ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 2 Juli 2026 | pp: 210-220
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i2.158

Klasifikasi Kelayakan Penerima Bantuan Menggunakan Metode Naive Bayes

Andara Cantika Effendy[1*], Roberto Kaban[2]

[1] Fakultas Sains dan Teknologi, Program Studi Rekayasa Perangkat Lunak, Institut Teknologi dan Bisnis Indonesia, Deli Serdang, Indonesia
[2] Fakultas Sains dan Teknologi, Program Studi Teknologi Informatika, Institut Teknologi dan Bisnis Indonesia, Deli Serdang, Indonesia
Email : [1*]andaracantikaeffendy12a@gmail.com, [2]roberto.kaban@yahoo.com
(* Email Corresponding Author: andaracantikaeffendy12a@gmail.com)
Dikirim: 28 Juni 2026 | Diterima: 25 Juni 2026 | Dipublish: 4 Juli 2026

Abstrak

Program bantuan merupakan salah satu upaya yang diselenggarakan untuk mendukung masyarakat dalam memenuhi kebutuhan hidup. Dalam proses penentuan penerima bantuan, diperlukan suatu metode yang mampu mengelompokkan data secara objektif berdasarkan informasi yang tersedia. Penelitian ini bertujuan untuk mengklasifikasikan kelayakan penerima bantuan menggunakan algoritma Naive Bayes berdasarkan karakteristik ekonomi setiap individu. Dataset yang digunakan berasal dari Adult Income Dataset yang diperoleh melalui Kaggle, dengan atribut meliputi usia, tingkat pendidikan, jenis pekerjaan, jumlah jam kerja per minggu, dan pendapatan. Tahapan penelitian mencakup seleksi data, preprocessing, transformasi data, serta penerapan algoritma Naive Bayes**.** Dataset dibagi menjadi data latih sebesar 70% dan data uji sebesar 30% untuk proses pelatihan serta evaluasi model. Berdasarkan hasil pengujian, model Naive Bayes memperoleh tingkat akurasi sebesar 76,78%**,** precision 78,70%**,** recall 95,25%, dan F1-score sebesar 86,19%**.** Temuan tersebut menunjukkan bahwa algoritma Naive Bayes memiliki kemampuan yang cukup baik dalam mengklasifikasikan kelayakan penerima bantuan berdasarkan kondisi ekonomi individu. Model yang dikembangkan mampu memberikan performa klasifikasi yang memadai pada dataset yang digunakan. Dengan demikian, metode Naive Bayes dapat dijadikan sebagai salah satu alternatif metode klasifikasi untuk mendukung proses pengambilan keputusan dalam menentukan kelayakan penerima bantuan.

Kata Kunci: Data Mining, Klasifikasi, Naive Bayes, Kelayakan Penerima Bantuan, Pendapatan

Abstract

Assistance programs are initiatives designed to help individuals meet their living needs. Determining eligibility for such assistance requires a method capable of objectively categorizing data based on available information. This study aims to classify assistance eligibility using the Naive Bayes algorithm, based on individuals' economic characteristics. The study utilizes the "Adult Income Dataset" obtained from Kaggle, featuring attributes such as age, education level, occupation, weekly working hours, and income. The research process involves data selection, preprocessing, data transformation, and the application of the Naive Bayes algorithm. The dataset is split into a training set (70%) and a test set (30%) for model training and evaluation. Test results indicate that the Naive Bayes model achieved an accuracy of 76.78%, precision of 78.70%, recall of 95.25%, and F1-score of 86.19%. These findings demonstrate that the Naive Bayes algorithm performs effectively in classifying assistance eligibility based on individual economic conditions. The developed model delivers satisfactory classification performance on the dataset used. Consequently, the Naive Bayes method serves as a viable classification alternative to support decision-making processes regarding assistance eligibility.

Keywords*: Data Mining, Classification, Naive Bayes, Eligibility of Aid Recipients, Income*

1. PENDAHULUAN

Bantuan merupakan salah satu bentuk program yang diberikan kepada masyarakat untuk membantu memenuhi kebutuhan hidup dan meningkatkan kesejahteraan [1]. Bantuan dapat diberikan dalam berbagai bentuk seperti bantuan pangan, bantuan tunai, bantuan pendidikan, maupun bantuan kesehatan. Dalam pelaksanaannya, proses penentuan penerima bantuan harus dilakukan secara tepat agar bantuan yang diberikan benar-benar diterima oleh pihak yang membutuhkan.

Permasalahan yang sering terjadi dalam proses penentuan penerima bantuan adalah ketidaktepatan sasaran penerima. Masih terdapat penerima yang sebenarnya kurang layak namun tetap mendapatkan bantuan, sedangkan pihak yang lebih membutuhkan justru tidak menerima bantuan. Selain itu, jumlah data dan banyaknya atribut yang digunakan dalam proses penentuan penerima bantuan dapat menyulitkan proses analisis dan pengambilan keputusan. Oleh karena itu, diperlukan suatu metode klasifikasi yang mampu mengolah data secara efektif untuk membantu menentukan kelayakan penerima bantuan berdasarkan

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 2 Juli 2026 | pp: 210-220
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i2.158

karakteristik data yang dimiliki. Perkembangan teknologi informasi saat ini memberikan peluang dalam membantu proses pengolahan data secara lebih efektif. Data mining merupakan salah satu teknologi yang sering digunakan untuk mendukung proses pengolahan data.

Data mining merupakan proses analisis dan pengolahan data untuk mengidentifikasi pola maupun informasi yang bernilai dari sekumpulan data sehingga hasilnya dapat dimanfaatkan sebagai pendukung dalam proses pengambilan keputusan [2]. Teknik data mining dapat digunakan dalam berbagai bidang seperti pendidikan, kesehatan, ekonomi, pemerintahan, dan sosial. Dalam bidang sosial, teknik ini dapat dimanfaatkan untuk membantu proses klasifikasi kelayakan penerima bantuan berdasarkan data yang tersedia. Naive Bayes adalah salah satu algoritma klasifikasi yang banyak diterapkan dalam data mining.

Naive Bayes merupakan teknik klasifikasi yang menerapkan pendekatan probabilistik khususnya Teorema Bayes untuk menghitung probabilitas suatu titik data termasuk dalam kelas tertentu [3]. Metode ini mampu mengklasifikasikan data dengan memanfaatkan probabilitas yang terkait dengan data tersebut. Selain itu, metode Naive Bayes memiliki kelebihan yaitu proses perhitungannya sederhana, mudah diterapkan, dan memiliki performa yang cukup baik dalam proses klasifikasi data. Oleh karena itu, Teknik ini banyak digunakan dalam penelitian klasifikasi data.

Algoritma Naive Bayes telah digunakan dalam berbagai penelitian terdahulu sebagai teknik klasifikasi untuk beragam kasus. Dalam salah satu penelitian tersebut, Damuri dkk. (2021) mengklasifikasikan kelayakan penerima manfaat bantuan pangan pokok menggunakan algoritma Naive Bayes, dan memperoleh hasil yang menunjukkan performa klasifikasi yang baik [4]. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes dapat digunakan sebagai metode untuk mengklasifikasikan penerima bantuan berdasarkan fitur-fitur data yang tersedia. Dalam studi lain, Nuraeni dkk. (2023) mengklasifikasikan penerima bantuan keuangan langsung yang didanai oleh dana desa dengan menggunakan pendekatan SMOTE dan Naive Bayes [5]. Tingkat akurasi yang tinggi dalam penelitian tersebut menunjukkan bahwa algoritma Naive Bayes dapat digunakan sebagai teknik klasifikasi untuk mengidentifikasi penerima bantuan. Selain itu, algoritma Decision Tree C4.5 digunakan dalam penelitian oleh Pratiwi dkk. (2022) untuk menilai kelayakan klien dalam menerima bantuan sosial finansial [6]. Berdasarkan temuan penelitian, penggunaan algoritma klasifikasi dapat meningkatkan efisiensi prosedur pemilihan penerima bantuan. Selanjutnya Metode Naive Bayes digunakan dalam penelitian Huriah dan Nuris (2023) untuk mengategorikan peserta penerima bantuan sosial UMKM [7]. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes dapat diterapkan sebagai teknik klasifikasi untuk mengidentifikasi kelompok penerima bantuan berdasarkan karakteristik data yang tersedia, serta menghasilkan kinerja klasifikasi yang baik. Penelitian lain oleh Ulumuddin dan Widodo (2025) mengevaluasi kinerja algoritma Naive Bayes dalam mengklasifikasikan data pasien tuberkulosis menggunakan pendekatan data mining [8]. Penelitian tersebut membuktikan bahwa algoritma Naive Bayes dapat digunakan untuk mengklasifikasikan data kesehatan dan mengambil keputusan terkait layanan kesehatan. Hasil penelitian juga mengungkapkan bahwa penggunaan algoritma klasifikasi dapat membantu proses penentuan penerima bantuan secara lebih efektif dan akurat.

Berdasarkan sejumlah penelitian terdahulu, dapat diketahui bahwa algoritma Naive Bayes memiliki kinerja yang baik dalam mengklasifikasikan data bantuan. Namun, sebagian besar penelitian sebelumnya menggunakan data bantuan secara spesifik seperti bantuan sembako, bantuan tunai, dan bantuan kesehatan. Selain itu, sebagian penelitian menggunakan dataset lokal tertentu dengan atribut yang cukup banyak dan kompleks. Oleh karena itu, penelitian ini memiliki perbedaan dengan penelitian sebelumnya yaitu menggunakan Adult Income Dataset yang diperoleh dari Kaggle [9]. Dataset tersebut digunakan sebagai sumber data dalam proses klasifikasi kelayakan penerima bantuan. Penggunaan Adult Income Dataset dipilih karena memiliki atribut yang berkaitan dengan kondisi ekonomi individu sehingga dapat digunakan untuk melakukan simulasi klasifikasi kelayakan penerima bantuan. Dataset tersebut merupakan dataset yang memuat karakteristik demografi dan ekonomi individu seperti usia, pendidikan, pekerjaan, jam kerja, dan pendapatan. Penelitian ini memanfaatkan data karakteristik ekonomi individu untuk melakukan klasifikasi kelayakan penerima bantuan menggunakan metode Naive Bayes. Dataset yang digunakan memiliki atribut pendapatan (income) yang selanjutnya ditransformasikan menjadi kelas Layak dan Tidak Layak sesuai dengan tujuan penelitian.

Pada penelitian ini dilakukan tahap seleksi data untuk memilih atribut yang relevan, kemudian dilakukan preprocessing dan transformasi label pendapatan menjadi kategori kelayakan bantuan. Label pendapatan <=50K dikategorikan sebagai layak menerima bantuan, sedangkan label >50K dikategorikan sebagai tidak layak menerima bantuan. Transformasi tersebut dilakukan untuk menyesuaikan label pada dataset dengan tujuan penelitian, yaitu klasifikasi kelayakan penerima bantuan berdasarkan kondisi ekonomi individu. Penelitian ini diharapkan dapat memberikan gambaran mengenai penerapan metode Naive Bayes dalam melakukan klasifikasi kelayakan penerima bantuan berdasarkan data yang digunakan. Selain itu, penelitian ini

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 2 Juli 2026 | pp: 210-220
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i2.158

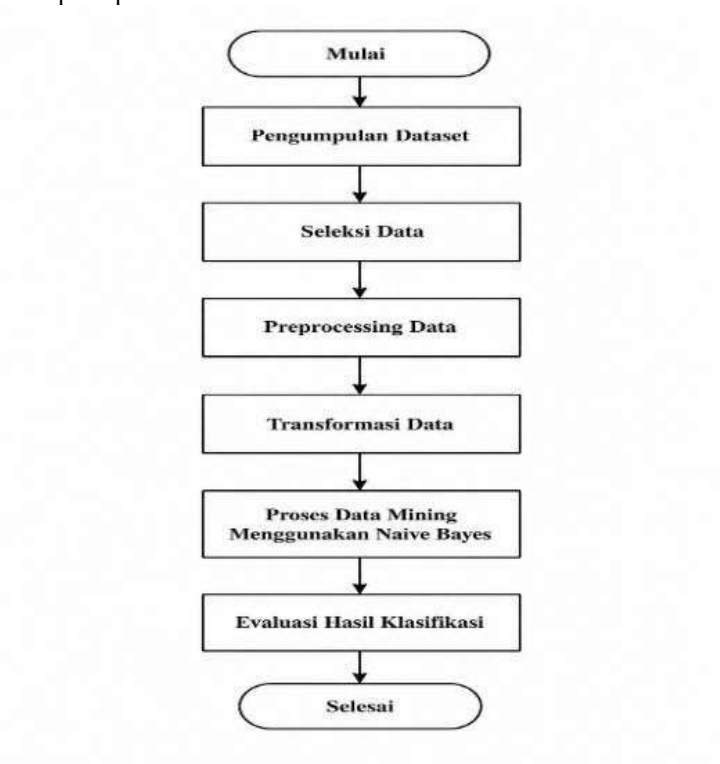
juga diharapkan dapat menjadi referensi dalam pengembangan penelitian yang memanfaatkan metode Naive Bayes pada permasalahan klasifikasi data.

Tujuan penelitian ini adalah menerapkan metode Naive Bayes untuk melakukan klasifikasi kelayakan penerima bantuan berdasarkan karakteristik ekonomi individu serta mengevaluasi performa model menggunakan accuracy, precision, recall, dan F1-score. Hasil penelitian diharapkan dapat memberikan informasi yang mendukung proses pengambilan keputusan dalam menentukan kelayakan penerima bantuan secara lebih objektif dan akurat.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan dalam penelitian ini menerapkan konsep Knowledge Discovery in Databases (KDD). KDD merupakan teknik pemrosesan data bertahap untuk memperoleh informasi atau pengetahuan baru dari sekumpulan data [10]. Tahapan penelitian dilakukan secara sistematis untuk membantu proses klasifikasi kelayakan penerima bantuan menggunakan algoritma Naive Bayes. Berikut merupakan flowchart tahapan penelitian yang dilakukan pada penelitian ini.



Gambar 1. Flowchart Tahapan Penelitian Metode KDD

Tahapan penelitian dimulai dari proses pengumpulan dataset yang diperoleh dari Kaggle, kemudian dilakukan seleksi data untuk memilih atribut yang relevan dengan tujuan penelitian. Selanjutnya dilakukan tahap preprocessing berupa pemeriksaan dan penyesuaian data sehingga data dapat digunakan pada proses klasifikasi. Proses data mining dilakukan menggunakan metode Naive Bayes untuk menentukan klasifikasi kelayakan penerima bantuan. Tingkat kinerja model akhir kemudian ditentukan dengan mengevaluasi hasil klasifikasi menggunakan confusion matrix, accuracy, precision, recall, dan F1-score.

2.2 Dataset dan Metode Penelitian

Penelitian ini memanfaatkan Adult Income Dataset yang diperoleh melalui platform Kaggle sebagai sumber data penelitian [9]. Dataset tersebut berasal dari *UCI Machine Learning Repository* dan berisi data demografi serta kondisi ekonomi individu, seperti usia, pendidikan, pekerjaan, jam kerja, dan pendapatan. Adult Income Dataset banyak digunakan sebagai dataset *benchmark* dalam penelitian *machine learning* dan klasifikasi data karena memiliki atribut kategorikal dan numerik yang beragam.

Dataset ini dipilih karena memiliki atribut pendapatan (income) yang dapat dimanfaatkan sebagai dasar dalam proses klasifikasi kelayakan penerima bantuan setelah dilakukan transformasi label sesuai dengan tujuan penelitian. Informasi mengenai struktur dataset yang digunakan dalam penelitian ini disajikan dalam Tabel 1.

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 2 Juli 2026 | pp: 210-220
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i2.158

Tabel 1. Struktur Dataset Adult Income

No	Age	Workclass	Fnlwgt	Education	Education-num	…	Income
1	90	?	77053	HS-grad	9	…	<=50K
2	82	Private	132870	HS-grad	9	…	<=50K
3	66	?	186061	Some-college	10	…	<=50K
4	54	Private	140359	7th-8th	4	…	<=50K
5	41	Private	264663	Some-college	10	…	<=50K
…	…	…	…	…	…	…	…
32.561	22	Private	201490	HS-grad	9	…	<=50K

Keterangan: Simbol "?" pada beberapa atribut menunjukkan nilai yang tidak diketahui (*missing value*) pada dataset asli yang diperoleh dari Adult Income Dataset. Berdasarkan Tabel 1, Adult Income Dataset terdiri atas 32.561 data dengan 15 atribut yang menggambarkan karakteristik demografi dan ekonomi individu. Setelah itu, prosedur pemilihan atribut digunakan untuk memilih karakteristik yang relevan dengan tujuan penelitian.

2.3 Seleksi Data

Atribut yang dianggap relevan dengan proses klasifikasi kelayakan penerima bantuan dipilih selama tahap seleksi data. Proses seleksi data bertujuan untuk mengurangi atribut yang tidak diperlukan sehingga proses pengolahan data menjadi lebih efektif dan mudah dilakukan [11].

Dataset asli Adult Income Dataset memiliki 15 atribut, namun pada penelitian ini hanya digunakan atribut yang berkaitan dengan kondisi ekonomi individu yaitu usia (*age*), pendidikan (*education*), pekerjaan (*occupation*), jam kerja (*hours-per-week*), dan pendapatan (*income*). Hasil proses seleksi atribut disajikan pada Tabel 2.

Tabel 2. Hasil Seleksi Atribut

No	Age	Education	Occupation	Hours-per-week	Income
1	90	HS-grad	?	40	<=50K
2	82	HS-grad	Exec-managerial	18	<=50K
3	66	Some-college	?	40	<=50K
4	54	7th-8th	Machine-op-inspct	40	<=50K
5	41	Some-college	Prof-specialty	40	<=50K
…	…	…	…	…	…
32.561	22	HS-grad	Adm-clerical	20	<=50K

Pemilihan atribut dilakukan karena atribut tersebut dianggap memiliki pengaruh dalam menentukan tingkat kelayakan penerima bantuan. Setelah proses seleksi data dilakukan, atribut yang telah dipilih selanjutnya digunakan pada tahap preprocessing dan transformasi data sebelum metode Naive Bayes digunakan untuk proses klasifikasi.

2.4 Preprocessing Data

Tahap preprocessing merupakan proses persiapan data sebelum dilakukan proses klasifikasi [12]. Tahap preprocessing dilakukan untuk mengorganisir dan menyesuaikan data sehingga proses klasifikasi menggunakan metode Naive Bayes dapat dilakukan dengan lebih baik [13]. Selain itu, dilakukan pemeriksaan terhadap dataset untuk mengidentifikasi adanya data yang tidak lengkap (*missing value*) serta memastikan bahwa data yang digunakan telah memenuhi kualitas yang diperlukan dalam penelitian. Hasil pemeriksaan menunjukkan bahwa terdapat beberapa atribut yang memiliki nilai tidak diketahui yang ditandai dengan simbol "?". Nilai tersebut dipertahankan sesuai dengan kondisi dataset asli dan tetap digunakan dalam proses penelitian karena tidak memengaruhi proses transformasi label maupun proses klasifikasi yang dilakukan. Selain itu, dilakukan penyesuaian data yang digunakan sesuai dengan kebutuhan proses klasifikasi menggunakan metode Naive Bayes.

2.5 Transformasi Data

Transformasi data merupakan salah satu tahapan dalam proses KDD yang bertujuan mengubah data ke dalam bentuk yang sesuai sehingga dapat digunakan dalam proses data mining dan klasifikasi [14]. Pada tahap ini dilakukan transformasi label pada atribut income. Label income pada dataset awal terdiri dari dua kategori yaitu <=50K dan >50K.

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 2 Juli 2026 | pp: 210-220
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i2.158

Dalam penelitian ini label tersebut ditransformasikan menjadi kategori kelayakan bantuan. Nilai <=50K dikategorikan sebagai Layak penerima bantuan, sedangkan nilai >50K dikategorikan sebagai Tidak Layak menerima bantuan. Hasil transformasi data ditampilkan pada Tabel 3.

Tabel 3. Transformasi Label Kelayakan

Income	Kelayakan
<=50K	Layak
>50K	Tidak Layak

Transformasi ini dilakukan karena penelitian berfokus pada klasifikasi kelayakan penerima bantuan berdasarkan kondisi ekonomi individu yang direpresentasikan melalui atribut pendapatan (income).

2.6 Metode Data Mining Naive Bayes

Data mining merupakan serangkaian proses analisis pengolahan data yang bertujuan mengidentifikasi pola maupun informasi bernilai dari kumpulan data dalam jumlah besar [15]. Salah satu teknik yang banyak diterapkan dalam data mining adalah klasifikasi, yaitu proses pengelompokan data ke dalam kategori tertentu berdasarkan karakteristik atau atribut yang dimiliki oleh setiap data. Dalam proses klasifikasi terdapat berbagai algoritma yang dapat digunakan, seperti Decision Tree, Naive Bayes, dan K-Nearest Neighbor (KNN). Pada penelitian ini algoritma yang digunakan untuk melakukan klasifikasi data adalah metode Naive Bayes.

Metode Naive Bayes merupakan metode klasifikasi berbasis probabilitas yang menggunakan konsep Teorema Bayes [16]. Metode ini digunakan untuk menentukan probabilitas suatu data terhadap kelas tertentu berdasarkan atribut yang dimiliki. Rumus Teorema Bayes ditunjukkan pada Persamaan (1).

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

Keterangan:
X = Data yang akan diklasifikasikan
H = Hipotesis atau kelas tertentu
P(H|X) = Probabilitas posterior dari H berdasarkan X
P(X|H) = Probabilitas data X jika hipotesis H benar (likelihood)
P(H) = Probabilitas prior dari hipotesis H
P(X) = Probabilitas data X
Dalam penelitian ini, metode Naive Bayes diterapkan untuk mengelompokkan data ke dalam dua kategori, yaitu Layak dan Tidak Layak.

2.7 Evaluasi Model

Tahap evaluasi dilakukan untuk mengetahui tingkat performa metode Naive Bayes dalam melakukan klasifikasi data. Pengukuran kinerja model dilakukan dengan memanfaatkan confusion matrix yang terdiri atas empat komponen, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Data yang diperoleh dari setiap komponen kemudian digunakan untuk menghitung nilai accuracy, precision, recall, dan F1-score sebagai indicator performa model [8].

Accuracy merupakan tingkat ketepatan model dalam mengklasifikasi seluruh data yang diuji. Precision merupakan tingkat ketepatan model dalam memprediksi data pada kelas positif. Recall merupakan tingkat peforma model dalam mengenali seluruh data yang berada dalam kategori positif. Selain itu, F1-score merupakan metrik evaluasi yang digunakan untuk mengukur keseimbangan antara precision dan recall [17]. Untuk menghitung nilai accuracy, precision, recall, dan F1-score dapat menggunakan persamaan berikut:

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (2)$$

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad (3)$$

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (4)$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

Keterangan:
TP (True Positive) = Data positif yang diprediksi benar
TN (True Negative) = Data negatif yang diprediksi benar
FP (False Positive) = Data negatif yang diprediksi positif
FN (False Negative) = Data positif yang diprediksi negatif

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 2 Juli 2026 | pp: 210-220
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i2.158

Hasil evaluasi model digunakan untuk mengetahui performa metode Naive Bayes dalam melakukan klasifikasi kelayakan penerima bantuan.

3. HASIL DAN PEMBAHASAN

3.1 Implementasi/Pengujian

Pada penelitian ini implementasi metode Naive Bayes dilakukan menggunakan Microsoft Excel dan Python. Microsoft Excel digunakan untuk melakukan perhitungan manual yang meliputi perhitungan probabilitas prior, nilai mean, standar deviasi, serta probabilitas Gaussian. Selanjutnya, Python digunakan untuk membangun model klasifikasi Naive Bayes, melakukan proses prediksi, serta mengevaluasi performa model menggunakan confusion matrix, accuracy, precision, recall, dan F1-score. Implementasi dan pengujian dilakukan menggunakan Adult Income Dataset yang telah melalui tahap seleksi atribut, preprocessing, dan transformasi data.

3.1.1 Perhitungan Manual Metode Naive Bayes Menggunakan Microsoft Excel

Perhitungan manual dilakukan menggunakan Microsoft Excel untuk mengetahui proses klasifikasi pada metode Naive Bayes. Perhitungan meliputi probabilitas prior, nilai mean, standar deviasi, probabilitas Gaussian, serta probabilitas posterior. Perhitungan dilakukan menggunakan seluruh data hasil transformasi yang berjumlah 32.561 data sehingga hasil yang diperoleh sesuai dengan dataset yang digunakan dalam penelitian.

a. Perhitungan Probabilitas Prior

Probabilitas prior digunakan untuk mengetahui peluang awal masing-masing kelas sebelum mempertimbangkan atribut yang digunakan dalam proses klasifikasi. Nilai probabilitas prior dihitung berdasarkan perbandingan jumlah data pada masing-masing kelas terhadap total data yang digunakan dalam penelitian. Hasil perhitungan tersebut ditunjukkan pada Tabel 4.

Tabel 4. Nilai Probabilitas Prior Berdasarkan Kelas

Kelas	Jumlah Data	Probabilitas
Layak	24.720	0,7592
Tidak Layak	7.841	0,2408
Total	32.561	1,0000

Berdasarkan Tabel 4 diperoleh jumlah data kelas Layak sebanyak 24.720 data dan kelas Tidak Layak sebanyak 7.841 data. Hasil perhitungan menunjukkan bahwa probabilitas prior kelas Layak sebesar 0,7592 dan probabilitas prior kelas Tidak Layak sebesar 0,2408.

b. Perhitungan Mean dan Standar Deviasi

Nilai mean digunakan untuk mengetahui nilai rata-rata atribut pada masing-masing kelas, sedangkan standar deviasi digunakan untuk mengetahui tingkat penyebaran data terhadap nilai rata-ratanya. Nilai mean dan standar deviasi digunakan dalam proses perhitungan probabilitas Gaussian pada metode Naive Bayes. Hasil Perhitungan Mean dan Standar Deviasi ditunjukkan pada Table 5.

Tabel 5. Nilai Mean dan Standar Deviasi Setiap Kelas

Kelas	Mean Age	Std. Age	Mean Hours-Per-Week	Std. Hours-Per-Week
Layak	36,7837	14,0198	38,8402	12,31871
Tidak Layak	44,2498	10,5184	45,4730	11,0123

Tabel 5 menunjukkan bahwa kelas Tidak Layak memiliki usia rata-rata yang lebih tinggi dari pada kelas Layak. Selain itu, rata-rata jam kerja perminggu kelas Tidak Layak lebih tinggi dari pada kelas Layak. Perhitungan probabilitas Gaussian kemudian dilakukan menggunakan nilai rata-rata dan deviasi standar ini.

c. Perhitungan Probabilitas Gaussian

Probabilitas Gaussian kemudian dihitung untuk setiap kelas setelah mean dan standar deviasi ditentukan. Perhitungan probabilitas Gaussian digunakan untuk menentukan peluang kemunculan suatu nilai atribut pada masing-masing kelas. Data uji yang digunakan dalam penelitian ini mencakup usia 40 tahun dan 40 jam kerja per minggu.

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 2 Juli 2026 | pp: 210-220
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i2.158

Tabel 6. Data Uji

Atribut	Nilai
Age	40
Hours-Per-Week	40

Probabilitas Gaussian dihitung untuk setiap atribut pada masing-masing kelas, yaitu kelas Layak dan Tidak Layak. Probabilitas posterior dihitung menggunakan nilai probabilitas yang diperoleh. Perhitungan probabilitas Gaussian dilakukan pada atribut usia dan jam per minggu untuk kelas Layak dan Tidak Layak. Hasil perhitungan probabilitas Gaussian ditunjukkan pada Tabel 7.

Tabel 7. Nilai Probabilitas Gaussian

Atribut	Layak	Tidak Layak
Age	0,027717	0,034955
Hours-per-week	0,032240	0,032020

Berdasarkan Tabel 7 diketahui bahwa probabilitas Gaussian untuk atribut age pada kelas Layak sebesar 0,027717 dan pada kelas Tidak Layak sebesar 0,034955. Sementara itu, probabilitas Gaussian untuk atribut hours-per-week pada kelas Layak sebesar 0,032240 dan pada kelas Tidak Layak sebesar 0,032020. Nilai probabilitas Gaussian tersebut selanjutnya digunakan dalam perhitungan probabilitas posterior.

d. Perhitungan Probabilitas Posterior

Probabilitas posterior dihitung dengan mengalikan probabilitas prior dengan seluruh probabilitas Gaussian dari atribut yang digunakan dalam proses klasifikasi. Perhitungan dilakukan untuk masing-masing kelas, yaitu kelas Layak dan kelas Tidak Layak.
Perhitungan probabilitas posterior untuk kelas Layak adalah sebagai berikut:
P(Layak|X) = P(Layak) × P(Age|Layak) × P(Hours-per-week|Layak)
P(Layak|X) = 0,7592 × 0,027717 × 0,032242
P(Layak|X) = 0,000678
Perhitungan probabilitas posterior untuk kelas Tidak Layak adalah sebagai berikut:
P(Tidak Layak|X) = P(Tidak Layak) × P(Age|Tidak Layak) × P(Hours-per-week|Tidak Layak)
P(Tidak Layak|X) = 0,2408 × 0,034955 × 0,032018
P(Tidak Layak|X) = 0,000270
Hasil perhitungan probabilitas posterior ditunjukkan pada Tabel 8.

Tabel 8. Hasil Perhitungan Probabilitas Posterior

Kelas	Nilai Probabilitas
Layak	0,000678
Tidak Layak	0,000270

Berdasarkan Tabel 8 diketahui bahwa nilai probabilitas posterior kelas Layak lebih besar dibandingkan kelas Tidak Layak.

e. Hasil Klasifikasi

Hasil klasifikasi ditentukan berdasarkan nilai probabilitas posterior terbesar yang diperoleh dari masing-masing kelas. Berdasarkan hasil perhitungan diperoleh nilai posterior kelas Layak sebesar 0,000678 dan kelas Tidak Layak sebesar 0,000270. Karena nilai posterior kelas Layak lebih besar dibandingkan kelas Tidak Layak, maka data uji dengan nilai age sebesar 40 tahun dan hours-per-week sebesar 40 jam kerja per minggu diklasifikasikan ke dalam kategori Layak menerima bantuan.

Hasil perhitungan manual menggunakan Microsoft Excel mengindikasikan bahwa algoritma Naive Bayes dapat melakukan klasifikasi berdasarkan atribut yang digunakan. Hasil tersebut selanjutnya digunakan sebagai dasar untuk membandingkan hasil implementasi menggunakan Python.

3.1.2 Hasil Pengolahan Dataset

Dataset yang digunakan pada penelitian ini adalah Adult Income Dataset yang diperoleh dari Kaggle. Pengolahan dataset dilakukan menggunakan bahasa pemrograman Python dengan bantuan library Pandas

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 2 Juli 2026 | pp: 210-220
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i2.158

untuk membaca dan mengelola data. Berdasarkan hasil pembacaan dataset diperoleh sebanyak 32.561 data dengan 15 atribut yang menggambarkan karakteristik demografi dan kondisi ekonomi individu.

Selanjutnya dilakukan proses seleksi atribut untuk memilih atribut yang relevan dengan tujuan penelitian. Hasil seleksi atribut menghasilkan lima atribut yang digunakan dalam proses klasifikasi, yaitu age, education, occupation, hours-per-week, dan income. Pemilihan atribut tersebut dilakukan karena dianggap mampu merepresentasikan kondisi ekonomi individu serta memiliki keterkaitan dengan proses klasifikasi kelayakan penerima bantuan. Informasi dataset yang digunakan dalam penelitian ditunjukkan pada Tabel 9.

Tabel 9. Informasi Dataset

Keterangan	Nilai
Jumlah Data	32.561
Jumlah Atribut Awal	15
Jumlah Atribut Terpilih	5

Berdasarkan hasil seleksi atribut, data yang telah diperoleh selanjutnya digunakan pada tahap transformasi label dan proses klasifikasi menggunakan metode Naive Bayes.

3.1.3 Transformasi Label

Transformasi dilakukan pada atribut income untuk membentuk kelas target yang digunakan dalam proses klasifikasi. Transformasi dilakukan untuk menyesuaikan label pada dataset dengan tujuan penelitian, yaitu klasifikasi kelayakan penerima bantuan. Pada penelitian ini, nilai income ≤50K dikategorikan sebagai Layak, sedangkan nilai income >50K dikategorikan sebagai Tidak Layak. Hasil Transformasi label disajikan pada Tabel 10.

Tabel 10. Transformasi Label kelas

Income	Kelayakan
<=50K	Layak
>50K	Tidak Layak

Hasil transformasi menunjukkan bahwa seluruh data berhasil dikelompokkan ke dalam dua kelas, yaitu Layak dan Tidak Layak. Kelas tersebut selanjutnya digunakan sebagai target pada proses klasifikasi menggunakan metode Naive Bayes.

3.1.4 Hasil Klasifikasi Naive Bayes

Library Scikit-Learn dan bahasa pemrograman Python digunakan untuk mengimplementasikan algoritma Naive Bayes. Metode train-test split diterapkan untuk membagi dataset menjadi 70% data pelatihan dan 30% data pengujian sebelum proses klasifikasi dilakukan. Pembagian data dilakukan agar model mampu mengenali pola dalam data pelatihan dan kemudian dievaluasi menggunakan data pengujian yang tidak disertakan dalam tahap pelatihan. Tabel 11 menampilkan hasil dari pembagian data tersebut.

Tabel 11. Pembagian Data

Jenis Data	Jumlah
Data Latih	22.792
Data Uji	9.769
Total	32.561

Sebanyak 22.792 data pelatihan dan 9.769 data pengujian diperoleh berdasarkan pembagian data. Model Naive Bayes dibangun menggunakan data pelatihan sedangkan data pengujian digunakan untuk menguji performa model terhadap data yang tidak digunakan pada proses pelatihan. Model yang telah dibangun kemudian digunakan untuk melakukan prediksi terhadap data Uji. Hasil prediksi selanjutnya dievaluasi menggunakan confusion matrix, accuracy, precision, recall, dan F1-score untuk mengetahui tingkat performa model dalam melakukan klasifikasi.

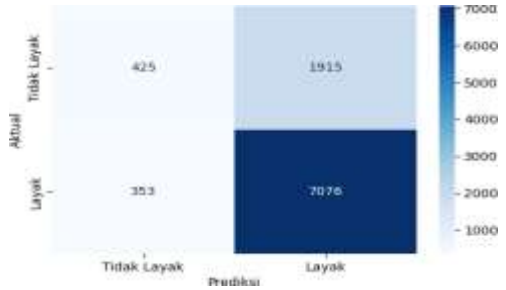
3.2 Evaluasi Model

3.2.1 Confusion Matrix

Confusion matrix serta metrik accuracyi, precision, recall, dan F-score digunakan untuk evaluasi dalam penelitian ini. Confusion matrix digunakan untuk membandingkan hasil prediksi model dengan data aktual sehingga dapat diketahui jumlah data yang berhasil diklasifikasikan dengan benar maupun yang

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 2 Juli 2026 | pp: 210-220
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i2.158

mengalami kesalahan klasifikasi. Hasil confusion matrix untuk pendekatan Naive Bayes disajikan pada Gambar 2.



Gambar 2. Confusion Matrix metode Naive Bayes

Berdasarkan confusion matrix pada Gambar 2, diperoleh hasil klasifikasi yang ditunjukkan pada Tabel 12.

Tabel 12. Hasil Confusion Matrix

	Prediksi Tidak Layak	**Prediksi Layak**
Aktual Tidak Layak	425	1915
Aktual Layak	353	7076

Berdasarkan hasil confusion matrix diperoleh sebanyak 7.076 data kelas Layak yang berhasil diklasifikasikan dengan benar dan 425 data kelas Tidak Layak yang berhasil diklasifikasikan dengan benar. Selain itu, terdapat 353 data kelas Layak yang diprediksi sebagai Tidak Layak dan 1.915 data kelas Tidak Layak yang diprediksi sebagai Layak. Hasil tersebut menunjukkan bahwa model Naive Bayes mampu melakukan klasifikasi dengan cukup baik meskipun masih terdapat sejumlah data yang mengalami kesalahan klasifikasi.

3.2.2 Accuracy, Precision, Recall, F1-Score

Berdasarkan hasil confusion matrix, dilakukan perhitungan nilai accuracy, precision, recall, dan F1-score untuk mengetahui tingkat performa model klasifikasi. Tabel 13 menampilkan hasil evaluasi.

Tabel 13. Hasil Evaluasi

Metrik	**Nilai**
Accuracy	76,78%
Precision	78,70%
Recall	95,25%
F1-Score	86,19%

Berdasarkan data evaluasi pada Tabel 13, algoritma Naive Bayes memperoleh accuracy sebesar 76,78%, presisi 78,70%, recall 95,25%, dan F1-score 86,19%. Angka tersebut menunjukkan bahwa model mampu mengategorikan data penelitian dengan baik terhadap data yang digunakan pada penelitian

3.3 Pembahasan

Berdasarkan implementasi serta hasil evaluasi pada penelitian ini, metode Naive Bayes berhasil diterapkan untuk melakukan klasifikasi kelayakan penerima bantuan berdasarkan karakteristik ekonomi individu yang terdapat pada Adult Income Dataset. Proses klasifikasi menggunakan atribut age, education, occupation, dan hours-per-week yang telah melalui tahap seleksi atribut, preprocessing, dan transformasi data. Atribut-atribut tersebut dipilih karena mencerminkan secara akurat kondisi keuangan masyarakat, yang menjadi dasar untuk menentukan siapa yang memenuhi syarat menerima bantuan.

Hasil evaluasi menunjukkan bahwa metode Naive Bayes memperoleh nilai accuracy sebesar 76,78%, precision sebesar 78,70%, recall sebesar 95,25%, dan F1-score sebesar 86,19%. Nilai accuracy menyatakan bahwa model tersebut mampu mengklasifikasikan sebagian besar data dengan benar. Nilai precision sebesar 78,70% mengindikasikan bahwa sebagian besar data yang diprediksi sebagai kelas Layak sesuai dengan kondisi sebenarnya. Perolehan recall sebesar 95,25% menunjukkan bahwa model memiliki kemampuan yang tinggi dalam mengidentifikasi data yang termasuk ke dalam kelas Layak. Nilai recall yang tinggi menunjukkan

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 2 Juli 2026 | pp: 210-220
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i2.158

bahwa model mampu mengidentifikasi Sebagian besar data dalam kelas target. Selain itu, nilai F1-score sebesar 86,19% menunjukkan adanya keseimbangan yang baik antara precision dan recall, sehingga performa klasifikasi yang dihasilkan dapat dikategorikan cukup baik.

Berdasarkan hasil confusion matrix, diperoleh sebanyak 7.076 data kelas Layak yang berhasil diklasifikasikan dengan benar (True Positive) dan 425 data kelas Tidak Layak yang berhasil diklasifikasikan dengan benar (True Negative). Namun, masih terdapat kesalahan klasifikasi berupa 353 data kelas Layak yang diprediksi sebagai Tidak Layak (False Negative) serta 1.915 data kelas Tidak Layak yang diprediksi sebagai Layak (False Positive). Kesalahan klasifikasi tersebut dapat disebabkan oleh keragaman karakteristik data yang menyebabkan sebagian pola belum dapat dipelajari secara optimal oleh model Naive Bayes.

Hasil penelitian yang memperoleh accuracy sebesar 76,78%, menunjukkan bahwa metode Naive Bayes mampu memberikan kinerja klasifikasi yang memadai dalam menentukan kelayakan penerima bantuan. Hasil ini sejalan dengan penelitian yang dilakukan oleh Damuri dkk. (2021), yang mengklasifikasikan kelayakan penerima bantuan pangan pokok menggunakan teknik Naive Bayes dan memperoleh hasil klasifikasi yang positif. Menurut penelitian Nuraeni dkk. (2023), terdapat tingkat akurasi yang tinggi dalam mengklasifikasikan penerima bantuan uang tunai langsung yang bersumber dari dana desa. Selain itu, penelitian oleh Huriah dan Nuris (2023) menunjukkan bahwa penerima bantuan sosial UMKM dapat diklasifikasikan secara efektif menggunakan metode Naive Bayes. Algoritma Naive Bayes sangat efektif dalam mengidentifikasi data sosial-ekonomi dan dapat digunakan sebagai alat pendukung keputusan untuk menentukan kelayakan penerima, sebagaimana ditunjukkan oleh hasil yang konsisten dari berbagai penelitian. Angka recall yang lebih tinggi dibandingkan nilai precision mengindikasikan bahwa model tersebut lebih baik dalam mengidentifikasi data yang termasuk dalam kelas Layak dibandingkan dalam menghindari kesalahan prediksi pada kelas tersebut. Kondisi ini dapat menjadi keuntungan apabila sistem digunakan sebagai tahap penyaringan awal calon penerima bantuan, karena sebagian besar individu yang layak dapat teridentifikasi oleh model. Namun demikian, masih terdapat sejumlah data Tidak Layak yang diprediksi sebagai Layak sehingga diperlukan proses verifikasi lanjutan sebelum keputusan akhir ditetapkan. Oleh karena itu, hasil klasifikasi yang diperoleh sebaiknya digunakan sebagai alat bantu pengambilan keputusan dan bukan sebagai satu-satunya dasar dalam menentukan penerima bantuan.

Hasil penelitian juga menyatakan bahwa metode Naive Bayes mampu menangani dataset berskala besar dengan beragam atribut serta menghasilkan kinerja klasifikasi yang memuaskan. Namun, pendekatan Naive Bayes didasarkan pada asumsi bahwa karakteristik-karakteristik yang ada tidak saling berkaitan. Dalam kondisi tertentu, hubungan antar atribut yang kuat dapat memengaruhi hasil klasifikasi sehingga performa model belum tentu optimal. Oleh karena itu, penelitian dimasa mendatang dapat mempertimbangkan penggunaan metode klasifikasi lain atau penerapan teknik seleksi fitur yang lebih optimal untuk meningkatkan performa model.
Secara keseluruhan, hasil penelitian membuktikan bahwa algoritma Naive Bayes merupakan teknik klasifikasi alternatif yang layak untuk menentukan kelayakan penerima bantuan berdasarkan karakteristik ekonomi individu. Keunggulan metode ini terletak pada proses perhitungannya yang sederhana, mudah diimplementasikan, serta mampu menghasilkan performa yang cukup baik pada dataset berukuran besar. Dengan demikian, metode Naive Bayes dapat membantu pengambilan keputusan mengenai kelayakan penerima bantuan yang lebih imparsial dan efektif.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, Adult Income Dataset yang diperoleh dari Kaggle berhasil diterapkan untuk melakukan klasifikasi kelayakan penerima bantuan menggunakan metode Naive Bayes. Penelitian dilakukan melalui beberapa tahapan, yaitu pembacaan dataset, seleksi atribut, transformasi data,membagi data menjadi data pelatihan dan pengujian dengan rasio 70:30, serta menggunakan metode Naive Bayes sebagai algoritma klasifikasi. Atribut yang digunakan dalam penelitian meliputi usia (age), pendidikan (education), pekerjaan (occupation), dan jam kerja per minggu (hours-per-week) yang dianggap mampu merepresentasikan kondisi ekonomi individu. Hasil pengujian menunjukkan bahwa, algoritma Naive Bayes mencapai accuracy sebesar 76,78%, precision 78,70%, recall 95,25%, dan F1-score 86,19%. Angka-angka tersebut menunjukkan bahwa model ini mampu mengategorikan data dengan baik berdasarkan atribut-atribut penelitian. Nilai recall yang tinggi menunjukkan kemampuan model dalam mengidentifikasi sebagian besar titik data yang termasuk dalam kelas Layak, yang mengindikasikan bahwa pendekatan ini dapat membantu proses pengambilan keputusan terkait kelayakan penerima bantuan. Dengan demikian, tujuan penelitian untuk menerapkan algoritma Naive Bayes dalam klasifikasi kelayakan penerima bantuan telah tercapai. Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan. Dataset yang digunakan merupakan Adult Income Dataset yang berasal dari Kaggle dan bukan dataset khusus penerima bantuan di Indonesia, sehingga

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 2 Juli 2026 | pp: 210-220
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i2.158

hasil klasifikasi yang diperoleh belum sepenuhnya menggambarkan kondisi penerima bantuan yang sebenarnya di Indonesia. Selain itu, karena hanya satu algoritma klasifikasi yaitu Naive Bayes yang digunakan dalam penelitian ini, kinerjanya belum dapat dibandingkan dengan algoritma klasifikasi lainnya seperti Decision Tree, Random Forest, atau Support Vector Machine. Oleh karena itu, penelitian selanjutnya dapat menggunakan dataset yang lebih spesifik dan disesuaikan dengan kondisi penerima bantuan, serta membandingkan berbagai pendekatan klasifikasi untuk mengembangkan model dengan kinerja yang lebih optimal.

REFERENCES

[1] F. F. B. Sembiring and R. Nababan, "Data Terpadu Kesejahteraan Sosial Terhadap Bantuan Sosial Bagi Kesejahteraan Ekonomi Masyarakat Kelurahan Simpang Selayang," *Innov. J. Soc. Sci. Res.*, vol. 4, no. 5, pp. 6779–6790, 2024.

[2] M. A. M. Nasir, *Data Mining: Algoritma dan Implementasi*. Yogyakarta: Andi, 2020.

[3] 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, 2023, doi: 10.36040/jati.v7i1.6303.

[4] 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, 2021, doi: 10.30865/jurikom.v8i6.3655.

[5] D. Kurniadi, F. Nuraeni, and M. Firmansyah, "Classification Of Society Recipients of Bantuan Langsung Tunai Dana Desa Using Naive Bayes Algorithm and SMOTE," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 36, pp. 1–11, 2023, doi: 10.25126/jtiik.2023106453.

[6] N. W. U. Pratiwi, "KLASIFIKASI PENENTUAN PENERIMA BANTUAN SOSIAL TUNAI (BST) MENGGUNAKAN ALGORITMA C4.5 DI DESA KERAMAS, GIANYAR BALI," vol. 4, no. 3, pp. 101–107, 2022.

[7] D. Azlil Huriah and N. Dienwati Nuris, "Klasifikasi Penerima Bantuan Sosial Umkm Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 360–365, 2023, doi: 10.36040/jati.v7i1.6300.

[8] W. P. Ulumudidin, "Analisis Kinerja Algoritma Naive Bayes dalam Klasifikasi Data pada Pasien Tuberkulosis Berbasis Data Mining," vol. 5, no. 1, pp. 75–81, 2025, doi: 10.47065/jogtc.v5i1.8999.

[9] Kaggle, "Adult Census Income." Accessed: May 15, 2026. [Online]. Available: https://www.kaggle.com/datasets/uciml/adult-census-income

[10] T. M. Marisa, F.; Maukar, A. L.; Akhriza, *Data Mining Konsep dan Penerapannya*. Yogyakarta: Deepublish, 2021.

[11] F. L. Hanis, U. Khaira, and D. Arsa, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Classification of Student Drop Out Risk Using Decision Tree C5.0 with Feature Selection Klasifikasi Status Mahasiswa Berisiko Drop Out Menggunakan Decision Tree C5.0 dengan Seleksi Fitur," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. October, pp. 1318–1330, 2025.

[12] T. Gori, A. Sunyoto, and H. Al Fatta, "Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 1, pp. 215–224, 2024, doi: 10.25126/jtiik.20241118074.

[13] Y. P. Astuti, A. R. Wibowo, E. Kartikadarma, E. R. Subhiyakto, N. A. Sri Winarsih, and M. S. Rohman, "Penerapan Metode Naïve Bayes Classifier Untuk Klasifikasi Sentimen Pada Judul Berita," *LogicLink*, vol. 1, no. 1, pp. 1–12, 2024, doi: 10.28918/logiclink.v1i1.7684.

[14] Y. Sari Siregar, D. Handoko, M. Khairani, N. I. Syahputri, and H. Harahap, "Implementasi Data Mining Klasifikasi Algoritma Chaid Dalam Menentukan Pola Penerima Mahasiswa Baru," *Digit. Transform. Technol.*, vol. 3, no. 2, pp. 978–989, 2024, doi: 10.47709/digitech.v3i2.3612.

[15] Prastyadi Wibawa Rahayu dkk, *Buku Ajar Data Mining*. Bandung: Sonpedia Publishing Indonesia, 2024.

[16] Y. Irfayanti, "Penerapan Metode Naive Bayes untuk Klasifikasi Status Pegawai pada Perusahaan Swasta," *Syntax Lit. J. Ilm. Indones.*, vol. 7, no. 10, pp. 17934–17941, 2022, doi: 10.36418/syntax-literate.v7i10.13230.

[17] S. A. Hicks *et al.*, "On evaluation metrics for medical applications of artificial intelligence," *Sci. Rep.*, vol. 12, no. 1, pp. 1–9, 2022, doi: 10.1038/s41598-022-09954-8.