# KLASIFIKASI PERSENTASE KEMISKINAN DI JAWA BARAT MENGGUNAKAN DATA MINING ALGORITMA K-NEAREST NEIGHBOR (KNN)

**Adi Duwo Jiwo Saputro, Adam Darmawan, Betha Nurina Sari**
Program Studi Teknik Informatika S1, Fakultas Ilmu Komputer
Universitas Singaperbangsa Karawang, Jl. HS. Ronggo Waluyo, Telukjambe Timur, Karawang, Indonesia
*1910631170060@student.unsika.ac.id*

**ABSTRAK**

Salah satu masalah utama yang masih dihadapi banyak negara, termasuk Indonesia, adalah kemiskinan. Keadaan dimana seseorang dipandang tidak mampu secara ekonomi untuk memenuhi kebutuhan. Data berdasarkan dari Badan Pusat Statistik (BPS) pada Provinsi Jawa Barat, persentase kemiskinan yang terjadi di Jawa Barat pada tahun 2016 – 2021, BPS mencatat jumlah persentase kemiskinan yang terjadi di Jawa Barat pada tahun 2016 = 8.95, 2017 = 8.71, 2018 = 7.45, 2019 = 6.91, 2020 = 7.88, 2021 = 8.40. Penelitian ini bertujuan untuk klasifikasi tingkat persentase kemiskinan di Jawa Barat pada tahun 2016 – 2021 menggunakan algoritma *K-Nearest Neighbor* (KNN). Cara kerja *K-Nearest Neighbor* yang berfungsi membedakan kelas data atau konsep. Hasil penelitian ini menunjukkan persentase kemiskinan 27 kabupaten/kota yang ada di Jawa Barat pada nilai k=3 dengan tingkat akurasi sebesar 96.30%, recall 100% dan presisi sebesar 92,31%. Dari penelitian yang dilakuakn Metode *K-Nearest Neighbor* (KNN) berhasil mengklasifikasikan dengan baik.

***Kata kunci:*** *Klasifikasi, Data Mining, Kemiskinan, K-Nearest Neighbor*

## 1. PENDAHULUAN

Kehidupan yang mempengaruhi banyak hal yang berbeda, termasuk hak atas pangan yang layak, perawatan kesehatan, pendidikan, dan pekerjaan, kemiskinan adalah suatu kondisi di mana orang dianggap tidak kompeten. Penduduk yang berpenghasilan rendah, tidak berpenghasilan tetap, dan penduduk yang tidak memiliki penghasilan sama sekali termasuk penduduk di bawah garis kemiskinan. Kemiskinan sering dikaitkan dengan kurangnya akses ke sumber daya pada tingkat ekonomi, sosial, budaya, dan politik, serta kurangnya partisipasi dalam masyarakat, kemiskinan secara teoritis dapat dilihat sebagai masalah multifaset. Ketidakmampuan memenuhi tuntutan yang tidak terkait langsung dengan pendapatan (non-income factors), seperti akses terhadap kebutuhan pokok, kesehatan, pendidikan, dan air bersih, oleh karena itu kemiskinan memiliki makna yang lebih dalam.

Permasalah pembangunan ekonomi di Indonesia menjadi masalah utama adalah kemiskinan. Badan Pusat Statistik (BPS) mencatat jumlah persentase kemiskinan di Indonesia di Provinsi Jawa Barat pada tahun 2016 – 2021. BPS mencatat jumlah persentase kemiskinan yang terjadi di Jawa Barat pada tahun 2016 = 8.95, 2017 = 8.71, 2018 = 7.45, 2019 = 6.91, 2020 = 7.88, 2021 = 8.40.

Istilah data mining memiliki berbagai sinonim yaitu "knowledge discovery" dan "pattern recognition". Istilah istilah tersebut memiliki makna dan ketepatan ketepatan pada setiap katanya. Karena tujuan mendasar dari penambangan data adalah untuk mengekstraksi pengetahuan yang masih tersembunyi dalam data, penemuan pengetahuan adalah istilah yang sangat relevan [1]. Penggunaan kata “pattern recognition” atau "pengenalan pola" juga berlaku karena informasi yang perlu diekstraksi benar-benar berbentuk pola, yang mungkin masih memerlukan ekstraksi dari beberapa potongan data yang masuk. Ada enam fungsi dalam data mining, fungsi asosiasi (association), fungsi klasifikasi (classification), fungsi pengelompokan (clustering), fungsi estimasi (estimation), dan fungsi prediksi (prediction).

Salah satu teknik untuk mengklasifikasikan objek berdasarkan data pembelajaran yang terdekat dengan objek yang diuji adalah algoritma *K-Nearest Neighbor* (KNN) [2]. Algoritma KNN memiliki sistemastis yang sederhana, permodelan yang dilakukan berdasarkan jarak terpendek. Data latih dalam KNN yang memiliki jumlah kerabat terbanyak di dalam rentang nilai terpilih akan dikelompokkan dengan hasil perhitungan. Persamaan euclidean distance digunakan untuk menghitung jarak antara *data training* dan *data testing* [3]. Beberapa penelitian sebelumnya tentang klasifikasi menggunakan metode *K-Nearest Neighbor* (KNN), diantaranya penelitian yang dilakukan oleh Fitra Kurnia dkk (2019). Metode Black Box dan UAT digunakan untuk menguji aplikasi yang telah dirancang. Kuesioner skala Likert digunakan untuk tes pada metode UAT, dan temuan menunjukkan nilai 75%, yang termasuk dalam kelompok sangat baik. Dengan rasio 90:10 pada k = 5, k = 7, dan k = 9, uji KNN untuk mengkategorikan rumah tangga miskin dalam 100 sampel data memiliki nilai akurasi yang paling baik sebesar 90% [4]. Haidah Putri dkk (2021) menggunakan metode Naïve Bayes dan *K-Nearest Neighbor* dengan hasil performa dari metode algoritma *Naïve Bayes* memperoleh hasil nilai akurasi sebesar 99.89% dan performa metode algoritma *K-Nearest Neighbor* 66.46% [5].

Pendekatan KNN dievaluasi dalam penelitian selanjutnya oleh Clariza Adelina Rachma (2022) menggunakan 38 data uji dan uji validasi silang 10 kali lipat dengan nilai parameter mulai dari k = 1 hingga k = 10. Temuan penelitian berupa perbandingan Hasil akurasi algoritma KNN berdasarkan nilai parameter

yang digunakan, k = 1 hingga k = 10. Berdasarkan hasil penelitian, nilai parameter k = 1 dan k = 2 memiliki tingkat akurasi tertinggi (76,67%). [6]. Penelitian juga dilakukan oleh Akbar Anung (2022), dengan hasil nilai akurasi pada metode *Naïve Bayes* sebesar 92.56%, *K-Nearest Neighbor* (KNN) sebesar 92.36, dan *Support Vector Machine* (SVM) sebesar 90.4% [7].

Penelitian (4-7) telah membahas penerapan algoritma KNN untuk permasalahan kemiskinan di beberapa provinsi tetapi belum membahas terkait presentase kemiskinan di Provinsi Jawa Barat selama beberapa tahun. Penelitian ini memodelkan presentase tingkat kemiskinan pada 27 kabupaten/kota di Provinsi Jawa Barat yang berguna untuk menentukan prioritas daerah pada program pengentasan kemiskinan.

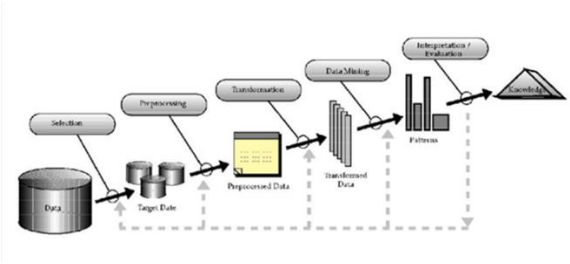
## 2. TINJAUAN PUSTAKA

### 2.1. Kemiskinan

Kemiskinan adalah salah satu masalah besar yang dihadapi pemerintah. Kemiskinan sering berulang di beberapa negara, terutama negara berkembang seperti Indonesia. Kemiskinan yang semakin meningkat di Indonesia dan konsekuensinya, yang mencakup masalah ekonomi, sosial, dan instabilitas politik, menjadikannya masalah yang sangat penting [12]. Kemiskinan bukan hanya masalah ekonomi semata, tetapi juga melibatkan aspek sosial, kesehatan, pendidikan, dan lainnya. Menurut Badan Pusat Statistik, ketika seseorang tidak dapat memenuhi kebutuhan hidup dasar mereka, disebut kemiskinan. Ketika seseorang berada di bawah garis nilai standar kebutuhan minimum, baik untuk makanan maupun non-makanan, itu dianggap kemiskinan. Garis ini disebut sebagai "garis kemiskinan" atau "batas kemiskinan" [11].

### 2.2. *Knowladge Discovery in Databases (KDD)*

Pendekatan sistematis untuk mengidentifikasi pengetahuan baru dari data *Knowladge Discovery in Databases* (KDD) melibatkan transformasi data mentah menjadi informasi yang berarti dan berguna. KDD menggabungkan ide dari berbagai bidang, seperti statistik, ilmu komputer, dan ilmu informasi. Salah satu bagian dari proses *Knowledge Discovery in Databases* (KDD) adalah *data mining*. Menurut Widaningsih (2019) KDD adalah proses mendapatkan informasi yang lebih berharga, lebih mudah dipahami, dan lebih baru dari penyimpanan data yang besar dan kompleks dengan menggabungkan ilmu lain [8]. Proses KDD memiliki beberapa tahapan yaitu pemilihan data, prosesing data, transformasi data, data mining, dan evaluasi. Tahapan tersebut bersifat sistematis atau harus dilakukan secara berurutan, dengan tujuan akhir adalah untuk memperoleh data yang dapat dimengerti




Gambar 1. Tahapan KDD

### 2.3. *Data Mining*

*Data mining* adalah suatu permodelan teknologi dengan menggabungkan metode analisisi dengan algoritma yang lebih modern dalam memproses data dengan volume besar [4]. *Data mining* merupakan teknik mengolah data berskala besar dengan tujuan memperoleh pengetahuan dari data tersebut [8]. Proses penambangan data menganalisis dan mengidentifikasi data menggunakan metode statistik, matematika, kecerdasan buatan, dan *mechine learning* untuk memperoleh informasi dan pengetahuan yang relevan dari berbagai data. Pada penelitian ini dilakukan pemilihan data yang relevan untuk proses klasifikasi dengan menerapkan algoritma *K-Nearest Neighbor* (K-NN).

### 2.4. *Algoritma K-Nearest Neighbor (K-NN)*

Algoritma K-NN atau K-Nearest Neighbor adalah metode klasifikasi pada data mining berdasarkan jarak terdekat dengan data uji [3]. Tahapan metode K-NN ini dengan mencari pola dan informasi yang menarik dari data dengan atribut k, yaitu jumlah tetangga yang dijadikan acuan pada data testing terhadap data training. Algoritma K-NN berjalan dengan menentukan data training sebagai index terbanyak sebagai bahan perhitungan dengan data testing [10]. Proses metode K-NN sebagai berikut:

a. Menyiapkan data presentase kemiskinan kabupaten/kota Provinsi Jawa Barat yang menjadi data training dan data testing
b. Menentukan nilai k
c. Menghitung jarak (D) dari data testing ke data training menggunakan rumus Euclidean Distance

$$Eucidean\ Distance = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2} \quad (1)$$

d. Mengurutkan data (D) menurut rangking terkecil
e. Klasifikasi data testing mayoritas

### 2.5. Evaluasi Data

Tahap evaluasi dilakukan menggunakan *confusion matrix*, hasil klasifikasi kemudian dinilai terhadap data yang diproses untuk menentukan efektivitas prosedur [9]. *Confusion matrix* adalah suatu teknik evaluasi yang digunakan untuk mengevaluasi kinerja model klasifikasi. *Confusion matrix* menggambarkan perbandingan antara kelas aktual dengan kelas yang diprediksi oleh model klasifikasi. Hasil evaluasi metode ini yang nantinya digunakan untuk mendapatkan kesimpulan dari hasil penelitian

yang telah diselesaikan. Parameter *confusion matrik* dapat dilihat pada tabel 1.

Table 1. *Confusion Matrix*

| | | *True Class* | |
|---|---|---|---|
| | | Positif | Negatif |
| *Predicted Class* | Positif | TP | FP |
| | Negatif | FN | TN |

Dengan keterangan sebagai berikut:
TP (True Positive): jumlah data pada poin label bernilai yes dengan nilai yang diidentifikasi benar
TN (True Negative): jumlah data pada poin label bernilai no dengan nilai yang diidentifikasi salah
FP (False Positive): jumlah data pada poin label bernilai yes dengan nilai yang diidentifikasi salah
FN (False Negative): jumlah data pada poin label bernilai no dengan nilai yang diidentifikasi benar

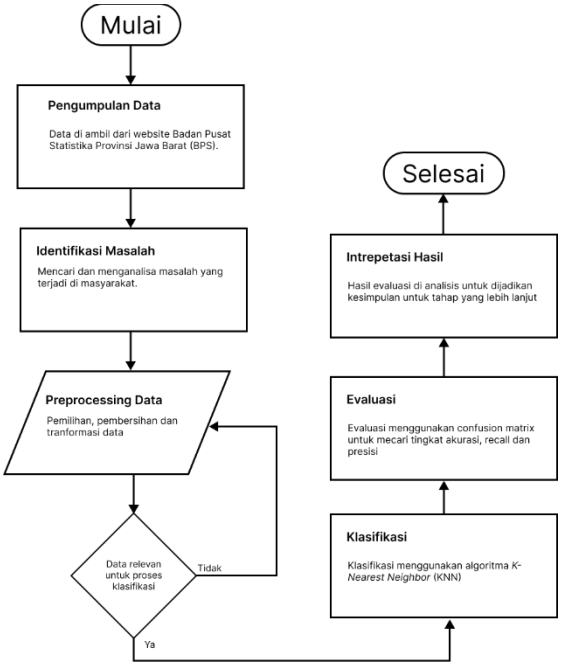
Setiap nilai yang berada pada *confusion matrix* dapat dihitung yang kemudian dijadikan sebagai bahan evaluasi kinerja model klasifikasi. Berikut evaluasi yang dapat dilakukan nilai pada tabel confusion matrix:

Table 2. Nilai Evaluasi Kinerja Klasifikasi

| No | Nilai | Perhitungan | |
|---|---|---|---|
| 1 | *Accuracy* | $\frac{(TP+TN)}{(TP+FP+FN+TN)}$ | (2) |
| 2 | *Misclassification Rate/Error Rate* | $\frac{(FP+FN)}{(TP+FP+FN+TN)}$ | (3) |
| 3 | *Recall* | $\frac{TP}{TP+FN}$ | (4) |
| 4 | *Precision* | $\frac{TP}{TP+FP}$ | (5) |
| 5 | *F-Measure* | $\frac{2 \times R \times P}{R+P}$ | (6) |

## 3. METODE PENELITIAN

Indentifikasi masalah perlu dilakukan untuk menganalisa masalah yang terjadi di Masyarakat, merumuskan masalah yang ada dan menentukan manfat dan tujuan penelitian dilakukan. Tahapan penelitian dapat dilihat pada alur penelitian dibawah ini:




Gambar 2. Flowcart Metode Penelitian

a. Mengidentifikasi Masalah
Tahapan yang dilakukan dalam penelitian untuk menemukan masalah yang terjadi pada masyarakat. Pada tahapan penulis mencari serta manganalisa masalah yang terjadi pada masyarakat. Dalam hal ini masalah yang terjadi adalah terkait kemiskinan yang terjadi di Provinsi Jawa Barat
b. Pengumpulan Data
Pengumpulan data dilakukan untuk mendapatkan berbagai insight dari berbagai sumber. Buku, jurnal, artikel, data dilapanagan serta studi literatue yang relevan dengan penelitian membantu dalam penyelesaian masalah yang akan dilakukan. Tujuan pengumpulan data adalah untuk memerdalam sumber masalah dan memperkuat dengan dasar teori teori yang ada.
c. Preprocessing data
Data yang terkumpul kemudian akan diolah dan dipersiapkan untuk dilakukan analisis. Preprocessing data yang dilakukan meliputi pemilihan data, pembersihan data, dan transformasi data sehingga relevan dengan perhitungan. Pada tahapan ini dilakukan labeling pada setiap data.
d. Implementasi Metode *K-Nearest Neighbor* (KNN)
Setelah melakukan tahapan preprocessing data selanjutnya adalah pengimplementasian metode data mining algoritma KNN. Pada metode ini akan dilakukan perhitungan pada setiap parameter yang ada pada data dilapangan. Metode KNN menggunakan pendekatan *Knowladge Discovery in Databases* (KDD) dengn tahapan yang sistematis.
e. Evaluasi Kerja
Setelah dilakukan perhitungan menggunakan metode data mining algoritma KNN selanjut nya dilakukan evaluasi kerja dimana pada tahapan ini

dilakukan evaluasi terhadap hasil yang didapat dari tahapan demi tahapan yang dilakukan. Pada tahapan ini evaluasi yang dilakukan menggunakan tabel *confusion matrix* dengan nilai evaluasi yang di hitung adalah tingkat akurasi, presisi dan recall, nilai tersebut berfungsi untuk mendapatkan tingkat peforma yang dihasilkan oleh algoritma KNN. Hasil yang didapat akan ditarik kesimpulan dari tahapan yang dilakukan dan berupa *knowledge* pada data mining.

f. Interpretasi hasil
Hasil penelitian akan diinterpretasikan dan disimpulkan untuk menghasilkan kesimpulan menunjukkan tingkat presentasi kemiskinan di Jawa Barat.

## 4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data presentase kemisikinan di provisi Jawa Barat yang diambil dari Badan Pusat Statistika Provinsi Jawa Barat (BPS) dengan data sebanyak 27 atribut yang digunakan adalah Nama Kabupaten/Kota dan tahun dari 2016 hingga 2021. Data tersebut kemudian ditransformasikan menjadi data yang disiapkan untuk diklasifikasikan menggunakan algoritma *K-Nearest Neighbor*. Data presentase kemiskinan Provinsi Jawa Barat dapat dilihat pada tabel 2.

Table 3. Presentase Kemiskinan Provinsi Jawa Barat

| Kabupaten/Kota | Tahun | | | | | |
|---|---|---|---|---|---|---|
| | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 |
| Bogor | 8,83 | 8,57 | 7,14 | 6,66 | 7,69 | 8,13 |
| Sukabumi | 8,13 | 8,4 | 6,76 | 6,22 | 7,09 | 7,7 |
| Cianjur | 11,62 | 11,41 | 9,81 | 9,15 | 10,36 | 11,18 |
| … | … | … | … | … | … | … |
| … | … | … | … | … | … | … |
| Kota Cimahi | 5,92 | 5,76 | 4,94 | 4,36 | 5,11 | 5,35 |
| Kota Tasikmalaya | 15,6 | 14,8 | 12,71 | 11,6 | 12,97 | 13,13 |
| Kota Banjar | 7,01 | 7,06 | 5,7 | 5,5 | 6,09 | 7,11 |

Data yang diperoleh kemudian diolah menggunakan data mining sebelum masuk ke dalam proses klasifikasi. Pada tahap transformasi data pada proses *data mining*, atribut-atribut ditransformasikan ke dalam format yang sesuai dengan perangkat lunak yang akan dijalankan. Tranformasi yang dilakukan pada data index presentase kemiskinan dibuat menjadi label/kelas. Dalam mendapatkan kelas dapat dilakukan dengan mencari rata rata presentase kemiskinan di Provinsi Jawa Barat. Adapun perhitungannya rata rata sebagai berikut:

$$Rata - rata = \frac{Total\ presentase\ kemiskinan}{Banyak\ data\ tahun} \quad (7)$$

$$Rata - rata = \frac{48.3}{6} = 8{,}05$$

Setelah memperoleh nilai presentase kemiskinan rata-rata, label atau klasifikasi dapat dibuat seperti berikut:

Table 4. Penentuan Label

| Label/Kelas | Presentase Kemiskinan (PK) |
|---|---|
| Kelas 0 | PK Rata-rata < 8,05 |
| Kelas 1 | PK Rata-rata >= 8,05 |

Tabel 3 menunjukan bahwa nilai rata rata suatu kota kurang dari 8,05 artinya data tersebut masuk ke dalam kelas 0. Dan apabila data suatu kota lebih dari sama dengan 8,05 data tersebut masuk ke dalam kelas 1. Sehingga diperoleh hasil pada tabel 4 dengan penambahan index untuk label/kelas dimana terdapat 12 kabupaten/kota termasuk ke dalam kelas 0 dan 15 kabupaten/kota masuk ke dalam kelas 1. Dengan memisahkan data yang telah disiapkan menjadi *data training* dan *data testing*, dilakukan prosedur klasifikasi algoritma *K-Nearest Neighbor* (K-NN).

Table 5. Data Training Presentase Kemiskinan

| Kabupaten/Kota | Tahun | | | | | | Kelas |
|---|---|---|---|---|---|---|---|
| | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | |
| Bogor | 8,83 | 8,57 | 7,14 | 6,66 | 7,69 | 8,13 | Kelas 0 |
| Sukabumi | 8,13 | 8,4 | 6,76 | 6,22 | 7,09 | 7,7 | Kelas 0 |
| Cianjur | 11,62 | 11,41 | 9,81 | 9,15 | 10,36 | 11,18 | Kelas 1 |
| … | … | … | … | … | … | … | …. |
| … | … | … | … | … | … | … | … |
| Kota Cimahi | 5,92 | 5,76 | 4,94 | 4,36 | 5,11 | 5,35 | Kelas 0 |
| Kota Tasikmalaya | 15,6 | 14,8 | 12,71 | 11,6 | 12,97 | 13,13 | Kelas 1 |
| Kota Banjar | 7,01 | 7,06 | 5,7 | 5,5 | 6,09 | 7,11 | Kelas 0 |

Table 6. Data Testing Presentase Kemiskinan

| Kabupaten/Kota | Tahun | | | | | | Kelas |
|---|---|---|---|---|---|---|---|
| | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | |
| Karawang | 10,07 | 10,25 | 8,06 | 7,39 | 8,26 | 8,95 | ? |

Setelah menentukan *data training* dan *data testing* tahap selanjutnya menentukan nilai k. Pada penelitian ini menggunakan k = 3 sehingga yang menjadi parameter pada hasil nanti adalah jarak 3 tetangga terdekat. Kemudian menghitung jarak yang dapat dilakukan dengan menggunakan *Euclidean Distance* menggunakan rumus di atas ke setiap *data training* ke *data testing*. Dari poses itu kemudian akan mendapatkan jarak dari masing masing *data training* seperti pada tabel 6. Untuk memperoleh jarak terdekat sesuai dengan tahapan algoritma K-NN maka data jarak tersebut diurutkan sesuai dengan rangking terkecil seperti pada tabel 7.

Table 7. Tabel Jarak

| Data Ke | Nilai Euclidean | Rangking | Label |
|---|---|---|---|
| 1 | 2,595 | 5 | Kelas 0 |
| 2 | 3,630 | 9 | Kelas 0 |
| 3 | 4,392 | 12 | Kelas 1 |
| … | … | … | … |
| … | … | … | … |
| 25 | 8,898 | 19 | Kelas 0 |
| 26 | 11,414 | 24 | Kelas 1 |
| 27 | 6,064 | 16 | Kelas 0 |

Tabel 8 Tabel Jarak Urut Sesuai Rangking

| Data Ke | Nilai Euclidean | Rangking | Label |
|---|---|---|---|
| 18 | 1,103 | 1 | Kelas 1 |
| 14 | 1,622 | 2 | Kelas 1 |
| 13 | 2,099 | 3 | Kelas 1 |
| … | … | … | … |
| … | … | … | … |
| 26 | 11,414 | 24 | Kelas 1 |
| 21 | 12,061 | 25 | Kelas 0 |
| 24 | 16,125 | 26 | Kelas 0 |

Setelah itu, setiap kelas diperiksa untuk memastikan bahwa inputan nilai k sesuai dengan tetangga terdekatnya. Pada penelitian ini, nilai k=3, sehingga data yang dikumpulkan berada di urutan ketiga teratas, termasuk data ke-18 dengan label kelas 1, data ke-14 dengan label kelas 1, dan data ke-13 dengan label kelas 1. Dari hasil tersebut diperoleh bahwa presentase kemiskinan kabupaten/kota Karawang masuk ke dalam kelas 1 artinya lebih dari rata rata.

Hasil klasifikasi yang telah dilakukan kemudian di evaluasi dengan menghitung nilai akurasi, recall dan presisinya menggunakan *confusion matrix* dengan tujuan untuk melihat performa dari algoritma K-NN ini dalam mengkasifikasikan presentase kemiskinan di Provinsi Jawa Barat. Data sebanyak 27 data kabupaten/kota menjadi bahan pengujian yang akan dilakukan. Hasil pengujian ini diperlihatkan dalam bentuk *confusion matrix* pada tabel 8.

Tabel 9 Confusion Matrix Hasil Klasifikasi

| | | *True Class* | |
|---|---|---|---|
| | | Positif | Negatif |
| *Predicted Class* | Positif | 12 | 1 |
| | Negatif | 0 | 14 |

Nilai akurasi yang didapatkan pada algoritma K-NN ini sebesar 96,3%, nilai recall sebesar 100% dan nilai presisi sebesar 92,31% berdasarkan nilai parameter yang dipakai adalah k=3. Perolehan nilai akurasi, recall dan presisi, perhitungan dapat dilihat sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$= \frac{12+14}{12+14+1+0} \times 100\%$$

$$= 96{,}3\%$$

$$Presisi = \frac{TP}{TP+FN} \quad (4)$$

$$= \frac{12}{12+0} \times 100\%$$

$$= 100\%$$

$$Presisi = \frac{TP}{TP+FP} \quad (5)$$

$$= \frac{12}{12+1} \times 100\%$$

$$= 92{,}31\%$$

## 5. KESIMPULAN DAN SARAN

Dari hasil penelitian di atas dapat disimpulkan bahwa presentase kemiskinan di Provinsi Jawa Barat dengan presentase terendah di setiap tahunnya ada pada kota Depok dan presentase kemiskinan tertinggi di setiap tahunnya pada kota Tasikmalaya. Terdapat 12 kabupaten/kota masuk ke dalam kelas 0 (kurang dari rata rata) dan terdapat 15 kabupaten/kota masuk ke dalam kelas 1 (di atas rata rata). Hasil klasifikasi yang dilakukan berdasarkan algoritma *K-Nearest Neighbor* berhasil dilakukan untuk klasifikasi presentase kemiskinan di Provinsi Jawa Barat dengan baik. Performa yang dihasilkan algoritma KNN memperoleh nilai akurasi sebesar 96,30%, nilai recall sebesar 100% dan nilai presisi sebesar 92,31% dengan parameter k = 3. Saran-saran untuk penelitian lebih lanjut penelitian ini sangat bisa dikembangkan menggunakan metodologi yang berbeda hingga mendapatkan hasil yang lebih baik, Seperti D*ecision Tree, k-Means, Naïve Bayes* dan lain lain.

## DAFTAR PUSTAKA

[1] Sani Susanto, Ph.D., Dedy Suryadi, S.T., M.S, 2010, PENGANTAR DATA MINING *Menggali Pengetahuan dari Bongkahan data*, C.V ANDI OFFSET, Yogyakarta.

[2] Harun, R., Pelangi, K. C., & Lasena, Y. (2020). Penerapan Data Mining untuk Menentukan Potensi Hujan Harian dengan Menggunakan Algoritma K Nearest Neighbor (KNN). *Jurnal Manajemen Informatika dan Sistem Informasi*, vol 3, 8-15.

[3] Baharuddin, M. M., Azis, H., & Hasanuddin, T. (2019). Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca. *ILKOM Jurnal Ilmiah*, vol 11, hal 269-274.

[4] Kurnia, F., Kurniawan, J., & Fahmi, I. (2019). Klasifikasi Keluarga Miskin Menggunakan Metode K-Nearest Neighbor Berbasis Euclidean Distance. *In Seminar Nasional Teknologi Informasi, Komunikasi dan Industri* (SNTIKI), Vol. 11, hal 230-239.

[5] Yusoff, M, Rahman, S.,A., Mutalib, S., and Mohammed, A. , 2006, Diagnosing Application

Development for Skin Disease Using Backpropagation Neural Network Technique, *Journal of Information Technology*, vol 18, hal 152-159.

[6] Rachma, C. A. (2022). *Implementasi Algoritma K-Neraest Neighbor dalam penentuan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi Jawa Timur*, Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim, Malang.

[7] Anung, A. (2022). *Analisa Perbandingan Algoritma Naïve Bayes, K-Nearest Neighbor (KNN) dan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Kemiskinan di Jawa Barat*, Doctoral dissertation, Universitas Mercu Buana, Jakarta.

[8] Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4, 5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, vol 13, hal 16-25.

[9] Akhmad, M. R. (2022). Implementasi K-Nearest Neighbor Dalam Memprediksi Keterlambatan Pembayaran Biaya Kuliah di Perguruan Tinggi. Progresif: *Jurnal Ilmiah Komputer*, vol 18, hal 185-192.

[10] Astuti, F. D., & Guntara, M. (2018). Analisis Performa Algoritma K-NN Dan C4. 5 Pada Klasifikasi Data Penduduk Miskin. *J Rekayasa Teknol Inf*, vol 2, Hal 135-142.

[11] Ferezagia, D.V., 2018. Analisis tingkat kemiskinan di Indonesia. *Jurnal Sosial Humaniora Terapan*, 1(1), p.1.

[12] Pratama, Y.C., 2014. Analisis faktor-faktor yang mempengaruhi kemiskinan di Indonesia.