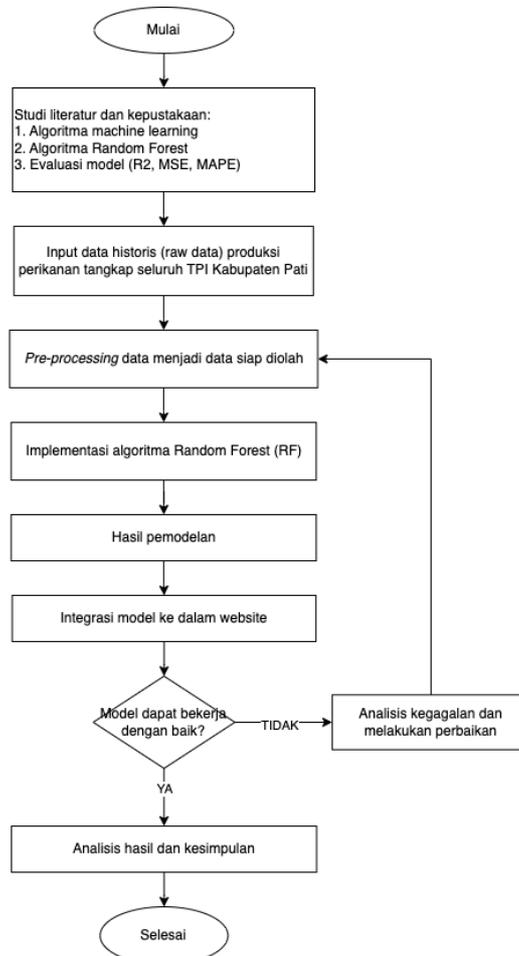


BAB III METODOLOGI PENELITIAN

3.1. Pendekatan/Desain Penelitian

Metode penelitian merupakan sebuah teknik yang digunakan agar dapat memperoleh dan mengolah data untuk memecahkan permasalahan tertentu. Penelitian yang dilakukan menggunakan metode kuantitatif. Metode penelitian kuantitatif merupakan teknik penelitian secara ilmiah atau *scientific* karena berhasil memenuhi pedoman ilmiah secara konkrit atau empiris, terukur, sistematis, objektif dan rasional (Sugiyono, 2018). Berikut alur penelitian dalam bentuk *flowchart* dapat dilihat pada Gambar 3.1.



Gambar 3.1. *Flowchart* Alur Tahapan Penelitian

3.2. Metode Penelitian

Penelitian ini juga menerapkan *data mining* sebagai kerangka dalam pengembangan sistem yang dibentuk, yaitu dengan metode *Knowledge Discovery in Databases* (KDD). KDD cocok digunakan untuk menganalisis data dalam jumlah besar dan menemukan korelasi antar dataset agar dapat membentuk pola untuk menentukan algoritma yang baik (Syahra *et al.*, 2018). Ketika memprediksi dengan algoritma *machine learning*, diperlukan tahapan *data mining* dalam pengolahan data (*raw data*) sebelum data tersebut dapat terbaca oleh model algoritma. Berikut merupakan tahapan proses KDD menurut Fayyad *et al.*, (1996):

1. *Data Selection*

Memilih data yang relevan dan menghilangkan data yang tidak relevan. Fitur atau variabel yang dihilangkan merupakan variabel yang tidak memiliki hubungan terhadap variabel independen (Samudra, 2019).

2. *Data Preparation*

Data yang sudah dipilih kemudian diubah menjadi CSV (*Comma Separated Values*) untuk mempermudah proses pengolahan data. Langkah berikutnya yaitu data *cleaning* untuk menghilangkan residu pada dataset, seperti *missing value*, *duplicate data*, *irrelevant*, *incomplete* data yang tidak terdistribusi normal dan *outlier* pada dataset.

3. *Data Transformation*

Data yang sudah bersih kemudian diubah menjadi format yang sama dengan *range* maupun nilai dengan skala atau rentang yang sama sehingga keseluruhan data dapat terbaca dan dapat diolah oleh algoritma *machine learning*. Selain itu, pada tahap ini terjadi pembagian (*splitting*) data latih atau *data train* dan data uji atau *data test*.

SIK UPI Kampus Serang

4. *Data Mining*

Proses ini mengimplementasikan algoritma *machine learning* untuk mengetahui algoritma yang cocok dilihat berdasarkan algoritma dengan nilai *error* lebih kecil.

5. *Interpretation/Evaluations*

Hasil yang diperoleh berdasarkan penggunaan algoritma *machine learning* tersebut dilakukan evaluasi dan analisis apakah nilai akhir tersebut sudah sesuai seperti yang diharapkan atau tidak.

3.3. Teknik Penelitian

3.3.1. Teknik Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan data yang dikumpulkan berasal dari pihak lain dan bukan hasil peneliti sendiri. Peneliti merupakan pihak kedua yang memanfaatkan data tersebut sebagai bahan penelitian. Data yang digunakan merupakan gabungan data arsip hasil tangkapan perikanan tangkap yang dikelola oleh DKP dan BPS Kabupaten Pati, tahun 2013 hingga tahun 2022. Sebelum melakukan penelitian, perlu adanya variabel sebab akibat. Berikut merupakan variabel yang digunakan:

1. Variabel bebas/*independent* (X)

Variabel bebas adalah variabel yang mempengaruhi sebab kemunculan dari variabel terikat. Berikut merupakan variabel bebas yang digunakan:

- a. Triwulan
- b. Tahun produksi
- c. Bulan produksi
- d. Jumlah keseluruhan produksi perbulan dalam bentuk kilogram (kg)

2. Variabel bebas/*independent* (X)

Variabel bebas adalah variabel yang mempengaruhi sebab kemunculan dari variabel terikat. Berikut merupakan variabel bebas yang digunakan:

- e. Triwulan
- f. Tahun produksi
- g. Bulan produksi
- h. Jumlah keseluruhan produksi perbulan dalam bentuk kilogram (kg)

3. Variabel terikat/*dependent* (Y)

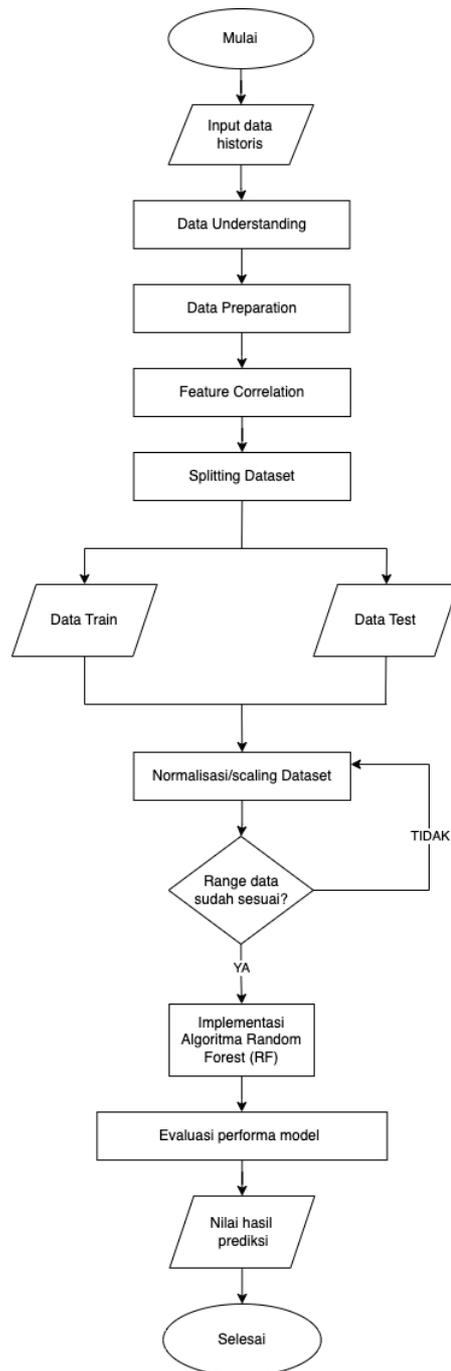
Variabel terikat adalah variabel yang dipengaruhi variabel bebas dan dapat berubah karena pengaruh variabel tersebut. Variabel terikat yang digunakan adalah harga jual tangkapan setiap bulan dalam bentuk (Rp).

Pengumpulan data tambahan untuk membantu penelitian berdasarkan catatan peneliti terdahulu disebut studi literatur (*library research*), yaitu mencari informasi melalui berbagai referensi terkait dengan nilai, budaya maupun norma yang berkembang dalam penelitian (Sugiyono, 2018). Studi literatur tidak hanya digunakan sebagai langkah awal dalam pembentukan kerangka penelitian, namun dapat juga dimanfaatkan untuk memperoleh data yang dibutuhkan pada penelitian (Zed, 2014).

3.3.2. Teknik Analisis Data

Tahapan pengolahan data dilakukan ketika data yang diperlukan sudah lengkap. Berikut merupakan *flowchart* tahapan *data mining* dalam pengolahan data menggunakan algoritma *machine learning Random Forest* (RF), dapat dilihat pada Gambar 3.2 dibawah ini:

SIK UPI Kampus Serang



Gambar 3.2. Flowchart Proses Algoritma *Random Forest*

Berdasarkan Gambar 3.2 di atas, berikut merupakan penjelasan lebih jelas terkait tahapan atau proses pengolahan data menggunakan algoritma pada *Data Mining*, yaitu dengan menggunakan algoritma *Random Forest*.

SIK UPI Kampus Serang

1. *Data Understanding*

Tahap ini dilakukan pembelajaran dan pemahaman informasi statistik yang terdapat pada keseluruhan fitur dalam dataset yang diperlukan menurut Rahmat *et al.*, (2023), yaitu mengetahui setiap tipe data pada fitur, memahami nilai seperti jumlah keseluruhan fitur, nilai minimum (min), nilai maksimum (max), rata-rata (*mean*), varians (*variance*) dan standar deviasi (*standar deviation*) dari setiap fitur.

2. *Data Preparation*

Tahap ini disebut juga sebagai tahap *preprocessing*, dimana data mentah (*raw data*) akan diperiksa kualitasnya sebelum diolah menjadi data siap pakai. Tahap ini penting dilakukan agar hasil akhir dari penelitian dapat tercapai dengan baik (Sari *et al.*, 2023). *Data preparation* dikatakan sebagai tahapan statistik, yaitu untuk menghilangkan residu pada *dataset* dengan *handling missing value*, *handling outlier* dan *label encoding* (Kuvalekar *et al.*, 2020). *Missing value* pada dataset dapat dihapus atau diisi dengan *nilai mean*, *median* atau *mode* sesuai kebutuhan.

3. *Features Correlation*

Feature correlation merupakan proses pemilihan fitur yang dilakukan oleh sistem atau manual untuk memilih fitur yang berkontribusi besar pada variabel (Amiruddin & Ishak, 2022). Seluruh fitur bukan target diuji korelasi menggunakan *heatmap* dengan variabel target (Y). Fitur dengan korelasi tinggi terhadap variabel target, disimpulkan bahwa fitur tersebut saling mempengaruhi. Jika fitur dengan korelasi rendah terhadap variabel target, fitur tersebut tidak memiliki pengaruh signifikan. Adanya *feature correlation* dilakukan agar ketika menggunakan metode *feature important* tidak terjadinya *overfitting* data pada model (Suliztia, 2020).

4. *Splitting Dataset*

Setelah *dataset* sudah menjadi data siap pakai, langkah selanjutnya adalah membagi dataset menjadi dua bagian, menjadi data latih (*data train*) dan data uji (*data test*) dengan beberapa perbandingan sesuai banyak dikitnya

SIK UPI Kampus Serang

data yaitu 10:90, 20:80: 30:70, 40:60, 50:50, 60:40, 70:30, 80:20 dan 90:10 (Nguyen *et al.*, 2021). Pembagian *dataset* menjadi data latih (*data train*) dan data uji (*data test*) digunakan untuk memisahkan data dalam melatih model dan untuk proses pengujian model (Muraina, 2022). Data latih (*data train*) digunakan untuk pembuatan model dengan algoritma *machine learning*. Sedangkan data uji (*data test*) digunakan untuk menguji performa kinerja model algoritma *machine learning* (Dalimunthe & Hakim, 2023). Penelitian ini menggunakan rasio *splitting* 60:40, 70:30 dan 80:20 menyesuaikan *dataset* yang digunakan dan mengacu pada penelitian Vrigazova (2021), *splitting dataset* dengan rasio perbandingan 50:50, 60:40, 70:30, 80:20 dapat membentuk model yang optimal.

5. *Scaling Dataset*

Scaling data terbagi menjadi dua, normalisasi dan standarisasi. Normalisasi data atau *normalization* merupakan proses mengubah data menjadi terdistribusi normal sehingga memiliki rentang nilai yang sama (Harmain *et al.*, 2021). Pengubahan nilai pada proses normalisasi data dilakukan tanpa merusak perbedaan rentang nilai (Arifin *et al.*, 2022). Normalisasi dapat menggunakan *min-max scaler*. Sedangkan standarisasi atau *standarization* merupakan penyesuaian *dataset* dengan memusatkan *mean* pada 0 dan standar deviasi 1 agar keseluruhan data dapat dibaca oleh algoritma (Raharjo, 2021). Standarisasi dapat menggunakan *z-score*.

6. *Algoritma Implementation*

Setelah melakukan normalisasi atau *scaling dataset*, langkah selanjutnya mengimplementasikan algoritma *Random Forest* kepada *dataset* untuk membentuk model yang dibutuhkan.

7. Model Evaluation

Tahap evaluasi dilakukan pengukuran akurasi hasil dari pemodelan menggunakan data uji atau *data testing* (Lestari & Astuti, 2022). Pengukuran evaluasi model dapat dilihat dari nilai akurasi atau ketepatan *error* yaitu menggunakan nilai r^2 , MSE, dan MAPE.

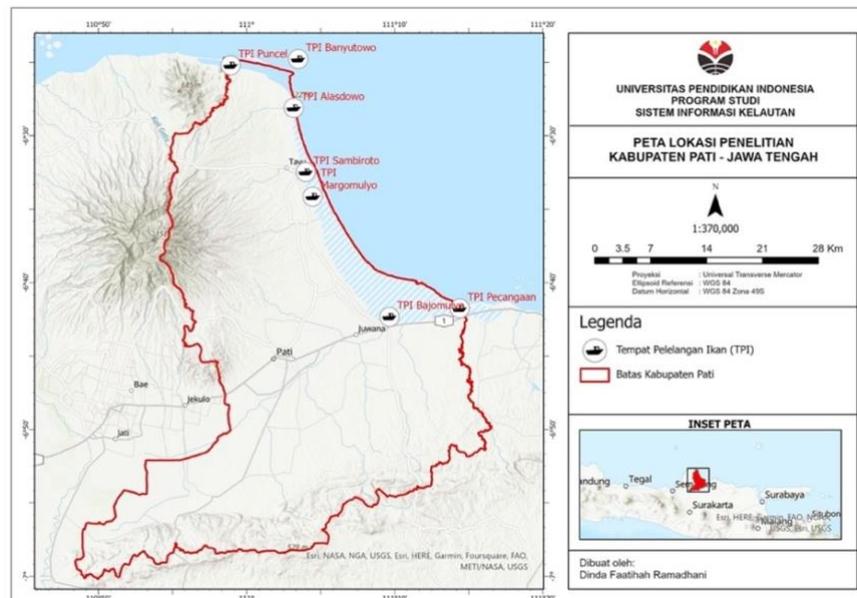
3.4. Latar/Setting Penelitian

3.4.1. Waktu Penelitian

Penelitian dilakukan dalam kurun waktu empat bulan. Penelitian dilakukan dimulai dari bulan Oktober hingga bulan Januari tahun 2024.

3.4.2. Tempat Penelitian

Data penelitian yang digunakan didapatkan dari tujuh TPI Kabupaten Pati. Berikut peta persebaran TPI Kabupaten Pati, pada Gambar 3.3.



Gambar 3.3. Peta Penelitian Wilayah Kabupaten Pati

3.5. Dataset Penelitian

Subjek penelitian terbagi menjadi dua yaitu data populasi dan data sampel. Menurut Sugiyono (2018), populasi diartikan sebagai gabungan antara objek dan subjek tertentu yang memenuhi syarat berdasarkan ketentuan peneliti. Menurut Sinaga (2014) populasi merupakan keseluruhan subjek penelitian yang mewakili suatu karakteristik tertentu dalam penelitian. Data populasi yang digunakan untuk penelitian adalah keseluruhan data latih (*data training*) yang sebelumnya sudah dilakukan pembagian *dataset* dengan *splitting dataset*, yaitu rasio 60:40, 70:30, dan 80:20. Data ini digunakan untuk melatih dan mempelajari pola pada *dataset* untuk membentuk model.

Sampel menurut Sugiyono (2018) merupakan bagian yang dipisahkan dari populasi. Pada penelitian ini, sampel yang digunakan adalah data uji atau *data testing*, yaitu data dengan jumlah nilai kecil, digunakan untuk menguji sekaligus mengevaluasi model yang telah dibentuk.

3.6. Prosedur Penelitian

Sebelum melakukan penelitian, terdapat beberapa prosedur dan sistematika tahapan penelitian yang jelas dan terstruktur saat melakukan penelitian. Tahapan penelitian terbagi menjadi:

1. Perencanaan

Tahap awal sebelum melakukan penelitian adalah mengumpulkan informasi pendukung terkait permasalahan dalam penelitian, menentukan rumusan masalah, tujuan, ruang lingkup penelitian, mengumpulkan data yang digunakan, menentukan variabel yang dibutuhkan untuk penelitian.

2. Pelaksanaan

Setelah melakukan perencanaan yang matang, dilakukan penelitian dengan mengambil data lapangan atau mengolah data sekunder yang didapat dari pihak terkait dengan penelitian menggunakan metode yang tepat untuk menghasilkan output yang baik.

3. Evaluasi

Setelah mendapatkan hasil, peneliti melakukan tahapan analisis data atau melakukan evaluasi untuk mengukur tingkat keakuratan atau keberhasilan penelitian tersebut menggunakan beberapa metode.

4. Penyusunan laporan

Tahap akhir dari penelitian adalah melakukan penyusunan dan penulisan laporan atas hasil penelitian yang sudah dilakukan.