

## Ensemble K-Nearest Neighbors Method to Predict Composite Stock Price Index (CSPI) in Indonesia

### Metode *Ensemble K-Nearest Neighbor* untuk Prediksi Indeks Harga Saham Gabungan (IHSG) di Indonesia

Moh. Jusman<sup>\*1</sup>, Nur'eni<sup>\*2</sup>, Lilies Handayani<sup>\*3</sup>

#### Abstract

The Composite Stock Price Index (CSPI) is a guide for investors to see the movement of stock prices as a whole from time to time. These movements always change from time to time, so it is necessary to use analytical methods to make predictions. The method that can be used to examine this is the K-Nearest Neighbor method. The combination of the results of several K-NN predictions is an effective way to get one final prediction result, namely the method ensemble K-NN. The response variable used in this study is the Composite Stock Price Index (CSPI), while the predictor variables are the gold price, the rupiah exchange rate against the dollar, and the Dow Jones Industrial Average (DJIA) index. The data used are 52 periods. The data used for training are 39 periods and the data used for testing is 13 periods. The prediction results from the ensemble have better results than the K-NN. The prediction results from the ensemble have better results than the single K-NN. The prediction results from the method are ensemble K-NN average of 6078, 634 with a MAPE value of 7,16% including high accuracy.

**Keywords:** Composite Stock Price Indeks, Ensemble K-NN, Prediction.

#### Abstrak

Indeks Harga Saham Gabungan (IHSG) pedoman bagi investor untuk melihat pergerakan harga saham secara keseluruhan dari waktu ke waktu. Pergerakan tersebut selalu berubah dari waktu ke waktu, sehingga diperlukan metode analisis untuk melakukan prediksi. Metode yang dapat digunakan untuk mengkaji hal tersebut adalah metode *K-Nearest Neighbor*. Kombinasi dari hasil beberapa prediksi K-NN menjadi salah satu cara efektif untuk mendapatkan satu buah hasil prediksi akhir, yaitu metode *ensemble* K-NN. Variabel respon yang digunakan pada penelitian ini yaitu Indeks Harga Saham Gabungan (IHSG), sedangkan variabel prediktor yaitu harga emas, nilai tukar rupiah terhadap dollar, dan indeks *Dow Jones Industrial Average* (DJIA). Data yang digunakan sebanyak 52 periode ( $t$ ). Data yang digunakan untuk *training* sebanyak 39 periode dan data yang digunakan untuk *testing* sebanyak 13 periode. Hasil prediksi dari metode *ensemble* K-NN memiliki hasil yang lebih baik dibandingkan K-NN tunggal. Hasil prediksi dari metode *ensemble* K-NN rata-rata sebesar 6078, 634 dengan nilai MAPE sebesar 7,16% masuk pada akurasi yang tinggi.

\* Program Studi S1 Statistika, Fakultas MIPA, Universitas Tadulako

Email address: [jusman.tamsar4@gmail.com](mailto:jusman.tamsar4@gmail.com)<sup>1</sup>, [eniocy@yahoo.com](mailto:eniocy@yahoo.com)<sup>2</sup>, [lilies.stath@gmail.com](mailto:lilies.stath@gmail.com)<sup>3</sup>



**Kata kunci:** *Ensemble* K-NN, Indeks Harga Saham Gabungan, Prediksi.

## 1. PENDAHULUAN

Pedoman bagi investor untuk melihat gambaran harga saham secara keseluruhan adalah Indeks Harga Saham Gabungan (IHSG). Rangkaian informasi historis mengenai pergerakan harga saham gabungan hingga suatu tanggal tertentu dapat dipantau melalui Indeks Harga Saham Gabungan (IHSG), namun pergerakannya selalu fluktuatif atau berubah dari waktu ke waktu [12]. Penelitian sebelumnya yang dilakukan oleh Khoiri & Arghawaty (2020) tentang faktor-faktor yang mempengaruhi nilai IHSG didapatkan hasil bahwa harga emas, nilai tukar rupiah terhadap dollar dan DJIA berpengaruh secara signifikan terhadap fluktuasi nilai IHSG [8]. Oleh karena itu, untuk menganalisis kemungkinan kenaikan atau penurunan harga saham di waktu akan datang (prediksi) diperlukan suatu metode analisis. Salah satu metode yang dapat digunakan untuk prediksi adalah *K-Nearest Neighbor* (K-NN).

Metode *K-Nearest Neighbor* (K-NN) merupakan suatu metode yang dapat diterapkan pada data deret waktu yang tidak memenuhi asumsi. Salah satu penyebabnya adalah adanya hubungan yang tidak linear antar peubahnya. Kelebihan lainnya adalah metode ini juga dapat melakukan prediksi pada data deret waktu dengan ukuran data yang besar dan peubah penjelas yang banyak. Selain itu, metode ini juga memiliki kekurangan dalam melakukan prediksi yaitu tidak dapat secara pasti pada perhitungan jarak  $k$  yang mana memiliki hasil yang mendekati nilai sebenarnya. Oleh karena itu, untuk mendapatkan hasil yang konsisten dan baik diperlukan optimasi dengan cara menambahkan suatu teknik. Teknik *ensemble* merupakan teknik yang akan digunakan untuk optimasi kinerja metode K-NN dalam melakukan prediksi [15].

Teknik *ensemble* menjadi salah satu teknik penting dalam peningkatan kemampuan prediksi dari berbagai hasil prediksi. Teknik ini tidak memilih satu prediksi terbaik dari sekian banyak kandidat prediksi dan kemudian melakukan pendugaan dari prediksi tersebut, namun menggabungkan dari berbagai prediksi yang ada menjadi suatu satu buah prediksi akhir. Artinya, metode *ensemble* dalam metode K-NN tidak perlu dilakukan pencarian nilai  $k$  (jumlah tetangga terdekat) yang optimal [13].

Penelitian sebelumnya yang dilakukan oleh Khaira, dkk (2019) tentang prediksi Indeks Harga Saham Gabungan (IHSG) menggunakan *algoritma Autoregressive Integrated Moving Average* (ARIMA) didapatkan hasil bahwa akan terjadi kenaikan harga saham yang sangat drastis di bulan Maret 2020 dan penurunan terjadi di bulan April hingga Mei. Dalam penelitian tersebut model prediksi yang dihasilkan cukup baik, hal ini diutarakan karena hasil pengujian skor MSE menggunakan *training* dataset dan *testing* dataset menunjukkan angka 0,004 dan 0,003 [7].

Berdasarkan uraian latar belakang yang telah diuraikan, maka pada penelitian ini akan dilakukan prediksi nilai Indeks Harga Saham Gabungan (IHSG) menggunakan metode *ensemble K-Nearest Neighbor*. Tujuan ini adalah untuk mengetahui akurasi metode *ensemble K-Nearest Neighbor* dalam melakukan prediksi nilai Indeks Harga Saham Gabungan (IHSG) di Indonesia.

### 1.1. Data Mining

Data Mining adalah sebuah proses yang menggunakan teknik statistika, matematika, kecerdasan buatan, pembelajaran mesin dalam mengidentifikasi dan mengumpulkan informasi dari data dengan jumlah yang sangat besar. Terdapat banyak metode yang dapat digunakan dalam data mining. Proses tersebut membutuhkan beberapa langkah untuk mendapatkan sebuah data yang diinginkan. *K-Nearest Neighbor* merupakan salah metode dalam data mining [2].

## 1.2. Prediksi

Prediksi dapat dilakukan dengan menggunakan berbagai informasi yang relevan pada waktu-waktu sebelumnya (historis) melalui suatu metode ilmiah dapat digunakan untuk menduga atau memperkirakan sesuatu yang akan terjadi di waktu mendatang. Informasi tentang apa yang akan terjadi di waktu mendatang diprediksi untuk meminimalisir tingkat *error* dan kesalahannya. Metode prediksi dapat dilakukan secara kualitatif dan kuantitatif [11].

Prediksi yang berasal dari data kuantitatif yang terjadi di masa lalu disebut prediksi kuantitatif. Pada penelitian ini, metode prediksi yang digunakan adalah prediksi kuantitatif karena data berasal dari data kuantitatif yang terjadi dimasa lalu [9]. Pada penelitian ini menggunakan prediksi kuantitatif karena data berasal dari data kuantitatif yang terjadi di masa lalu.

### 1.2. K-Nearest Neighbor (K-NN)

Metode yang menggunakan algoritma *supervised* salah satunya adalah K-Nearest Neighbor (K-NN). Dimana, data *testing* yang baru diklasifikasikan berdasarkan mayoritas kelas pada K-NN. Tujuan dari algoritma ini untuk mengklasifikasikan objek baru berdasarkan parameter dan data *training* [14].

Metode K-Nearest Neighbor (K-NN) biasa juga disebut dengan K-NN tunggal karena hanya menggunakan jumlah satu  $k$  tertentu. Cara kerja dari K-NN adalah mencari  $k$  objek dari data *training* yang memiliki jarak paling dekat dengan data *testing*. Jarak yang digunakan adalah jarak *euclidean*. Jarak *euclidean* adalah jarak yang paling banyak digunakan dalam data numerik [6].

$$d(x_{train,i}, x_{test,j}) = \sqrt{\sum_{i,j}^n (x_{train,i} - x_{test,j})^2} \quad (1.1)$$

Keterangan:

$d(x_{train,i}, x_{test,j})$	: jarak <i>euclidean</i>
$x_{train,i}$	: data <i>training</i> ke- $i$
$x_{test,j}$	: data <i>testing</i> ke- $j$
$n$	: banyaknya data
$i, j$	: 1, 2, 3, ..., $n$

Nilai *output* dari  $k$  data *training* yang terpilih sebagai tetangga terdekat digunakan untuk memprediksi nilai *output* dari data *testing* yang tidak diketahui. Berikut adalah algoritma dalam metode K-Nearest Neighbor (K-NN) [4]:

1. Menentukan nilai  $k$ .
2. Menghitung jarak antara data *testing* dengan data *training* menggunakan Persamaan (1.1).
3. Menentukan jarak terdekat sampai urutan  $k$ .
4. Menghitung prediksi dengan merata-ratakan dari nilai objek pada jumlah tetangga terdekat yang dirumuskan dalam Persamaan (1.2).

$$\bar{y}_i = \frac{1}{k} \sum_{i=1}^k y_i \quad (1.2)$$

Keterangan:

$\bar{y}_i$	: prediksi K-NN Tunggal dengan jumlah $k$ ke- $i$
$k$	: jumlah tetangga terdekat
$y_i$	: urutan jarak <i>euclidean</i> ke- $i$

Algoritma K-NN memiliki kelebihan dan kekurangan. Kelebihan dari algoritma K-NN, yaitu memiliki akurasi yang tinggi, *insentive* terhadap *outlier*, dan tidak ada asumsi terhadap data. Kelemahan dari metode K-NN, yaitu perlu untuk menentukan nilai  $k$  optimal, komputasi yang mahal, dan membutuhkan memori yang banyak [16].

### 1.3. Ensemble K-Nearest Neighbor (K-NN)

Beberapa variabel respon dari metode prediksi dapat digabungkan dengan teknik *ensemble*. Teknik ini merupakan salah satu metode prediksi yang biasa digunakan, terutama pada prediksi iklim. Studi terakhir telah menunjukkan bahwa gabungan beberapa prediksi dapat meningkatkan ketahanan dan kehandalan, yaitu *ensemble* [10].

*Weighted mean* (rata-rata terboboti) merupakan salah satu teknik *ensemble* yang dapat digunakan untuk memprediksi. Rumus dari teknik tersebut dapat diilustrasikan dalam Persamaan (1.3) [15].

$$\mu_i = \frac{\sum_{i=1}^n w_i \bar{y}_i}{\sum_{i=1}^n w_i} \quad (1.3)$$

dengan  $y_i$  merupakan nilai prediksi dari model ke- $i$  dan  $w_i$  adalah pembobotnya. Pembobot dari *ensemble* merupakan korelasi antara data sebenarnya dengan data hasil prediksi K-NN tunggal yang ditentukan melalui Persamaan (1.4) [15].

$$w_i = \frac{r_h}{\sum_{h=1}^n r_h} \quad (1.4)$$

Keterangan:

$w_i$  : Pembobot *ensemble* ke- $i$

$r_h$  : Korelasi antara data sebenarnya dengan data hasil prediksi K-NN Tunggal dengan  $k$  ke- $h$

$n$  : Banyaknya data

$h$  : 1, 2, 3, ...,  $s$

Korelasi *product moment* merupakan hubungan antara dua variabel yang akan diteliti, untuk mengetahui hubungan antara variabel prediktor ( $x$ ) dengan variabel respon ( $y$ ). Pada penelitian ini korelasi ingin diketahui adalah korelasi antara data sebenarnya dengan data hasil prediksi K-NN tunggal. Rumus yang digunakan terdapat pada Persamaan (1.5) [5].

$$r_h = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2} \sqrt{\sum(y-\bar{y})^2}} \quad (1.5)$$

Keterangan:

$x$  : data variabel  $x$

$\bar{x}$  : rata-rata data variabel  $x$

$y$  : data variabel  $y$

$\bar{y}$  : rata-rata data variabel  $y$

### 1.4. Mean Absolute Percentage Error (MAPE)

Hasil prediksi yang diperoleh, kemudian dievaluasi menggunakan sejumlah metode untuk mengetahui ketepatan hasil prediksi [1]. Metode evaluasi yang digunakan pada penelitian ini adalah dengan melihat nilai *Mean Absolute Percentaging Error* (MAPE). Ketika ukuran atau besar variabel itu penting dalam mengevaluasi ketetapan prediksi, maka dapat digunakan MAPE. MAPE dihitung dengan menggunakan kesalahan absolut pada setiap periode ( $e_i$ ) dibagi dengan

nilai pengamatan yang sebenarnya. Sehingga, besar kesalahan dalam prediksi yang dibandingkan dengan nilai sebenarnya dapat ditunjukkan melalui nilai MAPE. Semakin kecil (mendekati 0) nilai MAPE, maka semakin tepat hasil prediksi. MAPE dapat dihitung dengan menggunakan Persamaan (1.6) [13].

$$MAPE = \frac{\sum |e_i| \times 100\%}{n} = \frac{\sum_{t=1}^m \frac{|\hat{y}_t - y_t|}{y_t}}{n} \times 100\% \quad (1.6)$$

Dimana:

$\hat{y}_t$  : nilai sebenarnya ke- $t$   
 $y_t$  : nilai hasil prediksi ke- $t$   
 $n$  : banyaknya data

Akurasi hasil prediksi yang diperoleh berdasarkan nilai MAPE dapat dilihat pada Tabel 1.1 berikut [3]:

**Tabel 1.1** Akurasi Hasil Prediksi Berdasarkan Nilai MAPE

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

## 2. DATA DAN METODOLOGI

### 1.2. Data

Data dalam penelitian ini adalah data sekunder yang diperoleh dari sumber yang relevan. Variabel-variabel yang akan digunakan dalam penelitian ini sebagai variabel respon ( $y$ ) dan variabel prediktor ( $x$ ) disajikan dalam bentuk tabel sebagai berikut :

**Tabel 2.1.** Data penelitian

Variabel	Definisi	Sumber
Nilai IHSG ( $Y$ )	Informasi historis terkait harga seluruh saham dari waktu ke waktu	www.finance.yahoo.com
Harga Emas ( $X_1$ )	Harga dalam dollar Amerika dengan bobot menggunakan satuan <i>tray once</i> (TO)	id.investing.com
DJIA ( $X_2$ )	Indeks pasar saham di Amerika	www.finance.yahoo.com
Nilai tukar rupiah ( $X_3$ )	Harga satuan dollar Amerika (USD) yang dinilai ke dalam satuan rupiah	www.bi.go.id

## 1.2. Metodologi

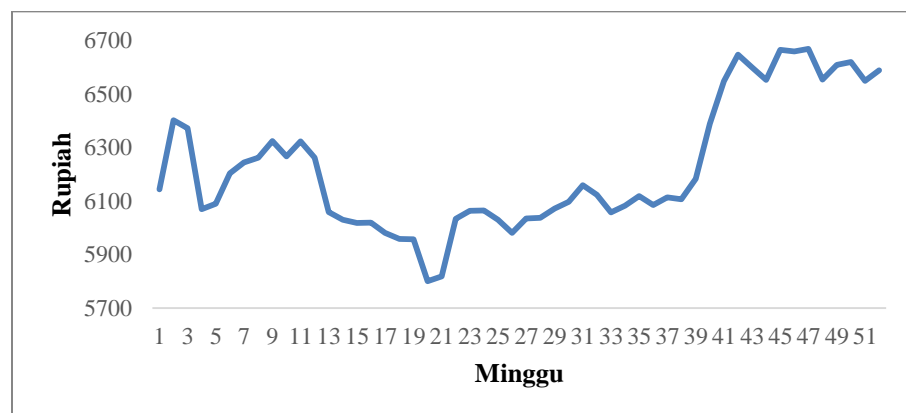
Metodologi dalam penelitian ini menggunakan *Ensemble K-Nearest Neighbor* (K-NN) dengan bantuan *software* RStudio dan *Microsoft Excel*. Tahapan-tahapan analisis yang dilakukan adalah sebagai berikut:

1. Pengumpulan data.
2. Eksplorasi data.
3. Normalisasi data.
4. Membagi data *training* dan data *testing*.
5. Menghitung jarak *euclidean* antara data *training* dan data *testing*. Kemudian, mengurutkan jarak *euclidean* tersebut dari jarak yang terkecil hingga terbesar.
6. Menentukan jumlah tetangga terdekat ( $k$ ) dengan menuju (Sinta, 2014) yaitu terdiri dari 3, 4, 6, 9, 10, 12, 15 dan 24.
7. Menghitung prediksi menggunakan metode K-NN dengan jumlah  $k$  yang telah ditentukan.
8. Menghitung prediksi menggunakan metode *ensemble* K-NN.
9. Melakukan evaluasi terhadap hasil prediksi nilai Indeks Harga Saham Gabungan (IHSG) berdasarkan nilai *Mean Absolute Percentage Error* (MAPE).
10. Kesimpulan.
11. Selesai.

## 3. HASIL DAN PEMBAHASAN

### 3.1. Ekplorasi Data

Beberapa variabel respon dari metode prediksi dapat digabungkan dengan teknik *ensemble*. Teknik ini merupakan salah satu metode prediksi yang biasa digunakan, terutama pada prediksi iklim. Studi terakhir telah menunjukkan bahwa gabungan beberapa prediksi dapat meningkatkan ketahanan dan kehandalan, yaitu *ensemble* (Lusia dan Suhartono, 2013).



**Gambar 3.1.** Grafik Data Indeks Harga Saham Gabungan (IHSG)

Pada Gambar 3.1 di atas dapat diketahui bahwa nilai Indeks Harga Saham Gabungan (IHSG) mengalami beberapa tren, baik tren naik maupun tren turun. Pada periode 11 hingga 20 nilai IHSG mengalami penurunan dari 6323,035 hingga 5799,911, kemudian terjadi kenaikan pada periode 21 hingga 42 dari 5871,479 hingga 6647,869. Pada periode selanjutnya IHSG sudah mulai stasioner. Nilai IHSG terendah terjadi pada periode 20 yaitu sebesar 5799,911, sedangkan

nilai tertinggi terjadi pada periode 47 yaitu sebesar 6669,088. Nilai IHSG memiliki rata-rata sebesar 6225,193.

### 3.2 Normalisasi Data

Proses ini digunakan menyamakan skala nilai dari tiga variabel prediktor (X) yang digunakan dalam penelitian sehingga dapat berada pada rentang tertentu. Normalisasi yang digunakan pada penelitian ini adalah normalisasi Min-Max. Normalisasi ini dilakukan dengan bantuan *Microsoft Excel*. Berikut adalah hasil normalisasi dari tiga variabel prediktor (X) tersebut:

**Tabel 3.1.** Normalisasi data

No.	Y	$X_1$	$X_2$	$X_3$
1	6143,878125	1	0	0,02842434
2	6403,107617	0,64160501	0,27078683	0,07686615
3	6372,493359	0,68539852	0,16448931	0,091874
4	6069,92998	0,67013206	0,20546318	0
5	6089,615137	0,55820923	0,12796763	0,03708561
6	6203,721802	0,5658971	0,07155731	0,1532841
7	6244,537695	0,36475602	0,06354218	0,17344257
8	6262,123047	0,35215885	0,26009501	0,16278941
9	6324,80459	0,03877494	0,54661669	0,13650562
10	6267,75	0,02097968	0,68319627	0,1532841
⋮	⋮	⋮	⋮	⋮
52	6588,986328	0,447878618	0,429042481	1

### 3.3 Jarak Euclidean

Jarak *euclidean* bertujuan untuk melihat jarak ketetanggaan paling dekat antara data *training* dan data *testing* pada jumlah  $k$  tertentu. Data yang digunakan untuk *training* sebanyak 39 data yaitu data pada periode 1 sampai 39. Sedangkan, data untuk *testing* sebanyak 13 yaitu data pada periode 40 sampai 52. Perhitungan jarak *euclidean* dilakukan dengan bantuan *software RStudio*, hasilnya adalah sebagai berikut:

**Tabel 3.2.** Jarak *euclidean*

	1	2	3	4	5	...	52
1	0	0,4518	0,36063	0,38966	0,46003	...	1,19703
2	0,4518	0	0,11594	0,10483	0,1701	...	0,95643
3	0,36063	0,11594	0	0,10175	0,14322	...	0,97524
4	0,38966	0,10483	0,10175	0	0,14109	...	1,04852
5	0,46003	0,1701	0,14322	0,14109	0	...	1,0149
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
52	1,19703	0,95643	0,97524	1,04852	1,0149	...	0

### 3.4 K-Nearest Neighbor (K-NN)

Pada penelitian ini, K-NN bertujuan mencari prediksi nilai IHSG dari data *testing* berdasarkan jarak yang paling dekat data *training* pada jumlah  $k$  tertentu. Jumlah  $k$  yang

digunakan pada penelitian ini sebanyak 8, yaitu  $k = 3, 4, 6, 9, 10, 12, 15, \text{ dan } 24$ . Perhitungan K-NN Tunggal dilakukan dengan bantuan *software* RStudio, hasilnya adalah sebagai berikut:

**Tabel 3.3.** Hasil prediksi K-NN Tunggal

No.	$k = 3$	$k = 4$	$k = 6$	$k = 9$	$k = 10$	$k = 12$	$k = 15$	$k = 24$
1	6134,5	6090,14	6095	6090,77	6077,51	6079,84	6066,84	6080,72
2	6052,3	6065,87	6085,16	6093,95	6096,93	6079,84	6066,84	6089,19
3	6052,3	6065,87	6094,09	6090,03	6096,93	6084,15	6074,51	6071,95
4	6052,3	6068,79	6081,91	6082,79	6092,79	6084,15	6078,32	6040,76
5	6087,08	6093,84	6102,8	6089,35	6098,7	6084,15	6078,32	6040,76
6	6095,45	6111,34	6090,54	6053,64	6057,93	6060,93	6029,76	6040,76
7	6072,05	6075,46	6030,67	6025,04	6033,95	6029,49	6029,76	6040,76
8	6106,03	6093,84	6081,91	6097,82	6098,7	6084,15	6074,51	6071,42
9	6118,07	6096,21	6083,48	6072,37	6077,51	6069,84	6066,84	6079,03
10	6099,56	6105,61	6106,21	6067,92	6069,7	6070,15	6062,18	6079,03
11	6099,56	6104,24	6104,4	6088,36	6077,58	6070,15	6067,96	6079,03
12	6111,52	6105,06	6102,8	6072,86	6075,23	6070,64	6060,09	6050,36
13	6120,98	6111,34	6102,8	6084,7	6085,89	6071,24	6078,63	6040,76

Berdasarkan Tabel 3.3 dapat dilihat bahwa hasil prediksi  $k = 3$  memiliki hasil yang paling mendekati nilai sebenarnya. Nilai IHSG sebenarnya pada data 1 sebesar 6389.24, setelah diprediksi hampir mendekati nilai sebenarnya yaitu sebesar 6134,5. Hal tersebut juga dapat dibuktikan melihat nilai MAPE pada Tabel 3.4. Semakin kecil nilai MAPE, maka semakin baik hasil prediksinya. Tabel 3.4 memperlihatkan bahwa  $k = 3$  menghasilkan nilai MAPE yang lebih kecil yaitu sebesar 7,52%.

**Tabel 3.4.** Evaluasi hasil prediksi K-NN Tunggal

K-NN	MAPE (%)
$k = 3$	7,52
$k = 4$	7,54
$k = 6$	7,57
$k = 9$	7,75
$k = 10$	7,71
$k = 12$	7,83
$k = 15$	7,95
$k = 24$	7,99

Berdasarkan Tabel 3.4 dapat dilihat bahwa nilai MAPE pada  $k = 4$  sebesar 7,54%, pada  $k = 6$  sebesar 7,57%, pada  $k = 9$  sebesar 7,75%, pada  $k = 10$  sebesar 7,71%, pada  $k = 12$  sebesar 7,83%, pada  $k = 15$  sebesar 7,95%, dan pada  $k = 24$  sebesar 7,99%.

### 3.5 Ensemble K-Nearest Neighbor (K-NN)

Teknik *ensemble* dilakukan untuk menggabungkan hasil prediksi K-NN tunggal menjadi suatu satu buah prediksi akhir. Teknik *ensemble* yang digunakan pada penelitian ini adalah



*weighted mean* (rata-rata terboboti). Pembobot dari teknik ini adalah korelasi antara data sebenarnya dengan data hasil prediksi K-NN tunggal. Korelasi tersebut didapatkan menggunakan rumus pada Persamaan (1.5), dengan bantuan *software Microsoft Excel*, hasilnya adalah sebagai berikut:

**Tabel 3.5.** Korelasi antara data sebenarnya dengan hasil prediksi K-NN Tunggal

$r_{y,k}$	Korelasi
$r_{y,k=3}$	-0,43182
$r_{y,k=4}$	0,00246
$r_{y,k=6}$	-0,30235
$r_{y,k=9}$	-0,35233
$r_{y,k=10}$	-0,18887
$r_{y,k=12}$	-0,32223
$r_{y,k=15}$	-0,30210
$r_{y,k=24}$	-0,30489

Setelah mendapatkan korelasi antara data sebenarnya dengan data hasil prediksi K-NN tunggal, maka dilakukan perhitungan untuk pembobot *ensemble*. Pembobot dari teknik *ensemble* didapatkan menggunakan rumus pada Persamaan (1.4), dengan bantuan *software Microsoft Excel*, hasilnya adalah sebagai berikut:

**Tabel 3.6.** Pembobot *ensemble*

$w_i$	Bobot
$w_1$	0,19609
$w_2$	-0,00112
$w_3$	0,13730
$w_4$	0,15999
$w_5$	0,08577
$w_6$	0,14633
$w_7$	0,13719
$w_8$	0,13845

Setelah pembobot dari teknik *ensemble* didapatkan, maka dilakukan prediksi menggunakan rumus pada Persamaan (1.3), dengan bantuan *software Microsoft Excel*, hasilnya adalah sebagai berikut:

**Tabel 3.7.** Hasil prediksi *ensemble* K-NN

No.	<i>Ensemble</i>	Aktual
1	6092,51	6389,24
2	6078,42	6548,46
3	6078,31	6647,87
4	6071,33	6600,05
5	6082,55	6553,3
6	6063,21	6665,55
7	6039,21	6659,95

8	6088,47	6669,09
9	6083,06	6554,44
10	6080,57	6610,07
11	6085,06	6620,3
12	6079,53	6549,89
13	6085,48	6588,99

Berdasarkan Tabel 3.7 dapat dilihat bahwa hasil prediksi menggunakan teknik memiliki hasil yang cukup baik dibandingkan K-NN tunggal. Hal tersebut juga dapat dibuktikan melihat nilai MAPE pada Tabel 3.8. Semakin kecil nilai MAPE, maka semakin baik hasil prediksinya. Tabel 3.8 memperlihatkan bahwa teknik *ensemble* menghasilkan nilai MAPE yang lebih kecil yaitu sebesar 7,16%.

**Tabel 3.8.** Evaluasi hasil Prediksi *ensemble* K-NN

K-NN	MAPE (%)
<i>Ensemble</i>	7,16

#### 4. KESIMPULAN

Berdasarkan nilai *Mean Absolute Percentage Error* (MAPE) hasil prediksi nilai Indeks Harga Saham Gabungan (IHSG) di Indonesia menggunakan metode *ensemble K-Nearest Neighbor* memiliki hasil yang baik dibandingkan K-NN tunggal yaitu sebesar 7,16%. Hasil tersebut menunjukkan bahwa hasil prediksi nilai Indeks Harga Saham Gabungan (IHSG) di Indonesia menggunakan metode *ensemble K-Nearest Neighbor* memiliki akurasi yang tinggi.

#### DAFTAR PUSTKA

- [1] Caraka, R. E., Chen, R. C., Yasin, H., Suhartono, Lee, Y., & Pardamean, B., 2021. Hybrid Vector Autoregression Feedforward Neural Network with Genetic Algorithm Model for Forecasting Space-Time Pollution Data. *Indonesian Journal of Science & Technology*, 243-266.
- [2] Gufron, Surarso, B., & Gernowo, R., 2019. Implementation of the K-Nearest Neighbor Method to determine the Classification of the Study Program Operational Budget in Higher Education. *1st International Conference of Health, Science & Technology (ICOHETECH) 2019*, 201-204.
- [3] Gustriansyah, R., 2017. Analisis Metode Single Exponential dengan Brown Exponential Smoothing Pada Studi Kasus Memprediksi Kuantiti Penjualan Produk Farmasidi Apotek. *Seminar Nasional Teknologi Informasi dan Multimedia*, Vol. 5, No. 1, 7-12.
- [4] Hidayati, Y. P., 2019. Prediksi Indeks Pembangunan Manusia (IPM) dengan Menggunakan Metode Ensemble K-Nearest Neighbor (K-NN).
- [5] Jainuddin, M., 2016. Hubungan Antara Ketersediaan Buku Referensi Perpustakaan Dengan Peningkatan Minat Baca Siswa Pada Perpustakaan SMPN 17 Kendari. *Jurnal Ilmu Komunikasi UHO*, Vol. 1, No. 2.

- [6] Kang, S., 2021. k-Nearest Neighbor Learning with Graph Neural Networks. *Mathematics*, 9, 830.
- [7] Khaira, U., Utomo, P. E., Suratno, T., & Gulo, P. C., 2019. Prediksi Indeks Harga Saham Gabungan (IHSG) Menggunakan Algoritma Autoregressive Integrated Moving Average (ARIMA). *Jurnal Sains dan Sistem Informasi*, Vol. 2, No. 2, 11-17.
- [8] Khoiri, H. A., & Arghawaty, E., 2020. Menganalisis Nilai IHSG Beserta Faktor-Faktor Yang Mempengaruhi Di Era Pandemi Covid-19. *Jurnal Riset Akuntansi & Keuangan Dewantara*, Vol. 3, No. 2, 110-121.
- [9] Kusumodestoni, R. H., & Sarwido., 2017. Komparasi Model Support Vector Machines (SVM) dan Neural Network Untuk Mengetahui Tingkat Akurasi Prediksi Tertinggi Harga Saham. *JURNAL INFORMATIKA UPGRIS*, Vol. 3, No. 1, 1-9.
- [10] Lusia, D. A., & Suhartono., 2013. Ensemble Method Based on Two Level ARIMAXFFNN for Rainfall Forecasting in Indonesia. *International Journal of Science and Research (IJSR)*, Vol. 2, No. 2, 144-149.
- [11] Nurmahaludin., 2014. Analisis Perbandingan Metode Jaringan Syaraf Tiruan Dan Regresi Linear Berganda Pada Prakiraan Cuaca. *Jurnal INTEKNA : Informasi Teknik dan Niaga*, Vol. 14, No. 2, 102-109.
- [12] Puspitasari, I., Suparti, & Wilandari, Y., 2012. Analisis Indeks Harga Saham Gabungan (IHSG) Dengan Menggunakan Model Regresi Kernel. *Jurnal Statistika Undip*, Vol. 1, No. 1, 93-103.
- [13] Satriya, R. H., Santoso, E., & Sutrisno., 2018. Implementasi Metode Ensemble K-Nearest Neighbor untuk Prediksi Nilai Tukar Rupiah Terhadap Dollar Amerika. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 2, No. 4, 1718-1725.
- [14] Saxena, R., Sharma, D. S., & Gupta, M., 2021. Role of K-nearest neighbour in detection of Diabetes Mellitus. *Turkish Journal of Computer and Mathematics Education*, Vol. 12, No. 10, 373-376.
- [15] Sinta, D., 2014. Ensemble K-Nearest Neighbors Method to Predict Rice Price in Indonesia. *Applied Mathematical Sciences*, Vol. 8, No. 160, 7993-8005.
- [16] Permana, T., 2020. Perbandingan Hasil Prediksi Kredit Macet Pada Koperasi Menggunakan Algoritma KNN dan C5.0. *Conference on Innovation and Application of Science and Technology (CIASTECH 2020)* (pp. 737-746). Malang: Universitas Widyagama Malang.