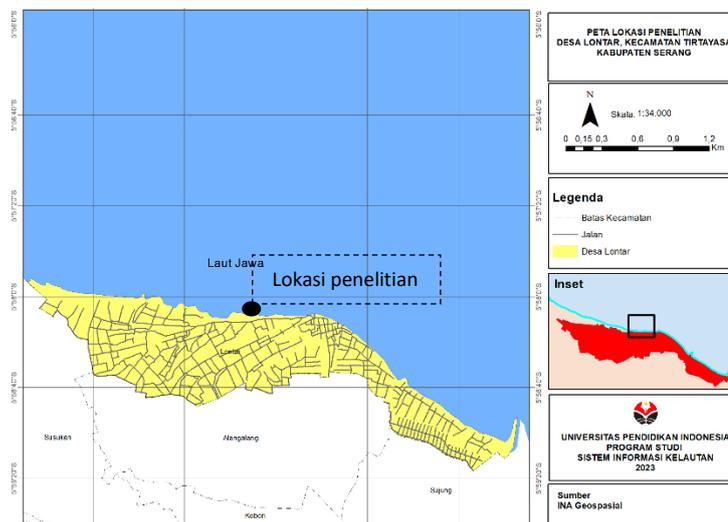


## BAB 3 METODE PENELITIAN

### A. Lokasi dan Waktu Penelitian

Penelitian ini termasuk kategori kuantitatif. Penelitian dilakukan di Desa Lontar, Kecamatan Tirtayasa, Kota Serang, Banten dengan koordinat  $-5.9683605519508784^{\circ}$  LS dan  $106.29757715737725^{\circ}$  BT. Penelitian dilakukan mulai dari bulan September 2023 – Mei 2024. Survey lapangan dilakukan pada bulan September 2023 dan Desember 2023. Peta lokasi penelitian tersaji pada Gambar 3.1.



Gambar 3. 1 Peta Desa Lontar

Desa Lontar berbatasan dengan Kecamatan Tanara di Timur, Kecamatan Pontang di Selatan, Kecamatan Domas di Barat, dan Laut Jawa di Utara. Desa ini merupakan sentra penghasil rumput laut terbesar di Provinsi Banten. Desa Lontar terdiri dari 9 kampung yaitu Lontar, Margiyasa, Kepaksan, Kampung Sawah, Berangbang, Kampung Baru, Sukadiri, Pekandangan, dan Kebalikan.

### B. Data Penelitian

Data rumput laut *Eucheuma cottonii* diperoleh dari survey lapangan di Desa Lontar. Data diperoleh langsung dari petani rumput laut *Eucheuma*

*cottonii* (tahun, bulan, tanggal, total produksi, dan modal) serta mengunduh dari situs penyedia data gratis milik NASA (suhu dan salinitas) dan Republik Perancis (arus). Diagram alir untuk mengunduh data tersedia pada Gambar 3.4, Gambar 3.5, dan Gambar 3.6. Data mentah diolah sehingga menjadi data yang siap digunakan untuk melakukan analisis prediksi oleh mesin. Terdapat 702 data mentah dan setelah dibersihkan menjadi 586 data siap pakai. Distribusi data terbagi menjadi 407 data latih dan 176 data uji.

### C. Variabel Penelitian

Variabel penelitian terbagi menjadi variabel dependen dan variabel independen. Variabel dependen adalah variabel yang dipengaruhi oleh variabel lain, sedangkan variabel independen adalah variabel yang tidak bergantung pada variabel lainnya. Variabel dependen terdiri dari total produksi. Variabel independen terdiri dari tahun, bulan, tanggal, arus, salinitas, suhu permukaan laut, dan modal.

### D. Teknik Analisis Data

Data diperoleh dari 2 sumber yaitu observasi lapangan dan mengunduh melalui internet. Data observasi mencakup tahun, bulan, tanggal, modal, dan total produksi. Data yang diunduh merupakan parameter oseanografi yaitu suhu dari *Website of Worldwide Energy Resource (POWER)* (<https://power.larc.nasa.gov/>), Salinitas dari *NASA Salinity* (<https://salinity.oceansciences.org/aq-salinity.htm>), dan arus laut dari *Aviso+ Satellite Altimetry Data* yang dimiliki oleh Republik Prancis (<https://www.aviso.altimetry.fr/en/data.html>).

Diperoleh 702 data dalam rentang 2011 sampai 2022. Data dirapikan ke dalam *excel* ekstensi *.csv*. Data masa panen rumput laut *Eucheuma cottonii* tersaji pada Tabel 3.1.

Tabel 3. 1 Data Mentah Masa Panen Rumput Laut

Tahun	Bulan	Tanggal	Arus (m/s)	Salinitas (ppt)	Suhu (°C)	Modal (Rp)	Total Prediksi (Kg)
2011	8	10	NaN	33	30	182520	1950
2011	8	11	NaN	33	30	89107	95200
2011	8	14	NaN	33	30	145735	1557
2011	8	14	NaN	33	30	114660	97800

Pengolahan data menggunakan aplikasi Visual Studio Code dan bahasa Python. Terdapat 5 tahapan yang digunakan yaitu memanggil *library*, pra-pemrosesan data, *modelling*, menyimpan ke dalam ekstensi *pickle*, dan membangun situs web. Berikut penjelasan tahapan yang dilakukan dalam menganalisis data.

#### 1. *Library*

*Library* yang digunakan yaitu *Pandas*, *Numpy*, *Matplotlib*, *Seaborn*, *Sklearn*, dan *Scipy*. *Library* berfungsi untuk menyediakan kumpulan kode yang telah dioptimalisasi sehingga membantu dalam efisiensi pengolahan. Berikut merupakan fungsi *library* tersaji pada Tabel 3.2.

Tabel 3. 2 *Library* yang Digunakan (Sial *et al.*, 2021)

<i>Library</i>	Fungsi
<i>Pandas</i>	Pengolahan data tabular
<i>Numpy</i>	Komputasi numerik
<i>Matplotlib</i>	Visualisasi grafik dan plot
<i>Seaborn</i>	Visualisasi statistik sama seperti <i>Matplotlib</i>
<i>Sklearn</i>	Pengolahan algoritma <i>machine learning</i> regresi
<i>Scipy</i>	Pengolahan dan analisis data bersama <i>Numpy</i>

## 2. Pra Pemrosesan Data

Tahap ini wajib dilakukan agar mempermudah dalam pengolahan data. Langkah ini dilakukan untuk memastikan data siap olah sehingga tidak menimbulkan bias. Tahapan yang digunakan dalam penelitian ini tersaji pada Tabel 3.3.

Tabel 3. 3 Tahapan Pra Pemrosesan Data

Tahapan	Fungsi
Mengisi nilai yang hilang ( <i>handling missing value</i> )	Menangani nilai yang hilang dengan <i>mean</i> .
Pembersihan data ( <i>data cleaning</i> )	Memperbaiki kesalahan data apabila terdapat nilai yang tidak valid.
Menghapus pencilan ( <i>handling outlier</i> )	Menghapus nilai pencilan yang dapat mengganggu kestabilan data.
Menangani kemiringan dan keruncingan data ( <i>handling skewness dan kurtosis</i> )	Menangani data supaya berdistribusi normal.
Pemisahan data ( <i>splitting data</i> )	Membagi data menjadi kelompok data latih ( <i>training</i> ) dan data uji ( <i>test</i> ).
Transformasi data	Merubah bentuk data agar sesuai dengan distribusi. Tahap ini melakukan standarisasi data menggunakan <i>StandardScaler</i> .
Reduksi fitur ( <i>Principal Component Analysis</i> )	Mengurangi dimensi <i>dataset</i> dan mempertahankan informasi yang relevan.

### 3. *Modelling*

Tahap ini mengoperasikan algoritma SVR untuk mengolah data. Algoritma ini dipadukan dengan algoritma *GridSearchCV* untuk menemukan *kernel* dan parameter terbaik. Proses penggunaan algoritma ini tersaji pada Gambar 3.7.

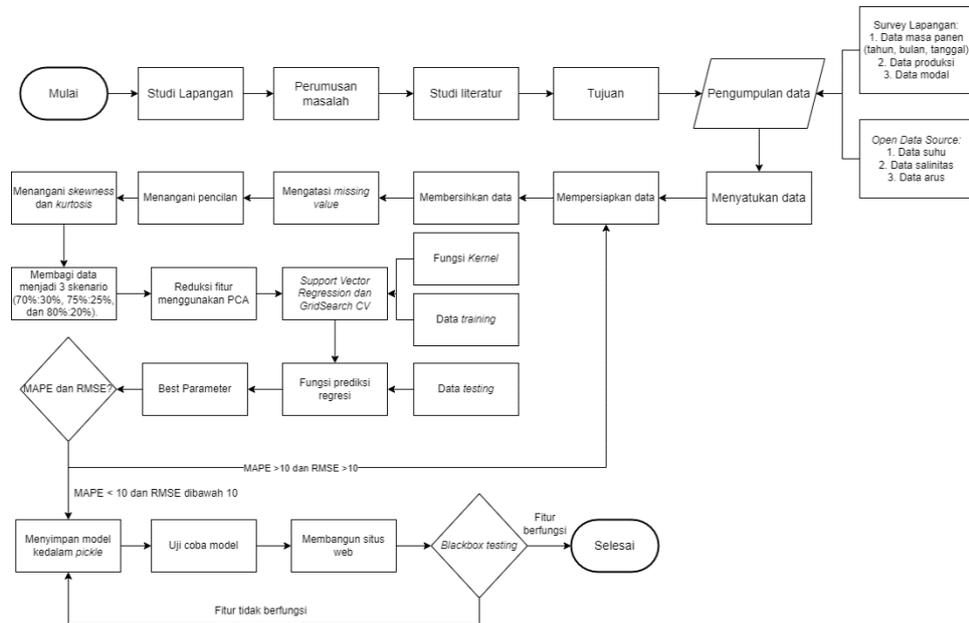
### 4. Akurasi

Akurasi dihitung berdasarkan nilai MAPE dan RMSE. Nilai MAPE yang diharapkan berada pada rentang 0% – 10% (sangat baik) dan nilai RMSE di bawah 10. Penjelasan terkait MAPE dan RMSE terdapat pada Bab 2 Poin 8. Angka tersebut menunjukkan bahwa model yang dihasilkan dari penjadohan *kernel* dan parameter oleh algoritma *GridSearchCV* baik. Model mampu menirukan pola data uji sehingga dapat digunakan untuk prediksi.

*Output* yang dihasilkan dari pengolahan *machine learning* berekstensi .pkl. *Output* tersebut digunakan untuk membangun situs web. Setelah rangkaian kode telah terbangun, seluruh *file* di unggah ke dalam lingkungan pengembangan terintegrasi (IDE) *Python Anywhere* dan siap dikembangkan menjadi situs web.

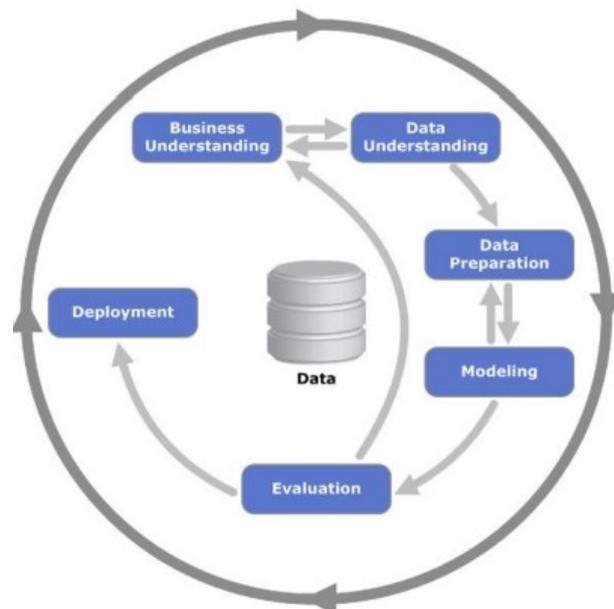
## E. Prosedur Penelitian

Prosedur penelitian yang dilakukan tertuang pada Gambar 3.2.



Gambar 3. 2 Prosedur Penelitian

Gambar 3.2 merupakan prosedur penelitian yang dilakukan dari awal hingga akhir. Penelitian dimulai dari studi lapangan untuk mengetahui proses bisnis dan fenomena yang terjadi. Prosedur penelitian mengacu pada metodologi *cross-industry standard process for data mining* (CRISP-DM). Metode CRISP-DM tertuang pada Gambar 3.3.



Gambar 3. 3 Metode CRISP-DM (Chapman, 2000)

Gambar 3.3 merupakan alur pengolahan data menggunakan metodologi CRISP-DM. Terdapat 7 langkah, yaitu *data source*, *business understanding*, *data understanding*, *data preparation*, *model building*, *test and evaluation*, *deployment*.

#### 1. Pemahaman Bisnis (*Business Understanding*)

Peneliti melakukan analisis data faktor-faktor yang mempengaruhi produksi rumput laut *Eucheuma cottonii* sehingga dapat dilakukan analisis prediksi total produksi rumput laut *Eucheuma cottonii* di masa depan. Pemilihan variabel dilakukan dengan mempertimbangkan berbagai aspek, seperti waktu pengambilan data, biaya, narasumber, ketersediaan data, keterbatasan alat, dan faktor lainnya. Pemahaman bisnis dapat dilakukan melalui studi literatur, diskusi, dan survey. Studi literatur dilakukan untuk memperdalam materi tentang *machine learning*, korelasi *rank spearman*, implementasi SVR, dan aplikasi *framework* Flask. Literatur bersumber dari buku, jurnal, artikel, situs web, dan media sosial.

#### 2. Pemahaman Data (*Data Understanding*)

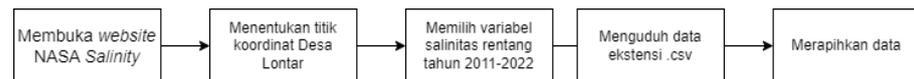
Peneliti melakukan pemahaman terhadap kebutuhan data terkait faktor yang mempengaruhi produksi rumput laut *Eucheuma cottonii* di Desa Lontar. Gambar 3.4 merupakan prosedur penelitian untuk mengunduh data suhu permukaan laut.



Gambar 3. 4 Alur Mengunduh Data Suhu Permukaan Laut

*NASA Prediction Of Worldwide Energy Resource* adalah proyek yang dikembangkan oleh NASA bertujuan untuk menyediakan akses data observasi cuaca dan iklim yang akurat, rinci, dan berkualitas tinggi (Nasution *et al.*, 2023). Gambar 3.4 merupakan alur mengunduh data suhu permukaan laut. Masukkan titik koordinat Desa Lontar yang diperoleh dari *Google Maps* atau geser peta hingga daerah permukaan

laut Desa Lontar. Pilih variabel suhu dengan rentang waktu tahun 2011-2022. Unduh data dengan ekstensi .csv. Sortir data yang dibutuhkan dan satukan dengan kumpulan data dari berbagai parameter dalam 1 file yang sama. Langkah selanjutnya yaitu mengunduh data salinitas tersaji pada Gambar 3.5.



Gambar 3. 5 Alur Mengunduh Data Salinitas

NASA *Salinity* menyediakan peta salinitas berdasarkan algoritma V5 yang diolah oleh *Aquarius Calibration/Validation Working Group* (NASA, 2023). Gambar 3.5 merupakan alur mengunduh data salinitas. Data yang ditampilkan berupa gambar dengan estimasi salinitas pada setiap wilayah yang divisualisasikan dengan warna yang berbeda. Kumpulkan data salinitas setiap tahun dalam 1 file yang sama dengan parameter lain. Langkah selanjutnya yaitu mengunduh data arus tersaji pada Gambar 3.6.



Gambar 3. 6 Alur Mengunduh Data Arus (Yuan *et al.*, 2021)

CNES (*Centre Nation d'Estudes Spatiales*) milik Republik Perancis adalah salah satu situs penyedia data satelit altimetri. Gambar 3.6 merupakan alur mengunduh data arus. Data yang tersedia merupakan data olahan yang telah diolah oleh *Aviso+ Altimetry* dengan persamaan:  $SLA = SSH - MSS - Koreksi$  (3.1)

Keterangan:

SSH = Tinggi permukaan air laut

MSS = Rata-rata tinggi muka air

SLA = Hasil yang telah dikoreksi

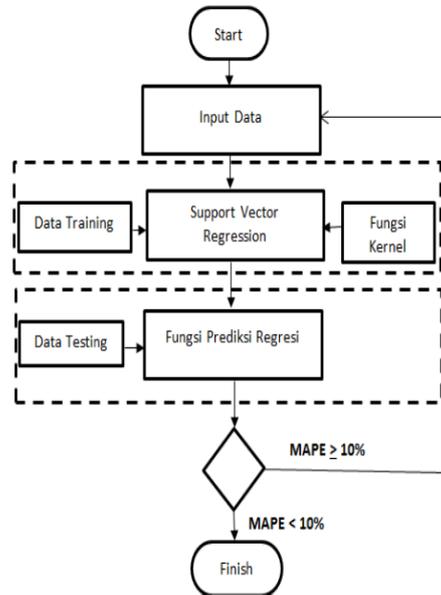
Koordinat Desa Lontar diperoleh dari situs web GOODS (*GNOME Online Oceanographic Data Server*) (Rufai *et al.*, 2023). Koordinat dimasukkan ke dalam *area of interest* pada *AVISO+ Satellite Altimetry Data* kemudian pilih rentang waktu 2011 – 2022. Unduh data dalam ekstensi .csv. Kumpulkan data arus setiap tahun dalam 1 file yang sama dengan parameter lain. Seluruh data yang telah diperoleh disatukan dan dipilah hingga memperoleh data yang dapat dibaca. *File* .csv dipanggil ke dalam baris program agar dapat diolah.

### 3. Persiapan Data (*Data Preparation*)

Setelah melakukan fase pemahaman data, data dipersiapkan untuk dapat diolah oleh mesin. Tahapan ini membangun dataset final berupa data mentah. Langkah-langkah yang harus dilakukan yaitu memanggil *library machine learning* (*Numpy, Pandas, Matplotlib, Seaborn, Sklearn, dan Scipy*), melihat tipe data, mengisi kolom kosong menggunakan mean, melihat distribusi data menggunakan *skewness* dan *kurtosis*, menangani data pencilan dengan IQR. Langkah tersebut membantu data untuk berdistribusi normal. Data yang telah berdistribusi normal dikorelasikan menggunakan *rank spearman*. Nilai korelasi tersaji pada Tabel 2.3.

### 4. Pemodelan (*Modelling*)

Tahap ini merupakan bagian inti pengolahan data yang tersaji pada Gambar 3.7.



Gambar 3. 7 Diagram Alir Algoritma SVR (Cahyono *et al.*, 2019)

Pemodelan yang dilakukan dalam penelitian ini adalah metode regresi menggunakan algoritma SVR. Gambar 3.7 merupakan tahapan pengolahan menggunakan algoritma SVR. Data yang telah melalui tahapan *preprocessing* akan diolah menggunakan algoritma SVR. Data akan dibagi ke dalam 2 kelompok yaitu data latih dan data uji. Pembagian dikelompokkan ke dalam 3 skenario, yaitu:

1. 80% data latih : 20% data uji;
2. 75% data latih : 25% data uji;
3. 70% data latih : 30% data uji.

Skenario tersebut berguna untuk menguji efektivitas model dari berbagai distribusi data. Data dinormalisasi menggunakan *standardscaler*. Agar model yang dihasilkan lebih optimal, perlu melakukan reduksi fitur menggunakan *Principle Component Analysis* (PCA). Pengolahan algoritma SVR dikombinasikan dengan algoritma *Gridsearch Cross Validation* (*GridsearchCV*) dengan memasang:

- a. *Kernel Linier*: rentang parameter "C": [0,001, 0.01, 0.01, 1, 10, 100, 1000, 10000].

- b. *Kernel Poly*: rentang parameter "C": [0,001, 0.01, 0.01,1, 10, 100, 1000, 10000], "degree": [2, 3, 4], "coef0": [0.0, 0.1, 0.5], "gamma": [0,001, 0.01, 0.01, 1, 10, 100, 1000, 10000].
- c. *Kernel Radial Basis Function*: rentang parameter "C": [0,001, 0.01, 0.01, 1, 10, 100, 1000, 10000], "gamma": [0,001, 0.01, 0.01, 1, 10, 100, 1000, 10000], "epsilon" : [0,001, 0.01, 0.01,, 10, 100, 1000, 10000].
- d. *Kernel Sigmoid* : rentang parameter "C": [0.01, 0.01,, 10, 100, 1000, 10000], "coef0": [0.0, 0.1, 0.5], "gamma": [0,001, 0.01, 0.01, 1, 10, 100, 1000, 10000].

#### 5. Evaluasi (*Evaluation*)

Tahap ini melihat nilai eror dari model yang dihasilkan menggunakan RMSE dan MAPE. Nilai RMSE cocok diaplikasikan pada SVR karena interpretasi yang mudah, rentang kesalahan yang sederhana, dan karakteristik model matematis yang cocok (Royal, 2024). Semakin kecil nilai RMSE semakin baik model dalam memprediksi nilai (Royal, 2024). Semakin kecil nilai MAPE (<10%) membuktikan model memiliki performa yang sangat baik (Sahwa Chanigia Viqri & Kurniati, 2023). Tabel nilai MAPE tersaji pada Tabel 2.6.

#### 6. *Deployment*

Setelah seluruh proses pemodelan menggunakan algoritma SVR selesai, selanjutnya adalah membangun situs web menggunakan *framework* Flask.