

Implementasi Algoritma *K-Nearest Neighbors Regression* Dalam Memprediksi Harga Saham

Divky Nurfauzan^{1*}, Titin Fatimah²

^{1,2}Fakultas Teknologi Infomasi, Mahasiswa Teknik Informatika, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹divky20@gmail.com, ²titin.fatimah@budiluhur.ac.id
(* : corresponding author)

Abstrak-Dunia investasi di Indonesia saat ini berkembang pesat, berinvestasi saham menjadi peran penting dalam perkembangan suatu perusahaan untuk terus mengembangkan suatu kinerja perusahaan. Pertambahan jumlah investor dari tahun ke tahun juga mempengaruhi perkembangan saham investasi di Indonesia. Dengan adanya pergerakan harga saham yang fluktuatif, perlu dilakukan pemodelan untuk melakukan prediksi kenaikan dan penurunan nilai saham. Melalui penggunaan statistik data, *machine learning* dilatih untuk membuat suatu prediksi. Nantinya akan membangun model *machine learning* yang sederhana dengan *dataset* yang digunakan sesuai dengan kemampuan yang dimiliki algoritma *k-nearest neighbors regression*. Dalam penelitian ini penulis memutuskan untuk memilih sample data dari PT Bank Negara Indonesia. Penelitian ini menggunakan empat variasi indikator fitur untuk memprediksi harga penutupan saham PT Bank Negara Indonesia Tbk. Satu indikator menggunakan fitur *close*, dua indikator menggunakan fitur *open* dan *close*, tiga indikator menggunakan fitur *open*, *high*, dan *close*, empat indicator menggunakan fitur *open*, *high*, *low* dan *close*. Penulis mengambil *dataset* dari platform *Yahoo Finance* sejak bulan 2 Mei 2013 sampai 12 Juli 2022 dengan jumlah *dataset* 2274 data. Hasil penelitian pemodelan menggunakan *k-nearest neighbors regression* dengan empat variasi indikator yang berbeda, prediksi dengan menggunakan satu indikator memiliki hasil yang paling baik dengan nilai *root mean squared error* 169 dengan akurasi 98.8% dengan rasio perbandingan 70% data latihan dan 30% data uji.

Kata Kunci: saham, prediksi, RMSE, KNN regression.

Implementasi Algoritma *K-Nearest Neighbors Regression* Dalam Memprediksi Harga Saham

Abstract-The world of investment in Indonesia is currently growing rapidly, investing in stocks is an important role in the development of a company to continue to develop a company's performance. The increase in the number of investors from year to year also affects the development of investment stocks in Indonesia. With the fluctuating movement of stock prices, it is necessary to model to predict the increase and decrease in the value of stocks. Through the use of data statistics, machine learning is trained to make predictions. Later will build a simple machine learning model with the dataset used according to the capabilities of the *k-nearest neighbors regression* algorithm. In this study, the author decided to choose a sample data from PT Bank Negara Indonesia. This study used four variations of feature indicators to predict the closing price of shares of PT Bank Negara Indonesia Tbk. One indicator uses the *close* feature, two indicators use the *open* and *close* features, three indicators use the *open*, *high*, and *close* features, four indicators use the *open*, *high*, *low* and *close* features. The author took a dataset from the *Yahoo Finance* platform from May 2, 2013 to July 12, 2022 with a total dataset of 2274 data. The results of the modeling study used *k-nearest neighbors regression* with four different variations of indicators, predictions using one indicator had the best results with a *root mean squared error* value of 169 with an accuracy of 98.8% with a ratio of 70% of training data and 30% of test data.

Keywords: stocks, predictions, RMSE, KNN regression.

1. PENDAHULUAN

Kecerdasan buatan merupakan teknologi yang mensimulasikan kecerdasan manusia ke dalam mesin (komputer) untuk menyelesaikan berbagai persoalan dan pekerjaan yang dilakukan oleh manusia, serta dapat memprediksi pola suatu hal yang terjadi dimasa yang akan datang. Dengan adanya pergerakan harga saham yang fluktuatif, perlu dilakukan pemodelan untuk melakukan prediksi dalam mempersiapkan strategi perusahaan dalam menghadapi kenaikan dan penurunan saham. Prediksi harga saham yang dapat naik dan turun menjadi lebih mudah dilakukan dengan adanya penganalisaan menggunakan *machine learning*. *Machine learning* itu sendiri merupakan ilmu komputer yang berfokus pada penggunaan data dan algoritma untuk meniru cara kerja manusia dalam belajar dan dapat meningkatkan akurasinya [1].

Seiring berkembangnya teknologi pada era modern ini, prediksi harga saham mendapatkan perhatian khusus oleh pakar ekonomi. Dengan adanya kemampuan prediksi atau menebak nilai harga saham, sebuah perusahaan dapat memudahkan dalam menganalisis dan memprediksi langkah kebijakan yang optimal untuk membuat keputusan pembelian/penjualan saham yang sesuai. Sehingga, prediksi menjadi alat bantu penting bagi

perencanaan yang lebih efektif dan efisien. Prediksi dilakukan dengan mencatatkan riwayat harga saham harian, mingguan, bulanan, bahkan tahunan[2].

Penelitian yang dilakukan oleh [3] Sistem Prediksi Pertumbuhan Jumlah Penduduk Kota Malang menggunakan Metode *K-Nearest Neighbor Regression*, penelitian ini, diimplementasikan metode KNN Regression untuk memprediksi jumlah penduduk kota Malang pada masa yang akan datang dengan menggunakan data time series jumlah penduduk kota Malang perbulan dengan total 73 data. Tahapan prediksi jumlah penduduk dimulai dengan melakukan *preprocessing*, perhitungan jarak antar seluruh data latih dengan data uji, kemudian perhitungan nilai prediksi berdasarkan k data latih terdekat. Penelitian yang dilakukan oleh [4] *LQ45 stock index prediction using k-nearest neighbors regression*, Penelitian ini menggunakan metode regresi kNN karena memprediksi data numerik. Hasil penelitian dalam pembuatan aplikasi prediksi indeks saham LQ45 telah berhasil dibangun. Penelitian yang dilakukan oleh [5] prediksi harga saham PT Bank *Central Asia Tbk* Metode *K-Nearest Neighbors*, penelitian ini dilakukan melalui tahap studi *literature* dengan melihat harga saham harian dimana atribut yang digunakan yaitu harga buka harian, tertinggi, terendah, tutup harian, kurs jual beli rupiah terhadap dolar Amerika Serikat dan data suku bunga 1 bulan. Penelitian yang dilakukan oleh [6] prediksi harga *cryptocurrency* dengan metode *K-Nearest Neighbors*, penelitian ini dilakukan untuk mencari *machine learning* yang paling optimal dalam memprediksi harga *cryptocurrency*. Penelitian yang dilakukan oleh [7] Implementasi Data Mining Pada Pelanggan Telkom Menggunakan Metode *K-Nearest Neighbor* untuk Memprediksi Status Pelayanan, penelitian ini bertujuan untuk klasifikasi status pelayanan berupa normal atau tidak normal pada data pelanggan. Penelitian yang dilakukan oleh [8] prediksi potensi siswa putus sekolah akibat pandemi covid-19 menggunakan algoritma *k-nearest neighbor*, hasil dari penelitian ini didapatkan pada algoritma KNN dalam mempredksi potensi siswa putus sekolah sebesar 87.4214%, dengan nilai *precision* 88.2%, *recall* 87.4% dan *F-Measure* 87%.

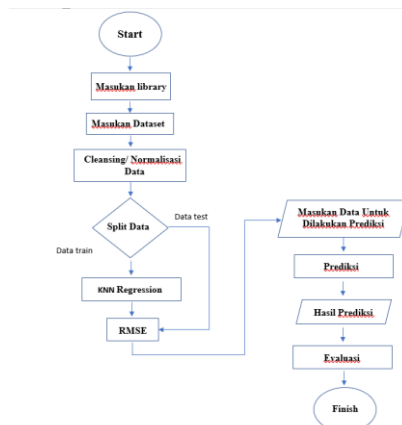
2. METODE PENELITIAN

2.1 Data Penelitian

Dalam metode pengumpulan data, dilakukan tiga tahapan penelitian. Tahapan tersebut dipaparkan pada sub-bab berikut. Pengumpulan data pertama yang dilakukan yaitu dengan menggunakan metode kuantitatif dengan pendekatan studi literasi. Dalam melakukan pengumpulan data ini, peneliti menggunakan platform *Google Scholar*. Pengumpulan data kedua mencari data sekunder, data sekunder di dapatkan dari platform *Yahoo Finance* dengan nama dataset BBNI.JK. Terdapat sekitar 2287 data yang digunakan dalam melakukan penelitian yang diambil hampir 10 tahun dari 25 april 2013 sampai 12 juli 2022. Pengumpulan data ketiga membangun *machine learning* dengan algoritma KNN *regression* serta menguji akurasi dari algoritma tersebut. Nantinya data yang dihasilkan bersifat primer yang merupakan hasil perhitungan yang dilakukan.

2.2 Metode K-Nearest Neighbors Regression

Metode penelitian merupakan langkah – langkah yang dilakukan untuk mencari data dan informasi yang nantinya data dan informasi tersebut akan diolah dan dianalisis secara ilmiah. Penelitian yang dilakukan pada kasus ini menggunakan metode penelitian kuantitatif. Setelah melakukan pencarian sumber literasi, selanjutnya dilakukan pencarian dataset yang berisi histori harga saham PT Bank Negara Indonesia untuk digunakan dalam melakukan prediksi menggunakan *machine learning*. Nantinya *machine learning* yang sudah dirancang akan melakukan proses prediksi harga saham PT Bank Negara Indonesia menggunakan algoritma KNN *regression* dengan dataset yang sudah diperoleh. Berikut cara membangun *machine learning* yang akan dirancang pada gambar 1.



Gambar 1. Rancangan Algoritma K nearest neighbors regression

Algoritma KNN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru. Jauh dekatnya data tetangga dihitung berdasarkan rumus 1 *Euclidean* berikut ini:

$$euc = \sqrt{\sum_{k=1}^d (a_k - b_k)^2} \quad (1)$$

Keterangan :

a = data latih

b = data uji

k = variable data

d = dimensi data

Cara kerja algoritma KNN diperlukan penentuan data latih, data uji, dan nilai k . Selanjutnya, data latih diurutkan berdasarkan hitungan jarak terdekat antara data uji dan data latih. Dan terakhir, diambil rata-rata data latih terkecil sesuai jumlah k untuk menentukan kelas regresi. Berikut proses alur perhitungan menggunakan metode *K-Nearest Neighbors* [9].

- Menentukan Parameter k
- Menghitung jarak antara data yang akan dievaluasi dengan semua pelatihan
- Mengurutkan jarak yang terbentuk
- Menentukan jarak yang terdekat sampai nilai k
- Memasangkan Kelas yang sesuai
- Mencari jumlah kelas yang terdekat dan menetapkan kelas tersebut untuk dievaluasi.

2.3 Root Mean Squared Error (RMSE)

RMSE merupakan metode yang digunakan untuk melakukan pengukuran dengan mengukur perbedaan nilai yang diambil dari prediksi sebuah model dalam mengestimasi nilai yang diobservasi. Keakuratan dalam estimasi sebuah nilai dibuktikan dengan adanya hasil nilai RMSE yang kecil. Semakin kecil nilai RMSE pada suatu estimasi dapat dikatakan akurat dibandingkan dengan hasil nilai RMSE yang lebih [10]. Penelitian dalam memprediksi data dengan menggunakan *Machine Learning* seperti ini sangat membutuhkan parameter RMSE untuk menilai akurasi suatu model. Berikut merupakan rumus 2.2 dari metode RMSE:

$$RMSE = \sqrt{\frac{1}{n} \left(\sum_{j=1}^n (y_j - \hat{y}_j)^2 \right)} \quad (2)$$

Keterangan :

y = data uji yang sebenarnya

\hat{y} = data hasil prediksi

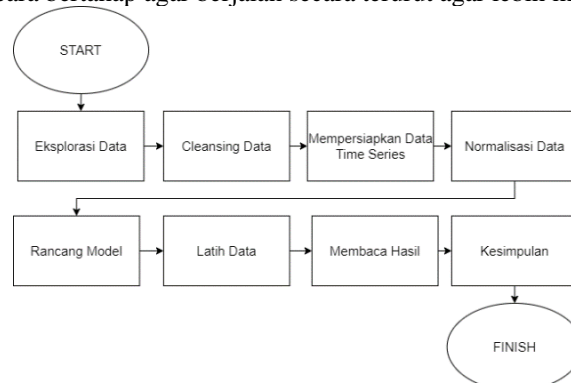
n = jumlah data uji

2.4 Tahapan Pengujian

Perancangan pengujian dilakukan untuk memberi gambaran terkait model yang akan dirancang. Rancangan pengujian dilakukan dengan membuat *flowchart* yang digunakan sebagai gambaran alur dari sebuah algoritma yang telah ditentukan.

- Rancangan Machine Learning

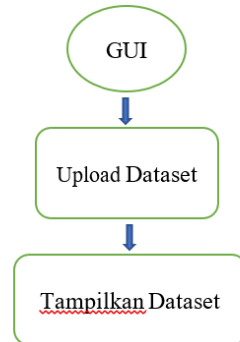
Rancangan yang dilakukan *machine learning* dalam melakukan prediksi harga saham Bank Negara Indonesia dilakukan secara bertahap agar berjalan secara teratur agar lebih mekanis.



Gambar 2. Rancangan *Machine Learning*

b. Rancangan Program Interface (GUI)

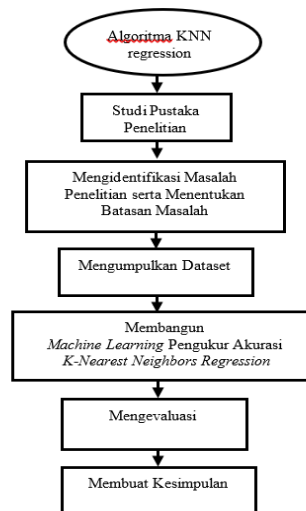
Rancangan yang dilakukan program *interface* (GUI) dalam melakukan prediksi harga saham Bank Negara Indonesia dilakukan untuk menampilkan dataset yang akan dimasukkan pada machine learning.



Gambar 3 Rancangan Algoritma Program *Interface* (GUI)

3. HASIL DAN PEMBAHASAN

Penelitian dapat berjalan dengan baik dengan adanya langkah – langkah yang dirancang secara sistematis dan berurutan. Langkah – langkah tersebut nantinya digunakan sebagai gambaran untuk melakukan penelitian. Terdapat langkah – langkah penelitian yang digambarkan pada gambar 4.



Gambar 4. Langkah - Langkah Penelitian

3.1 Dataset

Data yang digunakan dalam penelitian ini diambil dari situs Yahoo Finance dimana data tersebut merupakan histori data harga saham Bank Negara Indonesia dalam bentuk data harian. Data harian tersebut menjelaskan harga saham yang terjadi pada hari – hari sebelumnya. Sebelumnya, dari data yang diperoleh dalam penelitian ini terdapat beberapa fitur yang menjelaskan kategori harga saham seperti *Date*, *Open*, *High*, *Low*, *Close*, *Adj Close*, dan *Volume*.

3.2 Machine Learning

Dalam penelitian ini dilakukan perancangan sebuah model *machine learning* untuk memprediksi harga sebuah saham di hari berikutnya. Dimana *machine learning* yang dirancang melakukan proses regresi data. Metode yang digunakan dalam merancang *machine learning* ini termasuk ke dalam metode *supervised learning* karena dataset dalam memprediksi harga saham memiliki label tetap yang sudah dilakukan perubahan data sebelumnya menjadi data *time series*.

a. Hasil dataset

Awal dari melakukan persiapan data yaitu dengan mencari *dataset* menggunakan platform Yahoo Finance. Pada platform tersebut didapatkan *dataset* saham PT Bank Negara Indonesia Tbk dengan nama BBNI.JK yang

dimana data tersebut aktif berjalan memberikan informasi mengenai harga saham yang sedang terjadi dari hari – ke hari. Pada tabel 1 terdapat *dataset* yang digunakan dalam penelitian.

Tabel 1 Dataset Awal

Date	Open	High	Low	Close	Adj Close	Volume
2013-05-02	5550	5550	5350	5350	4064,941895	22442500
2013-05-03	5450	5450	5200	5300	4026,95166	21802000
2013-05-06	5300	5350	5200	5300	4026,95166	17589000
2013-05-07	5350	5450	5300	5400	4102,932617	27186500
2013-05-08	5450	5500	5300	5350	4064,941895	21303000
2013-05-10	5400	5450	5350	5400	4102,932617	32339000
2013-05-13	5450	5450	5150	5250	3988,962402	26239500

Selanjutnya *dataset* tersebut diolah terlebih dahulu melalui Microsoft Excel untuk dilakukan perubahan data ke dalam *time series*. Karena kegiatan dilakukan untuk bisa melakukan *time series* pemrograman prediksi saham. Pada tabel 2, terdapat *dataset* yang sudah diolah ke dalam *time series*.

Tabel 2 Dataset Time Series

Date	x(n-5)	x(n-4)	x(n-3)	x(n-2)	x(n-1)	x(n)	x(n+1)
2013-05-02	5250	5350	5300	5400	5550	5350	5300
2013-05-03	5350	5300	5400	5550	5350	5300	5300
2013-05-06	5300	5400	5550	5350	5300	5300	5400
2013-05-07	5400	5550	5350	5300	5300	5400	5350
2013-05-08	5550	5350	5300	5300	5400	5350	5400
2013-05-10	5350	5300	5300	5400	5350	5400	5250

b. Cleansing Data

Pada Tabel 3, terdapat tahap perancangan *machine learning* yang dilakukan dalam penelitian ini terdapat suatu kesalahan yang menyebabkan terjadinya kendala pada *machine learning*. Kendala yang terjadi yaitu dengan adanya kesalahan dalam membaca *dataset* sehingga ada beberapa *dataset* yang tidak terbaca oleh *machine learning* sehingga data bernilai NaN. Maka dari itu perlu dilakukannya proses *cleansing* data untuk memperbaiki data yang tidak terbaca.

Tabel 3 Cleansing Data

	x(n-5)	x(n-4)	x(n-3)	x(n-2)	x(n-1)	x(n)	x(n+1)
count	2274.000000	2274.000000	2274.000000	2274.000000	2274.000000	2274.000000	2274.000000
mean	6326.752419	6327.752858	6328.654354	6329.566843	6330.347405	6331.029024	6331.875550
std	1587.077946	1587.050450	1586.993880	1586.919176	1586.742156	1586.589982	1586.509501
min	3160.000000	3160.000000	3160.000000	3160.000000	3160.000000	3160.000000	3160.000000
25%	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000
50%	6050.000000	6050.000000	6050.000000	6050.000000	6062.500000	6075.000000	6075.000000
75%	7500.000000	7500.000000	7500.000000	7518.750000	7525.000000	7525.000000	7525.000000
max	10175.000000	10175.000000	10175.000000	10175.000000	10175.000000	10175.000000	10175.000000

c. Normalisasi Data

Pada tabel 4, terdapat data perancangan *machine learning* dalam membangun model prediksi terdapat kendala yang terjadi pada *dataset* yang digunakan. Hal tersebut perlu dilakukannya normalisasi data untuk memperbaiki *dataset* yang memiliki nilai kosong atau null.

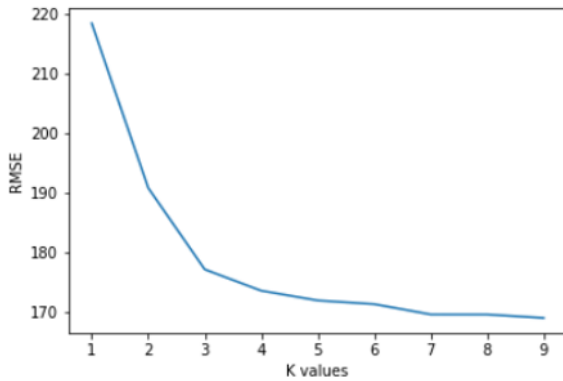
Tabel 4 Normalisasi Data

	x(n-5)	x(n-4)	x(n-3)	x(n-2)	x(n-1)	x(n)
0	5250.0	5250.0	5300.0	5350.0	5300.0	5200.0
1	5250.0	5300.0	5350.0	5300.0	5200.0	5250.0
2	5300.0	5350.0	5300.0	5200.0	5250.0	5350.0
3	5350.0	5300.0	5200.0	5250.0	5350.0	5300.0
4	5300.0	5200.0	5250.0	5350.0	5300.0	5400.0
...
1536	8675.0	8575.0	8600.0	8650.0	8500.0	8450.0
1544	8800.0	8850.0	8875.0	8925.0	8850.0	8900.0
1545	8850.0	8875.0	8925.0	8850.0	8900.0	9200.0
1546	8875.0	8925.0	8850.0	8900.0	9200.0	9375.0
1547	8925.0	8850.0	8900.0	9200.0	9375.0	9350.0

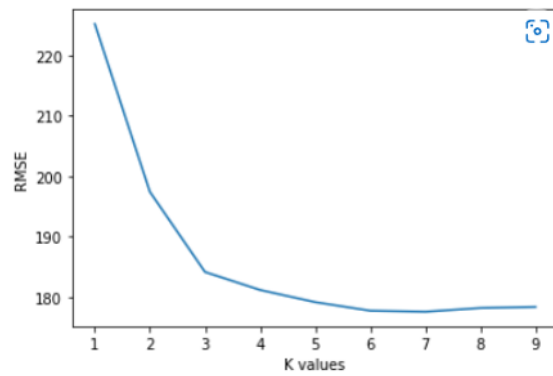
3.3 Pengujian

a. Pemodelan dengan Algoritma K-Nearest Neighbor Regression

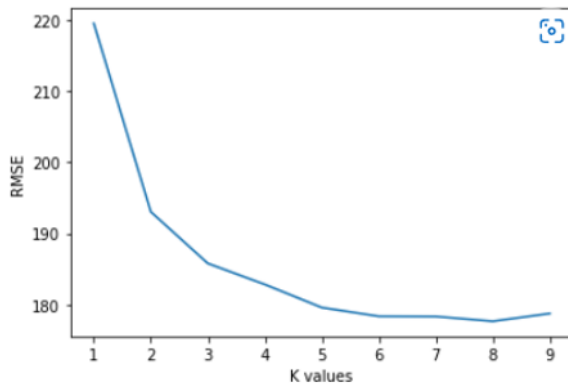
Langkah pertama yang dilakukan untuk pemodelan algoritma KNN regression yaitu dengan menentukan nilai K – optimal. Pencarian nilai K dilakukan dengan menghitung jarak antar Euclidean atau jarak antar tetangga menggunakan metode Elbow. Berikut hasil pencarian K – optimal dalam satu sampai empat indikator menggunakan metode Elbow.



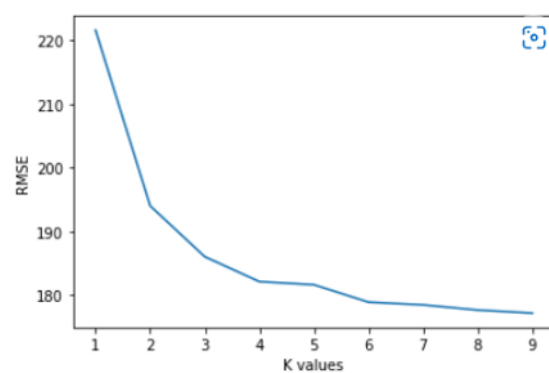
Gambar 5 Metode Elbow (K - Optimal) satu indikator



Gambar 6 Metode Elbow (K - Optimal) dua indikator



Gambar 7 Metode Elbow (K - Optimal) tiga indikator



Gambar 8 Metode Elbow (K - Optimal) empat indikator

Berdasarkan pada gambar 5, 6, 7 dan 8 grafik Metode Elbow (K – Optimal), jumlah nilai K yang memiliki nilai *mean squared error* (MSE) paling rendah merupakan nilai K yang paling optimal. Model menjadi lebih baik dengan adanya tingkat akurasi yang tinggi dan akurat apabila menggunakan nilai K – optimal. Setelah mendapatkan nilai K – optimal, selanjutnya dilakukan pemodelan KNN untuk mencari nilai akurasi dan tingkat *error* dengan menggunakan *root mean squared error* (RMSE) dengan hasil sebagai berikut.

Table 5. Hasil Pemodelan KNN

	1 Indikator	2 Indikator	3 Indikator	4 Indikator
K – Optimal	7	6	6	6
RMSE	168	177	177	177
Akurasi	99.17%	99.14%	99.14%	99.07%

Berdasarkan pada tabel 5, nilai K dengan jumlah indikator yang berbeda memiliki nilai yang sama yaitu berada pada angka 8 untuk dijadikan K – optimal. Tingkat RMSE yang dihasilkan dari ke empat dengan jumlah indikator yang berbeda ini menghasilkan nilai terendah dengan menggunakan satu indikator sebesar 168, dan nilai RMSE dengan dua, tiga dan empat indikator menghasilkan nilai yang sama yaitu 177. Dimana nilai RMSE merupakan jarak selisih rata – rata *error* antara data asli dengan data prediksi dibuat dalam bentuk desimal agar mudah untuk dianalisa dalam melihat jarak *error* tersebut. Hasil dari akurasi KNN terbesar di dapatkan dengan menggunakan satu indikator sebesar 99.14%, dan hasil menggunakan dua, tiga dan empat indikator masing – masing sebesar 99.14%, 99.14% dan 99.07%.

3.4 Perbandingan Rasio

Dalam melakukan prediksi dengan membangun model *machine learning* ini dilakukan beberapa tahap untuk proses pengujian *dataset*. Pada Tabel 6, terdapat data dibagi ke dalam beberapa rasio untuk menguji tingkat *error*.

Tabel 6. Perbandingan RMSE Evaluasi Data pada KNN

Rasio	1 Indikator	2 Indikator	3 Indikator	4 Indikator
70 : 30	168	177	177	177
80 : 20	182	194	190	192
90 : 10	230	239	236	237

3.5 Hasil Prediksi

Selanjutnya pada penelitian ini dilakukan prediksi harga saham PT Bank Negara Indonesia Tbk selama empat hari kedepan dengan menggunakan pemodelan *machine learning* dengan menggunakan algoritma KNN. Berikut merupakan hasil yang telah dilakukan pada penelitian ini dalam memprediksi harga saham PT Bank Negara Indonesia dalam beberapa hari kedepan.

Tabel 7. Hasil Prediksi Harga Saham

Tanggal	Prediksi			
	1 Indikator	2 Indikator	3 Indikator	4 Indikator
13 Juli 2022	7446	7391	7406	7375
14 Juli 2022	7435	7413	7525	7580
15 Juli 2022	7414	7458	7600	7569
18 Juli 2022	7464	7460	7625	7605

Pada Tabel 7, terdapat alur hasil prediksi harga saham yang dilakukan dalam lima hari kedepan. Prediksi dilakukan dengan menggunakan *dataset* harian dengan indikator yang berbeda – beda dimana pemodelan *machine learning* melakukan prediksi harga saham perhari. Prediksi harga saham selama lima hari ini dilakukan menggunakan data prediksi yang sudah dihasilkan pada satu hari untuk digunakan dalam memprediksi hari – hari berikutnya. Setiap prediksi harga pada masing – masing hari memiliki tingkat *error* selisih atau nilai RMSE berbeda – beda.

3.6 Tampilan Layar

Dalam penerapannya, penelitian ini dituangkan ke dalam bentuk aplikasi *web* dan *machine learning* menggunakan metode *Whitebox*. berikut beberapa tampilan layar dari aplikasi yang dibuat.

a. Tampilan Layar Machine Learning

Dalam pembahasan *Source Code*, dilakukan pemrograman dalam membangun *machine learning* untuk pemodelan prediksi harga menggunakan algoritma KNN. Kode program dibuat sesuai dengan kebutuhan yang diperlukan dalam membangun pemodelan *machine learning* pada penelitian ini. Selanjutnya kode program diuji dengan melihat hasil yang dikeluarkan sudah sesuai dengan spesifikasi yang dibutuhkan atau belum dengan menggunakan metode *Whitebox*. Pada Tabel 8, terdapat alur pengujian yang dilakukan.

Tabel 8. Source Code Program Machine Learning

No	Kode Program	Hasil Yang Diharapkan	Hasil Pengujian
1	<pre>1 from google.colab import drive 2 drive.mount('/content/drive')</pre>	Menghubungkan <i>machine learning</i> dengan Google Drive yang menyimpan data	Valid
2	<pre>1 import pandas as pd 2 import numpy as np 3 from sklearn.neighbors import KNeighborsRegressor as KNN_Reg 4 from sklearn.metrics import r2_score 5 from sklearn.metrics import mean_squared_error as mse 6 from matplotlib import pyplot as plt 7 from sklearn.model_selection import train_test_split 8 from sklearn.preprocessing import StandardScaler 9 import time</pre>	Melakukan <i>import library</i> yang akan digunakan	Valid
3	<pre>1 dataset = pd.read_excel('/content/drive/MyDrive/BBNI.JK.xlsx', "Sheet1")</pre>	Memanggil <i>dataset</i> dan memasukkannya	Valid
4	<pre>#Cleansing Data dataset = dataset.dropna(axis=0) dataset.describe()</pre>	Melakukan <i>Cleansing</i> data untuk menghapus data yang tidak terbaca	Valid

No	Kode Program	Hasil Yang Diharapkan	Hasil Pengujian
5	<pre> 1 #split dataset 2 X1 = dataset.iloc[:, 1: 7] 3 y1 = dataset.iloc[:, 7] 4 5 X = X1.to_numpy() 6 y = y1.to_numpy() 7 8 X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=1, test_size=0.3) </pre>	Melakukan Split data menjadi dua variable x dan y pada 1 indikator	Valid
6	<pre> 1 #Normalisasi 2 sc_X = StandardScaler() 3 4 X_train = sc_X.fit_transform(X_train) 5 X_test = sc_X.fit_transform(X_test) 6 </pre>	Melakukan Normalisasi data untuk membersihkan data yang bernilai nol	Valid
7	<pre> 1 # Vectorized approach to find the 2 start = time.process_time() 3 # We are setting a range of K values and calculating the RMSE for each of them. 4 k_list = [x for x in range(1,10)] 5 # Calculating the distance matrix using numpy broadcasting technique 6 distance = np.sqrt(((X_train[:, :, None] - X_test[:, :, None])** 2).sum(1)) 7 #Sorting each data points of the distance matrix to reduce computational effort 8 sorted_distance = np.argsort(distance, axis = 0) 9 #The knn function takes in the sorted distance and returns the RMSE of the 10 def knn(X_train,X_test,y_train,y_test,sorted_distance,k): 11 y_pred = np.zeros(y_test.shape) 12 for row in range(len(X_test)): 13 #Transforming the y_train values to adjust the scale. 14 y_pred[row] = y_train[sorted_distance[:,row][:k]].mean() * sigma_y + mu_y 15 RMSE = np.sqrt(np.mean((y_test - y_pred)**2)) 16 return RMSE 17 def knn1(X_train,xX_test,y_train,xy_test,distance,k): 18 xy_pred = np.zeros(xy_test.shape) 19 for row in range(len(xX_test)): 20 xy_pred[row] = y_train[sorted_distance[:,row][:k]].mean() * sigma_y + mu_y 21 hasil = xy_pred[row] 22 return hasil 23 #Storing the RMSE values in a list for each k value 24 rmse_list = [] 25 for i in k_list: 26 rmse_list.append(knn(X_train,X_test,y_train,y_test,sorted_distance,i)) 27 #Finding the optimal K value 28 min_rmse_k_value = k_list[rmse_list.index(min(rmse_list))] 29 #Finding the lowest possible RMSE 30 optimal_RMSE = knn(X_train,X_test,y_train,y_test,sorted_distance,min_rmse_k_value) 31 optimal_RMSE 32 </pre>	Memasukan algoritma KNN Regression kedalam machine learning	Valid
8	<pre> def Elbow(K): test_mse =[] for i in K: model = KNN_Reg(n_neighbors=i) model.fit(X_train,y_train) tmp = model.predict(X_test) tmp = mse(tmp,y_test) test_mse.append(tmp) return test_mse K = range(1,10) test = Elbow(K) plt.plot(K,test) plt.xlabel('K Neighbors') plt.ylabel('Mean Squared Error(MSE)') plt.title('Elbow Curve for Test') </pre>	Menentukan nilai K – optimal pada KNN menggunakan metode Elbow	Valid
9	<pre> #define model KNN model_knn= KNN_Reg(n_neighbors = 5) # training the model KNN model_knn.fit(X_train, y_train) #Akurasi acc1 = model_knn.score(X_test,y_test) # test for prediction test_predict_knn = model_knn.predict(X_test) #RMSE rmse_knn = sqrt(mean_squared_error(test_predict_knn,y_test)) print('RMSE K-NN: ', rmse_knn) print("Accuracy KNN: ", acc1*100,'%') </pre>	Melakukan pemodelan KNN untuk menentukan akurasi dan RMSE	Valid
10	<pre> 1 #Prediksi 1 Hari Selanjutnya 2 xa =np.array([[7750,7650,7600,7600,7500,7450]]) 3 xa 4 5 #Scaling Data 6 Xa = sc_X.transform(xa) 7 8 #Pemodelan Prediksi 9 ya_KNN = model_knn.predict(Xa) 10 11 12 #Hasil Prediksi 13 print ('Prediksi 13 Juli 2022') 14 print ('KNN : ', ya_KNN) 15 print ('Harga Aktual 13 Juli 2022') 16 print ('-') </pre>	Melakukan pemodelan prediksi harga dengan KNN untuk hari pertama menggunakan: 1. indikator	Valid

4. KESIMPULAN

Berdasarkan batasan masalah dan analisis yang dilakukan pada prediksi saham menggunakan algoritma *k nearest neighbors regression* dan mencari solusi untuk mengatasi masalah tersebut, maka dapat disimpulkan Algoritma KNN *regression* dengan seluruh variasi jumlah indikator yang berbeda memiliki tingkat *error* RMSE dengan nilai paling rendah yaitu pada rasio perbandingan 70:30. Dari hasil tersebut maka diambil perbandingan data dengan rasio 70:30 dengan nilai RMSE paling terendah pada seluruh variasi indikator. Hasil pemodelan menggunakan KNN *regression* dengan empat variasi indikator yang berbeda, prediksi dengan menggunakan satu indikator memiliki hasil yang paling baik dengan nilai RMSE 169 dengan akurasi 98.8%. Hasil prediksi dengan KNN memiliki selisih nilai sebesar 40 – 100 pada setiap variasi indikator. Saran yang diberikan untuk penelitian selanjutnya untuk dapat dikembangkan adalah dapat mengganti atau menambahkan indikator fitur yang dapat digunakan seperti *adj close*, dan *volume*. Dapat menambahkan beberapa algoritma machine learning lainnya seperti *Decision Tree*, *Line Regression*, *Support Vector Machine* dan lainnya sebagai perbandingan dalam melakukan prediksi. Dapat menggunakan dan menambahkan dataset yang akan digunakan machine learning dalam mempelajari sehingga dapat mendapatkan hasil yang lebih baik.

DAFTAR PUSTAKA

- [1] Algonz, R.D.B. (2022). *Machine Learning: Pengertian, Cara Kerja, Dan 3 Metodenya!*. Diakses Pada 20 April 2022. <https://www.ekrut.com/media/apa-itu-machine-learning>
- [2] A. Izzah And R. Widyastuti, "Prediksi Harga Saham Menggunakan Improved Multiple Linear Regression Untuk Pencegahan Data Outlier," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, Vol. 2, No. 3, Pp. 141–150, 2017, Doi: 10.22219/Kinetik.V2i3.268.
- [3] Seruni, D. S., Furqon, M. T., & Wihandika, R. C. (2020). Sistem Prediksi Pertumbuhan Jumlah Penduduk Kota Malang Menggunakan Metode K-Nearest Neighbor Regression. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 4(4), 1075–1082.
- [4] Tanuwijaya, J., & Hansun, S. (2019). LQ45 Stock Index Prediction Using K-Nearest Neighbors Regression. *International Journal Of Recent Technology And Engineering*, 8(3), 2388–2391.
- [5] Tauran, E. (2019). Berdasarkan Data Dari Bursa Efek Indonesia Prediction Of Stock Price Of Pt Bank Central Asia Tbk Based On Data From Indonesia Stock Exchange Using K- Nearest Neighbors (Knn) Method. *Jurnal Teika*, 11(2), 123–129.
- [6] Fatah, H., & Subekti, A. (2018). Prediksi Harga Cryptocurrency Dengan Metode K-Nearest Neighbours. *Jurnal Pilar Nusa Mandiri*, 14(2), 137.
- [7] F. Sains And U. M. Sidoarjo, "Implementasi Data Mining Pada Pelanggan Telkom Menggunakan Metode K-Nearest Neighbor Untuk Memprediksi Status Pelayanan," Pp. 115–119, 2022.
- [8] D. W. Farkhina Dwi Utari, Amril Mutoi Siregar *Et Al.*, "Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Student Graduation Prediction Using K-Nearest Neighbor (K-Nn)," *E-Proceeding Eng.*, Vol. 1, No. 1, Pp. 30–36, 2019.
- [9] Cahya, (2018). Contoh Implementasi Data Mining Algoritma K-Nearest Neighbors (K-NN) Menggunakan PHP Dan Mysql Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu. Diakses Pada 21 April 2022. <https://cahyadsn.phpindonesia.id/extra/knn.php#preproses>.
- [10] Khoiri, (2020). *Pengertian Dan Cara Menghitung Root Mean Square Error (RMSE)*. Diakses Pada 21 April 2022. <https://www.khoiri.com/2020/12/cara-menghitung-root-mean-square-error-rmse.html>