

Penerapan Algoritma ID3 dan Algoritma C4.5 Untuk Klasifikasi Penerima BPNT

Minhatin Nisaatus Sholikhah¹, Dinita Rahmalia², Mohammad Syaiful Pradana³

¹Universitas Islam Darul 'ulum Lamongan, minhatin@mhs.unisda.ac.id

²Universitas Islam Darul 'ulum Lamongan, dinitarahmalia@gmail.com

³Universitas Islam Darul 'ulum Lamongan, syaifulp@unisda.ac.id

Abstract. Non-Cash Food Assistance (BPNT) is social food assistance in the form of non-cash from the government which is given to Beneficiary Families (KPM) every month through an electronic account mechanism which is used only to buy food at traders or e-warongs. One of the difficulties that the government sometimes faces in distributing BPNT is that the distribution process is uneven and not on target. Therefore, it is necessary to carry out further analysis using a mathematical approach, so that we can determine the feasibility of a BPNT recipient prediction problem. Through the results of the data collection analysis, it can be seen whether residents are eligible to receive BPNT or not. Based on existing problems, a classification method is used to predict the eligibility of BPNT beneficiaries using two methods, namely the ID3 algorithm and the C4.5 algorithm. The ID3 algorithm produces an accuracy value of 90%, precision of 100%, and recall of 83.33%. The C4.5 algorithm produces an accuracy value of 80%, precision of 100%, and recall of 80%. The AUC/ROC value of the ID3 algorithm is 0.500, the classification is diagnosed in the AUC/ROC curve as failure or failure in classification. The C4.5 algorithm has an AUC/ROC value of 0.800, meaning that the classification is included in good classification. In this way, it can be concluded that the C4.5 algorithm has better results compared to the ID3 algorithm.

Keywords: ID3 Algorithm, C4.5 Algorithm, Classification, BPNT Feasibility.

Abstrak. Bantuan Pangan Non Tunai (BPNT) adalah bantuan sosial pangan dalam bentuk non tunai dari pemerintah yang diberikan kepada Keluarga Penerima Manfaat (KPM) setiap bulannya melalui mekanisme akun elektronik yang digunakan hanya untuk membeli bahan pangan di pedagang atau *e-warong*. Salah satu kesulitan yang terkadang dihadapi oleh pemerintah dalam penyaluran BPNT adalah proses pembagian yang tidak merata dan tidak tepat sasaran. Oleh karena itu perlu dilakukan analisis lanjutan dengan pendekatan matematis, sehingga dapat mengetahui kelayakan dari suatu permasalahan prediksi penerima BPNT. Melalui hasil analisa pendataannya, dapat diketahui warga layak atau tidak untuk menerima BPNT. Berdasarkan permasalahan yang ada, digunakan metode klasifikasi untuk memprediksi kelayakan penerima manfaat BPNT dengan menggunakan dua metode, yaitu algoritma ID3 dan algoritma C4.5. Algoritma ID3 menghasilkan nilai akurasi sebesar 90%, presisi 100%, dan recall 83,33%. Algoritma C4.5 menghasilkan nilai akurasi sebesar 80%, presisi 100%, dan recall 80%. Nilai AUC/ROC algoritma ID3 sebesar 0,500, klasifikasi tersebut di diagnosa dalam kurva AUC/ROC sebagai *failurer* atau kegagalan dalam klasifikasi. Algoritma C4.5 memiliki nilai AUC/ROC sebesar 0,800, artinya klasifikasi tersebut termasuk dalam *good clasification* atau klasifikasi yang baik. Dengan begitu, dapat disimpulkan bahwa Algoritma C4.5 memiliki hasil yang lebih baik dibandingkan dengan algoritma ID3.

Kata Kunci: Algoritma ID3, Algoritma C4.5, Klasifikasi, Kelayakan BPNT.

1 Pendahuluan

Kemiskinan merupakan masalah yang sering ditemukan di negara berkembang, termasuk Indonesia. Berbagai upaya telah dilakukan oleh pemerintah untuk menanggulangi kemiskinan, diantaranya dengan mengadakan program bantuan sosial untuk rakyat miskin [1]. Bantuan Pangan Non Tunai (BPNT) adalah bantuan sosial pangan dalam bentuk non tunai dari pemerintah yang diberikan kepada Keluarga Penerima Manfaat (KPM) setiap bulannya melalui mekanisme akun elektronik yang digunakan hanya untuk membeli bahan pangan di pedagang atau *e-warong* [2].

Beberapa kriteria yang menjadi bahan pertimbangan dalam penentuan penerima BPNT adalah umur, pekerjaan, kondisi tempat tinggal, pendapatan, dan jumlah tanggungan [3]. Untuk menentukan siapa saja yang berhak menerima BPNT, maka pihak desa maupun pihak dinas sosial harus lebih teliti dalam mempertimbangkan aspek-aspek yang dapat mendukung dalam mengambil keputusan. Aspek-aspek tersebut dapat berupa data umur, pekerjaan, kondisi tempat tinggal dan lain-lain. Dalam pengolahannya pun harus lebih diperhatikan setiap kriterianya, agar penyaluran BPNT dapat berlangsung secara efektif dan tepat sasaran. Terdapat banyak sekali teknik dalam pengolahan data. Salah satunya yaitu dengan menggunakan pendekatan matematis metode data mining.

Data Mining atau menambang data didefinisikan sebagai metode yang digunakan untuk mengekstraksi informasi prediktif tersembunyi pada database. Salah satu teknik dalam data mining yaitu teknik klasifikasi. Klasifikasi merupakan salah satu teknik pada data mining yang memetakan data ke dalam kelompok atau kelas yang telah ditentukan. Metode klasifikasi yang akan digunakan dalam hal ini yaitu *Decision Tree* metode algoritma ID3 dan algoritma C4.5. *Decision Tree* atau pohon keputusan merupakan pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan. Algoritma ID3 dan Algoritma C4.5 merupakan algoritma penginduksi *Decision Tree* berdasarkan kriteria *entropy*. Algoritma ID3 dan Algoritma C4.5 dalam analisisnya didasarkan pada pemilihan titik pemecahan yang memaksimalkan *informasi gain* (pengurangan *entropy* maksimal).

Dari penelitian yang dilakukan oleh [4] menyimpulkan metode klasifikasi menggunakan algoritma ID3 dan *software Rapidminer* menghasilkan akurasi sebesar 97,89% dengan 92 data, sehingga algoritma ini sangat baik digunakan dalam klasifikasi penentuan penerima program bantuan lanjut usia. Dari penelitian lain yang dilakukan oleh [5] diperoleh hasil bahwa metode *decision tree* menggunakan algoritma C4.5 dianggap sebagai algoritma yang sangat membantu dalam melakukan klasifikasi data karena karakteristik data yang diklasifikasikan dapat diperoleh dengan jelas, baik dalam bentuk struktur pohon keputusan (*decision tree*) maupun dalam bentuk aturan atau *rule if – then*, sehingga memudahkan pengguna dalam melakukan penggalian informasi terhadap data yang bersangkutan.

Berdasarkan uraian penelitian tersebut, dapat diketahui bahwa kedua algoritma yaitu ID3 dan C4.5 mempunyai kinerja yang baik dalam teknik mengklasifikasikan data. Namun, belum diketahui algoritma mana diantara keduanya yang lebih unggul kinerjanya. Oleh karena itu, dalam penelitian ini akan menentukan algoritma yang memiliki tingkat akurasi, presisi, recall, dan AUC yang lebih tinggi pada klasifikasi penerima BPNT di Desa XYZ.

2 Metode Penelitian

Pada tahap pertama penelitian yakni mencari, mengumpulkan dan mempelajari berbagai referensi dan informasi yang sesuai dengan permasalahan yang dikaji. Sumber referensi dapat berasal dari buku, dan jurnal ilmiah serta sumber informasi terkait bantuan ekonomi dari pemerintah. Data yang digunakan pada penelitian ini merupakan data sekunder berupa daftar penerima BPNT di Desa XYZ.

BPNT merupakan bantuan sosial dalam bentuk non tunai, bantuan tersebut berupa makanan pokok. Berdasarkan data yang diambil penulis dalam penelitian ini yaitu sebanyak 50 data, yang terdiri dari 32 data penerima BPNT dan 18 data yang tidak menerima BPNT. terdapat 11 variabel data penunjang yang digunakan dalam hal ini yaitu nama, jenis kelamin, status perkawinan, pekerjaan, kepemilikan tempat tinggal, sumber air, pemakaian bahan bakar/energi, jenis lantai, jenis dinding, jumlah pendapatan bersih, dan jumlah tanggungan.

Untuk mendapatkan data penunjang tersebut penulis melakukan observasi secara langsung lewat wawancara yang dilakukan bersama calon penerima BPNT selama kurun waktu 5 hari, dari tanggal 15 – 20 Juni 2022. Data penunjang yang dimaksud meliputi umur, pekerjaan, kondisi tempat tinggal, pendapatan, dan jumlah tanggungan. Data yang digunakan untuk penelitian ini sebanyak 32 data penerima BPNT dan 18 data pendaftar yang tidak lolos dalam pencairan dana Bantuan BPNT di Desa XYZ.

Proses klasifikasi menggunakan algoritma ID3 dan algoritma C4.5 melalui beberapa tahapan yaitu pengumpulan data, *preprocessing* data, proses analisis menggunakan rumus dari kedua algoritma, analisis menggunakan *software rapidminer*, menghitung akurasi, presisi, recall dan AUC/ROC menggunakan *software rapidminer*, dan penarikan kesimpulan

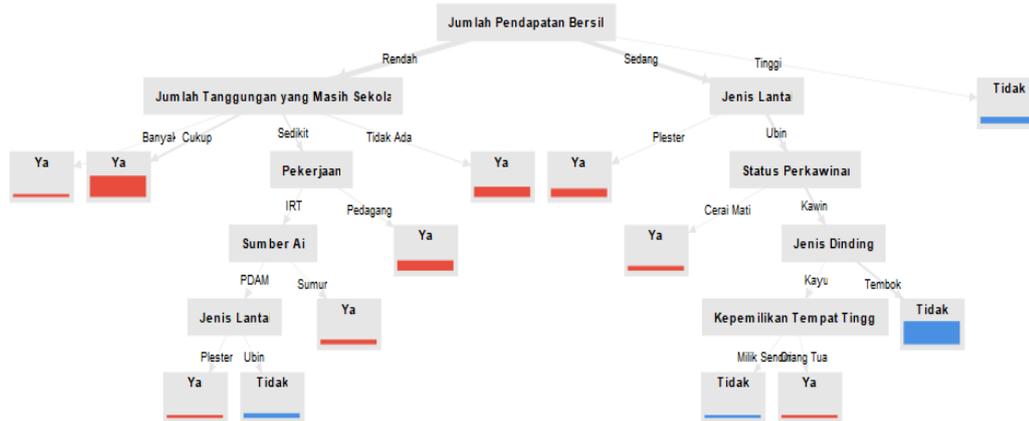
3 Hasil dan Pembahasan

Perancangan dan Pengujian

Data yang diperoleh pada penelitian ini sejumlah 50 data yang meliputi nama, jenis kelamin, status perkawinan, pekerjaan, kepemilikan tempat tinggal, sumber air, pemakaian bahan bakar/energi, jenis lantai, jenis dinding, jumlah pendapatan bersih, dan jumlah tanggungan. Label kelas yang digunakan dalam hal ini yaitu “Ya” untuk yang menerima BPNT, dan “Tidak” yang tidak menerima BPNT. Dalam tahap perancangan data ini, untuk mempermudah pengklasifikasian, maka data tersebut dilakukan *Preprocessing* data terlebih dahulu. *Preprocessing* data *preprocessing* data, tujuannya yaitu untuk memastikan bahwa data yang akan digunakan merupakan data yang memiliki kualitas yang baik sebelum digunakan untuk analisis data. Setelah dilakukan *preprocessing* data yang sebelumnya terdapat 11 variabel menjadi hanya 9 variabel. Variabel nama dihilangkan karena nama merupakan id label, dan variabel pemakaian bahan bakar/energi dihilangkan karena semua data yang diperoleh memiliki kriteria yang sama jadi tidak memiliki pengaruh untuk hasilnya nanti.

Pengujian Data dengan Algoritma ID3

Pengolahan algoritma ID3 tidak dapat memakai data numerik, maka data-data yang bernilai numerik diubah menjadi data kategorikal yaitu jumlah pendapatan yang memiliki tiga kriteria (rendah, sedang, dan tinggi), dan jumlah tanggungan yang memiliki 4 kriteria (tidak ada, sedikit, cukup, dan banyak). Hasil *decision tree* Algoritma ID3 ditunjukkan pada Gambar 1 berikut.



Gambar 1. *Decision Tree* Algoritma ID3

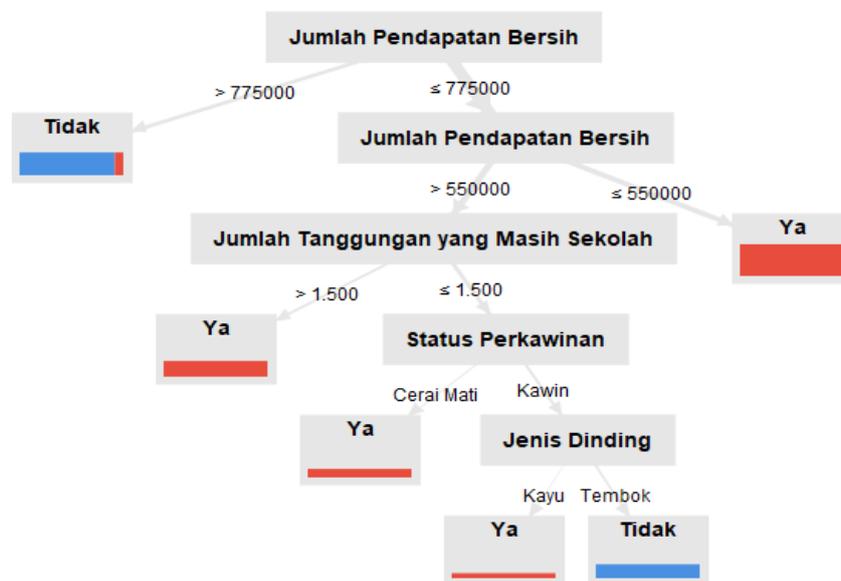
Pohon yang dihasilkan juga dapat dinyatakan dalam bentuk aturan IF THEN sebagai berikut:

1. IF Jumlah Pendapatan = Rendah AND Jumlah Tanggungan = Banyak THEN Menerima = Ya
2. IF Jumlah Pendapatan = Rendah AND Jumlah Tanggungan = Cukup THEN Menerima = Ya
3. IF Jumlah Pendapatan = Rendah AND Jumlah Tanggungan = Tidak Ada THEN Menerima = Ya
4. IF Jumlah Pendapatan = Rendah AND Jumlah Tanggungan = Sedikit AND Pekerjaan = IRT AND Sumber Air = PDAM AND Jenis Lantai = Plester THEN Menerima = Ya
5. IF Jumlah Pendapatan = Rendah AND Jumlah Tanggungan = Sedikit AND Pekerjaan = IRT AND Sumber Air = PDAM AND Jenis Lantai = Ubin THEN Menerima = Tidak
6. IF Jumlah Pendapatan = Rendah AND Jumlah Tanggungan = Sedikit AND Pekerjaan = IRT AND Sumber Air = Sumur THEN Menerima = Ya
7. IF Jumlah Pendapatan = Rendah AND Jumlah Tanggungan = Sedikit AND Pekerjaan = Pedagang THEN Menerima = Ya
8. IF Jumlah Pendapatan = Sedang AND Jenis Lantai = Plester THEN Menerima = Ya
9. IF Jumlah Pendapatan = Sedang AND Jenis Lantai = Ubin AND Status Perkawinan = Cerai Mati THEN Menerima = Ya
10. IF Jumlah Pendapatan = Sedang AND Jenis Lantai = Ubin AND Status Perkawinan = Kawin AND Jenis Dinding = Kayu AND Kepemilikan Rumah = Milik Sendiri THEN Menerima = Tidak

11. IF Jumlah Pendapatan = Sedang AND Jenis Lantai = Ubin AND Status Perkawinan = Kawin AND Jenis Dinding = Kayu AND Kepemilikan Rumah = Orang Tua THEN Menerima = Ya
12. IF Jumlah Pendapatan = Sedang AND Jenis Lantai = Ubin AND Status Perkawinan = Kawin AND Jenis Dinding = Tembok THEN Menerima = Tidak
13. IF Jumlah Pendapatan = Tinggi THEN Menerima = Tidak

Pengujian Data dengan Algoritma C4.5

Algoritma C4.5 dapat mengolah data numerik, jadi data yang digunakan yaitu data yang telah melalui tahap *preprocessing* tanpa mengubahnya menjadi kategori. Hasil *decision tree* Algoritma C4.5 ditunjukkan pada Gambar 2 berikut.



Gambar 2, Decision Tree Algoritma C4.5

Pohon yang dihasilkan juga dapat dinyatakan dalam bentuk aturan IF THEN sebagai berikut:

1. IF Jumlah Pendapatan = >775.000 THEN Menerima = Tidak
2. IF Jumlah Pendapatan = $775.000-550.000$ AND Jumlah Tanggungan = >1 THEN Menerima = Ya
3. IF Jumlah Pendapatan = $775.000-550.000$ AND Jumlah Tanggungan = ≤ 1 AND Status Perkawinan = Cerai Mati THEN Menerima = Ya
4. IF Jumlah Pendapatan = $775.000-550.000$ AND Jumlah Tanggungan = ≤ 1 AND Status Perkawinan = Cerai Mati AND Jenis Dinding = Kayu THEN Menerima = Ya
5. IF Jumlah Pendapatan = $775.000-550.000$ AND Jumlah Tanggungan = ≤ 1 AND Status Perkawinan = Cerai Mati AND Jenis Dinding = Tembok THEN Menerima = Tidak
6. IF Jumlah Pendapatan = ≤ 550.000 THEN Menerima = Ya

Perhitungan Akurasi, Presisi, Recall, dan Kurva AUC/ROC

Pengujian data dilakukan mulai dari 10% hingga 90% dari data keseluruhan 50 data. Pengujian data terbagi sebagai berikut ini 5 (10%), 10 (20%), 15 (30%), 35(70%), 40 (80%), dan 45 (90%) data. Hasil perhitungan akurasi, presisi, recall dan kurva AUC/ROC dari algoritma ID3 dan algoritma C4.5 ditunjukkan pada Tabel 1 berikut.

Tabel 1. Perbandingan Akurasi Data Testing dan Data Training

Perbandingan Akurasi	Training 90% Testing 10%	Training 80% Testing 20%	Training 70% Testing 30%
Algoritma ID3	80,00%	90,00%	73,33%
Algoritma C4.5	80,00%	70,00%	80,00%

Hasil pengujian menunjukkan bahwa Algoritma ID3 memiliki nilai akurasi tertinggi pada pengujian 40 (80%) data training, dan 10 (20%) data testing dengan hasil akurasi sebesar 90%. Sedangkan algoritma C4.5 memiliki nilai akurasi tertinggi pada pengujian 45 (90%) dan 35 (70%) data training, dan 5 (10%) dan 15 (30%) data testing dengan hasil akurasi sebesar 80%. Sehingga dari rata-rata yang didapatkan dari seluruh pengujian tersebut, algoritma ID3 memiliki nilai akurasi yang lebih tinggi dari pada algoritma C4.5.

Hasil perbandingan presisi dari algoritma ID3 dan algoritma C4.5 ditunjukkan pada Tabel 2 berikut.

Tabel 2. Perbandingan Presisi Data Testing dan Data Training

Perbandingan Presisi	Training 90% Testing 10%	Training 80% Testing 20%	Training 70% Testing 30%
Algoritma ID3	100,00%	100,00%	100,00%
Algoritma C4.5	100,00%	80,00%	88,89%

Hasil pengujian menunjukkan bahwa Algoritma ID3 memiliki nilai presisi tertinggi pada seluruh pengujian data training dan testing dengan nilai presisi sebesar 100%. Sedangkan algoritma C4.5 memiliki nilai akurasi tertinggi pada pengujian 45 (90%) data training, dan 5 (10%) data testing dengan hasil presisi sebesar 100%. Sehingga dari rata-rata yang didapatkan dari seluruh pengujian tersebut, algoritma ID3 memiliki nilai presisi yang lebih tinggi dari pada algoritma C4.5.

Hasil perbandingan recall dari algoritma ID3 dan algoritma C4.5 ditunjukkan pada Tabel 3 berikut.

Tabel 3. Perbandingan Recall Data Testing dan Data Training

Perbandingan Recall	Training 90% Testing 10%	Training 80% Testing 20%	Training 70% Testing 30%
Algoritma ID3	66,67%	83,33%	60,00%
Algoritma C4.5	66,67%	66,67%	80,00%

Hasil pengujian menunjukkan bahwa Algoritma ID3 memiliki nilai recall tertinggi pada pengujian 40 (80%) data training dan 10 (20%) data testing dengan nilai recall sebesar 83,33%. Sedangkan algoritma C4.5 memiliki nilai recall

tertinggi pada pengujian 35 (70%) data training, dan 15 (30%) data testing dengan hasil recall sebesar 80%. Sehingga dari rata-rata yang didapatkan dari seluruh pengujian tersebut, algoritma C4.5 memiliki nilai recall yang lebih tinggi dari pada algoritma ID3.

Hasil perbandingan nilai AUC dari algoritma ID3 dan algoritma C4.5 ditunjukkan pada Tabel 4 berikut.

Tabel 4. Perbandingan Recall Data Testing dan Data Training

Perbandingan AUC/ROC	Training 90% Testing 10%	Training 80% Testing 20%	Training 70% Testing 30%
Algoritma ID3	0.500	0.500	0.500
Algoritma C4.5	0.500	0.708	0.800

Hasil pengujian menunjukkan bahwa Algoritma ID3 memiliki nilai AUC/ROC tertinggi pada seluruh pengujian data training dan data testing dengan hasil AUC sebesar 0.500. Dalam pengukuran tingkat nilai diagnosa, nilai AUC/ROC sebesar 0.500 termasuk dalam kategori *failure* atau kegagalan dalam klasifikasi. Sedangkan Algoritma C4.5 memiliki nilai AUC/ROC tertinggi pada pengujian 35 (70%) data training dan 15 (30%) data testing dengan hasil AUC sebesar 0.800. Dalam pengukuran tingkat nilai diagnosa, nilai AUC/ROC sebesar 0.800 termasuk dalam kategori *good clasification* atau klasifikasi yang baik.

Dengan nilai akurasi, presisi, recall dan AUC/ROC tersebut maka dapat disimpulkan bahwa algoritma C4.5 memiliki nilai yang lebih baik dari pada algoritma ID3. Jadi rule/pola yang akan digunakan untuk implementasi sistem pendukung keputusan adalah rule/pola dari algoritma C4.5.

4 Kesimpulan

Dengan tahap yang sama algoritma C4.5 menghasilkan pohon keputusan yang lebih sederhana dari algoritma ID3. Algoritma ID3 memiliki nilai akurasi sebesar 90%, presisi 100%, dan recall 83,33%. Sedangkan algoritma C4.5 memiliki nilai akurasi sebesar 80%, presisi 100%, dan recall 80%. Nilai AUC/ROC algoritma ID3 sebesar 0.500, klasifikasi tersebut di diagnosa dalam kurva AUC/ROC sebagai *failurer* atau kegagalan dalam klasifikasi. Sedangkan algoritma C4.5 memiliki nilai AUC/ROC sebesar 0.800, artinya klasifikasi tersebut termasuk dalam *good clasification* atau klasifikasi yang baik. Sehingga algoritma C4.5 memiliki hasil yang lebih baik dibandingkan dengan algoritma ID3, sehingga rule atau pola yang dihasilkan oleh algoritma C4.5 dapat digunakan dalam implementasi pembuatan sistem pendukung keputusan.

Saran yang dapat diberikan penulis untuk penelitian lebih lanjut yaitu dengan menambahkan variable lain yang memiliki hubungan erat dengan masalah Bantuan Pangan Non Tunai (BPNT) dan dilanjutkan dengan menggunakan metode lainnya untuk mendapatkan perbandingan yang lebih baik.

4 Daftar Pustaka

- [1] D. Iskandar and Y. K. Suprpto, "Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan Antara Algoritma C 4.5 Dan Naïve Bayes," *Netw. Eng. Res. Oper.*, vol. 2, no. 1, 2016.

- [2] P. Julianto, “Implementasi Program Bantuan Pangan non Tunai (BPNT) di Kecamatan Sitinjau Laut Kabupaten Kerinci,” *Qawwam Leader’s Writ.*, vol. 1, no. 1, pp. 38–43, 2020.
- [3] M. Yusnita, D. Agustina, and W. Anggita, “E-Warong as One of The Solution to Eradicate Poverty in Society [E-Warong sebagai Salah Satu Solusi Mengentaskan Kemiskinan Masyarakat],” *Proceeding Community Dev.*, vol. 1, pp. 146–150, 2018.
- [4] D. K. Widiyati, M. Wati, and H. S. Pakpahan, “Penerapan Algoritma ID3 Decision Tree Pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara,” *J. Rekayasa Teknol. Inf.*, vol. 2, no. 2, p. 125, 2018.
- [5] L. N. Rani, “Klasifikasi Nasabah Menggunakan Algoritma C4. 5 Sebagai Dasar Pemberian Kredit,” *INOVTEK Polbeng-Seri Inform.*, vol. 1, no. 2, pp. 126–132, 2016.