

BAB III

METODE PENELITIAN

3.1. Objek Penelitian

Sasaran pada penelitian ini adalah data saham PT Telekomunikasi Indonesia dengan total 1.257 *record* data dengan tujuh atribut yang diperoleh dari *Website Yahoo Finance*. Data saham PT Telekomunikasi Indonesia tersebut akan diprediksi hingga menjadi salah satu pengetahuan baru mengenai saham sebelum memutuskan untuk menanam saham.

3.2. Peralatan Penelitian

Dalam melakukan penelitian, peralatan yang digunakan yaitu perangkat keras berupa spesifikasi laptop yang digunakan dan perangkat lunak.

Tabel 3.1. Perangkat Penelitian

Perangkat Keras	<i>Prosesor AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx 2.10 GHz</i>
	<i>RAM 8,00 GB (5,92 GB usable)</i>
	<i>System type 64-bit operating system, x64-based processor</i>
Perangkat Lunak	Sistem Operasi Windows 11
	Microsoft Word 2019, untuk penyusunan laporan.
	Micorosoft Excel 2019, untuk perhitungan model. <i>Google Colab</i> , digunakan untuk pengujian dengan menerapkan bahasa pemrograman <i>python</i>

3.3. Jenis dan Sumber Data

Jenis data yang digunakan pada penelitian tugas akhir ini yaitu data kuantitatif karena data tersebut bersifat angka (numerik) yang dapat diolah menggunakan metode statistika. Data saham PT Telekomunikasi Indonesia diunduh dari *Website Yahoo Finance* pada tanggal 07 Desember 2022 dengan rekaman data dari tanggal 01 Desember 2017 hingga 07 Desember 2022. Data diambil dari selama lima tahun terakhir karena dengan lima tahun terakhir diharapkan mampu mereflesikan kejadian terbaru karena pada masa periode tersebut terdapat kejadian-kejadian yang

Tabel 3.2. Deskripsi Atribut Saham

3.4. Lokasi dan Waktu Penelitian

Tabel 3.3. Waktu Penelitian

No	Keterangan	Nov 2022				Des 2022				Jan 2023				Feb 2023				Mar 2023			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Studi Literatur																				
2	Pengumpulan Data																				
3	Data <i>Pra-Processing</i>																				

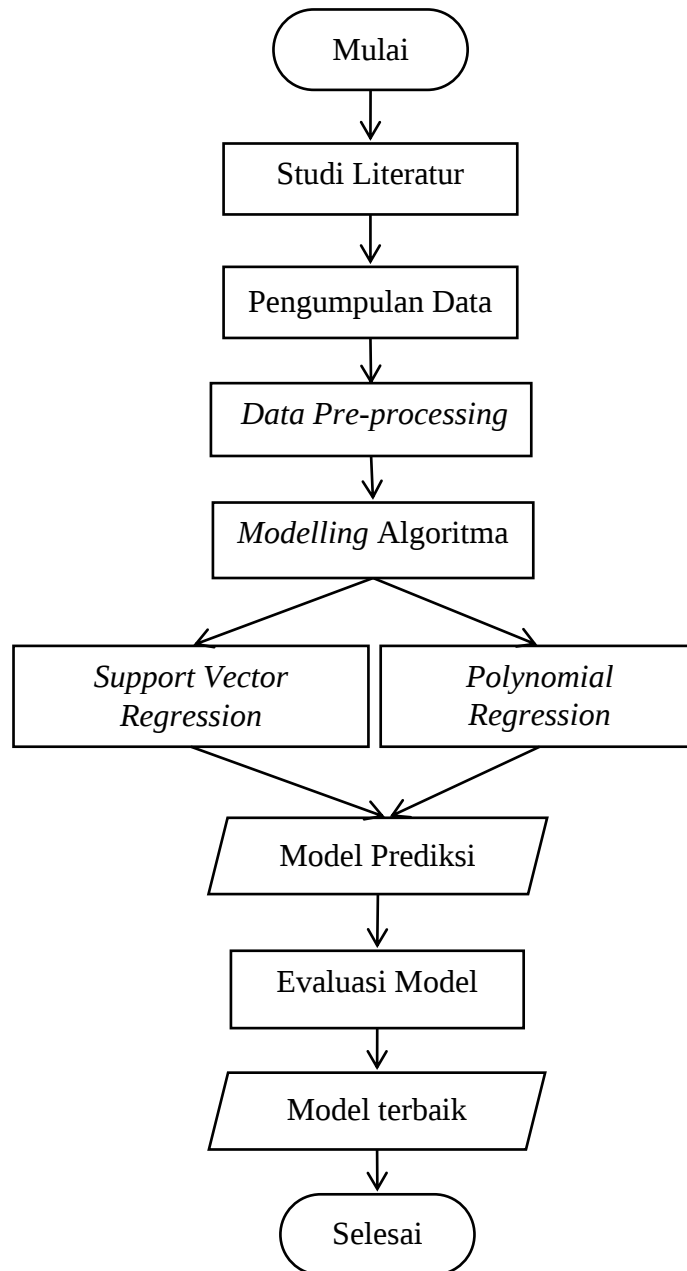
4	Modelling Algoritma		
5	Evaluasi Model		
6	Model Terbaik		

3.5. Analisis Data

Analisis data dilakukan karena fluktuasi harga saham yang tidak menentu, sehingga perlu adanya prediksi khususnya pada harga saham PT Telekomunikasi Indonesia untuk dijadikan sebagai bahan pertimbangan atau sebagai referensi masyarakat dalam keputusan berinvestasi saham. Data saham yang diperoleh akan dibersihkan atau diawali dengan *data pre-processing* hingga data siap digunakan. Data saham tersebut akan diprediksi menggunakan algoritma *Support Vector Regression* dan *Polynomial Regression* dengan teknik *data mining*. Output yang dihasilkan berupa grafik prediksi dan nilai *error* dari setiap algoritma yang digunakan untuk menunjukkan algoritma mana yang lebih cocok digunakan pada penelitian ini. Selain itu, *output* dari penelitian yang dilakukan diharapkan mampu menjadi bahan pertimbangan atau referensi dalam mengambil keputusan berinvestasi saham.

3.6. Prosedur Penelitian

Prosedur penelitian dalam memprediksi harga saham PT Telekomunikasi Indonesia terdiri dari beberapa tahapan yang meliputi pengumpulan data, *data pre-processing*, *modelling* algoritma menggunakan *Support Vector Regression* dan Algoritma *Polynomial Regression*, dan pengujian model. Prosedur penelitian ditunjukkan dalam bentuk diagram alir pada Gambar 3.1.



Gambar 3.1. Prosedur Penelitian

3.6.1. Pengumpulan Data

Teknik pengumpulan data pada penelitian ini tidak dilakukan secara langsung ke lapangan, melainkan diperoleh dari situs *website Yahoo finance*. Data yang diperoleh adalah data saham PT. Telekomunikasi Indonesia Tbk memiliki jumlah data sebanyak 1.257 *record data* dengan tujuh variabel. Data tersebut diunduh pada tanggal 07 Desember 2022 dan termasuk data *time series* dengan rentang data

dimulai dari tanggal 01 Desember 2017 sampai 07 Desember 2022. Data tersebut akan diolah dan dianalisis hingga menghasilkan informasi yang berguna dalam rangka mencapai tujuan penelitian. Adapun data yang digunakan ditunjukkan pada Tabel 3.4.

Tabel 3.4. Sampel Dataset

No	Date	Open	High	Low	Close	Adj Close	Volume
1	2017-12-01	4150.00000 0	4150.00000 0	4150.00000 0	4150.00000 0	3545.8889 16	0
2	2017-12-04	4180.00000 0	4230.00000 0	4160.00000 0	4200.00000 0	3588.6108 40	7804050 0
3	2017-12-05	4210.00000 0	4220.00000 0	4180.00000 0	4200.00000 0	3588.6108 40	9650480 0
4	2017-12-06	4160.00000 0	4210.00000 0	4160.00000 0	4200.00000 0	3588.6108 40	9212650 0
...							
1257	2022-12-06	3640.00000 0	3720.00000 0	3580.00000 0	3600.00000 0	3600.0000 00	4164687 00

3.6.2. Data Pre-Processing

Setelah pengumpulan data, tahapan selanjutnya adalah *data pre-processing* terhadap dataset, di mana dataset tersebut akan diolah terlebih dahulu sampai data siap digunakan untuk proses pemodelan menggunakan algoritma. Proses *data pre-processing* bertujuan untuk mengubah data mentah menjadi data yang lebih mudah untuk dipahami. *Data pre-processing* dilakukan untuk meminimalisir kesalahan data, seperti mengatasi *missing value* atau data yang tidak beraturan (konsisten). Adapun tahapan *data pre-processing* yang akan digunakan pada penelitian ini yaitu *Data Reduction* dan *Data Cleaning*.

1. Data Reduction

Data Reduction merupakan pengurangan sampel data dari dataset yang akan dipakai. Pada penelitian ini, tidak semua atribut digunakan, melainkan menggunakan atribut atau fitur yang dianggap relevan pada penelitian ini. Penelitian yang dilakukan akan memprediksi harga tertutup (*close*) saham berdasarkan hari dengan variabel *x*, maka atribut yang digunakan yaitu kolom *Date*, *Open*, dan *Close*. Data sebelum seleksi data ditunjukkan pada Tabel 3.5.

Tabel 3.5. Sebelum *Data Reduction*

<i>Date</i>	<i>Open</i>	<i>High</i>	<i>Low</i>	<i>Close</i>	<i>Adj Close</i>	<i>Volume</i>
2017-12-01	4150.000000	4150.000000	4150.000000	4150.000000	3545.888916	0
2017-12-04	4180.000000	4230.000000	4160.000000	4200.000000	3588.610840	78040500
2017-12-05	4210.000000	4220.000000	4180.000000	4200.000000	3588.610840	96504800

Sebelum *data reduction* terdapat tujuh atribut, diantaranya yaitu *Date*, *Open*, *High*, *Low*, *Close*, *Adj Close*, dan *Volume*. Atribut yang akan digunakan pada penelitian ini sebanyak tiga atribut, yaitu *Date*, *Open* dan *Close*, maka atribut *High*, *Low*, *Adj Close*, dan *Volume* akan dihapus. Contoh setelah *data reduction* ditunjukkan pada Tabel 3.6.

Tabel 3.6. Setelah *Data Reduction*

<i>Date</i>	<i>Open</i>	<i>Close</i>
2017-12-01	4.150.000.000	4.150.000.000
2017-12-04	4.180.000.000	4.200.000.000
2017-12-05	4.210.000.000	4.200.000.000
2017-12-06	4.160.000.000	4.200.000.000
2017-12-07	4.200.000.000	4.200.000.000

1. *Data Cleaning*

Setelah *data reduction* atau pengurangan atribut yang digunakan, selanjutnya yaitu proses *data cleaning*, proses ini akan mengatasi permasalahan data seperti *missing value* pada data, *outlier* atau menghilangkan data yang tidak efisien dan mengandung kesalahan. (Rohana, 2021). Proses *data cleaning* pada penelitian ini bertujuan untuk membuat data menjadi lebih bersih dan dapat dilanjutkan ke proses selanjutnya. Berikut adalah contoh data *missing value* ditunjukkan pada Tabel 3.7.

Tabel 3.7. Sebelum *Data Cleaning*

<i>Date</i>	<i>Open</i>	<i>Close</i>
2019-06-17	3.950.000.000	3.900.000.000
2019-06-18	3.890.000.000	3.980.000.000
2019-06-19	<i>null</i>	<i>null</i>

2019-06-20	4.110.000.000	4.040.000.000
2019-06-21	4.010.000.000	4.040.000.000

Dapat dilihat data harga saham pada tanggal 19 Juni 2019 terdapat nilai “null” pada kolom *Open* dan *Close*. Pada penelitian ini cara untuk mengatasi *missing value* adalah dengan menghapus data tersebut. Data sebelumnya berjumlah 1.257 *data record*, karena transaksi pada tanggal 19 Juni 2019 akan dihapus, maka jumlah data yang akan digunakan menjadi 1.256 *data record*. Berikut ilustrasi jika *missing value* pada atribut *Open* dan *Close* telah dihapus ditunjukkan pada Tabel 3.8.

Tabel 3.8. Setelah *Data Cleaning*

<i>Date</i>	<i>Open</i>	<i>Close</i>
2019-06-17	3.950.000.000	3.900.000.000
2019-06-18	3.890.000.000	3.980.000.000
2019-06-20	4.110.000.000	4.040.000.000
2019-06-21	4.010.000.000	4.040.000.000

1. *Type Casting*

Type casting bertujuan untuk merubah suatu tipe data menjadi tipe data lain yang dibutuhkan untuk memenuhi kebutuhan penelitian.

3.6.3. *Modelling Algoritma*

Setelah *data pre-preprocessing* selesai, tahap selanjutnya yaitu pemodelan algoritma menggunakan algoritma *Support Vector Regression* dan *Polynomial Regression* menggunakan bahasa pemrograman *python* hingga memperoleh model prediksi. Setelah proses pemodelan algoritma selesai, hasil model prediksi akan divisualisasikan ke dalam grafik untuk melihat pola antara hasil prediksi saham dengan pola data saham yang sebenarnya, sehingga lebih mudah untuk dipahami.

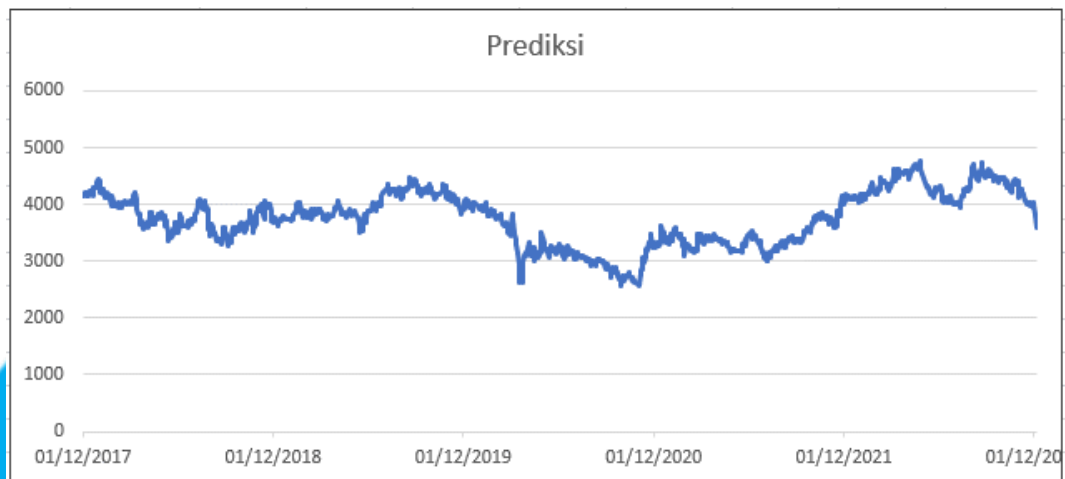
1. Langkah-langkah Algoritma *Support Vector Regression* pada pemrograman *python* diantaranya.

- Import library* SVR pada pemrograman *python*
- Import dataset* yang akan digunakan dengan format *.csv*

- a) *Data pre-processing* yang meliputi penghapusan atribut yang tidak digunakan dan penghapusan nilai *missing value*.
 - b) Pemodelan algoritma *Support Vector Regression* dengan kernel *Polynomial*
 - c) Evaluasi model algoritma *Support Vector Regression* menggunakan RMSE
1. Langkah-langkah Algoritma *Polynomial Regression* pada pemrograman *python* diantaranya.
- a) *Import library LinearRegression* dan *PolynomialFeatures*
 - b) *Import* dataset yang akan digunakan dengan format.csv
 - c) Menginisialisasikan variabel X dan variabel Y
 - d) Pemodelan algoritma *Polynomial Regression* dengan memasukkan fungsi *Linear Regression* terlebih dahulu.
 - e) Menentukan *degree* atau orde *Polynomial*
 - f) *Evaluasi* model algoritma *Support Vector Regression* menggunakan RMSE

3.6.4. Model Prediksi

Setelah rangkaian penerapan algoritma selesai, hasil yang dikeluarkan berupa angka, kemudian akan divisualisasikan ke dalam bentuk digram atau grafik yang sesuai dengan data dengan tujuan agar lebih mudah dalam memahami model prediksi. Hasil visualiasi model prediksi juga akan menunjukkan antara algoritma *Support Vector Regression* atau *Polynomial Regression* yang mendekati dengan data aktualnya. Berikut contoh visualisasi data menggunakan persamaan *Polynomial Regression* yang akan digunakan ditunjukan pada Gambar 3.2.



Gambar 3.2. Contoh Visualisasi Model Prediksi

3.6.5. Evaluasi Model

Setelah memperoleh hasil dari tahapan *modelling*, selanjutnya yaitu melakukan evaluasi untuk mengukur keakuratan antara model prediksi dengan model aktual berdasarkan algoritma yang digunakan. Evaluasi yang digunakan pada penelitian ini yaitu dengan menggunakan *R2 Score* dan *Root Mean Square Error* (RMSE) dan MAPE. *R2 Score* merupakan presentase yang menunjukkan kecocokan nilai variabel independent (y) terhadap variabel dependen (x) pada pengujian regresi dengan rentang 0 sampai 1, nilai *R2-Score* yang baik merupakan nilai yang mendekati nilai 1. Dengan *R2-Score*, hubungan antar variabel yang baik dapat dilihat dengan seberapa banyak variabel independen mampu mempengaruhi variabel dependen. Berikut rumus *R2 Score* yang ditunjukkan pada persamaan (5).

$$R2 - Score = 1 - \frac{SS_R}{SS_T} \quad (5)$$

$$SS_R = \sum_i (Y_i - \hat{Y}_i)^2$$

$$SS_T = \sum_i (Y_i - \bar{Y})^2$$

Keterangan :

SS_R = kuadrat dari selisih hasil prediksi dengan nilai rata-rata data aktual

SS_T = kuadrat dari selisih data aktual dengan rata-rata data aktual

$\bar{Y}$ = rata-rata data aktual

RMSE merupakan penjumlahan kuadrat *error* atau selisih antara nilai aktual dan nilai prediksi, di mana ketika nilai RMSE yang dihasilkan semakin kecil, maka

tingkat kesalahan dari hasil prediksi pun semakin rendah. Dengan kata lain, RMSE menunjukkan seberapa dekat atau jauh antara data aktual dengan prediksi. Rumus RMSE ditunjukkan pada persamaan (5) (Fadilah et al., 2020).

$$RMSE = \sqrt{\sum \frac{(Y' - Y)^2}{n}} \quad (6)$$

Keterangan;

Y' = nilai prediksi

Y = nilai aktual

n = jumlah data

σ = sigma

Selain menggunakan R^2 Score dan RMSE, untuk mengukur performa kedua algoritma pada penelitian ini yaitu menggunakan MAPE (*Mean Absolute Percentage Error*). MAPE biasanya digunakan untuk pengukuran persentase kesalahan antara data aktual dengan hasil prediksi. Hasil prediksi dikategorikan sangat baik apabila nilai MAPE kurang dari 10% dan dikategorikan baik apabila lebih dari 10% namun kurang dari 20%. Berikut rumus MAPE yang ditunjukkan pada persamaan (7) (Cahyono et al., 2019).

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{X_i - F_i}{X_i} \right|}{n} \times 100 \quad (7)$$

Keterangan:

X_i = data asli

F_i = data Prediksi

n = jumlah data

3.6.6. Model Terbaik

Setelah prosedur penelitian selesai dilakukan, selanjutnya adalah merekomendasikan algoritma yang tepat antara algoritma *Support Vector Regression* atau algoritma *Polynomial Regression* dalam memprediksi harga saham PT Telekomunikasi berdasarkan model prediksi yang memiliki nilai RMSE terendah, karena semakin kecil nilai RMSE yang dihasilkan maka tingkat kesalahannya pun semakin rendah.