ESCAF 4th 2025
p-ISSN : 2962-7710
e-ISSN : 3021-8594

Implementasi Machine Learning dalam Penentuan Penerima Program Indonesia Pintar (PIP) di PKBM Orange Muara Kelingi

Trio Anggoro[1], Davit Irawan[2], Novi Lestari[3]

[123]Program Studi Informatika, Universitas Bina Insan, Kota Lubuklinggau

e-mail : [1]**trioanggoro12345@gmail.com**, davit_irawan@univbinainsan.ac.id, novilestari@univbinainsan.ac.id

ABSTRAK

Pemerintah Indonesia melalui Program Indonesia Pintar (PIP) mendukung pendidikan anak-anak dari keluarga kurang mampu. Namun, seleksi penerima PIP di PKBM Orange Muara Kelingi masih dilakukan secara manual, yang memakan waktu lama dan rawan kesalahan. Penelitian ini bertujuan merancang sistem seleksi berbasis *machine learning* untuk meningkatkan efisiensi dan akurasi dalam menentukan siswa yang layak diusulkan menerima bantuan. Metode penelitian menggunakan algoritma *Naïve Bayes* dan *K-Nearest Neighbor (KNN)* sebagai teknik klasifikasi. Data siswa yang meliputi usia, penghasilan orang tua, jumlah tanggungan, dan status kepemilikan KIP/KPS digunakan sebagai atribut utama. Proses pengembangan sistem mengikuti pendekatan *Knowledge Discovery in Database (KDD)* yang meliputi pengumpulan data, pelabelan, transformasi, pemodelan, dan evaluasi. Algoritma terbaik digunakan untuk membuat halaman web sederhana yang mempermudah pengguna dalam melihat hasil klasifikasi secara *real-time*. Hasil penelitian menunjukkan bahwa algoritma *KNN* memiliki performa yang lebih baik dibandingkan dengan *Naïve Bayes* berdasarkan metrik akurasi, presisi, dan *recall*. Penelitian ini berkontribusi dalam mendukung digitalisasi sistem pendidikan di Indonesia, khususnya dalam proses penyaluran bantuan PIP.

Kata kunci : *Machine Learning, KDD , Naïve Bayes, KNN*, PIP

ABSTRACT

The Government of Indonesia, through the Program Indonesia Pintar (PIP), supports the education of children from underprivileged families. However, the selection process for PIP recipients at PKBM Orange Muara Kelingi is still conducted manually, which is time-consuming and prone to errors. This research aims to design a machine learning-based selection system to improve efficiency and accuracy in determining eligible students for assistance. The research utilizes the Naïve Bayes and K-Nearest Neighbor (KNN) algorithms as classification techniques. Student data, including age, parental income, number of dependents, and KIP/KPS ownership status, serve as the main attributes. The system development follows the Knowledge Discovery in Database (KDD) approach, encompassing data collection, labeling, transformation, modeling, and evaluation. The best-performing algorithm is implemented in a simple web page, allowing users to view classification results in real-time. The findings reveal that the KNN algorithm outperforms Naïve Bayes in terms of accuracy, precision, and recall. This research contributes to supporting the digitalization of Indonesia's education system, particularly in the distribution of PIP assistance.

Keyword : *Machine Learning, KDD , Naïve Bayes, KNN*, PIP

I. PENDAHULUAN

Pemerintah terus berupaya meningkatkan kualitas pendidikan, salah satunya melalui Program Indonesia Pintar (PIP), sebuah program bantuan dana yang disalurkan melalui Kartu Indonesia Pintar

ESCAF 4th 2025
p-ISSN : 2962-7710
e-ISSN : 3021-8594

(KIP). PIP merupakan bantuan yang ditujukan kepada siswa dari keluarga kurang mampu untuk mendukung kegiatan belajar di sekolah(Nata & Suparmadi, 2022). Program ini bertujuan membantu pembiayaan pendidikan personal bagi siswa miskin atau rentan miskin yang masih terdaftar di jenjang pendidikan dasar dan menengah. PIP dirancang untuk memastikan anak-anak usia sekolah tetap mendapatkan akses pendidikan hingga menyelesaikan pendidikan menengah, baik melalui jalur formal maupun non-formal. Namun, masih terdapat sekolah yang belum mampu menyeleksi siswa secara otomatis dan terkomputerisasi dalam menentukan calon penerima beasiswa(Uriyalita et al., 2020).

Pada tanggal 24 Mei 2017, didirikanlah PKBM Orange Muara Kelingi yang beralamat di Jl. Mawar RT 06, Kelurahan Muara Kelingi, Kecamatan Muara Kelingi. PKBM Orange Muara Kelingi merupakan lembaga pendidikan nonformal yang bergerak di bidang pendidikan, di bawah pengawasan dan bimbingan Dinas Pendidikan Kabupaten. Lembaga ini berlokasi di tingkat kecamatan dan dapat didirikan oleh pihak mana pun yang telah memenuhi persyaratan kelembagaan. Saat ini, PKBM Orange memiliki 245 siswa, sebagian besar berasal dari keluarga kurang mampu. Namun, hingga kini pihak sekolah belum pernah mengajukan bantuan Program Indonesia Pintar (PIP) bagi siswa yang membutuhkan. Hal ini disebabkan oleh kendala dalam proses seleksi siswa yang layak menerima bantuan. Salah satu metode seleksi yang bisa digunakan adalah secara manual dengan membandingkan data siswa secara individual, namun dapat memakan waktu lama dan cukup rumit untuk menghasilkan keputusan yang akurat. Dengan adanya permasalahan tersebut, diperlukan sistem atau metode yang lebih efisien untuk memastikan bantuan dapat tepat sasaran dan mendukung kesejahteraan siswa secara maksimal.

Alur penerimaan Program Indonesia Pintar (PIP) di PKBM Orange dimulai dengan sosialisasi kepada orang tua/wali dan siswa mengenai PIP serta prosedur pengajuan. Sekolah kemudian mengumpulkan data siswa yang memenuhi syarat, seperti siswa dari keluarga tidak mampu yang terdaftar di DTKS, pemegang KIP (Kartu Indonesia Pintar), siswa yatim/piatu, atau yang berkebutuhan khusus. Orang tua/wali menyerahkan dokumen pendukung seperti KIP, KPS, Kartu PKH, atau Surat Keterangan Tidak Mampu. Data calon penerima diverifikasi secara manual oleh pihak sekolah untuk memastikan kelengkapan dokumen dan kesesuaian kriteria. Operator sekolah selanjutnya mengusulkan data siswa yang layak melalui sistem Dapodik.

Oleh karena itu, berdasarkan permasalahan yang dihadapi di PKBM Orange, peneliti mengangkat judul "Implementasi *Machine Learning* dalam Penentuan Penerima Program Indonesia Pintar (PIP) di PKBM Orange Muara Kelingi" dengan mengusulkan solusi berupa penerapan teknik pembelajaran mesin untuk mempercepat dan meningkatkan akurasi dalam proses seleksi penentuan siswa yang akan diusulkan sebagai calon penerima beasiswa PIP. Dalam penelitian ini, peneliti menerapkan konsep data *mining* dengan dua Algoritma Klasifikasi, *Naïve Bayes (NB)*(Arfah Wahlil Pratama et al., 2023) dan *K-Nearest Neighbor (KNN)*(Arifin et al., 2019).

Tujuan dari penelitian ini adalah merancang dan membangun sistem yang dapat digunakan oleh PKBM Orange, khususnya di bidang Kesiswaan, untuk melakukan proses seleksi calon penerima beasiswa Program Indonesia Pintar (PIP) di PKBM Orange Muara Kelingi. Taknik *machine learning* dipilih karena dapat menyeleksi secara otomatis dan mempersingkat waktu(Juwita Sari et al., 2024). Model terbaik yang dihasilkan nantinya akan ditampilkan melalui satu halaman web sederhana, sehingga pengguna dapat dengan mudah melihat hasil klasifikasi secara *real-time*.

Machine Learning (ML) atau pembelajaran mesin merupakan pendekatan dalam *AI* yang banyak digunakan untuk menggantikan atau menirukan perilaku manusia untuk menyelesaikan masalah atau melakukan otomatisasi. Sesuai namanya, *ML* mencoba menirukan bagaimana proses manusia atau makhluk cerdas belajar dan

ESCAF 4th 2025
p-ISSN : 2962-7710
e-ISSN : 3021-8594

mengeneralisasi. Setidaknya ada dua aplikasi utama dalam *ML* yaitu, klasifikasi dan prediksi. Ciri khas dari *ML* adalah adanya proses pelatihan, pembelajaran, atau training. Oleh karena itu, *ML* membutuhkan data untuk dipelajari yang disebut sebagai data *training*(Alfarizi et al., 2023).

Data *mining* merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Data *mining* mulai ada sejak 1990-an sebagai cara yang benar dan tepat untuk mengambil pola dan informasi yang digunakan untuk menemukan hubungan antara data untuk melakukan pengelompokkan ke dalam satu atau lebih cluster, sehingga objek-objek yang berada dalam satu cluster akan mempunyai kesamaan yang tinggi antara satu dengan lainnya(Aprilia et al., 2023). Data *mining* merupakan bagian dari proses penemuan pengetahuan dari basis data *Knowledge Discovery in Databases*(Alkhairi & Windarto, 2019).

Naïve Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediski probabilitas keanggotaan suatu *class*. *Bayesian classification* didasarkan pada *teorema Bayes* yang memiliki kemampuan klasifikasi serupa dengan *decesion tree* dan *neural network*. *Bayesian classification* terbukti memiliki akurasai dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar(Nur et al., 2024). Berikut Persamaan dari *Teorema Bayes*:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \qquad (1)$$

Keterangan:
X : Data dengan *class* yang belum diketahui
H : Hipotesis data X merupakan suatu *class* spesifik
P(H|X) : Probabilitas hipotesis H berdasarkan kondisi x (*posteriori* prob.)
P(H) : Probabilitas hipotesis H
P(X|H) : Probabilitas X berdasarkan kondisi tersebut
P(X) : Probabilitas dari X (Annur, 2018).

Metode *K-Nearest Neighbor (KNN)* adalah metode klasifikasi yang menentukan kategori objek *berdasarkan* data pembelajaran yang memiliki jarak terdekat dengan objek tersebut. Metode ini bertujuan untuk mengklasifikasikan objek baru berdasarkan atribut dan sampel pelatihan. Nilai prediksi dari *query* ditentukan berdasarkan klasifikasi dari tetangga terdekatnya. Berdasarkan pengertian ini, *KNN* dapat diartikan sebagai metode klasifikasi yang menggunakan data terdekat, yaitu tetangga atau data sebelumnya, sebagai sampel untuk menentukan hasil akhir(Kushartanto & Aldisa, 2023). Kedekatan *didefinisikan* berdasarkan jarak metrik, seperti jarak *Euclidean*. Jarak *Euclidean* dapat dihitung menggunakan persamaan berikut:

$$D(a,b) = \sqrt{\sum_{k=1}^{d}(a_k - b_k)^2} \qquad (2)$$

Keterangan:
D(a,b) : Jarak (*Euclidean Distance*)
(ak) : Data a yang ke-k
(bk) : Data b yang ke-k
kkk : 1, 2, 3, ...n

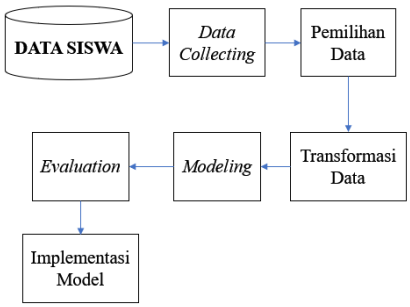
Program Indonesia Pintar (PIP) adalah salah satu inisiatif pemerintah Indonesia untuk *mendukung* pendidikan anak-anak dari keluarga kurang mampu. PIP memberikan bantuan dana pendidikan yang langsung disalurkan ke rekening Simpanan Pelajar (SimPel) milik siswa di bank seperti BRI, BNI, dan BSI. Pada tahun 2024, bantuan ini bervariasi sesuai jenjang pendidikan siswa, dengan nominal yang berbeda untuk siswa SD, SMP, dan SMA/SMK. Untuk memeriksa apakah Anda atau anak Anda termasuk penerima PIP 2024, dapat dilakukan dengan mengunjungi laman resmi PIP Kemendikbud Ristek dan memasukkan NIK serta NISN.

II. METODOLOGI PENELITIAN

Metode analisa data yang digunakan dalam penelitian ini adalah pendekatan *Knowledge Discovery in Database (KDD).* Untuk pengujian data dalam klasifikasi, digunakan metode *Naïve Bayes* dan *KNN*.

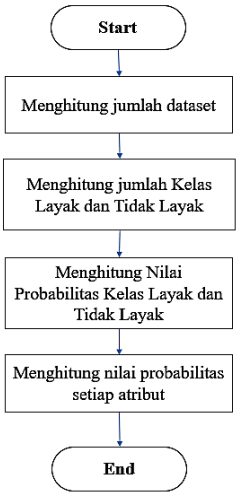
ESCAF 4th 2025
p-ISSN : 2962-7710
e-ISSN : 3021-8594

Tahapan pendekatan *KDD* yang diterapkan dalam penelitian dapat pada gambar 1 :



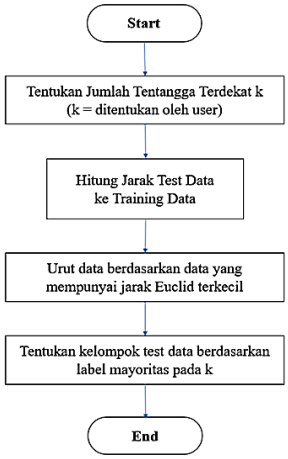
Gambar 1. Tahapan *KDD*

1. Pengumpulan Data (*Data Collection*)
   Data siswa PKBM Orange yang dikumpulkan antara lain nama, usia, penghasilan orang tua, jumlah tanggungan, status pemilik KIP dan KPS, dll. Data ini akan digunakan sebagai data latih dan data uji.
2. Pemilihan Data (*Data Selection*)
   Pemilihan fitur yang relevan seperti usia, penghasilan orang tua, jumlah tanggungan, pemilik KIP dan pemilik KPS dipilih untuk digunakan dalam pemodelan.
3. Pelabelan dan Transformasi Data
   Pada tahap pelabelan, sebagai contoh kolom "Status" akan ditambahkan yang selanjutnya akan digunakan sebagai variabel Y pada machine learning. Kolom label ditambahkan sesuai dengan aturan yang telah disepakati bersama oleh Kepala PKBM Orange Muara Kelingi untuk siswa yang akan diusulkan. Setelah itu, data akan ditransformasikan dan dikonversi dari data string menjadi numerik menggunakan fungsi *LabelEncoder* yang ada pada bahasa program Python agar atributnya sesuai dengan format yang diperlukan oleh algoritma *Naive Bayes (NB)* dan *K-Nearest Neighbors (KNN)*.
4. Proses Pemodelan (*Data Mining - Modeling*)
   Setelah data selesai melalui tahap pemilihan, pelabelan, dan transformasi, langkah berikutnya adalah melatih model machine learning. Dalam hal ini dua algoritma akan digunakan yaitu, *Naïve Bayes (NB)* dan *K-Nearest Neighbor (KNN)*. Dibawah ini adalah alur penerapan algoritma Naïve Bayes pada gambar 2.



Gambar 2. Alur Penerapan algoritma *Naive Bayes*

pada gambar 3 adalah alur penerapan algoritma *KNN*.

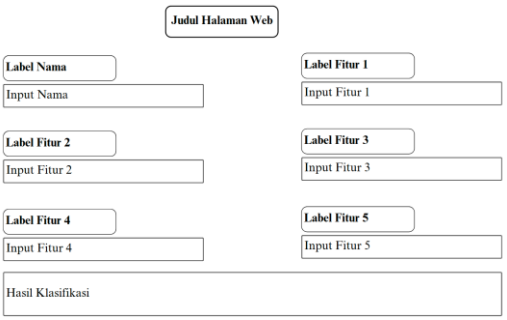


Gambar 3. Alur Penerapan Algoritma *KNN*

5. Evaluasi Model (*Evaluation*)
   Dua Model yang dihasilkan dievaluasi menggunakan confusion matrix, yang digunakan untuk menghitung accuracy, precision dan recall. Selain itu, K-fold cross-validation diterapkan untuk memastikan stabilitas dan keandalan kedua model.
6. Implementasi Model (*Deployment*)
   Model dengan hasil klasifikasi terbaik akan digunakan untuk membuat sebuah web sederhana agar nantinya pengguna bisa mengecak hasil siswa yang layak

ESCAF 4th 2025
p-ISSN : 2962-7710
e-ISSN : 3021-8594

atau tidak layak diusulkan sebagai calon penerima PIP secara real-time. Desain tampilan web klasifikasi PIP bisa dilihat pada gambar 4.



Gambar 4. Desain Web Klasifikasi PIP

III. HASIL DAN PEMBAHASAN

1. Pengumpulan Data (*Data Collection*)
   Data yang berhasil dikumpulkan sebanyak 245 data siswa dari PKBM Orange. Data ini diambil dari siswa yang terdaftar dalam sistem Dapodik. Sampel dataset siswa PKBM Orange bisa dilihat pada tabel 1.

Tabel 1. Sampel Dataset Siswa PKBM Orange

No	Nama	…	Usia	Penghasilan Orang Tua	Jumlah Tanggungan	Pemilik KIP	Pemilik KPS
1	AAS ARISKA	…	24	1500000	1	Tidak	Tidak
2	ABDUL GHOFUR	…	36	1500000	Lebih dari 3	Tidak	Tidak
3	ABDUL RAHMAN	…	39	1500000	2	Tidak	Tidak
4	ABU BAKAR	…	37	2500000	2	Tidak	Tidak
5	ADI MARLANGGA	…	17	3450000	Lebih dari 3	Tidak	Tidak
…	…	…	…	…	…	…	…
245	ZERUDAL SAPUTRA	…	19	2000000	Lebih dari 3	Tidak	Tidak

2. Pemilihan Data (*Data Selection*)
   Fitur yang dipilih dalam penelitian ini adalah usia, penghasilan orang tua, jumlah tanggungan, pemilik KIP dan pemilik KPS. Sampel hasil pemilihan data pada tabel 2.

Tabel 2. Data Fitur

Usia	Penghasilan Orang Tua	Jumlah Tanggungan	Pemilik KIP	Pemilik KPS
24	1500000	1	Tidak	Tidak
36	1500000	Lebih dari 3	Tidak	Tidak
39	1500000	2	Tidak	Tidak
37	2500000	2	Tidak	Tidak
17	3450000	Lebih dari 3	Tidak	Tidak

3. Pelabelan dan Transformasi Data
   Pada tahap pelabelan, kolom "Status" ditambahkan yang akan digunakan sebagai variabel Y pada machine learning. Selanjutnya, data akan ditransformasikan menjadi data kategori. Kemudian data dikonversi menjadi numerik agar sesuai dengan format yang diperlukan algoritma *Naive Bayes (NB)* dan *K-Nearest Neighbors (KNN)*. Hasil aturan kolom Status bisa dilihat pada tabel 3 dan sampel hasil pelabelan dan transformasi data bisa dilihat pada tabel 4 dibawah ini.

Tabel 3. Aturan Label Kategori Layak

Kriteria	Aturan	Status
Usia (C1)	dibawah 21	Layak
Penghasilan Orang Tua (C2)	dibawah UMR atau UMR jika tanggungan lebih dari 3	Layak
Jumlah Tanggungan (C3)	1, 2, 3, atau Lebih dari 3	Layak
Pemilik KIP (C4)/Pemilik KPS(C5)	Diabaikan jika Penghasilan dibawah UMR atau UMR jika tanggungan lebih dari 3	Layak

Tabel 4. Hasil Pelabelan

Usia (C1)	Penghasilan Orang Tua (C2)	Jumlah Tanggungan (C3)	Pemilik KIP (C4)	Pemilik KPS (C5)	Status *Actual*
diatas 21	dibawah UMR	1	Tidak	Tidak	Tidak Layak
diatas 21	dibawah UMR	Lebih dari 3	Tidak	Tidak	Tidak Layak
diatas 21	dibawah UMR	2	Tidak	Tidak	Tidak Layak
diatas 21	dibawah UMR	2	Tidak	Tidak	Tidak Layak
dibawah 21	UMR	Lebih dari 3	Tidak	Tidak	Layak

Data string akan dikonversi menjadi numerik menggunakan fungsi *LabelEncoder* pada Python. Tabel 5–10 menunjukkan hasil konversi setiap fitur, sedangkan tabel 11 menampilkan hasil keseluruhan.

Tabel 5. Konversi Data Usia (C1)

Usia (C1)	Konversi Angka
21	0
diatas 21	1
dibawah 21	2

Tabel 6. Konversi Data Penghasilan Orang Tua (C2)

Penghasilan Orang Tua (C2)	Konversi Angka
UMR	0
diatas UMR	1
dibawah UMR	2

ESCAF 4th 2025
p-ISSN : 2962-7710
e-ISSN : 3021-8594

Tabel 7. Konversi Data Jumlah Tanggungan (C3)

Jumlah Tanggungan (C3)	Konversi Angka
1	0
2	1
3	2
Lebih dari 3	3

Tabel 8. Konversi Data Pemilik KIP (C4)

Penerima KIP (C4)	Konversi Angka
Tidak	0
Ya	1

Tabel 9. Konversi Data Pemillik KPS (C5)

Penerima KPS (C5)	Konversi Angka
Tidak	0
Ya	1

Tabel 10. Konversi Data Status

Status	Konversi Angka
Tidak Layak	0
Layak	1

Tabel 11. Hasil Data Koversi

Usia (C1)	Penghasilan Orang Tua (C2)	Jumlah Tanggungan (C3)	Pemilik KIP (C4)	Pemilik KPS (C5)	Status *Actual*
1	2	0	0	0	0
1	2	3	0	0	0
1	2	1	0	0	0
1	2	1	0	0	0
2	0	3	0	0	1

Setelah itu Data akan dibagi menjadi 70% untuk pelatihan (171 data) dan 30% untuk pengujian (74 data). Sampel data pelatihan ditampilkan pada Tabel 12, sedangkan sampel data pengujian pada Tabel 13.

Tabel 12. Sampel Data Latih

Usia (C1)	Penghasilan Orang Tua (C2)	Jumlah Tanggungan (C3)	Pemilik KIP (C4)	Pemilik KPS (C5)	Status *Actual*
2	2	2	0	0	1
1	2	0	0	0	0
2	2	2	0	0	1
2	1	0	0	0	0
1	2	1	0	0	0

Tabel 13. Sampel Data Uji

Usia (C1)	Penghasilan Orang Tua (C2)	Jumlah Tanggungan (C3)	Pemilik KIP (C4)	Pemilik KPS (C5)	Status *Actual*
1	2	3	0	0	0
1	2	3	0	0	0
1	2	0	0	0	0
0	2	3	0	0	0
2	2	3	0	0	1

4. Proses Pemodelan (*Data Mining - Modeling*)
   a. Algoritma *Naïve Bayes (NB)*
   Berikut ini langkah-langkah penerapan algoritma *NB* :
   1) Menghitung jumlah 70% data latih dari dataset
   Jumlah data latih yang digunakan dalam penelitian ini adalah 171.
   2) Menghitung jumlah Kelas Layak dan Tidak Layak
   Dari 171 data latih didapatkan jumlah kelas Layak 90 dan kelas Tidak Layak 81 . Tabel 14 menampilkan pembagian kelas.

Tabel 14. Pembagian Kelas

Kelas	
Layak	**Tidak Layak**
90	81

   3) Menghitung Nilai Probabilitas Kelas Layak dan Tidak Layak

P(Ci)	P (Layak)	= 90/171 = 0,526
	P (Tidak Layak)	= 81/171 = 0,473

   4) Langkah selanjutnya adalah menghitung nilai probabilitas setiap atribut, nilai atribut yang akan dihitung antara lain :
   a) Hasil Probabilitas atribut Usia (C1) pada tabel 15

Tabel 15. Hasil Perhitungan atribut Usia (C1)

Usia	Probabilitas (C1)		Jumlah Kajadian "Dipilih"	
	Layak	Tidak	Layak	Tidak
21	0	0,17	0	14
diatas 21	0	0,77	0	63
dibawah 21	1	0,05	90	4

   b) Hasil Probabilitas atribut Penghasilan Orang Tua (C2) pada tabel 16

ESCAF 4th 2025
p-ISSN : 2962-7710
e-ISSN : 3021-8594

Tabel 16. Hasil Perhitungan atribut Penghasilan Orang Tua (C2)

Penghasilan Orang Tua (C2)	Probabilitas (C1)		Jumlah Kajadian "Dipilih"	
	Layak	Tidak	Layak	Tidak
UMR	0,1	0,02	9	2
diatas UMR	0,01	0,05	1	4
dibawah UMR	0,88	0,92	80	75

c) Hasil Probabilitas atribut Jumlah Tanggungan (C3) pada tabel 17

Tabel 17. Hasil Perhitungan atribut Jumlah Tanggungan (C3)

Jumlah Tanggungan (C3)	Probabilitas (C1)		Jumlah Kajadian "Dipilih"	
	Layak	Tidak	Layak	Tidak
1	0,19	0,20	17	16
2	0,23	0,24	21	20
3	0,35	0,33	32	27
Lebih dari 3	0,22	0,22	20	18

d) Hasil Probabilitas atribut Pemilik KIP (C4) pada tabel 18

Tabel 18. Hasil Perhitunngan atribut Pemilik KIP (C4)

Pemilik KIP (C4)	Probabilitas (C1)		Jumlah Kajadian "Dipilih"	
	Layak	Tidak	Layak	Tidak
Tidak	0,97	0,96	88	78
Ya	0,02	0,03	2	3

e) Hasil Probabilitas atribut Pemilik KPS (C5) pada tabel 19

Tabel 19. Hasil Perhitungan atribut Pemilik KPS (C5)

Pemilik KPS (C4)	Probabilitas (C1)		Jumlah Kajadian "Dipilih"	
	Layak	Tidak	Layak	Tidak
Tidak	0,97	0,97	88	79
Ya	0,02	0,02	2	2

Contoh perhitungan untuk menentukan kelas prediksi penerima PIP (Layak/Tidak Layak) pada Tabel 20 dijelaskan berikut ini.

Tabel 20. Data Pengujian NB

Objek	Usia (C1)	Penghasilan Orang Tua (C2)	Jumlah Tanggungan (C3)	Pemilik KIP (C4)	Pemilik KPS (C5)	Status
A1	21	dibawah UMR	1	Tidak	Tidak	..?

Dibawah ini adalah perhitungan klasifikasi *NB :*

- Hitung *Posterior* untuk "Layak"
  Rumus = C1xC2xC3xC4xC5
  Likelihood Layak
  = 0 x 0,88 x 0,19 x 0,97 x 0,97
  = 0
  Posterior Layak = 0 x 0,526 = 0
- Hitung *Posterior* untuk "Tidak Layak"
  Likelihood Tidak Layak
  = 0,17 x 0,92 x 0,20 x 0,96 x 0,97
  = 0,029143456
  Posterior Tidak Layak
  = 0,029143456 x 0,473
  = 0,013782361
- Hitung Total *Posterior*
  Total *Posterior*= *Posterior* Layak x *Posterior* Tidak Layak
  Total *Posterior*
  = 0 + 0.013782361
  = 0,013782361
- Normalisasi
  Probabilitas "Layak":
  P(Layak|Data) = *Posterior* Layak / Total *Posterior*
  P(Layak|Data)
  = 0 / 0,013782361
  = 0
  Probabilitas "Tidak Layak":
  P(Tidak Layak|Data)
  = *Posterior* Tidak Layak / Total *Posterior*
  P(Tidak Layak|Data)
  = 0,013782361 / 0,013782361
  = 1

Hasil akhir ditampilkan pada tabel 21

Tabel 21. Hasil Klasifikasi *NB*

Nilai Probabilitas Layak	Nilai Probabilitas Tidak Layak	Status *NB*	Status *Actual*
0	1	Tidak Layak	Tidak Layak

b. Algoritma *KNN*
Berikut ini langkah-langkah penerapan algoritma *KNN* :

1) Parameter K = 1
2) Menghitung kuadrat jarak *eucliden* antara objek pada data uji (tabel 22) terhadap data *training* (tabel 23). Hasil perhitungan jarak pada tabel 24.

ESCAF 4th 2025
p-ISSN : 2962-7710
e-ISSN : 3021-8594

Tabel 22. Data Pengujian *KNN*

Objek	Usia (C1)	Penghasilan Orang Tua (C2)	Jumlah Tanggungan (C3)	Pemilik KIP (C4)	Pemilik KPS (C5)	Status
A1	0	2	0	0	0	..?

Tabel 23. Data *Training*

Objek	Usia (C1)	Penghasilan Orang Tua (C2)	Jumlah Tanggungan (C3)	Pemilik KIP (C4)	Pemilik KPS (C5)	Status *Actual*
A1	2	2	2	0	0	Layak
A2	1	2	0	0	0	Tidak Layak
A3	2	2	2	0	0	Layak
A4	2	1	0	0	0	Tidak Layak

$$Dq = \sqrt{(a1 - b1)^2 + (a2 - b2)^2 + (an - bn)^2}$$

$$D1 = \sqrt{(0-2)^2 + (2-2)^2 + (0-2)^2 + (0-0)^2 + (0-0)^2}$$
$$= \sqrt{(-2)^2 + (0)^2 + (-2) + (0)^2 + (0)^2}$$
$$= \sqrt{4+0+4+0+0} = \sqrt{8}$$
$$= 2{,}82$$

$$D2 = \sqrt{(0-1)^2 + (2-2)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2}$$
$$= \sqrt{(-1)^2 + (0)^2 + (0)^2 + (0)^2 + (0)^2}$$
$$= \sqrt{1+0+0+0+0} = \sqrt{1}$$
$$= 1$$

$$D3 = \sqrt{(0-2)^2 + (2-2)^2 + (0-2)^2 + (0-0)^2 + (0-0)^2}$$
$$= \sqrt{(-2)^2 + (0)^2 + (-2)^2 + (0)^2 + (0)^2}$$
$$= \sqrt{4+0+4+0+0} = \sqrt{8}$$
$$= 2{,}82$$

$$D4 = \sqrt{(0-2)^2 + (2-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2}$$
$$= \sqrt{(-2)^2 + (1)^2 + (0)^2 + (0)^2 + (0)^2}$$
$$= \sqrt{4+1+0+0+0} = \sqrt{5}$$
$$= 2{,}23$$

Tabel 24. Hasil Perhitungan Jarak

No	Objek	Hasil Perhitungan Jarak
1	A1	2,82
3	A2	1
4	A3	2,82
5	A4	2,23

3) Hasil perhitungan jarak secara *ascending* pada tabel 25

Tabel 25. Hasil Ascending Jarak

No	Objek	Hasil Perhitungan Jarak	Status *Actual*
1	A2	1	Tidak Layak
3	A4	2,23	Tidak Layak
4	A1	2,82	Layak
5	A3	2,82	Layak

4) Mengumpulkan kategori Y berdasarkan nilai k = 1

Data dengan jarak terdekat k = 1 adalah objek A2 dengan nilai D (*distance*) = 1 . Hasil data nilai k = 1 pada tabel 26

Tabel 26. Hasil Data Nilai K = 1

No	Objek	Hasil Perhitungan Jarak	Status *KNN*
1	A2	1	Tidak Layak

5) Hasil klasifikasi *KNN* dengan nilai k = 1 pada tabel 27

Tabel 27. Hasil Klasifikasi *KNN* Nilai K = 1

Objek	Usia (C1)	Penghasilan Orang Tua (C2)	Jumlah Tanggungan (C3)	Pemilik KIP (C4)	Pemilik KPS (C5)	Status *KNN*	Status *Actual*
A1	21	dibawah UMR	1	Tidak	Tidak	Tidak Layak	Tidak Layak

5. Evaluasi Model (*Evaluation*)
Model yang sudah dilatih dievaluasi dengan *Confusion Matrix* menggunakan 30% data uji dari dataset yaitu 74 data Sampel data uji bisa dilihat pada tabel 13 diatas. Selanjutnya akan divalidasi menggunakan *K-Fold Cross Validation*, yang selanjutnya diuji dengan 10 data uji baru.

a. Algoritma *NB*
Tabel 28 menyajikan *Confusion Matrix* algoritma *NB* dari 74 data.

Tabel 28. Hasil *Confusion Matrix NB*

		Predicted Class	
		Tidak Layak	Layak
Actual Class	Tidak Layak	36	2
	Layak	0	36

Selanjutnya *confunsion matrix* digunakan untuk menghitung *accuracy, precision* dan *recall* sebagai berikut :

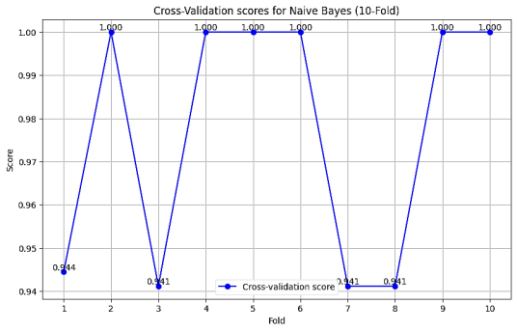
a) Akurasi
= (TP+TN) / (TP+TN+FP+FN)
= (36 + 36) / (36 + 36 + 0 + 2)
= 72 / 74 = 0, 97 x 100 = **97 %**

b) Presisi = TP / (TP+FP)
= 36 / (36 + 0)
= 1 x 100 = **100 %**

ESCAF 4th 2025
p-ISSN : 2962-7710
e-ISSN : 3021-8594

c) *Recall* = TP / (TP+FN)
= 36 / (36 + 2)
= 0,95 x 100 = **95 %**

Kemudian model divalidasi menggunakan *K-Fold Cross Validation* dengan nilai k = 10 . Hasil *Cross Validation* pada gambar 5



Gambar 5. Hasil *K-Fold Cross Validation NB*

b. Algoritma *KNN*
Tabel 29 menyajikan *Confusion Matrix* algoritma *KNN* dari 74 data.

Tabel 29. Hasil Confusion Matrix *KNN*

		Predicted Class	
		Tidak Layak	Layak
Actual Class	Tidak Layak	37	1
	Layak	0	36

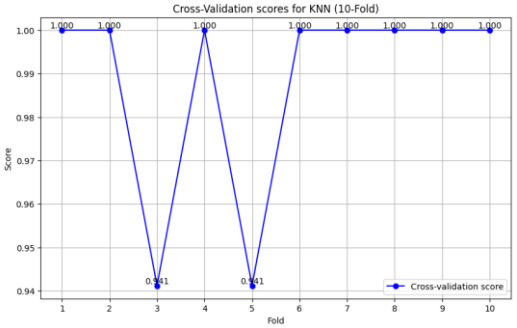
Selanjutnya confunsion matrix digunakan untuk menghitung *accuracy, precision* dan *recall* sebagai berikut :

a) Akurasi
= (TP+TN) / (TP+TN+FP+FN)
= (37+ 36) / (37 + 36 + 0 + 1)
= 73/ 74 = 0, 99 x 100 = **99 %**

b) Presisi = TP / (TP+FP)
= 37 / (37 + 0)
= 1 x 100 = **100 %**

c) *Recall* = TP / (TP+FN)
= 37 / (37 + 1) = 37 / 38
= 0,97 x 100 = **97 %**

Kemudian model divalidasi menggunakan *K-Fold Cross Validation* dengan nilai k = 10 . Hasil *Cross Validation* ditampilkan pada gambar 6



Gambar 6. Hasil *K-Fold Cross Validation KNN*

Setelah dievaluasi kedua algoritma akan diuji dengan 10 data uji baru. Tabel 30 menampilkan data uji yang akan digunakan.

Tabel 30. Data Uji Baru

Usia (C1)	Penghasilan Orang Tua (C2)	Jumlah Tanggungan (C3)	Pemilik KIP (C4)	Pemilik KPS (C5)	Status *Actual*
21	dibawah UMR	1	Tidak	Tidak	Tidak Layak
diatas 21	dibawah UMR	Lebih dari 3	Ya	Ya	Tidak Layak
dibawah 21	dibawah UMR	1	Tidak	Ya	Layak
dibawah 21	UMR	1	Ya	Ya	Layak
dibawah 21	UMR	Lebih dari 3	Tidak	Tidak	Layak
diatas 21	UMR	2	Tidak	Ya	Tidak Layak
dibawah 21	dibawah UMR	Lebih dari 3	Tidak	Tidak	Layak
diatas 21	dibawah UMR	1	Ya	Tidak	Tidak Layak
diatas 21	UMR	Lebih dari 3	Tidak	Tidak	Tidak Layak
dibawah 21	diatas UMR	2	Tidak	Tidak	Tidak Layak

Perbandingan hasil klasifikasi antara algoritma *Naïve Bayes* dan *KNN* ditampilkan pada tabel 31

Tabel 31. Hasil Klasifikasi *NB* dan *KNN*

Status *NB*	Status *KNN*	Status *Actual*
Tidak Layak	Tidak Layak	Tidak Layak
Tidak Layak	Tidak Layak	Tidak Layak
Layak	Layak	Layak
Layak	Layak	Layak
Layak	Layak	Layak
Tidak Layak	Tidak Layak	Tidak Layak
Layak	Layak	Layak
Tidak Layak	Tidak Layak	Tidak Layak
Tidak Layak	Tidak Layak	Tidak Layak
Layak	Tidak Layak	Tidak Layak

6. Implementasi Model (*Deployment*)
Dari hasil *accuracy, precision, recall* dan pengujian dengan 10 data uji baru pada tabel 31 algoritma *KNN* menunjukkan performa yang lebih unggul dibandingkan *Naïve Bayes*. Oleh karena

ESCAF 4th 2025
p-ISSN : 2962-7710
e-ISSN : 3021-8594

itu, algoritma *KNN* direkomendasikan sebagai pilihan terbaik untuk implementasi dalam sistem seleksi ini. Tampilan satu halaman web sederhananya bisa dilihat pada gambar 7.

Gambar 7. Satu Halaman Web Klasifikasi PIP

IV. KESIMPULAN

Dari penelitian yang telah dilakukan dapat diambil kesimpulan sebagai berikut :

1. Penelitian ini menunjukkan bahwa penerapan algoritma *Naïve Bayes* dan *K-Nearest Neighbor (KNN)* berhasil meningkatkan efisiensi dan akurasi dalam proses seleksi calon penerima Program Indonesia Pintar (PIP) di PKBM Orange Muara Kelingi.
2. Algoritma *KNN* menunjukkan performa yang lebih unggul dibandingkan *Naïve Bayes*. Algoritma *Naïve Bayes* mencapai akurasi sebesar 97%, presisi 100%, dan *recall* 95%, sedangkan algoritma *KNN* mencapai akurasi sebesar 99%, presisi 100%, dan *recall* 97%.
3. Model terbaik yaitu *KNN* telah diimplementasikan dalam bentuk sebuah web sederhana. Web ini memberikan kemudahan kepada pihak sekolah untuk mengakses hasil seleksi secara *real-time*.
4. Penelitian ini memberikan kontribusi nyata dalam mendukung digitalisasi proses seleksi bantuan pendidikan di Indonesia. Dengan menggunakan pendekatan *machine learning*, PKBM Orange kini memiliki solusi *modern* yang tidak hanya efektif, tetapi juga tepat sasaran.

V. SARAN

Saran yang dapat penulis berikan guna pengembangan sistem klasifikasi PIP berbasis *machine learning* ini agar lebih baik lagi adalah :

1. Agar sistem dapat menghasilkan keputusan yang lebih akurat dan relevan, peneliti menyarankan untuk menggunakan dataset yang lebih besar dan lebih beragam.
2. Selain menggunakan algoritma *Naive Bayes* dan *KNN*, ada baiknya untuk mencoba algoritma lain seperti *Random Forest* atau *Support Vector Machine (SVM)*. Dengan mengevaluasi algoritma-algoritma tersebut, kita dapat menemukan metode yang mungkin lebih unggul dalam menyelesaikan masalah klasifikasi ini.
3. Penelitian di masa depan sistem ini dapat dikembangkan lagi dengan mengimplementasikan model klasifikasi PIP pada website sekolah agar lebih mempermudah penggunaan web klasifikasi PIP.
4. Sistem ini perlu dipantau dan dievaluasi secara berkala untuk memastikan performanya tetap optimal. Langkah ini penting, terutama jika ada perubahan pada kebijakan atau kriteria penerimaan PIP yang ditetapkan oleh pemerintah.

VI. DAFTAR PUSTAKA

Alfarizi, M. R. S., Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan Python Sebagai Bahasa Pemrograman untuk Machine Learning dan Deep Learning. *Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID)*, *2*(1), 1–6.

Alkhairi, P., & Windarto, A. P. (2019). Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara. *Seminar Nasional Teknologi Komputer & Sains*, 762–767. https://seminar-id.com/prosiding/index.php/sainteks/article/view/228

Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode *Naive Bayes*. *ILKOM Jurnal Ilmiah*, *10*(2), 160–165. https://doi.org/10.33096/ilkom.v10i2.303.160-165

ESCAF 4th 2025
p-ISSN : 2962-7710
e-ISSN : 3021-8594

Aprilia, S., Elmayati, E., & Sobri, A. (2023). Data Mining Penerapan Metode K-Means Clustering Untuk Mengelompok kan Penerima Bantuan Kartu Indonesia Pintar (Studi Kasus …. *Escaf*, 1095–1100. https://semnas.univbinainsan.ac.id/index.php/escaf/article/view/487

Arfah Wahlil Pratama, M., fuad, M., Hazriani, & Yuyun. (2023). Penentuan Status Penerima Bantuan Indonesia Pintar Pada Smkn 9 Bulukumba Dengan Metode *Naive Bayes*. *Prosiding Seminar Nasional Sistem Informasi Dan Teknologi (SISFOTEK)*, 120–125.

Arifin, Z., Jafar Shudiq, W., & Magfiroh, S. (2019). Penerapan Metode *KNN* (K-Nearest Neighbor) Dalam Sistem Pendukung Keputusan Penerimaan Kip (Kartu Indonesia Pintar) Di Desa Pandean Berbasis Web Dan Mysql. *NJCA (Nusantara Journal of Computers and Its Applications)*, *4*(1). https://doi.org/10.36564/njca.v4i1.101

Juwita Sari, Sri Wulandari, Yollanda Putry, & Wiwin Handoko. (2024). Klasifikasi Kelayakan Penerima Program Indonesia Pintar Siswa MIS NU Dusun III Pinangripan. *Journal of Computers and Digital Business*, *3*(1), 1–8. https://doi.org/10.56427/jcbd.v3i1.239

Kushartanto, A. I., & Aldisa, R. T. (2023). Data Mining Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes dalam Prediksi Penerimaan Beasiswa. *Journal of Computer System and Informatics (JoSYC)*, *5*(1), 196–207. https://doi.org/10.47065/josyc.v5i1.4566

Nata, A., & Suparmadi, S. (2022). Analisis Sistem Pendukung Keputusan Dengan Model Klasifikasi Berbasis Machine Learning Dalam Penentuan Penerima Program Indonesia Pintar. *Journal of Science and Social Research*, *5*(3), 697. https://doi.org/10.54314/jssr.v5i3.1041

Nur, A. M., Nurhidayati, N., & Fathurrahman , I. (2024). Penerapan Metode Naïve Bayes Untuk Penentuan Penerima Beasiswa Program Indonesia Pintar (PIP). *Infotek: Jurnal Informatika Dan Teknologi*, *7*(1), 93–102. https://doi.org/10.29408/jit.v7i1.23995

Uriyalita, F., Syahrodi, J., & Sumanta. (2020). Evaluasi Program Indonesia Pintar (Pip) Telaah Tentang Aksesibilitas, Pencegahan Dan Penanggulangan Anak Putus Sekolah Di Wilayah Urban Fringe Harjamukti, Cirebon. *Edum Journal*, *3*(2), 179–199. https://doi.org/10.31943/edumjournal.v3i2.69