PERBANDINGAN TINGKAT AKURASI ALGORITMA DECISION TREE DAN RANDOM FOREST DALAM MENGKLASIFIKASI PENERIMA BANTUAN SOSIAL BPNT DI DESA SLANGIT

Aldiyansyah [1], Ade Irma Purnamasari [2], Irfan Ali [3]
[1,2] Teknik Informatika, STMIK IKMI CIREBON
[3] Rekayasa Perangkat Lunak, STMIK IKMI CIREBON
Jl. Perjuangan No. 10B Majasem, Kec. Kesambi, Kota Cirebon
smcaldy@gmail.com

ABSTRAK

Program Bantuan Sosial BPNT merupakan inisiatif pemerintah Indonesia yang memberikan bantuan pangan dalam bentuk nontunai kepada Keluarga Penerima Manfaat (KPM) melalui sistem perbankan. Kurang tepatnya penerima bantuan sosial tersebut menjadi permasalahan, serta kelayakan penerima KPM yang masih tidak tepat sasaran. Oleh karena itu, pada penelitian ini metode klasifikasi menggunakan algoritma decision tree dan random forest menjadi solusi yang tepat dalam untuk meningkatkan perbandingan tingkat akurasi, pada penerima bantuan sosial di Desa Slangit, Kecamatan Klangenan, Kabupaten Cirebon. Algoritma Decision tree dan random forest akan memberikan keputusan dengan model pohon keputusan dan menghasilkan tingkat akurasi yang baik. Data diperoleh dari Pemerintah Desa, Dinas Sosial Kabupaten Cirebon, dan Data Terpadu Kesejahteraan Sosial (DTKS) selama Juli dan Agustus 2023. Metode penelitian mencakup studi literatur, pengumpulan data, pra-pemrosesan data, implementasi model klasifikasi, *Knowledge Discovery in Database (KDD),* evaluasi model, dan kesimpulan. Evaluasi menunjukkan tingkat akurasi tinggi, yakni 97.86% untuk Decision Tree dan 97.86% untuk Random Forest. Rekomendasi penelitian ini mencakup pertimbangan terhadap metode alternatif serta perlunya penelitian lanjutan untuk memahami faktor-faktor sosial, ekonomi, dan demografis yang memengaruhi status kelayakan penerima BPNT. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan panduan bagi pengembangan program bantuan sosial di masa depan.

Kata kunci: *Klasifikasi, BPNT, Decision Tree, dan Random Forest.*

1. PENDAHULUAN

Kondisi Ekonomi yang terbatas merupakan tantangan dalam ranah sosial-ekonomi di Indonesia. Pemerintah telah mengimplementasikan program untuk mengurangi kemiskinan dengan memperhatikan standar rumah tangga yang kurang mampu dan standar tersebut dirumuskan berdasarkan informasi yang dikumpulkan[1]. Kemiskinan dapat diartikan sebagai keterbatasan dalam memenuhi berbagai kebutuhan, kendala masyarakat oleh norma-norma adat atau budaya, atau ketidakmampuan masyarakat menghadapi sistem sosial yang tidak adil. Penanganan kemiskinan membutuhkan penerapan kebijakan yang komprehensif melibatkan semua pihak terkait.[2]. Skema bantuan pangan non tunai adalah suatu program yang memberikan bantuan pangan kepada keluarga penerima manfaat atau KPM. Proses distribusi bahan pangan dilakukan dengan menggunakan Kartu Keluarga Sejahtera (KKS) secara elektronik, dengan tujuan meningkatkan taraf kesejahteraan masyarakat.[3]. Pengaliran Bantuan Pangan Non Tunai (BPNT) harus dilakukan dengan cermat, sesuai jadwal, dan transparan agar dapat memverifikasi bahwa penerima bantuan adalah mereka yang memang memerlukannya. [4]. Setiap bulan, program bantuan sembako diterapkan untuk memberikan dukungan kepada keluarga penerima manfaat, dan pendistribusiannya mencakup seluruh populasi di Indonesia sesuai dengan kriteria yang telah ditetapkan oleh pemerintah melalui Dinas Sosial.

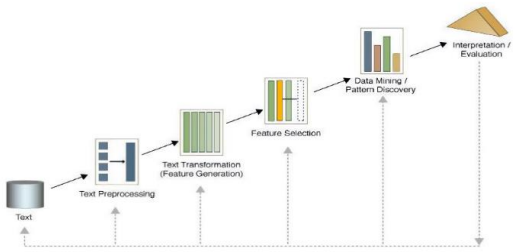
Penggunaan metode penambangan data, atau data mining, terbukti bermanfaat untuk mendapatkan informasi yang berharga dari berbagai data. Pendekatan ini melibatkan penerapan pengetahuan seperti statistika, matematika, dan pengenalan pola[5]. Klasifikasi merupakan proses pengelompokkan item ke dalam kategori yang berbeda. Fokus pada usaha mengklasifikasikan kelayakan telah menjadi perhatian dalam penelitian sebelumnya. Penelitian ini dilaksanakan untuk mengkategorikan penerima bantuan pangan non tunai, sehingga hasilnya dapat digunakan sebagai landasan dalam pengambilan keputusan mengenai apakah mereka memenuhi syarat atau tidak untuk menerima bantuan pangan non tunai[6]. Tujuan utama dari penelitian ini adalah untuk mengklasifikasikan atau mengelompokkan rumah tangga penerima Raskin dengan menggunakan pendekatan data mining. Selain itu, penelitian ini juga memasukkan perbandingan antara dua metode, yaitu Random Forest dan SVM *(Support Vector Machine)* [7]. Penelitian ini mengadakan perbandingan tingkat akurasi antara beberapa metode klasifikasi, seperti naïve bayes, decision tree, dan random forest. Hasil penelitian menunjukkan bahwa tingkat akurasi naïve bayes mencapai 90%, decision tree mencapai 83%, dan random forest mencapai 87%. Dengan

melakukan analisis sentimen terhadap klasifikasi komentar YouTube, diharapkan informasi ini dapat memberikan pencerahan kepada pemerintah tentang bagaimana masyarakat merespons dan menilai kebijakan yang diimplementasikan[8]. Berdasarkan evaluasi kinerja Random Forest dan Decision Tree C4.5 dalam meramal tingkat keberhasilan, diperoleh kesimpulan bahwa akurasi pengujian menggunakan Random Forest mencapai 87.20%, sementara Decision Tree C4.5 mencapai tingkat akurasi sebesar 85.90%[9].

2. TINJAUAN PUSTAKA

2.1. *Knowledge Discovery In Database (KDD)*

Knowledge Discovery in Database (KDD) adalah proses menganalisis terstruktur untuk memperoleh informasi yang benar, baru, dan menemukan pola dari data yang besar dan kompleks. Data mining menjadi inti dari proses Knowledge Discovery in Database (KDD) yaitu dengan menggunakan algoritma tertentu untuk mengeksplorasi data, membangun model, dan menemukan pola yang belum diketahui . Proses text mining dengan Knowledge Discovery in Database (KDD) sama dengan pada data mining [10].



Gambar 1 Knowledge Discovery in Database (KDD)

Berikut tahapan penelitian yang dilakukan menggunakan proses KDD (knowledge discovery in database) :

1. Text
   Pengumpulan data untuk penelitian ini memungkinkan untuk diperoleh dari berbagai sumber, termasuk data yang diberikan oleh pendamping kecamatan terkait BPNT, informasi demografis dari pemerintah desa, dan data dari puskesos desa Slangit.
2. Prepocessing
   Langkah ini dimaksudkan untuk meminimalkan dampak atribut yang memiliki sedikit relevansi dalam proses klasifikasi. Secara umum, pemrosesan teks melibatkan serangkaian tindakan, termasuk case folding, pembersihan data, konversi emotikon, tokenisasi, eliminasi kata penghenti, dan stemming.
3. Transformasi
   Dokumen diwujudkan oleh fitur-fitur (kata-kata) yang terdapat di dalamnya. Proses transformasi dokumen diperlukan agar dapat mengonversinya dari bentuk teks penuh menjadi bentuk vektor dokumen
4. Fitur Seleksi
   Tahap seleksi fitur berfungsi untuk mengurangi dimensi dari gangguan noise yang dapat memengaruhi hasil penelitian. Pentingnya pemilihan fitur muncul saat pembuatan model karena dataset seringkali mengandung berbagai fitur yang berlebihan dan tidak relevan.
5. Data mining
   Langkah-langkah yang diterapkan dalam text mining serupa dengan proses pada data mining, seperti klasifikasi, asosiasi, dan sebagainya. Dalam konteks penelitian ini, kami menggunakan metode klasifikasi dengan menerapkan algoritma decision tree dan random forest.
6. Evaluasi
   Evaluasi umumnya dilakukan terhadap hasil temuan pola, seringkali memanfaatkan suatu metrik performa sebagai referensi.

2.2. Decision Tree

Decision Tree adalah struktur yang tumbuh dari proses sekuensial, dimulai dari akar dan terus melakukan evaluasi fitur pada setiap tahap. Dalam setiap langkah, Decision Tree memilih satu dari dua cabang yang tersedia dan meneruskan evaluasi hingga mencapai cabang terakhir, yang dikenal sebagai daun. Pada umumnya, daun tersebut mencerminkan pencapaian target yang diinginkan selama proses tersebut[11]. Keunggulan decision tree terletak pada ketidakterpengaruhannya oleh nilai pada setiap fitur data, sehingga decision tree dapat bekerja efisien pada kumpulan data yang tidak melalui normalisasi pra-proses dengan *Scikit-learn*. *Gini Impurity* didefinisikan dengan rumus berikut

$$\int I\,\mathrm{gini}(x) = ?x \quad p(x \mid y)(1 ? p(x \mid y))$$

Scikit-learn memungkinkan pelatihan *binary* decision tree dengan ukuran *Gini* dan *cross-entropy* impurity. *Gini Impurity* diukur dengan rumus tertentu, di mana total selalu diperpanjang ke semua kelas yang ada. *Gini Impurity* sering dijadikan ukuran dan nilai bawaan di dalam *Scikit-Learn*. Dalam konteks sampel, *Gini Impurity* mengukur kemungkinan kesalahan klasifikasi suatu label jika label tersebut dipilih secara acak dengan probabilitas distribusi cabang.

$$I\,\mathrm{cross\text{-}entropy}(x) \equiv ??x \quad p(x \mid y)\log p(x \mid y)$$

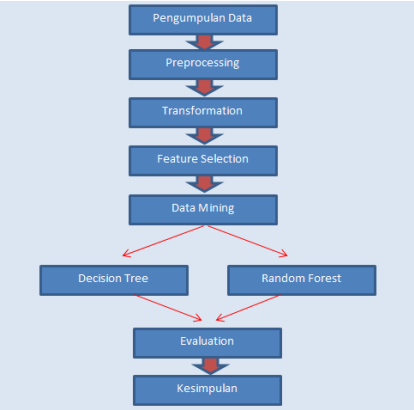
Ukuran ini diterapkan berdasarkan teori informasi dan mengasumsikan nilai minimum saat sampel dari suatu kelas hadir secara eksklusif, mencapai nilai maksimum ketika distribusinya merata di antara kelas-kelas. Meskipun mirip dengan Gini Impurity, perbedaannya terletak pada fakta bahwa *cross-entropy* memberikan keleluasaan kepada

pengguna untuk memilih pemisahan yang mengurangi ketidakpastian dalam klasifikasi, sedangkan Gini Impurity lebih berfokus pada upaya langsung untuk meminimalkan potensi kesalahan klasifikasi[11].

2.3. Random Forest

Menggunakan Random Forest Classifier untuk melatih model dalam konteks klasifikasi telah terbukti meningkatkan kemampuan model dalam mengenali permasalahan [12]. Analisis perbandingan ini akan memberikan pemahaman yang jelas tentang keunggulan dan batasan algoritma Decision Tree dibandingkan dengan Random Forest. Hasil penelitian diharapkan dapat memberikan panduan praktis bagi peneliti dan praktisi dalam memilih algoritma yang sesuai dengan kasus yang dihadapi.

3. METODE PENELITIAN



Gambar 2 Metode Penelitian

3.1. Pengumpulan Data

Data untuk penelitian ini dapat diperoleh dari berbagai sumber, termasuk data yang disediakan oleh pendamping kecamatan terkait BPNT, informasi demografis dari pemerintah desa.

3.2. Pre-prosessing

Langkah ini dimaksudkan untuk meminimalkan dampak atribut yang memiliki sedikit relevansi dalam proses klasifikasi. Secara umum, pemrosesan teks melibatkan serangkaian tindakan, termasuk case folding, pembersihan data, konversi emotikon, tokenisasi, eliminasi kata penghenti, dan stemming.

3.3. Transformasi

Dokumen diwujudkan oleh fitur-fitur (kata-kata) yang terdapat di dalamnya. Proses transformasi dokumen diperlukan agar dapat mengonversinya dari bentuk teks penuh menjadi bentuk vektor dokumen.

3.4. Fitur Seleksi

Tahap seleksi fitur berfungsi untuk mengurangi dimensi dari gangguan noise yang dapat memengaruhi hasil penelitian. Pentingnya pemilihan fitur muncul saat pembuatan model karena dataset seringkali mengandung berbagai fitur yang berlebihan dan tidak relevan.

3.5. Data Mining

Langkah-langkah yang diterapkan dalam text mining serupa dengan proses pada data mining, seperti klasifikasi, asosiasi, dan sebagainya. Dalam konteks penelitian ini, kami menggunakan metode klasifikasi dengan menerapkan algoritma decision tree dan random forest.

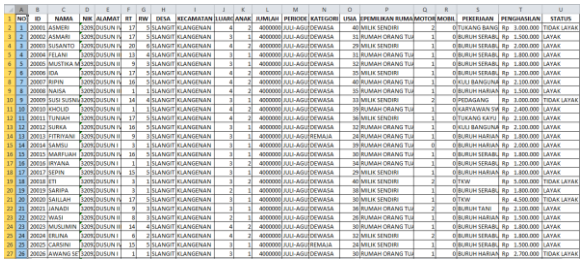
3.6. Evaluation

Evaluasi umumnya dilakukan terhadap hasil temuan pola, seringkali memanfaatkan suatu metrik performa sebagai referensi.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

Pada riset ini, terdapat satu set data mengenai penerima bantuan pangan non-tunai (BPNT) di Desa Slangit, terdiri dari 936 entri yang dikumpulkan pada bulan Juli dan Agustus 2023. Data tersebut mencakup berbagai indikator seperti Alamat, RT, RW, Keluarga, Anak, Kategori, Usia, Kepemilikan Rumah, Motor, Mobil, Pekerjaan, Penghasilan, dan Status Kelayakan. Rincian lebih lanjut dapat ditemukan dalam gambar berikut:



NO	ID	NAMA	NIK	ALAMAT	RT	RW	DESA	KECAMATAN	LUAR	ANAK	JUMLAH	PERIODE	KATEGORI	USIA	KEPEMILIKAN RUMAH	MOTOR	MOBIL	PEKERJAAN	PENGHASILAN	STATUS
1	20001	ASMERI	5205	DUSUN II	17	5	SLANGIT	KLANGENAN	4	2	4000000	JULI-AGU	DEWASA	40	MILIK SENDIRI		2	0 TUKANG BANG	Rp 1.000.000	TIDAK LAYAK
2	20002	ASMARI	5205	DUSUN II	17	5	SLANGIT	KLANGENAN	3	1	4000000	JULI-AGU	DEWASA	31	RUMAH ORANG TUA	1	0	BURUH SERABU	Rp 1.500.000	LAYAK
3	20003	SUSANTO	5205	DUSUN II	20	6	SLANGIT	KLANGENAN	4	2	4000000	JULI-AGU	DEWASA	29	MILIK SENDIRI	1	0	BURUH SERABU	Rp 2.000.000	LAYAK
4	20004	FELANI	5205	DUSUN II	13	4	SLANGIT	KLANGENAN	3	1	4000000	JULI-AGU	DEWASA	31	RUMAH ORANG TUA	1	0	BURUH SERABU	Rp 1.800.000	LAYAK
5	20005	MUSTIKA W	5205	DUSUN II	9	3	SLANGIT	KLANGENAN	3	1	4000000	JULI-AGU	DEWASA	32	RUMAH ORANG TUA	1	0	BURUH SERABU	Rp 1.800.000	LAYAK
6	20006	IDA	5205	DUSUN II	17	5	SLANGIT	KLANGENAN	4	2	4000000	JULI-AGU	DEWASA	35	MILIK SENDIRI	1	0	BURUH SERABU	Rp 1.200.000	LAYAK
7	20007	RIPIN	5205	DUSUN II	16	5	SLANGIT	KLANGENAN	4	2	4000000	JULI-AGU	DEWASA	40	RUMAH ORANG TUA	1	0	KLIJ BANGUNA	Rp 2.100.000	LAYAK
8	20008	NAISA	5205	DUSUN II	1	1	SLANGIT	KLANGENAN	4	2	4000000	JULI-AGU	DEWASA	33	RUMAH ORANG TUA	1	0	BURUH HARIAN	Rp 1.500.000	LAYAK
9	20009	SUSI SUSNA	5205	DUSUN I	14	4	SLANGIT	KLANGENAN	3	1	4000000	JULI-AGU	DEWASA	33	MILIK SENDIRI		2	0 PEDAGANG	Rp 3.000.000	TIDAK LAYAK
10	20010	KHOLID	5205	DUSUN II	1	1	SLANGIT	KLANGENAN	4	2	4000000	JULI-AGU	DEWASA	39	RUMAH ORANG TUA	1	0	KARYAWAN SW	Rp 2.400.000	LAYAK
11	20011	TUNAH	5205	DUSUN II	17	5	SLANGIT	KLANGENAN	4	2	4000000	JULI-AGU	DEWASA	36	MILIK SENDIRI	1	0	TUKANG KAYU	Rp 2.100.000	LAYAK
12	20012	SURIA	5205	DUSUN IV	16	5	SLANGIT	KLANGENAN	3	1	4000000	JULI-AGU	DEWASA	32	RUMAH ORANG TUA	1	0	KLIJ BANGUNA	Rp 2.100.000	LAYAK
13	20013	FITRIYANI	5205	DUSUN II	9	3	SLANGIT	KLANGENAN	3	1	4000000	JULI-AGU	REMAJA	24	RUMAH ORANG TUA	1	0	BURUH HARIAN	Rp 1.800.000	LAYAK
14	20014	SAMSU	5205	DUSUN I	3	1	SLANGIT	KLANGENAN	3	1	4000000	JULI-AGU	DEWASA	39	RUMAH ORANG TUA	0	0	BURUH HARIAN	Rp 2.000.000	LAYAK
15	20015	MARYUAH	5205	DUSUN IV	16	5	SLANGIT	KLANGENAN	3	1	4000000	JULI-AGU	DEWASA	30	RUMAH ORANG TUA	1	0	BURUH SERABU	Rp 1.800.000	LAYAK
16	20016	IRYANA	5205	DUSUN I	1	1	SLANGIT	KLANGENAN	3	2	4000000	JULI-AGU	DEWASA	34	RUMAH ORANG TUA	1	0	BURUH SERABU	Rp 1.200.000	LAYAK
17	20017	SEPIN	5205	DUSUN IV	15	5	SLANGIT	KLANGENAN	3	1	4000000	JULI-AGU	DEWASA	29	MILIK SENDIRI		1	0 BURUH HARIAN	Rp 1.800.000	LAYAK
18	20018	ETI	5205	DUSUN I	3	1	SLANGIT	KLANGENAN	3	2	4000000	JULI-AGU	DEWASA	40	MILIK SENDIRI		2	0 TKW	Rp 3.000.000	TIDAK LAYAK
19	20019	SARIPA	5205	DUSUN I	3	1	SLANGIT	KLANGENAN	2	1	4000000	JULI-AGU	DEWASA	38	MILIK SENDIRI		1	0 BURUH SERABU	Rp 1.800.000	LAYAK
20	20020	SAILAH	5205	DUSUN IV	17	5	SLANGIT	KLANGENAN	3	1	4000000	JULI-AGU	DEWASA	36	MILIK SENDIRI		1	0 TKW	Rp 4.500.000	TIDAK LAYAK
21	20021	JANADI	5205	DUSUN IV	9	3	SLANGIT	KLANGENAN	3	1	4000000	JULI-AGU	DEWASA	36	RUMAH ORANG TUA	2	0	BURUH TANI	Rp 2.100.000	LAYAK
22	20022	WASI	5205	DUSUN II	8	3	SLANGIT	KLANGENAN	2	1	4000000	JULI-AGU	DEWASA	26	RUMAH ORANG TUA	1	0	BURUH HARIAN	Rp 1.500.000	LAYAK
23	20023	MUSLIMIN	5205	DUSUN II	14	4	SLANGIT	KLANGENAN	4	2	4000000	JULI-AGU	DEWASA	36	RUMAH ORANG TUA	1	0	BURUH SERABU	Rp 1.800.000	LAYAK
24	20024	IRUNA	5205	DUSUN I	6	2	SLANGIT	KLANGENAN	4	2	4000000	JULI-AGU	DEWASA	32	MILIK SENDIRI		2	0 BURUH SERABU	Rp 1.800.000	LAYAK
25	20025	CARSINI	5205	DUSUN IV	15	5	SLANGIT	KLANGENAN	3	1	4000000	JULI-AGU	REMAJA	24	MILIK SENDIRI		1	0 BURUH SERABU	Rp 1.500.000	LAYAK
26	20026	AWANG SE	5205	DUSUN I	1	1	SLANGIT	KLANGENAN	3	1	4000000	JULI-AGU	DEWASA	30	RUMAH ORANG TUA	1	0	BURUH HARIAN	Rp 2.700.000	TIDAK LAYAK

Gambar 3 Dataset Penerima BPNT

4.2. Pengumpulan Data

Pada gambar diatas terdapat 21 variabel diantaranya Nama, NIK, Alamat, RT, RW, Desa, Kecamatan, Jumlah Keluarga, Anak, Jumlah, Periode, Kategori, Usia, Kepemilikan Rumah, Motor, Mobil, Pekerjaan, Penghasilan dan Status kelayakan. Pada fase ini, pengumpulan data dilakukan dari sumber yang efisien, diantaranya data dari Dinas Sosial, Tenaga Kesejahteraan Sosial Kecamatan (TKSK), Pemerintah Desa, dan Pusat Kesejahteraan Sosial (Puskesos) Desa Slangit.

4.3. Preprocessing

Data yang telah melewati validasi oleh seorang Ahli Bahasa Indonesia kemudian mengalami proses *text preprocessing*. Tujuan dari tahap ini adalah menghapus atribut yang memiliki dampak yang kurang signifikan terhadap hasil klasifikasi. Berikut adalah langkah-langkah *text preprocessing*. Hasil dari proses ini terdapat dalam tabel yang dapat dilihat di bawah ini :

Tabel 1 Preprocessing Data

Variabel	Sebelum	Sesudah
Alamat	Dusun I, II, III dan IV	1, 2, 3 dan 4
Kepemilkan Rumah	Milik Sendiri dan Rumah Orang Tua	0 dan 1
Status	Tidak Layak dan Layak	0 dan 1

4.4. Transformation

Dalam tahap ini, terjadi proses transformasi data menjadi format yang cocok untuk keperluan data mining. Tujuannya adalah untuk mempermudah koordinasi data yang diolah menggunakan Google Colaboratory dalam konteks klasifikasi dengan algoritma decision tree dan random forest.

```
   KELUARGA      ANAK  USIA  KEPEMILIKAN RUMAH  MOTOR  MOBIL  PENGHASILAN  \
0         4  0.285714    40                  0      2      0     0.473684
1         3  0.142857    31                  1      1      0     0.210526
2         4  0.285714    29                  0      1      0     0.298246
3         3  0.142857    31                  1      1      0     0.263158
4         3  0.142857    32                  1      1      0     0.263158

   STATUS
0     1.0
1     0.0
2     0.0
3     0.0
4     0.0
```

Gambar 4 Transformasi Data

4.5. Fitur Seleksi

Keberhasilan Fitur Seleksi terbukti dalam upaya meningkatkan hasil evaluasi pada tahap data mining. Dalam penelitian ini, algoritma Decision Tree dan Random Forest digunakan untuk menjalankan seleksi fitur. Proses seleksi fitur dilakukan setelah mengidentifikasi skenario dengan tingkat akurasi tertinggi pada tahap data mining. Akurasi tersebut selanjutnya dimanfaatkan sebagai nilai fungsi kecocokan *(fitness function)* dalam algoritma Decision Tree dan Random Forest.

4.6. Data Mining

Dalam penelitian ini, terdapat penerapan tahap Data Mining yang berfokus pada klasifikasi menggunakan algoritma Decision Tree dan Random Forest. Perhitungan klasifikasi dilakukan dengan melibatkan 8 variabel diantaranya jumlah keluarga, jumlah anak, usia, kepemilikan rumah, kepemilikan kendaraan motor, mobil, penghasilan dan status kelayakan. Dengan harapan dapat mencapai tingkat akurasi yang optimal. Proses pengujian menggunakan data uji sebesar 0.3 (30%) dari seluruh dataset.

```
[31] from sklearn.model_selection import train_test_split

     X = df.drop('STATUS', axis=1)
     y = df.STATUS

     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```

Gambar 5 Kode Presentase Data Testing

4.7. Decision Tree

```
Train Result:
================================================
Accuracy Score: 100.00%
________________________________________________
CLASSIFICATION REPORT:
                0.0   1.0  accuracy  macro avg  weighted avg
precision       1.0   1.0       1.0        1.0           1.0
recall          1.0   1.0       1.0        1.0           1.0
f1-score        1.0   1.0       1.0        1.0           1.0
support       581.0  74.0       1.0      655.0         655.0
________________________________________________
Confusion Matrix:
 [[581   0]
 [  0  74]]

Test Result:
================================================
Accuracy Score: 97.86%
________________________________________________
CLASSIFICATION REPORT:
                     0.0        1.0  accuracy   macro avg  weighted avg
precision       0.984314   0.923077  0.978648    0.953695      0.978212
recall          0.992095   0.857143  0.978648    0.924619      0.978648
f1-score        0.988189   0.888889  0.978648    0.938539      0.978294
support       253.000000  28.000000  0.978648  281.000000    281.000000
________________________________________________
Confusion Matrix:
 [[251   2]
 [  4  24]]
```

Gambar 6 Hasil Akurasi Decision Tree

Ilustrasi di atas menunjukkan bahwa pada dataset latihan, model berhasil mencapai tingkat akurasi yang optimal sebesar 100%, dan seluruh metrik lainnya juga mencapai nilai maksimum (1.0). Kemungkinan terdapat overfitting dapat diindikasikan, di mana model dapat terlalu terampil dalam menyesuaikan diri dengan data latihan tetapi kurang dapat melakukan generalisasi terhadap data baru. Pada dataset pengujian, kinerja model tetap sangat baik dengan tingkat akurasi sebesar 97.86%. Meskipun terdapat sedikit perbedaan antara nilai precision dan recall untuk kelas 1, secara keseluruhan, model ini memberikan hasil yang memuaskan pada data yang belum pernah dilihat sebelumnya.

4.8. Random Forest

```
Train Result:
================================================
Accuracy Score: 100.00%
________________________________________________
CLASSIFICATION REPORT:
                0.0   1.0  accuracy  macro avg  weighted avg
precision       1.0   1.0       1.0        1.0           1.0
recall          1.0   1.0       1.0        1.0           1.0
f1-score        1.0   1.0       1.0        1.0           1.0
support       581.0  74.0       1.0      655.0         655.0
________________________________________________
Confusion Matrix:
 [[581   0]
 [  0  74]]

Test Result:
================================================
Accuracy Score: 97.86%
________________________________________________
CLASSIFICATION REPORT:
                     0.0        1.0  accuracy   macro avg  weighted avg
precision       0.984314   0.923077  0.978648    0.953695      0.978212
recall          0.992095   0.857143  0.978648    0.924619      0.978648
f1-score        0.988189   0.888889  0.978648    0.938539      0.978294
support       253.000000  28.000000  0.978648  281.000000    281.000000
________________________________________________
Confusion Matrix:
 [[251   2]
 [  4  24]]
```

Gambar 7 Algoritma Random Forest

Gambar di atas menunjukkan bahwa model menunjukkan kinerja yang sangat baik pada data pelatihan dengan mencapai tingkat akurasi 100%. Perlu diperhatikan bahwa terdapat potensi overfitting, karena performa optimal pada data pelatihan tidak selalu mencerminkan kemampuan model untuk menghasilkan hasil yang baik pada data baru. Pada tahap pengujian, model mencapai tingkat akurasi yang tinggi, yakni 97.86%. Walaupun begitu, terdapat beberapa kesalahan dalam mengidentifikasi instance

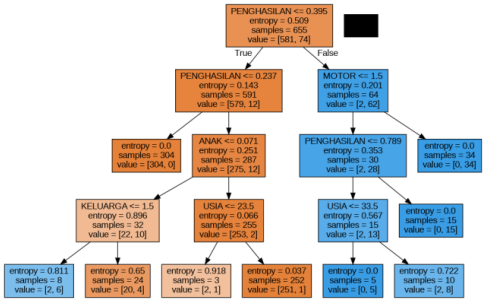
positif untuk kelas 1. Ini merupakan fokus area yang dapat diperbaiki untuk meningkatkan recall kelas 1 tanpa mengorbankan precision secara signifikan.

4.9. Evaluation

Evaluasi hasil tes dari metode Decision Tree mencapai 97.86%, dengan nilai presisi untuk kelas 0 dan kelas 1 sebesar 92.30%. Recall kelas 0 mencapai 99.20%, sementara kelas 1 mencapai 85.71%. F1-score kelas 0 sebesar 98.81%, dan kelas 1 sebesar 88.88%. Deteksi beberapa prediksi yang tidak tepat terlihat dari angka di luar diagonal utama pada matriks kebingungan, termasuk 2 false positive dan 4 false negative. Pada metode Random Forest, hasil evaluasi menunjukkan akurasi sebesar 100% untuk data pelatihan dan 97.86% untuk data uji. Presisi kelas 0 mencapai 98.43%, sedangkan kelas 1 mencapai 92.30%. Recall kelas 0 mencapai 99.20%, dan untuk kelas 1 mencapai 85.71%. F1-score kelas 0 mencapai 98.81%, sementara kelas 1 sebesar 88.88%. Analisis matriks kebingungan pada data pengujian mengungkapkan adanya 2 false positive dan 4 false negative, menunjukkan adanya kesalahan klasifikasi.

4.10. Pembahasan

Dalam penelitian ini, dilakukan klasifikasi terhadap penerima Bantuan Pangan Non Tunai (BPNT) dengan menggunakan algoritma decision tree dan random forest. Metode analisis yang diterapkan adalah KDD *(Knowledge Discovery In Database)*, dengan *Google Colaboratory* sebagai alat bantu. Proses analisis melibatkan beberapa tahap, termasuk Pengumpulan Data Teks, Preprocessing, Transformasi, Seleksi Fitur, Data Mining, dan Evaluasi. Data kemudian dipilih untuk digunakan dalam tahap klasifikasi, dengan beberapa variabel yang tidak dimasukkan dalam seleksi, seperti Nama, Desa, Kecamatan, Jumlah, Periode, dan Kategori. Pada tahap Data Mining, algoritma decision tree dan random forest diterapkan untuk memodelkan penerima bantuan sosial BPNT secara spesifik dari setiap variabel. Tingkat akurasi menggunakan algoritma decision tree mencapai 97.86%, sementara algoritma random forest menghasilkan nilai sebesar 97.86%.



Gambar 8 Visualisasi Algoritma Decision Tree

Pada gambar 8 yang menjadi root adalah variabel penghasilan dengan nilai kurang atau sama 0.395 berikut ada delapan peraturan yang berlaku pada decision tree dalam menentukan kelas data uji.

1. Apabila nilai PENGHASILAN berada di bawah atau sama dengan 0.237, data akan dikelompokkan sebagai penerima bantuan yang memenuhi syarat.
2. Apabila nilai PENGHASILAN kurang dari atau sama dengan 0.395, data akan diidentifikasi sebagai penerima bantuan yang memenuhi syarat.
3. Apabila nilai PENGHASILAN kurang dari atau sama dengan 0.395 dan nilai ANAK kurang dari atau sama dengan 0.071, data akan diklasifikasikan sebagai penerima bantuan yang memenuhi syarat.
4. Apabila nilai PENGHASILAN kurang dari atau sama dengan 0.395, nilai ANAK kurang dari atau sama dengan 0.071, dan nilai KELUARGA kurang dari atau sama dengan 1.5, data akan dikelompokkan sebagai penerima bantuan yang memenuhi syarat.
5. Apabila nilai PENGHASILAN kurang dari atau sama dengan 0.395, nilai ANAK kurang dari atau sama dengan 0.071, dan nilai USIA kurang dari atau sama dengan 2.35, data akan dianggap sebagai penerima bantuan yang tidak memenuhi syarat.
6. Apabila nilai PENGHASILAN kurang dari atau sama dengan 0.395 dan nilai MOTOR kurang dari atau sama dengan 1.5, data akan diklasifikasikan sebagai penerima bantuan yang tidak memenuhi syarat.
7. Apabila nilai PENGHASILAN kurang dari atau sama dengan 0.395, nilai MOTOR kurang dari atau sama dengan 1.5, dan nilai PENGHASILAN kurang dari atau sama dengan 0.789, data akan dikelompokkan sebagai penerima bantuan yang tidak memenuhi syarat.
8. Apabila nilai PENGHASILAN kurang dari atau sama dengan 0.395, nilai MOTOR kurang dari atau sama dengan 1.5, nilai PENGHASILAN kurang dari atau sama dengan 0.789, dan nilai USIA kurang dari atau sama dengan 33.5, data akan diidentifikasi sebagai penerima bantuan yang tidak memenuhi syarat.

5. KESIMPULAN

Penelitian ini memanfaatkan algoritma Decision Tree dan Random Forest untuk mengenali penerima Bantuan Pangan Non Tunai (BPNT) di Desa Slangit selama bulan Juli dan Agustus 2023. Proses ini melibatkan langkah-langkah seperti pengumpulan data, identifikasi sumber data, tinjauan literatur, perumusan masalah, pra-pemrosesan data, implementasi model klasifikasi, Knowledge Discovery in Database (KDD), evaluasi model, dan pembuatan kesimpulan. Data yang digunakan berasal dari Pemerintah Desa, Dinas Sosial Kabupaten Cirebon, dan Data Terpadu Kesejahteraan Sosial (DTKS). Hasil penelitian menunjukkan keberhasilan algoritma Decision Tree dan Random Forest dengan tingkat akurasi mencapai 97.86%. Evaluasi model

menunjukkan bahwa keduanya memberikan nilai precision, recall, dan F1-score yang memuaskan, menunjukkan efektivitas mereka dalam mengidentifikasi penerima BPNT. Temuan ini berpotensi mendukung proses penentuan penerima bantuan sosial dengan lebih efisien dan akurat.

DAFTAR PUSTAKA

[1] R. Syahputra and W. Riansah, "PREDIKSI PENERIMA BANTUAN PANGAN NON-TUNAI DENGAN METODE LEARNING VECTOR QUANTIZATION PADA DESA TANJUNG SELAMAT," *J. Educ. …*, 2021, [Online]. Available: https://journal.ipts.ac.id/index.php/ED/article/view/3042

[2] Y. G. Winarti and A. Yaskur, "PENGARUH BANTUAN PANGAN NON TUNAI TERHADAP TINGKAT KEMISKINAN DI KOTA MAGELANG MELALUI ANALISIS SIMULASI," *J. Jendela Inov. Drh.*, 2022, [Online]. Available: http://jurnal.magelangkota.go.id/index.php/cendelainovasi/article/view/118

[3] F. Bagi, I. Igirisa, and R. Tantu, "Implementasi Program Bantuan Pangan Non Tunai (BPNT) di Kelurahan Heledulaa Utara, Kecamatan Kota Timur," *ULIL ALBAB J. Ilm. …*, vol. 2, no. 8, 2023, [Online]. Available: https://journal-nusantara.com/index.php/JIM/article/view/1967

[4] R. A. Islahudin, S. Rahmatullah, A. Afandi, and ..., "ALGORITMA C4. 5 UNTUK MEMPREDIKSI KELAYAKAN PENERIMA BANTUAN PANGAN NON TUNAI," *J. …*, 2022, [Online]. Available: https://jurnal.darmajaya.ac.id/index.php/JurnalInformatika/article/view/3367

[5] A. Damuri, U. Riyanto, and ..., "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal …*, 2021, [Online]. Available: http://www.ejurnal.stmik-budidarma.ac.id/index.php/jurikom/article/view/3655

[6] N. Purnomo, S. Defit, and Y. Yuhandri, "Klasifikasi Penerima Bantuan Pangan Non Tunai Menggunakan Metode Decision Tree," *J. Inf. dan Teknol.*, 2021, [Online]. Available: https://www.jidt.org/jidt/article/view/148

[7] Q. Iman and A. W. Wijayanto, "Klasifikasi Rumah Tangga Penerima Beras Miskin (Raskin)/Beras Sejahtera (Rastra) Di Provinsi Jawa Barat Tahun 2017 Dengan Metode Random Forest Dan …," *JUSTIN (Jurnal Sist. dan Teknol. …*, 2021, [Online]. Available: https://jurnal.untan.ac.id/index.php/justin/article/view/44137

[8] M. Yasir and R. Suraji, "PERBANDINGAN METODE KLASIFIKASI NAÏVE BAYES, DECISION TREE, RANDOM FOREST TERHADAP ANALISIS SENTIMEN KENAIKAN BIAYA HAJI 2023 …," *J. Cahaya Mandalika*, 2023, [Online]. Available: https://ojs.cahayamandalika.com/index.php/JCM/article/view/1520

[9] A. Prabowo, S. Wardani, R. W. Dewantoro, and ..., "Komparasi Tingkat Akurasi Random Forest dan Decision Tree C4. 5 Pada Klasifikasi Data Penyakit Infertilitas," *KLIK Kaji. Ilm. …*, 2023, [Online]. Available: http://djournals.com/klik/article/view/1115

[10] D. Pajri, Y. Umaidah, and T. N. Padilah, "K-nearest neighbor berbasis particle swarm optimization untuk analisis sentimen terhadap Tokopedia," *J. Tek. Inform. dan Sist. …*, 2020, [Online]. Available: http://114.7.153.31/index.php/jutisi/article/view/2658

[11] D. H. Depari, Y. Widiastiwi, and ..., "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *Inform. J. Ilmu …*, 2022, [Online]. Available: https://ejournal.upnvj.ac.id/informatik/article/view/4694

[12] A. M. A. Rahim, I. Y. R. Pratiwi, and M. A. Fikri, "Klasifikasi Penyakit Jantung Menggunakan Metode Synthetic Minority Over-Sampling Technique Dan Random Forest Clasifier," *Indones. J. Comput. …*, 2023, [Online]. Available: http://3.8.6.95/ijcs/index.php/ijcs/article/view/3413