

# KLASIFIKASI KELAYAKAN PENERIMA BANTUAN PROGRAM KELUARGA HARAPAN DI KECAMATAN SINAR PENINJAUAN MENGGUNAKAN METODE NAÏVE BAYES

Devi Wulan Safitri<sup>\*1</sup>, Evi Fadilah<sup>2</sup>, Sri Rahayu<sup>3</sup>

<sup>1,2,3</sup>Universitas Islam Negeri Raden Fatah Palembang, Prodi Sistem Informasi  
e-mail : <sup>1</sup>deviulansapitri02@gmail.com, <sup>2</sup>Evifadilah\_uin@radenfatah.ac.id,  
<sup>3</sup>sriahayu@radenfatah.ac.id

**Abstrak** - Program Keluarga Harapan (PKH) merupakan bantuan sosial bersyarat yang ditujukan kepada keluarga kurang mampu. Namun, masih ditemukan ketidaktepatan dalam penentuan penerima bantuan, khususnya di Kecamatan Sinar Peninjauan. Penelitian ini bertujuan untuk mengklasifikasikan kelayakan penerima PKH menggunakan algoritma Naive Bayes Classifier berdasarkan data sekunder sebanyak 1.840 data dari tahun 2020 sampai 2024 yang mencakup 45 variabel. Hasil klasifikasi menunjukkan 323 individu (87,77%) layak menerima bantuan, sedangkan 45 individu (12,23%) tidak layak. Evaluasi model menggunakan 5-Fold Cross Validation menunjukkan akurasi tertinggi 100%, dengan presisi dan recall juga mencapai 100%. Sementara itu, evaluasi menggunakan Confusion Matrix menunjukkan akurasi 98,08%, recall 93,06%, dan presisi 91,50%. Temuan ini dapat menjadi dasar bagi Dinas Sosial dalam meningkatkan akurasi penentuan penerima PKH secara lebih objektif dan tepat sasaran.

**Kata Kunci:** Program Keluarga Harapan (PKH), Klasifikasi Kelayakan, Keluarga Penerima Manfaat (KPM), Naive Bayes.

## I. PENDAHULUAN

### A. Latar Belakang

Kemiskinan masih menjadi persoalan krusial di Indonesia sebagai negara berkembang. Meskipun pemerintah telah meluncurkan berbagai program bantuan sosial seperti BPNT, BLT, dan Program Keluarga Harapan (PKH), dampaknya belum optimal dalam menurunkan angka kemiskinan secara signifikan [1].

Program Keluarga Harapan (PKH) adalah sebuah inisiatif pemerintah Indonesia yang diadopsi dari keberhasilan program serupa di wilayah Amerika Latin dalam mengatasi kemiskinan yang bersifat kronis [2]. sebagai salah satu program andalan pemerintah yang dimulai sejak 2007, ditujukan untuk membantu keluarga prasejahtera melalui skema bantuan bersyarat. Namun, dalam implementasinya masih ditemukan berbagai

permasalahan, khususnya terkait ketidaktepatan sasaran penerima bantuan [3].

Di Kecamatan Sinar Peninjauan, masih ditemukan warga yang tidak layak justru menerima bantuan, sementara yang memenuhi kriteria tidak terdata sebagai penerima. Hal ini disebabkan oleh rendahnya akurasi pendataan, musyawarah desa yang subjektif, serta kurangnya pemahaman masyarakat terhadap indikator kelayakan sebagai Keluarga Penerima Manfaat (KPM). Dalam konteks ini, muncul pertanyaan utama: Bagaimana hasil klasifikasi kelayakan penerima bantuan PKH di Kecamatan Sinar Peninjauan jika dianalisis menggunakan metode *Naive Bayes* Dan sejauh mana tingkat akurasi metode tersebut dalam mendukung penentuan penerima bantuan secara objektif dan tepat sasaran.

Tujuan dari penelitian ini adalah untuk mengklasifikasikan kelayakan penerima bantuan PKH di Kecamatan Sinar Peninjauan dengan memanfaatkan metode *Naive Bayes*, serta mengevaluasi akurasinya melalui pengujian dengan teknik *Cross Validation*. Algoritma *Naive Bayes* dipilih karena kemampuannya dalam menangani data kategorik maupun numerik, serta efisiensinya dalam membangun model klasifikasi meskipun dengan data berukuran sedang [4].

Penelitian ini memiliki kontribusi signifikan dalam bidang sistem informasi, khususnya pada penerapan teknologi data mining dalam sektor pelayanan sosial. Berbeda dari studi sebelumnya yang berfokus pada perbandingan model klasifikasi atau penerapan dalam pinjaman kredit [5]. Penelitian ini menekankan pentingnya akurasi klasifikasi dalam konteks penyaluran bantuan sosial. Dengan pendekatan berbasis machine learning, diharapkan hasil penelitian ini dapat memberikan rekomendasi sistem pendukung keputusan bagi instansi sosial dalam meningkatkan efisiensi dan transparansi penyaluran bantuan sosial.

### B. Tinjauan Pustaka

#### 1. Data dan Data Mining

Data merupakan elemen fundamental dalam setiap proses analisis informasi dan pengambilan keputusan. Secara umum, data didefinisikan sebagai kumpulan fakta

atau nilai yang diperoleh melalui observasi, pengukuran, atau pencatatan terhadap suatu objek atau fenomena tertentu [6]. Data dapat berbentuk angka, teks, simbol, maupun atribut kategorikal yang mencerminkan karakteristik dari objek yang dikaji.

Data mining merupakan proses eksplorasi terhadap kumpulan data berskala besar guna menemukan pola, hubungan tersembunyi, serta pengetahuan yang bermakna. Proses ini menggabungkan teknik statistik, pembelajaran mesin (*machine learning*), dan kecerdasan buatan (AI) untuk mengekstraksi informasi yang sebelumnya tidak diketahui [7].

Tujuan utama data mining adalah membantu pengambilan keputusan secara lebih tepat dan berbasis data. Teknik ini digunakan secara luas dalam berbagai bidang, termasuk bisnis, keuangan, kesehatan, dan sektor publik. Proses data mining terdiri dari beberapa tahap, antara lain seleksi data, pembersihan data (*preprocessing*), transformasi, eksplorasi pola melalui algoritma tertentu (seperti klasifikasi, klusterisasi, atau asosiasi), serta interpretasi hasil [8].

Dalam praktiknya, data mining dapat dibagi menjadi dua pendekatan utama: deskriptif dan prediktif. Pendekatan deskriptif bertujuan untuk menggambarkan karakteristik atau struktur data, sementara pendekatan prediktif digunakan untuk memprediksi nilai atau kelas berdasarkan data historis [9]. Salah satu teknik prediktif yang paling umum digunakan adalah klasifikasi, di mana data dimasukkan ke dalam kelas tertentu berdasarkan model yang telah dilatih menggunakan dataset sebelumnya. Terdapat sejumlah teknik yang umum diterapkan dalam proses data mining yaitu Deskripsi, klasifikasi, klusterisasi, kaidah Asosiasi [10].

## 2. Knowledge Discovery in Database (KDD)

*Knowledge Discovery in Database* (KDD) adalah proses sistematis yang bertujuan untuk mengekstraksi pengetahuan baru dari kumpulan data berukuran besar. KDD tidak hanya mengacu pada proses penambangan data (*data mining*), tetapi mencakup keseluruhan tahapan mulai dari seleksi data, prapemrosesan, transformasi, penambangan data, hingga evaluasi dan interpretasi hasil [11].

Secara umum, proses KDD terdiri dari lima tahapan utama. Pertama, *selection*, yaitu memilih data yang relevan dari keseluruhan database. Kedua, *preprocessing*, yaitu membersihkan data dari duplikasi, noise, atau inkonsistensi. Ketiga, *transformation*, yaitu mengubah data ke dalam format yang sesuai untuk analisis lebih lanjut. Keempat, *data mining*, yaitu inti dari proses KDD di mana digunakan berbagai metode seperti klasifikasi, klusterisasi, dan asosiasi untuk menemukan pola-pola tersembunyi. Terakhir, *interpretation/evaluation*, yaitu mengevaluasi pola yang ditemukan dan menyajikannya dalam bentuk yang dapat dimengerti dan bermanfaat untuk pengambilan keputusan [12], [13]

Meskipun istilah *data mining* sering digunakan secara bergantian dengan KDD, secara teknis, data mining hanyalah satu bagian dari keseluruhan proses KDD. Oleh karena itu, dalam implementasi sistem berbasis data, penting untuk tidak hanya fokus pada pemilihan algoritma, tetapi juga memastikan kualitas data melalui

tahap-tahap awal KDD agar hasil analisis menjadi valid dan dapat diandalkan [14].

Penerapan KDD sangat relevan dalam sistem informasi modern, terutama dalam pengolahan data sosial seperti penyaluran bantuan sosial. Melalui pendekatan KDD, proses klasifikasi penerima bantuan dapat dilakukan secara objektif dan efisien karena berbasis data yang telah diproses dan dianalisis secara menyeluruh [10].

## 3. Program Keluarga Harapan

Program Keluarga Harapan (PKH) merupakan salah satu program bantuan sosial bersyarat (*conditional cash transfer*) yang diluncurkan oleh pemerintah Indonesia sejak tahun 2007. Tujuan utama dari program ini adalah untuk meningkatkan kesejahteraan masyarakat prasejahtera melalui dukungan finansial yang mendorong akses terhadap layanan pendidikan, kesehatan, dan kesejahteraan sosial lainnya [18]. PKH berfokus pada kelompok rentan seperti ibu hamil, anak usia dini, pelajar, serta lansia dan penyandang disabilitas berat.

Agar bantuan yang diberikan tepat sasaran, Kementerian Sosial menetapkan sejumlah kriteria bagi calon penerima yang dikenal sebagai Keluarga Penerima Manfaat (KPM). Penentuan KPM mengacu pada data terpadu kesejahteraan sosial (DTKS), namun dalam praktiknya seringkali masih terdapat ketidaksesuaian antara data dan kondisi riil di lapangan [1]. Hal ini diperparah oleh minimnya transparansi dalam proses verifikasi serta keterbatasan pemahaman masyarakat terhadap indikator kelayakan penerima bantuan [19].

Distribusi bantuan yang tidak tepat sasaran dapat menimbulkan ketimpangan sosial dan mengurangi efektivitas program. Oleh karena itu, dibutuhkan pendekatan berbasis data yang mampu mengevaluasi dan mengklasifikasikan kelayakan penerima secara objektif. Pemanfaatan metode klasifikasi dalam data mining, seperti *Naive Bayes*, dapat membantu pemerintah dalam mengidentifikasi KPM berdasarkan variabel-variabel sosial ekonomi secara lebih akurat dan efisien [17].

Penelitian sebelumnya menunjukkan bahwa integrasi teknologi informasi dalam program sosial seperti PKH berpotensi memperbaiki proses seleksi penerima bantuan, serta meningkatkan akuntabilitas dan efisiensi penyaluran bantuan. Dengan pendekatan ini, PKH dapat berfungsi lebih efektif sebagai instrumen pengentasan kemiskinan yang berkelanjutan.

## 4. Metode Naive Bayes

*Naive Bayes* merupakan salah satu algoritma klasifikasi yang berbasis pada teori probabilitas Bayes. Algoritma ini mengasumsikan bahwa setiap atribut dalam data bersifat independen satu sama lain terhadap kelas yang diprediksi. Meskipun asumsi ini jarang terpenuhi sepenuhnya dalam data dunia nyata, *Naive Bayes* tetap menunjukkan performa klasifikasi yang tinggi dalam berbagai kasus karena kesederhanaan dan efisiensinya [20].

Secara matematis, algoritma ini menghitung probabilitas posterior  $P(C|X)$ , yaitu probabilitas suatu data  $X$  termasuk dalam kelas  $C$ , berdasarkan distribusi probabilitas atribut terhadap kelas tersebut. Persamaan dasarnya merujuk pada Teorema Bayes:

$$P(X|H) = \frac{P(X|Ci) \cdot P(Ci)}{P(X)}$$

di mana P(C) adalah probabilitas awal kelas, P(X|C) adalah probabilitas data X muncul dalam kelas C, dan P(X) adalah probabilitas data X secara keseluruhan [9].

*Naive Bayes* memiliki keunggulan dalam hal kecepatan pelatihan, efisiensi memori, serta kemampuan menangani dataset berukuran besar dengan banyak atribut. Algoritma ini juga tahan terhadap noise dan bekerja cukup baik bahkan ketika data tidak memenuhi asumsi independensi secara sempurna.

Dalam konteks klasifikasi kelayakan penerima bantuan sosial seperti Program Keluarga Harapan (PKH), metode *Naive Bayes* terbukti mampu mengolah data dengan banyak variabel sosial ekonomi secara efektif. Beberapa studi terdahulu menunjukkan bahwa akurasi model ini dapat mencapai tingkat tinggi, khususnya ketika digunakan bersama teknik validasi silang (*Cross-Validation*) dan preprocessing yang optimal.

### 5. Cross Validasi

*Cross Validation* merupakan salah satu teknik evaluasi model dalam pembelajaran mesin (*machine learning*) yang digunakan untuk mengukur kemampuan generalisasi dari suatu algoritma terhadap data yang belum pernah dilihat sebelumnya. Teknik ini membagi dataset menjadi beberapa subset untuk memastikan bahwa setiap data digunakan sebagai data latih dan data uji secara bergantian, sehingga hasil evaluasi menjadi lebih objektif dan menghindari overfitting [21].

Salah satu bentuk umum dari *Cross Validation* adalah *K-Fold Cross Validation*, di mana data dibagi menjadi *k* bagian yang sama. Model akan dilatih sebanyak *k* kali, masing-masing menggunakan *k-1* bagian untuk pelatihan dan 1 bagian untuk pengujian. Nilai evaluasi akhir diperoleh dari rata-rata performa di semua *fold*, yang mencerminkan kestabilan dan keandalan model [22]. Dalam konteks penelitian ini, digunakan *5-Fold Cross Validation* untuk mengukur akurasi metode *Naive Bayes* dalam mengklasifikasikan kelayakan penerima bantuan PKH.

Teknik ini dipilih karena mampu memberikan estimasi performa model yang lebih konsisten dibandingkan dengan metode pembagian data sederhana seperti *holdout*

*validation*. Selain itu, *Cross Validation* juga cocok diterapkan pada dataset berukuran sedang, seperti data penerima bantuan sosial yang digunakan dalam penelitian ini.

### 6. Confusion Matrix

*Confusion Matrix* merupakan alat evaluasi performa model klasifikasi yang digunakan untuk mengukur tingkat akurasi berdasarkan perbandingan antara hasil prediksi dan kelas sebenarnya. Matriks ini berbentuk tabel dua dimensi yang menyajikan jumlah prediksi benar dan salah, serta mengklasifikasikannya ke dalam empat kategori: *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN) [23].

Tabel 1. Confusion Matrix

		Actual Visual	
		True	False
Prediction Values	True	TP (True Positive) Correct Result	FP(False Positive) Unexpected Result
	False	FN (False Negative) Missing Result	TN(True Negative) Correct Absence Of R

Dari *Confusion Matrix*, pengukuran *Accuracy*, *Recall*, *Precision* dari kelas positif dihitung menggunakan Persamaan berikut:

$$Accuracy$$

$$Recall$$

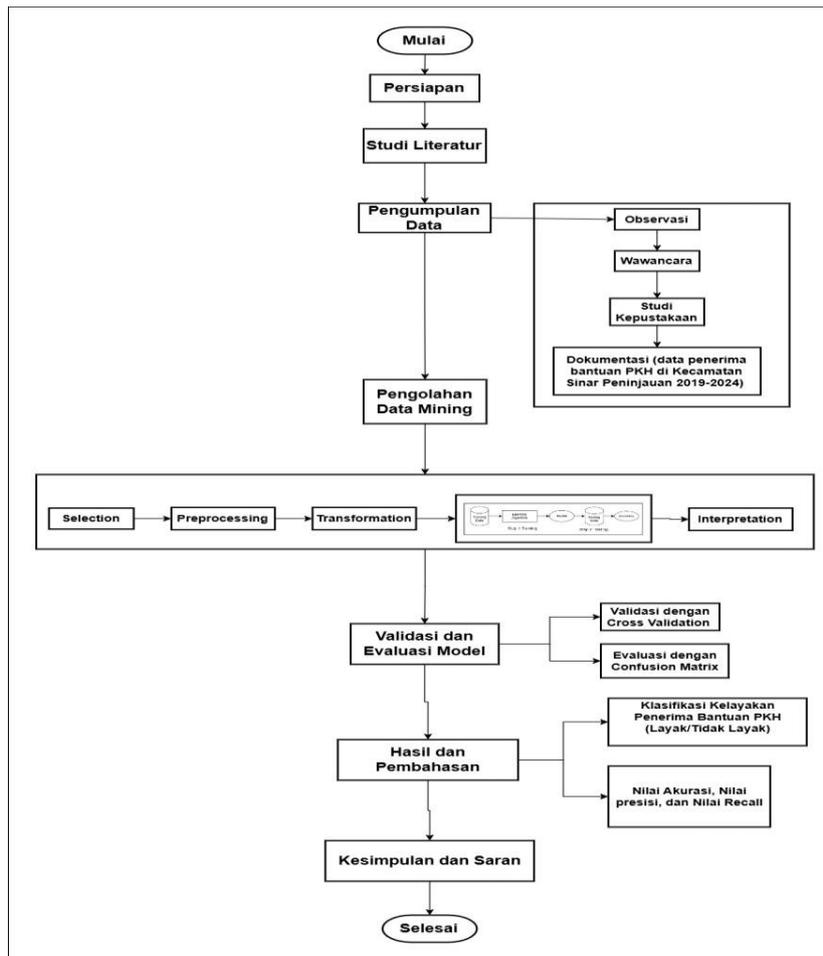
$$Precision$$

$$= \frac{TP}{TP + FP} \times 100\%$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\%$$

## II. METODE PENELITIAN

Penelitian ini dilaksanakan melalui tujuh tahapan utama yang disusun secara sistematis guna mencapai tujuan studi secara efektif, mulai dari persiapan hingga penarikan kesimpulan dan pemberian saran. Setiap tahapan dirancang untuk menghasilkan klasifikasi kelayakan penerima bantuan Program Keluarga Harapan (PKH) secara akurat dan berbasis data.



Gambar 1. Tahapan Penelitian

1. Persiapan

Tahap awal mencakup identifikasi masalah, penetapan ruang lingkup studi, dan penyusunan rencana kerja. Peneliti menentukan fokus masalah terkait ketidaktepatan distribusi bantuan PKH, khususnya di Kecamatan Sinar Peningjauan.

2. Studi Literatur

Literatur dikaji secara mendalam untuk memahami konsep data mining, algoritma *Naive Bayes*, serta indikator kelayakan PKH. Referensi diambil dari buku, jurnal, dan penelitian terdahulu untuk membangun dasar teoretis yang kuat.

3. Pengumpulan Data

Data dikumpulkan melalui observasi, wawancara dengan petugas PKH, studi pustaka, dan dokumentasi dari Dinas Sosial. Data yang digunakan mencakup 1.840 data penerima bantuan selama periode 2020–2024.

4. Pengolahan Data Mining (Tahapan KDD)

Pengolahan dilakukan menggunakan pendekatan *Knowledge Discovery in Database (KDD)*, yang mencakup:

- a. *Selection*: pemilihan atribut penting,
- b. *Preprocessing*: pembersihan dan validasi data,
- c. *Transformation*: konversi data ke format numerik,
- d. *Data Mining*: penerapan algoritma *Naive Bayes* untuk klasifikasi,
- e. *Interpretation/Evaluation*: penilaian terhadap pola klasifikasi yang ditemukan.

5. Validasi dan Evaluasi Model

Model divalidasi menggunakan teknik *5-Fold Cross Validation*, dan dievaluasi dengan *Confusion Matrix* yang menghasilkan nilai akurasi, presisi, recall, dan F1-score. Evaluasi ini mengukur performa model dalam mengklasifikasikan data secara akurat dan konsisten.

6. Hasil dan Pembahasan

Hasil klasifikasi menunjukkan distribusi data ke dalam kategori “Penerima” dan “Bukan Penerima”. Evaluasi metrik menunjukkan model memiliki kinerja tinggi, dengan interpretasi mendalam terhadap faktor-faktor yang mempengaruhi kelayakan penerima bantuan.

7. Kesimpulan dan Saran

Kesimpulan diambil berdasarkan kinerja model *Naive Bayes* yang terbukti efektif dalam klasifikasi kelayakan PKH. Saran diberikan untuk pengembangan sistem yang lebih adaptif dan berbasis data guna mendukung pengambilan keputusan dalam penyaluran bantuan sosial yang lebih adil dan tepat sasaran.

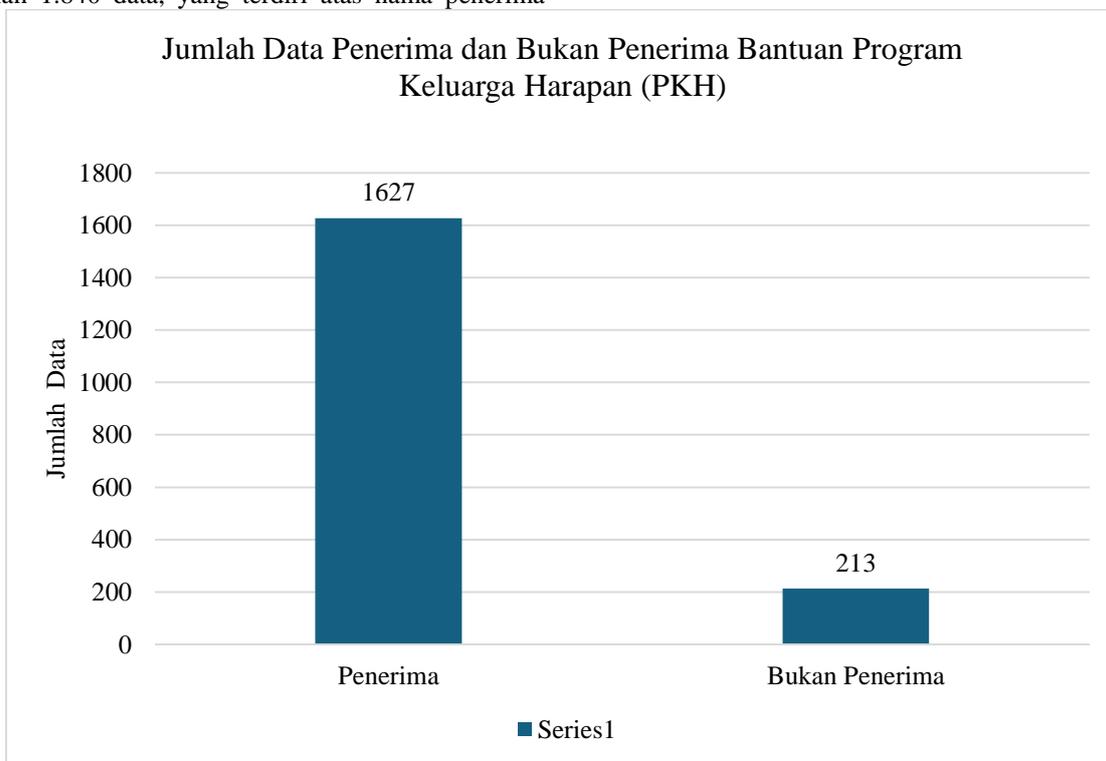
III. HASIL DAN PEMBAHASAN

A. Data Selection

Penelitian ini menggunakan data hasil rekapitulasi kuesioner Program Perlindungan Sosial (PPLS) yang mencakup beberapa program bantuan, antara lain Program Keluarga Harapan (PKH), Bantuan Pangan Non-Tunai (BPNT), Kartu Indonesia Pintar (KIP), dan Kartu Indonesia Sehat (KIS). Fokus utama penelitian tertuju

pada data penerima bantuan Program Keluarga Harapan (PKH) di Kecamatan Sinar Peninjauan, yang dikumpulkan selama periode 2020 hingga 2024. Jumlah Dataset yang digunakan dalam penelitian ini adalah berjumlah 1.840 data, yang terdiri atas nama penerima

dan bukan penerima bantuan dengan 45 variabel atribut yang digunakan. Dari total tersebut, terdiri 1.627 data nama penerima bantuan dan 213 data nama bukan penerima bantuan.



Gambar 1. Jumlah Data Penerima dan Bukan Penerima Bantuan Program Keluarga Harapan (PKH) Tahun 2020-2024

Sebelum masuk tahap data mining, dilakukan proses penyaringan untuk menghapus variabel yang tidak relevan atau tidak memberikan kontribusi signifikan terhadap klasifikasi. Pemilihan variabel dilakukan berdasarkan indikator resmi dari Badan Pusat Statistik (BPS) yang berkaitan langsung dengan kriteria kelayakan penerima bantuan. Setelah proses seleksi atribut, jumlah variabel disaring menjadi 24 variabel utama yang dinilai paling relevan, sementara 21 variabel lainnya dieliminasi karena kurang mendukung proses klasifikasi. Dataset yang telah terseleksi ini menjadi dasar dalam membangun model klasifikasi kelayakan penerima bantuan PKH. Daftar tabel variabel yang digunakan dalam klasifikasi pada tabel 2.

Tabel 2. Daftar dan Keterangan Variabel

No	Variabel/Atribut	Keterangan
1.	Jumlah Keluarga	Jumlah Orang Yang Tinggal Dalam Rumah
2.	Jumlah Tanggungan	Jumlah Tanggungan Yang Tinggal Dalam Rumah (Ibu Hamil/Balita, Anak Sekolah, Lansia/Penyandang Disabilitas)
3.	Penghasilan Perbulan	Penghasilan Dalam Sebulan Yang Didapatkan
4.	Bekerja Atau Tidak Bekerja	Keterangan Bekerja atau Tidak Bekerja
5.	Jenis Pekerjaan	Jenis Pekerjaan sehari-hari yang Dilakukan
6.	Status Penguasa Bangunan	Status Penguasaan Bangunan Tempat Tinggal yang Ditempati
7.	Jenis Lantai Terluas	Jenis Lantai Terluas Yang Di Gunakan Tempat Tinggal
8.	Jenis Dinding Terluas	Jenis Dinding Terluas Yang Di Gunakan Tempat Tinggal
9.	Kondisi Dinding	Kondisi Dinding Tempat Tinggal
10.	Jenis Atap Terluas	Jenis Atap Terluas Yang Di Gunakan Tempat Tinggal

No	Variabel/Atribut	Keterangan
11.	Kondisi Atap	Kondisi Atap Tempat Tinggal
12.	Sumber Air Minum	Sumber Air Minum Sehari-hari
13.	Sumber Penerangan	Sumber Penerangan Utama Yang Di Gunakan
14.	Daya Listrik Terpasang	Daya Listrik Yang Terpasang
15.	Bahan Bakar Masak	Bahan Bakar Untuk Memasak
16.	Fasilitas Pembuangan Air Besar	Penggunaan Fasilitas Tempat Buang Air Besar
17.	Tempat Pembuangan Akhir Tinja	Tempat Pembuangan Akhir Tinja
18.	Kepemilikan Tabung Gas 5,5 Kg atau Lebih	Rumah Tangga Memiliki Asset Tabung Gas 5,5 KG atau lebih
19.	Kepemilikan Lemari Es/Kulkas	Rumah Tangga Memiliki Asset Lemari Es/Kulkas
20.	Kepemilikan TV	Rumah Tangga Memiliki Asset TV
21.	Jumlah Kepemilikan Sepeda	Rumah Tangga Memiliki Asset Sepeda
22.	Jumlah Kepemilikan Motor	Rumah Tangga Memiliki Asset Motor
23.	Jumlah Kepemilikan Mobil	Rumah Tangga Memiliki Asset Mobil
24.	Keterangan - PKH	Keterangan: 1. Penerima 2. Bukan Penerima

Dataset penerima bantuan PKH dari tahun 2020-2024 dan yang akan dijadikan sebagai Data Training 80% dan Data Testing 20% dapat dilihat pada Tabel 3.

Tabel 3. Dataset dari tahun 2020-2024

No	Jumlah Keluarga	Jumlah Tanggungan	Penghasilan Perbulan	...	Keterangan PKH
1.	4	2	> 1 Juta Per Bulan < UMK	...	Penerima
2.	4	2	< 1 Juta Per Bulan	...	Penerima
3.	4	2	< 1 Juta Per Bulan	...	Penerima
4.	3	2	UMK	...	Penerima
5.	1	0	Tidak ada Penghasilan	...	Penerima
6.	4	2	> 1 Juta Per Bulan < UMK	...	Penerima
7.	4	3	UMK	...	Penerima
8.	2	1	> 1 Juta Per Bulan < UMK	...	Penerima
9.	2	1	< 1 Juta Per Bulan	...	Penerima
10.	2	1	> UMK - 10 Juta Per Bulan	...	Bukan Penerima
11.	3	2	> 1 Juta Per Bulan < UMK	...	Penerima
12.	4	2	< 1 Juta Per Bulan	...	Penerima
13.	3	2	< 1 Juta Per Bulan	...	Penerima
14.	5	3	> 1 Juta Per Bulan < UMK	...	Penerima
15.	6	4	> 1 Juta Per Bulan < UMK	...	Penerima
16.	4	2	< 1 Juta Per Bulan	...	Penerima
17.	3	1	> UMK - 10 Juta Per Bulan	...	Bukan Penerima
...	...	...	.....	...	.....
18 40.	3	2	> 1 Juta Per Bulan < UMK	...	Penerima

**B. Preprocessing**

Sebelum memasuki tahap utama dalam proses data mining, dilakukan tahapan preprocessing untuk memastikan kualitas data yang digunakan dalam kondisi optimal. Salah satu tahap penting dalam preprocessing adalah *data cleaning*, yaitu proses pembersihan data dari berbagai potensi kesalahan seperti nilai yang tidak konsisten, data tidak valid, atau kesalahan input. Berdasarkan hasil pemeriksaan menyeluruh terhadap dataset, tidak ditemukan adanya data duplikat, data kosong, maupun kesalahan lainnya. Oleh karena itu, seluruh data hasil seleksi sebanyak 1.840 data dapat digunakan secara utuh dalam proses klasifikasi tanpa ada penghapusan data tambahan.

**C. Transformation**

Tahapan transformasi data dilakukan setelah proses penyaringan dan pembersihan data selesai, dengan tujuan untuk mengonversi beberapa tipe data numerik menjadi kategorikal agar sesuai dengan kebutuhan analisis menggunakan metode klasifikasi data mining. Transformasi ini bertujuan untuk meningkatkan efektivitas pengolahan data dan mengadaptasikan format data terhadap algoritma yang digunakan. Beberapa variabel penting dikonversi berdasarkan acuan dari studi sebelumnya.

Beberapa contoh transformasi yang dilakukan antara lain:

1. Menentukan Nilai Kategori Jumlah keluarga daat dilihat pada Tabel 4.

Tabel 4. Nilai Jumlah Keluarga

No	Nilai Jumlah Keluarga	Kategori
1.	≤ 2	Rendah
2.	3 – 4	Sedang
3.	≥ 5	Tinggi

2. Menentukan Nilai Kategori Jumlah Tanggungan daat dilihat pada Tabel 5.

Tabel 5. Jumlah Tanggungan

No	Nilai Jumlah Tanggungan	Kategori
1.	<1	Rendah
2.	2	Sedang
3.	>3	Tinggi

3. Menentukan Nilai Kategori Penghasilan Per Bulan daat dilihat pada Tabel 6.

Tabel 6. Penghasilan Per Bulan

No	Nilai Penghasilan Per Bulan	Kategori
1.	≤ 1.000.000	Rendah
2.	1.000.000-3.000.000	Sedang
3.	≥ 3.100.000	Tinggi

4. Menentukan Nilai Kategori Jumlah Kepemilikan Sepeda daat dilihat pada Tabel 7.

Tabel 7. Kepemilikan Sepeda

No	Nilai Jumlah Kepemilikan Sepeda	Kategori
1.	0	Tidak
2.	1	Ya

5. Menentukan Nilai Kategori Jumlah Kepemilikan Motor daat dilihat pada Tabel 8.

Tabel 8. Jumlah Kepemilikan Motor

No	Nilai Jumlah Kepemilikan Motor	Kategori
1.	≤ 1	Rendah
2.	2 – 3	Sedang
	≥ 4	Tinggi

6. Menentukan Nilai Kategori Jumlah Kepemilikan Mobil daat dilihat pada Tabel 9.

Tabel 9. Kepemilikan Mobil

No	Nilai Jumlah Kepemilikan Motor	Hasil Transformasi
1.	0	Tidak
2.	1	Ya

**D. Data Mining (Naïve Bayes)**

Setelah proses klasifikasi dilakukan menggunakan algoritma *Naive Bayes* melalui perangkat lunak RapidMiner, diperoleh sejumlah hasil evaluasi yang mencerminkan kinerja model dalam mengklasifikasikan kelayakan penerima bantuan. Dataset dibagi menjadi dua bagian, yaitu 80% data pelatihan (1.472 data) dan 20% data pengujian (368 data), untuk menguji keakuratan dan efektivitas model.

Hasil evaluasi menunjukkan bahwa model menghasilkan tingkat akurasi sebesar 98,10%, yang mencerminkan tingkat kecocokan prediksi model terhadap data aktual secara keseluruhan. Selain itu, diperoleh nilai rata-rata recall sebesar 93,02%, yang

menunjukkan kemampuan model dalam mengenali data yang benar-benar layak menerima bantuan. Sementara itu, nilai rata-rata precision mencapai 90,91%, yang mengindikasikan tingkat ketepatan model dalam memprediksi bahwa suatu data termasuk dalam kategori layak menerima bantuan.

*E. Interpretation dan Evaluation*

Untuk menguji stabilitas dan konsistensi kinerja model klasifikasi berbasis algoritma Naive Bayes, dilakukan metode 5-Fold Cross Validation. Dataset dibagi menjadi lima bagian yang masing-masing bergiliran digunakan sebagai data uji, sementara empat bagian lainnya sebagai data latih. Hasil dari setiap fold menunjukkan performa model yang sangat baik, sebagaimana dijabarkan berikut:

Tabel 10. Hasil Evaluasi model menggunakan 5-Fold Cross Validation

No	Fold	Akurasi (%)	Presisi (%)	Recall (%)
1.	1	100.00	100.00	100.00
2.	2	99.46	97.14	97.14
3.	3	99.46	97.78	97.50
4.	4	95.37	85.06	77.78
5.	5	96.73	91.92	91.11
	Rata-rata	98.20	94.78	92.31

Secara umum, hasil validasi silang ini dengan menggunakan 5-Fold Cross Validation menunjukkan bahwa pada fold pertama diperoleh nilai akurasi, presisi dan recall masing-masing sebesar 100%. Perhitungan ini menandakan bahwa pada subset data tersebut, model klasifikasi mampu mengenali pola data dengan sangat baik tanpa menghasilkan kesalahan prediksi dan membuktikan bahwa model Naive Bayes yang dibangun memiliki konsistensi kinerja tinggi, serta mampu beradaptasi dengan baik terhadap variasi data pada setiap fold. Evaluasi ini memperkuat keandalan model dalam mengklasifikasikan kelayakan penerima bantuan Program Keluarga Harapan (PKH).

Evaluasi lanjutan terhadap model klasifikasi dilakukan dengan menggunakan Confusion Matrix guna memperoleh gambaran yang lebih rinci mengenai performa algoritma Naive Bayes, khususnya pada model dengan tingkat akurasi tertinggi sebesar 98,08%. Analisis ini bertujuan untuk mengukur sejauh mana model mampu mengklasifikasikan data secara akurat, serta mendeteksi kesalahan prediksi yang terjadi. Berdasarkan hasil pengujian, diketahui bahwa:

Tabel 11. Hasil Evaluasi Model Menggunakan Confusion Matrix

Accuracy : 98.08%				
No	Prediksi/Aktual	True Penerima	True Bukan Penerima	Class Precision
1.	Pred. Penerima	321	3	99.07%
2.	Pred. Bukan Penerima	4	40	90.91%
	Class Recall	98.77%	93.02%	

- a. Jumlah total data uji: 368 data
- b. Jumlah prediksi benar: 361 data
- c. Jumlah prediksi salah: 7 data

Hasil tersebut menunjukkan bahwa model memiliki tingkat akurasi yang sangat tinggi, dengan tingkat kesalahan prediksi yang minimal. Selain itu, nilai recall dan precision yang dihasilkan dari Confusion Matrix juga menunjukkan performa yang kuat, memperkuat temuan bahwa model mampu mengidentifikasi kategori data secara efektif dan tepat sasaran.

Visualisasi hasil Confusion Matrix, termasuk distribusi nilai true positive, false positive, true negative, dan false negative, disajikan pada gambar pendukung, yang memberikan ilustrasi komprehensif atas ketepatan dan keandalan klasifikasi.

Secara keseluruhan, hasil evaluasi ini mengonfirmasi bahwa algoritma Naive Bayes sangat efektif dalam mengklasifikasikan kelayakan penerima bantuan Program Keluarga Harapan (PKH), dengan tingkat kesalahan yang dapat diterima dalam konteks pengambilan keputusan berbasis data.

IV. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa penerapan algoritma Naive Bayes Classifier efektif dalam mengklasifikasikan kelayakan penerima bantuan Program Keluarga Harapan (PKH) di Kecamatan Sinar Peninjauan. Dengan menggunakan data sekunder sebanyak 1.840 data nama penerima bantuan dari periode 2020 hingga 2024, yang menunjukkan bahwa penerapan metode Naive Bayes dengan validasi 5-Fold Cross Validation berhasil mengklasifikasi kelayakan penerima bantuan Program Keluarga Harapan (PKH) dengan akurasi rata-rata yang tinggi. Dari hasil klasifikasi, sebanyak 87,77% nama penerima bantuan yang teridentifikasi layak menerima bantuan, sementara 12,23% dikategorikan tidak layak.

Hasil evaluasi model klasifikasi menggunakan Confusion Matrix mampu menghasilkan akurasi evaluasi sebesar 98,08%, recall 93,06%, dan precision 91,50%, yang mengaskan keandalan kinerja prediksi model dalam pengambilan keputusan berbasis data yang sangat tinggi dan stabil.

Model berhasil mengelompokkan masyarakat ke dalam dua kategori utama, yaitu “Penerima Bantuan” dan “Bukan Penerima Bantuan”, berdasarkan atribut-atribut sosial ekonomi yang relevan. Hasil ini menunjukkan bahwa metode klasifikasi berbasis data mining dapat dijadikan alternatif dalam memperbaiki sistem verifikasi kelayakan yang selama ini masih dilakukan secara manual dan subjektif.

Sebagai saran, penulis merekomendasikan agar instansi terkait, khususnya Dinas Sosial, mempertimbangkan pengembangan sistem informasi klasifikasi kelayakan berbasis algoritma Naive Bayes sebagai pendukung pengambilan keputusan. Selain itu, penelitian lanjutan disarankan untuk menguji komparasi dengan algoritma lain dan memperluas variabel yang digunakan, guna menghasilkan sistem klasifikasi yang lebih adaptif, akurat, dan berkelanjutan dalam distribusi bantuan sosial.

DAFTAR PUSTAKA

[1] D. Ispriyanti, A. Prahutama, and Mustafid, “Klasifikasi Kemiskinan Di Kota Semarang Menggunakan Algoritma Chisquare Automatic Interaction Detection

- (CHAID) Dan Classification And Regression Tree (CART),” *MEDIA STATISTIKA*, vol. 12, no. 1, pp. 63–72, Jul. 2019, doi: 10.14710/medstat.12.1.63-72.
- [2] S. M. Ulfah and P. Ratnasari, “Implementasi Program Keluarga Harapan,” *Jurnal Ilmu Sosial, Politik dan Pemerintah*, vol. 4, no. 1, pp. 56–67, 2021.
- [3] C. Fadlan, S. Ningsih, and A. P. Windarto, “Penerapan Metode Naive Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra,” *JUTIM*, vol. 3, no. 1, pp. 1–8, Jun. 2018.
- [4] Meilana, Y. Astuti, I. R. Wulandari, I. Sulistyowati, and B. A. M., “Algoritma Naive Bayes untuk Mengklasifikasikan Kepribadian Siswa SMP Berdasarkan Tipologi Hippocrates-Galenus,” *SISTEMASI: Jurnal Sistem Informasi*, vol. 10, pp. 480–489, May 2021, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [5] A. Bagja, K. Kusriani, and M. R. Arief, “Komparasi Algoritma Naive Bayes Dan Support Vector Machine (SVM) Untuk Klasifikasi Kelayakan Pemberian Pinjaman,” *Infotek: Jurnal Informatika dan Teknologi*, vol. 6, no. 2, pp. 513–523, Jul. 2023, doi: 10.29408/jit.v6i2.20059.
- [6] H. S. Situmorang and M. lufti, *ANALISIS DATA untuk Riset Manajemen dan Bisnis*, 3rd ed. USU Press, 2014. [Online]. Available: <https://www.researchgate.net/publication/353072388>
- [7] S. Adinugroho and Y. A. Sari, *Implementasi Data Mining Menggunakan Weka*. Universitas Brawijaya Press, 2018. Accessed: Jan. 21, 2025. [Online]. Available: [https://books.google.co.id/books/about/Implementasi\\_Data\\_Mining\\_Menggunakan\\_Weka.html?hl=id&id=p91qDwAAQBAJ&redir\\_esc=y](https://books.google.co.id/books/about/Implementasi_Data_Mining_Menggunakan_Weka.html?hl=id&id=p91qDwAAQBAJ&redir_esc=y)
- [8] Noviyanto, “Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia,” *Paradigma - Jurnal Informatika dan Komputer*, vol. 22, no. 2, pp. 183–188, Sep. 2020, doi: 10.31294/p.v21i2.
- [9] Suyanto, *Data Mining Untuk Klasifikasi Dan Klusterisasi Data, Revisi*. Bandung: Informatika Bandung, 2019.
- [10] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Deepublish, 2020.
- [11] Kusriani and E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta: Andi, STMIK Amikom, 2009. Accessed: Jan. 21, 2025. [Online]. Available: <https://books.google.co.id/books?id=-Ojclag73O8C&printsec=frontcover#v=onepage&q&f=false>
- [12] N. L. W. S. R. Ginantra et al., *Data Mining dan Penerapan Algoritma*. Yayasan Kita Menulis, 2021.
- [13] R. A. KPM Ramadhani and LiceFrense, K-Nears Neighbours Risa Helilintar, Risky Aswi Ramadhani Siti Rochana. Python “Belajar Pemrograman Python Dasar,” vol. 84. 2013. Accessed: Jan. 21, 2025. [Online]. Available: [http://www.askjansen.com/wp-content/uploads/2014/04/Kontroversi-Kalori\\_ebook.pdf%0Ahttp://ir.obihiro.ac.jp/dspace/handle/1032/2/3933](http://www.askjansen.com/wp-content/uploads/2014/04/Kontroversi-Kalori_ebook.pdf%0Ahttp://ir.obihiro.ac.jp/dspace/handle/1032/2/3933)
- [14] D. T. Larose and C. D. Larose, *Discovering Knowledge in Data a An Introduction to Data Mining*. In D. T. Larose (Ed.), *Discovering Knowledge in Data (Second Edi)*. Hoboken: John Wiley&Sons,Inc,Hoboken, 2014. doi: 10.1002/9781118874059.
- [15] J. Han, M. Kamber, and J. Pei, *Data Mining: Data Mining Concepts and Techniques, ThirdEdit*. In Elsevier, 2012. Accessed: Jan. 23, 2025. [Online]. Available: <https://doi.org/10.1109/ICMIRA.2013.45> dan <https://books.google.co.id/books?id=pQws07tdpjoC&printsec=copyright&hl=id#v=onepage&q&f=false>
- [16] A. F. Firdaus, R. Saedudin, and R. Andeswati, “Implementation of Naive Bayes Classification Method in Predicting Student Graduation,” *E-Proceeding of Engineering*, vol. 8, no. 5, pp. 9274–9279, 2021.
- [17] H. N. F. Fikrillah, S. Hudawiguna, and C. Juliane, “Klasifikasi Penerima Bansos Menggunakan Algoritma Naive Bayes,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, no. 1, pp. 683–695, Mar. 2023, [Online]. Available: <http://jurnal.mdp.ac.id>
- [18] Kementerian Sosial Republik Indonesia, “Pedoman Pelaksanaan Program Keluarga Harapan (PKH),” *Kemendrian Sosial RI*, Jakarta, 2021.
- [19] A. P. Giovani, Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *Jurnal Teknoinfo*, vol. 14, no. 2, pp. 116–124, Jul. 2020, doi: 10.33365/jti.v14i2.679.
- [20] R. Wati, “Penerapan Algoritma Naive Bayes Dan Particle Swarm Optimization Untuk Klasifikasi Berita Hoax Pada Media Sosial,” *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, vol. 5, no. 2, pp. 9–14, Feb. 2020, [Online]. Available: [www.bsi.ac.id](http://www.bsi.ac.id)
- [21] R. T. Handayanto and Herlawati, *Data Mining dan Machine Learning Menggunakan Matlab dan Python*. Bandung: Informasi Bandung, 2020.
- [22] T. Jo, *Text Mining: Concepts, Implementation, and Big Data Challenge*. Cham: Springer, 2019.
- [23] S. Visa, B. Ramsay, A. Ralescu, and E. Der Knap, “Edited by Sofia Visa, Atsushi Inoue, and Anca Ralescu,” *Maics*, vol. 710, pp. 120–127, 2011.