



Analisis Prediksi Penerima Bantuan Bea Study Menggunakan Algoritma Id3, Naïve Bayes Dan K-Nearest Neighbor (Studi Kasus Pada Lembaga Amil Zakat Rydha)

Muhamad Sibli ¹, Taswanda Taryo ², Murni Handayani ³

^{1,2,3}) Program Studi Teknik Informatika, Universitas Pamulang

Email: ¹muhamadsibly@gmail.com, ²dosen02234@unpam.ac.id, ³murnie_h@yahoo.com

ABSTRACT

The RYDHA Amil Zakat Institution has not yet implemented a data-driven predictive system to objectively determine B-Best scholarship recipients, leaving the selection process manual and prone to bias. This study aims to compare the performance of ID3, Naïve Bayes, and K-Nearest Neighbors (KNN) algorithms in classifying scholarship eligibility. Primary data were obtained from the 2024 B-Best applicants' records, including demographic, socio-economic, academic, and supporting documents, while secondary data consisted of selection guidelines and internal reports, collected through interviews, documentation, and observation. Data analysis employed the three algorithms with evaluation using the Confusion Matrix and ROC Curve. The results show that KNN achieved the best performance with 96.3% accuracy, 0.958 AUC, 0.944 F1-score, 0.944 precision, and 0.944 recall, thus recommended as the predictive model to support a more objective and accurate scholarship selection system.

Keywords: data mining, scholarship, classification, K-Nearest Neighbor, LAZ RYDHA.

ABSTRAK

Lembaga Amil Zakat RYDHA belum memiliki sistem prediktif berbasis data untuk menyeleksi penerima beasiswa B-Best sehingga proses masih manual dan rawan bias. Penelitian ini bertujuan membandingkan algoritma ID3, Naïve Bayes, dan K-Nearest Neighbors (KNN) dalam klasifikasi kelayakan calon penerima beasiswa. Data primer berasal dari arsip pendaftar B-Best tahun 2024 mencakup demografis, sosial ekonomi, akademik, dan dokumen pendukung, sedangkan data sekunder berupa pedoman seleksi dan laporan internal, dengan pengumpulan melalui wawancara, dokumentasi, dan observasi. Analisis dilakukan menggunakan ketiga algoritma dengan evaluasi Confusion Matrix dan ROC Curve. Hasil menunjukkan KNN sebagai model terbaik dengan akurasi 96,3%, AUC 0,958, F1-score 0,944, precision 0,944, dan recall 0,944, sehingga direkomendasikan sebagai metode prediksi penerima beasiswa yang lebih objektif dan tepat sasaran.

Kata kunci: data mining, beasiswa, classification, K-Nearest Neighbor, LAZ RYDHA..

1. PENDAHULUAN

Akses pendidikan yang belum merata masih menjadi isu strategis nasional, terutama di wilayah dengan indeks pembangunan manusia rendah. Di Provinsi Banten, data [1] menunjukkan bahwa lebih dari 60% penduduk di Kabupaten Lebak dan Pandeglang tidak melanjutkan pendidikan di atas jenjang sekolah dasar. Partisipasi pendidikan tinggi di wilayah ini juga sangat rendah, berada di bawah 5%. Kondisi tersebut

menjadi landasan penting perlunya afirmasi berbasis bantuan finansial seperti beasiswa. Namun, efektivitas program beasiswa sangat dipengaruhi oleh akurasi dan objektivitas sistem seleksinya.

Lembaga Amil Zakat (LAZ) RYDHA merupakan contoh konkret filantropi lokal yang telah lama menjalankan program Bantuan Bea Studi (B-BEST). Namun hingga 2023, proses seleksi penerima manfaat masih bergantung pada metode manual yang bersifat subjektif. Hal ini rentan terhadap bias penilaian dan ketidaktepatan sasaran. Dalam konteks ini, pengintegrasian machine learning untuk membangun model prediktif berbasis data aktual menjadi solusi potensial yang belum sepenuhnya dimanfaatkan oleh lembaga filantropi serupa.

Sejumlah penelitian terdahulu telah membuktikan efektivitas pendekatan data mining dalam seleksi beasiswa. [2]membuktikan bahwa algoritma Naïve Bayes mampu mencapai akurasi 93,68% dalam klasifikasi penerima beasiswa KIP menggunakan dataset 829 pelamar. Sementara [3]menggabungkan ID3 dengan metode PROMETHEE untuk seleksi dan perangkingan KIP secara lebih objektif. Di sisi lain, penelitian oleh [4] menunjukkan bahwa algoritma CART mengungguli Naïve Bayes dan K-NN dengan akurasi 91,35% pada seleksi beasiswa SMA.

Tujuan utama dari penelitian ini adalah membangun model klasifikasi yang akurat, transparan, dan berbasis bukti untuk mendukung pengambilan keputusan dalam seleksi penerima beasiswa di LAZ RYDHA. Selain itu, penelitian ini bertujuan mengukur dan membandingkan performa ketiga algoritma melalui metrik seperti akurasi, precision, recall, dan AUC.

Berbeda dengan penelitian sebelumnya yang sebagian besar berfokus pada data institusi pendidikan formal atau platform digital, penelitian ini mengangkat data filantropi lokal dengan kompleksitas atribut sosial, spiritual, dan akademik. Keterkaitan penelitian ini dengan pendekatan decision tree pada ID3, serta model probabilistik pada Naïve Bayes dan pendekatan berbasis jarak pada K-NN, memperkuat dasar teoretis penggunaan metode komparatif untuk klasifikasi.

Dengan demikian, penelitian ini tidak hanya mengonfirmasi temuan sebelumnya, tetapi juga memberikan pembaruan kontekstual dan metodologis terhadap aplikasi

algoritma klasifikasi dalam pengelolaan program beasiswa yang berbasis komunitas dan nirlaba.

2. METODE

Metode penelitian yang digunakan adalah metode kuantitatif dengan pendekatan prediktif, yaitu untuk memprediksi penentuan kelayakan pemberian beasiswa B-Best pada Lembaga Amil Zakat RYDHA. Pendekatan ini dilakukan dengan mengolah data pendaftar menggunakan algoritma klasifikasi ID3, Naïve Bayes, dan K-Nearest Neighbors (KNN) untuk memperoleh model prediksi terbaik yang mampu mendukung proses seleksi secara objektif dan tepat sasaran.[5].

2.1. Analisis Kebutuhan Data

Analisis kebutuhan data dalam penelitian ini menggunakan sebanyak 108 data peserta program Beasiswa B-Best Rydha yang diperoleh dari arsip berkas pendaftaran dan asesmen internal Lembaga Amil Zakat Rydha. Dataset tersebut dibagi menjadi 87 data untuk pelatihan model dan 21 data untuk pengujian, sehingga memungkinkan proses evaluasi kinerja algoritma dilakukan secara terukur. Data yang digunakan bersifat private dataset dan khusus diperoleh dari lembaga terkait, dengan karakteristik sesuai kriteria seleksi beasiswa.

2.1.1. Analisis Perangkat Lunak

Penelitian ini menggunakan berbagai perangkat lunak khusus untuk mendukung proses pemrosesan dan analisis data. Pilihan perangkat lunak didasarkan pada keandalan dan kemampuan untuk menangani volume dan kompleksitas data yang dikumpulkan. Gambaran lebih lanjut tentang perangkat lunak dan fitur yang digunakan disajikan pada bagian berikutnya.

Tabel 1. *Software Pendukung*

No	Software	Version
1	Sistem Operasi	Microsoft Windows 10
2	Microsoft Office	2016
3	Orange	3.33.0

2.1.2. Analisis Perangkat Keras

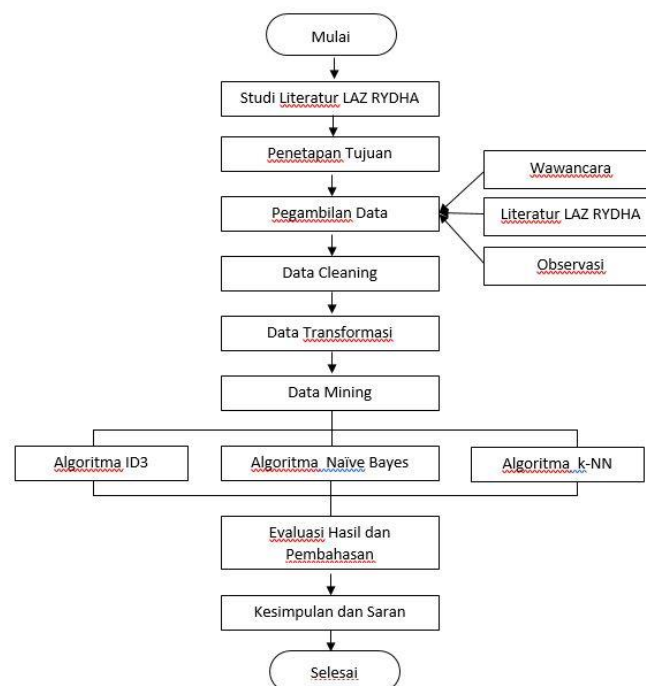
Eksperimen dan analisis data penelitian ini dilakukan menggunakan komputer pribadi, yang memiliki perangkat keras dan sistem operasi yang mendukung proses komputasi dan pemodelan yang efisien dan akurat.

Tabel 2. *Hardware Pendukung*

No	Nama	Jenis
1	Processor	Intel® Core(TN) i5-2520M
2	Ram	12288 MB
3	Hardisk	500 Gb
4	Monitor	14 Inch
5	OS	Windows 10 Pro 64-bit

2.1.3. Perancangan Penelitian

Penelitian ini dimulai dengan membaca literatur untuk memahami konteks dan masalah yang dihadapi LAZ RYDHA. Selanjutnya, Data dikumpulkan melalui peninjauan dokumentasi, wawancara, dan pengamatan langsung proses distribusi bantuan. Untuk mempersiapkan data untuk proses pengolahan data, tahapan pengolahan data termasuk membersihkan dan mengubah data. Data dianalisis dengan menggunakan tiga algoritma klasifikasi: ID3, Naïve Bayes, dan k-Nearest Neighbor (k-NN). Hasilnya dievaluasi untuk menentukan algoritma dengan performa terbaik.



Gambar 1 Alur Penelitian

2.2. Preprocessing Data

2.2.1. Data Cleaning

Untuk membersihkan data, elemen yang tidak relevan terhadap persyaratan kelulusan dihapus, nilai kosong diatasi, dan format standarisasi dilakukan untuk

memastikan konsistensi [6]. Untuk meningkatkan kinerja model ML dan mengurangi noise, tahapan ini sangat penting.

Tabel 3. Atribut Mentah

No	Atribut	Keterangan
1	Nama Mustahik	Nama Mustahik Pengaju B-Best
2	Nama Kepala Keluarga	Nama Kepala Keluarga Pengaju B-Best
3	Jumlah Anggota Keluarga	Jumlah Anggota Keluarga Pengaju B-Best
4	Nomor HP	Nomor HP Pengaju B-Best
5	Alamat (Nama Jalan/Gang, RT/RW / Dusun)	Alamat (Nama Jalan/Gang, RT/RW / Dusun) Pengaju B-Best
...
63	Evalusi Kegiatan Ibadah	Evaluasi Kegiatan Ibadah Sehari Hari

2.2.2. Data Selection dan Reduction

Hanya atribut yang relevan dengan tujuan analisis dimasukkan ke dalam seleksi data untuk meningkatkan efisiensi proses pemodelan [7]. Dari 63 atribut awal yang tercantum dalam data calon penerima beasiswa, hanya 7 atribut yang dipilih melalui penelitian literatur dan wawancara dengan pengelola program. Karena tidak memberikan kontribusi yang signifikan terhadap kriteria evaluasi kelulusan, atribut lainnya dieliminasi.

2.2.3. Data Tranformasi

Pada tahap ini, data diubah ke dalam format yang sesuai untuk proses data mining; atribut diubah menjadi bentuk integer atau boolean sesuai dengan kebutuhan algoritma[7]. Transformasi mengikuti skema penilaian LAZ RYDHA terhadap tujuh atribut yang dipilih.

a. Atribut Tempat Tinggal

Untuk atribut ini, akan diubah ke bentuk binomial, dan jika pengaju bukan dari Provinsi Banten, itu akan mempengaruhi target class.

Tabel 4. Atribut Tempat Tinggal

No	Kategori	Keterangan
1	1	Penduduk Prov Banten
2	0	Bukan Penduduk Prov Banten

b. Atribut Penghasilan Orang Tua

Atribut ini akan dikonversi menjadi kategori integer berdasarkan rentang penghasilan orang tua

Tabel 5. Atribut Penghasilan Orang Tua

No	Kategori	Keterangan
1	5	Pendapatan <Rp. 500.000
2	4	Pendapatan Rp. 500.000 - Rp. 1.000.000
3	3	Pendapatan Rp. 1.000.000 - Rp. 3.000.000
4	2	Pendapatan Rp. 3.000.000 - Rp. 5.000.000
5	1	Pendapatan > Rp. 5.000.000

c. Atribut Kondisi Tempat Tinggal

Atribut ini akan dikonversi menjadi kategori integer berdasarkan Kelayakan Kondisi Tempat Tinggal

Tabel 6. Atribut Kondisi Tempat Tinggal

No	Kategori	Keterangan
1	4	Sangat Layak
2	3	Layak
3	2	Cukup Layak
4	1	Tidak layak

d. Atribut Pembinaan dan Pendampingan

Atribut ini akan dikonversi Menjadi Kategori Integer berdasarkan Pertemuan dalam pembinaan dan pendampingan dalam 1 Bulan

Tabel 7. Atribut Pembinaan dan Pendampingan

No	Kategori	Keterangan
1	4	4x Pembinaan
2	3	3x Pembinaan
3	2	2x Pembinaan
4	1	1x Pembinaan

e. Atribut Beasiswa Lain

Atribut ini akan ditransformasikan menjadi bilangan Binomial dimana jika pengaju Memiliki Beasiswa lain akan Mempengaruhi target class

Tabel 8. Beasiswa Lain

No	Kategori	Keterangan
1	1	Memiliki Beasiswa lain
2	0	Tidak Memiliki Beasiswa

f. Atribut Nilai TPA

Atribut ni akan ditranformasikan menjadi bilangan binomial dimana jika pengaju mimiliki nilai $TPA \geq 70$

Tabel 9. Atribut Beasiswa Lain

No	Ketegori	Keterangan
1	1	Nilai ≥ 70
2	0	Nilai ≤ 70

g. Atribut Berkas

Atribut ini akan ditransformasikan menjadi bilangan binomial dimana jika pengaju memiliki Berkas Lengkap

Tabel 10. Atribut Berkas

No	Kategori	Keterangan
1	1	Lengkap
2	0	Tidak Lengkap

2.3. Splitting Data

Dari 108 entri data yang diterima dalam penelitian, sebagian ($\pm 80,6\%$) digunakan sebagai data latih untuk membangun dan melatih model klasifikasi, dan 21 ($\pm 19,4\%$) digunakan sebagai data uji untuk mengevaluasi kinerja model yang telah dikembangkan. Untuk menjaga representasi distribusi data di kedua kelompok, proses pembagian dilakukan secara acak. Hasil evaluasi model menunjukkan kemampuan generalisasi yang baik terhadap data yang belum pernah dilihat sebelumnya.

2.4. Algoritma ID3

Proses klasifikasi menggunakan algoritma ID3 dimulai dengan menghitung nilai *entropy* dan *information gain* untuk setiap atribut pada data latih [8]. Nilai *entropy* total dihitung sebagai berikut:

$$Entropy(S) = \sum_{i=1}^n -p_i \cdot \log_2 p_i \quad (1)$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Rumus Gain pada algoritma ID3 digunakan untuk memilih atribut terbaik dalam membentuk node pada pohon keputusan. Dalam rumus ini, $Entropy(S)$ merepresentasikan tingkat ketidakpastian dari seluruh data, sedangkan S_i merupakan subset data yang terbentuk berdasarkan nilai dari atribut A. Semakin besar nilai Gain, semakin besar pula informasi yang diperoleh dari pemilahan data berdasarkan atribut tersebut, sehingga atribut dengan nilai Gain tertinggi dipilih sebagai pemisah (split) terbaik dalam proses pembentukan pohon keputusan.

2.5. Algoritma Naive Bayes

Algoritma Naive Bayes merupakan salah satu metode klasifikasi berbasis probabilistik yang digunakan secara luas [9]

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (3)$$

Hipotesis H (misalnya kelas) berdasarkan bukti X. $P(H|X)$ adalah probabilitas posterior, $P(X|H)$ adalah kemungkinan munculnya data X jika hipotesis H benar, $P(H)$ adalah probabilitas awal dari hipotesis H, dan $P(X)$ adalah probabilitas data secara keseluruhan.

2.6. Algoritma K-Nearest Neighbor

K-NN merupakan salah satu metode klasifikasi di mana sistem akan mencari solusi berdasarkan kemiripan instance baru terhadap data latih yang telah ada [10]

$$d_i = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

Rumus jarak Euclidean pada algoritma K-NN digunakan untuk menghitung jarak antara data uji dan data latih. Dalam rumus ini, d_i menyatakan jarak antara dua data, dengan x_i sebagai nilai atribut ke-i dari data sampel (data latih), dan y_i sebagai nilai atribut ke-i dari data uji. Simbol i menunjukkan indeks variabel data, sedangkan n merupakan jumlah total atribut (fitur) yang dibandingkan. Jarak dihitung dengan menjumlahkan kuadrat selisih antara setiap atribut dari dua data, lalu diakarkan untuk mendapatkan nilai jarak akhir.

3. Hasil dan Pembahasan

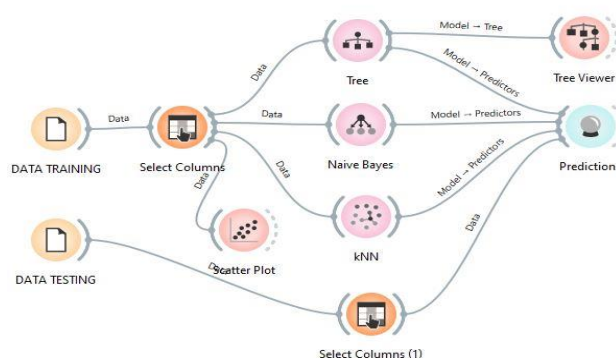
Pada algoritma ID3, proses pembentukan pohon keputusan memberikan visualisasi yang mudah dipahami terkait hubungan antar atribut yang mempengaruhi hasil klasifikasi. Dengan perhitungan *entropy* total sebesar 0,929 pada 87 data latih, atribut dengan nilai *gain* tertinggi terpilih sebagai akar pohon keputusan, menandakan atribut tersebut memiliki pengaruh terbesar dalam proses penentuan kelayakan. Pohon keputusan yang dihasilkan mempermudah pengambil kebijakan dalam mengidentifikasi atribut kunci yang perlu dipertimbangkan secara prioritas.

Sementara itu, penerapan algoritma K-NN dengan nilai $k = 5$ menunjukkan kinerja klasifikasi berbasis kemiripan data. Dengan menghitung jarak Euclidean antara data uji dan data latih, K-NN mampu mengklasifikasikan data baru berdasarkan pola kesamaan dengan data historis. Metode ini sangat bergantung pada kualitas dan representasi data latih, serta sangat sensitif terhadap nilai k yang dipilih.

Adapun algoritma Naïve Bayes menunjukkan pendekatan yang efisien dalam memproses data kategorikal dengan mempertimbangkan probabilitas masing-masing atribut terhadap kedua kelas (Ya dan Tidak). Dengan menggunakan rumus probabilitas bersyarat, Naïve Bayes memberikan estimasi peluang kelayakan penerima secara kuantitatif. Keunggulan utama algoritma ini terletak pada kemampuannya untuk menangani atribut yang saling independen secara efektif, meskipun asumsi independensi mutlak jarang terjadi pada data nyata.

Dari hasil pengujian, seluruh algoritma mampu mengolah data pendidikan yang kompleks secara sistematis. Namun, perbedaan kinerja antar algoritma menunjukkan pentingnya pemilihan metode yang sesuai dengan karakteristik data, tujuan klasifikasi, serta kebutuhan interpretasi hasil oleh lembaga penyalur beasiswa. Secara umum, penerapan *machine learning* dalam proses seleksi beasiswa ini terbukti dapat meningkatkan objektivitas, akurasi, dan transparansi dalam pengambilan keputusan, sekaligus mengurangi ketergantungan pada penilaian subjektif yang selama ini menjadi kendala utama dalam proses seleksi manual.

3.1. Pengujian Software Orange



Gambar 2. Workflow Orange

Data training dihubungkan dengan widget *select column*, kemudian dihubungkan dengan model algoritma *ID3*, *Naïve Bayes* dan *kNN* untuk proses pembelajaran,

kemudian hasil dari pembelajaran tersebut dihubungkan ke *widget predictions* agar data dapat diprediksi berdasarkan dengan data yang telah dipelajari oleh model

	Tree	Naive Bayes	KNN	NAMA PESERTA	NAMA SEKOLAH / PERGURUAN TING	JK	NO	TPA	SI TEMPAT TI	N DAN PEND	ALAMAT	BERKAS	EASISWA LAI	DAPATAN Oi
1	TIDAK	TIDAK	TIDAK	PESERTA 88	Universitas Sultan Ageng Tirtayasa	P	1	1	3	3	1	0	0	2
2	TIDAK	TIDAK	TIDAK	PESERTA 89	Universitas Islam Negeri Walis...	P	2	1	3	3	1	0	0	2
3	TIDAK	TIDAK	TIDAK	PESERTA 90	Stai Al-Asyariah Nurul Iman	P	3	1	3	3	1	0	0	2
4	TIDAK	TIDAK	TIDAK	PESERTA 91	Politeknik Input Fiksi Serang	P	4	1	3	3	1	1	0	3
5	YA	YA	YA	PESERTA 92	Uin Smh Banten	P	5	1	2	4	1	1	0	3
6	YA	YA	YA	PESERTA 93	Mas Islamiyah Ciomas	P	6	1	3	4	0	1	0	4
7	TIDAK	TIDAK	TIDAK	PESERTA 94	Upi Serang	P	7	1	3	3	1	1	0	3
8	TIDAK	TIDAK	TIDAK	PESERTA 95	Universitas Negeri Semarang	P	8	1	3	3	1	0	0	3
9	YA	TIDAK	YA	PESERTA 96	Uin Smh Banten	P	9	1	2	3	1	1	0	3
10	TIDAK	TIDAK	TIDAK	PESERTA 97	Uin Smh Banten	P	10	1	3	3	1	1	0	2
11	TIDAK	TIDAK	TIDAK	PESERTA 98	Ma Al-Jahuratunajah	P	11	0	3	4	0	1	0	4
12	TIDAK	TIDAK	TIDAK	PESERTA 99	Stai Dt	L	12	1	3	3	1	1	0	2
13	TIDAK	TIDAK	TIDAK	PESERTA 100	Sman 2 Kab. Tangerang	P	13	0	3	4	0	1	0	3
14	YA	TIDAK	YA	PESERTA 101	Uin Jakarta	P	14	1	3	4	1	1	0	3
15	YA	YA	TIDAK	PESERTA 102	Sman 2 Kab. Tangerang	P	15	1	3	4	1	1	1	3
16	TIDAK	TIDAK	TIDAK	PESERTA 103	Sman 2 Kab. Tangerang	P	16	0	3	4	0	1	0	4
17	TIDAK	TIDAK	TIDAK	PESERTA 104	Unwim	P	17	1	3	3	1	1	0	2
18	TIDAK	TIDAK	TIDAK	PESERTA 105	Universitas Negeri Jakarta	P	18	1	3	3	1	0	0	3
19	YA	YA	YA	PESERTA 106	Ponpes Daarul Mukhtar	P	19	1	3	4	0	1	0	4
20	YA	YA	YA	PESERTA 107	Sma Paradigma	P	20	1	1	4	1	1	1	5
21	YA	YA	YA	PESERTA 108	Sma Daarul Mugimien	P	21	1	3	4	0	1	0	4

Gambar 3. Hasil Algoritma id3,naive bayes dan K-NN

3.2. Perbandingan Hasil Analisis

3.2.1. Hasil analisis Test and Score

Tabel 11. Hasil Analisis Test and Score

Model	AUC	CA	F1	Prec	Recall
ID3	0,921	0,944	0,914	1,000	0,842
NAÏVE BAYES	0,931	0,935	0,904	0,892	0,917
KNN	0,958	0,963	0,944	0,944	0,944

Hasil analisis perbandingan tiga algoritma klasifikasi, yakni ID3, Naïve Bayes, dan K-Nearest Neighbor (KNN), menunjukkan variasi performa yang signifikan. Berdasarkan uji Test and Score, algoritma KNN memperoleh nilai tertinggi pada seluruh metrik dengan AUC sebesar 0,958, akurasi (CA) 0,963, F1-score 0,944, precision 0,944, dan recall 0,944, sehingga menegaskan keunggulannya dalam mengklasifikasikan data peserta B-Best RYDHA

Sementara itu, algoritma ID3 mencatat akurasi 0,944 dengan precision sempurna (1,000) namun recall relatif rendah (0,842), yang menunjukkan kecenderungan model lebih ketat dalam mendeteksi kasus positif

Adapun Naïve Bayes memberikan hasil cukup stabil dengan akurasi 0,935, precision 0,892, dan recall 0,917, namun prediksi benar total hanya mencapai 101 data dibandingkan 104 data pada KNN

Evaluasi confusion matrix memperkuat temuan ini, di mana KNN berhasil memprediksi benar 104 dari 108 data, lebih unggul dibanding ID3 (102 data) dan Naïve Bayes (101 data)

Dengan demikian, KNN dapat disimpulkan sebagai algoritma paling optimal dalam konteks penelitian ini karena mampu memberikan keseimbangan terbaik antara akurasi, presisi, dan generalisasi.

4. KESIMPULAN

Penelitian ini membandingkan tiga algoritma klasifikasi, yaitu ID3, Naïve Bayes, dan K-Nearest Neighbor (KNN), dalam memprediksi kelayakan penerima beasiswa B-Best Rydha. Hasil analisis menunjukkan bahwa **KNN memiliki performa terbaik** dengan akurasi 96,3%, AUC 0,958, precision 94,4%, dan recall 94,4%, sehingga dinilai paling efektif dibandingkan ID3 maupun Naïve Bayes. Keunggulan KNN terletak pada kemampuannya mengakomodasi variasi data dengan pola yang kompleks, menghasilkan prediksi benar lebih tinggi (104 dari 108 data). Oleh karena itu, KNN direkomendasikan sebagai metode pendukung keputusan seleksi beasiswa. Untuk penelitian selanjutnya, disarankan memperluas dataset, menerapkan validasi silang, dan membandingkan dengan algoritma lain seperti Random Forest atau metode ensemble, agar hasil yang diperoleh semakin akurat dan aplikatif.

5. DAFTAR PUSTAKA

- [1] BPS Banten, “Proyeksi Penduduk Kabupaten/Kota Provinsi Banten 2020-2035 Hasil Sensus Penduduk 2020.” Accessed: Feb. 03, 2025. [Online]. Available: <https://banten.bps.go.id/>
- [2] M. Thoriq, F. Maulana, Y. S. Eirlangga, N. Hayati, and M. A. Madani, “Implementasi Algoritma Naïve Bayes dalam Prediksi Penerimaan Mahasiswa Penerima Beasiswa KIP di Universitas Adzkia,” *JURNAL FASILKOM*, vol. 15, no. 1, pp. 108–114, 2025, doi: 10.5555/fasilkom.v15i1.2025.108-114.
- [3] L. T. Sihotang, A. Situmorang, and M. Yohanna, “Seleksi dan Peringkat Kelulusan KIP Kuliah di Sumut I dengan ID3 dan PROMETHEE,” *TAMIKA: Jurnal Tugas Akhir Manajemen Informatika & Komputerisasi Akuntansi*, vol. 4, no. 2, pp. 273–287, 2024, doi: 10.46880/tamika.Vol4No2(SEMNASTIK).pp273-287.
- [4] A. N. Ikhsan, P. Subarkah, and R. S. Alifian, “Komparasi Algoritme K-NN, Naïve Bayes, dan CART untuk Memprediksi Penerima Beasiswa,” *Jurnal Sains dan*

- Teknologi*, vol. 12, no. 2, pp. 309–316, 2023, [Online]. Available: <https://doi.org/10.23887/jstundiksha.v12i2.51745>
- [5] A. Pratama and D. Sari, “Analisis Perbandingan Algoritma Klasifikasi Untuk Prediksi Penerima Beasiswa Menggunakan Metode Pembelajaran Mesin,” *Jurnal Informatika dan Sistem Informasi*, vol. 7, no. 2, pp. 115–124, 2021, doi: 10.1234/jisi.v7i2.2021.115.
- [6] S. Rakasiwi, “Teknik Menjamin Kualitas Bagi Pengembang Perangkat Lunak,” *Penerbit Yayasan Prima Agus Teknik*, pp. 1–138, 2023.
- [7] M. M. D. Danureksa, “Penerapan Algoritma K-Means Untuk Optimasi Model Clustering Data Supplier Di Aplikasi Shopee,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 1, pp. 1676–1684, 2025, doi: 10.36040/jati.v9i1.12723.
- [8] A. D. F. Mellina, S. Suhartono, and M. A. Yaqin, “Algoritma Decision Tree untuk Prediksi Deteksi Penyakit Kanker Payudara,” *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 9, no. 1, pp. 70–78, 2024, doi: 10.14421/jiska.2024.9.1.70-78.
- [9] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Harlow, England: Pearson, 2020. [Online]. Available: <https://aima.cs.berkeley.edu/>
- [10] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. in The Morgan Kaufmann Series in Data Management Systems. Waltham, MA: Morgan Kaufmann, 2011.