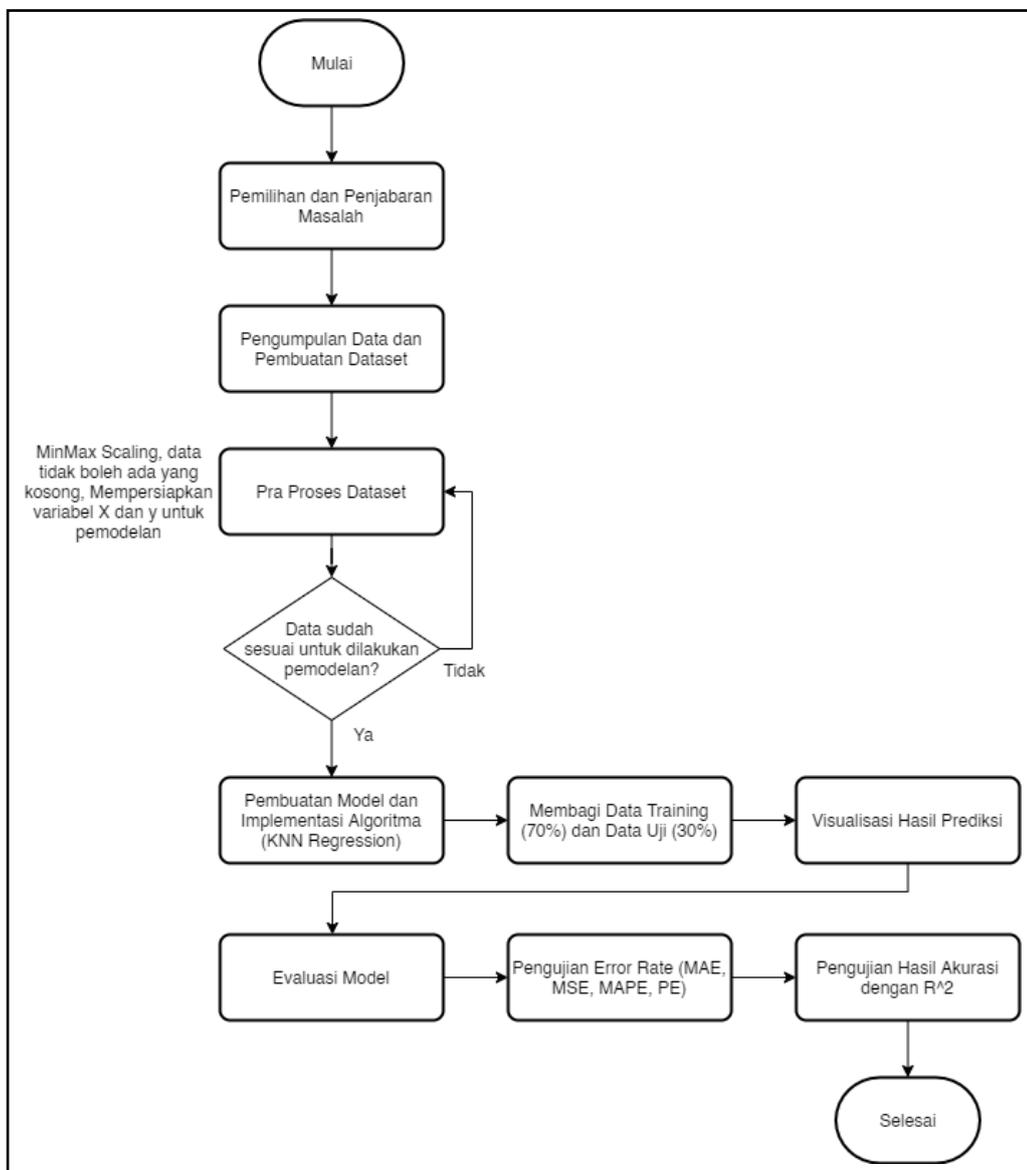


## BAB 3

### METODOLOGI PENELITIAN

#### 3.1. Metodologi Penelitian

Pada penelitian yang akan dilakukan, akan digunakan beberapa metode yang digambarkan pada *flowchart* sebagai berikut.



Gambar 3.1. *Flowchart* Metodologi Penelitian

### **3.1.1. Pemilihan dan Penjabaran Masalah**

Permasalahan yang dipilih adalah prediksi jumlah kasus Covid-19 di Kabupaten Karawang. Permasalahan akan diselesaikan menggunakan metode K-Nearest Neighbors Regression. Akan dibuat sebuah model untuk prediksi nilai dan kemudian dilakukan evaluasi model tersebut dengan pengujian akurasi dan *error rate*.

### **3.1.2. Pengumpulan Data dan Pembuatan Dataset**

Pengumpulan data akan dilakukan hanya di Kabupaten Karawang. Data akan diambil per hari dari web <https://covid19.karawangkab.go.id/> dan memuat beberapa variabel pengukur seperti jumlah kematian, jumlah pasien sembuh, dan jumlah kasus aktif. Data yang didapatkan akan dikumpulkan dalam bentuk *xlsx* untuk selanjutnya akan dilakukan pra proses menggunakan *Python*.

### **3.1.3. Pra Proses Dataset**

Pada pra proses, data yang didapatkan sebelumnya akan dinormalisasi menggunakan *min-max scaler*, serta *label encoding* sehingga mempermudah dalam pembuatan model. Pada tahap ini juga harus dipastikan bahwa kolom maupun baris yang ada di dataset tidak boleh ada yang kosong. Akan dibuat sebuah variabel baru untuk menampung variabel-variabel yang akan digunakan dalam pembuatan model. Variabel baru ini akan berbentuk vektor. Setelah semua terpenuhi, maka akan dilanjutkan ke pembuatan model.

### **3.1.4. Pembuatan Model dan Implementasi Algoritma**

Penelitian akan mengimplementasikan algoritma K-Nearest Neighbors Regression pada pembuatan model. Data hasil pra proses kemudian akan dibagi menjadi data *training* sebesar 70% dan data *testing* sebesar 30%. Pada pengaplikasiannya, akan dicari  $k$  sampel yang memiliki jarak terdekat dengan sampel uji. Pencarian jarak akan dilakukan menggunakan metode *Euclidian distance*. Kemudian akan dilakukan penghitungan prediksi nilai label untuk sampel uji dengan metode *mean method*.

### **3.1.5. Evaluasi Model**

Pada tahap ini akan diimplementasikan lima metode evaluasi, yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Percentage Prediction Error* (PE), *Mean Absolute Percentage Error* (MAPE), dan  $R^2$  *Metrics*. Metrik  $R^2$  akan digunakan untuk perhitungan akurasi dari model yang dibuat sedangkan MAE, MSE, MAPE, dan PE dapat digunakan untuk menghitung *error rate* pada model (Furqan *et al.*, 2020).

## **3.2. Perancangan Sistem**

### **3.2.1. Pembuatan Dataset**

*Dataset* yang dibuat diambil dari web resmi pemerintah Kabupaten Karawang yang khusus menangani kasus Covid-19 yaitu <https://covid19.karawangkab.go.id/data/>. Tampilan web dapat dilihat pada gambar 3.2. Data akan diambil per hari dan akan ditampung pada *dataset*.



Gambar 3.2. Web Covid-19 Kabupaten Karawang

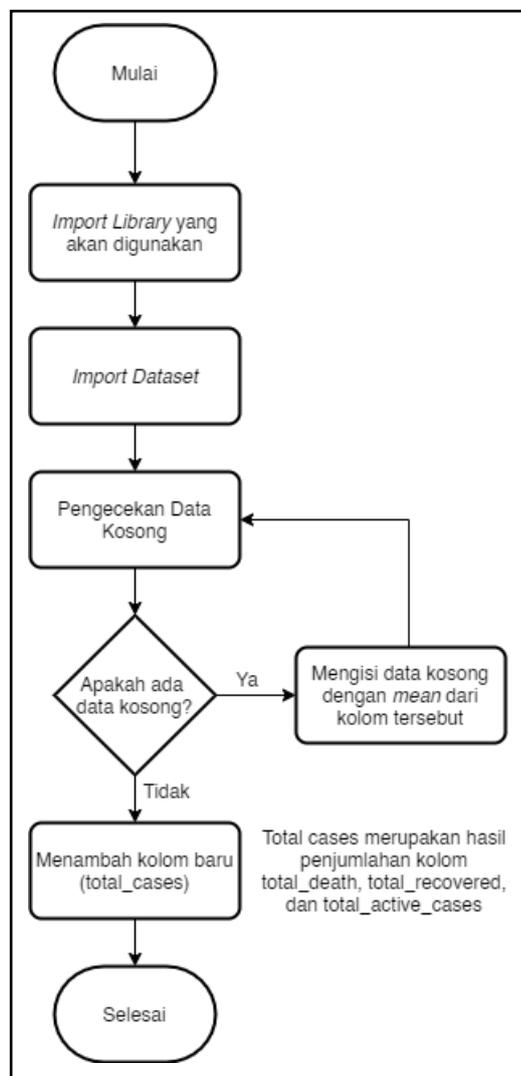
*Dataset* akan berbentuk *file xlsx*, dimana akan berisi variabel-variabel yang diambil dari web Covid-19 Kabupaten Karawang. Variabel yang diambil antara lain jumlah kasus baru per hari, jumlah kematian harian, jumlah kasus sembuh per hari, total kematian, total kasus sembuh, dan total kasus aktif. Dari *dataset* yang diambil, nantinya akan ditambahkan satu variabel baru bernama total kasus, dimana variabel ini merupakan hasil penjumlahan total kematian, total kasus sembuh, dan total kasus aktif. Perancangan *dataset* dapat dilihat pada tabel 3.1.

Tabel 3.1. Rancangan *Dataset* Covid-19 Kabupaten Karawang

Tanggal	Kasus baru per hari	Kematian baru per hari	Kasus sembuh baru per hari	Total kematian	Total kasus sembuh	Total kasus aktif	Total kasus
01/11/2020	22	8	19	61	1141	316	1518
02/11/2020	19	1	12	62	1153	322	1537
....	....	....	....	....	....	....	....
....	....	....	....	....	....	....	....
29/04/2021	26	2	46	588	17835	263	18686
30/04/2021	31	2	53	590	17888	239	18717

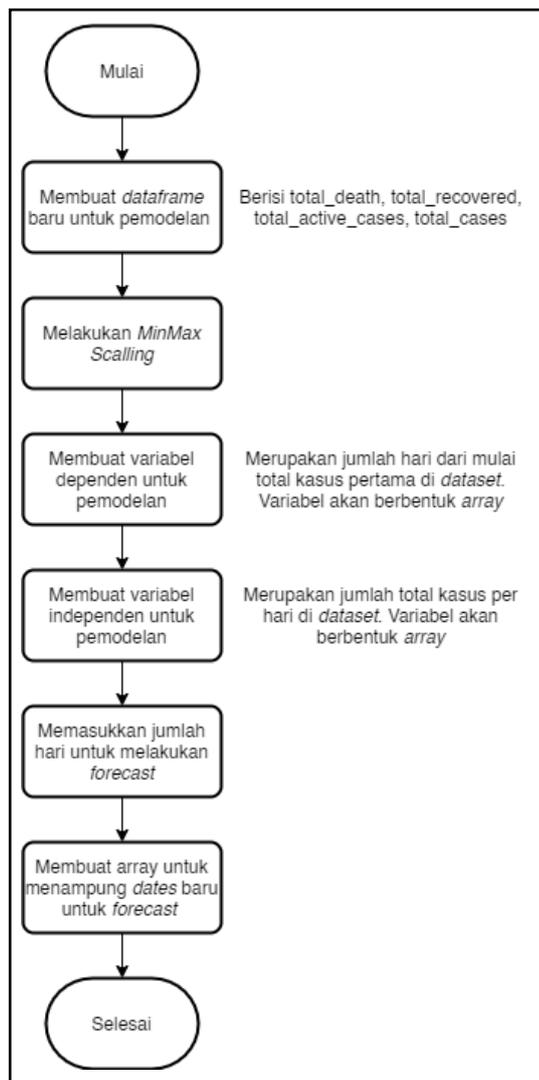
### 3.2.2. Flowchart Preprocessing

*Preprocessing* adalah fase pemrosesan awal *dataset* dan persiapan data-data yang diperlukan untuk pembentukan model. Pemrosesan awal *dataset* dapat dilihat pada gambar 3.3. Pada fase ini akan dilakukan *import* untuk setiap *library* yang akan dipakai, baik dalam pra proses, pembuatan model, maupun evaluasi. Selain itu akan dilakukan pengecekan *dataset* yang dipakai. Apabila terdapat data kosong pada *dataset*, maka akan dilakukan pengisian data kosong tersebut dengan menggantinya menjadi nilai rata-rata dari kolom yang bersangkutan.



Gambar 3.3. Pemrosesan Data Kosong pada *Dataset*

Setelah *dataset* dinilai sudah sesuai, maka akan dilanjutkan ke tahap pemrosesan untuk pemodelan. Pada tahap ini, akan disiapkan variabel – variabel untuk pemodelan, *scaling* data, serta variabel untuk melakukan *forecasting*. Detail tahap ini akan dijabarkan pada gambar 3.4.



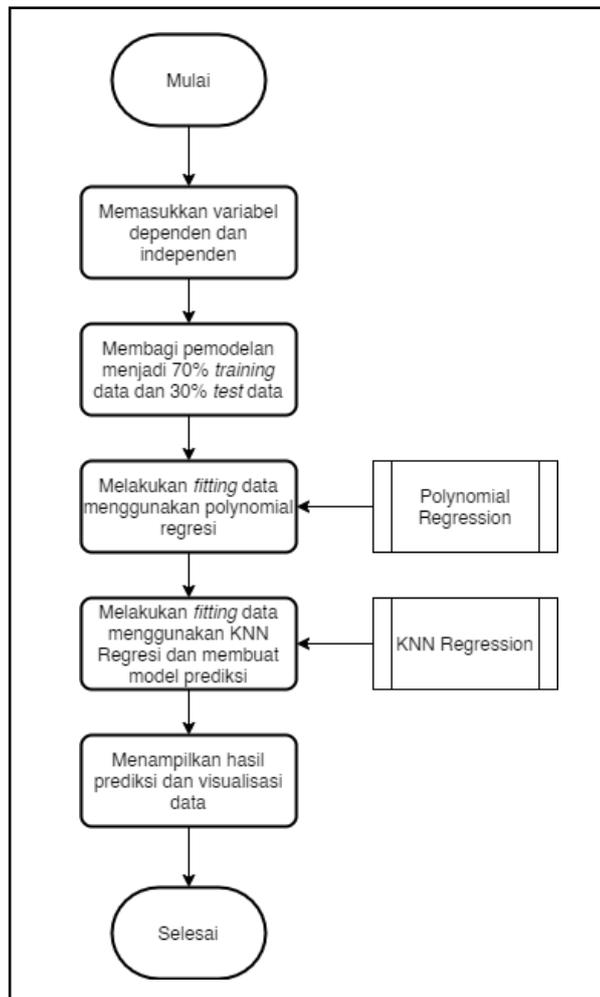
Gambar 3.4. Transformasi Data untuk Pemodelan

Sebelumnya akan dibuat sebuah *dataframe* baru yang berisi variabel – variabel yang dibutuhkan untuk pemodelan. Variabel tersebut yaitu jumlah total kasus, jumlah kematian, jumlah pasien sembuh, serta jumlah kasus aktif. Kolom

jumlah total kasus merupakan hasil penjumlahan dari ketiga kolom lainnya. Selanjutnya akan dilakukan *scalling* menggunakan *MinMax*. Alasan digunakannya *MinMax scalling* adalah karena nilai data cukup besar serta untuk visualisasi data sendiri, karena lebih mudah membaca angka kasus tertinggi atau terendah. Selain itu akan disiapkan variabel *dependen* dan variabel *independen* untuk pemodelan. Variabel *dependen* merupakan jumlah hari dari total kasus pertama di *dataset*, sedangkan variabel *independen* merupakan jumlah total kasus per hari. Masing-masing variabel akan berbentuk *array*. Selain itu, akan dibuat variabel berisi jumlah hari untuk melakukan *forecast*.

### **3.2.3. Flowchart Proses Regresi**

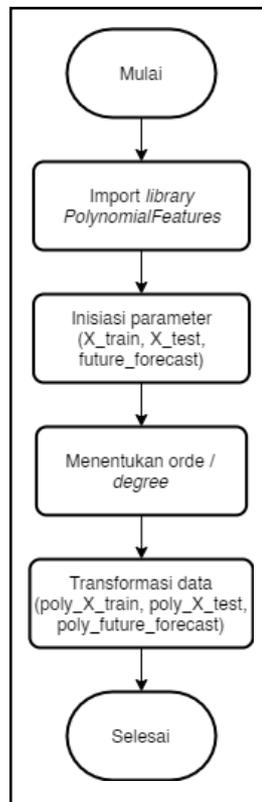
Pada pemodelan akan digunakan K-Nearest Neighbors Regression sebagai metode prediksi. Sebelum mengimplementasikan metode, akan dilakukan *fitting* terhadap data menggunakan Polynomial Regression dan kemudian dilanjutkan dengan KNN Regression. Pada pemodelan ini, data akan dibagi menjadi 70% sebagai data *training* serta 30% sebagai data *testing*. Alasan data *training* dibuat lebih besar karena metode KNN Regression membutuhkan data *training* yang lebih banyak agar data lebih terlatih dan meningkatkan akurasi model. Alur proses pemodelan dapat dilihat pada gambar 3.5.



Gambar 3.5. Alur Proses Pemodelan

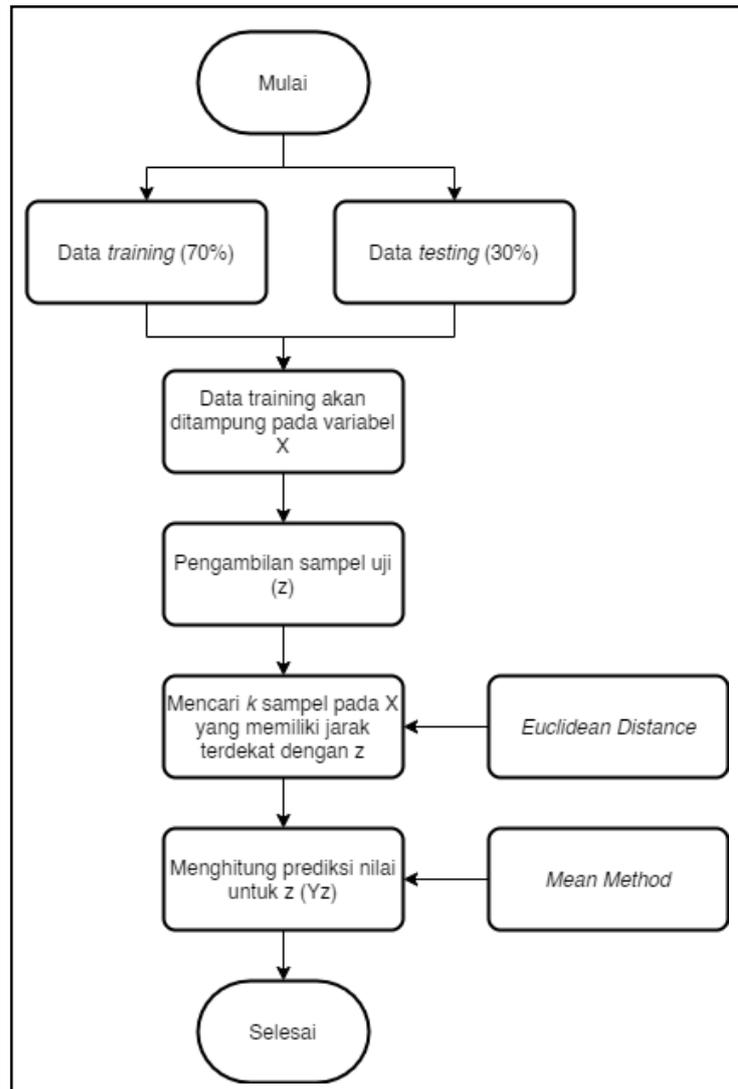
Regresi polynomial akan digunakan untuk melakukan *fitting* terhadap variabel sebelum digunakan untuk pemodelan menggunakan KNN Regression. Sebelumnya akan dilakukan inialisasi variabel yaitu  $X_{train}$ ,  $X_{test}$ , dan  $future\_forecast$ . Akan digunakan sebuah *library* bernama *PolynomialFeatures* dengan orde dua atau sering disebut sebagai fungsi *quadratic*, dimana fungsi ini akan memotong sumbu X sebanyak dua kali. Akan dilakukan transformasi data

terhadap variabel-variabel tersebut menjadi  $poly\_X\_train$ ,  $poly\_X\_test$ , dan  $poly\_future\_forecast$ . Alur proses regresi polynomial dapat dilihat pada gambar 3.6.



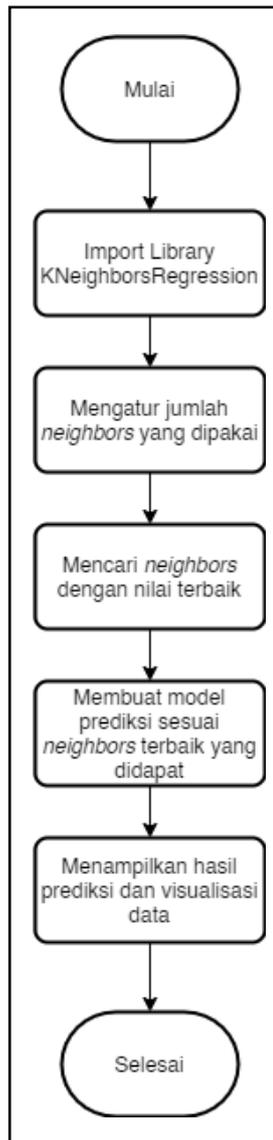
Gambar 3.6. Alur *Fitting* Data dengan Regresi Polynomial

KNN Regression merupakan metode yang digunakan untuk menganalisis nilai prediksi berdasarkan hubungan antara variabel *dependen* dan variabel *independen* (Xiao, Ma and Ding, 2018). Alur proses regresi menggunakan metode KNN Regression dapat dilihat pada gambar 3.7. Sebelumnya, data training akan ditampung pada variabel  $x$  serta data uji pada variabel  $z$ . Selain itu akan ditentukan jumlah  $k$  tetangga yang akan dipakai. Untuk mencari  $k$  tetangga yang memiliki jarak terdekat dengan  $z$  akan dilakukan perhitungan menggunakan *Euclidean distance*. Setelah telah mendapatkan  $k$  tetangga yang sesuai, maka akan dilakukan perhitungan nilai prediksi untuk sampel uji  $z$  menggunakan *mean method*.



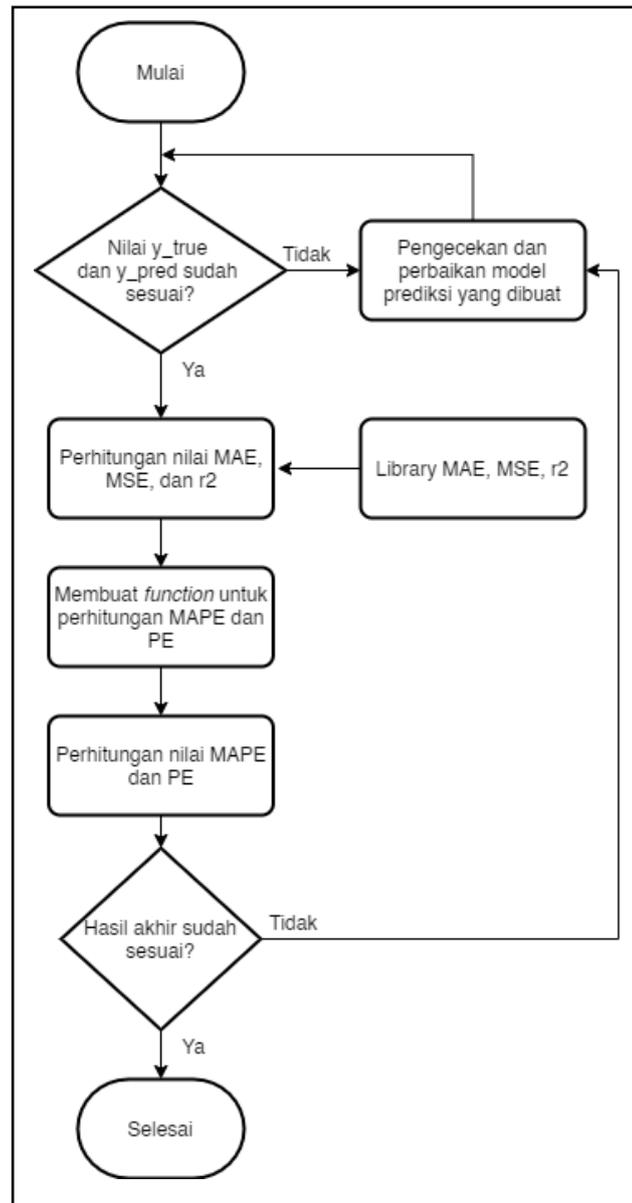
Gambar 3.7. Alur Kerja K-Nearest Neighbors Regression

Pada pengaplikasian metode KNN Regression di penelitian ini akan digunakan *library* bernama *KNeighborsRegressor*. Menggunakan *library* ini, dapat diatur jumlah  $k$  yang akan dipakai. Pada penelitian ini akan dilakukan perhitungan menggunakan  $k$  sebesar 5 dan 10. Akan dicari  $k$  terbaik untuk pemodelan menggunakan *best params*. Selanjutnya akan dibuat model prediksi serta visualisasi data yang dihasilkan. Alur kerja *library KNeighborsRegressor* dapat dilihat pada gambar 3.8.



Gambar 3.8. Proses Pemodelan dengan *Library KNeighborsRegressor*

Setelah didapatkan hasil prediksi, maka akan dilanjutkan evaluasi terhadap model. Untuk evaluasi menggunakan MAE, MSE, serta  $r^2$  akan digunakan *library* yang ada di *python*, sementara untuk evaluasi menggunakan MAPE dan PE akan dibuat sebuah *function* perhitungan, yang nantinya dapat langsung dipanggil saat diperlukan. Hal yang akan dibandingkan sendiri adalah nilai prediksi menggunakan KNN Regression dengan nilai aktual data. Detail model evaluasi dapat dilihat pada gambar 3.9.

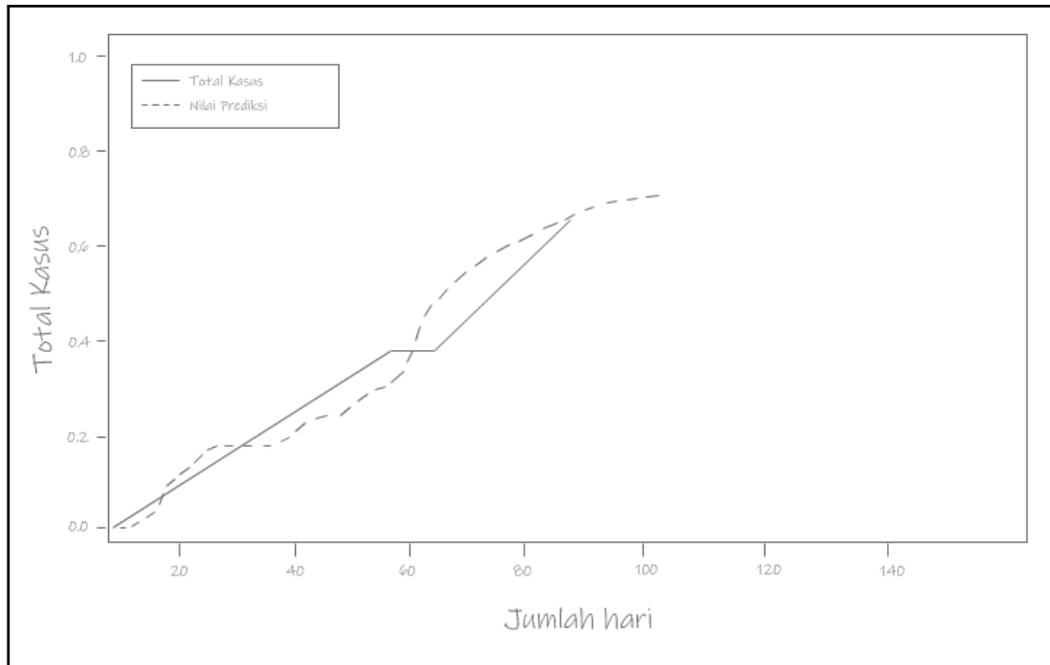


Gambar 3.9. Alur Evaluasi untuk Model

### 3.2.4. Tampilan Hasil Penelitian

Hasil penelitian akan ditampilkan dalam 2 bentuk visualisasi data, yaitu grafik dan tabel. Gambar 3.10 menunjukkan rancangan tampilan hasil penelitian

dalam bentuk grafik. Grafik akan menggambarkan perbandingan visualisasi antara data aktual dengan data hasil prediksi.



Gambar 3.10. Rancangan Tampilan Hasil Penelitian (Grafik)

Gambar 3.11 menunjukkan rancangan tampilan hasil penelitian dalam bentuk tabel. Tabel akan berisi tanggal dan hasil nilai prediksi yang didapat.

	Date	Nilai Prediksi
1	16/04/2021	17895.0
2	17/04/2021	17926.0
**	*****	*****
**	*****	*****

Gambar 3.11. Rancangan Tampilan Hasil Penelitian (Tabel)