JOURNAL OF COMPUTER SCIENCE AND INFORMATICS ENGINEERING
VOL. 05, NO. 3, JULY 2026: 338-346
Doi: https://doi.org/10.55537/cosie.v5i3.1780

*e-ISSN 2827-8356*
*p-ISSN 2827-9999*

# Analisis Komparatif Algoritma Klasifikasi dalam Prediksi Hasil Ujian Pelajar Berdasarkan Dataset Akademik

*Comparative Analysis of Classification Algorithms for Predicting Student Examination Outcomes Based on Academic Datasets*

Imam Sugiharto*[1], Sri Indri Wahyuni[2], Septian Tunijah Faradila,[3] Yunita[4,] Muhammad Ifan Rifani[5]
[1,2,3,4,5]Informatika, Universitas Bina Sarana Informatika Kampus Pontianak
[1]15240758@bsi.ac.id, [2]15240735@bsi.ac.id, [3]15240196@bsi.ac.id, [4]15240831@bsi.ac.id,
[5]ifan.mii@bsi.ac.id

Received: June 26, 2026 | Revised: July 1, 2026 | Accepted: July 13, 2026

**Abstrak**

Pemanfaatan data pendidikan saat ini masih terpaku pada evaluasi akhir, memicu keterlambatan intervensi bagi siswa berisiko gagal. Riset ini mengomparasikan algoritma Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), dan Random Forest untuk mengklasifikasikan capaian ujian sebagai landasan *Early Warning System*. Sebanyak 6.607 data *Student Exam Performance* diekstraksi dan diproses melalui *Mode Imputation*, *Label Encoding*, serta *StandardScaler*. Hasil pengujian menunjukkan bahwa model SVM ber-kernel RBF meraih akurasi tertinggi sebesar 76,40%, melampaui Random Forest (73,22%) dan K-NN (71,94%). Keunggulan performa ini sangat dipengaruhi oleh tingkat kepresisian SVM dalam merumuskan batas keputusan pada ruang data berdimensi tinggi. Oleh karena itu, model SVM direkomendasikan sebagai mesin analitik utama guna mendeteksi potensi kegagalan studi secara dini.

Kata kunci**:** *Big Data Analytics, Educational Data Mining, Early Warning System, Support Vector Machine, K-Nearest Neighbors, Random Forest.*

***Abstract***

*Current educational data utilization remains largely focused on final evaluations, resulting in delayed interventions for students at risk of academic failure. This study compares the performance of the Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Random Forest algorithms in classifying student examination outcomes as the foundation of an Early Warning System. A total of 6,607 Student Exam Performance records were extracted and processed using Mode Imputation, Label Encoding, and StandardScaler. The experimental results indicate that the SVM model with an RBF kernel achieved the highest accuracy of 76.40%, outperforming Random Forest (73.22%) and K-NN (71.94%). This superior performance is primarily attributed to SVM's ability to construct precise decision boundaries in high-dimensional feature spaces. Therefore, the SVM model is recommended as the primary analytical engine for the early detection of potential academic failure in educational institutions.*

*Keywords: Big Data Analytics, Educational Data Mining, Early Warning System, Support Vector Machine, K-Nearest Neighbors, Random Forest.*

## 1. PENDAHULUAN

Kualitas sistem pendidikan merupakan pilar esensial dalam mencetak sumber daya manusia yang kompeten [1]. Memasuki era digitalisasi, lembaga pendidikan secara alamiah mulai terintegrasi ke dalam lingkungan *Big Data*. Kondisi ini ditandai dengan penciptaan data

e-ISSN 2827-8356
p-ISSN 2827-9999

bervolume besar (*Volume*), bergulir cepat (*Velocity*), serta memiliki keragaman yang tinggi (*Variety*) mencakup profil demografis, rekam jejak akademis, hingga rutinitas belajar peserta didik [2]. Di tengah masifnya arus informasi tersebut, perolehan nilai serta ketepatan waktu kelulusan tetap menduduki posisi sebagai indikator utama dalam menilai kredibilitas sebuah institusi [3].

Data *Student Exam Performance* yang dianalisis pada studi ini menjadi cerminan langsung dari aspek keragaman (*Variety*) pada *Big Data* pendidikan, memuat puluhan variabel masukan. Berbagai fitur heterogen tersebut secara empiris memberikan dampak substansial terhadap naik-turunnya prestasi siswa [4]. Namun, potensi besar dari kumpulan data ini sering kali diabaikan. Banyak institusi hanya membiarkan tumpukan arsip akademik ini menganggur dan sekadar memanfaatkannya untuk administrasi pasca-evaluasi [5]. Walaupun sejumlah lembaga telah mengadopsi infrastruktur *Data Warehouse* beserta *Business Intelligence* standar, fungsinya sebagian besar masih sebatas pelaporan deskriptif, bukan sebagai instrumen pencegahan dini [6], [7], [8]. Hal ini berimbas pada lambatnya pendampingan bagi pelajar yang rentan mengalami kemerosotan nilai.

Beranjak dari permasalahan tersebut, penerapan analitik prediktif berbasis *Educational Data Mining* menjadi kebutuhan mendesak guna mentransformasikan data berdimensi tinggi menjadi sistem peringatan dini yang responsif [9]. Dalam upaya mengurai pola kompleksitas *Big Data* ini, metode klasifikasi pada *machine learning* mengambil peran sentral. Pendekatan *Support Vector Machine* (SVM) diakui sangat tangguh saat memproses klasifikasi data kompleks berkat kemampuannya merumuskan bidang pemisah (*hyperplane*) paling ideal [10]. Sebagai alternatif, *K-Nearest Neighbors* (K-NN) menyajikan metode kalkulasi jarak yang mumpuni dalam memetakan siswa berbekal kemiripan rekam jejak mereka, semisal tingkat kehadiran dan nilai masa lalu [11], [12]. Di pihak lain, *Random Forest* menawarkan kestabilan paripurna ketika dihadapkan pada variabel *Big Data* yang beraneka ragam, sekaligus piawai dalam menekan potensi terjadinya *overfitting* [4].

Meskipun beberapa studi terdahulu telah mengeksplorasi algoritma *machine learning* dalam konteks pendidikan, sebagian besar pendekatan tersebut masih bersifat parsial. Penelitian oleh Kurniawan & Farokhah (2025) [1] dan Rizqi et al. (2026) [10] membuktikan ketangguhan SVM, namun studi tersebut belum mengomparasikannya secara komprehensif dengan algoritma ansambel pada ruang data berdimensi tinggi. Di sisi lain, penggunaan K-NN dan Random Forest telah diuji oleh Fatah & Muasolli (2025) [11] serta Sulehu et al. (2025) [4], akan tetapi fokus analisisnya masih terbatas pada metrik akurasi dasar tanpa membedah ketahanan model terhadap batas keputusan antar-kelas yang tumpang tindih (*overlapping*). Secara kritis, literatur saat ini masih menyisakan celah terkait algoritma mana yang paling konsisten menangani transisi nilai siswa menengah (Grade B) yang sering kali kabur batasan fiturnya.

Untuk menjembatani kesenjangan tersebut, riset ini difokuskan pada perancangan model analitik prediktif dengan mengomparasikan performa SVM, K-NN, dan *Random Forest*. Tujuannya adalah mengolah serta mengelompokkan data pendidikan dengan heterogenitas tinggi ke dalam tiga kelas prestasi (*Grade A, B,* dan *C*). Guna memvalidasi keakuratan komputasinya, studi ini mengevaluasi ketiga algoritma secara komprehensif menggunakan parameter *Accuracy Score*, *Classification Report*, serta *Confusion Matrix*. Hasil komparasi ini diproyeksikan mampu menyoroti algoritma mana yang paling konsisten dan presisi, sehingga layak direkomendasikan sebagai motor analitik utama (*Big Data Analytics engine*) dalam membangun *Early Warning System*. Pada akhirnya, langkah ini akan memberdayakan pihak institusi untuk mendeteksi ancaman kegagalan akademis secara tajam dari kumpulan data raksasa, memungkinkan dilakukannya intervensi pembelajaran yang lebih dini, spesifik, dan didukung bukti nyata [13].

## 2. METODE PENELITIAN

e-ISSN 2827-8356
p-ISSN 2827-9999

Studi ini bersandar pada kerangka kuantitatif dengan menerapkan metode *Machine Learning* guna memetakan performa akademik pelajar. Seluruh rangkaian komputasi, bermula dari penarikan data mentah hingga tahap pengujian model, dieksekusi secara penuh menggunakan bahasa pemrograman Python di atas ekosistem *cloud* Google Colaboratory [14]. Sebagai pelengkap infrastruktur, layanan Google Drive difungsikan sebagai basis penyimpanan terintegrasi untuk menjamin keamanan dan efisiensi kolaborasi data [15]. Secara struktural, alur riset ini terbagi ke dalam lima fase utama.

*2.1. Pengumpulan Data (Data Acquisition)*

Data yang dianalisis pada penelitian ini bersumber dari *dataset Student Exam Performance* yang diakses secara terbuka melalui repositori Kaggle. *Dataset* tersebut merangkum 6.607 baris observasi dengan 20 fitur berbeda yang merepresentasikan dinamika sosial-ekonomi, kebiasaan belajar, serta profil akademik peserta didik. Guna menjamin keabsahan dan keutuhan data sebelum tahapan prapemrosesan, penarikan data mentah dilakukan secara terprogram menggunakan *Application Programming Interface* (API) resmi milik Kaggle.

*2.2 Prapemrosesan Data (Data Preprocessing)*

Sebelum memasuki fase pemodelan algoritmik, tahapan pembersihan data (*data cleaning*) diaplikasikan pada *dataset* mentah guna mengeliminasi keberadaan nilai yang kosong (*missing values*). Strategi imputasi rata-rata (*mean imputation*) diterapkan secara spesifik untuk mengatasi kekosongan pada fitur-fitur bertipe numerik. Sebaliknya, pada variabel kategorikal, pengisian data kosong merujuk pada nilai dengan frekuensi kemunculan tertinggi (*modus*). Langkah berikutnya adalah menjalankan *Label Encoding*, yakni proses mentransformasikan seluruh teks kategorikal menjadi representasi numerik sehingga selaras dengan kebutuhan operasional mesin *learning*.

*2.3. Rekayasa Fitur dan Pembagian Data (Feature Engineering & Data Splitting)*

Fokus inti dari rekayasa fitur ini adalah penetapan *Exam_Score* sebagai variabel target (Y). Melalui teknik diskritisasi (*binning*) menggunakan strategi *quantile*, skor tersebut dikonversi menjadi label kelas berjenjang, yakni Grade A, B, dan C [4]. Penggunaan strategi *quantile* ini secara otomatis mendistribusikan jumlah sampel secara merata ke dalam setiap kelas, sehingga masalah ketidakseimbangan kelas (*class imbalance*) dapat teratasi sejak tahap prapemrosesan. Di sisi lain, enam variabel terbaik diseleksi sebagai prediktor masukan (X), meliputi *Hours Studied, Attendance, Previous_Scores, Motivation_Level, Teacher Quality*, serta *Access_to_Resources*. Pemilihan keenam fitur ini didasarkan pada relevansi teoretis dalam *Educational Data Mining*, di mana perpaduan antara metrik kuantitatif akademik dan faktor pendukung (demografi/psikologis) terbukti memiliki korelasi prediktif yang kuat terhadap capaian akhir siswa [9].

Dataset kemudian dipartisi dengan proporsi 80% untuk pelatihan model (*training data*) dan alokasi 20% untuk pengujian (*testing data*). Mengingat populasi data yang cukup masif (6.607 baris), penggunaan *single train-test split* 80:20 dinilai sangat memadai untuk melatih model secara optimal sekaligus menyisakan jumlah data uji yang representatif (1.322 baris) guna memvalidasi performa algoritma tanpa risiko *overfitting* yang signifikan. Terakhir, untuk algoritma yang sangat sensitif terhadap rentang nilai (seperti K-NN dan SVM), metode *Feature Scaling* berbasis *StandardScaler* digunakan. Proses ini menormalisasi distribusi data sehingga memiliki varians satu dan rata-rata nol, memastikan perhitungan jarak berjalan optimal.

*2.4. Pemodelan Klasifikasi (Classification Modeling)*

Tiga algoritma *supervised learning* dikerahkan untuk memprediksi pengaruh berbagai

e-ISSN 2827-8356
p-ISSN 2827-9999

variabel independen terhadap tingkat capaian ujian:

1. *Support Vector Machine* (SVM)
   Mekanisme kerja utama dari algoritma ini berfokus pada penemuan bidang pemisah (*hyperplane*) paling ideal yang mampu memberikan margin terlebar antar-kelas data. Keunggulan utama SVM terletak pada ketangguhannya memproses klasifikasi di dalam ruang dimensi yang kompleks. Karakteristik ini menjadikannya sangat krusial dalam menciptakan garis batas keputusan (*decision boundary*) yang tegas, sehingga risiko bias atau tumpang tindih dalam prediksi dapat ditekan secara maksimal [1], [10]. Pada implementasi penelitian ini, SVM ditenagai oleh kernel *Radial Basis Function* (RBF) dengan pengaturan *hyperparameter default* (C=1.0, gamma='scale') yang dieksekusi di atas ruang fitur yang telah distandardisasi.
2. *K-Nearest Neighbors* (K-NN)
   Pendekatan K-NN bekerja dengan cara mengelompokkan data tak berlabel berdasarkan seberapa dekat jarak ukurnya (menggunakan metrik *Euclidean Distance*) terhadap titik data latih. Penentuan kelas akhir dieksekusi dengan melihat label mayoritas dari sejumlah *k* tetangga terdekat yang memiliki kemiripan pola tertinggi [11], [12]. Model K-NN pada studi ini dikonfigurasi dengan pengaturan *hyperparameter k* = 5 dan dilatih menggunakan data hasil *StandardScaler*.
1. *Random Forest* (RF)
   Sebagai metode *ensemble learning*, algoritma ini bekerja dengan cara mengintegrasikan hasil luaran dari sekumpulan pohon keputusan (*decision trees*). Penentuan klasifikasi akhirnya mengadopsi mekanisme suara mayoritas (*majority voting*). Pendekatan ansambel ini sangat menguntungkan karena menawarkan tingkat stabilitas komputasi yang unggul serta secara signifikan mampu meminimalkan risiko terjadinya *overfitting* saat menangani data pendidikan yang heterogen [3], [4], [9]. Model logikanya dibangun menggunakan distribusi data orisinal dengan pengaturan *hyperparameter n_estimators* = 100.

*2. 5. Evaluasi Model (Model Evaluation)*

Data uji berfungsi sebagai standar ukur untuk membandingkan dan mengevaluasi kapabilitas masing-masing model klasifikasi. Kinerja tersebut dianalisis melalui metrik evaluasi yang merangkum *Accuracy, Precision, Recall*, serta *F1-Score* ke dalam laporan klasifikasi terstruktur. Lebih lanjut, pemetaan *Confusion Matrix* divisualisasikan guna menginspeksi secara spesifik rasio perbandingan antara hasil prediksi mesin dengan kondisi aktual pada setiap kategori kelas (*Grade A, B,* dan *C*).

## 3. HASIL DAN PEMBAHASAN

*3.1 Hasil Pra-pemrosesan Data*

Tahap eksplorasi awal mengungkap bahwa dari 6.607 rekaman pada *dataset Student Exam Performance*, terdapat sejumlah observasi yang tidak lengkap (*missing values*). Permasalahan ini diselesaikan secara efektif dengan mengaplikasikan teknik imputasi *mean* (rata-rata) pada variabel numerik, serta penggunaan nilai modus pada kolom kategorikal, sehingga integritas data kembali utuh. Lebih lanjut, proses diskritisasi (*binning*) dimanfaatkan untuk memecah fitur target *Exam_Score* menjadi tiga kategori bertingkat: *Grade A* (Tinggi), *Grade B* (Sedang), dan *Grade C* (Rendah). Dari total data bersih tersebut, sebanyak 1.322 baris disisihkan secara khusus sebagai data uji (*testing data*) murni guna mengukur ketepatan prediksi model pada fase akhir eksperimen.

*3.2 Hasil Komparasi Kinerja Model*

engujian komputasi menggunakan data uji mengonfirmasi adanya perbedaan tingkat keandalan di antara ketiga algoritma. Perbandingan performa ini dianalisis dengan bersandar pada

e-ISSN 2827-8356
p-ISSN 2827-9999

empat indikator utama, yakni *Accuracy, Precision, Recall*, dan *F1-Score*, yang rinciannya dapat dicermati pada Tabel 1 berikut:

Tabel 1. Perbandingan Kinerja Algoritma Klasifikasi

| Model Klasifikasi | Akurasi (%) | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Support Vector Machine (SVM) | **76.40%** | **0.77** | **0.76** | **0.77** |
| Random Forest (RF) | 73.22% | 0.74 | 0.73 | 0.73 |
| K-Nearest Neighbors (K-NN) | 71.94% | 0.73 | 0.72 | 0.72 |

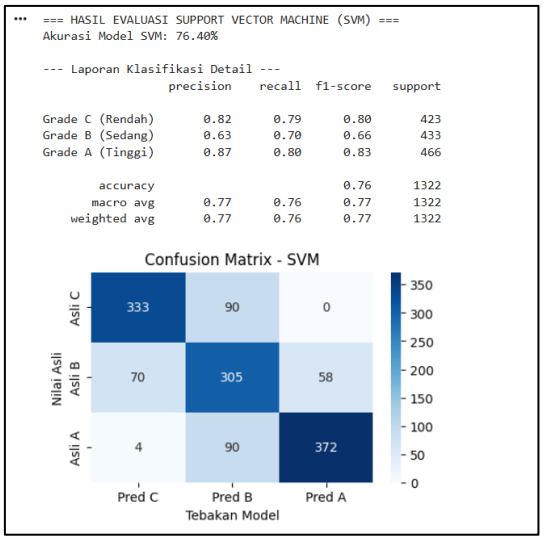
Merujuk pada distribusi angka di Tabel 1, *Support Vector Machine* (SVM) tampil sebagai model paling tangguh dengan membukukan puncak akurasi di angka 76,40%. Posisi kedua diamankan oleh *Random Forest* dengan torehan 73,22%, sementara *K-Nearest Neighbors* (K-NN) memberikan performa paling inferior dengan akurasi terbatas pada 71,94%.

Meskipun pengujian ini tidak menyertakan uji signifikansi statistik tradisional (seperti analisis varians), penggunaan kombinasi metrik evaluasi yang komprehensif (Akurasi, Presisi, *Recall*, dan *F1-Score*) pada *dataset* uji yang besar (1.322 observasi) telah memberikan tingkat kepercayaan (reliabilitas) yang kokoh terhadap hasil prediksi model. Metrik *F1-Score* yang stabil pada SVM (0,77) secara khusus membuktikan bahwa algoritma ini memiliki harmoni yang baik antara kemampuan mendeteksi kelas secara presisi dan meminimalkan *False Negative*, sehingga performanya dapat dipertanggungjawabkan.

*3.3 Analisis Confusion Matrix dan Pembahasan*

Guna menelaah lebih spesifik mengenai letak keberhasilan dan pola kekeliruan prediksi (*misclassification*) dari tiap algoritma, evaluasi dilanjutkan dengan membedah visualisasi *Confusion Matrix*.

Sebagai model paling superior, SVM membuktikan keandalannya dalam mengisolasi kelas-kelas yang berseberangan secara ekstrem, meskipun masih dihadapkan pada kendala di area transisi. Detail sebaran prediksi ini divisualisasikan pada matriks SVM di bawah ini:



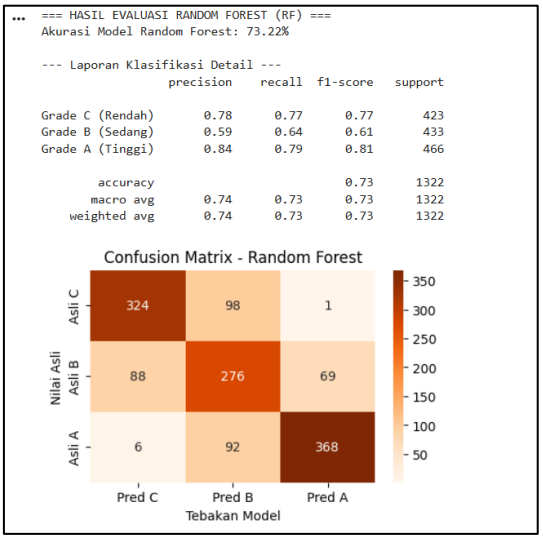

Gambar 1. Confusion Matrix - Support Vector Machine (SVM)

Berdasarkan Gambar 1, mesin SVM terbukti sangat kapabel dalam mendeteksi *True Positive*, yakni sukses memetakan 372 siswa berprestasi tinggi (*Grade A*) dan 333 siswa berprestasi rendah (*Grade C*). Tingkat kebingungan algoritma baru terlihat saat

e-ISSN 2827-8356
p-ISSN 2827-9999

mengklasifikasikan siswa pada kategori menengah (*Grade B*). Terdapat 70 observasi yang meleset diprediksi sebagai *Grade C*, dan 58 lainnya keliru terdeteksi sebagai *Grade A*. Kendati demikian, SVM mencatatkan angka kebocoran antarkelas paling minim dibandingkan dua pesaingnya. Fakta ini menegaskan bahwa perpaduan antara *StandardScaler* dan kernel *Radial Basis Function* (RBF) sangat solid untuk melukiskan garis pemisah (*decision boundary*) yang ketat pada ruang vektor berdimensi tinggi.

Di sisi lain, seperti yang tersaji pada Gambar 2, algoritma *Random Forest* (RF) memperlihatkan distribusi prediksi yang lebih terpencar:

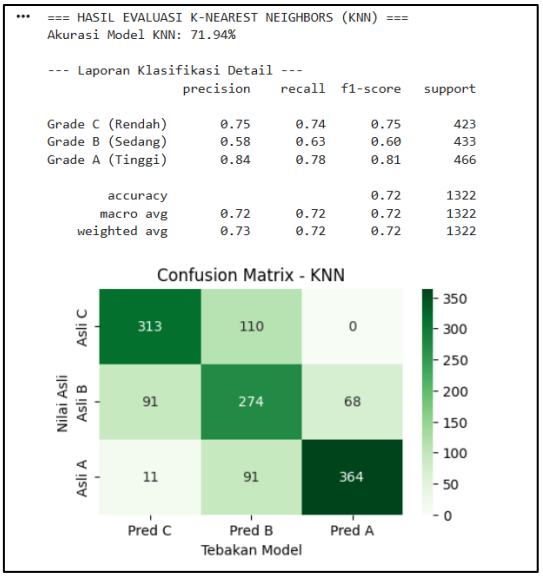


Gambar 2. Confusion Matrix - Random Forest (RF)

Merujuk pada matriks tersebut, RF sejatinya cukup tangguh bersaing dengan SVM pada area kelas ekstrem, dengan keberhasilan menebak 368 data *Grade A* dan 324 data *Grade C*. Sayangnya, algoritma *ensemble* berbasis pepohonan ini memproduksi lebih banyak *False Positive* saat menguji kelas medium (*Grade B*), dengan perolehan *True Positive* yang merosot menjadi 276 data. Meskipun RF sangat diandalkan untuk memproses data tabular campuran, mekanisme pemungutan suara mayoritas (*majority voting*) rupanya belum cukup sensitif untuk menarik batas tegas pada irisan fitur *Grade B*.

Pelemahan performa paling mencolok terjadi pada algoritma K-NN. Teknik kalkulasi jarak ini sangat kesulitan memilah karakteristik siswa yang saling tumpang tindih. Pola kesalahan ini terekam jelas pada matriks berikut:

e-ISSN 2827-8356
p-ISSN 2827-9999



Gambar 3. Confusion Matrix - K-Nearest Neighbors (K-NN)

Sesuai visualisasi pada Gambar 3, K-NN mendulang angka *True Positive* terendah (274 data) pada kategori *Grade B*. Angka prediksi meleset (*misclassification*) tergolong masif; di mana 91 siswa *Grade B* terlempar ke kelas *Grade C*, dan 68 sisanya diprediksi secara keliru sebagai *Grade A*. Tingginya rasio kesalahan ini mengindikasikan kerentanan K-NN terhadap fenomena *curse of dimensionality*. Mengingat indikator pembatas pada performa siswa kelas menengah sangat abu-abu, rumusan jarak *Euclidean* gagal membangun teritorial klasifikasi yang tajam, yang pada akhirnya mendegradasi tingkat akurasi secara keseluruhan.

Keunggulan SVM dalam penelitian ini sejalan dengan temuan Kurniawan & Farokhah (2025) [1] serta Rizqi et al. (2026) [10], yang juga menegaskan bahwa kemampuan SVM dalam membangun *hyperplane* optimal menjadikannya jauh lebih superior dibanding metode jarak seperti K-NN dalam memprediksi capaian belajar. Di sisi lain, performa Random Forest yang cukup kompetitif (73,22%) turut mengonfirmasi studi Sulehu et al. (2025) [4], di mana algoritma *ensemble* sangat andal menangani variasi data akademik, meski dalam kasus dataset ini, RF masih kalah sensitif dibanding *kernel* RBF milik SVM saat membelah area kelas yang tumpang tindih.

Terlepas dari akurasi model yang menjanjikan, penelitian ini memiliki sejumlah keterbatasan. Pertama, analitik model sangat bergantung pada ketersediaan variabel dari satu sumber dataset sekunder, sehingga generalisasi performa model pada karakteristik demografi siswa dari wilayah atau jenjang pendidikan lain perlu diuji lebih lanjut. Kedua, fitur masukan dominan bersifat kuantitatif, belum mengakomodasi variabel psikologis yang mendalam secara real-time (seperti tingkat stres siswa atau dinamika lingkungan keluarga) yang sejatinya memiliki korelasi kuat terhadap performa ujian. Keterbatasan ruang lingkup *dataset* ini menjadi ruang terbuka bagi penelitian selanjutnya untuk memperkaya variasi fitur dan menyempurnakan keandalan *Early Warning System*.

## 4. KESIMPULAN

Berpijak pada hasil komparasi ketiga algoritma *Machine Learning*, dapat ditarik konklusi bahwa Support Vector Machine (SVM) adalah solusi analitik paling optimal, secara spesifik pada karakteristik *dataset Student Exam Performance* yang dievaluasi dalam studi ini. Model ini mendominasi pengujian dengan capaian akurasi sebesar 76,40% dan indeks *F1-Score* 0,77, mengungguli kemampuan komputasi Random Forest (73,22%) maupun K-NN (71,94%).

e-ISSN 2827-8356
p-ISSN 2827-9999

Penemuan paling krusial dalam studi ini adalah teridentifikasinya pola kendala komputasi yang serupa di seluruh model saat berhadapan dengan klasifikasi Grade B (Sedang). Ambang batas fitur pada siswa kategori menengah ini memiliki area irisan (*overlap*) yang sangat pekat, baik dengan profil siswa unggulan maupun siswa yang tertinggal. Sebagai kontribusi ilmiah, penelitian ini membuktikan bahwa penggunaan *StandardScaler* yang dipadukan dengan kernel RBF pada SVM merupakan mekanisme paling andal untuk menembus dan memetakan batas keputusan (*decision boundary*) pada data pendidikan berdimensi tinggi yang memiliki irisan kelas yang pekat tersebut.

Secara implikasi praktis, institusi pendidikan dapat secara langsung mengadopsi model klasifikasi SVM ini sebagai motor penggerak analitik dalam membangun *Early Warning System*. Sistem ini akan memberdayakan pihak sekolah untuk mendeteksi potensi kegagalan siswa secara dini, sehingga intervensi dan pendampingan akademik dapat dilakukan jauh sebelum evaluasi akhir semester. Untuk menyempurnakan penelitian mendatang, sangat disarankan untuk mengimplementasikan metode *Feature Selection* guna menyaring variabel yang kurang deterministik, serta penerapan *Hyperparameter Tuning* (nilai C dan gamma) di algoritma SVM untuk mempertajam sensitivitas mesin.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Bapak Muhammad Ifan Rifani, S.Kom., M.Kom., selaku dosen pembimbing atas segala arahan, bimbingan keilmuan, dan masukan konstruktif yang diberikan selama proses penyusunan penelitian ini. Ucapan terima kasih juga ditujukan kepada segenap civitas academica Universitas Bina Sarana Informatika (UBSI) atas dukungan lingkungan akademik yang memfasilitasi terlaksananya riset ini dengan baik.

## KONFLIK KEPENTINGAN

Penulis menyatakan bahwa tidak ada konflik kepentingan finansial maupun personal yang dapat memengaruhi hasil penelitian maupun penulisan artikel ini.

## DAFTAR PUSTAKA

[1] F. B. Kurniawan dan L. Farokhah, "Aplikasi Cerdas Prediksi Kelulusan Mahasiswa Berbasis Website Menggunakan Metode Support Vector Machine (SVM)," *JURNAL FASILKOM*, vol. 15, no. 1, hlm. 155–162, 2025, doi: 10.37859/jf.v15i1.8767.

[2] C. Tandilino, R. Khaerany, G. T. Pontoh, dan A. Indrijawati, "Big Data and Business Intelligence in the Public Sector: Implementation and Benefits," *Jurnal Akuntansi Aktual*, vol. 12, no. 1, hlm. 27–47, 2025, doi: 10.17977/um004v12i12025p027.

[3] A. A. Sitanggang, M. Y. Lase, dan S. P. Sipayung, "Prediksi Kelulusan Mahasiswa Menggunakan Logistic Regression dan Random Forest Berdasarkan Data Akademik," *Jurnal Pendidikan Tambusai*, vol. 10, no. 1, hlm. 3503–3513, 2026, doi: 10.31004/jptam.v10i1.36623.

[4] M. Sulehu, Wisda, F. Wanita, dan Markani, "Optimasi Prediksi Kelulusan Mahasiswa Menggunakan Random Forest untuk Meningkatkan Tingkat Retensi," *Jurnal Minfo Polgan*, vol. 13, no. 2, hlm. 2364–2374, 2025, doi: 10.33395/jmp.v13i2.14472.

[5] R. Pahlevie, W. Purnomo, dan M. C. Saputra, "Pembangunan Dashboard Business Intelligence Data Akademik SMPN 1 Nguling," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 9, 2025, Tersedia pada: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/15295

e-ISSN 2827-8356
p-ISSN 2827-9999

[6] H. F. Ramadhan, A. Fauzi, C. N. Rupelu, D. P. Aprillia, N. D. Anjani, dan Halimatusadiah, "Pengaruh Business Intelligence Terhadap Perusahaan Dalam Pengambilan Keputusan: Business Intelligence, Arsitektur BI dan Data Warehouse (Kajian Studi Business Intelligence)," *JEMSI (Jurnal Ekonomi Manajemen Sistem Informasi)*, vol. 3, no. 6, hlm. 639–644, 2022, doi: 10.31933/jemsi.v3i6.

[7] D. Tribuana, A. D. H. Agustan, Hidayat, E. Halimah, K. Dianah, dan Isiswanty, "Membangun Taxonomy Riset Big Data Analytics dan Business Intelligence: Systematic Literature Review dalam Konteks Manajemen Informatika," *JTBC: Jurnal Teknologi dan Bisnis Cerdas*, vol. 1, no. 2, hlm. 140–154, 2025, doi: 10.64476/jtbc.v1i2.12.

[8] N. R. Khairani, N. Y. Setiawan, dan W. Purnomo, "Pengembangan Dashboard Business Intelligence untuk Monitoring Data Akademik Sekolah Menggunakan Metode Kimball (Studi Kasus: MTSS Bina Ihsan Mulia)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 9, 2025, Tersedia pada: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/15316

[9] M. A. Rachman, D. R. Maulana, B. Timothy, dan R. Annisa, "Aplikasi Prediksi Tingkat Kelulusan Mahasiswa Berdasarkan Data Akademik dan Demografi Menggunakan Algoritma Klasifikasi Random Forest," *Jurnal Media Informatika (JUMIN)*, vol. 7, no. 1, hlm. 247–255, 2026, doi: 10.55338/jumin.v7i1.7605.

[10] N. I. Rizqi, P. K. Farista, dan Agussalim, "Penerapan Metode Support Vector Machine untuk Memprediksi Kelulusan Siswa di SMA Bina Marga," *Jurnal Sistem Informasi, Teknik Komputer dan Teknologi Pendidikan (JUSTIKPEN)*, vol. 5, no. 2, hlm. 225–235, 2026, doi: 10.55338/justikpen.v5i2.388.

[11] Z. Fatah dan H. S. Muasolli, "Penerapan Algoritma K-Nearest Neighbor (K-NN) untuk Memprediksi Kelulusan Mahasiswa Menggunakan RapidMiner," *JAMASTIKA*, vol. 4, no. 2, hlm. 106–113, 2025, doi: 10.35473/jamastika.v4i2.4482.

[12] H. M. Nur, V. Maarif, F. F. D. Imaniawan, S. Sardiarinto, dan E. Saputro, "Algoritma K-Nearest Neighbor (K-NN) Untuk Prediksi Ketidaklulusan Mahasiswa Berdasar Presensi dan IPK," *Indonesian Journal Computer Science*, vol. 4, no. 1, hlm. 1–6, 2025, doi: 10.31294/3smve945.

[13] M. M. Triputra, A. Rifai, dan K. D. Tania, "Evaluasi Kualitas Pendidikan Dasar di Sumatera Selatan Menggunakan Business Intelligence Model," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 3, hlm. 601–613, 2025, doi: 10.52436/1.jpti.618.

[14] F. Wilyani, Q. N. Arif, dan F. Aslimar, "Pengenalan Dasar Pemrograman Python dengan Google Colaboratory," *Jurnal Pengabdian Pada Masyarakat Indonesia (JPPMI)*, vol. 3, no. 1, hlm. 8–14, 2024, doi: 10.55606/jppmi.v3i1.1087.

[15] H. P. Fitrian, I. Kusdiawati, R. R. Nursoleh, R. Ramdani, dan A. Rois, "Pemanfaatan Google Drive sebagai Layanan Cloud Computing untuk Penyimpanan Kolaborasi Data," *TAMIKA: Jurnal Tugas Akhir Manajemen Informatika dan Komputerisasi Akuntansi*, vol. 5, no. 2, hlm. 433–439, 2025, doi: 10.46880/tamika.Vol5No2.pp433-439.