

PREDIKSI HARGA SAHAM PT BANK CENTRAL ASIA TBK BERDASARKAN DATA DARI BURSA EFEK INDONESIA MENGUNAKAN METODE *K-NEAREST NEIGHBORS* (KNN)

Edward Ridley Tauran

Hope Channel Indonesia, Jakarta, Indonesia
e-mail: ridleyedward7@gmail.com

Abstrak

Prediksi saham menggunakan metode *K-Nearest Neighbors* dimaksudkan agar dapat memberikan prediksi yang dapat membantu masyarakat dan investor untuk mengetahui harga saham di waktu ke depan. Melalui tahap studi literatur, wawancara dan melihat data harga saham harian dimana atribut yang digunakan adalah harga buka saham harian, harga tertinggi, harga terendah, harga tutup saham harian, data kurs jual, kurs beli rupiah terhadap mata uang dolar Amerika, serta data suku bunga 1 bulan. Adapun hasil yang diperoleh dengan jumlah keseluruhan data adalah 1415 data, dengan total *data training* sebanyak 70% dan data testing sebanyak 30% menggunakan metode *K-Nearest Neighbors*, maka diperoleh tingkat akurasi sebanyak 61.79%. Melalui *confusion matrix* ditunjukkan nilai yang diberikan untuk *precision* atau tingkat ketepatan informasi yang diharapkan untuk data yang diklasifikasikan mengalami kenaikan adalah 62.03% dan data yang mengalami penurunan harga adalah 60.76%, sedangkan untuk nilai *recall* atau tingkat keberhasilan terhadap informasi yang ditemukan untuk data yang diklasifikasikan kenaikan harga adalah 87.35% dan data yang diklasifikasikan mengalami penurunan harga adalah sebesar 26.82%.

Kata Kunci: Prediksi harga saham, Kurs jual, Kurs beli, Suku bunga, *K-Nearest Neighbors*

PREDICTION OF STOCK PRICE OF PT BANK CENTRAL ASIA TBK BASED ON DATA FROM INDONESIA STOCK EXCHANGE USING K- NEAREST NEIGHBORS (KNN) METHOD

Abstract

Stock prediction using the K-Nearest Neighbors method is intended to be able to provide predictions that can help the public and investors to find out the share price in the future. Through the literature study stage, interviewing and looking at daily stock price data where the attributes used are daily stock open price, highest price, lowest price, daily stock closing price, selling exchange rate data, rupiah exchange rate against the US dollar, and interest rate data 1 month. The results obtained with the total data are 1415 data, with a total of 70% training data and 30% testing data using the K-Nearest Neighbors method, then the accuracy rate of 61.79% is obtained. Through the confusion matrix the value given for the precision or level of accuracy of the expected information for data classified as experiencing an increase is 62.03% and data experiencing a price reduction is 60.76%, while for the recall value or the success rate for information found for data classified as an increase the price was 87.35% and the data classified as having experienced a price decline was 26.82%.

Keywords: *Stock price prediction, selling rate, buying rate, interest rate, K-Nearest Neighbors*

1. Pendahuluan

Dunia investasi telah berkembang cukup marak di Indonesia. Saat ini investasi saham sudah menjadi bagian terpenting bagi perkembangan keuangan di sebuah perusahaan, masyarakat maupun investor [1]. Tak hanya investor Indonesia, investor yang berasal dari luar negeri melakukan investasi dan trading saham di Indonesia. Saham merupakan surat-surat berharga sebagai tanda kepemilikan atas suatu perusahaan.

Pasar modal bukan saja merupakan sarana yang digunakan oleh masyarakat atau para investor untuk melakukan investasi namun juga dapat mempermudah masyarakat dan para investor untuk mengetahui perkembangan tentang harga saham setiap perusahaan di Indonesia yang sudah *go public*, salah satunya adalah PT Bank Central Asia Tbk yang mana perusahaan ini Melalui beragam produk dan layanan yang berkualitas dan tepat sasaran, solusi finansial BCA mendukung perencanaan keuangan pribadi dan perkembangan nasabah bisnis. Didukung oleh kekuatan jaringan antar cabang, luasnya jaringan ATM, serta jaringan perbankan elektronik lainnya, siapa saja dapat menikmati kemudahan dan kenyamanan bertransaksi yang ditawarkan BCA. Masyarakat dan investor yang ingin mengetahuinya harga saham PT Bank Central Asia Tbk setiap harinya dapat melihatnya di bursa efek Indonesia dengan mengakses *website* <https://www.idx.co.id>.

Keterbukaan data harga saham yang dimiliki oleh PT Bank Central Asia Tbk pada pasar modal mengakibatkan masyarakat dan para investor berpeluang untuk memperoleh keuntungan. Keuntungan yang dimaksud adalah berupa Dividen (pembagian keuntungan yang diberikan perusahaan dan berasal dari keuntungan yang dihasilkan perusahaan – berupa uang tunai) dan Capital Gain (Capital Gain merupakan selisih antara harga beli dan harga jual). Namun, waktu yang tepat untuk melakukan penjualan dan pembelian saham yang diinginkan PT Bank Central Asia Tbk haruslah diketahui oleh masyarakat dan investor agar dapat memperoleh keuntungan. Harga saham PT Bank Central Asia Tbk yang terdapat pada pasar modal Indonesia atau BEI (Bursa Efek Indonesia) selalu menampilkan pergerakan harga saham yang selalu mengalami perubahan kenaikan dan penurunan di waktu yang tidak diketahui oleh masyarakat dan pihak investor. Oleh karena itu, rendah dan tingginya harga saham dapat diketahui dengan melihat harga saham baik dari harga tertinggi, harga terendah, harga penutup saham kemarin, frekuensi harga saham, dan sebagainya.

Dilansir dari laman resmi OJK (Otoritas Jasa Keuangan), terdapat faktor-faktor yang mempengaruhi naik dan turunnya harga saham, yang mana terdapat dua faktor utama yang mempengaruhi kondisi harga saham yaitu Faktor Eksternal dan Faktor Internal. Adanya banyak faktor yang ada, maka dalam penelitian ini, penulis melakukan riset terhadap naik dan turunnya harga saham dengan menambahkan atribut yang dimana mengacu pada faktor eksternal yaitu ekonomi makro dalam hal ini suku bunga serta kurs rupiah terhadap mata uang Amerika. Dengan dapat melakukan prediksi harga saham pada PT Bank Central Asia Tbk maka hal ini dapat memberikan keuntungan kepada para investor yang ingin melakukan penjualan dan pembelian saham di PT Bank Central Asia Tbk [2].

Analisis harga saham pada PT Bank Central Asia Tbk yang terdapat di Bursa Efek Indonesia dilakukan dengan 2 (dua) pendekatan [3], yaitu:

1. Analisis Fundamental: yang pada dasarnya adalah dengan melakukan analisis historis atau keuangan atau yang biasa disebut analisis perusahaan.
2. Analisis Teknikal: Analisis teknikal pada dasarnya dilakukan untuk mempelajari data-data saham di masa lampau yang terdapat di pasar modal Indonesia. Analisis teknikal lebih berfokus pada usaha-usaha untuk memperkirakan suatu harga saham berdasarkan data-data dan informasi yang masuk secara perlahan dalam harga saham yang terdapat di BEI.

Dalam melakukan penelitian ini, penulis memilih untuk menggunakan metode *K-Nearest Neighbors* dikarenakan metode ini memiliki beberapa kelebihan yaitu bahwa dia tangguh terhadap *training* data yang *noise* (data yang hilang atau berganda) dan efektif apabila *data training*-nya banyak serta adanya kesempatan untuk penggunaan metode *K-Nearest Neighbors* dikarenakan banyak penelitian terdahulu

yang belum banyak menggunakan metode ini. Adapun tujuan dari penelitian ini adalah bagaimana metode *K-Nearest Neighbors* dapat memprediksi kapan terjadinya kenaikan dan turun saham.

2. Metode Penelitian

Data mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [4].

Data mining adalah proses menemukan pola yang menarik, dan pengetahuan dari data yang berjumlah besar [5]. Algoritma *K-Nearest Neighbors* (KNN) merupakan sebuah metode untuk melakukan klasifikasi terhadap obyek baru berdasarkan (K) tetangga terdekatnya. KNN termasuk algoritma *supervised learning*, yang mana hasil dari query instance baru, diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Kelas yang paling banyak muncul, yang akan menjadi kelas hasil klasifikasi [6]

Ada beberapa cara untuk mencari tetangga terdekat yaitu *Euclidean distance* dan *manhattan distance* (*city block distance*). Dan yang sering digunakan adalah *euclidean distance* [7], yaitu:

$$d(X, Y) = \sqrt{\sum_i^n (a_i - b_i)^2 + (a_1 - b_1)^2, \dots ((a_n - b_n)^2} \quad (1)$$

Dimana $a = a_1, a_2, \dots, a_n$ dan $b = b_1, b_2, \dots, b_n$ mewakili n nilai atribut dari dua *record*. Adapun langkah-langkah untuk menghitung algoritma *K-Nearest Neighbors* antara lain :

1. Menentukan Parameter K (jumlah tetangga terdekat).
2. Menghitung kuadrat jarak *eucliden* (*queri instance*) masing-masing objek terhadap data sampel yang diberikan.
3. Kemudian mengurutkan jarak tersebut ke dalam kelompok yang mempunyai Euclid terkecil (mengurutkan hasil No 2 secara *ascending*).
4. Mengumpulkan kategori Y (Klasifikasi *Nearest Neighbor*) berdasarkan nilai K atau ambil data tetangga terdekat.
5. Dengan menggunakan kategori *Nearest Neighbor* yang paling mayoritas maka akan menghasilkan kelas data baru (prediksi). Nilai k yang terbaik untuk algoritma ini tergantung pada data. Secara umum, nilai k yang tinggi akan mengurangi efek *noise* pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi semakin kabur. Nilai k yang bagus dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan *Simple Unweighted Voting*.

Metode Akurasi Klasifikator

Confusion matrix adalah alat yang berguna untuk menganalisis seberapa baik classifier mengenali *tuple* dari kelas yang berbeda. TP dan TN memberikan informasi ketika *classifier* benar, sedangkan FP dan FN memberitahu ketika *classifier* salah [8].

		Predicted class		Total
		yes	no	
Actual class	yes	TP	FN	P
	no	FP	TN	N
Total		P'	N'	P + N

Gambar 1 *Confusion Matrix*

Keterangan :

TP : *True positives*, merupakan jumlah data dengan kelas positif yang diklasifikasikan positif.

TN : *True negatives*, merupakan jumlah data dengan kelas negatif yang diklasifikasikan negatif.

FP : *False positives*, merupakan jumlah data dengan kelas positif diklasifikasikan negatif.

FN : *False negatives*, merupakan jumlah data dengan kelas negatif diklasifikasikan positif.

Dikutip dari [9] terdapat 5 indikator pengukuran *performance* dari klasifikator yang dipakai, yaitu:

A. Accuracy

Merupakan rasio prediksi Benar (positif dan negatif) dengan keseluruhan data.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (2)$$

B. Precision

Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif.

$$Precision = \frac{(TP)}{(TP+FP)} \quad (3)$$

C. Recall (Sensitifitas)

Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif.

$$Recall = \frac{(TP)}{(TP+FN)} \quad (4)$$

D. Specificity

Merupakan kebenaran memprediksi negatif dibandingkan dengan keseluruhan data negatif.

$$Specificity = \frac{(TN)}{(TN+FP)} \quad (5)$$

E. F1 Score

F1 Score merupakan perbandingan rata-rata presisi dan *recall* yang dibobotkan.

$$F1\ Score = \frac{2(Recall*Precision)}{(Recall+Precision)} \quad (6)$$

RapidMiner

RapidMiner adalah perangkat lunak *open source* untuk *knowledge discovery* dan *data mining*. RapidMiner memiliki kurang lebih 400 prosedur (operator) *data mining*, termasuk operator untuk masukan, *output*, data preprocessing [10].

Beberapa fitur dari RapidMiner: Berlisensi gratis (*open source*)

1. Multiplatform karena diprogram dalam Bahasa Java.
2. Internal data berbasis XML sehingga memudahkan pertukaran data eksperimen.
3. Dilengkapi dengan *scripting language* untuk otomatisasi eksperimen.
4. Memiliki GUI (*Graphical User Interface*), command line mode (*batch mode*), dan Java API yang dapat dipanggil dari program lain.
5. Dapat dikembangkan dengan menambahkan *plugin* dan *extension*.
6. Fasilitas *plotting* untuk visualisasi dan multidimensi dan model.

3. Hasil

Berdasarkan hasil analisis yang telah dilakukan terhadap harga saham harian PT Bank Central Asia Tbk. beserta dengan suku bunga 1 bulan, kurs jual, dan kurs beli melalui data yang telah penulis kumpulkan dan dibantu melalui *software* RapidMiner, dengan melakukan pengujian terhadap 56 kali uji coba untuk nilai *k*, maka diambil 3 nilai *k* dengan performa yang baik.

1. Untuk nilai *k* = 37

Dari hasil pengujian yang dilakukan didapatkan akurasi sebesar 61.79% dan kesalahan klasifikasi sebesar 38.21%.

2. Untuk nilai *k* = 11 dan *k* = 15

Dari hasil pengujian yang dilakukan didapatkan akurasi sebesar 61.79% dan kesalahan klasifikasi sebesar 38.21%.

3. Untuk nilai *k* = 12 dan *k* = 17

Dari hasil pengujian yang dilakukan didapatkan akurasi sebesar 61.79% dan kesalahan klasifikasi sebesar 38.21%.

Setelah mengetahui hasil yang diperoleh dengan membandingkan nilai *k* serta melihat 3 nilai *k* yang memberikan nilai performa yang baik, maka diperoleh bahwa nilai *k* = 37 memberikan performa yang baik, yang dapat dilihat melalui tabel di bawah ini:

accuracy: 61.79%
classification_error: 38.21%

Tabel 1 Hasil Akurasi dan Error Klasifikasi pada K-Nearest Neighbors.

	true ya	true tidak	class precision
pred. ya	214	131	62.03%
pred. tidak	31	48	60.76%
class recall	87.35%	26.82%	

Melalui hasil yang diperoleh dengan bantuan *software* RapidMiner dengan jumlah keseluruhan data adalah 1415 data, dengan total *data training* sebanyak 70% dan data *testing* sebanyak 30% menggunakan metode *K-Nearest Neighbors*, maka diperoleh tingkat akurasi sebanyak 56.37%.

4. Pembahasan/Kesimpulan

Berdasarkan penelitian yang telah dilakukan penulis terhadap harga saham harian dari PT Bank Central Asia Tbk., kurs jual, dan kurs beli rupiah terhadap mata uang dollar amerika, dan suku bunga 1 bulan yang telah dikumpulkan dari tanggal 1 Januari 2014 sampai 31 Oktober 2019, maka dapat disimpulkan bahwa dengan metode *K-Nearest Neighbors* didapat hasil persentase akurasi sebesar 61.79%.

Melalui data yang terkumpul sebanyak 1415 jumlah data yang telah diklasifikasikan mengalami kenaikan harga saham, dan penurunan harga saham harian, dengan keterangan data yang diklasifikasikan ya sebanyak 817 data, dan data yang diklasifikasikan tidak sebanyak 598 data. Dengan keterangan 214 data yang diklasifikasikan mengalami kenaikan harga saham, dan terdapat 131 data yang seharusnya mengalami kenaikan harga saham, namun diklasifikasikan menjadi penurunan harga. Dan terdapat 48 data yang telah diklasifikasikan mengalami penurunan harga saham, namun terdapat 31 data yang seharusnya mengalami penurunan namun diklasifikasikan mengalami kenaikan harga.

Nilai yang diberikan untuk *precision* atau tingkat ketepatan informasi yang diharapkan penulis dengan jawaban yang diberikan oleh *software* RapidMiner untuk data yang diklasifikasikan mengalami kenaikan adalah 62.03% dan data yang mengalami penurunan harga adalah 60.76%, sedangkan untuk nilai *recall* atau tingkat keberhasilan terhadap informasi yang ditemukan oleh *software* RapidMiner untuk data yang diklasifikasikan kenaikan harga adalah 87.35% dan data yang diklasifikasikan mengalami penurunan harga adalah sebesar 26.82%.

5. Referensi

- [1] S. Andriani, "MINAT INVESTASI SAHAM PADA MAHASISWA," *Jurnal Ekonomi Dan Bisnis Indonesia*, vol. 4, no. 1, pp. 37-44, 2019.
- [2] V. C. Nugroho, E. Hulu dan G. S. Ugut, "FAKTOR YANG MEMPENGARUHI VOLUME TRANSAKSI PADA HARGA SAHAM PT. BANK CENTRAL ASIA TBK.," *Jurnal Manajemen Terapan dan Keuangan (Mankeu)*, vol. 10, no. 1, pp. 99-109, 2021.
- [3] H. V. Y. Sitorus, "ANALISIS HARGA SAHAM PT ASTRA INTERNASIONAL Tbk MENGGUNAKAN DATA BURSA EFEK INDONESIA DALAM JANGKA WAKTU PENDEK MENGGUNAKAN METODE NAIVE BAYES DAN DECISION TREE-J48," 2017.
- [4] E. Turban dan dkk., *Decision Support System and Intelligent System*, Yogyakarta: ANDI, 2005.
- [5] J. Han dan M. Kimber, *Data mining: Concepts and Techniques Second Edition.*, Morgan Kaufmann Publisher, 2006.
- [6] F. Gorunescu, *Data mining Concepts, Models and Techniques*, New York: Springer Verlag, 2011.
- [7] K. dan E. T. Lutfi, *Algoritma Data mining*, Yogyakarta: ANDI, 2009.
- [8] J. Han dan J. Pei, *Data mining Concepts and Techniques Third Edition*, Waltham: Elsevier Inc., 2011.

- [9] R. Arthana, "Mengenal Accuracy, Precision, *Recall* dan Specificity serta yang diprioritaskan dalam *Machine learning*," 18 November 2019. [Online]. Available: <https://medium.com/@rey1024/mengenal-accuracy-precision-recall-dan-specificity-serta-yang-diprioritaskan-b79ff4d77de8>.
- [10] F. Sulianta dan D. Juju, RapidMiner (YALE), PT. Elex Media Komputindo, 2010.