

KLASIFIKASI PENENTUAN STATUS KEMISKINAN PENDUDUK KELURAHAN KARANGPAWITAN KARAWANG MENGGUNAKAN METODE C4.5

Riliandhita, Iqbal Maulana, Purwanto

Teknik Informatika, Universitas Singaperbangsa Karawang
Jl. HS. Ronggowaluyo, Telukjambe Timur, Karawang - 41363
1910631170227@student.unsika.ac.id

ABSTRAK

Kemiskinan merujuk pada kondisi individu yang tidak mampu memenuhi kebutuhan dasar yang diperlukan untuk hidup layak, termasuk kebutuhan pangan dan non-pangan. Pemerintah telah melakukan upaya-upaya untuk mengatasi masalah kemiskinan melalui program-program bantuan sosial, namun sering kali menghadapi kendala-kendala tertentu. Salah satu permasalahan yang muncul adalah ketidakmerataan distribusi bantuan sosial dan kesulitan dalam memastikan bahwa bantuan tersebut tepat sasaran. Penelitian ini bertujuan untuk meningkatkan efektivitas dan relevansi program-program pengentasan kemiskinan penduduk Kelurahan Karangpawitan. Data terkait status kemiskinan di Kelurahan Karangpawitan dikumpulkan dan dianalisis guna menilai tingkat kemiskinan masyarakat. Metode klasifikasi dengan algoritma C4.5 dalam metodologi Knowledge Discovery in Databases (KDD) digunakan dalam penelitian ini. Data kemiskinan disubjekkan pada proses seleksi, preprocessing, transformasi, dan penggalian data, kemudian dianalisis untuk memahami faktor-faktor yang paling berpengaruh. Algoritma C4.5 digunakan untuk menghasilkan aturan yang dapat mengklasifikasikan status kemiskinan penduduk. Evaluasi menunjukkan bahwa algoritma C4.5 mampu mengklasifikasikan status kemiskinan dengan tingkat akurasi sebesar 99,02%. Koefisien kappa sebesar 0,980, mengindikasikan klasifikasi yang kuat dan dapat diandalkan. Hasil penelitian ini memberikan wawasan yang lebih baik mengenai status kemiskinan di Kelurahan Karangpawitan, yang dapat menjadi dasar dalam merumuskan kebijakan yang lebih efektif dalam penanggulangan kemiskinan.

Kata kunci : Status kemiskinan, Data Mining, C4.5

1. PENDAHULUAN

Indonesia adalah bagian dari negara berkembang di Asia, terutama Asia Tenggara yang menghadapi tantangan kemiskinan. Kemiskinan merupakan bagian permasalahan yang seringkali didapati oleh negara-negara berkembang. Menurut pemerintah Indonesia, permasalahan kemiskinan telah menjadi isu yang lama dan sukar untuk diatasi[1]. Berdasarkan data Badan Pusat Statistik (BPS), angka penduduk miskin di Indonesia pada September 2022 sebesar 26,36 juta orang atau sekitar 9,57% dari total populasi, jumlah ini mengalami peningkatan sebesar 0,20 juta orang (0,03%) dibandingkan dengan data pada Maret 2022 dan berkurang 0,14 juta orang (0,14%) terhadap September 2021.

Menurut data BPS Provinsi Jawa Barat, Kepala Badan Pusat Statistik Jawa Barat Marsudijono, memaparkan secara langsung pada tanggal 15 juli 2022 bahwa total penduduk miskin pada Maret 2022 mencapai 4,0juta orang. Angka ini mengalami peningkatan sebanyak 66,1 ribu orang dibandingkan dengan data September 2021,namun mengalami penurunan sebesar 124,4 ribu orang jika dibandingkan dengan data Maret 2021. Kecamatan Karawang Barat yang terdapat di Kabupaten Karawang, Jawa Barat terbagi atas delapan kelurahan salah satunya yaitu kelurahan karangpawitan yang hingga saat ini terus berupaya mengurangi tingkat kemiskinan. Data BPS menyebutkan bahwa Kabupaten Karawang memiliki jumlah penduduk

miskin yang setiap tahunnya mengalami peningkatan. Pada tahun 2019, jumlah penduduk miskin sekitar 173,66 ribu orang. Pada tahun 2020, jumlahnya meningkat menjadi 195,41 ribu orang, dan pada tahun 2021, jumlah penduduk miskin naik lagi menjadi 210,78 ribu orang. Pemerintah telah melakukan berbagai upaya dalam mengatasi kemiskinan, termasuk melalui program bantuan sosial, tetapi seringkali menghadapi kesulitan. Salah satu masalah yang timbul adalah ketidakmerataan dalam pembagian bantuan sosial dan kesulitan untuk memastikan bantuan tepat sasaran.

Pada penelitian sebelumnya terkait dengan data mining pada tingkat kemiskinan di Indonesia diantaranya. Penelitian yang dilakukan oleh Carissa Elma Purnomo & Rikendry yang berjudul Penerapan Metode C4.5 Untuk Klasifikasi Warga Miskin Pada Desa Mengandung Sari Berdasarkan penelitian tersebut didapatkan bahwa dengan algoritma C4.5, data warga desa mengandung sari dapat di kategorikan berdasarkan “miskin” dan “tidak miskin” dan hasil ketepatan pengkategorian data warga miskin menggunakan metode C4.5 dikategorikan “unggul” dengan hasil rata-rata ketepatan sebesar 91,6%[1]. Kemudian, penelitian yang dilakukan oleh Ni Wayan Oktha Pratiwi, dkk, penelitian dengan judul Klasifikasi Penentuan Penerima Bantuan Sosial Tunai (Bst) Menggunakan Algoritma C4.5 Di Desa Keramas, Gianyar Bali. Penelitian ini bertujuan untuk meningkatkan efektivitas dan relevansi program pengentasan kemiskinan. Peneliti melakukan

pengujian menggunakan lima atribut, yaitu riwayat menerima bantuan lain, pekerjaan, tingkat pendidikan, kepemilikan kartu miskin, dan jumlah tanggungan. pengujian ini menggunakan data sebanyak 2.074 sampel yang dipecah menjadi dua jenis, seperti data pelatihan (training) berjumlah 80% serta data pengujian (testing) berjumlah 20%. Hasil pengujian menyatakan tingkat ketepatan C4.5 sebesar 97,83%, dibandingkan dengan algoritma K-Nearest Neighbor yang mencapai akurasi sebesar 92,29%, dan Naïve Bayes yang mencapai akurasi sebesar 91,81%[2].

Penelitian ini bertujuan untuk meningkatkan efektivitas dan relevansi program-program pengentasan kemiskinan. Dari pemaparan yang dibuktikan melalui hasil penelitian sebelumnya, algoritma C4.5 terbukti memiliki kinerja yang baik dalam klasifikasi data, yang menghasilkan tingkat ketepatan yang akurat. Oleh karena itu, pada penelitian saat ini, algoritma C4.5 akan digunakan untuk mengklasifikasikan status kemiskinan penduduk di Kelurahan Karangpawitan, Karawang. Tahapan-tahapan KDD (*Knowledge Discovery in Database*) akan diterapkan pada penelitian ini. Metode evaluasi yang digunakan ialah metode *Confusion Matrix* dan nilai Kappa dengan menggunakan Software RapidMiner.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Pattern recognition dikenal sebagai istilah dari *data mining* yang berupa algoritma yang diterapkan untuk mengolah data agar menjumpai pola tersembunyi dari suatu data yang akan dikerjakan. Pada metode *data mining*, data yang diolah akan mengeluarkan suatu pengetahuan baru yang berasal dari data sebelumnya, hasil dari pengerjaan data tersebut dapat dipergunakan dalam pengambilan keputusan di masa mendatang[3]. Tiga jenis data yang digunakan pada *data mining*, yaitu jenis data numerik, jenis data klasifikasi, dan jenis data rentang waktu. Pada tipe data numerik terbagi mejadi dua yaitu rasio dan interval. Tipe data klasifikasi juga terbagi menjadi dua yaitu ordinal dan nominal[4]

2.2. Klasifikasi

Klasifikasi adalah proses pengelompokan data berdasarkan aturan tertentu yang diperoleh dari pembelajaran data latih. Klasifikasi data yang akan diuji dapat dimanfaatkan untuk mengkategorikan data uji. Proses klasifikasi ini termasuk dalam pembelajaran terawasi, di mana komputer menggunakan atribut sebagai penanda karakteristik data, yang sering disebut sebagai label atau kelas.

Tujuan klasifikasi pada data mining yaitu mengidentifikasi karakteristik kelompok untuk sampel data pada setiap kasus yang berbeda. Proses tersebut digunakan agar memahami data yang ada untuk memprediksi kasus baru Class atau kelompok pada setiap sampel data ditentukan oleh klasifikasi,

klasifikasi memprediksi nilai yang berupa label. Atribut dari sampel data disebut input dari model klasifikasi sedangkan class pada sampel data itu sendiri disebut outputnya[5].

2.3. Algoritma C4.5

Algoritma C4.5 merupakan versi perbaikan dari ID3, yang dikenal sebagai algoritma pembangunan pohon keputusan. Berbeda dengan ID3 yang hanya mendukung fitur tipe kategoris, C4.5 mampu menangani fitur numerik dengan mengubahnya menjadi fitur biner melalui pembagian data berdasarkan ambang batas. Sehingga, algoritma C4.5 memungkinkan penggunaan fitur numerik dalam pembangunan pohon keputusan[6].

Langkah-langkah proses dasar algoritma C4.5 dalam pembangunan pohon keputusan secara umum adalah sebagai berikut:

- a) Memilih atribut sebagai akar.
- b) Membuat cabang setiap nilai.
- c) Membagi kasus dalam cabang.
- d) Mengulangi proses setiap cabang hingga semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, diperlukan nilai gain terbesar dari atribut-atribut yang ada. Digunakan untuk menghitung gain, sebagai dalam persamaan berikut[7].

$$\text{Gain} = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \quad (1)$$

Keterangan:

- S : Himpunan pada kasus
- A : Atribut
- n : Jumlah kasus pada atribut A
- |S_i| : Jumlah kasus pada partisi ke i
- |S| : Jumlah kasus dalam S

Setelah Selepas memperoleh nilai gain, langkah berikutnya ialah menghitung *Entropy* untuk menemukan tingkat informasi dari atribut input dalam menghasilkan atribut output. Untuk memperoleh nilai *Entropy* dapat menggunakan persamaan :

$$\text{Entropy}(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (2)$$

Keterangan:

- S : Himpunan pada kasus
- n : Jumlah pada partisi S
- p_i : Proporsi dari S_i pada S

2.4. Knowledge Discovery in Databases (KDD)

Knowledge Discovery in Databases adalah proses mencari pengetahuan yang tersembunyi dan bermanfaat dari data yang ada dalam basis data KDD merupakan bidang multidisiplin yang menggabungkan teknik-teknik dari berbagai domain, seperti statistik, machine learning, manajemen basis data, dan visualisasi data. Tujuan dari KDD adalah untuk menghasilkan wawasan yang bermakna dan dapat diaplikasikan dari dataset yang besar dan kompleks. Proses KDD dimulai dengan pemahaman terhadap domain dan tujuan analisis, kemudian dilanjutkan dengan pemilihan dan preprocessing data,

penerapan algoritma data mining, dan interpretasi serta evaluasi pola yang ditemukan. Secara keseluruhan, KDD merupakan pendekatan yang kuat untuk menggali pengetahuan dari dataset yang besar dan telah menjadi alat yang penting untuk pengambilan keputusan dan pemecahan masalah dalam berbagai industri[8].

2.5. Evaluasi Model

Evaluasi model berfungsi membantu mengevaluasi kinerja model yang digunakan dalam penelitian. Dalam evaluasi kinerja model klasifikasi menggambarkan seberapa efektif sistem dalam mengklasifikasikan data dengan baik. Seberapa efektif sistem dalam mengklasifikasikan data menggambarkan Kinerja sistem klasifikasi. Evaluasi ini bertujuan untuk menguji model klasifikasi dalam data mining guna memahami seberapa baik kinerja sistem tersebut[9]. Metode untuk memeriksa kinerja kategori *data mining* yaitu :

2.5.1. Confusion Matrix

Confusion Matrix adalah teknik yang umumnya dipakai untuk menghitung akurasi pada konsep *data mining*. Confusion Matrix merupakan instrumen yang sangat berguna untuk menelaah sejauh mana pengkategorian dapat mengenali tuple dari kelas yang berlawanan dengan baik. Confusion Matrix ialah sebuah tabel yang terdiri dari dua tingkat, yaitu tingkat yang dianggap sebagai positif dan tingkat yang dianggap sebagai negatif[10]. Berikut ini ialah *confusion matrix* yang dipaparkan pada Tabel 1:

Table 1. Confusion Matrix

Confusion Matrix		Nilai Prediksi	
		Positif	Negatif
Nilai Sebenarnya	Positif	a (TP)	b (FP)
	Negatif	c (FN)	d (TN)

Berdasarkan tabel *Confusion Matrix* diatas:

- a. (TP) berarti hasil klasifikasi positif diprediksi dengan benar oleh pengklasifikasi.
- b. (FP) berarti data sebenarnya negatif namun diprediksi positif oleh pengklasifikasi.
- c. (FN) berarti data sebenarnya positif namun diprediksi negatif oleh pengklasifikasi.
- d. (TN) berarti hasil klasifikasi negatif diprediksi dengan benar oleh pengklasifikasi.

2.5.2. Kappa

Kappa adalah sebuah metode analisis statistik yang berfokus pada kelompok-kelompok yang berbeda. Semakin tinggi nilai kappa, semakin baik kinerja yang dihasilkan. Perhitungan koefisien Kappa ditunjukkan dalam persamaan berikut ini[11]:

$$K = \frac{P_o - P_e}{1 - P_e} \tag{3}$$

Keterangan:

- K : Koefisien Kappa
- P_o : Presentase kesepakatan yang diamati
- P_e : Jumlah baris dalam error matrix

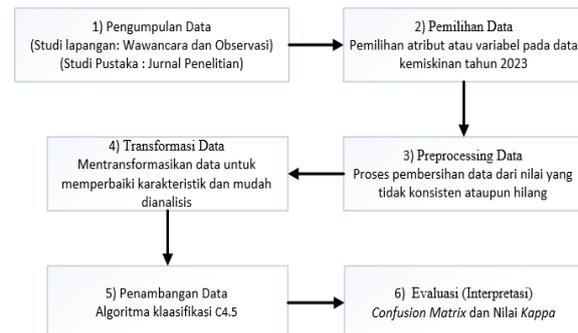
Berikut ini Klasifikasi nilai kappa dapat dilihat pada Tabel 2 :

Table 2. Klasifikasi Nilai Kappa

Nilai K	Keeratan Kesepakatan
< 0,20	Rendah
0,21 - 0,40	Lumayan
0,41 - 0,60	Cukup
0,61 - 0,80	Kuat
0,81 - 1,00	Sangat Kuat

3. METODE PENELITIAN

Metodologi penelitian ini yaitu menggunakan metode Knowledge Discovery in Database. KDD merupakan kumpulan tahapan dalam proses identifikasi data menjadi kumpulan informasi yang mengandung pola. Informasi ini terdapat dalam basis data berukuran besar yang sebelumnya belum diketahui manfaat potensial dari data tersebut. Proses yang dilakukan yaitu melalui tahapan data selection, data preprocessing, transformation, data mining, daninterpretation/evaluation.



Gambar 1. Rancangan Penelitian

4. HASIL DAN PEMBAHASAN

4.1. Hasil

Pada tahap pertama yaitu pengumpulan data, data yang diperoleh melalui survei lapangan dan wawancara penduduk Kelurahan Karangpawitan Karawang. Pada penelitian ini, jumlah dataset yang digunakan sebanyak 383, yang terbagi menjadi dua bagian, pertama data training dimana data yang sudah diketahui label – labelnya untuk membangun model atau fungsi sebesar 80% sebanyak 307 data, dan sisanya data testing yang digunakan untuk mengetahui akurasi model atau fungsi yang akan dibangun pada proses training sebesar 20% yaitu sebanyak 76 data. Berikut adalah sebagian dari dataset yang ditampilkan.

Table 3. Dataset

Luas Bangunan (m2)	Jenis Lantai	Jenis Dinding	Frekuensi mengonsumsi daging/minggu	Jenis Pekerjaan	Besar Penghasilan (per bulan)	Pendidikan	Ketersediaan WC	Aset Milik	Label
Ya	Keramik	Tembok	0-1x	Buruh Tani	50000	SD	Tidak Punya	<50000	Miskin
Tidak	Plester	Tembok	0-1x	Buruh Tani	50000	SD	Punya	>=50000	Tidak Miskin
Ya	Plester	Bambu	0-1x	Serabutan	50000	SD	Tidak Punya	<50000	Miskin
Tidak	Keramik	Tembok	>1x	Karyawan Swasta	450000	SMA	Punya	>=50000	Tidak Miskin
Tidak	Keramik	Tembok	>1x	Karyawan Swasta	450000	SMA	Punya	>=50000	Tidak Miskin
Tidak	Keramik	Tembok	0-1x	Pedagang	50000	SD	Tidak Punya	<50000	Miskin
Ya	Plester	Bambu	0-1x	Serabutan	50000	SD	Punya	<50000	Miskin
Ya	Keramik	Tembok	0-1x	Buruh Tani	50000	SD	Tidak Punya	>=50000	Miskin
Ya	Plester	Bambu	0-1x	Buruh Tani	50000	SD	Punya	<50000	Miskin
Ya	Plester	Bambu	0-1x	Serabutan	50000	SMP	Punya	<50000	Miskin
Ya	Keramik	Tembok	0-1x	Serabutan	50000	SMP	Tidak Punya	>=50000	Tidak Miskin
Ya	Keramik	Tembok	0-1x	Buruh Tani	50000	SD	Tidak Punya	<50000	Miskin
Ya	Keramik	Tembok	>1x	Karyawan Swasta	530000	SMA	Punya	>=50000	Tidak Miskin
Ya	Keramik	Tembok	0-1x	Buruh Tani	50000	SD	Tidak Sekolah	<50000	Miskin
Ya	Plester	Bambu	0-1x	Buruh Tani	50000	SD	Punya	<50000	Miskin
Tidak	Keramik	Tembok	>1x	Karyawan Swasta	400000	SMA	Punya	>=50000	Tidak Miskin
Ya	Keramik	Tembok	0-1x	Serabutan	40000	SD	Punya	<50000	Tidak Miskin
Tidak	Keramik	Tembok	0-1x	Serabutan	55000	SMA	Punya	>=50000	Tidak Miskin

Pada tahap kedua yaitu pemilihan data. Pada tahap data selection dilakukan proses penentuan variabel data yang digunakan dalam pembentukan decision tree. Data dipilih menjadi variabel keputusan atau variabel target dari pembentukan decision tree. Dan variabel target dari penelitian ini adalah atribut "Label" yang berisikan label "Miskin" dan "tidak Miskin".

Tahap ketiga adalah tahap Preprocessing data pada tahap ini data dilakukan pembersihan data untuk menghapus data yang tidak valid, duplikat, atau tidak relevan. Berikut pengecekan

Pada tahap ke empat yaitu data tranformation. Pada tahap transformasi untuk merubah bentuk data agar sesuai dengan proses data mining supaya dapat memahami bentuk pola dan hubungan data dalam interpretabilitas model. Berikut atribut yang akan di transformasi Frekuensi mengonsumsi daging/minggu. Dapat terlihat pada Tabel 3

Name	Type	Missing
Label	Binominal	0
Luas Bangunan (m2)	Binominal	0
Jenis Lantai	Binominal	0
Jenis Dinding	Binominal	0
Frekuensi mengonsumsi daging...	Binominal	0
Jenis Pekerjaan	Polynomial	0
Besar Penghasilan (per bulan)	Integer	0
Pendidikan	Polynomial	0
Ketersediaan WC	Binominal	0
Aset Milik	Polynomial	0

Gambar 2. Pemeriksaan missing value

Table 4. Hasil Transformasi Data

Atribut	Nilai	Keterangan
Frekuensi mengonsumsi daging/minggu	0-1x	1
	>1x	2

Pada tahap selanjutnya yaitu penambangan data, tahap penambangan data dilakukan dengan memilih metode algoritma untuk menemukan pola atau informasi saat mengklasifikasikan tingkat kepuasan. Pada penelitian ini, data yang telah dikumpulkan dibagi sebesar 80% untuk keperluan data latih dan 20% untuk data uji. Berikut hasil perhitungan entropy dan gain untuk data training adalah sebagai berikut:

Table 5. Hasil Perhitungan Entrophy dan Gain

ATRIBUT	JUMLAH KASUS	MISKIN	TIDAK MISKIN	ENTROPY	INFORMATION GAIN
TOTAL	307	137	170	0,991649	
Luas Bangunan (m2)	Ya	187	139	0.822	0.281
	Tidak	120	18	0.537	
Jenis Lantai	Keramik	173	16	0.113	0.540
	Plester	134	121	0.460	
Jenis Dinding	Tembok	92	88	0.258	0.372
	Bambu	215	49	0.774	
Frekuensi mengonsumsi daging/minggu	1	199	100	0.108	0.297
	2	108	2	0.999	
Jenis Pekerjaan	Karyawan Swasta	65	0	0.0000	0.2889
	Buruh Tani	150	106	0.873	
	Pedagang	17	6	0.937	
	Serabutan	75	25	0.9183	
Besarnya Penghasilan (per bulan)	300.000-499.999	138	14	0.474	0.286
	500.000-5.300.000	169	116	0.897	

NO	ATURAN IF	HASIL (THEN)
	575000	Miskin
10	Jenis Lantai = Keramik AND Ketersediaan WC = Tidak Punya AND Besar Penghasilan (per bulan) <= 575000 AND Pendidikan = SD	Miskin
11	Jenis Lantai = Keramik AND Ketersediaan WC = Tidak Punya AND Besar Penghasilan (per bulan) <= 575000 AND Pendidikan = SMP	Tidak Miskin

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dapat diperoleh kesimpulan Hasil evaluasi menunjukkan bahwa algoritma C4.5 berhasil mengklasifikasikan status kemiskinan dengan melihat accuracy, recall, precision dan kappa. Yaitu dengan tingkat akurasi sebesar 99,02%. Kemudian nilai recall yang didapat dengan hasil prediksi dari data yang ada, tingkat relevan benar miskin yang sesuai sebesar 100%. dan benar tidak miskin sebesar 98,24%. Lalu nilai precision yang didapat dengan hasil prediksi menggunakan rapidminer yaitu prediksi miskin sebesar 97.86% dan prediksi tidak miskin sebesar 100% dan Koefisien kappa yang dihasilkan sebesar 0,980. Dengan demikian hasil yang diperoleh mengindikasikan bahwa algoritma C4.5 berhasil dalam mengklasifikasi status kemiskinan masyarakat di desa Karangpawitan Karawang dengan sangat baik.

Adapun saran untuk pengembangan penelitian yaitu disarankan untuk mengikutsertakan responden dalam penelitian lanjutan dengan menyelidiki apakah mereka menerima bantuan dari pemerintah. Perolehan informasi mengenai sumber air minum, apakah berupa air galon isi ulang atau air mineral asli, dan bahan bakar untuk memasak, apakah menggunakan gas 3kg atau yang berukuran besar, juga disarankan untuk diperoleh. Keseluruhan informasi ini diharapkan dapat memberikan gambaran yang lebih komprehensif dan mendalam terkait dengan kondisi serta kebutuhan responden dalam kerangka penelitian ini. Kemudian diharapkan dapat dibuat versi website atau aplikasi agar lebih mudah dalam penggunaan dan aksesnya.

DAFTAR PUSTAKA

[1] Purnomo, C. E., & Rikendry, R. (2021). Penerapan Metode C4. 5 Untuk Klasifikasi Warga Miskin Pada Desa Mengandung Sari. *Jurnal Teknologi dan Sistem Informasi*, 2(3), 14-25.

[2] Pratiwi, N. W. O., Utami, N. W., & Putra, I. G. J. E. (2022). Klasifikasi Penentuan Penerima Bantuan Sosial Tunai (BST) Menggunakan Algoritma C4. 5 Di Desa Keramas, Gianyar, Bali. *Jurnal Informatika Teknologi dan Sains (Jinteks)*, 4(3), 101-107.

[3] Permatasari, I. A. (2021). Klasifikasi Penduduk Penerima Bantuan Covid-19 Di Kecamatan Kasomalang Menggunakan Metode *K-Nearest Neighbor*. (Skripsi Sarjana, Universitas Singaperbangsa).

[4] Sadewo, M. G., Windarto, A. P., & Wanto, A. (2018). Penerapan Algoritma *Clustering* Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/Mitigasi Bencana Alam Menurut Provinsi Dengan *K-Means*. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 2(1).

[5] Fitriani, E. (2020). Perbandingan Algoritma C4. 5 Dan *Naive Bayes* Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan. *Sistemasi*, 9 (1), 103.

[6] Gunawan, G., Hanes, H., & Catherine, C. (2021). C4. 5, *K-Nearest Neighbor*, *Naive Bayes*, and *Random Forest Algorithms Comparison to Predict Students' on TIME Graduation*. *Indonesian Journal of Artificial Intelligence and Data Mining*, 4(2), 62-71.

[7] Saragih, H., Buulolo, E., & Waruwu, F. T. (2017). Implementasi Data Mining Penyesuaian Jenis Lensa Terhadap Kebutuhan Pasien Dengan Menggunakan Algoritma C4. 5. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 1(1)

[8] W. Cholil, A. R. Dalimunthi, and L. Atika, "Model Data Mining Dalam Mengidentifikasi Pola Laju Pertumbuhan Antar Sektor Ekonomi di Provinsi Sumatera Selatan dan Bangka Belitung." *Teknika*, vol. 8, no. 2, pp. 103–109, 2019, doi: 10.34148/teknika.v8i2.181.

[9] Anam, C., & Santoso, H. B. (2018). Perbandingan Kinerja Algoritma C4. 5 dan *Naive Bayes* untuk Klasifikasi Penerima Beasiswa. *Energy-Jurnal Ilmiah Ilmu-Ilmu Teknik*, 8(1), 13-19.

[10] Kuntoro, A. Y., Hermanto, H., Asra, T., Syukmana, F., & Wahono, H. (2020). *Classification of Student Majors with C4. 5 and Naive Bayes Algorithms (Case Study: SMAN 2 Bekasi City)*. *Semesta Teknika*, 23(1), 1-13.

[11] Aprianti. B. (2022). klasifikasi Penerima Bantuan Langsung Tunai Dana Desa (Blt Dd) Pada Desa Warungbambu Menggunakan Algoritma *Naive Bayes*. (Skripsi Sarjana, Universitas Singaperbangsa).