

## KLASIFIKASI PROGRAM BANTUAN SOSIAL MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR* (K-NN) (STUDI KASUS KECAMATAN MALANGBONG KABUPATEN GARUT)

Hamzah Nurrifqi Fakhri Fikrillah\*<sup>1</sup>

<sup>1</sup>Universitas Garut, Kabupaten Garut  
Email: <sup>1</sup>hamzah.nurrifqi2@gmail.com

\*Penulis Korespondensi

(Naskah masuk: 5 Maret 2024, diterima untuk diterbitkan: 19 Juni 2025)

### Abstrak

Isu kemiskinan di Indonesia menjadi perhatian utama pemerintah baik di tingkat pusat maupun daerah. Pemerintah merespons permasalahan ini melalui beberapa program salah satunya Program Bantuan Sosial (BANSOS). Melalui inisiatif ini, pemerintah berusaha menurunkan tingkat kemiskinan. Namun sering kali penentuan atribut yang digunakan untuk menunjukkan bahwa masyarakat masuk dalam kategori miskin dan berhak mendapatkan BANSOS masih sering kali kurang tepat. Konsep *data mining* diharapkan dapat mempermudah penyelesaian masalah penentuan atribut BANSOS. Dengan menggunakan algoritma *K-Nearest Neighbor* dan metode *Cross-Industry Standard Process for Data Mining* (Crisp-DM), diharapkan dapat ditemukan solusi yang lebih akurat dan efisien dalam menentukan kriteria penduduk miskin untuk menerima BANSOS. Dari hasil penelitian yang telah dilakukan menggunakan aplikasi *Rapidminer* dengan jumlah data sebanyak 19.943 keluarga serta pembagian rasio 70% untuk data latih, 30% data uji dan nilai  $K=598$  atau 3% dari total *dataset*, diperoleh tingkat akurasi model sebesar 86,70%. Pada model yang telah dibuat, dilakukan validasi terhadap model tersebut dengan menggunakan 600 data keluarga di luar *dataset* dengan kondisi geografis wilayah pegunungan, perkotaan dan pesisir. Hasil validasi ini menghasilkan akurasi sebesar 85,92%.

**Kata Kunci:** Program BANSOS, CRISP-DM, *Data Mining*, *K-Nearest Neighbor*, Klasifikasi.

## CLASSIFICATION OF SOCIAL ASSISTANCE PROGRAMS USING *K-NEAREST NEIGHBOR* ALGORITHM (K-NN) (CASE STUDY: MALANGBONG DISTRICT, GARUT REGENCY)

### Abstract

The issue of poverty in Indonesia is a major concern of the government at both the central and regional levels. The government responds to this problem through several programs, one of which is the Social Assistance Program (BANSOS). Through this initiative, the government is trying to reduce poverty levels. However, often the determination of the attributes used to indicate that people are in the poor category and are entitled to BANSOS is often inaccurate. The concept of data mining hopefully it will make things easier to solve the problem of determining BANSOS attributes. By using the *K-Nearest Neighbor* algorithm and the *Cross-Industry Standard Process for Data Mining* (Crisp-DM) method, it is hoped that a more accurate and efficient solution can be found in determining the criteria for poor people to receive BANSOS. From the results of research conducted using the *Rapidminer* application with a total of 19,943 families of data and a division ratio of 70% for training data, 30% for test data and a value of  $K=598$  or 3% of the total dataset, a model accuracy rate of 86.70% was obtained. In the model that has been created, validation of the model was carried out using 600 family data outside the dataset with geographical conditions in mountainous, urban and coastal areas. The results of this validation produced an accuracy of 85.92%.

**Keywords:** Social Assistant Program, Classification, CRISP-DM, *Data Mining*, *K-Nearest Neighbor*.

### 1. PENDAHULUAN

Pemerintah memiliki berbagai macam program untuk mendorong penghapusan kemiskinan di Indonesia, salah satunya adalah program bantuan

sosial atau disingkat dengan BANSOS. Beberapa macam BANSOS di antaranya adalah Bantuan Pangan Non Tunai (BPNT) dan Program Keluarga Harapan (PKH). Program BANSOS tersebut diberikan kepada masyarakat miskin untuk

memenuhi kebutuhan dasar seperti makanan pokok dan juga kebutuhan lainnya seperti kebutuhan sekolah atau kebutuhan dasar untuk balita, ibu hamil, lansia serta penyandang disabilitas.

Data penerima BANSOS BPNT dan PKH dihimpun dalam Data Terpadu Kesejahteraan Sosial (DTKS) yang dikelola oleh Pusat Data dan Informasi Kesejahteraan Sosial (PUSDATIN KESOS). Pada praktiknya, masih terdapat permasalahan dalam penyelenggaraan Program BANSOS yang telah diberikan oleh pemerintah, salah satunya adanya masyarakat miskin yang belum tersentuh oleh program BANSOS. Sebagai usaha untuk menangani hal tersebut perlu adanya analisis data yang akurat agar program BANSOS ini benar-benar tertuju pada masyarakat yang layak untuk mendapatkan bantuan (Rahmansyah & et al, 2020). Oleh karena itu, dengan adanya permasalahan tersebut, penulis melakukan analisis klasifikasi terhadap data warga yang berhak menerima bantuan sosial dengan *data mining* menggunakan algoritma *K-Nearest Neighbor* (K-NN).

Penelitian terdahulu yang dilakukan oleh (Nurjanah & Rifai, 2023) dengan judul Penerapan Algoritma *K-Nearest Neighbor* Untuk Klasifikasi Kelayakan Status Penduduk Miskin Di Desa Susukan Tonggoh. Tujuan penelitian ini mengklasifikasikan penduduk layak dikategorikan miskin dengan algoritma *K-Nearest Neighbor* dengan menggunakan atribut nama, alamat, pekerjaan, tempat tinggal, tingkat pendapatan, sumber pendapatan, dan status penduduk miskin. Jumlah data yang digunakan sebanyak 72 dengan nilai  $K = 5$ , tingkat akurasi yang didapatkan setelah pengolahan data mencapai 94.67%. *Output* yang dihasilkan dalam penelitian ini yaitu menentukan apakah penduduk tersebut dikategorikan sebagai penduduk miskin atau tidak miskin.

Penelitian lain yang dilakukan oleh (Kamari & dkk, 2022) dengan judul *Verification of Covid-19 Social Assistance Recipients using Naïve Bayes Classifier*. Penelitian ini membandingkan algoritma K-NN dan algoritma Naïve bayes, dengan menggunakan atribut data ID, nama, nama desa, RT, RW, Nomor SIAK, NIK, Jenis Kelamin, Tempat Lahir, Tanggal Lahir, Keterangan Pekerjaan. Tujuan dari penelitian ini mencari nilai akurasi terbesar dari dua metode algoritma tersebut. Hasilnya Algoritma Naïve Bayes mendapatkan nilai akurasi 93% sedangkan K-NN mendapatkan nilai akurasi sebesar 90%.

Penelitian selanjutnya yang dilakukan oleh (Fikrillah & dkk, 2023) dengan judul Klasifikasi penerima Bansos menggunakan algoritma *Naïve Bayes*. Tujuan dari penelitian ini yaitu mengklasifikasikan program BANSOS yang sesuai dengan kondisi individu tersebut. atribut yang digunakan dalam penelitian tersebut Nama desa, Nomor KK, NIK, Nama, tanggal lahir, jenis kelamin, pekerjaan, hubungan keluarga, jumlah anggota

keluarga, program bansos yang didapatkan, jumlah data yang digunakan sebanyak 100.445 jiwa. Keluaran yang dihasilkan dalam penelitian tersebut yaitu menentukan program BANSOS dengan nilai probabilitas tertinggi, jumlah program bansos yang dihitung atau label yang dijadikan output sebanyak 7 program bansos yaitu BPNT, PKH, PBI, NON BANSOS dan irisan BPNT, PKH, PBI. Akurasi yang didapatkan dalam penelitian tersebut mencapai 62.51%.

Berdasarkan hasil kajian sebelumnya, penelitian ini bertujuan untuk memberikan klasifikasi serta rekomendasi terhadap keluarga yang dinyatakan layak menerima BANSOS berdasarkan kriteria Pekerjaan Kepala Keluarga, Jumlah Anggota Keluarga, Pendidikan Terakhir Kepala Keluarga, Kepemilikan Rumah, Memiliki Simpanan Uang/ Perhiasan/ Ternak/ Lainnya, Jenis Atap, Jenis Dinding, Jenis Lantai, Sumber Penerangan, Bahan Bakar Memasak, Sumber Air Minum, Memiliki fasilitas Buang Air Besar, Ada Komponen Dalam Keluarga, serta Range Pendapatan. Adapun kriteria BANSOS yang akan diklasifikasi serta direkomendasikan adalah program BANSOS BPNT dan PKH.

Penelitian ini menggunakan algoritma K-NN yang mana algoritma ini mencari pola berdasarkan mayoritas kedekatan jarak atau kemiripan dari karakteristik objek (Muhammad & dkk, 2018), (Firasari & dkk, 2020).

Data yang digunakan dalam penelitian ini menggunakan data hasil verifikasi dan validasi dari Kecamatan Malangbong, Kabupaten Garut. Hasil dari penelitian ini diharapkan dapat menjadi acuan bagi pemerintah daerah khususnya Dinas Sosial selaku pelaksana urusan pemerintah daerah dari Kementerian Sosial dalam menangani permasalahan penentuan kriteria penerima bantuan sosial tersebut.

## 2. LANDASAN PUSTAKA

### 2.1. *Data Mining*

*Data mining* didefinisikan sebagai proses menemukan pola-pola yang diinginkan secara otomatis maupun semi otomatis yang digunakan untuk menemukan pengetahuan yang tersembunyi dalam basis data seperti pengambilan keputusan di masa depan. Proses ini melibatkan penggunaan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan manfaat yang terkandung dalam basis data yang besar (Mustafa & dkk, 2017), (Muhammad & dkk, 2018).

### 2.2. Klasifikasi

Klasifikasi merupakan suatu konsep atau model yang digunakan untuk memisahkan dan mengkategorikan data ke dalam kelas atau kategori tertentu dengan tujuan memprediksi kelas dari objek

data yang belum memiliki label kelas. Klasifikasi ini umumnya dilatih menggunakan *dataset* historis atau *dataset* dengan label yang telah diketahui sebelumnya. (Purwanto & Darmadi, 2018).

Proses Klasifikasi memiliki 2 tahap, yaitu *Training* (Pelatihan) dan Evaluasi. Proses *Training* bertujuan untuk mengajarkan model klasifikasi untuk mengenali pola-pola dalam data latih agar dapat mengklasifikasikan data uji dengan menggunakan algoritma *data mining* yang sesuai. Proses Evaluasi atau validasi bertujuan untuk mengukur tingkat akurasi dan performa model yang telah dilatih dengan menggunakan data yang tidak digunakan dalam tahap *training* (Saputra & Primadasa, 2018).

### 2.3. Algoritma *K-Nearest Neighbor* (K-NN)

*K-Nearest Neighbor* (K-NN) adalah salah satu metode *data mining* yang digunakan untuk mencari pola berdasarkan mayoritas kedekatan jarak atau kemiripan dari karakteristik objek. Algoritma ini memiliki beberapa kelebihan, yaitu kesederhanaan dan kecepatan konvergensi yang relatif tinggi (Muhammad & dkk, 2018), (Firasari & dkk, 2020). Tahapan awal dari algoritma K-NN sebagai berikut (Supriatna & Astuti, 2019) :

1. Menentukan jumlah K atau jumlah tetangga yang akan digunakan dalam proses pengklasifikasian.
2. Menghitung jarak *Euclidean* antara objek yang akan diklasifikasikan (data uji) dengan setiap objek dalam dataset.
3. Mengurutkan data berdasarkan jarak *Euclidean* secara *ascending*.
4. Mengumpulkan kategori (label) dari K objek terdekat berdasarkan nilai K yang telah ditetapkan sebelumnya.
5. Menentukan kategori mayoritas sebagai hasil klasifikasi data uji.

Secara umum untuk mendefinisikan jarak antara dua objek  $x$  dan  $y$ , digunakan rumus *euclidean* sebagai berikut :

$$d_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan :

$d_{xy}$  = *distance* atau jarak *euclidean*

$x$  = data latih

$y$  = data uji

$i$  = fitur ke- $i$

$n$  = jumlah data

### 2.4. *Rapidminer*

*RapidMiner* adalah sebuah perangkat lunak bersifat terbuka (*Open Source*) yang dirancang untuk membantu dalam pemrosesan data dalam bidang *data mining*. Perangkat lunak ini menyediakan berbagai macam metode dan algoritma yang diperlukan untuk

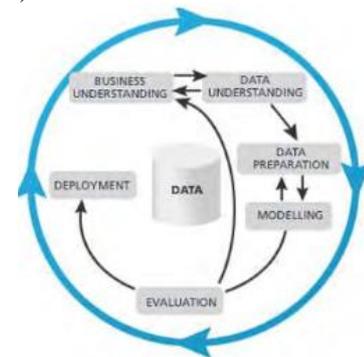
melakukan pengolahan *data mining* dalam satu paket aplikasi (Suyono & dkk, 2013).

### 2.5. *Cross Industry Standard Process For Data Mining*

*Cross Industry Standard Process For Data Mining* (SRISP-DM) adalah suatu metode atau kerangka kerja standar yang dikembangkan pada tahun 1996 untuk pengolahan *data mining*. Metode ini menyediakan proses yang terstruktur dan sistematis dalam memecahkan masalah menggunakan *data mining*.

CRISP-DM dirancang sebagai strategi pemecahan masalah yang dapat diterapkan di berbagai industri. Tujuan dari CRISP-DM adalah untuk menyediakan panduan yang jelas dan terstruktur dalam mengelola proyek *data mining* agar dapat mencapai hasil yang optimal (Wirth & Hipp, 2000).

CRISP-DM terdiri dari enam tahapan sebagaimana ditunjukkan pada gambar 2.1 (Astuti & dkk, 2014) :



Gambar 1. Model CRISP-DM

### 2.6. *Knowledge Data Discovery In Database*

*Knowledge Discovery in Database* (KDD) pendekatan ini merupakan metode untuk memperoleh pengetahuan yang belum digali dalam *database* atau basis data. Dalam basis data terdapat tabel tabel yang saling berelasi. Adapun tahapan KDD sebagaimana disajikan pada gambar 2 berikut (Mulyani & dkk, 2022).



Gambar 2 Tahapan Pemodelan KDD

### 2.7. *Data Terpadu Kesejahteraan Sosial*

Data Terpadu Kesejahteraan Sosial (DTKS) adalah data induk yang berisi data pemerlu pelayanan kesejahteraan sosial termasuk di dalamnya penerima Bantuan Sosial BPNT dan PKH. DTKS dihimpun dalam Sistem Informasi Kesejahteraan Sosial – *Next Generation* (SIKS-NG) yang merupakan suatu program aplikasi yang memberikan informasi terkait dalam penerima bantuan sosial serta pemerlu kesejahteraan sosial lainnya sesuai dengan data kependudukan yang terintegrasi secara langsung

dengan data induk kependudukan (Sulistyarini, 2021).

## 2.8. Data Pensasaran Percepatan Kemiskinan Ekstrem

Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) adalah kumpulan informasi dan data keluarga serta individu anggota keluarga hasil permuktakhiran basis data keluarga Indonesia dan keluarga berencana yang tersimpan dalam file elektronik dan sudah divalidasi NIK oleh DUKCAPIL serta memiliki status kesejahteraan (Desil) (MENKO PMK, 2022).

## 2.9. Bantuan Pangan Non Tunai

Bantuan Pangan Non Tunai (BPNT) adalah program transformasi dari program Beras Sejahtera (RASTRA) yang dimulai pada tahun 2017. Transformasi ini bertujuan untuk memenuhi prinsip 6T, yaitu Tepat Sasaran, Tepat Harga, Tepat Kualitas, Tepat Waktu, Tepat Jumlah, dan Tepat Administrasi. Program BPNT memberikan bantuan berupa bahan makanan termasuk beras dan bahan makanan pokok lainnya, yang memiliki kualitas dan jumlah yang sesuai dengan kebutuhan keluarga.

## 2.5. Program Keluarga Harapan

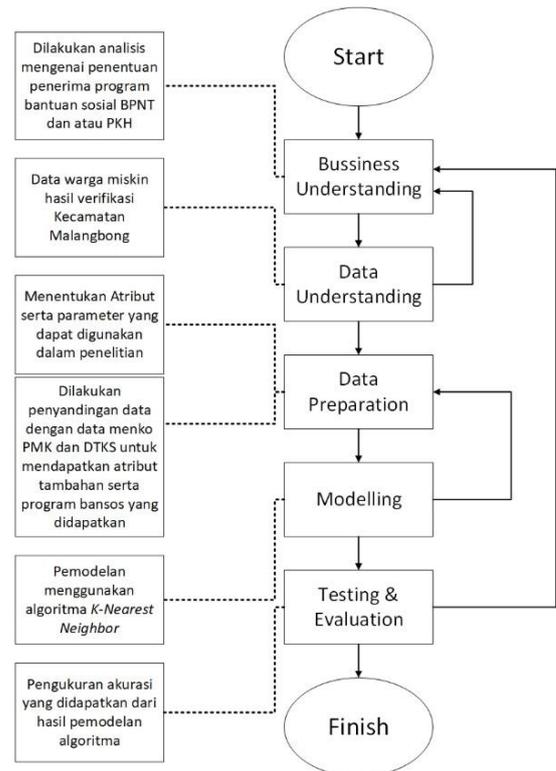
Program Bantuan Sosial Keluarga Harapan (PKH) adalah program yang diinisiasi oleh Kementerian Sosial untuk membantu masyarakat miskin dengan syarat tertentu. Melalui program PKH, pemerintah berhadapan memberikan akses kepada warga miskin dalam kesehatan, pendidikan, perawatan, pendampingan, serta perlindungan sosial lainnya kepada keluarga penerima PKH. Tujuan utama adalah menjadikan keluarga penerima PKH mandiri secara ekonomi, hal ini telah diatur dalam peraturan Menteri Sosial Nomor 1 Tahun 2018.

## 3. DESAIN PENELITIAN

Desain penelitian dengan menggunakan model CRISP-DM (*Cross Industry Standard Process-Data Mining*) memiliki langkah-langkah sebagai berikut:

### 3.1 Bussines Understanding

Bantuan sosial BPNT dan PKH adalah salah satu program bantuan sosial yang dikeluarkan oleh kementerian sosial bagi warga miskin, fakir miskin. Jumlah penduduk miskin di Kecamatan Malangbong Kabupaten Garut menurut Menko PMK dalam rentang desil 1 sampai desil 4 mencapai 69.150 jiwa dan 16.490 keluarga. Pada penelitian ini akan dilakukan analisis mengenai penentuan warga yang berhak menerima bantuan sosial BPNT dan atau PKH.



Gambar 3 Desain Penelitian

### 3.2 Data Understanding

Data yang didapat pada penelitian ini diperoleh dari hasil verifikasi Pihak Kecamatan Malangbong. Data tersebut Berisi atribut sebagai berikut: ID Individu, ID Keluarga, Nama, Alamat (Kampung, RT, RW, Desa), Status Hubungan Keluarga, Tempat Lahir, Tanggal Lahir, Jenis Kelamin, Pekerjaan, Nama Ibu Kandung, dan pendapatan per bulan.

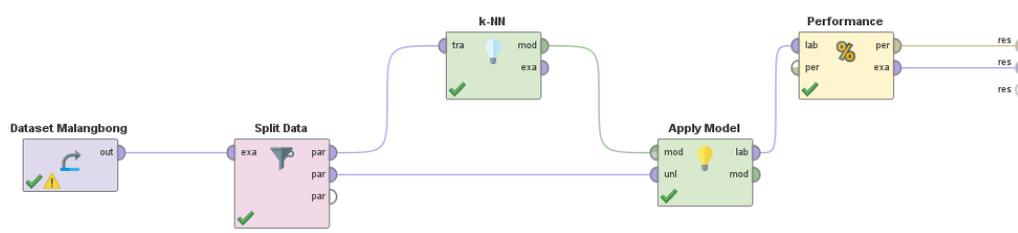
### 3.3 Data Preparation

Fase pengolahan data disiapkan data yang dibutuhkan yang akan digunakan pada fase Modeling atau pemodelan. Fase ini menggunakan pendekatan KDD.

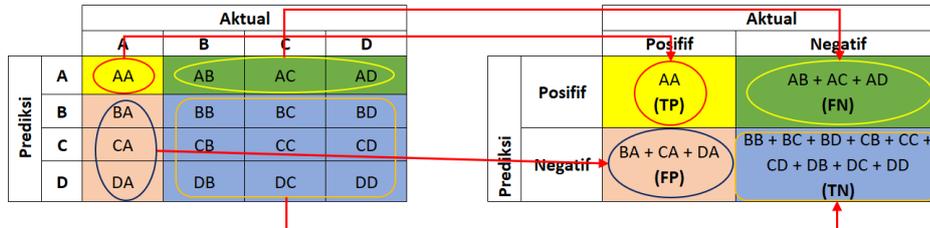
Pada fase ini perlu dilakukan secara teliti karena akan mempengaruhi hasil daripada pemodelan. Fase pengolahan data meliputi seluruh aktivitas yang dilakukan untuk membuat dataset yang akan digunakan pada fase pemodelan.

#### 1. Pembersihan Data

Pada tahap ini, dilakukan proses pembersihan data yang melibatkan data mentah sebanyak 102.045 jiwa. Langkah selanjutnya adalah mencocokkan data ini dengan data P3KE untuk menemukan ID keluarga P3KE dengan cara menyandingkan NIK pada data mentah dan data P3KE. Hasilnya, sebanyak 57.567 jiwa memiliki ID Keluarga P3KE, sementara 44.478 jiwa tidak memiliki ID keluarga P3KE. Data yang telah ditemukan memiliki ID Keluarga P3KE selanjutnya disandingkan kembali dengan data P3KE untuk mengidentifikasi hubungan antar anggota keluarga.



Gambar 3.2 Model Pengolahan Data Aplikasi *Rapidminer Studio*



Gambar 5 *Extend Confusion Matrix 4x4 ke 2x2*

Terakhir pemilahan data, proses ini memilah data dan menyisahkan satu anggota keluarga yang diambil untuk mewakili keluarga tersebut karena program BANSOS BPNT dan PKH ditujukan bagi keluarga secara keseluruhan, bukan individu tunggal. Data yang digunakan setelah proses pemilahan didapatkan sebanyak 19.943 jiwa, dan data lainnya tidak digunakan.

2. Penbandingan Data

Data hasil pemilahan dengan jumlah 19.943 jiwa selanjutnya dilakukan penbandingan dengan data P3KE untuk mendapatkan atribut informasi data keluarga. Data penbandingan ini dilakukan dengan cara mencari ID P3KE hasil proses pembersihan data dan dilakukan penbandingan serta pengambilan atribut pada database P3KE.

Data hasil pemilahan pada proses pembersihan data selanjutnya juga dilakukan penbandingan dengan Data Terpadu Kesejahteraan Sosial menggunakan aplikasi SIKS-NG. Pada tahap ini dilakukan penbandingan menggunakan NIK setiap individu yang telah melewati proses pemilahan data, yang mana data tersebut selanjutnya di unggah pada aplikasi SIKS-NG untuk dilakukan pemadanan dan mendapatkan informasi BANSOS yang didapatkan.

3. Normalisasi Atribut

Normalisasi atribut ini melibatkan proses seperti mengubah, menggabungkan atau menyusun ulang data sehingga dapat meningkatkan kualitas data. Normalisasi data juga dapat membantu menghindari dominasi atribut dengan rentang nilai yang besar dan memastikan bahwa setiap atribut dapat memberikan kontribusi yang seimbang terhadap analisis yang akan dilakukan.

4. Seleksi Atribut

Tujuan dari seleksi atribut ini untuk mengurangi dimensi data, mengurangi penggunaan memori pada perangkat pengolahan data, serta mempercepat proses pemodelan algoritma.

3.4 Modeling

Algoritma yang digunakan dalam penelitian ini menggunakan *K-Nearest Neighbor* dengan menggunakan perangkat lunak *Rapidminer Studio* Versi 9.10 serta menggunakan akun *Educational* yang mana akun ini diberikan kelebihan antara lain jumlah data yang dapat diproses tidak terbatas serta penggunaan *Logical Processor* yang lebih maksimal untuk mempercepat proses pemodelan.

Pembuatan model algoritma dari K-NN menggunakan aplikasi *Rapidminer* sebagaimana tersaji pada gambar 4.

Pada fungsi *split data* akan dipisahkan menjadi 2 tipe data yaitu data latih untuk digunakan untuk membuat model pada algoritma K-NN dan data uji untuk digunakan sebagai pengujian performa data latih. Adapun rasio pembagian data tersebut dibagi menjadi 2 yaitu 70% data latih, 30% data uji dan 80% data latih dan 20% data uji.

Setelah menentukan nilai rasio *split data*, selanjutnya menentukan nilai K pada algoritma K-NN, pada pengujian ini akan menggunakan variasi nilai K sebesar 3% dari populasi data, 4% dari populasi data dan 5% dari populasi data, penentuan nilai K ini tidak bersifat baku, namun harus memperhatikan apabila nilai K terlalu kecil kemungkinan model akan terlalu sensitif terhadap perubahan pola data, dan apabila nilai K terlalu besar model cenderung memberikan hasil prediksi yang lebih dominan terhadap pola dalam data.

3.5 *Testing & Evaluation*

Pada tahap ini disajikan *confusion matrix*, pada situasi *Multi-Class Classification* dimana terdapat lebih dari 2 kelas yang diprediksi dilakukan *Extend Confusion Matrix*. Pendekatan ini bertujuan untuk memberikan informasi yang lebih rinci terhadap performa setiap kelas yang akan diprediksi. Pada penelitian ini menggunakan *Extend Confusion Matrix 4x4* menjadi *2x2*.

Tabel 1 Perbandingan matriks perhitungan *split ratio* 70% data latih 30% data uji

K = 598 (3% dari populasi data)						
Metriks	PKH	BPNT	NON BANSOS	Irisan BANSOS	Rata-rata	F1-Score
Akurasi	85,33%	86,81%	80,83%	93,85%	86,70%	33,21%
Presisi	80,21%	68,08%	76,99%	64,27%	72,39%	
Recall	25,25%	25,63	27,05%	8,26%	21,55%	
K = 798 (4% dari populasi data)						
Metriks	PKH	BPNT	NON BANSOS	Irisan BANSOS	Rata-rata	F1-Score
Akurasi	84,77%	86,65%	80,38%	93,35%	86,29%	32,90%
Presisi	80,40%	68,29%	75,63%	59,56%	70,97%	
Recall	25,47%	25,75%	26,72%	7,70%	21,41%	
K = 997 (5% dari populasi data)						
Metriks	PKH	BPNT	NON BANSOS	Irisan BANSOS	Rata-rata	F1-Score
Akurasi	83,39%	86,36%	79,93%	92,43%	85,53%	32,32%
Presisi	79,78%	67,88%	75,52%	49,72%	68,22%	
Recall	25,70%	25,68%	26,83%	6,49%	21,18%	

Tabel 2 Perbandingan matriks perhitungan *split ratio* 80% data latih 20% data uji

K = 598 (3% dari populasi data)						
Metriks	PKH	BPNT	NON BANSOS	Irisan BANSOS	Rata-rata	F1-Score
Akurasi	85,06%	86,76%	80,39%	93,78%	86,50%	33,09%
Presisi	80,11%	68,23%	75,64%	63,83%	71,95%	
Recall	25,29%	25,69%	26,73%	8,21%	21,48%	
K = 798 (4% dari populasi data)						
Metriks	PKH	BPNT	NON BANSOS	Irisan BANSOS	Rata-rata	F1-Score
Akurasi	85,23%	86,48%	80,14%	93,63%	86,37%	32,97%
Presisi	81,05%	68,46%	74,85%	60,91%	71,32%	
Recall	25,54%	25,86%	26,53%	7,85%	21,44%	
K = 997 (5% dari populasi data)						
Metriks	PKH	BPNT	NON BANSOS	Irisan BANSOS	Rata-rata	F1-Score
Akurasi	84,35%	86,43%	79,69%	93,46%	85,98%	32,71%
Presisi	80,11%	68,07%	75,38%	56,34%	69,98%	
Recall	25,51%	25,73%	26,87%	7,27%	21,35%	

Setelah mendapatkan confusion matrix 2x2 setiap program BANSOS, selanjutnya menghitung akurasi, *Recall*, presisi dan *F1-Score*.

Setelah mendapatkan model algoritma dengan nilai akurasi, *recall*, presisi dan *F1-Score* yang tinggi, dilakukan validasi menggunakan model tersebut, adapun data validasi yang digunakan adalah data di luar *dataset* sejumlah 600 jiwa dengan wilayah geografis pegunungan, pesisir dan perkotaan. Data validasi terdiri dari 50 perwakilan setiap label, dengan masing-masing 200 data setiap wilayah geografis. Evaluasi ini akan memberikan gambaran terhadap kemampuan model dalam mengatasi variasi dalam data.

#### 4. HASIL DAN PEMBAHASAN

Berdasarkan hasil pemodelan yang telah dilakukan sebanyak 6 kali pemodelan yaitu rasio 70% data latih, 30% data uji dengan nilai K sebesar 3%, 4%, 5% dan rasio 80% data latih, 20% data uji dengan nilai K sebesar 3%, 4%, dan 5% dilakukan perhitungan rata-rata akurasi, presisi dan *recall* pada setiap hasil perhitungan, selanjutnya setiap nilai rata-rata nilai presisi dan *recall* akan dihitung nilai *F1-Score* untuk mencari keseimbangan antara nilai presisi dengan *recall*. Adapun hasil rata-rata matriks perhitungan disajikan pada Tabel 1 dan Tabel 2. Hasil perhitungan *F1-Score*, didapatkan nilai *F1-Score* tertinggi pada rasio 7:3 dengan nilai K = 598 sebesar 33,21%, adapun nilai akurasi yang didapat pada rasio dan nilai K tersebut sebesar 86,70% hal ini

menunjukkan bahwa model tersebut menghasilkan akurasi yang cukup tinggi. Selain mengukur tingkat akurasi, didapatkan juga nilai presisi sebesar 72,39%, nilai presisi ini menggambarkan sejauh mana hasil positif yang diprediksi oleh model adalah benar. Terakhir ada nilai *Recall* sebesar 21,55%, *Recall* ini mengukur sejauh mana algoritma dapat mengidentifikasi atau menemukan hasil positif, model dengan nilai K=598 akan digunakan sebagai referensi pada tahap validasi.

Tahap validasi data menggunakan 600 data di luar *dataset* sebelumnya dengan menggunakan nilai K=3% dari populasi data atau K=598 setelah dilakukan pembulatan. Hasil perhitungan akan dihitung rata-rata nilai akurasi, *recall* dan presisi serta menghitung nilai *F1-Score* untuk mencari keseimbangan antara nilai presisi dan *recall*. Adapun nilai rata-rata tersebut disajikan pada tabel 3 sebagai berikut

Nilai rata-rata akurasi yang didapatkan dari pengolahan validasi data dengan nilai K=598 mendapatkan nilai akurasi sebesar 85,92% yang menunjukkan bahwa model yang digunakan mampu memprediksi kelas-kelas dengan tingkat keseluruhan yang tinggi. Tingkat akurasi ini dapat memberikan keyakinan terhadap kemampuan model dalam memprediksi atau memberikan rekomendasi program BANSOS yang sesuai bagi masyarakat.

Tabel 3 Perhitungan nilai rata-rata metrik hasil validasi data

Metriks	PKH	BPNT	NON BANSOS	Irisan BANSOS	Rata-rata	F1-Score
Akurasi	89,17%	88,50%	83,00%	83,00%	85,92%	
Presisi	64,00%	59,33%	81,33%	82,67%	71,83%	32,53%
Recall	17,94%	16,76%	24,50%	24,90%	21,03%	

*Recall* mengukur sejauh mana algoritma dapat mengidentifikasi atau menentukan nilai positif. Nilai *recall* paling rendah ditemukan pada program BANSOS BPNT sebesar 16,76% dan BANSOS PKH sebesar 17,94%, hal ini menunjukkan bahwa model mengalami kesulitan dalam mengidentifikasi pola dari kelas-kelas yang diprediksi.

*F1-Score* digunakan untuk mencari keseimbangan antara nilai presisi dan *recall*, nilai *F1-Score* yang didapatkan sebesar 32,53%, hal ini menandakan adanya ketidakseimbangan antara presisi dengan *recall* terhadap data yang digunakan, sehingga algoritma mengalami kesulitan dalam mengevaluasi nilai positif yang didapatkan.

Hasil dari penelitian dimulai dari pemodelan sampai dengan tahap validasi memiliki perbedaan performa yang signifikan, hal ini disebabkan karena perbedaan jumlah data yang dominan pada satu kelas, contohnya pada atribut program bansos hasil penyandingan dengan DTKS yang mana nilai BANSOS BPNT lebih dominan daripada program BANSOS lainnya.

Setiap atribut memiliki relasi yang sangat mempengaruhi nilai matriks relasi setiap record dengan record lainnya berbeda-beda serta parameter setiap atribut yang variatif dapat mempengaruhi nilai matriks secara signifikan. Misalnya atribut *Range* Pendapatan dihilangkan menghasilkan matriks rata-rata nilai akurasi sebesar 73,58%, presisi sebesar 47,17% dan *Recall* sebesar 16,52%, hal ini dapat menunjukkan bahwa atribut *Range* pendapatan memiliki peran penting dalam proses prediksi.

Berdasarkan hasil pemodelan serta pengujian validasi yang telah dilakukan menggunakan data dengan geografis yang berbeda, algoritma K-NN dengan nilai  $K=3$  dari keseluruhan *dataset* dapat dijadikan suatu pertimbangan serta kebijakan dalam pengusulan program BANSOS bagi masyarakat. Meskipun nilai akurasi yang didapatkan cukup besar, namun bercermin dari nilai *F1-Score* yang memiliki nilai rendah bahwasanya perlu adanya evaluasi terhadap *dataset* yang digunakan untuk mengatasi ketidakseimbangan antara nilai presisi dan *Recall* sehingga program BANSOS yang diusulkan dapat lebih efektif dan adil sesuai dengan kondisi keluarga penerima BANSOS tersebut.

Terdapat beberapa metode yang dapat digunakan untuk mengatasi ketidakseimbangan antara nilai presisi dengan *recall*, antara lain mengurangi jumlah data yang memiliki relasi *record* yang sama pada *dataset* yang digunakan sehingga populasi setiap relasi antar atribut seimbang, cara ini dapat dilakukan pada saat normalisasi data, selanjutnya dengan metode *Ensemble Learning*, pendekatan ini

menggabungkan beberapa program menjadi satu kelas atribut dan melakukan pemodelan dengan *confusion matrix 2x2*, contoh melakukan pemodelan BANSOS PKH dengan BANSOS NON PKH yang mencakup penjumlahan data BPNT, Irisan dan NON BANSOS.

Hasil penelitian ini diharapkan dapat menjadi pertimbangan untuk pemerintah Kabupaten Garut maupun pemerintah pusat dalam menentukan program BANSOS.

## 5. KESIMPULAN DAN SARAN

### 5.1. Kesimpulan

1. Data warga hasil verifikasi yang telah dilakukan oleh Kecamatan Malangbong, TKSK dan PUSKESOS Kecamatan Malangbong Kabupaten Garut dengan jumlah total data bersih yang digunakan dalam penelitian ini sebanyak 19.943 Keluarga.
2. Atribut yang digunakan dalam penelitian ini berjumlah 16 meliputi ID, Pekerjaan Kepala Keluarga, Jumlah Anggota Keluarga, Pendidikan Terakhir Kepala Keluarga, Kepemilikan Rumah, Memiliki Simpanan Uang/ Perhiasan/ Ternak/ Lainnya, Jenis Atap, Jenis Dinding, Jenis Lantai, Sumber Penerangan, Bahan Bakar Memasak, Sumber Air Minum, Memiliki fasilitas Buang Air Besar, Ada Komponen Dalam Keluarga, *Range* Pendapatan, dan BANSOS.
3. Algoritma yang digunakan adalah *K-Nearest Neighbor* (K-NN) dengan menggunakan aplikasi *Rapidminer*. Penelitian ini dilakukan dengan 6 percobaan dengan metode *split data 70:30* dan 80:20, serta nilai K sebesar 3%, 4%, dan 5% dari jumlah *dataset*.
4. Hasil pengujian menunjukkan performa terbaik pada nilai K sebesar 3% dari total data atau  $K=598$ , dengan rasio data latih dan data uji sebesar 70% dan data uji sebesar 30% yang menghasilkan nilai rata-rata akurasi pada setiap program BANSOS sebesar 86,70%.
5. Dari hasil pengujian sebelumnya dilakukan validasi terhadap model yang telah dibuat dengan menggunakan data di luar wilayah *dataset* yang digunakan. Kategori data validasi didasarkan pada kondisi geografis wilayah yang melibatkan pegunungan, perkotaan, serta pesisir. Jumlah data yang divalidasi sebanyak 600 keluarga dengan pembagian 50 *dataset* setiap atribut pada setiap wilayah.
6. Hasil validasi menunjukkan rata-rata nilai akurasi setelah dilakukan evaluasi menggunakan

*confusin matrix* pada setiap program BANSOS sebesar 85,92%.

## 5.2. Saran

1. Melakukan uji coba menggunakan algoritma klasifikasi yang berbeda seperti *Naïve Bayes*, *Decision Tree*, *Random Forest*, atau *Gradient Boosting*.
2. Melakukan evaluasi terhadap *dataset* yang digunakan karena data terlalu bervariasi. Tujuan dari evaluasi ini untuk memberikan tingkat keseimbangan antara nilai akurasi, presisi serta nilai *Recall*.
3. Mengurangi atribut *oversampling* untuk menyeimbangkan dengan atribut *undersampling*.
4. Mengurangi atribut yang memiliki pola sama dengan kelas yang *oversampling*.
5. Melakukan pengujian dengan cara *ensemble learning* pada kelas yang hendak diuji.
6. Menggunakan aplikasi atau *software* lain seperti *Weka*, atau *R Studio* untuk menggali potensi serta informasi lebih rinci terhadap data yang diolah.

## DAFTAR PUSTAKA

- ASTUTI, K., & dkk. 2014. Data Mining Untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree C.45. *Program Studi Teknik Informatika-Fakultas Ilmu Komputer-Universitas Dian Nuswantoro*.
- FIKRILLAH, H. N., & dkk. 2023. Klasifikasi Penerima Bansos Menggunakan Algoritma Naive Bayes. *Jurnal Teknik Informatika dan Sistem Informasi*, 683-695.
- FIRASARI, E., & dkk. 2020. Comparison of K-Nearest Neighbor (K-NN) and Naive Bayes for the Classification of The Poor in Recipients of Social Assistance. *Journal of Physics : Conference Series*, 1742-6596.
- KAMARI, R., & dkk. 2022. Verification of Covid-19 Social Assistance Recipients Using Naive Bayes Classifier. *International Journal Of Emerging Multidisciplinaries : Computer Science & Artificial Intelligence*, 1-12.
- MENKO PMK. (2022, September 01). *P3KE MENKO PMK*. Dipetik Agustus 25, 2023, dari P3KE Menko PMK: <https://p3ke.kemenkopmk.go.id>
- MUHAMMAD, F., & dkk. 2018. Analisis Perbandingan CPU dan GPU (CUDA) Pada Klasifikasi Data Mining dengan Menggunakan Metode K-Nearest Neighbor Kernel Algorithm. *e-Proceeding of Engineering*, 7494.
- MULYANI, A., & dkk. 2022. The Prediction Of PPA and KIP-Kuliah Scholarship Reipient Using Naive Bayes Algorithm. *Jurnal teknik Informatika (JUTIF)*, 821-827.
- MUSTAFA, M. S., & dkk. 2017. Implementasi Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Citec Journal*.
- NURJANAH, A., & RIFAI, A. 2023. Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Kelayakan Status Penduduk Miskin Di Desa Susukan Tonggoh. *Jurnal Wahana Informatika*, 164-176.
- PURWANTO, A., & DARMADI, E. A. 2018. Perbandingan Minat Siswa SMU Pada Metode Klasifikasi Menggunakan 5 Algoritma. *Jurnal Komputer Dan Informatika*, 43-47.
- RAHMANSYAH, W., & et al. 2020. Pemetaan Permasalahan Penyaluran Bantuan Sosial Untuk Penganan Covid-19 Di Indonesia. *Jurnal Pajak dan Keuangan Negara*, 90-102.
- SAPUTRA, A. Y., & PRIMADASA, Y. 2018. Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Techno COM*, 395-403.
- SULISTYARINI, T. 2021. Implementasi Kebijakan Peraturan Bupati Ponogoro Nomor 60 Tahun 2020 Tentang Percepatan Penanggulangan Kemiskinan Berbasis Data Terpadu Kesejahteraan Sosial (DTKS) di Kecamatan Ngebel Kabupaten Ponogoro Provinsi Jawa Timur. *Jurnal Ilmu Administrasi Publik*, 139-155.
- SUPRIATNA, I. W., & ASTUTI, L. G. 2019. Implementasi K-Nearest Neighbor Pada Penentuan Keluarga Miskin Bagi Dinas Sosial Kabupaten Tabanan. *Jurnal Teknologi Informasi Dan Komputer*.
- SUYONO, H., & dkk. 2013. Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Jurnal EECCIS*.
- WIRTH, R., & HIPPEL, J. 2000. CRISP-DM: Towards a Standard Process Model for Data Mining. *Proceeding of the 4th international conference on the practical application of knowledge discovery and data mining* (hal. 29-39). Germany: Univerity of Tubingen.