



Klasifikasi Indeks Kedalaman Kemiskinan Provinsi Sulawesi Selatan Berbasis *Decision Tree*, *K-Nearest Neighbor*, *Naive Bayes*, *Neural Network*, dan *Random Forest*

Muhammad Faozan Mulad Khalik^{#1}, Fatchul Arifin^{#2}

[#]*Elektronika dan Informatika, Fakultas Teknik, Universitas Negeri Yogyakarta*

Jl. Colombo Yogyakarta No.1, Karang Malang, Kabupaten Sleman, Daerah Istimewa Yogyakarta

¹faozanfcb@gmail.com

²fatchul@uny.ac.id

Abstrak— Penelitian ini bertujuan untuk mendapatkan klasifikasi indeks kedalaman kemiskinan dengan metode terbaik untuk kabupaten/kota di Provinsi Sulawesi Selatan dengan membandingkan metode *Decision Tree*, *K-Nearest Neighbor*, *Naive Bayes*, *Neural Network*, dan *Random Forest*. Penelitian ini menggunakan metode kuantitatif dengan menggunakan data sekunder yang diperoleh dari situs resmi Badan Pusat Statistik Provinsi Sulawesi Selatan. Pada penelitian ini digunakan 168 data latih yang bersumber dari data tahun 2014 sampai dengan data tahun 2021, kemudian untuk data uji yang digunakan yaitu 24 data yang bersumber dari data tahun 2022. Hasil dari penelitian ini menunjukkan bahwa metode *K-NN* dan *Neural Network* memperoleh performa paling tinggi dibandingkan dengan metode lain tingkat akurasi 79,17%, *precision* 85,71%, *recall* 80%. Namun pada penilaian parameter *AUC*, metode *Neural Network* lebih unggul dibandingkan metode *K-NN* dengan skor *AUC* 0,837. Sehingga dapat disimpulkan bahwa metode *Neural Network* ini dapat dijadikan sebagai metode untuk melakukan klasifikasi indeks kedalaman kemiskinan kabupaten/kota Provinsi Sulawesi Selatan.

Kata kunci— *Decision Tree*, *K-Nearest Neighbor*, *Naive Bayes*, *Neural Network*, *Random Forest*, *Klasifikasi*

I. PENDAHULUAN

Kemiskinan merupakan persoalan yang dihadapi oleh berbagai negara, terutama negara-negara berkembang seperti Indonesia. Individu dianggap miskin ketika mereka tidak memiliki kemampuan untuk memenuhi kebutuhan dasar minimum yang layak, seperti sandang, pangan, perumahan, dan kebutuhan sosial. Kemiskinan juga mencerminkan ketidakmampuan masyarakat dalam memenuhi standar minimum kebutuhan primer dan sekunder yang penting [1]. Kemiskinan terjadi akibat terbatasnya aset dan pendapatan yang dapat mencukupi kebutuhan dasar seperti pakaian, makanan, perumahan, akses kesehatan, dan Pendidikan [2]. Selain berdampak pada kondisi materi dan kebutuhan dasar, kemiskinan juga

memiliki implikasi yang melampaui aspek ekonomi. Kemiskinan dapat membatasi akses individu terhadap kesempatan pendidikan yang layak, menghambat pertumbuhan ekonomi suatu negara, dan memperburuk kesenjangan sosial. Oleh karena itu, memahami dan mengatasi kemiskinan merupakan tantangan yang perlu diatasi dengan pendekatan yang holistik dan terintegrasi.

Di Indonesia, tingkat kemiskinan yang tinggi disebabkan oleh berbagai faktor, termasuk pengangguran, inflasi, pertumbuhan ekonomi, dan pendidikan. Tingkat kemiskinan yang tinggi juga berdampak negatif pada akses pendidikan yang berkualitas bagi anak-anak, kesulitan dalam membiayai kesehatan, keterbatasan tabungan, kesulitan mengakses pelayanan publik, keterbatasan perlindungan sosial untuk keluarga, dan peningkatan urbanisasi masyarakat [3]. Kemiskinan dapat dikenali melalui indikator seperti rendahnya tingkat pendidikan, kondisi kesehatan yang tidak memadai, serta tingkat kesejahteraan yang rendah di tengah masyarakat dengan pendapatan yang minim, sehingga menghasilkan ketidakseimbangan sosial.

Tingkat kemiskinan yang tinggi menjadi tantangan yang menderas. Provinsi Sulawesi Selatan ini memiliki potensi sumber daya alam yang besar, namun kesenjangan pembangunan antara daerah perkotaan dan pedesaan serta keterbatasan akses ke layanan dasar masih menjadi kendala yang harus diatasi. Mengklasifikasikan tingkat kedalaman kemiskinan di provinsi akan membantu pemerintah daerah dan *stakeholder* terkait merancang kebijakan yang lebih efektif dan tepat sasaran dalam penanggulangan kemiskinan.

Klasifikasi melibatkan langkah-langkah dalam mencari sekumpulan model (fungsi) yang mampu menggambarkan dan membedakan berbagai kelas data atau konsep, dengan tujuan menggunakan model tersebut untuk memprediksi kelas suatu objek yang belum diketahui kelasnya. [4].

Dalam konteks ini, terdapat beberapa metode yang dapat digunakan untuk mengklasifikasikan tingkat kedalaman kemiskinan di provinsi Sulawesi Selatan, antara lain metode *Decision Tree*, *K-NN*, *Naïve Bayes*, *Neural Network*, dan *Random Forest*. Dalam penelitian ini, perbandingan akurasi dari beberapa metode tersebut dapat memberikan kontribusi dalam pengembangan metode klasifikasi dan analisis data untuk pemodelan kemiskinan, sehingga dapat diterapkan dalam studi-studi selanjutnya dengan fokus yang serupa atau di wilayah lain yang relevan.

II. METODE PENELITIAN

Tujuan dari penelitian ini adalah untuk menentukan metode yang cocok digunakan untuk melakukan klasifikasi indeks kedalaman kemiskinan kabupaten/kota Provinsi Sulawesi Selatan. Penelitian ini menggunakan metode kuantitatif dengan menggunakan data sekunder yang diperoleh dari website resmi Badan Pusat Statistik Provinsi Sulawesi Selatan.

A. Dataset

Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari website resmi Badan Pusat Statistik Provinsi Sulawesi Selatan sebanyak 192 data. Data tersebut merupakan data yang akan digunakan sebagai data latih dan data uji. Pada penelitian ini terdapat, 3 variabel independen, dan 1 variabel dependen. Variabel independent meliputi (X1) tingkat pengangguran terbuka, (X2) indeks pembangunan manusia, dan (X3) pengeluaran perkapita, kemudian untuk variable dependen adalah (Y) indeks kedalaman kemiskinan. Data latih meliputi data dari tahun 2014 sampai dengan tahun 2021, tetapi tanpa data tahun 2016 karena salah satu variable tidak terdapat data pada tahun 2016. Kemudian untuk data uji adalah data tahun 2022.

B. Penentuan Label

Penentuan label dilakukan dengan kriteria jika P1 (indeks kedalaman kemiskinan) < rata-rata P0 (indeks kedalaman kemiskinan) maka diberikan kode “Rendah”. Sedangkan, jika P1 (indeks kedalaman kemiskinan) ≥ rata-rata P1 (indeks kedalaman kemiskinan) maka diberikan kode “Tinggi”.

C. Normalisasi Data

Pemilihan satuan pengukuran dalam data yang digunakan dapat berdampak pada analisis data. Sebagai contoh, jika variabel pengeluaran per kapita diukur dalam rupiah atau ribuan, sementara variabel lain diukur dalam persentase, hal tersebut dapat menghasilkan perbedaan dalam hasil analisis. Maka dari itu perlu dilakukan proses normalisasi data dengan menggunakan fungsi aktivasi sigmoid yang merupakan fungsi asimtotik (tidak pernah mencapai 0 ataupun 1). Sehingga normalisasi data yang dilakukan pada interval yang lebih kecil, misalnya pada interval [0,1 hingga 0,9] dengan menggunakan persamaan berikut :

$$X = \frac{0,8(x-a)}{b-a} \quad 0,1 \quad (1) [5]$$

Keterangan :

0,8 = Ketetapan

x = Nilai data ke-n

a = Nilai data yang paling rendah

b = Nilai data yang paling tinggi

Berikut ini merupakan hasil dari normalisasi data latih dan juga data uji yang digunakan :

TABEL I
HASIL NORMALISASI DATA LATIH

| No. | X1 | X2 | X3 | Y |
|-------|---------|---------|----------|--------|
| 1 | 0,20173 | 0,338 | 0,309207 | Tinggi |
| 2 | 0,21931 | 0,40816 | 0,424763 | Rendah |
| 3 | 0,26884 | 0,38439 | 0,512754 | Rendah |
| 4 | 0,17883 | 0,2173 | 0,320601 | Tinggi |
| 5 | 0,26138 | 0,33649 | 0,418221 | Rendah |
| 6 | 0,28109 | 0,43343 | 0,341845 | Rendah |
| 7 | 0,19108 | 0,33762 | 0,341919 | Rendah |
| 8 | 0,38762 | 0,43795 | 0,454167 | Tinggi |
| 9 | 0,36418 | 0,39269 | 0,489966 | Tinggi |
| 10 | 0,41105 | 0,46511 | 0,453064 | Rendah |
| | | | | |
| 159 | 0,2012 | 0,38175 | 0,428292 | Rendah |
| 160 | 0,12663 | 0,39873 | 0,330304 | Rendah |
| 161 | 0,3237 | 0,32216 | 0,287448 | Tinggi |
| 162 | 0,22783 | 0,23692 | 0,1 | Rendah |
| 163 | 0,14794 | 0,30556 | 0,422779 | Tinggi |
| 164 | 0,48349 | 0,41306 | 0,514959 | Rendah |
| 165 | 0,24913 | 0,27727 | 0,15447 | Tinggi |
| 166 | 0,63262 | 0,77515 | 0,751659 | Rendah |
| 167 | 0,43023 | 0,63597 | 0,576192 | Rendah |
| 168 | 0,48349 | 0,6356 | 0,504227 | Rendah |

TABEL II
HASIL NORMALISASI DATA UJI

| No. | X1 | X2 | X3 | Y |
|-----|----------|----------|----------|--------|
| 1 | 0,164769 | 0,243191 | 0,25166 | Tinggi |
| 2 | 0,148399 | 0,331684 | 0,373427 | Rendah |
| 3 | 0,252313 | 0,302779 | 0,470515 | Rendah |
| 4 | 0,216014 | 0,1 | 0,249949 | Tinggi |
| 5 | 0,245907 | 0,241412 | 0,357544 | Rendah |
| 6 | 0,290747 | 0,360589 | 0,28147 | Rendah |
| 7 | 0,186833 | 0,242301 | 0,274465 | Rendah |
| 8 | 0,417438 | 0,361034 | 0,411057 | Tinggi |
| 9 | 0,430961 | 0,307226 | 0,444777 | Tinggi |
| 10 | 0,437367 | 0,384603 | 0,400631 | Rendah |
| 11 | 0,220285 | 0,183602 | 0,237895 | Rendah |
| 12 | 0,300712 | 0,303224 | 0,276909 | Rendah |
| 13 | 0,239502 | 0,328127 | 0,519059 | Rendah |
| 14 | 0,3121 | 0,408171 | 0,490552 | Rendah |
| 15 | 0,257295 | 0,404169 | 0,46799 | Rendah |
| 16 | 0,1 | 0,467315 | 0,393138 | Tinggi |
| 17 | 0,33274 | 0,377043 | 0,321869 | Tinggi |
| 18 | 0,223843 | 0,311228 | 0,1 | Rendah |
| 19 | 0,258719 | 0,339244 | 0,468235 | Tinggi |
| 20 | 0,37758 | 0,490884 | 0,545856 | Rendah |
| 21 | 0,200356 | 0,332574 | 0,174119 | Tinggi |
| 22 | 0,9 | 0,9 | 0,9 | Rendah |
| 23 | 0,457295 | 0,696331 | 0,624781 | Rendah |
| 24 | 0,642349 | 0,712785 | 0,574038 | Tinggi |

D. Pengembangan Model Uji Klasifikasi

Dalam tahap pemodelan ini memproses klasifikasi data menggunakan lima metode, yaitu *Decision Tree*, *K-*

Nearest Neighbor, Naïve Bayes, Neural Network, Random Forest.

1) *Decision Tree*: adalah representasi visual dalam bentuk diagram aliran yang menyerupai sebuah pohon. *Decision tree* digunakan sebagai pendekatan untuk melakukan klasifikasi data. Dalam *decision tree*, kita dapat menemukan fitur dan pola dalam jumlah besar. Kemampuan ini membuat *decision tree* digunakan secara luas dalam eksplorasi data dan pemodelan prediktif selama lebih dari dua dekade. *Decision tree* memiliki dasar yang kuat dalam bidang pembelajaran mesin dan kecerdasan buatan [6].

2) *K-Nearest Neighbor* (K-NN) adalah suatu pendekatan atau metode dalam *data mining* yang digunakan untuk melakukan klasifikasi. Metode ini bertujuan untuk mencari atau memprediksi nilai atau kelas di masa depan [7]. Metode *K-Nearest Neighbor* (K-NN) adalah salah satu metode yang paling sederhana dalam data mining. Tujuan metode ini adalah untuk menentukan hasil klasifikasi berdasarkan jarak terdekat antara setiap objek data. Prosesnya dimulai dengan membagi data awal menjadi dua bagian, yaitu data training dan data testing. Kemudian, jarak antara setiap data testing dengan data training dihitung. Pada penelitian ini penentuan jarak yang digunakan yaitu dengan menggunakan metode *Chebyshev Distance*.

Chebyshev Distance adalah metode yang pengukuran jarak berdasarkan nilai absolut atau nilai mutlak dari selisih sepasang titik koordinat. Jika memiliki dua vektor dengan nilai berbeda untuk setiap elemen, jarak yang diukur oleh *chebyshev* adalah berdasarkan pada nilai mutlak dari perbedaan antara elemen-elemen pada vektor tersebut dan jumlah data secara otomatis harus sama [8].

3) *Naïve Bayes* adalah metode klasifikasi yang menggunakan peluang dengan menghitung frekuensi serta kombinasi nilai dalam dataset yang diberikan untuk menentukan peluang hasil. Metode ini adalah metode klasifikasi peluang sederhana yang menghitung sejumlah peluang dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan [9].

Algoritma *Naïve Bayes* adalah salah satu metode populer dalam *data mining* yang termasuk dalam sepuluh metode klasifikasi *data mining* yang paling terkenal di antara algoritma lainnya.[10]. *Data mining* ini berarti menganalisis suatu data yang berbeda dan meringkas data tersebut menjadi informasi yang berguna.

4) *Backpropagation Neural Network* (BNN) merupakan suatu metode pelatihan yang memiliki struktur dengan beberapa lapisan. Metode ini menggunakan nilai *error* dari *output* sebagai acuan untuk memperbarui bobotnya melalui proses yang bergerak ke belakang (*backward*). Untuk mendapatkan nilai tersebut, terlebih dahulu harus dilakukan tahap *forward propagation* [11]. Metode pengenalan adalah langkah awal dalam mempersiapkan data yang akan diproses oleh BNN. Di sisi lain, metode pelatihan merupakan proses di mana BNN

dilatih untuk mengenali data dan menyimpan pengetahuan atau informasi yang diperoleh ke dalam bobot-bobotnya.

Pelatihan BNN terdiri dari tiga tahap utama: tahap *feedforward*, tahap *backpropagation*, dan tahap modifikasi bobot. Pada tahap *feedforward*, pola input diproses secara berurutan, mulai dari lapisan *input* dan bergerak menuju lapisan *output*. Selama tahap *backpropagation*, setiap unit *output* menerima pola target yang sesuai dengan pola *input*, yang memungkinkan penghitungan kesalahan. Kesalahan ini kemudian dipropagasikan kembali melalui jaringan. Secara bersamaan, pada tahap modifikasi bobot, penyesuaian dilakukan pada bobot untuk meminimalkan terjadinya kesalahan. Ketiga tahap ini diulang secara iteratif sampai kriteria penghentian tertentu terpenuhi [12].

5) *Random Forest* adalah pengembangan dari metode *Decision Tree* yang menggabungkan beberapa *Decision Tree* secara simultan. *Random Forest* memiliki beberapa keunggulan, seperti kemampuannya dalam meningkatkan akurasi prediksi saat terdapat data yang hilang, serta dalam menangani data outlier. Selain itu, *Random Forest* juga efisien dalam penyimpanan data. Selain itu, *Random Forest* menggabungkan proses seleksi fitur, memungkinkannya untuk mengidentifikasi fitur yang paling relevan dan meningkatkan kinerja model klasifikasi. Kemampuan seleksi fitur ini memberdayakan *Random Forest* untuk secara efektif menangani *dataset* besar yang berisi parameter yang rumit [13].

Dalam penelitian ini, setiap metode klasifikasi yang digunakan telah ditentukan parameter-parameter yang spesifik. Berikut adalah parameter-parameter yang digunakan untuk masing-masing metode klasifikasi :

TABEL III
PARAMETER METODE

| No. | Metode | Parameter | Nilai |
|-----|----------------|--------------------------|--------------------|
| 1 | Decision Tree | Criterion | Accuracy |
| | | Maximal Depth | 100 |
| | | Confidence | 0,1 |
| | | Minimal Gain | 0,5 |
| | | Minimal Leaf Size | 10 |
| 2 | K-NN | k | 5 |
| | | Measure types | Numerical Measures |
| | | Numerical measure | Chebyshev Distance |
| 3 | Naïve Bayes | Tidak terdapat parameter | - |
| 4 | Neural Network | Hidden layer | 1 |
| | | Hidden layer neuron | 100 |
| | | Training cycles | 5000 |
| | | Learning rate | 0,05 |
| | | Momentum | 0,8 |
| 5 | Random Forest | Number of trees | 100 |
| | | Criterion | Accuracy |
| | | Maximal depth | 10 |
| | | Confidence | 0,5 |
| | | Minimal gain | 0,5 |
| | | Minimal leaf size | 2 |

E. Model Testing dan Validasi

Pada penelitian ini, untuk mengevaluasi model akan dilakukan dengan menghitung beberapa parameter validasi untuk mengukur kinerja masing-masing metode. Parameter yang dimaksud adalah akurasi, precision, recall, dan AUC. Persamaan (2), (3), dan (4) merupakan rumus untuk perhitungan akurasi, precision, dan recall.

$$average\ accuracy = \frac{1}{N} \sum_{i=1}^N \left(\frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \right) \quad (2)$$

$$Weighted\ Precision = \left(\frac{\sum_{i=1}^N Precision_i + Number_i}{\sum_{i=1}^N Number_i} \right) \quad (3)$$

$$Weighted\ Recall = \left(\frac{\sum_{i=1}^N Recall_i + Number_i}{\sum_{i=1}^N Number_i} \right) \quad (4)$$

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \quad (5)$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (6)$$

Keterangan :

N = jumlah class/jumlah label

i = kelas ke- i

$Number_i$ = Jumlah data pada kelas ke- i

Perhitungan yang dihasilkan dari persamaan (5) akan digunakan sebagai dasar untuk memperoleh solusi pada persamaan (3), sementara hasil perhitungan dari persamaan (6) akan menjadi faktor pendukung dalam perhitungan persamaan (4).

Perhitungan AUC juga sangat direkomendasikan untuk memprediksi label biner karena secara efektif menilai kemampuan model untuk membedakan antara kelas positif dan negatif pada berbagai ambang batas klasifikasi [16].

III. HASIL DAN PEMBAHASAN

A. Hasil Pengolahan Data

1. Decision Tree

Berikut ini merupakan hasil pengujian dengan menggunakan metode *Decision Tree* :

TABEL IV
HASIL KLASIFIKASI METODE DECISION TREE

| No. | Label | Prediksi | Confidence (Tinggi) | Confidence (Rendah) |
|-----|--------|----------|---------------------|---------------------|
| 1 | Tinggi | Rendah | 0,423 | 0,577 |
| 2 | Rendah | Rendah | 0,423 | 0,577 |
| 3 | Rendah | Rendah | 0,423 | 0,577 |
| 4 | Tinggi | Rendah | 0,423 | 0,577 |
| 5 | Rendah | Rendah | 0,423 | 0,577 |
| 6 | Rendah | Rendah | 0,423 | 0,577 |
| 7 | Rendah | Rendah | 0,423 | 0,577 |
| 8 | Tinggi | Rendah | 0,423 | 0,577 |
| 9 | Tinggi | Rendah | 0,423 | 0,577 |
| 10 | Rendah | Rendah | 0,423 | 0,577 |
| 11 | Rendah | Rendah | 0,423 | 0,577 |
| 12 | Rendah | Rendah | 0,423 | 0,577 |
| 13 | Rendah | Rendah | 0,423 | 0,577 |
| 14 | Rendah | Rendah | 0,423 | 0,577 |
| 15 | Rendah | Rendah | 0,423 | 0,577 |
| 16 | Tinggi | Rendah | 0,423 | 0,577 |

| No. | Label | Prediksi | Confidence (Tinggi) | Confidence (Rendah) |
|-----|--------|----------|---------------------|---------------------|
| 17 | Tinggi | Rendah | 0,423 | 0,577 |
| 18 | Rendah | Rendah | 0,423 | 0,577 |
| 19 | Tinggi | Rendah | 0,423 | 0,577 |
| 20 | Rendah | Rendah | 0,423 | 0,577 |
| 21 | Tinggi | Rendah | 0,423 | 0,577 |
| 22 | Rendah | Rendah | 0,423 | 0,577 |
| 23 | Rendah | Rendah | 0,423 | 0,577 |
| 24 | Tinggi | Rendah | 0,423 | 0,577 |

Berdasarkan tabel diatas, dapat dilihat bahwa dengan menggunakan metode *Decision Tree* jumlah prediksi yang benar adalah 15 dari 24 data, sehingga tingkat akurasi yang diperoleh adalah 62,5%. Kemudian untuk hasil parameter penilaiannya dapat dilihat pada tabel berikut :

TABEL V
HASIL PARAMETER PENILAIAN METODE DECISION TREE

| Parameter Penilaian | Skor |
|---------------------|-------|
| Akurasi | 62,5% |
| Precision | 62,5% |
| Recall | 100% |
| AUC | 0,500 |

2. K-NN (K-Nearest Neighbor)

Berikut ini merupakan hasil pengujian dengan menggunakan metode *K-NN* :

TABEL VI
HASIL KLASIFIKASI METODE K-NN

| No. | Label | Prediksi | Confidence (Tinggi) | Confidence (Rendah) |
|-----|--------|----------|---------------------|---------------------|
| 1 | Tinggi | Tinggi | 0,617 | 0,382 |
| 2 | Rendah | Rendah | 0,000 | 1,000 |
| 3 | Rendah | Rendah | 0,214 | 0,785 |
| 4 | Tinggi | Tinggi | 1,000 | 0,000 |
| 5 | Rendah | Rendah | 0,398 | 0,601 |
| 6 | Rendah | Rendah | 0,403 | 0,596 |
| 7 | Rendah | Rendah | 0,220 | 0,779 |
| 8 | Tinggi | Tinggi | 0,607 | 0,392 |
| 9 | Tinggi | Tinggi | 0,819 | 0,180 |
| 10 | Rendah | Rendah | 0,411 | 0,588 |
| 11 | Rendah | Rendah | 0,400 | 0,599 |
| 12 | Rendah | Tinggi | 0,584 | 0,415 |
| 13 | Rendah | Rendah | 0,186 | 0,813 |
| 14 | Rendah | Rendah | 0,407 | 0,592 |
| 15 | Rendah | Tinggi | 0,608 | 0,391 |
| 16 | Tinggi | Tinggi | 0,813 | 0,186 |
| 17 | Tinggi | Rendah | 0,215 | 0,784 |
| 18 | Rendah | Tinggi | 1,000 | 0,000 |
| 19 | Tinggi | Tinggi | 0,611 | 0,388 |
| 20 | Rendah | Rendah | 0,000 | 1,000 |
| 21 | Tinggi | Tinggi | 0,816 | 0,183 |
| 22 | Rendah | Rendah | 0,000 | 1,000 |
| 23 | Rendah | Rendah | 0,000 | 1,000 |
| 24 | Tinggi | Rendah | 0,401 | 0,598 |

Berdasarkan tabel diatas, dapat dilihat bahwa dengan menggunakan metode *K-NN* jumlah prediksi yang benar adalah 19 dari 24 data, sehingga tingkat akurasi yang diperoleh adalah 79,17%. Kemudian untuk hasil parameter penilaiannya dapat dilihat pada tabel berikut :

TABEL VII
HASIL PARAMETER PENILAIAN METODE K-NN

| Parameter Penilaian | Skor |
|---------------------|--------|
| Akurasi | 79,17% |
| Precision | 85,71% |
| Recall | 80% |
| AUC | 0,833 |

3. *Naïve Bayes*

Berikut ini merupakan hasil pengujian dengan menggunakan metode *Naïve Bayes* :

TABEL VIII
HASIL KLASIFIKASI METODE NAÏVE BAYES

| No. | Label | Prediksi | Confidence (Tinggi) | Confidence (Rendah) |
|-----|--------|----------|---------------------|---------------------|
| 1 | Tinggi | Tinggi | 0,831 | 0,169 |
| 2 | Rendah | Tinggi | 0,775 | 0,225 |
| 3 | Rendah | Tinggi | 0,628 | 0,372 |
| 4 | Tinggi | Tinggi | 0,672 | 0,328 |
| 5 | Rendah | Tinggi | 0,778 | 0,222 |
| 6 | Rendah | Tinggi | 0,786 | 0,214 |
| 7 | Rendah | Tinggi | 0,828 | 0,172 |
| 8 | Tinggi | Tinggi | 0,525 | 0,475 |
| 9 | Tinggi | Rendah | 0,470 | 0,530 |
| 10 | Rendah | Rendah | 0,480 | 0,520 |
| 11 | Rendah | Tinggi | 0,793 | 0,207 |
| 12 | Rendah | Tinggi | 0,802 | 0,198 |
| 13 | Rendah | Tinggi | 0,508 | 0,492 |
| 14 | Rendah | Rendah | 0,449 | 0,551 |
| 15 | Rendah | Tinggi | 0,550 | 0,450 |
| 16 | Tinggi | Tinggi | 0,569 | 0,431 |
| 17 | Tinggi | Tinggi | 0,725 | 0,275 |
| 18 | Rendah | Tinggi | 0,776 | 0,224 |
| 19 | Tinggi | Tinggi | 0,615 | 0,385 |
| 20 | Rendah | Rendah | 0,141 | 0,859 |
| 21 | Tinggi | Tinggi | 0,821 | 0,179 |
| 22 | Rendah | Rendah | 0,000 | 1,000 |
| 23 | Rendah | Rendah | 0,001 | 0,999 |
| 24 | Tinggi | Rendah | 0,000 | 1,000 |

Berdasarkan tabel diatas, dapat dilihat bahwa dengan menggunakan metode *Naïve Bayes* jumlah prediksi yang benar adalah 12 dari 24 data, sehingga tingkat akurasi yang diperoleh adalah 50%. Kemudian untuk hasil parameter penilaiannya dapat dilihat pada tabel berikut :

TABEL IX
HASIL PARAMETER PENILAIAN METODE NAÏVE BAYES

| Parameter Penilaian | Skor |
|---------------------|--------|
| Akurasi | 50% |
| Precision | 71,43% |
| Recall | 33,33% |
| AUC | 0,519 |

4. *Neural Network*

Berikut ini merupakan hasil pengujian dengan menggunakan metode *Neural Network* :

TABEL X
HASIL KLASIFIKASI METODE NEURAL NETWORK

| No. | Label | Prediksi | Confidence (Tinggi) | Confidence (Rendah) |
|-----|--------|----------|---------------------|---------------------|
| 1 | Tinggi | Tinggi | 0,803 | 0,197 |
| 2 | Rendah | Rendah | 0,265 | 0,735 |
| 3 | Rendah | Rendah | 0,447 | 0,553 |
| 4 | Tinggi | Tinggi | 0,544 | 0,456 |
| 5 | Rendah | Rendah | 0,475 | 0,525 |
| 6 | Rendah | Rendah | 0,370 | 0,630 |
| 7 | Rendah | Rendah | 0,429 | 0,571 |
| 8 | Tinggi | Tinggi | 0,533 | 0,467 |
| 9 | Tinggi | Tinggi | 0,926 | 0,074 |
| 10 | Rendah | Rendah | 0,146 | 0,854 |
| 11 | Rendah | Tinggi | 0,576 | 0,424 |
| 12 | Rendah | Rendah | 0,234 | 0,766 |
| 13 | Rendah | Rendah | 0,001 | 0,999 |
| 14 | Rendah | Tinggi | 0,562 | 0,438 |
| 15 | Rendah | Tinggi | 0,550 | 0,450 |
| 16 | Tinggi | Tinggi | 0,947 | 0,053 |
| 17 | Tinggi | Rendah | 0,355 | 0,645 |
| 18 | Rendah | Rendah | 0,354 | 0,646 |
| 19 | Tinggi | Tinggi | 0,500 | 0,500 |
| 20 | Rendah | Rendah | 0,000 | 1,000 |
| 21 | Tinggi | Tinggi | 0,726 | 0,274 |
| 22 | Rendah | Rendah | 0,000 | 0,999 |
| 23 | Rendah | Rendah | 0,000 | 1,000 |
| 24 | Tinggi | Rendah | 0,372 | 0,628 |

Berdasarkan tabel diatas, dapat dilihat bahwa dengan menggunakan metode *Neural Network* jumlah prediksi yang benar adalah 19 dari 24 data, sehingga tingkat akurasi yang diperoleh adalah 79,17%. Kemudian untuk hasil parameter penilaiannya dapat dilihat pada tabel berikut :

TABEL XI
HASIL PARAMETER PENILAIAN METODE NEURAL NETWORK

| Parameter Penilaian | Skor |
|---------------------|--------|
| Akurasi | 79,17% |
| Precision | 85,71% |
| Recall | 80% |
| AUC | 0,837 |

5. *Random Forest*

Berikut ini merupakan hasil pengujian dengan menggunakan metode *Random Forest* :

TABEL XII
HASIL KLASIFIKASI METODE RANDOM FOREST

| No. | Label | Prediksi | Confidence (Tinggi) | Confidence (Rendah) |
|-----|--------|----------|---------------------|---------------------|
| 1 | Tinggi | Tinggi | 0,756 | 0,244 |
| 2 | Rendah | Tinggi | 0,692 | 0,308 |
| 3 | Rendah | Rendah | 0,342 | 0,658 |
| 4 | Tinggi | Tinggi | 0,682 | 0,318 |
| 5 | Rendah | Rendah | 0,368 | 0,632 |
| 6 | Rendah | Rendah | 0,425 | 0,575 |
| 7 | Rendah | Tinggi | 0,638 | 0,362 |
| 8 | Tinggi | Rendah | 0,305 | 0,695 |
| 9 | Tinggi | Rendah | 0,312 | 0,688 |
| 10 | Rendah | Rendah | 0,315 | 0,685 |
| 11 | Rendah | Tinggi | 0,654 | 0,346 |
| 12 | Rendah | Rendah | 0,471 | 0,529 |
| 13 | Rendah | Rendah | 0,276 | 0,724 |

| No. | Label | Prediksi | Confidence (Tinggi) | Confidence (Rendah) |
|-----|--------|----------|---------------------|---------------------|
| 14 | Rendah | Rendah | 0,282 | 0,718 |
| 15 | Rendah | Rendah | 0,320 | 0,680 |
| 16 | Tinggi | Tinggi | 0,627 | 0,373 |
| 17 | Tinggi | Rendah | 0,356 | 0,644 |
| 18 | Rendah | Tinggi | 0,693 | 0,307 |
| 19 | Tinggi | Rendah | 0,324 | 0,676 |
| 20 | Rendah | Rendah | 0,266 | 0,734 |
| 21 | Tinggi | Tinggi | 0,750 | 0,250 |
| 22 | Rendah | Rendah | 0,278 | 0,722 |
| 23 | Rendah | Rendah | 0,266 | 0,734 |
| 24 | Tinggi | Rendah | 0,283 | 0,717 |

Berdasarkan tabel diatas, dapat dilihat bahwa dengan menggunakan metode *Random Forest* jumlah prediksi yang benar adalah 15 dari 24 data, sehingga tingkat akurasi yang diperoleh adalah 62,5%. Kemudian untuk hasil parameter penilaiannya dapat dilihat pada tabel berikut :

TABEL XIII
HASIL PARAMETER PENILAIAN METODE RANDOM FOREST

| Parameter Penilaian | Skor |
|---------------------|--------|
| Akurasi | 62,50% |
| Precision | 62,50% |
| Recall | 100% |
| AUC | 0,500 |

B. Pembahasan dan Diskusi

Setelah melakukan uji klasifikasi menggunakan masing-masing metode, langkah selanjutnya adalah membandingkan hasil-hasil uji tersebut. Berikut ini adalah perbandingan hasil uji yang dilakukan :

TABEL XIV
PERBANDINGAN HASIL PARAMETER TIAP METODE

| Metode | Akurasi | Precision | Recall | AUC |
|----------------|---------|-----------|--------|-------|
| Decision Tree | 62,50% | 62,50% | 100% | 0,500 |
| K-NN | 79,17% | 85,71% | 80% | 0,833 |
| Naïve Bayes | 50% | 71,43% | 33,33% | 0,519 |
| Neural Network | 79,17% | 85,71% | 80% | 0,837 |
| Random Forest | 62,50% | 62,50% | 100% | 0,500 |

Berdasarkan hasil yang diperoleh dari tabel 14, terdapat lima metode yang digunakan. Dalam perbandingan ini, metode *Naïve Bayes* menunjukkan hasil yang sangat rendah dibandingkan dengan metode-metode lainnya. Di sisi lain, metode *K-NN* dan *Neural Network* mencapai hasil yang sangat tinggi jika dibandingkan dengan metode-metode lainnya. Meskipun demikian, dalam parameter penilaian AUC, terlihat bahwa metode *Neural Network* sedikit lebih unggul daripada metode *K-NN*. Oleh karena itu, dalam konteks klasifikasi indeks kedalaman kemiskinan untuk kabupaten/kota di Provinsi Sulawesi Selatan, disarankan untuk menggunakan metode *Neural Network*.

IV. KESIMPULAN

Berdasarkan hasil pengujian yang dilakukan terhadap lima metode klasifikasi, metode *K-NN* dan *Neural Network* menunjukkan performa tertinggi dengan tingkat akurasi 79,17%, precision 85,71%, dan recall 80%. Namun, metode *Neural Network* lebih unggul dari metode *K-NN* pada skor AUC dengan nilai 0,837. Sehingga dapat disimpulkan bahwa diantara kelima metode *Decision Tree*, *K-NN*, *Naïve Bayes*, *Neural Network*, dan *Random Forest*, metode *Neural Network* dapat dijadikan sebagai metode untuk melakukan klasifikasi indeks kedalaman kemiskinan kabupaten/kota Provinsi Sulawesi Selatan.

REFERENSI

- [1] A. Bahauddin, A. Fatmawati, and P. F. Sari, "Analisis Clustering Provinsi Di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means," *Jurnal Manajemen Informatika & Sistem Informasi*, vol. 4, pp. 1–8, 2021.
- [2] D. Vita Ferezagia, "Analisis Tingkat Kemiskinan di Indonesia," 2018.
- [3] P. : Amset, I. Batusangkar, I. B. Press, N. Putu, N. Hendayanti, and M. Nurhidayati, "Regresi Logistik Biner dalam Penentuan Ketepatan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi- Provinsi di Indonesia," *Sainstek : Jurnal Sains dan Teknologi*, vol. 12, no. 2, pp. 63–70, Dec. 2020, Accessed: Jun. 27, 2023. [Online]. Available: <https://ojs.iainbatusangkar.ac.id/ojs/index.php/sainstek/article/view/2483>
- [4] D. F. Ristianti and S. Suparman, "Komparasi Algoritma Klasifikasi Data Mining," *Science, Technology, Engineering, Economics, Education, and Mathematics*, vol. 1, no. 1, Feb. 2020, Accessed: Jun. 27, 2023. [Online]. Available: <http://seminar.uad.ac.id/index.php/STEEEM/article/view/2877>
- [5] N. M. Sukarno, P. W. Wirawan, and S. Adhy, "Perancangan dan Implementasi Jaringan Syaraf Tiruan Backpropogation Untuk Mendiagnosa Penyakit Kulit," *Jurnal Masyarakat Informatika*, vol. 5, no. 10, 2014.
- [6] H. Sastypratiwi, H. Muhandi, and J. H. Hadari Nawawi, "Uji Komparasi Algoritma Naïve Bayes dan Decision Tree Classification Menggunakan Covid-19 Dataset," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 8, no. 1, pp. 1–6, Apr. 2022, Accessed: Jul. 12, 2023. [Online]. Available: <https://jurnal.untan.ac.id/index.php/jepin/article/view/49841>
- [7] M. A. Lusiandro, S. M. Nasution, and C. Setianingsih, "Implementation of the advanced traffic management system using k-nearest neighbor algorithm," *2020 International Conference on Information Technology Systems and Innovation, ICITSI 2020 - Proceedings*, pp. 149–154, Oct. 2020, doi: 10.1109/ICITSI50517.2020.9264952.
- [8] G. Putri, I. Rani, A. Aziz, M. Priyono, and T. Sulistyono, "IMPLEMENTASI EUCLIDEAN DAN CHEBYSHEV DISTANCE PADA K-MEDOIDS CLUSTERING," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 710–715, Oct. 2022, doi: 10.36040/JATI.V6I2.5443.
- [9] R. Y. Hayuningtyas, "Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita," *Jurnal Informatika*, vol. 6, no. 1, pp. 18–22, Apr. 2019, Accessed: Jul. 02, 2023. [Online]. Available: <https://ejournal.bsi.ac.id/ejournal/index.php/ji/article/view/4685>
- [10] R. Amegia Saputra, A. Ruslan Taufik, L. Saumi Ramdhani, R. Oktapiani, E. Marsusanti, and A. BSI Sukabumi, "Sistem Pendukung Keputusan Dalam Menentukan Metode Kontrasepsi Menggunakan Algoritma Naive Bayes," *SNIT 2018*, vol. 1, no. 1, pp. 106–111, Jul. 2018, Accessed: Jul. 02, 2023. [Online]. Available: <https://seminar.bsi.ac.id/snit/index.php/snit-2018/article/view/31>
- [11] M. F. Andrijasa and M. Mistianingsih, "Penerapan Jaringan Syaraf Tiruan Untuk Memprediksi Jumlah Pengangguran di

- Provinsi Kalimantan Timur Dengan Menggunakan Algoritma Pembelajaran Backpropagation,” *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, vol. 5, no. 1, pp. 50–54, Jun. 2016, Accessed: Jun. 27, 2023. [Online]. Available: <https://e-journals.unmul.ac.id/index.php/JIM/article/view/44>
- [12] A. Jumarwanto, R. Hartanto, and D. Prastiyanto, “APLIKASI JARINGAN SARAF TIRUAN BACKPROPAGATION UNTUK MEMREDIKSI PENYAKIT THT DI RUMAH SAKIT MARDI RAHAYU KUDUS,” *Jurnal Teknik Elektro*, vol. 1, no. 1, p. 11, 2009, Accessed: Jun. 27, 2023. [Online]. Available: <https://journal.unnes.ac.id/nju/index.php/jte/article/view/1601>
- [13] S. Devella, Y. Yohannes, and F. N. Rahmawati, “Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 7, no. 2, pp. 310–320, Aug. 2020, doi: 10.35957/JATISI.V7I2.289.
- [14] C. Leordeanu, C.-O. Truic, and alin Adrian LEORDEANU, “Classification of an Imbalanced Data Set using Decision Tree Algorithms CLASSIFICATION OF AN IMBALANCED DATA SET USING DECISION TREE ALGORITHMS,” *U.P.B. Sci. Bull., Series C*, vol. 79, 2017, [Online]. Available: <https://www.researchgate.net/publication/322103560>
- [15] L. Fu, P. Liang, X. Li, and C. Yang, “A machine learning based ensemble method for automatic multiclass classification of decisions,” in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Jun. 2021, pp. 40–49. doi: 10.1145/3463274.3463325.
- [16] M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, “Diabetes prediction using ensembling of different machine learning classifiers,” *IEEE Access*, vol. 8, pp. 76516–76531, 2020, doi: 10.1109/ACCESS.2020.2989857.