

BAB I

PENDAHULUAN

1.1 Latar Belakang

K-Nearest Neighbors (KNN) adalah salah satu algoritma yang paling sederhana dan mudah dipahami dalam *Machine Learning*, yang digunakan baik untuk masalah klasifikasi maupun regresi. Prinsip dasarnya adalah bahwa suatu titik data akan diklasifikasikan atau diprediksi berdasarkan mayoritas label atau nilai dari K titik data terdekat di sekitarnya. Dalam konteks klasifikasi, KNN memprediksi label kelas dari suatu data berdasarkan mayoritas dari tetangga terdekatnya, sementara dalam regresi, ia memprediksi nilai variabel kontinu dari suatu data (Zhang 2021).

KNN merupakan non parametrik, sebagai algoritma non-parametrik, KNN tidak membuat asumsi tertentu tentang distribusi data atau bentuk fungsi yang mendasari data. Sebaliknya, algoritma ini secara langsung mempelajari struktur data dari dataset pelatihan tanpa menghasilkan parameter-parameter yang harus diperkirakan terlebih dahulu. Berbeda dengan algoritma parametrik seperti regresi linier atau regresi logistik yang memperkirakan parameter-parameter (misalnya, koefisien dalam regresi linier) berdasarkan pada fungsi model yang telah ditentukan sebelumnya, KNN mengandalkan data pelatihan untuk membuat prediksi (Viswanatha et al. 2023).

Kekuatan dari pendekatan non-parametrik ini adalah kemampuannya untuk menangani data dengan struktur yang kompleks atau tidak teratur tanpa memerlukan asumsi tentang distribusi data. Dalam pembelajaran mesin tidak terlepas dari beberapa tantangan. Salah satunya adalah sensitivitas terhadap skala fitur, di mana perbedaan skala antar fitur dapat menghasilkan hasil prediksi yang tidak konsisten. Selain itu, pemilihan nilai K yang optimal juga menjadi pertimbangan krusial dalam penggunaan KNN, karena nilai K yang terlalu kecil dapat mengakibatkan model yang terlalu sensitif terhadap *noise*, sementara nilai K yang terlalu besar dapat membuat model menjadi terlalu umum. Selain itu,

kinerja KNN cenderung menurun ketika jumlah dimensi fitur meningkat. Tantangan lainnya termasuk keterbatasan dalam menangani data yang hilang atau tidak lengkap, serta keterbatasan dalam interpretasi hasil prediksi, terutama dalam konteks nilai K yang besar.

K-Nearest Neighbors (KNN), sebuah metode dalam pembelajaran mesin yang memanfaatkan tetangga terdekat, telah menjadi subjek banyak penelitian sebelumnya. Pada penelitian Julius Tanuwijaya dan Seng Hansun yang berjudul *LQ45 Stock Index Prediction using k-Nearest Neighbors Regression*, penelitian ini memprediksi harga saham pada indeks LQ45 di sektor properti, yaitu ADHI, BKSL, BSDE, LPKR, PTPP, WIKA, dan WSKT. Dalam penelitian ini, nilai K optimal dicari dari 2 hingga 10. Hasil penelitian menunjukkan bahwa RMSE pada ADHI dengan nilai K optimal 2 adalah 246.1999, untuk BKSL dengan $K = 2$ memiliki RMSE sebesar 23.98962, untuk BSDE dengan $K = 2$ memiliki RMSE sebesar 279.70222, untuk LPKR dengan $K = 2$ memiliki RMSE sebesar 130.3787, dan untuk WSKT dengan $K = 1$ memiliki RMSE sebesar 131.06629 (Tanuwijaya and Hansun 2019).

Penelitian (Bansal et al. 2022) dengan judul *Stock market prediction with high accuracy using machine learning techniques* menggunakan algoritma *K-Nearest Neighbor*, *Linear Regression*, *Support Vector Regression*, *Decision Tree Regression* dan *Long Sort-Term Memory* dalam memprediksi harga saham dalam 12 perusahaan di pasar saham India. Data yang digunakan adalah data dari Januari 2015 sampai April 2021 dengan atribut data set yaitu *previous close price*, *opening price*, *an all time high*, *an all time low*, *last price* dan *closing price*. Hasil prediksi terbaik yang diperoleh dengan kelima algoritma adalah dengan algoritma *Long Short-Term Memory* dengan *Symmetric Mean Absolute Percentage Error* (SMAPE) sebesar (1,59), *R-Squared Value* sebesar (-0,11) dan *Root Mean Square Error* (RMSE) sebesar (22,25), kemudian hasil kedua yang terbaik adalah *Support Vector Machine* dengan *Symmetric Mean Absolute Percentage Error* (SMAPE) sebesar (5,59), *R-Squared Value* sebesar (-1,69) dan *Root Mean Square Error* (RMSE) sebesar (46,36). Sementara hasil yang diberikan algoritma *Line-*

ar Regression dan Decision Tree Regression hampir setara, K-NN menunjukkan prediksi yang paling buruk.

Pada penelitian Adena Wahyu Gumelar dan Tacbir Hendro Pudjiantoro yang berjudul *NFT Coin Price Prediction (Non-Fungible Token) using K-Nearest Neighbors Method*, dilakukan prediksi harga penutupan (Close) dari koin NFT setiap harinya. Variabel yang digunakan dalam penelitian ini adalah *open*, *high*, *low*, *close*, *volume*, dan *marketcap*. Penelitian ini memilih nilai K optimal sebanyak 3 yaitu 3, 5, dan 7. Hasil penelitian menunjukkan bahwa akurasi tertinggi dicapai dengan menggunakan K=7, yaitu sebesar 65% (Gumelar et al. 2022).

Penelitian (Samruddhi and Kumar 2020) berjudul *Used car price prediction using k-nearest neighbor based model* bertujuan untuk memprediksi harga mobil bekas menggunakan model K-NN regression. Dalam penelitian ini, digunakan tiga jenis rasio sebagai fitur-fitur untuk melatih data guna menemukan nilai optimal dari parameter K, yang bervariasi dari 2 hingga 10. Hasil dari analisis menunjukkan bahwa tingkat akurasi dalam memprediksi harga mobil bekas mencapai 85%, dengan nilai K optimal adalah 4, RMSE (*Root Mean Square Error*) sebesar 4.01, dan MAE (*Mean Absolute Error*) sebesar 2.01. Dengan demikian, penelitian ini menunjukkan bahwa model K-NN regression berhasil memberikan prediksi harga mobil bekas yang cukup akurat.

Penelitian (Bhavsar and Gor 2022) berjudul *Stock Price Prediction Using Support Vector Regression and K-Nearest Neighbors: A Comparison* bertujuan untuk membandingkan kinerja algoritma K-Nearest Neighbors (K-NN) dan Support Vector Regression (SVR) dalam memprediksi harga saham. Penelitian ini menggunakan tujuh variabel, yaitu *open*, *high*, *low*, *close*, *adj close*, *volume*. Hasil penelitian menunjukkan bahwa model SVR memiliki Root Mean Square Error sebesar 12.458 dan akurasi pada data pengujian sebesar 99.57%, sedangkan model K-NN memiliki Root Mean Square Error yang sama sebesar 12.458 namun dengan akurasi sedikit lebih rendah, yaitu 99.37%. Kesimpulannya, meskipun kedua model menunjukkan performa yang baik, model SVR sedikit lebih akurat dibandingkan dengan model K-NN dalam prediksi harga saham.

Berdasarkan tinjauan literatur dan penelitian sebelumnya, terdapat beberapa alasan yang mendasari pemilihan algoritma K-Nearest Neighbor Regression untuk memprediksi harga penutupan saham. Pertama, karakteristik saham yang cenderung non-linear dan kompleks membuat algoritma non-parametrik seperti K-NN lebih sesuai karena tidak memerlukan asumsi distribusi data tertentu. Selain itu, penelitian sebelumnya menunjukkan bahwa K-NN dapat memberikan hasil prediksi yang baik dalam berbagai aplikasi, termasuk prediksi harga saham dan komoditas lainnya. Hal ini terutama disebabkan oleh sifatnya yang sederhana dan mudah dipahami, serta kemampuannya untuk menangani data berstruktur kompleks tanpa memerlukan asumsi distribusi data tertentu.

Memilih saham BBKA dan BBNI sebagai objek penelitian menggunakan algoritma K-Nearest Neighbor (K-NN) memiliki beberapa alasan yang kuat. Dilihat dari data BBKA pada tahun 2018 hingga 2024, saham ini mengalami kenaikan meskipun pada tahun 2019 dan 2020 terjadi pandemi COVID-19, yang tidak mengakibatkan penurunan signifikan. Sementara itu, saham BBNI pada tahun 2020 mengalami penurunan dibandingkan tahun 2018 dan mengalami kenaikan kembali pada tahun 2021. Berdasarkan data ini, kedua saham ini dipilih untuk menerapkan algoritma K-NN. Saham BBKA dan BBNI cenderung memiliki volatilitas yang lebih stabil dibandingkan saham perusahaan kecil atau sektor berisiko, sehingga mempermudah penerapan algoritma K-NN untuk prediksi yang akurat. Alasan lain pemilihan saham ini, dikarenakan saham BBKA dan BBNI termasuk dalam kategori saham blue chip, yaitu saham-saham dari perusahaan-perusahaan dengan reputasi kuat, jejak keuangan yang stabil, serta sejarah pertumbuhan yang konsisten. Jenis saham ini banyak dicari oleh investor karena berasal dari perusahaan-perusahaan yang memimpin di industri sejenis, memiliki pendapatan yang stabil, dan konsisten dalam pembayaran dividen (OJK).

Pemilihan variabel *open*, *high*, *low*, dan *close* didasarkan pada beberapa alasan. Variabel-variabel ini merupakan komponen utama dalam analisis teknikal yang digunakan untuk memahami perilaku dan pola pergerakan harga saham.

Algoritma K-Nearest Neighbor merupakan non-parametrik, yang sejalan dengan objek yang diteliti, di mana hubungan antar variabel independen saling memiliki keterkaitan yang kuat satu sama lain. Dengan menggunakan variabel-variabel ini, model yang dibangun dapat tetap sederhana namun efektif dalam memberikan prediksi yang akurat. Dengan demikian, penggunaan K-Nearest Neighbor Regression dan pemilihan variabel *open*, *high*, *low*, dan *close* diharapkan dapat memberikan hasil prediksi yang berguna bagi investor atau trader dalam pengambilan keputusan investasi.

Pemilihan algoritma *K-Nearest Neighbor* dibanding algoritma lain sebagai algoritma dalam penelitian ini karena algoritma yang mudah dipahami untuk klasifikasi atau regresi. Algoritma ini juga disebut sebagai algoritma yang malas atau *lazy algorithm* yang tidak menggunakan titik data pelatihan untuk melakukan generalisasi, dengan kata lain, ia menyimpan semua data pelatihan selama fase pengujian. Oleh karena itu, ukuran populasi menjadi perhatian utama untuk *K-Nearest Neighbor*, karena ukuran populasi yang besar dapat mengakibatkan kecepatan eksekusi yang lambat dan persyaratan memori yang besar (Amalia et al. 2021).

1.2 Identifikasi Masalah

Saham merupakan instrumen investasi yang menjanjikan keuntungan yang menggiurkan dibandingkan dengan instrumen investasi lainnya. Namun, di balik potensi keuntungan yang menarik, saham juga membawa risiko yang tinggi karena rentan terhadap fluktuasi harga yang tidak dapat diprediksi secara pasti. Karena sifatnya yang fluktuatif ini, harga saham seringkali menampilkan pola yang bersifat non-linear, yang membuatnya sulit untuk diidentifikasi dengan tepat. Oleh karena itu, untuk mengatasi tantangan ini, diperlukan pemodelan atau prediksi yang mampu mengestimasi atau memprediksi harga penutupan suatu saham.

1.3 Ruang Lingkup

Ruang lingkup pada penelitian ini adalah memprediksi harga penutupan saham Bank Central Asia dengan kode saham BBKA dan Bank Negara Indonesia dengan kode saham BBNI. Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari *Yahoo Finance*, berisi data historis saham Bank Central Asia (BBKA) dan Bank Negara Indonesia (BBNI) rentang waktu tanggal 2 Januari 2018 sampai 14 April 2024, dimana mencakup informasi harga pembukaan (*open price*), harga tertinggi (*high price*), harga terendah (*low price*), dan harga penutupan (*close price*). Penelitian ini menggunakan algoritma *K-Nearest Neighbor Regression* untuk memprediksi harga penutupan kedua saham tersebut

1.4 Batasan Masalah

1. Data yang digunakan yaitu harga *open*, *high*, *low*, dan *close* pada saham BBKA dan BBNI dengan data tanggal 2 Januari 2018 sampai 14 April 2024
2. Penelitian ini menggunakan algoritma *K-Nearest Neighbor Regression*
3. Data yang diprediksi adalah harga *close* hari selanjutnya.

1.5 Rumusan Masalah

1. Bagaimana menerapkan algoritma *K-Nearest Neighbor Regression* dalam memprediksi harga penutupan saham BBKA dan BBNI?
2. Bagaimana akurasi algoritma *K-Nearest Neighbor Regression* dalam memprediksi harga penutupan saham BBKA dan BBNI berdasarkan metrik evaluasi *R-Squared*, MAPE dan MAE?

1.6 Tujuan Penelitian

1. Untuk menerapkan algoritma *K-Nearest Neighbor Regression* dalam memprediksi harga penutupan saham BBKA dan BBNI

2. Untuk mengetahui akurasi algoritma *K-Nearest Neighbor Regression* dalam memprediksi harga penutupan saham BBKA dan BBNI berdasarkan metrik evaluasi *R-Squared*, MAPE dan MAE.

1.7 Manfaat Penelitian

1. Prediksi harga penutupan saham dengan algoritma *K-Nearest Neighbor Regression* diharapkan dapat membantu dalam pengembangan model dan teori dibidang analisis keuangan.
2. Menjadi sarana referensi bagi penelitian berikutnya.
3. Memberikan informasi kepada investor maupun trader, yang dapat digunakan referensi tambahan untuk mengambil keputusan investasi.
4. Bagi penulis, penulis dapat menerapkan ilmu pengetahuan yang diperoleh selama dibangku perkuliahan.