SYSTEMATIC LITERATUR REVIEW SELEKSI BEASISWA KIP DALAM MACHINE LEARNING

Muhammad Adha, Fitri Handayani
Teknik Informatika, Universitas Muhammadiyah Riau
Jalan Tuanku Tambusai Kota Pekanbaru, Indonesia
muhammadadha227@gmail.com

ABSTRAK

Proses seleksi penerima Beasiswa Kartu Indonesia Pintar (KIP) masih menghadapi tantangan dalam menjamin ketepatan sasaran, transparansi, dan konsistensi pengambilan keputusan. Tingginya jumlah pendaftar serta keberagaman indikator akademik dan sosial ekonomi seringkali menyebabkan proses evaluasi menjadi kurang efisien dan berpotensi subjektif. Untuk mengatasi permasalahan tersebut, pemanfaatan pendekatan berbasis *machine learning* menjadi alternatif solusi dalam mendukung sistem seleksi yang lebih objektif dan terukur. Penelitian ini menerapkan metode *Systematic Literature Review* (SLR) guna menelaah dan mensintesis berbagai studi yang membahas penerapan algoritma klasifikasi dalam seleksi Beasiswa KIP maupun program sejenis. Tahapan penelitian meliputi identifikasi sumber literatur, proses penyaringan berdasarkan kriteria inklusi, evaluasi kualitas publikasi, serta analisis komparatif terhadap metode yang digunakan. Hasil kajian menunjukkan bahwa algoritma Naïve Bayes dan Random Forest merupakan pendekatan yang paling banyak diterapkan, dengan Random Forest cenderung memberikan performa akurasi dan stabilitas yang lebih tinggi, sedangkan Naïve Bayes unggul dari sisi efisiensi komputasi dan kemudahan implementasi. Selain itu, penggunaan metrik evaluasi seperti *confusion matrix*, akurasi, presisi, recall, dan *k-fold cross validation* berperan penting dalam memastikan reliabilitas model. Penelitian ini merekomendasikan pengembangan sistem seleksi yang mengintegrasikan variabel akademik dan sosial ekonomi secara komprehensif serta menerapkan evaluasi multi-metrik untuk meningkatkan keadilan dan ketepatan keputusan.

Kata kunci : *Beasiswa KIP, Evaluasi Model, Klasifikasi, Machine Learning, Systematic Literature Review*

1. PENDAHULUAN

Program Beasiswa Kartu Indonesia Pintar (KIP) merupakan instrumen kebijakan pemerintah yang dirancang untuk memperluas kesempatan masyarakat dalam mengakses pendidikan tinggi, khususnya bagi mahasiswa yang berasal dari latar belakang ekonomi kurang mampu. Melalui skema beasiswa ini, pemerintah berupaya menghilangkan kendala finansial yang kerap menjadi penghalang bagi masyarakat untuk melanjutkan studi ke perguruan tinggi. Akan tetapi, pelaksanaan seleksi penerima Beasiswa KIP masih menghadapi sejumlah tantangan, antara lain tingginya jumlah pemohon, keterbatasan kapasitas pengelolaan, serta kemungkinan terjadinya subjektivitas dalam proses penentuan penerima. Situasi tersebut menunjukkan perlunya mekanisme seleksi yang lebih transparan, tepat, dan efektif [1].

Perkembangan pesat teknologi informasi turut mendorong pemanfaatan machine learning dalam berbagai proses pengambilan keputusan, termasuk dalam penentuan penerima beasiswa. Sejumlah studi terdahulu melaporkan bahwa penggunaan algoritma machine learning, seperti Naïve Bayes, Decision Tree, dan Random Forest, dapat mengklasifikasikan calon penerima beasiswa berdasarkan sejumlah indikator tertentu dengan tingkat akurasi yang memadai. Pendekatan yang berorientasi pada data ini dinilai mampu mengurangi potensi bias subjektif serta menghasilkan keputusan yang lebih konsisten dibandingkan dengan metode seleksi konvensional [1][2].

Meskipun demikian, kajian mengenai penerapan machine learning dalam seleksi Beasiswa KIP masih tersebar di berbagai publikasi dengan variasi fokus kajian, metode yang digunakan, serta hasil yang diperoleh. Hingga saat ini, masih terbatas penelitian yang secara terstruktur melakukan pemetaan, perbandingan, dan analisis terhadap tren penggunaan algoritma machine learning dalam konteks seleksi Beasiswa KIP maupun program beasiswa serupa. Oleh sebab itu, dibutuhkan suatu kajian menyeluruh yang dapat memberikan pemahaman komprehensif mengenai pendekatan yang digunakan, kelebihan, serta keterbatasan dari metode-metode tersebut [2].

Berdasarkan uraian latar belakang tersebut, penelitian ini bertujuan untuk mengkaji karakteristik penerapan machine learning dalam proses seleksi penerima Beasiswa KIP berdasarkan hasil penelitian sebelumnya, mengidentifikasi algoritma yang paling dominan digunakan, serta menganalisis kecenderungan kinerja metode yang diterapkan. Untuk mencapai tujuan tersebut, penelitian ini menerapkan pendekatan Systematic Literature Review (SLR) guna menelusuri, menilai, dan mensintesis temuan penelitian yang relevan secara sistematis dan terstruktur [3].

Berdasarkan gap tersebut, ada sejumlah hal metodologis yang sebaiknya dikaji dalam sebuah tinjauan literatur sistematis yang fokus pada Naïve Bayes dan Random Forest untuk seleksi KIP:

Kajian dalam tinjauan literatur sistematis ini difokuskan pada beberapa dimensi metodologis yang

saling berkaitan. Pertama, karakteristik dataset yang digunakan dalam setiap studi ditelaah meliputi ukuran sampel, proporsi kelas, tipe fitur, adanya missing values, serta kualitas atribut ekonomi dan sosial. Kedua, teknik pra-pemrosesan dan penyeimbangan data dikaji, termasuk penggunaan undersampling, oversampling, SMOTE, penyesuaian bobot kelas, serta feature engineering yang relevan. Ketiga, aspek pengaturan model dan validasi diperiksa, mencakup prosedur cross-validation, pembagian data latih-uji, tuning hyperparameter, serta pengukuran variabilitas performa. Keempat, metrik evaluasi yang digunakan dinilai secara komprehensif, tidak hanya terbatas pada akurasi, tetapi juga precision, recall, F1-score, AUC, confusion matrix, dan metrik kalibrasi. Kelima, analisis interpretabilitas model ditelaah untuk mengetahui apakah studi menilai kontribusi atribut, menerapkan teknik explainable-AI, atau menyediakan interpretasi yang dapat dipertanggungjawabkan kepada pemangku kebijakan. Keenam, aspek operasional dan etika dipertimbangkan, mencakup waktu komputasi, kebutuhan sumber daya, kemudahan integrasi ke sistem administrasi kampus, serta potensi bias atau dampak diskriminatif. Ketujuh, kajian ini menghasilkan rekomendasi praktis berupa pedoman pemilihan algoritma yang sesuai dengan kondisi data, di mana Naïve Bayes lebih cocok untuk data berskala kecil dengan fitur yang relatif independen, sementara Random Forest lebih tepat diterapkan pada data berfitur kompleks dan non-linear, disertai kombinasi pipeline terbaik yang mencakup pra-pemrosesan, pemilihan algoritma, dan evaluasi model secara terpadu.

2. TINJAUAN PUSTAKA

Beasiswa Kartu Indonesia Pintar (KIP) adalah salah satu bentuk intervensi kebijakan pemerintah Indonesia di bidang pendidikan tinggi yang ditujukan bagi mahasiswa yang berasal dari kelompok ekonomi tidak mampu. Sebagai perluasan dari program KIP pada jenjang pendidikan dasar dan menengah, beasiswa ini hadir untuk menekan angka mahasiswa yang terpaksa berhenti kuliah karena terkendala biaya [1].

Bertambahnya jumlah pendaftar dari tahun ke tahun membuat proses penentuan penerima beasiswa KIP semakin rumit, sehingga diperlukan pendekatan seleksi yang lebih terukur, adil, dan dapat dipertanggungjawabkan [2].

Machine learning adalah salah satu cabang ilmu kecerdasan buatan yang berfokus pada pengembangan sistem yang mampu belajar dan beradaptasi secara mandiri berdasarkan pola dalam data, tanpa bergantung pada instruksi pemrograman yang bersifat eksplisit [3].

Dalam penerapannya untuk tugas klasifikasi, pendekatan ini dimanfaatkan untuk mengelompokkan data ke dalam kelas atau kategori tertentu berdasarkan keteraturan yang ditemukan selama proses pelatihan. Algoritma yang lazim digunakan dalam penelitian akademik antara lain Naïve Bayes, Decision Tree, Random Forest, Support Vector Machine (SVM), serta K-Nearest Neighbor (KNN), masing-masing dengan keunggulan dan keterbatasan tersendiri [5][6].

Naïve Bayes adalah algoritma klasifikasi berbasis probabilitas yang bekerja dengan menerapkan Teorema Bayes sambil mengasumsikan bahwa setiap fitur tidak saling mempengaruhi satu sama lain. Kelebihan utama algoritma ini terletak pada kemudahannya dalam diimplementasikan, kecepatan proses pelatihannya, serta kemampuannya menghasilkan prediksi yang cukup andal meskipun diterapkan pada dataset berukuran terbatas [2][3].

Sejumlah penelitian di bidang seleksi beasiswa mencatat bahwa Naïve Bayes dapat menghasilkan tingkat akurasi yang memuaskan, khususnya ketika fitur-fitur yang digunakan—seperti data akademik dan kondisi ekonomi mahasiswa—tidak memiliki ketergantungan yang kuat satu dengan lainnya [10][12].

Random Forest termasuk dalam kategori ensemble learning, yakni pendekatan yang menggabungkan prediksi dari banyak model untuk menghasilkan keputusan yang lebih kuat. Algoritma ini membangun sejumlah pohon keputusan secara bersamaan menggunakan teknik bootstrap aggregating (bagging) disertai pemilihan subset fitur secara acak pada setiap proses pemisahan node [5][9].

Hasil prediksi akhir diperoleh melalui mekanisme voting dari seluruh pohon yang terbentuk. Dengan cara kerja seperti ini, Random Forest secara efektif meredam tendensi overfitting yang sering menjadi kelemahan pohon keputusan tunggal, sehingga menghasilkan model yang lebih konsisten dan tangguh ketika berhadapan dengan data yang memiliki pola hubungan antar fitur yang rumit [9][11].

Systematic Literature Review (SLR) adalah pendekatan penelitian yang dirancang untuk menelusuri, menyaring, dan merangkum temuan-temuan dari berbagai studi yang telah ada secara terstruktur dan dapat diverifikasi ulang [3].

Berbeda dari kajian pustaka konvensional yang cenderung bersifat naratif dan selektif, SLR mensyaratkan adanya protokol pencarian yang baku, kriteria penerimaan dan penolakan artikel yang telah ditetapkan sejak awal, serta proses penilaian kualitas publikasi yang konsisten. Keunggulan metode ini terletak pada kemampuannya meminimalkan bias peneliti dalam proses seleksi literatur, sehingga hasil sintesis yang diperoleh lebih representatif dan dapat dipercaya [1][7].

Penilaian kualitas model machine learning dalam tugas klasifikasi tidak cukup hanya bertumpu pada satu angka akurasi semata, melainkan memerlukan serangkaian metrik yang saling melengkapi. Confusion matrix menjadi dasar dari hampir semua metrik evaluasi karena mencatat secara rinci distribusi prediksi benar dan salah ke dalam empat kategori: true positive, true negative, false positive, dan false negative. Dari matriks tersebut kemudian diturunkan

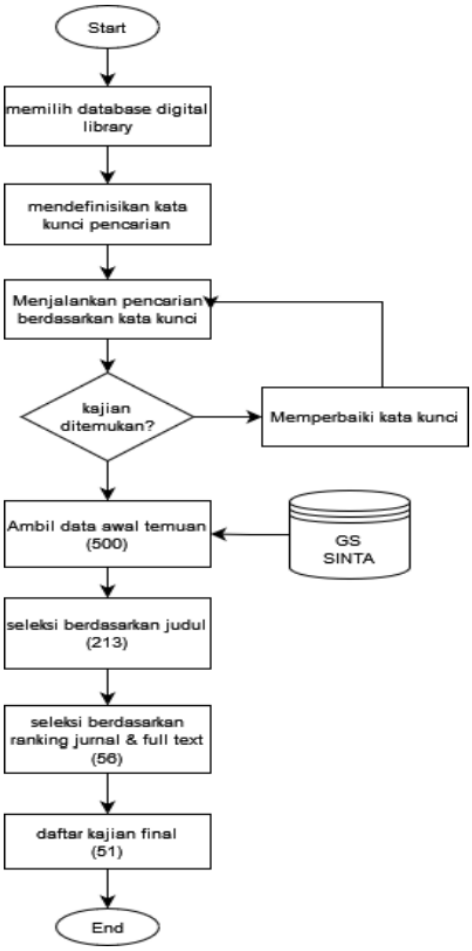
nilai presisi, recall, serta F1-score yang memberikan gambaran lebih lengkap tentang performa model [12][13].

Di sisi lain, k-fold cross validation diterapkan untuk memvalidasi kemampuan generalisasi model secara lebih andal dengan cara merotasi data uji dan data latih sebanyak k kali sehingga setiap data berkesempatan menjadi bagian pengujian [11].

Hal ini menjadi krusial dalam konteks seleksi beasiswa, mengingat distribusi data antara kelompok penerima dan bukan penerima seringkali tidak seimbang [12].

3. METODE PENELITIAN

Diagram alir pada gambar berikut menunjukkan rangkaian prosedur terstruktur yang diterapkan dalam pelaksanaan Systematic Literature Review (SLR) untuk menelaah pemanfaatan metode machine learning pada proses penentuan penerima Beasiswa Kartu Indonesia Pintar (KIP). Setiap langkah dalam alur tersebut disusun secara sistematis guna menjamin bahwa sumber pustaka yang dianalisis memiliki relevansi, mutu akademik, serta kesesuaian dengan sasaran penelitian.



Gambar 1. Strategi Pencarian dan Seleksi Artikel

Pada Gambar 1 menampilkan tahapan yang runtut dalam proses penelusuran dan penyaringan literatur pada sebuah penelitian, terutama pada bagian kajian pustaka atau Systematic Literature Review (SLR). Proses diawali dari tahap permulaan (Start) dengan menentukan pangkalan data digital yang relevan dan memiliki kredibilitas tinggi sebagai sumber referensi ilmiah. Pada diagram terlihat bahwa sumber yang dimanfaatkan adalah Google Scholar (GS) dan SINTA, dua basis data publikasi ilmiah yang umum digunakan dalam penelitian akademik di Indonesia. Tahap pemilihan basis data menjadi sangat penting karena berpengaruh terhadap mutu, relevansi, serta cakupan artikel yang berhasil dihimpun [1].

Sesudah menetapkan basis data, peneliti menyusun kata kunci pencarian yang selaras dengan fokus penelitian. Penyusunan kata kunci dilakukan secara sistematis dan spesifik agar hasil pencarian lebih terarah serta mampu mewakili variabel maupun konsep utama yang diteliti. Langkah selanjutnya adalah melakukan pencarian pada basis data terpilih menggunakan kata kunci tersebut. Jika pada tahap evaluasi ditemukan bahwa literatur yang diperoleh belum memadai atau belum sepenuhnya relevan, maka dilakukan penyempurnaan kata kunci. Proses ini bersifat berulang (iteratif) hingga diperoleh hasil yang sesuai dengan kebutuhan penelitian [2][3].

Setelah literatur yang relevan berhasil diidentifikasi, peneliti mengumpulkan sebanyak 500 artikel sebagai populasi awal. Jumlah ini merupakan hasil pencarian awal yang masih bersifat luas dan belum melalui tahap penyaringan mendalam. Tahap seleksi pertama dilakukan berdasarkan judul artikel, sehingga jumlahnya menyusut menjadi 213 artikel yang dianggap memiliki keterkaitan dengan topik penelitian. Tujuan tahap ini adalah mengeliminasi artikel yang secara eksplisit tidak sesuai dengan fokus kajian [2].

Tahap berikutnya adalah seleksi kedua yang mempertimbangkan peringkat jurnal serta ketersediaan teks lengkap (full text). Pada tahap ini, aspek kualitas sumber menjadi perhatian utama, termasuk reputasi jurnal, tingkat akreditasi, dan akses terhadap isi artikel secara menyeluruh guna menjamin kelengkapan data. Setelah proses ini, jumlah artikel berkurang menjadi 56. Tahap akhir berupa penetapan literatur final dilakukan melalui penelaahan mendalam terhadap isi, metode penelitian, serta kesesuaiannya dengan kriteria inklusi dan eksklusi yang telah ditetapkan. Dari proses tersebut, diperoleh 51 artikel yang dinilai layak dan relevan untuk dianalisis lebih lanjut [1].

Secara umum, diagram tersebut menggambarkan mekanisme seleksi literatur yang tersusun secara sistematis dan bertahap, dimulai dari identifikasi sumber, penyaringan awal, evaluasi kualitas, hingga penetapan referensi akhir. Alur ini menegaskan pentingnya transparansi dan ketelitian dalam memilih sumber pustaka agar penelitian memiliki landasan teoretis yang kuat, sahih, dan dapat dipertanggungjawabkan secara akademik [2].

Berdasarkan temuan dari Systematic Literature Review (SLR) yang telah dilaksanakan, dapat

disimpulkan bahwa pemanfaatan teknik machine learning dalam proses penentuan penerima Beasiswa Kartu Indonesia Pintar (KIP) berkontribusi signifikan terhadap peningkatan kualitas pengambilan keputusan, khususnya dari aspek efisiensi dan objektivitas. Melalui tahapan SLR yang dilakukan secara terencana meliputi penentuan basis data digital, perumusan kata kunci pencarian, serta penyaringan literatur secara bertahap kajian ini berhasil mengumpulkan dan memilih artikel ilmiah yang relevan serta memiliki mutu akademik yang sesuai dengan fokus penelitian [3][4].

Hasil integrasi dan analisis literatur menunjukkan bahwa berbagai algoritma machine learning, termasuk Naïve Bayes dan metode klasifikasi lainnya, banyak diaplikasikan dalam konteks seleksi beasiswa karena kemampuannya menangani data multikriteria secara efektif dan menghasilkan tingkat akurasi yang relatif tinggi. Studi yang dilakukan oleh Hidayat et al. mengungkapkan bahwa algoritma Naïve Bayes mampu mengelompokkan calon penerima Beasiswa KIP secara optimal dengan mempertimbangkan indikator akademik maupun kondisi sosial ekonomi [5].

Temuan tersebut sejalan dengan hasil penelitian Zirby et al. yang menegaskan bahwa pendekatan berbasis machine learning memiliki keunggulan dalam meningkatkan ketepatan prediksi penerima beasiswa dibandingkan dengan metode seleksi tradisional [6].

Lebih lanjut, penerapan metode Systematic Literature Review dalam penelitian ini menyediakan kerangka analisis yang sistematis dan dapat direplikasi untuk menelaah perkembangan riset mengenai seleksi beasiswa berbasis machine learning. Pendekatan SLR memungkinkan peneliti untuk memetakan arah tren penelitian, melakukan perbandingan antar algoritma yang digunakan, serta menilai kelebihan dan keterbatasan masing-masing metode yang telah diimplementasikan dalam studi sebelumnya [7].

Secara umum, hasil kajian ini mengindikasikan bahwa machine learning memiliki potensi yang besar dalam mendukung proses seleksi Beasiswa KIP yang lebih transparan, adil, dan efisien. Meskipun demikian, peluang pengembangan masih terbuka lebar, terutama terkait pemanfaatan dataset berskala lebih besar, pengujian performa algoritma secara lebih menyeluruh, serta eksplorasi metode machine learning yang lebih mutakhir. Oleh sebab itu, penelitian lanjutan disarankan untuk melakukan studi empiris yang lebih mendalam guna menghasilkan model seleksi Beasiswa KIP yang semakin optimal dan berkeadilan [8].

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian secara sistematis sekaligus membahasnya secara analitis berdasarkan tiga pertanyaan penelitian (RQ1–RQ3). Pembahasan dilakukan dengan mengaitkan temuan empiris, teori pembelajaran mesin, serta studi terdahulu yang relevan dalam konteks seleksi beasiswa berbasis machine learning.

Hasil ekstraksi literatur terpilih kemudian disajikan dalam bentuk tabel.

4.1. RQ1: Variabel apa saja yang paling sering digunakan dalam penelitian seleksi penerima Beasiswa KIP berbasis *machine learning*?

RQ1 diarahkan untuk menelaah serta mengklasifikasikan atribut-atribut yang digunakan dalam proses seleksi penerima Beasiswa KIP berdasarkan studi-studi sebelumnya. Mengacu pada Tabel 2, variabel yang teridentifikasi terbagi ke dalam dua kategori utama, yaitu kategori akademik dan kategori sosial ekonomi. Pada kategori akademik, variabel yang paling dominan adalah Indeks Prestasi (IPK) dengan estimasi frekuensi penggunaan sebesar 95%, diikuti nilai akademik (rapor/ujian) sebesar 85%. Kedua variabel ini merepresentasikan konsistensi dan kompetensi dasar calon penerima dalam aspek performa akademik. Sementara itu, pada kategori sosial ekonomi, pendapatan orang tua menjadi atribut dengan tingkat penggunaan tertinggi (98%), diikuti jumlah tanggungan keluarga (90%) serta latar belakang sosial ekonomi (80%) yang mencakup status kepemilikan rumah dan kondisi wilayah tempat tinggal [5].

Identifikasi dan pengelompokan variabel tersebut menjadi krusial karena pemilihan fitur (feature selection) yang tepat berpengaruh langsung terhadap akurasi dan stabilitas model machine learning dalam mengklasifikasikan kelayakan penerima bantuan pendidikan. Studi-studi terbaru dalam lima tahun terakhir menunjukkan bahwa integrasi variabel akademik dan sosial ekonomi mampu meningkatkan performa klasifikasi dibandingkan hanya menggunakan satu kelompok variabel saja [7][12].

Penelitian oleh Hidayat et al. [12] juga menegaskan bahwa atribut pendapatan orang tua dan IPK merupakan faktor paling signifikan dalam model klasifikasi beasiswa berbasis Naïve Bayes, karena keduanya merepresentasikan keseimbangan antara kebutuhan ekonomi dan capaian akademik.

Temuan tersebut diperkuat oleh Thoriq et al. [9] yang menyatakan bahwa penambahan variabel sosial ekonomi seperti jumlah tanggungan keluarga dan status kepemilikan rumah pada model Random Forest secara signifikan meningkatkan nilai akurasi dan precision dalam proses seleksi penerima bantuan pendidikan [6].

Dengan demikian, selaras dengan distribusi persentase pada Tabel 2, dapat disimpulkan bahwa kombinasi variabel akademik dan sosial ekonomi menjadi pendekatan paling relevan dan efektif dalam membangun model prediksi kelayakan Beasiswa KIP, karena mampu mencerminkan aspek meritokrasi sekaligus urgensi kebutuhan finansial secara komprehensif [7].

Tabel 1. Persentase Distribusi Variabel Seleksi Beasiswa KIP

Kategori Utama	Variabel Spesifik	Estimasi Frekuensi Penggunaan (%)	Sitasi
Akademik	Indeks Prestasi (IPK)	95%	[7][12]
	Nilai Akademik (Rapor/Ujian)	85%	[2][3]
	Semester Aktif	70%	[3][12]
	Status Akademik (aktif/non-aktif)	65%	[10][12]
	Jumlah SKS Tempuh	60%	[7][12]
	Riwayat Prestasi Akademik	55%	[5][7]
Non Akademik	Pendapatan Orang Tua	98%	[3][6]
	Pekerjaan Orang Tua	85%	[4][6]
	Jumlah Tanggungan Keluarga	90%	[4][6]
	Status Kepemilikan Rumah	75%	[2][5]
	Kondisi Tempat Tinggal	70%	[5][6]
	Kepemilikan Aset (kendaraan/tanah)	65%	[2][5]
	Bantuan Sosial Lain (PKH/KKS)	60%	[3][6]
	Tagihan Listrik / Daya Listrik	55%	[4][6]
	Wilayah Tempat Tinggal (desa/kota)	50%	[2][5]

4.2. RQ2: Metode *machine learning* apa saja yang paling banyak digunakan dan bagaimana karakteristik kinerjanya dalam seleksi Beasiswa KIP?

RQ2 difokuskan pada penelusuran algoritma machine learning yang paling sering dimanfaatkan dalam proses penentuan penerima Beasiswa Kartu Indonesia Pintar (KIP), sekaligus mengevaluasi karakteristik performanya berdasarkan bukti empiris dari literatur yang telah diseleksi. Dari total 51 artikel yang lolos kriteria inklusi, metode yang paling banyak digunakan adalah Naïve Bayes dan Random Forest, kemudian disusul oleh Support Vector Machine (SVM) serta beberapa pendekatan klasifikasi lain dalam ranah data mining.

Algoritma Naïve Bayes kerap dipilih dalam sistem seleksi beasiswa karena memiliki model probabilistik yang sederhana, proses komputasi yang efisien, serta mudah diintegrasikan ke dalam sistem pendukung keputusan berbasis web [2][3][10][12].

Pendekatan ini berlandaskan konsep probabilitas bersyarat dengan asumsi bahwa setiap atribut bersifat independen. Pada konteks seleksi KIP, metode ini cukup efektif diterapkan pada dataset berskala kecil hingga menengah yang memuat variabel akademik dan sosial ekonomi yang relatif terstruktur.

Studi oleh Suganda et al. [3] mengungkapkan bahwa Naïve Bayes mampu menghasilkan tingkat akurasi yang memadai dalam menentukan penerima KIP Kuliah.

Temuan tersebut selaras dengan penelitian Yunita et al. [10] serta Hidayat et al. [12], yang merekomendasikan algoritma ini sebagai baseline model karena konsistensi kinerjanya. Meskipun demikian, kelemahan utamanya terletak pada asumsi independensi antar variabel, yang dalam praktiknya sering kali tidak sepenuhnya sesuai dengan karakteristik data sosial ekonomi mahasiswa.

Sebaliknya, Random Forest dalam berbagai studi komparatif menunjukkan performa yang lebih kompetitif, terutama karena kemampuannya menangani interaksi multivariat serta pola hubungan non-linear antar atribut [5][9].

Metode ini termasuk dalam kategori ensemble learning yang membangun sejumlah pohon keputusan dan menggabungkan hasil prediksi untuk meningkatkan akurasi sekaligus meminimalkan risiko overfitting.

Penelitian yang dilakukan oleh Thoriq et al. [9] menunjukkan bahwa Random Forest memberikan nilai akurasi dan recall yang lebih tinggi dibandingkan Naïve Bayes pada kasus seleksi KIP di perguruan tinggi.

Hasil tersebut diperkuat oleh Musa [5], yang menyatakan bahwa algoritma ini mampu mengelola variabel akademik dan ekonomi secara bersamaan dengan performa klasifikasi yang lebih stabil. Namun demikian, Random Forest memerlukan sumber daya komputasi yang lebih besar serta proses penyetelan parameter (hyperparameter tuning) yang relatif lebih kompleks.

Selain dua algoritma utama tersebut, beberapa studi juga mengimplementasikan Support Vector Machine (SVM) sebagai alternatif pendekatan klasifikasi [13].

SVM dikenal efektif dalam menangani data berdimensi tinggi serta mampu membentuk batas pemisah kelas yang optimal.

Kendati demikian, penggunaannya dalam konteks seleksi Beasiswa KIP masih belum seintensif Naïve Bayes maupun Random Forest. Zirby et al. [11] menegaskan bahwa tidak ada satu algoritma yang secara universal unggul untuk seluruh jenis data. Pemilihan metode sebaiknya mempertimbangkan karakteristik dataset, distribusi kelas, jumlah variabel, serta kebutuhan operasional sistem seleksi.

Secara umum, hasil sintesis literatur memperlihatkan bahwa Naïve Bayes lebih unggul dari segi efisiensi dan kemudahan implementasi, sementara Random Forest lebih menonjol dalam hal akurasi dan kemampuan mengelola kompleksitas data. Oleh sebab itu, penentuan algoritma pada sistem seleksi Beasiswa KIP perlu disesuaikan dengan kebutuhan institusi, skala dan karakteristik data, serta tujuan implementasi, sehingga proses seleksi dapat berlangsung secara

objektif, transparan, dan berbasis evidensi empiris [6][9][11].

Tabel 3. Karakteristik dan Kinerja Algoritma Machine Learning dalam Seleksi Beasiswa KIP

No	Algoritma	Karakteristik Model	Kecenderungan Kinerja (Literatur)	Sitasi
1	Support Vector Machine (SVM)	Model klasifikasi berbasis hyperplane dengan margin optimal	Akurasi tinggi pada dataset terstruktur; masih terbatas penggunaannya pada studi KIP	[13]
2	Naïve Bayes	Model probabilistik berbasis Teorema Bayes dengan asumsi independensi fitur	Stabil dan sering dijadikan baseline; akurasi cukup baik (>80% pada beberapa studi)	[2][3] [10] [12]
3	Decision Tree	Model berbasis struktur pohon dengan aturan keputusan berbentuk if–then	Efektif untuk analisis awal dan interpretasi variabel	[5][6]
4	K-Nearest Neighbor (KNN)	Metode berbasis jarak (distance-based classification)	Performa bergantung pada nilai K dan normalisasi data	[6]
5	Random Forest	Ensemble learning berbasis banyak Decision Tree dengan teknik bootstrap dan random feature selection	Konsisten menghasilkan akurasi dan recall tertinggi dalam studi komparatif KIP	[5][9] [11]
6	Logistic Regression	Model statistik berbasis fungsi logit untuk klasifikasi biner	Cukup stabil pada data terstruktur; sering digunakan sebagai pembanding statistik	[5]

4.3. RQ3: Metode pengujian dan evaluasi apa yang digunakan untuk menilai kinerja model *machine learning* dalam seleksi Beasiswa KIP?

RQ3 berfokus pada metode pengujian dan evaluasi yang digunakan untuk menilai kinerja model machine learning dalam seleksi Beasiswa KIP. Evaluasi model sangat penting untuk memastikan bahwa sistem seleksi yang dibangun memiliki tingkat akurasi dan keandalan yang dapat dipertanggungjawabkan.

Metode pengujian yang paling umum digunakan dalam penelitian adalah confusion matrix, k-fold cross validation, serta pengukuran metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Beberapa penelitian juga menerapkan pembagian data latih dan data uji untuk menghindari overfitting.

Nurhadi dan Lestari menjelaskan bahwa penggunaan k-fold cross validation mampu memberikan evaluasi model yang lebih stabil dan objektif dibandingkan pembagian data sederhana [11].

Selain itu, Putri dan Wijaya menekankan bahwa penggunaan lebih dari satu metrik evaluasi diperlukan agar kinerja model machine learning dapat dinilai secara komprehensif, khususnya pada data seleksi beasiswa yang tidak seimbang [12][13].

Tabel 4. Metode Evaluasi Model Machine Learning Dalam Seleksi Beasiswa KIP

No	Metode Evaluasi	Deskripsi dan Fungsi	Keunggulan dalam Seleksi KIP	Jurnal Pendukung
1	Confusion Matrix	Matriks evaluasi yang membandingkan hasil prediksi dengan kondisi aktual (TP, TN, FP, FN)	Memberikan gambaran detail kesalahan klasifikasi (false positive & false negative) dalam menentukan kelayakan penerima	[12]
2	Naïve Bayes	Model probabilistik berbasis Teorema Bayes dengan asumsi independensi fitur	Stabil dan sering dijadikan baseline; akurasi cukup baik (>80% pada beberapa studi)	[2][3][10] [12]
3	Decision Tree	Model berbasis struktur pohon dengan aturan keputusan berbentuk if–then	Efektif untuk analisis awal dan interpretasi variabel	[5][6]
4	K-Nearest Neighbor (KNN)	Metode berbasis jarak (distance-based classification)	Performa bergantung pada nilai K dan normalisasi data	[6]
5	Random Forest	Ensemble learning berbasis banyak Decision Tree dengan teknik bootstrap dan random feature selection	Konsisten menghasilkan akurasi dan recall tertinggi dalam studi komparatif KIP	[5][9][11]
6	Logistic Regression	Model statistik berbasis fungsi logit untuk klasifikasi biner	Cukup stabil pada data terstruktur; sering digunakan sebagai pembanding statistik	[5]

5. KESIMPULAN

Berdasarkan temuan Systematic Literature Review (SLR) yang disusun mengacu pada Research Question (RQ) yang telah ditetapkan, dapat disimpulkan bahwa efektivitas penerapan metode machine learning dalam proses seleksi penerima Beasiswa Kartu Indonesia Pintar (KIP) sangat dipengaruhi oleh ketepatan pemilihan variabel, jenis algoritma yang digunakan, serta pendekatan pengujian dan evaluasi model yang diterapkan.

Hasil kajian pada RQ1 menunjukkan bahwa variabel yang paling sering dan berpengaruh dalam sistem seleksi Beasiswa KIP berasal dari aspek akademik dan kondisi sosial ekonomi mahasiswa. Indikator seperti indeks prestasi, nilai akademik, pendapatan orang tua, jumlah tanggungan keluarga, serta latar belakang sosial ekonomi terbukti memiliki kontribusi signifikan dalam menentukan hasil klasifikasi kelayakan penerima beasiswa. Penggunaan kombinasi variabel akademik dan ekonomi secara konsisten menghasilkan kinerja model yang lebih optimal dibandingkan penggunaan satu jenis variabel saja, sehingga pemilihan serta kualitas variabel input menjadi faktor krusial dalam keberhasilan model machine learning.

Berdasarkan analisis RQ2, algoritma machine learning yang paling banyak diimplementasikan dalam penelitian seleksi Beasiswa KIP adalah Naïve Bayes dan Random Forest. Naïve Bayes banyak digunakan karena memiliki struktur yang sederhana, efisiensi komputasi yang tinggi, serta kemudahan dalam implementasi, sehingga sering dijadikan sebagai model pembanding atau baseline. Di sisi lain, Random Forest menunjukkan performa yang lebih baik dalam hal akurasi dan kemampuan menangani hubungan data yang kompleks, multivariat, dan non-linear, meskipun membutuhkan sumber daya komputasi yang lebih besar. Temuan ini mengindikasikan bahwa pemilihan algoritma harus mempertimbangkan karakteristik data, skala sistem, serta tujuan penerapan seleksi Beasiswa KIP.

Sementara itu, hasil kajian pada RQ3 mengungkapkan bahwa metode evaluasi yang paling umum digunakan meliputi confusion matrix dengan metrik akurasi, presisi, recall, dan F1-score, serta penerapan teknik k-fold cross validation. Penggunaan berbagai metrik evaluasi dinilai penting untuk memperoleh pemahaman yang lebih menyeluruh terhadap kinerja model, terutama pada permasalahan seleksi beasiswa yang umumnya menghadapi ketidakseimbangan kelas data. Selain itu, penerapan cross validation mampu menghasilkan evaluasi yang lebih reliabel dan meminimalkan risiko terjadinya overfitting.

Secara keseluruhan, sintesis hasil RQ1, RQ2, dan RQ3 menunjukkan bahwa keberhasilan implementasi machine learning dalam seleksi Beasiswa KIP tidak hanya bergantung pada pemilihan algoritma, tetapi juga ditentukan oleh kualitas variabel input serta ketepatan metode evaluasi yang digunakan. Oleh karena itu, penelitian selanjutnya disarankan untuk mengembangkan model seleksi Beasiswa KIP dengan cakupan variabel yang lebih luas, melakukan perbandingan algoritma secara empiris, serta menerapkan teknik evaluasi yang lebih kuat agar sistem seleksi yang dihasilkan menjadi lebih objektif, akurat, dan aplikatif dalam skala yang lebih luas.

DAFTAR PUSTAKA

[1] Andika, I., Nevile, S., Satya, R., Farisi, A., Informasi, S., & Ilmu Komputer Dan Rekayasa, F. (2024). Analisis Sistem Informasi Manajemen Proyek: Systematic Literature Review Information System Management Project Analysis : Systematic Literature Review (Vol. 11, Number 1). Http://Jurnal.Mdp.Ac.Id

[2] Badriyah, N., Satyareni, D. H., & Anugrah, C. S. (2025). Sistem Pendukung Keputusan Kelayakan Beasiswa KIP-K Berbasis Web Menggunakan Algoritma Gaussian-Naïve Bayes (Vol. 4).

[3] Gagan Suganda, Marsani Asfi, Ridho Taufiq Subagio, & Ricky Perdana Kusuma. (2022). Penentuan Penerima Bantuan Beasiswa Kartu Indonesia Pintar (Kip) Kuliah Menggunakan Naïve Bayes Classifier. Jsii (Jurnal Sistem Informasi), 9(2), 193–199. Https://Doi.Org/10.30656/Jsii.V9i2.4376

[4] [4] Juniawan, M. A., & Sudrajat, A. (2025). Prediksi Seleksi Penerima Beasiswa Kip Kuliah Menggunakan Algoritma Naïve Bayes (Studi Kasus Politeknik Tedc Bandung). In Jurnal Mahasiswa Teknik Informatika) (Vol. 9, Number 6).

[5] Musa, A. B. (2024). Understanding Student Performance In Foundation Year: Insights From Logistic Regression, Naïve Bayes, And Random Forest Models. International Journal Of Information And Education Technology, 14(12), 1716–1723. Https://Doi.Org/10.18178/Ijiet.2024.14.12.2202

[6] Nata, A., & Royal, S. (2022). Analisis Sistem Pendukung Keputusan Dengan Model Klasifikasi Berbasis Machine Learning Dalam Penentuan Penerima Program Indonesia Pintar. In Journal Of Science And Social Research (Number 3). Http://Jurnal.Goretanpena.Com/Index.Php/JSSR

[7] Rizki Siregar, A., & Furqan, Mhd. (2025). Classification Of Scholarships For Students In Schools Using The Naïve Bayes Method. Journal Of Computer Networks, Architecture And High Performance Computing, 7(1), 278–289. Https://Doi.Org/10.47709/Cnahpc.V7i1.5417

[8] Sharief, S., Balaji, M. B., & Prof, A. (N.D.). Scholarship Prediction System Using Machine Learning. International Journal Of Scientific Research And Engineering Development, 8. Retrieved Www.Ijsred.Com

[9] Thoriq, M., Maulana, F., Qurrata Ayun, A., & Adzkia, U. (2025). Perbandingan Naïve Bayes

Dan Random Forest Pada Seleksi KIP Di Universitas Adzkia. Random Forest, Pembelajaran Mesin, Prediksi Beasiswa, Seleksi KIP, 5(3), 2809–4069. Https://Doi.Org/10.55382/Jurnalpustakaai.V5i3.1385

[10] Yunita, Yunita, N. Hartono, & Erfina, E. (2025). Penerapan Data Mining Dalam Sistem Pendukung Keputusan Penerimaan Beasiswa KIP-Kuliah Menggunakan Algoritma Naive Bayes. Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer, 5(1), 154–175.

[11] Zirby, Q., Nafi'iyah, N., & Budi, A. S. (2025). *Penerapan Machine Learning Untuk Prediksi Penerima Beasiswa Pada Universitas Muhammadiyah Pringsewu*. Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI), 8(5).

[12] H. Hidayat, M. Musliadi KH, I. Kusmanto, M. Kadir, & K. Kristian, *"Klasifikasi Mahasiswa Calon Penerima Beasiswa KIP Menggunakan Algoritma Naive Bayes Di Universitas Tomakaka Mamuju"*, Jurnal FASILKOM (Teknologi Informasi Dan Ilmu Komputer), Vol. 15, No. 3, 2025.

[13] I. Amelia, S. Sugiyono, F. M. Sarimole, & Tundo, *"Analisis Sentimen Tanggapan Pengguna Media Sosial X Terhadap Program Beasiswa KIP-Kuliah Dengan Menggunakan Algoritma Support Vector Machine (SVM)"*, Jurnal Indonesia: Manajemen Informatika Dan Komunikasi, Vol. 5, No. 3, 2024.