Analisis Komparatif Algoritma C4.5, Naïve Bayes, dan K-Nearest Neighbor untuk Prediksi Kesejahteraan Masyarakat di Kabupaten Temanggung

Kartika Sari Sunarno[1], Deden Istiawan[2]✉
[1,2]Institut Teknologi Statistika dan Bisnis Muhammadiyah Semarang, Indonesia

✉Corresponding Author: deden.istiawan@itesa.ac.id

ABSTRAK

Kesejahteraan penduduk dapat meningkat seiring dengan menurunnya tingkat kemiskinan. Pada Maret 2019, tingkat kemiskinan di Kabupaten Temanggung mencapai 9,42 persen atau sekitar 72,6 ribu jiwa. Meskipun mengalami penurunan dibanding tahun sebelumnya, terdapat kekhawatiran bahwa penduduk rentan miskin dapat jatuh ke dalam kemiskinan apabila garis kemiskinan meningkat sebesar 10 persen, yang berpotensi menyebabkan lonjakan angka kemiskinan. Oleh karena itu, strategi penanggulangan kemiskinan memerlukan data yang akurat sesuai dengan karakteristik kesejahteraan masyarakat agar kebijakan yang diambil lebih tepat sasaran. Penelitian ini mengusulkan perbandingan beberapa algoritma klasifikasi, yaitu C4.5, Naïve Bayes, dan k-Nearest Neighbor, untuk menentukan algoritma dengan performa terbaik dalam menganalisis Data Kemiskinan Daerah (DKD) Kabupaten Temanggung. Fokus penelitian ini adalah pada prediksi kelas positif, yaitu keluarga sejahtera. Berdasarkan evaluasi menggunakan metrik AUC, precision, recall/sensitivity, dan F-measure pada kelas positif, hasil penelitian menunjukkan bahwa algoritma Naïve Bayes memiliki kinerja terbaik. Sebaliknya, algoritma C4.5 memiliki performa terendah karena memperoleh nilai precision, recall, dan F-measure yang paling kecil.

Kata kunci: algoritma klasifikasi, kemiskinan, perbandingan algoritma

A. Pendahuluan

Kemiskinan berdampak negatif terhadap berbagai aspek kehidupan, termasuk penurunan kualitas hidup, penghambatan pembentukan sumber daya manusia yang unggul, peningkatan beban sosial ekonomi, lonjakan tingkat kriminalitas, serta berkurangnya ketertiban umum [1]. Ketidakmampuan dalam membangun sumber daya manusia yang berkualitas dapat berkontribusi pada penurunan kesejahteraan masyarakat [2]. Todaro & Smith (2015) menegaskan bahwa kemiskinan yang meluas dan berjumlah besar merupakan inti dari berbagai permasalahan pembangunan. Hal ini diperkuat oleh Sukwika (2018) yang menyatakan bahwa tantangan utama pembangunan di Indonesia saat ini adalah mengatasi ketimpangan, yang tidak hanya terjadi pada tingkat individu atau rumah tangga, tetapi juga dalam skala wilayah.

Kabupaten Temanggung terletak di bagian tengah Provinsi Jawa Tengah dan memiliki persentase penduduk miskin sebesar 9,42 persen atau sekitar 72,6 ribu jiwa pada Maret 2019 [5]. Angka ini menunjukkan penurunan dibandingkan tahun 2018, yang mencapai 9,87 persen atau sekitar 75,4 ribu jiwa [5]. Berdasarkan hasil Susenas 2019, Setiadi (2020) melaporkan bahwa jumlah penduduk rentan miskin pada tahun tersebut mencapai 17,08 persen atau sekitar 131.565 jiwa, dengan garis kemiskinan sebesar Rp310.176 per kapita per bulan, meningkat dari Rp287.981 per kapita per bulan pada 2018. Kelompok ini berisiko jatuh ke dalam kemiskinan apabila garis kemiskinan meningkat atau daya beli mereka menurun. Setiadi (2020)

juga menggambarkan bahwa jika garis kemiskinan naik 10 persen, jumlah penduduk miskin akan bertambah menjadi 158.358 jiwa, meningkat sebesar 118 persen atau hampir dua kali lipat dibandingkan kondisi garis kemiskinan normal. Awalnya, persentase penduduk miskin sebesar 9,42 persen akan melonjak menjadi 20,55 persen jika garis kemiskinan mengalami kenaikan tersebut. Kondisi ini menjadi peringatan bagi Pemerintah Kabupaten Temanggung untuk segera mengambil langkah strategis dalam mengatasi permasalahan ini.

Berdasarkan Perpres Nomor 15 Tahun 2010, pemerintah Indonesia membentuk Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K) yang diketuai oleh Wakil Presiden RI. Tim ini bertugas menyusun kebijakan dan program yang bertujuan untuk menyinergikan berbagai kegiatan penanggulangan kemiskinan di tingkat kementerian atau lembaga, serta melakukan pengawasan dan pengendalian dalam implementasinya [7]. Salah satu aspek krusial dalam mendukung strategi penanggulangan kemiskinan adalah penyediaan data kemiskinan yang akurat dan tepat sasaran [8]. Untuk memastikan efektivitas program tersebut, diperlukan perencanaan strategis yang komprehensif, terutama dalam hal pemetaan wilayah serta karakteristik kemiskinan [9]. Oleh karena itu, klasifikasi prediktif dalam menentukan rumah tangga miskin menjadi langkah penting bagi pemerintah dalam mengidentifikasi serta merancang kebijakan yang lebih spesifik, dengan melibatkan konsultasi bersama pemangku kepentingan terkait [10].

Klasifikasi merupakan salah satu metode dalam data mining yang populer dan paling sering digunakan [11]. Metode ini memungkinkan untuk mengkategorikan atau memprediksi label kelas berdasarkan data sampel yang tersedia [12]. Berbagai teknik dapat digunakan dalam membangun model klasifikasi, seperti Naïve Bayes, Support Vector Machines (SVM), k-Nearest Neighbor (k-NN) [13], serta Decision Trees dan metode lainnya (Gorunescu, 2011). Menurut Wu & Kumar (2009) menyebutkan bahwa algoritma klasifikasi yang paling umum digunakan adalah C4.5, Naïve Bayes, dan k-Nearest Neighbor (k-NN).

Beberapa penelitian sebelumnya telah menggunakan metode klasifikasi untuk mengkategorikan data kemiskinan, seperti penelitian yang dilakukan oleh Redjeki, Guntara, & Anggoro (2015) serta Annur (2018), yang mengaplikasikan algoritma Naïve Bayes. Algoritma Naïve Bayes terbukti memiliki akurasi tinggi ketika diterapkan pada basis data yang besar [13]. Selain itu, Naïve Bayes juga memiliki keunggulan dalam kecepatan prediksi kelas dibandingkan dengan algoritma klasifikasi lainnya [17]. Namun demikian, kelemahan dari Naïve Bayes Classifier adalah adanya asumsi independensi antara fitur dan kelas atribut yang digunakan [18].

Penelitian lain yang dilakukan oleh Hariati, Wati, & Cahyono (2018) menerapkan algoritma C4.5 untuk program penerima bantuan pemerintah daerah Kabupaten Kutai Kartanegara dalam upaya menanggulangi kemiskinan. Hasil penelitian menunjukkan bahwa algoritma C4.5 termasuk dalam kategori klasifikasi yang sangat baik (excellent classification). Algoritma C4.5 memiliki beberapa kelebihan, antara lain sifatnya yang sederhana [20], mudah dimengerti, mudah diimplementasikan, membutuhkan sedikit waktu, mampu menangani data numerik maupun kategorik [21], serta dapat digunakan untuk mengolah dataset besar [22]. Namun, kelemahan dari algoritma ini adalah apabila dataset memiliki banyak atribut, maka pohon keputusan yang dihasilkan akan menjadi sangat besar, sehingga menyulitkan pembacaan dan pemahaman [20].

Penelitian lain mengenai pengkategorian tingkat kemiskinan dilakukan oleh Santoso & Irawan (2016) dengan membandingkan algoritma klasifikasi k-Nearest Neighbor (k-NN) dan Learning Vector Quantization (LVQ). Hasil penelitian menunjukkan bahwa algoritma k-NN memiliki akurasi yang lebih baik dibandingkan dengan LVQ. Algoritma k-NN memiliki beberapa kelebihan, seperti kesederhanaannya [24], kemudahan dalam penerapan, serta kemampuan menghasilkan kinerja yang baik [25]. Namun, algoritma ini menghadapi masalah dalam menentukan nilai k untuk tetangga terdekat pada titik query dari dataset yang digunakan [26]. Oh & Kim (2018) juga menyatakan bahwa penggunaan nilai k yang tetap atau pasti tidak selalu menjamin estimasi yang akurat dalam berbagai kondisi dataset.

Pendekatan yang digunakan dalam penelitian ini berbeda dari penelitian sebelumnya dalam beberapa aspek utama, terutama dari segi pendekatan klasifikasi, cakupan wilayah, dan variabel yang digunakan. Selain itu, penelitian ini difokuskan pada Kabupaten Temanggung, yang memiliki karakteristik kemiskinan spesifik, berbeda dari wilayah yang pernah diteliti sebelumnya. Penelitian ini juga mengutamakan validasi model yang lebih komprehensif dengan berbagai metrik evaluasi, tidak hanya akurasi tetapi juga precision, recall, dan F1-score. Di samping itu, penelitian ini menghubungkan hasil klasifikasi dengan rekomendasi kebijakan yang lebih aplikatif, sehingga dapat digunakan oleh pemerintah daerah untuk menyusun strategi intervensi yang lebih tepat sasaran. Dengan demikian, penelitian ini tidak hanya berkontribusi dalam aspek teknis klasifikasi kemiskinan, tetapi juga memiliki dampak nyata dalam perencanaan kebijakan penanggulangan kemiskinan di Kabupaten Temanggung.

B. Perumusan Masalah

Berdasarkan berbagai penelitian yang telah dilakukan sebelumnya, dapat disimpulkan bahwa terdapat perbedaan dalam penggunaan algoritma untuk pengkategorian tingkat kemiskinan. Tidak ada konsensus yang jelas mengenai algoritma klasifikasi mana yang paling baik diterapkan untuk data kemiskinan. Oleh karena itu, penelitian ini bertujuan untuk membandingkan algoritma klasifikasi C4.5, Naïve Bayes (NB), dan k-Nearest Neighbor (k-NN) pada Data Kemiskinan Daerah (DKD) Kabupaten Temanggung, dengan harapan dapat menemukan algoritma dengan performa terbaik. Dengan demikian, diharapkan kebijakan yang diambil oleh pemerintah akan lebih tepat sasaran dan dapat mencegah peningkatan jumlah penduduk miskin di Kabupaten Temanggung.

C. Metode

Metode yang diusulkan dapat dilihat pada Gambar 1. Kerangka penelitiannya sebagai berikut. 1) *Dataset*, 2) Algoritma Klasifikasi, 3) Model Validasi, 4) Model Evaluasi.

1. Dataset

Data pada penelitian ini adalah data sekunder yang diperoleh dari Dinas Sosial Kabupaten Temanggung berupa hasil Verifikasi dan Validasi Data Kemiskinan Daerah Kabupaten Temanggung tahun 2019 dengan variabel yang digunakan berjumlah 39 atribut ditambah 1 label. Label yang digunakan merupakan hasil dari perhitungan oleh TNP2K dengan metode *Proxy Means Testing* (PMT) [28]. Sedangkan *dataset* yang digunakan sebanyak 74871 *record*.

Tabel 1. Variabel yang Digunakan

No	Nama Variabel	Jenis Variabel	Skala data
1	Jumlah ART	Atribut	Rasio
2	Jumlah keluarga	Atribut	Rasio
3	Status bangunan	Atribut	Nominal
4	Status lahan	Atribut	Nominal
5	Luas lantai	Atribut	Rasio
6	Lantai terluas	Atribut	Nominal
7	Dinding terluas	Atribut	Nominal
8	Kondisi dinding	Atribut	Nominal
9	Atap terluas	Atribut	Nominal
10	Kondisi atap	Atribut	Nominal
11	Jumlah kamar tidur	Atribut	Rasio
12	Sumber air minum	Atribut	Nominal
13	Cara memperoleh air minum	Atribut	Nominal
14	Sumber penerangan utama	Atribut	Nominal
15	Daya terpasang	Atribut	Nominal
16	Bahan bakar memasak	Atribut	Nominal
17	Fasilitas BAB	Atribut	Nominal
18	Jenis kloset	Atribut	Nominal
19	Tempat buang tinja	Atribut	Nominal
20	Kepemilikan tabung gas 5,5 kg	Atribut	Nominal
21	Kepemilikan lemari es	Atribut	Nominal
22	Kepemilikan ac	Atribut	Nominal
23	Kepemilikan pemanas air	Atribut	Nominal
24	Kepemilikan telepon rumah	Atribut	Nominal
25	Kepemilikan televisi	Atribut	Nominal
26	Kepemilikan emas dan tabungan	Atribut	Nominal
27	Kepemilikan laptop/komputer	Atribut	Nominal
28	Kepemilikan sepeda	Atribut	Nominal
29	Kepemilikan sepeda motor	Atribut	Nominal
30	Kepemilikan mobil	Atribut	Nominal
31	Kepemilikan aset tidak bergerak	Atribut	Nominal
32	Luas aset tidak bergerak	Atribut	Rasio
33	Kepemilikan rumah di tempat lain	Atribut	Nominal
34	Jumlah sapi yang dimiliki	Atribut	Rasio
35	Jumlah kerbau yang dimiliki	Atribut	Rasio
36	Jumlah kuda yang dimiliki	Atribut	Rasio
37	Jumlah babi yang dimiliki	Atribut	Rasio
38	Jumlah kambing/domba yang dimiliki	Atribut	Rasio
39	Status ART yang memiliki usaha	Atribut	Nominal
40	Tingkat kesejahteraan keluarga	Label	Nominal

2. Algoritma C4.5

Algoritma C4.5 adalah algoritma klasifikasi berbasis pohon keputusan [29] yang sangat popular dan sering kali digunakan [30]. Algoritma C4.5 merupakan pengembangan dari ID3 yang ditemukan oleh J. Ross Quinlan [22]. Konsep data dalam C4.5 adalah data dinyatakan dalam bentuk tabel yang terdiri dari atribut dan *record*. Atribut digunakan sebagai parameter yang dibuat sebagai kriteria dalam pembuatan pohon, untuk pohon pertama adalah nilai *Gain* tertinggi dan berulang sampai tidak ada cabang lagi [31]. Pohon keputusan mirip sebuah struktur pohon dimana terdapat node internal (bukan daun) yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas [32].

3. Algoritma Naïve Bayes (NB)

Naïve Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas [33]. *Naïve Bayes* didasarkan pada Teorema Bayes yang dikemukakan seorang ahli matematika yang juga menteri Prebysterian Inggris, Thomas Bayes [34]. *Naïve Bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya [35] dengan asumsi independensi (ketidaktergantungan) yang kuat (naif) pada atributnya [36]. Maknanya, setiap atribut bersifat saling terpisah atau tidak berkaitan dengan atribut yang lainnya pada suatu data yang sama [31]. Algoritma *Naive Bayes* memberikan suatu cara mengkombinasikan peluang terdahulu dengan syarat kemungkinan menjadi sebuah formula yang dapat digunakan untuk menghitung peluang dari tiap kemungkinan yang terjadi [37].

4. Algoritma k-Nearest Neighbor (k-NN)

Algoritma k-*Nearest Neighbor* (k-NN) dikemukakan pertama kali pada awal tahun 1950 [13]. K-*Nearest Neighbor* (k-NN) merupakan salah satu metode klasifikasi yang sederhana dengan menggunakan prinsip ketetanggaan dalam memprediksi kelas data baru [38]. Prinsip kerja k-NN adalah menemukan jarak terpendek antara data *testing* dengan tetangga terdekat sejumlah k dalam data *training* [23], kemudian dikategorikan berdasarkan kelas dengan probabilitas paling besar [39]. Peringkat untuk tetangga terdekat k diukur berdasarkan jarak antara dua tetangga (data *testing* dengan antar data *training*) [40]. Terdapat beberapa metode untuk menentukan jarak antara dua tetangga, antara lain *euclidian distance, minkowski distance*, dan *manhattan distance* [13]. Metode penentuan jarak yang paling sering digunakan adalah *euclidian distance* [41]. Sedangkan kunci utama agar performa algoritma k-NN optimal yaitu pada nilai k yang digunakan [25]. Jika nilai k yang terlalu kecil, maka hasil klasifikasi akan lebih terpengaruh oleh *noise*. Di sisi lain, jika nilai k terlalu besar akan membuat batasan antara setiap klasifikasi menjadi lebih kabur. Nilai k yang bagus dapat dipilih dengan optimasi parameter yaitu dengan *cross validation* [14].

D. Hasil dan Pembahasan

Eksperimen dan perhitungan dilakukan untuk mengetahui performa terbaik di antara ketiga algoritma klasifikasi, yaitu C4.5, Naïve Bayes, dan k-Nearest Neighbor pada Data Kemiskinan Daerah Kabupaten Temanggung ini. *Tools* yang digunakan pada penelitian ini adalah *RapidMiner Studio version 9.7* dengan metode pengujian 10-*fold cross validation*. *Dataset* yang digunakan sejumlah 74871 *record* dengan 39 atribut dan 1 label yang terdiri dari 2 kelas, yaitu kelas Keluarga Tidak Sejahtera dan kelas Keluarga Sejahtera.

Tabel 2. Hasil Perbandingan Algoritma Klasifikasi

Algoritma	AUC	Precision	Recall/ Sensitivity	Specificity	F-Measure
C4.5	0.694	25.81%	0.11%	99.97%	0.00224
NB	0.823	30.18%	60.72%	85.27%	0.40319
k-NN	0.623	37.68%	4.31%	99.25%	0.07728

Berdasarkan Tabel 2 di atas, algoritma C4.5 sangat baik dalam memprediksi kelas keluarga

tidak sejahtera, dapat dilihat dari nilai *specificity* yang sangat tinggi sebesar 99.97%. Namun algoritma ini sangat buruk dalam memprediksi kelas keluarga sejahtera, dilihat dari nilai *recall* atau *sensitivity* dan *F-measure* yaitu sebesar 0.11% yang benar diprediksi sebagai keluarga sejahtera dari total keluarga sejahtera sebenarnya dan secara rata-rata hanya mampu memprediksi kelas positif sebesar 0.00224.

Algoritma *Naïve Bayes* cukup baik dalam memprediksi kelas keluarga sejahtera dapat dilihat dari nilai *recall* atau *sensitivity* yang cukup tinggi sebesar 60.72% yang benar diprediksi sebagai keluarga sejahtera dari total keluarga sejahtera sebenarnya. Dapat dilihat pula pada nilai *F-measure*, bahwa algoritma ini kurang baik dalam memprediksi kelas keluarga sejahtera secara rata-rata. Namun nilai *F-measure* ini jauh lebih baik dibandingkan algoritma sebelumnya yaitu C4.5 yang hanya sebesar 0.00224. Selain itu, algoritma *Naïve Bayes* juga baik dalam memprediksi kelas keluarga tidak sejahtera dapat dilihat dari nilai *specificity* yang tinggi yaitu sebesar 85.27%, namun masih lebih tinggi algoritma C4.5 yaitu sebesar 99.97%.

Algoritma k-*Nearest Neighbor* sangat baik dalam memprediksi kelas keluarga tidak sejahtera dapat dilihat dari nilai *specificity* yang sangat tinggi sebesar 99.25%, namun masih lebih tinggi nilai *specificity* algoritma C4.5 yaitu sebesar 99.97%. Di sisi lain algoritma ini sangat buruk dalam memprediksi kelas keluarga sejahtera, dilihat dari nilai *recall* atau *sensitivity* yaitu sebesar 4.31% yang benar diprediksi sebagai keluarga sejahtera dari total keluarga sejahtera sebenarnya dan nilai *F-measure* yang secara rata-rata hanya mampu memprediksi kelas positif sebesar 0.07728.

Penelitian ini difokuskan pada kelas positif, yaitu kelas keluarga sejahtera. Maka, evaluasi yang digunakan difokuskan pada evaluasi untuk prediksi kelas positif pula, yaitu p*recision*, *recall*, *F-measure*, dan AUC. Secara umum, dari hasil penelitian ini nilai AUC, *recall/sensitivity*, dan *F-measure* tertinggi ditunjukkan oleh algoritma *Naïve Bayes*. Dilihat dari nilai AUC, algoritma *Naïve Bayes* mampu menunjukkan performa terbaik dibandingkan algoritma lainnya secara keseluruhan. Sedangkan dilihat dari nilai *recall/sensitivity*, dan *F-measure* menunjukkan bahwa algoritma *Naïve Bayes* mampu memprediksi kelas positif lebih baik dibandingkan algoritma lainnya. Sedangkan pada nilai p*recision* algoritma *Naïve Bayes* bukanlah nilai yang tertinggi, namun nilai p*recision* algoritma *Naïve Bayes* tidak jauh berbeda dengan algoritma k-*Nearest Neighbor* yang memiliki nilai *precision* tertinggi. Sehingga dapat dikatakan bahwa algoritma *Naïve Bayes* memiliki performa terbaik untuk memprediksi tingkat kesejahteraan pada Data Kemiskinan Daerah (DKD) Kabupaten Temanggung. Hal ini sesuai dengan hasil penelitian Redjeki, Guntara, & Anggoro (2015) dan pernyataan Han et al., (2012) bahwa algoritma *Naïve Bayes* menunjukkan nilai akurasi yang tinggi. Sedangkan algoritma dengan nilai evaluasi kelas positif atau kelas keluarga sejahtera terendah yaitu algoritma C4.5 karena memiliki nilai p*recision*, *recall*, dan *F-measure* terkecil.

E. Simpulan

Dilihat nilai evaluasi kelas positif atau kelas keluarga sejahtera, yaitu AUC, *precision*, *recall/sensitivity*, dan *F-measure,* algoritma *Naïve Bayes* menunjukkan nilai yang tertinggi dibandingkan algoritma lainnya. Nilai *precision* algoritma *Naïve Bayes* bukanlah yang tertinggi, namun nilai *precision* algoritma *Naïve Bayes* tidak jauh berbeda dengan nilai *precision* tertinggi yang ditunjukkan oleh algoritma k-*Nearest Neighbor*. Sedangkan dilihat

dari *specificity*, nilai tertinggi ditunjukkan oleh algoritma C4.5. Sehingga, dapat disimpulkan bahwa algoritma *Naïve Bayes* memiliki performa terbaik untuk memprediksi tingkat kesejahteraan pada Data Kemiskinan Daerah (DKD) Kabupaten Temanggung ini

Pada penelitian ini, algoritma *Naïve Bayes* mampu memprediksi kelas positif atau kelas keluarga sejahtera dengan baik. Namun, tidak mampu memprediksi kelas negatif atau kelas keluarga tidak sejahtera dengan maksimal. Nilai *specificity* algoritma *Naïve Bayes* tidaklah buruk, namun terendah dibandingkan algoritma C4.5 dan k-*Nearest Neighbor.* Sehingga dapat dikatakan bahwa pada penelitian ini tidak terdapat algoritma yang mampu memprediksi kelas positif (kelas keluarga sejahtera) dan kelas negatif (kelas keluarga tidak sejahtera) secara maksimal.

Daftar Pustaka

[1] D. Septiadi and M. Nursan, "Pengentasan Kemiskinan Indonesia: Analisis Indikator Makroekonomi dan Kebijakan Pertanian," *J. Hexagro*, vol. 4, no. 1, pp. 1–14, 2020.
[2] A. S. Rini and L. Sugiharti, "Faktor-Faktor Penentu Kemiskinan di Indonesia: Analisis Rumah Tangga," *J. Ilmu Ekon. Terap.*, vol. 01, no. 2, pp. 17–33, 2016.
[3] M. P. Todaro and S. C. Smith, *Economic Development*, 12th editi. New Jersey: Pearson Education, Inc., 2015.
[4] T. Sukwika, "Peran Pembangunan Infrastruktur terhadap Ketimpangan Ekonomi Antarwilayah di Indonesia," *J. Wil. dan Lingkung.*, vol. 6, no. 2, pp. 115–130, 2018, doi: 10.14710/jwl.6.2.115-130.
[5] BPS Temanggung, *Kabupaten Temanggung Dalam Angka 2020*. Temanggung: BPS Kabupaten Temanggung, 2020.
[6] W. Setiadi, "Prestasi dan Peringatan Kemiskinan Temanggung," Pemerintah Kabupaten Temanggung.
[7] TNP2K, *Standar Pengelolaan Basis Data Terpadu Untuk Program Perlindungan Sosial*, 1st ed. Jakarta Pusat: TNP2K, 2015.
[8] BPS, *Data dan Informasi Kemiskinan Kabupaten/Kota*. Jakarta: BPS, 2019.
[9] G. Anuraga, "Hierarchical Clustering Multiscale Bootstrap Untuk Pengelompokan Kemiskinan di Jawa Timur," *J. Stat.*, vol. 1, no. 3, pp. 27–33, 2015.
[10] N. S. Sani, M. A. Rahman, A. A. Bakar, S. Sahran, and H. M. Sarim, "Machine Learning Approach for Bottom 40 Percent Households (B40) Poverty Classification," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 4, pp. 1698–1705, 2018, doi: 10.18517/IJASEIT.8.4-2.6829.
[11] B. Cigsar and D. Unal, "Comparison of Data Mining Classification Algorithms Determining the Default Risk," *J. Sci. Program.*, vol. 2019, pp. 1–8, 2019, doi: 10.1155/2019/8706505.
[12] L. Yu, G. Chen, A. Koronios, S. Zhu, and X. Guo, "Application and Comparison of Classification Techniques in Controlling Credit Risk," in *Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications*, vol. 6, World Scientific, 2007, pp. 111–145. doi: 10.1142/9789812779861_0002.
[13] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Third Edit. Morgan Kaufmann, 2012. doi: 10.1016/B978-0-12-381479-1.00001-0.
[14] X. Wu and V. Kumar, *The Top Ten Algorithms in Data Mining*, First Edit. Minnesota: CRC Press, 2009.
[15] S. Redjeki, M. Guntara, and P. Anggoro, "Naive Bayes Classifier Algorithm Approach for Mapping Poor Families Potential," *Int. J. Adv. Res. Artif. Intell.*, vol. 4, no. 12, pp. 29–33, 2015.

[16] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018.
[17] P. Kaviani and S. Dhotre, "Short Survey on Naive Bayes Algorithm," *Int. J. Adv. Eng. Res. Dev.*, vol. 4, no. 11, pp. 607–611, 2017.
[18] C. Zhang, C. Liu, X. Zhang, and G. Almpanidis, "An Up-to-Date Comparison of State-of-the-Art Classification Algorithms," *Expert Syst. Appl.*, 2017, doi: 10.1016/j.eswa.2017.04.003.
[19] Hariati, M. Wati, and B. Cahyono, "Penerapan Algoritma C4.5 pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara," *J. Rekayasa Teknol. Inf.*, vol. 2, no. 2, pp. 106–114, 2018.
[20] H. Dahan, S. Cohen, L. Rokach, and O. Maimon, *Proactive Data Mining with Decision Trees*. Tel Aviv: Springer, 2014. doi: 10.1007/978-1-4939-0539-3.
[21] D. Farid, L. Zhang, C. M. Rahman, M. A. Hossain, and R. Strachan, "Hybrid Decision Tree and Naïve Bayes Classifiers for Multi-Class Classification Tasks," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1937–1946, 2014, doi: 10.1016/j.eswa.2013.08.089.
[22] L. Rokach and O. Maimon, *Data Mining with Decision Trees : Theory and Applications*, Second Edi. 5 Toh Tuck Link: World Scientific Publishing Co. Pte. Ltd, 2015.
[23] Santoso and M. I. Irawan, "Classification of Poverty Levels Using k-Nearest Neighbor and Learning Vector Quantization," *Int. J. Comput. Sci. Appl. Math.*, vol. 2, no. 1, pp. 8–13, 2016, doi: 10.12962/j24775401.v2i1.1578.
[24] N. Bhatia and Vandana, "Survey of Nearest Neighbor Techniques," *Int. J. Comput. Sci. Inf. Secur.*, vol. 8, no. 2, pp. 302–305, 2010.
[25] M. W. Berry and M. Browne, *Lecture Notes in Data Mining*. Knoxville: World Scientific Publishing Co. Pte. Ltd. 5, 2006.
[26] Y. Liaw, C. Wu, and M. Leou, "Fast K-Nearest Neighbors Search Using Modified Principal Axis Search Tree," *Digit. Signal Process.*, vol. 20, no. 5, pp. 1494–1501, 2010, doi: 10.1016/j.dsp.2010.01.009.
[27] J. Oh and J. Kim, "Adaptive K-Nearest Neighbour Algorithm for WiFi Fingerprint Positioning," *ICT Express*, pp. 1–3, 2018, doi: 10.1016/j.icte.2018.04.004.
[28] TNP2K, *Petunjuk Pelaksanaan Verifikasi/Validasi Data Rumah Tangga dalam Mekanisme Pemutakhiran Mandiri (MPM) Data Terpadu Program Penanganan Fakir Miskin*. Jakarta Pusat: TNP2K, 2017.
[29] D. H. Kamagi and S. Hansun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *Ultim. J.*, vol. 6, no. June 2014, pp. 15–20, 2014, doi: 10.31937/ti.v6i1.327.
[30] L. Marlina, Muslim, and A. P. U. Siahaan, "Data Mining Classification Comparison (Naïve Bayes and C4.5 Algorithms)," *Int. J. Eng. Trends Technol.*, vol. 38, no. 7, pp. 380–383, 2016, doi: 10.14445/22315381/IJETT-V38P268.
[31] N. Khotimah and D. Istiawan, "Perbandingan Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbour untuk Prediksi Lahan Kritis di Kabupaten Pemalang," in *The 7th University Research Colloqium 2018*, 2018, pp. 41–50.
[32] S. Haryati, A. Sudarsono, and E. Suryana, "Implementasi Data Minig untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
[33] N. Yahya and A. Jananto, "Komparasi Kinerja Algoritma C4.5 dan Naive Bayes untuk Prediksi Kegiatan Penerimaan Mahasiswa Baru (Studi kasus: Universitas Stikubank Semarang)," in *Seminar Nasional Multi Disiplin Ilmu*, 2019, pp. 221–228.
[34] F. Gorunescu, *Data Mining: Concepts, Models, and Techniques*, Twelfth Ed. Craiova: Springer, 2011. doi: 10.1007/978-3-642-19721-5.

[35] Bustami, "Penerapan Algoritma Naive Bayes untuk Mengklasifikasi Data Nasabah Asuransi," *J. Inform.*, vol. 8, no. 1, pp. 884–898, 2014.
[36] S. Suprianto, "Implementasi Algoritma Naive Bayes Untuk Menentukan Lokasi Strategis Dalam Membuka Usaha Menengah Ke Bawah di Kota Medan (Studi Kasus : Disperindag Kota Medan)," *J. Sist. Komput. dan Inform.*, vol. 1, no. 2, pp. 125–130, 2020, doi: 10.30865/json.v1i2.1939.
[37] R. A. Saputra, "Komparasi Algoritma Klasifikasi Data Mining untuk Memprediksi Penyakit Tuberculosis (TB): Studi Kasus Puskesmas Karawang Sukabumi," in *Seminar Nasional Inovasi dan Tren*, 2014, pp. 1–8.
[38] R. Siringoringo, "Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan k-Nearest Neighbor," *J. ISD*, vol. 3, no. 1, pp. 44–49, 2018.
[39] K. S. Kim, H. H. Choi, C. S. Moon, and C. W. Mun, "Comparison of k -Nearest Neighbor, Quadratic Discriminant and Linear Discriminant Analysis in Classification of Electromyogram Signals Based On The Wrist-motion Directions," *Curr. Appl. Phys.*, vol. 11, no. 3, pp. 740–745, 2011, doi: 10.1016/j.cap.2010.11.051.
[40] J. Sreemathy and P. S. Balamurugan, "An Efficient Text Classification Using kNN and Naive Bayesian," *Int. J. Comput. Sci. Eng.*, vol. 4, no. 03, pp. 392–396, 2012.
[41] Y. Zhang, G. Cao, B. Wang, and X. Li, "A Novel Ensemble Method for k-Nearest Neighbor," *Pattern Recognit.*, 2018, doi: 10.1016/j.patcog.2018.08.003.