

BAB 2

LANDASAN TEORI

2.1 Pasar Saham

Pasar saham merupakan bagian dari pasar modal yang mempertemukan pihak yang membutuhkan dana dengan pihak yang memiliki dana melalui aktivitas jual beli saham. Saham menunjukkan kepemilikan terhadap suatu perusahaan, sehingga perubahan harga saham dapat mencerminkan dinamika nilai perusahaan di pasar. Harga saham dapat berubah dari waktu ke waktu karena dipengaruhi oleh berbagai faktor, seperti kondisi ekonomi, kebijakan moneter, faktor global, karakteristik perusahaan, serta sentimen pasar [1].

Pada sektor perbankan, harga saham dapat mengalami volatilitas karena sektor ini berkaitan dengan kondisi ekonomi dan kebijakan keuangan. Oleh karena itu, saham sektor perbankan dapat digunakan sebagai objek penelitian dalam kajian prediksi harga saham.

Salah satu saham sektor perbankan yang digunakan dalam penelitian ini adalah saham PT Bank Central Asia Tbk (BBCA). Saham BBCA telah digunakan dalam penelitian terkait analisis harga saham sektor finansial, sehingga relevan digunakan sebagai objek penelitian prediksi harga saham [2].

2.2 Harga Saham

Harga saham merupakan nilai yang terbentuk dari aktivitas jual beli saham di pasar. Nilai tersebut dapat berubah sesuai dengan permintaan dan penawaran yang terjadi selama periode perdagangan. Perubahan harga saham menunjukkan adanya dinamika pasar yang dapat dipengaruhi oleh kondisi perusahaan, kondisi ekonomi, serta faktor eksternal lainnya [2].

Dalam penelitian prediksi harga saham, data harga saham digunakan sebagai dasar untuk membentuk variabel masukan dan target prediksi. Data harga saham dapat dianalisis berdasarkan pola historis yang terbentuk dari waktu ke waktu. Namun, pola pergerakan harga saham tidak selalu tetap karena harga saham dapat dipengaruhi oleh banyak faktor, baik faktor internal perusahaan maupun faktor eksternal pasar.

Pada penelitian ini, data harga saham yang digunakan berasal dari PT Bank Central Asia Tbk dengan kode saham BBCA. Fokus utama penelitian adalah harga

penutupan atau *close price*, karena nilai tersebut digunakan sebagai target prediksi dalam proses pemodelan.

2.3 Data Historis Saham

Data historis saham merupakan data harga saham yang tercatat dalam periode waktu tertentu. Data ini disusun berdasarkan urutan waktu dan dapat digunakan untuk menganalisis pergerakan harga saham dari periode sebelumnya. Dalam konteks prediksi harga saham, data historis digunakan untuk membentuk pola yang dapat dipelajari oleh model prediksi [3].

Data historis saham umumnya terdiri dari beberapa atribut, seperti *Open*, *High*, *Low*, *Close*, *Adj Close*, dan *Volume*. Atribut *Open* menunjukkan harga pembukaan saham, *High* menunjukkan harga tertinggi, *Low* menunjukkan harga terendah, *Close* menunjukkan harga penutupan, *Adj Close* menunjukkan harga penutupan yang telah disesuaikan, dan *Volume* menunjukkan jumlah saham yang diperdagangkan.

Dalam penelitian ini, data historis saham digunakan sebagai dasar dalam pembentukan fitur. Atribut yang digunakan adalah *Close*, karena penelitian ini berfokus pada prediksi harga penutupan saham BBKA. Data historis tersebut kemudian diolah menjadi beberapa fitur, yaitu MA5, MA10, *Return*, serta *lag* 1 sampai dengan *lag* 7.

2.4 Harga Penutupan (*Close Price*)

Harga penutupan atau *close price* merupakan harga terakhir yang tercatat pada akhir periode perdagangan. Dalam penelitian prediksi saham, *close price* sering digunakan sebagai target prediksi karena nilai tersebut merepresentasikan harga akhir saham pada suatu periode. Beberapa penelitian terkait prediksi harga penutupan saham BBKA juga menggunakan harga penutupan sebagai target prediksi [5, 8].

Pada penelitian ini, *close price* digunakan sebagai variabel target. Nilai *close price* diprediksi berdasarkan fitur yang dibentuk dari data historis harga penutupan sebelumnya. Dengan demikian, model tidak menggunakan informasi dari periode yang sedang diprediksi sebagai variabel masukan.

2.5 Time Series

Data *time series* merupakan data yang dikumpulkan dan disusun berdasarkan urutan waktu. Pada data harga saham, setiap nilai harga memiliki keterkaitan dengan periode waktu tertentu, sehingga urutan data menjadi hal yang penting dalam proses analisis dan prediksi [3].

Dalam penelitian prediksi harga saham, urutan waktu tidak dapat diabaikan karena data pada periode sebelumnya dapat digunakan untuk membentuk fitur yang berkaitan dengan periode yang akan diprediksi. Oleh karena itu, proses pembagian data dalam penelitian ini dilakukan tanpa pengacakan atau *shuffle*. Data latih diambil dari bagian awal data, sedangkan data uji diambil dari bagian akhir data, sehingga urutan kronologis data tetap dipertahankan.

2.6 Prediksi Rekursif

Prediksi rekursif merupakan proses prediksi beberapa periode ke depan dengan menggunakan hasil prediksi sebelumnya sebagai bagian dari masukan untuk prediksi periode berikutnya. Pada prediksi satu periode ke depan, model menggunakan data historis terakhir yang tersedia. Selanjutnya, hasil prediksi tersebut digunakan kembali untuk membentuk fitur pada periode berikutnya sampai jumlah periode prediksi yang ditentukan tercapai [9].

Dalam penelitian ini, *forecasting* 7 hari kerja setelah data terakhir dilakukan secara rekursif. Prediksi hari pertama dibuat berdasarkan data historis terakhir. Prediksi hari kedua menggunakan hasil prediksi hari pertama sebagai bagian dari data historis untuk membentuk fitur MA5, MA10, *Return*, dan *lag*. Proses tersebut dilakukan sampai diperoleh hasil *forecasting* untuk 7 hari kerja setelah data terakhir.

Prediksi rekursif dapat menghasilkan akumulasi kesalahan apabila periode prediksi semakin panjang. Oleh karena itu, penelitian ini membatasi *forecasting* masa depan pada 7 hari kerja setelah data terakhir.

2.7 Feature Engineering

Feature engineering merupakan proses pembentukan atau pengolahan variabel dari data awal agar dapat digunakan sebagai masukan bagi model *machine learning*. Dalam prediksi harga saham, fitur dapat dibentuk dari data historis untuk merepresentasikan pola pergerakan harga pada periode sebelumnya [3, 4].

Pada penelitian ini, *feature engineering* dilakukan dengan membentuk fitur *moving average*, *return*, dan *lag*. Fitur-fitur tersebut dibentuk dari data historis harga penutupan sebelum periode yang diprediksi. Hal ini dilakukan agar model hanya menggunakan informasi yang telah tersedia sebelum nilai target diprediksi.

Pembentukan fitur dilakukan dengan memperhatikan urutan waktu karena data saham termasuk data *time series*. Fitur yang digunakan dalam model tidak boleh dibentuk dari data pada periode yang sedang diprediksi. Hal ini bertujuan untuk menghindari terjadinya *data leakage*, yaitu kondisi ketika informasi dari target prediksi ikut masuk ke dalam variabel masukan.

2.7.1 Moving Average

Moving average merupakan fitur yang digunakan untuk menggambarkan rata-rata harga dalam periode tertentu. Pada data harga saham, *moving average* dapat digunakan untuk melihat kecenderungan harga berdasarkan beberapa periode sebelumnya. Dalam penelitian ini digunakan dua fitur *moving average*, yaitu MA5 dan MA10.

MA5 merupakan rata-rata harga penutupan selama 5 hari sebelumnya. MA10 merupakan rata-rata harga penutupan selama 10 hari sebelumnya. Kedua fitur tersebut digunakan untuk merepresentasikan rata-rata historis harga penutupan dalam periode yang berbeda.

Secara umum, pembentukan MA5 dapat dituliskan pada Persamaan (2.1):

$$MA5_t = \frac{Close_{t-1} + Close_{t-2} + Close_{t-3} + Close_{t-4} + Close_{t-5}}{5} \quad (2.1)$$

Sementara itu, pembentukan MA10 dapat dituliskan pada Persamaan (2.2):

$$MA10_t = \frac{1}{10} \sum_{i=1}^{10} Close_{t-i} \quad (2.2)$$

di mana $Close_{t-i}$ menunjukkan harga penutupan pada periode sebelum hari yang diprediksi.

2.7.2 Return

Return merupakan fitur yang digunakan untuk menggambarkan perubahan harga saham dari satu periode ke periode berikutnya. Pada data harga saham, *return* menunjukkan besarnya perubahan harga penutupan secara relatif terhadap harga pada periode sebelumnya.

Dalam penelitian ini, *return* dihitung berdasarkan persentase perubahan harga penutupan harian sebelum hari yang diprediksi. Rumus *return* yang digunakan dapat dituliskan pada Persamaan (2.3):

$$Return_t = \frac{Close_{t-1} - Close_{t-2}}{Close_{t-2}} \quad (2.3)$$

di mana $Return_t$ adalah nilai *return* yang digunakan untuk memprediksi periode t , $Close_{t-1}$ adalah harga penutupan satu periode sebelumnya, dan $Close_{t-2}$ adalah harga penutupan dua periode sebelumnya.

2.7.3 Lag

Fitur *lag* merupakan fitur yang digunakan untuk merepresentasikan nilai pada periode sebelumnya. Pada data *time series*, fitur *lag* digunakan agar model dapat mempelajari hubungan antara nilai historis dengan nilai yang akan diprediksi.

Dalam penelitian ini digunakan 7 fitur *lag*, yaitu *lag* 1 sampai dengan *lag* 7. Fitur tersebut menunjukkan harga penutupan pada 1 hari sebelumnya sampai 7 hari sebelumnya.

Secara umum, fitur *lag* dapat dituliskan pada Persamaan (2.4):

$$Lag_i = Close_{t-i} \quad (2.4)$$

di mana Lag_i adalah fitur *lag* ke- i , dan $Close_{t-i}$ adalah harga penutupan pada i periode sebelum hari yang diprediksi.

Fitur *lag* yang digunakan dalam penelitian ini adalah sebagai berikut:

1. *lag* 1, yaitu harga penutupan 1 hari sebelumnya.
2. *lag* 2, yaitu harga penutupan 2 hari sebelumnya.
3. *lag* 3, yaitu harga penutupan 3 hari sebelumnya.

4. *lag* 4, yaitu harga penutupan 4 hari sebelumnya.
5. *lag* 5, yaitu harga penutupan 5 hari sebelumnya.
6. *lag* 6, yaitu harga penutupan 6 hari sebelumnya.
7. *lag* 7, yaitu harga penutupan 7 hari sebelumnya.

2.8 Machine Learning

Machine learning merupakan metode komputasi yang memungkinkan model mempelajari pola dari data dan menghasilkan prediksi terhadap data baru. Dalam konteks prediksi harga saham, *machine learning* digunakan untuk melatih model berdasarkan data historis sehingga model mampu mengenali pola hubungan antara variabel masukan dan variabel target [3].

Metode *machine learning* banyak digunakan dalam prediksi harga saham karena dapat diterapkan pada data yang memiliki pola kompleks. Namun, kinerja model dapat berbeda tergantung pada karakteristik data, fitur yang digunakan, algoritma yang diterapkan, dan metode evaluasi yang digunakan [4]. Oleh karena itu, penelitian ini membandingkan dua algoritma, yaitu *Linear Regression* dan *Support Vector Regression*, untuk mengetahui perbedaan tingkat kesalahan prediksi pada data saham BBKA.

2.9 Linear Regression

Linear Regression merupakan metode regresi yang digunakan untuk memodelkan hubungan linear antara variabel masukan dan variabel target. Pada prediksi harga saham, metode ini dapat digunakan untuk memperkirakan nilai harga saham berdasarkan fitur yang dibentuk dari data historis [5].

Persamaan umum *Linear Regression* ditunjukkan pada Persamaan (2.5):

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (2.5)$$

di mana $\hat{y}$ adalah hasil prediksi, $x_1, x_2, \dots, x_n$ adalah variabel masukan, β_0 adalah konstanta, dan $\beta_1, \beta_2, \dots, \beta_n$ adalah koefisien model.

Penerapan *Linear Regression* dalam prediksi harga saham BCA juga telah digunakan pada beberapa penelitian sebelumnya, seperti penelitian Zapar dkk. [10]

dan Satriyo dkk. [11]. Berdasarkan penelitian tersebut, *Linear Regression* dapat digunakan sebagai salah satu model pembandingan dalam penelitian prediksi harga saham.

2.10 *Support Vector Regression*

Support Vector Regression (SVR) merupakan metode regresi yang dikembangkan dari konsep *Support Vector Machine*. SVR digunakan untuk memprediksi nilai numerik dengan membentuk fungsi regresi yang memiliki batas toleransi kesalahan tertentu. Dalam konteks prediksi saham, SVR dapat digunakan untuk memodelkan hubungan antara fitur historis dan harga saham yang diprediksi [7].

Secara umum, fungsi SVR dapat dituliskan pada Persamaan (2.6):

$$f(x) = w \cdot x + b \quad (2.6)$$

di mana $f(x)$ adalah fungsi prediksi, w adalah bobot model, x adalah variabel masukan, dan b adalah bias. SVR menggunakan parameter ε sebagai batas toleransi kesalahan dalam pembentukan fungsi regresi.

Pada SVR dengan fungsi *kernel*, hasil prediksi dapat dituliskan dalam bentuk umum seperti pada Persamaan (2.7):

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (2.7)$$

di mana $f(x)$ adalah fungsi prediksi, α_i dan α_i^* adalah koefisien hasil pelatihan, $K(x_i, x)$ adalah fungsi *kernel*, x_i adalah data latih yang menjadi *support vector*, x adalah data masukan yang akan diprediksi, dan b adalah bias.

Pada penelitian ini, SVR digunakan karena memiliki pendekatan yang berbeda dari *Linear Regression*. Jika *Linear Regression* memodelkan hubungan secara linear, maka SVR dapat menggunakan fungsi *kernel* untuk memodelkan hubungan data yang tidak linear.

2.11 Fungsi Kernel

Fungsi *kernel* digunakan pada SVR untuk membantu memodelkan hubungan data yang tidak linear. Salah satu fungsi *kernel* yang digunakan dalam penelitian ini adalah *Radial Basis Function* (RBF). Penggunaan SVR dengan fungsi *kernel* dalam prediksi saham telah digunakan pada penelitian terkait prediksi harga penutupan saham BCA [8].

Fungsi RBF dapat dituliskan pada Persamaan (2.8):

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (2.8)$$

di mana $K(x_i, x_j)$ adalah nilai fungsi *kernel*, x_i dan x_j adalah pasangan data masukan, serta γ adalah parameter yang mengatur pengaruh suatu data terhadap pembentukan model.

Dalam penelitian ini, SVR menggunakan *kernel* RBF dengan parameter $C = 1000$, $\gamma = 0.001$, dan $\varepsilon = 0.01$. Parameter tersebut digunakan dalam proses pelatihan model untuk menghasilkan prediksi harga penutupan saham BBKA.

2.12 Evaluasi Model

Evaluasi model dilakukan untuk mengukur tingkat kesalahan prediksi yang dihasilkan oleh model. Dalam penelitian ini, metode evaluasi utama yang digunakan adalah *Root Mean Square Error* (RMSE). RMSE digunakan untuk mengukur selisih antara nilai aktual dan nilai prediksi dalam satuan yang sama dengan variabel target [12].

Rumus RMSE ditunjukkan pada Persamaan (2.9):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.9)$$

di mana y_i adalah nilai aktual, $\hat{y}_i$ adalah nilai prediksi, dan n adalah jumlah data.

Nilai RMSE yang lebih kecil menunjukkan bahwa hasil prediksi memiliki tingkat kesalahan yang lebih rendah pada data yang dievaluasi. Oleh karena itu, dalam penelitian ini RMSE digunakan sebagai dasar utama untuk membandingkan kinerja model *Linear Regression* dan SVR.

Selain RMSE, penelitian ini juga menggunakan $RMSE\%$ sebagai ukuran tambahan untuk menunjukkan besar kesalahan prediksi dalam bentuk persentase. $RMSE\%$ dihitung dengan membandingkan nilai RMSE terhadap rata-rata nilai aktual pada data yang dievaluasi. Rumus $RMSE\%$ ditunjukkan pada Persamaan (2.10):

$$RMSE\% = \frac{RMSE}{\bar{y}} \times 100\% \quad (2.10)$$

di mana $RMSE\%$ adalah persentase kesalahan berbasis RMSE, $RMSE$ adalah nilai kesalahan prediksi, dan $\bar{y}$ adalah rata-rata nilai aktual pada data yang dievaluasi.

Akurasi estimasi dalam penelitian ini dihitung sebagai ukuran tambahan berdasarkan nilai $RMSE\%$. Rumus akurasi estimasi ditunjukkan pada Persamaan (2.11):

$$\text{Akurasi Estimasi} = 100\% - RMSE\% \quad (2.11)$$

Akurasi estimasi pada penelitian ini tidak dimaknai sebagai akurasi klasifikasi, melainkan sebagai ukuran tambahan untuk menunjukkan tingkat kedekatan hasil prediksi terhadap nilai aktual berdasarkan $RMSE\%$.

Pada penelitian ini, evaluasi dilakukan pada dua bagian. Evaluasi pertama dilakukan pada data uji historis untuk membandingkan kinerja model berdasarkan 200 data uji. Evaluasi kedua dilakukan pada hasil *forecasting* 7 hari kerja setelah data terakhir. Nilai aktual pada periode *forecasting* digunakan sebagai pembanding setelah proses *forecasting* selesai dilakukan, sehingga RMSE, $RMSE\%$, dan akurasi estimasi juga dapat dihitung pada periode tersebut.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A