

Klasifikasi Pengeluaran per Kapita di Tiga Provinsi Sulawesi menggunakan K-Nearest Neighbor

Ismi Rizqa Lina ⁽¹⁾, Dia Cahya Wati ⁽²⁾

^{1,2}Prodi Sains Data, Universitas Insan Cita Indonesia

Graha Binakarsa, 10th floor, Jl. H.R Rasuna Said, Kota Jakarta Selatan, DKI Jakarta

e-mail: irizqalina@gmail.com⁽¹⁾, diacahyawati@gmail.com⁽²⁾

ABSTRAK

Klasifikasi pengeluaran per kapita merupakan analisis pasar yang penting bagi banyak perusahaan untuk menentukan Kabupaten/kota mana yang paling cocok untuk menjual suatu produk di perusahaan tersebut. KNN dapat digunakan untuk berbagai jenis data, termasuk data ekonomi seperti pengeluaran per kapita. Pada penelitian ini, 56 data pengeluaran per kapita di provinsi Sulawesi Selatan, Sulawesi Utara, dan Sulawesi Tenggara pada tahun 2022 diklasifikasi dengan algoritma KNN. Proses klasifikasi menggunakan algoritma KNN diawali dengan melakukan *pre-procecing* data dan menghitung jarak antara data pelatihan (data *training*) dengan data uji (data *testing*). Dalam perhitungan jarak, digunakan metrik *Euclidean* dan metrik *Manhattan*. Selanjutnya, dilakukan perhitungan nilai prediksi berdasarkan *k* data *training* terdekat. Hasil pengujian pada penelitian ini menunjukkan akurasi tertinggi pada $k = 9$ untuk jarak *Eucledian* sebesar 76,47% yang berarti klasifikasi cukup dan untuk jarak *Manhattan* sebesar 94,12% yang berarti klasifikasi sangat baik. Dari hasil akurasi kedua jarak tersebut dapat disimpulkan bahwa jarak *Manhattan* lebih baik daripada jarak *Eucledian*.

Kata kunci: K-Nearest Neighbor; Euclidean; Manhattan; Pengeluaran per Kapita; Sulawesi

ABSTRACT

*Classification of expenditure per capita is an important market analysis for many companies to determine which district/city is most suitable for selling products in the company. KNN can be used for various types of data, including economic data such as expenditure per capita. In this study, 56 per capita expenditure data in the provinces of South Sulawesi, North Sulawesi and Southeast Sulawesi in 2022 were classified using the KNN algorithm. The classification process using the KNN algorithm begins with pre-processing the data and calculating the distance between the training data and the test data. In calculating distances, the Euclidean metric and the Manhattan metric are used. Furthermore, the predicted value is calculated based on the closest *k* training data. The test results in this study showed the highest accuracy at $k = 9$ for the Eucledian distance of 76.47% which means that the classification is sufficient and for the Manhattan distance it is 94.12% which means the classification is very good. Based on the accuracy results of both distances, the conclusion can be drawn that the Manhattan distance outperforms the Euclidean distance.*

Keywords: K-Nearet Neighbor; Eucledian; Manhattan; Expenditure per Capita; Sulawesi

PENDAHULUAN

Pengeluaran per kapita adalah salah satu indikator ekonomi yang berguna dalam menilai tingkat produktivitas dan kesejahteraan masyarakat. Perbedaan antara desil terkaya dan termiskin dalam masyarakat tercermin dari distribusi pendapatan dan tingkat ketimpangan pendapatan, yang dipengaruhi oleh faktor-faktor struktural ekonomi dan situasi sosial dalam lingkungan masyarakat [1]. Pernyataan ini juga di dukung oleh penelitian yang berkaitan dengan pertumbuhan populasi, semakin banyak penduduk dalam suatu wilayah akan berdampak pada perubahan jumlah penduduk dan status ekonomi di wilayah tersebut [2].

Pertumbuhan ekonomi Sulawesi Selatan pada triwulan I tahun 2013 mencapai 7,79%, yang merupakan angka yang sedikit lebih rendah daripada pertumbuhan ekonomi pada triwulan I tahun 2012 yang mencapai 7,95%. Namun, pertumbuhan ekonomi Sulawesi Selatan tersebut lebih tinggi dibandingkan dengan pertumbuhan ekonomi secara nasional pada periode yang sama tahun 2013, yang hanya mencapai 0,02% [3]. Provinsi Sulawesi Selatan menjadi wilayah yang paling berkembang dan berperan sebagai pendorong utama pertumbuhan ekonomi, sehingga menyebabkan disparitas antara daerah-daerah yang lebih besar [4]. Dalam konteks indeks Gini, nilai indeks Gini yang berada antara 0 hingga 1 menunjukkan sejauh mana distribusi pendapatan yang merata. Nilai indeks Gini yang sama dengan 0 mengindikasikan distribusi pendapatan yang sangat merata, sedangkan nilai indeks Gini yang sama dengan 1 menunjukkan distribusi pendapatan yang sangat tidak merata [5]. Hal serupa dilakukan oleh penelitian [6] bahwa ketimpangan perekonomian dan pembangunan juga terjadi di Kawasan Timur seperti Sulawesi Selatan, Sulawesi Utara, dan Sulawesi Tenggara.

Penggunaan Algoritma *K-Neares Neighbor* (KNN) sesuai untuk melakukan klasifikasi pengeluaran per kapita di wilayah-wilayah Sulawesi Selatan, Sulawesi Utara, dan Sulawesi Tenggara. KNN merupakan suatu teknik klasifikasi yang digunakan dalam data mining dan termasuk dalam kelompok metode pembelajaran berbasis contoh (*instance-based learning*). Dalam KNN, pencarian dilakukan untuk menemukan k objek terdekat dari data pelatihan yang paling mirip dengan objek pada data uji [7]. Penelitian [8] melakukan prediksi nilai tanah menggunakan algoritma KNN yang menghasilkan tingkat akurasi prediksi data testing sebesar 80%. penggunaan Algoritma KNN dengan metrik jarak Manhattan dapat membantu penetapan dalam pemenangan lelang dengan tingkat akurasi sebesar 0,8 [9].

Berdasarkan paparan tersebut penulis ingin mengklasifikasikan pengeluaran per-kapita di Sulawesi Selatan, Sulawesi Utara, dan Sulawesi Tenggara dengan KNN. Metode klasifikasi K-NN dengan menggunakan metrik jarak Euclidean dan Manhattan telah diteliti oleh [10]. Hasil penelitiannya menunjukkan bahwa dalam klasifikasi transportasi bus, metrik jarak Manhattan lebih akurat dibandingkan dengan metrik jarak Euclidean. Penelitian yang dilakukan oleh [11] menunjukkan hasil yang berbeda. Dalam penelitian tersebut, metrik jarak *Euclidean* dan *Minkowski* pada algoritma KNN pada data dengan representasi vektor dari kalimat menunjukkan akurasi terbaik, yang lebih baik daripada menggunakan metrik jarak *Chebyshev* maupun *Manhattan*.

METODE

Penelitian ini menggunakan data sekunder yang diperoleh dari situs resmi Badan Pusat Statistik (BPS) Sulawesi Selatan (<https://sulsel.bps.go.id/>), Sulawesi Utara

(<https://sulut.bps.go.id/>), dan Sulawesi Tenggara (<https://sultra.bps.go.id/>) untuk tahun 2022. Adapun variabel-variabel yang menjadi fokus dalam penelitian ini sebagai berikut.

Tabel 1. Struktur data

Variabel	Notasi	Keterangan
Pengeluaran per kapita	y	Kategorik : 1 : Tinggi (≥ 10610) 0 : Rendah (< 10610)
HLS	x_1	Numerik
IPM	x_2	Numerik
PDRB	x_3	Numerik
Mi	x_4	Numerik
RLS	x_5	Numerik
UHH	x_6	Numerik

Keterangan.

HLS : Harapan Lama Sekolah

IPM : Indeks Pembangunan manusia

PDRB : Produk Domestik Regional Bruto

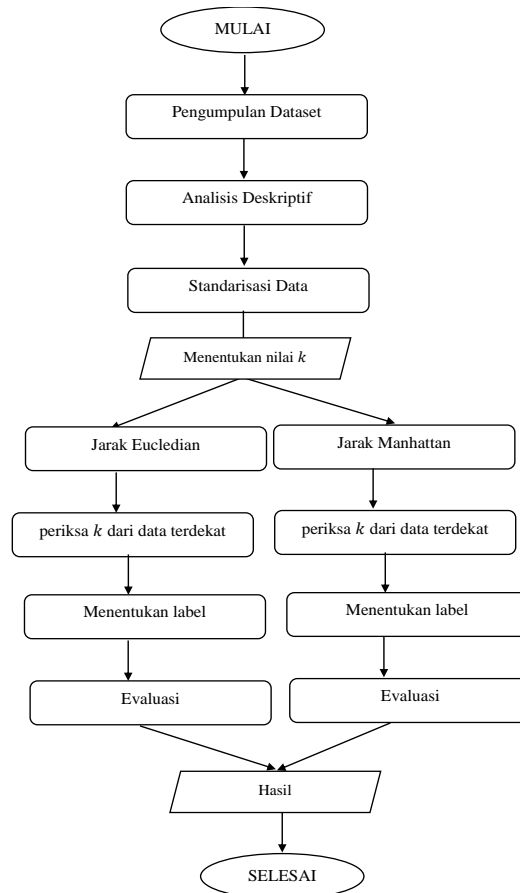
Mi : Presentase penduduk miskin menurut kabupaten/kota

RLS : Rata-rata lama sekolah

UHH : Umur harapan hidup.

Penelitian ini direpresentasikan melalui sebuah flowchart yang menunjukkan rangkaian tahapan pelaksanaan secara visual dan terstruktur yang ditunjukkan pada **Gambar 1**.

Tahapan awal penelitian ini dimulai dengan pengumpulan dataset dari sumber-sumber yang relevan. Selanjutnya, dilakukan analisis deskriptif untuk memahami karakteristik data. Jika satuan data berbeda, standarisasi data digunakan untuk menyamakan skala variabel. Selanjutnya, dataset dibagi menjadi dua bagian, yaitu data training (data pelatihan) dan data testing (data uji), agar model KNN dapat diuji pada data yang tidak digunakan selama pelatihan. Pada tahap klasifikasi, algoritma KNN digunakan untuk melakukan klasifikasi data berdasarkan tetangga terdekat. KNN adalah salah satu algoritma machine learning yang digunakan untuk analisis klasifikasi dan regresi. Prinsip kerjanya sederhana, yaitu melakukan klasifikasi pada objek baru berdasarkan jumlah k tetangga terdekatnya [12]. Dalam ruang dua dimensi, perhitungan jarak antara dua titik digunakan untuk mengidentifikasi tetangga yang jauh atau dekat dari suatu titik tertentu. Salah satunya, yaitu jarak *Euclidean* dan jarak *Manhattan* [13].



Gambar 1. Flowchart Algoritma KNN

Jarak *Euclidean* merupakan metode perhitungan jarak yang digunakan untuk mengukur jarak antara dua buah titik dalam ruang *Euclidean*. Metode ini pertama kali diperkenalkan oleh Euclid, seorang matematikawan terkemuka dari Yunani kuno. Secara umum, jarak *Euclidean* didefinisikan sebagai panjang garis lurus yang menghubungkan dua titik tersebut. Jarak *Euclidean* antara dua titik (x_1, y_1) dan (x_2, y_2) dalam ruang dua dimensi dihitung menggunakan rumus berikut [12].

$$d_E(x_1, x_2) = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2 + (y_{1i} - y_{2i})^2 + \dots}, \tag{1}$$

dengan

$d_E(x_1, x_2)$: jarak *Euclidean* dari data *trainig* ke-1 ke data *testing* ke-2

x_1 : koordinat data *training* ke-1

x_2 : koordinat data *testing* ke 2

Rumus ini menghitung panjang garis lurus (jarak *Euclidean*) antara dua titik dalam ruang dua dimensi yang diwakili oleh data training ke-1 (x_1) dan data testing ke-2 (x_2). Hasil perhitungan jarak *Euclidean* ini akan digunakan dalam algoritma KNN untuk menentukan k tetangga terdekat dari data testing (x_2) berdasarkan data training (x_1). Semakin kecil nilai jarak *Euclidean*

$d_E(x_1, x_2)$, semakin dekat data training (x_1) dengan data testing (x_2), dan semakin besar nilai jarak *Euclidean*, semakin jauh keduanya.

Jarak *Manhattan*, juga dikenal sebagai *City Distance*, adalah metode perhitungan jarak yang digunakan dalam algoritma KNN untuk mengidentifikasi kasus yang paling cocok dari basis kasus dengan mengukur jumlah bobot absolut dari perbedaan antara kasus yang sedang diuji dan kasus lain dalam basis kasus. Untuk menghitung bobot (jarak *Manhattan*) antara dua data x dan y dalam ruang dua dimensi, digunakan persamaan berikut:

$$d_M(x, y) = \sum_{i=0}^n |x_i - y_i|, \tag{2}$$

x dan y adalah dua data dalam ruang dua dimensi, dan n adalah jumlah dimensi dalam ruang tersebut. Perhitungan bobot menggunakan jarak Manhattan ini melibatkan penjumlahan dari selisih absolut antara koordinat data x_i dan y_i dalam masing-masing dimensi. Hasil bobot ini akan digunakan dalam algoritma KNN untuk menentukan k tetangga terdekat dari data testing x_i berdasarkan data training y_i . Semakin kecil nilai bobot (jarak *Manhattan*), semakin dekat data x_i dengan data y_i dalam ruang berdimensi dua, dan semakin besar nilai bobot, semakin jauh keduanya.

Setelah menghitung jarak *Euclidean* dan *Manhattan* serta melakukan prediksi pada data *testing* menggunakan algoritma *KNN*, langkah selanjutnya adalah melakukan evaluasi model untuk membandingkan nilai aktual dan nilai hasil prediksi. Evaluasi yang digunakan dalam penelitian ini adalah *confusion matrix* (matriks kebingungan). *Confusion matrix* adalah tabel yang digunakan untuk menggambarkan kinerja model klasifikasi. Berikut bentuk perhitungan dari *confusion matrix* dan menghasilkan empat nilai.

Tabel 2. *Confusion matrix* pengklasifikasi biner

Predict	Actual	
	True	False
True	True Positif	False Negatif
False	False Positif	True Negatif

Keterangan :

- True Positive (TP): Jumlah data positif yang benar-benar diprediksi dengan benar, artinya model berhasil mengklasifikasikan data positif secara tepat.
- False Positive (FP): Jumlah data negatif yang salah diprediksi sebagai positif, artinya model keliru mengklasifikasikan data negatif sebagai positif.
- True Negative (TN): Jumlah data negatif yang benar-benar diprediksi dengan benar, artinya model berhasil mengklasifikasikan data negatif secara tepat.
- False Negative (FN): Jumlah data positif yang salah diprediksi sebagai negatif, artinya model keliru mengklasifikasikan data positif sebagai negatif.

Dari *confusion matrix* dapat dihitung berbagai metrik evaluasi seperti akurasi, presisi, *recall*, dan F1-score yang memberikan informasi lebih lanjut tentang performa model dalam melakukan

klasifikasi. Evaluasi model ini membantu dalam menilai sejauh mana model KNN telah berhasil dalam memprediksi data dengan benar.

Accuracy merupakan salah satu metrik evaluasi yang digunakan untuk mengukur kinerja model klasifikasi. *Accuracy* menggambarkan seberapa akurat model dalam mengidentifikasi kasus dengan benar dari seluruh kasus yang ada dalam dataset. Rumus untuk menghitung *Accuracy* adalah:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \times 100\%. \quad (4)$$

Presisi merupakan salah satu metrik evaluasi yang digunakan untuk mengukur kinerja model klasifikasi, khususnya dalam kasus klasifikasi positif. Presisi mengukur proporsi dari kasus dengan hasil positif (positif yang diprediksi) yang benar-benar diidentifikasi dengan benar dari seluruh kasus yang diprediksi sebagai positif. Rumus untuk menghitung Presisi sebagai berikut.

$$Precision = \frac{TP}{(TP + FP)} \times 100\% \quad (5)$$

Presisi memberikan informasi tentang seberapa akurat model dalam mengidentifikasi kasus positif. Semakin tinggi nilai presisi, semakin sedikit data negatif yang salah diprediksi sebagai positif, dan semakin tepat model dalam mengklasifikasikan data sebagai positif.

Recall yang juga dikenal sebagai *Sensitivity* atau *True Positive Rate* (TPR), adalah salah satu metrik evaluasi yang digunakan untuk mengukur kinerja model klasifikasi, khususnya dalam kasus klasifikasi positif. *Recall* mengukur proporsi kasus positif yang benar-benar diidentifikasi dengan benar (True Positive) dari seluruh kasus positif yang ada dalam dataset. Berikut formula untuk menghitung Recall.

$$Recall = \frac{TP}{(TP + FN)} \times 100\% \quad (6)$$

Recall memberikan informasi tentang seberapa baik model dapat mengidentifikasi kasus positif dengan benar. Semakin tinggi nilai *Recall*, semakin sedikit kasus positif yang terlewatkan (*False Negative*), dan semakin baik model dalam mengenali data positif.

F-measure, juga dikenal sebagai F1-score, adalah metrik evaluasi yang mengkombinasikan pengukuran Precision dan Recall menjadi suatu nilai tunggal yang mencerminkan keseimbangan antara keduanya. F-measure dihitung dengan menggunakan rumus berikut:

$$F = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \times 100\% \quad (7)$$

F-measure mencakup *Precision* dan *Recall* dalam perhitungannya, sehingga memberikan gambaran tentang kinerja model klasifikasi yang lebih holistik. Metrik ini berguna dalam situasi di mana kita ingin menemukan keseimbangan antara ketepatan (*Precision*) dan kemampuan mengenali positif (*Recall*) dari model klasifikasi.

F-measure memberikan nilai dari 0 hingga 1, di mana nilai 1 menunjukkan performa yang sempurna (keseimbangan yang ideal antara *Precision* dan *Recall*) dan nilai 0 menunjukkan performa yang sangat buruk.

Dalam klasifikasi data mining, nilai akurasi umumnya dapat digolongkan ke dalam beberapa kategori sebagai berikut [14]:

- 0.90 – 1.00: Klasifikasi sangat baik, yang berarti model memiliki tingkat keakuratan yang sangat tinggi dalam melakukan klasifikasi. Model mampu memprediksi data dengan sangat tepat dan memiliki performa yang sangat baik.
- 0.80 – 0.90: Klasifikasi baik, menunjukkan bahwa model memiliki tingkat keakuratan yang baik dalam melakukan klasifikasi. Model mampu memprediksi data dengan baik dan memberikan hasil yang dapat diandalkan.
- 0.70 – 0.80: Klasifikasi cukup, yang berarti model memiliki tingkat keakuratan yang cukup dalam melakukan klasifikasi. Meskipun performanya cukup baik, masih ada ruang untuk perbaikan.
- 0.60 – 0.70: Klasifikasi buruk, menunjukkan bahwa model memiliki performa yang buruk dalam melakukan klasifikasi. Model mungkin mengalami kesulitan dalam memprediksi data dengan tepat.
- 0.50 – 0.60: Klasifikasi salah, yang berarti model memiliki performa yang sangat buruk dan hasil prediksinya hampir tidak lebih baik daripada tebakan acak.

HASIL DAN DISKUSI

Perhitungan nilai prediksi dilakukan pada data sekunder berdasarkan waktu tahunan, yaitu tahun 2022. Data tersebut diprediksi menggunakan *forecasting*. Setelah data di prediksi kemudian dilakukan analisis deskriptif. Analisis Deskriptif dalam statistika digunakan untuk memperoleh sedikit gambaran umum tentang karakteristik data pengeluaran per-kapita di Sulawesi Selatan, Sulawesi Utara, dan Sulawesi Tenggara berjumlah 56 Kabupaten/Kota.

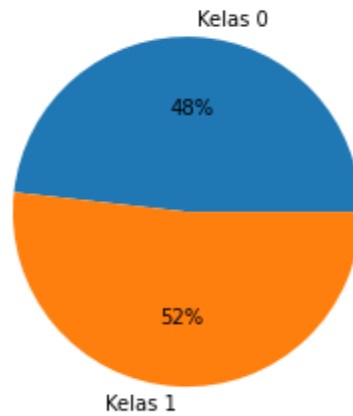
Tabel 3. Deskripsi Data

No	Variabel	Rata-rata	N	Min	Maks
1	HLS	13.20	56	11.61	16.90
2	IPM	71.26	56	65.13	84.51
3	PDRB	43.11	56	2.50	145.89
4	Mi	19.89	56	4.57	80.34
5	RLS	8.79	56	6.75	12.52
6	UHH	70.11	56	64.74	73.93
7	Pengeluaran per-kapita	10610.38	56	7192	17406

Berdasarkan **Tabel 3**, rata-rata angka pengeluaran per-kapita di Sulawesi Selatan, Sulawesi Utara, dan Sulawesi Tenggara sebesar 10610.38. Untuk HLS terendah pada Bolaang Mongondow sebesar 11,61 dan tertinggi pada Kota Kendari sebesar 16,9. Untuk IPM terendah pada Jeneponto sebesar 12.11 dan tertinggi pada Kota Kendari 84,51. Untuk PDRB terendah pada Bolaang

Mongondow sebesar 2,5 dan tertinggi pada Kota Makassar sebesar 145,89. Untuk Mi terendah pada Kota Kendari 4,57 dan tertinggi pada Bone sebesar 80,34. Untuk RLS terendah pada Jenepoto sebesar 6,75 dan tertinggi sebesar Kota Kendari 12,52. Untuk UHH terendah pada Bolaang Mongondow sebesar 64,74 dan tertinggi pada Kota Kendari sebesar 73,93. Untuk pengeluaran perkapita terendah pada Konawe Kepulauan sebesar 7192 dan tertinggi pada Kota Makassar sebesar 17406.

Klasifikasi pengeluaran per-kapita



Gambar 2. Klasifikasi pengeluaran per-kapita menurut BPS

Berdasarkan **Gambar 2.** ada 29 Kabupaten/Kota atau 52% yang tergolong di kelas 1 dan 27 Kabupaten/Kota atau 48% tergolong di kelas 0. Hal ini menunjukkan bahwa angka pengeluaran per-kapita pada kelas 1 atau kategori tinggi sedikit lebih besar dibandingkan kelas 0 atau kategori rendah. Penerapan standarisasi menjadi penting ketika data dalam sebuah penelitian memiliki unit yang berbeda-beda. Standarisasi data bertujuan agar data tersebut mempunyai jarak yang tidak terlalu jauh dan membuat kisaran nilai data dengan format tertentu sehingga seluruh data menjadi standar. Proses standarisasi pada data asli dilakukan dengan menggunakan nilai Z-Score, yang dihitung menggunakan rumus berikut.

$$Z = \frac{x - \mu}{\sigma}$$

dengan

Z : Z-Score

x : data yang diamati

μ : rata-rata populasi

σ : standar deviasi.

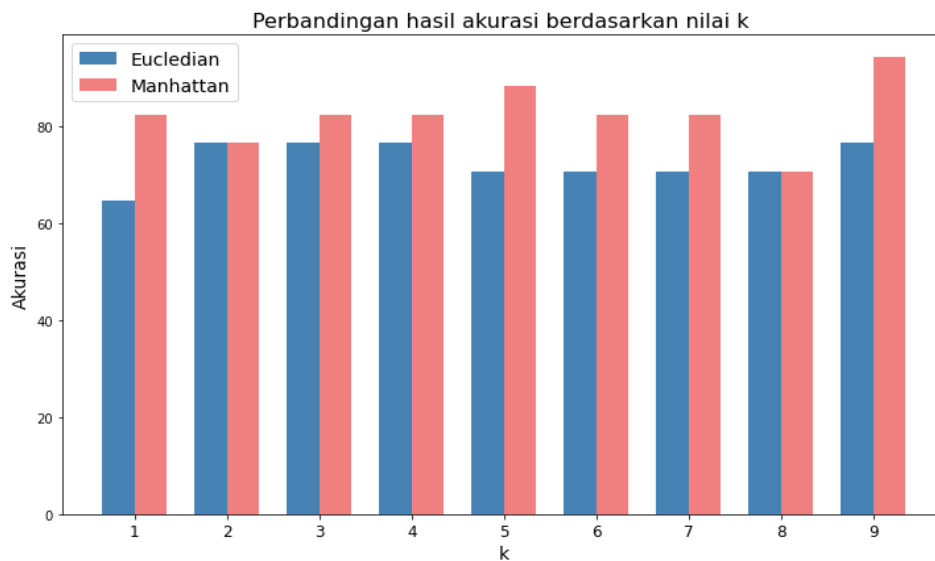
Hasil standarisasi data sebagai berikut.

Tabel 4. Data yang telah distandarisasi

Kab/ Kota	x1	x2	x3	x4	x5	x6
Kep. Selayar	-5.5494	-6.9678	3.7029	-1.9329	-5.7573	-7.1276
Bulukumba	2.3114	-2.2005	-1.4712	6.9873	-6.4107	-8.7671
Bantaeng	-9.4275	-3.7577	6.2205	-1.6386	-1.6212	4.1844
⋮						
Bitung	-5.7590	8.2443	-7.8208	-8.2477	9.1092	6.9168
Tomohon	1.0801	1.4808	-1.3242	-8.7381	1.5970	1.2709
Kotamobagu	-3.2435	6.4715	-1.3457	-8.7504	1.2458	4.7309

Pada penelitian ini, split ratio yang digunakan adalah 8:2, yang berarti 80% dari dataset digunakan sebagai data training dan 20% dari dataset digunakan sebagai data testing. Penggunaan split ratio ini penting dalam proses pembagian dataset menjadi dua bagian yang berbeda untuk tujuan evaluasi model. Dengan menggunakan 80% data training, model akan belajar dari mayoritas data untuk melakukan klasifikasi. Sedangkan, data testing sebesar 20% digunakan untuk menguji performa model terhadap data yang belum pernah dilihat sebelumnya. Penggunaan split ratio yang tepat akan membantu menghindari overfitting atau underfitting, serta memberikan evaluasi yang lebih kredibel terhadap performa model dalam keadaan yang lebih realistis.

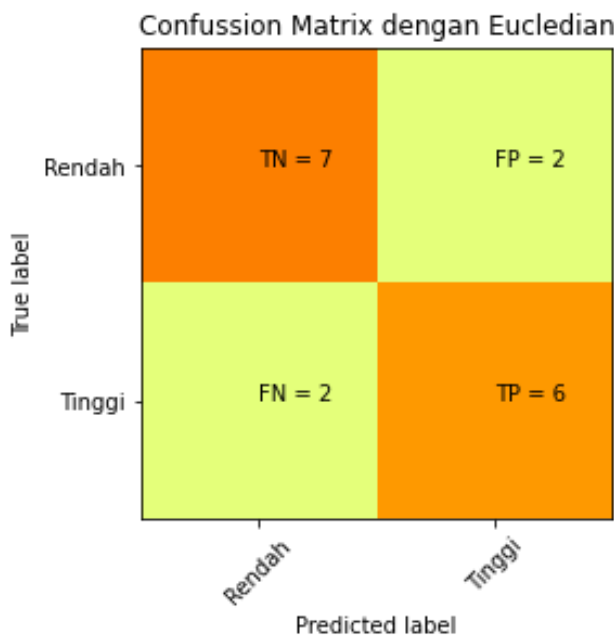
Pada penelitian ini, nilai k divariasikan dari 1 hingga 9 untuk menghasilkan prediksi berdasarkan k data terdekat yang memiliki jarak Euclidean dan Manhattan paling kecil dengan data yang akan diprediksi. Dengan melakukan variasi nilai k , penelitian ini bertujuan untuk mengeksplorasi pengaruh jumlah tetangga terdekat pada performa prediksi algoritma KNN. Dengan cara ini, model dapat menyesuaikan diri dengan berbagai situasi dan memilih k yang optimal untuk mencapai hasil prediksi yang akurat dan konsisten.



Gambar 3. Hasil akurasi berdasarkan nilai k dengan *Euclidean* dan *Manhattan*

Berdasarkan Gambar 3. memperlihatkan hasil perbandingan dari beberapa percobaan nilai k dengan jarak Euclidian dan Manhattan. Nilai akurasi yang paling tinggi untuk kedua jarak tersebut terdapat pada $k = 9$. Akurasi pada jarak Euclidian dengan $k = 9$ sebesar 76,47% yang berarti hasil prediksi kelompok angka pengeluaran per kapita yang tergolong cukup dengan metode KNN sesuai dengan dataset sebelum diprediksi. Sedangkan akurasi dengan jarak Manhattan pada $k = 9$ sebesar 94,12%. Hal ini berarti hasil prediksi kelompok angka pengeluaran per kapita dengan metode KNN tergolong sangat baik sesuai dengan dataset sebelum diprediksi.

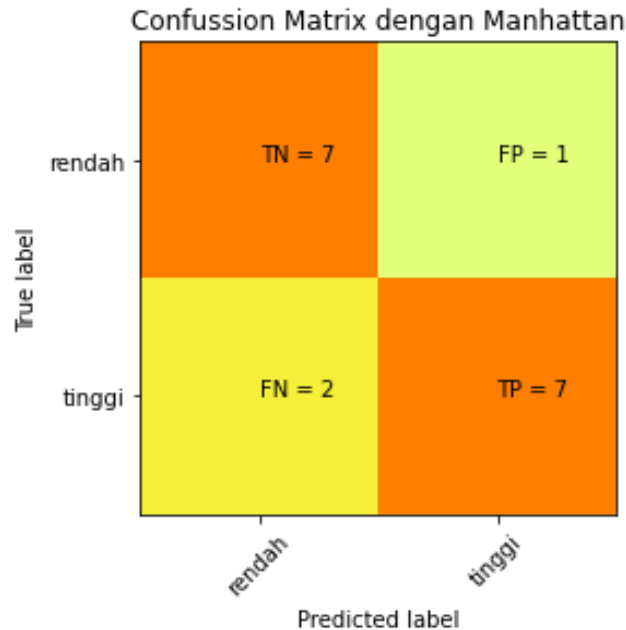
Setelah data diklasifikasi selanjutnya dievaluasi dengan *Confusion matrix*. Empat indikator utama yang digunakan sebagai hasil evaluasi, yaitu akurasi, presisi, *recall*, dan F-1 score. Hasil akurasi *Confusion matrix* dengan *Euclidian* menggunakan nilai $k = 9$ ditunjukkan pada **Gambar 4**.



Gambar 4. Hasil *Confusion* dengan *Matrix Euclidian*

Berdasarkan **Gambar 4**, terdapat 7 data yang diprediksi dengan benar sebagai angka pengeluaran per kapita yang tergolong rendah, dan 2 data yang salah diprediksi sebagai rendah. Selanjutnya, terdapat 6 data yang diprediksi dengan benar sebagai angka pengeluaran per kapita yang tergolong tinggi, dan 2 data yang salah diprediksi sebagai tinggi.

Dengan demikian, hasil evaluasi model menunjukkan nilai precision, recall, dan F1-score yang sama, yaitu sebesar 0.78. Hal ini berarti model memiliki tingkat keakuratan yang cukup baik dalam mengidentifikasi kasus positif dan negatif secara benar.



Gambar 5. Hasil *Confussion Matrix* dengan *Manhattan*

Berdasarkan **Gambar 5**, terdapat 7 data yang diprediksi dengan benar sebagai angka pengeluaran per kapita yang tergolong rendah, dan 1 data yang salah diprediksi sebagai rendah. Selanjutnya, terdapat 7 data yang diprediksi dengan benar sebagai angka pengeluaran per kapita yang tergolong tinggi, dan 2 data yang salah diprediksi sebagai tinggi.

Dengan demikian, hasil evaluasi model dengan menggunakan jarak *Manhattan* dan nilai $k = 9$ menunjukkan adanya peningkatan dalam jumlah data yang diprediksi dengan benar pada kategori angka pengeluaran per kapita yang tergolong rendah (dari 6 menjadi 7) dan kategori yang tergolong tinggi (dari 6 menjadi 7) dibandingkan dengan hasil pada **Gambar 4**.

Berdasarkan hasil evaluasi model, nilai *precision*, *recall*, dan F1-score berturut-turut adalah 0.78, 0.88, dan 0.82. Nilai-nilai ini menunjukkan bahwa model memiliki kinerja yang baik dalam mengidentifikasi kasus secara benar. Dari hasil evaluasi tersebut, dapat disimpulkan bahwa metode KNN dengan $k = 9$ dan jarak *Manhattan* tergolong klasifikasi sangat baik, karena memiliki nilai *precision*, *recall*, dan F1-score yang tinggi. Hal ini menunjukkan bahwa model mampu mengklasifikasikan data dengan baik dan akurat.

Sementara itu, jarak *Euclidean* tergolong klasifikasi cukup dengan nilai *precision*, *recall*, dan F1-score yang lebih rendah dibandingkan dengan jarak *Manhattan*. Meskipun masih memberikan hasil yang cukup baik, performa KNN dengan jarak *Manhattan* lebih unggul dalam mengklasifikasikan data.

KESIMPULAN

Dari hasil dan diskusi yang telah disajikan, klasifikasi angka pengeluaran per kapita kabupaten/kota di tiga provinsi Sulawesi pada tahun 2022, menggunakan metode KNN menghasilkan nilai k tertinggi, yaitu $k = 9$, dengan akurasi terbaik. Hasil akurasi yang diperoleh adalah sebesar 76,47% dengan jarak *Euclidean*, menunjukkan hasil klasifikasinya dapat dianggap

cukup baik. Selain itu, dengan jarak Manhattan, diperoleh akurasi sebesar 94,12%, yang berarti hasil klasifikasinya tergolong sangat baik.

Dengan menggunakan jarak *Manhattan*, model KNN mampu mengklasifikasikan data dengan akurasi yang lebih tinggi dan memberikan hasil prediksi yang lebih akurat. Hal ini menunjukkan bahwa jarak *Manhattan* lebih sesuai dengan karakteristik data pengeluaran per kapita di tiga provinsi Sulawesi pada tahun 2022, sehingga memberikan kinerja model yang lebih optimal dalam mengklasifikasikan data tersebut.

DAFTAR PUSTAKA

- [1] U. Hasannah and H. Ahmadi, "Pengaruh Ketimpangan Pendapatan, Pendapatan Per Kapita, Dan Pengeluaran Pemerintah Di Bidang Kesehatan Terhadap Sektor Kesehatan Di Indonesia," *J. Ilmu Ekon. Terap.*, vol. 2, no. 1, pp. 1–18, 2017.
- [2] R. F. K. Dewi, O. Obert, and R. Gusmana, "Implementasi Metode K-Nearest Neighbor (KNN) dalam Pengelompokan Status Ekonomi Warga," *J. Big Data Anal. Artif. Intell.*, vol. 4, no. 1, pp. 15–22, 2018.
- [3] A. Hakib, "Pengaruh Konsumsi Rumah Tangga Dan Pengeluaran Pemerintah Terhadap Pertumbuhan Ekonomi Di Sulawesi Selatan Periode 2012-2016," *J. Ekon. Balanc. Fak. Ekon. Dan Bisnis*, vol. 15, no. 1, 2019.
- [4] J. Zakaria, "Analisis Pertumbuhan Ekonomi dan Disparitas Pendapatan Terhadap Kesejahteraan Antar Daerah Kabupaten/Kota di Provinsi Sulawesi Selatan Tahun 2005-2019," *Tata Kelola*, vol. 8, no. 1, pp. 2–24, 2021.
- [5] S. I. S. Dai, S. Canon, and D. O. Bauty, "Analisis Pengaruh Rls, Pengeluaran Perkapita, Uhh, Dan Tingkat Kemiskinan Terhadap Ketimpangan Distribusi Pendapatan Di Kbi Dan Kti," *Jesya (Jurnal Ekon. dan Ekon. Syariah)*, vol. 6, no. 1, pp. 535–544, 2023.
- [6] A. N. Insany, M. Fajri, and others, "Pemodelan IPM Di Kawasan Timur Indonesia Menggunakan Multivariate Adaptive Regression Spline (MARS)," *Nat. Sci. J. Sci. Technol.*, vol. 8, no. 2, pp. 94–98, 2019.
- [7] M. Y. Darsyah, "Lasifikasi indeks pembangunan manusia (ipm) dengan pendekatan k-nearest neighbor (k-nn)," in *Prosiding Seminar Nasional & Internasional*, 2017.
- [8] W. Yustanti, "Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah," *J. Mat. Stat. dan Komputasi*, vol. 9, no. 1, pp. 57–68, 2012.
- [9] K. Latifah, "Kombinasi Algorithma K-NN dan Manhattan Distance untuk Menentukan Pemenang Lelang," *J. Inform. Upgris*, vol. 1, no. 1 Juni, 2015.
- [10] R. K. Dinata, H. Akbar, and N. Hasdyna, "Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 104–111, 2020.
- [11] W. Wahyono, I. N. P. Trisna, S. L. Sariwening, M. Fajar, and D. Wijayanto, "Perbandingan penghitungan jarak pada k-nearest neighbour dalam klasifikasi data tekstual," *J. Teknol. dan Sist. Komput.*, vol. 8, no. 1, pp. 54–58, 2020.
- [12] L. Hakim and A. Saefudin, *Introduction to Machine Learning Using R*, 1st ed. Bogor: IPB Press, 2022.
- [13] T. Wahyono, "Fundamental of Python for Machine Learning," *Yoyakarta Penerbit Gava Media*, 2018.
- [14] F. Gorunescu, *Data Mining: Concepts, models and techniques*, vol. 12. Springer Science & Business Media, 2011.