

BAB I

PENDAHULUAN

1.1 Latar Belakang

Saham merupakan jenis investasi sekaligus instrumen keuangan yang menggambarkan kepemilikan individu/badan usaha pada suatu perusahaan terbuka yang sahamnya tercatat serta terdaftar di bursa efek. Investor yang membeli saham menjadi sebagian pemilik perusahaan tersebut dan memiliki hak atas penghasilan perusahaan (berupa dividen), potensi kenaikan harga, yakni *capital gain*, serta hak suara dalam aktivitas perusahaan (Maruddani, 2019). Beberapa faktor yang mempengaruhi harga saham meliputi; faktor fundamental perusahaan, hukum permintaan dan penawaran, tingkat suku bunga, kurs valas, dana asing di bursa, indeks harga saham, berita dan rumor yang berkembang, dividen, laba perusahaan, dan faktor lain (Aprilianti *et al.*, 2024). Sama halnya seperti harga barang dan komoditas di pasar, harga saham bersifat fluktuatif yang berarti sifat ini mencerminkan tingginya kemungkinan terjadinya kenaikan dan penurunan harga dari waktu ke waktu.

Data harga saham pada dasarnya merupakan data *time series* dengan observasi dicatat secara berurutan pada interval waktu tertentu (seperti harian, mingguan, atau bulanan). Zhao *et al.* (2023) mengatakan bahwa data harga saham sebagai data *time series* mengandung dependensi antar waktu yang kuat, dengan nilai pada suatu periode waktu juga dipengaruhi oleh nilai-nilai pada periode sebelumnya. Hal tersebut menjadikan prediksi harga saham menarik dilakukan, khususnya melalui pendekatan *time series* yang memanfaatkan nilai-nilai masa lalu

sebagai variabel prediktor dalam membentuk model peramalan. Oleh karena itu, pemilihan metode analisis yang sesuai dengan karakteristik data *time series* menjadi penting dalam prediksi harga saham.

Perkembangan analisis data *time series* telah melahirkan berbagai pendekatan dalam memodelkan dan memprediksi data *time series*. Secara umum, metode prediksi *time series* dapat diklasifikasikan ke dalam dua kelompok besar, yaitu metode parametrik dan metode nonparametrik (Gautam & Singh, 2020). Sebagai salah satu metode parametrik yang banyak digunakan, *Auto-Regressive Integrated Moving Average* (ARIMA) dibangun atas asumsi-asumsi yang ketat. Asumsi-asumsi kunci model ARIMA mencakup stasioneritas data, asumsi bahwa residual atau error model bersifat acak, tidak saling berkorelasi, memiliki rata-rata nol dan varians konstan atau yang biasa disebut asumsi *white noise*, asumsi residual berdistribusi normal (Tsay, 2005; Fattah *et al.*, 2018).

Meskipun metode parametrik seperti ARIMA masih banyak digunakan karena interpretabilitas dan kemudahan implementasinya, berbagai penelitian menunjukkan bahwa asumsi-asumsi dasar model parametrik tidak sepenuhnya terpenuhi dalam data harga saham. Penelitian oleh Sudipa *et al.* (2023) pada saham BMRI menemukan pelanggaran asumsi independensi dan normalitas residual. Sementara Maulidya *et al.* (2024) melaporkan bahwa beberapa model ARIMA yang diidentifikasi juga gagal memenuhi asumsi independensi residual. Temuan serupa disampaikan oleh Rahmawati *et al.* (2024) pada pemodelan harga saham BBCA, yang menunjukkan ketidakpatuhan terhadap asumsi normalitas residual. Pelanggaran terhadap asumsi-asumsi tersebut berpotensi menghasilkan estimasi dan keputusan yang menyesatkan jika tetap digunakan (Suparti *et al.*, 2023),

sehingga menunjukkan bahwa dalam praktiknya pendekatan parametrik seperti ARIMA sering menghadapi keterbatasan ketika diterapkan pada data pasar saham yang bersifat kompleks dan volatil.

Berbeda dari metode parametrik, metode nonparametrik merupakan pendekatan alternatif yang bisa digunakan dalam pemodelan data *time series*. Menurut Hardle *et al.* (1997), dalam analisis *time series*, metode nonparametrik dapat digunakan dengan memanfaatkan nilai *lag* atau nilai masa lalu sebagai variabel prediktor. Pendekatan ini juga memungkinkan pemodelan yang lebih fleksibel, yakni tidak memerlukan bentuk fungsional yang ditentukan sebelumnya. Oleh karena itu, regresi nonparametrik sebaiknya digunakan ketika pola hubungan antara variabel respon dan variabel prediktor belum diketahui secara pasti, tidak dapat dijelaskan dengan bentuk fungsi matematis tertentu, dan atau ketika data menunjukkan kecenderungan hubungan yang kompleks dan nonlinier. Selain itu, Rica *et al.* (2016) mengungkapkan bahwa metode nonparametrik juga dapat menjadi alternatif apabila salah satu asumsi pemodelan parametrik tidak terpenuhi. Kelebihan tersebut dapat memberikan kegunaan yang dapat diandalkan dalam analisis data *time series*.

Salah satu model nonparametrik yang dapat digunakan dalam pemodelan data *time series* adalah regresi kernel. Sama seperti model nonparametrik pada umumnya, regresi kernel merupakan model statistika yang digunakan untuk mengestimasi hubungan antara variabel respon dan variabel prediktor berdasarkan data pengamatan dengan menggunakan teknik pemulusan, tanpa menspesifikasikan bentuk fungsi tertentu. Prinsip dasar regresi kernel adalah mengestimasi nilai ekspektasi bersyarat variabel respon pada suatu titik dengan memberikan bobot yang

lebih besar pada observasi yang berada di sekitar titik tersebut (Sadek & Mohammed, 2024). Pendekatan ini memungkinkan data menentukan sendiri bentuk hubungan yang ada, sehingga sangat sesuai untuk data nonlinier maupun data acak.

Regresi kernel terbagi menjadi 2 bagian berdasarkan jumlah variabel prediktor-nya, yakni regresi kernel univariabel dan regresi kernel multivariabel. Regresi kernel univariabel meliputi satu variabel prediktor sedangkan regresi kernel multivariabel meliputi lebih dari satu variabel prediktor.

Dalam pemodelan regresi kernel, salah satu estimator fungsi regresi kernel yang biasa digunakan adalah estimator Nadaraya-Watson (Anisa *et al.*, 2019). Estimator Nadaraya-Watson bekerja dengan mengestimasi fungsi regresi sebagai rata-rata tertimbang secara lokal dengan menggunakan fungsi kernel sebagai pembobot (Karimuse *et al.*, 2023). Terdapat beberapa jenis fungsi kernel, antara lain *Uniform*, *Triangle*, *Biweight*, *Triweight*, *Gaussian*, dan *Epanechnikov* (Deng *et al.*, 2011). Fungsi kernel *Gaussian* merupakan fungsi yang paling sering digunakan dalam pemodelan regresi kernel dan juga pada fungsi ini domainnya terbentang ke semua nilai riil. Untuk mendapatkan estimasi yang baik, terdapat parameter lain yang memainkan peran penting dalam regresi kernel selain fungsi kernel, yakni *bandwidth* atau parameter pemulus. Nilai *bandwidth* yang terlalu kecil akan menyebabkan fungsi yang diestimasi menjadi sangat kasar sehingga menghasilkan variansi yang tinggi dan memiliki potensi bias yang rendah sedangkan nilai *bandwidth* yang terlalu besar akan menghasilkan fungsi yang terlalu mulus menyebabkan variansi-nya rendah dan potensi bias yang tinggi. Menurut Komang *et al.* (2012), dalam regresi kernel yang lebih penting adalah memilih nilai *bandwidth* yang optimal, bukan pemilihan fungsi kernel. Hal ini dikarenakan

penggunaan fungsi kernel yang berbeda menggunakan *bandwidth* yang optimal akan menghasilkan estimasi fungsi regresi yang hampir sama.

Pencarian nilai *bandwidth* yang optimal membantu dalam menemukan model terbaik yang dapat meningkatkan kemampuan model dalam mengestimasi nilai aktual. Beberapa kriteria optimasi yang umum digunakan dalam optimasi *bandwidth* adalah *Mean Squared Error* (MSE), *Cross-Validation* (CV), dan *Generalized Cross-Validation* (GCV). Dalam penerapannya, kriteria optimasi menggunakan MSE seringkali cenderung menghasilkan pemilihan *bandwidth* yang sangat kecil yang terbukti sangat baik pada *in-sample*, namun hal ini menyebabkan terjadinya *overfitting* dan memiliki kinerja yang tidak lebih baik dari kriteria optimasi lainnya pada *out-sample*. Menurut Kohler *et al.* (2011), kriteria optimasi CV dapat menghasilkan performa yang bagus, namun seiring dengan bertambahnya jumlah data, kriteria ini dapat digantikan oleh kriteria GCV. Namun demikian, kriteria optimasi CV dan GCV, memiliki kinerja yang hampir ekuivalen pada jumlah data yang digunakan/ditentukan.

Terdapat beberapa penelitian terdahulu yang menggunakan regresi kernel dalam memprediksi suatu nilai. Penelitian yang dilakukan Anisa *et al.* (2019) yang menggunakan regresi kernel pada pemodelan pengaruh angka harapan hidup terhadap data indeks pembangunan manusia di Indonesia menggunakan fungsi kernel Gaussian dan optimasi *bandwidth* CV. Diperoleh koefisien determinasi 0,632 dan MAPE sebesar 2,5%. Penelitian oleh Lamusu *et al.* (2020) yang memodelkan data produksi jagung dengan variabel respon Luas Tanam, Luas Panen, dan Pupuk dengan menggunakan optimasi CV dan GCV dengan fungsi kernel Gaussian.

Didapat hasil pada metode CV nilai koefisien determinasi sebesar 0,995349 dan pada metode GCV didapat nilai koefisien determinasi sebesar 0,9998944.

Lebih khusus, beberapa penelitian juga dilakukan dalam pemodelan *time series* regresi kernel. Penelitian yang dilakukan oleh Suparti *et al.* (2023) yang menggunakan model kernel dalam memodelkan data harga saham MDKA dengan menggunakan *lag*-1 sebagai prediktor menggunakan fungsi kernel Gaussian dan optimasi GCV. Didapat hasil koefisien determinasi sebesar 0,9828372 dan juga MAPE *out-sample* sebesar 1,985681% menunjukkan kinerja model sangat bagus pada data *out-sample*. Penelitian yang lain dilakukan oleh Pembargi *et al.* (2023) yang memodelkan pendapatan asli daerah di Kabupaten Lombok Tengah menggunakan *lag*-1 dan *lag*-12 sebagai prediktor dengan menggunakan fungsi kernel Gaussian dan optimasi GCV. Didapatkan koefisien determinasi 0,8755 dengan MAPE pada *out-sample* sebesar 5,4%.

Berdasarkan pemaparan yang telah disampaikan, penelitian yang dilakukan adalah pemodelan regresi nonparametrik kernel *time series* multivariabel dengan optimasi GCV pada harga saham BBKA, mengingat bahwa penggunaan lebih dari 1 *lag* sebagai prediktor dalam regresi nonparametrik kernel *time series* masih sangat sedikit dilakukan.

BBKA merupakan kode saham perusahaan yang bergerak di bidang perbankan, yakni PT Bank Central Asia Tbk. Menurut laporan IDX *Company Fact Sheet* per 31 Juli 2025 BBKA merupakan salah satu saham yang paling aktif diperjualbelikan dan perusahaan dengan kapitalisasi market terbesar ke-2 di Indonesia. Tingginya frekuensi transaksi dan besarnya kapitalisasi pasar

menunjukkan saham BBKA merupakan saham yang likuid, berpengaruh di pasar modal, serta banyak diminati investor. Kondisi tersebut menjadikan saham BBKA relevan untuk dijadikan objek penelitian. Selain itu, ketersediaan data historis harga saham BBKA juga mendukung dilakukannya pemodelan ini.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka dapat ditentukan beberapa rumusan, yakni:

1. Bagaimana model regresi nonparametrik *time series* kernel multivariabel pada peramalan harga saham BBKA dengan penentuan *bandwidth* optimal menggunakan kriteria optimasi *Generalized Cross Validation* (GCV)?
2. Seberapa baik kinerja model regresi nonparametrik *time series* kernel multivariabel dengan *bandwidth* kriteria GCV optimal dalam peramalan harga saham BBKA ditinjau dari ukuran kebaikan model seperti koefisien determinasi yang disesuaikan (*Adjusted R²*) dan MAPE?

1.3 Batasan Masalah

Pembatasan masalah perlu dilakukan dalam suatu penelitian untuk membantu peneliti agar terarah dan lebih fokus sesuai dengan topik penelitian. Pada penelitian ini, analisis yang digunakan adalah analisis regresi nonparameterik *time series* multivariabel menggunakan estimator Nadaraya-Watson dengan memanfaatkan fungsi kernel Gaussian sebagai fungsi pembobotnya dan kriteria optimasi GCV. Data yang digunakan adalah data harga penutupan saham BBKA dari tanggal 1 November 2021 sampai dengan 30 Oktober 2025. Data tersebut

dibagi menjadi menjadi data *training (in-sample)* sebanyak 85% dan data *testing (out-sample)* sebanyak 15%.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, tujuan yang ingin dicapai dari penelitian ini yaitu:

1. Memperoleh model terbaik pada data harga saham BBCA dengan menggunakan optimasi GCV.
2. Memperoleh ukuran kebaikan model dan evaluasi kinerja model terbaik berupa koefisien determinasi dan MAPE yang dihasilkan.