JURNAL ILMIAH SISTEM INFORMASI (JUSI)
VOL 4, No. 1, Januari 2025, pp. 191 - 204
p-ISSN : 2803-1507 (media cetak) | e-ISSN : 2803-1531 (media online)
Link Publikasi: https://ejurnal.provisi.ac.id/index.php/JUISI/issue/view/126

Implementation of Naïve Bayes Algorithm on the Eligibility of Kartu Indonesia Pintar Scholarship (Case Study: University of Sepuluh Nopember Papua)

Daniel Bienfield Manahan Siahaan*[1], Estevina Carolina Bagre[2], Jered Imanuel Wanda[3], Grisye Silahooy[4], Heru Sutejo[5]
[1,2,3,4,5] Program Studi Sistem Informasi, Universitas Sepuluh Nopember Papua
Jl. Ardipura Raya No. 22B, 99222, Ardipura, Distrik Jayapura Selatan, Kota Jayapura, Papua, Indonesia
e-mail: danielsiahaan2603@gmail.com*[1], estevinacarolina@gmail.com[2], imanuellwanda@gmail.com[3], silahooyisye@gmail.com[4], heru.sutejo01@gmail.com[5]

ARTICLE INFO

Article History:

Received: 22-10-2025
Revised: 11-11-2025
Accepted: 02-12-2025
Published: 30-12-2025

ABSTRACT

Indonesia's KIP Kuliah scholarship aims to widen equitable access to higher education, yet manual campus-level screening can be subjective and difficult to audit. This study proposes a data-driven selection aid using a Naïve Bayes classifier to support decisions at Universitas Sepuluh Nopember Papua. The objectives are twofold: (1) to design and implement a transparent, replicable model for predicting scholarship eligibility and (2) to evaluate its performance against standard classification metrics. We adopt a KDD workflow implemented in RapidMiner that covers data import, quality control, missing-value imputation, role assignment, and encoding. The dataset contains 543 applicants from the 2024–2025 intake, with socio-economic attributes including parents' occupations and income brackets, number of dependents, DTKS registration, and P3KE deciles; historical eligibility serves as the target label. Model training uses Laplace smoothing; evaluation relies on a held-out test set of 50 records. The model attains 94% accuracy, with a confusion matrix of TP=45, FP=2, FN=1, TN=2, and precision and recall for the Eligible class of 95.74% and 97.83%, respectively, and an AUC of 0.891, indicating strong class separability. Findings show that the approach reliably identifies eligible candidates while highlighting residual sensitivity in the non-eligible minority class. The contribution is a lightweight, auditable pipeline that accelerates screening, reduces subjectivity, and strengthens accountability through measurable outputs. Practically, the model can serve as a first-stage filter to focus committee review, with future gains expected from class balancing, threshold tuning, and periodic retraining to sustain fairness and efficiency.

Keywords: *KIP Kuliah Scholarship, Naïve Bayes, Data Mining, KDD, RapidMiner.*

DOI: https://doi.org/10.51903/rdzdm469

Abstrak

Program KIP Kuliah bertujuan memperluas akses pendidikan tinggi yang adil, namun seleksi manual di tingkat kampus kerap subjektif dan sukar diaudit. Penelitian ini menawarkan dukungan seleksi berbasis data menggunakan klasifikator Naive Bayes untuk membantu pengambilan keputusan di Universitas Sepuluh Nopember Papua. Tujuan penelitian adalah: (1) merancang dan mengimplementasikan model yang transparan dan replikabel untuk memprediksi kelayakan beasiswa, serta (2) mengevaluasi kinerjanya dengan metrik klasifikasi standar. Metode yang digunakan mengikuti alur KDD di RapidMiner, mencakup impor data, kendali mutu, imputasi nilai hilang, penetapan peran atribut, dan penyandian fitur; pelatihan model menggunakan smoothing Laplace. Dataset berisi 543 pendaftar periode 2024–2025 dengan atribut sosioekonomi (pekerjaan dan penghasilan orang tua, jumlah tanggungan, status DTKS, desil P3KE), sementara label target berupa kelayakan historis. Evaluasi dilakukan pada himpunan uji berukuran 50. Hasil menunjukkan akurasi 94% dengan matriks kebingungan TP=45, FP=2, FN=1, TN=2; untuk kelas Layak, presisi 95,74% dan recall 97,83%; AUC 0,891 mengindikasikan pemisahan kelas yang kuat. Temuan ini membuktikan pendekatan mampu mengenali kandidat layak secara andal, seraya menyoroti keterbatasan sensitivitas pada kelas Tidak Layak yang minoritas. Kontribusi penelitian adalah rancangan pipa analitik yang ringan dan dapat diaudit, yang mempercepat penyaringan, mengurangi subjektivitas, serta memperkuat akuntabilitas melalui keluaran terukur. Implikasinya, model dapat berperan sebagai filter tahap awal untuk memfokuskan telaah komite; peningkatan ke depan mencakup penyeimbangan kelas, penyetelan ambang, dan pelatihan ulang berkala demi menjaga keadilan dan efisiensi**.**

Kata Kunci: *KIP Kuliah, Naive Bayes, Data Mining, KDD, RapidMiner.*

1. PENDAHULUAN

Pendidikan memiliki posisi strategis dalam menyiapkan sumber daya manusia yang andal bagi pembangunan nasional pada era global [1]. Dalam konteks kompetisi pengetahuan dan teknologi, perguruan tinggi dituntut melahirkan lulusan yang adaptif serta berdaya saing. Namun, di tengah dorongan untuk menempuh pendidikan tinggi, masih terdapat kelompok masyarakat yang terhambat melanjutkan studi, terutama karena keterbatasan ekonomi [2]. Kondisi ini menimbulkan kesenjangan partisipasi yang jika tidak ditangani akan memperlebar ketidaksetaraan. Untuk memperkecil kesenjangan akses tersebut, pemerintah menyelenggarakan program Kartu Indonesia Pintar Kuliah sebagai intervensi afirmatif yang menempatkan pemerataan kesempatan belajar sebagai prioritas kebijakan [3]. Program ini memberikan dukungan biaya dan mendorong keberlanjutan studi sehingga potensi putus kuliah dapat ditekan.

Pada tataran kelembagaan, Universitas Sepuluh Nopember Papua membuka akses penerimaan mahasiswa melalui skema Kartu Indonesia Pintar Kuliah dan menugaskan unit Penerimaan Mahasiswa Baru untuk melakukan seleksi serta pengelolaan program. Di level operasional, proses seleksi memerlukan konsistensi kriteria, ketertelusuran keputusan, dan dokumentasi yang baik. Prosedur seleksi yang masih bersifat manual berpotensi menimbulkan subjektivitas dan ketidaktepatan sasaran sehingga sebagian calon yang layak tidak teridentifikasi, sementara pihak yang kurang memenuhi kriteria kemungkinan justru ditetapkan sebagai penerima [4]. Implikasi dari kondisi tersebut bukan hanya ketidakadilan bagi calon mahasiswa, tetapi juga inefisiensi alokasi sumber daya beasiswa.

Untuk meningkatkan akurasi dan akuntabilitas keputusan, penelitian ini mengajukan pemanfaatan algoritma *Naïve Bayes* sebagai model klasifikasi yang memanfaatkan data historis beserta atribut pendaftar [5]. Pendekatan ini relevan karena mampu menghasilkan estimasi probabilistik berbasis bukti data, mudah diimplementasikan, serta efisien pada ukuran data menengah. Melalui estimasi probabilistik yang sederhana namun efektif, model diharapkan memberikan rekomendasi penetapan kelayakan yang konsisten, transparan, serta mudah direplikasi pada siklus seleksi berikutnya. Selain itu, keluaran model dapat menjadi bahan umpan balik untuk penyempurnaan kriteria seleksi di periode mendatang.

Rumusan masalah penelitian adalah sebagai berikut. Pertama, bagaimana merancang dan mengimplementasikan model *Naïve Bayes* guna memprediksi kelayakan penerima beasiswa Kartu Indonesia Pintar Kuliah di Universitas Sepuluh Nopember Papua. Kedua, seberapa baik kinerja model tersebut ditinjau dari akurasi, presisi, *recall*, dan *F1-score* dalam meningkatkan ketepatan keputusan dibandingkan prosedur seleksi manual yang selama ini diterapkan [6]. Perumusan ini menegaskan fokus pada aspek metodologis (perancangan model) sekaligus aspek evaluatif (pembuktian kinerja terhadap tolok ukur yang dapat diaudit).

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Tujuan penelitian sejalan dengan rumusan masalah tersebut. Pertama, membangun model *Naïve Bayes* untuk klasifikasi kelayakan penerima Kartu Indonesia Pintar Kuliah berdasarkan indikator akademik dan sosioekonomi. Kedua, mengevaluasi performa model pada data institusi menggunakan metrik akurasi, presisi, *recall,* dan *F1-score* [7]. Manfaat penelitian mencakup tersedianya alat bantu keputusan yang objektif dan terukur bagi unit Penerimaan Mahasiswa Baru, peningkatan ketepatan sasaran program sehingga lebih berkeadilan, serta kontribusi metodologis berupa penerapan pembelajaran mesin yang dapat direplikasi pada seleksi beasiswa serupa di perguruan tinggi lain. Secara praktis, hasil penelitian juga diharapkan memperkuat tata kelola seleksi melalui dokumentasi proses dan pelaporan metrik yang mudah dipahami.

Ruang lingkup dan batasan penelitian mencakup penggunaan 543 data pendaftar Kartu Indonesia Pintar Kuliah di Universitas Sepuluh Nopember Papua pada periode 2024 sampai 2025. Pengelolaan data menggunakan aplikasi *RapidMiner*, meliputi impor, pembersihan dan imputasi nilai hilang, penyandian atau pendiskretan variabel sesuai kebutuhan *Naïve Bayes* [8], pembagian data latih–uji atau k-fold cross validation, pelatihan model, serta evaluasi dengan akurasi, presisi, *recall*, *F1-score*, dan *confusion matrix* [9]. Indikator penilaian meliputi pekerjaan ayah dan ibu, penghasilan ayah dan ibu, kepemilikan atau nomor KKS, status DTKS, serta status P3KE beserta desil; kelayakan KIP historis digunakan sebagai label. Kolom nama siswa berfungsi sebagai identitas dan tidak dijadikan fitur. Penghasilan dan desil P3KE diperlakukan sebagai kategori berjenjang, sedangkan pekerjaan, DTKS, dan KKS sebagai kategori. Seluruh proses mematuhi etika penelitian, termasuk persetujuan institusi, anonimisasi data, dan pembatasan akses.

2. TINJAUAN PUSTAKA

2.1. *Kartu Indonesia Pintar Kuliah (KIP Kuliah)*

Berbagai penelitian menunjukkan bahwa KIP Kuliah dimanfaatkan sebagai instrumen kebijakan untuk memperluas akses pendidikan tinggi bagi mahasiswa dari latar belakang ekonomi rentan serta menekan risiko putus studi akibat keterbatasan finansial [10]. Pada praktiknya, proses seleksi di perguruan tinggi melibatkan verifikasi administratif dan penilaian kondisi sosial-ekonomi, namun pendekatan manual dilaporkan berpotensi menimbulkan ketidakkonsistenan keputusan dan kesalahan sasaran [11]. Sejalan dengan temuan tersebut, sejumlah studi mengusulkan pemanfaatan pendekatan analitik berbasis data untuk mendukung proses seleksi agar lebih objektif, transparan, dan dapat diaudit. Dalam penelitian ini, kelayakan KIP Kuliah dimodelkan sebagai masalah klasifikasi biner dengan memanfaatkan variabel sosioekonomi mahasiswa dan rumah tangga, serta menerapkan algoritma *Multinomial Naïve Bayes* yang sesuai dengan karakteristik data kategorikal dan ordinal.

2.2. *Basis Data Sosial: KKS, DTKS, dan P3KE*

Literatur terkait menunjukkan bahwa basis data sosial nasional sering digunakan untuk meningkatkan ketepatan sasaran program bantuan, termasuk dalam konteks pendidikan tinggi. Kepemilikan KKS banyak dimanfaatkan sebagai indikator administratif awal untuk mengidentifikasi rumah tangga dengan tingkat kerentanan ekonomi tertentu dalam berbagai studi penargetan bantuan [12]. Selain itu, status keterdaftaran dalam DTKS dipandang relevan karena data tersebut telah melalui proses verifikasi dan pemutakhiran resmi, sehingga sering dijadikan rujukan dalam penelitian maupun kebijakan [13]. P3KE melengkapi informasi tersebut melalui pemetaan kesejahteraan berbasis desil yang memungkinkan diferensiasi tingkat kerentanan secara lebih granular, sehingga penggunaannya dinilai efektif dalam mendukung analisis prioritas intervensi [14].

2.3. *Research Gap*

Sejumlah penelitian telah mengkaji pendekatan klasifikasi, baik pada konteks umum maupun pendidikan [15], [16], [17]. Namun, kajian yang secara spesifik mengembangkan model prediksi kelayakan KIP Kuliah berbasis data mahasiswa pada konteks institusi perguruan tinggi lokal dengan evaluasi kinerja yang terstandar masih terbatas. Selain itu, aspek keadilan model atau *fairness* antar subkelompok mahasiswa, seperti jurusan, asal daerah, dan gender, belum banyak dianalisis secara sistematis dalam penelitian terdahulu, meskipun isu tersebut berpotensi memengaruhi tingkat salah tolak penerima bantuan. Penelitian sebelumnya juga jarang membahas pengaruh *class imbalance* serta belum banyak melakukan pembandingan terstruktur dengan model *baseline* dan evaluasi kalibrasi probabilitas, sehingga implikasi operasional hasil prediksi terhadap pengambilan keputusan masih belum sepenuhnya terjelaskan.

2.4. *Data Mining*

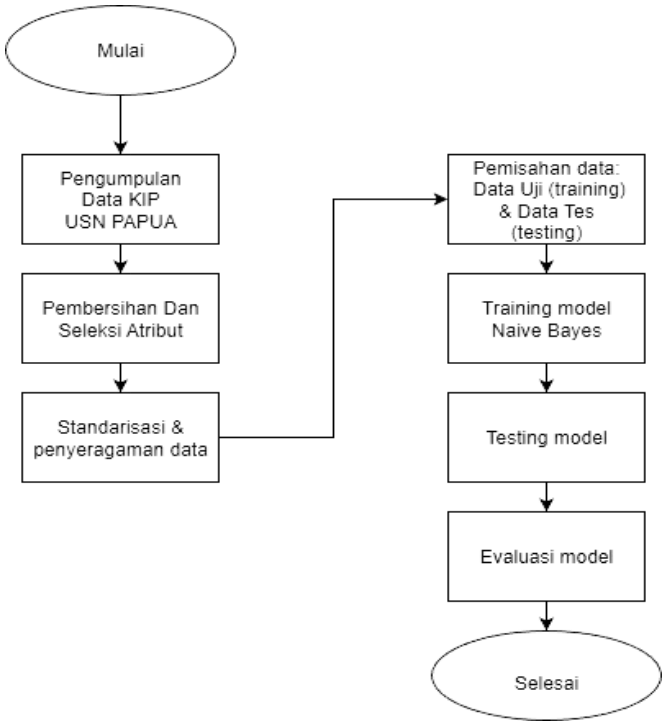
Pendekatan *data mining* telah banyak diterapkan dalam penelitian untuk mengekstraksi pola dan pengetahuan dari kumpulan data berukuran besar yang sulit dianalisis secara manual. Sejumlah studi memanfaatkan kerangka *Knowledge Discovery in Databases* (KDD) sebagai proses sistematis yang mencakup tahap prapemrosesan, pemodelan, dan evaluasi hasil [18]. Dalam bidang pendidikan dan bantuan sosial, *data mining* digunakan untuk mengklasifikasikan penerima bantuan, memprediksi risiko akademik, serta mendukung pengambilan keputusan berbasis data. Relevansi pendekatan ini dalam penelitian terletak pada kemampuannya mengolah data sosioekonomi mahasiswa secara terstruktur dan menghasilkan model prediktif yang dapat diuji serta direplikasi.

2.5. *Rapid Miner*

Beberapa penelitian empiris melaporkan penggunaan RapidMiner sebagai platform analitik karena kemampuannya mendukung seluruh tahapan *data mining* dalam satu lingkungan terpadu. RapidMiner banyak dimanfaatkan pada tahap prapemrosesan, termasuk penanganan nilai hilang, transformasi atribut, dan penyandian variabel kategorikal, yang krusial dalam pengolahan data sosial. Dalam pemodelan dan evaluasi, studi-studi terdahulu menggunakan skema pembagian data seperti *hold-out* dan *k-fold cross validation* untuk memperoleh estimasi kinerja yang lebih andal. Ketersediaan operator *Naïve Bayes* beserta metrik evaluasi seperti akurasi, presisi, recall, *F1-score*, dan *confusion matrix* menjadikan RapidMiner relevan untuk mendukung eksperimen yang transparan, replikatif, dan mudah diaudit [19].

3. METODOLOGI PENELITIAN

Bagian ini menguraikan tahapan metodologis yang digunakan untuk membangun dan mengevaluasi model klasifikasi kelayakan KIP Kuliah. Seluruh proses disusun secara sistematis agar setiap langkah dapat ditelusuri kembali serta memberikan hasil analitis yang dapat dipertanggungjawabkan. Tahapan dimulai dari pengumpulan data pendaftar yang menjadi dasar analisis, dilanjutkan dengan proses pembersihan serta transformasi untuk memastikan kualitas dan konsistensi data. Setelah itu, model dibangun menggunakan algoritma *Naïve Bayes* yang dipilih karena kesederhanaan, efisiensi, dan kemampuannya menangani data kategorikal. Terakhir, model yang dihasilkan diuji menggunakan skema evaluasi terstruktur untuk menilai tingkat keakuratan dan ketepatannya [20]. Alur keterkaitan antar tahapan metodologis tersebut divisualisasikan dalam bentuk diagram alir sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alir Naïve Bayes [21]

3.1. *Pengumpulan Data*

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Proses dimulai dari tahapan pengumpulan data sebagaimana ditunjukkan pada alur awal dalam *flowchart*. Pada tahap ini, dataset berisi 543 rekaman pendaftar KIP Kuliah Universitas Sepuluh Nopember Papua untuk periode 2024–2025 dihimpun dari arsip resmi Unit Penerimaan Mahasiswa Baru. Setiap rekaman memuat atribut yang relevan dengan proses penilaian kelayakan bantuan pendidikan. Variabel-variabel tersebut mencakup pekerjaan ayah, pekerjaan ibu, penghasilan ayah, penghasilan ibu, status ketercakupan pada DTKS, serta status P3KE berikut desil yang menyertainya sebagai indikator kesejahteraan keluarga. Seluruh atribut ini disiapkan sebagai variabel prediktor, sedangkan variabel keluarannya berupa riwayat keputusan kelayakan KIP yang terdiri atas dua kelas: layak dan tidak layak. Seluruh proses mulai dari pemanggilan data, pengelolaan atribut, hingga tahap evaluasi selanjutnya dikerjakan menggunakan perangkat lunak *RapidMiner* [22].

3.2. *Pembersihan dan Transformasi Data*

Setelah pengumpulan, data dipersiapkan melalui satu tahap ringkas yang mencakup pembersihan dan transformasi agar konsisten serta layak dimodelkan: entri ganda dihapus, format nilai diseragamkan, dan kekosongan ditangani dengan imputasi yang sesuai karakter variabel; kategori juga diharmonisasi dengan menyatukan penamaan pekerjaan, mengurutkan rentang penghasilan, serta menegaskan konsistensi label desil P3KE, kemudian seluruh fitur kategorikal dikodekan agar kompatibel dengan *Naïve Bayes*. Peran atribut ditetapkan jelas, yaitu prediktor sebagai *regular attributes*, target kelayakan sebagai *label*, sedangkan identitas seperti nama siswa disimpan hanya sebagai referensi dan tidak ikut dilatih [23]. Untuk menjaga keandalan estimasi kinerja, evaluasi utama menggunakan validasi silang k lipat, dengan skema bagi latih dan uji yang proporsional sebagai opsi pembanding bila diperlukan [24]. Selain itu, proses transformasi data juga mencakup penyesuaian skala nilai numerik agar setiap atribut memiliki rentang yang sebanding sehingga tidak mendominasi perhitungan probabilitas pada metode *Naïve Bayes*.

3.3. *Penerapan Metode Naïve Bayes pada Penentuan Kelayakan Beasiswa KIP*

Dalam penelitian ini, penentuan kelayakan penerima Beasiswa Kartu Indonesia Pintar (KIP) di Universitas Sepuluh Nopember Papua dilakukan dengan memanfaatkan metode *Naïve Bayes* sebagai alat klasifikasi. Metode ini digunakan untuk memproses data pendaftar berdasarkan atribut-atribut yang berkaitan dengan kondisi akademik dan sosial ekonomi. Pendekatan *Naïve Bayes* memanfaatkan prinsip probabilistik untuk mengaitkan karakteristik calon mahasiswa dengan kelas kelayakan yang tersedia. Proses klasifikasi dilakukan berdasarkan pola hubungan antara atribut dan kelas yang terbentuk dari data historis.

Metode *Naïve Bayes* mengelompokkan calon penerima beasiswa ke dalam dua kelas utama, yaitu c_1 sebagai kategori *Layak* menerima Beasiswa KIP dan c_2 sebagai kategori *Tidak Layak* menerima Beasiswa KIP. Setiap calon penerima beasiswa direpresentasikan dalam bentuk sebuah vektor fitur $x = (x_1,x_2,\ldots,x_n)$ yang menggambarkan karakteristik atau kriteria penilaian masing-masing calon. Representasi vektor ini digunakan sebagai dasar perhitungan probabilitas pada setiap kelas. Setiap atribut dalam vektor diasumsikan saling bebas bersyarat terhadap kelas, sesuai dengan asumsi dasar *Naïve Bayes*.

Fitur-fitur yang digunakan dalam penelitian ini mencerminkan aspek akademik dan kondisi sosial ekonomi calon penerima beasiswa. Atribut tersebut meliputi kondisi ekonomi orang tua (x_1), besar penghasilan orang tua (x_2), jumlah anggota keluarga yang menjadi tanggungan (x_3), nilai rapor atau Indeks Prestasi Kumulatif (IPK) (x_4), serta prestasi akademik maupun non-akademik (x_5). Selain atribut tersebut, dimungkinkan adanya atribut lain yang relevan dengan ketentuan seleksi Beasiswa KIP. Seluruh atribut diperlakukan sebagai variabel input dalam proses perhitungan probabilitas.

Perhitungan probabilitas kelas *Layak* dilakukan dengan menggunakan pendekatan teorema Bayes. Perhitungan ini melibatkan probabilitas awal kelas dan probabilitas bersyarat dari setiap atribut terhadap kelas tersebut. Peluang seorang calon mahasiswa dengan profil x termasuk ke dalam kelas *Layak* dinyatakan sebagai berikut:

$$P(Layak|x_1 \ldots, x_n) \propto P(Layak) \prod_{i=1}^{n} P(x_i|Layak) \quad (1)$$

Persamaan tersebut menunjukkan hubungan antara probabilitas prior kelas *Layak* dan kontribusi setiap atribut terhadap kelas tersebut. Nilai probabilitas yang dihasilkan mencerminkan tingkat keterkaitan profil calon mahasiswa dengan kelas *Layak*. Perhitungan ini dilakukan untuk setiap data pendaftar yang dianalisis oleh model.

Perhitungan probabilitas untuk kelas *Tidak Layak* dilakukan dengan prinsip yang sama seperti pada kelas *Layak*. Perbedaan terletak pada probabilitas prior kelas dan probabilitas bersyarat atribut terhadap kelas *Tidak Layak*. Peluang tersebut dirumuskan sebagai berikut:

$$P(Tidak\ Layak|x_1, \dots, x_n \propto P(Tidak\ Layak) \prod_{i=1}^{n} P(x_i|Tidak\ Layak) \quad (2)$$

Nilai probabilitas yang diperoleh dari persamaan ini menggambarkan tingkat keterkaitan profil calon mahasiswa dengan kelas *Tidak Layak*. Probabilitas tersebut dihitung untuk setiap atribut yang dimiliki calon mahasiswa. Hasil perhitungan digunakan sebagai dasar pembandingan dengan probabilitas kelas *Layak*.

Penentuan kelas kelayakan dilakukan dengan membandingkan nilai probabilitas dari kedua kelas yang tersedia. Proses ini mengikuti aturan pemilihan kelas dengan nilai probabilitas terbesar. Aturan pengambilan keputusan dalam metode *Naïve Bayes* dinyatakan sebagai berikut:

$$\hat{y} = \arg \underset{c \in \{Layak, Tidak\ Layak\}}{max} \left[P(c) \prod_{i=1}^{n} P(x_i|c) \right] \quad (3)$$

Persamaan ini menggambarkan mekanisme pemilihan kelas berdasarkan hasil perhitungan probabilitas. Setiap kelas dievaluasi menggunakan probabilitas prior dan probabilitas bersyarat atribut. Proses ini dilakukan secara konsisten untuk seluruh data pendaftar.

Secara operasional, klasifikasi kelayakan dilakukan dengan membandingkan nilai probabilitas kelas *Layak* dan kelas *Tidak Layak*. Kondisi perbandingan tersebut dinyatakan dalam bentuk ketidaksamaan sebagai berikut:

$$P(Layak) \prod_{i=1}^{n} P(x_i|Layak) \geq P(Tidak\ Layak) \prod_{i=1}^{n} P(x_i|Tidak\ Layak) \quad (4)$$

Ketidaksamaan ini digunakan sebagai kriteria dalam proses klasifikasi kelayakan. Apabila kondisi tersebut terpenuhi, data pendaftar ditempatkan pada kelas *Layak*. Apabila kondisi tersebut tidak terpenuhi, data pendaftar ditempatkan pada kelas *Tidak Layak*.

3.4. *Pemodelan*

Pada tahap pelatihan, algoritma Naive Bayes diimplementasikan menggunakan *RapidMiner* [25]. Mengingat mayoritas fitur bersifat kategorikal atau ordinal, penelitian ini menetapkan Multinomial Naive Bayes sebagai model utama; *varian Bernoulli* dipertimbangkan ketika penyandian menghasilkan fitur biner, sedangkan *Gaussian Naive Bayes* digunakan pada skenario eksploratif yang tetap mempertahankan atribut kontinu [26]. Untuk mencegah probabilitas nol pada kategori yang jarang, diterapkan *Laplace smoothing* sehingga setiap nilai kategori tetap memiliki peluang kontribusi pada estimasi. Hasil pelatihan berupa pemetaan probabilistik yang mengklasifikasikan profil pendaftar ke kelas layak atau tidak layak berdasarkan pola statistik pada data historis. Penulisan tipe Naive Bayes secara spesifik perlu ditegaskan karena memengaruhi asumsi terhadap fitur serta cara menafsirkan keluaran model.

3.5. *Evaluasi*

Langkah terakhir yang ditunjukkan dalam diagram alir adalah "Evaluasi model", yaitu proses penilaian kemampuan model dalam membedakan kedua kelas kelayakan secara objektif. Evaluasi dilakukan menggunakan validasi silang untuk memperoleh estimasi kinerja yang konsisten dan tidak bias terhadap pembagian data tertentu [27]. Proses evaluasi ini bertujuan untuk memastikan bahwa model mampu mempertahankan performa yang stabil ketika diterapkan pada subset data yang berbeda selama proses pengujian. Selain itu, penggunaan validasi silang memungkinkan pemanfaatan data secara lebih optimal karena setiap data berkesempatan menjadi data latih dan data uji.

Prediksi yang dihasilkan oleh model kemudian dibandingkan dengan label aktual menggunakan matriks kebingungan, yang berisi empat komponen utama: benar positif, salah positif, salah negatif, dan benar

p-ISSN : 2803-1507 e-ISSN : 2803-1531

negatif. Berdasarkan matriks tersebut dihitung metrik evaluasi utama, meliputi akurasi sebagai ukuran ketepatan keseluruhan, presisi sebagai indikator ketepatan prediksi pada kelas layak, *recall* sebagai ukuran kemampuan model menangkap seluruh pendaftar yang seharusnya layak, serta skor F1 yang merupakan kombinasi harmonis antara presisi dan *recall* sehingga memberikan gambaran menyeluruh tentang performa model saat kedua aspek tersebut dipertimbangkan bersamaan [28]. Setiap metrik evaluasi memberikan sudut pandang yang berbeda terhadap kinerja model dalam mengklasifikasikan data pendaftar beasiswa. Nilai metrik tersebut digunakan untuk menganalisis kekuatan dan keterbatasan model dalam mengenali pola kelayakan penerima Beasiswa KIP**.**

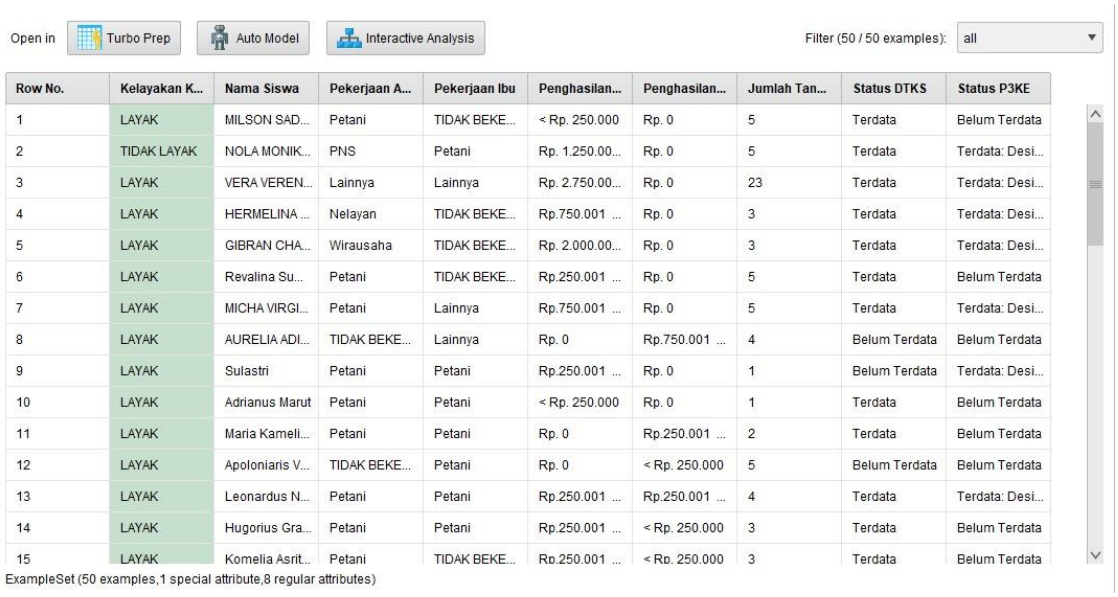
4. HASIL DAN PEMBAHASAN

4.1. *Pengumpulan Data*

Pada penelitian ini, data dikumpulkan dari arsip resmi Universitas Sepuluh Nopember Papua yang memuat 543 pendaftar KIP Kuliah periode 2024–2025 beserta informasi sosial ekonomi. Seluruh identitas pribadi seperti nama, NIK, nomor KK, alamat, dan nomor telepon dikecualikan dari himpunan analisis untuk menjaga kerahasiaan peserta dan memfokuskan analisis pada faktor penentu kelayakan. *Label* yang diprediksi adalah status kelayakan KIP dengan dua kategori, yaitu LAYAK dan TIDAK LAYAK, yang merefleksikan hasil verifikasi universitas pada periode sebelumnya. *Prediktor* yang digunakan mencakup pekerjaan ayah dan ibu (kategorikal), penghasilan orang tua dalam rentang pendapatan bulanan (ordinal), jumlah tanggungan sebagai indikator beban ekonomi (numerik), serta indikator *registri* sosial berupa status DTKS dan status P3KE beserta *desil* kesejahteraannya. Komposisi *fitur* tersebut dinilai memadai untuk merepresentasikan kondisi sosial ekonomi pendaftar tanpa menyertakan data identitas pribadi. Tabel 1 (Lampiran Tabel 1) memperlihatkan cuplikan data awal mahasiswa KIP Universitas Sepuluh Nopember Papua beserta atribut-atribut yang digunakan dalam penelitian.

4.2. *Pembersihan Data*

Pada tahap pascapengumpulan, hanya atribut yang relevan untuk pemodelan yang dipertahankan. Peran kolom ditetapkan tegas: Kelayakan KIP menjadi label keluaran, sedangkan Nama Siswa disimpan sebagai identitas sehingga tidak memengaruhi proses pelatihan. Prediktor yang digunakan meliputi Pekerjaan Ayah, Pekerjaan Ibu, Penghasilan Ayah, Penghasilan Ibu, Jumlah Tanggungan, Status DTKS, dan Status P3KE. Dari total 543 rekaman, dataset dibagi dengan skema holdout menjadi 109 data latih dan 434 data uji; tampilan ExampleSet berisi 50 baris pada RapidMiner hanyalah cuplikan antarmuka untuk ilustrasi, bukan jumlah keseluruhan data. Seluruh variabel kemudian disejajarkan formatnya agar siap diproses: kategori seperti jenis pekerjaan, DTKS, dan P3KE diseragamkan penulisannya; rentang penghasilan diperlakukan sebagai kategori berjenjang agar selaras dengan asumsi Naive Bayes dan tetap mudah ditafsirkan; Jumlah Tanggungan dipastikan bertipe numerik dan diimputasi bila kosong. Setelah itu, peran atribut diatur pada *RapidMiner* sebagai label, id, dan regular, serta dilakukan penyandian nominal bila diperlukan agar seluruh fitur kompatibel dengan operator Naive Bayes, sehingga data siap ditransformasikan dan dilatih tanpa membawa bias dari identitas maupun format yang tidak konsisten. Gambar 2 menunjukkan data yang telah dibersihkan dan siap untuk analisis.



Open in Turbo Prep Auto Model Interactive Analysis Filter (50 / 50 examples): all

Row No.	Kelayakan K...	Nama Siswa	Pekerjaan A...	Pekerjaan Ibu	Penghasilan...	Penghasilan...	Jumlah Tan...	Status DTKS	Status P3KE
1	LAYAK	MILSON SAD...	Petani	TIDAK BEKE...	< Rp. 250.000	Rp. 0	5	Terdata	Belum Terdata
2	TIDAK LAYAK	NOLA MONIK...	PNS	Petani	Rp. 1.250.00...	Rp. 0	5	Terdata	Terdata: Desi...
3	LAYAK	VERA VEREN...	Lainnya	Lainnya	Rp. 2.750.00...	Rp. 0	23	Terdata	Terdata: Desi...
4	LAYAK	HERMELINA ...	Nelayan	TIDAK BEKE...	Rp.750.001 ...	Rp. 0	3	Terdata	Terdata: Desi...
5	LAYAK	GIBRAN CHA...	Wirausaha	TIDAK BEKE...	Rp. 2.000.00...	Rp. 0	3	Terdata	Terdata: Desi...
6	LAYAK	Revalina Su...	Petani	TIDAK BEKE...	Rp.250.001 ...	Rp. 0	5	Terdata	Belum Terdata
7	LAYAK	MICHA VIRGI...	Petani	Lainnya	Rp.750.001 ...	Rp. 0	5	Terdata	Terdata: Desi...
8	LAYAK	AURELIA ADI...	TIDAK BEKE...	Lainnya	Rp. 0	Rp.750.001 ...	4	Belum Terdata	Belum Terdata
9	LAYAK	Sulastri	Petani	Petani	Rp.250.001 ...	Rp. 0	1	Belum Terdata	Terdata: Desi...
10	LAYAK	Adrianus Marut	Petani	Petani	< Rp. 250.000	Rp. 0	1	Terdata	Belum Terdata
11	LAYAK	Maria Kameli...	Petani	Petani	Rp. 0	Rp.250.001 ...	2	Terdata	Belum Terdata
12	LAYAK	Apoloniaris V...	TIDAK BEKE...	Petani	Rp. 0	< Rp. 250.000	5	Belum Terdata	Belum Terdata
13	LAYAK	Leonardus N...	Petani	Petani	Rp.250.001 ...	Rp.250.001 ...	4	Terdata	Terdata: Desi...
14	LAYAK	Hugorius Gra...	Petani	Petani	Rp.250.001 ...	< Rp. 250.000	3	Terdata	Belum Terdata
15	LAYAK	Komelia Asrit...	Petani	TIDAK BEKE...	Rp.250.001 ...	< Rp. 250.000	3	Terdata	Belum Terdata

ExampleSet (50 examples,1 special attribute,8 regular attributes)

Gambar 2. Data yang Telah Bersih

4.3. *Perhitungan Manual Menggunakan Rumus Naïve Bayes*

Pada bagian ini dilakukan perhitungan manual untuk menjelaskan mekanisme kerja metode *Naïve Bayes* dalam proses penentuan kelayakan penerima Beasiswa Kartu Indonesia Pintar (KIP). Perhitungan manual ini disusun untuk memperlihatkan tahapan penghitungan probabilitas yang digunakan dalam proses klasifikasi data. Tahapan tersebut mencakup perhitungan probabilitas prior setiap kelas, probabilitas bersyarat setiap atribut terhadap kelas, serta perhitungan probabilitas posterior. Penyajian dilakukan secara rinci agar proses pengolahan data dapat ditelusuri secara matematis.

Data yang digunakan sebagai contoh perhitungan merupakan satu sampel dari dataset penelitian yang memuat atribut pekerjaan orang tua, penghasilan orang tua, jumlah tanggungan keluarga, status DTKS, dan status P3KE. Atribut-atribut tersebut direpresentasikan sebagai variabel acak yang diasumsikan saling bebas bersyarat terhadap kelas kelayakan. Nilai atribut dari sampel ini dimasukkan ke dalam rumus *Naïve Bayes* sesuai dengan persamaan probabilitas bersyarat. Hasil perhitungan probabilitas selanjutnya digunakan untuk menentukan kelas kelayakan dengan nilai posterior tertinggi.

Sampel data yang digunakan dalam perhitungan manual memiliki karakteristik sebagai berikut: pekerjaan ayah adalah petani, pekerjaan ibu adalah petani, penghasilan ayah kurang dari Rp250.000, penghasilan ibu sebesar Rp0, jumlah tanggungan sebanyak tiga orang, status DTKS terdata, dan status P3KE belum terdata. Seluruh atribut tersebut dinotasikan sebagai X_1 hingga X_7, yaitu pekerjaan ayah, pekerjaan ibu, penghasilan ayah, penghasilan ibu, jumlah tanggungan, status DTKS, dan status P3KE. Notasi ini digunakan untuk mempermudah proses substitusi nilai ke dalam perhitungan probabilitas. Representasi variabel dinyatakan pada persamaan (6).

$$X_1\text{-}X_7=\{\text{Pek. Ayah, Pek. Ibu, Pengh. Ayah, Pengh. Ibu, Jumlah Tanggungan, DTKS, P3KE}\} \quad (6)$$

4.3.1. *Probabilitas Prior Kelas*

Probabilitas prior kelas merepresentasikan peluang awal suatu data termasuk ke dalam masing-masing kelas sebelum mempertimbangkan atribut-atribut prediktor. Nilai probabilitas prior diperoleh dari proporsi jumlah data pada setiap kelas terhadap keseluruhan data dalam dataset pelatihan. Pada dataset yang digunakan, jumlah data keseluruhan adalah 50 data, dengan 48 data berlabel *Layak* dan 2 data berlabel *Tidak Layak*. Berdasarkan distribusi tersebut, probabilitas prior untuk masing-masing kelas dihitung sebagai berikut:

$$P(\text{"Layak"}) = \frac{48}{50} = 0{,}96$$

$$P(\text{"Tidak Layak"}) = \frac{2}{50} = 0{,}04$$

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Gambar 3. Probabilitas Prior Kelas

Nilai probabilitas prior ini digunakan sebagai faktor pengali awal dalam perhitungan skor kelas pada metode *Naïve Bayes*. Probabilitas prior mencerminkan distribusi kelas dalam data historis yang digunakan sebagai dasar pembelajaran. Nilai tersebut kemudian dikombinasikan dengan probabilitas bersyarat setiap atribut pada tahap perhitungan berikutnya. Dengan struktur ini, kontribusi probabilitas awal terhadap hasil klasifikasi dapat diamati secara eksplisit.

4.3.2. *Probabilitas Bersyarat Kelas Layak*

Perhitungan probabilitas bersyarat untuk kelas *Layak* dilakukan dengan menerapkan *Laplace smoothing* untuk menghindari nilai probabilitas nol pada kategori dengan frekuensi rendah. Probabilitas setiap atribut dihitung berdasarkan distribusi data pada kelas *Layak* dalam dataset pelatihan. Proses perhitungan dilakukan secara terpisah untuk setiap atribut yang terlibat dalam model. Nilai probabilitas bersyarat kelas *Layak* diperoleh sebagai berikut:

$$\begin{aligned}
P(x_1 = \text{Petani} \mid \text{Layak}) &= \tfrac{18+1}{48+7} = \tfrac{19}{55} \approx 0{,}3455 \\
P(x_2 = \text{Petani} \mid \text{Layak}) &= \tfrac{12+1}{48+6} = \tfrac{13}{54} \approx 0{,}2407 \\
P(x_3 = < 250\text{rb} \mid \text{Layak}) &= \tfrac{9+1}{48+13} = \tfrac{10}{61} \approx 0{,}1639 \\
P(x_4 = \text{Rp.0} \mid \text{Layak}) &= \tfrac{30+1}{48+6} = \tfrac{31}{54} \approx 0{,}5741 \\
P(x_5 = 3\ \text{org} \mid \text{Layak}) &= \tfrac{9+1}{48+12} = \tfrac{10}{60} = 0{,}1667 \\
P(x_6 = \text{Terdata} \mid \text{Layak}) &= \tfrac{38+1}{48+2} = \tfrac{39}{50} = 0{,}78 \\
P(x_7 = \text{Belum Terdata} \mid \text{Layak}) &= \tfrac{32+1}{48+7} = \tfrac{33}{55} = 0{,}60
\end{aligned} \tag{7}$$

4.3.3. *Probabilitas Bersyarat Kelas Tidak Layak*

Perhitungan probabilitas bersyarat untuk kelas *Tidak Layak* dilakukan menggunakan prosedur yang sama dengan penerapan *Laplace smoothing*. Nilai probabilitas dihitung berdasarkan distribusi data pada kelas *Tidak Layak* dalam dataset pelatihan. Proses ini memberikan nilai peluang masing-masing atribut terhadap kelas *Tidak Layak*. Hasil perhitungan probabilitas bersyarat kelas *Tidak Layak* adalah sebagai berikut:

$$\begin{aligned}
P(x_1 = \text{Petani} \mid \text{Tidak}) &= \frac{0+1}{2+7} = \frac{1}{9} \approx 0{,}1111 \\
P(x_2 = \text{Petani} \mid \text{Tidak}) &= \frac{0+1}{2+6} = \frac{1}{8} = 0{,}125 \\
P(x_3 = < 250\text{rb} \mid \text{Tidak}) &= \frac{0+1}{2+13} = \frac{1}{15} \approx 0{,}0667 \\
P(x_4 = \text{Rp.0} \mid \text{Tidak}) &= \frac{2+1}{2+6} = \frac{3}{8} = 0{,}375 \\
P(x_5 = 3\ \text{org} \mid \text{Tidak}) &= \frac{0+1}{2+12} = \frac{1}{14} \approx 0{,}0714 \\
P(x_6 = \text{Terdata} \mid \text{Tidak}) &= \frac{1+1}{2+2} = \frac{2}{4} = 0{,}5 \\
P(x_7 = \text{Belum Terdata} \mid \text{Tidak}) &= \frac{2+1}{2+7} = \frac{3}{9} \approx 0{,}3333
\end{aligned}$$

Gambar 4. Probabilitas Bersyarat Kelas Tidak Layak

4.3.4. *Perhitungan Skor Kelas dan Probabilitas Posterior*

Setelah seluruh probabilitas bersyarat diperoleh, dilakukan perhitungan skor kelas dengan mengalikan probabilitas prior kelas dengan seluruh probabilitas bersyarat atribut. Probabilitas prior yang digunakan adalah 0,96 untuk kelas *Layak* dan 0,04 untuk kelas *Tidak Layak*. Perhitungan skor kelas dilakukan sesuai dengan prinsip dasar *Naïve Bayes*. Nilai skor kelas dihitung sebagai berikut:

$$\begin{aligned}\text{Skor}_{\text{Layak}} &= P(\text{Layak})\prod_{i=1}^{7} P(x_i \mid \text{Layak}) \\ &= 0{,}96 \times 0{,}3455 \times 0{,}2407 \times 0{,}1639 \times 0{,}5741 \times 0{,}1667 \times 0{,}78 \times 0{,}60 \\ &\approx 5{,}86 \times 10^{-4}\end{aligned} \tag{8}$$

$$\begin{aligned}\text{Skor}_{\text{Tidak}} &= P(\text{Tidak Layak})\prod_{i=1}^{7} P(x_i \mid \text{Tidak Layak}) \\ &= 0{,}04 \times 0{,}1111 \times 0{,}125 \times 0{,}0667 \times 0{,}375 \times 0{,}0714 \times 0{,}5 \times 0{,}3333 \\ &\approx 1{,}65 \times 10^{-7}\end{aligned}$$

Probabilitas posterior diperoleh dengan melakukan normalisasi terhadap skor kelas *Layak* dan *Tidak Layak*. Normalisasi dilakukan dengan membagi skor masing-masing kelas dengan jumlah total skor kedua kelas. Perhitungan probabilitas posterior digunakan untuk melihat proporsi peluang relatif dari masing-masing kelas. Nilai probabilitas posterior dihitung sebagai berikut:

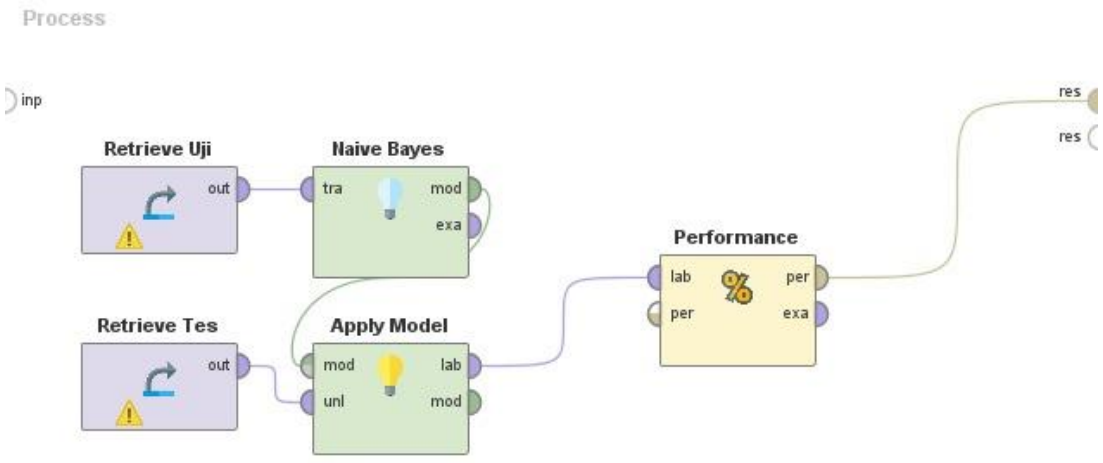
$$P(\text{Layak} \mid x) = \frac{\text{Skor}_{\text{Layak}}}{\text{Skor}_{\text{Layak}} + \text{Skor}_{\text{Tidak}}} \approx \frac{5{,}86 \times 10^{-4}}{5{,}86 \times 10^{-4} + 1{,}65 \times 10^{-7}} \approx 0{,}9997$$

$$P(\text{Tidak Layak} \mid x) \approx 0{,}0003 \tag{9}$$

4.4. *Pemodelan Data*

Rangkaian proses di *RapidMiner* menggambarkan tahapan klasifikasi mulai dari pemanggilan data hingga pengukuran kinerja model. Mula-mula, operator *Retrieve* Uji memuat data untuk pelatihan sambil menetapkan peran setiap atribut secara tepat: Kelayakan KIP dijadikan label, sedangkan atribut identitas dikeluarkan dari fitur. Dataset tersebut kemudian dipakai oleh operator Naive Bayes untuk membangun model, dengan pengaturan koreksi *Laplace* dan penanganan nilai hilang yang disesuaikan dengan karakteristik data. Proses ini memastikan bahwa setiap atribut diolah sesuai tipenya sehingga hasil prediksi lebih konsisten.

Berikutnya, operator Retrieve Tes memanggil data uji yang tidak digunakan pada proses pelatihan. Model hasil pelatihan diterapkan ke data uji melalui *Apply Model*, sehingga dihasilkan prediksi kelas dan, bila diaktifkan, skor keyakinan. Tahap akhir menggunakan operator *Performance Classification* untuk menilai kesesuaian prediksi terhadap label aktual melalui metrik akurasi, presisi, *recall, F1-score*, serta *confusion matrix.* Dengan alur tersebut, mutu model dapat dinilai secara terukur dan prosesnya mudah direplikasi. Hasil implementasi operator dan evaluasi model dapat dilihat pada Gambar 5.



Gambar 5. Pemodelan Data

4.5. *Evaluasi*

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Pada data latih berisi 109 baris, model mencapai akurasi 91,74 persen: dari 79 kasus LAYAK, 78 terklasifikasi benar dan 1 salah ditolak; dari 30 kasus TIDAK LAYAK, 22 terdeteksi benar dan 8 keliru lolos sebagai LAYAK. Metriknya menunjukkan LAYAK sangat kuat (*precision* 90,70 persen; *recall* 98,73 persen), sedangkan TIDAK LAYAK memiliki *precision* 95,65 persen namun *recall* 73,33 persen, dengan *balanced accuracy* sekitar 86,03 persen. Cakupan TIDAK LAYAK yang lebih rendah terutama dipengaruhi ketidakseimbangan kelas (79 LAYAK vs 30 TIDAK LAYAK) yang mendorong *prior* dan keputusan mengikuti mayoritas; kategori atau kombinasi fitur yang jarang muncul (misalnya jenis pekerjaan tertentu, rentang penghasilan ekstrem, atau desil P3KE rendah) membuat estimasi probabilitas kelas minor rapuh; asumsi independensi pada *Naive Bayes* tidak selalu menangkap interaksi penting seperti pendapatan rendah bersamaan dengan tanggungan tinggi dan desil P3KE rendah; ukuran sampel minoritas yang kecil meningkatkan variansi *recall*; serta penggunaan *default* (*argmax*) membuat kasus *borderline* lebih mudah diklasifikasikan sebagai LAYAK. Hasil evaluasi ini untuk model uji ditampilkan secara visual pada Gambar 6.

accuracy: 91.74%

	true LAYAK	true TIDAK LAYAK	class precision
pred. LAYAK	78	8	90.70%
pred. TIDAK LAYAK	1	22	95.65%
class recall	98.73%	73.33%	

Gambar 6. Hasil Model Uji

AUC atau luas di bawah kurva *ROC* menggambarkan kemampuan model memberi peringkat kasus positif—dalam hal ini tidak layak—lebih tinggi daripada kasus negatif di berbagai ambang keputusan; nilai 0,5 setara tebakan acak, sedangkan 1,0 menandakan pemisahan sempurna. Nilai *AUC* 0,978 berarti sekitar 97,8 persen kemungkinan seorang pendaftar tidak layak akan memperoleh skor probabilitas lebih tinggi dibanding pendaftar layak yang dipilih acak, sehingga daya pisah model tergolong sangat kuat. Konsistensi nilai ini pada estimasi optimistis, standar, dan pesimistis menunjukkan pemeringkatan yang stabil dan tidak sensitif terhadap cara perhitungan kurva. Secara praktis, meskipun *recall* untuk kelas tidak layak pada ambang saat ini belum optimal, *AUC* yang tinggi menandakan masih ada ruang untuk menyesuaikan ambang agar cakupan tidak layak meningkat sambil menjaga laju positif palsu tetap terkendali. Gambar 7 memperlihatkan performance vector untuk memvisualisasikan hasil klasifikasi pada masing-masing fitur.

PerformanceVector

```
PerformanceVector:
accuracy: 91.74%
ConfusionMatrix:
True:   LAYAK   TIDAK LAYAK
LAYAK:  78      8
TIDAK LAYAK:    1       22
precision: 95.65% (positive class: TIDAK LAYAK)
ConfusionMatrix:
True:   LAYAK   TIDAK LAYAK
LAYAK:  78      8
TIDAK LAYAK:    1       22
recall: 73.33% (positive class: TIDAK LAYAK)
ConfusionMatrix:
True:   LAYAK   TIDAK LAYAK
LAYAK:  78      8
TIDAK LAYAK:    1       22
AUC (optimistic): 0.978 (positive class: TIDAK LAYAK)
AUC: 0.978 (positive class: TIDAK LAYAK)
AUC (pessimistic): 0.978 (positive class: TIDAK LAYAK)
```

Gambar 7. Hasil Performance Vector

5. DISKUSI

Hasil penelitian menunjukkan bahwa penerapan algoritma *Naïve Bayes* mampu memodelkan kelayakan penerima KIP Kuliah secara probabilistik dengan tingkat akurasi dan nilai *AUC* yang tinggi. Pola klasifikasi yang terbentuk memperlihatkan keterkaitan kuat antara atribut sosial ekonomi, seperti penghasilan orang tua

dan status DTKS, dengan label kelayakan sebagaimana juga dijelaskan dalam [1], [3]. Nilai presisi dan *recall* yang dominan pada kelas *Layak* mencerminkan kemampuan model dalam mengenali karakteristik mayoritas data. Temuan ini menunjukkan bahwa pendekatan *classification* berbasis probabilitas dapat merepresentasikan proses seleksi secara konsisten.

Jika dibandingkan dengan penelitian sebelumnya, performa *Naïve Bayes* pada konteks seleksi bantuan pendidikan menunjukkan kesesuaian dengan hasil yang dilaporkan dalam [4], [9], khususnya pada pengolahan data kategorikal. Beberapa kajian lain mencatat stabilitas metode ini ketika digunakan pada dataset administratif dengan struktur fitur yang serupa. Perbedaan muncul pada kemampuan model dalam mengklasifikasikan kelas minoritas, yang pada penelitian lain dilaporkan memiliki keseimbangan metrik yang lebih baik. Variasi tersebut mengindikasikan bahwa distribusi data institusional memengaruhi sensitivitas model terhadap kelas tertentu.

Analisis lebih lanjut memperlihatkan adanya perbedaan kinerja antara kelas *Layak* dan *Tidak Layak*, terutama pada nilai *recall*. Kondisi ini berkaitan dengan ketidakseimbangan distribusi kelas, di mana jumlah data *Layak* lebih dominan sehingga memengaruhi probabilitas prior dalam perhitungan model. Selain itu, asumsi independensi antar fitur pada *Naïve Bayes* berpotensi mengurangi kemampuan model dalam menangkap hubungan kompleks antar variabel, sebagaimana juga dibahas dalam [2], [14]. Pola ini memperlihatkan bahwa karakteristik data memiliki peran penting dalam menentukan stabilitas hasil klasifikasi.

Meskipun hasil evaluasi menunjukkan kinerja model yang kuat, penelitian ini memiliki keterbatasan yang perlu diperhatikan. Ketidakseimbangan kelas berpengaruh terhadap representasi performa model pada kategori *Tidak Layak*, sehingga interpretasi hasil perlu mempertimbangkan proporsi data yang digunakan. Atribut yang dianalisis masih terbatas pada indikator sosial ekonomi administratif dan belum mencakup faktor non-administratif yang berpotensi relevan. Ruang lingkup data yang spesifik pada satu institusi juga membatasi generalisasi temuan ke konteks yang lebih luas.

6. KESIMPULAN DAN SARAN

6.1 *Kesimpulan*

Model *Naïve Bayes* yang dibangun pada 543 pendaftar, dengan pembagian 109 data latih dan 434 data uji, menunjukkan kinerja kuat pada data latih (akurasi 91,74 persen; *AUC* 0,978) dan efektif sebagai alat bantu penyaringan untuk mempercepat serta menstandarkan identifikasi penerima KIP Kuliah. Meski demikian, beberapa batasan perlu ditegaskan: distribusi kelas tidak seimbang (kasus LAYAK jauh lebih banyak daripada TIDAK LAYAK) sehingga recall untuk kelas TIDAK LAYAK tertekan dan metrik sensitif terhadap pemilihan ambang; ukuran evaluasi pada data uji yang disajikan relatif kecil atau belum dilaporkan lengkap sehingga konfirmasi kemampuan generalisasi belum kuat; serta atribut cenderung homogen karena didominasi variabel sosial ekonomi dan pekerjaan dan penghasilan orang tua, jumlah tanggungan, DTKS, dan P3KE yang dapat membatasi pemisahan pola terutama pada kasus TIDAK LAYAK. Dengan menyatakan keterbatasan ini secara eksplisit, hasil penelitian tetap proporsional dan memberikan pijakan yang jelas bagi penguatan evaluasi serta pengayaan fitur pada studi berikutnya. Analisis lanjutan dapat meninjau stabilitas parameter model serta sensitivitas hasil terhadap variasi proporsi data latih dan data uji untuk melihat *robustness* model pada skenario distribusi data yang berbeda.

6.2 *Saran*

Sebagai tindak lanjut, kami merekomendasikan pengembangan yang lebih konkret dengan menyeimbangkan kelas pada cabang pelatihan melalui *class weighting* dan *oversampling* (mis. *SMOTE* atau *SMOTE-NC* untuk fitur campuran), lalu mengoptimalkan ambang keputusan berbasis kurva *precision–recall* agar cakupan kelas TIDAK LAYAK meningkat sesuai biaya kesalahan. Setelah itu lakukan kalibrasi probabilitas menggunakan *Platt scaling* atau *isotonic regression* dan terapkan ambang berbasis probabilitas terkalibrasi. Jadikan *logistic regression* sebagai *baseline* utama dengan pembobotan kelas dan sertakan *decision tree* sebagai pembanding, dievaluasi melalui *stratified k-fold* atau *nested cross-validation*. Perkaya fitur dengan rasio pendapatan per tanggungan serta penyeragaman makna "belum terdata" agar konsisten, dan laporkan metrik tambahan seperti *Brier score* serta uji *fairness* antar subkelompok. Secara operasional, pertahankan pendekatan *human-in-the-loop*, lakukan *monitoring* berkala, dan jadwalkan *retraining* periodik agar kinerja tetap stabil.

DAFTAR PUSTAKA

[1] M. Thoriq, F. Maulan, Y. Septi Eirlangga, N. Hayati, M. Ashim Madani, and F. Maulana,

p-ISSN : 2803-1507 e-ISSN : 2803-1531

"Implementasi Algoritma Naïve Bayes dalam Prediksi Penerimaan Mahasiswa Penerima Beasiswa KIP di Universitas Adzkia," *J. Fasilkom*, vol. 15, no. 1, pp. 108–114, May 2025, doi: 10.37859/jf.v15i1.8959.

[2] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, p. 219, Dec. 2021, doi: 10.30865/jurikom.v8i6.3655.

[3] Z. Rohman, A. Homaidi, and Z. Fatah, "A Application of the Naïve Bayes Method to Determine Smart Indonesia Card (KIP) Recipients," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1806–1815, Jul. 2024, doi: 10.33379/gtech.v8i3.4573.

[4] C. Fitri, N. Halizah, and M. Kartikasari, "Implementasi Algoritma Naïve Bayes dalam Penentuan Pemberian Beasiswa KIP Kuliah (Studi Kasus STIKI Malang)," *J. Simantec*, vol. 12, no. 2, pp. 1–6, Jul. 2024, doi: 10.21107/simantec.v12i2.21889.

[5] I. Priyanto, E. M. Dewanti, T. Tundo, M. Nurdin, and R. Kasiono, "Penerapan Algoritma Metode Naïve Bayes untuk Penentuan Penerimaan Bantuan Program Indonesia Pintar (PIP)," *J. Manajamen Inform. Jayakarta*, vol. 4, no. 2, pp. 162–172, Apr. 2024, doi: 10.52362/jmijayakarta.v4i2.1355.

[6] B. A. Pangestu, "Penerapan Metode Naive Bayes Classifier pada Klasifikasi Penerimaan Beasiswa Kartu Indonesia Pintar," Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, 2024. [Online]. Available: http://repository.uin-suska.ac.id/82412/

[7] I. P. Oktavia and N. L. Anggreini, "Implementasi Algoritma Naive Bayes dengan Metode Klasifikasi dalam Menentukan Siswa Penerima Bantuan Program Indonesia Pintar (Studi Kasus: SMPN 3 Cihampelas)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 6, pp. 11152–11158, Nov. 2024, doi: 10.36040/jati.v8i6.11241.

[8] D. A. Setiawan, R. Halilintar, and L. S. Wahyuniar, "Penerapan Metode Naive Bayes untuk Klasifikasi Penentuan Penerima Bantuan PKH," *Pros. SEMNAS INOTEK (Seminar Nas. Inov. Teknol.*, vol. 5, no. 2, pp. 249–254, Aug. 2021, doi: 10.29407/inotek.v5i2.1137.

[9] G. Suganda, M. Asfi, R. T. Subagio, and R. P. Kusuma, "Penentuan Penerima Bantuan Beasiswa Kartu Indonesia Pintar (KIP) Kuliah Menggunakan NaãVe Bayes Classifier," *JSiI (Jurnal Sist. Informasi)*, vol. 9, no. 2, pp. 193–199, Sep. 2022, doi: 10.30656/jsii.v9i2.4376.

[10] A. C. Junior and I. Pangaribuan, "Penerapan Algoritma Naive Bayes untuk Seleksi Beasiswa Pemprov dan BAWAKU di Universitas Komputer Indonesia," Universitas Komputer Indonesia, 2019. [Online]. Available: https://www.academia.edu/87700051/penerapan

[11] M. L. Najikh and A. Subowo, "Evaluasi Dampak Program Kartu Indonesia Pintar (KIP) Kuliah di Universitas Diponegoro," *J. Public Policy Manag. Rev.*, vol. 14, no. 2, pp. 550–564, Apr. 2025, doi: 10.14710/jppmr.v14i2.50657.

[12] E. F. Sihombing and D. Y. Niska, "Implementasi Algoritma Naïve Bayes Classifier dalam Klasifikasi Calon Penerima Beasiswa KIP di SMK Negeri 5 Medan Berbasis Website," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 6, pp. 9836–9844, Nov. 2025, doi: 10.36040/jati.v9i6.15837.

[13] A. Anggleni, "Implementasi Kebijakan Program Kartu Keluarga Sejahtera (KKS) dalam Meningkatkan Kesejahteraan Masyarakat Miskin di Kelurahan Sekip Jaya Kecamatan Kemuning Kota Palembang," *J. Ilmu Adm. dan Stud. Kebijak.*, vol. 1, no. 1, pp. 24–39, Sep. 2018, doi: 10.48093/jiask.v1i1.3.

[14] A. Irsyada, E. Haerani, M. Irsyad, F. Wulandari, and L. Afriyanti, "Penerapan Algoritma Naïve Bayes Terhadap Klasifikasi Penerima Bantuan Program Keluarga Harapan (PKH)," *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, pp. 457–465, 2023, doi: 10.30865/json.v5i2.7203.

[15] N. Orinos, Q. Onola, and O. Ben Chistoff, "Zero-Shot Learning for Multilingual Document Classification in Low-Resource Languages," *J. Technol. Informatics Eng.*, vol. 4, no. 3, pp. 391–402, Dec. 2025, doi: 10.51903/jtie.v4i3.446.

[16] U. Karolina, N. Wijaya, and D. Alamsyah, "Klasifikasi Potensi Anak SD Berdasarkan Hobi

Menggunakan Metode Naive Bayes dengan Teknik Undersampling," *J. Ilm. Sist. Inf.*, vol. 4, no. 3, pp. 340–350, Jul. 2025, doi: 10.51903/wxnxrm24.

[17] D. Santi, R. D. Harahap, and S. U. Siregar, "Scientific Attitude Analysis on Students in Class XI-IPA at SMA Negeri 2 Bilah Hulu Regarding Human Blood Circulation System Material," *J. Manag. Informatics*, vol. 2, no. 1, pp. 01–07, Apr. 2023, doi: 10.22216/jcc.2019.v4i3.4.

[18] M. Rasyida, "Naïve Bayes Classification untuk Penentuan Status Penduduk Miskin," *J. Inform. Kaputama*, vol. 4, no. 2, pp. 175–180, Jul. 2020, doi: 10.59697/jik.v4i2.329.

[19] A. P. Ayudhitama and U. Pujianto, "Analisa 4 Algoritma dalam Klasifikasi Liver Menggunakan RapidMiner," *J. Inform. Polinema*, vol. 6, no. 2, pp. 1–9, Feb. 2020, doi: 10.33795/jip.v6i2.274.

[20] E. W. (Endah) Ningsih and H. (Hardiyan) Hardiyan, "Penerapan Algoritma Naïve Bayes dalam Penentuan Kelayakan Penerima Kartu Jakarta Pintar Plus," *J. Khatulistiwa Inform.*, vol. 6, no. 1, pp. 15–20, Jan. 2020, doi: 10.31294/jtk.v6i1.6680.

[21] I. B. Setiawan, J. Maulindar, and N. Nurchim, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Pada Aplikasi Kesehatan Digital," *G-Tech J. Teknol. Terap.*, vol. 8, no. 4, pp. 2301–2312, Oct. 2024, doi: 10.70609/gtech.v8i4.5020.

[22] A. Pebdika, R. Herdiana, and D. Solihudin, "Klasifikasi Menggunakan Metode Naive Bayes untuk Menentukan Calon Penerima PIP," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 452–458, Mar. 2023, doi: 10.36040/jati.v7i1.6303.

[23] F. Fauziah and A. S. Ramadhantya, "Penggunaan Rapidminer untuk Memprediksi Kelulusan Mahasiswa Dengan Algorithm Naive Bayes," *J. Teknol. Inf.*, vol. 10, no. 1, pp. 52–60, Jun. 2024, doi: 10.52643/jti.v10i1.4671.

[24] H. F. Putro, R. T. Vulandar, and W. L. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *J. Teknol. Inf. dan Komun.*, vol. 8, no. 2, pp. 19–24, Oct. 2020, doi: 10.30646/tikomsin.v8i2.500.

[25] A. N. Kirana, B. Nurhakim, S. E. Permana, W. Prihartono, and G. Dwilestari, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Cuaca Menggunakan RapidMiner," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1637–1642, 2024, doi: 10.36040/jati.v8i2.8967.

[26] A. A. A. Arifin, W. Handoko, and Z. Efendi, "Implementasi Metode Naive Bayes Untuk Klasifikasi Penerima Program Keluarga Harapan," *J-Com (Journal Comput.*, vol. 2, no. 1, pp. 21–26, 2022, doi: 10.33330/j-com.v2i1.1577.

[27] A. Fatmawati, A. I. Purnamasari, and I. Ali, "Implementasi Naïve Bayes untuk Klasifikasi Kelayakan Penerima Bantuan Sosial," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 745–750, 2024, doi: 10.36040/jati.v8i1.8714.

[28] W. Ningsih, B. Budiman, and I. Umami, "Implementasi Algoritma Naïve Bayes Untuk Menentukan Calon Penerima Beasiswa Di SMK YPM 14 Sumobito Jombang," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 4, no. 2, pp. 446–454, Jul. 2022, doi: 10.47233/jteksis.v4i2.570.

Lampiran Tabel 1. Data Awal Mahasiswa KIP Universitas Sepuluh Nopember Papua

Nama Siswa	Status DTKS	Status P3KE	Asal Sekolah	Pekerjaan Ayah	Penghasilan Ayah	Pekerjaan Ibu	Penghasilan Ibu	Seleksi Mandiri PTS
MILSON SADRAK EFAKE	Terdata	Belum Terdata	20220320 - SMAS KRISTEN YPK SATU BAKTI	Petani	< Rp. 250.000	TIDAK BEKERJA	Tidak Berpenghasilan	Universitas Sepuluh Nopember Papua - Hukum, Teknik Informatika
NOLA MONIKA POVAY	Terdata	Terdata: Desil 2	20403143 - SMAS KATOLIK SANG TIMUR	PNS	Rp. 1.250.001 - Rp. 1.500.000	Petani	Tidak Berpenghasilan	Universitas Sepuluh Nopember Papua - Sistem Informasi, Hukum
VERA VERENA WORUMI	Terdata	Terdata: Desil 5	20403167 - SMKS BOPKRI 1 YOGYAKARTA	Lainnya	Rp. 2.750.001 - Rp. 3.000.000	Lainnya	Tidak Berpenghasilan	Universitas Sepuluh Nopember Papua - Teknik Informatika,
HERMELINA RENI DAWIR	Terdata	Terdata: Desil 3	20534382 - SMAN 8 KEDIRI	Nelayan	Rp. 750.001 - Rp. 1.000.000	TIDAK BEKERJA	Tidak Berpenghasilan	Universitas Sepuluh Nopember Papua - Teknik Informatika,
GIBRAN CHALIQ FAHREZY	Terdata	Terdata: Desil 4	20577359 - SMKS MODERN AL-RIFA`IE GONDANGLEGI	Wirausaha	Rp. 2.000.001 - Rp. 2.250.000	TIDAK BEKERJA	Tidak Berpenghasilan	Universitas Sepuluh Nopember Papua - Teknik Informatika,
Revalina Sumanti	Terdata	Belum Terdata	40100885 - SMAN 1 TONDANO	Petani	Rp. 250.001 - Rp. 500.000	TIDAK BEKERJA	Tidak Berpenghasilan	Universitas Sepuluh Nopember Papua - Manajemen, Teknologi Pangan dan Hasil Pertanian
MICHA VIRGITA	Terdata	Terdata: Desil 2	40306527 - SMKN 1 TANA TORAJA	Petani	Rp. 750.001 - Rp. 1.000.000	Lainnya	Tidak Berpenghasilan	Universitas Sepuluh Nopember Papua - Sistem Informasi, Teknologi Pangan dan Hasil Pertanian
AURELIA ADINDA PAWAN	Belum Terdata	Belum Terdata	40307402 - SMKS KARTIKA WIRABUANA 1 MAKASSAR	TIDAK BEKERJA	Tidak Berpenghasilan	Lainnya	Rp. 750.001 - Rp. 1.000.000	Universitas Sepuluh Nopember Papua - Sistem Informasi,
….	….	….	…..	…	….	…	….	….