

KLASIFIKASI KABUPATEN/KOTA DI INDONESIA BERDASARKAN TINGKAT KEDALAMAN DAN KEPARAHAN KEMISKINAN MENGGUNAKAN *NAÏVE BAYES* *CLASSIFIER* DAN *K-NEAREST NEIGHBOR*

Erma Shofi Utami¹, Yudi Setyawan^{2*}, Noeryanti³

Jurusan Statistika, Institut Sains dan Teknologi AKPRIND Yogyakarta
email: ¹ermashofi2509@gmail.com, ²setyawan@akprind.ac.id, ³snoeryanti@akprind.ac.id
*Corresponding Author: setyawan@akprind.ac.id

ABSTRACT

Poverty is a complex problem, because it involves many other fields. Indonesia as a developing country with 514 districts/cities so far has not been able to solve the problem of poverty. The poverty level can be seen from the value of the poverty depth index and the poverty severity index. It is necessary to classify it, to find out which districts/cities have a high or low poverty rate seen from the two index values. The poverty rate can also be influenced by several factors such as the number of poor people, the percentage of poor people, adjusted per capita expenditure, and the average length of schooling. In this study, the classification is carried out using the Naïve Bayes Classifier and K-nearest Neighbor methods. Both methods are carried out by dividing the training data and testing data by 60%:40%, 70%:30%, and 80%:20%. In addition, samples were randomized using the functions *set.seed(6)*, *set.seed(99)*, and *set.seed(123)*. In K-nearest Neighbor using the number of neighbors $k = 3$, $k = 5$, $k = 7$, and $k = 9$ with Euclid distance and Manhattan distance. The results of the analysis with both methods obtained that the Naïve Bayes Classifier method is better used than the K-nearest Neighbor. The results of the highest average accuracy on the poverty depth index and poverty severity index occur in the data splitting of 60%:40% with an average accuracy of 89.47% and 87.80%, respectively.

Keywords: K-nearest Neighbor, Naïve Bayes Classifier, Poverty Depth Level, Poverty Severity Level

INTISARI

Kemiskinan merupakan masalah yang kompleks, karena menyangkut berbagai bidang lainnya. Indonesia sebagai negara berkembang dengan 514 kabupaten/kota hingga saat ini belum dapat menyelesaikan masalah kemiskinan. Tingkat kemiskinan dapat dilihat dari nilai indeks kedalaman kemiskinan dan indeks keparahan kemiskinan. Hal tersebut perlu dilakukan klasifikasi, untuk mengetahui kabupaten/kota yang memiliki tingkat kemiskinan tinggi atau rendah dilihat dari kedua nilai indeks. Tingkat kemiskinan juga dapat dipengaruhi oleh beberapa faktor seperti jumlah penduduk miskin, persentase penduduk miskin, pengeluaran per kapita disesuaikan, serta rata-rata lama sekolah. Pada penelitian ini, klasifikasi dilakukan dengan metode Naïve Bayes Classifier dan K-nearest Neighbor. Kedua metode dilakukan dengan membagi data training dan data testing sebesar 60%:40%, 70%:30%, dan 80%:20%. Selain itu, dilakukan pengacakan sampel dengan fungsi *set.seed(6)*, *set.seed(99)*, serta *set.seed(123)*. Pada K-nearest Neighbor menggunakan jumlah ketetanggaan $k = 3$, $k = 5$, $k = 7$, dan $k = 9$ dengan jarak Euclid dan jarak Manhattan. Hasil analisis dengan kedua metode diperoleh metode Naïve Bayes Classifier lebih baik digunakan dibandingkan dengan K-nearest Neighbor. Hasil rata-rata akurasi tertinggi pada indeks kedalaman kemiskinan dan indeks keparahan kemiskinan terjadi pada pembagian data 60%:40% dengan rata-rata akurasi berturut-turut sebesar 89,47% dan 87,80%.

Kata Kunci: K-nearest Neighbor, Naïve Bayes Classifier, Tingkat Kedalaman Kemiskinan, Tingkat Keparahannya Kemiskinan

1. PENDAHULUAN

Kemiskinan menjadi masalah yang banyak dihadapi oleh negara berkembang salah satunya Indonesia. Kemiskinan di Indonesia hingga saat ini belum dapat diselesaikan, karena merupakan masalah kompleks yang tidak dapat dipisahkan dari masalah lainnya seperti masalah pendidikan, pengangguran, kesehatan dan masalah-masalah lainnya yang secara eksplisit mempengaruhi kemiskinan serta bersifat multidimensional. Tingkat kemiskinan dapat dilihat dari nilai indeks kedalaman kemiskinan dan nilai indeks keparahan kemiskinan. Indonesia merupakan negara dengan 514 kabupaten/kota di 34 provinsi yang memiliki tingkat kedalaman dan keparahan kemiskinan yang berbeda-beda. Hal tersebut yang mendasari bahwa perlu dilakukan klasifikasi untuk mengetahui tingkat kemiskinan dilihat dari kedua nilai indeks tersebut, apakah termasuk dalam kategori rendah atau tinggi. Nilai indeks dikategorikan rendah apabila berada di bawah nilai rata-rata dari keseluruhan nilai indeks. Sedangkan

nilai indeks dikategorikan tinggi apabila berada di atas nilai rata-rata dari keseluruhan nilai indeks. Untuk mengetahui klasifikasi tersebut maka perlu dilakukan analisis dengan menggunakan metode klasifikasi. Metode klasifikasi merupakan proses penempatan objek tertentu dalam satu set kategori berdasarkan masing-masing objek *property* (Septiani, 2017). Metode klasifikasi yang dapat digunakan yaitu *Naïve Bayes Classifier* dan *K-nearest Neighbor*. Kedua metode tersebut nantinya akan dipilih metode yang terbaik dilihat dari nilai rata-rata akurasi tertinggi.

Penelitian terdahulu terkait penggunaan metode *Naïve Bayes Classifier* dan *K-nearest Neighbor* antara lain:

- Penelitian yang dilakukan oleh Devita, Herwanto, dan Wibawa, (2018) dengan judul “Perbandingan Kinerja Metode *Naïve Bayes* dan *K-nearest Neighbor* untuk Klasifikasi Artikel Berbahasa Indonesia”.
- Penelitian yang dilakukan oleh Sodiq dan Sela, (2019) dengan judul “Perbandingan Metode *Naïve Bayes* dan *K-nearest Neighbor* pada Klasifikasi Kualitas Udara di DKI Jakarta”.
- Penelitian yang dilakukan oleh Dita, Chairunisya, dan Mesran, (2021) dengan judul “Penerapan *Naïve Bayes Classifier* dalam Penyeleksian Beasiswa PPA”.
- Penelitian yang dilakukan oleh Tempola, Muhammad, dan Khairan, (2018) dengan judul “Perbandingan Klasifikasi antara KNN dan *Naïve Bayes* pada Penentuan Status Gunung Berapi dengan *K-Fold Cross Validation*”.
- Penelitian yang dilakukan oleh Yusra, Olivita, dan Vitriani, (2016) dengan judul “Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika Menggunakan Metode *Naïve Bayes Classifier* dan *K-nearest Neighbor*”.

Berdasarkan uraian di atas maka dilakukan klasifikasi dengan menggunakan metode *Naïve Bayes Classifier* dan *K-nearest Neighbor* untuk mengetahui klasifikasi tingkat kemiskinan dilihat dari nilai indeks kedalaman dan indeks keparahan kemiskinan kabupaten/kota di Indonesia tahun 2021. Pada penelitian ini menggunakan beberapa cara pengacakan data dengan fungsi *set.seed*, pembagian data, jumlah nilai ketetanggaan (*k*), serta jarak yang berbeda-beda agar diperoleh nilai akurasi yang paling akurat. Penelitian ini diharapkan dapat membantu pemerintah dalam mengambil sebuah kebijakan terkait tingkat kemiskinan dilihat dari nilai indeks kedalaman dan keparahan kemiskinan.

1.1 Kedalaman dan Keparahannya Kemiskinan

Kemiskinan secara etimologi berasal dari kata “miskin” yang berarti tidak berharta dan serba kekurangan (berpenghasilan sangat rendah). Menurut Tim Badan Pusat Statistik (2021) kemiskinan adalah kondisi dimana kebutuhan pokok atau kebutuhan standar minimum tidak dapat terpenuhi dengan baik seperti makanan, pakaian, rumah, pendidikan, dan kesehatan. Penduduk digolongkan sebagai penduduk miskin apabila rata-rata pengeluaran per kapita per bulan di bawah garis kemiskinan.

Tingkat kemiskinan dapat dilihat dari nilai indeks kedalaman kemiskinan dan nilai indeks keparahan kemiskinan. Nilai indeks kedalaman kemiskinan (*Poverty Gap Index-P1*) adalah ukuran rata-rata kesenjangan pengeluaran masing-masing penduduk miskin terhadap garis kemiskinan. Semakin tinggi nilai indeks maka semakin jauh rata-rata pengeluaran dari garis kemiskinan. Nilai indeks kedalaman kemiskinan dapat dihitung dengan menggunakan rumus sebagai berikut (Tim Badan Pusat Statistik, 2021):

$$P_1 = \frac{1}{n} \sum_{i=1}^q \left[\frac{z-y_i}{z} \right]^1 \quad (1)$$

Sedangkan indeks keparahan kemiskinan (*Proverty Severity Index-P2*) adalah penyebaran pengeluaran di antara penduduk miskin. Semakin tinggi nilai indeks keparahan kemiskinan, maka semakin tinggi ketimpangan pengeluaran di antara penduduk miskin. Indeks keparahan kemiskinan dapat dihitung dengan menggunakan rumus sebagai berikut (Tim Badan Pusat Statistik, 2021):

$$P_2 = \frac{1}{n} \sum_{i=1}^q \left[\frac{z-y_i}{z} \right]^2 \quad (2)$$

Keterangan:

P_1 : Indeks Kedalaman Kemiskinan/*Poverty Gap Index*

P_2 : Indeks Keparahannya Kemiskinan/*Proverty Severity Index*

z : Garis Kemiskinan

y_i : Rata-rata pengeluaran per kapita sebulan penduduk yang berada di bawah garis kemiskinan ($i = 1, 2, 3, \dots, q$), $y_i < z$

q : Banyaknya penduduk yang berada di bawah garis kemiskinan

n : Jumlah penduduk

1.2 Klasifikasi dengan *Naïve Bayes Classifier* dan *K-nearest Neighbor*

Klasifikasi merupakan salah satu metode *supervised learning* atau pembelajaran terawasi. Klasifikasi juga didefinisikan sebagai proses untuk menemukan model atau fungsi yang dapat menggambarkan dan membedakan kelas data atau konsep dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kelas yang belum diketahui dari suatu objek pengamatan. Terdapat dua langkah dalam melakukan klasifikasi (Yulianti dan Sihombing, 2021):

- 1) Langkah belajar atau *training*, pada langkah ini melakukan analisis *training* data untuk membangun model klasifikasi.
- 2) Langkah klasifikasi, pada langkah ini, model yang telah dibangun digunakan untuk memprediksi label kelas untuk data *testing* dan mengetahui kinerja model klasifikasi.

Oleh karena itu, pada algoritma *supervised learning* yaitu metode klasifikasi, *dataset* yang akan diamati terlebih dahulu dilakukan pembagian data menjadi data *training* dan data *testing*. Data *training* digunakan untuk membuat model dan data *testing* digunakan untuk melakukan evaluasi kinerja model.

Naïve Bayes Classifier adalah pengklasifikasi probabilitas sederhana pada teorema Bayes. *Naïve Bayes Classifier* menggunakan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes yang digunakan untuk memprediksi peluang di masa depan berdasarkan kejadian di masa sebelumnya sehingga dikenal dengan teorema Bayes. Teorema Bayes dikombinasikan dengan “*Naïve*” yang artinya atribut/variabel bersifat bebas (*independent*). Ciri *Naïve Bayes* yang utama yaitu masing-masing variabel saling independen. Persamaan teorema Bayes dapat ditulis seperti pada persamaan berikut (Yulianti dan Sihombing, 2021):

$$P(Y|X) = \frac{P(Y)+P(X|Y)}{P(X)} \quad (3)$$

Keterangan:

- X : Data dengan kelas yang belum diketahui
 Y : Hipotesis data X merupakan suatu kelas spesifik
 P(Y|X) : Probabilitas hipotesis Y dengan syarat X (*posterior probability*)
 P(Y) : Probabilitas hipotesis Y (*prior probability*)
 P(X|Y) : Probabilitas X dengan syarat hipotesis Y
 P(X) : Probabilitas X

Gaussian Naïve Bayes adalah metode *Naïve Bayes* Adapun rumus distribusi *Gauss*. Metode ini digunakan pada tipe data kontinu, adapun rumusnya seperti berikut:

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi} \sigma_{i,j}} \exp \left\{ -\frac{(x_i - \mu_{i,j})^2}{2\sigma_{i,j}^2} \right\} \quad (4)$$

Keterangan:

- P : Peluang
 X_i : Atribut ke i
 x_i : Nilai atribut ke i
 Y : Kelas yang dicari
 y_j : Sub kelas yang dicari
 $\mu_{i,j}$: Rata – rata untuk $X_i = x_i$ untuk $Y = y_j$
 $\sigma_{i,j}$: Standar deviasi untuk $X_i = x_i$ untuk $Y = y_j$
 $\sigma_{i,j}^2$: Varian untuk $X_i = x_i$ untuk $Y = y_j$

Berikut langkah perhitungan dengan *Naïve Bayes Classifier* menggunakan data kontinu:

- a) Menghitung probabilitas (*prior*) tiap kelas yang ada
- b) Menghitung rata-rata (*mean*) dan standar deviasi tiap fitur.

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad (5)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n-1)}} \quad (6)$$

- c) Setelah itu, menghitung nilai *likelihood*
- d) Menghitung *posterior probability*
- e) Melakukan normalisasi nilai probabilitas

$$NP(Y_1|X_i) = \frac{P(Y_1|X_i)}{P(Y_1|X_i)+P(Y_2|X_i)} \quad (7)$$

$$NP(Y_2|X_i) = \frac{P(Y_2|X_i)}{P(Y_1|X_i)+P(Y_2|X_i)} \quad (8)$$

Keterangan:

$NP(Y_1|X_i)$: Normalisasi nilai probabilitas kategori (*class*) pertama

$NP(Y_2|X_i)$: Normalisasi nilai probabilitas kategori (*class*) kedua

K-nearest Neighbor adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran (*neighbor*) yang jaraknya paling dekat dengan objek tersebut. Prinsip kerja *K-nearest Neighbor* yaitu mencari jarak terdekat antara data yang akan di evaluasi dengan *k-nearest (neighbor)* terdekatnya dalam data *training* (Baharuddin, 2019). Jarak yang biasa digunakan adalah jarak Euclid. Jarak Euclid adalah perhitungan jarak dari dua titik dalam *euclidean space*. Adapun rumus jarak Euclid sebagai berikut (Istighfarani, dkk., 2020):

$$d_{(x,y)} = \sqrt{\sum_{j=1}^p (x_i - y_i)^2} \quad (9)$$

Keterangan:

$d_{(x,y)}$: Jarak antar objek ke- x dan objek ke- y

p : Jumlah variabel

j : 1,2,...,j

x_i : Data latih

y_i : Data uji

Selain itu, juga dapat menggunakan perhitungan jarak Manhattan atau biasa disebut dengan *city block distance* (blok absolut) yang berarti salah satu perhitungan yang digunakan untuk menghitung perbedaan absolut dengan perbedaan kuadrat. Rumus jarak Manhattan sebagai berikut (Istighfarani, dkk., (2020):

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (10)$$

Keterangan:

$d_{(x,y)}$: Jarak antar objek ke- x dan objek ke- y

x_i : Titik x pada nilai variabel i

y_i : Titik y pada nilai variabel i

n : Jumlah variabel

i : 1,2,...,n

Langkah-langkah dalam melakukan analisis dengan *K-nearest Neighbor* adalah sebagai berikut (Mahardika, dkk., 2019):

- 1) Memasukkan data latih, data uji, dan menentukan parameter k
- 2) Menghitung jarak antar data uji dengan semua data latih, dengan menggunakan jarak Euclid maupun jarak Manhattan
- 3) Mengurutkan jarak dari jarak terkecil
- 4) Menentukan tetangga terdekat berdasarkan jarak minimum nilai k
- 5) Menentukan kategori tetangga terdekat
- 6) Mengkategorikan kelas data uji berdasarkan mayoritas kategori tetangga terdekat.

Pada metode klasifikasi, model di evaluasi dengan nilai akurasi. Nilai akurasi diperoleh dari nilai *confusion matrix*. Nilai akurasi dapat dihitung dengan jumlah prediksi benar dibagi dengan keseluruhan data. Akurasi dapat dihitung dengan rumus:

$$Akurasi = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (11)$$

Keterangan:

TP (*True Positif*) : Apabila data aktualnya Ya dan hasil prediksinya pun juga Ya

FP (*False Positif*) : Apabila data aktualnya Tidak namun hasil prediksinya Ya

FN (*False Negatif*) : Apabila data aktualnya Ya namun hasil prediksinya Tidak

TN (*True Negatif*) : Apabila data aktualnya Tidak dan hasil prediksinya pun benar Tidak.

Pada penelitian ini, analisis dilakukan dengan bantuan software Rstudio. *Packages* yang digunakan untuk melakukan analisis *Naïve Bayes Classifier* yaitu *packages caret, naivebayes, e1071*, serta *caTools* sedangkan untuk analisis *K-nearest Neighbor* menggunakan *packages tidymodels*.

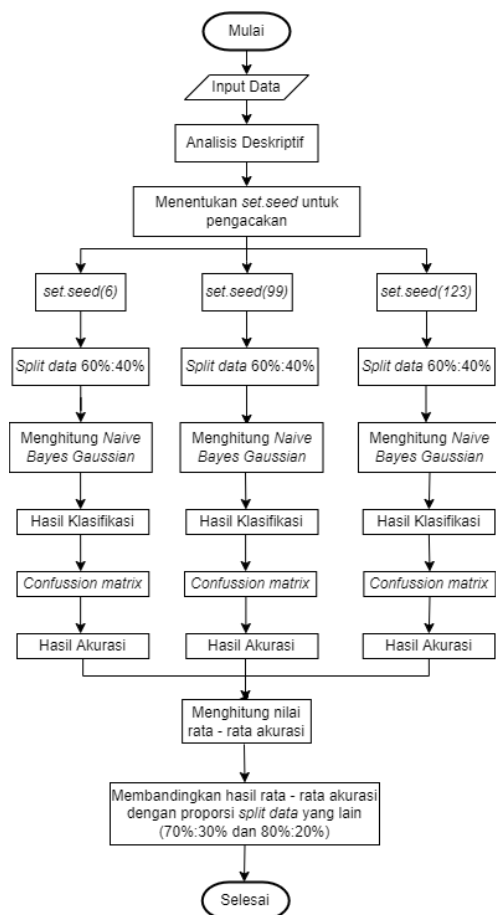
2. METODE PENELITIAN

Penelitian ini menggunakan data sekunder yang diperoleh dari *website* resmi Badan Pusat Statistik (BPS) pada laman www.bps.co.id. Data yang digunakan adalah data tahun 2021 sebanyak 514. Adapun variabel dependen yang digunakan yaitu indeks kedalaman kemiskinan (Y_1) dan indeks keparahan kemiskinan (Y_2). Sedangkan variabel independen terdiri dari garis kemiskinan (X_1), pengeluaran per kapita yang disesuaikan (X_2), rata-rata lama sekolah (X_3), persentasi penduduk miskin (X_4), serta jumlah penduduk miskin (X_5). Pada penelitian ini, klasifikasi terdiri dari dua kategori yaitu kategori tinggi, apabila nilai indeks sama atau di atas nilai rata-rata pada

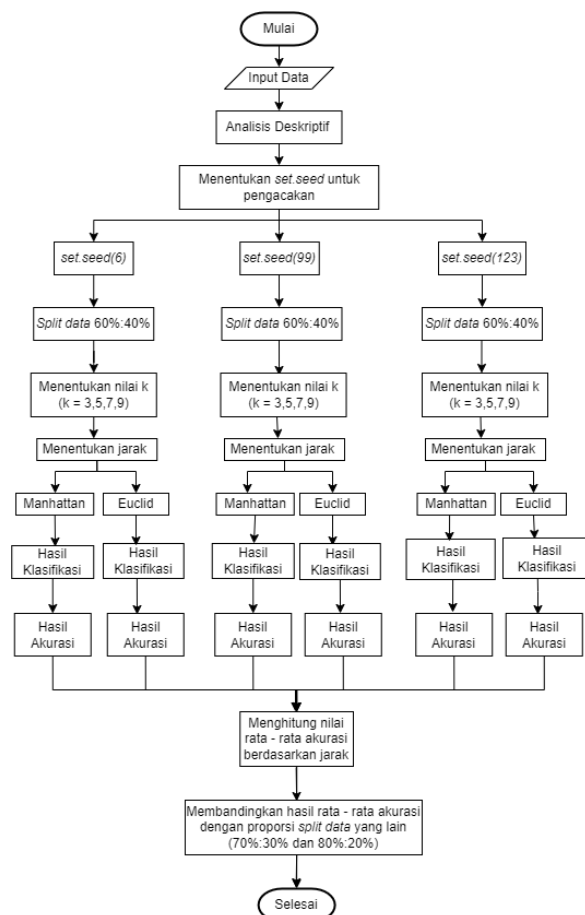
masing-masing variabel dependen serta kategori rendah, apabila nilai indeks di bawah nilai rata-ratanya. Langkah analisis yang dilakukan pada penelitian ini adalah sebagai berikut:

- 1) Mencari dan mengumpulkan data menjadi *dataset* yang siap dilakukan analisis.
- 2) Melakukan analisis statistik deskriptif untuk mengetahui karakteristik data
- 3) Melakukan pengacakan sampel dengan fungsi *set.seed*. Pada penelitian ini menggunakan beberapa fungsi *set.seed* yaitu *set.seed (6)*, *set.seed (99)*, dan *set.seed (123)*.
- 4) Membagi data menjadi data *training* dan data *testing* pada kedua metode *Naïve Bayes Classifier* dan *K-nearest Neighbor*. Pada penelitian ini data *training* dan data *testing* dibagi dengan beberapa perbandingan yaitu 80% data *training* 20% data *testing*, 70% data *training* 30% data *testing*, serta 60% data *training* 40% data *testing*
- 5) Melakukan analisis *Naïve Bayes Classifier* dengan menggunakan distribusi *Gauss*
- 6) Melakukan proses *K-nearest Neighbor* menggunakan nilai *k* atau jumlah ketetanggaan 3, 5, 7, dan 9 serta dengan dua jarak yaitu jarak Euclid dan jarak Manhattan
- 7) Membuat *matrix confusion* pada metode *Naïve Bayes Classifier* dan *K-nearest Neighbor*
- 8) Menghitung nilai akurasi model berdasarkan *matrix confusion* dan membandingkan hasilnya dikedua model.

Flowchart analisis *Naïve Bayes Classifier* disajikan pada **Gambar 1** sedangkan untuk *K-nearest Neighbor* disajikan pada **Gambar 2**.



Gambar 1. Flowchart NBC



Gambar 2. Flowchart Knearest Neighbor

Pada **Gambar 1** dan **Gambar 2** merupakan tahapan untuk melakukan analisis pada pembagian data 60%:40%, langkah ini berlaku untuk pembagian data 70%:30% dan 80%:20%.

3. HASIL DAN PEMBAHASAN

3.1 Deskriptif Data Penelitian

Statistik deskriptif digunakan untuk mengetahui gambaran karakteristik tingkat kedalaman dan keparahan kemiskinan kabupaten/kota di Indonesia tahun 2021 yang disajikan pada **Tabel 1**.

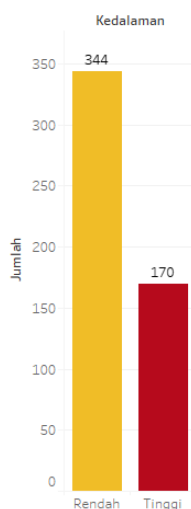
Tabel 1. Statistik Deskriptif Variabel – Variabel Penelitian

	X ₁	X ₂	X ₃	X ₄	X ₅	Y ₁	Y ₂
N	514	514	514	514	514	514	514
Nilai Minimum	253.877	3.976	1,42	2,38	1,49	0,19	0,03
Nilai Maksimum	1.051.297	23.888	12,83	41,66	491,24	14,87	8,12
Rata – rata	460.588,59	10.324,79	8,44	12,27	53,59	2,08	0,57
Standar deviasi	110.745,83	2.717,14	1,63	7,46	60,23	1,93	0,80

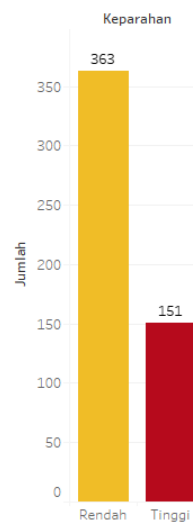
Keterangan:

- X₁ : Garis kemiskinan (Rupiah/Kapita/Bulan)
- X₂ : Pengeluaran per kapita yang disesuaikan (Ribu Rupiah/Kapita/Tahun)
- X₃ : Rata-rata lama sekolah (Tahun)
- X₄ : Persentase penduduk miskin (Persen)
- X₅ : Jumlah penduduk miskin (Ribu Jiwa)
- Y₁ : Indeks kedalaman kemiskinan
- Y₂ : Indeks keparahan kemiskinan

Kemudian pada variabel target yaitu indeks kedalaman kemiskinan (Y₁) dan indeks keparahan kemiskinan (Y₂) yang semula berupa data numerik diubah menjadi data kategorik dengan dua kategori yaitu rendah ($x < \text{rata-rata}$) dan tinggi ($x \geq \text{rata-rata}$). Diperoleh jumlah kabupaten/kota yang memiliki nilai indeks kedalaman kemiskinan tinggi dan rendah seperti pada **Gambar 3** dan untuk nilai indeks keparahan kemiskinan tinggi dan rendah disajikan pada **Gambar 4**.



Gambar 3. Statistik Deskriptif Indeks Kedalaman Kemiskinan



Gambar 4. Statistik Deskriptif Indeks Keparahan Kemiskinan

Berdasarkan **Gambar 3** diperoleh sebanyak 170 kabupaten/kota memiliki kategori tinggi dan 344 kabupaten/kota dengan kategori rendah. Selanjutnya, untuk jumlah kabupaten/kota yang memiliki nilai indeks keparahan kemiskinan dengan kategori tinggi sebanyak 151 kabupaten/kota dan sisanya 363 kabupaten/kota termasuk ke dalam kategori rendah.

3.2 Hasil Klasifikasi

Proses klasifikasi dilakukan menggunakan Rstudio, dengan *software* tersebut perlu proses pengacakan sampel menggunakan fungsi *set.seed*, supaya pengacakan data yang dihasilkan sama setiap saat. Proses pengacakan menggunakan *set.seed(6)*, *set.seed(99)*, dan *set.seed(123)*. Selanjutnya dilakukan pembagian data menjadi data *training* (data latih) dan data *testing* (data uji). Data *training* digunakan untuk membentuk model, sedangkan data *testing* digunakan untuk mencari nilai akurasi dari hasil model. Pada penelitian ini, dilakukan dengan menggunakan beberapa perbandingan yaitu 60%:40%, 70%:30%, dan 80%:20%. Proporsi pembagian data dilakukan secara bebas, tidak ada aturan khusus, pembagian proporsi data *training* lebih besar dari data *testing*. Setelah membagi data selanjutnya dilakukan standarisasi data. Hal ini dilakukan karena variabel pada data penelitian yang digunakan memiliki satuan data yang berbeda-beda sehingga perlu dilakukan standarisasi data supaya sama. Proses standarisasi menggunakan fungsi *scale()*.

Hasil klasifikasi menggunakan metode *Naïve Bayes Classifier* pada variabel indeks kedalaman kemiskinan (Y_1) diperoleh akurasi yang disajikan pada **Tabel 2**.

Tabel 2. Hasil Akurasi Indeks Kedalaman Kemiskinan NBC

Pengacakan	Pembagian data		
	60%:40%	70%:30%	80%:20%
<i>set.seed(6)</i>	89,32%	88,31%	89,32%
<i>set.seed(99)</i>	88,32%	86,36%	84,47%
<i>set.seed(123)</i>	90,78%	90,26%	90,29%
Rata – rata	89,47%	88,31%	88,03%

Hasil analisis menggunakan metode *Naïve Bayes Classifier* pada variabel indeks keparahan kemiskinan (Y_2) diperoleh akurasi disajikan pada **Tabel 3**.

Tabel 3. Hasil Akurasi Indeks Keparahannya Kemiskinan NBC

Pengacakan	Pembagian data		
	60%:40%	70%:30%	80%:20%
<i>set.seed(6)</i>	86,34%	85,06%	83,5%
<i>set.seed(99)</i>	88,29%	88,31%	86,41%
<i>set.seed(123)</i>	88,78%	88,31%	83,5%
Rata – rata	87,80%	87,23%	84,47%

Berdasarkan hasil analisis di atas diperoleh bahwa dengan menggunakan pengacakan (*set.seed*) yang berbeda-beda akan menghasilkan nilai akurasi yang berbeda pula. Selain itu, nilai proporsi perbandingan pada pembagian data *training* dan data *testing* juga dapat mempengaruhi hasil akurasi model. Selanjutnya, untuk hasil klasifikasi dengan *K-nearest Neighbor* pada variabel indeks kedalaman kemiskinan (Y_1) menggunakan jarak Euclid diperoleh nilai akurasi seperti pada **Tabel 4**.

Tabel 4. Hasil Akurasi KNN Jarak Euclid Indeks Kedalaman Kemiskinan

Split data	60%:40%				70%:30%			
	3	5	7	9	3	5	7	9
<i>set.seed(6)</i>	86,9%	87,9%	86,4%	87,4%	83,3%	83,3%	84,6%	82,7%
<i>set.seed(99)</i>	84,5%	85%	85,4%	83%	89,1%	87,2%	86,5%	89,7%
<i>set.seed(123)</i>	86,4%	86,4%	87,4%	86,4%	88,5%	89,7%	89,1%	87,8%
Rata - rata	85,93%	86,43%	86,4%	85,6%	86,97%	86,73%	86,73%	86,73%

Split data	80%:20%			
	3	5	7	9
<i>set.seed(6)</i>	90,3%	91,3%	92,2%	92,2%
<i>set.seed(99)</i>	87,4%	85,4%	83,5%	83,5%
<i>set.seed(123)</i>	85,4%	85,4%	85,4%	86,4%
Rata – rata	87,7%	87,37%	87,03%	87,37%

Kemudian untuk hasil klasifikasi dengan *K-nearest Neighbor* pada variabel indeks keparahan kemiskinan (Y_2) menggunakan jarak Manhattan diperoleh nilai akurasi seperti pada **Tabel 5**.

Tabel 5 Hasil Akurasi KNN Jarak Manhattan Indeks Kedalaman Kemiskinan

Split data	60%:40%				70%:30%			
	3	5	7	9	3	5	7	9
<i>set.seed(6)</i>	86,9%	87,4%	86,9%	86,4%	85,9%	85,9%	86,5%	84,6%
<i>set.seed(99)</i>	86,4%	87,4%	85,9%	84,5%	90,4%	90,4%	87,2%	90,4%
<i>set.seed(123)</i>	88,3%	87,9%	87,4%	87,9%	89,1%	91%	90,4%	88,5%
Rata – rata	87,2%	87,57%	86,73%	86,27%	88,47%	89,1%	88,03%	87,83%

Split data	80%:20%			
	Pengacakan			
	3	5	7	9
set.seed(6)	90,3%	92,2%	92,2%	91,3%
set.seed(99)	88,3%	86,4%	87,4%	85,4%
set.seed(123)	87,4%	87,4%	88,3%	86,4%
Rata - rata	88,67%	88,67%	89,3%	87,7%

Hasil klasifikasi variabel indeks keparahan kemiskinan (Y_2) dengan *K-nearest Neighbor* menggunakan jarak Euclid diperoleh nilai akurasi seperti pada **Tabel 6**.

Tabel 6. Hasil Akurasi KNN Jarak Euclid Indeks Keparahhan Kemiskinan

Split data	60%:40%				70%:30%			
	Pengacakan							
	3	5	7	9	3	5	7	9
set.seed(6)	81,2%	83,6%	86,5%	84,5%	84,5%	87,1%	85,2%	86,5%
set.seed(99)	84,1%	85,5%	86%	85,5%	83,2%	83,9%	83,2%	83,9%
set.seed(123)	82,6%	83,6%	84,1%	84,5%	85,2%	85,8%	83,2%	82,6%
Rata - rata	82,63%	84,23%	85,53%	84,83%	84,30%	85,60%	83,87%	84,33%

Split data	80%:20%			
	Pengacakan			
	3	5	7	9
set.seed(6)	80,8%	81,7%	79,8%	83,7%
set.seed(99)	86,5%	88,5%	86,5%	87,5%
set.seed(123)	77,9%	78,8%	78,8%	76,9%
Rata - rata	81,73%	83%	81,7%	82,7%

Terakhir, untuk hasil klasifikasi variabel indeks keparahan kemiskinan (Y_2) dengan *K-nearest Neighbor* menggunakan jarak Manhattan diperoleh nilai akurasi seperti pada **Tabel 7**.

Tabel 7. Hasil Akurasi KNN Jarak Manhattan Indeks Keparahhan Kemiskinan

Split data	60%:40%				70%:30%			
	Pengacakan							
	3	5	7	9	3	5	7	9
set.seed(6)	84,5%	86%	86%	88,9%	85,2%	87,7%	87,7%	87,1%
set.seed(99)	86%	87,4%	86%	85,5%	86,5%	86,5%	86,5%	85,8%
set.seed(123)	81,6%	84,1%	83,6%	86%	89%	87,1%	83,9%	85,8%
Rata - rata	84,03%	85,83%	85,2%	86,8%	86,9%	87,1%	86,03%	86,23%

Split data	80%:20%			
	Pengacakan			
	3	5	7	9
set.seed(6)	82,7%	79,8%	80,8%	82,7%
set.seed(99)	87,5%	89,4%	91,3%	88,5%
set.seed(123)	77,9%	78,8%	79,8%	78,8%
Rata - rata	82,7%	82,67%	83,97%	83,33%

Berdasarkan analisis dengan *K-nearest Neighbor* diperoleh bahwa dengan menggunakan *set.seed* dan pembagian proporsi data yang berbeda akan menghasilkan nilai akurasi yang berbeda pula. Selain itu, penggunaan jarak serta jumlah ketetanggaan (k) juga dapat mempengaruhi hasil akurasi. Pada penelitian ini terdapat dua jarak yaitu jarak Euclid dan jarak Manhattan yang diperoleh bahwa ternyata jarak Manhattan lebih baik digunakan karena memiliki nilai akurasi yang lebih baik.

4. KESIMPULAN

Hasil analisis diperoleh jumlah kabupaten/kota yang memiliki nilai indeks kedalaman kemiskinan tinggi sebanyak 170 kabupaten/kota dan 344 kabupaten/kota dengan kategori rendah. Kemudian pada indeks keparahan kemiskinan terdapat 151 kabupaten/kota dengan kategori tinggi dan sisanya 363 kabupaten/kota masuk ke dalam kategori rendah. Selain itu, diperoleh hasil dengan menggunakan metode *Naïve Bayes Classifier* dan *K-nearest Neighbor* pada pembagian data *training* dan data *testing* 60%:40%, 70%:30%, dan 80%:20% serta pengacakan sampel yang berbeda-beda, diperoleh rata-rata nilai akurasi pada indeks kedalaman kemiskinan tertinggi yaitu

metode *Naïve Bayes Classifier* dengan pembagian data *training* dan *testing* 60%:40% diperoleh rata-rata akurasi sebesar 89,47%. Sedangkan pada indeks keparahan kemiskinan nilai rata-rata akurasi tertinggi yaitu metode *Naïve Bayes Classifier* pada pembagian data *training* dan *testing* 60%:40% dengan nilai sebesar 87,80%.

Saran untuk penelitian selanjutnya, dalam melakukan pengacakan sampel dengan *set.seed()* dapat dilakukan lebih banyak lagi sehingga nilai akurasi lebih konsisten dan akurat. Hal tersebut dapat menggunakan fungsi *looping* dalam Rstudio agar lebih menghemat waktu analisis. Selain itu, peneliti menyarankan untuk dapat membandingkan dengan metode klasifikasi lain seperti SVM, *Decision Tree*, *Random Forest*, *Modified K-nearest Neighbor*, atau lainnya.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah terlibat dalam pembuatan dan penelitian jurnal ini.

DAFTAR PUSTAKA

- Baharuddin, M. M., Hasanuddin, T., & Azis, H. (2019). Analisis Performa Metode K-nearest Neighbor untuk Identifikasi Jenis Kaca. *ILKOM Jurnal Ilmiah*, 11(3), 269-274.
- Devita, R. N., Herwanto, H. W., & Wibawa, A. P. (2018). Perbandingan Kinerja Metode Naive Bayes dan Knearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 5(4), 427-434.
- Dita, C. A., Chairunisyah, P., & Mesran. (2021). Penerapan Naive Bayes Classifier dalam Penyeleksian Beasiswa PPA. *Journal of Computer System and Informatics (JoSYC)*, 2(2), 194-198.
- Istiqfarani, W. A., Cholissodin, I., & Bachtiar, F. A. (2020). Klasifikasi Penyakit Dental Caries Menggunakan Algoritma Modified K-nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 4(5), 1499-1506.
- Mahardika, A. N., Widodo, A. W., & Rahman, M. A. (2019). Diagnosis Penyakit Mata menggunakan Metode Improved K-nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(11), 10531-10537.
- Septiani, W. D. (2017). Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 dan Naive Bayes untuk Prediksi Penyakit Hepatitis. *Jurnal Pilar Nusa Mandiri*, 13(1), 76-84.
- Sodiq, M. J., & Sela, E. I. (2019). Perbandingan Metode Naive Bayes dan K-nearest Neighbor pada Klasifikasi Kualitas Udara di DKI Jakarta .
- Tempola, F., Muhammad, M., & Khairan, A. (2018). Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-fold Cross Validation. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 5(5), 577-584.
- Tim Badan Pusat Statistik, (2021), *Buku Profil Kemiskinan Kabupaten/Kota 2021*, Jakarta: Badan Pusat Statistik.
- Yuliati, I. F., & Sihombing, P. R. (2021). Penerapan Metode Machine Learning dalam Klasifikasi Resiko Kejadian Berat Badan Lahir Rendah di Indonesia. *Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, 20(2), 417-426.
- Yusra, Olivita, D., & Vitriani, Y. (2016). Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika Menggunakan Metode Naive Bayes Classifier dan K-nearest Neighbor. *Jurnal Sains Teknologi dan Industri*, 14(1), 79-85.