

PREDIKSI INDEKS HARGA SAHAM GABUNGAN (IHSG) MENGUNAKAN REGRESI K-NEAREST NEIGHBORS

JAKARTA COMPOSITE INDEX (JCI) PREDICTION USING K-NEAREST NEIGHBORS REGRESSION

Najibullah¹, Azlina², Ricky Ariansyah³, Lisnawati⁴

¹Analisis Keuangan/Politeknik Kutaraja, Banda Aceh, Indonesia

²Analisis Keuangan, Politeknik Kutaraja, Banda Aceh, Indonesia

³Administrasi Perkantoran, Politeknik Kutaraja, Banda Aceh, Indonesia

⁴Analisis Keuangan, Politeknik Kutaraja, Banda Aceh, Indonesia

Korespondensi: ¹ najibullah@poltekkutaraja.ac.id

ABSTRACT

This paper aims to evaluate the accuracy of the k-Nearest Neighbors regression method in predicting the Jakarta Composite Index (JCI), thus adding evidence of the application of k-NN regression in predicting noisy financial time series data. The dataset consists of daily data from several time series variables in 2022. The result suggests that the accuracy of k-NN regression in predicting JCI fluctuation is very good, indicated by Mean Absolute Percentage Error (MAPE) being less than 2% for $k = 2, 3, \dots, 10$.

Keywords : k-NN, Regression, Prediction, Machine Learning, JCI

ABSTRAK

Penelitian ini bertujuan untuk mengevaluasi keakuratan metode regresi k-Nearest Neighbors dalam memprediksi Indeks Harga Saham Gabungan (IHSG), sehingga menambah bukti penerapan regresi k-NN dalam memprediksi data time series keuangan yang noisy. Dataset tersebut terdiri dari data harian dari beberapa variabel time series selama tahun 2022. Hasil penelitian menunjukkan bahwa keakuratan regresi k-NN dalam memprediksi fluktuasi IHSG sangat baik, hal ini ditunjukkan dengan mean absolute percentage error (MAPE) yang kurang dari 2% untuk $k = 2, 3, \dots, 10$.

Kata Kunci : k-NN, Regresi, Prediksi, Machine Learning, IHSG

1. PENDAHULUAN

Pemahaman mengenai pergerakan pasar ekuitas di suatu negara merupakan hal yang sangat krusial bagi partisipan pasar modal untuk menentukan bauran kelas aset yang sesuai di dalam portofolio. Salah satu ukuran yang sering digunakan untuk mengukur kondisi pasar ekuitas adalah indeks harga yang merupakan indikator yang mencerminkan keadaan pasar. Beberapa indeks harga saham yang terkenal adalah Indeks Standard &



Poor 500 (SP500), Dow Jones Industrial Average (DJIA), Financial Times Stock Exchange (FTSE) index, Hang Seng Stock Index (HSI), Nikkei (N225) dan Indeks Harga Saham Gabungan (IHSG). Masing-masing indeks tersebut menggambarkan keadaan pasar saham di negara terkait. Para praktisi pasar saham umumnya menggunakan analisis teknikal dan kualitatif untuk meramalkan pergerakan pasar saham di masa yang akan datang.

Perkembangan algoritma *machine learning* selama beberapa dekade terakhir telah memungkinkan praktisi pasar keuangan untuk mengaplikasikan algoritma tersebut untuk memprediksi keadaan pasar maupun harga saham tertentu di masa yang akan datang, setidaknya untuk jangka pendek. Satriya, dkk (2018) berusaha meramalkan nilai tukar rupiah dengan menggunakan *ensemble k-NN* memperoleh hasil prediksi yang sangat akurat dengan nilai *Mean Absolute Percentage Error* (MAPE) dibawah 5%[1]. Penelitian lain yang juga menerapkan metode *ensemble k-NN* untuk memprediksi IHSG sebagai target variabel dengan indeks DJIA, harga emas, dan nilai tukar rupiah mampu memperoleh akurasi prediksi yang tinggi (MAPE < 10%)[2]. Priambodo, dkk (2019) menggunakan sampel yang relatif kecil menemukan bahwa metode regresi k-NN mampu memprediksi PDB Indonesia dengan nilai MAPE kurang dari 7%, lebih baik jika dibandingkan metode *Neural Network* (NN) dan *Multiple Linear Regression* (MLR) yang masing-masing memiliki nilai MAPE 29,80 dan 91,59[3]. Sebaliknya, Maulida, dkk dalam penelitiannya yang berusaha memprediksi harga saham PT Telkom Indonesia Tbk (TLKM) menemukan bahwa metode NN memiliki tingkat keakuratan yang lebih baik jika dibandingkan dengan metode k-NN[4]. Meskipun membutuhkan biaya komputasi yang lebih besar, algoritma k-NN mampu menghasilkan prediksi yang lebih akurat terutama ketika diaplikasikan pada sampel kecil. Fitur input yang digunakan dalam k-NN akan menentukan akurasi yang diperoleh dalam regresi k-NN sehingga diperlukan pemahaman mengenai faktor-faktor yang mempengaruhi variabel target. Khoiry dan Arghawaty (2020) menemukan bahwa IHSG dipengaruhi secara signifikan oleh harga emas dan nilai tukar Rupiah terhadap US Dollar[5]. Sementara Najibullah (2022) menemukan bahwa beberapa indeks global seperti S&P500 dan HSI berpengaruh positif dan signifikan terhadap arah pergerakan IHSG[6].

Penelitian ini bertujuan untuk memperkaya literatur mengenai efektifitas algoritma regresi k-NN dalam memprediksi variabel target khususnya pada data deret waktu keuangan yang umumnya memiliki *noise* yang tinggi. Secara spesifik, penelitian ini mengukur akurasi dari algoritma k-NN dalam memprediksi IHSG dengan menggunakan tambahan beberapa fitur input yang berbeda dari penelitian sebelumnya.

2. TINJAUAN PUSTAKA

2.1. Regresi k-Nearest Neighbor

Metode k-NN memiliki beberapa istilah yang hampir sama namun mengacu kepada



algoritma yang berbeda. Algoritma *k*-NN classifier, merupakan algoritma supervised dengan model probabilita yang digunakan ketika variabel target merupakan variabel kategorikal dimana algoritma akan mengklasifikasikan data testing ke dalam kategori tertentu berdasarkan data training. Sementara algoritma *k*-NN regression merupakan salah satu algoritma supervised dalam kelas model deterministik dimana nilai variabel target yang diprediksi merupakan nilai rata-rata dari *k* data training yang terpilih sebagai tetangga terdekat. Kedua metode merupakan metode non-parametrika yang tidak memerlukan pemenuhan asumsi mengenai struktur data.

Metode *k*-NN juga memiliki kelebihan seperti akurasi yang tinggi dan tidak mudah dipengaruhi oleh outlier[2]. Hanya saja, sebagaimana metode non-parametrika lainnya, metode ini tidak dapat menghasilkan output yang stabil dan tidak dapat memberikan informasi mengenai derajat kepentingan dari masing-masing fitur input. Metode ini juga membutuhkan komputasi yang lebih kompleks dan mengkonsumsi banyak memori jika dibandingkan dengan metode lainnya. Dikarenakan target variabel dalam penelitian ini merupakan data kontinu, penelitian ini menggunakan algoritma regresi *k*-NN untuk memprediksi nilai IHSG.

Algoritma regresi *k*-NN mencari *k* objek dari data training yang memiliki jarak paling dekat dengan data testing dimana ukuran jarak yang sering digunakan adalah jarak euclidean[2].

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

Keterangan:

- $d()$: jarak euclidean
- $x_{train,i}$: data training ke-*i*
- $x_{test,j}$: data testing ke-*j*
- n : jumlah observasi
- $i, j : 1, 2, 3 \dots n$

Secara spesifik, detail algoritma dalam metode regresi *k*-NN dapat dijabarkan sebagai berikut [2]:

1. Menentukan nilai *k*
2. Menghitung jarak euclidean dengan persamaan (2.1)
3. Menentukan jarak terdekat sampai urutan *k*
4. Menentukan nilai prediksi dengan menghitung rata-rata dari nilai objek pada jumlah tetangga terdekat seperti persamaan berikut:

$$\bar{y}_i = \frac{1}{k} \sum_{i=1}^k y_i \tag{2}$$



Keterangan:

- \hat{y}_i : prediksi k -NN dengan jumlah k ke- i
- k : jumlah tetangga terdekat
- y_i : urutan jarak *euclidean* ke- i

Secara manual, penentuan nilai k yang optimal akan menambah kompleksitas komputasi algoritma k -NN terutama ketika jumlah sampel dan fitur input yang digunakan semakin besar. Dalam penelitian ini, penentuan nilai k yang optimal dapat dilakukan secara otomatis dengan menggunakan bahasa pemrograman Julia via modul *NearestNeighbor.jl*.

2.2. Mean Absolute Percentage Error (MAPE)

Derajat ketepatan prediksi dari metode k -NN akan diukur dengan menggunakan MAPE. Menurut Hyndman & Athanasopoulos (2021), MAPE memiliki kelebihan dalam hal tidak terikat oleh unit pengukuran dan mudah diinterpretasikan. Namun demikian, ukuran ini memberikan bobot lebih besar pada error negatif dibandingkan error positif[7]. Ukuran ini menghitung kesalahan pada setiap periode (e_i) dalam *term* absolute, dibagi dengan nilai observasi aktual. Jika dikalikan 100 maka nilai MAPE dapat diartikan sebagai persentase error rata-rata. Persamaan (2.3) berikut menyajikan formula perhitungan MAPE:

$$MAPE = \left(\frac{1}{n} \sum_{i=1}^n e_i \right) \times 100\% = \left(\frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \right) \times 100\% \tag{3}$$

dimana:

- \hat{y}_i : nilai prediksi ke- i
- y_i : nilai aktual ke- i
- n : jumlah observasi

Nilai MAPE yang lebih rendah menunjukkan persentase kesalahan prediksi yang lebih sedikit. Beberapa peneliti terdahulu mengklasifikasikan nilai MAPE kedalam beberapa kategori sebagai berikut[2], [8]:

Tabel 1. Klasifikasi tingkat akurasi berdasarkan nilai MAPE

Nilai MAPE	Akurasi
MAPE < 10%	tinggi
10% < MAPE < 20%	baik
20% < MAPE < 50%	moderat
MAPE > 50%	rendah



3. METODE

3.1. Data dan Operasional Variabel

Data dalam penelitian ini diperoleh dari sumber yang relevan dan kredibel dengan rincian sebagai berikut:

Tabel 2. Rincian variabel target dan fitur input

Variabel	Definisi	Unit ukur
IHSG (Y)	Informasi historis nilai penutupan akhir hari indeks	indeks
Harga emas (X_1)	harga penutupan akhir hari emas dalam USD/troy ounce	USD/troy ounce
Nilai tukar Rupiah (X_2)	harga penutupan akhir hari USD /IDR	Rupiah
S&P500(X_3)	Informasi historis nilai penutupan akhir hari indeks hari sebelumnya	indeks
HSI (X_4)	Informasi historis nilai penutupan akhir hari indeks	indeks
N225 (X_5)	Informasi historis nilai penutupan akhir hari indeks	indeks

3.2. Langkah-langkah regresi k -NN

Metode yang digunakan dalam penelitian ini adalah regresi k -NN standar yang akan dihitung menggunakan bahasa pemrograman Julia via kerangka kerja *MLJ* dan modul (*library*) *NearestNeighbors.jl*. *MLJ* sendiri merupakan kerangka kerja *Machine Learning* dalam bahasa pemrograman Julia. Seluruh tahapan dalam penelitian ini dilakukan sepenuhnya menggunakan bahasa pemrograman Julia.

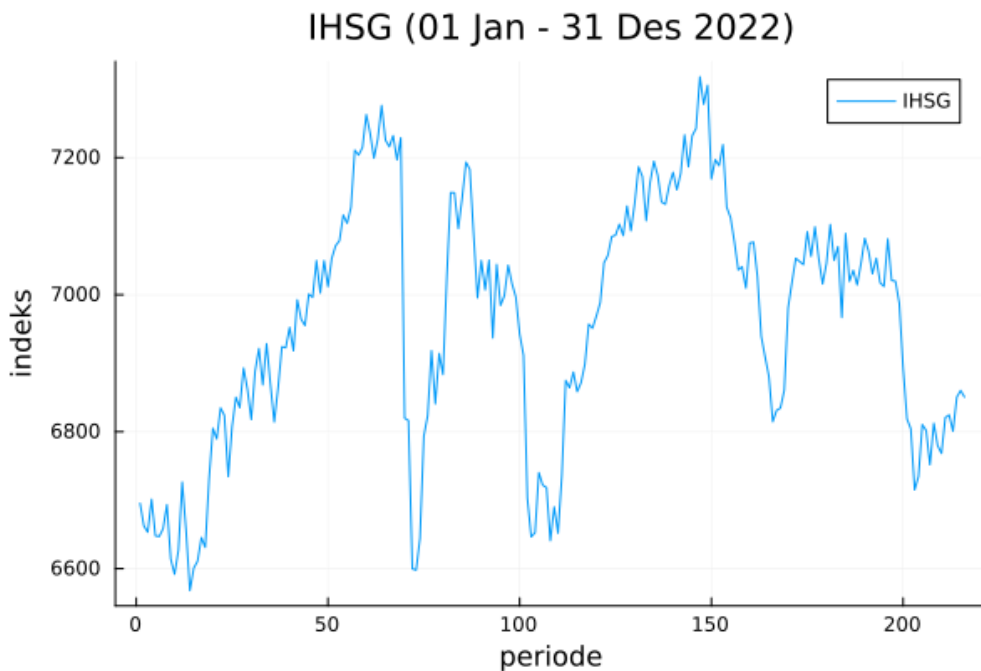
1. Pengumpulan data.
2. Eksplorasi data.
3. Konversi fitur input ke skala z (0 hingga 1).
4. Pemisahan data menjadi data *training* dan *testing*.
5. Menghitung nilai prediksi untuk nilai beberapa k tertentu.
6. Evaluasi akurasi prediksi menggunakan MAPE
7. Kesimpulan



4. HASIL DAN PEMBAHASAN

4.1. Eksplorasi data

Variabel respon (target) dalam penelitian ini adalah nilai penutupan akhir hari dari IHSB. Nilai indeks ini senantiasa berubah seiring berubahnya harga dari setiap saham yang terdaftar di Bursa Efek Indonesia (BEI).



Gambar 1. Grafik historis IHSB (01 Jan– Des 2022)

Sebagaimana disajikan pada Gambar 1, selama tahun 2022, IHSB cenderung bergerak *sideways* (tidak memiliki tren) pada range nilai indeks 6500 hingga 7400. Hal ini mencerminkan perilaku partisipan pasar yang relatif *wait and see* terkait dengan prospek ekonomi di tahun mendatang disebabkan adanya upaya normalisasi suku bunga oleh bank sentral Amerika Serikat yang akan berdampak pada prospek perekonomian global. Tabel berikut menyajikan statistik deskriptif untuk seluruh variabel yang terdapat dalam penelitian ini:

Tabel 3. Statistik deskriptif

Variabel	Mean	Std	Min	Q1	Mediar	Q3	Max
hsi_t	202	23	146	190	201	217	249
ihs_g_t	69	1	65	68	69	70	73
$n225_t$	272	8	247	266	273	279	293
$emas_t$	17		16	172	17	184	20
$idrx_t$	148	4	142	143	148	152	157
$sp500_{t-1}$	40	2	35	38	40	43	47

4.2. Konversi fitur input ke skala z (0 s.d 1)

Penelitian-penelitian sebelumnya telah melaporkan bahwa penyetaraan skala di antara fitur (variabel) input yang digunakan akan berpengaruh signifikan terhadap derajat akurasi prediksi yang dihasilkan oleh algoritma k -NN[2], [9]. Konversi ini dilakukan dengan menggunakan persamaan berikut:

$$z_i = \frac{x_i - \min X}{\max X - \min X} \tag{4}$$

dimana:

- z_i : variabel observasi ke- i dalam skala 0 s.d 1
- X : vektor variabel X
- x_i : bilangan ke- i dari variabel X
- $\min X$: nilai minimum dari variabel X
- $\max X$: nilai maximum dari variabel X

4.3. Regresi k -NN

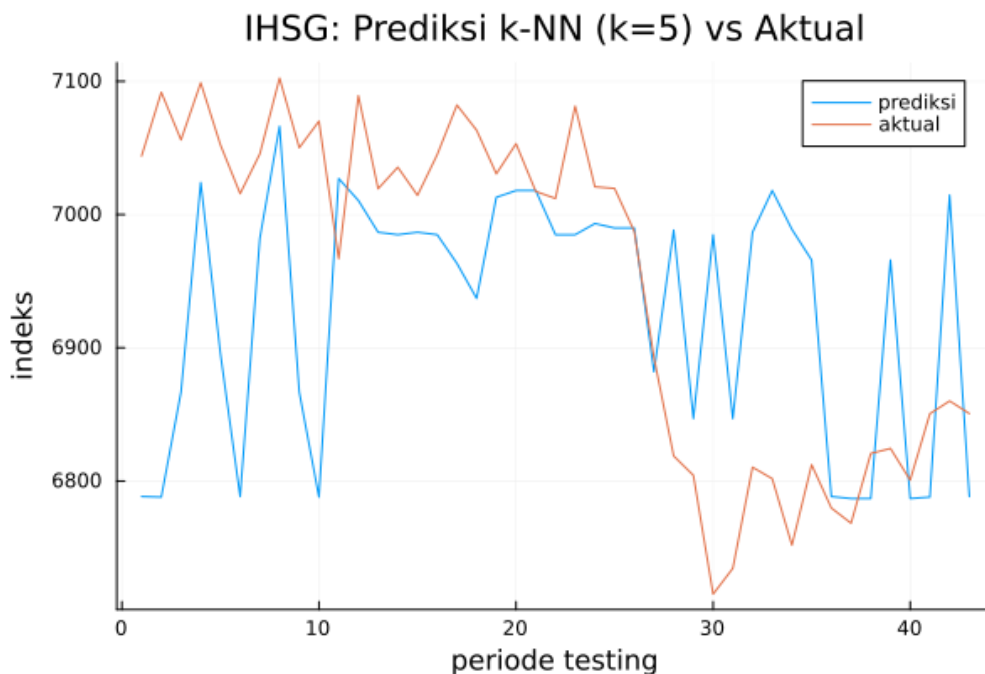
Sebelum implementasi regresi k -NN, data dalam penelitian ini yang berjumlah sebanyak 216 observasi, dibagi menjadi 80% untuk estimasi regresi k -NN (proses *training*) dan 20% untuk pengujian. Data *testing* digunakan untuk mengevaluasi akurasi dari hasil *fitting* regresi k -NN menggunakan data *training*. Secara manual, nilai k yang digunakan dalam penelitian ini, yaitu $k = 2, 3, \dots, 10$. Sementara pemilihan nilai k optimal akan ditentukan berdasarkan nilai MAPE terendah. Penelitian sebelumnya, menunjukkan bahwa nilai k yang besar cenderung meningkatkan error [2].



Tabel 4. Hasil evaluasi prediksi *k*-NN standar

Jumlah <i>k</i>	MAPE(%)
2	1,78
3	1,81
4	1,59
5	1,49
6	1,50
7	1,58
8	1,69
9	1,62
10	1,64

Sebagaimana terlihat pada Tabel 4, nilai MAPE berada pada kisaran yang sangat rendah yaitu 1,4% hingga 1,8% untuk nilai $k = 2, 3, \dots, 10$. Jika mengacu pada Tabel 1, maka dapat disimpulkan bahwa metode *k*-NN memiliki kemampuan prediksi yang sangat baik bahkan untuk data pasar keuangan yang cenderung mengandung banyak *noise*. Tabel 4 juga menunjukkan bahwa nilai $k = 5$ merupakan nilai parameter *k* yang optimal berdasarkan nilai MAPE. Gambar 2 berikut menyajikan data hasil prediksi menggunakan *k*-NN dengan parameter $k = 5$.



Gambar 2. Grafik IHSG, prediksi vs aktual

5. KESIMPULAN

Berdasarkan nilai MAPE, hasil prediksi nilai IHSG pada berbagai nilai k , menunjukkan derajat akurasi yang tinggi yang diindikasikan oleh nilai $MAPE < 2\%$, yang mana lebih kecil dari 10%. Namun demikian, secara praktis, metode ini sulit untuk diaplikasikan disebabkan prediksi membutuhkan data fitur input masa yang akan datang yang tidak tersedia pada saat prediksi dilakukan. Penelitian selanjutnya dapat menggunakan nilai lag dari fitur input untuk memprediksi nilai indeks sehingga hasil prediksi dapat digunakan sebagai acuan pengambilan keputusan dalam perdagangan saham dan investasi.

DAFTAR PUSTAKA

- R. H. D. Satriya, E. Santoso, dan Sutrisno, "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, 2018.
- Moh. Jusman, N. Nur'eni, dan L. Handayani, "Ensemble K-Nearest Neighbors Method to Predict Composite Stock Price Index (CSPI) in Indonesia," *J*, vol. 18, no. 3, hlm. 423–433, Mei 2022, doi: 10.20956/j.v18i3.19641.
- B. Priambodo *dkk.*, "Predicting GDP of Indonesia Using K-Nearest Neighbour Regression," *J. Phys.: Conf. Ser.*, vol. 1339, no. 1, hlm. 012040, Des 2019, doi: 10.1088/1742-6596/1339/1/012040.
- R. Maulida, R. Septima, dan H. Syahputra, "Perbandingan Metode K-Nearest Neighbor (K-NN) Dan Neural Network (NN) Dalam Prediksi Saham PT. Telkom Indonesia Tbk (TLKM)," *Jurnal Teknik Elektro dan Informatika*, vol. 5, no. 1, hlm. 71–83, 2023.
- H. A. Khoiri dan E. Arghawaty, "Menganalisis Nilai IHSG Beserta Faktor-Faktor Yang Mempengaruhi Di Era Pandemi Covid-19," *Jurnal Riset Akuntansi dan Keuangan Dewantara*, vol. 3, no. 2, 2020.
- Najibullah, "Pengaruh Indeks Global Terhadap Harga Saham Gabungan (IHSG): Bukti Empiris Menggunakan Data Harian," *Jurnal Mahasiswa Akuntansi Samudra (JMAS)*, vol. 4, no. 4, hlm. 171–178, 2023.
- R. J. Hyndman dan G. Athanasopoulos, *Forecasting: Principles and Practice (3rd ed)*, 3rd edition. Melbourne, Australia: Otexts, 2021. Diakses: 12 Januari 2024. [Daring]. Tersedia pada: <https://otexts.com/fpp3/>
- R. Fadhillah, D. Kusnandar, dan N. M. Huda, "Pemodelan ARIMA-ANN pada Harga Saham Bank Mandiri," *Bimaster*, vol. 13, no. 1, hlm. 117–126, 2024.
- J. Tanuwijaya*, S. Hansun, dan Informatics Department, Universitas Multimedia Nusantara, Tangerang, Indonesia., "LQ45 Stock Index Prediction using k-Nearest Neighbors Regression," *IJRTE*, vol. 8, no. 3, hlm. 2388–2391, Sep 2019, doi: 10.35940/ijrte.C4663.098319.

