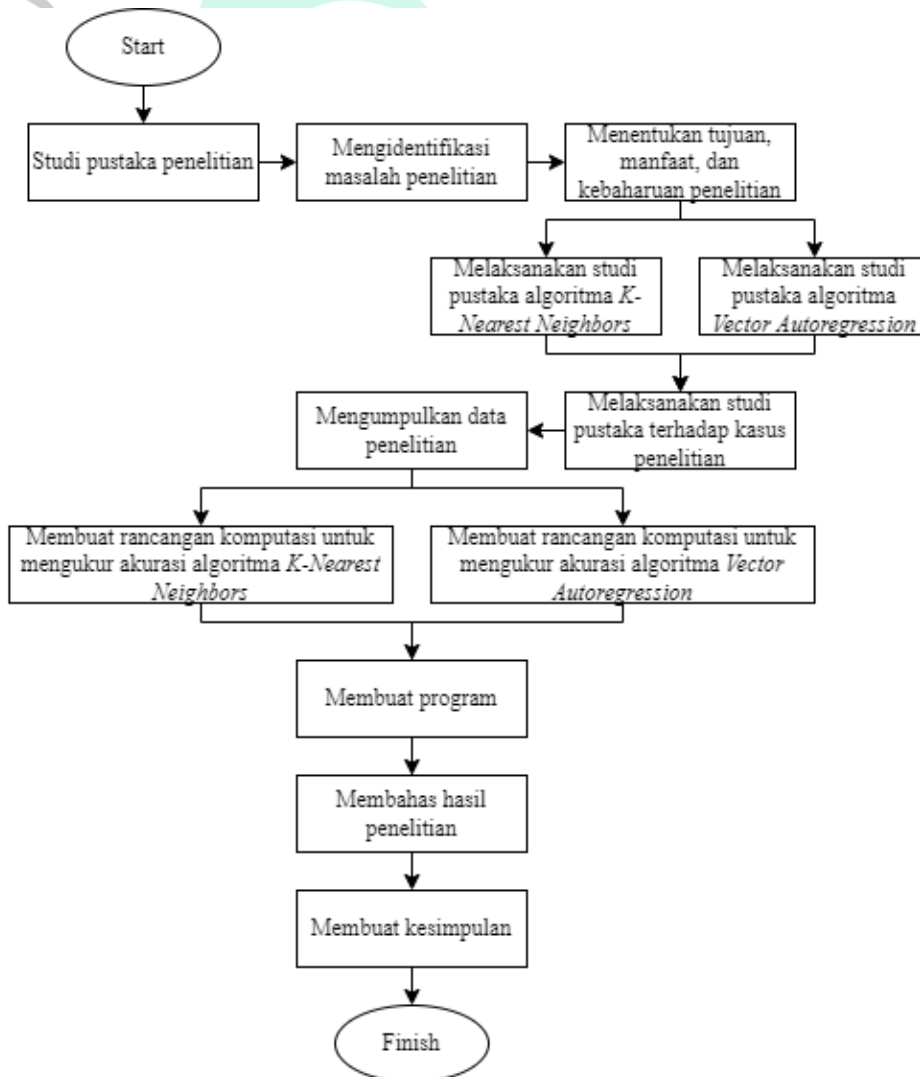


BAB IV PERENCANAAN

Skripsi ini dilaksanakan dengan menggunakan berbagai langkah dan rancangan penelitian. Setiap langkah dan rancangan yang dibuat merupakan pedoman untuk melakukan penelitian secara berurutan sesuai tujuan skripsi ini.

4.1. Langkah – Langkah Penelitian

Penelitian tidak dapat berjalan dengan baik tanpa adanya langkah-langkah penelitian yang berurutan dan sistematis. Langkah-langkah tersebut dapat berguna sebagai gambaran dan pedoman bagi penulis untuk melaksanakan penelitian. Apapun langkah-langkah penelitian ini tertuang pada diagram berikut.



Gambar 4.1 Diagram Alur Penelitian

Berdasarkan gambar 4.1 di atas, langkah-langkah penelitian dapat diuraikan sebagai berikut.

- (1) Studi pustaka penelitian dilaksanakan dengan mencari topik yang akan diteliti. Topik dibuat dengan melakukan riset pada jurnal, artikel, maupun skripsi yang telah terbit secara daring dan memiliki kredibilitas baik. Berdasarkan riset tersebut dapat tergambar garis besar yang akan diteliti, lalu dipersempit dengan menentukan detail seperti penentuan algoritma yang digunakan dan studi kasus penelitian.
- (2) Mengidentifikasi masalah penelitian dengan merumuskan masalah dan membuat batasan yang akan diteliti. Setiap masalah yang dirumuskan merupakan penelitian yang dirancang untuk dilaksanakan.
- (3) Menentukan tujuan dari dilaksanakannya penelitian, manfaat untuk masyarakat, ilmu pengetahuan, dan peneliti, serta kebaruan berdasarkan penelitian sebelumnya yang sesuai dengan topik skripsi.
- (4) Melaksanakan studi pustaka terkait algoritma *K-Nearest Neighbors* dan *Vector Autoregression* seperti definisi, alur kerja, proses komputasi, tingkat akurasi kinerja, dan lain-lain. Studi pustaka dilakukan sebagai wujud pembelajaran untuk mencari keterkaitan antara algoritma *machine learning* dengan topik penelitian.
- (5) Melaksanakan studi pustaka terhadap kasus yang digunakan pada penelitian. Hal tersebut diperlukan sebagai wujud pengetahuan kepada penulis terkait gambaran secara besar sampai dengan detail mengenai batubara Newcastle berjangka.
- (6) Mengumpulkan data penelitian berupa kumpulan baris historis harga batubara Newcastle berjangka yang diperoleh dari halaman web investing.com sesuai dengan batasan penelitian.
- (7) Membuat rancangan komputasi untuk mengukur akurasi antar algoritma. Rancangan komputasi dibuat dalam bentuk *flowchart* berupa cara kerja algoritma *K-Nearest Neighbors* dan *Vector Autoregression*. Komputasi dibuat dengan menyesuaikan kondisi data dan bahasa pemrograman yang digunakan.

- (8) Mengumpulkan data hasil perbandingan dari algoritma *K-Nearest Neighbors* dan *Vector Autoregression* berdasarkan data historis yang diperoleh dan hasil yang diinginkan sesuai dengan tujuan penelitian. Data hasil tersebut disajikan berdasarkan data latih dan data uji yang sudah ditentukan untuk dilakukannya pemodelan prediksi.
- (9) Membahas hasil penelitian dengan menganalisis diantara dua algoritma yang digunakan. Setiap langkah-langkah komputasi diperhatikan hasilnya sesuai dengan topik penelitian.
- (10) Membuat kesimpulan berdasarkan hasil penelitian yang telah dibahas. Kesimpulan pada penelitian ini akan dapat mengetahui algoritma mana yang lebih akurat dalam membuat pemodelan pergerakan harga batubara Newcastle berjangka. Hasil akurasi akan disajikan berupa nilai *Root Mean Squared Error (RMSE)* dan *Mean Absolute Percentage Error (MAPE)* sebagai perbandingan antar algoritma. Selain itu, pada bagian ini diberikan saran penelitian oleh peneliti yang dapat membangun kepada pembaca supaya dapat melakukan penelitian lanjutan atau dijadikan sebagai sumber referensi.

Langkah-langkah penelitian dibuat berdasarkan kebutuhan yang diperlukan peneliti dalam melaksanakan seluruh rangkaian pengerjaan skripsi. Langkah-langkah tersebut dikerjakan secara berurutan. Pada setiap langkah memiliki kompleksitas pendalaman yang tinggi dalam melakukan penelitian.

4.2. Rancangan Pengujian

Prediksi harga batubara Newcastle berjangka yang dilihat pada halaman web investing.com hanya memberikan kesimpulan bahwa apakah pada hari ini baik untuk dibeli atau tidak. Halaman web tersebut tidak menyajikan harga di masa yang akan mendatang. Walaupun setiap pengguna situs tersebut dapat beropini mengenai prediksi harga di hari yang akan datang. Namun, tidak dapat dipastikan keakuratan hasilnya karena tidak dihitung secara terkomputerisasi dengan menggunakan sistem cerdas. Dengan demikian, investor akan ragu untuk dapat berinvestasi. Oleh karena itu, pada penelitian ini dikembangkan prediksi harga dengan membuat suatu pemodelan pergerakan harga batubara Newcastle berjangka berdasarkan harga

historis yang diperoleh dari situs investing.com sesuai dengan batasan penelitian. Data tersebut akan dimodelkan menggunakan algoritma *K-Nearest Neighbors* dan *Vector Autoregression*.

Penelitian dalam membuat pemodelan pergerakan harga batubara Newcastle berjangka menggunakan algoritma *K-Nearest Neighbors* dan *Vector Autoregression* dilakukan dengan berbagai langkah komputasi. Komputasi pengujian algoritma tersebut dirancang menggunakan diagram perancangan. Rancangan diperlukan sebagai wujud dari suatu perencanaan yang terarah dalam melakukan penelitian. Setiap algoritma memiliki rancangan komputasi untuk pengujian yang berbeda-beda. Kebutuhan seperti *library* dan *tools* yang digunakan memiliki perbedaan di antara algoritma *K-Nearest Neighbors* dengan *Vector Autoregression*. Oleh karena itu, hasil akurasi kedua algoritma tersebut akan beragam.

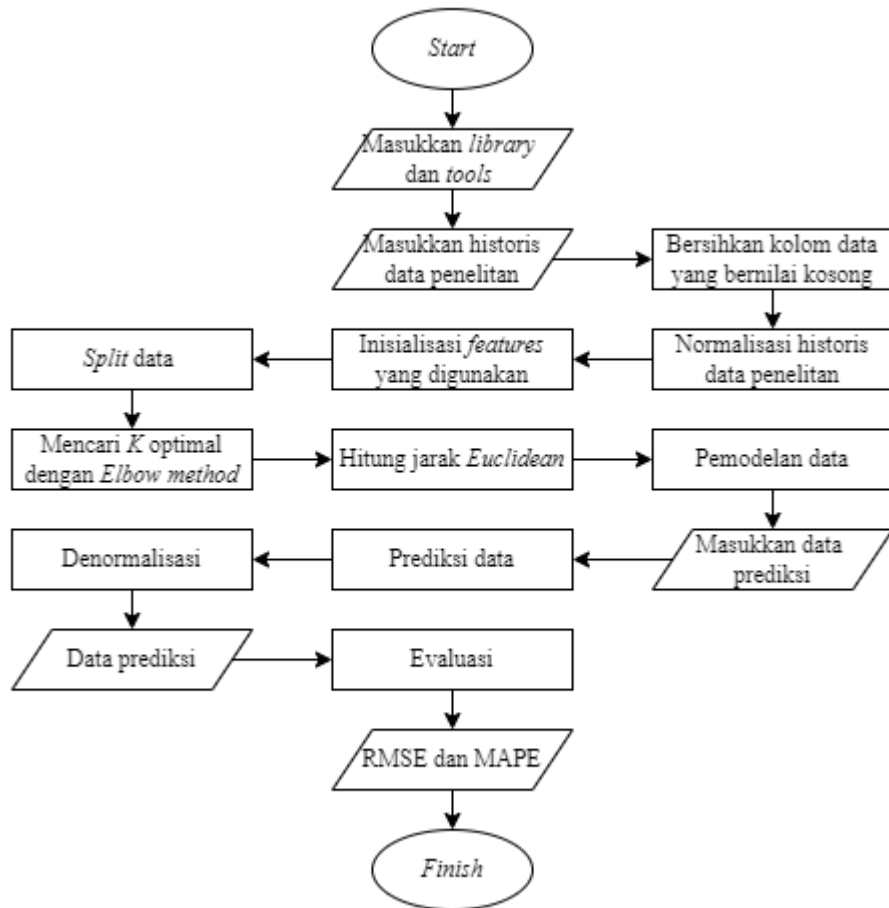
4.2.1. Rancangan Pengolahan Data

Pengolahan data pada penelitian ini dilakukan untuk dapat mencapai hasil yang sesuai dengan permasalahan yang diangkat. Adapun rancangan pengolahan data diuraikan sebagai berikut.

- (1) Membersihkan kolom data yang bernilai kosong. Setiap baris data akan diperiksa oleh sistem apakah masih terdapat salah satu kolom yang tidak terisi.
- (2) Normalisasi data menggunakan metode *MinMax scaler*. Proses ini diperlukan untuk menstabilkan setiap angka menjadi setara supaya dapat diolah dengan baik.
- (3) Proses komputasi algoritma *K-Nearest Neighbors* dan *Vector Autoregression*. Proses ini dibahas pada sub sub bab berikutnya mengenai rancangan *K-Nearest Neighbors* dan *Vector Autoregression*.
- (4) Denormalisasi hasil pemodelan yang didapatkan untuk memperoleh angka yang sebenarnya. Tahap ini merupakan tahap terakhir dalam membuat pemodelan pergerakan harga batubara Newcastle berjangka menggunakan algoritma *machine learning*.

4.2.2. Rancangan *K-Nearest Neighbors*

K-Nearest Neighbors merupakan algoritma *machine learning* yang digunakan untuk memprediksi suatu data dengan menggunakan nilai K optimal. Algoritma tersebut memiliki rancangan komputasi untuk pengujian pada diagram berikut ini.



Gambar 4.2 Diagram Alur Algoritma *K-Nearest Neighbors*

Berdasarkan gambar 4.2 pada algoritma *K-Nearest Neighbors* di atas, maka dapat diuraikan secara detail sebagai berikut.

- (1) Memasukkan *library* dan *tools* yang dibutuhkan untuk melakukan komputasi pada algoritma *K-Nearest Neighbors*. *Library* yang digunakan yaitu *pandas* (analisis data), *numpy* (komputasi data numerik), *matplotlib, pyplot* (visualisasi data dalam bentuk grafik), dan *math* (mengolah operasi matematika). *Tools* yang digunakan pada algoritma *K-Nearest Neighbors* yaitu *sklearn* (melakukan normalisasi data, pembagian

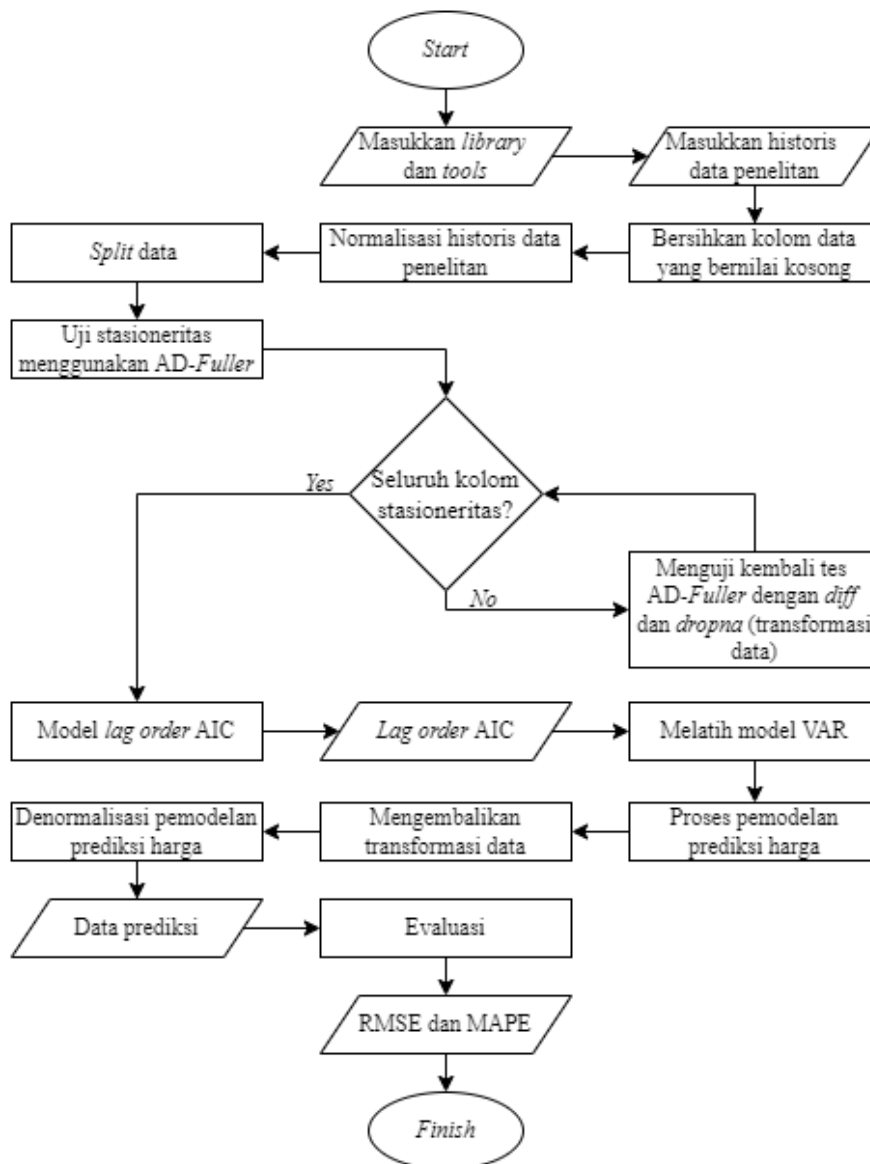
data, proses prediksi, dan evaluasi) dan *warnings* (membuang luaran yang tidak diperlukan untuk pengguna).

- (2) Memasukkan data historis harga batubara Newcastle Berjangka yang telah diunduh melalui *investing.com* dengan periode sesuai batasan penelitian dan disimpan di folder Google Drive.
- (3) Membersihkan kolom data yang bernilai kosong menggunakan *dropna*. Data kosong perlu dihapus untuk menghindari adanya pengolahan data yang kurang tepat.
- (4) Data yang telah dimasukkan akan dinormalisasi dengan menggunakan metode *MinMax scaler*. Normalisasi berfungsi untuk menormalkan seluruh harga pada setiap baris agar memiliki status nilai yang sama.
- (5) Inisialisasi fitur yang digunakan pada pemodelan prediksi. Pada penelitian ini, fitur yang diinisialisasikan sebagai x merupakan kolom data *Open*, *High*, *Low*, dan *Vol* pada hari itu (d). Selain itu, fitur yang diinisialisasikan sebagai y berisikan kolom data *High* ($d+1$) dan *Low* ($d+1$).
- (6) Variabel x dan y yang telah diinisialisasikan dibagi menjadi data latih dan data uji. Data uji memiliki porsi yang lebih sedikit daripada data latih. Data uji pada penelitian ini adalah 10% dari total banyaknya data atau setara dengan rasio 90:10.
- (7) Nilai K optimal dicari menggunakan metode Elbow. Nilai K yang optimal tersebut dihitung jarak *Euclidean* untuk dibuatkan pemodelan data. Pemodelan data dilakukan dengan cara menginisialisasi isi kolom pada variabel baru.
- (8) Buat variabel yang berisikan data prediksi untuk melakukan proses prediksi dengan berdasarkan nilai K optimal.
- (9) Denormalisasi data yang telah diprediksi menggunakan rumus denormalisasi pada setiap hasil prediksi.
- (10) Hasil prediksi (*High* ($d+1$) dan *Low* ($d+1$)) akan dikeluarkan setelah proses denormalisasi harga batubara Newcastle berjangka selesai.
- (11) Evaluasi model menggunakan *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE). Evaluasi dilakukan untuk

melihat skor RMSE dan MAPE yang akan dibandingkan dengan algoritma *Vector Autoregression*.

4.2.3. Rancangan *Vector Autoregression*

Vector Autoregression (VAR) merupakan algoritma yang dapat bekerja dengan menggunakan data *time series*. Algoritma tersebut memiliki rancangan komputasi yang tersaji pada diagram berikut ini.



Gambar 4.3 Diagram Alur Algoritma *Vector Autoregression*

Berdasarkan gambar 4.3 algoritma *Vector Autoregression* di atas, maka dapat diuraikan secara terperinci sebagai berikut ini.

- (1) Memasukkan *library* dan *tools* pada program pemodelan pergerakan harga batubara Newcastle berjangka. *Library* yang digunakan yaitu *numpy* (komputasi data numerik) dan *pandas* (analisis data). *Tools* yang dipakai yaitu *Statsmodels* (memproses model algoritma, mengecek stasioneritas, dan melakukan evaluasi hasil), *sklearn* (mengolah normalisasi data dengan metode *MinMax scaler*), dan *warnings* (membuang luaran yang tidak diperlukan untuk pengguna).
- (2) Memasukkan data historis harga batubara Newcastle berjangka yang diperoleh dari halaman web *investing.com*. Data historis tersebut diunduh lalu dimasukkan ke dalam folder *Google Drive* yang akan dipanggil melalui program.
- (3) Membersihkan kolom data yang bernilai kosong menggunakan *dropna*.
- (4) Data yang telah dibersihkan akan dinormalisasi dengan menggunakan metode *MinMax scaler*.
- (5) Bagi data menjadi dua yaitu data latih dan data uji berdasarkan *range* yang telah ditentukan yaitu pada rasio 90:10 terhadap data latih dan data uji.
- (6) Menguji stasioneritas menggunakan *Augmented Dickey-Fuller* (*AD-Fuller*). *Augmented Dickey-Fuller* merupakan uji statistik untuk menguji apakah setiap kolom pada data latih tersebut stasioner atau tidak. Nilai ADF dapat dikatakan stasioner jika memiliki angka yang dihasilkan kurang dari 0,05 (ketentuan standar *AD-Fuller*).
- (7) Memeriksa apakah seluruh kolom sudah stasioner atau belum. Jika terdapat salah satu kolom yang belum stasioner, maka dilakukan pengujian *Augmented Dickey-Fuller* kembali dengan menggunakan *diff* dan *dropna* pada *pandas* (proses transformasi data). Setelah itu diulang kembali tahap ke-6 sampai seluruh kolom pada data telah stasioner dan dilanjutkan ke tahap selanjutnya.
- (8) Memilih model *lag order Vector Autoregression* (VAR). Pada penelitian yang menggunakan algoritma *Vector Autoregression* umumnya menggunakan *lag order* AIC. Hal tersebut dikarenakan AIC memiliki hasil

nilai terbaik diantara model lainnya seperti BIC dan *Vuong's Test* (Pho et al., 2019, p. 299). Selain itu, AIC dapat memproses data deret waktu yang diolah menggunakan *machine learning*.

- (9) Melihat dan menganalisis hasil *Akaike Information Criterion* (AIC) terbaik.
- (10) Melatih model pada algoritma *Vector Autoregression* (VAR) berdasarkan *lag order* yang telah ditentukan sebelumnya.
- (11) Membuat pemodelan prediksi harga batubara Newcastle berjangka berdasarkan batasan periode yang telah ditentukan.
- (12) Mengembalikan data prediksi harga yang telah ditransformasi menjadi data yang sebenarnya.
- (13) Data yang telah dikembalikan selanjutnya didenormalisasi untuk mendapatkan nilai yang sesuai dengan kondisi nyata. Pada tahap ini prediksi harga (*High (d+1)* dan *Low (d+1)*) akan ditampilkan.
- (14) Evaluasi tingkat akurasi algoritma *Vector Autoregression* menggunakan *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE). Algoritma akan dikatakan memiliki nilai akurasi yang baik jika memiliki skor RMSE dan MAPE terendah daripada skor pada pengujian dengan menggunakan algoritma *machine learning* lainnya.