

CLASSIFICATION OF FAMILY HOPE PROGRAM RECIPIENTS USING NAIVE BAYES AND C4.5 METHODS

Farras Ahmad Fauzi¹, Tatang Rohana², Ayu Ratna Juwita³, Deden Wahiddin⁴

^{1,2,3,4}*Informatics Engineering, Computer Science Faculty, Buana Perjuangan Karawang University, Indonesia*
Email: ¹farrasfauzi@mhs.ubpkarawang.ac.id, ²tatang.rohana@ubpkarawang.ac.id, ³ayurj@ubpkarawang.ac.id, ⁴deden.wahiddin@ubpkarawang.ac.id

(Article received: September 18, 2024; Revision: September 22, 2024; published: October 25, 2024)

Abstract

Receiving PKH assistance in Rawamerta District does not always go well, so there are people who are not entitled to receive assistance. This is because there is still no system that can facilitate the process of classifying PKH assistance recipients. The application of data mining can facilitate classification with high speed and accuracy. The purpose of this study is to classify PKH assistance recipients using the Naïve Bayes and C4.5 methods to determine the eligibility of PKH for people facing social welfare problems. The data used is PKH data in Rawamerta District, Karawang Regency in 2023, totaling 1834 data. The results of naive bayes accuracy of 98.89%, precision 98.25%, recall 98.51%, F1-score 98.89%, and AUC 1.00 are included in the excellent classification because they are in the range of 0.90-1.00, while the C4.5 algorithm produces Accuracy values of 99.26%, Precision 99.25%, Recall 99.25%, F1-score 99.25% and AUC 0.99 are included in the excellent classification because they are in the range of 0.90-1.00. The C4.5 algorithm is superior to Naive Bayes, because the accuracy produced is higher.

Keywords: C4.5; Confusion Matrix; Classification; Naive Bayes; ROC Curve

KLASIFIKASI PENERIMA PROGRAM KELUARGA HARAPAN MENGGUNAKAN METODE NAIVE BAYES DAN C4.5

Abstrak

Penerimaan bantuan PKH di Kecamatan Rawamerta tidak selalu berjalan dengan baik, sehingga terdapat orang yang tidak berhak menerima bantuan. Hal ini disebabkan karena masih belum tersedia sistem yang mampu memudahkan proses klasifikasi penerima bantuan PKH. Penerapan data mining dapat mempermudah klasifikasi dengan kecepatan dan akurasi yang tinggi. Adapun tujuan penelitian ini untuk mengklasifikasikan penerima bantuan PKH menggunakan metode *Naïve Bayes* dan C4.5 untuk menentukan kelayakan PKH bagi masyarakat yang menghadapi masalah kesejahteraan sosial. Data yang digunakan berupa data PKH di Kecamatan Rawamerta Kabupaten Karawang pada tahun 2023 yang berjumlah 1834 data. Hasil akurasi *naive bayes* sebesar 98.89%, presisi 98.25%, recall 98.51%, *F1-score* 98.89%, dan AUC 1.00 termasuk ke dalam *excellent classification* karena berada pada rentang 0.90-1.00, sedangkan algoritma C4.5 menghasilkan nilai *Accuracy* sebesar 99.26%, *Precision* 99.25%, *Recall* 99.25%, *F1-score* 99.25% dan AUC 0.99 termasuk ke dalam *excellent classification* karena berada pada rentang 0,90-1.00. Algoritma C4.5 lebih unggul dibandingkan Naive Bayes, karena akurasi yang dihasilkan lebih tinggi.

Kata Kunci: C4.5; Confusion Matrix; Klasifikasi; Naive Bayes; Kurva ROC

1. PENDAHULUAN

Program Keluarga Harapan (PKH) adalah salah satu inisiatif pemerintah dalam Upaya mengentaskan kemiskinan dan meningkatkan kesejahteraan masyarakat [1]. PKH yang dikenal sebagai *Conditional Cash Transfer* (CCT) di dunia internasional, telah menjadi salah satu program bantuan sosial paling sukses di Indonesia. Program ini diluncurkan pada tahun 2007 dengan tujuan

untuk mengurangi kemiskinan dan mendorong kemandirian penerima bansos [2]. Program ini memiliki fokus utama pada keluarga-keluarga yang berada dalam kondisi ekonomi rendah, dengan memberikan bantuan sosial berupa tunjangan tunai secara berkala. Dengan memberikan dukungan finansial kepada keluarga-keluarga yang membutuhkan, PKH diharapkan dapat menciptakan perubahan signifikan dalam peningkatan kesejahteraan, akses pendidikan, dan pelayanan

Kesehatan [3]. Pada tahun 2023, PKH sudah mencakup 514 kabupaten dan 6.709 kecamatan di seluruh Indonesia. Hal ini menunjukkan bahwa PKH semakin diperluas dan diterima di masyarakat bagi penyaluran PKH 9,8 juta KPM [4].

Berdasarkan data BPS Karawang pada tahun 2023, Kecamatan Rawamerta adalah kecamatan yang terletak di Kabupaten Karawang, Jawa Barat, memiliki luas wilayah yang cukup luas, yaitu sekitar 5.032,41 hektare. Pada tahun 2022, jumlah penduduknya mencapai hampir 83.987 jiwa. Dari segi jenis kelamin, jumlah penduduknya terbagi menjadi dua jenis kelamin, yaitu 27.846 laki-laki dan 56.141 perempuan [5]. Pada tahun 2021, angka kemiskinan di Kabupaten Karawang termasuk Kecamatan Rawamerta mencapai 494.201 orang, atau 1,20% dari total populasi. Angka ini naik hingga tahun 2022, menjadi 521,158 orang, atau 2,70% dari total populasi. Artinya, selama 2 tahun, angka kemiskinan di Kabupaten Karawang naik sekitar 26.957 orang [6].

Penerimaan bantuan PKH di Kecamatan Rawamerta pada proses verifikasi dan validasi data penerima bantuan tidak selalu berjalan dengan baik, sehingga terdapat orang yang tidak berhak menerima bantuan. Karena belum ada sistem yang dapat memudahkan proses klasifikasi penerima bantuan PKH. Oleh karena itu, penyaluran bantuan PKH di Kecamatan Rawamerta Kabupaten Karawang dinilai perlu diperbaiki [7].

Berdasarkan permasalahan penerima bantuan PKH, masih menjadi tantangan. Hal ini dikarenakan pengolahan data masih dilakukan secara manual, hal tersebut membutuhkan waktu yang cukup lama karena banyaknya data harus diproses dan kriteria yang harus dipenuhi. Hal ini juga meningkatkan risiko kehilangan atau kerusakan dokumen. Akibatnya, penyaluran Program PKH dianggap kurang optimal.

Data mining diperlukan untuk membantu para ketua pendamping PKH di dalam menentukan Keluarga Penerima Manfaat (KPM) bantuan PKH. Data *mining* dalam sebuah basis data bertujuan proses mengekstraksi informasi dan pola yang berguna dari kumpulan data yang besar [8]. Penerapan data mining tersebut dapat menggunakan metode klasifikasi *Naive Bayes* dan C4.5.

Klasifikasi merupakan tahapan untuk mendapatkan model yang mampu membedakan kelas data Adapun Algoritma C4.5 merupakan sebuah model berbentuk pohon yang digunakan untuk menganalisis data dan membuat keputusan berdasarkan informasi yang tersedia dari data tersebut [9].

Kedua algoritma *Naive Bayes* dan C4.5 relatif sederhana dan memiliki kinerja tinggi. Kedua algoritma tersebut dibandingkan untuk mengetahui akurasi dan kecepatan komputasi yang lebih tinggi [10]. Alasan penggunaan Algoritma *Naive Bayes*

dan C4.5 karena kinerjanya sangat baik dalam mengklasifikasikan data dan memiliki keunggulan pada komputasinya yang sederhana serta tingkat akurasi yang tinggi.

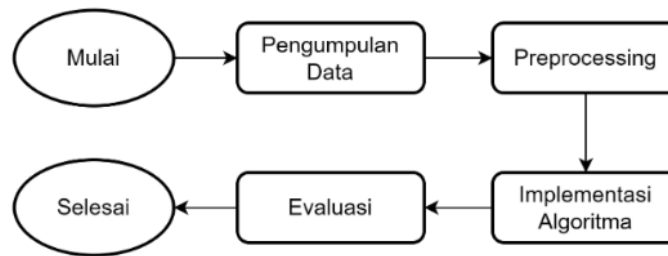
Pada Penelitian sebelumnya yang dilakukan oleh Amelia Irsyada, dkk (2023), tentang klasifikasi PKH menggunakan *Naive Bayes*. Dari hasil klasifikasi ini, delapan variabel digunakan untuk mengklasifikasikan penerima bantuan. Akurasi terbaik adalah 99%. [11]. Selanjutnya, penelitian Nabilah Iftah Nella, dkk (2022), menerapkan metode C4.5, jumlah data 357 dari 7 desa, dalam proses klasifikasi penerima PKH. Pengujian yang dilakukan menghasilkan *confusion matrix* dengan akurasi 71,1%, presisi 90,9%, recall 70,4%, dan *f1-score* 79,93% [12]. Sedangkan, penelitian Aviv Fitria Yulia, dkk (2023), menerapkan *K-Means* dalam proses klasifikasi data menentukan kelayakan penerima bantuan bagi masyarakat miskin di desa Sukoharjo III. Hasil pengujian proses pengolahan data seleksi penerimaan dana bantuan dengan algoritma *K-Means* menghasilkan indeks Davies bouldin sebesar 0,738. Hasil tersebut dinilai cukup baik karena semakin mendekati angka nol maka penelitian ini menghasilkan tiga kelompok yaitu: 1.218 warga layak menerima bantuan, 2.514 warga dianggap menerima bantuan, 1.040 warga tidak layak menerima bantuan [13]. Pada penelitian Muhammad Syarif Hartawan, dkk (2023), metode yang digunakan KNN, NBC, dan PNN digunakan untuk klasifikasi Program Keluarga Harapan (PKH). Hasil pengujian Metode *Naive Bayes* (NBC) memiliki akurasi tertinggi sebesar 80%, lebih tinggi dari metode KNN sebesar 20% dan PNN sebesar 10%. Oleh karena itu, metode NBC ditetapkan sebagai metode terbaik untuk kasus klasifikasi penerima bantuan PKH Kota Pekanbaru [14].

Penelitian ini bertujuan untuk membandingkan metode *Naive Bayes* dan C4.5 dalam mengklasifikasi penerima bantuan PKH. Dari hasil Implementasi metode klasifikasi ini diharapkan dapat memberikan dampak positif terhadap kebijakan pemerintah dalam menyalurkan bantuan PKH secara lebih tepat sasaran kepada keluarga yang membutuhkan.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Pada penelitian ini, tahapan penelitian dimulai dari studi literatur, pengumpulan dataset preprocessing, implementasi algoritma dengan membandingkan algoritma Naïve Bayes dan C4.5. Tahap akhir adalah evaluasi. Berikut adalah gambaran tahapan pada penelitian ini.



Gambar 1. Tahapan Penelitian

2.1.2 Pengumpulan Data

Dataset yang diperoleh berasal dari Kantor Kecamatan Rawamerta. Data yang digunakan merupakan data PKH di Kecamatan Rawamerta Kabupaten Karawang pada tahun 2023 yang berjumlah 1834 data dan 10 atribut diantaranya id, kabupaten, kecamatan, desa, nama, jumlah tanggungan, kondisi rumah, status kepemilikan, status kesejahteraan dan keterangan. Adapun isi dataset pada gambar 2.

id	kabupaten	kecamatan	desa	nama	jumlah_tanggungan	kondisi_rumah	status_kepemilikan	status_kesejahteraan	Keterangan
1	KARAWANG	RAWAMERTA	BALONGSARI	WARSIH	4	Baik	Milik Sendiri	Mampu	Tidak Layak
2	KARAWANG	RAWAMERTA	BALONGSARI	ENAH	3	sederhana	Milik Sendiri	Miskin	Layak
3	KARAWANG	RAWAMERTA	BALONGSARI	AMAH	5	Buruk	Sewa	Miskin	Layak
4	KARAWANG	RAWAMERTA	BALONGSARI	ROHAYATI	2	Baik	Milik Sendiri	Mampu	Tidak Layak
5	KARAWANG	RAWAMERTA	BALONGSARI	NANA ROHANA	3	sederhana	Sewa	Miskin	Layak
...
1831	KARAWANG	RAWAMERTA	SUKARAJA	DALI	2	buruk	sewa	sangat Miskin	Layak
1832	KARAWANG	RAWAMERTA	SUKARAJA	AMIH	1	baik	Milik Sendiri	Mampu	Tidak Layak
1833	KARAWANG	RAWAMERTA	SUKARAJA	SARINAH	3	buruk	sewa	sangat Miskin	Layak
1834	KARAWANG	RAWAMERTA	SUKARAJA	ENJUM	2	sederhana	sewa	Miskin	Layak
1835	KARAWANG	RAWAMERTA	SUKARAJA	ENTO	1	baik	Milik Sendiri	Mampu	Tidak Layak

Gambar 2. Data Program Keluarga Harapan (PKH)

2.1.3 Preprocessing Data

Tujuan dari *preprocessing* yaitu untuk menyeleksi data terhadap dataset yang digunakan, serta membuang data yang tidak diperlukan seperti menghapus beberapa komponen yang tidak digunakan sehingga akan menghasilkan sebuah data yang relevan dan lebih ringkas [15]. Proses ini mencakup beberapa tahapan, termasuk data *selection*, data *cleaning*, transformasi data, dan normalisasi data.

a. *Data Selection*

Data selection bertujuan untuk menentukan atribut yang relevan, karena data yang didapatkan tidak semuanya digunakan.

b. *Data cleaning*

Data cleaning bertujuan untuk membersihkan data yang terdapat *missing value* dan duplikasi data.

c. *Transformasi Data*

Transformasi data bertujuan untuk merubah bentuk suatu data numerik menjadi kategorial ataupun sebaliknya.

d. *Normalisasi Data*

Normalisasi data merupakan proses mengubah nilai variabel sehingga dapat diukur dalam skala umum.

2.1.4 Implementasi Algoritma

Algoritma *Naïve Bayes* adalah suatu Teknik klasifikasi yang mengelompokkan masalah ke dalam kategori label dengan memanfaatkan probabilitas dan statistik [16]. Teorema *Naïve Bayes* menjadi

metode yang efisien dalam penggunaannya, terutama ketika ketersediaan data pelatihan terbatas [17]. Berikut adalah tahapan algoritma *Naïve Bayes* [18].

1. Menghitung kelas
2. Selanjutnya, kalikan semua hasil berdasarkan data X yang akan diklasifikasikan.
3. Terakhir, bandingkan hasil probabilitas per kelas. Kelas dengan probabilitas tertinggi merupakan prediksi yang dihasilkan oleh model

Adapun persamaan teorema bayes adalah sebagai berikut.

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

Setelah mengimplementasikan Algoritma *Naïve Bayes*, kemudian melakukan perhitungan menggunakan C4.5. Algoritma C4.5 adalah teknik klasifikasi untuk membangun struktur pohon keputusan. Kelebihan Algoritma C4.5 adalah mudah dipahami [19]. Adapun tahapan dari Algoritma C4.5 adalah sebagai berikut.

1. Menentukan atribut sebagai akar
2. Membuat cabang
3. Membagi kasus menjadi beberapa cabang.
4. Mengulangi proses hingga cabang tersebut memiliki kelas yang sama.

Rumus Algoritma C4.5 dibagi menjadi dua bagian. Bagian pertama, perhitungan *Entropy*. Berikut adalah persamaan untuk menghitung *entropy* [20].

$$Entropy(S) = \sum_{i=1}^n p_i * \log_2 p_i \tag{2}$$

Adapun persamaan untuk menghitung *gain* adalah sebagai berikut.

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \tag{3}$$

2.1.5 Evaluasi

Tahap evaluasi menggunakan *Confusion Matrix* dan kurva ROC. *Confusion Matrix* adalah metode

untuk mengevaluasi akurasi dalam konteks data mining, Dalam *Confusion Matrix* terdapat beberapa metrik evaluasi, termasuk *precision*, *recall*, dan *accuracy*. *Precision* mengukur kemampuan model dalam mengklasifikasikan data positif dengan tepat, *recall* menunjukkan sejauh mana model dapat mendeteksi seluruh data positif yang seharusnya terklasifikasikan[21].

Dengan menggunakan *Confusion Matrix*, evaluasi model menjadi lebih komprehensif karena mempertimbangkan berbagai aspek klasifikasi. *Precision*, *recall*, dan *accuracy* adalah metrik yang penting dalam evaluasi model klasifikasi. Berikut adalah rumus untuk menghitung masing-masing metrik tersebut [22]

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

Kurva ROC digunakan untuk mengevaluasi kinerja klasifikasi yang melibatkan dua kelas keputusan. Setiap objek diklasifikasikan sebagai positif atau negatif. Pada gambar kurva ROC, *True Positive rate* dipetakan pada sumbu Y dan *False Positive rate* dipetakan pada sumbu X [23]. Hasil AUC (*Area Under Curve*) dapat dikelompokkan menjadi beberapa kategori berdasarkan metode klasifikasi dalam data mining.

- 0.90 - 1.00 = Klasifikasi sangat baik
- 0.80 - 0.90 = Klasifikasi baik
- 0.70 - 0.80 = Klasifikasi cukup
- 0.60 - 0.70 = Klasifikasi buruk
- 0.50 - 0.60 = Gagal

3. HASIL DAN PEMBAHASAN

3.1 Preprocessing Data

Preprocessing data terdiri dari data *selection*, data *cleaning*, transformasi data, dan normalisasi data.

3.1.1 Data Selection

Pada proses ini, atribut id, kabupaten, kecamatan, desa, dan nama akan dihapus karena tidak berpengaruh pada klasifikasi. Berikut adalah data *selection* pada tabel 1.

Tabel 1. Data Selection

Jumlah tanggungan	Kondisi rumah	Status kepemilikan	Status kesejahteraan	Keterangan
4	Baik	Milik sendiri	Mampu	Tidak layak
3	Sederhana	Milik sendiri	Miskin	Layak
5	Buruk	Sewa	Miskin	Layak
2	Baik	Milik sendiri	Mampu	Tidak layak
3	Sederhana	Sewa	Miskin	Layak

Berdasarkan tabel 1, hasil seleksi data yang akan digunakan sebanyak 5 atribut yaitu jumlah tanggungan, kondisi rumah, status kepemilikan, status kesejahteraan dan keterangan.

3.1.2 Data Cleaning

Data *cleaning* pada penelitian ini dilakukan untuk menghilangkan *missing value* dan duplikasi data. Berikut adalah data *cleaning* pada gambar 3.

```

===== Pre-Processing =====
jumlah_tanggungan      False
kondisi_rumah          False
status_kepemilikan     False
status_kesejahteraan   False
Keterangan             False
dtype: bool
jumlah_tanggungan      0
kondisi_rumah          0
status_kepemilikan     0
status_kesejahteraan   0
Keterangan             0
dtype: int64

#Cek Duplikasi Data
print(df.duplicated().any())
print(df.duplicated().sum())

False
0
    
```

Gambar 3. Data *Cleaning*

Berdasarkan gambar 3. tidak terdapat *missing value* dan duplikasi data, sehingga dapat dilanjut ke tahap berikutnya.

3.1.3 Transformasi Data

Transformasi data bertujuan untuk mengubah data kategorikal menjadi numerik. Berikut adalah transformasi data pada tabel 2.

Tabel 2. Transformasi Data

Jumlah tanggungan	Kondisi rumah	Status kepemilikan	Status kesejahteraan	keterangan
3	0	0	0	1
2	5	0	1	0
4	1	1	1	0
1	0	0	0	1
2	5	1	1	0

Berdasarkan tabel 2. Menunjukkan transformasi data kategorikal menjadi numerik pada atribut jumlah tanggungan, kondisi rumah, status kepemilikan dan status kesejahteraan.

3.1.4 Normalisasi Data

Normalisasi data dilakukan dengan menggunakan *min-max*. Berikut adalah normalisasi data pada tabel 3.

Tabel 3. Normalisasi Data

Jumlah tanggungan	Kondisi rumah	Status kepemilikan	Status kesejahteraan	Keterangan
0.6	0.0	0.000000	0.00	1.0
0.4	1.0	0.000000	0.25	0.0
0.8	0.2	0.333333	0.25	0.0
0.2	0.0	0.000000	0.00	1.0
0.4	1.0	0.333333	0.25	0.0

3.2 Implementasi Algoritma

3.2.1 Membuat Variable x dan y

Setelah data di transformasi, proses selanjutnya yaitu menentukan variable x dan y. Pada penelitian ini variable x ditentukan berdasarkan semua kolom, kecuali kolom keterangan, karena termasuk kedalam variabel y.

3.2.2 Split Data

Pada penelitian ini, split data dilakukan beberapa kali diantaranya, 70:30, 80:20 dan 90:10. Adapun split data dapat dilihat pada tabel 4.

Tabel 4. Jumlah data *training* dan *testing*

	70:30	80:20	90:10
Data Training	1284	1468	1651
Data Testing	551	367	184

Pada tabel 4. Split data dilakukan beberapa kali, karena untuk mengevaluasi performa model secara konsisten, dan menemukan rasio optimal yang memberikan keseimbangan terbaik antara pelatihan dan pengujian.

3.2.3 Algoritma Naive Bayes

Setelah split data dilakukan, tahap selanjutnya yaitu pengujian algoritma *Naive Bayes*, dengan menggunakan fungsi dari *Gaussian Naive Bayes* menggunakan *import library sklearn* pada *python*. Dengan *source code* pada gambar 4.

```
# Inisialisasi model Naive Bayes
naive_bayes_model = GaussianNB()

# Latih model dengan data pelatihan
naive_bayes_model.fit(X_train, y_train)
```

DecisionTreeClassifier
GaussianNB()

Gambar 4. Model *Naive Bayes*

Pada penelitian ini, percobaan klasifikasi dilakukan dengan membandingkan rasio. Berikut adalah nilai akurasi pada tabel 5.

Tabel 5. Perbandingan Hasil Akurasi

Data rasio	Akurasi
70:30	98.64%
80:20	98.52%
90:10	98.89%

Berdasarkan tabel 5 nilai akurasi terbaik dengan menggunakan rasio 90:10 dan menghasilkan akurasi 98.89%.

3.2.4 Algoritma C4.5

Tahap selanjutnya yaitu pengujian algoritma C4.5, dengan menggunakan fungsi dari *DecisionTreeClassifier*. Dengan *source code* pada gambar 5.

```
tree_dataset = DecisionTreeClassifier(random_state=0)
tree_dataset.fit(X_train, y_train)
```

DecisionTreeClassifier
DecisionTreeClassifier(random_state=0)

Gambar 5. Model C4.5

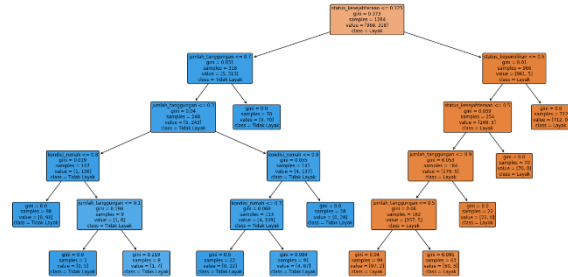
Pada penelitian ini, percobaan klasifikasi dilakukan dengan membandingkan rasio. Berikut adalah nilai akurasi pada tabel 6.

Tabel 6. Perbandingan Hasil Akurasi

Data rasio	Akurasi
70:30	99.26%
80:20	99.07%
90:10	99.26%

Berdasarkan tabel 6 nilai akurasi terbaik dengan menggunakan rasio 90:10 dan menghasilkan akurasi 99.26%.

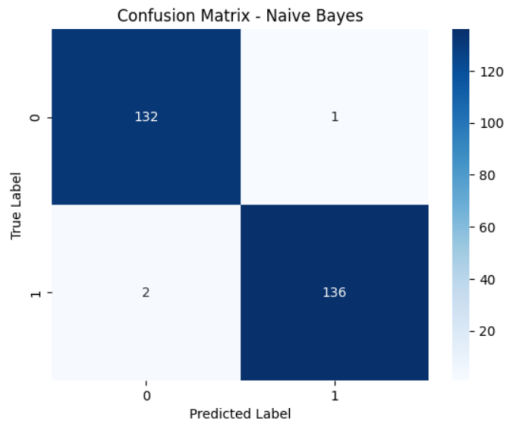
Adapun visualisasi pohon keputusan yang dibuat menggunakan pustaka *matplotlib* dan *scikit-learn tree* dari bahasa pemrograman *python* untuk menganalisis dan memodelkan data, dapat dilihat pada gambar 6.



Gambar 6. Pohon Keputusan C4.5

3.3 Evaluasi Algoritma Naive Bayes

Evaluasi pada penelitian ini menggunakan *confusion matrix* yang ditunjukkan pada gambar 7.



Gambar 7. Confusion matrix Algoritma Naive Bayes

Berdasarkan gambar 7 nilai *true* positif sebanyak 132, *true* negatif sebanyak 136, *false* positif sebanyak 1 dan *false* negatif sebanyak 2. Setelah mendapatkan nilai TP, TN, FP dan FN, maka selanjutnya mendapatkan nilai akurasi, presisi dan *recall* seperti Gambar 8.

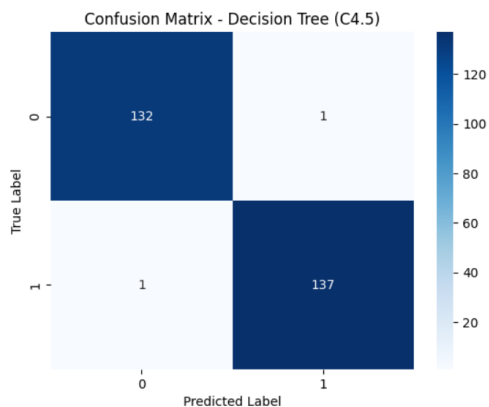
	precision	recall	f1-score	support
0	0.99	0.99	0.99	133
1	0.99	0.99	0.99	138
accuracy			0.99	271
macro avg	0.99	0.99	0.99	271
weighted avg	0.99	0.99	0.99	271

Gambar 8. Confusion matrix Algoritma Naive Bayes

Model *Confusion Matrix* dengan menggunakan klasifikasi algoritma *Naive Bayes* menghasilkan *Accuracy* sebesar 98.89%, *Precision* 99.25%, *Recall* 98.51%, dan *F1-score* 98.88%.

3.4 Evaluasi Algoritma C4.5

Evaluasi pada penelitian ini menggunakan *confusion matrix* yang ditunjukkan pada gambar 9.



Gambar 9. Confusion matrix Algoritma C4.5

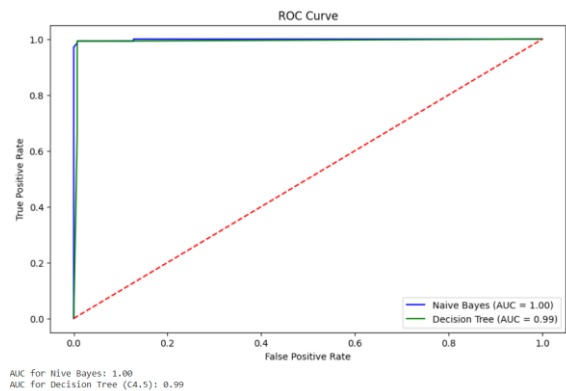
Berdasarkan gambar 9 nilai *true* positif sebanyak 132, *true* negatif sebanyak 137, *false* positif sebanyak 1 dan *false* negatif sebanyak 1. Setelah mendapatkan nilai TP, TN, FP dan FN, maka selanjutnya mendapatkan nilai akurasi, presisi dan *recall* seperti gambar 10.

	precision	recall	f1-score	support
0	0.99	0.99	0.99	133
1	0.99	0.99	0.99	138
accuracy			0.99	271
macro avg	0.99	0.99	0.99	271
weighted avg	0.99	0.99	0.99	271

Gambar 10. Confusion matrix Algoritma C4.5

Model *Confusion Matrix* dengan menggunakan klasifikasi C4.5 menghasilkan akurasi sebesar 99.26%, presisi 99.25%, *recall* 99.25%, dan *F1-score* sekitar 99.25%.

Evaluasi algoritma *Naive Bayes* dan C4.5 menggunakan kurva ROC/AUC dapat dilihat pada gambar 11.



Gambar 11. Kurva ROC/AUC

Berdasarkan gambar 11, menunjukkan perbandingan kinerja antara dua model klasifikasi, yaitu *Naive Bayes* dan C4.5. Kedua model menunjukkan kinerja yang sangat baik. Garis merah putus-putus merupakan performa model acak. Adapun kurva roc pada garis warna biru merupakan model *naive bayes* dan garis warna hijau merupakan model C4.5 berada pada sudut kiri atas. Adapun nilai AUC algoritma *naive bayes* sebesar 1.00 dan algoritma C4.5 sebesar 0.99. Performa dari model *Naive Bayes* dan C4.5 yang dihasilkan berdasarkan kurva ROC termasuk kedalam kategori kategori *Excellent Classification* karena berada pada rentang 0.90 - 1.00

3.5 Analisis Perbandingan Hasil Evaluasi

Dari hasil evaluasi *confusion matrix* dan kurva ROC diperoleh hasil C4.5 lebih tinggi. Karena

menghasilkan *Accuracy* sebesar 99.26%, *Precision* 99.25%, *Recall* 99.25%, *F1-score* 99.25% dan AUC 0.99. Sedangkan hasil pengujian *Naive Bayes* nilai *Accuracy* sebesar 98.89%, *Precision* 98.25%, *Recall* 98.51%, *F1-score* 98.88% dan AUC 1.00. Adapun perbandingan hasil evaluasi pada Tabel 7.

Tabel 7. Perbandingan Hasil Evaluasi

	Naive Bayes	C4.5
Accuracy	98.89%	99.26%
Precision	98.25%	99.25%
Recall	98.51%	99.25%
F1-score	98.88%	99.25%
AUC	1.00	0.99

KESIMPULAN

Pengujian model dalam penelitian ini menggunakan *Naive Bayes* dan C4.5. Adapun data yang digunakan berupa data PKH di Kecamatan Rawamerta Kabupaten Karawang pada tahun 2023 yang berjumlah 1834 data. Hasil evaluasi diperoleh nilai *accuracy* dari algoritma *naive bayes* sebesar 98.89%, *Precision* 98.25%, *Recall* 98.51%, *F1-score* 98.88%, dan AUC 1.00 termasuk kedalam kategori *Excellent classification* karena berada pada rentang 0.90-1.00, sedangkan algoritma C4.5 menghasilkan nilai *Accuracy* sebesar 99.26%, *Precision* 99.25%, *Recall* 99.25%, *F1-score* 99.25% dan AUC 0.99 termasuk kedalam kategori *Excellent classification* karena pada rentang 0.90-1.00. Berdasarkan perbandingan hasil evaluasi tersebut diperoleh kesimpulan bahwa algoritma C4.5 merupakan algoritma terbaik untuk menentukan penerima bantuan PKH. Dari hasil penelitian ini dapat digunakan sebagai acuan bagi pemerintah dalam menentukan metode terbaik untuk klasifikasi penerima PKH serta dapat memberikan kontribusi positif terhadap kebijakan pemerintah dalam mendistribusikan bantuan PKH kepada keluarga yang membutuhkan dengan lebih tepat sasaran. Adapun saran untuk penelitian selanjutnya yaitu dapat menggunakan metode lain atau menggabungkan lebih dari dua algoritma untuk mendapatkan hasil yang lebih akurat.

DAFTAR PUSTAKA

- [1] A. Triayudi, "Penerapan Data Mining Untuk Klasifikasi Penerima Dana Bantuan Sosial Dengan Menggunakan Algoritma K-Nearest Neighbor," *Building of Informatics, Technology and Science (BITS)*, vol. 5, no. 2, Sep. 2023, doi: 10.47065/bits.v5i2.3972.
- [2] F. A. Lubis, N. Ahmadi, B. Rahmani, I. K. Putri, U. I. Negeri, and S. Utara, "Strategi Pemberdayaan Usaha Mikro Kecil Menengah (UMKM) Melalui Program Mekaar Oleh PT. PNM Kota Medan Perspektif Ekonomi Islam".
- [3] A. Ramdani, C. D. Sofyan, F. Ramdani, M. F. A. Tama, and M. A. Rachmatsyah, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Masyarakat dalam Menerima Bantuan Sosial", *Jurnal Ilmiah Sistem Informasi*, vol. 1, no. 2, pp. 39–47, 2022.
- [4] M. Eka, P. Sompie, A. Aminudin, and J. Yogopriyatno, "Analisis Peran Pendamping Sosial Program Keluarga Harapan (PKH) Di Kecamatan Sungai Serut," *Jurnal Multidisiplin Dehasen*, vol. 1, no. 3, pp. 93–102, 2022.
- [5] J. Ilmu *et al.*, "AKSARA: Jurnal Ilmu Pendidikan Nonformal 909," vol. 09, no. 2, 2023, doi: 10.37905/aksara.9.2.909-922.2023.
- [6] N. Alfiah, "Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode Naive Bayes," *Jurnal Teknologi Informasi*, vol. 16, no. 1, 2021.
- [7] C. Cholifah, H. Hikmahyanti, and A. Ratna Juwita, "Analisis Sentimen Twitter Terhadap Omnibuslaw Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," *Jurnal Informatika Teknologi dan Sains*, vol. 4, no. 4, 2022.
- [8] W. Hadikristanto and I. Nasai, "Penerapan Algoritma Genetika dalam Memprediksi Penerima Program Keluarga Harapan Dengan Metode Naive Bayes," *Jurnal Teknologi Pelita Bangsa*, vol. 10, 2019.
- [9] I. Nurjanah, J. Karaman, I. Widaningrum, and D. Mustikasari, "Penggunaan Algoritma Naive Bayes Untuk Menentukan Pemberian Kredit Pada Koperasi Desa," *Journal of Computer Science and Information Technology E-ISSN*, vol. 3, no. 2, 2023.
- [10] L. Firdaus and T. Setiadi, "Perbandingan Algoritma Naive Bayes, Decision Tree, dan KNN untuk Klasifikasi Produk Populer Adidas US dengan Confusion Matrix," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 5, no. 2, p. 185, Dec. 2023, doi: 10.30865/json.v5i2.6124.
- [11] R. Pasya Sidik, T. Rohana and Rahmat, "Klasifikasi Data Pokok Pendidikan Sekolah Dasar dan Sekolah Menengah Peratama Menggunakan Algoritma Naive Bayes serta C4.5," *Scientific Student Journal for Information, Technology and Science*, vol. 5, no. 1, 2024.
- [12] A. Dina, I. Permana, F. Muttakin, and I. Maita, "Perbandingan Algoritma NBC, KNN, dan C4.5 Untuk Klasifikasi Penerima Bantuan Program Keluarga Harapan," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 3, p. 1079, Jul. 2023, doi: 10.30865/mib.v7i3.6316.
- [13] A. Irsyada, E. Haerani, M. Irsyad, F. Wulandari, and L. Afriyanti, "Penerapan Algoritma Naive Bayes Terhadap Klasifikasi Penerima Bantuan Program Keluarga Harapan (PKH)," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 5, no. 2, p. 457, Dec. 2023, doi: 10.30865/json.v5i2.7203.
- [14] N. Iftah Nella, N. Yudi Setiawan, and D. Eka Ratnawati, "Klasifikasi Penerima Bantuan Program Keluarga Harapan

- menggunakan Algoritme Decision Tree C4.5 (Studi Kasus: Desa Mlirip Kabupaten Mojokerto),” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 3, pp. 1332–1339, 2022.
- [15] A. Fitria Yulia and P. Bintoro, “K-Means Clustering in Determining the Eligibility of Recipients of Assistance for the Poor Case Study of Village Sukoharjo III,” *International Journal of Software Engineering and Informatics*, vol. 1, no. 1, pp. 63–70, 2023, [Online]. Available: <https://journal.aisyahuniversity.ac.id/index.php/IJosei>
- [16] M. S. Hartawan, Moh. Erkamim, S. Rachmawati, N. C. Santi, L. Legito, and S. Sepriano, “Penerapan Algoritma Supervised Learning untuk Klasifikasi Program Keluarga Harapan,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, 2023.
- [17] I. Nurjanah, J. Karaman, I. Widaningrum, and D. Mustikasari, “Penggunaan Algoritma Naive Bayes Untuk Menentukan Pemberian Kredit Pada Koperasi Desa,” *Journal of Computer Science and Information Technology E-ISSN*, vol. 3, no. 2, p. 77, 2023.
- [18] A. Nur, A. Rohim, A. I. Purnamasari, and I. Ali, “Komparasi Efektivitas Algoritma C4.5 dan Naive Bayes Untuk Menentukan Kelayakan Penerima Manfaat Program Keluarga Harapan,” *Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, 2024.
- [19] T. D. Ramadhan, D. Wahiddin, and E. E. Awal, “Klasifikasi Sentimen Terhadap Pinjaman Online (Pinjol) Menggunakan Algoritma Naive Bayes,” vol. IV, no. 1, 2023, [Online]. Available: www.tripadvisor.com
- [20] R. Yati, T. Rohana, and A. R. Pratama, “Klasifikasi Jenis Mangga Menggunakan Algoritma Convolutional Neural Network,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 3, p. 1265, Jul. 2023, doi: 10.30865/mib.v7i3.6445.
- [21] R. Rifaldi *et al.*, “Analisis Sentimen Pemboikotan Produk dengan Pendekatan Algoritma Naive Bayes Media Sosial X,” *Journal of Information System Research (JOSH)*, vol. 5, no. 4, pp. 940–946, 2024, doi: 10.47065/josh.v5i4.5420.
- [22] A. P. Pristiawati, I. Permana, Z. Zarnelly, and F. Muttakin, “Klasifikasi Penerima Bantuan Beras Miskin Menggunakan Algoritma K-NN, NBC dan C4.5,” *Building of Informatics, Technology and Science (BITS)*, vol. 5, no. 1, Jun. 2023, doi: 10.47065/bits.v5i1.3617.
- [23] F. Noer Azzahra *et al.*, “Penerapan Metode Naive Bayes Dalam Klasifikasi Spam SMS Menggunakan Fitur Teks Untuk Mengatasi Ancaman Pada Pengguna,” *Journal of*
- Information System Research (JOSH)*, vol. 5, no. 3, p. 880, 2024, doi: 10.47065/josh.v5i3.5070.