E-ISSN: 3025-0994
Halaman 4258-4266
Volume 4 Nomor 4 Tahun 2026

Implementasi Data Mining Penentuan Prioritas Program Pelatihan SKB Lebak Menggunakan Random Forest Berbasis Web

Rangga Maulana[1], Kussigit Santosa[2]

[1,2] Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan
e-mail: ranggamlana15@gmail.com , dosen00202@unpam.ac.id

Abstrak

Sanggar Kegiatan Belajar Kabupaten Lebak menyelenggarakan berbagai program pelatihan bagi masyarakat, tetapi penentuan prioritas programnya masih dilakukan secara manual sehingga berpotensi menimbulkan subjektivitas dalam pengambilan keputusan. Penelitian ini bertujuan menerapkan algoritma Random Forest untuk mengklasifikasikan prioritas program pelatihan ke dalam kategori tinggi, sedang, dan rendah melalui sistem berbasis web. Metode yang digunakan adalah Knowledge Discovery in Databases yang terdiri atas seleksi data, praproses, transformasi, penambangan data, dan evaluasi terhadap 920 data program dan peserta pelatihan dengan sembilan atribut, yaitu usia, jenis kelamin, pendidikan terakhir, pekerjaan, status ekonomi, kategori program, jumlah pendaftar, kuota, dan durasi pelatihan. Sistem dikembangkan menggunakan Python, Flask, dan MySQL, sedangkan kinerja model dievaluasi dengan confusion matrix, accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model memperoleh accuracy 78,35%, precision 78,76%, recall 78,35%, dan F1-score 77,14%. Kinerja model baik pada kelas sedang dan tinggi, tetapi masih lemah pada kelas rendah dengan recall 33,33%. Jumlah pendaftar menjadi atribut yang paling berpengaruh dengan nilai feature importance 0,202, diikuti usia dan status ekonomi. Pengujian black box terhadap empat belas skenario menunjukkan seluruh fungsi sistem berjalan sesuai rancangan. Penelitian ini dilakukan pada satu lembaga sehingga hasilnya perlu dikembangkan dengan data yang lebih banyak. Simpulan penelitian ini adalah bahwa Random Forest dapat dijadikan informasi pendukung bagi pengelola dalam menentukan prioritas program pelatihan secara lebih objektif dan berbasis data.

Kata Kunci: Data Mining, Random Forest, Prioritas Program Pelatihan, Pendidikan Nonformal, Sistem Berbasis Web

Abstract

Sanggar Kegiatan Belajar of Lebak Regency provides various training programs for the community, yet the priority of these programs was still determined manually, which could introduce subjectivity into decision making. This study aimed to apply the Random Forest algorithm to classify training program priority into high, medium, and low categories through a web-based system. The Knowledge Discovery in Databases method was used, consisting of data selection, preprocessing, transformation, data mining, and evaluation on 920 records of training programs and participants with nine attributes, namely age, gender, last education, occupation, economic status, program category, number of applicants, quota, and program duration. The system was developed using Python, Flask, and MySQL, and the model performance was evaluated using a confusion matrix, accuracy, precision, recall, and F1-score. The results show that the model achieves an accuracy of 78.35%, a precision of 78.76%, a recall of 78.35%, and an F1-score of 77.14%. The model performs well on the medium and high classes but remains weak on the low class, with a recall of 33.33%. The number of applicants is the most influential attribute with a feature importance value of 0.202, followed by age and economic status. Black box testing on fourteen scenarios shows that all system functions work as designed. This study was conducted at a single institution, so the

results need to be extended with more data. In conclusion, Random Forest can serve as supporting information for managers in determining training program priorities more objectively and based on data.

Keywords: Data Mining, Random Forest, Training Program Priority, Non-Formal Education, Web-Based System

LATAR BELAKANG

Sanggar Kegiatan Belajar (SKB) merupakan satuan pendidikan nonformal (Kementerian Pendidikan dan Kebudayaan, 2016) yang menyelenggarakan program pelatihan bagi masyarakat. Di SKB Kabupaten Lebak, jumlah program yang dapat dijalankan setiap tahun dibatasi oleh anggaran, sarana, instruktur, dan waktu, sementara kebutuhan masyarakat terhadap pelatihan beragam dan terus berkembang. Kondisi ini menuntut mekanisme penentuan prioritas program yang objektif, sistematis, dan berbasis data.

Selama ini prioritas program ditentukan secara manual melalui rapat internal dengan mempertimbangkan pengalaman tahun sebelumnya dan persepsi pengelola. Cara ini berpotensi menimbulkan bias dan kurang memanfaatkan data historis peserta dan program yang sebenarnya telah tersedia. Pendekatan pengambilan keputusan berbasis data (*data-driven decision making*) melalui teknik *data mining* dapat menjadi solusi untuk memetakan prioritas pelatihan secara lebih objektif.

Knowledge Discovery in Databases (KDD) adalah proses menemukan pengetahuan dari kumpulan data melalui lima tahap, yaitu *data selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*. Pemilihan fitur (*feature selection*) merupakan tahap yang lazim dalam persiapan data untuk klasifikasi (Pintas dkk., 2021). Pada tahap *data mining*, algoritma *Random Forest* dipilih karena merupakan metode *ensemble learning* yang membangun banyak pohon keputusan dari sampel data yang diambil secara acak (*bootstrap aggregating* atau *bagging*), lalu menggabungkan prediksi seluruh pohon dengan pemungutan suara mayoritas pada kasus klasifikasi (Sarker, 2021). Setiap pohon dilatih dengan subset fitur acak sehingga struktur antarpohon berbeda dan risiko *overfitting* menurun. Prediksi akhir ditunjukkan pada Persamaan (1).

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \ldots, h_n(x)\} \quad (1)$$

dengan $\hat{y}$ adalah kelas hasil prediksi, $h_i(x)$ adalah prediksi pohon ke-*i*, dan *n* adalah jumlah pohon. Algoritma ini juga menghasilkan nilai *feature importance* yang menunjukkan kontribusi relatif tiap atribut terhadap prediksi (Musolf dkk., 2022). Pengembangan algoritma ini terus dilakukan, misalnya melalui optimasi berbasis Spark (Wang dkk., 2022) dan kombinasi dengan *Particle Swarm Optimization* untuk prediksi kelulusan tepat waktu (Rahman dkk., 2025).

Random Forest telah banyak diterapkan di bidang pendidikan. Andrianof dkk. (2025) memperoleh akurasi 87,5% pada prediksi kelulusan mahasiswa, sedangkan Putra dan Harahap (2024) memperoleh akurasi 88% dan mengimplementasikan model pada aplikasi web. Afandi dkk. (2025) melaporkan akurasi 97,50% pada pembagian data 80:20, lebih tinggi daripada *Decision Tree* dan *Naïve Bayes*, dan Aprilia dkk. (2026) menerapkan *Random Forest* pada klasifikasi predikat kelulusan mahasiswa dengan tahap pembersihan dan transformasi data serta evaluasi menggunakan *confusion matrix*. Amin dkk. (2025) serta Arrohman dkk. (2025) melaporkan bahwa *Random Forest* mengungguli *K-Nearest Neighbors* pada data kinerja dan kelulusan siswa, sedangkan Saputri dkk. (2025) membangun aplikasi web berbasis Python dan Flask untuk memprediksi prestasi akademik siswa dengan akurasi 87,40% pada 136 data uji. Pratama dkk. (2023) menaikkan akurasi klasifikasi diabetes dari 78,2% menjadi 82,1% melalui optimasi, Gunawan dan Paputungan (2025) menerapkan *preprocessing* berupa penghapusan data tidak relevan dan *encoding* variabel kategorikal sebelum pemodelan, Hidayah dkk. (2025) menerapkan *hyperparameter tuning* pada prediksi gaya belajar berbasis web, dan Nurhafiz dkk. (2026) menerapkan *Random Forest* untuk mengukur kepuasan peserta pelatihan guru.

Namun, penelitian-penelitian tersebut belum membahas penentuan prioritas program pelatihan pada lembaga pendidikan nonformal seperti SKB. Masalah penelitian ini adalah bagaimana mengolah data peserta dan program pelatihan SKB Lebak menjadi dataset yang dapat diklasifikasikan, bagaimana menerapkan *Random Forest* untuk menentukan prioritas program, dan bagaimana membangun sistem berbasis web yang mengintegrasikan proses tersebut. Pemecahan masalah dilakukan melalui tahapan KDD yang diimplementasikan pada sistem web berbasis Flask. Penelitian ini bertujuan menyusun dataset terstruktur, mengklasifikasikan prioritas program pelatihan ke dalam kategori Tinggi, Sedang, dan Rendah dengan *Random Forest*, serta membangun sistem yang mengintegrasikan pengolahan data, pelatihan dan evaluasi model, dan prediksi prioritas. Hasil klasifikasi diposisikan sebagai informasi pendukung, bukan pengganti keputusan pengelola SKB Lebak.

METODE

Penelitian ini menggunakan pendekatan kuantitatif karena melibatkan pengolahan data yang dapat diukur secara numerik. Data berasal dari dokumentasi dan arsip pelaksanaan program pelatihan SKB Kabupaten Lebak dengan jumlah 920 data. Pengumpulan data dilakukan melalui dokumentasi untuk memperoleh data peserta dan program, observasi untuk memahami proses pengelolaan program, wawancara dengan pihak terkait untuk menggali kebutuhan sistem, dan studi literatur untuk memperoleh landasan teori dan penelitian terdahulu. Sembilan atribut digunakan sebagai variabel prediktor dan atribut prioritas sebagai variabel target dengan tiga kelas, seperti ditunjukkan pada Tabel 1, sedangkan contoh tiga data ditampilkan pada Tabel 2.

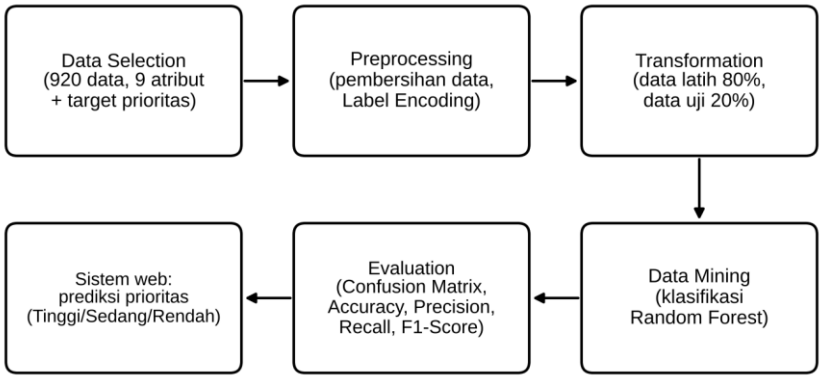
Tabel 1. Variabel penelitian

No	Variabel	Jenis	Keterangan
1	Usia	Numerik	Umur peserta (tahun)
2	Jenis kelamin	Kategorikal	Laki-laki atau perempuan
3	Pendidikan terakhir	Kategorikal	SD, SMP, SMA, SMK
4	Pekerjaan	Kategorikal	Pelajar, mahasiswa, bekerja, belum bekerja
5	Status ekonomi	Kategorikal	Rendah, sedang, tinggi
6	Kategori program	Kategorikal	Komputer, menjahit, desain grafis, tata boga, otomotif, dll.
7	Jumlah pendaftar	Numerik	Peserta yang mendaftar pada program
8	Kuota	Numerik	Kapasitas peserta program
9	Durasi (bulan)	Numerik	Lama pelaksanaan program
10	Prioritas (target)	Kategorikal	Tinggi, Sedang, Rendah

Tabel 2. Contoh data penelitian

Atribut	Data 1	Data 2	Data 3
Usia	20	25	30
Jenis kelamin	L	P	P
Pendidikan terakhir	SMA	SMA	SMA
Pekerjaan	Belum bekerja	Belum bekerja	Wirausaha
Status ekonomi	Sedang	Rendah	Sedang
Kategori program	Komputer	Menjahit	Tata rias
Jumlah pendaftar	120	60	75
Kuota	30	20	25
Durasi (bulan)	3	2	3
Prioritas	Tinggi	Rendah	Sedang

Penelitian mengikuti tahapan KDD pada Gambar 1. Pada *data selection* dipilih atribut yang relevan. Tahap *preprocessing* meliputi penghapusan data duplikat, penanganan *missing value*, serta *Label Encoding* pada atribut kategorikal dan variabel target (Rendah = 0, Sedang = 1, Tinggi = 2). Pada *transformation*, dataset disusun menjadi matriks fitur dan target, lalu dibagi menjadi 80% data latih dan 20% data uji. Model *Random Forest* dibangun dari data latih pada tahap *data mining*, kemudian diuji dengan data uji pada tahap *evaluation*.



Gambar 1. Alur penelitian berbasis KDD

Kinerja klasifikasi diukur menggunakan *confusion matrix* (Heydarian dkk., 2022) yang memuat jumlah data yang diklasifikasikan benar dan salah pada tiap kelas, yaitu *true positive* (TP), *false positive* (FP), *false negative* (FN), dan *true negative* (TN). *Accuracy* adalah proporsi data yang diklasifikasikan benar terhadap seluruh data uji, sedangkan *precision*, *recall*, dan *F1-score* dihitung untuk tiap kelas dengan Persamaan (2) sampai (4).

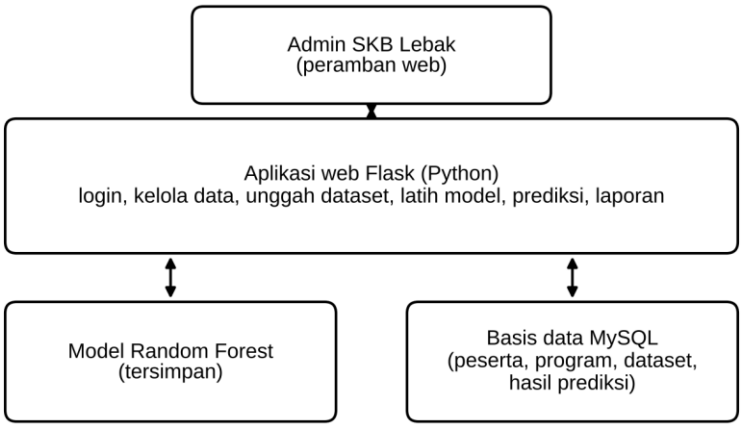
$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F1} - \text{score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{4}$$

Nilai keseluruhan diperoleh sebagai rata-rata terbobot menurut jumlah data tiap kelas.

Sistem dikembangkan dengan model *Waterfall* yang mencakup analisis kebutuhan, perancangan, implementasi, pengujian, dan pemeliharaan. Pada proses yang berjalan, data peserta dan informasi program direkap sebagai bahan pertimbangan tanpa model klasifikasi, dan pengolahan data, pembentukan model, evaluasi, serta prediksi belum terintegrasi. Perancangan sistem usulan menggunakan *flowchart* dan diagram *Unified Modeling Language* berupa *use case diagram*, *activity diagram*, dan *sequence diagram* (Wayahdi & Ruziq, 2023), *entity relationship diagram* untuk perancangan basis data (Afiifah dkk., 2022), serta rancangan antarmuka. Arsitektur sistem ditunjukkan pada Gambar 2. Implementasi menggunakan bahasa Python dengan *framework* Flask (Darmawan & Iswari, 2022) dan basis data MySQL. Fungsi sistem diuji dengan *black box testing* untuk memastikan setiap fitur menghasilkan keluaran sesuai yang diharapkan, dan dengan *white box testing* pada bagian kode yang memuat kondisi validasi.



Gambar 2. Arsitektur sistem

HASIL DAN PEMBAHASAN

Sistem yang dihasilkan memiliki fitur *login* admin, dasbor, pengelolaan data peserta dan program pelatihan, unggah *dataset*, pelatihan model, prediksi prioritas, riwayat prediksi, serta laporan yang dapat dicetak. Admin mengunggah *dataset*, menjalankan pelatihan model dengan menentukan jumlah pohon (*n_estimators*) dan porsi data uji, kemudian melihat hasil evaluasi. Model yang telah dilatih disimpan sehingga prediksi dapat dilakukan tanpa pelatihan ulang. Pada halaman prediksi, pengguna mengisi sembilan atribut dan sistem menampilkan kategori prioritas Tinggi, Sedang, atau Rendah.

Hasil klasifikasi pada data uji ditunjukkan oleh *confusion matrix* pada Tabel 3. Sebanyak 152 data diklasifikasikan dengan benar, yaitu 8 data kelas Rendah, 89 data kelas Sedang, dan 55 data kelas Tinggi.

Tabel 3. *Confusion matrix* model *Random Forest*

Aktual \ Prediksi	Rendah	Sedang	Tinggi
Rendah	8	16	0
Sedang	2	89	11
Tinggi	0	13	55

Nilai TP, FP, FN, dan TN tiap kelas yang diturunkan dari Tabel 3 ditunjukkan pada Tabel 4. Sebagai contoh, pada kelas Rendah terdapat 8 data yang benar diprediksi Rendah (TP), 2 data kelas lain yang salah diprediksi Rendah (FP), dan 16 data Rendah yang salah diprediksi sebagai kelas lain (FN).

Tabel 4. Nilai TP, FP, FN, dan TN per kelas

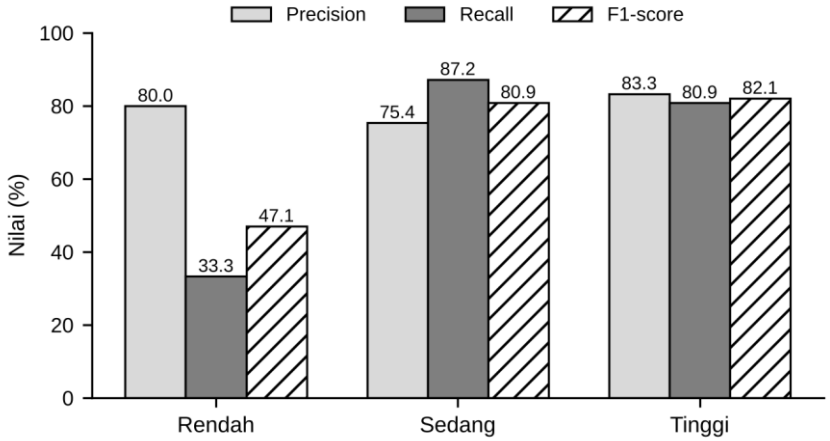
Kelas	TP	FP	FN	TN
Rendah	8	2	16	168
Sedang	89	29	13	63
Tinggi	55	11	13	115

Berdasarkan Tabel 3 dan Tabel 4, model memperoleh *accuracy* 78,35%, *precision* 78,76%, *recall* 78,35%, dan *F1-score* 77,14%. Nilai *precision*, *recall*, dan *F1-score* tersebut merupakan rata-rata terbobot menurut jumlah data tiap kelas, sesuai hasil perhitungan ulang dari Tabel 3. Kinerja tiap kelas dirinci pada Tabel 5.

Tabel 5. Kinerja model per kelas (%)

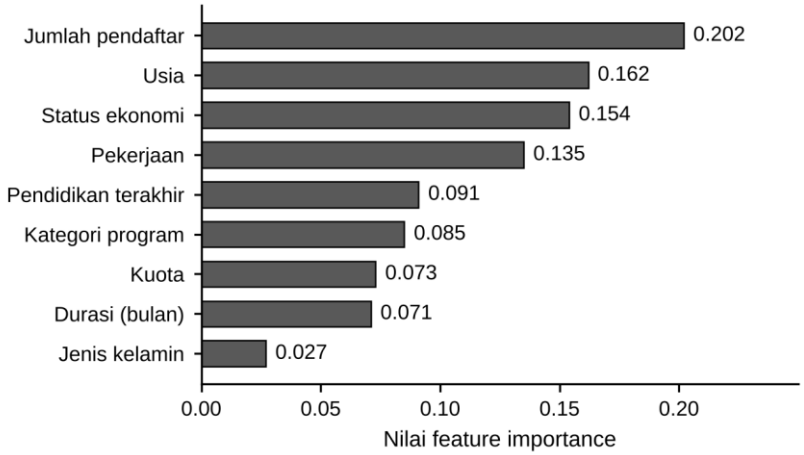
Kelas	Precision	Recall	F1-score	Data
Rendah	80,00	33,33	47,06	24
Sedang	75,42	87,25	80,91	102
Tinggi	83,33	80,88	82,09	68
Rata-rata terbobot	78,76	78,35	77,14	194

Kelas Sedang dan Tinggi dikenali dengan baik (*recall* 87,25% dan 80,88%), sedangkan *recall* kelas Rendah hanya 33,33% karena 16 dari 24 data kelas Rendah diprediksi sebagai Sedang. Kelas Rendah memiliki jumlah data paling sedikit sehingga model kemungkinan kurang mendapat contoh untuk mempelajari polanya. Dugaan ini perlu diuji lebih lanjut dengan penambahan data atau teknik penyeimbangan data. Perbandingan ketiga metrik pada tiap kelas divisualisasikan pada Gambar 3.



Gambar 3. Perbandingan *precision*, *recall*, dan *F1-score* tiap kelas

Gambar 4 menunjukkan bahwa jumlah pendaftar merupakan atribut paling berpengaruh (0,202), diikuti usia (0,162), status ekonomi (0,154), dan pekerjaan (0,135), sedangkan jenis kelamin paling kecil (0,027). Temuan ini menunjukkan bahwa tingkat minat masyarakat yang tercermin pada jumlah pendaftar berperan penting dalam penentuan prioritas program. Perlu dicatat bahwa *feature importance* menunjukkan besarnya kontribusi atribut, bukan arah hubungannya terhadap kelas prioritas.



Gambar 4. Nilai *feature importance* atribut

Pengujian *black box* dilakukan pada 14 skenario yang mencakup *login*, pengelolaan data peserta dan program, unggah *dataset*, pelatihan dan evaluasi model, prediksi, riwayat

prediksi, laporan, dan *logout*, seperti ditunjukkan pada Tabel 6. Seluruh skenario menghasilkan keluaran sesuai dengan yang diharapkan.

Tabel 6. Ringkasan pengujian *black box*

ID	Skenario pengujian	Hasil
TC-01	*Login* dengan data valid	Berhasil
TC-02	*Login* dengan data tidak valid	Berhasil
TC-03	Tambah data peserta	Berhasil
TC-04	Ubah data peserta	Berhasil
TC-05	Hapus data peserta	Berhasil
TC-06	Tambah program pelatihan	Berhasil
TC-07	Unggah *dataset*	Berhasil
TC-08	Pelatihan model *Random Forest*	Berhasil
TC-09	Tampil evaluasi model	Berhasil
TC-10	Prediksi prioritas program	Berhasil
TC-11	Tampil riwayat prediksi	Berhasil
TC-12	Tampil laporan	Berhasil
TC-13	Cetak laporan	Berhasil
TC-14	*Logout*	Berhasil

Pengujian *white box* dilakukan pada bagian kode yang memuat kondisi validasi, yaitu *login*, unggah *dataset*, pelatihan model, dan prediksi. Pada *login*, admin diarahkan ke dasbor apabila data sesuai dan menerima pesan kesalahan apabila tidak sesuai. Pada unggah *dataset*, sistem memeriksa keberadaan file, format .csv, dan kelengkapan kolom wajib sebelum data diproses.

Temuan tersebut menjawab masalah penelitian. Data peserta dan program SKB Lebak berhasil disusun menjadi dataset dengan sembilan atribut dan satu variabel target, *Random Forest* mampu mengklasifikasikan prioritas program dengan *accuracy* 78,35%, dan seluruh proses terintegrasi dalam sistem web yang fungsinya berjalan sesuai rancangan. Nilai *accuracy* tersebut lebih rendah daripada hasil sejumlah penelitian terdahulu pada prediksi kelulusan dan prestasi akademik, yaitu 87,40% hingga 97,50% (Andrianof dkk., 2025; Putra & Harahap, 2024; Saputri dkk., 2025; Afandi dkk., 2025). Perbedaan tersebut wajar karena objek, atribut, jumlah data, dan label yang digunakan berbeda sehingga angkanya tidak dapat dibandingkan secara langsung. Nilai ini justru mendekati akurasi *Random Forest* sebelum optimasi pada klasifikasi diabetes (78,2%) yang kemudian meningkat menjadi 82,1% setelah dioptimasi (Pratama dkk., 2023), sehingga penyetelan parameter layak dicoba pada data SKB Lebak. Perbandingan dengan penelitian terdahulu disajikan pada Tabel 7.

Tabel 7. Perbandingan akurasi dengan penelitian terdahulu

Peneliti	Objek penelitian	Akurasi (%)
Andrianof dkk. (2025)	Kelulusan mahasiswa	87,5
Putra & Harahap (2024)	Kelulusan mahasiswa	88
Afandi dkk. (2025)	Kelulusan mahasiswa	97,50
Saputri dkk. (2025)	Prestasi akademik siswa	87,40
Pratama dkk. (2023)	Klasifikasi diabetes	78,2 (82,1 setelah optimasi)
Penelitian ini	Prioritas program pelatihan	78,35

Pola kesalahan pada Tabel 3 menunjukkan bahwa tidak ada data kelas Rendah yang diprediksi Tinggi maupun sebaliknya. Seluruh kesalahan terjadi antarkelas yang bersebelahan, terutama dari Rendah ke Sedang (16 data), sehingga model jarang melakukan kesalahan ekstrem. Komposisi data uji terdiri atas 24 data Rendah (12,4%), 102 data Sedang

(52,6%), dan 68 data Tinggi (35,1%), sehingga model kemungkinan condong ke kelas mayoritas. Karena label prioritas berasal dari data historis, model mempelajari pola penetapan prioritas yang sudah ada, dan kualitas label tersebut turut memengaruhi hasil. Dominasi jumlah pendaftar (0,202) menunjukkan bahwa tingkat minat masyarakat menjadi sinyal terkuat bagi model, sedangkan usia, status ekonomi, dan pekerjaan dengan nilai 0,135 hingga 0,162 menunjukkan bahwa karakteristik peserta turut membentuk klasifikasi.

Sistem yang dibangun menyatukan pengolahan data, pelatihan model, evaluasi, dan prediksi dalam satu antarmuka web sehingga admin tidak perlu menjalankan kode program. Karena masih ada data yang salah diklasifikasikan, terutama pada kelas Rendah, hasil prediksi digunakan sebagai informasi pendukung, sedangkan keputusan akhir tetap berada pada pengelola SKB Lebak. Penelitian ini juga terbatas pada satu lembaga dan satu algoritma tanpa pembandingan dengan algoritma klasifikasi lain.

SIMPULAN

Penelitian ini berhasil mengimplementasikan *Random Forest* pada sistem berbasis web untuk mengklasifikasikan prioritas program pelatihan SKB Lebak dengan *accuracy* 78,35% dan *F1-score* 77,14%, serta jumlah pendaftar sebagai atribut paling berpengaruh. Model baik pada kelas Sedang dan Tinggi, tetapi lemah pada kelas Rendah. Peneliti selanjutnya disarankan menambah data kelas Rendah, menyeimbangkan kelas, menyetel parameter, dan membandingkan algoritma lain, sedangkan pengelola SKB Lebak disarankan menggunakan hasil prediksi sebagai bahan pertimbangan, bukan satu-satunya dasar keputusan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Kepala beserta seluruh jajaran SKB Kabupaten Lebak yang telah memberikan izin penelitian dan menyediakan data, serta kepada Program Studi Teknik Informatika Universitas Pamulang atas dukungan selama penelitian ini.

DAFTAR PUSTAKA

Afandi, M.C., Indahyanti, U., Setiawan, H. & Kautsar, I.A. 2025. Perbandingan Algoritma Machine Learning Dalam Memprediksi Kelulusan Mahasiswa. *Semantik: Teknik Informasi*, 11(2). Https://Doi.Org/10.55679/Semantik.V11i2.154

Afiifah, K., Azzahra, Z.F. & Anggoro, A.D. 2022. Analisis Teknik Entity-Relationship Diagram Dalam Perancangan Database: Sebuah Literature Review. *INTECH*, 3(1): 8–11. Https://Doi.Org/10.54895/Intech.V3i1.1261

Amin, M.S., Astrida, D.N. & Sekar, E. 2025. Analisis Prediktif Terhadap Kinerja Siswa Dalam Ujian Menggunakan Algoritma Random Forest Dan K-Nearest Neighbors. *Jurnal Teknologi Pendidikan*, 3(1). Https://Doi.Org/10.47134/Jtp.V3i1.1835

Andrianof, H., Gusman, A.P. & Putra, O.A. 2025. Implementasi Algoritma Random Forest Untuk Prediksi Kelulusan Mahasiswa Berdasarkan Data Akademik: Studi Kasus Di Perguruan Tinggi Indonesia. *Jurnal Sains Informatika Terapan*, 4(1): 24–28. Https://Doi.Org/10.62357/Jsit.V4i2.464

Aprilia, K., Rosihan & Abdullah, S.D. 2026. Penerapan Algoritma Random Forest Untuk Klasifikasi Predikat Kelulusan Mahasiswa Universitas Khairun. *JATI (Jurnal Jaringan Dan Teknologi Informasi)*, 4(2): 23–30.

Arrohman, M.R., Khoiriyah, N., Fawaida, A. & Fithri, D.L. 2025. Analisis Klasifikasi Kelulusan Siswa Menggunakan Metode KNN Random Forest Pada Dataset Students Performance. *Jurnal Mahasiswa Sistem Informasi*, 6(2). Https://Doi.Org/10.24127/Jmsi.V6i2.9102

Darmawan, A.C. & Iswari, L. 2022. Pengembangan Aplikasi Berbasis Web Dengan Python Flask Untuk Klasifikasi Data Menggunakan Metode Decision Tree C4.5. *Jurnal Pendidikan Dan Konseling*, 4(5): 5351–5362. Https://Doi.Org/10.31004/Jpdk.V4i5.7492

Gunawan, P.H. & Paputungan, I.V. 2025. Detection Of Graduation Potential In Prospective Students Using The Random Forest Algorithm. *Sistemasi: Jurnal Sistem Informasi*, 14(5): 2055–2065.

Heydarian, M., Doyle, T.E. & Samavi, R. 2022. MLCM: Multi-Label Confusion Matrix. *IEEE Access*, 10: 19083–19095. Https://Doi.Org/10.1109/ACCESS.2022.3151048

Hidayah, R.B., Latifah, K. & Herlambang, B.A. 2025. Prediksi Gaya Belajar Menggunakan Algoritma Random Forest Dengan Optimasi Hyperparameter Tuning Berbasis Web. *RABIT: Jurnal Teknologi Dan Sistem Informasi Univrab*, 10(2): 864–873. Https://Doi.Org/10.36341/Rabit.V10i2.6393

Kementerian Pendidikan Dan Kebudayaan Republik Indonesia. 2016. *Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 4 Tahun 2016 Tentang Alih Fungsi Sanggar Kegiatan Belajar Menjadi Satuan Pendidikan Nonformal Sejenis*.

Musolf, A.M., Holzinger, E.R., Malley, J.D. Dkk. 2022. What Makes A Good Prediction? Feature Importance And Beginning To Open The Black Box Of Machine Learning In Genetics. *Human Genetics*, 141: 1515–1528. Https://Doi.Org/10.1007/S00439-021-02402-Z

Nurhafiz, R.A., Marleny, F.D. & Ningrum, A.A. 2026. Optimasi Algoritma Random Forest Dalam Mengukur Kepuasan Peserta Pelatihan Guru Pada Lembaga HAFECS. *Jurnal Media Informatika*, 7(1): 302–311. Https://Doi.Org/10.55338/Jumin.V7i1.8314

Pintas, J.T., Fernandes, L.A.F. & Garcia, A.C.B. 2021. Feature Selection Methods For Text Classification: A Systematic Literature Review. *Artificial Intelligence Review*, 54: 6149–6200. Https://Doi.Org/10.1007/S10462-021-09970-6

Pratama, P.F., Rahmadani, N., Nahampun, H., Harmutika, S., Rahmadeyan, A. & Evizal. 2023. Random Forest Optimization Using Particle Swarm Optimization For Diabetes Classification. *Public Research Journal Of Engineering, Data Technology And Computer Science*, 1(1). Https://Doi.Org/10.57152/Predatecs.V1i1.809

Putra, M. & Harahap, E. 2024. Machine Learning Pada Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Random Forest. *Jurnal Riset Matematika*, 4(2): 127–136. Https://Doi.Org/10.29313/Jrm.V4i2.5102

Rahman, A., Mahdiana, D. & Fauzi, A. 2025. Predicting Student On-Time Graduation Using Particle Swarm Optimization And Random Forest Algorithms. *Indonesian Journal Of Artificial Intelligence And Data Mining*, 8(1): 161–173. Https://Doi.Org/10.24014/Ijaidm.V8i1.33577

Saputri, R.A., Asrianda & Rosnita, L. 2025. A Random Forest-Based Predictive Model For Student Academic Performance: A Case Study In Indonesian Public High Schools. *Journal Of Applied Informatics And Computing*, 9(3): 1042–1049.

Sarker, I.H. 2021. Machine Learning: Algorithms, Real-World Applications And Research Directions. *SN Computer Science*, 2(3): 160. Https://Doi.Org/10.1007/S42979-021-00592-X

Wang, S., Zhang, Z., Geng, S. & Pang, C. 2022. Research On Optimization Of Random Forest Algorithm Based On Spark. *Computers, Materials & Continua*, 71(2). Https://Doi.Org/10.32604/Cmc.2022.015378

Wayahdi, M.R. & Ruziq, F. 2023. Pemodelan Sistem Penerimaan Anggota Baru Dengan Unified Modeling Language (UML) (Studi Kasus: Programmer Association Of Battuta). *Jurnal Minfo Polgan*, 12(1). Https://Doi.Org/10.33395/Jmp.V12i1.12870