

Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan (PKH) Kecamatan Gondanglegi Menggunakan Metode *Naïve Bayes*

Muhammad Rifat dan Tubagus Mohammad Akhriza

STMIK PPKIA Pradnya Paramita (STIMATA)

Jl. Laksda Adi Sucipto No.249a, Pandanwangi, Kec. Blimbing, Kota Malang, Jawa Timur 65126

E-mail: rifat_23510065@stimata.ac.id, akhriza@stimata.ac.id ac.id*)

Abstrak

Program Keluarga Harapan (PKH) merupakan program bantuan sosial penting untuk mengurangi kemiskinan di Indonesia, namun penyaluran yang tidak tepat sasaran menjadi kendala utama. Penelitian ini bertujuan meningkatkan akurasi klasifikasi penerima PKH di Kecamatan Gondanglegi menggunakan metode *Naïve Bayes*, dengan fokus pada data sosial ekonomi. Penelitian menggunakan algoritma Gaussian *Naïve Bayes*, melibatkan 1000 data penerima PKH dengan 18 atribut yang dikelompokkan ke dalam empat aspek kesehatan, pendidikan, kesejahteraan sosial, dan kemiskinan. Proses meliputi persiapan data, pemilihan atribut, pembagian data menjadi rasio 80:20, 70:30, dan 60:40, serta pengkodean variabel kategorikal menggunakan *One-Hot Encoding*. Evaluasi model dilakukan dengan *Confusion Matrix* untuk menghitung akurasi, presisi, recall, dan F1-score. Hasil menunjukkan model Gaussian *Naïve Bayes* dengan rasio data uji 80:20 memberikan akurasi tertinggi sebesar 96,5%. Model ini terbukti efektif dalam klasifikasi penerima bantuan PKH dengan tingkat ketepatan tinggi, sehingga dapat mendukung pengambilan keputusan yang lebih tepat sasaran oleh pemerintah daerah. Dengan demikian, implementasi metode ini dapat meningkatkan efisiensi dan efektivitas penyaluran bantuan sosial dalam program PKH.

Kata kunci : *Naïve Bayes*, PKH, klasifikasi, pembelajaran mesin, bantuan sosial.

Pendahuluan

Kemiskinan merupakan permasalahan utama dalam bidang ekonomi di negara Indonesia [1]. Ketimpangan ekonomi di setiap daerah menjadi pemicu adanya warga miskin dan warga sejahtera. Pemerintah Indonesia telah melakukan berbagai cara untuk mengentaskan kemiskinan di Indonesia. Melalui Kementerian Sosial Republik Indonesia telah meluncurkan Program Keluarga Harapan (PKH) sebagai salah satu program perlindungan sosial di Indonesia dalam bentuk bantuan sosial. Bantuan ini diberikan kepada keluarga miskin dan rentan miskin dengan persyaratan tertentu di mana mereka terdaftar dalam Data Terpadu Kesejahteraan Sosial (DTKS).

PKH merupakan salah satu upaya pemerintah dalam percepatan penanggulangan kemiskinan dan secara khusus bertujuan untuk memutus rantai kemiskinan antargenerasi [2]. Sejak diluncurkan pada tahun 2007, PKH telah berkontribusi dalam menekan angka kemiskinan dan mendorong kemandirian penerima bansos, yang selanjutnya disebut sebagai Keluarga Penerima Manfaat (KPM) [3].

Menurut Peraturan Menteri Sosial Nomor 1 Tahun 2018 terdapat beberapa komponen kriteria penerima manfaat PKM diantaranya komponen kesehatan yang menyangkut ibu hamil/nifas/menyusui dan anak usia dini, komponen pendidikan terdiri dari anak dengan usia 6 sampai 21 tahun yang belum menyelesaikan wajib belajar, yang menempuh tingkat pendidikan SD, SMP, SMA atau sederajat dan yang terakhir yaitu komponen kesejahteraan sosial yang menyangkut lanjut usia dan penyandang disabilitas [4].

Pelaksanaan Program Keluarga Harapan (PKH) di Kabupaten Malang dimulai sejak bulan Oktober 2013 dan berdasarkan Keputusan Direktur Jenderal Jaminan Sosial pada Kementerian Sosial Republik Indonesia tanggal 4 Februari 2014 Nomor : 22/LJS/02/2013 tentang Penetapan Kabupaten/Kota Lokasi Pengembangan PKH di Provinsi Pelaksana PKH Tahun 2013. Kecamatan Gondanglegi merupakan salah satu daerah yang merasakan dampak program tersebut.

Proses seleksi Keluarga Penerima Manfaat (KPM) pada Kecamatan Gondanglegi masih dilakukan secara manual yaitu dengan melakukan ver-

ifikasi dengan melakukan survey ke tiap-tiap desa kepada calon penerima program PKH, kemudian dilakukan pembahasan dalam forum musyawarah Desa (Musdes) untuk menentukan usulan tersebut. Kendala yang sering dialami yaitu tidak sedikit warga miskin yang seharusnya masuk ke dalam data warga miskin justru tidak masuk, begitu juga sebaliknya, banyak warga yang sudah sejahtera dan tergolong tidak miskin justru masih terdaftar sebagai warga miskin. Hal ini tentu saja sangat berpengaruh terhadap ketepatan sasaran program-program kemiskinan. Anggaran besar yang dikeluarkan untuk program-program penanggulangan kemiskinan menjadi sia-sia tatkala program tersebut diberikan kepada orang yang salah. Atau sebaliknya, orang yang seharusnya mendapat program pengentasan kemiskinan justru tidak pernah mendapat manfaat dari program tersebut.

Berdasarkan uraian permasalahan di atas, tujuan penelitian ini adalah untuk mengetahui akurasi data dari warga yang tidak mampu di Kecamatan Gondanglegi yang berhak untuk mendapatkan bantuan PKH dari pemerintah menggunakan metode Naïve Bayes serta dapat membantu penerimaan Keluarga Penerima Manfaat (KPM) menjadi lebih efektif dan meningkatkan efisiensi waktu.

Beberapa penelitian mengenai klasifikasi penerima bantuan sosial program kesejahteraan sosial (PKH) juga telah banyak diterapkan. Seperti sebuah studi yang ditulis oleh [5] menggunakan Naive Bayes untuk mengklasifikasikan penerima PKH di Kecamatan Bungaraya dengan 8 atribut menunjukkan akurasi hingga 99% dengan *Google Colab* dan 94% dengan *Rapid Miner*. Penelitian ini berhasil mengatasi ketidaktepatan sasaran akibat pemilihan manual. Hasil ini menunjukkan potensi penggunaan Naive Bayes untuk klasifikasi penerima bantuan yang lebih efektif dan akurat di berbagai wilayah.

Studi dari [5] dengan judul “Penerapan Algoritma *Naïve Bayes* dalam Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan” mengkaji ketidaktepatan sasaran penerima manfaat PKH di Desa Petataru Kecamatan Batubara yang bertujuan untuk mengatasinya. Penelitian ini menggunakan 100 data latih dan 20 data uji dengan 8 atribut. Saat menerapkan metode *Naive Bayes*, *Rapid Miner* dan *Confusion Matrix* memberikan tingkat akurasi 100% dan error 0%. Hasil tersebut menunjukkan bahwa *Naive Bayes* dapat meningkatkan akurasi penentuan penerima bantuan PKH.

Selanjutnya studi dari [6] dalam penelitian berjudul "Komparasi Efektivitas Algoritma C4.5 dan Naïve Bayes untuk Menentukan Kelayakan Penerima Manfaat Program Keluarga Harapan (studi kasus : kecamatan cicalengka kabupaten bandung) “ melakukan komparasi dua algoritma, yaitu C4.5 dan *Naïve Bayes*, dalam mengklasifikasikan kelayakan penerima bantuan PKH di Kecamatan Cicalengka, Kabupaten Bandung meng-

gunakan dataset sebanyak 774 dengan 17 atribut. Algoritma C4.5, yang merupakan algoritma Decision Tree, dan *Naïve Bayes*, yang berbasis probabilitas, diuji menggunakan metode CRISP-DM. Hasil evaluasi menunjukkan bahwa *Naïve Bayes* memiliki akurasi tertinggi sebesar 99,87% dan AUC 1,000, sedangkan algoritma C4.5 memiliki akurasi 99,61% dan AUC 0,743. Hal ini mengindikasikan bahwa *Naïve Bayes* lebih unggul dalam klasifikasi dengan hasil AUC yang menunjukkan klasifikasi sangat baik (*Excellent classification*) dibandingkan C4.5 yang hanya menghasilkan klasifikasi adil (*Fair*).

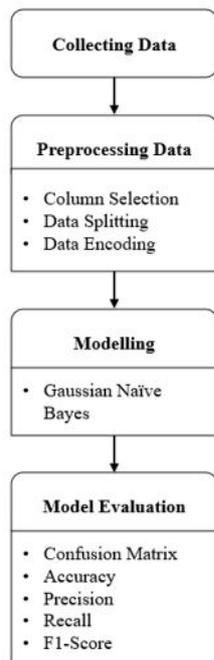
Dari berbagai penelitian yang telah disebutkan, jelas bahwa penggunaan algoritma *Naïve Bayes* dalam klasifikasi penerima bantuan PKH menunjukkan hasil yang menjanjikan. Hal ini memberikan gambaran bahwa tujuan melakukan klasifikasi dengan metode ini dapat menjadi alternatif yang efektif dalam meningkatkan akurasi dan efisiensi dalam penentuan penerima bantuan sosial.

Dengan latar belakang tersebut, penelitian ini akan fokus pada penerapan metode *Naïve Bayes* untuk menentukan kelayakan warga yang berhak menerima bantuan PKH di Kecamatan Gondanglegi yang menggunakan dataset 1000 yang terdiri dari 18 atribut yang akan diklasifikasi, serta menganalisis hasil yang diperoleh untuk memastikan bahwa bantuan sosial disalurkan kepada mereka yang benar-benar membutuhkan. Melalui penelitian ini, diharapkan dapat memberikan kontribusi nyata bagi upaya pemerintah dalam mengurangi angka kemiskinan dan meningkatkan kesejahteraan masyarakat.

Pemilihan metode *Naïve Bayes* untuk klasifikasi penerima PKH karena beberapa alasan utama. Pertama, metode ini sangat efisien secara komputasi, baik untuk melatih model maupun membuat prediksi, menjadikannya pilihan yang baik jika memiliki dataset besar atau keterbatasan sumber daya. Kedua, *Naïve Bayes* menunjukkan performa yang baik bahkan dengan data yang terbatas dan dapat menangani banyak fitur dengan efektif. Ketiga, meskipun didasarkan pada asumsi independensi yang sederhana, metode ini mudah diinterpretasikan, memungkinkan untuk memahami kontribusi setiap faktor dalam keputusan klasifikasi. Keempat, *Naïve Bayes* cenderung tidak mengalami *overfitting*, sehingga hasilnya lebih umum dan dapat diandalkan pada data baru.

Metode Penelitian

Ada enam tahapan dalam penelitian ini, yaitu: *Collecting Data*, *Preprocessing Data*, *Modelling* dan *Model Evaluation* sebagaimana di ilustrasikan pada Gambar 1.



Gambar 1: Tahapan Penelitian

Collecting Data

Pada tahap ini data yang dikumpulkan adalah data *By Name by Address* (BNBA) dari penerima manfaat pada Data Terpadu Kesejahteraan Sosial (DTKS) yang berada di aplikasi Siks-Ng. Pemilihan data pada proses seleksi data mencakup 21 atribut yang dikelompokkan menjadi 4 aspek yaitu aspek kemiskinan, aspek kesehatan, aspek pendidikan dan aspek kesejahteraan sosial yang ditampilkan pada Tabel 1.

Selanjutnya, data sebanyak 1000 penerima PKH diambil dan dipilih berdasarkan aspek-aspek yang tercantum pada Tabel 1.

Preprocessing Data

Data yang telah dikumpulkan masih dalam bentuk mentah dan perlu dilakukan preprocessing agar dapat diolah menggunakan metode *Naïve Bayes*. Tahapan *preprocessing* yang dilakukan meliputi seleksi kolom, pembagian data (*data splitting*), dan pengkodean data (*data encoding*).

Tabel 1: Hasil Uji Coba Perangkat

Kode	Atribut	Aspek
A1	NIK	
A2	Nama	
A3	Status	
A4	Memiliki Balita	Kesehatan
A5	Terdapat Ibu Hamil	
A6	Memiliki anak sekolah SD s/d SMA atau sederajat	Pendidikan
A7	Terdapat Lansia	Kesejahteraan Sosial
A8	Terdapat Disabilitas	
A9	Sudah / Tidak Bekerja	
A10	Status Kepemilikan Rumah	
A11	Pendapatan Perbulan	
A12	Terdapat Tempat Membuang Air Besar/Kecil	Kemiskinan
A13	Sebagian besar dinding terbuat dari bambu, kawat atau kayu	
A14	Sebagian besar lantai tempat tinggal terbuat dari tanah	
A15	Pernah khawatir tidak memiliki makanan untuk disantap	
A16	Melakukan kegiatan bekerja yang menghasilkan uang dalam seminggu terakhir	
A17	Membeli pakaian untuk diri sendiri atau keluarga dalam setahun terakhir	
A18	Tinggal bersama anggota keluarga lain	
A19	Sumber penerangan rumah dari listrik 450 watt ataupun bukan Listrik	
A20	Mempunyai tempat berteduh tetep sehari-hari	
A21	Sebagian besar pengeluaran digunakan untuk membeli makanan (lebih dari dua pertiga penghasilan)	

Column Selection

Dari total 21 atribut yang tersedia, hanya 18 atribut yang dipilih untuk diolah menggunakan metode *Naïve Bayes*. Dua variabel yang tidak diikutsertakan dalam analisis adalah Nama dan NIK, karena keduanya tidak relevan untuk klasifikasi. Selain itu, atribut Status ditetapkan sebagai atribut target, sedangkan atribut lainnya berfungsi seba-

gai fitur yang akan digunakan dalam proses klasifikasi/prediksi.

Data Splitting

Data splitting merupakan proses membagi dataset menjadi dua bagian, yakni data latih (*training data*) dan data uji (*testing data*). Data latih digunakan untuk melatih model, sementara data uji

digunakan untuk mengevaluasi kinerja model setelah pelatihan [7]. Dalam penelitian ini, telah dilakukan beberapa percobaan dengan berbagai rasio pembagian data, yaitu 80:20, 70:30, dan 60:40.

Dalam tahap ini juga menerapkan dua fungsi penting dalam proses ini yaitu *shuffle* dan *stratify*. Fungsi *shuffle* digunakan untuk mengacak urutan data sebelum melalui proses data splitting untuk menghindari bias yang mungkin timbul jika data disusun dalam urutan tertentu seperti berdasarkan waktu atau kategori, dengan mengacak data maka memastikan bahwa setiap sampel memiliki peluang yang sama untuk masuk ke dalam data training atau data *testing* yang membantu model belajar pola yang lebih umum dan tidak *overfit* pada urutan data tertentu. Di sisi lain, fungsi *stratify* sangat penting ketika berhadapan dengan *dataset* yang memiliki distribusi kelas yang tidak seimbang. Dengan menggunakan *stratify* dapat menjaga proporsi kelas yang sama dalam kedua *dataset*, sehingga model dapat dilatih dan diuji pada data yang representatif. Misalnya, jika *dataset* asli memiliki 70% sampel dari kelas A dan 30% dari kelas B, *stratify* akan memastikan bahwa proporsi ini dipertahankan dalam data *training* dan *testing*.

Kombinasi penggunaan *shuffle* dan *stratify* dalam data splitting membantu dalam menciptakan model *Naive Bayes* yang lebih akurat dengan menghindari bias dan memastikan representasi kelas yang konsisten. Tujuan penelitian ini adalah mencari pembagian yang menghasilkan kinerja terbaik untuk model *Naive Bayes*.

Data Encoding

Data encoding adalah langkah untuk mengubah format data agar dapat dipahami oleh algoritma *machine learning* [8]. Algoritma seperti *Naive Bayes* hanya bisa digunakan dengan data numerik, sehingga fitur-fitur kategorikal perlu diubah menjadi numerik.

Dalam penelitian ini menggunakan metode *one-hot encoding* karena semua fitur yang digunakan bersifat kategorik. *One-hot encoding* merupakan teknik untuk mengubah data kategorikal menjadi representasi biner di mana setiap kategori fitur akan diwakili oleh satu kolom biner (0 atau 1) [9]. Setiap kategori akan memiliki kolomnya sendiri, dan hanya satu kolom akan berisi angka 1, sedangkan kolom lainnya akan bernilai 0. Ini penting agar algoritma dapat memproses atribut kategorikal dengan benar tanpa memberikan urutan atau prioritas numerik di antara kategori tersebut.

Modelling

Setelah proses *preprocessing* selesai, langkah berikutnya adalah pemodelan. Pada tahap ini, algoritma yang digunakan adalah *Naive Bayes*.

Seperti yang dikemukakan ilmuwan Inggris yaitu Thomas Bayes bahwa *Naive Bayes* merupakan

pengklasifikasian dengan metode probabilitas dan statistik [10]. Menurut Olson Delen (2008), *Naive Bayes* memperhitungkan setiap jenis keputusan, dengan menghitung probabilitas tergantung pada asumsi bahwa atribut objek bersifat independen [11].

Sedangkan Klasifikasi *Naive Bayes* adalah suatu implementasi dari algoritma *Naive Bayes* yang digunakan untuk memprediksi kategori atau kelas dari suatu data dengan mengklasifikasikannya berdasarkan probabilitas [12].

Metode *Naive Bayes* yang memiliki penghitungan berdasarkan probabilitas hubungan antara atribut dan label tertinggi memiliki rumus (1) [13]:

$$P(H|x) = \frac{P(x|H)P(H)}{P(x)}, P(x) > 0 \quad (1)$$

Dimana $P(H|x)$ adalah Probabilitas bahwa hipotesis H benar (kelas) diberikan data x , $P(x|H)$ adalah Probabilitas data x diberikan bahwa hipotesis H benar, $P(H)$ adalah Probabilitas awal dari hipotesis H (prior probability) dan $P(x)$ adalah Probabilitas data x (normalisasi).

Keuntungan menggunakan metode *Naive Bayes* antara lain relatif mudah dalam penerapannya karena tidak menggunakan optimasi numerik, perhitungan matriks dan sejenisnya, efektif dalam pelatihan dan penggunaan, data biner atau polinomial dapat digunakan, dan dianggap independen, metode ini memungkinkan implementasi dengan banyak tipe kumpulan data yang berbeda [10].

Dalam penelitian ini, *Naive Bayes* diimplementasikan menggunakan *library scikit-learn* dalam Python untuk klasifikasi.

Model Evaluation

Setelah tahap pemodelan selesai, langkah berikutnya adalah evaluasi model. Tahap evaluasi merupakan tahap yang dilakukan untuk melihat seberapa baik kinerja algoritma klasifikasi yang digunakan dalam penelitian [7]. Evaluasi merupakan tahap penting dalam menilai kinerja model agar dapat diketahui bahwa tidak hanya baik pada data tertentu saja [14].

Saat mengevaluasi kinerja model klasifikasi, metrik penting seperti akurasi, presisi, *recall* dan *f1-score* digunakan untuk mengukur efektivitas model. Dalam menghitung metrik-metrik ini, digunakan alat yang disebut *confusion matrix*. *Confusion Matrix* adalah alat evaluasi yang sangat berguna untuk model klasifikasi, terutama dalam konteks klasifikasi biner seperti yang diterapkan dalam studi ini. Dengan membandingkan prediksi model dengan label kelas yang sebenarnya (dikenal sebagai "*ground truth*"), *Confusion Matrix* memberikan informasi tentang jumlah prediksi yang benar dan salah yang dibuat oleh model [15] seperti pada Tabel 2.

Tabel 2: *Confusion Matrix*

Kelas Sebenarnya	Kelas Prediksi		
		Positif	Negatif
	Positif	TP	FN
Negatif	FP	TN	

Berdasarkan Tabel 2, maka dapat dilakukan perhitungan akurasi, presisi, *recall* dan *f1-score*. Rumus (2) sampai (5) adalah masing-masing matriks tersebut :

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

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$Presisi = \frac{TP}{TP + FP} \quad (4)$$

$$F1Score = 2 * \frac{Presisi * recall}{Presisi + Recall} \quad (5)$$

Akurasi mengukur seberapa tepat prediksi model secara keseluruhan, presisi menilai seberapa akurat prediksi positif yang diberikan, *recall* melihat sejauh mana model dapat mendeteksi semua *instance* positif, dan *F1-score* memberikan keseimbangan antara presisi dan recall, terutama ketika ada ketidakseimbangan data [15]. Metrik-metrik ini memberikan gambaran menyeluruh tentang performa model dalam klasifikasi data penerima PKH.

Hasil dan Pembahasan

Data

Setelah tahap pemodelan selesai, langkah berikutnya Penelitian ini melibatkan 1000 data penerima Program Keluarga Harapan (PKH) di Kecamatan Gondanglegi, yang dipilih secara acak. Dataset terdiri dari 21 atribut, di mana terdapat 2 atribut identitas, yaitu "NIK" dan "NAMA", yang tidak digunakan dalam pemodelan. Selain itu, ada 1 atribut target yang disebut "Status" dengan dua nilai klasifikasi: Layak dan Tidak Layak. Sisa 18 atribut lain berperan sebagai fitur dan dikelompokkan ke dalam 4 aspek utama: Kesehatan, Pendidikan, Kesejahteraan Sosial, dan Kemiskinan. Tabel 1 t adalah ciri-ciri lengkap dari kumpulan data yang digunakan untuk pemodelan. Tipe data dan nilai kategori masing-masing atribut berdasarkan kode yang telah diberikan disajikan pada Tabel 3 yang mengacu pada Tabel 1.

Tabel 3: Karakteristik data yang digunakan

Jenis Atribut	Kode Atribut	Tipe	Nilai
Identitas	NIK	Numerik	-
	Nama	String	-
Target	Status	Kategorikal	"Layak" & "Tidak Layak"
	A1 – A5 dan A9 – A18	Kategorikal	"Ya" & "Tidak"
Fitur	A6	Kategorikal	"Tidak Bekerja" & "Sudah Bekerja"
	A7	Kategorikal	"Milik Sendiri", "Sewa" & "Menumpang"
	A8	Kategorikal	"Rendah", "Sedang" & "Tinggi"

Hasil *Preprocessing Data*

Hasil Data Encoding

Agar data dapat diolah oleh model *Naïve Bayes*, perlu dilakukan proses *encoding* pada variabel-variabel kategorik. Dalam penelitian ini, metode *encoding* yang digunakan adalah *one-hot encoding*. Berikut adalah contoh data sebelum dan sesudah dilakukan proses *encoding* dengan *one-hot* yang disajikan pada Tabel 4 dan Tabel 5.

Tabel 4: Data sebelum *encoding*

Status	Memiliki Balita
Layak	Ya
Tidak Layak	Tidak

Tabel 5: Data sesudah *encoding*

Status Layak	Status Tidak Layak	Memiliki Balita (Ya)	Memiliki Balita (Tidak)
1	0	1	0
0	1	0	1

Tabel sesudah *encoding* menunjukkan bagaimana setiap kategori diubah menjadi nilai biner (0 dan 1) untuk memudahkan model *machine learning* dalam memproses data kategorikal

Hasil *Data Splitting*

Tabel 6 menunjukkan jumlah data *training* dan *testing* berdasarkan rasio pembagian data 80:20, 70:30, dan 60:40 dengan total data 1000. Rasio pembagian data latih dan data uji untuk masing-masing skenario pemodelan disajikan pada Tabel 6.

Tabel 6: Rasio perbandingan *data training* dan *data testing*

Rasio	Jumlah Data Training	Jumlah Data Testing
80:20	800	200
70:30	700	300
60:40	600	400

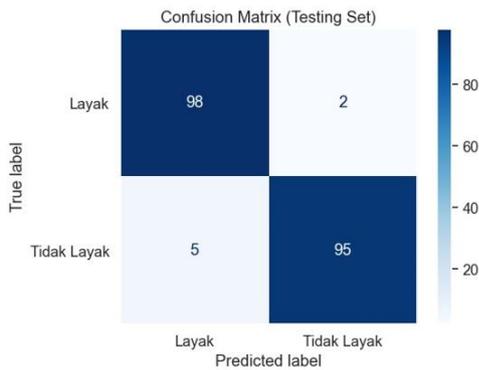
Pada Tabel 6, jumlah data *training* adalah bagian dari total data yang digunakan untuk melatih model, sedangkan data *testing* digunakan untuk menguji performa model.

Modeling Naïve Bayes

Pemodelan klasifikasi dilakukan dengan menggunakan *library Naïve Bayes Gaussian* dari *scikit learn*. Selanjutnya, data latih akan dimasukkan ke dalam fungsi *GaussianNB*. Setelah itu, akan dilakukan prediksi hasil dari data uji untuk menentukan probabilitas prediksi.

Hasil Evaluasi Model

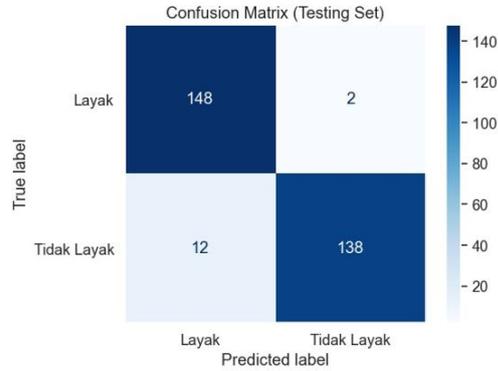
Setelah proses klasifikasi selesai, langkah selanjutnya adalah melakukan evaluasi untuk menilai performa model yang telah dibangun. Evaluasi ini bertujuan untuk melihat sejauh mana model dapat mengklasifikasikan data dengan benar. Dalam penelitian ini menggunakan beberapa metrik utama untuk mengevaluasi kinerja model, termasuk *Confusion Matrix*, akurasi, presisi, *recall*, dan *F1-score*. Metrik-metrik ini dihitung berdasarkan hasil klasifikasi pada data uji dengan rasio yang telah ditentukan sebelumnya.



Gambar 2: *Confusion Matrix data testing 20%*

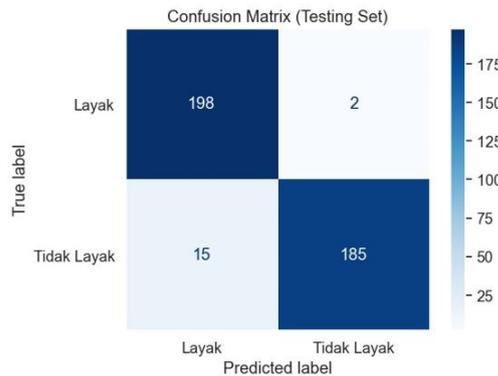
Gambar 2 menampilkan pengujian dengan data testing sebesar 20%, model menghasilkan hasil evaluasi dengan *Confusion Matrix* sebagai berikut: *True Positive* (TP) = 98, *True Negative* (TN) = 95, *False Positive* (FP) = 5, dan *False Negative* (FN) = 2. Dari hasil ini, dapat dilihat bahwa model mampu mengklasifikasikan 98 data dengan

benar sebagai layak menerima bantuan (TP), dan 95 data dengan benar sebagai tidak layak menerima bantuan (TN). Namun, terdapat 5 data yang diklasifikasikan salah sebagai layak menerima bantuan padahal sebenarnya tidak layak (FP), dan 2 data yang salah diklasifikasikan sebagai tidak layak padahal sebenarnya layak (FN).



Gambar 3: *Confusion Matrix data testing 30%*

Selanjutnya pada Gambar 3 menampilkan pengujian dengan data testing sebesar 30%, model menghasilkan hasil evaluasi dengan *Confusion Matrix* sebagai berikut: *True Positive* (TP) = 148, *True Negative* (TN) = 138, *False Positive* (FP) = 12, dan *False Negative* (FN) = 2. Dari hasil ini, dapat dilihat bahwa model mampu mengklasifikasikan 148 data dengan benar sebagai layak menerima bantuan (TP), dan 138 data dengan benar sebagai tidak layak menerima bantuan (TN). Namun, terdapat 12 data yang diklasifikasikan salah sebagai layak menerima bantuan padahal sebenarnya tidak layak (FP), dan 2 data yang salah diklasifikasikan sebagai tidak layak padahal sebenarnya layak (FN).



Gambar 4: *Confusion Matrix data testing 40%*

Selanjutnya pada Gambar 4 menunjukkan pengujian dengan data testing sebesar 40%, model menghasilkan hasil evaluasi dengan *Confusion Matrix* sebagai berikut: *True Positive* (TP) = 198,

True Negative (TN) = 185, *False Positive* (FP) = 15, dan *False Negative* (FN) = 2. Dari hasil ini, dapat dilihat bahwa model mampu mengklasifikasikan 198 data dengan benar sebagai layak menerima bantuan (TP), dan 185 data dengan benar sebagai tidak layak menerima bantuan (TN). Namun, terdapat 15 data yang diklasifikasikan salah sebagai layak menerima bantuan padahal sebenarnya tidak layak (FP), dan 2 data yang salah diklasifikasikan sebagai tidak layak padahal sebenarnya layak (FN). Ringkasan hasil evaluasi performa model untuk masing-masing rasio data uji disajikan pada Tabel 7.

Tabel 7: *Performance data testing*

Rasio Data Uji	Akurasi	Presisi	Recall	F1-Score
80:20	96,5%	96,54%	96,5%	96,49%
70:30	95,33%	95,53%	95,53%	95,32%
60:40	95,75%	95,94%	95,75%	95,74%

Berdasarkan pengujian dengan berbagai rasio data uji, model evaluasi menunjukkan kinerja yang konsisten dan sangat baik. Pada rasio 80:20, model mencapai akurasi, presisi, *recall*, dan *F1-score* masing-masing sebesar 96,5%, yang mencerminkan kemampuan model dalam mengklasifikasikan data dengan tingkat kesalahan yang sangat rendah.

Selanjutnya, pada rasio 70:30, meskipun porsi data uji lebih besar, model masih menghasilkan akurasi sebesar 95,33%, dengan presisi dan *recall* masing-masing 95,53%, serta *F1-score* 95,32%, yang menunjukkan keseimbangan yang baik antara prediksi positif dan negatif yang benar.

Pada rasio 60:40, model berhasil mempertahankan performanya dengan akurasi 95,75%, presisi 95,94%, *recall* 95,75%, dan *F1-score* 95,74%. Hasil ini menunjukkan model tetap kokoh meskipun persentase data uji ditingkatkan, memberikan hasil yang konsisten dalam berbagai skenario pengujian.

Selain hasil evaluasi model berdasarkan *confusion matrix* dan metrik klasifikasi, perlu dijelaskan pula bagaimana proses kerja model *Naïve Bayes* dalam menghasilkan prediksi klasifikasi pada *dataset* yang digunakan. Berikut adalah analisis dan interpretasi dari model *Naïve Bayes* yang diterapkan pada penelitian ini.

Model *Naïve Bayes* bekerja dengan menghitung probabilitas suatu data masuk ke dalam kelas tertentu, yaitu Layak atau Tidak Layak, berdasarkan atribut-atribut yang dimiliki. Setiap atribut pada data, seperti memiliki balita, ibu hamil, dan pendapatan, memiliki peluang kemunculan dalam masing-masing kelas. Model kemudian mengalikan probabilitas dari seluruh atribut tersebut dengan probabilitas awal kelas (*prior probability*), untuk menentukan kelas dengan kemungkinan tertinggi.

Sebagai contoh, jika terdapat data penerima PKH dengan ciri memiliki balita, ibu hamil, dan pendapatan rendah, model akan menghitung peluang data tersebut termasuk dalam kelas Layak atau

Tidak Layak berdasarkan frekuensi kemunculan ciri-ciri tersebut di data pelatihan. Jika probabilitas masuk ke kelas Layak lebih tinggi dibandingkan Tidak Layak, maka data tersebut akan diprediksi sebagai Layak.

Melalui pendekatan ini, model *Naïve Bayes* tidak hanya memberikan hasil prediksi, tetapi juga memperlihatkan bagaimana kontribusi masing-masing atribut memengaruhi klasifikasi. Analisis ini melengkapi hasil evaluasi model, sehingga tidak hanya diketahui kinerja model secara keseluruhan, tetapi juga bagaimana data digunakan dalam proses klasifikasi.

Penutup

Berdasarkan penelitian yang telah dilakukan, kesimpulannya adalah model klasifikasi menggunakan algoritma *Naïve Bayes* berhasil diterapkan untuk menentukan kelayakan penerima bantuan PKH di Kecamatan Gondanglegi. Melalui proses *preprocessing* dan evaluasi model, hasil evaluasi menunjukkan performa model yang baik dengan tingkat akurasi, presisi, *recall*, dan *F1-score* yang tinggi. Meskipun semua rasio menunjukkan performa yang baik, rasio 80:20 memberikan hasil terbaik di semua metrik. Secara keseluruhan, rasio data uji sangat mempengaruhi akurasi model klasifikasi, memilih rasio yang tepat adalah langkah penting dalam pengembangan model yang efektif dan dapat diandalkan untuk meningkatkan efisiensi dan efektivitas dalam menentukan penerima bantuan yang tepat sasaran.

Hasil prediksi dapat digunakan oleh pemerintah daerah sebagai panduan dalam pengambilan keputusan terkait penentuan penerima bantuan sosial. Dengan demikian, penelitian ini turut berperan membantu pemerintah dalam menangani ketidakakuratan dalam penyaluran bantuan serta meningkatkan kesejahteraan masyarakat.

Daftar Pustaka

- [1] A. Fatmawati, A. Irma Purnamasari, dan I. Ali, "Implementasi *Naïve Bayes* Untuk Klasifikasi Kelayakan Penerima Bantuan Sosial," *Jurnal Mahasiswa Teknik Informatika*, Vol. 8, No. 1, pp. 745–750, doi: 10.36040/jati.v8i1.8714, Januari 2024.
- [2] S. Sulfadli, G. Susanti, M. T. Abdullah, dan R. Pauzi, "Evaluasi Dampak Program: Studi Kasus Program Keluarga Harapan (PKH) di Kabupaten Enrekang," *Dev. Policy Manag. Rev.*, Vol. 3, No. 1, pp. 1–20, doi: 10.61731/dpmr.v3i1.26674, 2023.
- [3] A. H. Ramadhan dan A. Hidir, "Lansia Penerima Program Keluarga Harapan Di Desa Batang Tumu Kecamatan Mandah Kabupaten

- Indragiri Hilir,” *Cross-border*, Vol. 4, No. 1, pp. 166–180, 2021.
- [4] Anonim, “Peraturan Menteri Sosial Republik Indonesia Nomor: 1 Tahun 2018 Tentang Program Keluarga Harapan,” Kementerian Sosial Republik Indonesia, 2018. diakses daring pada [https://peraturan.bpk.go.id/Home/Download/120868/PERMENSOS NOMOR 1 TAHUN 2018.pdf](https://peraturan.bpk.go.id/Home/Download/120868/PERMENSOS%20NOMOR%201%20TAHUN%202018.pdf).
- [5] A. Irsyada, E. Haerani, M. Irsyad, F. Wulandari, dan L. Afriyanti, “Penerapan Algoritma Naïve Bayes Terhadap Klasifikasi Penerima Bantuan Program Keluarga Harapan (PKH),” *Jurnal Sistem Komputer dan Informatika*, Vol. 5, No. 2, p. 457, doi: 10.30865/json.v5i2.7203, 2023.
- [6] A. Nur, A. Rohim, A.I. Purnamasari, dan I. Ali, “Komparasi Efektifitas Algoritma C4.5 dan Naïve Bayes untuk Menentukan Kelayakan Penerima Manfaat Program Keluarga Harapan (Studi Kasus: Kecamatan Cicalengka Kabupaten Bandung),” *Jurnal Mahasiswa Teknik Informatika*, Vol. 8, No. 2, pp. 2355–2362, 2024.
- [7] M.S. Haris, A.N. Khudori, dan W.T. Kusuma, “Perbandingan Metode Supervised Machine Learning untuk Prediksi Prevalensi Stunting di Provinsi Jawa Timur,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, Vol. 9, No. 7, p. 1571, doi: 10.25126/jtiik.2022976744, 2022.
- [8] E.P. Febtiawan, L.A. Syamsul, I. Akbar, dan A.S. Rachman, “Forecasting Produksi Energi Photovoltaic Menggunakan Algoritma Random Forest Classification,” *Journal of Information Systems Research*, Vol. 5, No. 4, pp. 1053–1062, doi: 10.47065/josh.v5i4.5514, 2024.
- [9] G.P. Insany, P. Somantri, dan A. Putri, “Implementasi Sistem Rekomendasi dengan Content Based Filtering dan Teknologi Virtual Tour untuk Strategi Pemasaran pada Website,” *Technol. Sci.*, Vol. 6, No. 1, pp. 300–313, doi: 10.47065/bits.v6i1.5358, 2024.
- [10] N. Wikayati, “Model Pendeteksi Kematangan Buah Apel dengan Metode Naive Bayes Classifier dan Sensor Warna TCS3200 Berbasis ArduinoNANO,” *Dinamika Dotcom Jurnal Pengembangan Manajemen Informasi dan Komputer*, Vol. 13, pp. 149–161, 2022.
- [11] F. K. Pratama, D. W. Widodo, dan N. Shofia, “Implementasi Metode Naïve Bayes dalam Mengklasifikasi Penerima Program Keluarga Harapan (PKH) Desa Minggiran Kediri,” *Semin. Nas. Inov. Teknol. UN PGRI Kediri*, pp. 23–28, 2021.
- [12] Windy, D. Rudiaman Sijabat, dan F. Eka Purwiantono, “Implementasi Algoritma Naïve Bayes untuk Memprediksi Lama Masa Studi dan Predikat Kelulusan Mahasiswa,” *Jurnal Dinamika DOTCOM*, Vol. 10, No. 1, pp. 19–28, 2019.
- [13] M.F. Essy Rahma Meilaniwati, “Klasifikasi Penduduk Miskin Penerima PKH Menggunakan Metode Naïve Bayes dan KNN,” *Jurnal Kajian dan Terapan Matematika*, Vol. 8, No. 2, pp. 75–84, 2022.
- [14] Kaka Kamaludin, Woro Isti Rahayu, dan M.Y. Helmi Setywan, “Transfer Learning to Predict Genre Based on Anime Posters,” *Jurnal Teknologi Informasi*, Vol. 4, No. 5, pp. 1041–1052, 2023, doi: 10.52436/1.jutif.2023.4.5.860.
- [15] A. Jalil, A. Homaidi, dan Z. Fatah, “Implementasi Algoritma Support Vector Machine untuk Klasifikasi Status Stunting pada Balita,” *G-Tech Jurnal Teknologi Terapan*, Vol. 8, No. 3, pp. 2070–2079, doi: 10.33379/gtech.v8i3.4811, 2024.