JURNAL ILMIAH SISTEM INFORMASI (JUSI)
VOL 5, No. 3, November 2026, pp. 47 - 61
p-ISSN : 2803-1507 (media cetak) | e-ISSN : 2803-1531 (media online)
Link Publikasi: https://ejurnal.provisi.ac.id/index.php/JUISI/issue/view/175

Classification of Social Assistance Recipient Eligibility in Heram Urban Village Using the Naive Bayes Method

Helson Matuan*[1], Heru Sutejo[2] , Esau Dude[3] .
[1,2,3] Sistem Informasi, Universitas Sepuluh Nopember Papua.
Jl. Ardipura Raya No. 22B, 99222, Ardipura, Distrik Jayapura Selatan, Kota Jayapura, Papua, Indonesia
e-mail: helsonmatuan94@gmail.com[1], heru.sutejo01@gmail.com [2] , esaudude19@gmail.com[3]

ARTICLE INFO
Article History:

Received: 03-05-2026
Revised: 12-07-2026
Accepted: 09-09-2026
Published: 29-10-2026

ABSTRACT
Determining the eligibility of social assistance recipients is an important stage in the government's effort to distribute aid accurately to people who truly need it. In practice, however, the selection process often encounters problems, particularly inaccurate targeting caused by manual assessment. This study aims to implement the Naive Bayes algorithm to determine the eligibility of social assistance recipients in Heram Village. The dataset used in this study consisted of 358 records, divided into 251 training data and 107 testing data. Before the modelling process, the data were cleaned and selected so that only relevant variables were used in the classification process. The attributes applied in this study included gender, age, occupation, number of dependents, dependent category, priority category, monthly income, income category, housing condition, asset ownership, and employment status, with actual eligibility as the class label. The stages of data processing, modelling, and evaluation were carried out using RapidMiner software. The results showed that the Naive Bayes model achieved an accuracy of 85.05%, precision of 83.33%, recall of 86.54%, and AUC of 0.924, which is categorized as excellent classification. These findings indicate that the Naive Bayes algorithm is effective and feasible to be used as a decision-support method in determining social assistance recipient eligibility in a more objective, systematic, and measurable manner in Heram Village.

Keywords: *Naive Bayes, social assistance, classification, recipient eligibility, RapidMiner.*

Abstrak

Penentuan kelayakan penerima bantuan sosial merupakan salah satu tahapan penting dalam upaya pemerintah menyalurkan bantuan secara tepat sasaran kepada masyarakat yang benar-benar membutuhkan. Namun, dalam pelaksanaannya, proses seleksi penerima bantuan sosial masih sering menghadapi kendala, seperti ketidaktepatan sasaran akibat penilaian yang dilakukan secara manual. Penelitian ini bertujuan untuk mengimplementasikan algoritma *Naive Bayes* dalam menentukan kelayakan penerima bantuan sosial di Kelurahan Heram. Data yang digunakan dalam penelitian ini berjumlah 358 dataset, yang terdiri atas 251 data latih dan 107 data uji. Sebelum dilakukan pemodelan, data terlebih dahulu melalui tahap pembersihan dan seleksi atribut agar hanya variabel yang relevan yang digunakan dalam proses klasifikasi. Atribut yang digunakan meliputi Jenis Kelamin, Usia, Pekerjaan, Jumlah Tanggungan, Kategori Tanggungan, Kategori Prioritas, Penghasilan per Bulan, Kategori Penghasilan, Kondisi Rumah, Kepemilikan Aset, dan Status Pekerjaan, dengan Kelayakan Aktual sebagai label kelas Proses pengolahan data, pemodelan, dan evaluasi dilakukan menggunakan perangkat lunak *RapidMiner*. Hasil penelitian menunjukkan bahwa model Naive Bayes memperoleh nilai akurasi sebesar 85,05%, *precision* sebesar 83,33%, *recall* sebesar 86,54%, serta AUC sebesar 0,924 yang termasuk dalam kategori *excellent classification*. Berdasarkan hasil tersebut, algoritma *Naive Bayes* dinilai efektif dan layak digunakan sebagai pendukung pengambilan keputusan dalam

DOI: https://doi.org/10.51903/h1g15q44

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

menentukan kelayakan penerima bantuan sosial secara lebih objektif, sistematis, dan terukur di Kelurahan Heram.

Kata Kunci: *Naive Bayes,* bantuan sosial, klasifikasi, kelayakan penerima*, RapidMiner.*

1. PENDAHULUAN

Kemiskinan masih menjadi salah satu permasalahan sosial yang dihadapi berbagai daerah di Indonesia, termasuk pada tingkat desa atau kelurahan [1]. Kondisi ini dipengaruhi oleh beberapa faktor, seperti terbatasnya lapangan pekerjaan, rendahnya tingkat pendapatan, besarnya jumlah tanggungan keluarga, serta kondisi tempat tinggal yang kurang memadai. Keadaan tersebut menyebabkan sebagian masyarakat mengalami kesulitan dalam memenuhi kebutuhan hidup sehari-hari, sehingga membutuhkan perhatian melalui program bantuan sosial dari pemerintah [2]. Bantuan sosial merupakan bentuk upaya pemerintah dalam membantu masyarakat yang kurang mampu agar dapat meringankan beban ekonomi mereka. Akan tetapi, dalam pelaksanaannya, proses penentuan calon penerima bantuan sosial masih sering menghadapi berbagai kendala. Salah satu masalah yang kerap muncul adalah penyaluran bantuan yang belum sepenuhnya tepat sasaran, sehingga masih ada masyarakat yang sebenarnya layak menerima bantuan tetapi belum terdata, sementara sebagian penerima lainnya justru kurang sesuai dengan kriteria yang telah ditetapkan. Kondisi ini menunjukkan bahwa proses seleksi penerima bantuan sosial perlu dilakukan secara lebih objektif, sistematis, dan berdasarkan data yang akurat [3].

Perkembangan teknologi informasi dapat dimanfaatkan untuk mendukung proses tersebut, salah satunya melalui penerapan teknik data mining, khususnya metode klasifikasi. Dengan metode ini, data masyarakat dapat diolah menjadi informasi yang membantu dalam menentukan kelayakan penerima bantuan sosial secara lebih terstruktur dan terukur dibandingkan dengan cara manual [4]. Dalam penelitian ini, proses pengolahan data dan pengujian model dilakukan menggunakan *RapidMiner* sebagai alat bantu analisis, karena perangkat lunak ini mampu mempermudah proses pengolahan dataset, penerapan algoritma, serta evaluasi hasil klasifikasi [5]. Dalam penelitian ini, metode yang digunakan adalah algoritma *Naive Bayes* untuk menentukan kelayakan penerima bantuan sosial. Algoritma ini dipilih karena memiliki konsep yang sederhana, mudah diterapkan, serta mampu melakukan klasifikasi berdasarkan nilai probabilitas dari data yang tersedia [6]. Selain itu, *Naive Bayes* juga dikenal efisien dalam proses pengolahan data dan dapat memberikan hasil prediksi dengan cepat, sehingga sesuai digunakan untuk mendukung pengambilan keputusan dalam penentuan penerima bantuan sosial.

Penelitian ini dilakukan pada Kelurahan Heram dengan menggunakan sejumlah atribut yang dianggap mewakili kondisi sosial dan ekonomi masyarakat, yaitu jenis kelamin, usia, pekerjaan, jumlah tanggungan, kategori tanggungan, kategori prioritas, penghasilan per bulan, kategori penghasilan, kondisi rumah, kepemilikan aset, status pekerjaan, dan kelayakan aktual. Data yang digunakan berjumlah 358 dataset, yang dibagi menjadi 251 data latih dan 107 data uji. Data latih digunakan untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengukur kemampuan model dalam memprediksi kelayakan penerima bantuan sosial dengan bantuan *RapidMiner.*

Melalui implementasi algoritma *Naive Bayes*, penelitian ini diharapkan mampu memberikan solusi dalam proses penentuan penerima bantuan sosial di Kelurahan Heram. Sistem yang dihasilkan diharapkan dapat membantu pihak terkait dalam mengambil keputusan secara lebih cepat, objektif, dan tepat sasaran, sehingga bantuan sosial benar-benar disalurkan kepada masyarakat yang membutuhkan. Dengan demikian, hasil penelitian ini diharapkan dapat menjadi acuan dalam mendukung proses seleksi penerima bantuan sosial yang lebih efektif, akurat, dan sesuai dengan kondisi nyata di lapangan.

2. TINJAUAN PUSTAKA

2.1. Kajian Penelitian tentang Penyaluran Bantuan Sosial

Penelitian terdahulu mengenai penyaluran bantuan sosial menunjukkan bahwa proses pemberian bantuan kepada masyarakat kurang mampu masih sering menghadapi permasalahan ketidaktepatan sasaran. Bantuan sosial seharusnya diberikan kepada masyarakat yang benar-benar membutuhkan, namun dalam praktiknya masih ditemukan penerima yang kurang sesuai dengan kriteria, sementara masyarakat yang layak justru belum memperoleh bantuan. Kondisi tersebut menunjukkan bahwa proses

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

seleksi penerima bantuan sosial perlu dilakukan secara lebih objektif dan berbasis data agar hasil penyaluran bantuan menjadi lebih tepat [7]. Penelitian sebelumnya juga menjelaskan bahwa data sosial ekonomi masyarakat dapat digunakan sebagai dasar dalam mendukung proses pengambilan keputusan. Data tersebut dapat membantu pihak terkait dalam menilai kondisi calon penerima bantuan secara lebih sistematis, sehingga proses seleksi tidak hanya bergantung pada penilaian manual. Berdasarkan kajian tersebut, penelitian ini menggunakan data masyarakat Kelurahan Heram sebagai dasar untuk membangun model klasifikasi kelayakan penerima bantuan sosial menggunakan algoritma *Naive Bayes*.

2.2. Analisis Indikator Kelayakan Penerima Bantuan Sosial

Penelitian terdahulu tentang kelayakan penerima bantuan sosial menunjukkan bahwa penentuan penerima bantuan umumnya didasarkan pada beberapa indikator sosial dan ekonomi. Indikator yang sering digunakan meliputi usia, pekerjaan, jumlah tanggungan, tingkat pendapatan, kondisi tempat tinggal, kepemilikan aset, serta status pekerjaan. Atribut-atribut tersebut dinilai penting karena dapat menggambarkan kondisi ekonomi dan kebutuhan masyarakat yang menjadi calon penerima bantuan sosial [8]. Beberapa penelitian juga menunjukkan bahwa kesalahan dalam menentukan kelayakan penerima dapat berdampak pada tidak tepatnya sasaran bantuan.

Kesalahan tersebut dapat terjadi ketika masyarakat yang sebenarnya layak tidak terpilih sebagai penerima, atau sebaliknya masyarakat yang tidak layak justru ditetapkan sebagai penerima bantuan. Oleh karena itu, penelitian terdahulu menekankan perlunya sistem penilaian yang lebih terarah dan objektif agar proses seleksi penerima bantuan sosial menjadi lebih akurat [9]. Penelitian ini mengacu pada pendekatan tersebut dengan menggunakan atribut sosial ekonomi masyarakat Kelurahan Heram sebagai dasar klasifikasi.

2.3. Pemanfaatan Data *Mining* dalam Klasifikasi Data Sosial

Penelitian terdahulu menunjukkan bahwa data *mining* banyak digunakan untuk membantu proses pengambilan keputusan pada data sosial. Salah satu teknik yang sering digunakan adalah klasifikasi, yaitu proses pengelompokan data ke dalam kelas tertentu berdasarkan pola atribut yang dimiliki. Dalam konteks bantuan sosial, klasifikasi digunakan untuk membedakan masyarakat yang termasuk kategori layak dan tidak layak menerima bantuan [10]. Penerapan data mining pada penelitian sebelumnya dinilai mampu membantu pengolahan data masyarakat secara lebih cepat, terstruktur, dan konsisten dibandingkan proses manual. Dengan pendekatan ini, data yang sebelumnya hanya berupa catatan administrasi dapat diolah menjadi informasi yang bermanfaat untuk mendukung pengambilan keputusan [11]. Berdasarkan penelitian tersebut, penelitian ini menggunakan pendekatan data mining untuk membangun model klasifikasi kelayakan penerima bantuan sosial di Kelurahan Heram.

2.4. Penerapan Algoritma *Naive Bayes* pada Penelitian Terdahulu

Beberapa penelitian terdahulu telah menerapkan algoritma *Naive Bayes* dalam kasus klasifikasi, termasuk pada penentuan kelayakan penerima bantuan sosial. *Naive Bayes* digunakan karena mampu menghitung peluang suatu data masuk ke dalam kelas tertentu berdasarkan atribut yang dimiliki. Dalam penelitian sebelumnya, algoritma ini dinilai cukup baik untuk mengolah data terstruktur dan menghasilkan keputusan klasifikasi secara sederhana serta efisien [12]. Penelitian lain juga menunjukkan bahwa *Naive Bayes* dapat digunakan untuk mengolah atribut sosial dan ekonomi dalam menentukan kelayakan penerima bantuan. Metode ini mampu menghasilkan prediksi berdasarkan probabilitas dari setiap atribut terhadap kelas yang telah ditentukan, seperti Layak dan Tidak Layak [13]. Oleh karena itu, penelitian ini menggunakan algoritma *Naive Bayes* untuk mengklasifikasikan kelayakan penerima bantuan sosial di Kelurahan Heram berdasarkan data yang tersedia.

2.5. Penggunaan *RapidMiner* dalam Pengolahan Data Penelitian

Penelitian terdahulu yang menggunakan pendekatan data mining banyak memanfaatkan perangkat lunak *RapidMiner* sebagai alat bantu analisis. *RapidMiner* digunakan karena menyediakan fasilitas untuk mengelola dataset, memilih atribut, menerapkan algoritma klasifikasi, serta mengevaluasi hasil model yang dibangun [14]. Dengan bantuan perangkat lunak tersebut, proses penelitian dapat dilakukan secara lebih sistematis mulai dari tahap input data hingga evaluasi performa model. Pada penelitian sebelumnya, *RapidMiner* juga digunakan untuk mempermudah proses pembagian data latih dan data

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

uji, penerapan algoritma, serta perhitungan nilai evaluasi seperti *accuracy, precision, recall*, dan AUC [15]. Berdasarkan hal tersebut, penelitian ini menggunakan *RapidMiner* untuk membantu proses pemodelan *Naive Bayes* pada data kelayakan penerima bantuan sosial di Kelurahan Heram. Penggunaan *RapidMiner* diharapkan dapat membuat proses pengujian model menjadi lebih terukur dan mudah dianalisis.

2.6. Atribut Penentu Kelayakan Penerima Bantuan Sosial

Penilaian kelayakan penerima bantuan sosial memerlukan atribut yang dapat mencerminkan keadaan sosial dan ekonomi masyarakat secara nyata [16]. Pada penelitian ini, atribut yang digunakan meliputi jenis kelamin, usia, pekerjaan, jumlah tanggungan, kategori tanggungan, kategori prioritas, penghasilan per bulan, kategori penghasilan, kondisi rumah, kepemilikan aset, status pekerjaan, serta kelayakan aktual. Pemilihan atribut tersebut didasarkan pada pertimbangan bahwa seluruhnya dapat mewakili kondisi masyarakat di Kelurahan Heram. Melalui atribut-atribut ini, model klasifikasi dapat mempelajari pola dari 358 dataset, yang terdiri atas 251 data latih dan 107 data uji, sehingga hasil prediksi mengenai kelayakan penerima bantuan sosial diharapkan menjadi lebih akurat dan sesuai dengan kondisi lapangan.

2.7. *Research Gap*

Berdasarkan kajian pustaka, penelitian mengenai klasifikasi kelayakan penerima bantuan sosial dengan algoritma *Naive Bayes* telah dilakukan dalam berbagai konteks. Penelitian Luthfy Budhy Adzy membahas klasifikasi kelayakan penerima Bantuan Iuran Jaminan Kesehatan Pemerintah Daerah Kabupaten Sukabumi menggunakan algoritma *Naive Bayes* dengan pendekatan KDD untuk membantu proses seleksi calon penerima secara lebih objektif [17]. Penelitian Nur Madia dkk. juga menerapkan metode *Naive Bayes* untuk menentukan kelayakan masyarakat miskin penerima bantuan di Kabupaten Penajam Paser Utara [18]. Selain itu, penelitian mengenai Program Keluarga Harapan (PKH) menunjukkan bahwa metode Naive Bayes dapat digunakan untuk membantu penentuan penerima bantuan yang berhak, sedangkan pada penelitian tentang Bantuan Pangan Non Tunai (BPNT), algoritma yang sama juga dimanfaatkan untuk mengklasifikasikan kelayakan penerima bantuan [19]. Dari beberapa penelitian tersebut dapat dilihat bahwa *Naive Bayes* telah banyak digunakan untuk menyelesaikan permasalahan klasifikasi penerima bantuan sosial pada berbagai jenis program dan wilayah penelitian.

Adapun penelitian ini memiliki perbedaan dan kebaruan pada objek, atribut, serta implementasi model yang digunakan. Penelitian ini berfokus pada penentuan kelayakan menerima bantuan sosial di Kelurahan Heram, yang belum banyak dibahas secara khusus pada penelitian sebelumnya. Selain itu, penelitian ini menggunakan kombinasi atribut sosial dan ekonomi yang meliputi jenis kelamin, usia, pekerjaan, jumlah tanggungan, kategori tanggungan, kategori prioritas, penghasilan per bulan, kategori penghasilan, kondisi rumah, kepemilikan aset, dan status pekerjaan untuk menggambarkan kondisi nyata masyarakat setempat. Proses pengolahan data dilakukan menggunakan *RapidMiner* terhadap 358 dataset yang dibagi menjadi 251 data latih dan 107 data uji. Dengan demikian, penelitian ini diarahkan untuk memberikan model klasifikasi yang lebih kontekstual, terukur, dan sesuai dengan kondisi masyarakat di Kelurahan Heram, sehingga dapat membantu proses penentuan penerima bantuan sosial secara lebih tepat sasaran.

3. METODOLOGI PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah metode klasifikasi dengan algoritma *Naive Bayes*. Alur penelitian ditunjukkan dalam Gambar 1, yang memuat tahapan pengumpulan data, pembersihan dan seleksi atribut, pemisahan data latih dan data uji, pelatihan model, serta evaluasi hasil. Metode ini dipilih karena mampu mengelompokkan data berdasarkan probabilitas setiap atribut terhadap kelas Layak dan Tidak Layak, sehingga dapat digunakan sebagai alat bantu awal dalam proses penentuan kelayakan penerima bantuan sosial secara lebih objektif dan terukur.

3.1. Pengumpulan Data

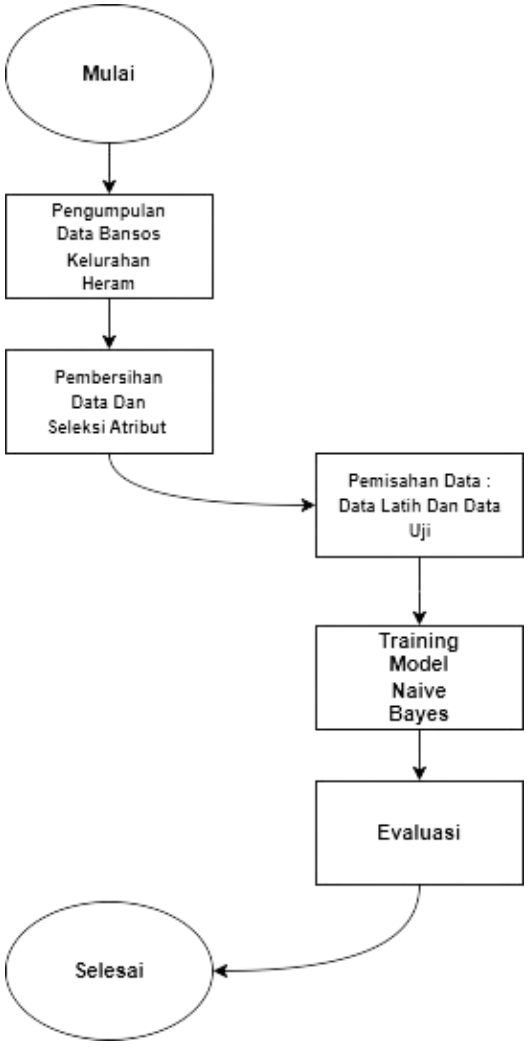
Proses penelitian diawali dengan tahap pengumpulan data sebagaimana ditunjukkan pada alur penelitian. Pada tahap ini, data bantuan sosial masyarakat Kelurahan Heram dikumpulkan sebagai bahan

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

utama dalam proses klasifikasi kelayakan penerima bantuan sosial. Data yang dihimpun merupakan data awal atau data mentah yang masih memuat seluruh atribut hasil pendataan lapangan. Atribut pada data awal tersebut terdiri atas No, ID_KK, NIK, Nama Penerima, Jenis Kelamin, Usia, Pekerjaan, Jumlah Tanggungan, Kategori Tanggungan, Kategori Prioritas, Penghasilan per Bulan, Kategori Penghasilan, Kondisi Rumah, Kepemilikan Aset, dan Status Pekerjaan. Seluruh atribut tersebut dikumpulkan untuk menggambarkan kondisi sosial dan ekonomi masyarakat yang menjadi objek penelitian. Jumlah data yang digunakan dalam penelitian ini sebanyak 358 dataset, yang selanjutnya diproses pada tahap berikutnya menggunakan perangkat lunak *RapidMiner* untuk mendukung implementasi algoritma *Naive Bayes* [20], [21].

3.2. Pembersihan Data dan Seleksi Atribut

Setelah tahap pengumpulan data selesai dilakukan, langkah berikutnya adalah pembersihan data dan seleksi atribut. Tahap ini bertujuan untuk memastikan bahwa data yang digunakan dalam penelitian berada dalam kondisi baik, konsisten, dan sesuai untuk proses klasifikasi. Pada proses pembersihan data, dilakukan pemeriksaan terhadap data yang tidak lengkap, data ganda, serta ketidaksesuaian penulisan pada setiap atribut[22], [23]. Selain itu, data juga disesuaikan agar formatnya seragam sehingga memudahkan proses pengolahan pada tahap berikutnya.



Gambar 1. Alur Penelitian Menggunakan *Naive Bayes*

Pada tahap ini, atribut yang bersifat identitas seperti No, ID_KK, NIK, dan Nama Penerima tidak digunakan dalam proses klasifikasi, karena atribut tersebut tidak memiliki pengaruh langsung terhadap penentuan kelayakan penerima bantuan sosial. Setelah proses seleksi atribut dilakukan, data yang

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

digunakan dalam penelitian ini meliputi Jenis Kelamin, Usia, Pekerjaan, Jumlah Tanggungan, Kategori Tanggungan, Kategori Prioritas, Penghasilan per Bulan, Kategori Penghasilan, Kondisi Rumah, Kepemilikan Aset, dan Status Pekerjaan, dengan Kelayakan Aktual sebagai label kelas. Hasil dari tahap ini adalah dataset yang telah bersih dan siap digunakan dalam proses pemodelan.

3.3. Pemisahan Data Latih dan Data Uji

Setelah data dibersihkan dan atribut yang digunakan telah ditentukan, tahap berikutnya adalah pemisahan data menjadi data latih dan data uji. Pemisahan data dilakukan untuk membangun model klasifikasi sekaligus menguji kemampuan model dalam memprediksi data yang belum pernah dipelajari sebelumnya [24]. Dalam penelitian ini, jumlah keseluruhan data yang digunakan adalah 358 dataset, yang kemudian dibagi menjadi 251 data latih dan 107 data uji. Data latih digunakan untuk membentuk model klasifikasi berdasarkan pola hubungan antara atribut-atribut yang digunakan dengan label kelas. Sementara itu, data uji digunakan untuk mengukur kemampuan model dalam memprediksi kelayakan penerima bantuan sosial pada data baru [25]. Pembagian data ini dilakukan agar hasil penelitian dapat menunjukkan performa algoritma *Naive Bayes* secara lebih objektif dalam menyelesaikan permasalahan klasifikasi kelayakan penerima bantuan sosial di Kelurahan Heram.

3.4. Penerapan Metode *Naive Bayes*

Metode yang digunakan dalam penelitian ini adalah algoritma *Naive Bayes*. *Naive Bayes* merupakan metode klasifikasi yang didasarkan pada teori probabilitas dan teorema Bayes. Metode ini digunakan untuk menentukan kelas suatu data berdasarkan peluang kemunculan atribut-atribut yang dimiliki terhadap kelas tertentu [26]. Dalam penelitian ini, kelas yang digunakan adalah Layak dan Tidak Layak menerima bantuan sosial. Setiap data masyarakat direpresentasikan ke dalam atribut-atribut yang digunakan dalam penelitian, yaitu Jenis Kelamin, Usia, Pekerjaan, Jumlah Tanggungan, Kategori Tanggungan, Kategori Prioritas, Penghasilan per Bulan, Kategori Penghasilan, Kondisi Rumah, Kepemilikan Aset, dan Status Pekerjaan.

Tabel 1. Distribusi Kelas pada Data Latih dan Data Uji

Jenis Data	Layak	Tidak Layak	Jumlah
Data Latih	129	122	251
Data Uji	55	52	107
Total	184	174	358

Berdasarkan atribut-atribut tersebut, algoritma *Naive Bayes* menghitung probabilitas setiap data termasuk ke dalam kelas Layak atau Tidak Layak. Kelas dengan nilai probabilitas terbesar akan menjadi hasil akhir dari proses klasifikasi. Secara umum, perhitungan *Naive Bayes* dinyatakan dengan rumus sebagai berikut[27][28], [29]:

$$P(C \mid X) = \frac{P(X \mid C) \cdot P(C)}{P(X)} \qquad (1)$$

Di mana, $P(C \mid X)$ merupakan probabilitas suatu data X termasuk ke dalam kelas C, $P(X \mid C)$merupakan probabilitas kemunculan atribut atau karakteristik data X pada kelas C, $P(C)$ merupakan probabilitas awal (prior probability) dari kelas C, sedangkan $P(X)$merupakan probabilitas munculnya data X secara keseluruhan. Untuk mempermudah penghitungan manual, metode *Naive Bayes* dalam penelitian ini dijabarkan ke dalam beberapa rumus. Probabilitas prior adalah peluang awal dari masing-masing kelas berdasarkan jumlah data pada tiap kelas.

$$P(C) = \frac{N_C}{N} \qquad (2)$$

dimana $P(C)$ merupakan probabilitas prior kelas, N_C adalah jumlah data pada kelas tertent, dan N merupakan jumlah seluruh data latih. Probabilitas likelihood digunakan untuk menghitung peluang kemunculan suatu atribut tertentu pada kelas tertentu.

$$P(X_i \mid C) = \frac{N_{X_i,C}}{N_C} \qquad (3)$$

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

Dimana, $P(X_i \mid C)$ adalah probabilitas atribut X_i pada kelas C. $N_{X_i,C}$ merupakan jumlah kemunculan atribut X_i pada kelas C, sedangkan N_C merupakan jumlah data pada kelas C. Untuk menentukan peluang suatu data termasuk ke dalam kelas Layak, digunakan perhitungan sebagai berikut:

$$P(\text{Layak} \mid X) \propto P(\text{Layak}) \times \prod_{i=1}^{n} P(X_i \mid \text{Layak}) \quad (4)$$

Di mana, P(Layak|X) adalah probabilitas data X termasuk kelas Layak, sedangkan P(Layak) merupakan probabilitas awal kelas Layak. P(x_i|Layak) merupakan probabilitas atribut ke-i pada kelas Layak, dan n merupakan jumlah atribut. Untuk menentukan peluang suatu data termasuk ke dalam kelas Tidak Layak, digunakan rumus berikut:

$$P(\text{Tidak Layak} \mid X) \propto P(\text{Tidak Layak}) \times \prod_{i=1}^{n} P(X_i \mid \text{Tidak Layak}) \quad (5)$$

Di mana, P(Tidak Layak|X) merupakan probabilitas data X termasuk ke dalam kelas Tidak Layak, sedangkan P(Tidak Layak) merupakan probabilitas awal kelas Tidak Layak. Selanjutnya, P(x_i|Tidak Layak) merupakan probabilitas atribut ke-i pada kelas Tidak Layak, dan n merupakan jumlah atribut. Setelah nilai probabilitas untuk kedua kelas diperoleh, hasil klasifikasi ditentukan dengan memilih kelas yang memiliki nilai probabilitas paling besar.

$$\hat{Y} = \underset{C \in \{\text{Layak, Tidak Layak}\}}{\arg\max} P(C) \prod_{i=1}^{n} P(X_i \mid C) \quad (6)$$

Apabila nilai probabilitas kelas Layak lebih besar daripada kelas Tidak Layak, maka data diklasifikasikan sebagai Layak menerima bantuan sosial. Sebaliknya, jika probabilitas kelas Tidak Layak lebih besar, maka data diklasifikasikan sebagai Tidak Layak. Dalam penerapannya, proses klasifikasi dilakukan menggunakan 251 data latih yang telah dipisahkan sebelumnya. Data tersebut diolah menggunakan perangkat lunak *RapidMiner* dengan menerapkan algoritma *Naive Bayes*. Proses ini bertujuan untuk mempelajari pola hubungan antara atribut-atribut yang digunakan dengan hasil klasifikasi kelayakan penerima bantuan sosial. Melalui proses tersebut, sistem membentuk model klasifikasi berdasarkan data historis yang tersedia.

Model yang dihasilkan dari proses pelatihan akan mengenali pola tertentu yang membedakan masyarakat yang termasuk kategori Layak dan Tidak Layak menerima bantuan sosial. Setelah model terbentuk, model tersebut digunakan untuk melakukan proses klasifikasi pada data uji. Dengan demikian, penerapan metode *Naive Bayes* dalam penelitian ini diharapkan mampu membantu proses penentuan kelayakan penerima bantuan sosial secara lebih sistematis, objektif, dan terukur.

Tabel 2. Data Set

No	ID_KK	NIK	.....	Pekerjaan	Penghasilan per Bulan	Kategori Penghasilan	Kondisi Rumah	Kelayakan Aktual
1	KK-HDM-0001	9,17106E+15	.....	Tenaga Kebersihan	Rp980.325	Rendah	Sederhana	Layak
2	KK-HDM-0002	9,17106E+15	....	Buruh Harian Lepas	Rp1.296.561	Sedang	Sangat Sederhana	Layak
3	KK-HDM-0003	9,17105E+15	.....	Ojek	Rp1.489.115	Sedang	Sederhana	Layak
4	KK-HDM-0004	9,17105E+15	....	Tenaga Kebersihan	Rp1.294.991	Sedang	Sederhana	Tidak Layak
5	KK-HDM-0005	9,17105E+15	.....	Tenaga Kebersihan	Rp1.289.714	Sedang	Sederhana	Layak
6	KK-HDM-0006	9,17105E+15	.....	Pedagang Kecil	Rp1.658.598	Tinggi	Sederhana	Layak
7	KK-HDM-0007	9,17105E+15	.....	Petani	Rp785.945	Rendah	Sederhana	Layak
8	KK-HDM-0008	9,17105E+15	.....	Penjaga Kios	Rp1.527.631	Tinggi	Sederhana	Tidak Layak
9	KK-HDM-0009	9,17106E+15	.....	Lansia/Tidak Bekerja	Rp400.734	Rendah	Sangat Sederhana	Layak
10	KK-HDM-0010	9,17105E+15	.....	Ibu Rumah Tangga	Rp456.506	Rendah	Sangat Sederhana	Layak

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

18	KK-HDM-0018	9,17105E+15	.....	Nelayan	Rp828.113	Rendah	Layak	Tidak Layak
...	....	....	....	....	....	....	...	....

4. PEMBAHASAN

4.1. Pengumpulan Data

Pada tahap pengumpulan data, peneliti mengumpulkan data bantuan sosial masyarakat Kelurahan Heram yang akan digunakan sebagai bahan utama dalam proses klasifikasi kelayakan penerima bantuan sosial. Data yang diperoleh merupakan data awal atau data mentah yang masih memuat atribut identitas dan atribut penilaian. Berdasarkan data yang ditampilkan, atribut yang tersedia meliputi No, ID_KK, NIK, Nama Penerima, Jenis Kelamin, Usia, Pekerjaan, Jumlah Tanggungan, Penghasilan per Bulan, Kategori Penghasilan, Kondisi Rumah, Kepemilikan Aset, Status Pekerjaan, dan Kelayakan Aktual. Data awal ini digunakan untuk menggambarkan kondisi sosial dan ekonomi masyarakat yang menjadi objek penelitian. Setiap baris data merepresentasikan satu penerima atau calon penerima bantuan sosial dengan karakteristik yang berbeda-beda. Sebagai contoh, terdapat data masyarakat dengan pekerjaan Tenaga Kebersihan, Buruh Harian Lepas, Ojek, Pedagang Kecil, Penjaga Kios, Nelayan, Petani, Ibu Rumah Tangga, dan Lansia/Tidak Bekerja. Selain itu, data juga memuat informasi tentang usia, jumlah tanggungan, pendapatan bulanan, kondisi rumah, kepemilikan aset, dan status pekerjaan yang menjadi faktor pendukung dalam penentuan kelayakan menerima bantuan sosial.

Berdasarkan dataset yang telah dikumpulkan, jumlah keseluruhan data dalam penelitian ini adalah 358 data. Data tersebut kemudian digunakan sebagai sumber utama untuk proses selanjutnya, yaitu pembersihan data, seleksi atribut, pembagian data latih dan data uji, serta penerapan algoritma *Naive Bayes*. Pada tahap ini, data yang dikumpulkan masih dalam bentuk data mentah sehingga belum seluruh atribut langsung digunakan dalam proses klasifikasi. Tahap pengumpulan data menjadi bagian yang sangat penting dalam penelitian ini karena kualitas data yang diperoleh akan memengaruhi hasil klasifikasi yang dilakukan. Semakin lengkap dan sesuai data yang dikumpulkan, maka semakin baik pula model yang dibangun dalam menentukan apakah seseorang termasuk kategori Layak atau Tidak Layak menerima bantuan sosial. Oleh karena itu, data bantuan sosial dari Kelurahan Heram dijadikan dasar dalam pembangunan model klasifikasi pada penelitian ini.

4.2. Pembersihan Data

Setelah tahap pengumpulan data selesai dilakukan, data kemudian melalui proses pembersihan agar sesuai dengan kebutuhan penelitian. Pada tahap ini, atribut identitas seperti No, ID_KK, NIK, dan Nama Penerima dikeluarkan dari proses klasifikasi. Pemeriksaan data dilakukan terhadap kemungkinan *missing value*, duplikasi, perbedaan penulisan kategori, dan format nilai numerik. Data yang tidak lengkap pada atribut penting tidak digunakan, data ganda diperiksa berdasarkan identitas, dan penulisan kategori seperti kondisi rumah, kepemilikan aset, status pekerjaan, kategori penghasilan, serta kategori tanggungan diseragamkan. Hasil data yang telah siap diuji dapat dilihat pada Gambar 2. Berdasarkan hasil pembersihan data, jumlah dataset yang digunakan tetap sebanyak 358 data.

Atribut final yang masuk ke dalam model adalah Jenis Kelamin, Usia, Pekerjaan, Jumlah Tanggungan, Kategori Tanggungan, Kategori Prioritas, Penghasilan per Bulan, Kategori Penghasilan, Kondisi Rumah, Kepemilikan Aset, dan Status Pekerjaan, dengan Kelayakan Aktual sebagai label kelas. Dataset tersebut dibagi menjadi 251 data latih dan 107 data uji. Distribusi kelas Layak dan Tidak Layak pada masing-masing bagian data telah dijelaskan pada Tabel 2 agar pembaca dapat menilai keseimbangan data yang digunakan dalam pengujian model.

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

Data yang telah dibersihkan selanjutnya digunakan dalam perhitungan probabilitas Naive Bayes. Setiap atribut dihitung peluang kemunculannya terhadap kelas Layak dan Tidak Layak berdasarkan data latih, kemudian pola tersebut diterapkan pada data uji. Tahap pembersihan dan seleksi atribut menjadi penting karena kesalahan format, duplikasi, atau kategori yang tidak konsisten dapat menyebabkan probabilitas yang dihitung menjadi kurang tepat dan berdampak pada hasil klasifikasi.

Row No.	Pekerjaan	Kelayakan A...	Jumlah Tan...	Kategori Ta...	Kategori Pri...	Penghasilan...	Kategori Pe...	Kondisi Rum...	Kepemilikan...	Status Peke...
1	Buruh Harian...	Layak	4	Sedang	Keluarga Re...	1245201	Sedang	Sangat Seder...	Tidak Ada	Informal
2	Petani	Layak	4	Sedang	Keluarga Re...	1193991	Sedang	Sederhana	Tidak Ada	Informal
3	Tenaga Kebe...	Layak	3	Sedang	Disabilitas Ri...	1289714	Sedang	Sederhana	Sedikit	Informal
4	Lansia/Tidak ...	Layak	2	Rendah	Kepala Keluа...	331908	Rendah	Sederhana	Sedikit	Tidak Tetap
5	Lansia/Tidak ...	Layak	3	Sedang	Lansia	546381	Rendah	Sangat Seder...	Tidak Ada	Tidak Tetap
6	Ibu Rumah T...	Layak	4	Sedang	Kepala Keluа...	895775	Rendah	Sederhana	Sedikit	Tidak Tetap
7	Petani	Layak	6	Tinggi	Keluarga Re...	1109186	Sedang	Sederhana	Tidak Ada	Informal
8	Tidak/Belum ...	Layak	6	Tinggi	Keluarga Re...	779260	Rendah	Sangat Seder...	Tidak Ada	Tidak Tetap
9	Tidak/Belum ...	Layak	3	Sedang	Janda/Duda	480636	Rendah	Sederhana	Tidak Ada	Tidak Tetap
10	Penjaga Kios	Tidak Layak	2	Rendah	Keluarga Re...	1311676	Sedang	Sederhana	Sedikit	Informal
11	Nelayan	Layak	5	Tinggi	Janda/Duda	1464999	Sedang	Sederhana	Sedikit	Informal
12	Nelayan	Tidak Layak	2	Rendah	Kepala Keluа...	1005956	Sedang	Layak	Sedikit	Informal
13	Lansia/Tidak ...	Layak	5	Tinggi	Janda/Duda	464930	Rendah	Sangat Seder...	Tidak Ada	Tidak Tetap
14	Penjaga Kios	Layak	5	Tinggi	Lansia	1044619	Sedang	Sederhana	Sedikit	Informal
15	Ibu Rumah T...	Layak	6	Tinggi	Janda/Duda	884559	Rendah	Sangat Seder...	Sedikit	Tidak Tetap
16	Nelayan	Layak	5	Tinggi	Keluarga Re...	963525	Rendah	Sederhana	Tidak Ada	Informal
17	Petani	Tidak Layak	1	Rendah	Lansia	1030196	Sedang	Sederhana	Sedikit	Informal
18	Lansia/Tidak ...	Layak	6	Tinggi	Disabilitas Ri...	354287	Rendah	Sangat Seder...	Tidak Ada	Tidak Tetap
19	Ibu Rumah T...	Layak	6	Tinggi	Keluarga Re...	456506	Rendah	Sangat Seder...	Tidak Ada	Tidak Tetap
20	Penjaga Kios	Tidak Layak	6	Tinggi	Kepala Keluа...	1672463	Tinggi	Sederhana	Sedikit	Informal

Gambar 2. Data Set Yang Siap Di Uji

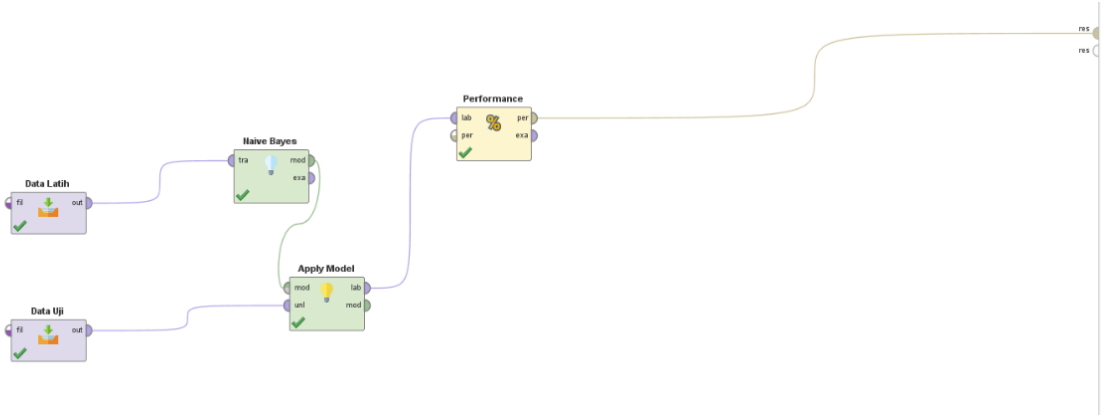
4.3. Pemodelan Data Menggunakan Alat *Rapidminer*

Pada tahap pemodelan, proses klasifikasi dilakukan menggunakan perangkat lunak RapidMiner. Rangkaian operator yang digunakan ditampilkan pada Gambar 3, yaitu Data Latih, *Naive Bayes, Apply Model, Retrieve* Data Uji, dan *Performance.* Data latih sebanyak 251 data dimasukkan ke operator Naive Bayes untuk membentuk model klasifikasi. Pada tahap ini, algoritma mempelajari pola hubungan antara atribut final dengan label kelas Kelayakan Aktual yang terdiri dari kategori Layak dan Tidak Layak. Setelah model terbentuk, tahap berikutnya adalah penerapan model menggunakan operator *Apply Model*.

Pada proses ini, model yang dihasilkan dari data latih diterapkan pada 107 data uji yang sebelumnya telah dipisahkan. Tujuan dari tahap ini adalah untuk mengetahui kemampuan model *Naive Bayes* dalam mengklasifikasikan data yang belum pernah dipelajari sebelumnya. Setiap data uji akan diproses berdasarkan probabilitas masing-masing atribut terhadap kelas Layak dan Tidak Layak, kemudian sistem akan memberikan hasil prediksi sesuai dengan kelas yang memiliki nilai probabilitas terbesar. Dengan demikian, proses pemodelan tidak hanya membentuk model, tetapi juga langsung menguji penerapannya pada data uji.

Tahap terakhir dalam pemodelan adalah evaluasi menggunakan operator Performance sebagaimana terlihat pada Gambar 3. Pada tahap ini, hasil prediksi model dibandingkan dengan label aktual pada data uji untuk mengetahui tingkat kinerja model. Evaluasi dilakukan menggunakan *confusion matrix, accuracy, precision*, *recall*, serta AUC agar hasil model dapat dinilai dari sisi ketepatan prediksi dan kemampuan membedakan kelas.

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



Gambar 3. Model Pengujian

4.4. Evaluasi

Tahap evaluasi dilakukan untuk mengetahui kinerja model *Naive Bayes* berdasarkan 107 data uji. Hasil *confusion matrix* ditampilkan pada Gambar 4. Berdasarkan gambar tersebut, model menghasilkan 45 data Tidak Layak yang diprediksi benar sebagai Tidak Layak dan 46 data Layak yang diprediksi benar sebagai Layak. Sementara itu, terdapat 9 data Layak yang diprediksi sebagai Tidak Layak dan 7 data Tidak Layak yang diprediksi sebagai Layak. Kesalahan klasifikasi pada Gambar 4 perlu dibahas secara kritis karena berhubungan langsung dengan konteks bantuan sosial. Sembilan data Layak yang diprediksi Tidak Layak berpotensi menyebabkan warga yang seharusnya menerima bantuan tidak terpilih apabila model digunakan tanpa verifikasi tambahan. Sebaliknya, tujuh data Tidak Layak yang diprediksi Layak berpotensi membuat bantuan diberikan kepada warga yang kurang sesuai dengan kriteria. Oleh karena itu, hasil model tidak boleh digunakan sebagai penentu akhir, melainkan sebagai alat bantu awal yang perlu dikombinasikan dengan verifikasi administrasi dan peninjauan lapangan.

Table View Plot View

accuracy: 85.05%

	true Tidak Layak	true Layak	class precision
pred. Tidak Layak	45	9	83.33%
pred. Layak	7	46	86.79%
class recall	86.54%	83.64%	

Gambar 4. *Confusion Matrix*

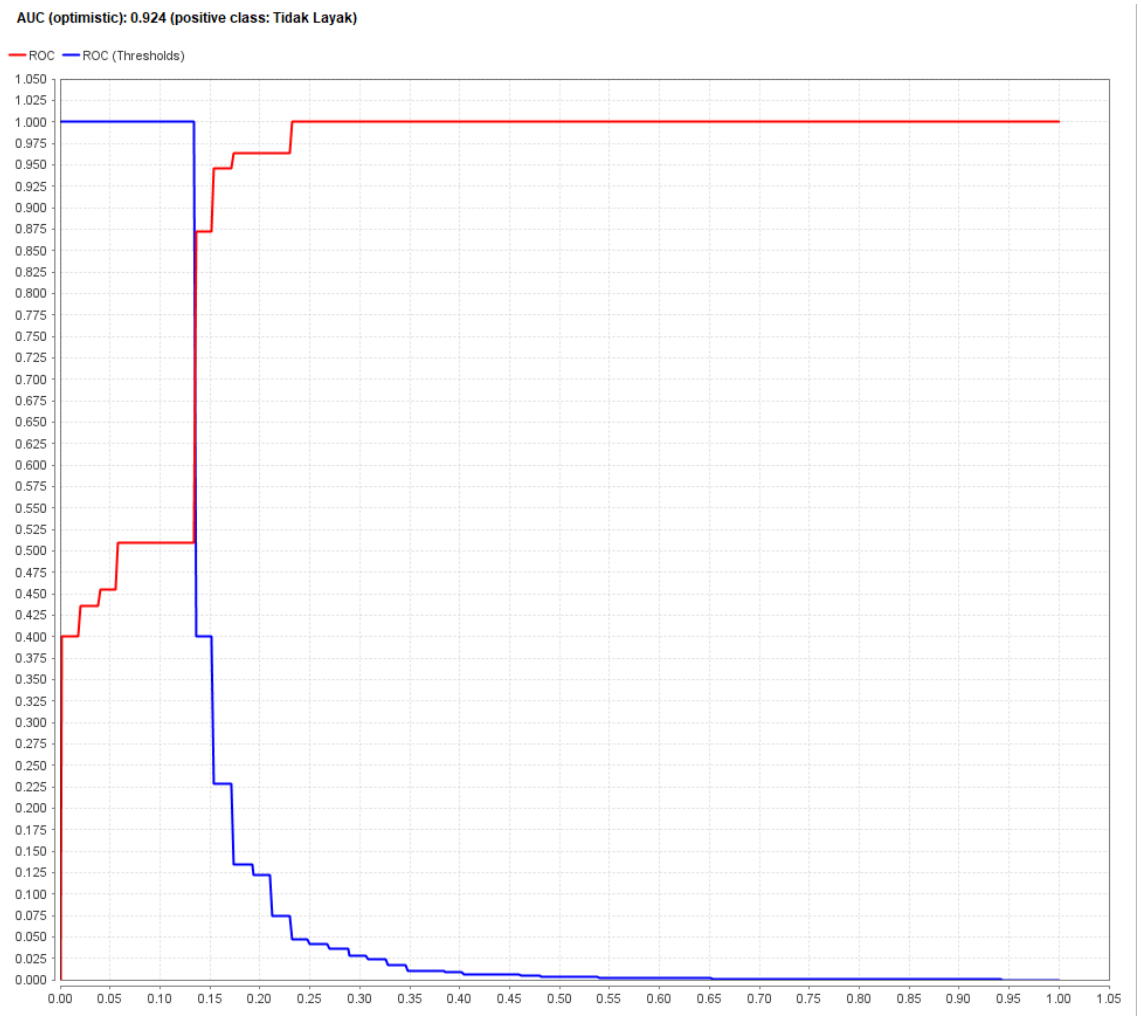
Secara kuantitatif, model memperoleh akurasi 85,05%. Pada positive class Tidak Layak, precision sebesar 83,33% menunjukkan bahwa sebagian besar data yang diprediksi Tidak Layak memang benar termasuk kelas tersebut, sedangkan recall sebesar 86,54% menunjukkan kemampuan model dalam menemukan data Tidak Layak dari seluruh data Tidak Layak yang ada. Nilai-nilai ini menunjukkan performa yang baik, tetapi interpretasinya tetap harus memperhatikan biaya sosial dari kesalahan klasifikasi.

Selain menggunakan *confusion matrix* dan nilai akurasi, evaluasi model dalam penelitian ini juga dilakukan melalui kurva ROC (*Receiver Operating Characteristic*) dan nilai AUC (*Area Under Curve*). Berdasarkan hasil pengujian menggunakan RapidMiner, diperoleh nilai AUC sebesar 0,924 dengan positive class yaitu Tidak Layak. Nilai tersebut menunjukkan bahwa model *Naive Bayes* memiliki kemampuan yang sangat baik dalam membedakan data yang termasuk ke dalam kelas Tidak Layak dan Layak. Semakin mendekati nilai 1,00, maka semakin baik kemampuan model dalam melakukan klasifikasi.

Oleh karena itu, nilai AUC sebesar 0,924 menunjukkan bahwa model yang dibangun memiliki tingkat performa klasifikasi yang tinggi. Selain *confusion matrix* pada Gambar 4, evaluasi juga dilakukan

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

menggunakan kurva ROC dan nilai AUC sebagaimana ditampilkan pada Gambar 5. Berdasarkan hasil pengujian menggunakan RapidMiner, diperoleh nilai AUC sebesar 0,924 dengan positive class Tidak Layak. Nilai tersebut menunjukkan bahwa model memiliki kemampuan diskriminasi kelas yang sangat baik karena mampu membedakan data Layak dan Tidak Layak pada berbagai ambang keputusan.



Gambar 5. Kurva ROC (*Receiver Operating Characteristic*) Dan Nilai AUC (*Area Under Curve*)

Kurva ROC pada Gambar 5 memperlihatkan bahwa garis model berada jauh di atas garis diagonal dan cenderung mendekati sudut kiri atas. Kondisi tersebut mengindikasikan bahwa model memiliki kemampuan yang baik dalam membedakan kedua kelas. Namun, nilai AUC yang tinggi tetap perlu dilengkapi dengan analisis *confusion matrix* karena AUC tidak secara langsung menunjukkan dampak kesalahan prediksi terhadap warga penerima bantuan. Berdasarkan nilai AUC sebesar 0,924, model *Naive Bayes* dalam penelitian ini dapat dikategorikan sebagai *excellent classification* atau kemampuan klasifikasi yang sangat baik. Hasil tersebut menunjukkan bahwa model tidak hanya memiliki tingkat akurasi yang tinggi, tetapi juga memiliki kemampuan diskriminasi kelas yang kuat. Dengan demikian, algoritma *Naive Bayes* yang diterapkan pada data bantuan sosial di Kelurahan Heram dinilai mampu mendukung proses penentuan kelayakan penerima bantuan sosial secara lebih objektif, terukur, dan tepat sasaran.

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

PerformanceVector

```
PerformanceVector:
accuracy: 85.05%
ConfusionMatrix:
True:   Tidak Layak     Layak
Tidak Layak:    45      9
Layak:  7       46
precision: 83.33% (positive class: Tidak Layak)
ConfusionMatrix:
True:   Layak   Tidak Layak
Layak:  46      7
Tidak Layak:    9       45
recall: 86.54% (positive class: Tidak Layak)
ConfusionMatrix:
True:   Layak   Tidak Layak
Layak:  46      7
Tidak Layak:    9       45
AUC (optimistic): 0.924 (positive class: Tidak Layak)
AUC: 0.924 (positive class: Tidak Layak)
AUC (pessimistic): 0.924 (positive class: Tidak Layak)
```

Gambar 6. *Performance Vector*

Gambar 6 menampilkan ringkasan *Performance Vector dari RapidMiner*, yang memuat akurasi, *confusion matrix, precision, recall*, dan AUC. Ringkasan tersebut memperkuat hasil evaluasi sebelumnya bahwa model mampu menghasilkan jumlah prediksi benar yang lebih besar dibandingkan prediksi salah. Informasi pada Gambar 6 juga menunjukkan bahwa performa model perlu dibaca bersama dengan jumlah *false classification* agar evaluasi tidak hanya mengulang angka akurasi. Berdasarkan *Performance Vector* pada Gambar 6, jumlah kesalahan klasifikasi terdiri atas 9 data Layak yang diprediksi Tidak Layak dan 7 data Tidak Layak yang diprediksi Layak. Kesalahan pertama lebih berisiko dari sisi perlindungan sosial karena dapat menghambat akses bantuan bagi warga yang membutuhkan. Kesalahan kedua berisiko terhadap efisiensi penyaluran karena bantuan dapat diberikan kepada warga yang tidak memenuhi kriteria. Mitigasi yang dapat dilakukan adalah menetapkan hasil prediksi sebagai rekomendasi awal, meninjau ulang data dengan nilai probabilitas yang berdekatan, serta melakukan validasi lapangan untuk kasus yang berada di batas keputusan.

Dengan mempertimbangkan akurasi, precision, recall, AUC, distribusi kelas, dan kesalahan klasifikasi, model Naive Bayes dalam penelitian ini dinilai memiliki performa yang baik sebagai alat bantu seleksi awal. Akan tetapi, kekuatan klaim penelitian tetap dibatasi pada dataset yang digunakan. Perbandingan dengan metode lain seperti *Decision Tree, K-Nearest Neighbor*, atau *Random Forest* belum dilakukan dalam penelitian ini sehingga belum dapat dipastikan bahwa Naive Bayes merupakan metode terbaik untuk kasus ini. Secara keseluruhan, proses penelitian menunjukkan bahwa *Naive Bayes* dapat membantu penyusunan rekomendasi kelayakan penerima bantuan sosial secara lebih sistematis. Hasil model perlu diposisikan sebagai bahan pertimbangan bagi pihak kelurahan atau instansi terkait, bukan sebagai keputusan final. Pendekatan ini penting agar hasil klasifikasi tetap dikombinasikan dengan kebijakan, kelengkapan dokumen, dan kondisi nyata masyarakat di lapangan.

5. DISKUSI

Berdasarkan hasil penelitian, algoritma *Naive Bayes* menunjukkan performa yang baik dalam menentukan kelayakan penerima bantuan sosial di Kelurahan Heram. Hal ini terlihat dari nilai akurasi sebesar 85,05% dan nilai AUC sebesar 0,924, yang menunjukkan bahwa model memiliki kemampuan klasifikasi yang tinggi dan termasuk dalam kategori *excellent classification*. Hasil tersebut mengindikasikan bahwa metode *Naive Bayes* mampu mengolah atribut sosial dan ekonomi masyarakat, seperti pekerjaan, jumlah tanggungan, kategori prioritas, penghasilan, kondisi rumah, kepemilikan aset, dan status pekerjaan, untuk membedakan penerima yang termasuk kategori Layak dan Tidak Layak. Dengan demikian, penggunaan metode ini dinilai mampu mendukung proses penentuan penerima bantuan sosial secara lebih objektif, sistematis, dan terukur dibandingkan dengan penilaian manual.

Meskipun model menunjukkan hasil yang baik, masih terdapat beberapa kesalahan klasifikasi pada data uji, yaitu adanya data yang seharusnya Layak tetapi diprediksi Tidak Layak, serta data yang seharusnya

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

Tidak Layak namun diprediksi Layak. Kondisi ini menunjukkan bahwa masih terdapat kemiripan karakteristik antar data pada kedua kelas, sehingga model belum mampu membedakan seluruh data secara sempurna. Selain itu, hasil klasifikasi juga dipengaruhi oleh kualitas data yang telah dibersihkan, pemilihan atribut yang digunakan, serta pembagian data menjadi 251 data latih dan 107 data uji. Oleh karena itu, meskipun algoritma *Naive Bayes* dalam penelitian ini telah terbukti layak digunakan sebagai pendukung pengambilan keputusan, pengembangan lebih lanjut masih dapat dilakukan, misalnya dengan menambah jumlah data, memperluas atribut penelitian, atau membandingkannya dengan metode klasifikasi lain agar hasil yang diperoleh menjadi lebih optimal.

6. KESIMPULAN DAN SARAN

6.1. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma *Naive Bayes* berhasil diterapkan untuk menentukan kelayakan penerima bantuan sosial di Kelurahan Heram. Proses penelitian dimulai dari pengumpulan data, pembersihan data, pembagian dataset menjadi 251 data latih dan 107 data uji, kemudian dilanjutkan dengan pemodelan dan pengujian menggunakan perangkat lunak RapidMiner. Atribut yang digunakan dalam proses klasifikasi meliputi pekerjaan, jumlah tanggungan, kategori tanggungan, kategori prioritas, penghasilan per bulan, kategori penghasilan, kondisi rumah, kepemilikan aset, dan status pekerjaan, dengan kelayakan aktual sebagai label kelas.

Hasil pengujian menunjukkan bahwa model *Naive Bayes* memiliki performa yang baik dalam mengklasifikasikan data ke dalam kategori Layak dan Tidak Layak. Hal ini dibuktikan dengan nilai akurasi sebesar 85,05%, yang menunjukkan bahwa sebagian besar data uji dapat diprediksi dengan benar. Selain itu, hasil *confusion matrix* menunjukkan bahwa model mampu mengklasifikasikan 45 data Tidak Layak dan 46 data Layak secara tepat, meskipun masih terdapat beberapa kesalahan klasifikasi. Nilai *precision* sebesar 83,33%, recall sebesar 86,54%, serta AUC sebesar 0,924 juga menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik dan termasuk dalam kategori *excellent classification*.

Dengan demikian, penelitian ini membuktikan bahwa algoritma *Naive Bayes* dapat digunakan sebagai metode yang efektif untuk mendukung proses penentuan kelayakan penerima bantuan sosial secara lebih objektif, sistematis, dan terukur. Penerapan metode ini diharapkan dapat membantu pihak terkait dalam mengambil keputusan yang lebih tepat sasaran, sehingga bantuan sosial dapat disalurkan kepada masyarakat yang benar-benar membutuhkan di Kelurahan Heram.

6.2. Saran

Berdasarkan hasil penelitian yang telah diperoleh, terdapat beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya. Penelitian berikutnya disarankan untuk menambah jumlah dataset agar model dapat mempelajari pola data dengan lebih baik dan menghasilkan tingkat akurasi yang lebih tinggi. Selain itu, atribut yang digunakan juga dapat diperluas dengan menambahkan variabel lain yang dianggap berpengaruh terhadap kelayakan penerima bantuan sosial, sehingga hasil klasifikasi dapat menjadi lebih rinci dan akurat. Selain itu, penelitian selanjutnya juga dapat membandingkan algoritma Naive Bayes dengan metode klasifikasi lainnya, seperti *Decision Tree, K-Nearest Neighbor*, atau *Support Vector Machine*, untuk mengetahui metode yang paling sesuai dalam menentukan kelayakan penerima bantuan sosial. Dari sisi implementasi, sistem yang telah dibangun juga dapat dikembangkan menjadi aplikasi yang dapat digunakan secara langsung oleh pihak kelurahan atau instansi terkait, sehingga proses seleksi penerima bantuan sosial dapat dilakukan dengan lebih cepat, efisien, dan tepat sasaran.

DAFTAR PUSTAKA

[1] I. Arfanda et al., “Naive Bayes dalam Menentukan Penerima Bantuan Langsung Tunai,” *itscience-indexing.comI Arfanda, W Ramdhan, RA YusdaDigital Transformation Technology, 2021•itscience-indexing.com*, Accessed: Apr. 07, 2026. [Online]. Available: https://itscience-indexing.com/jurnal/index.php/digitech/article/view/1091

[2] D. Setiawan et al., “Penerapan Metode Naive Bayes untuk Klasifikasi Penentuan Penerima Bantuan PKH,” *proceeding.unpkediri.ac.idDA Setiawan, R Halilintar, LS WahyuniarProsiding SEMNAS*

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

INOTEK (Seminar Nasional Inovasi Teknologi), 2021•proceeding.unpkediri.ac.id, Accessed: Apr. 07, 2026. [Online]. Available: https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/1137

[3] A. Musyafa, A. Januar, and E. Rilvani, "Penerapan Algoritma Naïve Bayes untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *Kohesi: Jurnal Sains dan Teknologi*, vol. 10, no. 2, pp. 71–80, Aug. 2025, doi: 10.2238/EC3WFN06.

[4] D. W. Safitri et al., "Klasifikasi Kelayakan Penerima Bantuan Program Keluarga Harapan di Kecamatan Sinar Peninjauan Menggunakan Metode Naïve," *ejournal.catursakti.ac.idDW Safitri, E Fadihilah, S RahayuSimtek: jurnal sistem informasi dan teknik komputer, 2025•ejournal.catursakti.ac.id*, vol. 10, no. 2, 2025, Accessed: Apr. 07, 2026. [Online]. Available: https://ejournal.catursakti.ac.id/index.php/simtek/article/view/1631

[5] Septia, Mega et al., "Penerapan Data Mining Klasifikasi Penerima Bantuan Pkh Di Desa Rambung Sialang Tengah Dengan Metode Naive Bayes," *journal.pelitanusantara.ac.idA SitohangJournal Data Science Penusa (JDSP), 2024•journal.pelitanusantara.ac.id*, Accessed: Apr. 07, 2026. [Online]. Available: http://journal.pelitanusantara.ac.id/index.php/jdsp/article/download/320/125

[6] T. Desi Nst, "Data Mining Kelayakan Penerima Bantuan Sembako di desa Bangun Purba Barat Menggunakan Metode Algoritma Naïve Bayes (Studi Kasus: Desa Bangun Purba Barat)," May 2024.

[7] A. Arifin, W. Handoko et al., "Implementasi Metode Naive Bayes untuk Klasifikasi Penerima Program Keluarga Harapan," *academia.edu*, Accessed: Apr. 07, 2026. [Online]. Available: https://www.academia.edu/download/86645177/812.pdf

[8] A. Damuri, U. Riyanto, "Implementasi Data Mining dengan Algoritma Naïve Bayes untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *academia.edu*, Accessed: Apr. 07, 2026. [Online]. Available: https://www.academia.edu/download/122710411/2403.pdf

[9] I. Hadianti, R. Astuti, "Implementasi Algoritma Naïve Bayes untuk Prediksi Penerima Bantuan Sosial," *ejournal.itn.ac.id*, vol. 7, no. 6, 2023, Accessed: Apr. 07, 2026. [Online]. Available: https://www.ejournal.itn.ac.id/jati/article/view/6302

[10] N. Huda, M. Hasbi, T. Susyanto, "Naive Bayes dalam Menentukan Penerima Bantuan Langsung Tunai," *itscience-indexing.com*, vol. 19, no. 1, p. 39, 2021, doi: 10.30646/sinus.v19i1.525.

[11] Pohan, Tiara, "Implementasi Algoritma Naïve Bayes untuk Klasifikasi Penerima Bantuan BPJS di Rantauprapat Labuhanbatu," 2025, Accessed: Apr. 07, 2026. [Online]. Available: http://repository.ulb.ac.id/id/eprint/1827

[12] L. N. Darussalam, R. Kurniawan, Y. A. Wijaya, and T. Suprapti, "Algoritma Naive Bayes untuk Meningkatkan Model Klasifikasi Penerima Program Indonesia Pintar Di SDN 2 Purwawinangun," *journal.eng.unila.ac.id*, vol. 13, no. 1, pp. 2830–7062, doi: 10.23960/jitet.v13i1.5882.

[13] D. Bienfield, M. Siahaan, E. Carolina Bagre, J. I. Wanda, G. Silahooy, and H. Sutejo, "Implementation of Naïve Bayes Algorithm on the Eligibility of Kartu Indonesia Pintar Scholarship (Case Study: University of Sepuluh Nopember Papua)," *Jurnal Ilmiah Sistem Informasi*, vol. 4, no. 1, pp. 191–204, Dec. 2025, doi: 10.51903/RDZDM469.

[14] M. Hofmann and R. Klinkenberg, "RapidMiner: Data Mining Use Cases and Business Analytics Applications," 2016, Accessed: Apr. 15, 2026. [Online].

[15] V. Kotu and B. Deshpande, "Predictive Analytics and Data Mining: Concepts and Practice with Rapidminer," 2014, Accessed: Apr. 15, 2026. [Online].:

[16] M. F. Nugraha and S. B. Rahayu, "Penerapan Naïve Bayes dalam Mengklasifikasi Calon Penerima Bantuan Pangan Non Tunai di Desa Nanjung Mekar," *INTERNAL (Information System Journal)*, vol. 5, no. 2, pp. 137–146, Dec. 2022, doi: 10.32627/INTERNAL.V5I2.634.

[17] L. Adzy, A. Asriyanik, A. P.-J. Mnemonic, "Algoritma Naïve Bayes untuk Klasifikasi Kelayakan Penerima Bantuan Iuran Jaminan Kesehatan Pemerintah Daerah Kabupaten Sukabumi," *ejournal.itn.ac.idLB Adzy, A Asriyanik, A PambudiJurnal Mnemonic, 2023•ejournal.itn.ac.id*, Accessed: Apr. 07, 2026. [Online]. Available: https://www.ejournal.itn.ac.id/index.php/mnemonic/article/view/5714

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

[18] A. Septiarini et al., "Penentuan Kelayakan Masyarakat Miskin Penerima Bantuan menggunakan Metode Naïve Bayes (Studi Kasus: Kabupaten Penajam Paser Utara)," *ejournal.uin-suka.ac.idN Madia, A Septiarini, HR Hatta, H Hamdani, M WatiJISKA (Jurnal Informatika Sunan Kalijaga), 2023•ejournal.uin-suka.ac.id*, Accessed: Apr. 07, 2026. [Online]. Available: https://ejournal.uin-suka.ac.id/saintek/JISKA/article/view/3752

[19] J. S. Azizah et al., "Penerapan Metode Naive Bayes Klasifikasi Kelayakan Penerima Bantuan Pangan Non Tunai (BPNT)," *ejournal.itn.ac.id*, vol. 8, no. 1, 2025, Accessed: Apr. 05, 2026. [Online]. Available: https://ejournal.itn.ac.id/index.php/mnemonic/article/download/12778/7153

[20] F. Pratama et al., "Implementasi Metode Naïve Bayes dalam Mengklasifikasi Penerima Program Keluarga Harapan (PKH) Desa Minggiran Kediri," *proceeding.unpkediri.ac.id*, Accessed: Apr. 05, 2026. [Online]. Available: https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/1072

[21] G. Wibawa., "Penerapan Metode Naive Bayes untuk Identifikasi Penerima Bantuan Program Keluarga Harapan (PKH) di Kabupaten Buton Tengah," *jmks.uho.ac.idGNA WibawaJurnal Matematika, Komputasi, dan Statistika, 2025•jmks.uho.ac.id*, Accessed: Apr. 07, 2026. [Online]. Available: https://jmks.uho.ac.id/index.php/journal/article/view/147

[22] D. L.Simorangkir and R. N. Sari, "Penerapan Data Mining dengan Algoritma SMOTE dan Naive Bayes untuk Klasifikasi Kelayakan Penerimaan Bantuan Sosial pada Kantor Camat Medan Timur," *Prosiding Seminar Nasional Multi Disiplin Ilmu (SENADIMU)*, vol. 2, no. 1, pp. 801–813, Dec. 2025, Accessed: Apr. 07, 2026. [Online]. Available: https://senadimu.potensi-utama.org/index.php/home/article/view/624

[23] B. Khotimah, W. Kustina, and Y. Kustiyahningsih, "Implementation of the Naive Bayes and Simple Additive Weighting Methods in the Feasibility Selection of Village Cash Assistance Recipients," 2023, Accessed: Apr. 07, 2026. [Online]. Available: https://www.academia.edu/download/103512674/3985.pdf

[24] N. Riyanah et al., "Penerapan Algoritma Naive Bayes untuk Klasifikasi Penerima Bantuan Surat Keterangan Tidak Mampu," *scholar.archive.orgN Riyanah, F FatmawatiJurnal Teknologi Informasi dan Multimedia, 2021•scholar.archive.org*, vol. 2, no. 4, pp. 206–213, 2021, Accessed: Apr. 07, 2026. [Online]. Available: https://scholar.archive.org/work/zicdp2vfjzfctk6vdjz5ohxuwm/access/wayback/https://journal.sekawan-org.id/index.php/jtim/article/download/117/82

[25] V. Savitri, H. Setiawan, Z., "Classification Eligibility Recipient BPJS in Ward Sendang Sari Using the Naive Bayes Method," *ejournal.isha.or.id*, Accessed: Apr. 07, 2026. [Online]. Available: https://www.ejournal.isha.or.id/index.php/Mandiri/article/view/405

[26] A. Siregar, D. Irmayani, M., "Application of Data Mining to Classify Receiving Social Assistance Using the Naïve Bayes Method," *jurnal.itsm.ac.id*, Accessed: Apr. 07, 2026. [Online]. Available: https://jurnal.itsm.ac.id/index.php/inside/article/view/881

[27] H. Nurrifqi Fakhri Fikrillah, S. Hudawiguna, C. Juliane, "Klasifikasi Penerima Bansos Menggunakan Algoritma Naive Bayes," *JATISI*, vol. 10, no. 1, pp. 683–695, Mar. 2023, doi: 10.35957/JATISI.V10I1.3624.

[28] A. Surahman and U. Hayati, "Implementasi Algoritma Naïve Bayes untuk Prediksi Penerima Bantuan Sosial," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 347–352, Mar. 2023, doi: 10.36040/JATI.V7I1.6302.

[29] L. Adzy, A. Asriyanik, "Algoritma Naïve Bayes untuk Klasifikasi Kelayakan Penerima Bantuan Iuran Jaminan Kesehatan Pemerintah Daerah Kabupaten Sukabumi," *ejournal.itn.ac.id*, vol. 6, no. 1, 2023, Accessed: Apr. 05, 2026. [Online]. Available: https://www.ejournal.itn.ac.id/index.php/mnemonic/article/view/5714