

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Dalam penelitian mengenai prediksi harga saham menggunakan algoritma Support Vector Regression (SVR) pada PT Kalbe Farma Tbk, peneliti mendapatkan dan menggunakan sumber informasi untuk acuan diambil dari jurnal penelitian sebelumnya. Penelitian sebelumnya dengan algoritma dan persoalan yang sama, digunakan sebagai pendukung untuk memperkuat isi penelitian dan menambah literatur, berikut adalah penelitian terdahulu sebagai berikut :

Penelitian pertama, berjudul Analisis Kinerja Algoritma Support Vector Machine (SVM) Guna Pengambilan Keputusan Beli/Jual Pada Saham PT Elnusa Tbk. (ELSA), oleh Elfa Aufa Nida, Agung Budi Prasetyo, dan Risma Septiana menjelaskan diperlukannya analisis sebelum melakukan jual beli saham. Hasil uji saham dalam penelitian dari peneliti ini sebesar 92,34% [10].

Penelitian selanjutnya, berjudul Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode Support Vector Machine, oleh Widya Rizka Ulul Fadilah, Dewi Agfiannisa, dan Yufis Azhar menjelaskan bahwa pada penelitian ini menggunakan perbandingan data *testing* 20% dan 80% *training* dari keseluruhan dataset pada pengujian dengan algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN). Hasil akurasi yang diperoleh pada pengujian algoritma SVM sebesar 0,9641 dan RMSE sebesar 0,0932. Diperoleh hasil kesimpulan bahwa hasil prediksi algoritma SVM lebih tinggi dibandingkan algoritma KNN [21].

Penelitian berikutnya oleh Eka Patriya dengan judul Implementasi Support Vector Machine Pada Prediksi Harga Saham Gabungan (IHSG), menjelaskan penggunaan data input dengan perbandingan data *testing* 80% *training* dari Januari 2014 hingga Maret 2018 dan 20% *testing* dari April 2018 hingga April 2019. Data yang telah dibagi kemudian masuk tahap pemodelan algoritma SVR yang juga menggunakan tahap *hyperparameter* tuning dengan digunakan *hyperparameter* C, epsilon, dan gamma. Hasil yang diperoleh dari *training* dan *testing* adalah MAPE dengan hasil 0,211% dan 0,251%, dan RMSE dengan hasil 14,334 dan 20,281 [22].

Penelitian selanjutnya, oleh Vina Putri Damartya, Deni Saepudin, dan Putu Harry Gunawan berjudul Optimasi Portofolio Saham LQ45 dengan Mempertimbangkan Prediksi *Return* Menggunakan Metode Support Vector Regression (SVR) menjelaskan *history* saham cenderung berubah, diperlukan untuk memperoleh keuntungan yang maksimal. Hasil penelitiannya diperoleh prediksi SVR dengan RMSE sebesar 0,34973 [23].

Penelitian berikutnya, dilakukan oleh Ratih Febrilia Tri Wulandari dan Dian Anubhakti dengan judul Implementasi Algoritma Support Vector Machine (SVM) dalam Memprediksi Harga Saham PT. Garuda Indonesia Tbk memaparkan hasil akurasi prediksi perhitungan harga saham PT. Garuda Indonesia. Hasil yang diperoleh pada penelitian ini adalah akurasi sebesar 0,545 dari 520 data [14].

Penelitian selanjutnya, dilakukan A. Hermawan, I W. Mangku, dan N.K.K. Ardana berjudul Analisis Support Vector Regression Dengan Algoritma *Grid Search* Untuk Memprediksi Harga Saham, menjelaskan ketidakpastian risiko investasi saham, yang mana investasi banyak bermanfaat dalam perekonomian. Hasil penelitiannya diperoleh saham untuk INDF dan MYOR dengan hasil

MAPE sebesar 5,570% dan R2 79,9%, serta MAPE sebesar 2,954% dan R2 96% [24].

Penelitian berikutnya, dilakukan oleh Decesa Erla Krislianti, Etik Zukhronah, dan Yuliana Susanti berjudul Peramalan Harga Minyak Menggunakan Support Vector Regression menjelaskan harga minyak mentah yang tidak menentu berakibat pada perkembangan ekonomi di Indonesia. Hasil SVR terbaik dengan kernel sigmoid diperoleh nilai $p = 10$ dan parameter $C = 1000$, $\varepsilon = 0,0001$, $\gamma = 0,01$, serta MAPE sebesar 3,618% [25].

Penelitian selanjutnya, dilakukan I Made Gananta, I Nyoman Purnama, dan Ketut Queena Fredlina berjudul Optimasi Prediksi Harga Emas Dengan Metode Support Vector Regression (SVR) Menggunakan Algoritma *Grid Search*. menjelaskan emas sebagai contoh investasi dinilai seimbang dan terkenal, walaupun dengan adanya pandemi. Hasil penelitiannya adalah SVR dengan grid search dan tanpa *grid search* dengan MAPE sebesar 3,37% dan 4,35% [26].

Penelitian berikutnya, dilakukan oleh Aanisah Waliy Ishlah, Sudarno, Puspita Kartikasari berjudul Implementasi *Gridsearchcv* Pada Support Vector Regression (SVR) Untuk Peramalan Harga Saham menjelaskan bahwa penelitian ini memakai data saham harian PT Anabatic Technologies Tbk mulai Januari 2017 hingga 31 Desember 2021 kemudian menggunakan metode *Grid Search* dengan kernel linear untuk melakukan analisis dengan algoritma SVR. Hasil yang diperoleh dari model SVR adalah MSE sebesar 0,001237173; R-Square sebesar 0,9206643; and MAPE sebesar 0,1167301 pada model penggunaan linear kernel dengan cost (C) = 1 dan epsilon = 0,11 [27].

Penelitian terakhir, dilakukan Lala Nur Faiza dan Dina Agustina berjudul Aplikasi *Machine Learning* dalam Prediksi Harga Saham Jakarta Islamic Index (JII) Menggunakan Metode Support

Vector Regression menjelaskan data saham JII yang digunakan mulai Desember 2021 hingga November 2022 untuk prediksi algoritma SVR pada 2 kernel yaitu kernel Linear dan kernel RBF dan membagi data saham untuk pemodelan SVR 70 banding 30. Hasil RMSE setelah dilakukan penelitian dari saham ANTM, BRIS, dan BRPT 0,0004 [28].

Tabel 2.1 Penelitian Terdahulu

No	Judul	Penulis	Masalah	Hasil Penelitian
1.	Analisis Kinerja Algoritma Support Vector Machine (SVM) Guna Pengambilan Keputusan Beli/Jual Pada Saham PT Elnusa Tbk. (ELSA) [10].	Elfa Aufa Nida, Agung Budi Prasetyo, dan Risma Septiana.	Diperlukannya analisis sebelum melakukan jual beli saham.	Diperoleh hasil uji saham dalam penelitian dari PT Elnusa Tbk. (ELSA) sebesar 92,34%.
2.	Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode Support Vector Machine [21].	Widya Rizka Ulul Fadilah, Dewi Agfiannisa, dan Yufis Azhar.	Perubahan harga saham yang terus berevolusi, yang disimpan dalam history harga saham dapat digunakan untuk prediksi. Hal ini membantu investor memantau pola harga saham yang bergerak.	Hasil akurasi dari uji saham dengan algoritma SVM diperoleh sebesar 0,9641 dan RMSE sebesar 0,0932. Hasil akurasi lainnya dengan algoritma KNN diperoleh sebesar 0,945 dan RMSE sebesar 0,1162. Diperoleh kesimpulan, hasil akurasi algoritma SVM lebih tinggi dan <i>error</i> lebih rendah dari algoritma KNN.
3.	Implementasi Support Vector Machine Pada Prediksi Harga Saham Gabungan (IHSG) [22]	Eka Patriya.	Transaksi jual beli saham mengalami harga naik turun yang tidak menentu dengan risiko yang besar.	Hasil untuk RMSE dari penelitian ini dari tahap <i>training</i> dan <i>testing</i> adalah sebesar 14,334 dan 20,281, MAPE untuk <i>training</i> dan <i>testing</i> sebesar 0,211% dan 0,251%.
4.	Optimasi Portofolio Saham LQ45 dengan Mempertimbangkan Prediksi Return Menggunakan Metode	Vina Putri Damartya, Deni Saepudin, dan Putu Harry	<i>History</i> saham cenderung berubah, diperlukan untuk memperoleh keuntungan maksimal.	Hasil SVR dengan RMSE sebesar 0,34973.

No	Judul	Penulis	Masalah	Hasil Penelitian
	Support Vector Regression (SVR) [23].	Gunawan.		
5.	Implementasi Algoritma Support Vector Machine (SVM) dalam Memprediksi Harga Saham PT. Garuda Indonesia TBK [14].	Ratih Febrilia Tri Wulandari, dan Dian Anubhakti.	Saham memiliki risiko yang tinggi, harga saham yang bergerak tidak dengan garis lurus.	Penelitian melewati 2 tahap sebelum prediksi yaitu tahap <i>testing</i> dan <i>training</i> dengan akurasi prediksi sebesar 0,545 dari 520 dataset.
6.	Analisis Support Vector Regression Dengan Algoritma Grid Search Untuk Memprediksi Harga Saham [24].	A. Hermawan, I W. Mangku, dan N.K.K. Ardana.	Ketidakpastian dari risiko investasi saham, yang mana investasi bermanfaat banyak dalam perekonomian.	Saham untuk INDF dan MYOR diperoleh hasil MAPE sebesar 5,570% dan R^2 79,9%, serta MAPE sebesar 2,954% dan R^2 96%.
7.	Peramalan Harga Minyak Menggunakan Support Vector Regression [25].	Decesa Erla Krislianti, Etik Zukhronah, dan Yuliana Susanti.	Harga minyak mentah yang tidak menentu berakibat pada perkembangan ekonomi di Indonesia.	Hasil SVR terbaik dengan kernel sigmoid diperoleh nilai sebesar $p=10$ dan parameter $C=1000$, $\epsilon=0,0001$, $\gamma=0,01$, serta MAPE sebesar 3,618%
8.	Optimasi Prediksi Harga Emas Dengan Metode Support Vector Regression (SVR) Menggunakan Algoritma Grid Search [26].	I Made Gananta, I Nyoman Purnama, dan Ketut Queena Fredlina.	Emas sebagai contoh investasi dinilai seimbang dan terkenal, walaupun dengan adanya COVID-19.	Hasil SVR dengan <i>grid search</i> dan tanpa <i>grid search</i> dengan MAPE sebesar 3,37% dan 4,35%.
9.	Implementasi Gridsearchcv Pada Support Vector Regression (SVR) Untuk Peramalan Harga Saham [27].	Aanisah Waliy Ishlah, Sudarno, dan Puspita Kartikasari.	Harga saham yang sempat naik dan turun yang berarti bersifat fluktuatif.	Hasil SVR dengan kernel linear dengan <i>hyperparameter</i> berupa $C=1$ and $\epsilon=0,11$ diperoleh MSE 0,001237173; R^2 0,9206643; dan SMAPE 0,1167301.
10.	Aplikasi <i>Machine Learning</i>	Lala Nur Faiza,	Ketidakpastian dari risiko	Hasil RMSE setelah dilakukan

No	Judul	Penulis	Masalah	Hasil Penelitian
	dalam Prediksi Harga Saham Jakarta Islamic Index (JII) Menggunakan Metode Support Vector Regression [28]	Dina Agustina.	investasi saham.	penelitian dari saham ANTM, BRIS, dan BRPT 0,0004.

Penelitian sebelumnya pada Tabel 2.1, terdapat tujuan penelitian yang sama dengan perbedaan penelitian terdapat pada tujuan saham yang diprediksi, penggunaan *GridSearch* [27], dan besar pembagian data untuk model SVR. Penelitian sebelumnya pada nomor 3 oleh Eka Patriya [22], dijadikan sebagai rujukan utama dalam penelitian prediksi harga saham menggunakan algoritma Support Vector Regression (SVR) pada PT. Kalbe Farma Tbk.

2.2. Landasan Teori

2.2.1 Saham

Saham adalah bukti yang dimiliki oleh seseorang yang memiliki kepemilikan pada perusahaan, semakin besar nilai saham yang dimiliki membuat tingkat kendali kuasanya menjadi semakin kuat dan berpengaruh pada perusahaan [29]. Saham adalah contoh aset investasi terkenal dari pasar modal dan aset investasi. Harga suatu saham bersifat naik turun (fluktuatif) dipengaruhi oleh banyak faktor, diantaranya ada faktor ekonomi, kinerja perusahaan, dan politik [17][30][31]. Investasi saham memiliki hasil keuntungan menarik, namun ikut diimbangi dengan risiko yang sama besar (*high risk high return*) [17][31].

2.2.2 Data Mining

Data mining adalah rangkaian langkah dimulainya persiapan data, pemakaian algoritma dan menampilkan hasil proses secara visual. *Data mining* sesuai artinya yaitu menggali dan menambang digunakan untuk mendapatkan data dari kumpulan data yang berjumlah besar. Istilah yang singkat pengertian dari data mining adalah proses ekstraksi data berupa angka, informasi, pola, dan lainnya yang penting untuk diambil kesimpulan [8][32]. *Data mining* dalam tahapan prosesnya mengolah data dari data yang memiliki ukuran besar untuk mencari pola (*trend*) sesuai dengan tujuan awal *data mining*.

Tahap paling awal yang diperlukan dalam *data mining* adalah mencari dan mempersiapkan data [8].

Tantangan yang kerap dialami dalam melakukan teknik penambangan data adalah awal persiapan data sebelum menuju proses data *mining* karena semakin besar data maka sebanding dengan waktu yang diperlukan untuk menguraikan, menjabarkan dan analisis data. Tantangan dalam *data mining* memiliki keuntungan yang besar sesuai dengan manfaatnya untuk menghasilkan pencarian informasi yang lebih berbobot, lebih penting, dan lebih koheren dari data yang berjumlah cenderung besar. Terdapat 3 algoritma pada *data mining*, yaitu *supervised learning*, *unsupervised learning*, dan *semi-supervised* [8].

Algoritma *supervised learning* digunakan pada bentuk data yang memiliki kelas atau label, selain itu algoritma ini memerlukan proses pembelajaran dari data pelatihan untuk melakukan pelatihan terus menerus hingga diperoleh hasil akurasi yang dianggap sesuai. Algoritma *unsupervised learning*, algoritma yang bekerja dengan data belum diketahui kelas atau labelnya, dan banyak digunakan dalam pengelompokan data berdasarkan kesamaan dan kesesuaian, selain itu algoritma ini tidak memerlukan proses pembelajaran seperti algoritma *supervised learning*, dan algoritma bekerja untuk menemukan struktur data. Selanjutnya untuk algoritma *semi supervised learning*, sebagian data dengan label atau kelas dan sebagian lainnya belum diketahui label atau kelasnya, dan bekerja dengan cara di antara kedua algoritma sebelumnya.

2.2.3 Prediksi

Prediksi adalah usaha dengan dugaan untuk estimasi dari kejadian yang muncul dimasa depan namun belum terjadi di masa depan dengan mencoba menerawang, meramal, dan

perkiraan [6][7]. Tujuan dilakukan prediksi adalah mendapatkan informasi tentang yang terjadi di masa depan dengan kemungkinan yang besar [33]. Kebanyakan dalam melakukan prediksi menggunakan data bersifat kuantitatif. Hasil yang diperoleh dari prediksi tidak wajib menghasilkan output yang tepat atau pasti dengan hal yang sesungguhnya. Prediksi dibedakan berdasarkan kurun waktu terjadinya menjadi 3, yaitu prediksi dalam kurun waktu pendek, prediksi dalam kurun waktu menengah, dan prediksi dalam kurun waktu panjang. Prediksi jangka pendek dilakukan dengan memperhatikan pola data yang memiliki jangka waktu pendek yang berubah terhadap faktor yang membentuk pola data. Sementara prediksi jangka menengah dan prediksi jangka panjang banyak digunakan untuk membuat rencana strategis, menyiapkan perluasan usaha, dan ketersediaan kebutuhan. Prediksi jangka panjang bermanfaat untuk menjamin jumlah ketersediaan kebutuhan di masa depan.



Gambar 2.1 Riwayat saham dari tahun 2016 [5]

Prediksi adalah contoh dari sistem pengolahan dalam *machine learning* yang memerlukan tahap pelatihan terlebih dahulu (*supervised learning*) yang masuk dalam klasifikasi. Beberapa algoritma yang digunakan dalam melakukan prediksi

diantaranya adalah K-Nearest Neighbor (KNN) [8], Linear Regression (LR) [34], Long Short Term Memory (LSTM) [9], dan Support Vector Machine (SVM) [10].

2.2.4 Preprocessing

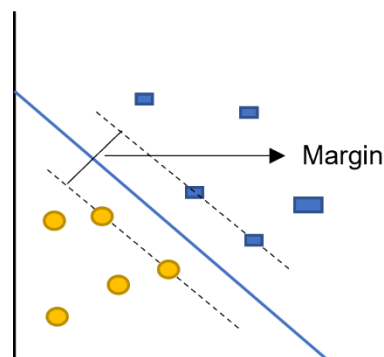
Tahap *preprocessing* adalah tahapan persiapan dan proses pembersihan data yang telah diperoleh, contohnya penyaringan data dari *noise* atau mengecek format data [32]. Pembersihan data yang diambil pada tahapan ini dilakukan pada kondisi data seperti, *missing value*, *double data*, dan *outlier* pada data (data yang terpercil atau memiliki jarak jauh dengan data lainnya) [8]. Tahap *preprocessing* menjadi langkah penting dalam melakukan proses *data mining*, penentuan kualitas, dan keandalan dari data dan model yang diolah [10]. Tahap *preprocessing* dilakukan sebelum data diolah dengan algoritma untuk dijadikan model prediksi. Tahap *preprocessing* berfungsi memperbaharui data belum siap pakai atau berantakan menjadi data yang telah tersusun baik dan efisien dengan membersihkan data, menyeleksi kembali data yang tidak sesuai sebelum menuju tahap selanjutnya.

2.2.5 Support Vector Regression (SVR)

Support Vector Machine (SVM) adalah algoritma yang diperlukan untuk klasifikasi yang terstruktur karena perlu adanya data latih (*training*) untuk melakukan pembelajaran [35]. Algoritma SVM dikenalkan oleh Vapnik tahun 1992 dalam bidang materi pattern recognition (pengenalan pola) [1][36]. Algoritma SVM menjadi salah satu algoritma untuk melakukan prediksi, yaitu untuk persoalan atau permasalahan klasifikasi dan regresi [36][37][38]. Selain itu, algoritma SVM bertugas pada 2 jenis pengklasifikasian data linear maupun data non-linear [8]. Aturan dalam model algoritma SVM menggunakan

tipe klasifikasi linear dan digunakan untuk data non-linear dengan bantuan kernel [39][40]. Algoritma SVM dalam melakukan prediksi tidak dipengaruhi oleh jumlah data *training* dan proses *training*.

Pada Gambar 2.2 memperlihatkan ada beberapa gambar kotak dan lingkaran kecil yang dimisalkan sebagai *pattern*, yang mana untuk gambar kotak sebagai *pattern* positif dan gambar lingkaran kecil sebagai *pattern* negatif.



Gambar 2.2 Hyperplane pada algoritma SVM [41]

Hyperplane untuk klasifikasi SVM dirumuskan pada persamaan (1) sebagai berikut :

$$f(x) = w^T x + b \quad (2.1)$$

Konsep dasar aturan algoritma SVM dirumuskan pada persamaan (2) dan persamaan (3) sebagai berikut :

$$\min \frac{1}{2} |W|^2 \quad (2.2)$$

$$\text{s.t } y_i (X_i \cdot w + b) - 1 \geq 0 \quad (2.3)$$

Keterangan :

$(X_i \cdot w + b) \geq 1$ = digunakan untuk kelas 1.

$(X_i \cdot w + b) \leq -1$ = digunakan untuk kelas 2.

X_i = dataset.

y_i = hasil keluaran dari data X_i dan w .

b = parameter yang dicari.

Algoritma Support Vector Regression (SVR) menjadi bagian dari algoritma SVM dengan tugas khusus untuk persoalan fungsi regresi dengan input linear dan variabel, karena memiliki tugas pengklasifikasian yang berbeda [42]. Algoritma SVR dalam teori diciptakan dan memiliki cara kerja untuk menyelesaikan persoalan regresi dijelaskan dengan ada dua kelas berupa untuk dipisahkan menjadi kelompok data dengan garis (*hyperplane*) terbaik [1][40]. Kelas yang dimaksud berupa kelas di luar garis *hyperplane* dan di dalam *hyperplane*. *Hyperplane* adalah langkah untuk fungsi regresi menentukan keputusan hasil prediksi dengan memaksimalkan *margin*, mencari fungsi regresi antara fitur data dengan nilai target yang ada di dalam *hyperplane*. Validasi yang digunakan pada algoritma SVR diantaranya *Mean Square Error* (MSE), *Root Mean Square Error* (RMSE), dan *R-Square*.

MSE adalah metode evaluasi dalam ilmu data yang menghitung nilai *mean* dari perbedaan nilai absolut antara nilai prediksi pada Y dengan nilai aktual pada Y' , berapa nilai mean kesalahan kuadrat dalam prediksi. Hasil MSE pada suatu model dikatakan baik apabila nilai MSE yang diperoleh semakin kecil [43].

$$\text{MSE} = \frac{(Y' - Y)^2}{n} \quad (2.4)$$

RMSE adalah metode evaluasi dalam ilmu data yang menghitung nilai *mean* dari perbedaan nilai absolut antara nilai prediksi pada Y dengan nilai aktual pada Y' , dengan beda rata-rata kuadrat antara nilai prediksi dan nilai aktual selanjutnya dihitung prediksi nilai akar kuadrat. Hasil RMSE pada suatu memperoleh hasil baik apabila nilai RMSE memperoleh nilai yang semakin kecil. RMSE dijelaskan menggunakan rumus perhitungan sebagai berikut [43].

$$\text{RMSE} = \sqrt{\frac{(Y' - Y)^2}{n}} \quad (2.5)$$

R-Square atau koefisien determinasi adalah metode yang menghitung kesesuaian dari model algoritma antara garis regresi dengan kecocokan data dari selisih kuadrat dengan jumlah total kuadrat. Hasil *R-Square* pada suatu model dikatakan baik apabila hasil *R-Square* semakin besar. *R-Square* dijelaskan menggunakan rumus perhitungan sebagai berikut [27].

$$R\text{-Square} = \frac{JKR}{JKT} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.6)$$

Keterangan *Mean Square Error* (MSE), *Root Mean Square Error* (RMSE), dan *R-Square* untuk pengujian SVR dirumuskan pada persamaan (2.4), (2.5), dan (2.6) sebagai berikut :

Y = nilai yang diprediksi.

Y' = nilai sebenarnya.

n = jumlah data.