

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Indeks Pembangunan Manusia**

Kekayaan dan kesejahteraan suatu bangsa dapat dilihat dari manusianya sendiri. Menurut Badan Pusat Statistik (2017) pembangunan manusia adalah proses yang bertujuan untuk memperluas pilihan hidup, termasuk meningkatkan harapan hidup, menjaga kesehatan, dan memperoleh pengetahuan serta dapat hidup dengan layak. Pembangunan manusia menetapkan manusia untuk bebas memilih agar dapat hidup dengan bebas dan bermartabat.

Menurut Badan Pusat Statistik (2017) pembangunan manusia memiliki tujuan utama yaitu mewujudkan kawasan yang mengharuskan masyarakatnya untuk bisa merasakan usia panjang dan hidup yang sehat, dan dapat memperoleh ilmu pengetahuan serta dapat hidup dengan layak. Menurut Badan Pusat Statistik (2017) Indeks Pembangunan Manusia merupakan komponen penting untuk menggambarkan atau mengukur tingkat keberhasilan suatu daerah dalam usaha membangun kualitas penduduknya yaitu dari segi kesehatan, pengetahuan dan lainnya. Di tahun 1990, *United Nation Development Programme (UNDP)* meluncurkan alat untuk menilai pembangunan manusia yang dikenal dengan nama Indeks Pembangunan Manusia. Kualitas hidup dan kinerja pembangunan di suatu daerah dapat diukur dari Indeks Pembangunan Manusiannya.

Menurut Badan Pusat Statistik (2017) dalam pembentukan Indeks Pembangunan Manusia terdapat tiga dimensi yang menjadi tolak ukur diantaranya:

- a. Umur Harapan Hidup saat lahir (UHH) adalah pengrefleksian dari umur panjang dan hidup sehat
- b. Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS) adalah pengrefleksian dari dimensi pengetahuan
- c. Pengeluaran per kapita yang disesuaikan adalah pengrefleksian dari Standar hidup yang layak

Menurut Badan Pusat Statistik (2017) perhitungan Indeks Pembangunan Manusia (IPM) yaitu sebagai berikut:

$$IPM = \sqrt[3]{I_{UHH} \times I_{pendidikan} \times I_{pengeluaran}} \times 100 \quad (2.1)$$

Dimana:

IPM = Indeks dari Pembangunan Manusia

$I_{UHH}$  = Indeks dari Umur Harapan Hidup

$I_{pendidikan}$  = Indeks dari Pendidikan

$I_{pengeluaran}$  = Indeks dari Pengeluaran

## 2.2 Wahdatul Ulum

Indeks Pembangunan Manusia ialah suatu alat ukur yang menggambarkan tingkat keberhasilan suatu daerah dalam usaha membangun kualitas penduduknya yaitu dari segi kesehatan, pendidikan dan sebagainya. Di 10 provinsi yang berada di Pulau Sumatera, Indeks Pembangunan Manusia di Provinsi Sumatera Utara masih jauh dibawah Provinsi Kepulauan Riau. Untuk mempermudah pemerintah Provinsi Sumatera Utara dalam memprediksi Indeks Pembangunan Manusia-nya, dibutuhkan metode yang tepat yaitu metode Ensemble K-Nearest Neighbor. Upaya Pemerintah dalam menanggulangi keberagaman Indeks Pembangunan Manusia harus dapat dilakukan secara adil agar setiap masyarakat dapat hidup dengan sejahtera. Seperti yang tertuang pada Firman Allah SWT dalam dalam al-Qur'an Surah An-Nisa ayat 58 :

إِنَّ اللَّهَ يَأْمُرُكُمْ أَنْ تُؤَدُّوا الْأَمَانَاتِ إِلَىٰ أَهْلِهَا وَإِذَا حَكَمْتُمْ بَيْنَ النَّاسِ أَنْ تَحْكُمُوا بِالْعَدْلِ ۗ إِنَّ اللَّهَ نِعِمَّا يَعِظُكُمْ بِهِ ۗ إِنَّ اللَّهَ كَانَ سَمِيعًا بَصِيرًا ﴿٥٨﴾

Artinya: “Sungguh, Allah menyuruhmu menyampaikan amanat kepada yang berhak menerimanya, dan apabila kamu menetapkan hukum di antara manusia hendaknya

*kamu menetakannya dengan adil. Sungguh, Allah sebaik-baik yang memberi pengajaran kepadamu. Sungguh, Allah Maha Mendengar, Maha Melihat.”*

Menurut tafsir dari ayat di atas bahwasannya amanat dari seseorang terhadap sesama manusia harus dilakukan secara adil seperti peran pemerintah daerah terhadap masyarakat/rakyatnya dalam menanggulangi keberagaman Indeks Pembangunan Manusia di Provinsi Sumatera Utara harus dilakukan dengan adil dengan tidak membedakan kelompok yang satu dengan yang lainnya agar meningkatnya kualitas manusia sehingga dapat hidup lebih sejahtera (Hidayati, 2019).

Relevansi ayat Al-Qur'an yang berkaitan dengan pembangunan manusia selanjutnya dijelaskan dalam Al-Qur'an Surat Ar-Ra'd Ayat 11

إِنَّ اللَّهَ لَا يُغَيِّرُ مَا بِقَوْمٍ حَتَّىٰ يُغَيِّرُوا مَا بِأَنفُسِهِمْ وَإِذَا أَرَادَ اللَّهُ بِقَوْمٍ سُوءًا فَلَا مَرَدَّ لَهُ وَمَا لَهُمْ مِّنْ دُونِهِ مِنِّ وَّالٍ ﴿١١﴾

Artinya: “...Sesungguhnya Allah tidak akan mengubah keadaan suatu kaum sebelum mereka mengubah keadaan diri mereka sendiri. Dan apabila Allah menghendaki keburukan terhadap suatu kaum, maka tak ada yang dapat menolaknya dan tidak ada pelindung bagi mereka selain Dia.”

Menurut tafsir dari ayat di atas menjelaskan bahwa Allah tidak akan mengubah keadaan suatu manusia sebelum manusianya sendiri yang dapat mengubah keadaan diri mereka. Oleh karena itu, perubahan harus dimulai masyarakat itu sendiri dengan kemampuan berfikir manusia untuk dapat hidup yang jauh lebih baik dengan meningkatkan taraf hidup manusia sehingga dapat hidup lebih sejahtera, sehingga pembangunan manusia di Indonesia kedepannya dapat jauh lebih baik lagi (Hidayati, 2019).

### 2.3 Dimensi Kesehatan

Menurut Badan Pusat Statistik (2017) dimensi kesehatan dipresentasikan oleh Umur Harapan Hidup saat lahir (UHH). Perkiraan banyaknya tahun yang dapat

dijalani seseorang sejak seseorang itu lahir disebut dengan Umur Harapan Hidup. Berikut adalah perhitungan indeks kesehatan:

$$I_{UHH} = \frac{UHH - UHH_{min}}{UHH_{maks} - UHH_{min}} \quad (2.2)$$

Dimana:

$I_{UHH}$  = Indeks dari Umur Harapan Hidup

$UHH$  = Nilai dari Umur Harapan Hidup seseorang

$UHH_{min}$  = Nilai minimal dari Umur Harapan Hidup seseorang

$UHH_{maks}$  = Nilai maksimal dari Umur Harapan Hidup seseorang

Untuk melihat perkembangan kesehatan di suatu wilayah dapat ditinjau dari Umur Harapan Hidupnya. Di negara berkembang dan negara maju Umur Harapan Hidup bervariasi karena dipengaruhi oleh negara itu sendiri, di negara berkembang tingkat kematian jauh lebih tinggi dibanding negara maju. Indikator Umur Harapan Hidup sangat berguna untuk mengidentifikasi kesehatan penduduk karena kesehatan memiliki dampak besar pada keberlangsungan hidup seseorang.

#### 2.4 Dimensi Pendidikan

Menurut Badan Pusat Statistik (2019) Harapan lama sekolah dan Rata-rata lama sekolah adalah penggambaran dari dimensi pendidikan. Kedua indikator ini berfungsi untuk mengetahui gambaran kemajuan pendidikan disuatu daerah. Adapun rumus indeks pendidikan menurut UNDP adalah sebagai berikut:

$$I_{HLS} = \frac{HLS - HLS_{min}}{HLS_{maks} - HLS_{min}} \quad (2.3)$$

Dimana:

$I_{HLS}$  = Indeks harapan dari lamanya seseorang bersekolah

$HLS$  = Harapan dari lamanya seseorang bersekolah

$HLS_{min}$  = Nilai minimal dari harapan lamanya seseorang bersekolah

$HLS_{maks}$  = Nilai maksimal dari harapan lamanya seseorang bersekolah

$$I_{RLS} = \frac{RLS - RLS_{min}}{RLS_{maks} - RLS_{min}} \quad (2.4)$$

Dimana:

$I_{RLS}$  = Indeks Rata-rata dari lamanya seseorang bersekolah

$RLS$  = Rata-rata dari lamanya seseorang bersekolah

$RLS_{min}$  = Nilai minimal dari rata-rata lamanya seseorang bersekolah

$RLS_{maks}$  = Nilai maksimal dari harapan lamanya seseorang bersekolah

$$I_{pendidikan} = \frac{I_{HLS} + I_{RLS}}{2} \quad (2.5)$$

Dimana:

$I_{pendidikan}$  = Indeks Pendidikan

$HLS$  = Harapan dari lamanya seseorang bersekolah

$RLS$  = Rata-rata dari lamanya seseorang bersekolah

#### 2.4.1 Rata-rata Lama Sekolah (RLS)

Menurut BPS (2019) rata-rata lama sekolah menunjukkan rata-rata lamanya seseorang bersekolah untuk menempuh pendidikan. Perhitungan rata-rata lama sekolah dinilai pada masyarakat yang berusia 25 tahun keatas, dilihat dari banyaknya tahun saat mereka menjalani pendidikan formal. Komponen variabel dari indeks pendidikan data yang digunakan adalah jenjang pendidikan yang pernah ditempuh, angka partisipasi sekolah, kelas yang saat ini ditempuh dan tingkat pendidikan yang diselesaikan.

Menurut Badan Pusat Statistik (2017) perhitungan rata-rata lama sekolah terdiri dari enam langkah yaitu meliputi:

- a. Memilih masyarakat yang berusia lebih dari 25 tahun.
- b. Menggolongkan tingkat pendidikan yang sudah ditempuh ataupun yang sedang ditempuh.
- c. Menggolongkan ijazah terakhir yang dimiliki setiap individu.
- d. Memvariasikan durasi pendidikan berdasarkan ijazah terakhir yang dimiliki.

- e. Menilai durasi pendidikan seseorang hingga mencapai jenjang pendidikan terakhir.
- f. Mengukur total waktu yang dihabiskan seseorang untuk bersekolah.

#### **2.4.2 Harapan Lama Sekolah (HLS)**

Menurut BPS (2017) indeks harapan lama sekolah ialah harapan lamanya seorang anak bersekolah. Masyarakat yang tergolong dalam indeks harapan lama sekolah ialah masyarakat yang berusia 7 tahun ke atas. Angka tersebut dianggap memiliki kemungkinan bahwa anak yang berusia tersebut dapat tetap melanjutkan sekolah pada usia selanjutnya. Tujuan dari perhitungan indeks harapan lama sekolah yaitu untuk mengetahui keadaan pembangunan di setiap tahap pendidikan yang dapat dilihat dari harapan lamanya dalam tahun seorang anak dapat menempuh pendidikan.

Menurut Badan Pusat Statistik (2017) Harapan lama sekolah dihitung dengan melalui tiga langkah:

- a. Menghitung jumlah masyarakat berdasarkan usia lebih dari 7 tahun.
- b. Menghitung jumlah masyarakat yang sedang bersekolah berdasarkan usia lebih dari 7 tahun.
- c. Mengitung rasio masyarakat yang sedang bersekolah berdasarkan usia lebih dari 7 tahun.

Perhitungan Harapan lama sekolah dapat dilakukan dengan menjumlahkan semua rasio masyarakat yang masih bersekolah berdasarkan usia 7 tahun ke atas.

#### **2.5 Dimensi Pengeluaran**

Kemampuan suatu penduduk dalam membeli kebutuhan konsumsi dapat berubah-ubah bisa naik maupun turun. Ketidakpastian kesanggupan penduduk untuk membeli kebutuhan konsumsi dipengaruhi beberapa hal yaitu harga barang, pendapatan penduduk, dan selera penduduk, serta kualitas dari barang tersebut.

Menurut Badan Pusat Statistik (2017) kondisi perekonomian suatu penduduk dapat diamati dari tingginya kemampuan daya beli. Kondisi perekonomian yang baik

dapat dilihat dari tingginya kemampuan daya beli penduduk dalam membeli suatu barang, begitupun sebaliknya. Kemampuan daya beli penduduk dipengaruhi oleh beberapa hal, yaitu tingkat pendapatan penduduk, acuan konsumsi penduduk, dan harga barang.

Pengeluaran per kapita yang disesuaikan adalah penggambaran dari dimensi standar hidup yang layak. Standar hidup layak digambarkan melalui kesejahteraan penduduk dari segi perekonomian. Menurut Badan Pusat Statistik (2017) rumus indeks pengeluaran adalah sebagai berikut:

$$I_{pengeluaran} = \frac{\ln(pengeluaran) - \ln(pengeluaran_{min})}{\ln(pengeluaran_{maks}) - \ln(pengeluaran_{min})} \quad (2.6)$$

Dimana,

$I_{pengeluaran}$  = Indeks Pengeluaran

$I_n$  = Indeks Komponen

$pengeluaran_{min}$  = Pengeluaran Minimal

$pengeluaran_{maks}$  = Pengeluaran Maksimal

## 2.6 Kemiskinan

Menurut Badan Pusat Statistik (2019) kemiskinan diukur oleh jumlah penduduk miskin. Kemiskinan memiliki peranan yang penting dalam menentukan keberhasilan pembangunan di suatu daerah, karena terdapat sejumlah elemen dasar dalam Indeks Pembangunan Manusia yang mempengaruhi kualitas hidup seseorang. Sejalan dengan hal tersebut, kemiskinan menjadi masalah yang sangat erat kaitannya dengan pembangunan. Terjadinya kemiskinan disebabkan karena ketidakmampuan penduduk dalam memenuhi seluruh aspek kehidupannya sehingga mereka tidak ikut andil dalam proses pembangunan maupun menikmati hasil dari pembangunan tersebut. Kemiskinan telah membuat ratusan bahkan jutaan anak tidak dapat menempuh pendidikan, membuat masyarakat tidak mampu membiayai fasilitas kesehatan, kurangnya akses kepelayanan publik, kurangnya tabungan dan yang lebih

parah yaitu keterbatasan mereka dalam memenuhi seluruh aspek kehidupan mereka baik dari segi memperoleh makanan, pakaian maupun lainnya.

Kemiskinan dipahami sebagai ketidakmampuan ekonomi masyarakat dalam memenuhi seluruh kebutuhan hidupnya yang dilihat dari sisi pengeluarannya. Yang dikategorikan sebagai penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran perkapita per bulan mereka di bawah garis kemiskinan.

Menurut Badan Pusat Statistik (2023) perhitungan Garis kemiskinan yaitu sebagai berikut :

$$GK = GKM + GKNM \quad (2.7)$$

Dimana:

GK = Nilai dari garis kemiskinan

GKM = Nilai dari garis kemiskinan yang dinilai dari segi makanan

GKNM = Nilai dari garis kemiskinan yang dinilai dari segi non-makanan

## 2.7 Praproses Data

Teknik Praproses data adalah langkah penting dalam penelitian, yang mencakup berbagai prosedur, salah satunya adalah transformasi data. Transformasi data melibatkan perubahan atau penyesuaian data untuk membuat proses mining lebih efisien dan lebih mudah dipahami. Salah satu metode transformasi data adalah normalisasi. Normalisasi yaitu proses penyesuaian nilai atribut agar sesuai dengan skala tertentu (Putra *et al.* 2023).

Normalisasi Min-Max adalah normalisasi yang digunakan pada penelitian ini. Dengan menerapkan normalisasi Min-Max dapat menghasilkan keseimbangan nilai. Adapun rumus mendapatkan nilai normalisasi Min-Max adalah sebagai berikut: (Putra *et al.* 2023).

$$X_{normalisasi} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2.8)$$

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

## 2.8 *K-Nearest Neighbor*

*K-Nearest Neighbor* ialah metode yang menentukan label (kelas) suatu objek baru dengan mengacu pada kelas yang paling umum di antara *K-Nearest Neighbor* dalam data latih. *K-Nearest Neighbor* berbasis kesamaan yang dimiliki objek didasarkan pada jarak antara objek yang akan ditentukan dengan objek yang telah ada sebelumnya. *K-Nearest Neighbor* termasuk juga dalam kategori regresi yang bisa digunakan untuk memprediksi. *K-Nearest Neighbor* digolongkan kedalam algoritma *supervised learning* yang berbasis pada jarak (Arhami & M. Nasir,2020).

Metode *K-Nearest Neighbor* disebut juga dengan metode K-NN tunggal. K-NN ialah metode yang diterapkan dalam memprediksi variabel output melalui pendekatan klasifikasi. Pada pendekatannya, data dibagi menjadi dua set yaitu data *training* dan data *testing*. Metode *K-Nearest Neighbor* menentukan nilai output dengan meninjau  $k$  tetangga terdekat dari data *testing*. Nilai *output* dari  $k$  tersebut diterapkan dalam memprediksi nilai *output* yang berasal dari data *testing* yang belum diketahui (Arhami & M. Nasir,2020).

Nilai  $k$  merupakan suatu parameter yang merujuk kepada jumlah tetangga paling dekat dengan objek yang diprediksi kelasnya sehingga dapat ditentukan tetangga yang mayoritas bagi suatu objek. Pemilihan jumlah  $k$  yang ideal harus dipertimbangkan dengan hati-hati untuk mengurangi tingkat kesalahan. Umumnya, nilai  $k$  meningkat seiring dengan jumlah dataset. Nilai  $k$  terbaik tergantung pada karakteristik dataset tersebut, sehingga seringkali perlu dilakukan eksperimen untuk menemukan nilai  $k$  yang paling optimal (Primartha, 2021).

Pada penerapan metode *K-Nearest Neighbor* ini menggunakan ukuran kemiripan dalam membandingkan antara data testing dengan data training. Jarak

*euclidean* adalah ukuran kemiripan yang digunakan metode *K-Nearest Neighbor* ini, adapun tujuannya untuk mengukur jarak antara data *training* dengan data *testing* (Wibowo *et al.* 2023).

Adapun rumus perhitungan jarak *Euclidean* yaitu sebagai berikut (Wibowo *et al.* 2023).

$$dx_{train-i, x_{test-j}} = \sqrt{\sum_{i,j=1}^n (X_{train-i} - X_{test-j})^2} \quad (2.9)$$

Dimana:

$dx_{train-i, x_{test-j}}$	= jarak <i>euclidean</i>
$x_{train-i}$	= data <i>training</i> dengan nilai ke-i
$x_{test-j}$	= data <i>testing</i> dengan nilai ke-j
$n$	= jumlah semua data
$i, j$	= 1,2,3,...,n

Metode *K-Nearest Neighbor* memiliki kelebihan maupun kekurangan. Kelebihan metode *K-Nearest Neighbor* ini seperti menggeneralisasi himpunan data *training* yang relatif kecil, efektivitasnya, dan penerapannya yang luas dalam berbagai kasus penelitian, baik untuk klasifikasi maupun prediksi sedangkan, kekurangan *K-Nearest Neighbor* yaitu memerlukan ketepatan untuk menentukan nilai dari parameter *k* atau biasa dikenal dengan istilah jumlah tetangga terdekat (Wibowo *et al.* 2023). Adapun algoritma dalam metode *K-Nearest Neighbor* yaitu diantaranya sebagai berikut (Wibowo *et al.* 2023):

1. Menghitung nilai dari jarak *euclidean* antara data *testing* (data uji) dan data *training* (data latih) dengan menerapkan persamaan

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

Kemudian, menentukan jarak terdekat sampai urutan *k*.

2. Menentukan nilai *k*.

3. Menghitung prediksi metode K-NN dengan merata-ratakan nilai dari objek yang telah diperoleh pada jumlah tetangga terdekat. Adapun perhitungan rata-ratanya yaitu sebagai berikut (Bishop, 1995):

$$Y = \frac{1}{k} \sum_{i=1}^k y_i \quad (2.10)$$

Dimana:

$Y$  = prediksi dari metode K-NN tunggal pada jumlah  $k$  ke- $i$

$k$  = banyaknya nilai tetangga terdekat yang telah ditentukan

$y_i$  = urutan dari jarak berdasarkan hasil Euclidean

4. Mengevaluasi hasil prediksi metode K-NN tunggal menggunakan nilai MAPE

Setelah melakukan prediksi menggunakan metode K-Nearest Neighbor Tunggal selanjutnya dilakukan pengoptimalan kinerja K-Nearest Neighbor dengan suatu teknik yaitu teknik *Ensemble*. Pada penerapannya teknik *ensemble* dalam metode *ensemble* K-NN ini melibatkan pembobotan sehingga, tidak diperlukan lagi pencarian nilai  $k$  yang optimal (Wibowo *et al.* 2023). Adapun dilakukan teknik *ensemble* ini bertujuan untuk melihat manakah yang jauh lebih baik dalam memprediksi Indeks Pembangunan Manusia diantara metode *K-Nearest Neighbor* Tunggal dengan *Ensemble K-Nearest Neighbor*.

## 2.9 Teknik *Ensemble*

Pada penerapan teknik *ensemble* ini dilakukan tidak hanya memilih suatu model yang terbaik dari sejumlah model yang tersedia melainkan menggabungkan hasil prediksi dari berbagai model pada bobot tertentu untuk mendapatkan hasil prediksi yang jauh lebih akurat. Teknik *ensemble* adalah metode yang tidak hanya memilih satu model terbaik dari sejumlah model yang tersedia dan menggunakan model tersebut untuk prediksi, sebaliknya mengintegrasikan hasil prediksi dari berbagai model dengan bobot tertentu. Teknik ini berguna untuk meningkatkan kemampuan prediksi dibandingkan dengan model-model standar lainnya. Pada

intinya, teknik *ensemble* adalah metode prediksi yang mengintegrasikan berbagai hasil output dari berbagai metode. Penelitian terbaru memperlihatkan bahwa kombinasi dari berbagai model dapat meningkatkan akurasi serta kinerja dengan melalui metode *ensemble* (Cohen, 2013).

Salah satu teknik *ensemble* yang diterapkan dalam memprediksi adalah *Weighted Mean* (rata-rata terboboti). Adapun perhitungannya sebagai berikut (Cohen, 2013).

$$\mu = \frac{\sum_{i=1}^s w_i y_i}{\sum_{i=1}^s w_i} \quad (2.11)$$

Dimana:

$s$  = banyaknya data

$y_i$  = Hasil nilai dari prediksi pada model ke- $i$

$w_i$  = Pembobot.

Pada prinsipnya teknik *ensemble* melakukan pembobotan, adapun pembobotnya berguna untuk mengukur korelasi diantara data aktual dan data dari hasil prediksi menggunakan *K-Nearest Neighbor* tunggal, adapun perhitungannya sebagai berikut (Cohen, 2013).

$$W_i = \frac{r_i}{\sum_{j=1}^s r_i} \quad (2.12)$$

Dimana:

$W_i$  = Pembobot pada ensemble ke-  $i$

$r_i$  = Nilai korelasi yang diperoleh antara data aktual terhadap data hasil prediksi *K-Nearest Neighbor*

$s$  = Jumlah data yang digunakan

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

Dalam mengukur hubungan antara variabel X dan Y yang akan diteliti dapat diketahui dengan mencari nilai koefisien korelasi product moment. Adapun korelasi yang ingin dianalisis yaitu hubungan antara data aktual dengan data dari hasil

prediksi menggunakan metode K-NN tunggal. Adapun perhitungannya adalah sebagai berikut (Santosa & Muliawan, 2007):

$$r_i = \frac{n \sum \bar{y}_i \hat{y}_t - (\sum \bar{y}_i)(\sum \hat{y}_t)}{\sqrt{(n \sum \bar{y}_i^2 - (\sum \bar{y}_i)^2)(n \sum \hat{y}_t^2 - (\sum \hat{y}_t)^2)}} \quad (2.13)$$

Dimana:

$r_i$  = Koefisien korelasi product moment

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

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

Adapun analisis data dalam teknik *ensemble* adalah sebagai berikut (Cohen, 2013).

1. Menggunakan metode K-NN tunggal dengan nilai  $k$ , terdiri dari 3, 4, 5, 10, 13, 15, 18, dan 24. Dimana Nilai-nilai  $k$  yang diterapkan bertujuan untuk mengevaluasi dampak nilai tersebut terhadap hasil prediksi yang diperoleh.
2. Menghitung korelasi, dimana korelasi mencerminkan besarnya hubungan diantara data aktual dengan prediksi K-NN tunggal. Adapun perhitungannya yaitu :

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

Dengan,  $r_i$  = koefisien korelasi product moment,  $\bar{y}_i$  = nilai prediksi dari K-NN tunggal pada jumlah  $k$  ke  $-i$ , dan  $\hat{y}_t$  = nilai sebenarnya ke  $-t$ .

3. Menghitung pembobot *ensemble* untuk mencari nilai ( $W_i$ ), menggunakan rumus

$$W_i = \frac{r_i}{\sum_{j=1}^s r_i}$$

Dengan,

$W_i$  = nilai dari Pembobot pada ensemble ke-  $i$

$r_i$  = Nilai korelasi yang diperoleh antara data aktual terhadap data hasil prediksi K-Nearest Neighbor

$s$  = Jumlah semua data yang digunakan

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

4. Menggabungkan hasil prediksi yang telah diperoleh tersebut dengan menerapkan rumus

$$\mu = \frac{\sum_{i=1}^s w_i y_i}{\sum_{i=1}^s w_i}$$

Dengan,

$s$  = banyaknya data yang diteliti

$y_i$  = Hasil nilai dari prediksi pada model ke- $i$

$w_i$  = Pembobot.

5. Melakukan evaluasi terhadap hasil prediksi menggunakan nilai MAPE.

### 2.10 Mean Absolute Percentage Error (MAPE)

Perhitungan *Mean Absolute Percentage Error* dilakukan dengan membagi kesalahan yang terjadi dalam suatu periode tertentu terhadap nilai observasi yang sebenarnya pada periode tersebut. MAPE berfungsi untuk mengukur seberapa besar ketepatan dari suatu prediksi jika dibandingkan dengan nilai yang sesuai kenyataan. Dengan demikian, keakuratan suatu prediksi dapat dilihat dengan membandingkan nilai data aktual yang diukur menggunakan nilai MAPE. Jika memperoleh nilai MAPE yang kecil maka semakin akurat hasil prediksinya. Adapun perhitungan dari Evaluasi MAPE yaitu sebagai berikut (Mendenhall *et al.* 1993)

$$MAPE = \sum_{i=1}^n \frac{1}{m} \cdot \frac{|x_i - y_i|}{x_i} \times 100\% \quad (2.14)$$

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

Nilai *Mean Absolute Percentage Error* memiliki akurasi hasil prediksi yang dapat dilihat pada tabel berikut (Mendenhall *et al.* 1993)

**Tabel 2.1 Akurasi Hasil Prediksi Berdasarkan Nilai MAPE**

Nilai MAPE	Kategori Akurasi
$MAPE \leq 10\%$	Tinggi
$10\% < MAPE < 20\%$	Baik
$20\% < MAPE < 50\%$	Sedang
$MAPE > 50\%$	Rendah

### 2.11 Penelitian Yang Relevan

1. Pada penelitian yang dilaksanakan oleh Muhamad Fajri, Syafriandi, D. Vionanda, dan Zilrahmi (2023) yang berjudul “Prediksi Harga Minyak Mentah Dunia Menggunakan Metode *Ensemble K-Nearest Neighbor*”.

Dalam penelitian ini, ditemukan bahwa metode *Ensemble K-Nearest Neighbor* memiliki tingkat akurasi prediksi sebesar 11,11 %, sedangkan metode K-NN tunggal menghasilkan MAPE sebesar 12,8 %. Prediksi harga minyak mentah menggunakan metode *Ensemble K-Nearest Neighbor* untuk periode Januari hingga Desember 2023 menunjukkan fluktuasi harga, dengan tren naik dan turun. Pada Januari 2023, harga minyak berada di angka 93,95 USD per barel, tetapi meningkat menjadi sekitar 98,66 USD per barel pada Februari 2023.

2. Penelitian yang dilaksanakan oleh Moh. Jusman, Nur’eni, dan Lilies Handayani (2022) yang berjudul “Metode *Ensemble K-Nearest Neighbor* untuk Prediksi Indeks Harga Saham Gabungan (IHSG) di Indonesia.”

Pada penelitian ini dapat disimpulkan bahwa di Indonesia hasil prediksi dari nilai Indeks Harga Saham Gabungan dengan menerapkan metode *Ensemble K-Nearest Neighbor* menunjukkan rata-rata sebesar 6078,634, dan berdasarkan

nilai MAPE diperoleh nilai 7,16%. Angka ini mengindikasikan bahwa prediksi nilai IHSG memiliki akurasi yang tinggi.

3. Penelitian yang dilakukan oleh Moh. Khoiru Alfin, Aang Alim Murtopo, dan Nurul Fadilah (2022) yang berjudul “Penerapan Metode Clustering untuk Prediksi Produksi Bawang Merah ( Ensemble K-Nearest Neighbors)”.

Dari penelitian ini, dapat disimpulkan 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. Prediksi harga bawang merah dengan metode ini hampir sesuai dengan harga sebenarnya. Pada tahun 2022, prediksi harga bawang merah berkisar antara Rp. 23.400 hingga Rp. 23.700 per kilogram, yang mendekati harga sebenarnya yang berkisar antara Rp. 23.100 hingga Rp. 23.800 per kilogram.

4. Penelitian yang dilaksanakan oleh Meila Ika Pradipta (2020) dengan judul “Klasifikasi Curah Hujan Menggunakan Metode *Ensemble Subset K-Nearest Neighbor*”.

Pada penelitian ini dapat disimpulkan bahwa berdasarkan nilai k yang digunakan yaitu k=3,4,5 dan 6 diketahui bahwa k=3 dan k=6 memiliki tingkat akurasi dengan nilai 74%. Hal ini disebabkan oleh ketepatan k=3 dan k=6 dibandingkan dengan k=4 dan k=5. Semakin tinggi ketepatan klasifikasi, semakin tinggi pula tingkat akurasi antara hasil prediksi dan data aktual.

5. Penelitian yang dilakukan oleh Yussy Puspita Hidayati (2019) yang berjudul “Prediksi Indeks Pembangunan Manusia (IPM) dengan Menggunakan Metode Ensemble KNN”.

Pada penelitian ini dapat ditarik kesimpulan bahwa di Provinsi Jawa Timur nilai prediksi Indeks Pembangunan Manusia sebesar 69,83 dan hasil MAPE nya ada data training masuk kedalam kategori akurasi prediksi tinggi yaitu sebesar 0,076% begitu juga dengan data testingnya masuk kedalam kategori akurasi prediksi tinggi sebesar 0,082%.

6. Penelitian yang dilakukan oleh Rezza Hary Dwi Satriya, Edy Santoso, dan Sutrisno (2018) yang berjudul “Implementasi Metode *Ensemble K-Nearest Neighbor* untuk Prediksi Nilai Tukar Rupiah Terhadap Dollar Amerika”.

Pada penelitian ini dapat disimpulkan bahwa hasil prediksi tidak berbeda jauh dari kurs aslinya. Untuk prediksi kurs jual pada tahun 2016 berkisar Rp.13.506,96 sampai Rp. 14.022,89 sementara itu, kurs jual aktual berada dalam kisaran Rp.13.063 sampai Rp.13.915. Prediksi untuk kurs beli berkisar antara Rp.13.372,80 sampai Rp.13.882,62 sedangkan, kurs beli sebenarnya berada dalam rentang Rp.12.933 sampai Rp. 13.777.

7. Penelitian yang dilaksanakan oleh Dewi Sinta (2015) yang berjudul “Metode *Ensemble K-Nearest Neighbor* untuk Prediksi Harga Beras di Indonesia”.

Pada penelitian ini memperoleh nilai MAPE berkisar 1.46 dan hasil prediksi yang diperoleh harga beras yang hampir sama dengan harga aslinya. Tercatat pada tahun 2012 prediksi harga beras berkisar Rp. 8.400 sampai dengan Rp. 8.700 per kilogram. Pada prediksi harga beras juga memiliki pola *trend* yang hampir sama dengan harga aslinya. Dengan hasil dari nilai MAPE, MAE, RMSEP dapat diketahui bahwa metode *ensemble kNN* memiliki performa yang lebih baik dibanding hanya menggunakan metode *kNN* tunggal saja.