

# Penerapan Algoritma Naïve Bayes Terhadap Klasifikasi Penerima Bantuan Program Keluarga Harapan (PKH)

Amelia Irsyada, Elin Haerani\*, Muhammad Irsyad, Fitri Wulandari, Liza Afriyanti

Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: <sup>1</sup>11950121670@students.uin-suska.ac.id, <sup>2\*</sup>elin.haerani@uin-suska.ac.id, <sup>3</sup>irsyadtech@uin-suska.ac.id,

<sup>4</sup>fitri\_wulandari@uin-suska.ac.id, <sup>5</sup>liza.afriyanti@uin-suska.ac.id

Email Penulis Korespondensi: elin.haerani@uin-suska.ac.id

Submitted: 20/12/2023; Accepted: 31/12/2023; Published: 31/12/2023

**Abstrak**—Kemiskinan di Indonesia merupakan salah satu masalah sosial yang kompleks. Sebagai bentuk kepedulian pemerintah terhadap kasus kemiskinan di Indonesia, maka pemerintah membuat program-program bantuan yang ditujukan untuk masyarakat miskin. Salah satu program yang pemerintah jalankan untuk mengentas kemiskinan di Indonesia yaitu Program Keluarga Harapan (PKH). PKH merupakan program bantuan sosial bersyarat yang diberikan kepada masyarakat miskin. Proses seleksi penerima bantuan yang masih dilakukan secara manual menjadi kurang ideal dan terdapat beberapa isu tentang pembagian bantuan yang tidak tepat sasaran, maka pada penelitian ini akan menerapkan algoritma Naïve Bayes terhadap klasifikasi penerima bantuan PKH di Kecamatan Bungaraya, Kabupaten Siak. Data yang digunakan sebanyak 560 data, yang merupakan data penerima bantuan PKH serta data calon penerima dari beberapa desa yang terdapat di Kecamatan Bungaraya tahun 2022. Atribut yang digunakan pada penelitian ini yaitu usia, penghasilan, jumlah tanggungan, tanggungan anak sekolah, tanggungan disabilitas, status rumah, jenis lantai dan jenis dinding. Nilai akurasi terbaik yang diperoleh melalui perhitungan yang dilakukan pada Google Colab adalah 99% untuk perbandingan 80:20. Sedangkan nilai akurasi yang dilakukan pada Rapid Miner sebesar 94%.

**Kata Kunci:** Kemiskinan; Program Keluarga Harapan; Data Mining; Klasifikasi; Algoritma Naïve Bayes

**Abstract**—Poverty in Indonesia is one of the complex social issues. As a manifestation of the government's concern about poverty in the country, various assistance programs have been established to target the impoverished population. One such program aimed at alleviating poverty in Indonesia is the Family Hope Program (Program Keluarga Harapan or PKH). PKH is a conditional cash transfer program provided to the impoverished community. The manual selection process for aid recipients is considered less than ideal, leading to issues of improper distribution. In this study, the Naïve Bayes algorithm is applied to classify PKH aid recipients in the Bungaraya Subdistrict, Siak Regency, as part of the government's efforts to tackle poverty. The dataset used consists of 560 records, including data on existing PKH aid recipients and potential recipients from various villages in the Bungaraya Subdistrict for the year 2022. The attributes considered in this research include age, income, number of dependents, dependents attending school, dependents with disabilities, housing status, floor type, and wall type. The highest accuracy obtained through calculations on Google Colab is 99% for an 80:20 ratio, while the accuracy obtained using RapidMiner is 94%.

**Keywords:** Poverty; Family of Hope Program; Data Mining; Classification; Naïve Bayes Algorithm

## 1. PENDAHULUAN

Kemiskinan merupakan salah satu masalah yang banyak dan rumit di Indonesia. Meski Indonesia memiliki potensi ekonomi yang besar dan kaya akan sumber daya alam (SDA), banyak masyarakat yang masih hidup di bawah garis kemiskinan. Jumlah penduduk miskin di Indonesia meningkat sebanyak 0,16 juta orang di perkotaan pada September 2022, meningkat dari 11,82 juta orang pada Maret 2022 menjadi 11,98 juta orang pada September 2022, menurut data yang tercatat di Badan Pusat Statistik (BPS). Pada saat yang sama, jumlah penduduk miskin di perdesaan meningkat sebesar 0,04 juta jiwa (dari 14,34 juta jiwa pada Maret 2022 menjadi 14,38 juta jiwa pada September 2022)[1]. Sebagai bentuk kepedulian pemerintah terhadap kasus kemiskinan di Indonesia, maka pemerintah membuat program-program bantuan yang ditujukan untuk masyarakat miskin. Salah satu program yang dilaksanakan pemerintah untuk mengentaskan kemiskinan di Indonesia adalah Program Keluarga Harapan (PKH).

PKH adalah program bantuan sosial bersyarat untuk orang yang tidak mampu[2][3]. Peraturan Menteri Sosial (Permensos) RI No.1 2018 terkait Program Keluarga Harapan menyatakan bahwa Program Keluarga Harapan (PKH) adalah program bersyarat yang ditujukan untuk memberikan bantuan sosial kepada keluarga atau individu miskin atau rentan yang terdaftar dalam data terpadu program penanggulangan kemiskinan, diolah oleh pusat data dan informasi perlindungan sosial dan ditetapkan sebagai Program Keluarga Harapan (PKH). Komponen pendidikan, kesejahteraan sosial, dan kesehatan termasuk dalam kriteria PKH[4]. Dalam jangka pendek, diharapkan tujuan PKH dapat mengurangi beban rumah tangga (dampak konsumsi langsung), sedangkan dalam jangka panjang, program ini dapat menjadi investasi dalam kesehatan dan pendidikan yang lebih baik untuk generasi berikutnya (dampak pengembangan modal manusia)[5].

Program Keluarga Harapan sudah dicanangkan pemerintah sejak tahun 2007[6] mulai terlaksana di Kabupaten Siak sejak tahun 2014, dan terlaksanan di seluruh kecamatan yang ada di kabupaten Siak pada akhir 2015. Kecamatan Bungaraya merupakan salah satu daerah yang merasakan dampak program tersebut. Proses seleksi Keluarga Penerima Manfaat (KPM) dapat dilakukan bervariasi diberbagai daerah, namun pada Kecamatan Bungaraya, proses seleksi penerima bantuan masih dilakukan secara manual, yaitu dengan melakukan musyawarah

kampung. Dalam kenyataannya, hal itu menjadi kurang ideal. Berdasarkan wawancara yang dilakukan kepada koordinator PKH di Bungaraya, mengatakan bahwa, masih ada terdengar isu-isu pemberian bantuan yang tidak tepat sasaran. Untuk memberikan informasi tentang penerima bantuan PKH kepada pihak terkait, maka, peneliti akan menggunakan data mining untuk mengidentifikasi penerima bantuan Program Keluarga Harapan di Kecamatan Bungaraya.

Data mining adalah upaya untuk menemukan pola dalam suatu database dengan menggunakan metode seperti pengajaran mesin, statistik matematika, dan kecerdasan buatan[7]. Mengidentifikasi model (atau fungsi) yang menjelaskan dan membedakan kelas data atau konsep untuk dapat menggunakan model tersebut untuk memprediksi kelas objek yang label kelasnya tidak diketahui. Istilah "klasifikasi" mengacu pada proses ini. Model turunan bergantung pada analisis data latih[8]. Peneliti menggunakan algoritma klasifikasi Naïve Bayes. Berdasarkan teorema Bayes, metode klasifikasi Naïve Bayes berfokus pada asumsi yang sangat kuat (naif) bahwa setiap kondisi atau kejadian independen[9]. Metode Naïve Bayes merupakan metode bayes yang sangat disederhanakan[10].

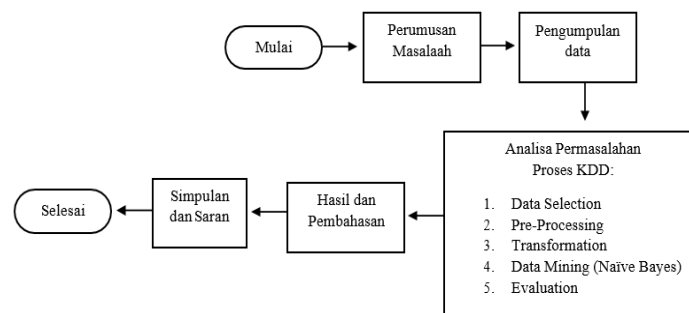
Pada penelitian sebelumnya yang dilakukan oleh Nurul Alfiah, dkk (2021), dengan tulisannya yang berjudul "Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Algoritma Naïve Bayes" mendapatkan kesimpulan nilai akurasi pada confusion matrix sebesar 84,2411% yang termasuk dalam kategori Good Classification[11]. Selain itu, Algoritma Naïve Bayes juga digunakan untuk berbagai penelitian pada kasus lain, seperti Klasifikasi kebakaran hutan dan lahan yang dilakukan oleh Trya Ayu Pratiwi, dkk, (2021) dalam penelitiannya yang berjudul "Klasifikasi Kebakaran Hutan dan Lahan Menggunakan Algoritma Naïve Bayes di Kabupaten Pelalawan" mendapatkan kesimpulan bahwa suhu, curah hujan, kecepatan angin dan kelembaban yang menjadi faktor paling mempengaruhi penyebab kebakaran hutan dan lahan[12]. Penelitian lain yang dilakukan Mardi Yudhi Putra dan Dwi Ismiyana Putri (2022), yang berjudul "Pemanfaatan Algoritma Naïve Bayes dan K-Nearest Neighbor untuk Klasifikasi Jurusan Siswa Kelas XI" dari penelitian tersebut, pada algoritma Naïve Bayes didapat nilai akurasi sebesar 81,82% sementara nilai akurasi yang didapat pada algoritma K-Nearest Neighbor sebesar 92,73%. Sehingga dapat ditarik kesimpulan bahwa algoritma K-Nearest Neighbor lebih baik, karena memiliki nilai akurasi yang tinggi, yang berarti mendekati keadaan sebenarnya[12]. Penelitian lain yang dilakukan oleh Amat Danuari, dkk (2021) dengan judul "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Kelayakan Penerima Bantuan Sembako" dari penelitian tersebut didapat nilai akurasi sebesar 86%, recall 85% dan presisi 88% dengan menggunakan data training sebanyak 135 data dan data testing 40 data[14].

Perbedaan penelitian ini dengan penelitian sebelumnya terdapat pada atribut dan data yang akan digunakan untuk klasifikasi penerima bantuan Program Keluarga Harapan di Kecamatan Bungaraya. Atribut yang digunakan pada penelitian ini yaitu umur, penghasilan, jumlah tanggungan, jumlah tanggungan anak sekolah, tanggungan disabilitas, status rumah, jenis dinding rumah, dan jenis lantai rumah. Kemudian, untuk proses evaluasi pada penelitian ini, penulis menggunakan confusion matrix, sebagai metode pengukur akurasi, presisi dan recall.

Tujuan dari penelitian ini ialah untuk menerapkan algoritma Naïve Bayes dalam klasifikasi penerima bantuan PKH di Kecamatan Bungaraya serta dapat menghasilkan model yang terbaik. Penelitian ini juga diharapkan dapat membantu penerimaan Keluarga Penerima Manfaat (KPM) menjadi lebih efisien dan efektif.

## 2. METODOLOGI PENELITIAN

Tahapan dari penelitian yang direncanakan digambarkan dalam Gambar 1 sebagai berikut:



**Gambar 1.** Metodologi Penelitian

### 2.1 Perumusan Masalah

Proses awal dalam melakukan penelitian adalah menentukan fokus penelitian atau perumusan masalah. Studi ini menemukan rumusan masalah tentang cara menggunakan data mining untuk mengklasifikasi penerima bantuan sosial Program Keluarga Harapan (PKH).

### 2.2 Pengumpulan Data

Peneliti melakukan pengumpulan data dalam dua tahap yakni studi pustaka dan observasi.

a. Studi pustaka

Proses mengumpulkan literatur yang diperlukan untuk penelitian dikenal sebagai studi pustaka. Salah satu cara untuk menerapkan metode ini adalah dengan mencari referensi pada jurnal, buku, atau penelitian lain yang terkait dengan masalah.

b. Observasi

Dalam proses observasi ini, melakukan melihat, mempelajari secara langsung serta mengumpulkan data penerima bantuan PKH di Kecamatan Bungaraya. Data yang dikumpulkan terdiri dari data primer, yang mencakup semua penerima bantuan PKH di Kecamatan Bungaraya pada tahun 2022, serta calon penerima dari beberapa desa di Kecamatan Bungaraya. File.xlsx dari data ini akan digunakan untuk pengujian data training dan testing. Atribut yang terdapat pada penelitian ini yaitu nama, usia, penghasilan, jumlah tanggungan, jumlah tanggungan anak sekolah, penyandang disabilitas, status rumah, jenis lantai dan jenis dinding, status PKH.

**2.3 Analisa Permasalahan**

Salah satu langkah penting yang biasa dilakukan setelah pengumpulan data adalah analisis. Pada titik ini, pemahaman terhadap data yang telah dikumpulkan diperlukan untuk menarik kesimpulan dan menginterpretasikan temuan. Berikut adalah penjelasan tentang langkah-langkah yang digunakan untuk menemukan klasifikasi dari kumpulan data penerima bantuan PKH dengan menggunakan algoritma NaïveBayes:

1. Data Selection

Data selection merupakan tahapan seleksi data. Pemilihan data pada proses seleksi data melibatkan pemilihan 8 atribut yang akan digunakan, yakni usia, penghasilan, jumlah tanggungan, tanggungan anak sekolah, tanggungan disabilitas, status rumah, jenis lantai dan jenis dinding. Data PKH kecamatan Bungaraya dapat dilihat pada tabel 1 berikut.

**Tabel 1.** Data PKH

Nama	Usia	Penghasilan	Jumlah Tanggungan	Tanggungan Anak Sekolah	Tanggungan Disabilitas	Status Rumah	Jenis Dinding	Jenis Lantai	Status PKH
Kartiyem	65	1.500.000	3	1	0	Pribadi	Tembok	Plester	YA
Sopini	69	0	1	0	0	Rumah anak	Tembok	Plester	YA
Misiran	69	1.300.000	2	0	0	Pribadi	Tembok	Plester	YA
Istiqomah	45	1.200.000	4	2	0	Sewa	Tembok	Plester	YA
Sri Tami	52	1.300.000	5	2	0	Pribadi	Semi Tembok	Plester	YA
Bayu	28	2.000.000	4	1	0	Pribadi	Semi tembok	Plester	TIDAK
Ponidah	28	2.000.000	3	1	0	Sewa	Semi Tembok	Keramik	TIDAK
Samsul	26	2.500.000	2	1	0	Rumah Orang Tua	Semi Tembok	Plester	TIDAK
Slamet	29	1.900.000	3	0	0	Pribadi	Semi Tembok	Plester	TIDAK
Jamal	27	1.900.000	2	0	0	Sewa	Tembok	Keramik	TIDAK
Marinem	31	1.000.000	3	0	0	Pribadi	Tembok	Plester	TIDAK
Dian	28	2.000.000	3	0	0	Pribadi	Tembok	Plester	TIDAK
Marsidi	77	1.500.000	2	0	0	Pribadi	Tembok	Keramik	TIDAK
Sapri	36	2.300.000	4	1	0	Pribadi	Tembok	Plester	TIDAK
Sutarwan	34	2.500.000	3	2	0	Pribadi	Tembok	Plester	TIDAK

2. Pre-processing

Pre-processing yakni membersihkan data yang memiliki nilai kosong (value missing), inkonsistensi, atau nilai yang tidak sesuai [15].

3. Tranformation

Dalam penelitian ini, fitur tipe data kategorial diubah untuk melakukan transformasi data, ditunjukkan pada Tabel 2 berikut ini

**Tabel 2.** Transformasi data

Atribut	Tipe Atribut	Tipe Atribut setelah Transformasi
Usia	Numerik	Kategorial (Rendah, sedang, Tinggi)
Penghasilan	Numerik	Kategorial (Rendah, sedang, Tinggi)

Atribut	Tipe Atribut	Tipe Atribut setelah Transformasi
Jumlah tanggungan	Numerik	Kategorial (Rendah, sedang, Tinggi)
Jumlah tanggungan anak sekolah	Numerik	Kategorial (Rendah, sedang, Tinggi)
Penyandang disabilitas	Numerik	Kategorial (Rendah, sedang, Tinggi)
Status rumah	Kategorial	Kategorial (Pribadi, Sewa, Rumah Orang tua, Rumah Anak)
Jenis lantai	Kategorial	Kategorial (Keramik, Plester, Semi Plester)
Jenis dinding	Kategorial	Kategorial (Tembok, Papan, Semi tembok)

Tabel 2 menunjukkan transformasi data numerik menjadi menjadi kategorial. Tranformasi data dilakukan pada atribut Usia, Penghasilan, Jumlah tanggungan, Jumlah tanggungan anak sekolah dan Penyandang disabilitas.

#### 4. Data Mining

Langkah pertama dalam proses perhitungan Naive Bayes adalah mengumpulkan data latih dari data penerima bantuan PKH. Berikut ini adalah variabel penentu yang digunakan untuk mengklasifikasikan data penerima bantuan PKH:

- Usia merupakan variabel yang dikelompokkan menjadi 3 kategori, yaitu rendah ( $\leq 40$  tahun), sedang (41-59 tahun), tinggi ( $\geq 60$  tahun).
- Penghasilan merupakan pendapatan/kemasukan yang diperoleh per-bulan. variabel ini dikelompokkan menjadi 3 kategori, yaitu rendah ( $\leq 1.000.000$ ), sedang (1.100.000-2.900.000), tinggi ( $\geq 3.000.000$ ).
- Jumlah Tanggungan merupakan jumlah anggota keluarga yang terdapat di Kartu keluarga. variabel yang dikelompokkan sebanyak 3 (tiga) kategori, yakni rendah( $\leq 2$ ), sedang(3), dan tinggi( $\geq 4$ ).
- Tanggungan Anak Sekolah merupakan jumlah tanggungan anak yang masih bersekolah. Variabel yang dikelompokkan sebanyak 3 (tiga) kategori, yakni rendah( $\leq 1$ ), sedang(2), dan tinggi( $\geq 3$ ).
- Tanggungan disabilitas merupakan anggota keluarga yang memiliki keterbatasan fisik atau mental. variabel yang dikelompokkan menjadi 3 kategori, yaitu rendah ( $\leq 1$ ), sedang (2), tinggi ( $\geq 3$ ).
- Status Rumah merupakan variable yang dikelompokkan menjadi 4 kategori, yaitu rumah sewa, rumah pribadi, rumah orang tua, dan rumah anak.
- Jenis Lantai merupakan variabel yang dikelompokkan menjadi 3 kategori, yaitu Keramik, plester, dan semi plester.
- Jenis Dinding merupakan variabel yang dikelompokkan menjadi 3 kategori, yaitu tembok, papan dan semi papan.

#### 5. Evaluation

Tahap ini dilakukan guna membandingkan hasil dari pengerjaan pada Google Collab dan pada tools Rapid Miner. Sehingga dapat dijadikan perbandingan, dan menemukan pola atau data baru yang menarik dan menguntungkan.

### 2.4 Perhitungan Algoritma Naive Bayes

Berikut adalah penjabaran tentang perhitungan algoritma Naive Bayes pada data penerima bantuan PKH di Kecamatan Bungaraya. Perhitungan klasifikasi menggunakan algoritma Naive Bayes dapat dilakukan menggunakan persamaan berikut[16][17]:

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

#### Keterangan

- X : Data yang memiliki class tidak diketahui
- H : Hipotesis data X yang merupakan suatu class yang lebih spesifik
- P (H|X) : Probabilitas hipotesis H berdasarkan kondisi X (posteriori probability)
- P (H) : Probabilitas hipotesis H (prior probability)
- P (X|H) : Probabilitas X bergantung kondisi hipotesis H
- P (X) : Probabilitas x

#### a. Menginput Dataset

Dataset yang digunakan pada penelitian ini merupakan data PKH di Kecamatan Bungaraya, yang berisi data penerima dan bukan penerima bantuan PKH. Dataset awal yang digunakan, dapat dilihat pada table berikut.

**Tabel 3.** Dataset

Nama	Usia	Penghasilan	Jumlah Tanggungan	Tanggungan Anak Sekolah	Tanggungan Disabilitas	Status Rumah	Jenis Dinding	Jenis Lantai	Status PKH
Kartiyem	Tinggi	Sedang	Sedang	Rendah	Rendah	Pribadi	Tembok	Plester	YA
Sopini	Tinggi	Rendah	Rendah	Rendah	Rendah	Rumah anak	Tembok	Plester	YA
Misiran	Tinggi	Sedang	Rendah	Rendah	Rendah	Pribadi	Tembok	Plester	YA

Nama	Usia	Penghasilan	Jumlah Tanggungan	Tanggungan Anak Sekolah	Tanggungan Disabilitas	Status Rumah	Jenis Dinding	Jenis Lantai	Status PKH
Istiqomah	Sedang	Sedang	Tinggi	Sedang	Rendah	Sewa	Tembok	Plester	YA
Sri Tami	Sedang	Sedang	Tinggi	Sedang	Rendah	Pribadi	Semi Tembok	Plester	YA
Bayu	Rendah	Sedang	Tinggi	Rendah	Rendah	Pribadi	Semi tembok	Plester	TIDAK
Ponidah	Rendah	Sedang	Sedang	Rendah	Rendah	Sewa	Semi Tembok	Keramik	TIDAK
Samsul	Rendah	Sedang	Rendah	Rendah	Rendah	Rumah Orang Tua	Semi Tembok	Plester	TIDAK
Slamet	Rendah	Sedang	Sedang	Rendah	Rendah	Pribadi	Semi Tembok	Plester	TIDAK
Jamal	Rendah	Sedang	Rendah	Rendah	Rendah	Sewa	Tembok	Keramik	TIDAK
Marinem	Rendah	Rendah	Sedang	Rendah	Rendah	Pribadi	Tembok	Plester	TIDAK
Dian	Rendah	Sedang	Sedang	Rendah	Rendah	Pribadi	Tembok	Plester	TIDAK
Marsidi	Tinggi	Sedang	Rendah	Rendah	Rendah	Pribadi	Tembok	Keramik	TIDAK
Sapri	Rendah	Sedang	Tinggi	Rendah	Rendah	Pribadi	Tembok	Plester	TIDAK
Sutarwan	Rendah	Sedang	Tinggi	Sedang	Rendah	Pribadi	Tembok	Plester	TIDAK

Tabel 3 di atas merupakan dataset yang di dapat dari data PKH di Kecamatan Bungaraya dan sudah dilakukan peoses transformasi sebelumnya.

**b. Menghitung probabilitas kelas**

Dalam penelitian ini menggunakan data set sebanyak 15 data, 5 sebagai penerima dan 10 tidak menerima. Probabilitas kemunculan nilai atribut label pada penelitian ini diketahui sebagai berikut:

$$1. P(Ya) = \frac{5}{15} = 0,3333$$

$$2. P(Tidak) = \frac{10}{15} = 0,6667$$

Dari perhitungan di atas didapatkan nilai probabilitas untuk kelas penerima sebesar 0,3333 dan probabilitas untuk kelas tidak menerima sebesar 0,6667. Perbandingan nilai kelas penerima dan tidak menerima yaitu  $0,333 < 0,6667$ , yang artinya peluang penerima pada data set tersebut lebih sedikit.

**c. Menghitung probabilitas atribut**

Perhitungan probabilitas atribut dilakukan pada setiap masing-masing kategori pada setiap subset, pada masing-masing kelas, yang terdapat pada data set. Probabilitas kemunculan nilai atributnya diketahui sebagai berikut:

**Tabel 4.** Probabilitas Kategori

ATRIBUT	SUBSET	YA	TIDAK
Usia	Rendah	0	0,7
	Sedang	0,4	0,2
	Tinggi	0,6	0,1
Penghasilan	Rendah	0,2	0,1
	Sedang	0,6	0,7
	Tinggi	0	0,2
Jumlah tanggungan	Rendah	0,2	0
	Sedang	0,4	0,8
	Tinggi	0,4	0,2
Tanggungan Anak sekolah	Rendah	0,4	0,8
	Sedang	0,6	0,2
	Tinggi	0	0
Tanggungan Disabilitas	Rendah	1	1
	Sedang	0	0
	Tinggi	0	0
Status rumah	Rumah Pribadi	0,6	0,7
	Rumah Sewa	0,2	0,1
	Rumah Orang tua	0	0,2
	Rumah Anak	0,2	0
Jenis dinding	Tembok	0,8	0,4
	Semi Tembok	0,2	0,6
	Papan	0	0
Jenis lantai	Plester	1	0,7
	Semi Plester	0	0

ATRIBUT	SUBSET	YA	TIDAK
	Keramik	0	0,3

Table 4 di atas menunjukkan nilai subset dari masing-masing atribut pada kelas Ya dan kelas Tidak. Nilai tersebut akan digunakan untuk perhitungan pada probabilitas kelas yang akan terpilih.

**d. Menghitung nilai probabilitas pada setiap kelas**

Pada tahap ini, menghitung probabilitas setiap nilai kelas berdasarkan nilai data uji dan tabel probabilitas. Penjabaran perhitungan dapat dilihat seperti dibawah ini:

$$P(Ya) = (UMUR | Tinggi) * (PENGASILAN | Sedang) * (JML TANGGUNGAN | Sedang) * (TANGGUNGAN ANAK SEKOLAH | Sedang) * (TANGGUNGAN DISABILITAS | Rendah) * (STATUS RUMAH | Pribadi) * (JENIS DINDING | Tembok) * (JENIS LANTAI |Plester)$$

$$P(Ya) = 0,3 \times 0,6 \times 0,6 \times 0,4 \times 0,6 \times 1 \times 0,6 \times 0,8 \times 1 = 0,0124$$

$$P(Tidak) = (UMUR | Tinggi) * (PENGASILAN | Sedang) * (JML TANGGUNGAN | Sedang) * (TANGGUNGAN ANAK SEKOLAH | Sedang) * (TANGGUNGAN DISABILITAS | Rendah) * (STATUS RUMAH | Pribadi) * (JENIS DINDING | Tembok) * (JENIS LANTAI |Plester)$$

$$P(Tidak) = 0,7 \times 0,1 \times 0,7 \times 0,8 \times 0,2 \times 1 \times 0,7 \times 0,4 \times 0,7 = 0,0061$$

Berdasarkan perhitungan tersebut, bandingkan hasil antara kelas YA dan kelas TIDAK. Nilai tertinggi adalah label yang terpilih. Sehingga hasil dari data uji adalah Kelas YA.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Data

Dalam penelitian ini, peneliti menggunakan data mengenai penerima dan calon bantuan Program Keluarga Harapan (PKH) tahun 2022 di Kecamatan Bungaraya, dengan jumlah sebanyak 560 data penerima dan bukan penerima bantuan PKH. Atribut yang digunakan pada penelitian ini yaitu usia, penghasilan, jumlah tanggungan, tanggungan anak sekolah, tanggungan disabilitas, status rumah, jenis lantai dan jenis dinding, yang akan ditunjukkan pada tabel 4 berikut:

**Tabel 5.** Atribut Data

No	Atribut	Keterangan
1.	Usia	Usia
2.	Penghasilan	Penghasilan yang bersangkutan perbulan
3.	Jumlah Tanggungan	Jumlah anggota keluarga dalam Kartu Keluarga (KK)
4.	Tanggungan Anak Sekolah	Jumlah anak yang masih bersekolah
5.	Tanggungan Disabilitas	Jumlah keluarga yang mengalami disabilitas
6.	Status Rumah	Status rumah rumah yang dihuni
7.	Jenis Dinding	Jenis dinding rumah yang dihuni
8.	Jenis Lantai	Jenis lantai rumah yang dihuni

Pada tabel 5 berisi seluruh atribut yang ada pada data penerima dan calon penerima bantuan PKH di Bungaraya dengan penjelasan dari masing-masing atribut, yang berfungsi sebagai kelas atau target. Perbandingan data yang digunakan dapat dilihat pada gambar 3.



**Gambar 2.** Perbandingan data antar kelas

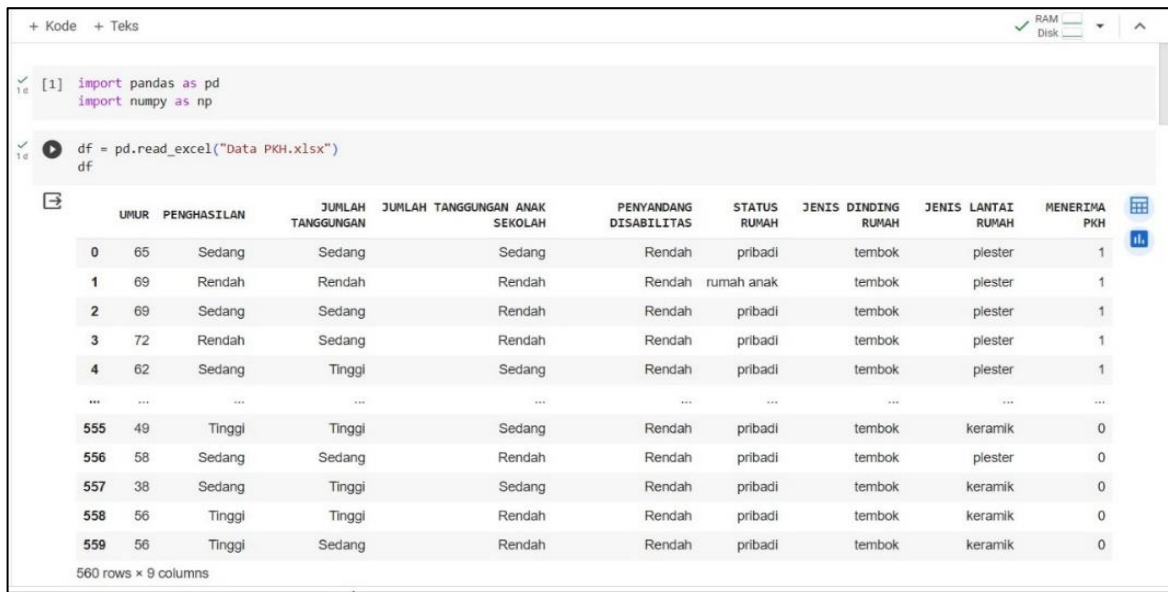
Pada gambar 3, dapat mengetahui perbandingan jumlah data yang tersedia pada penelitian ini. dari data yang tersedia, diagram berwarna biru menunjukkan jumlah data penerima bantuan yang berjumlah 205 penerima, sementara sisanya merupakan data bukan penerima bantuan PKH ditunjukkan dengan diagram berwarna merah.

### 3.2 Pembagian Data

Proses pemisahan antara data latih dan data uji dilakukan melalui penerapan validasi terpisah atau disebut split validation, di mana sebagian data akan diidentifikasi sebagai data training dan sisanya akan dianggap sebagai data testing untuk keperluan analisis.

### 3.3 Pembentukan Model Klasifikasi

Dalam pembuatan model klasifikasi pada penelitian ini memanfaatkan Google Colab dengan Bahasa pemrograman python. Berikut adalah contoh pernyataan untuk membangun model klasifikasi penerima bantuan PKH menggunakan algoritma Naive Bayes.



**Gambar 4.** Proses pengolahan data di Google colab

Gambar 4 menunjukkan proses input data dan isi data. Selanjutnya, algoritma Naive Bayes digunakan untuk mengolah data pada Google Colab. Dalam penelitian ini, percobaan klasifikasi dilakukan beberapa kali pada Google Colab dan Rapid Miner dengan membandingkan berbagai jenis data training dan testing sebesar 70:30, 80:20 dan 90:10. Untuk mengukur sejauh mana model mampu melakukan klasifikasi data yang benar, dapat dilihat melalui nilai presisi, recall dan akurasi yang disajikan pada table berikut.

**Tabel 6.** Nilai Presisi dan Recall pada Google colab

Perbandingan data Training dan data Testing	Kelas (Label)	Presisi	Recall
70:30:00	TIDAK	99%	85%
	YA	78%	98%
80:20:00	TIDAK	100%	98%
	YA	98%	100%
90:10:00	TIDAK	100%	92%
	YA	86%	100%

Tabel 6 menunjukkan nilai presisi dan recall yang diperoleh melalui perhitungan yang dilakukan pada google colab. Nilai presisi dan recall terbaik terdapat pada saat menggunakan perbandingan data training dan data testing 80:20 dengan nilai rata-rata sebesar 99%.

**Tabel 7.** Nilai Presisi dan Recall pada Rapid Miner

Perbandingan data Training dan data Testing	Kelas (Label)	Presisi	Recall
70:30:00	TIDAK	93,58%	95,33%
	YA	91,53%	88,52%
80:20:00	TIDAK	93%	94,37%
	YA	90%	87,80%
90:10:00	TIDAK	96,88%	88,57%
	YA	82,61%	95%

Tabel 7 menunjukkan nilai presisi dan recall yang diperoleh melalui perhitungan yang dilakukan pada Rapid Miner. Nilai presisi dan recall terbaik terdapat pada saat menggunakan perbandingan data training dan data testing 70:30 dengan nilai rata-rata sebesar 92%. Nilai presisi digunakan untuk mengukur sejauh mana prediksi positif

model benar. Presisi dapat dihitung dengan membagi nilai jumlah prediksi positif yang benar dengan total jumlah prediksi positif. Sedangkan nilai recall merupakan nilai yang mengukur sejauh mana model dapat mengidentifikasi secara benar contoh-contoh positif. Nilai recall dihitung dengan cara membagi jumlah prediksi positif yang benar dengan total jumlah contoh dalam dataset[18]. Nilai akurasi yang didapat pada penelitian ini tersaji pada table berikut.

**Tabel 8.** Nilai Akurasi

Perbandingan data Training dan data Testing	Akurasi pada google colab	Akurasi pada Rapid Miner
70:30:00	90%	92%
80:20:00	99%	94%
90:10:00	95%	91%

Tabel 8 menunjukkan hasil percobaan bahwa jumlah data antar masing-masing kelas sasaran mempengaruhi nilai akurasi yang diperoleh. Akurasi terbaik diperoleh ketika menggunakan perbandingan data training dan data testing 80:20. Semakin seimbang jumlah data antara setiap kelas target, semakin baik nilai yang dihasilkan. Untuk mengetahui keakuratan model yang dibuat, dapat dilihat melalui nilai akurasi, presisi dan recall[19]. Sebagaimana diketahui bahwa nilai akurasi dipresentasikan dari 0 hingga 100%, artinya semakin tinggi nilai akurasi yang didapat, maka akan baik pula model tersebut[20].

#### 4. KESIMPULAN

Penelitian ini berhasil menggunakan algoritma Naïve Bayes untuk mengklasifikasi penerima bantuan PKH di Kecamatan Bungaraya. Atribut yang digunakan meliputi usia, penghasilan, jumlah tanggungan, jumlah tanggungan anak sekolah, tanggungan disabilitas, status rumah, jenis dinding rumah, dan jenis lantai rumah. Nilai akurasi terbaik yang diperoleh melalui perhitungan yang dilakukan pada Google Colab pada perbandingan 80:20, sebesar 99%. Dari hasil tersebut menunjukkan bahwa terdapat 65 data bukan penerima bantuan yang diklasifikasikan sebagai data bukan penerima, dan terdapat 1 data penerima yang diklasifikasikan sebagai data bukan penerima. Kemudian terdapat 46 data penerima yang seluruhnya diklasifikasikan sebagai data penerima. Metode klasifikasi dengan algoritma Naïve Bayes ini dapat diaplikasikan untuk menentukan calon penerima bantuan PKH di Kecamatan Bungaraya di masa depan. Penelitian yang akan datang diharapkan dapat mengembangkan aplikasi berbasis web yang menggunakan Naïve Bayes untuk menampilkan hasil klasifikasi penerima bantuan PKH.

#### REFERENCES

- [1] BADAN PUSAT STATISTIK, “Profil Kemiskinan di Indonesia Maret 2023,” 2023. <https://www.bps.go.id/pressrelease/2023/07/17/2016/profil-kemiskinan-di-indonesia-maret-2023.html> (accessed Jul. 12, 2023).
- [2] N. Apriani, D. Tampubolon, and Y. Maulida, “Analisis Pelaksanaan Program Keluarga Harapan ( Pkh ) Dalam Meningkatkan Ekonomi Masyarakat Miskin ( Studi Pada Kecamatan Tenayan Raya Kota Pekanbaru ),” vol. 5, no. 2, pp. 132–138, 2023.
- [3] D. Utami and P. A. R. Devi, “Klasifikasi Kelayakan Penerima Bantuan Program Keluarga Harapan (Pkh) Menggunakan Metode Weighted Naïve Bayes Dengan Laplace Smoothing,” JPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform., vol. 7, no. 4, pp. 1373–1384, 2022, doi: 10.29100/jpi.v7i4.3592.
- [4] Menteri Sosial Republik Indonesia, Peraturan Menteri Sosial Republik Indonesia Nomor 1 Tahun 2018 Tentang Program Keluarga Harapan, no. 1. 2018, pp. 430–439.
- [5] A. Arifin, W. Handoko, and Z. Efendi, “Implementasi Metode Naive Bayes Untuk Klasifikasi Penerima Program Keluarga Harapan,” J-Com (Journal Comput., vol. 2, no. 1, pp. 21–26, 2022, doi: 10.33330/j-com.v2i1.1577.
- [6] N. Nuryanti, “Pemanfaatan Bantuan Program Keluarga Harapan untuk Menyukkseskan Wajib Belajar XII Tahun di Kecamatan Pusako Siak,” Indones. J. Islam. Educ. Manag., vol. 2, no. 1, p. 26, 2019, doi: 10.24014/ijiem.v2i1.7122.
- [7] D. . U. and M. Mesran, “Analisis Komparasi Metode Klassifikasi Data Mining dan Reduksi Atribut pada Data Set Penyakit Jantung,” J. MEDIA Inform. BUDIDARMA, vol. 4, no. 2, pp. 437–444, 2020.
- [8] M. Kamber and J. Han, Data Mining: Concepts and Techniques : Concepts and Techniques. 2018.
- [9] N. I. Nella, N. Y. Setiawan, and ..., “Klasifikasi Penerima Bantuan Program Keluarga Harapan menggunakan Algoritme Decision Tree C4. 5 (Studi Kasus: Desa Mlirip Kabupaten Mojokerto),” ... Teknol. Inf. dan ..., vol. 6, no. 3, pp. 1332–1339, 2022, [Online]. Available: <https://jptiik.multi.web.id/index.php/j-ptiik/article/view/10801%0Ahttps://jptiik.multi.web.id/index.php/j-ptiik/article/download/10801/4769>.
- [10] C. Prianto, Ema Ainun Novia, Woro Isti Rahayu, Sistem Perbandingan Algoritma K-Means Dan Naive Bayes Untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan. Bandung: Kreatif Industri Nusantara, 2020.
- [11] N. Alfiah and P. S. Informasi, “Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode Naive Bayes,” vol. XVI, pp. 32–40, 2021.
- [12] T. A. Pratiwi, M. Irsyad, R. Kurniawan, S. Agustian, and B. S. Negara, “Klasifikasi Kebakaran Hutan Dan Lahan Menggunakan Algoritma Naïve Bayes Di Kabupaten Pelalawan,” CESS (Journal Comput. Eng. Syst. Sci., vol. 6, no. 1, p. 139, 2021, doi: 10.24114/cess.v6i1.22555.



- [13] P. M. Yudhi and D. Ismiyana Putri, “Pemanfaatan Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Kelas XI,” *J. Tekno Kompak*, vol. 16, no. 2, pp. 176–187, 2022.
- [14] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, “Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, p. 219, 2021, doi: 10.30865/jurikom.v8i6.3655.
- [15] M. Y. H. S. Dinda Ayu Pratiwi, Rolly Maulana Awangga, *Seleksi Calon Kelulusan Tepat Waktu Mahasiswa Teknik Informatika Menggunakan Metode Naive Bayes*. Bandung: Kreatif Industri Nusantara, 2020.
- [16] J. Homepage et al., “IJIRSE: Indonesian Journal of Informatic Research and Software Engineering Naïve Bayes Classifier (NBC) Algorithm Analysis for Prediction Medical Device Sales,” vol. 3, no. 2, pp. 119–126, 2023.
- [17] A. Ridwan, “Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus,” *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21, 2020, doi: 10.47970/siskom-kb.v4i1.169.
- [18] D. Hasanudin Sirait., Rini Mayasari., Nono Heryana., *Metode Dan Penerapan Sistem Pakar*. Get prees Indonesia, 2023.
- [19] Kunchahyo Setyo Nugroho, “Condusion Matrix untuk Evaluasi Model pada Supervised Learning,” *Medium*, 2019. .
- [20] M. L. P. Putra, “Menghitung Akurasi Dengan Confusion Matrix,” *Medium*, 2020.