

IMPLEMENTASI ALGORITMA NAÏVE BAYES CLASSIFIER DALAM KLASIFIKASI CALON PENERIMA BEASISWA KIP DI SMK NEGERI 5 MEDAN BERBASIS WEBSITE

Esra Fransiska Sihombing, Debi Yandra Niska

Ilmu Komputer, Universitas Negeri Medan

Jalan Willem Iskandar PS. V, Kenangan Baru, Percut Sei Tuan, Deli Serdang, Sumatera Utara

esra0206002@gmail.com

ABSTRAK

Pendidikan memiliki peran strategis dalam meningkatkan kualitas hidup dan kemajuan bangsa. Namun, biaya pendidikan yang tinggi menjadi kendala bagi sebagian masyarakat. Pemerintah meluncurkan program Kartu Indonesia Pintar (KIP) untuk membantu siswa kurang mampu, termasuk di SMK Negeri 5 Medan. Sekolah ini memberikan kesempatan bagi siswanya memanfaatkan bantuan KIP, namun proses seleksi masih dilakukan manual sehingga memakan waktu dan menyulitkan pencarian data lama. Penelitian ini mengusulkan sistem seleksi berbasis teknologi dengan metode klasifikasi menggunakan algoritma Naïve Bayes. Algoritma ini menghitung probabilitas antar fitur yang diasumsikan saling bebas. Data penelitian diperoleh dari 250 siswa kelas XI SMK Negeri 5 Medan melalui kuesioner, mencakup 9 variabel: Pendapatan Ayah, Pendapatan Ibu, Pekerjaan Ayah, Pekerjaan Ibu, Jumlah Tanggungan, Status Kepemilikan Rumah, Kondisi Rumah, Status Keluarga Berdasarkan Keberadaan Orangtua, dan Status Penerima KIP. Hasil evaluasi menunjukkan akurasi 92%, presisi 89%, recall 94%, dan F1-Score 92%. Temuan ini membuktikan bahwa sistem berbasis Naïve Bayes dapat meningkatkan efisiensi dan akurasi seleksi penerima beasiswa KIP di SMK Negeri 5 Medan.

Kata kunci : *Naïve Bayes, klasifikasi, machine learning, KIP, website*

1. PENDAHULUAN

Pendidikan memiliki peran penting dalam meningkatkan kualitas hidup masyarakat sekaligus mendorong pembangunan bangsa. Namun, akses terhadap pendidikan yang berkualitas masih menjadi tantangan bagi sebagian kelompok masyarakat. Salah tantangan yang memengaruhi adalah mahalnya biaya pendidikan. Menurut data yang dirilis oleh Badan Pusat Statistik (BPS) berdasarkan Survei Ekonomi Nasional (Susenas) tahun 2021, sebanyak 76% keluarga menyatakan bahwa anak mereka putus sekolah karena faktor ekonomi. Dari jumlah tersebut, mayoritas 67,0% tidak mampu membayar biaya pendidikan, sementara 8,7% lainnya terpaksa bekerja untuk memenuhi kebutuhan hidup. Untuk mengatasi permasalahan ini, pemerintah Indonesia meluncurkan program Beasiswa Kartu Indonesia Pintar (KIP) sebagai upaya memberikan akses pendidikan yang lebih merata dan mendukung siswa dari keluarga kurang mampu.

SMK Negeri 5 Medan merupakan salah satu sekolah kejuruan yang memberikan kesempatan kepada siswa untuk melanjutkan pendidikan dengan memanfaatkan bantuan dari program Kartu Indonesia Pintar (KIP). Dimana pada SMK Negeri 5 Medan ini telah berhasil memberikan beasiswa KIP sebanyak 241 siswa dari total 436 siswa yang memenuhi kriteria [1].

Berdasarkan observasi bersama Wakil Kepala Kesiswaan, proses seleksi beasiswa KIP di SMK Negeri 5 Medan masih dilakukan manual melalui pengumpulan dan pemeriksaan berkas satu per satu. Cara ini memakan waktu lama, menyulitkan pencarian data lama, serta tidak adanya sistem penyimpanan

membuat proses pencarian informasi menjadi lambat [2]. Untuk mengatasi permasalahan tersebut, penelitian ini bertujuan mengembangkan sebuah website sebagai alat bantu seleksi penerima KIP yang terintegrasi dengan algoritma *Naïve Bayes Classifier* untuk mengklasifikasikan kelayakan calon penerima beasiswa. Sistem berbasis teknologi ini dirancang untuk mengelola data pendaftar secara efisien, mempercepat proses penilaian, dan memungkinkan pengambilan keputusan yang objektif berdasarkan kriteria seperti pendapatan dan pekerjaan orang tua, jumlah tanggungan, status kepemilikan dan kondisi rumah, serta status yatim, sehingga proses seleksi dapat dilakukan lebih cepat, akurat, dan minim kesalahan [3].

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Banyak penelitian terdahulu membuktikan efektivitas algoritma *Naïve Bayes Classifier* dalam menentukan calon penerima beasiswa KIP. Penelitian oleh Gagan Suganda menunjukkan algoritma ini mampu mengklasifikasikan dengan baik, menghasilkan tingkat kesesuaian 88,21% [4].

Riyyan dan Firdaus membandingkan kinerja *Naïve Bayes* dan KNN pada data penerimaan beasiswa, di mana *Naïve Bayes* memperoleh akurasi 80%, sementara KNN hanya mencapai akurasi 78,97% [5].

Fauziah dkk menerapkan *Naïve Bayes* untuk mengklasifikasikan status kesejahteraan masyarakat Desa Gunungsari dengan target kategori “Sejahtera”

dan “Tidak Sejahtera” dan memperoleh akurasi 93,69% [6].

Asri dkk juga menggunakan algoritma ini untuk klasifikasi penerima Program Keluarga Harapan dengan hasil akurasi 88% [7].

2.2. Kartu Indonesia Pintar (KIP)

Pemerintah menyediakan bantuan pendidikan berupa Kartu Indonesia Pintar (KIP) bagi pelajar yang berasal dari keluarga dengan kondisi ekonomi lemah, sebagai upaya untuk mencegah anak-anak berhenti sekolah. Menurut keterangan dari Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K), program ini memberikan dana tunai kepada peserta didik berusia 6 hingga 21 tahun yang memiliki Kartu Keluarga Sejahtera (KKS) atau yang memenuhi persyaratan lain sesuai ketentuan pemerintah. Berdasarkan regulasi yang tertuang dalam Permendikbud Republik Indonesia Nomor 12 Tahun 2015, Program Indonesia Pintar (PIP) merupakan bantuan finansial dalam bentuk uang tunai yang ditujukan kepada siswa dari keluarga dengan keterbatasan biaya pendidikan, agar mereka tetap dapat melanjutkan sekolah.

2.3. Machine Learning

Machine learning adalah salah satu bidang dalam artificial intelligence yang memungkinkan komputer atau sistem mempelajari dan mengenali pola dari sejumlah data secara mandiri, sehingga dapat menjalankan suatu fungsi tertentu tanpa harus diprogram secara spesifik untuk tugas tersebut [8].

Machine learning merupakan bidang yang sangat terhubung dengan statistika berbasis komputasi, di mana fokus utamanya adalah memanfaatkan kekuatan komputasi untuk melakukan prediksi dari data yang tersedia. Pada prosesnya algoritma pada *Machine learning* akan mencari pola tertentu pada setiap data atau yang disebut dengan dataset untuk menghasilkan sebuah rules (aturan).

2.4. Klasifikasi

Klasifikasi adalah suatu proses pembelajaran yang bertujuan untuk mengidentifikasi dan menetapkan setiap himpunan atribut yang dimiliki oleh suatu objek. Dengan kata lain, klasifikasi dapat diartikan sebagai upaya sistematis dalam mengelompokkan data berdasarkan karakteristik atau atribut tertentu. Pengelompokkan merupakan proses sistematis untuk mengidentifikasi dan menempatkan suatu entitas ke dalam kategori tertentu yang memiliki kesamaan karakteristik [9].

Dalam pendekatan ini, data yang digunakan harus memiliki atribut target yang jelas. Sasaran utama klasifikasi adalah memprediksi label kelas berdasarkan pola-pola yang ditemukan dalam data historis [10].

2.5. Naïve Bayes Classifier

Naïve Bayes merupakan metode klasifikasi yang berasal dari teori probabilitas yang dikembangkan oleh Thomas Bayes, seorang ilmuwan asal Inggris. Pendekatan ini memanfaatkan konsep Bayes *Theorem* untuk memperkirakan probabilitas suatu kelas berdasarkan data sebelumnya, sehingga memungkinkan prediksi terhadap kejadian yang belum terjadi dengan menggunakan informasi historis yang tersedia [11].

Berikut persamaan rumus dari naïve bayes:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (1)$$

Keterangan:

$P(A|B)$: Probabilitas posterior (probabilitas A setelah mengetahui B).

$P(B|A)$: Likelihood (probabilitas B jika A benar).

$P(A)$: Prior (probabilitas awal kejadian A).

$P(B)$: Evidence (probabilitas B terjadi secara keseluruhan).

2.6. Evaluasi Model

Evaluasi terhadap performa model merupakan bagian penting dalam tahapan machine learning untuk memastikan pemilihan model yang paling optimal. Dalam konteks klasifikasi, pendekatan evaluasi yang paling sering digunakan adalah melalui *confusion matrix*. *Confusion matrix* adalah sebuah tabel yang digunakan untuk menunjukkan hasil evaluasi performa model, dengan menampilkan jumlah data uji yang diklasifikasikan dengan benar dan jumlah data uji yang salah [12].

Tabel ini berfungsi sebagai alat untuk mengevaluasi sejauh mana model mampu membedakan dan mengenali data dalam berbagai kelas. Selain itu, *confusion matrix* juga digunakan sebagai indikator untuk mengukur tingkat akurasi, kesalahan, dan kinerja keseluruhan dari proses klasifikasi.

2.7. Website

Website adalah platform digital yang tersusun dari beberapa halaman yang terintegrasi dan berperan sebagai media penyampaian informasi secara *online*. Setiap halaman dapat diakses melalui koneksi internet dan umumnya memuat beragam konten seperti tulisan, visual, audio, serta animasi, sehingga memungkinkan penyajian informasi secara lebih menarik, dinamis, dan mudah dipahami oleh pengguna,

3. METODE PENELITIAN

Berikut ini adalah alur penelitian yang digunakan dalam penelitian ini:



Gambar 1. Skema Kerja Penelitian

3.1. Pengumpulan Data

Proses pengumpulan data dilakukan dengan penyebaran kuesioner. Informasi yang dijadikan dasar analisis merupakan data primer yang bersumber langsung dari siswa kelas XI SMK Negeri 5 Medan. Data tersebut digunakan dalam proses klasifikasi untuk mengidentifikasi siswa yang berpotensi menerima bantuan beasiswa KIP.

3.2. Preprocessing Data

Kemudian dilakukan preprocessing data untuk menyiapkan data agar bisa digunakan oleh model. Proses yang dilakukan pada tahap preprocessing ini antara lain adalah pengecekan *missing value*, melakukan data *balancing* melakukan *label encoding* dan melakukan *split* data.

3.3. Klasifikasi Dengan Naïve Bayes

Setelah dilakukan preprocessing data, tahap selanjutnya dilakukan klasifikasi dengan model naïve bayes untuk mengklasifikasikan calon penerima beasiswa KIP

3.4. Evaluasi Model

Evaluasi model dilakukan dengan melihat hasil pengujian menggunakan *Confusion Matrix*. Kemudian dihitung nilai akurasi, presisi, *recall*, dan *F1-score*.

3.5. Perancangan Sistem

Dilakukan perancangan dan desain sistem yang menggambarkan tampilan dan fungsi situs web untuk

mengklasifikasikan calon penerima beasiswa KIP. Proses ini mencakup perancangan antarmuka yang sederhana dan efisien, serta mengidentifikasi fitur-fitur yang relevan untuk memenuhi kebutuhan pengguna dalam mengklasifikasikan data siswa.

3.6. Penerapan Sistem

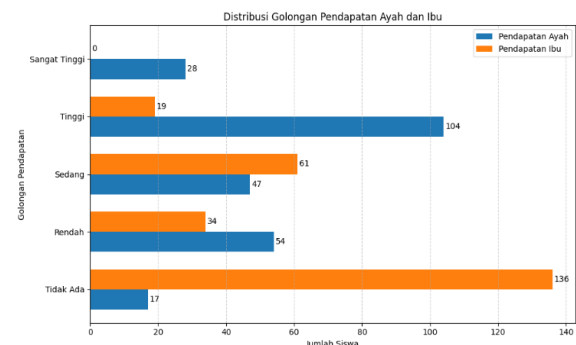
Pada tahap ini, sistem yang sudah dirancang sebelumnya di implementasikan ke dalam bentuk *web* menggunakan bahasa pemrograman PHP dan MySQL sebagai database.

4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan purposive sampling dengan subjek siswa kelas XI SMK Negeri 5 Medan, dipilih karena dianggap lebih memahami kondisi sosial ekonomi keluarga. Dari populasi tersebut, diperoleh 250 data melalui angket. Dataset memuat 9 variabel: pendapatan ayah (A_1), pendapatan ibu (A_2), pekerjaan ayah (A_3), pekerjaan ibu (A_4), jumlah tanggungan (A_5), status kepemilikan rumah (A_6), kondisi rumah (A_7), status keluarga berdasarkan keberadaan orang tua (A_8), dan status penerima KIP (y) sebagai target klasifikasi.

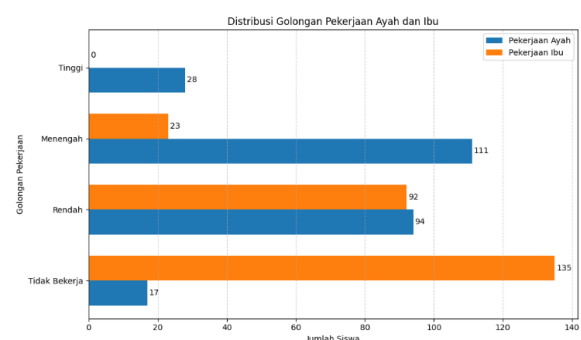
4.1. Deskripsi Data

Berikut adalah deskripsi data dari 9 variabel yang akan digunakan.



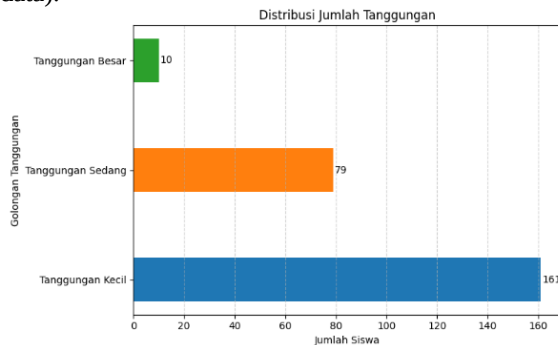
Gambar 2. Diagram Variabel Pendapatan Ayah & Ibu

Berdasarkan gambar, pendapatan ayah terbanyak berada pada rentang di atas Rp2.500.000–Rp3.500.000 (104 data) dan terendah pada Rp0 (17 data). Pendapatan ibu terbanyak pada Rp0 (136 data) dan terendah pada kategori sangat tinggi (0 data).



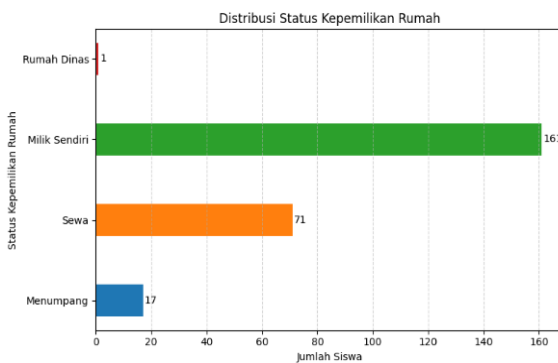
Gambar 3. Diagram Variabel Pekerjaan Ayah & Ibu

Berdasarkan gambar, pekerjaan ayah terbanyak berada pada kategori berpenghasilan menengah (111 data) dan terendah pada tidak bekerja (17 data). Pekerjaan ibu terbanyak pada kategori tidak bekerja (135 data) dan terendah pada berpenghasilan tinggi (0 data).



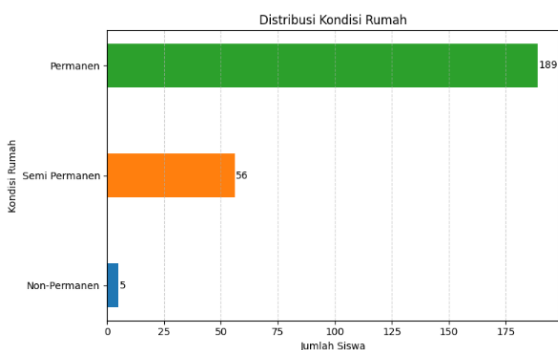
Gambar 4. Diagram Variabel Jumlah Tanggungan

Berdasarkan gambar, jumlah tanggungan terbanyak berada pada kategori 1–3 orang (161 data) dan terendah pada kategori 7–9 orang (10 data).



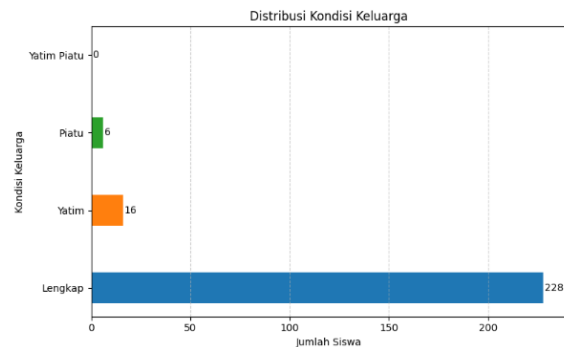
Gambar 5. Status Kepemilikan Rumah

Berdasarkan gambar, status kepemilikan rumah terbanyak adalah milik sendiri (161 data) dan terendah rumah dinas (1 data).



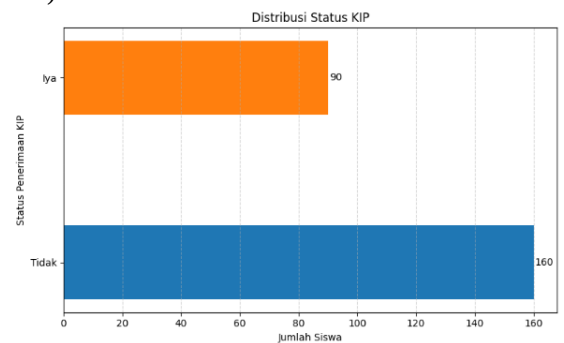
Gambar 6. Diagram Variabel Kondisi Rumah

Berdasarkan gambar, kondisi rumah terbanyak adalah permanen (189 data) dan terendah non-permanen (5 data).



Gambar 7. Diagram Variabel Kondisi Keluarga

Berdasarkan gambar, kondisi keluarga terbanyak adalah lengkap (228 data) dan terendah yatim piatu (0 data).



Gambar 8. Diagram Status Penerima KIP

Berdasarkan gambar, status penerima KIP terbanyak adalah tidak menerima KIP (160 data) dan terendah menerima KIP (90 data).

4.2. Pengecekan Missing Value

Pada tahap *preprocessing* akan dilakukan analisis terhadap jumlah data yang hilang (*missing values*) pada masing-masing atribut. Tujuannya untuk menangani data yang hilang, mengingat data tanpa nilai tidak dapat digunakan dalam perhitungan. Hasil identifikasi terhadap missing value dapat dilihat pada Gambar 9.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 250 entries, 0 to 249
Data columns (total 14 columns):
#   Column                Non-Null Count  Dtype
---  -
0   k_id                   250 non-null   int64
1   k_nama                 250 non-null   object
2   k_nis                  250 non-null   int64
3   k_jurusan              250 non-null   object
4   k_jeniskelamin         250 non-null   object
5   pend_ayah              250 non-null   object
6   pend_ibu               250 non-null   object
7   peke_ayah              250 non-null   object
8   peke_ibu               250 non-null   object
9   jmlh_tanggungan        250 non-null   object
10  status_kr              250 non-null   object
11  kond_rumah             250 non-null   object
12  kond_keluarga          250 non-null   object
13  hasil_klasifikasi      250 non-null   object
dtypes: int64(2), object(12)
```

Gambar 9. Hasil Cek Missing Value

4.3. Data Balancing

Tahap ini dilakukan penyeimbangan data karena distribusi kelas tidak seimbang: kelas “tidak” memiliki 160 data, sedangkan kelas “iya” hanya 90. Ketimpangan ini dapat menimbulkan bias pada model. Untuk mengatasinya digunakan metode undersampling, sehingga jumlah data kedua kelas menjadi sama, yaitu 90 data.

4.4. Label Encoding

Setelah dilakukan pengecekan *missing value* tahap *preprocessing* selanjutnya adalah melakukan *Label Encoding*. *Label Encoding* adalah proses mengubah nilai kategori menjadi angka (*integer*). Berikut transformasi data variabel dari kategorial menjadi angka.

Tabel 1. *Label Encoding* Variabel Pendapatan Ayah & Ibu

Pendapatan Ayah dan Ibu	Label
Tidak Ada	0
Golongan Pendapatan Rendah	1
Golongan Pendapatan Sedang	2
Golongan Pendapatan Tinggi	3
Golongan Pendapatan Sangat Tinggi	4

Tabel 2. *Label Encoding* Variabel Pekerjaan Ayah & Ibu

Pekerjaan Ayah dan Ibu	Label
Tidak Bekerja	0
Pekerjaan Berpendapatan Rendah	1
Pekerjaan Berpendapatan Sedang	2
Pekerjaan Berpendapatan Tinggi	3

Tabel 3. *Label Encoding* Variabel Jumlah Tanggungan

Jumlah Tanggungan	Label
Tanggungan Besar	0
Tanggungan Sedang	1
Tanggungan Kecil	2

Tabel 4. *Label Encoding* Status Kepemilikan Rumah

Status Kepemilikan Rumah	Label
Menumpang	0
Sewa	1
Rumah Dinas	2
Milik Sendiri	3

Tabel 5. *Label Encoding* Kondisi Rumah

Kondisi Rumah	Label
Non-Permanen	0
Semi Permanen	1
Permanen	2

Tabel 6. *Label Encoding* Kondisi Keluarga

Kondisi Keluarga	Label
Yatim Piatu	0
Yatim	1
Piatu	2
Lengkap	3

Tabel 7. *Label Encoding* Status KIP

Status KIP	Label
Iya	0
Tidak	1

4.5. Klasifikasi dengan Naïve Bayes

Pada tahap ini, proses pelatihan model naïve bayes dilakukan dengan menggunakan proporsi pembagian data sebesar 80:20. Proses klasifikasi dengan algoritma Naïve Bayes melibatkan beberapa tahapan penting, di antaranya adalah sebagai berikut:

4.6. Menghitung Probabilitas Prior

Probabilitas prior merupakan probabilitas awal dari suatu kejadian sebelum mempertimbangkan adanya informasi atau bukti tambahan. Dalam data *train* dari 144 data, dari jumlah tersebut terdapat 72 data yang termasuk dalam kelas penerima KIP dan 72 data dalam kelas tidak menerima KIP. Untuk menghitung probabilitas *prior* masing-masing kelas, digunakan rumus sebagai berikut:

- Probabilitas Kelas Penerima Kip (B_1)

$$P(B_1) = \frac{\text{Jumlah seluruh data penerima kip}}{\text{Jumlah seluruh data}} = \frac{72}{144} = 0,5$$
- Probabilitas Kelas Tidak Penerima KIP (B_2)

$$P(B_2) = \frac{\text{Jumlah data tidak penerima kip}}{\text{Jumlah seluruh data}} = \frac{72}{144} = 0,5$$

4.7. Menghitung Probabilitas Likelihood

Probabilitas *likelihood* merupakan probabilitas kemunculan nilai-nilai tertentu dari atribut (variabel) dengan asumsi bahwa data tersebut berasal dari suatu kelas tertentu, dalam hal ini kelas penerima KIP maupun tidak penerima KIP. Secara umum, rumus yang digunakan adalah:

- Probabilitas A_1 = Tidak Ada Dalam Kelas Penerima Kip ($P(A_1|B_1)$)

$$(P(A_1|B_1) = \frac{\text{Jumlah data } A_1 \text{ pada kelas penerima kip}}{\text{Jumlah data penerima kip}} = \frac{12}{72} = 0,167$$
- Probabilitas A_1 = Tidak Ada Dalam Kelas Tidak Penerima Kip ($P(A_2|B_2)$)

$$(P(A_2|B_2) = \frac{\text{Jumlah data } A_1 \text{ pada non penerima kip}}{\text{Jumlah data tidak penerima kip}} = \frac{1}{72} = 0,014$$

Setelah seluruh data dihitung menggunakan metode perhitungan yang telah dijelaskan sebelumnya, hasil probabilitas *likelihood* untuk variabel pendapatan ayah disajikan pada Gambar 4.9.

Tabel 8. Probabilitas *Likelihood* Pendapatan Ayah

Golongan Pendapatan Ayah (A_1)	Total Kejadian		Probabilitas ($P(A_1 B_i)$)	
	Layak (B_1)	Tidak Layak (B_2)	Layak (B_1)	Tidak Layak (B_2)
Tidak Ada	12	1	0,1667	0,0139
Rendah	46	0	0,6389	0
Sedang	14	12	0,1944	0,1667
Tinggi	0	47	0	0,6528
Sangat Tinggi	0	12	0	0,1667
Total	72	72	1	1

Sehingga dengan menggunakan cara yang sama seperti pada variabel pendapatan ayah didapat hasil perhitungan probabilitas likelihood untuk variabel lain, yaitu sebagai berikut:

Tabel 9. Probabilitas *Likelihood* Pendapatan Ibu

Golongan Pendapatan Ibu (A_2)	Total Kejadian		Probabilitas ($P(A_2 B_i)$)	
	Layak (B_1)	Tidak Layak (B_2)	Layak (B_1)	Tidak Layak (B_2)
Tidak Ada	49	39	0,6806	0,5417
Rendah	14	5	0,1944	0,0694
Sedang	8	20	0,1111	0,2778
Tinggi	1	8	0,0139	0,1111
Sangat Tinggi	0	0	0	0
Total	72	72	1	1

Tabel 10. Probabilitas *Likelihood* Pekerjaan Ayah

Golongan Pekerjaan Ayah (A_3)	Total Kejadian		Probabilitas ($P(A_3 B_i)$)	
	Layak (B_1)	Tidak Layak (B_2)	Layak (B_1)	Tidak Layak (B_2)
Tidak Bekerja	12	1	0,1667	0,0139
Rendah	58	10	0,8056	0,1389
Sedang	2	49	0,0278	0,6806
Tinggi	0	12	0	0,1667
Total	72	72	1	1

Tabel 11. Probabilitas *Likelihood* Pekerjaan Ibu

Golongan Pekerjaan Ibu (A_4)	Total Kejadian		Probabilitas ($P(A_4 B_i)$)	
	Layak (B_1)	Tidak Layak (B_2)	Layak (B_1)	Tidak Layak (B_2)
Tidak Bekerja	48	39	0,6667	0,5417
Rendah	22	24	0,3056	0,3333
Sedang	2	9	0,0278	0,125
Tinggi	0	0	0	0
Total	72	72	1	1

Tabel 12. Probabilitas *Likelihood* Jumlah Tnggungan

Jumlah Tanggungan (A_5)	Total Kejadian		Probabilitas ($P(A_5 B_i)$)	
	Layak (B_1)	Tidak Layak (B_2)	Layak (B_1)	Tidak Layak (B_2)
Besar	4	3	0,0556	0,0417
Sedang	30	18	0,4167	0,25
Kecil	38	51	0,5278	0,7083
Total	72	72	1	1

Tabel 13. Probabilitas *Likelihood* Status Rumah

Status Rumah (A_6)	Total Kejadian		Probabilitas ($P(A_6 B_i)$)	
	Layak (B_1)	Tidak Layak (B_2)	Layak (B_1)	Tidak Layak (B_2)
Menumpang	16	1	0,2222	0,0139
Sewa	24	18	0,3333	0,25
Rumah Dinas	0	0	0	0
Milik Sendiri	32	53	0,4444	0,7361
Total	72	72	1	1

Tabel 14. Probabilitas *Likelihood* Kondisi Rumah

Kondisi Rumah (A_7)	Total Kejadian		Probabilitas ($P(A_7 B_i)$)	
	Layak (B_1)	Tidak Layak (B_2)	Layak (B_1)	Tidak Layak (B_2)
Non-Permanen	5	0	0,0694	0
Semi Permanen	20	12	0,2778	0,1667
Permanen	47	60	0,6528	0,8333
Total	72	72	1	1

Tabel 15. Probabilitas *Likelihood* Kondisi Keluarga

Kondisi Keluarga (A_8)	Total Kejadian		Probabilitas ($P(A_8 B_i)$)	
	Layak (B_1)	Tidak Layak (B_2)	Layak (B_1)	Tidak Layak (B_2)
Yatim Piatu	0	0	0	0
Yatim	11	1	0,1528	0,0139
Piatu	4	0	0,0556	0
Lengkap	57	71	0,7917	0,9861
Total	72	72	1	1

4.8. Menghitung Probabilitas Posterior

Maka hasil perhitungan probabilitas posterior berdasarkan data uji yang digunakan dapat disajikan sebagai berikut:

Nama siswa = M.Riski Fauzan; Pendapatan ayah = Rp. 2.000.000; Pendapatan ibu = Rp. 0; Pekerjaan ayah = Gojek; Pekerjaan ibu = ibu rumah tangga; Jumlah tanggungan = 2; Status kepemilikan rumah = milik sendiri; Kondisi rumah = permanen; Kondisi keluarga = lengkap. Maka untuk nilai probabilitas prior dan likelihood dapat dilihat dari langkah sebelumnya.

a. Probabilitas *prior*

- Kelas penerima kip (B_1) = 0,5
- Kelas tidak penerima kip (B_2) = 0,5

b. Probabilitas *Likelihood*

- Pendapatan Ayah (A_1)
 $P(A_1|B_1) = 0,1944$
 $P(A_1|B_2) = 0,1667$
- Pendapatan Ibu (A_2)
 $P(A_2|B_1) = 0,6806$
 $P(A_2|B_2) = 0,5417$
- Pekerjaan Ayah (A_3)
 $P(A_3|B_1) = 0,8056$
 $P(A_3|B_2) = 0,1389$
- Pekerjaan Ibu (A_4)
 $P(A_4|B_1) = 0,6667$
 $P(A_4|B_2) = 0,5417$
- Jumlah Tanggungan (A_5)
 $P(A_5|B_1) = 0,5278$
 $P(A_5|B_2) = 0,7083$
- Status Kepemilikan Rumah (A_6)
 $P(A_6|B_1) = 0,4444$
 $P(A_6|B_2) = 0,7361$
- Kondisi Rumah (A_7)
 $P(A_7|B_1) = 0,6528$
 $P(A_7|B_2) = 0,8333$

- Kondisi Keluarga (A_8)
 $P(A_8|B_1) = 0,7917$
 $P(A_8|B_2) = 0,9861$
- c. Probabilitas Posterior
 - Probabilitas *posterior* kelas layak

$$P(B_1|A_i) = \frac{P(A_i|B_1) \times \dots \times P(A_i|B_1) \times P(B_1)}{(P(A_i|B_1) \times P(B_1)) + (P(A_i|B_2) \times P(B_2))}$$

$$= \frac{0,008614 \times 0,5}{(0,008614 \times 0,5) + (0,002911 \times 0,5)}$$

$$= \frac{0,004307}{0,004307 + 0,001456}$$

$$= \frac{0,005763}{0,005763}$$

$$= 0,7474$$
 - Probabilitas *posterior* kelas tidak layak

$$P(B_2|A_i) = \frac{P(A_i|B_2) \times \dots \times P(A_i|B_2) \times P(B_2)}{(P(A_i|B_1) \times P(B_1)) + (P(A_i|B_2) \times P(B_2))}$$

$$= \frac{0,002911 \times 0,5}{(0,008614 \times 0,5) + (0,002911 \times 0,5)}$$

$$= \frac{0,001456}{0,004307 + 0,001456}$$

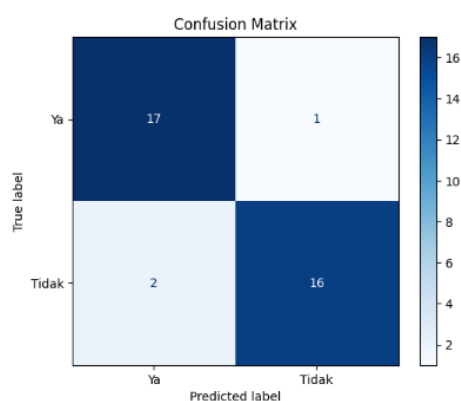
$$= \frac{0,001456}{0,005763}$$

$$= 0,2526$$

Berdasarkan hasil perhitungan probabilitas posterior yang telah dilakukan, dapat disimpulkan bahwa nilai $P(B_1|A_i) > P(B_2|A_i)$. Oleh karena itu data uji tersebut diklasifikasikan ke dalam kelas Layak.

4.9. Evaluasi Model

Setelah model Naïve Bayes berhasil dibangun, tahap selanjutnya adalah melakukan evaluasi kinerja model untuk mengetahui seberapa baik model tersebut bekerja dalam melakukan klasifikasi. Evaluasi ini dilakukan dengan menggunakan *confusion matrix*. Adapun hasil dari *confusion matrix* yang diperoleh setelah model diterapkan akan disajikan berikut ini.



Gambar 10. Hasil *Confusion Matrix*

Melalui perhitungan menggunakan evaluasi model diperoleh tingkat akurasi sebesar 92%, nilai presisi mencapai 89%, nilai *recall* sebesar 94% dan nilai *F-1 score* 92%. Nilai-nilai ini diambil dari hasil *confusion matrix* yang memberikan gambaran mengenai kinerja klasifikasi. Untuk menilai performa

model secara menyeluruh, digunakan *classification report* yang mencakup metrik presisi, recall, dan *f1-score* pada masing-masing kelas. Berikut hasil *classification report*.

Classification Report:				
	precision	recall	f1-score	support
Ya	0.89	0.94	0.92	18
Tidak	0.94	0.89	0.91	18
accuracy			0.92	36
macro avg	0.92	0.92	0.92	36
weighted avg	0.92	0.92	0.92	36

Gambar 11. Hasil *Classification Report*

4.10. Halaman Login

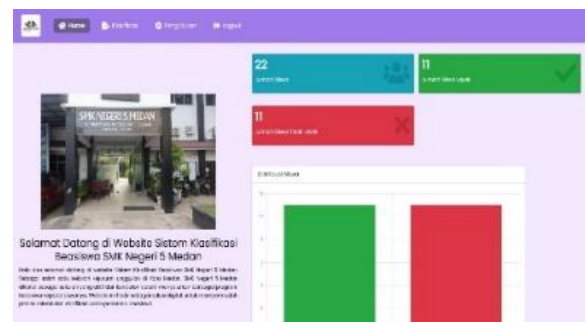
Tampilan awal saat pengguna pertama kali mengakses website sistem klasifikasi calon penerima Beasiswa KIP.



Gambar 12. Halaman *Login*

4.11. Halaman Dashboard

Halaman ini memuat informasi rekap data dalam bentuk panel warna dan terdapat diagram distribusi siswa yang memberi gambaran visual terkait hasil klasifikasi.



Gambar 13. Halaman *Dashboard*

4.12. Halaman Tabel Klasifikasi

No	Nama Lengkap	NIS	Jurusan	Jenis Kelamin	Daftar Pengantar Ayah	Daftar Pengantar Ibu	Daftar Pengantar Ayah	Daftar Pengantar Ibu	Jumlah Tanggungan	Status Kependidikan	Kondisi Rumah
1	AL-ISTIS/151006	220604	Teknik Grafika Bangunan	Laki-laki	Siswa	Tidak Ada	Siswa	Tidak Ada	1 orang	Belum	Normal
2	AL-ISTIS/151006	220605	Teknik Grafika Bangunan	Laki-laki	Siswa	Tidak Ada	Siswa	Tidak Ada	1 orang	Belum	Normal

Gambar 14. Halaman Tabel Klasifikasi

Menampilkan prediksi kelayakan Beasiswa KIP dari data input dalam tabel lengkap, serta fitur pencarian dan hapus data.

4.13. Halaman Tambah Siswa

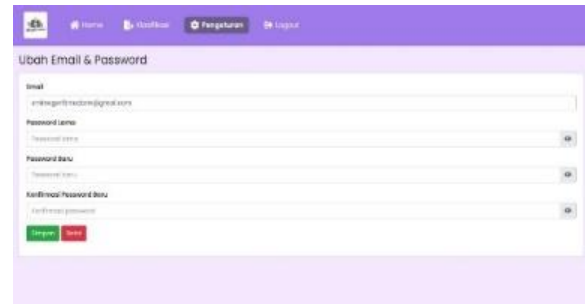
Form ini memungkinkan admin menambah data siswa sebagai acuan kelayakan KIP, dengan tombol "Simpan" dan "Batal".



Gambar 15. Halaman Tambah Data Siswa

4.14. Halaman Pengaturan

Mengelola email dan kata sandi pengguna untuk menjaga keamanan akun.



Gambar 16. Halaman Pengaturan

4.15. Pengujian Sistem

Dalam penelitian ini, pengujian sistem dilakukan menggunakan metode *Black Box Testing*.

Tabel 1. Pengujian Black Box

No.	Fitur Pengujian	Output yang diinginkan	Keterangan
1.	Halaman Login	Menampilkan form login, mengarahkan ke dashboard jika benar, atau pesan error jika salah.	☒ Valid ☐ Invalid
2.	Halaman Dashboard	Menampilkan profil sekolah, total siswa, rincian status, dan grafik perbandingan.	☒ Valid ☐ Invalid
3.	Halaman Tabel Klasifikasi	Menampilkan tabel hasil klasifikasi siswa beserta status kelayakannya masing-masing siswa.	☒ Valid ☐ Invalid
5.	Halaman Tambah Siswa	Menampilkan form yang menjadi acuan sistem dalam menentukan kelayakan siswa.	☒ Valid ☐ Invalid
6.	Halaman Pengaturan	Menampilkan form untuk mengubah email dan kata sandi guna memperbarui informasi akun pengguna.	☒ Valid ☐ Invalid

5. KESIMPULAN DAN SARAN

Penerapan sistem berbasis website untuk proses seleksi calon penerima Kartu Indonesia Pintar (KIP) di SMK Negeri 5 Medan terbukti mampu mengoptimalkan proses seleksi dengan mengurangi waktu yang dibutuhkan dibandingkan metode manual. Sistem ini dilengkapi fitur manajemen data terstruktur dan terintegrasi yang memudahkan penyimpanan, pengelolaan, serta pencarian data historis secara efisien, sekaligus menjaga konsistensi dan keakuratan informasi. Dengan menerapkan algoritma Naïve Bayes Classifier, proses seleksi menjadi lebih objektif karena data diklasifikasikan secara otomatis berdasarkan variabel relevan, menghasilkan keputusan yang adil dan berbasis data. Hasil evaluasi menunjukkan performa yang baik dengan akurasi 92%, presisi 89%, dan recall 94%. Saran pengembangan penelitian selanjutnya adalah membandingkan beberapa algoritma klasifikasi untuk menentukan metode paling efektif, serta menggunakan dataset lebih besar dan bervariasi agar model lebih general dan akurat.

DAFTAR PUSTAKA

- [1] “Data Penyaluran PIP Perwilayah,” dikdasmen. [Online]. Available: <https://pip.dikdasmen.go.id/penyaluran/jenjang>

- sekolah/076015/SMK/?tahun=2024
- [2] S. P. Sipayung, T. P. Sihalofo, A. Purba, and J. R. Tarigan, "Analisa Algoritma C.45 Terhadap penentuan Rekomendasi Penerima Beasiswa SMP Swasta Methodist-8 Medan," *LOFIAN J. Teknol. Inf. dan Komun.*, vol. 3, no. 1, pp. 20–24, 2023, doi: 10.58918/lofian.v3i1.217.
 - [3] 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, 2022, doi: 10.47233/jteksis.v4i2.570.
 - [4] Gagan Suganda, Marsani Asfi, Ridho Taufiq Subagio, and Ricky Perdana Kusuma, "Penentuan Penerima Bantuan Beasiswa Kartu Indonesia Pintar (Kip) Kuliah Menggunakan Naïve Bayes Classifier," *JSII (Jurnal Sist. Informasi)*, vol. 9, no. 2, pp. 193–199, 2022, doi: 10.30656/jsii.v9i2.4376.
 - [5] M. Riyyan and H. Firdaus, "PERBANDINGAN ALGORITME NAA• VE BAYES DAN KNN TERHADAP DATA PENERIMAAN BEASISWA (Studi Kasus Lembaga Beasiswa Baznas Jabar)," *J. Inform. dan Rekayasa Elektron.*, vol. 5, no. 1, pp. 1–10, 2022, doi: 10.36595/jire.v5i1.547.

- [6] Fauziah, Nur Syifa and R. D. Dana, "Implementasi Algoritma Naive bayes dalam Klasifikasi Status Kesejahteraan Masyarakat Desa Gunungsari," *Blend Sains J. Tek.*, vol. 1, no. 4, pp. 295–305, 2023, doi: 10.56211/blendsains.v1i4.234.
- [7] A. Asri, A. Arifin, and W. Handoko, "IMPLEMENTASI METODE NAIVE BAYES UNTUK KLASIFIKASI PENERIMA PROGRAM KELUARGA HARAPAN," vol. 2, no. 1, pp. 21–26, 2022.
- [8] B. Mahesh, "Machine Learning Algorithms - A Review," *Int. J. Sci. Res.*, vol. 9, no. 1, pp. 381–386, 2020, doi: 10.21275/art20203995.
- [9] N. M. Putry, "Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus," *EVOLUSI J. Sains dan Manaj.*, vol. 10, no. 1, 2022, doi: 10.31294/evolusi.v10i1.12514.
- [10] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32–39, 2020, doi: 10.47970/siskom-kb.v4i1.173.
- [11] Meriyam Yunita & Tri Widodo, "Sistem Pakar Diagnosa Gangguan jiwa Menggunakan Metode Naïve Bayes Berbasis Web," *Informatika*, vol. 4, no. 1, pp. 166–174, 2021.
- [12] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.