accuracy rate among classification

algorithms. The accuracy of an algorithm is a measure of how precisely it classifies a sample. Evaluation results are compared in the form of validation accuracy and standard deviation. The comparison is done to determine the best algorithm based on accuracy and stability. The results showed that the SVM algorithm has the highest validation accuracy with a value of 0.987410 which shows the best performance in predicting classes and the lowest standard deviation value of 0.005132 which shows a more stable and consistent performance, compared to other algorithms. This indicates that SVM excels in predicting the correct class with stable performance. Based on the results and analysis, it is concluded that the selected SVM algorithm is used to develop a classification model of students' academic achievement in the form of a python program that is still simple but has high accuracy, stable and consistent. This program has become a tool for SMP X in identifying students' academic achievement and as a material for reporting students' learning outcomes to parents.

Keywords: Academic achievement, classification, machine learning, learners

1. PENDAHULUAN

Kemajuan teknologi berdampak pada kehidupan manusia, karena kehadiran teknologi menawarkan banyak keuntungan dan kemudahan di berbagai sektor, terutama dalam sektor pendidikan [1]. Sektor pendidikan memiliki peran penting dalam pengembangan SDM, mulai dari sekolah tingkat dasar hingga ke tingkat sekolah tinggi, Sekolah Menengah Pertama (SMP) X juga salah satunya. SMP X memiliki total peserta didik sebanyak 656, termasuk kelas 7, 8 dan 9 pada tahun ajaran 2023/2024. Setiap tahun ajaran memiliki nilai akademik yang digunakan sebagai pelaporan hasil belajar akhir semester ke orang tua yang biasanya masih dalam bentuk pengolahan manual menggunakan rumus *excel* sederhana, dan pengolahan nilai yang tidak terkontrol dimana pengolah nilai hanya mendapatkan nilai mentah saja tiap semesternya yang mengakibatkan tambahan beban kerja dalam menentukan ketuntasan nilai peserta didik. Data nilai ini akan selalu bertambah setiap waktu, sehingga perlu adanya pengelolaan nilai yang efektif, khususnya dalam mengklasifikasi ketuntasan belajar atau prestasi akademik peserta didik.

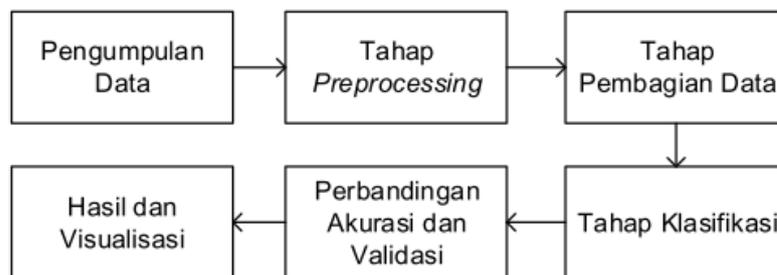
Berdasarkan permasalahan tersebut, SMP X memerlukan alat alternatif sebagai alat bantu mengelola nilai agar dapat melakukan klasifikasi nilai dengan banyaknya data yang tiap waktunya akan selalu bertambah dan beragam, *machine learning* sangat sesuai untuk klasifikasi prestasi akademik peserta didik, karena kemampuan dalam menganalisa banyak data yang kompleks secara efisien, selain itu *machine learning* dapat mengidentifikasi pola dan tren yang mungkin tidak teridentifikasi dengan cara manual sehingga dapat meningkatkan akurasi dalam penilaian prestasi peserta didik.

Machine learning (ML) salah satu bidang ilmu yang berfokus pada perancangan dan pengembangan model algoritma untuk menciptakan perilaku berdasarkan data yang tersedia [2]. Prestasi akademik adalah metrik yang digunakan untuk penilaian lembaga pendidikan berkualitas [3]. Dengan menggunakan data akademik peserta didik dan algoritma dari *machine learning* seperti *Random Forest*, *Naïve Bayes*, *k-Nearest Neighbours* (k-NN), dan *Support Vector Machine* (SVM), sehingga penelitian ini dapat melakukan klasifikasi prestasi akademik peserta didik di SMP X dengan optimal dilihat dari perbandingan tingkat akurasi yang paling baik diantara algoritma klasifikasi. Keakuratan suatu algoritma adalah ukuran seberapa tepat algoritma tersebut mengklasifikasikan suatu sampel [4].

2. METODE

Klasifikasi termasuk ke dalam *supervised method* dalam *machine learning* dimana model dibangun menggunakan data latih yang terdiri dari objek data kelasnya [5]. Beberapa model yang terdapat dalam klasifikasi antara lain; *Naïve Bayes* adalah klasifikasi dengan metode yang mengandalkan skor probabilitas dan statistik dengan memprediksi peluang yang akan muncul berdasarkan data pengalaman sebelum-sebelumnya, dan *naïve bayes* juga dapat bekerja dengan cepat pada *dataset* yang besar [6]. *k- NN* adalah teknik klasifikasi data berdasarkan pada klasifikasi data pada *data training* yang memiliki kedekatan dengan *data testing*, dengan sejumlah *k* data yang terdekat, sehingga menghasilkan klasifikasi yang baik berdasarkan pada kedekatan data [5]. SVM merupakan algoritma klasifikasi yang biasanya digunakan untuk analisis *sentiment*, algoritma ini akan membuat sebuah garis pemisah atau disebut *hyperplane* yang berperan dalam memisahkan kelas dengan label *positif* dan kelas dengan label *negative*, sehingga sangat cocok untuk *dataset* yang kompleks [7]. *Random Forest* merupakan gabungan dari berbagai cabang *decision tree* yang berkualitas, yang kemudian digabungkan menjadi satu model. Model ini memiliki ketergantungan pada nilai vektor acak dengan distribusi yang seragam di semua cabang, di mana masing-masing cabang memiliki kedalaman maksimum [8]. Dengan membandingkan keempat algoritma, akan menghasilkan keandalan dan validitas yang paling sesuai berdasarkan *dataset* penelitian.

Tahapan penelitian yang dilakukan melalui beberapa tahap, [9] yaitu:



Gambar 1. Diagram tahapan penelitian

Pengumpulan Data

Tahap ini dimaksudkan untuk memperoleh data, fakta, dan informasi yang relevan [10]. Dalam penelitian ini, pengumpulan data (*dataset*) dilakukan dengan menggunakan *data primer* yang dikumpulkan penulis secara langsung dari sumber nyata/aslinya. Data tersebut secara khusus dikumpulkan selama tahun ajaran 2023/2024 pada semester 1 dan 2 semua jenjang dan rombel kelas di SMP X, dengan total sebanyak 2.621 baris dan 18 kolom.

Tahap Preprocessing

Pra-pemrosesan mengacu pada tugas persiapan data, seperti transformasi, reduksi, imputasi, dan penyeimbangan data [11]. Data akan dipahami dan dinormalisasi untuk tahap analisis. Data nilai peserta didik, memiliki data yang terdiri dari 12 kolom nilai, tetapi hanya 11 nilai yang tersedia di setiap jenjangnya, karena satu mata pelajaran hanya ditempuh secara bergantian yaitu antara 'sby', 'tik' dan 'pra', maka nilai kosong tersebut diberikan *value* yaitu *mean* (rata-rata) dari data nilai, sehingga tidak ada nilai *null* atau NaN pada data nilai. Pada kolom 'jk', 'kls', dan 'rom' masih bertipe data kategorikal, sehingga perlu dikonversi menjadi *numerik* (int64) menggunakan '*LabelEncoder*' pada tiap kolom 'jk', 'kls', dan 'rom'. Kemudian menentukan data target (variabel dependen)

adalah kolom 'status' dan data fitur (variabel independen) adalah kolom 'jk', 'kls', 'rom', 'sem', 'ind', 'agm', 'mtk', 'pkn', 'ing', 'ips', 'ipa', 'pjk', 'bjw', 'sby', 'tik', dan 'pra'.

Tahap Pembagian Data

Data dibagi menjadi dua bagian yaitu 80% digunakan untuk *data training* dan 20% untuk *data testing* [12]. *Data training* adalah data yang diproses menggunakan model yang akan diterapkan, dan hasilnya akan digunakan untuk melakukan prediksi atau klasifikasi pada data pengujian. Sementara itu, *data testing* adalah data yang akan diuji dan diprediksi atau diklasifikasikan [9]. *Data Training* sejumlah 2.096 data dan sisanya sebagai *data testing* sejumlah 525 data. Selain itu dilakukan proses penentuan nilai k untuk algoritma k -NN, nilai k harus ditentukan untuk mengetahui berapa banyak jumlah tetangga terdekat ($n_neighbors$) yang paling baik digunakan. 'GridSearchCV' untuk menguji nilai k terbaik antara 1-20. Pada tahap ini didapatkan nilai k terbaik adalah 19, sehingga untuk nilai k pada proses k -NN menggunakan nilai $k = 19$.

Tahap Klasifikasi

Dari data training yang telah dibagi, data akan diuji menggunakan algoritma machine learning untuk proses klasifikasi, ada 4 (empat) algoritma, yaitu:

Random Forest

Algoritma ini digunakan dalam pohon keputusan (*decision tree*). Algoritma ini menciptakan *decision tree* yang memiliki bagian-bagian yang bernama *root node*, *internal node*, dan *leaf node*, dimana dengan cara mengacak pemilihan atribut serta data sesuai dengan ketentuan yang telah ditetapkan. *Root node* berperan dalam mengumpulkan informasi, sementara *node internal* memuat pertanyaan terkait data informasi. *Leaf node* melakukan penyelesaian masalah dan membuat keputusan. Proses ini dimulai dengan menghitung nilai *entropy* untuk menetapkan tingkat ketidakmurnian atribut serta nilai penghasilan informasi. Atribut yang memiliki penghasilan informasi tertinggi akan diangkat sebagai *root node* [13].

$$Entropy(S) = \sum_{i=1}^n -p_i \log_2 p_i \quad (1)$$

Keterangan:

S = Himpunan kasus

n = Jumlah anggota dari S

p_i = proporsi dari S_i dan S

$$Information\ Gain(Y, X) = E(Y) - E(Y|X) \quad (2)$$

Keterangan:

$E(Y)$ = *entropi* dari atribut target

$E(Y|X)$ = *entropi* rata-rata dari atribut prediktor terhadap target

X = atribut prediktor

Y = atribut target

Naïve Bayes

Metode klasifikasi *machine learning* ini bergantung pada nilai probabilitas dan statistik. Metode ini memprediksi kemungkinan di masa depan berdasarkan data pengalaman yang telah ada sebelumnya, dan lebih dikenal dengan sebutan *Teorema Bayes* [6]. Persamaan *Teorema* [13] dapat dilihat pada persamaan 3.

$$P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad (3)$$

Keterangan:

- $P(C_i|X)$ = probabilitas dari kelas C berdasarkan kondisi X
- $P(X|C_i)$ = probabilitas dari X berdasarkan kondisi kelas C
- $P(C_i)$ = probabilitas kelas C (probabilitas *prior*)
- $P(X)$ = probabilitas data X
- i = kelas
- X = data kelas tidak diketahui
- C = kelas spesifik data X

***k*-Nearest Neighbours (*k*-NN)**

k-NN merupakan metode klasifikasi yang baik dengan kedekatan yang dekat dengan *data mining* berdasarkan suatu objek dan istilah-istilahnya. Metode ini melakukan pengujian dan perbandingan antara *data training* dan *data testing* sehingga dapat mempermudah penyelesaian suatu permasalahan dan memiliki akurasi yang tinggi [14]. Data yang telah dihitung jarak kedekatannya (disebut *k*) kemudian diurutkan dari yang terdekat hingga yang terjauh, dan sejumlah nilai *k* diambil dari data tersebut. Untuk menghitung jarak, dapat digunakan rumus jarak *Euclidean Distance* [15].

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

Keterangan:

- $d(x, y)$ = Jarak *Euclidean Distance*
- x_i = *data testing*
- y_i = *data training*
- i = variable data
- n = dimensi data

pada penelitian ini parameter nilai *k* sesuai hasil optimalisasi *GridSearchCV* dalam menentukan banyaknya tetangga terdekat yaitu 19.

Support Vector Machine (SVM)

SVM bekerja dengan mengidentifikasi *hyperplane* optimal yang memisahkan beberapa kelas dengan mengukur jarak *hyperplane* dan mencari titik maksimum. Jarak atau *margin* adalah ruang antara *hyperplane* dan data terdekat dari setiap kelas, di mana data terdekat ini dikenal sebagai *support vector* [13]. *Kernel SVM Gaussian* adalah jenis kernel yang digunakan dalam SVM untuk menangani masalah yang tidak dapat diselesaikan secara *linier*. Fungsi *Kernel SVM Gaussian* dapat dinyatakan dalam persamaan berikut.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (5)$$

dengan x_i dan x_j merupakan vektor dari data set, γ adalah parameter untuk mengendalikan proses *learning*, dan \exp merupakan logaritma alami.

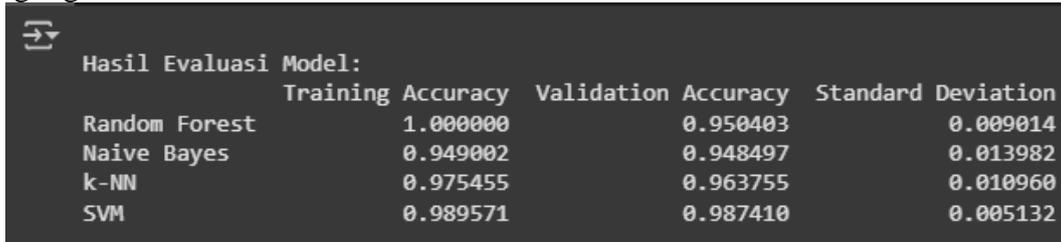
Perbandingan Akurasi dan Validasi

Tahap ini dilakukan untuk memperoleh akurasi dalam proses pengujian. *Cross validation* (CV) adalah teknik validasi model yang digunakan untuk menilai sejauh mana hasil analisis statistik dapat digeneralisasi ke kumpulan data independen. *k-fold Cross validation* merupakan salah satu dari metode CV dimana metode ini merupakan cara yang efektif untuk menggabungkan pemilihan fitur dan penyetelan parameter pembelajaran

mesin untuk melatih model yang optimal, data dibagi menjadi k lipatan luar dan kemudian lipatan dalam dibuat di setiap pelatihan luar untuk memilih fitur, menyetel parameter, dan melatih model, nilai k -fold yang direkomendasikan yaitu 10 -fold [12] [16].

Hasil dan Visualisasi

Tahap ini dilakukan visualisasi yang diperoleh berdasarkan hasil dari perbandingan akurasi validasi dan *standard deviation* dari keempat algoritma klasifikasi dengan visual yang informatif. Setelah *dataframe* dilatih dan dievaluasi menghasilkan nilai *Accuracy* dan *Standard Deviation* yang dapat dibandingkan agar terlihat perbedaan antar algoritma yang digunakan.



Hasil Evaluasi Model:			
	Training Accuracy	Validation Accuracy	Standard Deviation
Random Forest	1.000000	0.950403	0.009014
Naive Bayes	0.949002	0.948497	0.013982
k-NN	0.975455	0.963755	0.010960
SVM	0.989571	0.987410	0.005132

Gambar 2. Hasil evaluasi model

Gambar 2 menampilkan hasil evaluasi model yang telah dilatih dan dievaluasi menggunakan pemrograman *python* dengan *library sklearn*. Kinerja algoritma dievaluasi berdasarkan dua metrik utama, *Validation Accuracy* dan *Standard Deviation*. *Validation Accuracy* mengukur akurasi model dalam memprediksi kelas yang benar pada data baru, sedangkan *Standard Deviation* mengukur variasi atau sebaran akurasi validasi.

3. HASIL DAN PEMBAHASAN

Dari hasil implementasi, didapatkan hasil perbandingan algoritma antara *Random Forest*, *Naive Bayes*, *k-NN*, dan *SVM* untuk klasifikasi prestasi akademik peserta didik yang dapat disederhanakan menjadi Tabel 1.

Tabel 1. Hasil perbandingan *validation accuracy* dan *standard deviation*

Algoritma	Validation Accuracy	Standard Deviation
Random Forest	0.951929	0.011466
Naive Bayes	0.948497	0.013982
k-NN	0.963755	0.010960
SVM	0.987410	0.005132

Dari Tabel 1 diatas, dapat dilihat bahwa nilai *Random Forest* yaitu 0.951929 dan *Standard Deviation* adalah 0.011466. Algoritma *Naive Bayes* menghasilkan nilai *Validation Accuracy* 0.948497 dan *Standard Deviation* 0.013982. Algoritma *k-Nearest Neighbours* (*k-NN*) dengan nilai $k = 19$ menghasilkan nilai *Validation Accuracy* 0.963755 dan *Standard Deviation* 0.010960. Algoritma *Support Vector Machine* (*SVM*) menghasilkan nilai *Validation Accuracy* 0.987410 dan *Standard Deviation* 0.005132.

Algoritma yang memberikan performa terbaik adalah *SVM* (*Support Vector Machine*). Hal ini menunjukkan bahwa *SVM* mampu memetakan data ke ruang dimensi yang lebih tinggi dan menemukan *hyperplane* optimal yang memisahkan peserta didik dengan status tuntas dan tidak tuntas. Kemampuan *SVM* untuk memaksimalkan *margin* antara dua kelas memungkinkan algoritma ini untuk menghasilkan prediksi yang akurat,

bahkan dengan variabel yang memiliki korelasi lemah. Dengan demikian, algoritma SVM dapat digunakan sebagai alat alternatif yang handal untuk mengidentifikasi peserta didik yang berpotensi tidak tuntas dan memberikan intervensi yang tepat untuk meningkatkan peluang kelulusan mereka.

4. KESIMPULAN DAN SARAN

Berdasarkan dari hasil dan analisa disimpulkan bahwa algoritma SVM yang terpilih digunakan untuk mengembangkan model klasifikasi prestasi akademik peserta didik dalam bentuk program *python* yang masih sederhana namun memiliki akurasi yang tinggi, stabil dan konsisten. Program ini telah menjadi alat bantu bagi pihak SMP X dalam mengidentifikasi prestasi akademik peserta didik dan sebagai bahan pelaporan hasil belajar peserta didik kepada orang tua.

Saran dari penulis agar pihak sekolah dapat mengimplementasikan model ini kedalam sebuah sistem atau aplikasi yang lebih informatif untuk digunakan oleh pihak SMP X atau sekolah lain dan melakukan evaluasi secara berkala terhadap kinerja model serta lakukan penyesuaian jika diperlukan. Selain itu perlu melakukan analisis lebih lanjut untuk memahami faktor-faktor yang paling berpengaruh dalam prediksi model. Ini dapat membantu dalam memberikan rekomendasi yang lebih spesifik kepada pihak SMP X.

5. DAFTAR PUSTAKA

- [1] I. Oktadiani, H. Fitriawan, M. Nurwahidin, and Herpratiwi, "Penerapan Machine Learning Untuk Prediksi Masa Studi Mahasiswa di Perguruan Tinggi X," *Jurnal Rekayasa dan Teknologi Elektro*, vol. 17, no. 3, pp. 242–252, Sep. 2023.
- [2] A. F. A. Naibaho and A. Zahra, "Prediksi Kelulusan Siswa Sekolah Menengah Pertama Menggunakan Machine Learning," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 3, pp. 267–281, Aug. 2023, doi: 10.23960/jitet.v11i3.3056.
- [3] E. P. Saputra, M. Maulidah, N. Hidayati, and A. Saryoko, "Komparasi Evaluasi Kinerja Siswa Belajar dengan Menggunakan Algoritma Machine Learning," *Jurnal Media Informatika Budidarma*, vol. 6, no. 4, pp. 2239–2246, Oct. 2022, doi: 10.30865/mib.v6i4.4786.
- [4] W. Menna Eligo, C. Leng, A. Elias Kurika, and A. Basu, "Comparing Supervised Machine Learning Algorithms on Classification Efficiency of multiclass classifications problem," *International Journal of Emerging Trends in Engineering Research*, vol. 10, no. 6, pp. 346–360, Jun. 2022, doi: 10.30534/ijeter/2022/081062022.
- [5] S. Widaningsih and S. Yusuf, "Penerapan Data Mining Untuk Memprediksi Siswa Berprestasi Dengan Menggunakan Algoritma K Nearest Neighbor," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 3, pp. 2598–2611, Sep. 2022.
- [6] A. Chusyairi, T. Haryanto, and R. N. Hayat, "Prediksi Perubahan Iklim Untuk Pertumbuhan Tanaman Jeruk Keprok Menggunakan Naïve Bayes," *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, vol. 18, no. 1, p. 23, Feb. 2023, doi: 10.30872/jim.v18i1.9352.
- [7] Y. Ansori and K. F. H. Holle, "Perbandingan Metode Machine Learning dalam Analisis Sentimen Twitter," *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 10, no. 4, pp. 429–434, Oct. 2022, doi: 10.26418/justin.v10i4.51784.
- [8] F. M. Almufqi and A. Voutama, "Perbandingan Metode Data Mining untuk Memprediksi Prestasi Akademik Siswa," *Jurnal Teknika*, vol. 15, no. 1, pp. 61–66, Apr. 2023, doi: 10.30736/jt.v15i1.929.

- [9] R. Wahyudi and G. Kusumawardhana, “Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine,” *Jurnal Informatika*, vol. 8, no. 2, pp. 200–207, Sep. 2021, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [10] Dewa Ayu Indah Cahya Dewi and Dewa Ayu Kadek Pramita, “Optimalisasi Pengelompokan Data Tingkat Hunian Hotel dengan Algoritma K-Medoid,” *Jurnal Teknologi Informasi dan Komputer*, vol. 8, no. 3, pp. 240–246, Oct. 2022.
- [11] J. Delgadillo, “Machine learning: A primer for psychotherapy researchers,” *Psychotherapy Research*, vol. 31, no. 1, pp. 1–4, Dec. 2020, doi: 10.1080/10503307.2020.1859638.
- [12] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, “Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung,” *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 81–86, Aug. 2020, doi: 10.33096/ilkom.v12i2.507.81-86.
- [13] P. R. Sihombing and I. F. Yuliati, “Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia,” *Matrik: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 20, no. 2, pp. 417–426, May 2021, doi: 10.30812/matrik.v20i2.1174.
- [14] N. F. Munazhif, G. J. Yanris, and M. N. S. Hasibuan, “Implementation of the K-Nearest Neighbor (kNN) Method to Determine Outstanding Student Classes,” *Sinkron : Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 2, pp. 719–732, Apr. 2023, doi: 10.33395/sinkron.v8i2.12227.
- [15] F. Mandita and R. Kurnia Pratama, “Klasifikasi Penerimaan Tenaga Kerja Tertutup Menggunakan Algoritma K-Nearest Neighbor (KNN),” *Informatic and Computational Intelligent Journal*, vol. 06, no. 1, pp. 42–61, May 2024, doi: <http://dx.doi.org/10.30587/indexia.v6i1.7224>.
- [16] A. Nugroho and A. Amrullah, “Evaluasi Kinerja Algoritma K-NN menggunakan K-Fold Cross Validation pada Data Debitur KSP Galih Manunggal,” *Jurnal Informatika Teknologi dan Sains*, vol. 5, no. 2, pp. 294–300, May 2023.