

Prediksi Saham PT. Aneka Tambang Tbk. dengan K-Nearest Neighbors

¹ Lathifah Alfat, ²Hendi Hermawan, ³Aviana Z. Rustandiputri, ⁴Rizky M. Y. Inzhagi, ⁵Refi Tandjilal

^{1,2,3,4,5} Universitas Pembangunan Jaya, Tangerang Selatan, Indonesia

¹lathifah.alfat@upj.ac.id; ²hendi.hermawan@upj.ac.id;

Article Info

Article history:

Received, 2022-10-20

Revised, 2022-11-09

Accepted, 2022-11-22

Kata Kunci:

prediksi saham
KNN
K-Nearest Neighbors
penambangan data
Aneka Tambang

Keywords:

stock prediction
KNN
K-Nearest Neighbors
data mining
Aneka Tambang

ABSTRAK

Teknologi informasi yang semakin berkembang membuat data yang dihasilkan turut tumbuh menjadi *big data*. Data tersebut dapat dimanfaatkan dengan disimpan, dikumpulkan, dan ditambang sehingga menghasilkan informasi dan pengetahuan yang bernilai. Penambangan data merupakan proses menerapkan algoritma *machine learning* untuk mengambil informasi dari data. Prediksi merupakan bentuk penambangan data untuk mendapatkan informasi di masa depan dari data sebelumnya. Algoritma K-Nearest Neighbors termasuk algoritma yang handal untuk prediksi. KNN menentukan nilai K, dan menghitung jarak data di sekitarnya, sehingga hasil prediksi lebih akurat. Prediksi saham PT Aneka Tambang dibutuhkan investor untuk mengurangi resiko finansial. Penelitian ini memodelkan prediksi saham PT Aneka Tambang menggunakan algoritma K-Nearest Neighbors. Data yang diolah berasal dari sumber terbuka dari tahun 2018 sampai 2021. Hasil prediksi membuktikan bahwa model saham PT Aneka Tambang optimal diperoleh pada perbandingan data *training:testing sebesar 9:1*, dengan nilai akurasi sebesar 93,36%. Maka dari itu, pemodelan saham PT Aneka Tambang dengan KNN dapat diterapkan untuk memprediksi nilainya di masa depan.

ABSTRACT

The development of information technology makes the resulting data also grow into big data. The data could be utilized by being stored, warehoused, and mined to produce valuable information and knowledge. Data mining is the process of applying machine learning algorithms to extract information from data. Prediction is a form of data mining to obtain future information from previous data. The K-Nearest Neighbors algorithm is a reliable algorithm for predictions. KNN determines the value of K, and calculates the distance of the surrounding data, so that the prediction results are more accurate. PT Aneka Tambang's stock prediction is needed by investors to reduce financial risk. This study models PT Aneka Tambang's stock predictions using the K-Nearest Neighbors algorithm. The processed data comes from open sources from 2018 to 2021. The prediction results prove that the optimal PT Aneka Tambang stock model is obtained at a training: testing data comparison of 9:1, with an accuracy value of 93.36%. Therefore, PT Aneka Tambang's stock modeling with KNN can be applied to predict its value in the future.

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



Penulis Korespondensi:

Lathifah Alfat,
Program Studi Informatika,
Universitas Pembangunan Jaya,
Email: lathifah.alfat@upj.ac.id

1. PENDAHULUAN

Perkembangan teknologi informasi pasca revolusi industri 4.0 menghasilkan data yang begitu besar. Data tersebut dapat disimpan, dikumpulkan, dan ditambang, sehingga menghasilkan informasi dan pengetahuan. Informasi dari data memberikan manfaat kemudahan berbagai aspek kehidupan manusia. *Data mining* (penambangan data) merupakan gabungan dari basis data, kecerdasan buatan, *machine learning*, ekstraksi informasi, dan statistika. *Data mining* menganalisa kumpulan data untuk memperoleh pola tertentu. Pada

prakteknya, *data mining* dapat melakukan beberapa tugas penting, seperti deskripsi, estimasi, prediksi, klasifikasi, pengklasteran, dan asosiasi [1].

Pasar saham termasuk lingkungan stokastik dan menantang. Pada kenyataannya, pergerakan saham melibatkan pengaruh berbagai faktor tertentu [2][3]. Data yang melimpah dihasilkan dari pergerakan saham di seluruh dunia. Peningkatan data ini meliputi volume, kecepatan, keragaman, dan kebenaran [2][4]. *Big data* saham dapat dianalisa di sisi fundamental dan teknikal. Analisa fundamental terpusat pada tren lokal dan internasional, keuangan, dan sentimen publik. Sedangkan analisa teknikal mempertimbangkan pergerakan saham di masa lalu atau data historis [2].

Berbagai faktor telah mempengaruhi pergerakan pasar saham di Indonesia. PT. Aneka Tambang Tbk. merupakan salah satu perusahaan yang terdaftar pada Bursa Efek Indonesia [5]. Perusahaan dengan kode ANTM ini membagi kepemilikan sahamnya sebesar 65% untuk pemerintah Indonesia dan 35% untuk masyarakat [6]. Fluktuasi saham PT Aneka Tambang merupakan kondisi yang rumit. Nilai saham dapat berubah naik turun akibat dari faktor internal maupun eksternal perusahaan [7]. Hal ini mengakibatkan perkiraan atau prediksi harga saham membutuhkan perhitungan yang kompleks [8]. Melihat permasalahan tersebut, perlu dilakukan pemodelan untuk memprediksi fluktuasi harga saham. Dengan prediksi yang akurat, investor dapat mempersiapkan strategi ketika menghadapi kenaikan dan penurunan harga saham PT. Aneka Tambang.

Algoritma K-Nearest Neighbors (KNN) merupakan algoritma yang umum digunakan untuk prediksi. Beberapa penelitian telah menerapkan algoritma KNN untuk memprediksi suatu kondisi di masa depan. Penelitian pertama [9] memanfaatkan algoritma KNN dan *linear regression* untuk membangun model analisis. Hasil penelitian menunjukkan performa model *linear regression* lebih baik daripada KNN pada data set tersebut [9]. Penelitian kedua [10] melakukan prediksi pasar saham menggunakan *machine learning* dengan algoritma Decision Trees (DT), Support Vector Machine (SVM) dan Neural Network (NN). Data yang diolah dari emiten Ghana Stock Exchange (GSE), Johannesburg Stock Exchange (JSE), Bombay Stock Exchange (BSE-SENSEX) dan New York Stock Exchange (NYSE). Pengujian RMSE terendah dihasilkan oleh teknik *stacking* pada nilai 0.0001–0.001 [10]. Penelitian ketiga [11] berupaya mengurangi resiko tren dengan memprediksi menggunakan algoritma *machine learning* dan *deep learning*. Algoritma yang diuji dalam penelitian adalah DT, Random Forest, Adaboost, XGBoost, Support Vector Classifier, Naïve Bayes, KNN, Logistic Regression dan Artificial Neural Network dan metode Deep learning Recurrent Neural Network (RNN) dan Long short-term memory (LSTM). Hasilnya, RNN dan LSTM dinyatakan sebagai algoritma yang unggul performanya [11]. Penelitian keempat [12] mengimplementasikan teknik penambangan data di bidang pendidikan dengan algoritma KNN and Naïve Bayes. Data yang digunakan adalah dari sekolah menengah dari kementerian pendidikan di Jalur Gaza, tahun 2015. Hasil eksperimen menunjukkan hasil bahwa algoritma Naïve Bayes lebih baik dari KNN dengan nilai akurasi tertinggi sebesar 93.6% [12]. Penelitian kelima [13] meningkatkan kecerdasan bisnis dengan prediksi penjualan. Penelitian menganalisa data penjualan Rossmann menggunakan model prediktif *linear regression* dan KNN *regression*. Penelitian menggunakan RMSE dan MAPE. Hasil menunjukkan metode yang lebih sesuai untuk prediksi penjualan [13].

Penelitian ini dirancang dengan tujuan memberikan informasi tentang penerapan KNN pada model pergerakan harga saham. Prediksi yang dilakukan menggunakan data *time series* tanpa mempertimbangkan aspek sosial, politik, dan ekonomi. Selain itu, hasil penelitian berkontribusi untuk menambah literatur mengenai penerapan algoritma KNN pada data sejenis. Penelitian ini berupaya untuk memberikan sumbangsih dalam ilmu pengetahuan mengenai penerapan algoritma KNN. Penelitian juga bertujuan untuk mengurangi kerugian investor ketika mengambil keputusan terkait saham PT. Aneka Tambang.

2. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini dibagi dalam beberapa tahap. Pertama, identifikasi masalah dilanjutkan pengumpulan data dari sumber yang berakses terbuka. Setelah itu, data masuk ke tahap *pre-processing* agar data layak dan siap untuk tahap pengolahan. Kemudian, pengolahan data dilakukan dengan menggunakan *library* dari bahasa pemrograman Python. Data selanjutnya dibagi menjadi data *testing* dan data *training*. Data ini kemudian masuk ke tahap prediksi dengan menerapkan algoritma KNN. Terakhir, model prediksi masuk ke tahap evaluasi [14]. *Flowchart* yang memuat metode penelitian digambarkan pada Gambar 1.

2.1 Identifikasi Masalah

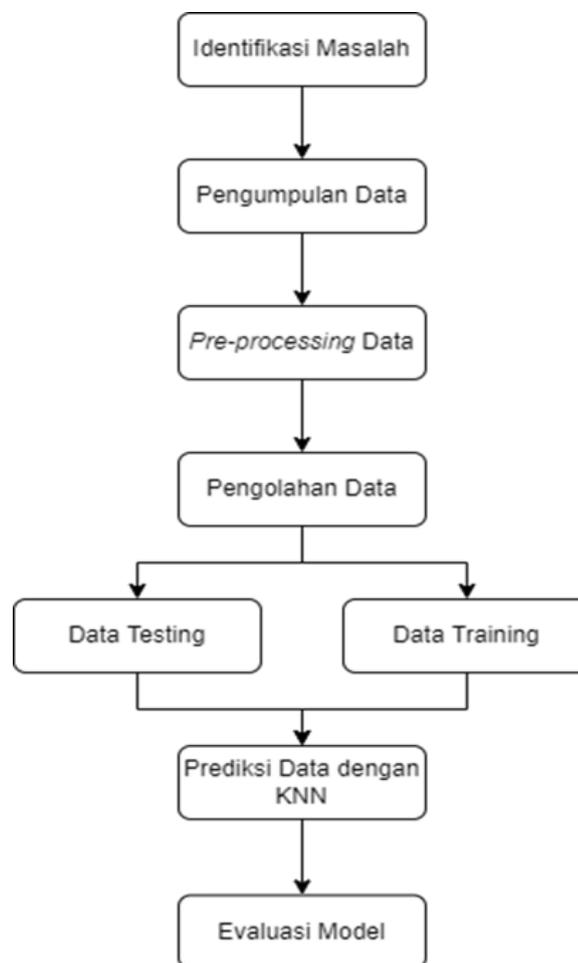
Berdasarkan sumber jurnal [9][10][11] diidentifikasi beberapa masalah yang mendasari penelitian mengenai prediksi saham. Masalah yang tersebut di antaranya adalah investor mengalami kesulitan dalam menyalurkan dana ke suatu emiten saham dan masyarakat membutuhkan model prediksi yang efektif dan

efisien. Masalah lain adalah perlunya suatu program untuk melihat resiko naik turun saham akibat tren di masyarakat.

Salah satu saham dengan emiten terbitan BUMN di Indonesia adalah ANTM, milik PT Aneka Tambang. Saham PT. Aneka Tambang termasuk saham dengan kinerja yang baik dan menunjang portofolio optimal. Saham ini menempati indeks IDX-BUMN20, merupakan indeks saham dengan kategori yang efisien dimana nilai *return* yang dihasilkan lebih besar daripada ekpektasi [15]. Maka dari itu, model prediksi kinerja saham PT Aneka Tambang menjadi hal dibutuhkan oleh masyarakat.

2.2 Pengumpulan Data

Data yang digunakan untuk penelitian bersumber dari situs Yahoo Finance [16]. Data harga saham yang dikumpulkan berasal dari data tahun 2018 hingga 2021. Data yang diperoleh berjumlah 747 baris dan 7 kolom. Atribut data di antara tujuh kolom tersebut adalah Date, Open, High, Close, Adj Close, dan Volume. Date merupakan tanggal, Open merupakan harga ketika perdagangan dibuka, High adalah harga tertinggi selama perdagangan saham, Close adalah harga ketika perdagangan ditutup. Volume adalah jumlah transaksi yang dilakukan dalam satu hari. Data tersebut disimpan dengan format .csv. Setelah tahap pengumpulan data, tahapan berikutnya adalah melakukan *pre-processing* data.



Gambar 1. Flowchart Penelitian

2.3 Pre-processing Data

Pemodelan prediksi memanfaatkan *machine learning* dengan menerapkan bahasa pemrograman Python. Python mempunyai *library* atau perpustakaan yang berisi modul dan paket yang mempunyai fungsi untuk memudahkan penulisan kode program. Pustaka yang digunakan adalah sebagai berikut:

1. Pandas; adalah *tool* untuk menganalisa dan mengubah data dari sumber terbuka yang cepat, tangguh, fleksibel dan mudah digunakan [17]. Pandas dalam penelitian digunakan untuk memuat atau mengolah file dalam format .csv.

2. Numpy; adalah proyek *open source* yang bertujuan untuk memungkinkan komputasi numerik dengan Python. Numpy dibuat pada tahun 2005, dibangun di awal penggunaan pustaka Numeric dan Numarray [18]. Penelitian memanfaatkan Numpy untuk mengelola *array*, operasi matriks dan vektor.
3. Matplotlib; adalah pustaka komprehensif untuk membuat visualisasi gambar, animasi, dan interaksi dengan Python [19]. Dalam penelitian ini, Matplotlib digunakan untuk membuat visualisasi data seperti grafik.
4. Scikit-learn; adalah pustaka yang bersumber terbuka yang mendukung *supervised* dan *unsupervised learning*. Scikit-learn juga menyediakan beragam *tools* untuk memuat model, pemrosesan data, seleksi model, evaluasi model, dan utilitas lainnya [20]. Pustaka Scikit-learn dimanfaatkan dalam penelitian untuk pemodelan algoritma.

Jumlah data yang dikumpulkan adalah 747 baris dan tujuh kolom. Data tersebut dimuat pada Google Colaboratory. Tahap *pre-processing* menghilangkan data yang tidak lengkap menggunakan fungsi “drop”. Karena sebaris data kurang lengkap, maka 746 baris data masuk ke tahap pengolahan data.

2.4 Pengolahan Data

Pengolahan data dimulai dengan menentukan fitur. Fitur yang dipilih adalah tiga atribut; Open, High, Low. Fitur kemudian dimasukan ke variabel *x* sebagai masukan. Pada penentuan keluaran atau *y*, dipilih data pada atribut Close. Tahapan menuju ke pembagian data menjadi data *testing* dan data *training*. Persentase data *testing* dan *training* menentukan model yang sesuai dengan algoritma dan besarnya akurasi model. Pembagian data optimal diperoleh pada presentase data *training* : data *testing* sebesar 90:10.

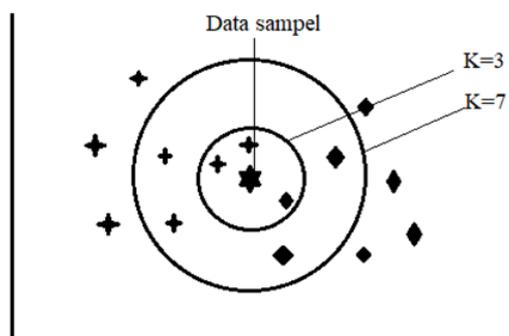
2.5 Prediksi Data dengan KNN

Penelitian menerapkan algoritma KNN sebagai metode pemodelan. Dua properti biasanya disarankan untuk KNN, karena tidak ada asumsi mengenai distribusi data dengan KNN. Penelitian dilanjutkan dengan memilih nilai *K* dan menentukan fungsi jarak yang akan digunakan. Kemudian, memilih sampel pengujian dan menghitung jaraknya dari sampel data *training* *n*. Selanjutnya, mengurutkan jarak yang diperoleh dan mengambil sampel data *K* terdekat[11]. Akhirnya, menetapkan kelas *testing* ke sampel pada bagian mayoritas dari *K* terdekat. Gambar 2 merupakan ilustrasi metode KNN [11].

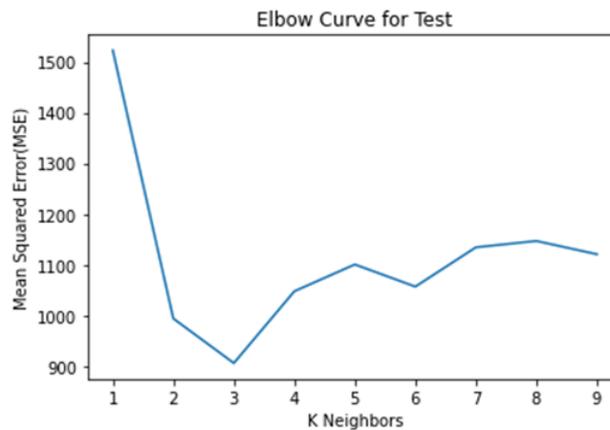
Jauh dekatnya jarak data *K* terdekat dihitung berdasarkan rumus Euclidean seuai persamaan (1) berikut [21]:

$$d_{euclidean(x,y)} = \sqrt{\left(\sum_i^n (x_i - y_i)^2\right)} \tag{1}$$

Pada persamaan (1), *x*₁ adalah data latih, *y*₁ adalah data uji, *i* adalah variabel data, dan *n* adalah dimensi data. Nilai *K* terdekat ditentukan oleh jarak dengan data lainnya.



Gambar 2. Ilustrasi KNN



Gambar 3. Kurva Elbow Pengujian

Pencarian K-optimal menggunakan metode Elbow. Hasil dari pencarian K-optimal ditunjukkan pada grafik Gambar 3. Berdasarkan Gambar 3, diperoleh bahwa nilai tiga merupakan K yang mempunyai nilai *Mean Square Error* (MSE) terkecil. Maka dari itu, untuk melakukan pemodelan, diisikan nilai K senilai tiga. Jika nilai MSE pada suatu K kecil, semakin baik tingkat akurasi suatu model. Akhirnya, pemodelan dijalankan dengan fungsi “model_knn”. Fungsi diiringi dengan perintah “fit” untuk melakukan *training model*. Kemudian, Python menjalankan perintah “predict” untuk melakukan pengujian atau tes terhadap model.

2.6 Evaluasi Model

Dalam melakukan prediksi data, dibutuhkan parameter evaluasi untuk memeriksa akurasi model. Penelitian menggunakan parameter penilaian *Root Mean Squared Error* atau RMSE. Kalkulasi RMSE melibatkan akar kuadrat dari rata-rata perbedaan kuadrat antara prediksi dan observasi aktual [22]. Semakin kecil hasil RMSE, maka akan semakin kecil tingkat *error* yang dihasilkan. Rumus perhitungan parameter RMSE [22] ditunjukkan pada persamaan (2) berikut:

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

Pada persamaan (2), y adalah data testing sebenarnya, \hat{y} adalah data hasil prediksi, dan n adalah jumlah data testing.

Selain itu, digunakan parameter R Square (R^2) untuk mengukur koefisien determinasi yang sesuai dari persamaan regresi. R^2 diartikan sebagai pembagian variansi pada variabel dependen yang mampu diprediksi dari variabel independen. Nilai R^2 antara nol hingga satu, berarti kecocokan model dinyatakan lebih baik jika nilainya semakin mendekati angka satu. Parameter R^2 dituliskan pada persamaan (4) berikut [23]:

$$\bar{Y} = \frac{1}{m} \sum_{i=1}^m Y_i$$
(3)

$$R^2 = 1 - \left(\sum_{i=1}^m (X_i - Y_i)^2 \right) / \left(\sum_{i=1}^m (\bar{Y}_i - Y_i)^2 \right)$$
(4)

Pada persamaan (4), X_i adalah nilai prediksi ke i . Pada persamaan (3) dan (4), Y_i adalah nilai aktual ke i .

3. HASIL DAN ANALISIS

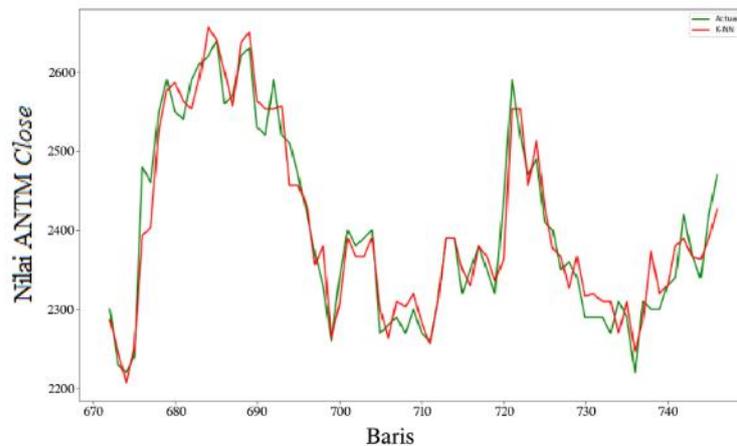
Dari penelitian yang telah dilaksanakan, diperoleh beberapa temuan hasil. Hasil dari pemodelan harga saham dengan menerapkan algoritma KNN ditunjukkan pada Tabel 1.

Tabel 1. Hasil Pemodelan Saham PT. Aneka Tambang

Nilai K	RMSE	R2	Akurasi
3	30,29	0,931	93,36%

Berdasarkan Tabel 1, pemodelan KNN mencapai tingkat akurasi sebesar 93,36%. Performa dari model termasuk dalam kategori baik berdasarkan hasil akurasi yang melebihi nilai 80%. Nilai R2 sebesar 0,931 atau mendekati satu. Berdasarkan analisa jika semakin tinggi nilai R2, menunjukkan fitur yang dipilih mempunyai keterkaitan yang tinggi. Nilai RMSE sebesar 30,29 menunjukkan model memiliki nilai kesalahan yang kecil.

Grafik untuk mempermudah pembacaan terhadap perbandingan harga aktual dan harga hasil prediksi KNN. Grafik dibuat dengan fungsi “plot” dari library matplotlib. Data test yang digunakan dalam pengujian merupakan indeks data 672-746. Hasil dari grafik tersebut ditunjukkan pada Gambar 4. Pergerakan harga dari saham dari tahun 2018 s.d 2021 ditunjukkan pada garis warna hijau. Hasil prediksi harga menggunakan KNN ditunjukkan dengan garis warna merah pada Gambar 4.



Gambar 4. Grafik Data Prediksi vs Aktual

Kemudian pengujian dilakukan oleh peneliti untuk melihat hasil prediksi harga saham menggunakan data test. Hasil pengujian diperoleh harga prediksi saham saat “Close”. Perbandingan dapat dilakukan dengan harga aktual dari data test yang digunakan. Hasil pengujian ditunjukkan pada Tabel 2.

Tabel 2. Pengujian Harga Saham PT. Aneka Tambang

Baris	Open	High	Low	Close (Aktual)	Close (Prediksi)
672	2240.0	2310.0	2230.0	2300.0	2286.6
673	2280.0	2310.0	2200.0	2230.0	2246.6
674	2230.0	2250.0	2190.0	2220.0	2206.6
675	2230.0	2260.0	2220.0	2240.0	2250.0
676	2250.0	2510.0	2240.0	2480.0	2393.3
677	2450.0	2470.0	2380.0	2460.0	2403.3
678	2460.0	2610.0	2460.0	2550.0	2526.6
679	2560.0	2630.0	2530.0	2590.0	2576.6
680	2610.0	2670.0	2520.0	2550.0	2586.6
681	2560.0	2610.0	2500.0	2540.0	2563.3

4. KESIMPULAN

PT. Aneka Tambang Tbk merupakan perusahaan BUMN yang memiliki tren nilai saham yang baik. Kenaikan nilai saham pada perusahaan ini menjadikan ketertarikan investor untuk memiliki saham dari perusahaan. Namun, nilai saham yang bersifat fluktuatif menyebabkan para investor mengalami kesulitan dalam menganalisis nilai saham yang berjalan apakah akan selalu naik atau turun. Dengan adanya pemodelan yang dirancang oleh peneliti, maka nilai saham pada perusahaan dapat dijadikan informasi untuk membuat sistem prediksi dengan mudah dan akurat. Pembuatan model harga prediksi saham PT Aneka Tambang yang telah dirancang menggunakan perbandingan data *training*:data *testing* sebesar 9:1. Model prediksi menerapkan algoritma KNN dapat dikatakan baik karena menghasilkan nilai akurasi sebesar 93,36%. Maka, tingkat akurasi

nilai saham secara aktual dibandingkan hasil prediksi KNN tidak jauh berbeda. Sehingga kenaikan dan turunnya nilai saham dapat diprediksi dengan baik. Nantinya, pemodelan ini bisa dimanfaatkan oleh pengembang aplikasi untuk dapat membuat sistem prediksi nilai saham perusahaan dan dapat dijadikan alat untuk menganalisis nilai saham di masa depan.

REFERENSI

- [1] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, Feb. 2017.
- [2] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions," *Artif. Intell. Rev.*, vol. 53, no. 4, pp. 3007–3057, Apr. 2020.
- [3] C. Bousono-Calzon, J. Bustarviejo-Munoz, P. Aceituno-Aceituno, and J. J. Escudero-Garzas, "On the economic significance of stock market prediction and the no free lunch theorem," *IEEE Access*, vol. 7, pp. 75177–75188, 2019.
- [4] K. O. Nti, A. Adekoya, and B. Weyori, "Random forest based feature selection of macroeconomic variables for stock market prediction," *SSRN Electron. J.*, 2019.
- [5] K. I. Lestari, *Analisis Kinerja Keuangan Berdasarkan Metode Du Pont Pada PT. Aneka Tambang (Persero) Tbk*. 2016.
- [6] S. Pareda, C. E. Mongi, and C. E. J. C. Montolalu, "Sistem Pendukung Keputusan Pemilihan Karyawan Teladan di PT. Aneka Tambang (ANTAM) Tbk Unit Bisnis Pertambangan Buli Menggunakan Metode Simple Additive weight (SAW)," *dC*, vol. 8, no. 1, p. 1, Jan. 2019.
- [7] S. O. Nellawati and Y. Isbanah, "Analisis Pengaruh Faktor Internal Dan Faktor Eksternal Terhadap Pergerakan Indeks Harga Saham Gabungan (Ihsg) Di Bursa Efek Indonesia Periode 2012 -- 2016," *J. Ilmu Manaj.*, vol. 7, no. 1, pp. 113–123, 2012.
- [8] B. Jange, "Prediksi Harga Saham Bank BCA Menggunakan Prophet," *J. Trends Econ. Account. Res.*, vol. 2, no. 1, pp. 1–5, 2021.
- [9] S. P. Poornima, C. N. Priyanka, P. Reshma, and K. N. Suraj Kr Jaiswal, "Stock Price Prediction Using {KNN} and Linear Regression," *Blue Eyes Intell. Eng. Sci. Publ.*, vol. 8, pp. 142–145, 2019.
- [10] I. Nti, A. F. Adekoya, and B. A. Weyori, "A comprehensive evaluation of ensemble learning for stock market prediction," *J. Big Data*, vol. 7, no. 20, pp. 1–40, 2020.
- [11] M. Nabipour, P. Nayyeri, H. Jabani, Shahab, and A. Mosavi, "Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; A comparative analysis," *IEEE Access*, vol. 8, pp. 150199–150212, 2020.
- [12] I. A. Abu Amra and A. Y. A. Maghari, "Students performance prediction using KNN and Naive Bayesian," in *2017 8th International Conference on Information Technology (ICIT)*, May 2017.
- [13] S. Kohli, G. T. Godwin, and S. Urolagin, "Sales prediction using linear and KNN regression," in *Algorithms for Intelligent Systems*, Singapore: Springer Singapore, 2021, pp. 321–329.
- [14] W. R. U. Fadilah, D. Agfiannisa, and Y. Azhar, "Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode Support Vector Machine," *Fountain Informatics J.*, vol. 5, no. 2, pp. 45–51, 2020.
- [15] A. D. Prasetya, "Analisis Pembentukan Portofolio Optimal Dengan Menggunakan Capital Asset Pricing Model (CAPM) (Studi Empiris pada Saham yang Masuk dalam Indeks IDX-BUMN20 di Bursa Efek Indonesia)," in *Seminar Nasional Manajemen, Ekonomi dan Akuntansi Fakultas Ekonomi*, pp. 1269–1278, 2021.
- [16] "PT Aneka Tambang Tbk (ANTM.JK) Stock Historical Prices Data," *Yahoo Finance*. <https://finance.yahoo.com/quote/ANTM.JK/history?period1=1539820800&period2=1634515200&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true> (accessed Oct. 10, 2022).
- [17] "Pandas - Python Data Analysis Library," *Pandas*. <https://pandas.pydata.org/> (accessed Oct. 10, 2022).
- [18] "Numpy - About Us," *Numpy*. <https://numpy.org/about/> (accessed Oct. 10, 2022).
- [19] "Matplotlib - visualization with python," *Matplotlib*. <https://matplotlib.org/> (accessed Oct. 10, 2022).
- [20] "Getting Started," *Scikit Learn*. https://scikit-learn.org/stable/getting_started.html (accessed Oct. 10,

2022).

- [21] S. Mulyati and S. M. Husein, *Rancang Bangun Aplikasi Data Mining Prediksi Kelulusan Ujian Nasional Menggunakan Algoritma (KNN) K-Nearest Neighbor Dengan Metode Euclidean Distance Pada Smpn 2 Pagedangan''*. Januari, 2020.
- [22] E. Sutoyo and A. Almaarif, "Educational Data Mining for Predicting Student Graduation Using the Naïve Bayes Classifier Algorithm," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 1, pp. 95–101, Feb. 2020, doi: 10.29207/resti.v4i1.1502.
- [23] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, no. e623, p. e623, Jul. 2021.