

Penerapan Metode *Ensemble K-Nearest Neighbor* untuk Memprediksi Indeks Pembangunan Manusia di Provinsi Sumatera Utara

¹Oktaviana, ²Riri Syafitri Lubis, ³Rina Widyasari

¹²³Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara, Indonesia
¹oktaviana22medan@gmail.com, ²riri_syafitri@uinsu.ac.id, ³rina_widyasari@uinsu.ac.id

ARTICLE INFO

Article History:

Diterima : 04-10-2024
Disetujui : 13-12-2024

Keywords:

Ensemble K-Nearest Neighbor; Human Development Index; Prediction.



ABSTRACT

Abstract: *Human Development Index is one of the important indicators to describe the level of success of a region in an effort to build the quality of its population. The purpose of this study is to determine the results of predicting the value of the Human Development Index in North Sumatra Province using the Ensemble K-Nearest Neighbor method and see how much accuracy using the Ensemble K-Nearest Neighbor method to predict the Human Development Index. One method that can be used in predicting the Human Development Index is the Ensemble K-Nearest Neighbor method. K-Nearest Neighbor is a method that determines the label (class) based on distance. To optimize the performance of the K-NN method in making predictions, ensemble techniques are needed. The results of research using the K-Nearest Neighbor Ensemble method have an average prediction of the Human Development Index of 72.97 with a MAPE value of 2.1%, which is much better than using only the Single K-NN method which has a MAPE value of 2.38%.*

Abstrak: Indeks Pembangunan Manusia adalah salah satu indikator penting untuk menggambarkan tingkat keberhasilan suatu daerah dalam usaha membangun kualitas penduduknya. Tujuan dari penelitian ini yaitu untuk mengetahui hasil prediksi nilai Indeks Pembangunan Manusia di Provinsi Sumatera Utara menggunakan metode *Ensemble K-Nearest Neighbor* serta melihat seberapa besar tingkat akurasi menggunakan metode *Ensemble K-Nearest Neighbor* untuk memprediksi Indeks Pembangunan Manusia. Salah satu metode yang dapat digunakan dalam memprediksi Indeks Pembangunan Manusia adalah metode *Ensemble K-Nearest Neighbor*. *K-Nearest Neighbor* adalah salah satu metode yang menentukan label (class) berdasarkan jarak. Untuk mengoptimalkan kinerja metode K-NN dalam melakukan prediksi diperlukan teknik *ensemble*. Hasil penelitian menggunakan metode *Ensemble K-Nearest Neighbor* memiliki prediksi rata-rata Indeks Pembangunan Manusia sebesar 72,97 dengan nilai MAPE 2,1% hal tersebut jauh lebih baik jika dibandingkan hanya menggunakan metode K-NN Tunggal yang memiliki nilai MAPE sebesar 2,38%.



<https://doi.org/10.31764/justek.vXIY.ZZZ>

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



A. LATAR BELAKANG

Keberhasilan pembangunan manusia disuatu daerah dapat ditinjau dari besarnya permasalahan yang ada pada penduduk yang dapat teratasi. Permasalahan yang ada pada penduduk diantaranya meliputi kemiskinan, pengangguran dan sulitnya mendapatkan fasilitas kesehatan dan pendidikan (Fajaryanto & Alamsyah, 2016).

Menurut Badan Pusat Statistik keberhasilan pembangunan manusia disuatu wilayah dapat diukur dengan cara menghitung Indeks Pembangunan Manusia. Indeks Pembangunan Manusia adalah suatu alat ukur yang dapat menggambarkan tingkat pencapaian atau tingkat keberhasilan suatu daerah dalam usaha membangun kualitas penduduknya yaitu dari segi kesehatan, pendidikan dan pengeluaran perkapitanya (Dewi, 2019).

Tercatat di Badan Pusat Statistik pada tahun 2023 Indeks Pembangunan Manusia di Indonesia yang tertinggi berada di DKI Jakarta dengan nilai IPM 83,55 kemudian, disusul Daerah Istimewa Yogyakarta dengan nilai IPM 81,09 sedangkan, dari 10 Provinsi yang terdapat di Pulau Sumatera Indeks Pembangunan Manusia yang tertinggi berada di Kepulauan Riau dengan nilai IPM 79,08, sedangkan pada Provinsi Sumatera Utara Indeks Pembangunan Manusiannya masih jauh di bawah Provinsi Kepulauan Riau. Berlandaskan tingkat Kabupaten/Kota Indeks Pembangunan Manusia di Provinsi Sumatera Utara sangat beragam, diantaranya termasuk kedalam kategori sedang, rendah dan tinggi. Hal demikian diakibatkan masih kurangnya peran pemerintah dalam usaha memperbaiki pembangunan manusia disuatu wilayah sehingga mengakibatkan Indeks Pembangunan Manusia disuatu wilayah ada yang jauh tertinggal jika dibandingkan dengan daerah lain.

Perlunya dilakukan prediksi terhadap IPM karena prediksi Indeks Pembangunan Manusia dapat membantu pemerintah dalam menentukan kebijakan ataupun strategi kedepan agar dapat meningkatkan kualitas penduduknya dari segi kesehatan, pendidikan, dan pendapatan perkapita. Selain sistem untuk memprediksi Indeks Pembangunan Manusia, telah banyak dilakukan prediksi terhadap bidang lainnya seperti prediksi terhadap nilai tukar rupiah, prediksi terhadap curah hujan, dan lain sebagainya. Dari berbagai masalah dalam prediksi menggunakan data deret waktu, diperlukan metode yang sesuai untuk penerapannya. Metode yang dinilai paling tepat dalam memprediksi IPM yaitu metode Ensemble K-Nearest Neighbor.

Dalam penerapannya pada masalah prediksi, *k-nearest neighbor* telah banyak digunakan untuk penelitian bermacam-macam sistem prediksi yang dilakukan terhadap objek berdasarkan nilai k tetangga terdekat. K-NN adalah metode yang menentukan label (kelas) dari suatu objek baru berdasarkan kelas yang paling sering muncul di antara k -nearest neighbor dalam data latih. K-NN berbasis kesamaan yang dimiliki objek didasarkan pada jarak antara objek yang akan ditentukan dengan objek yang telah ada sebelumnya (Arhami& Nasir, 2020). Meskipun metode *k-nearest neighbor* dianggap sederhana, ia memiliki kelebihan seperti kemampuannya untuk menggeneralisasi himpunan data training yang relatif kecil, efektivitasnya, dan penerapannya yang luas dalam berbagai kasus penelitian, baik untuk klasifikasi maupun prediksi (Wibowo et al., 2013). Namun, metode ini juga memiliki kekurangan, yaitu memerlukan kecermatan untuk menetapkan nilai dari parameter k atau yang biasa disebut dengan jumlah tetangga terdekat. Untuk mengoptimalkan kinerja K-NN digunakan salah satu metode yang dinilai cukup baik dalam melakukan prediksi, adapun digunakan teknik ini karena bertujuan untuk mengoptimalkan kinerja dari metode *k-NN*, teknik ini dinamakan dengan teknik *ensemble*. Penggunaan teknik *ensemble* dilakukan karena metode ini

menerapkan pembobotan atau biasa dikenal dengan *weighted means*, sehingga dalam penerapannya metode ensemble k-NN ini tidak diperlukan lagi pencarian nilai k yang optimal sehingga menghasilkan prediksi yang jauh lebih akurat dibanding hanya menggunakan menggunakan metode k-nn tunggal (Satriya et al., 2018).

Penelitian sebelumnya yang dilakukan oleh Alfin, dkk (2022) dengan judul penerapan metode clustering untuk prediksi produksi bawang merah (ensemble k-nearest neighbor), dalam penelitian tersebut diperoleh hasil bahwa metode ensemble k-nearest neighbor menghasilkan nilai MAPE sebesar 1,46. Metode ensemble k-nearest neighbor ini memberikan prediksi yang lebih akurat dibandingkan dengan metode k-nn tunggal (Alfin et al., 2022).

Berdasarkan uraian di atas, peneliti bermaksud ingin melaksanakan penelitian dengan menerapkan metode ensemble k-nearest neighbor untuk memprediksi indeks pembangunan manusia di Provinsi Sumatera Utara. Adapun dilakukan penelitian ini yaitu bertujuan agar mengetahui hasil prediksi nilai Indeks Pembangunan Manusia di Provinsi Sumatera Utara menggunakan metode ensemble k-nearest neighbor serta melihat seberapa besar tingkat akurasi metode ensemble k-nearest neighbor untuk memprediksi Indeks Pembangunan Manusia.

B. METODE PENELITIAN

Penelitian ini dilakukan di Provinsi Sumatera Utara. Penelitian ini menggunakan pendekatan kuantitatif. Data yang digunakan adalah data sekunder yang diperoleh dari Badan Pusat Statistik Provinsi Sumatera Utara. Data yang digunakan terdiri dari data indeks pembangunan manusia, umur harapan hidup, harapan lama sekolah, rata-rata lama sekolah, pengeluaran per kapita yang disesuaikan, dan jumlah penduduk miskin pada tahun 2022-2023.

Pada penelitian ini menggunakan metode *ensemble k-nearest neighbor*. Adapun tahapan analisis yang dilakukan dengan menggunakan metode *ensemble k-nearest neighbor* adalah sebagai berikut:

1. Mengidentifikasi Masalah
2. Mengumpulkan data mengenai Indeks Pembangunan Manusia yang diperoleh dari Badan Pusat Statistik Provinsi Sumatera Utara.
3. Melakukan perhitungan Normalisasi data menggunakan persamaan (1) (Putra et al., 2023).

$$X_{normalisasi} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Dimana:

$X_{normalisasi}$ = Perolehan hasil data yang telah dinormalisasi

X = Data yang akan dilakukan normalisasi

X_{min} = Nilai minimum dari data yang akan dinormalisasi

X_{max} = Nilai maksimum dari data yang akan dinormalisasi

4. Membagi data menjadi data *training* (data uji) dan data *testing* (data latih) (Jusman et al., 2022). Dimana data dari hasil normalisasi pada periode 1 sampai 33 sebagai data *training* dan data hasil normalisasi pada periode 34 sampai 66 sebagai data *testing*.
5. Perhitungan nilai jarak *euclidean* antara data *testing* (data uji) dan data *training* (data latih) menggunakan persamaan (2) (Kang, 2021)

$$dx_{train-i}, x_{test-j} = \sqrt{\sum_{i,j=1}^n (X_{train-i} - X_{test-j})^2} \quad (2)$$

Dimana:

$dx_{train-i}, x_{test-j}$ = Jarak Euclidean
 $x_{train-i}$ = data *training* dengan nilai ke-i
 x_{test-j} = data *testing* dengan nilai ke-j
 n = jumlah semua data
 i, j = 1,2,3,...,n

Kemudian, menentukan jarak terdekat sampai urutan k

6. Menentukan nilai k , dimana nilai k terdiri dari 3, 4, 5, 10, 13, 15, 18 dan 24.
7. Melakukan perhitungan rata-rata nilai dari k terdekat untuk mengetahui hasil prediksi K-Nearest Neighbor Tunggal (Hidayati, 2019). Adapun perhitungannya menggunakan persamaan (3)

$$Y = \frac{1}{k} \sum_{i=1}^k y_i \quad (3)$$

Dimana:

Y = prediksi menggunakan K-NN tunggal pada k ke-i
 k = banyaknya tetangga terdekat
 y_i = urutan jarak berdasarkan hasil Euclidean

8. Mengevaluasi hasil prediksi dari K-NN tunggal menggunakan nilai MAPE (Mendenhall et al., 1993). Adapun perhitungannya menggunakan persamaan (4)

$$MAPE = \sum_{i=1}^n \frac{1}{m} \cdot \frac{|x_i - y_i|}{x_i} \times 100\% \quad (4)$$

Dimana:

x_i = Nilai data sebenarnya ke - i
 y_i = Nilai prediksi ke - i
 n = banyaknya data yang diteliti
 m = jumlah data yang akan diprediksi

9. Menghitung korelasi, korelasi menggambarkan besarnya hubungan antara data aktual dengan prediksi K-NN tunggal (Santosa & Muliawan, 2007). Adapun perhitungannya menggunakan persamaan (5)

$$r_i = \frac{n \sum \bar{y}_i \hat{y}_t - (\sum \bar{y}_i)(\sum \hat{y}_t)}{\sqrt{(n \sum \bar{y}_i^2 - (\sum \bar{y}_i)^2)(n \sum \hat{y}_t^2 - (\sum \hat{y}_t)^2)}} \quad (5)$$

Dimana:

r_i = koefisien korelasi product moment

\bar{y}_i = hasil prediksi dari K-NN tunggal pada jumlah k ke $-i$,

\hat{y}_t = nilai sebenarnya ke $-t$.

10. Menghitung pembobot *ensemble* (Cohen, 2013) menggunakan rumus (6)

$$W_i = \frac{r_i}{\sum_{j=1}^s r_j} \quad (6)$$

Dimana:

W_i = pembobot ensemble ke- i

r_i = korelasi antara data sebenarnya dengan data hasil prediksi K-Nearest Neighbor

s = banyaknya data yang digunakan

j = 3, 4, 5, 10, 13, 15, 18, 24.

11. Menggabungkan hasil prediksi yang diperoleh (Sinta et al., 2014) dengan menerapkan rumus (7)

$$\mu = \frac{\sum_{i=1}^s w_i y_i}{\sum_{i=1}^s w_i} \quad (7)$$

Dimana:

s = jumlah semua data yang digunakan

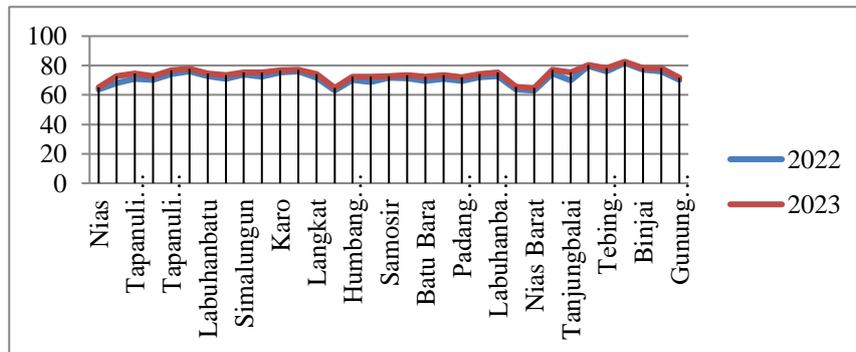
y_i = perolehan nilai dari prediksi pada model ke- i

w_i = merupakan pembobotnya.

12. Melakukan evaluasi terhadap hasil prediksi berlandaskan nilai MAPE (Mendenhall et al., 1993). Adapun perhitungannya menggunakan rumus (4)
13. Kesimpulan
14. Selesai.

C. HASIL DAN PEMBAHASAN

Data Indeks Pembangunan Manusia di Provinsi Sumatera Utara pada tahun 2022-2023 disajikan dalam Gambar 1. Pada Gambar 1 dapat dilihat bahwa nilai Indeks Pembangunan Manusia di Sumatera Utara mengalami tren naik dan juga tren turun. Di Sumatera Utara nilai Indeks Pembangunan Manusia yang tertinggi berada di Kota Medan, memperoleh nilai IPM sebesar 81,76 di tahun 2022. Sedangkan nilai Indeks Pembangunan Manusia yang terendah berada di Nias Selatan. Hal tersebut juga terjadi pada tahun 2023, Indeks Pembangunan Manusia di Provinsi Sumatera tertinggi berada di Kota Medan dan yang terendah berada pada Kabupaten Nias Selatan. Nilai Indeks Pembangunan Manusia yang tinggi dipengaruhi oleh jumlah pendapatan yang tinggi sehingga menyebabkan jumlah pengeluaran juga tinggi. Nilai Indeks Pembangunan Manusia yang tinggi salah satunya dipengaruhi oleh pengeluaran perkapita. Selain itu juga, indeks pendidikan, kesehatan, dan jumlah penduduk miskin juga mempengaruhi tingginya nilai Indeks Pembangunan Manusia (Hidayati, 2019).



Gambar 1. Indeks Pembangunan Manusia di Provinsi Sumatera Utara tahun 2022-2023
Langkah pertama yang dilakukan untuk memprediksi Indeks Pembangunan Manusia menggunakan metode *ensemble k-nearest neighbor* adalah dengan melakukan normalisasi data. Normalisasi data bertujuan untuk menyelaraskan nilai variabel prediktor (variabel x) sehingga berada pada range tertentu (Satriya et al., 2018). Normalisasi yang diterapkan adalah metode normalisasi data Min-Max. Perhitungan normalisasi Min-Max ini dilakukan menggunakan rumus dari persamaan (1) dan dibantu menggunakan *Microsoft Excel*. Untuk hasil perhitungan normalisasi dapat dilihat pada tabel 1.

Tabel 1. Normalisasi Data

No	Kabupaten/Kota	IPM	UHH	RLS	HLS	PPD	JPM
1	Nias	63,69	0,5986	0	0,2435	0,1096	0,1046
2	Mandailing Natal	68,05	0	0,5017	0,5957	0,4105	0,2012
3	Tapanuli Selatan	70,92	0,1904	0,6028	0,387	0,5855	0,1036
4	Tapanuli Tengah	70,31	0,3868	0,5192	0,3304	0,4561	0,2344
5	Tapanuli Utara	74,14	0,5218	0,7178	0,5391	0,6028	0,1277
6	Toba	75,96	0,6584	0,8188	0,4783	0,664	0,0679
7	Labuhanbatu	72,92	0,6251	0,6132	0,2043	0,5589	0,2137
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
66	Gunung Sitoli 2	71,55	0,9377	0,4826	0,5652	0,2608	0,0981

Setelah dilakukan normalisasi data, selanjutnya menghitung jarak *Euclidean*. Jarak *euclidean* digunakan untuk mengukur jarak terdekat antara data *training* dan data *testing* (Jusman et al., 2022). Hasil yang diperoleh pada normalisasi data periode 1 sampai 33 digunakan sebagai data *training*, sedangkan data pada periode 33 sampai 66 sebagai data *testing*. Adapun perhitungan untuk jarak *Euclidean* ini dilakukan menggunakan rumus dari persamaan (2) dan dibantu menggunakan *software R studio*. Hasil perhitungan jarak *Euclidean* dapat dilihat pada tabel 2.

Tabel 2. Jarak *Euclidean*

No	Kabupaten/Kota	Nias	Mandailing Natal	Tapanuli Selatan	...	Gunung Sitoli 2
1	Nias	0	0.9132345	0.8814728	...	0.6886349
2	Mandailing Natal	0.9132345	0	0.3607845	...	0.9558062
3	Tapanuli Selatan	0.8814728	0.3607845	0	...	0.8426427
⋮	⋮	⋮	⋮	⋮	...	⋮
66	Gunung Sitoli 2	0.6886349	0.9558062	0.8426427	...	0

Langkah selanjutnya, menentukan nilai k . Pada penelitian ini menggunakan nilai $k = 3, 4, 5, 10, 13, 15, 18,$ dan 24 . Dari hasil perhitungan jarak *Euclidean*, maka dapat diketahui jarak terdekat dari masing-masing data *training* untuk data *testing*. Adapun perhitungan untuk jarak terdekat untuk setiap data testing dilakukan menggunakan bantuan *software* R studio. Hasil perhitungannya dapat dilihat pada tabel 3.

Tabel 3. Jarak terdekat untuk data *testing* 1

Jumlah k	Data Jarak Terdekat
$k = 3$	1, 24 dan 25
$k = 4$	1, 24, 25 dan 14
$k = 5$	1, 24, 25, 14 dan 33
$k = 10$	1, 24, 25, 14, 33, 17, 19, 4, 21 dan 15.
$k = 13$	1, 24, 25, 14, 33, 17, 19, 4, 21, 15, 23, 22 dan 18
$k = 15$	1, 24, 25, 14, 33, 17, 19, 4, 21, 15, 23, 22, 18, 10 dan 7
$k = 18$	1, 24, 25, 14, 33, 17, 19, 4, 21, 15, 23, 22, 18, 10, 7, 20, 13 dan 16
$k = 24$	1, 24, 25, 14, 33, 17, 19, 4, 21, 15, 23, 22, 18, 10, 7, 20, 13, 16, 8, 9, 5, 11, 3 dan 26.

Begitu seterusnya, sampai data jarak terdekat untuk data *testing* 33. Selanjutnya, untuk memperoleh hasil prediksi variabel prediktor yang dipilih pada setiap nilai k dirata-ratakan menggunakan rumus dalam persamaan (3). Perhitungan hasil prediksi dilakukan dengan menggunakan *Microsoft excel*.

$$Y = \frac{1}{k} \sum_{i=1}^k y_i$$

$$\begin{aligned}
 Y_{pred\ 1} &= \frac{63,69+63,75+62,93}{3} \\
 &= \frac{190,37}{3} \\
 &= 63,46
 \end{aligned}$$

Begitu seterusnya, sampai Y prediksi ke 33. Adapun hasil perhitungan prediksi menggunakan metode *k-nearest neighbor* dapat dilihat pada tabel 4 berikut.

Tabel 4. Hasil Prediksi K-Nearest Neighbor Tunggal

No.	$k = 3$	$k = 4$	$k = 5$	$k = 10$	$k = 13$	$k = 15$	$k = 18$	$k = 24$
1	63,46	63,39	64,75	67,52	68,56	69,12	69,36	70,35
2	72,01	72,2	72,27	72,23	72,35	72,57	72,91	72,72
3	74,22	73,86	74,03	73,92	73,99	73,54	73,15	72,81
4	72,33	72,48	72,03	72,59	72,64	72,74	72,59	72,51
5	76,09	76,12	75,23	75,34	74,57	74,35	74,03	73,17
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
30	79,22	78,43	78,13	76,17	75,8	75,39	74,72	73,69
31	77,57	77,17	76,56	75,1	74,62	74,38	74,03	73,14
32	77,57	77,17	76,56	74,82	74,62	74,11	73,68	73,58
33	70,74	71,2	72,35	72,71	72,18	72,46	71,73	71,55

Menurut Tabel 4. didapat bahwa hasil prediksi $k = 4$ memiliki nilai dengan akurasi yang paling tinggi. Setelah melakukan prediksi menggunakan metode *k-nearest neighbor*

Tunggal selanjutnya dilakukan pengoptimalan kinerja K-Nearest Neighbor dengan suatu teknik yaitu teknik *ensemble*. Tujuan dari teknik *ensemble* yaitu untuk menggabungkan hasil prediksi dari K-NN tunggal menjadi suatu prediksi akhir (Satriya et al., 2018). *Weighted mean* (rata-rata terboboti) merupakan teknik *ensemble* yang diterapkan pada penelitian ini. Bobot untuk teknik ini yaitu korelasi antara data sebenarnya dengan data hasil prediksi K-NN tunggal, yang dihitung dengan menerapkan rumus pada persamaan (5) dan dibantu menggunakan *Microsoft excel*. Adapun hasil perhitungannya dapat dilihat pada tabel 5.

Tabel 5. Korelasi antara data aktual dengan hasil prediksi menggunakan metode K-Nearest Neighbor

r_h	Korelasi
$r_{k=3}$	0,975527
$r_{k=4}$	0,971242
$r_{k=5}$	0,9732998
$r_{k=10}$	0,964590
$r_{k=13}$	0,959216
$r_{k=15}$	0,961953
$r_{k=18}$	0,95621
$r_{k=24}$	0,940805
Jumlah	7,702843

Setelah dilakukan perhitungan nilai korelasi dari data aktual terhadap data hasil prediksi K-NN tunggal maka selanjutnya dilakukan perhitungan pembobot *ensemble* dengan menggunakan rumus pada persamaan (6) dan dibantu menggunakan *Microsoft excel*. Adapun hasil perhitungannya dapat dilihat pada tabel 6.

Tabel 6. Pembobot *Ensemble*

w_i	Hasil Pembobot <i>Ensemble</i>
w_1	0,126645
w_2	0,1260888
w_3	0,126356
w_4	0,125225
w_5	0,1245276
w_6	0,1248828
w_7	0,1241372
w_8	0,1221373
Jumlah	1

Setelah dilakukan perhitungan terhadap pembobot *ensemble* maka selanjutnya melakukan prediksi menggunakan metode *Ensemble* K-NN menggunakan rumus dari persamaan (7) dan dibantu menggunakan *Microsoft excel*. Adapun hasil perhitungannya dapat dilihat pada tabel 7.

Tabel 7. Hasil Prediksi *Ensemble* K-NN

No	Kabupaten/Kota	<i>Ensemble</i> K-NN
1	Nias	67,04
2	Mandailing Natal	72,41
3	Tapanuli Selatan	73,69
4	Tapanuli Tengah	72,49

5	Tapanuli Utara	74,87
6	Toba	75,03
7	Labuhanbatu	73,9
8	Asahan	73
9	Simalungun	73,83
10	Dairi	74,1
11	Karo	74,59
12	Deli Serdang	74,39
13	Langkat	73,18
14	Nias Selatan	67,14
15	Humbang Hasundutan	72,63
16	Pakpak Bharat	72,33
17	Samosir	72,66
18	Serdang Bedagai	73,01
19	Batu Bara	72,73
20	Padang Lawas Utara	73,01
21	Padang Lawas	72,07
22	Labuhanbatu Selatan	73,39
23	Labuhanbatu Utara	74,05
24	Nias Utara	67,27
25	Nias Barat	67,07
26	Sibolga	74,83
27	Tanjungbalai	74,1
28	Pematangsiantar	75,56
29	Tebing Tinggi	74,73
30	Medan	76,46
31	Binjai	75,34
32	Padang Sidempuan	75,28
33	Gunung Sitoli	71,86
	Rata-rata	72,97

Berdasarkan Hasil Prediksi Indeks Pembangunan Manusia di Provinsi Sumatera Utara menggunakan metode *Ensemble* K-NN ini dinilai cukup baik jika dibandingkan hanya menggunakan metode K-NN Tunggal saja. Hasil prediksi menggunakan metode *Ensemble* K-Nearest Neighbor ini memiliki rata-rata sebesar 72,97. Hal tersebut juga dapat dibuktikan dengan menghitung evaluasi hasil prediksi *Ensemble* K-NN menggunakan rumus dari persamaan (4) dan dibantu menggunakan *microsoft excel*. Adapun hasil perhitungannya dapat dilihat pada tabel 8.

Tabel 8. Perbandingan evaluasi hasil prediksi *K- Nearest Neighbor* Tunggal dan *Ensemble K- Nearest Neighbor*

KNN	MAPE %	Akurasi	Rata-Rata	
$k = 3$	1,54 %	Tinggi	2,38 %	
$k = 4$	1,29 %	Tinggi		
$k = 5$	1,35 %	Tinggi		
$k = 10$	2,23 %	Tinggi		
$k = 13$	2,69 %	Tinggi		
$k = 15$	2,89 %	Tinggi		
$k = 18$	3,21 %	Tinggi		
$k = 24$	3,83 %	Tinggi		
<i>Ensemble K-Nearest</i>	2,1 %	Tinggi		2,1 %

Neighbor

Semakin kecil perolehan nilai MAPE, maka akan semakin tinggi tingkat akurasi hasil prediksi. Tabel 8. di atas menunjukkan bahwa teknik *ensemble* memperoleh hasil nilai MAPE yang lebih kecil yaitu sebesar 2,1%. Hal ini lebih baik dibandingkan hanya menerapkan metode K-NN Tunggal, yang memiliki nilai MAPE sebesar 2,38%.

D. SIMPULAN DAN SARAN

Hasil penelitian ini menunjukkan bahwa prediksi nilai Indeks Pembangunan Manusia di Provinsi Sumatera Utara dengan menggunakan metode *ensemble k-nearest neighbor* memiliki rata-rata sebesar 72,97. Berdasarkan metode *ensemble k-nearest neighbor* yang digunakan pada penelitian ini diperoleh hasil prediksi nilai MAPE Indeks Pembangunan Manusia di Provinsi Sumatera Utara termasuk kedalam kategori akurasi yang tinggi yaitu sebesar 2,1 %. Hasil Prediksi menggunakan metode Ensemble K-NN dinilai cukup baik jika dibandingkan hanya menggunakan metode K-NN saja yang memiliki nilai MAPE nya sebesar 2,38 %.

Adapun saran dalam penelitian ini untuk penelitian berikutnya yaitu untuk dapat mengoptimalkan metode *ensemble K-Nearest Neighbor* dengan menambahkan variabel lainnya yang dapat mempengaruhi Indeks Pembangunan Manusia agar mendapatkan hasil yang jauh lebih baik, serta disarankan menambahkan metode lain dalam memprediksi untuk dapat membandingkan metode mana yang paling optimal ataupun akurat ketika digunakan.

UCAPAN TERIMA KASIH

Terima kasih kepada Universitas Islam Negeri Sumatera Utara dan Badan Pusat Statistik Provinsi Sumatera Utara yang telah membantu dalam penelitian ini.

REFERENSI

- Alfin, M. K., Murtopo, A. A., & Fadilah, Nurul. (2022). Penerapan Metode Clustering Untuk Prediksi Produksi Bawang Merah (*Ensemble K-Nearest Neighbors*). *IJIR*, 3(2),30-37. <https://journal.peradaban.ac.id/index.php/ijir/article/view/1210/821>
- Arhami, M., & M. Nasir. (2020). *Data Mining Algoritma dan Implementasi*. Yogyakarta: Penerbit Andi.
- BPS. (2017). *Indeks Pembangunan Manusia Kabupaten Sidoarjo*. Sidoarjo: BPS Kabupaten Sidoarjo.
- BPS. (2022). *Provinsi Sumatera Utara dalam Angka 2022*. Sumatera Utara: BPS Provinsi Sumatera Utara
- BPS. (2023). *Provinsi Sumatera Utara dalam Angka 2023*. Sumatera Utara: BPS Provinsi Sumatera Utara.
- Cohen, Barry H. (2013). *Explaining Psychological Statistic*. United States of America: Wiley.
- Dewi, Nur Laely. (2019). *Identifikasi Faktor-Faktor Signifikan Yang Memengaruhi Tingkat Indeks Pembangunan Manusia Di Provinsi Sumatera Utara Tahun 2017 Menggunakan Regresi Nonparametrik Spline*. Makasar: UIN Alauddin Makasar
- Fajaryanto, Adit Dwi, & Alamsyah, Andry. (2016). *Prediksi Indeks Pembangunan Manusia (IPM) Berdasarkan Jumlah Pengguna Internet, Produk Domestik Regional Bruto (PDRB), Jumlah Pekerja, Jumlah Penduduk Miskin Dan Populasi Penduduk Pada Tahun 2010*. E-

- Proceeding Of Management*, 3(3), 2938-2946. DOI: file:///C:/Users/Hyper/Downloads/16.04.2043_jurnal_eproc.pdf
- Hidayati, Yussy Puspita. (2019). *Prediksi IPM dengan menggunakan Metode Ensemble KNN*. Surabaya: UIN Sunan Ampel Surabaya
- Jusman, M., Nur'eni, & Handayani, L. (2022). Metode *Ensemble K-Nearest Neighbor* untuk Prediksi Indeks Harga Saham Gabungan (IHSG) di Indonesia. *Jurnal Matematika, Statistika & Komputasi*, 18(3). 423-433. DOI: <https://doi.org/10.20956/j.v18i3.19641>
- Kang, S. (2021). K-Nearest Neighbor Learning with Graph Neural Networks. *mathematics*, 9(8), 830. DOI: <https://doi.org/10.3390/math9080830>
- Mendenhall, W., Reinmuth J.E., & Beaver R.J. (1993). *Statistics for Management and Economics*. California: South-Western College Pub.
- Putra, Randi Farmana, R.S.Y. Zebua, Budiman, P. W. Rahayu & P. Choirina. (2023). *Data Mining Algoritma dan Penerapannya*. Jambi: PT Sonpedia Publishing Indonesia.
- Santosa, Purbayu Budi, & Muliawan Hamdani. (2007). *Statistika Deskriptif dalam Bidang Ekonomi dan Niaga*. Semarang: PenerbitErlangga.
- Satriya, Rezza Hary, Santoso, Edy & Sutrisno. (2018). Implementasi Metode Ensemble K-Nearest Neighbor Untuk Prediksi Nilai Tukar Rupiah Terhadap Dollar Amerika. *Jurnal Pengembangan Teknologi Informasi dan IlmuKomputer*, 2(4), 1718-1725. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1361>
- Sinta, D., H. Wijayanto, & B. Sartono. (2014). Ensemble K-Nearest Neighbors Method to predict rice in Indonesia. *Appl. Math. Sci.*, 8(160), 7993-8005. DOI: <http://dx.doi.org/10.12988/ams.2014.49721>
- Wibowo, Guntur W. N., A. A. Permana, Wahyuddin S., L. W. Santoso, & Abdurasyid. (2013). *Machine Learning*. Padang: PT. Global Eksekutif Teknologi.