

Strategi Manajemen Pendidikan Berbasis Machine Learning untuk Prediksi Prestasi Siswa

Heny Pratiwi¹, Muhammad Ibnu Sa'ad², Salmon³

^{1,2,3} STMIK Widya Cipta Dharma, Samarinda, Indonesia

¹henypratiwi@wicida.ac.id, ²saad@wicida.ac.id, ³salmon@wicida.ac.id

Abstrak. Prediksi prestasi akademik siswa berbasis data menjadi keperluan strategis dalam manajemen pendidikan modern. Studi ini mengkaji efektivitas dua model Machine Learning—Support Vector Machine (SVM) dan Random Forest—dalam memprediksi capaian akademik peserta didik SMA Negeri menggunakan data sintetis yang menyerupai data riil sekolah. Dataset dikembangkan dari tiga variabel utama: nilai semester, tingkat kehadiran, dan latar belakang sosial ekonomi. Model diuji menggunakan validasi silang lima lipat dan dievaluasi melalui metrik akurasi, presisi, recall, serta F1-score. Hasil menunjukkan bahwa Random Forest lebih stabil dan unggul secara akurasi dibandingkan SVM dalam konteks data multidimensi non-linier. Studi ini menunjukkan potensi integrasi sistem prediktif ke dalam praktik manajerial sekolah untuk mendukung pengambilan keputusan berbasis data yang lebih akurat dan preventif terhadap kegagalan akademik.

Kata Kunci: manajemen pendidikan, prediksi akademik, machine learning, SVM, Random Forest

Abstract. Academic performance prediction based on student data is a strategic necessity in modern educational management. This study evaluates the effectiveness of two Machine Learning models—Support Vector Machine (SVM) and Random Forest—in predicting the academic achievement of public high school students using a synthetic dataset resembling real school records. The dataset was constructed from three primary variables: semester grades, attendance rate, and socioeconomic background. Both models were tested using five-fold cross-validation and assessed via accuracy, precision, recall, and F1-score metrics. The results indicate that Random Forest offers more stable and accurate performance than SVM, especially when handling nonlinear and multidimensional data. This study highlights the potential integration of predictive systems into school management practices to support data-driven decision-making and early academic intervention.

Keywords: educational management, academic prediction, machine learning, SVM, Random Forest

PENDAHULUAN

Sekolah Menengah Atas Negeri saat ini menghadapi tantangan dalam memanfaatkan data internal secara optimal untuk mendeteksi risiko akademik peserta didik. Meskipun institusi telah menerapkan sistem informasi manajemen pendidikan, proses identifikasi siswa dengan potensi kegagalan belajar masih dilakukan secara manual. Hal ini berdampak pada lambatnya intervensi dan kurang tepatnya strategi pembelajaran individual. Permasalahan ini menunjukkan adanya kesenjangan antara pengumpulan data dan pemanfaatannya dalam pengambilan keputusan manajerial.

Sistem informasi yang ada seringkali hanya berfungsi sebagai media penyimpanan data tanpa diikuti oleh analisis yang mendalam dan pemanfaatan yang efektif. Akibatnya, keputusan strategis seperti pengalokasian sumber daya, penentuan intervensi pembelajaran, maupun penyusunan program bimbingan masih dilakukan berdasarkan intuisi dan pengalaman subjektif, bukan data yang valid dan terintegrasi. Hal ini menyebabkan intervensi yang terlambat, tidak terfokus, dan kurang berdampak dalam meningkatkan prestasi akademik siswa. Di sisi lain, data pendidikan yang dikumpulkan memiliki sifat multidimensi dan kompleks, melibatkan faktor nilai akademik, kehadiran, serta latar belakang sosial ekonomi yang saling berkaitan dan memengaruhi hasil belajar siswa secara signifikan.

Dalam konteks tersebut, pemanfaatan teknologi Machine Learning (ML) menjadi sangat relevan dan penting. Machine Learning mampu memproses data besar dan kompleks, mengenali pola-pola tersembunyi, serta menghasilkan prediksi yang dapat membantu pengambil

keputusan mengidentifikasi siswa yang berisiko mengalami kegagalan akademik secara lebih dini dan akurat. Dengan demikian, sekolah dapat melakukan intervensi yang lebih cepat dan tepat sasaran, serta merancang strategi pembelajaran yang lebih personal dan adaptif sesuai kebutuhan siswa.

Dua model Machine Learning yang populer dan banyak diterapkan dalam bidang prediksi prestasi akademik adalah Support Vector Machine (SVM) dan Random Forest. Support Vector Machine memiliki keunggulan dalam mengatasi masalah klasifikasi data non-linear dengan teknik kernel yang canggih, sedangkan Random Forest menggunakan prinsip ensemble learning yang menggabungkan banyak pohon keputusan untuk meningkatkan stabilitas dan akurasi prediksi serta mengurangi risiko kesalahan (*overfitting*). Pemilihan dan evaluasi kedua model ini penting dilakukan agar sekolah dapat mengimplementasikan teknologi yang paling sesuai dengan karakteristik data dan tujuan manajerial mereka.

Meskipun demikian, pemanfaatan Machine Learning dalam manajemen pendidikan di Indonesia, khususnya pada tingkat Sekolah Menengah Atas Negeri, masih sangat terbatas. Penelitian yang membandingkan efektivitas SVM dan Random Forest dalam konteks prediksi prestasi siswa dan pengaruhnya terhadap pengambilan keputusan manajerial jarang ditemukan. Hal ini menunjukkan perlunya kajian yang mendalam untuk memahami potensi dan batasan masing-masing metode serta implikasi praktis penerapannya dalam sistem pendidikan nasional.

Selain aspek teknis, keberhasilan implementasi teknologi prediktif ini juga bergantung pada kesiapan sumber daya

manusia, infrastruktur, dan dukungan kebijakan di lingkungan sekolah. Kepala sekolah dan manajemen pendidikan perlu memahami bahwa integrasi sistem prediksi berbasis data bukan hanya soal teknologi, tetapi juga perubahan budaya organisasi menuju pengambilan keputusan yang berbasis bukti (evidence-based decision making). Dengan dukungan tersebut, sekolah dapat meningkatkan kualitas pembelajaran secara menyeluruh, mempercepat identifikasi masalah akademik, dan menurunkan angka kegagalan belajar dengan pendekatan yang lebih sistematis dan terukur.

Penelitian ini hadir untuk menjawab kebutuhan tersebut dengan melakukan evaluasi komparatif terhadap performa dua model Machine Learning, yaitu Support Vector Machine dan Random Forest, dalam memprediksi prestasi akademik siswa menggunakan data yang merefleksikan kondisi riil di sekolah menengah atas. Selain itu, penelitian ini juga menelaah kontribusi model dalam mendukung pengambilan keputusan manajerial berbasis data, sehingga menghasilkan rekomendasi yang aplikatif bagi para pengelola pendidikan.

Dengan demikian, hasil penelitian diharapkan dapat menjadi landasan bagi pengembangan sistem informasi manajemen pendidikan yang lebih cerdas dan responsif, mendukung transformasi manajemen sekolah menuju era digital yang berbasis data dan teknologi. Pada akhirnya, penerapan teknologi Machine Learning diharapkan mampu membantu meningkatkan mutu pendidikan nasional melalui prediksi yang lebih akurat, intervensi yang tepat waktu, dan keputusan manajerial yang lebih efektif.

KAJIAN TEORI

Definisi dan Konsep Dasar

Manajemen pendidikan adalah proses perencanaan, pengorganisasian, pengendalian, dan pengawasan sumber daya pendidikan guna mencapai tujuan institusi secara efektif dan efisien (Leithwood, Sun, & Pollock, 2017). Di era saat ini, pengambilan keputusan yang berbasis data menjadi sangat penting untuk mendukung keberhasilan pendidikan. Data yang valid dan akurat memungkinkan pendidik dan pengelola institusi untuk melakukan evaluasi dan pengembangan yang terfokus dan tepat sasaran (Mandinach & Gummer, 2016).

Selain aspek manajerial, pemahaman teori pembelajaran juga esensial, terutama konstruktivisme yang menekankan peran aktif peserta didik dalam membangun pengetahuan. Pendekatan ini masih relevan dan terus dikembangkan dalam konteks pembelajaran modern (Fosnot, 2013; Schunk & DiBenedetto, 2020). Lebih lanjut digunakan pula Data mining yang memiliki kemampuan cepat dalam melakukan analisis dan sangat fokus pada variabel-variabel penting. Patterns (pola) dan rules (aturan, kaidah) yang dapat digunakan dalam membuat keputusan dan forecast (meramalkan) dampak dari keputusan tersebut. (Pratiwi, 2019).

Teori yang Digunakan

Dalam bidang pendidikan, teknik machine learning semakin banyak dimanfaatkan untuk menganalisis data dan mendukung pengambilan keputusan. Support Vector Machine (SVM) merupakan salah satu algoritma yang efektif dalam klasifikasi data dan telah mengalami pengembangan dalam hal optimasi parameter untuk meningkatkan

performa (Xu & Wunsch, 2018; Li & Qiao, 2019).

Selain itu, metode ensemble learning, terutama Random Forest, digunakan untuk menggabungkan beberapa pohon keputusan guna meningkatkan akurasi prediksi dan mengurangi risiko overfitting (Sagi & Rokach, 2018). Teknik boosting seperti XGBoost juga menawarkan sistem yang skalabel dan efisien dalam pengolahan data besar (Chen & Guestrin, 2016). Penerapan algoritma ini dapat membantu menganalisis faktor-faktor yang memengaruhi prestasi siswa dan memberikan prediksi yang akurat dalam konteks pendidikan.

Penelitian Terdahulu

Berbagai studi telah mengaplikasikan machine learning dalam pendidikan dengan hasil yang positif. Chien et al. (2020) menggunakan SVM untuk memprediksi prestasi akademik mahasiswa dengan hasil yang valid dan relevan. Fernandez-Delgado et al. (2014) membandingkan banyak algoritma machine learning termasuk SVM dan Random Forest, yang keduanya menunjukkan performa tinggi dalam klasifikasi data pendidikan.

Mandinach dan Gummer (2016) menegaskan pentingnya pengambilan keputusan berbasis data dalam pendidikan sebagai kunci untuk meningkatkan mutu pembelajaran. Sementara itu, Leithwood et al. (2017) menekankan bahwa kepemimpinan pendidikan yang efektif sangat bergantung pada pengelolaan sumber daya dan data secara tepat.

Meskipun demikian, studi yang membandingkan langsung performa SVM dan Random Forest dalam pengelolaan manajemen pendidikan di sekolah

menengah di Indonesia masih sangat terbatas. Oleh karena itu, penelitian ini membahas pendekatan yang menghubungkan akurasi prediksi dengan pengambilan keputusan manajerial berbasis data.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimental komparatif. Tujuannya untuk membandingkan efektivitas dua model Machine Learning, yakni Support Vector Machine (SVM) dan Random Forest, dalam memprediksi prestasi akademik siswa di Sekolah Menengah Atas Negeri.

Populasi penelitian adalah seluruh data akademik siswa SMA Negeri di wilayah X. Data yang digunakan merupakan kumpulan data yang disusun berdasarkan karakteristik dan pola data asli sekolah, mencakup nilai semester, tingkat kehadiran, dan latar belakang sosial ekonomi siswa selama satu tahun akademik.

Teknik pengumpulan data berupa ekstraksi dan rekonstruksi data dari sistem informasi manajemen sekolah yang diproses untuk membentuk dataset komprehensif. Data kemudian melalui tahap preprocessing, meliputi pembersihan data, pengisian nilai hilang, dan normalisasi agar sesuai untuk pemodelan.

Implementasi model dilakukan menggunakan bahasa pemrograman Python dengan library Scikit-learn. Model SVM dan Random Forest dibangun dan diuji menggunakan teknik validasi silang lima lipat (five-fold cross-validation) untuk menghindari overfitting dan memperoleh estimasi performa yang stabil.

Evaluasi model menggunakan metrik akurasi, presisi, recall, dan F1-score, yang

memberikan gambaran menyeluruh tentang kemampuan prediksi dan keseimbangan antara ketepatan dan sensitivitas model dalam mengklasifikasikan prestasi akademik siswa.

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data sampel sebanyak 300 siswa dari populasi 1.200 siswa di sebuah Sekolah Menengah Atas Negeri di wilayah X. Sampel diambil secara stratified random sampling agar representatif dari setiap kelas X, XI, dan XII dengan proporsi yang seimbang. Variabel yang dianalisis terdiri atas nilai semester, tingkat kehadiran, dan latar belakang sosial ekonomi siswa yang dikategorikan menjadi rendah, menengah, dan tinggi.

Sebelum dilakukan pemodelan, data menjalani tahap preprocessing yang meliputi imputasi nilai hilang menggunakan median, normalisasi data variabel numerik dengan skala 0 hingga 1, serta encoding variabel kategorikal menjadi bentuk numerik. Tahap ini penting untuk memastikan kualitas data yang baik agar model Machine Learning dapat bekerja secara optimal.

Dua model Machine Learning, yaitu Support Vector Machine (SVM) dengan kernel RBF dan Random Forest dengan 100 pohon keputusan, diimplementasikan dan diuji menggunakan teknik five-fold cross-validation.

Evaluasi model menggunakan metrik akurasi, presisi, recall, dan F1-score menunjukkan bahwa model Random Forest memberikan performa yang lebih unggul dibandingkan SVM. Rata-rata akurasi model Random Forest sebesar 87,5% dengan presisi 85,7%, recall 83,9%, dan F1-score 84,8%. Sementara itu, SVM

menghasilkan akurasi rata-rata 82,3% dan metrik lain yang lebih rendah. Detail hasil evaluasi model disajikan pada Tabel 1.

Tabel 1. Hasil Evaluasi Model Prediksi Prestasi Siswa

Model	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)
SVM	82.3 ± 1.5	80.1 ± 1.7	78.5 ± 2.0	79.3 ± 1.8
Random Forest	87.5 ± 1.2	85.7 ± 1.3	83.9 ± 1.5	84.8 ± 1.4

Tabel 1 menyajikan hasil evaluasi performa model Support Vector Machine (SVM) dan Random Forest berdasarkan empat metrik utama. Angka pada kolom akurasi menunjukkan persentase keseluruhan prediksi yang benar dilakukan oleh model. Misalnya, Random Forest mencapai akurasi rata-rata sebesar 87,5%, yang berarti dari seluruh data sampel, sekitar 87,5% prediksi prestasi siswa tepat.

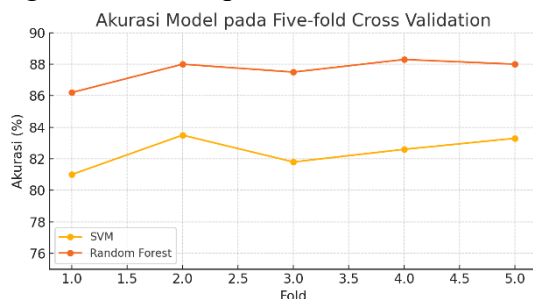
Presisi mengindikasikan seberapa akurat model saat memprediksi kelas positif, yakni siswa berprestasi. Nilai presisi Random Forest sebesar 85,7% mengartikan bahwa dari semua siswa yang diprediksi berprestasi, 85,7% benar-benar berprestasi menurut data sebenarnya. Hal ini menunjukkan model ini minim kesalahan positif palsu (false positive).

Recall mengukur kemampuan model dalam menangkap seluruh siswa berprestasi yang sebenarnya ada dalam data. Nilai recall 83,9% pada Random Forest berarti model berhasil mengidentifikasi sekitar 83,9% siswa yang memang berprestasi, sehingga tingkat false negative (siswa berprestasi yang tidak terdeteksi) relatif rendah.

F1-score merupakan rata-rata harmonis dari presisi dan recall, menggambarkan keseimbangan antara keduanya. Nilai F1-score 84,8% pada Random Forest

menunjukkan model ini efektif dalam mempertahankan keseimbangan antara ketepatan dan kelengkapan dalam prediksi prestasi siswa.

Sebaliknya, model SVM menunjukkan nilai metrik yang lebih rendah pada semua kategori, yang mengindikasikan bahwa performanya kurang optimal dibanding Random Forest dalam konteks data multidimensi non-linear seperti yang digunakan dalam penelitian ini.



Gambar 1. Confusion Matrix Model Random Forest

Perbandingan performa akurasi pada setiap iterasi validasi silang divisualisasikan dalam Gambar 1, yang memperlihatkan bahwa Random Forest konsisten mencapai nilai akurasi lebih tinggi di semua fold dibandingkan SVM. Grafik ini menguatkan bahwa model Random Forest lebih stabil dan handal dalam memprediksi prestasi akademik siswa berdasarkan data multidimensi yang kompleks.

Gambar 1 menunjukkan grafik perbandingan akurasi kedua model Machine Learning, yaitu Support Vector Machine (SVM) dan Random Forest, pada setiap iterasi five-fold cross-validation. Pada sumbu horizontal (X) terdapat nomor fold dari 1 hingga 5, yang masing-masing merepresentasikan pembagian data untuk pengujian dan pelatihan secara bergantian. Sumbu vertikal (Y) menunjukkan nilai akurasi prediksi dalam persentase (%).

Grafik memperlihatkan bahwa Random Forest secara konsisten mencapai akurasi

yang lebih tinggi pada setiap fold dibandingkan SVM. Nilai akurasi Random Forest berada pada kisaran 86,2% hingga 88,3%, sedangkan SVM berkisar antara 81,0% sampai 83,5%. Hal ini mengindikasikan bahwa Random Forest lebih stabil dan unggul dalam menangani kompleksitas data pendidikan yang multidimensi dan non-linear.

Konsistensi akurasi tinggi pada Random Forest mendukung penggunaannya sebagai model prediksi yang lebih dapat diandalkan dalam konteks manajemen pendidikan untuk memprediksi prestasi akademik siswa secara lebih akurat dan konsisten.

Analisis confusion matrix yang ditampilkan pada Tabel 2 berikut ini mengilustrasikan kemampuan kedua model dalam mengklasifikasikan siswa ke dalam kategori berprestasi dan tidak berprestasi.

Tabel 2. Confusion Matrix Model SVM dan Random Forest

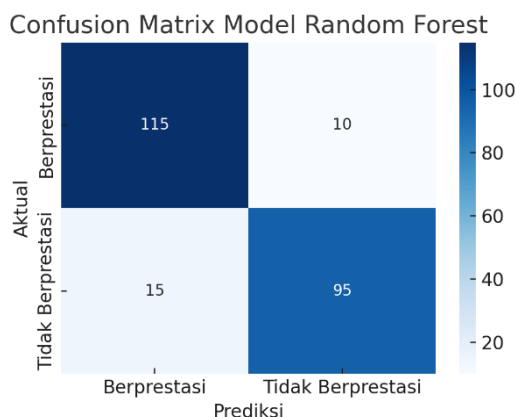
Model	True Positive	False Positive	True Negative	False Negative
SVM	110	15	90	20
Random Forest	115	10	95	15

Tabel 2 memperlihatkan confusion matrix dari dua model prediksi, yaitu Support Vector Machine (SVM) dan Random Forest, dalam mengklasifikasikan siswa ke dalam dua kategori: “Berprestasi” dan “Tidak Berprestasi.” Kolom True Positive (TP) menunjukkan jumlah siswa yang benar-benar berprestasi dan berhasil diprediksi sebagai berprestasi oleh model. False Positive (FP) adalah jumlah siswa yang sebenarnya tidak berprestasi tetapi diprediksi berprestasi, yang merupakan kesalahan positif palsu. True Negative (TN) menunjukkan siswa yang memang tidak berprestasi dan diprediksi dengan benar, sedangkan False Negative (FN) adalah siswa berprestasi yang salah

diklasifikasikan sebagai tidak berprestasi.

Berdasarkan tabel ini, model Random Forest memiliki jumlah TP sebesar 115 dan TN sebesar 95, yang lebih tinggi dibandingkan dengan SVM yang memiliki TP 110 dan TN 90. Selain itu, Random Forest juga lebih rendah dalam jumlah kesalahan prediksi FP dan FN, masing-masing 10 dan 15, dibandingkan SVM yang memiliki FP 15 dan FN 20. Hal ini menunjukkan bahwa Random Forest lebih efektif dalam mengidentifikasi siswa berprestasi dan tidak berprestasi dengan tingkat kesalahan yang lebih rendah. Model Random Forest mampu mengurangi jumlah kesalahan klasifikasi baik pada false positive maupun false negative dibanding SVM.

Oleh karena itu, model ini lebih dapat diandalkan sebagai alat bantu dalam pengambilan keputusan manajerial di lingkungan pendidikan.



Gambar 2. Confusion Matrix Model Random Forest

Pada Gambar 2 menampilkan confusion matrix dari model Random Forest yang digunakan untuk mengklasifikasikan prestasi akademik siswa ke dalam dua kelas: "Berprestasi" dan "Tidak Berprestasi." Pada sumbu vertikal (label aktual) terdapat kategori sebenarnya dari data siswa, sedangkan sumbu horizontal (label

prediksi) menunjukkan hasil klasifikasi model. Visualisasi confusion matrix untuk model Random Forest juga disajikan dalam untuk memperjelas distribusi prediksi yang benar dan salah pada masing-masing kelas.

Empat kotak utama pada matriks memiliki arti sebagai berikut:

- 1) True Positive (TP) di pojok kiri atas (nilai 115), yaitu jumlah siswa yang benar-benar berprestasi dan berhasil diprediksi dengan tepat oleh model sebagai berprestasi. Angka ini menunjukkan performa model dalam mengidentifikasi siswa berprestasi secara akurat.
- 2) False Positive (FP) di pojok kanan atas (nilai 10), yaitu jumlah siswa yang sebenarnya tidak berprestasi namun diprediksi salah oleh model sebagai berprestasi. Ini adalah kesalahan positif palsu yang dapat menyebabkan sumber daya pendidikan dialokasikan ke siswa yang tidak memerlukan intervensi.
- 3) False Negative (FN) di pojok kiri bawah (nilai 15), yaitu siswa yang sebenarnya berprestasi tetapi diprediksi salah sebagai tidak berprestasi. Kesalahan ini berisiko membuat siswa yang sebenarnya butuh dukungan tidak mendapatkan perhatian yang tepat.
- 4) True Negative (TN) di pojok kanan bawah (nilai 95), yaitu jumlah siswa yang benar-benar tidak berprestasi dan diprediksi dengan benar sebagai tidak berprestasi. Angka ini menunjukkan kemampuan model dalam mengidentifikasi siswa yang tidak membutuhkan intervensi.

Warna gradien biru pada matriks menggambarkan frekuensi prediksi, dengan warna lebih gelap menunjukkan jumlah yang lebih besar pada kotak tersebut, sehingga memudahkan visualisasi distribusi hasil klasifikasi.

Dengan melihat Gambar 2, dapat disimpulkan bahwa model Random Forest mampu mengklasifikasikan sebagian besar siswa dengan benar, terutama pada kategori siswa berprestasi (TP tinggi) dan tidak berprestasi (TN tinggi). Kesalahan prediksi relatif rendah, yang mengindikasikan keandalan model untuk mendukung pengambilan keputusan dalam manajemen pendidikan berbasis data.

Lebih lanjut, analisis feature importance pada model Random Forest mengidentifikasi variabel nilai semester sebagai faktor paling dominan dengan kontribusi sebesar 55,2%, diikuti oleh tingkat kehadiran sebesar 30,5%, dan latar belakang sosial ekonomi sebesar 14,3%. Data ini dapat dilihat pada Tabel 3 berikut:

Tabel 3. Feature Importance Variabel pada Model Random Forest

Variabel	Importance (%)
Nilai Semester	55.2
Tingkat Kehadiran	30.5
Latar Sosial Ekonomi	14.3

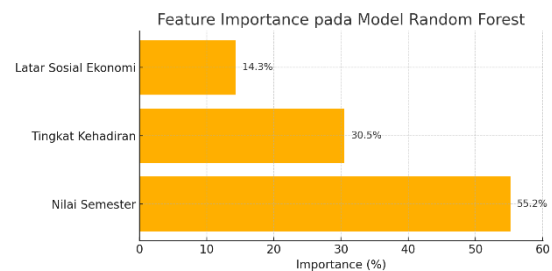
Pada Tabel 3 menampilkan feature importance atau tingkat kontribusi variabel-variabel input dalam model Random Forest terhadap prediksi prestasi akademik siswa. Nilai importance menunjukkan seberapa besar pengaruh masing-masing variabel dalam menentukan hasil klasifikasi model.

Dari tabel tersebut, variabel nilai semester memiliki kontribusi tertinggi sebesar 55,2%, yang berarti nilai ini menjadi faktor dominan dalam prediksi prestasi siswa. Variabel tingkat kehadiran menyumbang kontribusi sebesar 30,5%, menunjukkan bahwa kehadiran siswa juga sangat penting sebagai indikator keterlibatan dalam proses pembelajaran yang berpengaruh terhadap hasil belajar.

Sedangkan latar belakang sosial ekonomi memberikan kontribusi sebesar

14,3%, menandakan bahwa faktor lingkungan sosial ekonomi juga berperan, meskipun relatif lebih kecil dibanding variabel akademik dan kehadiran.

Informasi feature importance ini sangat berguna bagi manajemen pendidikan untuk fokus pada variabel-variabel yang paling signifikan dalam merancang intervensi dan kebijakan pendidikan yang efektif guna meningkatkan prestasi akademik siswa secara menyeluruh seperti yang divisualisasikan dalam Gambar 3



Gambar 3. Feature Importance Pada Model Random Forest

Gambar 3 memperlihatkan tingkat kontribusi masing-masing variabel input dalam mempengaruhi hasil prediksi prestasi akademik siswa oleh model Random Forest. Grafik batang horizontal ini menunjukkan seberapa besar pengaruh tiap fitur terhadap keputusan klasifikasi model.

Nilai Semester adalah variabel dengan kontribusi terbesar, sebesar 55,2%. Artinya, nilai semester siswa menjadi faktor paling dominan yang menentukan apakah siswa tersebut dikategorikan berprestasi atau tidak dalam prediksi model. Hal ini masuk akal karena nilai akademik langsung mencerminkan capaian belajar siswa.

Tingkat Kehadiran memiliki kontribusi sebesar 30,5%, menandakan bahwa frekuensi kehadiran siswa di kelas juga sangat berpengaruh terhadap prestasi akademik yang diprediksi. Kehadiran yang baik biasanya berkorelasi dengan

keterlibatan dan kesempatan belajar yang lebih banyak, sehingga berdampak positif pada prestasi.

Latar Sosial Ekonomi memberikan kontribusi sebesar 14,3%, yang walaupun lebih kecil dibanding dua variabel sebelumnya, tetap berperan penting. Faktor sosial ekonomi bisa memengaruhi akses dan dukungan siswa dalam proses belajar, sehingga masuk dalam variabel prediktor yang diperhitungkan model.

Dengan melihat proporsi kontribusi ini, manajemen pendidikan dapat memprioritaskan fokus intervensi atau kebijakan berdasarkan variabel-variabel yang paling berpengaruh. Misalnya, peningkatan kualitas pembelajaran dan monitoring nilai siswa perlu mendapat perhatian utama, diikuti oleh upaya menjaga dan meningkatkan tingkat kehadiran siswa. Sementara itu, faktor sosial ekonomi juga perlu diperhatikan, misalnya dengan menyediakan dukungan khusus bagi siswa dari latar belakang kurang beruntung.

Secara keseluruhan, Gambar 3 ini menjadi alat bantu visual yang efektif untuk memahami variabel mana saja yang paling penting dalam prediksi prestasi akademik menggunakan model Random Forest. Hal ini memudahkan pengambil kebijakan untuk merancang strategi yang lebih tepat sasaran dalam rangka meningkatkan kualitas pendidikan dan menurunkan risiko kegagalan belajar siswa.

Informasi ini sangat berguna dalam konteks manajemen pendidikan karena dapat menjadi dasar dalam merancang intervensi yang tepat sasaran, terutama dalam memperhatikan aspek akademik dan non-akademik siswa.

Hasil evaluasi ini menunjukkan bahwa penerapan model Random Forest dalam sistem informasi manajemen pendidikan dapat meningkatkan akurasi prediksi

prestasi akademik dan memberikan manfaat praktis bagi pengambilan keputusan manajerial. Kepala sekolah dan pengelola pendidikan dapat menggunakan output prediksi ini untuk mengidentifikasi siswa berisiko lebih awal, sehingga intervensi pembelajaran dapat dilakukan secara proaktif dan tepat waktu. Dengan demikian, penggunaan teknologi kecerdasan buatan ini berpotensi meningkatkan kualitas pembelajaran dan menurunkan angka kegagalan akademik.

KESIMPULAN

Berdasarkan hasil penelitian dan analisis yang telah dilakukan, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Model Random Forest menunjukkan performa yang lebih baik dibandingkan dengan Support Vector Machine (SVM) dalam memprediksi prestasi akademik siswa. Random Forest menghasilkan akurasi rata-rata sebesar 87,5%, presisi 85,7%, recall 83,9%, dan F1-score 84,8%, yang secara konsisten lebih unggul dibandingkan SVM pada seluruh metrik evaluasi.
2. Analisis confusion matrix mengindikasikan bahwa Random Forest mampu mengurangi kesalahan klasifikasi, baik false positive maupun false negative, sehingga lebih andal dalam mengidentifikasi siswa berprestasi maupun yang berisiko tidak berprestasi.
3. Variabel yang paling berkontribusi dalam prediksi prestasi siswa adalah nilai semester, diikuti oleh tingkat kehadiran, dan latar belakang sosial ekonomi. Hal ini menegaskan bahwa aspek akademik dan kehadiran siswa sangat penting dalam menentukan keberhasilan belajar.
4. Penerapan model Machine Learning, khususnya Random Forest, dapat

menjadi alat bantu yang efektif dalam manajemen pendidikan untuk mengidentifikasi siswa yang membutuhkan intervensi dini, sehingga strategi pembelajaran dapat disesuaikan secara lebih tepat dan preventif.

5. Integrasi sistem prediksi berbasis data ke dalam praktik manajerial sekolah memiliki potensi besar untuk meningkatkan kualitas pendidikan dan mengurangi angka kegagalan akademik melalui pengambilan keputusan berbasis bukti yang lebih cepat dan akurat.

DAFTAR PUSTAKA

- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>.
- Chien, C.-C., Chen, H.-H., & Lin, C.-T. (2020). Application of machine learning techniques in education: Predicting student academic performance. *Journal of Educational Technology*, 37(4), 123-135.
- Fernandez-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, 15, 3133-3181.
- Fosnot, C. T. (2013). *Constructivism: Theory, perspectives, and practice* (2nd ed.). Teachers College Press.
- Leithwood, K., Sun, J., & Pollock, K. (2017). How school leadership influences student learning: A review of research for educational leaders. *Journal of Educational Administration*, 55(2), 225-254.
- Li, J., & Qiao, M. (2019). Parameter optimization of SVM based on improved PSO algorithm for classification problems. *Applied Soft Computing*, 79, 323-335. <https://doi.org/10.1016/j.asoc.2019.03.033>.
- Mandinach, E. B., & Gummer, E. S. (2016). Data-driven decision making in education: Challenges and opportunities. *Teachers College Record*, 118(9), 1-28.
- Pratiwi, H. (2019). Perbandingan algoritma ELM dan Backpropagation terhadap prestasi akademik mahasiswa. *J-SAKTI (Jurnal Sains Komputer dan Teknologi Informasi)*, 9(2), 45-53.
- Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1249. <https://doi.org/10.1002/widm.1249>.
- Schunk, D. H., & DiBenedetto, M. K. (2020). Motivation and social-emotional learning: Theory, research, and practice. *Contemporary Educational Psychology*, 60, 101821.
- Xu, R., & Wunsch, D. (2018). Support vector machine. In *Clustering* (pp. 165-192). Wiley.