

# Prediksi Pengangguran Menggunakan Decision Tree Dengan Algoritma C5.0 Pada Data Penduduk Kecamatan Caringin Kabupaten Bogor

Muhammad Dzatul Kahfi<sup>1</sup>, Fajri Rakhmat Umbara<sup>2</sup>, Herdi Ashaury<sup>3</sup>

<sup>1,2,3</sup> Universitas Jenderal Achmad Yani Cimahi, Jl. Terusan Jend. Sudirman, Kota Cimahi 40531, Indonesia

e-mail: mdzatulkahfi17@if.unjani.ac.id<sup>1</sup>, fajri.rakhmat@lecture.unjani.ac.id<sup>2</sup>, herdi.ashaury@lecture.unjani.ac.id<sup>3</sup>

INFORMASI ARTIKEL	ABSTRAK
<p><i>Sejarah Artikel:</i> Diterima Redaksi : 27 Februari 2022 Revisi Akhir : 29 Juni 2022 Diterbitkan Online : 30 November 2022</p> <p><i>Kata Kunci:</i> C5.0, Data mining, Decision tree, Pengangguran, Prediksi.</p> <p><i>Korespondensi:</i> Telepon / Hp : +62 91908332330 E-mail : : mdzatulkahfi17@if.unjani.ac.id</p>	<p>Tingkat kesejahteraan dalam kehidupan bermasyarakat dapat dilihat dari tingkat penganggurannya. Pemerintah daerah biasanya mengadakan sebuah program untuk membantu mengurangi jumlah pengangguran, entah itu dengan mengadakan sebuah pelatihan atau hal lain yang dapat mendorong kreativitas masyarakat dan meningkatkan kemampuan hardskill agar dapat bersaing di dunia kerja. Ada banyak penelitian yang memprediksi tingkat pengangguran dan juga ada penelitian yang menggunakan algoritma C5.0 untuk melakukan prediksi, namun belum ada penelitian yang menggabungkan subjek dan metode tersebut. Tujuan dari penelitian ini adalah untuk membuat sebuah model prediksi menggunakan algoritma C5.0 terhadap data penduduk kecamatan caringin dan mencari skenario dengan hasil akurasi yang paling tinggi. namun terdapat beberapa permasalahan yang harus dihadapi seperti bagaimana tingkat akurasi Model klasifikasi Decision Tree dengan algoritma C5.0 terhadap dataset penduduk Kecamatan Caringin dan Bagaimana rasio data latih data uji dan penggunaan pruning mempengaruhi tingkat akurasi prediksi yang akan dilakukan. Penelitian ini dievaluasi menggunakan beberapa skenario rasio data latih dan data uji yang berbeda beda dan penggunaan pruning yang berbeda. Hasil dari penelitian ini adalah model prediksi pengangguran berhasil dibuat dengan tingkat akurasi paling tinggi yaitu pada skenario data latih 70% dan data uji 30% dengan menerapkan teknik post pruning.</p>

## 1. PENDAHULUAN

Tingkat kesejahteraan dalam kehidupan bermasyarakat dapat dilihat dari tingkat penganggurannya [1], dan tingkat pengangguran yang tinggi di suatu daerah dapat menjadi suatu masalah di sektor sosial maupun ekonomi suatu daerah [1], pemerintah daerah biasanya mengadakan sebuah program untuk membantu mengurangi jumlah pengangguran, entah itu dengan mengadakan sebuah pelatihan atau hal lain yang dapat mendorong kreativitas masyarakat dan meningkatkan kemampuan agar dapat bersaing di dunia kerja.

Dalam bidang pengolahan data, teknik klasifikasi merupakan salah satu metode yang dapat digunakan untuk melakukan prediksi [2]. Klasifikasi adalah proses menemukan model yang menggambarkan dan membedakan kelas atau konsep data. Model diturunkan berdasarkan analisis sekumpulan data latih (yaitu, objek data yang label kelasnya diketahui). Model digunakan untuk memprediksi label kelas dari objek yang label kelasnya tidak diketahui.[3]. salah satu teknik klasifikasi adalah decision tree, Decision tree merupakan model prediksi menggunakan struktur pohon atau struktur

hirarki. Konsep dari decision tree adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan [4]. Decision Tree (DT) juga dapat disebut sebagai pembelajaran induktif. Untuk kumpulan data tertentu, tujuannya adalah untuk membangun model yang menangkap mekanisme yang memunculkan data. Proses penyusunan model berjenis "Supervised learning" karena pelatihan diawasi oleh variabel outcome yang disebut target [5]. Salah satu algoritma decision tree adalah algoritma C5.0, algoritma C5.0 merupakan penyempurnaan dari algoritma sebelumnya yaitu, ID3 dan C4.5 yang diperkenalkan terlebih dahulu oleh J.Ross Quinlan pada tahun 1987 [6].

Dari beberapa penelitian yang menggunakan algoritma C5.0 untuk melakukan prediksi atau klasifikasi, terdapat berbagai tepok penelitian seperti penggunaan algoritma C5.0 di bidang kesehatan [7][8]–[11], lalu ada beberapa penelitian yang menggunakan algoritma C5.0 di bidang pendidikan khususnya memprediksi lama studi mahasiswa [12]–[14], dan ada juga penelitian yang menggunakan algoritma C5.0 di bidang sosial masyarakat seperti klasifikasi pendapatan masyarakat, klasifikasi penentuan penerima bantuan

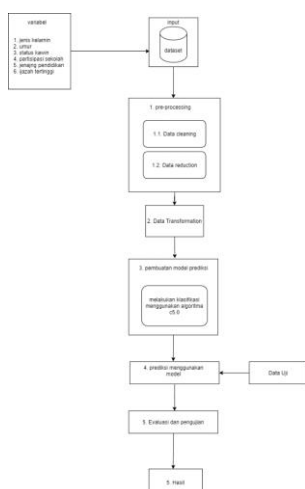
sosial, dan klasifikasi status kemiskinan rumah tangga [15]–[17]. Namun penelitian yang menggunakan algoritma C5.0 untuk memprediksi status pengangguran belum pernah dilakukan.

Berdasarkan permasalahan dan penelitian terkait diatas penelitian ini mengusulkan untuk menggunakan algoritma C5.0. untuk membuat model prediksi yang memiliki tingkat akurasi tinggi berdasarkan data penduduk kecamatan caringin. Decision tree dipilih sebagai metode klasifikasi dalam penelitian ini karena dari beberapa penelitian yang membandingkan berbagai metode klasifikasi dengan metode decision tree [18] [19] [20] , metode decision tree menjadi metode yang memiliki tingkat akurasi tertinggi. Algoritma C5.0 dipilih karena algoritma ini merupakan pengembangan dari algoritma C4.5 dan ID3 dimana algoritma C5.0 menggunakan lebih sedikit memori dan tree yang dihasilkan lebih ringkas [6]. Penelitian ini akan melakukan perbandingan evaluasi terhadap pembagian data latih data uji dan penggunaan pruning yang berbeda beda. pembagian data latih dan data uji yang akan digunakan yaitu 60:40, 70:30, 80:20 dan 90:10. Lalu setiap pembagian data latih tersebut akan diterapkan teknik pruning. Tujuannya adalah untuk mencari kombinasi pembagian data dan penggunaan pruning yang optimal untuk mendapatkan nilai akurasi tertinggi.

Isi pendahuluan mengandung latar belakang, tujuan, identifikasi masalah dan metode penelitian, yang dipaparkan secara tersirat (implisit). Kecuali bab Pendahuluan dan bab Kesimpulan, penulisan judul-judul bab sebaiknya eksplisit menyesuaikan isinya. Tidak harus implisit dinyatakan sebagai dasar teori, perancangan, dan sebagainya.

## 2. METODE PENELITIAN

Adapun tahapan metode penelitian yang digunakan pada penelitian ini seperti yang dapat dilihat pada gambar 1.



Gambar 1. Metode penelitian

### 2.1. Pre-processing

Tahap pre-processing dilakukan untuk menyiapkan data agar bisa diolah oleh sistem dengan cara mengatasi

data yang tidak lengkap, menghilangkan atribut yang tidak dibutuhkan, dan merubah data kedalam format yang sama. Dalam penelitian ini dilakukan beberapa teknik pre-processing yaitu:

Data Cleaning untuk menghilangkan atau mengisi nilai yang kosong dengan nilai yang sesuai. Dalam tahap ini penduduk dengan usia diluar angkatan kerja yaitu penduduk dengan usia dibawah 15 tahun dan diatas 65 tahun akan dihilangkan.

Data reduction untuk menghilangkan atribut yang tidak akan digunakan dalam pembuatan model prediksi, pada awalnya dataset memiliki 22 atribut. Menurut penelitian [21] [22] beberapa faktor yang memiliki pengaruh yang signifikan terhadap peluang pengangguran diantaranya yaitu pendidikan, umur dan status perkawinan. Apabila disesuaikan dengan atribut yang terdapat pada dataset maka terdapat 6 atribut atribut yang akan dipakai sebagai penentu, yaitu jenis kelamin, usia, status kawin, partisipasi sekolah, jenjang pendidikan, ijazah tertinggi.

### 2.2. Data Transformation

Pada tahap ini bentuk data diubah menjadi format yang sama. Karena penelitian ini menggunakan metode decision tree maka bentuk data numerik akan diubah menjadi kategorial, seperti Atribut usia diubah menjadi rentang usia. contoh penerapannya dapat dilihat pada tabel 1.

Tabel 1 Transformasi atribut usia

Atribut	Data Asli	Setelah Transformasi
Usia	23	26-35
	31	26-35

### 2.3. Algoritma C5.0

Algoritma C5.0 merupakan algoritma berbasis decision tree yang merupakan penyempurnaan dari algoritma ID3 dan C4.5. Dalam algoritma ini pemilihan atribut yang akan diproses menggunakan information gain, sehingga kesalahan yang ditimbulkan dalam pengambilan keputusan lebih diminimalkan [16].

Setelah melalui tahapan *preprocessing* data yang cocok untuk dipakai dalam pemodelan berjumlah 3.532 record data yang kemudian akan dilakukan proses pengklasifikasian dengan menggunakan algoritma C5.0, namun untuk yang akan dilakukan secara manual, diasumsikan terdapat 20 record data dengan atribut yang sama. berikut adalah tahapan tahapan yang dilakukan.

Tahapan pertama dalam algoritma C5.0 yaitu menghitung *Entropy* dengan persamaan (1).

$$Entropy(total) = \left( \left( -\frac{6}{20} \times \frac{6}{20} \right) + \left( -\frac{14}{20} \times \frac{14}{20} \right) \right) = 0.881 \tag{1}$$

Setelah dilakukan perhitungan *entropy* dilakukan perhitungan *gain* dengan persamaan (2).

$$\begin{aligned}
 \text{Gain}(\text{jenis kelamin}) &= \text{Entropy}(\text{total}) \\
 &- \left( \left( \frac{9}{20} \right) \times \text{Entropy}(\text{laki laki}) \right) \\
 &+ \left( \left( \frac{11}{20} \right) \times \text{Entropy}(\text{perempuan}) \right) \\
 &= 0.05908553 \quad (2)
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain Ratio}(\text{jenis Kelamin}) &= \frac{\text{Gain}(\text{jenis kelamin})}{\text{Entropy}(\text{laki laki}) + \text{Entropy}(\text{perempuan})} \\
 &= \frac{0.05908553}{0.99107606 + 0.684038436} \\
 &= 0.035272534 \quad (3)
 \end{aligned}$$

Untuk hasil perhitungan lengkap yang telah dilakukan terhadap semua atribut dapat dilihat pada tabel 2.

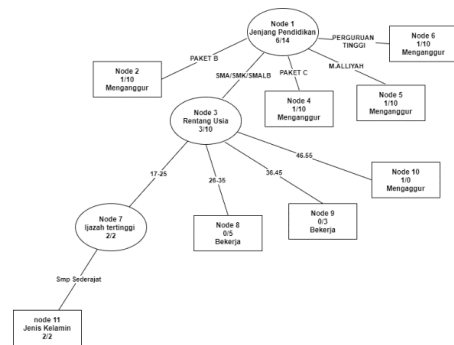
Setelah didapatkan nilai *gain* dilakukan perhitungan terhadap *gain ratio* dengan perhitungan (3).

Tabel 2 tabel perhitungan lengkap

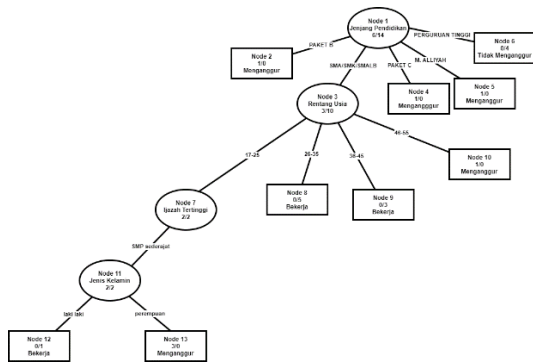
	jumlah kasus	Menganggur	Bekerja	entropy	Gain	Gain ratio
Total	20	6	14	0.8813		
Jenis kelamin	Laki laki	9	4	0.9911	0.0591	0.0352
	perempuan	11	2	0.6840		
Rentang usia	17-25	5	3	0.9709	0.2041	0.0836
	26-35	10	1	0.4690		
	46-55	4	2	1		
	56-65	1	0	0		

### 2.4. Pruning

Ketika decision tree dibuat banyak terjadi anomali dalam pembagian cabang dalam data latih karena noise atau outlier.[23] Fase *pruning* bertujuan untuk mengontrol kompleksitas *decision tree* dan menggeneralisasi kan *decision tree*. Fase *pruning* bertujuan untuk menghapus leaf yang tidak diperlukan dan menghindari *overfitting* [24]. Ada berbagai macam jenis pruning, pruning yang digunakan pada penelitian ini adalah *post-pruning*. Dalam *post-pruning* pemangkasan tree dilakukan setelah model tree dibuat,



Gambar SEQ Gambar \\* ARABIC 3 Visualisasi tree setelah pruning



Gambar SEQ Gambar \\* ARABIC 2 Visualisasi tree sebelum pruning

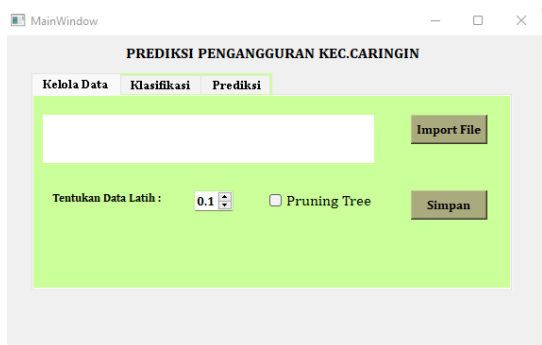
pemangkasan tree biasanya menghilangkan node atau branch yang tidak signifikan. Contoh visualisasi tree dapat dilihat pada gambar 2 dan gambar 3.

Tahapan penelitian yang dijelaskan diterapkan ke beberapa skenario pembagian data latih dan data uji yaitu dengan rasio pembagian 60:40, 70:30, 80:20 dan 90:10.

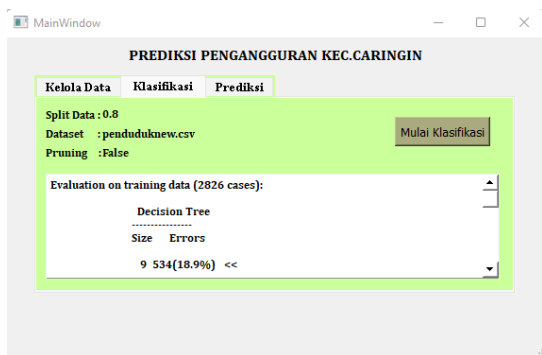
### 3. HASIL DAN PEMBAHASAN

Skenario penelitian yang dibahas dilakukan terhadap beberapa skenario pembagian data latih dan data uji dengan rasio pembagian 60:40, 70:30, 80:20, 90:10. Evaluasi akurasi dilakukan menggunakan *Confusion matrix*.

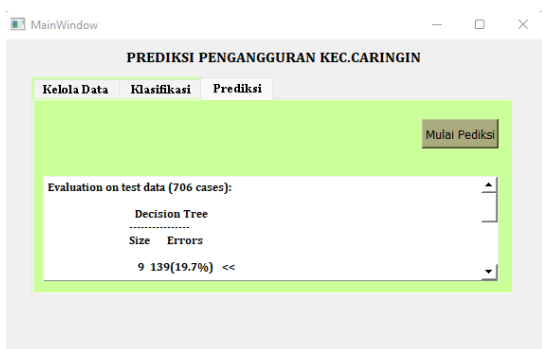
Dalam penelitian ini proses perhitungan dilakukan menggunakan perangkat lunak berbasis desktop menggunakan bahasa pemrograman *python*. Hasil implementasi perangkat lunak dapat dilihat pada gambar 3, gambar 4 dan gambar 5.



Gambar 4 Tampilan antarmuka kelola dataset



Gambar 5 Antarmuka kelola klasifikasi



Gambar 6 Antarmuka kelola prediksi

Setelah dilakukan evaluasi untuk berbagai skenario yaitu pembagian data latih data uji dan penggunaan pruning didapatkan hasil perbandingan akurasi seperti bisa dilihat pada tabel 3.

Tabel 3 Hasil evaluasi akurasi

Pembagian split data	Dengan Pruning (%)	Tanpa Pruning (%)
60% / 40%	80.5	80.5
70% / 30%	81.5	80.1
80% / 20%	81.3	81.1
90% / 10%	80.7	80.9

Berdasarkan tabel 3 diantara semua evaluasi skenario yang dilakukan, skenario yang memiliki tingkat akurasi paling tinggi adalah pembagian split data 70:30 dengan tingkat akurasi sebesar 81,5%, dan skenario dengan akurasi paling rendah adalah skenario dengan pembagian split data 60:40 dengan penggunaan pruning dan tanpa pruning memiliki hasil yang sama yaitu sebesar 80,5%.

Untuk evaluasi terhadap data uji beberapa skenario mendapatkan hasil akurasi lebih tinggi jika digunakan pruning terhadap model tree yaitu pada skenario 70:30 dan 80:20, namun ada juga yang tingkat akurasinya menurun ketika dilakukan pruning yaitu skenario pembagian split data 90:10, hal ini bisa disebabkan karena pembagian data latih dan data uji yang tidak seimbang atau karena kurangnya rasio data uji yang dievaluasi membuat data uji yang kurang representatif, atau karena model terlalu baik dalam mempelajari data latih sehingga mempengaruhi performa model ketika dihadapkan dengan data uji. Ada pula skenario yang tingkat akurasinya tidak berubah setelah dilakukan pruning yaitu skenario pembagian split data 60:40 yang merupakan hasil akurasi yang paling rendah. Skenario pembagian split data ini memiliki akurasi paling rendah bisa disebabkan karena training data tidak mewakili keseluruhan data yang digunakan atau biasa disebut dengan *underfitting*.

#### 4. KESIMPULAN

Dalam penelitian Prediksi Pengangguran Menggunakan Decision Tree Dengan Algoritma C5.0 Pada Data Penduduk Kecamatan Caringin Kabupaten Bogor, menghasilkan sebuah model prediksi berdasarkan data penduduk dengan atribut yang telah didefinisikan berdasarkan penelitian penelitian yang membahas tentang faktor faktor yang mempengaruhi tingkat pengangguran.

Hasil klasifikasi menggunakan algoritma C5.0 mendapatkan tingkat akurasi yang bervariasi berdasarkan rasio split data yang dan penggunaan pruning. Dengan hasil split data 70:30 yang menggunakan pruning memiliki tingkat akurasi paling tinggi dengan 81,5% dan hasil akurasi paling rendah yaitu 60:40 dengan 80,5%. Jika melihat pada penelitian sebelumnya yang menggunakan algoritma C5.0 untuk mengklasifikasikan jenis data yang sama yaitu yang bertipe nominal dengan hasil akurasi tertinggi di 80,5%, hasil akurasi di penelitian ini memiliki tingkat akurasi yang lebih baik. Hal ini mengindikasikan bahwa pembagian split data yang seimbang dapat mempengaruhi tingkat akurasi model, pembagian data latih dan data uji yang terlalu sedikit atau terlalu banyak dapat menyebabkan terjadinya *underfitting* dan *overfitting* terhadap model yang dibuat.

Dalam penelitian ini pemilihan atribut diambil berdasarkan faktor faktor yang mempengaruhi tingkat pengangguran. Atribut yang dapat digunakan pada dataset untuk penelitian ini adalah atribut yang memiliki kaitan dengan pendidikan, status kawin, usia dan jenis kelamin. namun ada faktor faktor lain yang dapat mempengaruhi tingkat pengangguran seperti faktor ekonomi. Dan juga pada penelitian ini jenis pruning yang digunakan dibatasi dengan menggunakan *post-pruning* Oleh karena itu untuk penelitian lebih lanjut perlu ditambahkan atribut atribut yang menjadi faktor lain yang mempengaruhi tingkat pengangguran dan menerapkan jenis pruning lain.

## DAFTAR PUSTAKA

- [1] S. Hermuningsih, A. D. Rahmawati, and M. Mujino, "Faktor-Faktor Yang Mempengaruhi Return Saham," *J. Ekon. dan Bisnis*, vol. 19, no. 3, p. 78, 2018, doi: 10.30659/ekobis.19.3.78-89.
- [2] B. Zhao and B. Xue, "Genomics Improving prediction accuracy using decision-tree-based meta-strategy and multi-threshold sequential-voting exemplified by miRNA target prediction," *Genomics*, 2017, doi: 10.1016/j.ygeno.2017.04.003.
- [3] S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014.
- [4] D. M. Informatika, F. I. Terapan, U. Telkom, J. Telekomunikasi, and B. Batu, "PREDIKSI NILAI PROYEK AKHIR MAHASISWA MENGGUNAKAN," no. November, pp. 2-3, 2015.
- [5] L. Vanfretti and V. S. N. Arava, "Electrical Power and Energy Systems Decision tree-based classification of multiple operating conditions for power system voltage stability assessment," *Electr. Power Energy Syst.*, vol. 123, no. April, p. 106251, 2020, doi: 10.1016/j.ijepes.2020.106251.
- [6] C. Hutabarat, "Penerapan Data Mining Untuk Memprediksi Permintaan Produk Kartu Perdana Internet Menggunakan Algoritma C5.0 (Studi Kasus: Vidha Ponsel)," *J. Pelita Inform.*, vol. 6, no. 4, pp. 419-424, 2018.
- [7] A. Kantono, I. Y. Purbasari, and F. T. Anggraeny, "Penerapan pruning pada algoritma c5.0 untuk mendiagnosis penyakit diabetes melitus 1," no. September, pp. 184-189, 2019.
- [8] M. Santoso and P. Musa, "REKOMENDASI KESEHATAN JANIN DENGAN PENERAPAN ALGORITMA C5.0 MENGGUNAKAN CLASSIFYING CARDIOTOGRAPHY DATASET," vol. 9, no. 2, pp. 65-76, 2021.
- [9] S. Nurnaningsih, Y. Susanti, and S. S. Handajani, "Implementasi Algoritma C5.0 Untuk Klasifikasi Penyakit Gagal Ginjal Kronik," *INTEK J. Inform. dan Teknol. Inf.*, vol. 2, no. 1, pp. 26-31, 2019, doi: 10.37729/intek.v2i1.89.
- [10] M. Riadi, Y. Azhar, and G. W. Wicaksono, "Implementasi Algoritma C5.0 Dan K-Medoids Untuk Klasterisasi Ibu Hamil Beresiko Tinggi," *J. Repos.*, vol. 2, no. 4, p. 511, 2020, doi: 10.22219/repositor.v2i4.696.
- [11] U. S. Aayi, T. W. Diwangkara, and R. T. Kurniawan, "Diagnosa Penyakit Disk Hernia Dan Spondylolisthesis Menggunakan Algoritma C5," *Telematika*, vol. 16, no. 2, p. 81, 2020, doi: 10.31315/telematika.v16i2.3181.
- [12] V. Rahmayanti, Y. Azhar, and A. E. Pramudita, "Penerapan algoritma C5.0 pada analisis faktor-faktor pengaruh kelulusan tepat waktu mahasiswa Teknik Informatika UMM," *J. Repos.*, vol. 1, no. 2, p. 131, 2020, doi: 10.22219/repositor.v1i2.545.
- [13] M. I. A. Vachroni and A. Hermawan, "Prediksi kelulusan mahasiswa menggunakan metode backpropagation," *Univ. Teknol. Yogyakarta*, vol. 4, pp. 230-238, 2021.
- [14] A. Nusari, I. Purbasari, and E. Puspaningrum, "PREDIKSI KETEPATAN WAKTU KELULUSAN MAHASISWA DENGAN METODE ALGORITMA C5.0," vol. 1, no. 3, pp. 817-825, 2020.
- [15] F. N. Umma, B. Warsito, and D. A. I. Maruddani, "Klasifikasi Status Kemiskinan Rumah Tangga Dengan Algoritma C5.0 Di Kabupaten Pematang," *J. Gaussian*, vol. 10, no. 2, pp. 221-229, 2021, doi: 10.14710/j.gauss.v10i2.29934.
- [16] D. Wintana, H. Hikmatullah, N. Ichsan, J. J. Purnama, and A. Rahmawati, "KLASIFIKASI PENENTUAN PENERIMA MANFAAT PROGRAM KELUARGA HARAPAN (PKH) MENGGUNAKAN ALGORITMA C5.0 (Studi kasus: Desa Sukamaju, Kec.Kadudampit)," *Klik - Kumpul. J. Ilmu Komputer.*, vol. 6, no. 3, p. 254, 2019, doi: 10.20527/klik.v6i3.206.
- [17] A. C. Wijaya, N. A. Hasibuan, and P. Ramadhani, "Implementasi Algoritma C5.0 Dalam Klasifikasi Pendapatan Masyarakat ( Studi Kasus: Kelurahan Mesjid Kecamatan Medan Kota)," *Inf. dan Teknol. Ilm.*, vol. 13, pp. 192-198, 2018.
- [18] S. Wahyuningsih and D. R. Utari, "Perbandingan Metode K-Nearest Neighbor, Naive Bayes dan Decision Tree untuk Prediksi Kelayakan Pemberian Kredit," *Konf. Nas. Sist. Inf. 2018 STMIK Atma Luhur Pangkalpinang, 8-9 Maret 2018*, pp. 619-623, 2018.
- [19] D. Sartika and D. Indra, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian," *J. Tek. Inform. Dan Sist. Inf.*, vol. 1, no. 2, pp. 151-161, 2017.
- [20] I. C. Wibowo, A. C. Fauzan, M. D. P. Yustiana, and FAQ habib, "Komparasi Algoritma Naive Bayes dan Decision Tree Untuk Memprediksi Lama Studi Mahasiswa," *Ilk. J. Comput. Sci. Appl. Informatics*, vol. 1, no. 2, pp. 65-74, 2019, doi: 10.28926/ilkomnika.v1i2.21.
- [21] A. Wardhana, B. Kharisma, and Y. F. Ibrahim, "Pengangguran Usia Muda Di Jawa Barat (Menggunakan Data Sakernas)," *E-Jurnal Ekon. dan Bisnis Univ. Udayana*, vol. 9, p. 1049, 2019, doi: 10.24843/eeb.2019.v08.i09.p04.
- [22] S. Prawira, "Pengaruh Pertumbuhan Ekonomi, Upah Minimum Provinsi, Dan Tingkat Pendidikan Terhadap Pengangguran Terbuka Di Indonesia," *J. Ecogen*, vol. 1, no. 4, p. 162, 2018, doi: 10.24036/jmpe.v1i1.4735.
- [23] P. Kapoor, R. Rani, and R. JMIT, "Efficient

Decision Tree Algorithm Using J48 and Reduced Error Pruning,” *Int. J. Eng. Res. Gen. Sci*, vol. 3, no. 3, pp. 1613–1621, 2015.

- [24] J. Ye *et al.*, “A Chi-MIC Based Adaptive Multi-Branch Decision Tree,” *IEEE Access*, vol. 9, pp. 78962–78972, 2021, doi: 10.1109/ACCESS.2021.3077125.