

Implementasi Algoritma *K-Nearest Neighbor* Dalam Klasifikasi Indeks Pembangunan Manusia Provinsi Sumatera Selatan Tahun 2023

Ira Safitri^{1*}, Ardika Satria², Rizty Maulida Badri³

^{1,2,3}Institut Teknologi Sumatera

¹ira.safitri@sd.itera.ac.id, ²ardika.satria@sd.itera.ac.id, ³rizty.badri@sd.itera.ac.id



Histori Artikel:

Diajukan: 31 August 2024

Disetujui: 18 September 2024

Dipublikasi: 30 September 2024

Kata Kunci:

Indeks Pembangunan Manusia; K-Nearest Neighbour; Validasi Silang; Standar Hidup; LOOCV

Digital Transformation Technology (Digitech) is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak

Indeks Pembangunan Manusia (IPM) merupakan indikator yang sangat penting dalam mengukur tingkat kesejahteraan suatu daerah. IPM memiliki tiga dimensi penting seperti kesehatan, pengetahuan dan standar hidup yang layak. Dimensi kesehatan dapat diukur melalui angka harapan hidup saat lahir, sementara dimensi pengetahuan menggunakan kombinasi indikator harapan lama sekolah dan rata-rata tahun sekolah. Sedangkan dimensi standar hidup yang layak mengacu pada kemampuan daya beli masyarakat dalam memenuhi kebutuhan pokok, dilihat dari rata-rata pengeluaran per kapita yang disesuaikan. Pada penelitian ini, IPM Tahun 2023 di Provinsi Sumatera Selatan diklasifikasikan menjadi 2 kategori yaitu IPM rendah (IPM < 70) dan IPM tinggi (IPM > 70). Klasifikasi kategori IPM dilakukan menggunakan metode KNN dengan teknik validasi silang *Leave-One-Out Cross Validation* (LOOCV). Metode ini dilakukan dengan mencari k objek dalam data latih yang paling dekat dengan objek pada data uji. Penelitian ini memiliki tujuan agar mengetahui akurasi klasifikasi IPM dengan metode k-NN dengan nilai k sebesar 1, 3, dan 5. Berdasarkan hasil evaluasi model KNN menggunakan *confusion matrix* didapatkan nilai k = 1 dengan akurasi sebesar 88,89%, sedangkan k = 3 tingkat akurasi 83,33% kemudian nilai k = 5 didapatkan akurasi 77,78%. Hasil evaluasi model KNN didapatkan nilai k = 1 untuk mendapatkan klasifikasi data yang optimal dengan tingkat akurasi sebesar 88,89%.

PENDAHULUAN

Indonesia memiliki jumlah penduduk sebesar 257.9.12.349 jiwa sehingga berada di posisi ke 4 terbesar di dunia. Peningkatan jumlah penduduk menyebabkan permasalahan seperti kemiskinan, pengangguran dan kualitas sumber daya manusia yang rendah (Abdy et al., 2019). Provinsi Sumatera Selatan merupakan provinsi di Indonesia yang memiliki jumlah penduduk mengalami peningkatan yang sangat pesat. Berdasarkan sensus penduduk 2020 terdapat penambahan jumlah penduduk mencapai 1,02 juta jiwa (Badan Pusat Statistik Kota Palembang, 2020). Peningkatan jumlah penduduk akan mempengaruhi keberhasilan pembangunan nasional.

Pembangunan di suatu daerah dapat diukur dengan suatu indikator yang disebut sebagai Indeks Pembangunan Manusia (IPM) (Putra, 2015). Indeks Pembangunan Manusia memiliki tiga komponen dasar seperti dimensi kesehatan, dimensi pengetahuan, serta dimensi hidup layak. Menurut Badan Pusat Statistik, Indeks Pembangunan Manusia terbagi 4 kategori antara lain Indeks Pembangunan Manusia (IPM) rendah jika < 60, sedang $60 \leq \text{IPM} < 70$, tinggi $70 \leq \text{IPM} < 80$ serta sangat tinggi ≥ 80 (Finaka, n.d.). Pembangunan di Indonesia tidak merata jika IPM di setiap daerah terutama di kabupaten maupun kota sehingga akan sangat beragam (Darsyah, 2014). Indeks Pembangunan Manusia (IPM) di Provinsi Sumatera Selatan di setiap kabupaten/kota mengalami peningkatan di tahun 2023 dibandingkan tahun sebelumnya (Caron & Markusen, 2016).

Beberapa penelitian yang sudah dilakukan oleh peneliti sebelumnya terkait dengan metode *K-Nearest Neighbor* berjudul Perbandingan Algoritma *Naive Bayes* dan *K-Nearest Neighbor (KNN)* untuk Mengetahui Keakuratan Diagnosa Penyakit Diabetes yang menghasilkan akurasi klasifikasi 71% (Puteri et al., 2023). Penelitian yang dilakukan oleh (Fauzi, 2014) menghasilkan akurasi klasifikasi dengan metode KNN sebesar 91.64%. Penelitian terkait IPM yang dilakukan oleh (Rizqa Lina & Wati, 2023) menggunakan KNN menghasilkan akurasi klasifikasi dengan jarak *Manhattan* sebesar 94,12% dan jarak *Euclidian* sebesar 76,47%.

Pada penelitian ini, dilakukan klasifikasi kategori IPM Tahun 2023 di Provinsi Sumatera Selatan menggunakan metode KNN dengan teknik validasi silang *Leave-One-Out Cross Validation* (LOOCV). Metode ini dilakukan dengan mencari k objek dalam data uji yang paling dekat dengan objek pada data latih. Penelitian ini memiliki tujuan agar mengetahui akurasi klasifikasi IPM dengan metode k-NN dengan nilai k sebesar 1, 3, dan 5 sehingga akan mendapatkan akurasi klasifikasi IPM Tahun 2023 paling terbaik di Provinsi Sumatera Selatan.

STUDI LITERATUR

Klasifikasi merupakan pengelompokkan data menjadi kelas berdasarkan kelas yang sudah ada. *K-Nearest Neighbour (k-NN)* termasuk ke dalam metode klasifikasi yang terdapat pada data mining selain *Support Vector Machine (SVM)*. k-NN dilakukan dengan mencari k objek dalam data latih yang paling dekat dengan objek pada data uji. Metode KNN berfungsi dalam mengklasifikasikan data berdasarkan kategori yang sudah ditentukan sebelumnya dengan memperhatikan jarak antar data yang akan diklasifikasikan dengan data latih. Adapun cara yang digunakan dalam mengukur jarak antara data uji dengan data latih seperti *euclidean distance* (Fauziah et al., 2022). Rumus perhitungan pengukuran jarak *Euclidean* sebagai berikut.

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2} \tag{1}$$

dengan $a = a_1, a_2, \dots, a_n$ dan $b = b_1, b_2, \dots, b_n$ mewakili n nilai atribut dari dua *record*.

$$D(X, Y) = \sqrt{\sum_{i=1}^p (X_i - Y_i)^2} \tag{2}$$

Keterangan:

- D (X, Y) : Jarak Euclidean
- X_i : Sampel data
- Y_i : Data uji
- p : Dimensi data
- i : Variabel data

K-Fold Cross Validation merupakan teknik pembagian data menjadi K subset dengan ukuran yang sama yaitu data pelatihan dan data validasi (aldi prajela, syafriandi, dony permana, n.d.). Setiap subset digunakan secara bergantian sebagai data validasi, sementara subset yang lain digunakan sebagai data latih. Dalam prosesnya pengulangan ini dilakukan sebanyak k kali, yang kemudian membentuk subset pada setiap pengulangan. *K-fold cross validation* juga digunakan untuk mengevaluasi kinerja dan kemampuan pada model *K-Nearest Neighbor*, serta membantu mengurangi *overfitting* dan *underfitting* (Hafid, 2023).

Leave-one-out Cross-Validation (LOOCV) merupakan versi LpOCV yang dapat disederhanakan. Aplikasi utama validasi silang agar mampu mengevaluasi kinerja model pembelajaran mesin sehingga dapat membandingkan metode pembelajaran mesin dan menentukan model terbaik dalam pemecahan untuk masalah. Dalam hal ini *Cross Validation* memiliki peran terhadap kepastian model sudah baik sehingga siap masuk ke tahap pengujian yang memiliki data relatif berbeda dengan data latih sebab tujuan tahap pengujian menjadi landasan apakah model dapat digunakan secara luas (Berrar, 2019).

Confusion matrix dapat digambarkan berupa sebaran kesalahan antara kelas atau kategori yang berbeda selama proses klasifikasi. *Confusion matrix* memiliki tabel angka yang menunjukkan kesalahan yang terjadi ketika memprediksi kelas-kelas yang berbeda. *Confusion matrix* sangat penting dalam memperhatikan kesalahan dalam memprediksi satu kelas tidak lebih besar daripada yang lain (Barranco-Chamorro & Carrillo-García, 2021). Jika ada perbedaan yang besar, dapat penyebab nya berupa kesalahan cara melakukan prediksi atau mengenali kelas-kelas tersebut. *Confusion matrix* memiliki empat komponen yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, *False Negative (FN)*. Matriks evaluasi kinerja model klasifikasi yang dihitung dengan *confusion matrix* yaitu ada akurasi (*accuracy*) untuk mengukur seberapa sering model benar memprediksi kelas, presisi (*precision*) untuk mengukur seberapa sering prediksi positif dari model benar, sensitivitas (*recall*) untuk mengukur seberapa sering model dapat menemukan kelas positif sebenarnya, dan nilai F1 (*F1-score*) untuk mengukur keseimbangan antara presisi dan recall (Suryadewiansyah et al., 2020). Untuk menghitung metrik evaluasi ini menggunakan persamaan :

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$\text{Presisi} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{F1-Score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{5}$$

Keterangan :

- True Positive* : Jumlah observasi yang benar diprediksi positif oleh model

- True Negative* : Jumlah observasi yang benar diprediksi negatif oleh model
- False Negative* : Jumlah Prediksi negatif yang salah oleh model
- False Positive* : Jumlah prediksi positif yang salah oleh model
- Precision* : Persentase dari prediksi positif dari model benar
- Recall* : Persentase kasus positif yang terdeteksi dari total kasus positif yang sebenarnya

METODE

Penelitian ini menggunakan data sekunder yang bersumber dari BPS Provinsi Sumatera Selatan. Variabel respon yang digunakan yaitu Indeks Pembangunan Manusia Tahun 2023 menggunakan UHH SP2020 (IPM UHH SP2020) (BPS, n.d.). Dalam penelitian ini diukur dimensi yang terdapat di IPM melalui Umur Harapan Hidup (UHH), Rata-rata Lama Sekolah (RLS), Harapan Lama sekolah (HLS), dan Pengeluaran Per kapita. Adapun deskripsi data penelitian sebagai berikut.

Tabel 1. Deskripsi data penelitian

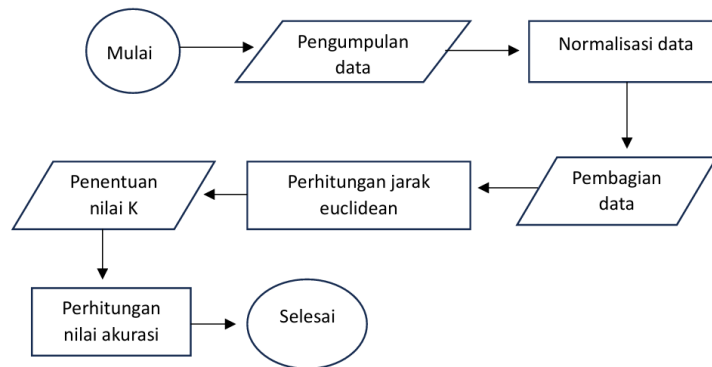
Variabel	Definisi Operasional	Kategori	Keterangan
Y	Indeks Pembangunan Manusia	0 = IPM rendah 1 = IPM Tinggi	Kategorik
X ₁	Umur Harapan Hidup	-	Numerik
X ₂	Rata-rata lama sekolah	-	Numerik
X ₃	Harapan lama sekolah	-	Numerik
X ₄	Pengeluaran per kapita	-	Numerik

Tabel 2. Dataset penelitian

No.	Kabupaten/Kota	IPM	UHH	RLS	HLS	PPK	Kategori
1	Ogan Komering Ulu	73,17	68,85	8,5	12,99	10859	Tinggi
2	Ogan Komering Ilir	70,8	69,31	8,85	11,97	11736	Tinggi
3	Muara Enim	72,33	69,72	7,08	12,01	11683	Tinggi
4	Lahat	72,15	66,87	8,14	12,44	10811	Tinggi
5	Musi Rawas	70,52	68,95	8,56	12,22	10429	Tinggi
6	Musi Banyuasin	71,3	69,51	7,56	12,31	10862	Tinggi
7	Banyu Asin	70,67	69,76	7,68	12,09	10731	Tinggi
8	Ogan Komering Ulu Selatan	69,05	67,88	7,46	11,79	9176	Rendah
9	Ogan Komering Ulu Timur	73,05	69,78	8,05	12,47	12046	Tinggi
10	Ogan Ilir	72,01	66,33	8,07	12,41	11456	Tinggi
11	Empat Lawang	69,07	65,73	8,08	12,1	10375	Rendah
12	Penukal Abab Lematang Ilir	68,78	68,96	7,66	12,49	9139	Rendah
13	Musi Rawas Utara	69,63	66,37	7,08	11,7	10631	Rendah
14	Kota Palembang	81,72	71,99	7,5	14,45	15592	Tinggi
15	Kota Prabumulih	78,07	71,27	10,92	13,31	13494	Tinggi
16	Kota Pagar Alam	73,25	67,7	10,35	13,49	9683	Tinggi
17	Kota Lubuklinggau	78,36	70,25	9,43	13,55	14211	Tinggi
18	Sumatera Selatan	73,18	70,66	9,93	12,63	11472	Tinggi

Berdasarkan klasifikasi diperoleh 2 kategori IPM Tahun 2023 pada kabupaten/kota di Provinsi Sumatera Selatan, yaitu IPM tinggi dan IPM rendah . Berdasarkan hal tersebut, pada penelitian ini klasifikasi dari IPM disederhanakan menjadi 2 kategori, yaitu IPM rendah (IPM < 70) dan IPM tinggi (IPM >70). Klasifikasi yang dilakukan menghasilkan 14 kabupaten/kota yang tergolong dalam kategori IPM tinggi dan 4 kabupaten/kota dengan kategori IPM rendah. Hasil dari klasifikasi ditunjukkan pada Tabel 2.

Klasifikasi Indeks Pembangunan Manusia (IPM) Tahun 2023 di Provinsi Sumatera Selatan menggunakan algoritma *K-Nearest Neighbour* (*k-NN*). Proses pengolahan data dilakukan menggunakan bahasa pemrograman python. Diagram alir penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

Adapun tahapan penelitian ini pada Gambar 1 dapat dijelaskan sebagai berikut:

1. Tahapan pertama merupakan proses awal dalam melakukan penelitian ini yaitu pengumpulan data penelitian. Data yang digunakan dalam penelitian ini bersifat data sekunder yang bersumber dari BPS Sumatera Selatan. Pada tahapan ini, data dari setiap variabel yang digunakan digabungkan menjadi satu buah data. Penggabungan data penelitian ini disusun berdasarkan kabupaten/kota yang ada di Provinsi Sumatera Selatan tahun 2023.
2. Tahapan kedua yaitu proses yang dilakukan setelah pengumpulan data berupa proses normalisasi data. Normalisasi data dilakukan dalam proses ini bertujuan agar data dapat diseragamkan untuk menghilangkan bias yang dapat timbul akibat perbedaan skala. Pada proses normalisasi data dengan metode *min max scaling* menggunakan persamaan 6 dibawah ini.

$$x_{baru} = \frac{x_{lama} - x_{min}}{x_{max} - x_{min}} \quad (6)$$

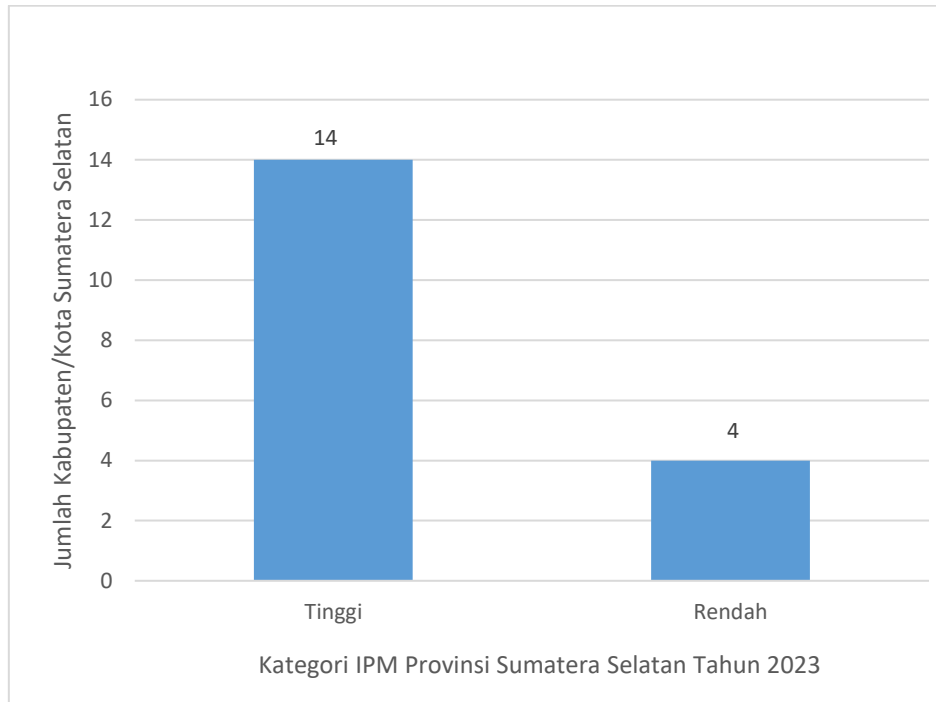
Dimana,

- X_{baru} = Nilai setelah dinormalisasi
- X_{lama} = Nilai awal sebelum dinormalisasi
- X_{max} = Nilai tertinggi dari variabel
- X_{min} = Nilai terendah dari variabel

3. Tahapan selanjutnya adalah proses pembagian data. Dalam proses ini, terlebih dahulu melakukan pemilihan data uji. Pentingnya pemilihan data uji dalam validasi silang metode LOOCV, data uji yang digunakan hanya satu. Pemilihan satu data uji ini dari seluruh dataset lalu baris data lainnya akan menjadi data latih. Kemudian data uji yang telah dipilih akan dipisahkan dari data lainnya sehingga dalam proses pembagian data ini terbagi menjadi 2 kategori yaitu data uji dan data latih.
4. Selanjutnya dilakukan proses perhitungan jarak *euclidean*. Tujuan dilakukan proses ini untuk mengetahui seberapa dekat atau jauh setiap data latih dengan data uji. Jika data latih yang memiliki nilai jarak Euclidean kecil dengan data uji maka dapat di anggap sebagai tetangga terdekat dari data uji tersebut.
5. Proses selanjutnya adalah penentuan nilai k. Dalam proses ini terlebih dahulu dapat klasifikasi data uji.maka tahapan selanjutnya dapat menentukan kategori dari suatu data uji dilakukan dengan mempertimbangkan kelas mayoritas dari k data latih terdekat.
6. Tahapan akhir, adalah proses perhitungan nilai akurasi. Proses ini akan menghasilkan nilai prediksi klasifikasi dari setiap data uji dibandingkan dengan klasifikasi yang sebenarnya. Jumlah prediksi yang benar dan salah dimasukkan kedalam tabel confusion matriks. Akurasi diukur dari seberapa sering model membuat prediksi yang benar.

HASIL

Klasifikasi IPM Tahun 2023 Provinsi Sumatera Selatan pada penelitian ini menghasilkan bahwa mayoritas wilayah kabupaten/kota di provinsi Sumatera Selatan termasuk kategori IPM berstatus tinggi dengan jumlah 14 dari 18 kabupaten/kota dan memiliki berstatus rendah terdapat 4 kabupaten/kota. Jumlah dari kabupaten/kota dari klasifikasi tersebut dapat dilihat pada Gambar 2.



Gambar 2. Jumlah Kabupaten/Kota di Provinsi Sumatera Selatan Berdasarkan Kategori IPM

Berdasarkan Gambar 2 menunjukkan bahwa nilai IPM tertinggi terdapat pada Kota Palembang yaitu sebesar 81,72 sedangkan wilayah yang memiliki kategori IPM rendah meliputi Ogan Komering Ulu Selatan, Empat Lawang, Penukal Abab Lematang Ilir, dan Musi Rawas Utara.

Selanjutnya, variabel dari setiap data pada penelitian ini dilakukan normalisasi agar menghasilkan nilai-nilai data dengan skala yang sama. Hal penting dalam kalsifikasi untuk memastikan hasil klasifikasi yang konsisten. Berikut ini data yang dihasilkan setelah dilakukan normalisasi ditunjukkan pada Tabel 3.

Tabel 3. Data normalisasi

No.	Kabupaten/Kota	IPM	UHH	RLS	HLS	Pengeluaran per kapita
1	Ogan Komering Ulu	0,339	0,498	0,370	0,469	0,267
2	Ogan Komering Ilir	0,156	0,572	0,461	0,098	0,402
3	Muara Enim	0,274	0,637	0,000	0,113	0,394
4	Lahat	0,260	0,182	0,276	0,269	0,259
5	Musi Rawas	0,134	0,514	0,385	0,189	0,200
6	Musi Banyuasin	0,195	0,604	0,125	0,222	0,267
7	Banyu Asin	0,146	0,644	0,156	0,142	0,247
8	Ogan Komering Ulu Selatan	0,021	0,343	0,099	0,033	0,006
9	Ogan Komering Ulu Timur	0,330	0,647	0,253	0,280	0,450
10	Ogan Ilir	0,250	0,096	0,258	0,258	0,359
11	Empat Lawang	0,022	0,000	0,260	0,145	0,192
12	Penukal Abab Lematang Ilir	0,000	0,516	0,151	0,287	0,000
13	Musi Rawas Utara	0,066	0,102	0,000	0,000	0,231
14	Kota Palembang	1,000	1,000	0,109	1,000	1,000
15	Kota Prabumulih	0,718	0,885	1,000	0,585	0,675
16	Kota Pagar Alam	0,345	0,315	0,852	0,651	0,084
17	Kota Lubuklinggau	0,740	0,722	0,612	0,673	0,786
18	Sumatera Selatan	0,340	0,788	0,742	0,338	0,362

Data yang telah dinormalisasi menggunakan metode *min max scaling* dengan rentang angka 0 dan 1. Nilai normalisasi yang mendekati 1 menunjukkan bahwa data tersebut mewakili nilai maksimum pada variabel, jika dibandingkan nilai normalisasi menghasilkan angka 0 mewakili nilai minimum pada variabel. Selanjutnya, data dinormalisasi akan dilakukan pembagian data yang bertujuan agar dapat di analisis lebih lanjut. Pembagian data pada penelitian ini menggunakan metode *Leave-One-Out Cross Validation (LOOCV)*. Pada penelitian ini metode LOOCV akan melakukan iterasi sebanyak 18 kali dan pada setiap iterasi akan diambil satu data sebagai data uji dan sisa baris data lainnya menjadi set data latih.

PEMBAHASAN

Selanjutnya, penentuan tetangga atau objek yang terdekat dengan data uji didapatkan dengan menghitung jarak antara data uji dengan semua data latih menggunakan perhitungan *euclidean distance*. Tujuan dilakukan perhitungan jarak *Euclidean* pada setiap baris data dapat memberikan hasil klasifikasi yang lebih akurat. Hasil perhitungan *euclidean distance* dari setiap iterasi dapat dilihat pada Tabel 4. dengan angka pada penamaan D1, D2, ..., D18 yang dapat disimbolkan sebagai baris data uji yang digunakan untuk menghitung jarak *Euclidean*.

Tabel 4. Euclidean Distance pada setiap iterasi

D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18
-	0,86	0,81	0,56	0,71	0,73	0,74	0,36	0,93	0,52	0,36	0,61	0,26	2,00	1,76	1,17	1,59	1,24
0,89	-	0,81	0,56	0,71	0,73	0,74	0,36	0,93	0,52	0,36	0,61	0,26	2,00	1,76	1,17	1,59	1,24
0,89	0,86	-	0,56	0,71	0,73	0,74	0,36	0,93	0,52	0,36	0,61	0,26	2,00	1,76	1,17	1,59	1,24
0,89	0,86	0,81	-	0,71	0,73	0,74	0,36	0,93	0,52	0,36	0,61	0,26	2,00	1,76	1,17	1,59	1,24
0,89	0,86	0,81	0,56	-	0,73	0,74	0,36	0,93	0,52	0,36	0,61	0,26	2,00	1,76	1,17	1,59	1,24
0,89	0,86	0,81	0,56	0,71	-	0,74	0,36	0,93	0,52	0,36	0,61	0,26	2,00	1,76	1,17	1,59	1,24
0,89	0,86	0,81	0,56	0,71	0,73	-	0,36	0,93	0,52	0,36	0,61	0,26	2,00	1,76	1,17	1,59	1,24
0,89	0,86	0,81	0,56	0,71	0,73	0,74	-	0,93	0,52	0,36	0,61	0,26	2,00	1,76	1,17	1,59	1,24
0,89	0,86	0,81	0,56	0,71	0,73	0,74	0,36	-	0,52	0,36	0,61	0,26	2,00	1,76	1,17	1,59	1,24
0,89	0,86	0,81	0,56	0,71	0,73	0,74	0,36	0,93	-	0,36	0,61	0,26	2,00	1,76	1,17	1,59	1,24
0,89	0,86	0,81	0,56	0,71	0,73	0,74	0,36	0,93	0,52	-	0,61	0,26	2,00	1,76	1,17	1,59	1,24
0,89	0,86	0,81	0,56	0,71	0,73	0,74	0,36	0,93	0,52	0,36	-	0,26	2,00	1,76	1,17	1,59	1,24
0,89	0,86	0,81	0,56	0,71	0,73	0,74	0,36	0,93	0,52	0,36	0,61	-	2,00	1,76	1,17	1,59	1,24
0,89	0,86	0,81	0,56	0,71	0,73	0,74	0,36	0,93	0,52	0,36	0,61	0,26	-	1,76	1,17	1,59	1,24
0,89	0,86	0,81	0,56	0,71	0,73	0,74	0,36	0,93	0,52	0,36	0,61	0,26	2,00	-	1,17	1,59	1,24
0,89	0,86	0,81	0,56	0,71	0,73	0,74	0,36	0,93	0,52	0,36	0,61	0,26	2,00	1,76	-	1,59	1,24
0,89	0,86	0,81	0,56	0,71	0,73	0,74	0,36	0,93	0,52	0,36	0,61	0,26	2,00	1,76	1,17	-	1,24
0,89	0,86	0,81	0,56	0,71	0,73	0,74	0,36	0,93	0,52	0,36	0,61	0,26	2,00	1,76	1,17	1,59	-

Berdasarkan nilai-nilai pada Tabel 4, objek terdekat dipilih dari nilai dengan *euclidean distance* terkecil hingga mencapai jumlah k tetangga. Penentuan kategori dari nilai IPM ditentukan kelas mayoritas objek terdekat dengan jumlah k tetangga. Hasil klasifikasi Indeks Pembangunan Manusia (IPM) menggunakan validasi silang LOOCV pada algoritma KNN dengan nilai K = 1, 3, 5 direpresentasikan dalam bentuk confusion matrix pada Tabel 5 sebagai berikut:

Tabel 5. Confusion matriks dengan nilai k (1, 3, 5)

Nilai k	Aktual		Prediksi		Akurasi (%)
			Rendah	Tinggi	
K= 1	Y	Rendah	2	0	88,89
		Tinggi	2	14	
K= 3	Y	Rendah	2	1	83,33
		Tinggi	2	13	

K= 5	Y	Rendah	0	0	77,78
		Tinggi	4	14	
		Tinggi	4	14	

Berikut penjelasan dari setiap pemilihan nilai k terhadap hasil klasifikasi pada data uji:

- Nilai K=1, pada model ini didapatkan akurasi yang tinggi dengan nilai akurasi mencapai 88,89%. Hal ini menunjukkan bahwa dengan 1 objek tetangga akan memberikan model hasil klasifikasi yang optimal.
- Nilai K=3, model yang menggunakan 3 objek tetangga terdekat memberikan nilai akurasi yang cukup baik yaitu sebesar 83,33%. model ini cukup efektif dalam mengklasifikasikan data pada studi penelitian ini.
- Nilai K=5, terdapat penurunan tingkat akurasi sebesar 77,78% pada penambahan objek tetangga terdekat menjadi 5 hal ini bisa terjadi akibat dari beberapa objek yang memberikan informasi yang *noise*.

Berdasarkan Tabel 5 hasil penelitian ini menunjukkan ketika $k = 1$ menghasilkan nilai akurasi sebesar 88,89% tetapi pada saat nilai K ditingkatkan dari 1 menjadi 3, dan 5 membuat nilai dari akurasi klasifikasi menurun. Hal tersebut dapat terjadi karena penambahan objek tetangga memberikan informasi yang tidak akurat sehingga nilai $k = 3, 5$ belum cukup baik dalam memprediksi tingkat IPM di Provinsi Sumatera Selatan pada tahun 2023.

KESIMPULAN

Penelitian ini melakukan klasifikasi pada data Indeks Pembangunan Manusia (IPM) Tahun 2023 di Provinsi Sumatera Selatan dengan menggunakan algoritma *K-Nearest Neighbors (KNN)* dan validasi silang *Leave-One Out Cross-Validation (LOOCV)*. Pengklasifikasian Indeks Pembangunan Manusia (IPM) menggunakan variabel Umur Harapan Hidup (UHH), Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Pengeluaran Perkapita dari 18 Kabupaten/Kota di Provinsi Sumatera Selatan pada tahun 2023. Hasil penelitian menunjukkan bahwa ketika $k = 1$ menghasilkan nilai akurasi sebesar 88,89% tetapi pada saat nilai K ditingkatkan dari 1 menjadi 3, dan 5 membuat nilai dari akurasi klasifikasi menurun. Hal tersebut dapat terjadi karena penambahan objek tetangga memberikan informasi yang tidak akurat.

REFERENSI

- Abdy, M., Syam, R., & Haryanensi, E. (2019). Metode Automatic clustering-fuzzy logical relationships pada Peramalan Jumlah Penduduk di Kota Makassar. *Journal of Mathematics, Computations, and Statistics*, 1(2), 193. <https://doi.org/10.35580/jmathcos.v1i2.9242>
- aldi prajela, syafriandi, dony permana, dina fitria. (n.d.). *Analisis Sentimen Data Twitter Calon Presiden 2024 Menggunakan Algoritma Naïve Bayes Classifier Dengan Metode K-Fold Cross Validation.pdf*.
- Badan Pusat Statistik Kota Palembang. (2020). Hasil Sensus Penduduk Provinsi Sumatera Selatan 2020. *BPS Provinsi Sumatera Selatan*, 1–12. <https://sumsel.bps.go.id/pressrelease/2021/01/21/549/hasil-sensus-penduduk-provinsi-sumatera-selatan-2020.html>
- Barranco-Chamorro, I., & Carrillo-García, R. M. (2021). Techniques to deal with off-diagonal elements in confusion matrices. *Mathematics*, 9(24). <https://doi.org/10.3390/math9243233>
- Berrar, D. (2019). *Cross-Validation* (S. Ranganathan, M. Gribskov, K. Nakai, & C. B. T.-E. of B. and C. B. Schönbach (eds.); pp. 542–545). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-809633-8.20349-X>
- BPS. (n.d.). *Indeks Pembangunan Manusia*. BPS Sumatera Selatan. Retrieved August 30, 2024, from <https://sumsel.bps.go.id/id/statistics-table/2/MjA5IzI=/indeks-pembangunan-manusia.html>
- Caron, J., & Markusen, J. R. (2016). 濟無No Title No Title No Title. [Metode Baru] Indeks Pembangunan Manusia per Kabupaten/Kota Se Sumatera Selatan, 2022-2023. <https://sumsel.bps.go.id/id/statistics-table/2/MjA5IzI=/indeks-pembangunan-manusia.html>
- Darsyah, M. Y. (2014). Klasifikasi Indeks Pembangunan Manusia (IPM) dengan Pendekatan K-Nearset Neighbor (K-NN). *Seminar Nasional Pendidikan, Sains Dan Teknologi Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang*, 29–35. https://www.researchgate.net/publication/339470520_Klasifikasi_Indeks_Pembangunan_Manusia_Ipm_Dengan_Pendekatan_K-Nearset_Neighbor_K-Nn

- Fauzi, F. (2014). *Jurnal MIPA*. 37(2), 105–114.
- Fauziah, Tiro, M. A., & Ruliana. (2022). Comparison of k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) Methods for Classification of Poverty Data in Papua. *ARRUS Journal of Mathematics and Applied Science*, 2(2), 83–91. <https://doi.org/10.35877/mathscience741>
- Finaka, A. W. (n.d.). *Indeks Pembangunan Manusia Terus Meningkatkan*. Indonesia Baik. Retrieved August 30, 2024, from <https://indonesiabaik.id/infografis/indeks-pembangunan-manusia-terus-meningkat>
- Hafid, H. (2023). Penerapan K-Fold Cross Validation untuk Menganalisis Kinerja Algoritma K-Nearest Neighbor pada Data Kasus Covid-19 di Indonesia. *Journal of Mathematics*, 6(2), 161–168. <http://www.ojs.unm.ac.id/jmathcos>
- Puteri, Q. A., Sagirani, T., & Lemantara, J. (2023). Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbor (KNN) untuk Mengetahui Keakuratan Diagnosa Penyakit Diabetes. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 9(3), 247–254. <https://doi.org/10.25077/teknosi.v9i3.2023.247-254>
- Putra, D. M. (2015). *Provinsi Jawa Timur Dengan Menggunakan Modelling Human Development Index (Hdi) of East Java Province Using*. 96, 22.
- Rizqa Lina, I., & Wati, D. C. (2023). Klasifikasi Pengeluaran per Kapita di Tiga Provinsi Sulawesi menggunakan K-Nearest Neighbor. *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, 16(1), 395–406. <https://doi.org/10.36456/jstat.vol16.no1.a7193>
- Suryadewiansyah, M. K., Endra, T., & Tju, E. (2020). Naïve Bayes dan Confusion Matrix untuk Efisiensi Analisa Intrusion Detection System Alert. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 8(2), 81–88.