Analisis Penentuan Kelayakan Penerima Bantuan Program Keluarga Harapan (PKH) dengan Metode *Naïve Bayes* di Kelurahan Tanjung Gusta Kecamatan Medan Helvetia

Agatha Anggriani Tumanggor[1*],Marlina Setia Sinaga[2]
[1,2] Matematika, Universitas Negeri Medan, Kota Medan, 20221, Indonesia
*Corresponding Author: agathatumanggor07@gmail.com

Abstrak

Program Keluarga Harapan (PKH) di Kelurahan Tanjung Gusta, Kecamatan Medan Helvetia masih menghadapi masalah ketidaktepatan sasaran akibat proses verifikasi manual yang rawan kesalahan. Penelitian ini bertujuan menganalisis penentuan kelayakan penerima bantuan PKH dengan menerapkan metode Naïve Bayes serta mengukur tingkat akurasi, presisi, recall, dan error. Pendekatan kuantitatif digunakan dengan data lapangan yang melibatkan 623 keluarga, dan 240 data dipilih sebagai sampel menggunakan formula Isaac dan Michael melalui teknik random sampling. Variabel penelitian mencakup 13 indikator sosial-ekonomi seperti jenis kelamin, jumlah tanggungan, penghasilan, kepemilikan rumah, kondisi rumah, jumlah sepeda motor, jumlah ibu hamil, anak usia dini, anak SD, anak SMP, anak SMA, lansia, dan disabilitas. Data dibagi menjadi training (70%) dan testing (30%), kemudian dianalisis menggunakan algoritma Naïve Bayes dengan Python dan scikit-learn. Hasil penelitian menunjukkan model mampu mengklasifikasikan penerima PKH dengan akurasi 83,3%, error 16,7%, presisi 82%, dan recall 98%. Dari 72 data testing, 60 terklasifikasi dengan benar. Nilai recall tinggi menunjukkan kemampuan model mengidentifikasi keluarga layak secara optimal. Dengan demikian, metode Naïve Bayes efektif sebagai sistem pendukung keputusan yang objektif dan efisien dibandingkan verifikasi manual.
Kata kunci: Naïve Bayes, klasifikasi, bantuan sosial, PKH, akurasi, data mining

Abstract

The Family Hope Program (PKH) in Tanjung Gusta Village, Medan Helvetia District, still faces issues of mistargeting due to manual verification processes that are prone to error. This study aims to analyze the eligibility determination of PKH beneficiaries by applying the Naïve Bayes method and to measure the levels of accuracy, precision, recall, and error. A quantitative approach was used with field data involving 623 families, of which 240 were selected as samples using the Isaac and Michael formula through random sampling techniques. The research variables include 13 socio-economic indicators such as gender, number of dependents, income, home ownership status, housing condition, number of motorcycles, number of pregnant women, early childhood, elementary, junior high, and senior high school children, elderly, and disabilities. The data were divided into training (70%) and testing (30%) sets, then analyzed using the Naïve Bayes algorithm through Python and the scikit-learn library. The results showed that the model classified PKH recipients with 83.3% accuracy, 16.7% error, 82% precision, and 98% recall. Of 72 testing data, 60 were correctly classified. The high recall value indicates the model's strong ability to identify eligible families. Thus, the Naïve Bayes method proves effective as a decision support system that is objective and efficient compared to manual verification.
Keywords: Naïve Bayes, classification, social assistance, PKH, accuracy, data mining

Citation: Tumanggor, A. A. & Sinaga, M. S. (2025). Analisis Penentuan Kelayakan Penerima Bantuan Program Keluarga Harapan (PKH) Dengan Metode Naïve Bayes di Kelurahan Tanjung Gusta Kecamatan Medan Helvetia. *Prodising Seminar Nasional Jurusan Matematika 2025*. 424 – 435.

PENDAHULUAN

Tingginya angka kemiskinan merupakan salah satu faktor penyebab keterbelakangan dan menghambat pertumbuhan suatu negara, sehingga menjadikannya isu publik yang kompleks. Persoalan ekonomi yang menghambat akses terhadap keperluan hidup fundamental telah menjadi tantangan utama bagi bangsa Indonesia. Kondisi ini terjadi ketika individu atau komunitas, tanpa memandang jenis kelamin, mengalami keterbatasan dalam mencapai standar hidup yang memadai. Mereka yang tidak mampu mengakses fasilitas esensial seperti sandang, pangan, papan, pembelajaran, serta pelayanan medis dikategorikan mengalami kemiskinan. Status ini muncul karena hilangnya jaminan terhadap hak-hak fundamental yang seharusnya mendukung keberlangsungan dan peningkatan kualitas hidup manusia (Guntur, M, 2009). Menurut data Badan Pusat Statistik (BPS), 25,22 juta penduduk Indonesia atau 9,03% dari total penduduk negara ini hidup dalam kemiskinan pada tahun 2024. Menurut survey, keluarga kurang mampu atau disebut juga dengan Rumah Tangga Miskin (RTM) rata-rata mempunyai anggota keluarga yang lebih besar daripada keluarga yang mampu, dengan banyaknya tanggungan tersebut berpengaruh pada pendidikan anak-anaknya sehingga anak dari keluarga kurang mampu tidak mendapatkan pendidikan yang baik dikarenakan masalah biaya.

Data BPS menunjukkan tingkat keikutsertaan pendidikan di Sumatera Utara tahun 2024 mengalami penurunan bertahap seiring kenaikan jenjang akademik. Untuk kelompok usia 7-12 tahun di tingkat pendidikan dasar mencapai

99.50, usia 13-15 tahun pada jenjang menengah pertama sebesar 97.35, kelompok 16-18 tahun di pendidikan menengah atas 78.59 dan usia 19-23 tahun di jenjang perguruan tinggi 30.22. Statistik ini memperlihatkan adanya penurunan progresif partisipasi pendidikan pada level yang lebih tinggi di wilayah Sumatera Utara. Kondisi tersebut berdampak pada merosotnya mutu sumber daya manusia generasi mendatang, yang pada akhirnya menciptakan siklus keterpurukan ekonomi berkelanjutan. Kondisi yang seperti itu, maka pemerintah Indonesia membuat suatu program guna mengurangi angka kemiskinan di negara ini. Pemerintah Indonesia, melalui Kementerian Sosial, meluncurkan Program Keluarga Harapan (PKH) pada tahun 2007 dengan tujuan mempercepat penanggulangan kemiskinan dan mengembangkan program perlindungan sosial. PKH merupakan program bantuan sosial bersyarat dari pemerintah yang memberikan transfer tunai kepada rumah tangga kurang mampu yang tercatat dalam basis data DTKS dan ditetapkan sebagai sasaran program. Melalui skema ini, diharapkan dapat menurunkan beban pengeluaran keluarga miskin dalam jangka pendek dan berkontribusi pada upaya pengentasan kemiskinan secara jangka panjang (Fitriani, 2020).

Dalam prosesnya, PKH memiliki kendala ketidaktepatan sasaran dalam penyaluran bantuan, di mana keluarga yang layak sering terlewat dan yang kurang layak justru terdata sebagai penerima (Sriwibowo et al., 2020). Menurut Noerkaisar (2021), hal ini disebabkan oleh berbagai faktor seperti pendistribusian yang tidak merata, penyelewengan dana, pungutan liar, serta kesalahan inclusion dan exclusion. Hal ini menggambarkan perlunya proses pengambilan keputusan yang tidak memihak dan berdasarkan data untuk meningkatkan ketepatan dalam mengidentifikasi kelayakan pengguna bantuan. Penelitian ini menjadi penting karena memberikan solusi berbasis klasifikasi data untuk membantu Kementerian Sosial dan aparatur kelurahan dalam menentukan kelayakan penerima secara lebih akurat dan efisien.

Diperlukan suatu teknik untuk memberikan bobot dan mengurutkan berbagai kriteria komponen berdasarkan tingkat kepentingannya dari yang tertinggi ke terendah. Salah satu cara yang dapat diterapkan adalah pengelompokan atau pengenalan pola karakteristik serupa dalam dataset serta memprediksi kategori objek yang belum memiliki label. Algoritma pengelompokan yang populer digunakan adalah *Naïve Bayes*. Algoritma *Naïve Bayes* merupakan teknik pengelompokan probabilistik yang berdasar pada Teorema Bayes dengan prinsip dasar bahwa setiap atribut bersifat independen dan tidak saling mempengaruhi. Metode ini digunakan untuk memperkirakan kemungkinan suatu objek termasuk dalam kategori tertentu berdasarkan data yang ada. Metode *Naïve Bayes* cocok digunakan untuk memprediksi kelayakan penerima bantuan PKH karena mampu menghitung probabilitas berdasarkan data sebelumnya secara sederhana dan efisien. Metode ini secara akurat mengklasifikasikan keluarga sebagai memenuhi syarat atau tidak untuk menerima bantuan dengan menerapkan data *training* untuk menghasilkan kemungkinan setiap kriteria untuk setiap kelas. Selain minim kesalahan, keunggulan utama *Naïve Bayes* adalah kemudahan implementasinya dalam mengevaluasi probabilitas bersyarat.

Penelitian ini menerapkan pembagian data 70% untuk *training* dan 30% untuk *testing*, sesuai praktik standar dalam data mining untuk menjaga keseimbangan antara pembelajaran dan evaluasi model. Rasio ini memungkinkan model mengenali pola secara optimal sekaligus menguji kemampuannya dalam menggeneralisasi data baru. Arifin et al. (2022) membuktikan efektivitas rasio ini dalam klasifikasi penerima PKH dengan akurasi 88%, presisi 91%, dan *recall* 88%. Validasi model klasifikasi dilaksanakan menerapkan *confusion matrix* untuk menghitung akurasi, presisi, *recall*, dan *error rate*, guna menilai ketepatan prediksi model terhadap kelas layak atau tidak layak. Penentuan kelayakan penerima PKH dalam penelitian ini didasarkan pada indikator yang mencerminkan kondisi demografi (jenis kelamin kepala keluarga dan jumlah tanggungan keluarga), kondisi ekonomi (penghasilan/bulan, status kepemilikan rumah dan kondisi rumah), dan indikator khusus penerima (jumlah sepeda motor, jumlah ibu hamil, jumlah anak usia dini, jumlah anak sekolah, jumlah lansia dan jumlah disabilitas). Kategori layak diberikan kepada rumah tangga yang secara kumulatif menunjukkan kondisi tidak mampu berdasarkan kombinasi indikator tersebut. Penentuan ini sejalan dengan Peraturan Wali Kota Medan No. 33 Tahun 2021, di mana seseorang dikategorikan sebagai fakir miskin jika memenuhi minimal 60% dari 20 kriteria kemiskinan yang ditetapkan pemerintah daerah.

Salah satu kelurahan yang menjalankan bantuan PKH adalah kelurahan Tanjung Gusta. Dari hasil wawancara dengan lurah Tanjung Gusta, dan bapak Revormanuel Duha sebagai pendamping PKH kelurahan Tanjung Gusta. Kelurahan Tanjung Gusta ini merupakan salah satu kelurahan di kecamatan Medan Helvetia. Menurut data administrasi kependudukan tahun 2025, Kelurahan Tanjung Gusta memiliki wilayah dengan luas dan dihuni oleh populasi sejumlah 32.642 jiwa yang terbagi dalam 9.722 kepala keluarga. Jumlah penerima PKH sampai dengan tahun 2025 adalah berjumlah 508 KK, dengan tambahan bukan penerima yang diusulkan sebanyak 115 KK. Maka, jumlah total KK yang berhubungan langsung dengan program PKH adalah 623 KK. Dari total populasi 623 KK tersebut, penelitian ini mengambil sampel sebanyak 240 data keluarga penerima dan bukan penerima. Penentuan jumlah sampel dilakukan merujuk pada

rumus Isaac dan Michael dengan tingkat kesalahan 5%, sehingga hasil klasifikasi dapat mewakili keseluruhan populasi secara akurat.

Kecamatan Tanjung Gusta dipilih sebagai fokus penelitian karena memiliki konsentrasi penerima PKH tertinggi di Kecamatan Medan Helvetia. Banyaknya anak dari keluarga kurang mampu di Kelurahan Tanjung Gusta yang putus sekolah atau tidak melanjutkan pendidikan bahkan masih ada anak disana yang tidak menempuh dunia pendidikan dikarenakan masalah biaya. Namun permasalahan yang terjadi saat ini ialah bantuan PKH belum tersalurkan secara optimal dan tepat sasaran (Nella et al., 2022). Studi mengenai Program Keluarga Harapan menggunakan algoritma *Naive Bayes* telah dilaksanakan sebelumnya oleh para akademisi. Riset yang dikerjakan Harsyiah et al., (2024) tentang penentuan kategori penerima bantuan PKH melalui pendekatan analisis diskriminan dan *Naïve Bayes* mencatat tingkat error mencapai 27,6%. Hasil kategorisasi penerima bantuan PKH menggunakan teknik tersebut menunjukkan tingkat presisi yang baik untuk kasus penentuan status penerima bantuan PKH di wilayah NTB.

Studi oleh Nurul di Desa Petatal untuk menganalisis kriteria kelayakan penerima bantuan PKH melalui algoritma *Naïve Bayes* memperoleh presisi sempurna 100% dan tingkat kesalahan 0% dengan memanfaatkan Rapid Miner serta pendekatan *Confusion Matrix* (Rahmadani et al., 2023). Riset berikutnya oleh (Arifin et al., 2022). Teknik pengelompokan *Naïve Bayes* diterapkan untuk membangun sistem kategorisasi penerima PKH. Hasil penelitiannya mencapai akurasi 88% yang termasuk dalam kategori klasifikasi berkualitas baik. Studi tersebut menggunakan 82 data dengan pembagian rasio 70:30. Studi oleh Prata di Desa Minggiran Kediri tentang pengelompokan penerima PKH menerapkan algoritma *Naïve Bayes*. Data pelatihan berjumlah 16 sampel, mencakup 6 warga yang tidak memperoleh bantuan dan 9 warga yang mendapat bantuan. Penelitian ini menggunakan 4 parameter, meliputi status pernikahan, jumlah anggota keluarga, usia, dan pendapatan. Hasil riset menunjukkan tingkat presisi mencapai 93,33%, sehingga layak diimplementasikan di Desa Minggiran Kediri untuk membantu aparat desa dalam menetapkan kandidat penerima bantuan PKH (Pratama et al., 2021). Berdasarkan pemaparan latar belakang sebelumnya, diperlukannya sebuah metode yang dapat membantu pihak Kementerian Sosial dalam menentukan warga penerima bantuan PKH secara tepat dengan waktu yang lebih efisien di Kelurahan Tanjung Gusta, Kecamatan Medan Helvetia. Metode pengambilan keputusan tersebut akan mengklasifikasikan penerima dan bukan penerima dengan layak atau tidak layak untuk menerima bantuan sosial PKH.

METODE PENELITIAN

Jenis penelitian ini adalah penelitian kuantitatif dengan studi lapangan (*field research*) yang menggunakan sumber data primer. Metode klasifikasi diterapkan untuk mengkategorikan penerima dan bukan penerima menjadi dua kelompok yaitu layak dan tidak layak mendapatkan bantuan. Hasilnya diukur secara matematis menggunakan *confusion matrix*. Penelitian ini akan dilaksanakan di Kelurahan Tanjung Gusta, Kecamatan Medan Helvetia, Kota Medan, Provinsi Sumatera Utara. Proses pengambilan data untuk penelitian ini dilaksanakan selama dua bulan, yaitu pada periode Juni hingga Juli 2025. Populasi dalam penelitian ini adalah masyarakat yang berhubungan langsung dengan Program Keluarga Harapan (PKH) pada tahun 2025. Total populasi yang menjadi acuan adalah sebanyak 623 Kepala Keluarga (KK), yang terdiri dar penerima dan bukan penerima PKH. Dari total populasi tersebut, sampel penelitian yang diambil adalah sebanyak 240 data keluarga. Jumlah sampel ini ditentukan menggunakan rumus *Isaac* dan *Michael* dengan tingkat kesalahan toleransi sebesar 5%. Teknik pegambilan sampel yang digunakan adalah teknik *random sampling* untuk memastikan setiap anggota populasi memiliki kesempatan yang sama untuk dipilih.

Variabel dependen (Y) pada penelitian ini ialah kelayakan penerima bantuan PKH (layak / tidak layak). Variabel independen-nya adalah: Kepala Keluarga (X_1), Jumlah Tanggungan Keluarga (X_2), Penghasilan/bulan (X_3), Status Kepemilikan Rumah (X_4), Kondisi Rumah (X_5), Jumlah Sepeda Motor (X_6), Jumlah Ibu Hamil (X_7), Jumlah Anak Usia Dini (X_8), Jumlah Anak SD (X_9), Jumlah Anak SMP (X_{10}), Jumlah Anak SMA (X_{11}), Jumlah Lansia (X_{12}), Jumlah Disabilitas (X_{13}). Penelitian ini mengunakan instrumen penelitian dengan panduan wawancara terstruktur. Panduan wawancara berisi beberapa pertanyaan yang telah disusun secara sistematis berdasarkan indikator kriteria kelayakan penerima PKH seperti karakteristik demografi, kondisi ekonomi dan indikator khusus penerima. Setiap pertanyaan akan memiliki opsi jawaban yang sudah dikodekan secara numerik untuk memudahkan proses pengolahan data menggunakan metode *Naïve Bayes*. Selama wawancara, peneliti akan mencatat jawaban responden ke dalam lembar isian yang disediakan.

Teknik pengumpulan data dalam penelitian ini menggunakan Observasi lapangan dan wawancara. Prosedur penelitian ini yaitu Identifikasi masalah, Menentukan tujuan, Studi literatur, Pengumpulan data, Data *Processing*, Klarifikasi *Naïve Bayes, Confusion Matrix.* Berikut ialah tahapan analisis data pada studi ini: *Preprocessing* Data, melakukan pengujian metode *Naïve Bayes,* Evaluasi model klasifikasi menggunakan *Confusion Matrix* (Akurasi/*Accuracy, Error*,

Presisi*/ Precision, Recall).* Dalam melakukan proses perhitungan, *software* yang akan diterapkan yaitu: *Microsoft Office Excel* dan *Python*,

HASIL DAN PEMBAHASAN

Penelitian ini dilaksanakan di Kelurahan Tanjung Gusta, Kecamatan Medan Helvetia, Kota Medan, yang memiliki jumlah penerima Program Keluarga Harapan (PKH) terbanyak di wilayah tersebut. Berdasarkan data tahun 2025, terdapat 623 kepala keluarga (KK), terdiri atas 508 penerima dan 115 bukan penerima PKH. Dari jumlah tersebut, diambil 240 data sebagai sampel penelitian, terdiri atas 150 penerima dan 90 bukan penerima. Pemilihan sampel dilakukan menggunakan rumus Isaac dan Michael dengan teknik *random sampling* pada tingkat kesalahan 5%. Data diperoleh melalui wawancara terstruktur, observasi lapangan, serta dokumentasi dari pihak kelurahan dan pendamping PKH. Data penelitian kemudian dibagi menjadi dua bagian, yaitu 70% (168 data) untuk *training* dan 30% (72 data) untuk *testing* dengan metode *stratified random sampling* agar proporsi kelas tetap seimbang. Analisis dilakukan dengan algoritma *Naïve Bayes* menggunakan 13 variabel independen yang mencerminkan kondisi sosial-ekonomi keluarga, seperti jenis kelamin, jumlah tanggungan, penghasilan, kepemilikan dan kondisi rumah, jumlah kendaraan, serta jumlah anggota keluarga pada kategori khusus (ibu hamil, anak usia sekolah, lansia, dan disabilitas). Variabel dependen yang digunakan adalah status kelayakan penerima PKH (layak/tidak layak). Berikut merupakan data penerima dan bukan penerima bantuan sosial PKH Kelurahan Tanjun Gusta.

Tabel 1. Penduduk Kelurahan Tanjung Gusta

Jenis Kelamin (A_1)	Kepemilikan Rumah (A_4)	Kondisi Rumah (A_5)	Sepeda Motor (A_6)	Jumlah Anak			Lansia (A_{12})	Disabilitas (A_{13})	Staus (asli) (A_{14})
				SD (A_9)	SMP (A_{10})	SMA (A_{11})			
PR	Sewa	Permanen	Tidak Punya	√	-	-	-	-	LAYAK
LK	Sewa	Permanen	Tidak Punya	-	-	-	√	-	LAYAK
LK	Bebas Sewa	Permanen	Tidak Punya	-	√	-	-	-	LAYAK
LK	Milik Sendiri	Permanen	Cukup	-	-	-	-	-	LAYAK
LK	Bebas Sewa	Permanen	Tidak Punya	√	-	-	√	-	LAYAK
LK	Milik Sendiri	Permanen	Cukup	-	-	-	√	-	LAYAK
:	:	:	:	:	:	:	:	:	:

Dari 168 tabel 1 hanya menampilkan 6 data saja, dimana dari 6 data tersebut jumlah tanggungan (A_2) kategori *sedikit*, penghasilan (A_3) juga dalam kategori *sedikit*, tidak ada kehamilan (A_7) serta tidak memiliki anak usia dini (A_8).

Setelah data dikumpulkan, dilakukan tahap *preprocessing* untuk menyiapkan data sebelum diproses menggunakan metode *Naïve Bayes*. Seluruh variabel diubah dari bentuk kategorikal menjadi numerik agar dapat diolah oleh algoritma. Proses *preprocessing* meliputi pemeriksaan kelengkapan data guna memastikan tidak terdapat *missing value* atau format yang tidak sesuai, dengan perbaikan dilakukan berdasarkan hasil wawancara dan data resmi dari kelurahan. Selanjutnya, data dibagi menjadi dua bagian menggunakan teknik *stratified random sampling* agar proporsi kelas layak dan tidak layak tetap seimbang, yaitu 168 data (70%) sebagai *training data* dan 72 data (30%) sebagai *testing data*. Tahap akhir adalah pengubahan tipe atribut kategorikal menjadi numerik baik pada data *training* maupun *testing* untuk memastikan kesesuaian dengan kebutuhan model.

Tabel 2. merupakan hasil *preprocessing* data *training*

A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}	A_{13}	A_{14}
2	1	1	3	1	1	2	2	2	2	2	1	2	0
1	2	2	2	1	2	2	2	1	1	2	2	2	0
1	1	2	3	1	2	2	1	1	2	2	2	2	0
2	1	1	3	1	1	2	2	2	2	2	2	2	1
1	2	2	2	2	2	2	2	1	2	1	2	2	0
:	:	:	:	:	:	:	:	:	:	:	:	:	:
1	1	2	3	1	2	2	2	2	1	2	2	2	0

Keterangan: Data hasil *preprocessing* (menampilkan 5 baris awal, titik-titik mewakili baris tengah, dan baris terakhir adalah data ke-168).

Tabel 3. Hasil *preprocessing* data *testing*

A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}	A_{13}	A_{14}
1	2	1	1	1	1	2	2	2	2	2	1	2	1
1	1	2	1	1	1	2	2	2	2	2	2	2	0
2	1	1	3	1	1	2	2	2	2	2	1	2	0
1	1	1	3	1	2	2	2	2	2	2	1	2	0
1	2	3	2	1	2	2	1	1	1	2	2	2	0

:	:	:	:	:	:	:	:	:	:	:	:	:	:
1	1	1	2	1	2	2	1	1	2	2	2	2	0

Keterangan: Data hasil *preprocessing* (menampilkan 5 baris awal, titik-titik mewakili baris tengah, dan baris terakhir adalah data ke-72).

Selanjutnya Perhitungan *Naive Bayes Classifier*, merujuk pada temuan perhitungan probabilitas *prior* pada tabel 4 diketahui bahwasanya pemodelan ini menghasilkan dua kelas yaitu kelas layak (C_0) dengan nilai probabilitas 0.583 atau 58.3% dan kelas tidak layak (C_1) dengan nilai probabilitas 0.417 atau 41.7%. Sehingga ketika kedua kelas tersebut dijumlahkan akan menghasilkan nilai 100%. Hasil ini menunjukkan bahwa dalam dataset *training*, mayoritas keluarga di Kelurahan Tanjung Gusta (58.3%) memenuhi kriteria untuk menerima bantuan PKH, yang mengindikasikan kondisi sosial ekonomi masyarakat yang sebagian besar berada dalam kategori kurang mampu. *Prior probability* ini akan menjadi *baseline* dalam proses klasifikasi *Naïve Bayes*. Setelah memperoleh probabilitas *prior*, langkah selanjutnya adalah menghitung probabilitas *likelihood* untuk setiap variabel berdasarkan kelas target.

Tabel 4. *Probabilitas prior*

Perhitungan	Probabilitas *Prior*	Hasil	Persentase
$P(C_0)$	$\frac{98}{168}$	0.583	58.3%
$P(C_1)$	$\frac{70}{168}$	0.417	41.7%
	Total	1	100%

Tabel 5 adalah perhitungan setiap kemungkinan hasil kriteria $P(X_{\text{Kepala Keluarga}}|C_i)$. Analisis terhadap variabel kepala keluarga pada tabel 4 menunjukkan bahwa kepala keluarga 'Laki-laki' memiliki probabilitas tinggi di kedua kelas, namun sedikit lebih dominan pada kelas 'Tidak Layak' sebanyak (0.80). Sebaliknya, kepala keluarga 'Perempuan' menunjukkan probabilitas yang sedikit lebih tinggi untuk diklasifikasikan sebagai 'Layak' sebanyak (0.26).

Tabel 5. Probabilitas Kepala Keluarga

Jenis Kelamin	Jumlah	Kejadian	Probabilitas	
	Layak	Tidak Layak	Layak (C_0)	Tidak Layak (C_1)
Laki-laki	73	56	0.74	0.80
Perempuan	25	14	0.26	0.20
Jumlah	98	70	1	1

Berikut ialah perhitungan setiap kemungkinan hasil kriteria $P(X_{\text{Jumlah Tanggungan Keluarga}}|C_i)$. Dari data jumlah tanggungan keluarga pada tabel 6, terlihat pola bahwa jumlah tanggungan kategori 'Sedikit' sangat dominan pada kelas 'Tidak Layak' sebanyak (0.71). Sementara jumlah tanggungan kategori 'Sedang' lebih cenderung berada pada kelas 'Layak' sebanyak (0.36).

Tabel 6. Probabilitas Jumlah Tanggungan Keluarga

Jumlah Tanggungan Keluarga	Jumlah Kejadian		Probabilitas	
	Layak	Tidak Layak	Layak (C_0)	Tidak Layak (C_1)
Sedikit	62	50	0.63	0.71
Sedang	35	18	0.36	0.26
Banyak	1	2	0.01	0.03
Jumlah	98	70	1	1

Sebuah anomali menarik terlihat pada kategori 'Banyak' untuk jumlah tanggungan. Berdasarkan data, probabilitas untuk kelas 'Tidak Layak' sebanyak (0.03) justru lebih tinggi dibandingkan kelas 'Layak'. Hal ini tampak berlawanan dengan asumsi umum bahwa lebih banyak tanggungan seharusnya meningkatkan kelayakan. Namun, hasil ini dapat dijelaskan oleh kelangkaan data pada kategori tersebut. Dalam 168 data *training*, hanya terdapat 3 kasus keluarga dengan tanggungan 'Banyak', dimana 2 diantaranya tergolong 'Tidak Layak'. Oleh karena itu, hasil probabilitas untuk kategori ini lebih mencerminkan keunikan sampel data secara terbatas daripada representasi populasi secara umum. Berikut Adalah (tabel 7) perhitungan setiap kemungkinan hasil kriteria $P(X_{\text{Penghasilan}}|C_i)$.

Tabel 7. Probabilitas Penghasilan

Penghasilan / bulan	Jumlah Kejadian		Probabilitas	
	Layak	Tidak Layak	Layak (C_0)	Tidak Layak (C_1)
Sedikit	58	35	0.59	0.50
Sedang	36	32	0.37	0.46
Banyak	4	3	0.04	0.04
Jumlah	98	70	1	1

Tabel 7 menyajikan probabilitas untuk variabel penghasilan, dimana kategori penghasilan 'Sedikit' menjadi indikator kuat untuk kelas 'Layak' dengan probabilitas sebesar (0.59). Sementara itu, penghasilan 'Sedang' lebih cenderung berada pada kelas 'Tidak Layak' sebesar (0.46). Sedangkan, penghasilan 'Banyak' probabilitasnya sama untuk kedua kelas yaitu (0.04). Hal ini terjadi karena jumlah sampel untuk kategori 'Banyak' memiliki jumlah yang sedikit dalam data *training* yaitu 4 kasus layak dan 3 kasus tidak layak. Dengan data yang minim, model tidak dapat menemukan pola yang kuat, sehingga probabilitasnya menjadi hampir seimbang. Berikut ialah perhitungan setiap kemungkinan hasil kriteria $P(X_{\text{Status Kepemilikan Rumah}}|C_i)$.

Tabel 8. Probabilitas Status Kepemilikan Rumah

Status	Jumlah Kejadian		Probabilitas	
Kepemilikan Rumah	Layak	Tidak Layak	Layak (C_0)	Tidak Layak (C_1)
Sewa	15	16	0.16	0.23
Bebas Sewa	62	39	0.63	0.56
Milik Sendiri	21	15	0.21	0.21
Jumlah	98	70	1	1

Untuk variabel status kepemilikan rumah, dapat diamati bahwa keluarga dengan status rumah kategori 'Sewa' secara proporsional memiliki probabilitas yang lebih tinggi pada kelas 'Tidak Layak' sebesar (0.23). Hal ini bukan disebabkan oleh jumlah absolutnya, melainkan karena efek pembagian terhadap total data di masing-masing kelas yang berbeda (98 vs 70). Sementara itu, kategori 'Bebas Sewa' memiliki probabilitas tertinggi pada kelas 'Layak' sebesar (0.63). Sedangkan, kategori 'Milik Sendiri' menghasilkan nilai probabilitas yang sama-sama (0.21) untuk kedua kelas. Akibatnya, khusus untuk kategori 'Milik Sendiri', status kepemilikan rumah menjadi indikator yang netral dalam penentuan klasifikasi. Berikut ialah perhitungan setiap kemungkinan hasil kriteria $P(X_{\text{Kondisi Rumah}}|C_i)$.

Tabel 9. Probabilitas Kondisi Rumah

Kondisi	Jumlah Kejadian		Probabilitas	
Rumah	Layak	Tidak Layak	Layak (C_0)	Tidak Layak (C_1)
Permanen	91	57	0.93	0.81
Semi Permanen	7	13	0.07	0.19
Jumlah	98	70	1	1

Tabel 9 menyajikan hasil yang paling signifikan dan berlawanan dengan intuisi umum. Kondisi rumah 'semi permanen' yang secara logika seharusnya menjadi indikator kuat kelayakan justru memiliki probabilitas lebih tinggi pada kelas 'Tidak Layak' sebesar (0.19). Fenomena ini kemungkinan besar disebabkan oleh karaktenstik unik pada data sampel, dimana 13 keluarga dalam kondisi rumah ini ternyata tidak memenuhi syarat kelayakan karena perbedaan indikator lain yang lebih signifikan (seperti penghasilan atau komponen keluarga). Sebaliknya, kondisi rumah 'Permanen' justru sedikit lebih dominan pada kelas 'Layak' sebanyak (0.93). Hal ini mengindikasikan bahwa di lokasi penelitian, memiliki rumah 'Permanen' adalah standar umum dan bukan merupakan indikator yg andal untuk membedakan tingkat kesejahteraan. Berikut ialah perhitungan setiap kemungkinan hasil kriteria $P(X_{\text{Jumlah Sepeda Motor}}|C_i)$.

Tabel 10. Probabilitas Jumlah Sepeda Motor

Jumlah Sepeda	Jumlah Kejadian		Probabilitas	
Motor	Layak	Tidak Layak	Layak (C_0)	Tidak Layak (C_1)
Tidak Punya	37	34	0.38	0.49
Cukup	55	33	0.56	0.47
Mampu	6	3	0.06	0.04
Jumlah	98	70	1	1

Tabel 10 menunjukkan pola yang menarik dan berlawanan degan asumsi umum terkait kepemilikan aset. Kategori 'Tidak Punya' sepeda motor justru memiliki probabilitas lebih tinggi pada kelas 'Tidak Layak' sebesar (0.49). Hal ini kemungkinan disebabkan oleh karakteristik lain pada kelas tersebut dimana kepemilikan motor bukan menjadi kebutuhan utama. Kategori yang paling signifikan adalah kategori 'Cukup' dan kategori 'Mampu' justru lebih dominan pada kelas 'Layak' (masing-masing 0.56 dan 0.06). Ini mengindikasikan bahwa di lokasi penelitian, sepeda motor tidak dapat dianggap sebagai barang mewah, melainkan sebagai alat produksi atau aset yang sangat penting untuk mencari nafkah. Bagi banyak keluarga kurang mampu, kepemilikan sepeda motor adalah syarat utama untuk dapat bekerja dan memperoleh penghasilan, sehingga keberadaannya justru menjadi karakteristik umum pada keluarga yang layak menerima bantuan. Berikut ialah perhitungan setiap kemungkinan hasil kriteria $P(X_{\text{Jumlah Ibu Hamil}}|C_i)$.

Tabel 11. Probabilitas Jumlah Ibu Hamil

Jumlah	Jumlah Kejadian		Probabilitas	
Ibu Hamil	Layak	Tidak Layak	Layak (C_0)	Tidak Layak (C_1)

Ada	0	1	0	0.01
Tidak Ada	98	69	1	0.99
Jumlah	98	70	1	1

Tabel 11 menyajikan hasil paling kritis dalam penelitian ini, yang dikenal sebagai masalah frekuensi nol. Secara aturan, ibu hamil adalah salah satu komponen utama penerima PKH, namun data menunjukkan probabilitas nol untuk kelas 'Layak'. Hal ini terjadi karena dalam 168 data training, tidak ditemukan satu pun kasus keluarga 'Layak' yang memiliki Ibu Hamil. Jadi pada kelas 'Tidak Layak' mendapatkan nilai yg lebih tinggi sebesar (0.01). Meskipun kategori 'Tidak Ada' pada variabel Ibu Hamil menunjukkan probabilitas (1) pada kelas 'Layak', ini tidak boleh diartikan bahwa ketiadaan ibu hamil menjadi penyebab kelayakan. Sebaliknya, hasil ini mencerminkan kondisi demografis yang paling umum di lokasi penelitian, dimana mayoritas keluarga secara alami tidak memiliki anggota yang sedang hamil pada satu waktu. Berikut ialah perhitungan setiap kemungkinan hasil kriteria $P(X_{\text{Jumlah Anak Usia Dini}}|C_i)$.

Tabel 12. Probabilitas Jumlah Anak Usia Dini

Jumlah Anak Usia Dini	Jumlah Kejadian		Probabilitas	
	Layak	Tidak Layak	Layak (C_0)	Tidak Layak (C_1)
Ada	15	13	0.15	0.19
Tidak Ada	83	57	0.85	0.81
Jumlah	98	70	1	1

Sama seperti variabel sebelumnya, Tabel 12 menunjukkan pola yang berlawanan dengan intuisi. Kategori 'Ada' Anak Usia Dini secara proporsional justru lebih dominan pada kelas 'Tidak Layak' sebesar (0.19). Hal ini disebabkan oleh adanya faktor lain yang membuat keluarga tersebut tidak memenuhi syarat, meskipun memiliki komponen Anak Usia Dini. Di sisi lain, kategori 'Tidak Ada' Anak Usia Dini sangat dominan pada kelas 'Layak' sebesar (0.85). Ini tidak berarti ketiadaan Anak Usia Dini menjadi penyebab kelayakan, melainkan mencerminkan bahwa mayoritas keluarga 'Layak' dalam sampel ini memenuhi syarat kelayakan melalui komponen PKH lainnya sepeti memiliki anak sekolah. Berikut ialah perhitungan setiap kemungkinan hasil kriteria $P(X_{\text{Jumlah Anak SD}}|C_i)$.

Tabel 13. Probabilitas Jumlah Anak SD

Jumlah Anak SD	Jumlah Kejadian		Probabilitas	
	Layak	Tidak Layak	Layak (C_0)	Tidak Layak (C_1)
Ada	28	18	0.29	0.26
Tidak Ada	70	52	0.71	0.74
Jumlah	98	70	1	1

Analisis terhadap variabel Jumlah Anak SD pada tabel 13 menunjukkan pola yang sejalan dengan kriteria Program Keluarga Harapan (PKH). Keberadaan Anak SD dalam keluarga memiliki probabilitas yang lebih tinggi untuk diklasifikasikan sebagai 'Layak' sebesar (0.29). Sedangkan pada kategori 'Tidak Ada' menunjukkan probabilitas yang sedikit lebih tinggi untuk diklasifikasikan sebagai 'Tidak layak' sebesar (0.74). Berikut ialah perhitungan setiap kemungkinan hasil kriteria $P(X_{\text{Jumlah Anak SMP}}|C_i)$.

Tabel 14. Probabilitas Jumlah Anak SMP

Jumlah Anak SMP	Jumlah Kejadian		Probabilitas	
	Layak	Tidak Layak	Layak (C_0)	Tidak Layak (C_1)
Ada	27	11	0.28	0.16
Tidak Ada	71	59	0.72	0.84
Jumlah	98	70	1	1

Dari data jumlah Anak SMP pada tabel 14, terlihat pola bahwa jumlah Anak SMP kategori 'Ada' sangat dominan pada kelas 'Layak' sebesar (0.28). Sedangkan pada kategori 'Tidak Ada' menunjukkan probabilitas yang lebih tinggi untuk diklasifikasikan sebagai 'Tidak Layak' sebesar (0.84). Berikut ialah perhitungan setiap kemungkinan hasil kriteria $P(X_{\text{Jumlah Anak SMA}}|C_i)$.

Tabel 15. Probabilitas Jumlah Anak SMA

Jumlah Anak SMA	Jumlah Kejadian		Probabilitas	
	Layak	Tidak Layak	Layak (C_0)	Tidak Layak (C_1)
Ada	26	14	0.27	0.20
Tidak Ada	72	56	0.73	0.80
Jumlah	98	70	1	1

Untuk variabel jumlah Anak SMA pada tabel 15, dapat diamati bahwa keluarga dengan jumlah anak SMA kategori 'Ada' secara proporsional memiliki probabilitas yang lebih tinggi pada kelas 'Layak' sebesar (0.27). Sedangkan pada

kategori 'Tidak Ada' menunjukkan probabilitas yang lebih tinggi pada kelas 'Tidak Layak' (0.80). Berikut ialah perhitungan setiap kemungkinan hasil kriteria $P(X_{\text{Jumlah Lansia}}|C_i)$.

Tabel 16. Probabilitas Jumlah Lansia

Jumlah	Jumlah Kejadian		Probabilitas	
Lansia	Layak	Tidak Layak	Layak (C_0)	Tidak Layak (C_1)
Ada	31	24	0.32	0.34
Tidak Ada	67	46	0.68	0.66
Jumlah	98	70	1	1

Tabel 16 menunjukkan hasil yang menarik, dimana kategori 'Ada' Lansia secara proporsional memiliki probabilitas yang sedikit lebih tinggi pada kelas 'Tidak Layak' sebesar (0.34). Hal ini mengindikasikan bahwa keberadaan lansia saja tidak cukup untuk menjamin kelayakan jika tidak didukung oleh komponen lain. Sebaliknya, kategori 'Tidak Ada' Lansia justu sedikit lebih dominan pada kelas 'Layak' sebesar (0.68). Hal ini tidak berarti ketiadaan lansia menjadi penyebab kelayakan. Hasil ini lebih mencerminkan bahwa mayoritas keluarga 'Layak' dalam sampel penelitian ini memenuhi syarat kelayakan melalui komponen PKH yang paling umum, yaitu kepemilikan anak sekolah. Berikut ialah perhitungan setiap kemungkinan hasil kriteria $P(X_{\text{Jumlah Disabilitas}}|C_i)$.

Tabel 17. Probabilitas Jumlah Disabilitas

Jumlah	Jumlah Kejadian		Probabilitas	
Disabilitas	Layak	Tidak Layak	Layak (C_0)	Tidak Layak (C_1)
Ada	2	0	0.02	0
Tidak Ada	96	70	0.98	1
Jumlah	98	70	1	1

Analisis terhadap Tabel 17 menunjukkan bahwa variabel jumlah disabilitas dengan kategori 'Ada' memiliki probabilitas sedikit lebih tinggi pada kelas 'Layak' sebesar (0.02). Sedangkan pada kategori 'Tidak Ada' memiliki probabilitas yang sedikit lebih tinggi juga pada kelas 'Tidak Layak' sebesar (1). Perhitungan probabilitas posterior dilakukan untuk menentukan kelas akhir pada setiap data uji, yaitu apakah keluarga tergolong *Layak* atau *Tidak Layak* menerima bantuan PKH. Probabilitas posterior dihitung dengan mengalikan nilai *likelihood* dari setiap atribut dengan probabilitas *prior* dari masing-masing kelas. Kelas dengan nilai posterior tertinggi akan ditetapkan sebagai hasil klasifikasi. Dari total 72 data uji, dua data dipilih sebagai sampel untuk perhitungan manual guna menunjukkan proses klasifikasi menggunakan metode *Naïve Bayes*.

Pada kasus pertama dengan karakteristik keluarga berpenghasilan rendah, rumah sewa, dan jumlah tanggungan sedang, diperoleh hasil perhitungan bahwa probabilitas posterior untuk kelas *Layak* ($P(C_0|X)$=0.00003069) lebih besar daripada kelas *Tidak Layak* ($P(C_1|X)$=0.00001635). Dengan demikian, data uji tersebut diklasifikasikan sebagai keluarga *Layak* menerima bantuan PKH. Hasil ini menunjukkan bahwa model mampu mengenali ciri-ciri keluarga berpenghasilan rendah secara akurat. Sebaliknya, pada kasus kedua dengan kondisi ekonomi lebih baik seperti kepemilikan rumah sendiri dan jumlah kendaraan cukup, diperoleh nilai $P(C_0|X)$=0.00016757 dan $P(C_1|X)$=0.00018893. Karena nilai $P(C_1|X)$ lebih besar, maka data uji tersebut dikategorikan *Tidak Layak* menerima bantuan PKH. Hasil ini membuktikan bahwa metode *Naïve Bayes* dapat membedakan dengan baik antara keluarga yang layak dan tidak layak berdasarkan kombinasi variabel sosial-ekonomi yang dianalisis. Kemudian

Implementasi *Python* untuk *Confusion Matrix, Accuracy* dan *Classification Report*:

```
# ========================================
# 8. Evaluasi
# ========================================
print("Akurasi:", accuracy_score(y_test, y_pred))
print("\nConfusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))
```

Gambar 1. *Implementasi Python*

```
Akurasi: 0.8333333333333334

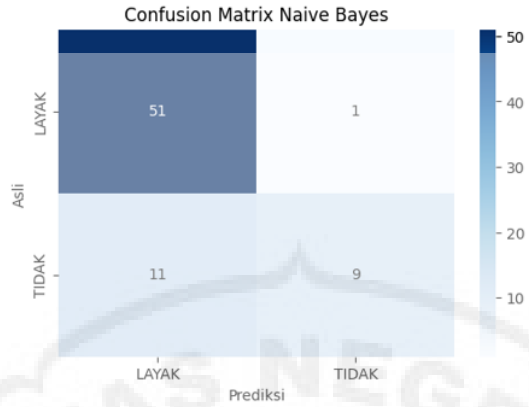
Confusion Matrix:
 [[51  1]
 [11  9]]

Classification Report:
              precision    recall  f1-score   support

           0       0.82      0.98      0.89        52
           1       0.90      0.45      0.60        20

    accuracy                           0.83        72
   macro avg       0.86      0.72      0.75        72
weighted avg       0.84      0.83      0.81        72
```

Gambar 2. Output Hasil



Gambar 3. Output Hasil *Confusion Matrix*

Tabel 18. Hasil *Confusion Matrix*

Status Asli	Prediksi: Layak	Prediksi: Tidak Layak
Layak	51 (TP)	1 (FN)
Tidak Layak	11 (FP)	9 (TN)

Berdasarkan hasil pengujian menggunakan *confusion matrix*, metode *Naïve Bayes* menunjukkan performa yang baik dalam mengklasifikasikan kelayakan penerima bantuan sosial Program Keluarga Harapan (PKH). Dari total 72 data uji, model berhasil memprediksi 60 data secara benar dengan nilai *accuracy* sebesar 83,3%, yang berarti sebagian besar data dapat diklasifikasikan secara tepat antara kategori *Layak* dan *Tidak Layak*. Hasil ini menunjukkan bahwa algoritma *Naïve Bayes* mampu bekerja secara efektif dalam mengenali pola dari variabel-variabel yang digunakan, seperti jumlah tanggungan, penghasilan, dan kondisi rumah. Sementara itu, tingkat kesalahan (*error rate*) yang dihasilkan sebesar 16,7% mengindikasikan bahwa hanya sebagian kecil data yang mengalami kekeliruan prediksi. Nilai ini masih tergolong rendah, sehingga model dapat dianggap memiliki tingkat reliabilitas yang cukup tinggi dalam konteks penelitian sosial.

Selain itu, hasil evaluasi menunjukkan nilai *precision* sebesar 82% dan *recall* sebesar 98%. Nilai *precision* yang tinggi menandakan bahwa sebagian besar keluarga yang diprediksi sebagai *Layak* memang benar-benar layak menerima bantuan, sehingga tingkat salah sasaran relatif kecil. Sementara itu, nilai *recall* yang mendekati sempurna menunjukkan kemampuan model yang sangat baik dalam mengidentifikasi seluruh keluarga yang benar-benar membutuhkan. Hal ini penting dalam konteks pemerataan bantuan sosial, karena model mampu meminimalkan risiko keluarga miskin yang terlewat. Secara keseluruhan, hasil ini membuktikan bahwa algoritma *Naïve Bayes* dapat menjadi alat bantu yang efektif, efisien, dan akurat dalam proses penentuan penerima PKH secara objektif dan berbasis data.

Pembahasan

Berdasarkan hasil penelitian saya, penggunaan algoritma *Naïve Bayes* dalam proses penentuan kelayakan penerima PKH pada wilayah Kelurahan Tanjung Gusta memperoleh tingkat ketepatan klasifikasi sejumlah 83.3%. Ini berarti model berhasil mengklasifikasikan dengan benar 60 dari 72 data *testing*, yang menunjukkan performa baik dalam konteks klasifikasi biner untuk sistem pendukung keputusan. Namun, perlu diperhatikan bahwa tingkat kesalahan sebesar 16.7% masih ada, dan ini memerlukan evaluasi lebih lanjut, mengingat dampak sosial yang signifikan dari kesalahan klasifikasi dalam program bantuan sosial khususnya PKH. Hasil akurasi sebesar 83.3% yang diperoleh dalam penelitian ini menunjukkan kinerja yang 'Baik' dan sejalan dengan penelitian relevan sebelumnya yang dibahas pada latar belakang. Sebagai perbandingan, penelitian oleh Arifin et al. (2022) yang juga menerapkan *Naïve Bayes* untuk klasifikasi penerima PKH mendapatkan akurasi sebesar 88%. Sementara itu, penelitian Harsyiah et al. (2024) mencatat tingkat *error* 27,6% yang berarti akurasinya sekitar 72,4%. Hasil penelitian ini berada di antara kedua studi tersebut, mengonfirmasi bahwa metode *Naïve Bayes* merupakan pendekatan yang efektif dan dapat diandalkan untuk kasus penentuan kelayakan penerima bantuan sosial khususnya PKH.

Distribusi probabilitas *prior* yang dihasilkan dari data *training* menunjukkan karakteristik demografis yang penting, di mana 58.3% keluarga tergolong layak menerima PKH, sementara 41.7% tidak layak. Proporsi ini mencerminkan kondisi sosial ekonomi masyarakat Kelurahan Tanjung Gusta, yang mayoritas berada dalam kategori kurang mampu, sejalan dengan tujuan PKH sebagai program pengentasan kemiskinan. Temuan ini konsisten dengan data BPS Kota Medan yang menunjukkan bahwa pada tahun 2024, terdapat 7.94% penduduk miskin di Kota Medan, atau sebanyak 187,04 ribu jiwa.

Ini mengindikasikan bahwa wilayah urban marginal seperti Tanjung Gusta memiliki konsentrasi kemiskinan yang cukup signifikan. Analisis terhadap nilai *precision* sebesar 82% menunjukkan bahwa model memiliki tingkat kepercayaan yang tinggi dalam mengidentifikasi penerima yang benar-benar memenuhi kriteria PKH. Dari semua prediksi "layak" yang dibuat oleh model, 82% diantaranya terbukti benar-benar layak menerima bantuan. Hal ini mengindikasikan bahwa model berhasil meminimalkan kesalahan *false positive*, yaitu pemberian bantuan kepada keluarga yang sebenarnya tidak memenuhi kriteria. Dalam konteks program bantuan sosial, minimalisasi *false positive* sangat penting untuk memastikan efisiensi penggunaan anggaran dan ketepatan sasaran program.

Sementara itu, nilai *recall* sebesar 98% menunjukkan kemampuan model dalam mengidentifikasi keluarga yang benar-benar berhak menerima PKH. Model berhasil mendeteksi 98% dari seluruh keluarga yang seharusnya layak mendapat bantuan, meskipun masih terdapat 2% keluarga layak yang tidak teridentifikasi (*false negative*). Dalam konteks keadilan sosial, *false negative* memiliki implikasi yang serius karena dapat merugikan keluarga miskin yang seharusnya mendapat bantuan. Nilai *precision* dan *recall* yang baik menunjukkan bahwa model *Naïve Bayes* cukup efektif dalam mengklasifikasi penerima PKH. Model ini seimbang, tidak terlalu ketat sehingga keluarga layak tidak terlewat, dan tidak terlalu longgar sehingga bantuan tidak salah sasaran. Hal ini penting untuk memastikan bantuan benar-benar tepat sasaran tanpa mengabaikan keluarga yang membutuhkan.

Kinerja model dengan akurasi 83.3% ini mendukung teori klasifikasi data mining yang menyatakan bahwa *Naïve Bayes* efektif untuk dataset dengan atribut numerik dan ukuran data yang terbatas. Hasil ini secara praktis mengonfirmasi beberapa keunggulan Naive Bayes, yaitu kemampuannya memberikan hasil yang stabil meskipun hanya dilatih dengan dataset yang relatif kecil (168 data), sejalan dengan teori yang menyatakan bahwa *Naïve Bayes* tidak memerlukan data pelatihan dalam jumlah besar. Kemampuannya untuk menangani data kategorikal dan numerik secara bersamaan juga terbukti efektif dalam memproses 13 variabel penelitian yang beragam. Selain itu, kecepatan komputasi yang tinggi dari algoritma ini menjadikannya praktis untuk implementasi dalam sistem informasi kelurahan yang memiliki keterbatasan sumber daya komputasi. Analisis terhadap dua kasus *testing* yang dipresentasikan mengungkapkan pola menarik dalam proses klasifikasi. Pada kasus pertama, meskipun memiliki beberapa karakteristik yang mengindikasikan kemiskinan (kepala keluarga laki-laki, status kepemilikan rumah sewa, kondisi rumah permanen, tidak punya sepeda motor), model tetap mengklasifikasikan sebagai "layak" dengan nilai probabilitas akhir yang lebih tinggi. Sebaliknya, kasus kedua dengan kondisi yang relatif lebih baik (kepala keluarga laki-laki, status kepemilikan rumah milik sendiri, kondisi rumah permanen, punya sepeda motor) diklasifikasikan sebagai "tidak layak". Hal ini menunjukkan bahwa model berhasil menangkap kompleksitas multidimensional dari kemiskinan yang tidak hanya bergantung pada satu atau dua indikator, melainkan kombinasi dari seluruh variabel yang digunakan.

Analisis lebih lanjut terhadap nilai-nilai probabilitas *likelihood* menunjukkan adanya beberapa variabel yang bersifat dominan dalam menentukan hasil klasifikasi. Variabel kondisi Rumah kategori 'Permanen' dan variabel Jumlah Anak SMP kategori 'Ada' merupakan variabel yang paling dominan karena keduanya memberikan selisih probabilitas terbesar yaitu (0.12) untuk mendukung kelayakan. Sedangkan yang menjadi pendorong kuat ketiga adalah variabel Jumlah Tanggungan kategori 'Sedang' karena memiliki selisih probabilitas (0.10). Variabel-variabel inilah yang memberikan sinyal pembeda paling kuat buat bagi model dalam membuat keputusan. Sebaliknya, model ini juga memiliki beberapa variabel yang bersifat lemah, yaitu variabel yang gagal menjadi pembeda yang kuat karena selisih probabilitasnya sangat kecil antara kedua kelas. Variabel terlemah dalam model ini adalah Jumlah Lansia dengan selisih probabilitas hanya (0.02), diikuti oleh Jumlah Anak SD dengan selisih probabilitas (0.03), dan selanjutnya Jumlah Anak Usia Dini dengan selisih probabilitas (0.04). Kecilnya selisih ini menunjukkan bahwa variabel tersebut hampir tidak memberikan pengaruh signifikan terhadap kelas klasifikasi akhir. Penelitian ini memberikan kontribusi teoretis dalam pengembangan aplikasi data mining untuk sektor publik, khususnya dalam bidang kesejahteraan sosial. Implementasi *Naïve Bayes* untuk klasifikasi penerima bantuan sosial menambah literatur tentang penerapan *machine learning* dalam tata kelola dan pengentasan kemiskinan. Secara praktis, penelitian ini menawarkan solusi teknologi untuk meningkatkan efektivitas dan efisiensi program PKH melalui sistem pendukung keputusan yang dapat membantu petugas kelurahan dalam melakukan seleksi awal calon penerima.

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, penerapan metode *Naïve Bayes* dalam penentuan kelayakan penerima bantuan Program Keluarga Harapan (PKH) di Kelurahan Tanjung Gusta terbukti efektif. Model ini memanfaatkan 13 variabel independen yang mencerminkan kondisi sosial, ekonomi, dan demografis keluarga

untuk membangun sistem prediksi berbasis probabilitas yang objektif. Melalui tahapan *preprocessing* data dan perhitungan probabilitas prior, likelihood, serta posterior, sistem berhasil mengidentifikasi pola karakteristik yang membedakan antara keluarga layak dan tidak layak menerima bantuan. Hasil distribusi probabilitas prior menunjukkan bahwa 58,3% keluarga tergolong layak menerima PKH, mencerminkan kondisi sosial ekonomi masyarakat yang sebagian besar masih membutuhkan dukungan. Variabel seperti kondisi rumah, jumlah tanggungan, dan keberadaan anak usia SMP menjadi faktor yang paling dominan dalam menentukan kelayakan penerimaan bantuan. Dengan demikian, metode *Naïve Bayes* layak dijadikan alat bantu dalam sistem pendukung keputusan untuk melakukan *screening* awal calon penerima PKH secara konsisten dan efisien.

Evaluasi kinerja model menunjukkan hasil yang sangat baik dengan tingkat *accuracy* sebesar 83,3%, *error rate* sebesar 16,7%, *precision* sebesar 82%, dan *recall* sebesar 98%. Nilai *accuracy* yang tinggi mengindikasikan kemampuan model untuk melakukan klasifikasi dengan tepat terhadap sebagian besar data uji, sementara *precision* yang mencapai 82% mencerminkan tingkat keandalan model dalam memprediksi penerima yang benar-benar layak. Selain itu, nilai *recall* yang sangat tinggi menunjukkan efektivitas model dalam mendeteksi keluarga yang berhak menerima bantuan tanpa banyak yang terlewat. Secara keseluruhan, kombinasi metrik evaluasi tersebut menegaskan bahwa algoritma *Naïve Bayes* memiliki potensi besar untuk digunakan sebagai sistem pendukung keputusan berbasis data dalam program bantuan sosial, guna meningkatkan ketepatan sasaran dan transparansi dalam penyaluran bantuan pemerintah.

Untuk penelitian mendatang, disarankan agar menggunakan himpunan data yang lebih banyak serta melibatkan cakupan wilayah yang lebih luas, sehingga temuan yang diperoleh dapat memberikan gambaran yang lebih representatif. Penerapan teknik *cross-validation* diperlukan untuk meningkatkan ketepatan estimasi kinerja model. Dalam penentuan penerima PKH, metode pembobotan seperti *Analytical Hierarchy Process* (AHP) atau *Simple Additive Weighting* (SAW) dapat dikombinasikan dengan algoritma *Naïve Bayes* guna mengetahui faktor-faktor yang memengaruhi kelayakan penerima bantuan PKH. Untuk implementasi, perlu dikembangkan sistem yang mudah digunakan, disertai pelatihan bagi petugas, serta dilakukan monitoring secara berkala. Hasil penelitian ini juga berpotensi menjadi dasar penyusunan kebijakan dan prosedur penentuan penerima bantuan sosial yang lebih objektif, transparan, serta tetap menjaga keamanan dan privasi data Masyarakat.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak Kelurahan Tanjung Gusta, khususnya Bapak Ade, yang telah memfasilitasi izin riset dan akses data. Penghargaan juga disampaikan kepada Ibu Marlina Setia Sinaga, S.Si., M.Si. selaku dosen pembimbing atas arahan dan bimbingannya. Teristimewa, rasa terima kasih dipersembahkan kepada orang tua penulis, Alm. Bapak Budi Saputra Rafael Tumanggor dan Ibu Demeria Marpaung, serta adik-adik tercinta (Audia, Ana, Aurel, dan Aceline) atas doa dan dukungan moral yang tak ternilai selama penyelesaian penelitian ini.

DAFTAR PUSTAKA

Arifin, A. A. A., Handoko, W., & Efendi, Z. (2022). Implementasi Metode Naive Bayes untuk Klasifikasi Penerima Program Keluarga Harapan. J-Com (Journal of Computer), 2(1), 21-26.

Badan Pusat Statistik Indonesia. (1 Juli 2024). Persentase Penduduk Miskin Maret 2024 turun menjadi 9,03 persen.

Badan Pusat Statistik Indonesia. (2 Desember 2024). Angka Partisipasi Sekolah (APS) Menurut Provinsi dan Kelompok Umur, 2024.

Fitriani, E. (2020). Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan. SISTEMASI: Jurnal Sistem Informasi, 9(1), 103-116.

Harsyiah, L., Hadijati, M., & Fitriyani, N. (2024). Perbandingan Analisis Diskriminan dan Naïve Bayes dalam Pengklasifikasian Status Penerima Bantuan Program Keluarga Harapan di NTB. Jurnal Matematika UNAND, 13(4), 296–308.

Kementerian Sosial Republik Indonesia. (2021). Pedoman Pelaksanaan Program Keluarga Harapan. Direktorat Jenderal Perlindungan dan Jaminan Sosial, Direktorat Jaminan Sosial Keluarga.

Nella, N. I., Setiawan, N. Y., & Ratnawati, D. E. (2022). Klasifikasi Penerima Bantuan Program Keluarga Harapan menggunakan Algoritme Decision Tree C4.5 (Studi Kasus: Desa Mlirip Kabupaten Mojokerto). Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 6(3), 1332–1339.

Noerkaisar, N. (2021). Efektivitas Penyaluran Bantuan Sosial Pemerintah untuk Mengatasi Dampak Covid-19 di Indonesia. Jurnal Manajemen Perbendaharaan, *2*(1), 83-104.

Pratama, F. K., Widodo, D. W., & Shofia, N. (2021). Implementasi Metode Naïve Bayes dalam Mengklasifikasi Penerima Program Keluarga Harapan (PKH) Desa Minggiran Kediri. Seminar Nasional Inovasi Teknologi UN PGRI Kediri, 23-28.

Ramadani, S., Sekar Ayu, N. Z., Nurhayati, N., Azzahra, F., & Windarto, A. P. (2020). Analisis Data Mining Naive Bayes Klasifikasi Pada Kelayakan Penerima PKH. KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer), 4(1), 374-381.

Sriwibowo, A. W. S., Alwi, A., & Sugianti, S. (2020). Implementation of Naïve Bayes Algorithm on Prediction of Acceptance of Keluarga Harapan Program (PKH). JICTE (Journal of Information and Computer Technology Education), 4(1), 1-7.

Tarigan, P. M. S., Hardinata, J. T., Qurniawan, H., Safii, M., & Winanjaya, R. (2022). Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus: Toko Sinar Harahap). Just IT: Jurnal Sistem Informasi dan Teknologi Informasi, 12(2), 51-61.

Wayuni, S. A., Pangemanan, S. E., & Singkoh, F. (2021). Kontribusi Program Keluarga Harapan dalam menunjang pendidikan siswa kurang mampu di Kecamatan Tuminting. Jurnal Governance, 1(2), 1-12.

Zahra, F., Khalif, A., & Sari, B. N. (2024). Pengelompokan Tingkat Kemiskinan di Setiap Provinsi di Indonesia Menggunakan Algoritma K-Medoids. Jurnal Informatika dan Teknik Elektro Terapan, 12(2), 1243-1249