

## BAB III METODE PENELITIAN

### 3.1 Objek Penelitian

Sasaran penelitian ini yaitu pada data harga rumah di Karawang tahun 2019 hingga 2022 yang didapatkan dari 3 *website* yaitu *olx.co.id*, *rumah123.com* dan *realoka.com* dengan jumlah data sebanyak 1732 data dengan 9 *variable* yaitu tanggal, judul, alamat, tipe properti, luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, dan harga rumah.

### 3.2 Peralatan Penelitian

Dalam melaksanakan penelitian, peralatan yang digunakan seperti perangkat keras dan perangkat lunak pada laptop dengan spesifikasi ditunjukkan pada Tabel 3.1.

Tabel 3.1 Perangkat Penelitian

Spesifikasi Perangkat Keras	<i>Processor Intel(R) Core(TM) i5-8300H CPU @ 2.30GHz (8CPUs), 2.3GHz. Memory 8192MB RAM Render NVIDIA GeForce GTX 1050</i>
Spesifikasi Perangkat Lunak	<i>Operating System Windows 11 Home Single Language 64-bit (10.0, Build 22621) Microsoft Word 2021 64-bit, untuk penyusunan laporan. Microsoft Excel 2021 64-bit, untuk perhitungan model. Google Colaboratory, sebagai pengujian dengan menerapkan bahasa pemrograman <i>python</i>.</i>

### 3.3 Jenis dan Sumber Data

Jenis data yang digunakan pada penelitian tugas akhir ini yaitu data Kuantitatif. Data kuantitatif merupakan data yang bersifat angka (numerik) yang akan diolah

menggunakan metode statistika. Sedangkan sumber data didapatkan berasal dari 3 Website penjualan properti yaitu olx.co.id, rumah123.com dan realoka.com. Atribut pada data dan penjelasannya ditunjukkan pada Tabel 3.2.

Tabel 3.2 Atribut Penelitian

Atribut	Keterangan
tanggal	Tanggal Penjualan Rumah
judul	Judul penjualan
alamat	Alamat rumah
tipe_properti	Tipe properti yang dijual
luas_tanah	Luas tanah dalam satuan m <sup>2</sup>
luas_bangunan	Luas bangunan dalam satuan m <sup>2</sup>
kamar_tidur	Jumlah kamar tidur
kamar_mandi	Jumlah kamar mandi
harga_juta	Harga rumah

### 3.4 Lokasi dan Waktu Penelitian

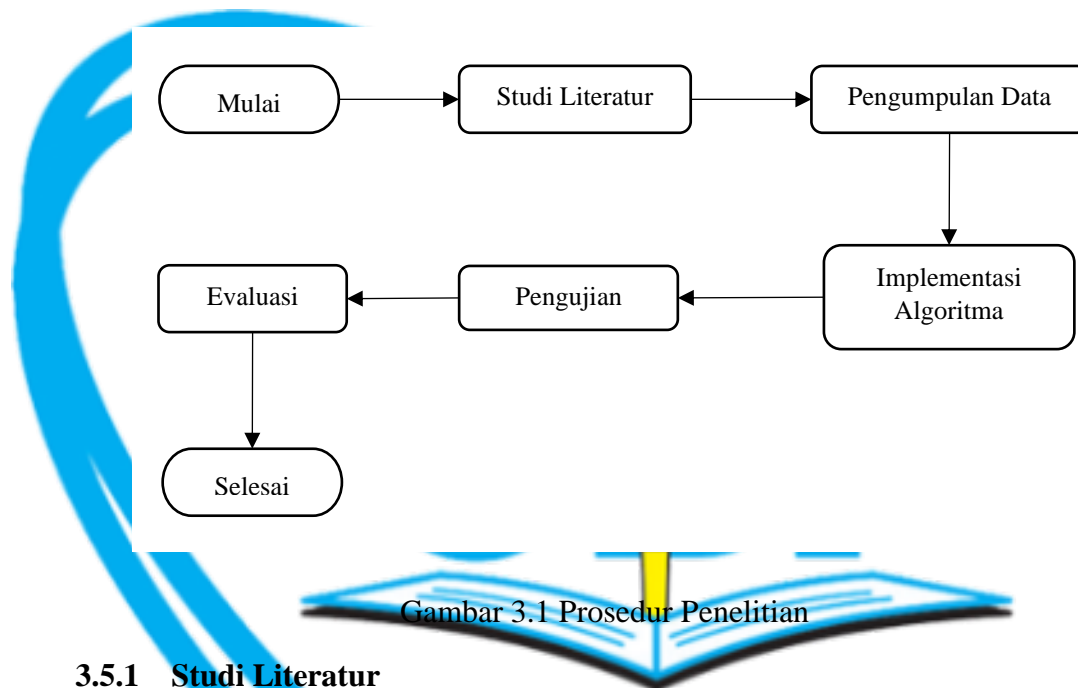
Waktu pencarian data dan pengolahan data harga rumah dilakukan selama kurang lebih satu bulan dari bulan Oktober hingga Nopember 2022. Penelitian dilakukan di Universitas Buana Perjuangan. Berikut merupakan rangkaian pelaksanaan penelitian tugas akhir, terlampir pada Tabel 3.3.

Tabel 3.3 Waktu Pelaksanaan dan Penyusunan Tugas Akhir

No	Keterangan	Nop 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	<i>Pre-processing</i>					█															
4	<i>Modeling</i> Algoritma									█											
5	Pengujian													█							
6	Evaluasi																	█			

### 3.5 Prosedur Penelitian

Penelitian ini diawali dengan studi literatur, pengumpulan data, dan selanjutnya ditampilkan pada Gambar 3.1. Kemudian dilanjutkan pada tahapan implementasi algoritma *Multiple Linear Regression* dan *Support Vector Regression*. Selanjutnya pada pengujian menggunakan data harga rumah di Karawang, setelah diuji akan dilakukan evaluasi menggunakan RMSE.



Gambar 3.1 Prosedur Penelitian

#### 3.5.1 Studi Literatur

Penelitian ini diawali dengan mempelajari jurnal ilmiah yang membahas tentang teknik data mining dengan metode prediksi menggunakan algoritma *Multiple Linear Regression* dan *Support Vector Regression* yang didapatkan dari internet.

#### 3.5.2 Pengumpulan Data

Pengumpulan data pada penelitian ini dengan melakukan *web scrapping* untuk mengumpulkan data, pengumpulan data untuk prediksi harga rumah dilakukan dengan studi dokumenter pada 3 *website* penjualan properti yaitu *olx.co.id*, *rumah123.com* dan *realoka.com*. Pada penelitian ini data yang terkumpul sebanyak 1732 data yang diperoleh melalui *web scrapping* dari 3 *website* penjualan properti. Pada Gambar 3.2 menunjukkan data rumah yang dikumpulkan berisi kolom tanggal, judul, alamat, luas tanah, luas bangunan, kamar tidur, kamar mandi, dan harga rumah. Langkah – langkah pada *web scrapping* (Djufri 2020) diantaranya:

- 1) Install *web scrapping* pada ekstensi Chrome “*Web Scraper*”
- 2) Aktifkan ekstensi yang telah di install
- 3) Buka halaman website yang akan di-*scrape* di Chrome
- 4) Tentukan elemen yang diambil dari halaman website tersebut
- 5) Ekstrak data dari elemen tersebut
- 6) Simpan data yang di ekstrak dalam format yang diinginkan (CSV, JSON atau EXCEL)
- 7) Ulangi langkah ke 4 – 6 untuk pindah ke halaman berikutnya.



	tanggal	judul	alamat	tipe_properti	luas_tanah	luas_bangunan	kamar_tidur	kamar_mandi	harga_juta
0	11/21/2019	Rumah Cantik di Pusat Kota Karawang	Teluk Jambe Timur	Rumah	30	36	2	1	440
1	8/2/2019	Cluster Palumbon City Biaya hanya 2 juta Terim...	Karawang Timur	Rumah	72	48	2	1	451
2	8/18/2019	Rumah Ready Siap Huni di Karawang, Biaya 2 jut...	Majalaya	Rumah	60	42	2	1	326
3	5/1/2019	Subsidi di Karawang Murah Bisa Langsung Huni	Karawang Barat	Rumah	60	30	2	1	125
4	1/14/2019	Tanpa Uang Muka Data Dibantu Sampai Punya Ruma...	Purwasari	Rumah	60	30	2	1	129
...	...	...	...	...	...	...	...	...	...
1727	9/3/2022	DP 2 juta dan Cicilan Mulai 1 Jutaan Perumahan...	Cikampek	Rumah	60	30	2	1	268
1728	3/6/2022	Green Mezzani Karawang Lokasi Klari	Klari	Rumah	60	32	2	1	379
1729	1/23/2022	Rumah Siap Huni Full Renov	Karawang Timur	Rumah	78	36	2	1	350
1730	9/16/2022	Perumahan Subsidi Rasa Komersil di Karawang	Purwasari	Rumah	30	60	2	1	150
1731	4/5/2022	Perumahan Subsidi Temurah dan Termewah di Pusa...	Purwasari	Rumah	30	60	2	1	150

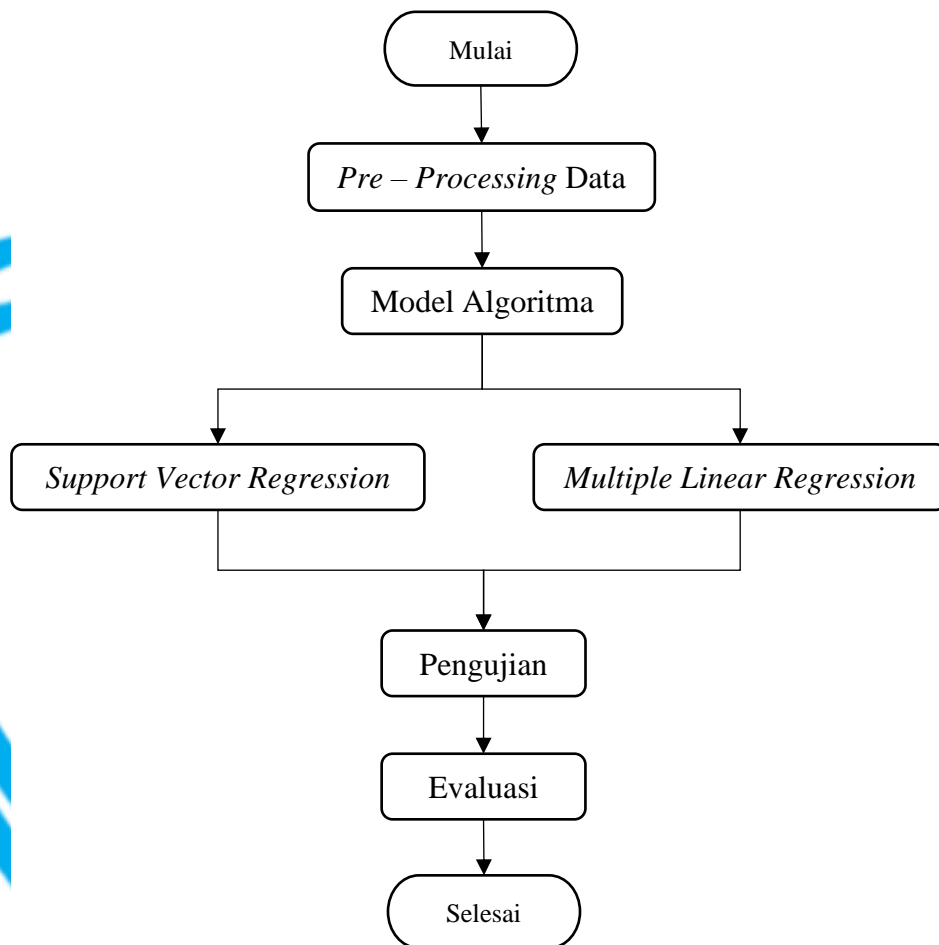
1732 rows x 9 columns

Gambar 3.2 Data Yang Dikumpulkan Dengan *Web Scrapping*



### 3.5.3 Implementasi Algoritma

Berikut model pengolahan prediksi yang ditampilkan pada Gambar 3.3



Gambar 3.3 Implementasi Algoritma

#### 1) *Pre-Processing Data*

Pada penelitian ini, *dataset* akan melalui beberapa tahapan data *pre-processing* untuk mempersiapkan data sebelum digunakan dalam proses pengolahan model prediksi menggunakan algoritma. Berikut tahapan *pre-processing* data yang dilakukan pada penelitian ini:

##### a) Transformasi Data

Transformasi data dalam penelitian ini dilakukan menggunakan metode label *encoding* untuk mengubah *variable* kategorikal menjadi *variable* numerik agar dapat digunakan dalam model prediksi. Pada penelitian ini atribut yang akan di transformasi yaitu judul, alamat dan tipe properti terlihat pada Gambar 3.2 sebelum

transformasi data dan sesudah transformasi data seperti pada Tabel 3.5.

Tabel 3.4 Sesudah Transformasi Data

Judul	Alamat	Tipe Properti	Luas Lahan	Luas Bangunan	Harga Juta
284	17	0	30	30	440
40	5	0	72	48	451
352	11	0	60	42	326
416	4	0	60	30	125
422	12	0	60	30	129

b) Seleksi Data

Pada seleksi data akan pemilihan atribut karena tidak semua atribut digunakan dalam penelitian ini. Pada awalnya, terdapat 9 atribut dalam *dataset*, setelah melakukan seleksi data hanya 6 atribut yang dipilih sebagai atribut yang digunakan dalam penelitian ini. Atribut yang tidak digunakan dalam penelitian ini yaitu “tanggal”, “judul” dan “tipe properti”. Berikut atribut yang akan digunakan seperti pada Tabel 3.4.

Tabel 3.5 Hasil Seleksi Data

Alamat	Luas Lahan	Luas Bangunan	Kamar Tidur	Kamar Mandi	Harga Juta
Telukjambe Timur	30	30	2	1	440
Karawang Timur	72	48	2	1	451
Majalaya	60	42	2	1	326
Karawang Barat	60	30	2	1	125
Purwasari	60	30	2	1	129

c) Pembersihan Data

Pembersihan data dilakukan untuk mengatasi permasalahan data seperti *missing value*, *outlier*, atau mengatasi permasalahan data lainnya yang dapat menghambat pemrosesan data. Setelah data

dilakukan pembersihan data, dan data sudah bersih dan dapat dilanjutkan ke proses selanjutnya. Pada penelitian ini data akan dilakukan pengecekan *missing value* dan *outlier* pada data. Berikut contoh data sebelum pembersihan data terlihat pada tabel 3.6 dan sesudah pembersihan data pada Tabel 3.7.

Tabel 3.6 Sebelum Pembersihan Data

Tanah	Bangunan	Kamar Tidur	Kamar Mandi	Harga
30	36	2	1	173
72	48	3	2	451
<i>Null</i>	<i>Null</i>	<i>Null</i>	<i>Null</i>	<i>Null</i>
40	30	2	1	440
50	36	2	1	245

Tabel 3.7 Sesudah Pembersihan Data

Tanah	Bangunan	Kamar Tidur	Kamar Mandi	Harga
35	30	2	1	173
72	48	3	2	451
45	38	2	1	232
40	30	2	1	440
50	36	2	1	245

Setelah Menangani *missing value*, selanjutnya mengidentifikasi *outlier* dalam data. *Outlier* data yang memiliki nilai yang jauh berbeda dari sebagian besar data dalam sekumpulan data tersebut. Berikut contoh data yang memiliki *outlier* seperti pada Tabel 3.8.

Tabel 3.8 Melihat *Outlier*

Tanah	Bangunan	Kamar Tidur	Kamar Mandi	Harga
91	45	2	2	450
78	42	2	1	344
0	50	2	1	379
30	36	2	1	387

91                      75                      3                      2                      478

---

Pada Tabel 3.8, dapat dilihat bahwa pada atribut tanah terdapat nilai 0 yang dapat diidentifikasi sebagai *outlier* dalam data. Nilai 0 tersebut berbeda dari rentang nilai pada umumnya ada dalam data tersebut, nilai 0 tersebut dapat mempengaruhi model prediksi yang dibangun.

## 2) Model Algoritma

Pada tahap ini model menggunakan 2 algoritma yaitu *Multiple Linear regression* dan *Support Vector Regression* yang menggunakan *Google Colab* yang akan memudahkan dalam melakukan penelitian ini.

Berikut langkah – langkah model algoritma *Multiple Linear Regression* dan *Support Vector Regression* seperti pada Tabel 3.9.

Tabel 3.9 Alur Model Algoritma MLR dan SVR

<i>Multiple Linear Regression</i>	<i>Support Vector Regression</i>
Import <i>library</i> yang digunakan	Import <i>library</i> yang digunakan
Import data yang digunakan	Import data yang digunakan
Korelasi fitur	Korelasi fitur
<i>Pre – Processing</i> data	<i>Pre – Processing</i> data
Menentukan data target dan data prediksi	Menentukan data target dan data prediksi
Membuat model MLR	Membuat model SVR dengan memilih kernel “linear”
Melakukan Prediksi	Melakukan prediksi
Evaluasi model menggunakan RMSE	Evaluasi model menggunakan RMSE

### 3.5.4 Pengujian

Pengujian pada algoritma *Multiple Linear Regression* dan *Support Vektor Regression* dengan menggunakan data harga rumah di Karawang dari tahun 2019 sampai dengan 2022 dan membagi data menjadi *variable dependent* yaitu harga rumah dan *variable independent* yaitu alamat, luas tanah, luas bangunan, jumlah



kamar tidur dan jumlah kamar mandi. Kemudian model akan diuji untuk melakukan prediksi dengan spesifikasi rumah yang ditentukan yaitu alamat, luas lahan, luas bangunan, jumlah kamar tidur dan jumlah kamar mandi.

### 3.5.5 Evaluasi

Evaluasi menggunakan RMSE (*Root Mean Squared Error*) dimana RMSE sering digunakan untuk evaluasi kinerja prediksi yang dipakai untuk mengukur tingkat akurasi dari hasil prediksi. Jika nilai RMSE kecil, maka model prediksi mendekati nilai aslinya dan sebaliknya, jika nilai RMSE besar maka model prediksi yang dihasilkan jauh dari nilai aslinya (Hidayanti *et al.* 2021). Rumus RMSE pada persamaan 3.1 (Bode 2019) seperti berikut.

$$RMSE = \sqrt{\frac{\sum (Y' - Y)^2}{n}} \quad (3.1)$$

Dimana:

$Y'$  = nilai prediksi

$Y$  = nilai sebenarnya

$n$  = jumlah data

