

Klasifikasi Sekolah Potensial untuk Promosi Kampus UMBJM Menggunakan Random Forest

Ahmad Aqli¹, Finki Dona Marleny², Windarsyah³

^{1,2,3}Teknik Informatika, Universitas Muhammadiyah Banjarmasin, Barito Kuala, Indonesia

Email: ¹azure2757@gmail.com

Email Penulis Korespondensi: ¹1955201110001@umbjm.ac.id, ²finkidona@umbjm.ac.id, ³windarsyah@umbjm.ac.id

Abstrak—Setiap perguruan tinggi melakukan promosi rutin setiap tahun untuk menarik minat calon mahasiswa baru. Selain memasang spanduk di berbagai lokasi dalam kota, promosi juga dapat dilakukan langsung ke sekolah-sekolah yang dianggap potensial. Penelitian ini bertujuan mengidentifikasi sekolah asal mahasiswa yang memiliki potensi sebagai target promosi kampus Universitas Muhammadiyah Banjarmasin (UMBJM). Metode klasifikasi dengan pendekatan *supervised learning* digunakan dalam penelitian ini, dengan algoritma *Random Forest* yang diterapkan pada data pendaftaran mahasiswa UMBJM tahun 2019–2024. Setelah melalui proses filtering dan preprocessing data, model dilatih untuk menghasilkan prediksi dengan akurasi sebesar 81.60%. Hasil penelitian ini diharapkan dapat membantu pihak kampus dalam merancang strategi promosi yang lebih efisien dan tepat sasaran.

Kata Kunci: Klasifikasi, Random Forest, Promosi Kampus, Sekolah Potensial, Data Pendaftaran Mahasiswa

Abstract- Every university conducts annual promotional activities to attract prospective new students. In addition to placing banners in various locations around the city, promotions can also be carried out directly to schools considered to have potential. This study aims to identify the origin schools of students that have the potential to be targeted for campus promotion by Universitas Muhammadiyah Banjarmasin (UMBJM). A classification method with a supervised learning approach was used in this study, applying the Random Forest algorithm to UMBJM student registration data from 2019 to 2024. After data filtering and preprocessing, the model was trained to achieve predictions with an accuracy of 81,60%. The results of this study are expected to help the university design more efficient and targeted promotional strategies.

Keywords: Classification, Random Forest, Campus Promotion, Potential Schools, Student Registration Data

1. PENDAHULUAN

Penerimaan mahasiswa baru merupakan salah satu aspek krusial dalam pengelolaan perguruan tinggi. Proses ini tidak hanya memengaruhi jumlah pendaftar yang diterima, tetapi juga keragaman dan kualitas akademik calon mahasiswa. Di tengah persaingan yang semakin ketat antarperguruan tinggi, strategi promosi kampus harus dirancang secara tepat sasaran agar mampu menjangkau calon mahasiswa potensial. Salah satu pendekatan yang dapat diterapkan adalah dengan menganalisis data historis pendaftaran mahasiswa baru, terutama informasi mengenai asal sekolah mereka. Data ini memiliki potensi strategis dalam membantu perguruan tinggi mengenali sekolah-sekolah yang secara konsisten mengirimkan lulusannya untuk mendaftar. Penggunaan data historis ini menjadi semakin penting karena dapat memberikan gambaran yang lebih komprehensif tentang pola-pola yang terjadi dalam proses pendaftaran mahasiswa baru, serta membantu perguruan tinggi dalam merencanakan dan mengalokasikan sumber daya secara lebih efisien. Dalam praktiknya, masih banyak perguruan tinggi yang melakukan promosi tanpa strategi berbasis data atau hanya berdasarkan intuisi, sehingga berisiko menghamburkan sumber daya promosi tanpa hasil optimal. Oleh karena itu, pemanfaatan data historis sebagai dasar dalam pengambilan keputusan promosi menjadi semakin relevan, terutama dalam era digital dan data-driven seperti saat ini. Selain itu, penelitian sebelumnya juga menunjukkan bahwa teknik K-Means Clustering dapat digunakan secara efektif untuk menentukan strategi promosi mahasiswa baru berdasarkan data penerimaan, seperti yang dilakukan oleh Agneresaa et al. [1]

Berbagai penelitian sebelumnya telah membuktikan bahwa teknik data mining dan machine learning memiliki kemampuan yang kuat dalam menganalisis data pendidikan. Pratama et al. [2], misalnya, menerapkan algoritma Random Forest untuk klasifikasi pemilihan jurusan berdasarkan riwayat nilai siswa, lalu dilaur lingkungan pendidikan juga digunakan dalam klasifikasi produk seperti yang dilakukan oleh Pratiwi dan Nugroho [3]. Sementara itu, Rahmalinda dan Jananto [4] menggunakan metode K-Means Clustering untuk menentukan strategi promosi berdasarkan data penerimaan mahasiswa baru. Yusuf et al. [15] melakukan pengelompokan data mahasiswa untuk optimalisasi penerimaan mahasiswa baru menggunakan metode clustering. Selain itu, Breiman [5] mengembangkan algoritma Random Forest, yang memiliki akurasi tinggi dalam klasifikasi data kompleks. Hasil dari studi-studi tersebut menunjukkan bahwa algoritma klasifikasi dan pendekatan spasial dapat menjadi alat yang efektif dalam memahami pola perilaku pendaftar dan asal-usul mereka. Selain Random Forest, metode klasifikasi lainnya seperti Naïve Bayes dan Decision Tree juga telah diterapkan dalam konteks pendidikan. Nugroho dan Prihandoyo [6] memanfaatkan kedua metode tersebut untuk memprediksi pilihan program studi calon mahasiswa baru, menunjukkan bahwa pemanfaatan algoritma klasifikasi dapat membantu dalam memahami preferensi calon mahasiswa. Berbagai pendekatan ini, meskipun memiliki metodologi yang berbeda, semakin memperlihatkan potensi besar algoritma dalam meningkatkan efisiensi pengambilan keputusan berbasis data pada tingkat

institusi pendidikan tinggi. Penelitian yang lebih lanjut terus memperkaya literatur ini dengan berbagai aplikasi algoritma lain yang semakin berkembang, yang berfokus pada pengelolaan data yang lebih akurat dan efisien.

Pendekatan klasifikasi juga diterapkan oleh Abdurrohman et al. [7] untuk mengelompokkan data pendaftaran mahasiswa baru menggunakan metode K-Means, guna mengetahui karakteristik calon mahasiswa secara lebih mendalam. Selain itu, Yobioktabera dan Wibowo [8] memanfaatkan algoritma K-NN untuk memprediksi penerimaan mahasiswa baru pada fakultas kedokteran, sementara Budiman dan Niqotaini [9] membandingkan algoritma klasifikasi data mining untuk memprediksi minat studi calon mahasiswa baru. Di sisi lain, Maukar et al. [10] menunjukkan bahwa pendekatan berbasis K-Means Clustering juga dapat diterapkan untuk menganalisis data penerimaan mahasiswa baru, memperkuat keyakinan akan peran algoritma klasifikasi dalam konteks pendidikan tinggi. Meskipun metode yang digunakan berbeda, semua penelitian tersebut mengarah pada pemanfaatan data historis untuk pengambilan keputusan yang lebih efektif. Namun, penelitian khusus yang mengkaji klasifikasi sekolah asal calon mahasiswa untuk tujuan promosi kampus masih terbatas, terutama di Universitas Muhammadiyah Banjarmasin (UMBJM). Kesenjangan ini semakin nyata ketika mempertimbangkan belum adanya kajian yang menerapkan metode klasifikasi untuk menentukan sekolah asal potensial berdasarkan data historis pendaftaran mahasiswa dalam konteks lokal. Oleh karena itu, penelitian ini hadir untuk mengisi kekosongan tersebut dengan mengembangkan pendekatan berbasis data yang kontekstual dan dapat diimplementasikan langsung oleh institusi terkait. Penelitian ini tidak hanya penting untuk pengembangan akademis, tetapi juga memiliki dampak praktis yang signifikan dalam merancang strategi promosi yang lebih efektif dan efisien. Hal ini menjadi sangat relevan, mengingat bahwa strategi promosi berbasis data akan lebih adaptif terhadap perubahan dinamika sosial dan ekonomi yang memengaruhi pola pendaftaran mahasiswa setiap tahunnya.

Pada penelitian ini, algoritma Random Forest digunakan untuk mengklasifikasikan sekolah asal mahasiswa baru UMBJM selama periode 2019–2024. Random Forest adalah metode ensemble learning yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi prediksi [5]. Selain itu, teknik k-fold cross-validation digunakan untuk mengevaluasi model sebagaimana disarankan oleh Kohavi [11], guna memastikan bahwa hasil prediksi memiliki akurasi yang stabil dan tidak mengalami overfitting. Teknik pengolahan data lainnya seperti encoding variabel kategori juga digunakan sebagaimana disarankan oleh Bolikulov et al. [12]. Pendekatan clustering juga digunakan dalam beberapa studi terkait untuk mendukung pengambilan keputusan promosi. Udariansyah dan Ibrahim [13], serta Hematang et al. [14] menunjukkan bahwa metode K-Means dan algoritma Random Forest dapat digunakan untuk mengelompokkan sekolah berdasarkan kesamaan karakteristik, sehingga strategi promosi dapat lebih terarah. Meskipun pendekatan berbeda, prinsip dasarnya tetap sama, yaitu pemanfaatan data historis untuk memperkuat pengambilan keputusan promosi.

2. METODOLOGI PENELITIAN

Penelitian ini dilakukan dengan menggunakan data pendaftaran mahasiswa Universitas Muhammadiyah Banjarmasin (UMBJM) dari periode 2019 hingga 2024 dengan tujuan untuk mengidentifikasi sekolah asal mahasiswa yang berpotensi menjadi target promosi. Proses metodologi penelitian ini terdiri dari beberapa tahapan yang dimulai dari pra-pemrosesan data hingga evaluasi model klasifikasi.



Gambar 1. Alur metode penelitian

2.1. Pra-pemrosesan Data

Data pendaftaran mahasiswa baru yang diteliti mencakup atribut seperti “Sekolah Asal”, “Provinsi/Kabupaten/Kecamatan Peserta”, “Nama”, dan atribut lain yang relevan dengan pendaftaran. Tahap pra-pemrosesan meliputi tiga langkah utama:

- 1) Filtering data, menghapus data pendaftar yang tidak relevan dengan tujuan penelitian, seperti entri yang berisi kata “universitas”, “keperawatan”, “S1”, dan sebagainya.
- 2) Penandaan validitas sekolah, memisahkan entri sekolah antara yang valid seperti “SMA Darul Hijrah” dengan yang tidak valid seperti “SMA” menggunakan pola teks tertentu.
- 3) Pembuatan label target, sekolah dengan jumlah pendaftar berjumlah 10 atau lebih selama periode 2019-2024 akan ditandai sebagai “Sekolah potensial” (label 1), dan yang jumlahnya kurang dari 10 diberi label 0.

Untuk mempersiapkan data kategorikal menjadi input numerik bagi model machine learning, dilakukan konversi ke dummy variables menggunakan metode one-hot encoding. Teknik ini telah digunakan secara luas dalam penelitian pendidikan untuk mengubah data kategorikal menjadi format numerik yang dapat diproses oleh algoritma machine learning [12]. Setiap kategori diubah menjadi representasi biner (0/1) agar dapat diproses oleh algoritma klasifikasi.

2.2. Pemodelan

Setelah tahap pra-pemrosesan selesai, langkah berikutnya adalah pemodelan. Metode klasifikasi seperti Random Forest digunakan karena kemampuannya dalam menangani data berukuran besar dan fitur yang kompleks [5]. Teknik ini telah banyak digunakan dalam berbagai bidang pendidikan untuk memprediksi dan mengklasifikasikan preferensi calon mahasiswa berdasarkan variabel-variabel input yang tersedia [2]. Pemilihan Random Forest didasarkan pada kemampuannya untuk menangani data besar, termasuk data dengan missing values, serta kemampuannya dalam menghasilkan model yang stabil dengan tingkat akurasi tinggi.

Untuk meningkatkan akurasi dan mengurangi kemungkinan bias dalam model, dilakukan Stratified K-Fold Cross Validation dengan 5 fold. Teknik ini penting untuk memastikan bahwa distribusi kelas tetap proporsional dalam setiap subset data yang digunakan untuk pelatihan dan pengujian [11]. Proses ini membantu menjaga keadilan dalam pengujian model, sehingga dapat menghasilkan estimasi akurasi yang lebih akurat dan stabil, serta mengurangi kemungkinan overfitting yang sering terjadi pada dataset kecil hingga menengah.

2.3. Evaluasi Model

Evaluasi model merupakan tahap terakhir dalam proses penelitian. Model dievaluasi dengan mengukur akurasi rata-rata dari hasil cross-validation. Akurasi ini dihitung dengan mengambil rata-rata dari hasil validasi pada setiap fold yang digunakan dalam Stratified K-Fold. Selain akurasi, model juga dievaluasi menggunakan classification report, yang mencakup metrik seperti precision, recall, f1-score, dan support. Metrik-metrik ini memberikan gambaran yang lebih menyeluruh tentang kinerja model dalam memprediksi sekolah-sekolah potensial.

Setelah proses pelatihan selesai, prediksi dapat dilakukan dengan menggunakan seluruh data yang valid. Prediksi ini bertujuan untuk mengidentifikasi sekolah-sekolah yang memiliki potensi besar untuk dijadikan target promosi kampus. Visualisasi hasil prediksi akan dibuat dengan menggunakan diagram batang untuk menampilkan 10 sekolah teratas yang memiliki potensi terbesar sebagai tempat promosi.

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengidentifikasi sekolah asal mahasiswa yang berpotensi tinggi untuk dijadikan target promosi kampus. Proses dimulai dengan mengumpulkan data pendaftaran mahasiswa baru Universitas Muhammadiyah Banjarmasin (UMBJM) dari tahun 2019 hingga 2024. Data yang digunakan telah melalui proses pra-pemrosesan yang ketat, termasuk penyaringan data tidak valid, normalisasi nama sekolah, konversi fitur kategorikal menggunakan one-hot encoding, serta pelabelan target berdasarkan jumlah mahasiswa dari setiap sekolah. Tahap pra-pemrosesan ini sangat penting untuk memastikan bahwa data yang dimasukkan ke dalam model klasifikasi benar-benar representatif dan bebas dari kesalahan yang dapat mempengaruhi hasil akhir. Proses penyaringan data melibatkan penghapusan entri yang tidak relevan, seperti data dengan kata kunci “universitas”, “keperawatan”, atau format tidak standar lainnya. Hal ini penting karena data yang tidak sesuai dapat mempengaruhi akurasi model secara keseluruhan. Selain itu, normalisasi nama sekolah dilakukan dengan menyamakan semua huruf menjadi kapital dan menghapus karakter khusus yang tidak diperlukan. Hal ini bertujuan untuk mengurangi duplikasi akibat perbedaan penulisan nama sekolah dalam data mentah. Konversi fitur kategorikal menggunakan one-hot encoding bertujuan untuk mengubah data berbentuk teks menjadi representasi numerik yang dapat diproses oleh algoritma machine learning. Teknik ini telah digunakan dalam berbagai penelitian serupa karena mampu meningkatkan kompatibilitas data dengan model klasifikasi berbasis pohon keputusan seperti Random Forest. Dengan menggunakan one-hot encoding, setiap kategori pada fitur geografis seperti kabupaten dan kota diubah menjadi bentuk biner (0 atau 1), sehingga informasi dapat diekstraksi secara optimal oleh model.

Langkah selanjutnya adalah pelabelan target. Dalam penelitian ini, sekolah yang memiliki jumlah pendaftar sebanyak 10 atau lebih selama periode penelitian diberi label “potensial” (1), sementara sekolah dengan jumlah pendaftar

kurang dari 10 diberi label "tidak potensial" (0). Kriteria ini dipilih untuk memudahkan analisis statistik dan pemetaan sekolah yang berkontribusi secara signifikan terhadap jumlah mahasiswa baru di UMBJM. Setelah tahap pra-pemrosesan selesai, model Random Forest digunakan sebagai metode klasifikasi utama. Pemilihan Random Forest didasarkan pada beberapa pertimbangan: pertama, model ini memiliki kemampuan menangani data berdimensi tinggi dengan kompleksitas fitur yang cukup besar. Kedua, Random Forest relatif tahan terhadap overfitting karena memanfaatkan kombinasi dari beberapa pohon keputusan yang dilatih secara independen. Ketiga, interpretabilitas model ini dianggap cukup baik dalam konteks klasifikasi berbasis fitur kategorikal, sehingga hasil prediksi dapat dijelaskan secara logis. Untuk memastikan model bekerja dengan baik, dilakukan evaluasi menggunakan teknik Stratified K-Fold Cross Validation sebanyak 5 fold. Teknik ini digunakan agar distribusi kelas pada setiap fold tetap seimbang, sehingga hasil validasi lebih stabil dan tidak bias. Pada setiap iterasi, model dilatih dan diuji menggunakan subset data yang berbeda, dan hasilnya dihitung secara rata-rata. Dari proses validasi ini, model menghasilkan akurasi rata-rata sebesar 81,60%, yang dianggap cukup baik dalam konteks klasifikasi sekolah potensial. Selain akurasi, metrik evaluasi lainnya seperti precision, recall, dan f1-score juga dihitung untuk menilai kemampuan model dalam mengidentifikasi sekolah potensial secara akurat. Precision mengukur seberapa banyak prediksi sekolah potensial yang benar-benar tepat sasaran, sedangkan recall mengukur sejauh mana model dapat menangkap semua sekolah potensial dari keseluruhan data yang ada. F1-score, sebagai harmonisasi dari precision dan recall, menunjukkan keseimbangan antara kedua metrik tersebut. Meskipun hasil rinci dari metrik-metrik ini tidak ditampilkan dalam bagian ini, secara umum semua metrik mendukung kesimpulan bahwa model yang dibangun memiliki performa yang memadai dalam memprediksi sekolah asal potensial.

Hasil yang diperoleh dari model ini memberikan keyakinan bahwa data geografis dan asal sekolah memiliki peranan penting dalam memprediksi potensi pendaftaran mahasiswa baru. Dengan demikian, model Random Forest yang diterapkan dalam penelitian ini terbukti mampu mengolah data pendidikan secara efektif dan memberikan hasil prediksi yang akurat.

3.1. Hasil Prediksi Sekolah Potensial

Setelah model Random Forest dikembangkan dan dievaluasi, seluruh data valid yang tersedia digunakan untuk melakukan prediksi. Model ini bekerja dengan mengklasifikasikan setiap baris data berdasarkan fitur-fitur input yang telah ditentukan, seperti "Kabupaten/Kota Pendidikan Sebelumnya" dan variabel-variabel relevan lainnya. Hasil dari klasifikasi ini berupa label yang menunjukkan apakah suatu sekolah tergolong potensial (label 1) atau tidak potensial (label 0). Tahap ini merupakan implementasi akhir dari model klasifikasi yang bertujuan untuk mengidentifikasi sekolah-sekolah dengan jumlah calon mahasiswa baru yang signifikan. Setelah klasifikasi dilakukan, data yang tergolong sebagai sekolah potensial (label 1) dianalisis lebih lanjut. Salah satu permasalahan yang ditemukan pada tahap ini adalah adanya variasi penulisan nama sekolah dalam data pendaftaran. Misalnya, beberapa entri menggunakan format huruf kapital penuh seperti "SMA NEGERI 1 BANJARMASIN", sementara yang lain menggunakan format campuran seperti "SMAN 1 Banjarmasin". Untuk mengatasi masalah ini, dilakukan proses normalisasi nama sekolah dengan mengonversi semua teks menjadi huruf kapital. Langkah ini penting agar data yang identik tetapi ditulis dengan cara berbeda dapat disatukan, sehingga tidak terjadi duplikasi dalam proses analisis. Selanjutnya, dilakukan penghitungan frekuensi kemunculan nama sekolah potensial dalam dataset hasil klasifikasi. Data frekuensi ini diurutkan secara menurun untuk memperoleh 10 besar sekolah yang paling sering muncul sebagai sumber calon mahasiswa potensial. Tabel 1 di bawah ini menyajikan hasil pengelompokan tersebut:

Tabel 1. Hasil top 10 sekolah asal potensial

No	Sekolah Asal	Jumlah Mahasiswa	Kabupaten/Kota Pendidikan Sebelumnya
1	SMK ISFI BANJARMASIN	35	KOTA BANJARMASIN
2	MAN 1 MURUNG RAYA	31	KABUPATEN MURUNG RAYA
3	SMAN 1 SIMPANG EMPAT	20	KABUPATEN TANAH BUMBU
4	MAN BARITO UTARA	17	KABUPATEN BARITO UTARA
5	SMAN 1 MARABAHAN	17	KABUPATEN BARITO KUALA
6	MAN 2 HULU SUNGAI UTARA	16	KABUPATEN HULU SUNGAI UTARA
7	SMAN 1 SUNGAI PANDAN	15	KABUPATEN HULU SUNGAI UTARA
8	SMK UNGGULAN HUSADA	15	KOTA BANJARMASIN
9	SMA NEGERI 2 BANJARMASIN	15	KOTA BANJARMASIN
10	MAN 3 BANJARMASIN	15	KOTA BANJARMASIN

Dari tabel tersebut dapat dilihat bahwa tiga sekolah dengan jumlah mahasiswa potensial terbanyak berasal dari Kota Banjarmasin dan Kabupaten Murung Raya. Fakta ini mengindikasikan bahwa wilayah-wilayah tersebut memiliki kontribusi besar dalam penyediaan calon mahasiswa baru bagi Universitas Muhammadiyah Banjarmasin (UMBJM). Untuk memastikan keakuratan data, dilakukan pengecekan silang terhadap data mentah dari pendaftaran mahasiswa baru.

Langkah ini penting untuk memastikan bahwa tidak ada data ganda atau kesalahan identifikasi sekolah. Hasil validasi menunjukkan bahwa sekolah-sekolah yang berada dalam daftar 10 besar tersebut memang secara konsisten menyumbangkan jumlah pendaftar yang signifikan selama periode 2019-2024. Temuan ini memiliki implikasi penting bagi strategi promosi kampus. Dengan memanfaatkan data sekolah potensial yang telah diidentifikasi, pihak kampus dapat mengalokasikan sumber daya promosi secara lebih efektif dan tepat sasaran. Misalnya, kegiatan sosialisasi dan promosi dapat difokuskan pada sekolah-sekolah dengan kontribusi tinggi untuk memaksimalkan daya tarik kampus. Selain itu, kerjasama strategis dengan pihak sekolah juga dapat ditingkatkan agar informasi mengenai program studi unggulan UMBJM lebih mudah diakses oleh calon mahasiswa. Secara keseluruhan, hasil prediksi ini menunjukkan bahwa model Random Forest yang diterapkan mampu menghasilkan klasifikasi yang akurat dan informatif. Dengan demikian, strategi promosi kampus dapat didasarkan pada data konkret yang menggambarkan asal-usul calon mahasiswa potensial, sehingga kampus dapat meningkatkan efektivitas perekrutan dengan menasar sekolah-sekolah yang memiliki rekam jejak kontribusi tinggi.

3.2. Interpretasi Hasil dan Keterkaitannya dengan Penelitian Sebelumnya

Hasil prediksi pada penelitian ini menunjukkan bahwa model Random Forest mampu mengklasifikasikan sekolah potensial dengan akurasi sebesar 81,60%, yang dapat dianggap tinggi dalam konteks prediksi pendidikan. Temuan ini memperkuat gagasan bahwa pendekatan berbasis machine learning mampu memberikan kontribusi nyata dalam strategi promosi kampus. Penggunaan fitur “Kabupaten/Kota Pendidikan Sebelumnya” sebagai variabel input utama terbukti informatif dalam mengidentifikasi sekolah yang cenderung menghasilkan lebih banyak pendaftar. Jika dibandingkan dengan penelitian terdahulu, pendekatan dalam studi ini memiliki beberapa perbedaan penting.

Pratama et al. [2] menggunakan algoritma Random Forest untuk klasifikasi pemilihan jurusan berdasarkan riwayat nilai siswa. Fokus mereka adalah membantu calon mahasiswa dalam menentukan jurusan yang paling sesuai dengan kemampuan akademiknya, sementara penelitian ini lebih berfokus pada identifikasi sekolah potensial sebagai target promosi kampus. Meskipun berbeda konteks, kesamaan metode menunjukkan bahwa Random Forest efektif dalam menangani data pendidikan dengan fitur yang kompleks. Rahmalinda dan Jananto [4] menggunakan metode K-Means Clustering untuk menentukan strategi promosi berdasarkan data penerimaan mahasiswa baru. Meskipun mereka menggunakan pendekatan unsupervised learning yang berbeda dari penelitian ini, tujuan akhirnya tetap sama, yaitu menghasilkan strategi promosi yang lebih terarah berdasarkan data historis. Penelitian ini lebih menekankan pada pemanfaatan klasifikasi terawasi untuk langsung memetakan sekolah potensial berdasarkan data historis pendaftaran. Studi oleh Yusuf et al. [15] menggunakan metode K-Means Clustering untuk mengelompokkan data mahasiswa dalam rangka optimalisasi penerimaan. Meskipun objek prediksinya berbeda (program studi vs. sekolah potensial), prinsip penggunaan data historis dan teknik klasifikasi untuk mendukung pengambilan keputusan di institusi pendidikan tetap sejalan.

Sementara itu, Abdurrohman et al. [7] juga menerapkan metode K-Means Clustering untuk mengelompokkan data pendaftaran mahasiswa baru guna mengetahui karakteristik calon mahasiswa secara lebih mendalam. Walau berbeda dalam tujuan utama, kedua penelitian ini sama-sama memanfaatkan teknik pengelompokan untuk mendukung strategi pengambilan keputusan. Selain itu, Yobioktabera dan Wibowo [8] memanfaatkan algoritma K-NN untuk memprediksi penerimaan mahasiswa baru pada fakultas kedokteran. Meskipun algoritma dan objek prediksi berbeda, kedua penelitian tersebut sama-sama menggunakan data historis untuk memprediksi tren pendaftaran. Di sisi lain, Budiman dan Niqotaini [9] melakukan perbandingan beberapa algoritma klasifikasi (Naïve Bayes, Decision Tree, dan K-NN) untuk memprediksi minat calon mahasiswa baru. Temuan mereka menunjukkan bahwa meskipun algoritma berbeda, semua metode mampu mengolah data pendidikan secara efektif. Hal ini sejalan dengan penelitian ini yang menggunakan Random Forest sebagai model utama dengan akurasi tinggi. Maukar et al. [10] juga menunjukkan efektivitas K-Means Clustering dalam menganalisis data penerimaan mahasiswa baru, memperkuat keyakinan akan peran algoritma berbasis data dalam pengambilan keputusan promosi.

Dengan demikian, dapat disimpulkan bahwa meskipun pendekatan dalam penelitian ini berbeda dalam fokus dan objek dari beberapa penelitian sebelumnya, semuanya mengarah pada pemanfaatan data historis dan teknik machine learning untuk mendukung pengambilan keputusan strategis di bidang pendidikan. Kelebihan utama dari penelitian ini terletak pada penggunaan model prediktif dengan label eksplisit (supervised classification) terhadap entitas sekolah, yang menawarkan nilai praktis langsung dalam penyusunan prioritas promosi kampus.

3.3. Implikasi Strategis

Hasil dari penelitian ini dapat memberikan dasar empiris bagi Universitas Muhammadiyah Banjarmasin (UMBJM) dalam menyusun strategi promosi berbasis data yang lebih efektif. Dengan mengidentifikasi sekolah-sekolah potensial secara kuantitatif, pihak kampus memiliki peluang untuk melakukan pendekatan yang lebih personal dan efisien. Salah satu strategi yang dapat diterapkan adalah melakukan kunjungan langsung ke sekolah-sekolah tersebut dengan membawa materi promosi yang terstruktur dan menarik. Melalui kunjungan langsung, pihak kampus tidak hanya dapat memperkenalkan program studi unggulan, tetapi juga membangun hubungan yang lebih erat dengan pihak sekolah, sehingga potensi perekrutan mahasiswa baru dapat meningkat secara signifikan. Selain kunjungan, pemberian beasiswa khusus kepada siswa dari sekolah-sekolah potensial juga dapat menjadi langkah strategis untuk menarik minat calon

mahasiswa baru. Beasiswa ini tidak hanya akan menjadi daya tarik tersendiri, tetapi juga dapat meningkatkan citra positif kampus di mata calon mahasiswa dan pihak sekolah.

Selain itu, penyelenggaraan kegiatan sosialisasi akademik seperti seminar karir, workshop tentang dunia perkuliahan, dan kompetisi antar-sekolah juga dapat menjadi sarana untuk meningkatkan minat siswa terhadap UMBJM. Dengan demikian, kampus dapat memposisikan dirinya sebagai pilihan utama bagi siswa dari sekolah-sekolah yang telah teridentifikasi potensial. Di sisi lain, hasil penelitian ini juga dapat menjadi masukan penting dalam pengembangan sistem informasi promosi yang terotomatisasi dan berbasis kecerdasan buatan (AI). Sistem ini dapat memanfaatkan data historis penerimaan mahasiswa untuk secara otomatis mengidentifikasi sekolah dengan kontribusi tinggi dan menyarankan strategi promosi yang sesuai. Misalnya, platform promosi berbasis AI dapat mengelompokkan sekolah berdasarkan jumlah pendaftar dan mengirimkan notifikasi atau laporan otomatis kepada tim promosi kampus. Hal ini akan meningkatkan efisiensi dalam perencanaan promosi sekaligus mempermudah pemantauan efektivitas dari strategi yang diterapkan. Lebih lanjut, penggunaan machine learning dalam konteks promosi kampus menunjukkan adanya tren baru dalam manajemen pendidikan tinggi, di mana pendekatan berbasis data mulai menggantikan intuisi semata. Implementasi metode ini secara konsisten dapat menghasilkan peningkatan jumlah mahasiswa baru dari sekolah-sekolah sasaran yang telah diidentifikasi. Dalam jangka panjang, hal ini akan memberikan dampak positif tidak hanya pada peningkatan jumlah mahasiswa, tetapi juga pada keberlanjutan kampus dalam menghadapi persaingan yang semakin ketat antarperguruan tinggi.

Tidak hanya untuk kampus UMBJM, metode serupa juga dapat diterapkan oleh perguruan tinggi lain yang menghadapi tantangan dalam menarik mahasiswa baru. Melalui pemanfaatan data historis yang lebih terstruktur, kampus-kampus dapat mengidentifikasi daerah atau sekolah yang memberikan kontribusi tinggi dan mengoptimalkan alokasi sumber daya promosi. Strategi berbasis data ini memungkinkan kampus melakukan perencanaan jangka panjang dengan lebih terarah dan dapat diukur tingkat keberhasilannya. Oleh karena itu, penelitian ini tidak hanya bermanfaat bagi UMBJM, tetapi juga dapat menjadi model bagi institusi pendidikan lainnya dalam menyusun kebijakan promosi yang lebih inovatif dan responsif terhadap perubahan tren pendidikan.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model klasifikasi machine learning menggunakan algoritma random forest untuk mengidentifikasi sekolah-sekolah yang potensial dalam menyumbang mahasiswa baru bagi Universitas Muhammadiyah Banjarmasin. Model dilatih berdasarkan fitur “Kabupaten/Kota Pendidikan Sebelumnya” dan divalidasi menggunakan teknik cross validation 5-fold, yang menghasilkan akurasi rata-rata sebesar 81.60%. Hasil ini menunjukkan bahwa pendekatan machine learning berbasis data historis memiliki potensi besar dalam mendukung strategi promosi kampus yang lebih terarah dan efektif. Dari hasil prediksi, sepuluh sekolah asal dengan kontribusi terbesar terhadap label “potensial” berhasil diidentifikasi. Sekolah-sekolah ini sebagian besar berasal dari Kota Banjarmasin serta kabupaten sekitar seperti Murung Raya, Hulu Sungai Utara, dan Barito Kuala, menunjukkan bahwa kedekatan geografis dan jaringan historis menjadi faktor penting dalam penentuan potensi sekolah. Temuan ini diharapkan dapat menjadi landasan awal bagi kebijakan promosi yang lebih berbasis data.

Meski demikian, penelitian ini memiliki keterbatasan dalam aspek normalisasi data sekolah, yang dapat mempengaruhi akurasi klasifikasi. Oleh karena itu, untuk pengembangan ke depan, disarankan integrasi fitur tambahan seperti asal jurusan siswa atau data sosio-demografis serta pemanfaatan teknik natural language processing untuk pembersihan dan penyamaan nama sekolah guna meningkatkan akurasi dan ketepatan strategi promosi yang dihasilkan. Penambahan mekanisme evaluasi lainnya seperti precision, recall, dan F1-score juga direkomendasikan agar hasil analisis semakin komprehensif dan dapat dipertanggungjawabkan secara ilmiah.

REFERENCES

- [1] Agneresa, A. L. Hananto, S. S. Hilabi, A. Hananto, and Tukino, “Strategi Promosi Penerapan Data Mining Mahasiswa Baru Dengan Metode K-Means Clustering,” *Dirgamaya J. Manaj. dan Sist. Inf.*, vol. 2, no. 2, pp. 25–34, 2022, doi: 10.35969/dirgamaya.v2i2.275, [Online]. Available: <https://ejournal.rosma.ac.id/index.php/dirgamaya/article/view/275/224>
- [2] A. R. Pratama, R. R. Aryanto, and A. T. M. Pratama, “Model Klasifikasi Calon Mahasiswa Baru Untuk Sistem Rekomendasi Program Studi Sarjana Berbasis Machine Learning,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 4, pp. 725–734, Aug. 2022, doi: 10.25126/jtiik.2022934311, [Online]. Available: <https://jtiik.ub.ac.id/index.php/jtiik/article/view/4311>
- [3] G. E. Pratiwi and A. Nugroho, “Implementasi metode random forest untuk klasifikasi penjualan produk sabun paling laris,” *J. TEKINKOM*, vol. 7, no. 2, pp. 531–540, 2024, doi: 10.37600/tekinkom.v7i2.1610, [Online]. Available: <https://jurnal.murnisadar.ac.id/index.php/Tekinkom/article/view/1610/780>
- [4] N. A. Rahmalinda and A. Jananto, “Penerapan Metode K-Means Clustering Dalam Menentukan Strategi Promosi Berdasarkan Data Penerimaan Mahasiswa Baru,” *J. Tekno Kompak*, vol. 16, no. 2, p. 163, 2022, doi: 10.33365/jtk.v16i2.1971, [Online]. Available: <https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/1971/954>



- [5] L. Breiman, "Random Forests," in *Machine Learning*, vol. 45, R. E. Schapire, Ed., Kluwer Academic Publishers, 2001, pp. 5–32. [Online]. Available: <https://link.springer.com/article/10.1023/A:1010933404324>
- [6] W. E. Nugroho and M. T. Prihandoyo, "Menentukan Program Studi Bagi Calon Mahasiswa Baru Menggunakan Metode Naïve Bayes Dan Metode Decision Tree," *J. Inform. Univ. Pamulang*, vol. 7, no. 3, pp. 672–678, 2022, [Online]. Available: <https://openjournal.unpam.ac.id/index.php/informatika/article/view/12563/pdf>.
- [7] M. H. Abdurrohman, E. Haerani, F. Syafria, and L. Oktavia, "IMPLEMENTASI K-MEANS CLUSTERING PADA DATA PENGELOMPOKAN PENDAFTARAN MAHASISWA BARU (STUDI KASUS UNIVERSITAS ABDURRAB)," *RABIT J. Teknol. dan Sist. Inf. Univrab*, vol. 9, no. 1, pp. 138–147, 2024, [Online]. Available: <https://jurnal.univrab.ac.id/index.php/rabit/article/view/4255/1754>
- [8] A. Yobioktabera and A. W. Wibowo, "Penerapan Data Mining Untuk Memprediksi Penerimaan Calon Mahasiswa Baru Fakultas Kedokteran Menggunakan Algoritma K-NN," *JTET (Jurnal Tek. Elektro Ter.)*, vol. 10, no. 1, pp. 16–19, 2021, [Online]. Available: <https://jurnal.polines.ac.id/index.php/jtet/article/view/2550/pdf>
- [9] Budiman and Z. Niqotaini, "Perbandingan Algoritma Klasifikasi Data Mining untuk Penelusuran Minat Calon Mahasiswa Baru," *Nuansa Inform.*, vol. 15, no. 2, pp. 37–52, 2021, doi: 10.25134/nuansa.v15i2.4162, [Online]. Available: <https://journal.uniku.ac.id/index.php/ilkom/article/view/4162/2676>
- [10] A. L. Maukar, F. Marisa, A. A. Widodo, N. Kamilaningtyas, and N. D. Nugraha, "Analisis Data Penerimaan Mahasiswa Baru Berbasis K-Means," *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 2, pp. 142–147, 2022, doi: 10.26798/jiko.v6i2.558, [Online]. Available: <https://ejournal.akakom.ac.id/index.php/jiko/article/view/558/pdf>
- [11] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *Proc. 14th Int. Jt. Conf. Artif. Intell.*, vol. 2, pp. 1137–1143, 1993, [Online]. Available: <https://www.ijcai.org/Proceedings/95-2/Papers/016.pdf>
- [12] F. Bolikulov, R. Nasimov, A. Rashidov, F. Akhmedov, and Y. I. Cho, "Effective Methods of Categorical Data Encoding for Artificial Intelligence Algorithms," *Mathematics*, vol. 12, no. 16, Aug. 2024, doi: 10.3390/math12162553. [Online]. Available: <https://doi.org/10.3390/math12162553>
- [13] D. Udariansyah and D. Rahmat Ibrahim, "Klasifikasi Data Penerimaan Mahasiswa Baru Pada Universitas Bina Darma Menggunakan Algoritma K-Means Clustering," *J. Pendidik. dan Konseling*, vol. 4, no. 4, pp. 2692–2701, 2022, [Online]. Available: <https://journal.universitaspahlawan.ac.id/index.php/jpdk/article/view/5879>
- [14] A. Kurniawati Hematang, N. Widya Utami, and A. A. I. Ita Paramitha, "Perbandingan Model Prediksi Calon Mahasiswa Baru Menggunakan Algoritma Id3 Dan Random Forest," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 6, pp. 3427–3434, 2023, doi: 10.36040/jati.v7i6.8193, [Online]. Available: <https://ejournal.itn.ac.id/index.php/jati/article/view/8193/4847>
- [15] D. Yusuf, E. Sestri, and F. Razi, "PENGELOMPOKKAN DATA MAHASISWA MENGGUNAKAN CLUSTERING UNTUK OPTIMALISASI PENERIMAAN MAHASISWA BARU," *JIKA (Jurnal Inform. Univ. Muhammadiyah Tangerang)*, vol. 8, no. 4, p. 484–490, 2024, [Online]. Available: <https://jurnal.umt.ac.id/index.php/jika/article/view/12637/5733>