

# APPLYING KNN, NBC, AND C4.5 ALGORITHMS TO IDENTIFY ELIGIBILITY FOR NON-CASH FOOD AID

## PENERAPAN ALGORITMA KNN, NBC, DAN C 4.5 UNTUK KLASIFIKASI PENERIMA BANTUAN PANGAN NON-TUNAI

Rizki Pratama Putra Agri<sup>1</sup>, Inggih Permana<sup>2</sup>, Febi Nur Salisah<sup>3</sup>, Muhammad Jazman<sup>4</sup>, M. Afdal<sup>5</sup>  
<sup>1,2,3,4,5</sup> Universitas Islam Negeri Sultan Syarif Kasim Riau, Jl. HR. Soebrantas, No. 155, Km. 15  
Simpang Baru, 28293, Pekanbaru, Riau, Indonesia

Email: 11950311579@students.uin-suska.ac.id<sup>1</sup>, inggihpermana@uin-suska.ac.id<sup>2</sup>, febinursalisah@uin-suska.ac.id<sup>3</sup>, jazman@uin-suska.ac.id<sup>4</sup>, m.afdal@uin-suska.ac.id<sup>5</sup>

**Abstract** - The Indonesian government has implemented the Non-Cash Food Assistance (BPNT) program as an effort to improve people's welfare. However, in its implementation, there are still obstacles in the process of determining the right beneficiaries. Determining the right BPNT recipients is important to ensure that the assistance is received by people who really need it and to prevent budget misuse. This research aims to help the government to easily process data using three classification algorithms, namely K-Nearest Neighbour (K-NN), Naïve Bayes Classifier (NBC), and C4.5 in classifying BPNT recipient data in Air Molek Village, Indragiri Hulu Regency. K-NN, NBC, and C4.5 were chosen because they represent different approaches: K-NN is distance-based, NBC is probability-based, and C4.5 uses decision trees. The stages of the methodology used include data collection, data preprocessing, data splitting (Hold-Out), data balancing and model testing. The results showed that the K-NN algorithm got an accuracy of 70.45%, precision 68.34% recall 72.42%, NBC got an accuracy of 60.58%, precision 58.21%, recall 85.42%, and C 4.5 with an accuracy of 62.56%, precision 59.17%, recall 63.33%. The results of this study can help the government in developing a more objective and data-based decision support system for determining BPNT recipients. The limitation of this research is the use of data that is limited to only one of the data sources.

**Keywords** - BPNT, K-Nearest Neighbour, Naïve Bayes Classifier, C 4.5, Classification.

**Abstrak** - Pemerintah Indonesia telah menjalankan program Bantuan Pangan Non-Tunai (BPNT) sebagai upaya meningkatkan kesejahteraan masyarakat. Namun, dalam implementasinya, masih terdapat kendala dalam proses penentuan penerima bantuan yang tepat sasaran. Penentuan penerima BPNT yang tepat sasaran penting untuk memastikan bantuan diterima oleh masyarakat yang benar-benar membutuhkan dan mencegah penyalahgunaan anggaran. Penelitian ini bertujuan untuk membantu pemerintah agar mudah mengelola data menggunakan tiga algoritma klasifikasi, yaitu K-Nearest Neighbor (K-NN), Naïve Bayes Classifier (NBC), dan C4.5 dalam mengklasifikasikan data penerima BPNT di Kelurahan Air Molek, Kabupaten Indragiri Hulu. K-NN, NBC, dan C4.5 dipilih karena mewakili metode pendekatan yang berbeda: K-NN berbasis jarak, NBC berbasis probabilitas, dan C4.5 menggunakan pohon keputusan. Tahapan metodologi yang digunakan meliputi pengumpulan data, melakukan praproses data, membagi data (Hold-Out), penyeimbangan data serta pengujian model. Hasil penelitian menunjukkan bahwa Algoritma K-NN mendapatkan akurasi sebesar 70.45% presisi 68.34% recall 72.42%, NBC mendapatkan akurasi 60.58% presisi 58.21% recall 85.42%, dan C 4.5 dengan akurasi 62.56% presisi 59.17% recall 63.33%. Hasil penelitian ini dapat menjadi membantu pemerintah dalam mengembangkan sistem pendukung keputusan yang lebih objektif dan berbasis data untuk penentuan penerima BPNT. Keterbatasan penelitian ini yaitu penggunaan data yang terbatas hanya dari satu wilayah dan jumlah data yang relatif kecil.

**Kata Kunci** - BPNT, K-Nearest Neighbour, Naïve Bayes Classifier, C 4.5, Klasifikasi.

## I. PENDAHULUAN

Kemiskinan merupakan isu yang memiliki banyak dampak dan dipengaruhi oleh berbagai faktor, diantaranya faktor ekonomi, sosiologi, antropologi, kebijakan, teknologi, dan perubahan global. Dampak dari kemiskinan muncul dari berbagai bidang, termasuk pendidikan, kesehatan, kemampuan ekonomi, dan partisipasi masyarakat. Untuk menjawab tantangan ini, Pemerintah Indonesia telah mengimplementasikan beberapa program bantuan sosial, salah satunya yaitu Bantuan Pangan Non-Tunai (BPNT) [1]. Pemerintah telah mengimplementasikan program BPNT untuk membantu masyarakat prasejahtera. Program ini menggunakan mekanisme kartu elektronik atau rekening untuk menyalurkan bantuan setiap bulan kepada para penerima bantuan yang terdaftar pada bagian Program Keluarga Harapan (PKH) yang biasa disebut dengan Keluarga Penerima Manfaat (KPM). Dana tersebut secara khusus dialokasikan guna untuk bagian bidang pangan masyarakat di E-Warong dan pada pedagang pangan terdaftar yang harus terpenuhi pada syarat yang telah ditentukan serta dalam pekerjaan sama seperti Bank Pemerintah seperti Bank Rakyat Indonesia, Bank Tabungan Negara, dan lainnya [2].

Tujuan dari program ini yaitu untuk dapat menaikkan kualitas pada sumber daya manusia yang ada di Indonesia Dimana harapan inisiatif ini dapat menghasilkan generasi yang cerdas, bermental kuat, dan memiliki semangat untuk maju pada dunia kerja. Hal ini juga diharapkan dapat berkesinambungan pada pengurangan angka kemiskinan di Indonesia [3]. Permasalahan disini ialah penentuan penerima BPNT yang ada di Kelurahan Air Molek, Kabupaten Indragiri Hulu masih belum efektif. Terdapat beberapa kasus di mana bantuan ini tidak tepat sasaran. Misalnya, individu yang memiliki harta benda yang cukup banyak seperti sudah memiliki rumah, kendaraan, dan hewan ternak, tetap menerima bantuan, sementara individu lain yang memenuhi kriteria serta telah ditetapkan oleh pemerintah tidak menerima bantuan tersebut.

Tujuan pada penelitian ini yaitu mengembangkan model untuk mengklasifikasi data penerima BPNT dengan tujuan untuk meningkatkan keakuratan proses verifikasi data warga. Penerapan sistem klasifikasi yang efektif dan penting dalam memastikan bantuan pada BPNT dialokasikan terhadap rumah tangga miskin yang memenuhi syarat. Penerapan sistem klasifikasi akan memudahkan penentuan prioritas penerima manfaat BPNT berdasarkan kriteria yang telah ditetapkan. Data diperoleh melalui pesan teks dengan salah satu pegawai yang memiliki informasi relevan terkait topik penelitian dan data yang digunakan pada penelitian ini bersumber dari data sekunder dari pemerintah yaitu Data Terpadu Kesejahteraan Sosial (DTKS). Hipotesis penelitian ini yaitu Penerapan algoritma klasifikasi K-NN, NBC, dan C4.5 mampu meningkatkan akurasi dan efektivitas proses verifikasi data penerima BPNT dibandingkan metode manual atau konvensional. Hasil akhir dari pengolahan data ini dapat membantu pemerintah, khususnya di Kelurahan Air Molek, Kabupaten Indragiri Hulu, untuk menentukan kategori penerima BPNT berdasarkan data yang ada. Salah satu pendekatan untuk mengembangkan model klasifikasi adalah dengan menggunakan teknik *machine learning*, yaitu menggunakan beberapa algoritma seperti *Naïve Bayes Classifier* (NBC), *K-Nearest Neighbor* (K-NN), dan C4.5 [4].

Algoritma *K-Nearest Neighbor* (K-NN) adalah salah satu dari teknik klasifikasi yang sering digunakan dan dipopulerkan metodenya oleh Fix dan Hodges pada tahun 1951 dan telah diakui sebagai algoritma sederhana terbaik. K-NN adalah bagian klasifikasi yang sering digunakan untuk pengelompokan data dengan algoritma *supervised* [5]. Algoritma K-NN sendiri merupakan bagian dari *supervised learning* yang merupakan metode klasifikasi berdasarkan jarak data dengan data lainnya [6]. Metode K-NN ini berdasarkan dari kedekatan jarak terhadap data dalam ruang n-dimensi dengan pilihan ini mudah dipelajari dan model bisa dibentuk secara cepat dan tepat. Meskipun memiliki kelemahan yaitu harus menemukan nilai k yang terbaik, sehingga hal ini bisa dilakukan dengan pengulangan pencarian klasifikasi terhadap nilai k berbeda sampai ditemukan pengklasifikasian dengan akurasi terbaik [7]. K-NN dipilih karena kedekatan data dapat bekerja

baik untuk data yang memiliki pola spasial yang jelas, tetapi cenderung lambat saat digunakan pada dataset besar dan sangat sensitif terhadap pemilihan parameter K.

Algoritma *Naïve Bayes* adalah salah satu metode pengklasifikasian didasarkan pada teorema Bayes yang diperkenalkan oleh seorang ilmuwan Inggris yaitu Thomas Bayes. Metode ini biasa di kenal dengan sebutan Bayesian atau yang sering disebut *Naïve Bayes*. Metode ini menggunakan nilai probabilitas dan statistik yang diperuntukan dalam memprediksi suatu data yang belum teridentifikasi kelasnya, data yang akan datang selanjutnya di olah dengan data yang sudah ada, atau data sebelumnya [3]. Dari Olson Delen (2008), menyampaikan bahwa *Naïve Bayes* melakukan perhitungan probabilitas untuk masing-masing kelas keputusan yang memiliki asumsi bahwa atribut objek bersifat independen. Probabilitas yang terlibat dalam menghasilkan estimasi akhir dihitung sebagai jumlah frekuensi dari tabel keputusan “master” [8]. *Naive Bayes* dipilih karena unggul dalam kecepatan dan efisiensi komputasi, menjadikannya cocok untuk dataset besar, meskipun performanya sangat bergantung pada asumsi independensi antar fitur.

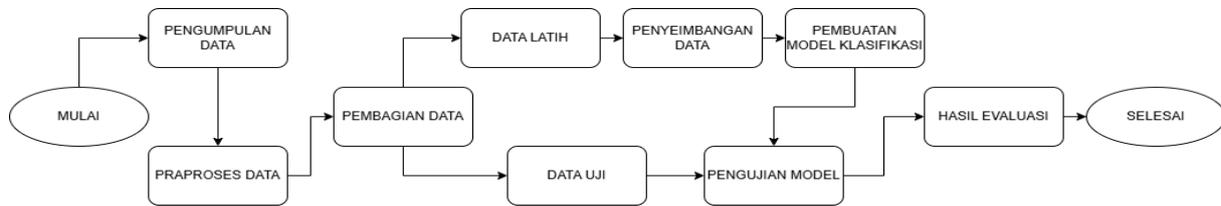
Algoritma C4.5 yaitu sebuah metode klasifikasi dimana metode ini memakai sebuah pohon keputusan untuk memodelkan proses klasifikasi. Algoritma ini digunakan untuk melakukan klasifikasi prediksi atau segmentasi, di mana pohon keputusan terdiri dari node akar, node internal, dan daun yang merepresentasikan nilai target dari salah satu kelas [9]. Algoritma C4.5 memiliki keunggulan dalam menangani nilai yang hilang dengan menggunakan data numerik dan diskrit, serta mengatasi pemangkasan. Algoritma ini merupakan modifikasi lebih lanjut dari algoritma ID3 dikembangkan oleh J. Ross Quinlan [10]. Algoritma ini termasuk kedalam *supervised learning*, sehingga membutuhkan sebuah klasifikasi variabel target/sasaran. Satu set data latih yang disiapkan dalam melatih algoritma terhadap nilai-nilai variabel target/sasaran. Algoritma ini dievaluasi dengan menggunakan pendekatan yang statistik seperti perhitungan entropi dan *information gain*. *Information gain* merupakan informasi atau ukuran keefektifan suatu atribut untuk klasifikasi data [11]. Ketiga algoritma yang telah disebutkan sebelumnya dipilih karena memiliki kinerja tinggi dalam mengklasifikasikan data. Ketiga algoritma ini juga memiliki kelebihan dibidang komputasi sederhana serta memiliki akurasi cukup baik.

Penelitian yang dilakukan oleh Aulia Dina (2023) tentang perbandingan algoritma K-NN, NBC, dan C 4.5 untuk klasifikasi PKH menghasilkan nilai Recall dari algoritma NBC sebesar 38.03% [4]. Penelitian yang dilakukan oleh Rizal Amegia Saputra (2021) tentang kepuasan siswa terhadap kinerja guru menggunakan C 4.5 menghasilkan akurasi sebesar 91.54 % dan AUC sebesar 0.986 [12]. Penelitian yang dilakukan oleh Yuris Alkhalifi (2020) tentang perbandingan NBC dan C 4.5 untuk penerima BPNT menghasilkan nilai akurasi NBC adalah sebesar 98.89% dan nilai akurasi C.45 adalah sebesar 95.00% [13]. Penelitian yang dilakukan oleh Hanna Syajida (2024) tentang klasifikasi penerima BPNT menggunakan K-NN mendapatkan hasil akurasi sebesar 93.14%, *weighted mean precision* sebesar 79.53%, dan *weighted mean recall* sebesar 74.49% [14]. Penelitian yang dilakukan oleh Haidah Putri (2021) mengenai penerima BPNT menggunakan NBC dan K-NN mendapatkan hasil akurasi NBC sebanyak 99.89% dan K-NN 66.46% [15].

## II. SIGNIFIKASI STUDI

Penelitian ini sejalan dengan arah kebijakan pemerintah Indonesia dalam meningkatkan tata kelola pemerintahan berbasis data melalui pemanfaatan big data dan kecerdasan buatan (AI), sebagaimana tertuang dalam Peraturan Presiden Nomor 39 Tahun 2019 tentang Satu Data Indonesia. Penggunaan algoritma klasifikasi dalam penelitian ini mendukung upaya digitalisasi proses verifikasi data penerima bantuan sosial, khususnya dalam memperkuat akurasi dan ketepatan sasaran program

Bantuan Pangan Non Tunai (BPNT) yang menjadi bagian dari sistem Data Terpadu Kesejahteraan Sosial (DTKS) di bawah naungan Kementerian Sosial. Dengan demikian, hasil penelitian ini dapat menjadi rekomendasi teknis bagi pengambil kebijakan dalam meningkatkan transparansi dan efisiensi program bantuan sosial berbasis data.



**Gambar 1** Metode Penelitian

### A. Pengambilan Data

Data penerima BPNT diperoleh melalui wawancara daring yang dilakukan wawancara dengan salah satu pegawai yang ada di Kelurahan Air Molek. Data yang didapatkan adalah Data Terpadu Kesejahteraan Sosial (DTKS) pada tahun 2024 di Kelurahan Air Molek.

### B. Praproses Data

Pada tahapan praproses data ini dilakukan beberapa langkah penting, seperti membersihkan data *null* untuk menghilangkan entri yang tidak memiliki nilai. Kemudian melakukan pelabelan data dan transformasi data, yaitu mengubah data kategorikal menjadi numerik agar bisa diproses oleh algoritma. Lalu, untuk upaya menjaga kerahasiaan negara dan melindungi privasi individu, nama-nama warga negara diubah menjadi kode unik yang tidak mengandung informasi pribadi.

### C. Pembagian Data

Pembagian data ini merupakan membagi data menjadi dua, yaitu data latih dan data uji dilakukan dengan rasio *Hold-Out validation*. *Hold-Out validation* adalah pembagian data set yaitu data uji dan data latih, misal 0.2, 20% data dipakai untuk data uji dan sisanya 80% data akan dipakai untuk data latih [16]. *Hold-Out* yang digunakan pada penelitian ini ialah 80:20.

### D. Penyeimbangan Data

Penyeimbangan data ini dilakukan dengan menggunakan *Synthetic Minority Over-Sampling Technique* (SMOTE) berguna dalam menyeimbangkan data pada setiap kelas dalam data latih [17].

### E. Pembuatan dan Pengujian Model

Pengujian model ini dilakukan dengan menggunakan tiga algoritma, yaitu *Naïve Bayes Classifier* (NBC), *Algoritma K-Nearest Neighbor* (K-NN), dan C4.5 dan dijalankan menggunakan Google Colab, yaitu platform berbasis cloud untuk menjalankan kode secara online. Nilai K yang digunakan dalam algoritma K-NN adalah 1. Naive Bayes unggul dalam kecepatan dan efisiensi komputasi, menjadikannya cocok untuk dataset besar, meskipun performanya sangat bergantung pada asumsi independensi antar fitur. Di sisi lain, K-NN yang berbasis pada kedekatan data dapat bekerja baik untuk data yang memiliki pola spasial yang jelas, tetapi cenderung lambat saat digunakan pada dataset besar dan sangat sensitif terhadap pemilihan parameter K.

### F. Hasil dan Evaluasi

Untuk setiap model klasifikasi, hasil pengukuran kinerja dihitung menggunakan *confusion matrix*. Lalu dilakukan pencarian nilai yang presisi, akurasi, dan recall. Selanjutnya, melakukan analisis perbandingan antara presisi, akurasi, dan recall dari tiga algoritma yang berbeda (NBC, K-NN, dan C4.5). Algoritma K-NN diharuskan untuk melakukan perbandingan pada nilai K dari algoritma K-

NN, karena sangat penting untuk memastikan nilai parameter yang paling tepat. Berikut rumus menghitung nilai presisi, akurasi, dan recall.

$$\text{Akurasi} = \frac{(TP+TN)}{TP+TN+FP+FN} \times 100\%$$

$$\text{Presisi} = \frac{(TP+TN)}{TP+FP} \times 100\%$$

$$\text{Recall} = \frac{(TP)}{TP+FN} \times 100\%$$

---

Ref: A. Dina, I. Permana, F. Muttakin, and ..., "Perbandingan Algoritma NBC, KNN, dan C4.5 Untuk Klasifikasi Penerima Bantuan Program Keluarga Harapan," *J. Media ...*, vol. 7, no. 3, pp. 1079–1087, 2023, doi: 10.30865/mib.v7i3.6316

Penelitian ini mengisi kekosongan dalam literatur terkait penerapan algoritma klasifikasi untuk mendukung penyaluran bantuan pemerintah yang lebih tepat sasaran. Berbeda dari studi sebelumnya, penelitian ini menggunakan *Naïve Bayes Classifier* (NBC), *Algoritma K-Nearest Neighbor* (K-NN), dan C4.5 untuk memprediksi kelayakan penerima bantuan menggunakan data yang sudah ada. Inovasi dari penelitian ini terletak pada pemanfaatan teknik seperti praproses data, pembagian data, penyeimbangan datadadn seterusnya untuk meningkatkan akurasi klasifikasi, serta penerapan model ini pada data nyata yang merepresentasikan kondisi sosial masyarakat.

Penelitian ini memiliki potensi dampak sosial yang signifikan, terutama dalam meningkatkan efektivitas distribusi bantuan sosial seperti Bantuan Pangan Non-Tunai (BPNT). Dengan menggunakan algoritma klasifikasi, identifikasi penerima bantuan dapat dilakukan lebih akurat, objektif, dan transparan, sehingga mengurangi risiko salah sasaran dan kebocoran anggaran. Peningkatan ketepatan dalam distribusi ini dapat mempercepat penyaluran bantuan kepada yang membutuhkan, memperkuat perlindungan sosial, dan meningkatkan kualitas hidup penerima BPNT melalui akses pangan yang lebih terjamin. Secara keseluruhan, penelitian ini berpotensi mengurangi kemiskinan dan memberikan kontribusi positif bagi kesejahteraan masyarakat. Pemilihan algoritma K-NN, Naïve Bayes, dan C4.5 dalam penelitian ini memberikan keuntungan besar dalam hal akurasi dan efisiensi dalam klasifikasi dibandingkan dengan metode tradisional yang lebih umum, seperti penilaian berbasis kriteria manual atau metode berbasis aturan. Metode tradisional sering kali melibatkan penilaian subjektif dan proses manual yang lebih rentan terhadap kesalahan manusia dan kesalahan alokasi bantuan. Sebaliknya, algoritma klasifikasi yang digunakan memiliki keunggulan dalam memproses data besar dan menganalisis pola yang tersembunyi dalam data yang kompleks, memungkinkan pengambilan keputusan lebih cepat dan akurat.

Penelitian ini dapat menjadi dasar penting untuk eksplorasi lebih lanjut, seperti pengembangan model ensemble yang menggabungkan beberapa algoritma demi meningkatkan akurasi klasifikasi, penerapan teknik praproses lanjutan seperti seleksi fitur dan normalisasi untuk meningkatkan kinerja model, serta validasi pada konteks geografis dan demografis yang berbeda. Selain itu, studi ini membuka peluang untuk penerapan lebih luas di bidang lain seperti klasifikasi penerima bantuan pendidikan atau subsidi energi, serta integrasi dengan sistem informasi geografis (GIS) untuk mendukung pengambilan keputusan berbasis lokasi.

### III. HASIL DAN PEMBAHASAN

#### A. Implementasi Algoritma

Data yang didapatkan merupakan data pada tahun 2024 dan disajikan dalam format Ms. Excel (.xlsx) yang mencakup total 204 calon penerima bantuan BPNT di Desa Air Molek. Data ini menggunakan lebih kurang 33 atribut yang akan digunakan untuk menentukan kelayakan warga dalam menerima bantuan sosial. Sebanyak 85 warga berhak mendapatkan bantuan BPNT, sementara 119 warga dinyatakan tidak berhak mendapatkan bantuan BPNT. Warga yang dinyatakan tidak layak menerima bantuan BPNT dikategorikan ke dalam dua kelas yang berbeda, yaitu kelas "Yes", yang mencakup warga yang berhak mendapatkan bantuan BPNT, dan kelas "No" yang mencakup warga yang tidak berhak mendapatkan bantuan BPNT. Berikut ini adalah penjelasan mengenai atribut-atribut data yang diambil dan akan diolah dari data Data Terpadu Kesejahteraan Sosial (Tabel 1). Pada tabel 2 merupakan sampel data calon penerima bantuan BPNT.

Tabel 1 Penjelasan Mengenai Atribut Data

Atribut	Kode	Keterangan	Deskripsi					
Nama	A1	Nama warga	Nama Warga					
Jumlah ART	A2	Total anggota rumah tangga	Sesuai dengan banyaknya anggota keluarga disetiap rumah					
Status Kepemilikan Bangunan	A3	Status bangunan tempat tinggal yang ditempati	1. Milik Sendiri	2. Kontrak/Sewa	3. Bebas Sewa	4. Dinas		
Status Lahan	A4	Status lahan tempat tinggal yang ditempati	1. Milik sendiri	2. Milik orang lain	3. Tanah negara	4. Lainnya		
Jenis Lantai	A5	Jenis lantai terluas	1. Marmer/granit	2. Keramik	3. Parket/vinil/permadani	4. Ubin/tegel/teraso		
Jenis Dinding	A6	Jenis dinding terluas	5. Kayu/papan berkualitas tinggi	6. Semestara/bata	7. Bambu	8. Kayu/papan		
			9. Tanah	10. Lainnya	1. Tembok	2. Plasteran anyaman bambu		
			3. Kayu	4. Anyaman bambu	5. Batang kayu	6. Bambu		
			7. Lainnya					
Kondisi Dinding	A7	Kondisi dinding terluas	1. Bagus/kualitas tinggi	2. Jelek/kualitas rendah				
Atap	A8	Jenis atap terluas	1. Beton/genteng beton	2. Genteng keramik	3. Genteng metal	4. Genteng tanah liat		
Kondisi Atap	A9	Kondisi atap terluas	5. Asbes	6. Seng	7. Sirap	8. Bambu		
			9. Jerami/ijuk/daun	10. Lainnya	1. Bagus/kualitas tinggi	2. Jelek/kualitas rendah		
Jumlah Kamar	A10	Total jumlah kamar tidur	Menyesuaikan jumlah kamar pada setiap rumah					
Sumber Air Minum	A11	Sumber air minum	1. Air kemasan bermerk	2. Air isi ulang	3. Leding meteran	4. Leding eceran		
Sumber Penerangan	A12	Sumber penerangan utama	5. Sumur bor/pompa	6. Sumur terlindung	7. Sumur tak terlindung	8. Mata air terlindung		
			9. Mata air tidak terlindung	10. Air sungai/danau/waduk	11. Air hujan	12. Lainnya		
			1. Listrik PLN	2. Listrik non PLN	3. Bukan listrik			

Daya	A13	Daya yang terpasang	1. 450 watt 3. 1.300 watt 5. >2.200 watt	2. 900 watt 4. 2.200 watt 6. Tanpa penerangan
Bahan Bakar masak	A14	Bahan bakar utama untuk memasak	1. Listrik 3. Gas > 3 kg 5. Minyak tanah 7. Arang 9. Tidak memasak dirumah	2. Gas 3 kg 4. Gas kota/biogas 6. Briket 8. Kayu bakar
Fasilitas BAB	A15	Penggunaan fasilitas tempat BAB	1. Sendiri 3. Tidak ada	2. Bersama
Kloset	A16	Jenis kloset	1. Leher angsa 3. Cemplung/cubluk	2. Plengsengan 4. Tidak pakai
Buang Tinja	A17	Tempat pembuangan akhir tinja	1. Tangki 3. Lubang tanah 5. Pantai/tanah lapang/kebun	2. Plengsengan 4. Kolam/sawah/sungai/danau/laut 6. Lainnya
Tabung Gas	A18	Memiliki tabung gas 5,5 atau lebih	1. Ya	2. Tidak
Lemari Es	A19	Memiliki lemari es/kulkas	1. Ya	2. Tidak
AC	A20	Memiliki AC	1. Ya	2. Tidak
Pemanas Air	A21	Memiliki pemanas air ( <i>water heater</i> )	1. Ya	2. Tidak
Telepon	A22	Memiliki telepon rumah (PSTN)	1. Ya	2. Tidak
TV	A23	Memiliki televisi	1. Ya	2. Tidak
Emas	A24	Memiliki perhiasan & tabungan (senilai 10 gr emas)	1. Ya	2. Tidak
Komputer/Laptop	A25	Memiliki komputer/laptop	1. Ya	2. Tidak
Sepeda	A26	Memiliki sepeda	1. Ya	2. Tidak
Motor	A27	Memiliki motor	1. Ya	2. Tidak
Mobil	A28	Memiliki mobil	1. Ya	2. Tidak
Perahu	A29	Memiliki perahu	1. Ya	2. Tidak
Motor Tempel	A30	Memiliki motor tempel	1. Ya	2. Tidak
Kapal	A31	Memiliki kapal	1. Ya	2. Tidak
Aset Tidak Bergerak	A32	Rumah tangga memiliki aset tidak bergerak	a. Lahan 1. Ya	2. Tidak
Rumah Lain	A33	Rumah tangga memiliki aset rumah di tempat lain	b. Rumah ditempat lain 1. Ya	2. Tidak
Status	Kelas atau Label	Jika YES menerima BPNT	YES	NO

Jika NO tidak menerima BPNT

**Tabel 2** Sampel Data Warga BPNT

No	A1	A2	A3	A4	A5	A6	A7	A8	A9	...	A30	A31	A32	A33	STATUS
1	XXXX	0	0	0	0	0	0	0	0	...	0	0	0	0	NO
2	XXXX	2	1	1	2	1	1	6	1	...	2	2	2	4	NO
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
203	XXXX	3	1	1	6	1	1	6	1	...	2	2	2	4	YES
204	XXXX	3	1	1	2	1	1	6	2	...	2	2	2	4	NO

**B. Praproses Data**

Pada data pertama yang totalnya 204 data, dilakukan pembersihan data pada 41 baris yang memiliki nilai *null*, lalu bagian tersebut dihapus, dan kemudian data warga yang awalnya 204 menjadi 164 baris data. Hasil data menampilkan bahwa ada 72 warga yang layak mendapatkan bantuan BPNT dan 92 warga yang tidak layak mendapatkan bantuan BPNT. Berikut merupakan hasil dari pembersihan data pada tabel 3.

**Tabel 3** Sampel Cleaning Data

No	A1	A2	A3	A4	A5	A6	A7	A8	A9	...	A30	A31	A32	A33	STATUS
1	w1	2	1	1	2	1	1	6	1	...	2	2	2	4	NO
2	w2	3	1	1	6	1	6	2	2	...	2	2	2	4	YES
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
163	w163	3	1	1	6	1	1	6	1	...	2	2	2	4	YES
164	w164	3	1	1	2	1	1	6	2	...	1	2	2	4	NO

**C. Pembagian Data**

Dalam penelitian ini, dua teknik validasi digunakan untuk memastikan kualitas dan keandalan model klasifikasi yang dibangun, yaitu *hold-out validation* dan *cross-validation*. *Hold-out validation* digunakan untuk membagi data menjadi dua set: satu set untuk pelatihan dan satu set untuk pengujian. Teknik ini relatif sederhana dan cepat karena hanya memerlukan satu pembagian data. Namun, kelemahannya adalah hasil model dapat sangat bergantung pada bagaimana data dibagi, sehingga dapat menyebabkan ketidakstabilan jika pembagian data tidak representatif. Selain itu, hanya sebagian data yang digunakan untuk evaluasi, yang mengurangi seberapa menyeluruh model diuji. Di sisi lain, *cross-validation*, khususnya *k-fold cross-validation*, membagi data menjadi beberapa bagian (*fold*), dan model dilatih serta diuji pada setiap *fold* secara bergantian. Teknik ini memberikan evaluasi yang lebih *robust* karena model diuji pada berbagai subset data, serta memastikan bahwa seluruh data digunakan untuk pelatihan dan pengujian. Meskipun memberikan gambaran yang lebih stabil dan akurat tentang performa model, *cross-validation* memerlukan waktu komputasi yang lebih lama dan proses implementasi yang lebih kompleks. Kedua teknik ini saling melengkapi, dengan *hold-out validation* memberikan kecepatan dan kesederhanaan, sementara *cross-validation* memberikan evaluasi yang lebih andal dan akurat, memungkinkan peneliti untuk memperoleh gambaran yang lebih komprehensif mengenai kemampuan model klasifikasi.

**D. Penyeimbangan Data**

Penelitian ini menggunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) dilakukan karena data yang digunakan dalam penelitian mengalami ketidakseimbangan kelas, di mana jumlah data pada kelas mayoritas (penerima yang tidak layak) jauh lebih besar dibandingkan kelas minoritas (penerima yang layak). Sebelum penerapan SMOTE, model menunjukkan akurasi dan presisi yang rendah, namun recall yang tinggi, karena ketidakseimbangan kelas membuat model

lebih cenderung mengidentifikasi kelas minoritas meskipun dengan tingkat kesalahan yang tinggi. Setelah SMOTE diterapkan, yang menyeimbangkan data dengan menambah sampel sintesis kelas minoritas, terdapat peningkatan signifikan dalam akurasi dan presisi, sambil tetap menjaga recall yang baik, sehingga model lebih akurat dalam mengidentifikasi penerima bantuan yang tepat sasaran. Berikut beberapa hasil sebelum (Tabel 4) dan setelah menggunakan SMOTE (Tabel 5).

**Tabel 4** Sebelum SMOTE

Nilai K	Akurasi	Presisi	Recall
1	58.66%	56.10%	75.61%
3	61.23%	57.66%	79.09%
5	62.10%	59.50%	80.91%

**Table 5** Setelah SMOTE

Nilai K	Akurasi	Presisi	Recall
1	70.45%	68.34%	72.42%
3	62.54%	61.31%	60.50%
5	62.56%	63.45%	60.58%

**E. Hasil Klasifikasi**

Berdasarkan hasil percobaan, algoritma K-NN (K=1) mendapatkan nilai akurasi paling tinggi daripada algoritma lainnya yakni sebesar 70.45%, dan nilai presisinya sebesar 68.34%, namun nilai *recall* cukup kecil dengan NBC yakni sebesar 72.42%. Algoritma NBC mendapatkan nilai yang akurasinya dibawah K-NN yakni sebesar 60.58%, dan nilai presisinya sebesar 58.21%, namun nilai recallnya paling besar daripada algoritma lain yakni sebesar 85.42%. Sedangkan algoritma C4.5 mendapatkan nilai diantara kedua algoritma tersebut dengan nilai akurasi sebesar 63.23%, nilai presisinya sebesar 59.14%, dan nilai recallnya sebesar 64.58%. Berikut *Confusion matrix* untuk K-NN, NBC, dan C 4.5 yang ditampilkan pada Tabel 6, Tabel 7, dan Tabel 8.

**Tabel 6** Confusion Matrix K-NN

K-NN		
	Prediksi Negatif	Prediksi Positif
Aktual Negatif	8	8
Actual Positif	6	11

**Tabel 7** Confusion Matrix NBC

NBC		
	Prediksi Negatif	Prediksi Positif
Aktual Negatif	3	13
Actual Positif	3	14

**Tabel 8** Confusion Matrix C 4.5

C 4.5		
	Prediksi Negatif	Prediksi Positif
Aktual Negatif	9	7
Actual Positif	9	8

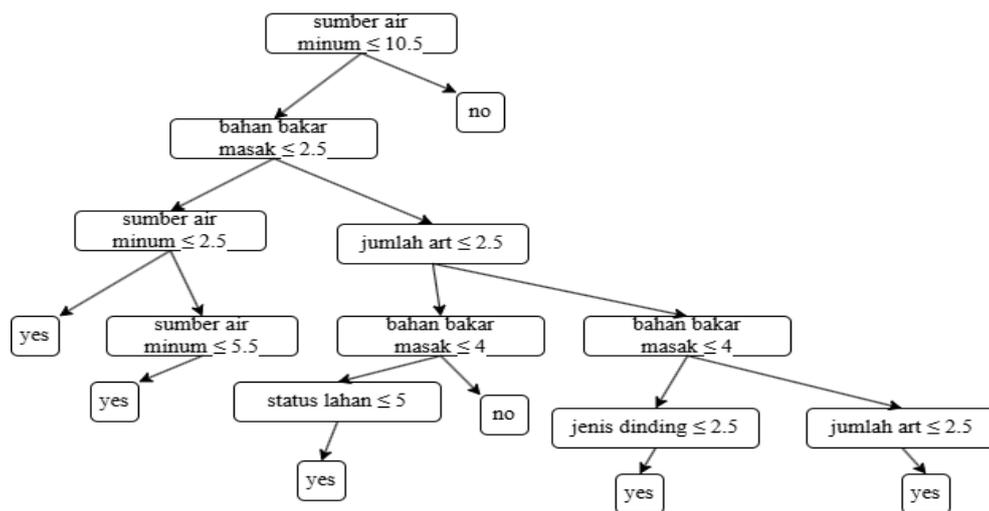
Untuk algoritma K-NN akan dibuat 10 kali percobaan dimana pada nilai K yang berbeda, yaitu 1, 3, 5, 7, 9, 11, 13, 15, 17, dan 19. Hasil dari nilai percobaan yang didapat bisa dilihat pada Tabel 9. Dari hasil tersebut, K terbaik dalam penelitian ini adalah K = 1. Nilai K=1 dipilih karena memberikan akurasi tertinggi pada model. Meskipun nilai K yang lebih kecil dapat menyebabkan overfitting dan nilai K yang lebih besar dapat menyebabkan underfitting, K=1 terbukti memberikan performa terbaik dalam hal akurasi pada dataset ini. Pemilihan nilai K yang tepat sangat bergantung pada karakteristik data, dan dalam kasus ini, K=1 menghasilkan hasil yang optimal.

**Tabel 9** Pencarian Nilai K Terbaik

Nilai K	Akurasi	Presisi	Recall
---------	---------	---------	--------

1	70.45%	68.34%	72.42%
3	62.54%	61.31%	60.50%
5	62.56%	63.45%	60.58%
7	61.23%	61.29%	59.25%
9	61.27%	60.41%	64.50%
11	59.89%	58.65%	63.08%
13	64.49%	64.52%	67.08%
15	61.91%	62.45%	64.42%
17	61.27%	61.24%	65.75%
19	62.56%	61.87%	69.75%

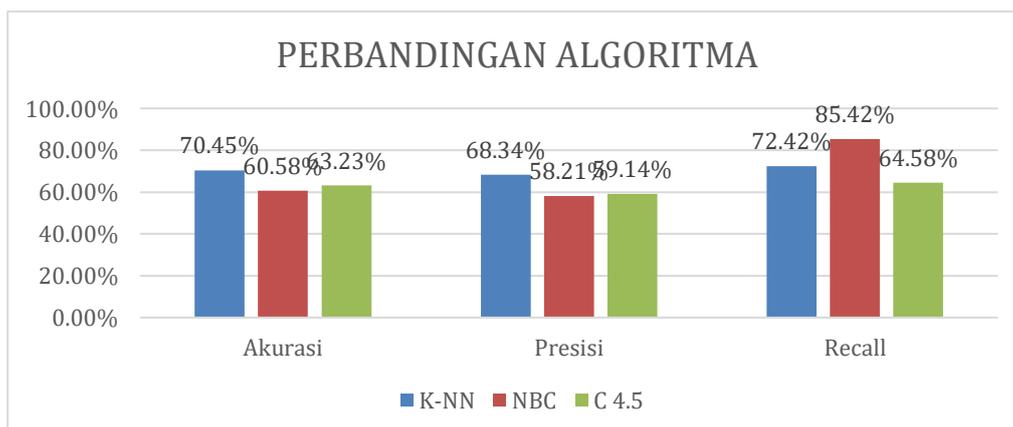
Pada Gambar 2 dibawah merupakan pohon keputusan yang dihasilkan oleh algoritma C 4.5. Penggunaan atribut di pohon keputusan hanya 5 atribut saja, yaitu: jumlah art, status lahan, jenis dinding, sumber air minum, dan bahan bakar masak.



Gambar 2 Pohon Keputusan

F. Perbandingan Performa

Dari hasil percobaan, penelitian ini mengutamakan hasil yang didapatkan oleh algoritma K-NN sebagai algoritma terbaik, karena K-NN menghasilkan akurasi tertinggi dibandingkan dengan algoritma NBC dan C4.5. Artinya, model yang dihasilkan oleh algoritma K-NN lebih mampu mengklasifikasikan dengan tepat keluarga yang layak dan tidak layak mendapatkan bantuan. Dalam konteks ini, akurasi diprioritaskan karena keseimbangan antara kesalahan positif dan negatif sama-sama penting, sehingga diperlukan model yang memiliki performa keseluruhan terbaik dalam membedakan kedua kategori tersebut. Grafik perbandingan kinerja model klasifikasi dapat dilihat pada Gambar 3.



Gambar 3 Grafik Perbandingan Algoritma

#### IV. KESIMPULAN

Penelitian ini menganalisis 204 data calon penerima BPNT di Desa Air Molek. Setelah proses pembersihan data, 40 baris data dihapus karena mengandung nilai nol, sehingga menyisakan 164 data yang terdiri dari 72 warga yang layak mendapatkan bantuan (*Yes*) dan ada 92 warga yang tidak layak mendapatkan bantuan (*No*). Hasil ini didapat dari evaluasi ketiga algoritma klasifikasi, yakni algoritma K-NN (K=1) mendapatkan hasil akurasi sebesar 70.45%, nilai dari presisi sebesar 68.34%, dan nilai recall sebesar 72.42%. Algoritma NBC memiliki akurasi 60.58%, presisi 58.21% dan recall 85.42%. Sementara itu, algoritma C4.5 memiliki akurasi 63.23%, presisi 59.14%, dan recall 64.58%. Algoritma K-NN menunjukkan performa terbaik pada nilai K=1, sementara algoritma C4.5 mampu mengecilkan jumlah atribut data yang awalnya dari 33 menjadi 5, sehingga dapat mengurangi kompleksitas proses klasifikasi. Jika dibandingkan dengan penelitian sebelumnya, hasil penelitian ini sejalan dengan temuan-temuan yang ada dalam literatur. Misalnya, penelitian oleh Aulia Dina (2023) yang membandingkan algoritma K-NN, NBC, dan C4.5 untuk klasifikasi PKH, menghasilkan nilai recall sebesar 38.03% untuk NBC. Sedangkan penelitian Yuris Alkhalifi (2020) yang membandingkan NBC dan C4.5 untuk penerima BPNT menghasilkan akurasi 98.89% untuk NBC dan 95% untuk C4.5. Hasil penelitian ini menunjukkan bahwa algoritma-algoritma tersebut dapat menghasilkan akurasi tinggi, namun penelitian ini menambah kontribusi baru dengan penerapan teknik SMOTE untuk menyeimbangkan data dan meningkatkan kinerja model. Selain itu, penelitian ini juga memperkenalkan penggunaan dua teknik validasi, hold-out dan cross-validation, yang memberikan evaluasi yang lebih komprehensif mengenai kestabilan model. Hasil penelitian ini secara langsung mendukung tujuan utama yang telah dinyatakan sebelumnya, yaitu mengembangkan model klasifikasi untuk meningkatkan akurasi verifikasi data penerima BPNT di Kelurahan Air Molek. Dengan membandingkan performa algoritma K-NN, NBC, dan C4.5, serta menerapkan teknik SMOTE untuk menangani ketidakseimbangan data, penelitian ini berhasil menunjukkan bahwa metode klasifikasi berbasis machine learning dapat memberikan hasil yang lebih akurat dan adil dibandingkan metode manual. Dengan demikian, tujuan penelitian untuk menciptakan sistem yang lebih objektif dan tepat sasaran dalam penyaluran bantuan sosial telah tercapai. Hasil akhir dari pengolahan data ini dapat membantu pemerintah, khususnya di Kelurahan Air Molek, Kabupaten Indragiri Hulu, untuk menentukan kategori penerima BPNT berdasarkan data yang ada. Penelitian ini membantu meningkatkan efektivitas distribusi BPNT dengan memastikan penerima bantuan lebih akurat dan tepat sasaran melalui algoritma klasifikasi. Hal ini mendukung kebijakan berbasis data, mengurangi kesalahan penyaluran, dan memperkuat perlindungan sosial. Keterbatasan penelitian ini yaitu penggunaan data yang terbatas hanya dari satu wilayah dan jumlah data yang relatif kecil. Saran bagi penelitian berikutnya yaitu menggunakan data yang lebih beragam atau memperluas cakupan wilayah agar model yang dihasilkan dapat diaplikasikan secara lebih luas dan akurat.

#### REFERENSI

- [1] L. R. Nadhifah and N. H. Mustofa, "Pengaruh PKH dan BPNT terhadap Kemiskinan dengan Pertumbuhan Ekonomi Sebagai Variabel Moderasi," *Al Maal J. Islam. Econ. Bank.*, vol. 3, no. 1, p. 12, 2021, doi: 10.31000/almaal.v3i1.4510.
- [2] P. Julianto, "Implementasi Program Bantuan Pangan Non Tunai (Bpnt) Di Kecamatan Sitingau Laut Kabupaten Kerinci," *Qawwam Leader's Writ.*, vol. 1, no. 1, pp. 38–43, 2020, doi: 10.32939/qawwam.v1i1.77.
- [3] D. A. Setiawan, R. Halilintar, and L. S. Wahyuniar, "Penerapan Metode Naive Bayes Untuk Klasifikasi Penentuan Penerima Bantuan PKH," *Pros. SEMNAS INOTEK (Seminar Nas. Inov. Teknol.*, vol. 5, no. 2, pp. 249–254, 2021.

- [4] A. Dina, I. Permana, F. Muttakin, and ..., "Perbandingan Algoritma NBC, KNN, dan C4. 5 Untuk Klasifikasi Penerima Bantuan Program Keluarga Harapan," *J. Media ...*, vol. 7, no. 3, pp. 1079–1087, 2023, doi: 10.30865/mib.v7i3.6316.
- [5] P. D. Rinanda, B. Delvika, S. Nurhidayarnis, N. Abror, and A. Hidayat, "Perbandingan Klasifikasi Antara Naive Bayes dan K-Nearest Neighbor Terhadap Resiko Diabetes pada Ibu Hamil," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 68–75, 2022, doi: 10.57152/malcom.v2i2.432.
- [6] S. K. P. Loka and A. Marsal, "Perbandingan Algoritma K-Nearest Neighbor dan Naive Bayes Classifier untuk Klasifikasi Status Gizi Pada Balita," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 8–14, 2023, doi: 10.57152/malcom.v3i1.474.
- [7] E. Rahma Meilaniwati and M. Fauzan, "Klasifikasi Penduduk Miskin Penerima PKH Menggunakan Metode Naive Bayes dan KNN," *J. Kaji. dan Terap. Mat.*, vol. 8, no. 2, pp. 75–84, 2022, [Online].
- [8] F. K. Pratama, D. W. Widodo, and N. Shofia, "Implementasi Metode Naive Bayes dalam Mengklasifikasi Penerima Program Keluarga Harapan ( PKH ) Desa Minggiran Kediri," *Semin. Nas. Inov. Teknol. UN PGRI Kediri*, pp. 23–28, 2021.
- [9] F. Fatmawati and N. Narti, "Perbandingan Algoritma C4.5 dan Naive Bayes Dalam Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Daring," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 4, no. 1, pp. 1–12, 2022, doi: 10.35746/jtim.v4i1.196.
- [10] N. I. Nella, N. Y. Setiawan, and D. E. Dian, "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,
- [11] A. Junaidi, Y. Yunita, S. Agustyani, C. I. Agustyaningrum, and Y. T. Arifin, "Klasifikasi Penerima Bantuan Sosial Menggunakan Algoritma C 4.5," *J. Tek. Komput.*, vol. 9, no. 1, pp. 77–82, 2023, doi: 10.31294/jtk.v9i1.14378.
- [12] R. A. Saputra, S. Wasiyanti, and D. Pribadi, "Information Gain Pada Algoritma C4.5 Untuk Klasifikasi Penerimaan Bantuan Pangan Non Tunai (Bpnt)," *Indones. J. Bus. Intell.*, vol. 4, no. 1, p. 25, 2021, doi: 10.21927/ijubi.v4i1.1757.
- [13] Y. Alkhalifi, A. Zumarniansyah, R. Ardianto, N. Hardi, and A. E. Augustia, "Comparison of Naive Bayes Algorithm and C.45 Algorithm in Classification of Poor Communities Receiving Non Cash Food Assistance in Wanasari Village Karawang Regency," *J. Techno Nusa Mandiri*, vol. 17, no. 1, pp. 37–42, 2020, doi: 10.33480/techno.v17i1.1191.
- [14] H. Syajida, A. Irma Purnamasari, and T. Suprapti, "Klasifikasi Penerima Bantuan Bpnt Menggunakan Algoritma K-Nn," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 608–614, 2024, doi: 10.36040/jati.v8i1.8357.
- [15] H. Putri, A. I. Purnamasari, A. R. Dikananda, O. Nurdiawan, and S. Anwar, "Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode NAIVE BAYES dan KNN," *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 331–337, 2021, doi: 10.47065/bits.v3i3.1093.
- [16] A. I. Shafarindu, E. Patimah, Y. M. Siahaan, A. W. Wardhana, B. V. Haekal, and D. S. Prasvita, "Klasifikasi Data Penjualan pada Supermarket dengan Metode Decision Tree," *Senamika*, vol. 2, no. 1, pp. 660–667, 2021
- [17] D. Kurniadi, F. Nuraeni, and M. Firmansyah, "Klasifikasi Masyarakat Penerima Bantuan Langsung Tunai Dana Desa Menggunakan Naive Bayes dan SMOTE," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 2, pp. 309–320, 2023, doi: 10.25126/jtiik.20231026453.