

## BAB III METODE PENELITIAN

### 3.1. Objek Penelitian

Data saham Bank Rakyat Indonesia dari tanggal 2 Januari 2018 hingga 30 Desember 2022 dengan total objek sebesar 1.235 data dengan tujuh variabel yaitu *date*, *open*, *high*, *low*, *close*, *adj close*, dan *volume*. Data diambil pada *website yahoo finance* yang diakses pada tanggal 2 Januari 2023.

### 3.2. Peralatan Penelitian

Dalam melakukan penelitian, peralatan yang digunakan yaitu perangkat keras berupa spesifikasi laptop yang digunakan dan perangkat lunak seperti Tabel 3.1 berikut.

Tabel 3.1 Peralatan Penelitian

Spesifikasi Perangkat Keras	Processor Intel Core i5 Gen 8
	VGA NVIDIA GeForce MX230
	Resolusi Layar 1920 x 1080 px
	Sistem Operasi Windows 10
Perangkat Lunak	Microsoft Word 2019, untuk penyusunan laporan.
	Micorosoft Excel 2019, untuk perhitungan model.
	Google Colab, digunakan untuk mengkode program.

### 3.3. Jenis dan Sumber Data

Jenis data yang digunakan dalam penelitian tugas akhir adalah data kuantitatif. Data kuantitatif berupa data numerik diolah metode statistik. Sedangkan sumber data saham Bank Rakyat Indonesia yang didapatkan berasal dari *Website Yahoo Finance* dan diakses pada tanggal 2 Januari 2023. Atribut pada data dan penjelasannya ditunjukkan pada Tabel 3.2.

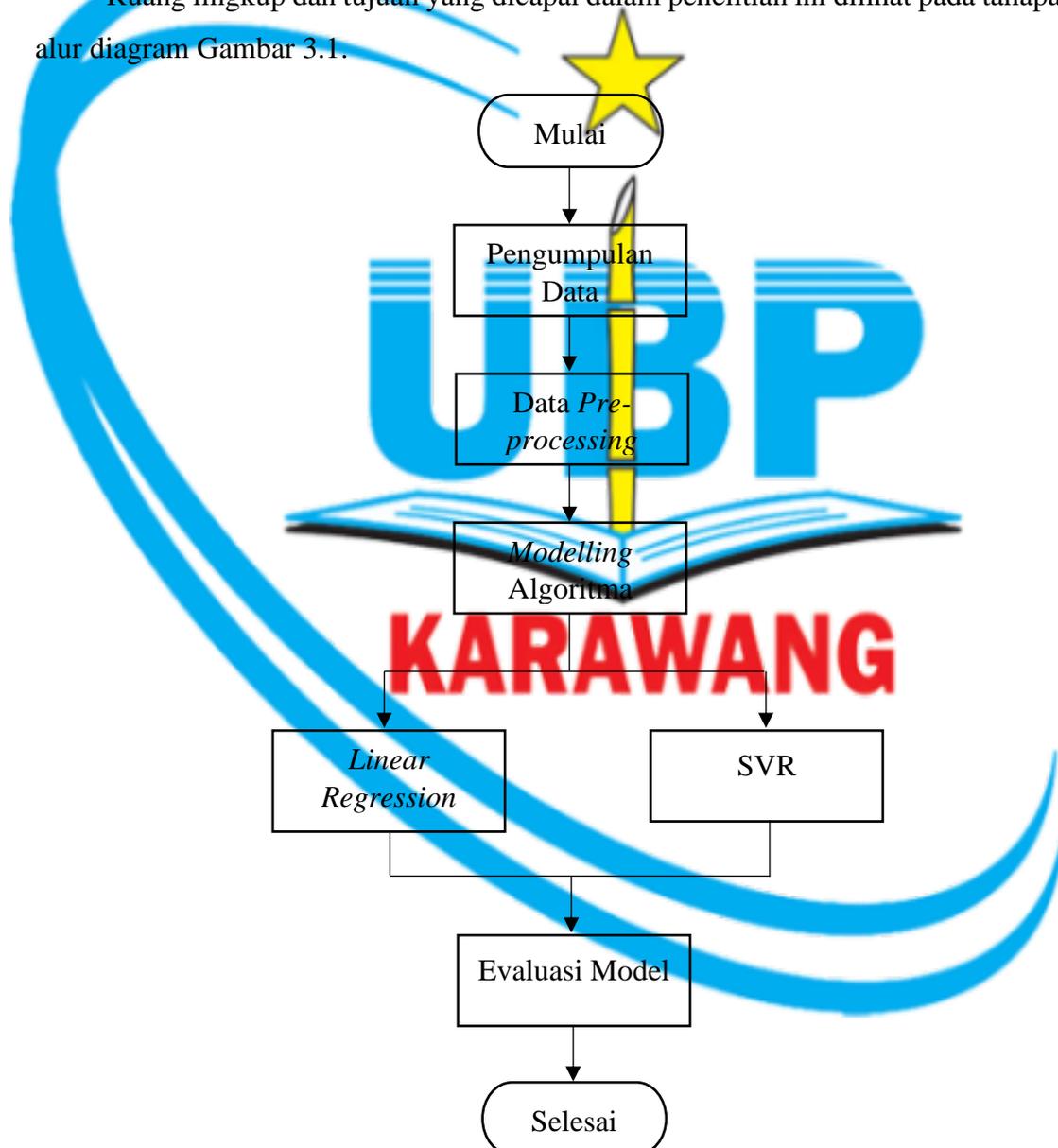


### 3.5. Analisis Data

Penelitian berdasarkan pada analisis data yang dilakukan melalui studi literatur tentang prediksi nilai saham dengan metode regresi. Analisis data menggunakan teknik *data mining* yang dimulai dengan tahap *pre-processing* dan *modeling* sehingga menghasilkan informasi terbaru.

### 3.6. Prosedur Penelitian

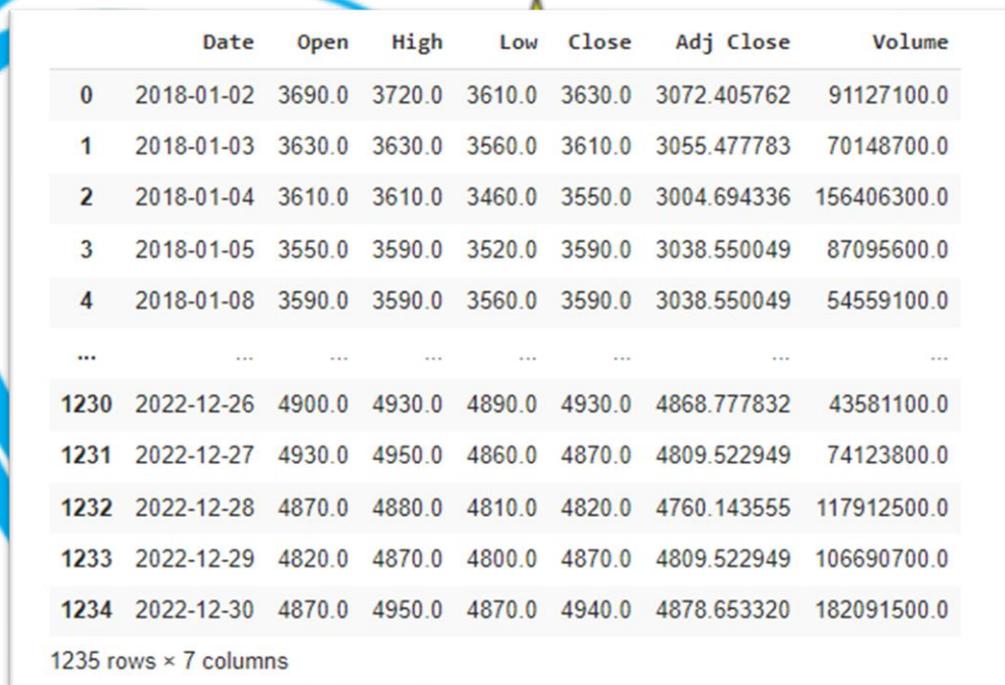
Ruang lingkup dan tujuan yang dicapai dalam penelitian ini dilihat pada tahapan alur diagram Gambar 3.1.



Gambar 3.1 Prosedur Penelitian

### 3.6.1. Pengumpulan Data

Penelitian ini dimulai dengan pengumpulan dataset harga saham Bank Rakyat Indonesia pada *website yahoo finance*. Pada penelitian ini data yang terkumpul sebanyak 1.235 *records* data saham. Pada Gambar 3.2 menunjukkan data saham yang dikumpulkan berisi kolom *date*, *open*, *high*, *low*, *close*, *adj close*, dan *volume* untuk perhitungan model prediksi harga saham menggunakan algoritma *Linear Regression* dan *SVR*.



	Date	Open	High	Low	Close	Adj Close	Volume
0	2018-01-02	3690.0	3720.0	3610.0	3630.0	3072.405762	91127100.0
1	2018-01-03	3630.0	3630.0	3560.0	3610.0	3055.477783	70148700.0
2	2018-01-04	3610.0	3610.0	3460.0	3550.0	3004.694336	156406300.0
3	2018-01-05	3550.0	3590.0	3520.0	3590.0	3038.550049	87095600.0
4	2018-01-08	3590.0	3590.0	3560.0	3590.0	3038.550049	54559100.0
...	...	...	...	...	...	...	...
1230	2022-12-26	4900.0	4930.0	4890.0	4930.0	4868.777832	43581100.0
1231	2022-12-27	4930.0	4950.0	4860.0	4870.0	4809.522949	74123800.0
1232	2022-12-28	4870.0	4880.0	4810.0	4820.0	4760.143555	117912500.0
1233	2022-12-29	4820.0	4870.0	4800.0	4870.0	4809.522949	106690700.0
1234	2022-12-30	4870.0	4950.0	4870.0	4940.0	4878.653320	182091500.0

1235 rows × 7 columns

Gambar 3.2 Dataset Harga Saham

### 3.6.2. Data Pre-processing

Tahapan setelah pengumpulan data adalah pelaksanaan *pre-processing* terhadap data yang telah didapatkan. Sebelum dilakukannya prediksi, data akan melalui proses *pre-processing*. *Pre-processing* data dilakukan dengan tujuan mengubah data mentah menjadi data yang lebih bersih dan efisien. *Pre-processing* data sangat penting karena kesalahan, redundansi, *missing value*, dan data yang tidak konsisten mengurangi keakuratan hasil analisis. Terdapat beberapa tahapan pada *Pre-Processing* sebagai berikut.

#### 1. Reduksi Data

Tahapan pengurangan jumlah data dengan metode *Attribute Subset Selection* yaitu untuk memilih atribut yang paling cocok dengan menghapus atribut-

atribut yang tidak diperlukan dan beberapa atribut yang tidak relevan dengan melihat korelasi antar fitur.

## 2. *Data Cleaning*

Tahapan awal dari *pre-processing*. Pada tahap ini, dataset yang telah dimasukan akan dibersihkan. Tujuan dari *data cleaning* adalah membuat data lebih bersih dan berkualitas sebelum dianalisis dengan cara menghapus data kosong, data tidak lengkap, *outlier data*, dan *missing value*.

## 3. Transformasi Data

Tahapan ini dilakukan untuk mengubah bentuk tipe dari data yang akan digunakan agar dapat dipahami.

### 3.6.3. Modeling Algoritma

Pada tahap ini dilakukan setelah tahap *pre-processing* selesai. Tahapan *modeling* diuji menggunakan algoritma *Linear Regression* dan SVR. Pengujian setiap algoritma menggunakan *Google Colab* kode program *python*. Pada tahapan ini juga dilakukan visualisasi dari hasil uji prediksi dengan menggunakan algoritma *Linear Regression* dan SVR. Langkah-langkah dari algoritma *Linear Regression* dan SVR ditunjukkan pada Tabel 3.4.

Tabel 3.4 Langkah-langkah Algoritma

<i>Linear Regression</i>	<i>Support Vector Regression</i>
Import library yang digunakan	Import library yang digunakan
Import data yang digunakan berupa format csv	Import data yang digunakan berupa format csv
Analisis korelasi antar fitur	Analisis korelasi antar fitur
Seleksi data dengan pemilihan fitur yang memiliki korelasi yang cukup kuat dengan variabel dependen	Seleksi data dengan pemilihan fitur yang memiliki korelasi yang cukup kuat dengan variabel dependen
Data <i>cleaning</i> atau pembersihan data dan transformasi bentuk tipe data yang digunakan	Data <i>cleaning</i> atau pembersihan data dan transformasi bentuk tipe data yang digunakan
Menentukan variabel dependen dan independen yang digunakan	Menentukan variabel dependen dan independen yang digunakan
Membuat model algoritma linear regresi	Memilih kernel yang digunakan "linear"
Menentukan nilai intercept dan koefisien	Membuat model algoritma SVR
Melakukan prediksi algoritma <i>linear regression</i>	Melakukan prediksi algoritma SVR
Evaluasi model menggunakan RMSE	Evaluasi model menggunakan RMSE

### 3.6.4. Evaluasi Model

Setelah melakukan tahapan pemodelan, langkah selanjutnya adalah evaluasi untuk mengukur keakuratan algoritma yang digunakan. Evaluasi diperoleh dari perhitungan nilai *Root Mean Square Error* (RMSE). *Root Mean Square Error* (RMSE) merupakan evaluasi dengan penjumlahan kuadrat *error* atau selisih antara nilai aktual dan nilai prediksi. Semakin kecil nilai RMSE, semakin akurat algoritma yang digunakan. Berdasarkan hasil evaluasi akan didapatkan model algoritma terbaik untuk penelitian ini. Rumus RMSE didefinisikan pada persamaan 5.

$$RMSE = \sqrt{\sum \frac{(Y' - Y)^2}{n}} \quad (5)$$

