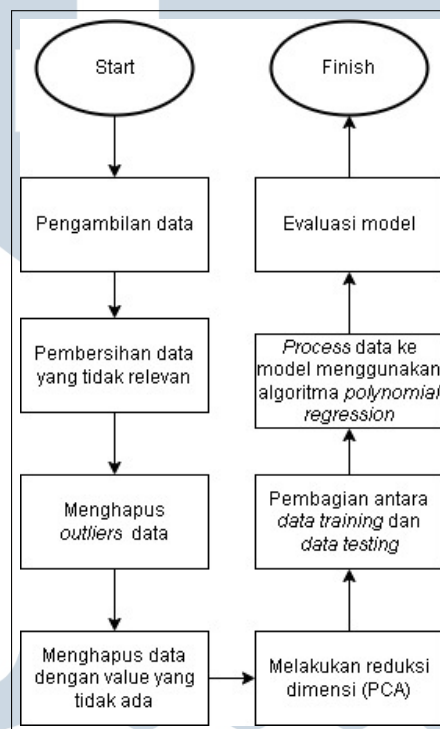


### BAB 3 METODOLOGI PENELITIAN

Untuk mencapai tujuan penelitian ini, terdapat metodologi penelitian yang digunakan untuk melakukan penelitian ini. Metode penelitian yang diambil memiliki beberapa langkah yang dijalankan, seperti studi literatur tentang *Polynomial Regression*, pencarian dataset, pengambilan data, pembersihan data, menghapus *ouliers*, menghapus data kosong, melakukan reduksi dimensi, pembagian data, proses data dan evaluasi model. Proses pembuatan model terlampir pada Gambar 3.1 dalam bentuk *flowchart*.



Gambar 3.1. Metode Penelitian Skripsi

#### 3.1 Studi Literatur

Studi literatur dilakukan dengan mencari dan membaca jurnal atau *paper* terkait dari algoritma yang dipakai yaitu *polynomial regression*. Studi literatur bertujuan juga untuk mencari informasi terkait pengambilan data, pembersihan data, penghapusan data, dan lain-lain. Studi literatur dilakukan untuk menambah pengetahuan dan mencari pembandingan antara hasil penelitian yang sedang

dijalankan dengan penelitian yang sudah ada sehingga bisa dijadikan sumber referensi dalam penelitian. *Paper* dan jurnal yang dicari merupakan *paper* dan jurnal dari rentang tahun 2017-2024. *Paper* dan jurnal tersebut memuat penelitian terdahulu dan menggunakan teknologi yang tidak ketinggalan zaman.

### 3.2 Pencarian Dataset

Dalam tahap ini, data dari sebuah rumah di Belgia yang berisi kelembapan ruangan, suhu ruangan, suhu luar ruangan, jumlah bohlam lampu, dan lain-lain. *Dataset* ini digunakan untuk membantu memprediksi angka penggunaan listrik rumah sehingga banyaknya karbon yang keluar dari rumah tersebut dapat diprediksi. Data yang ditemukan dicatat dalam ukuran per 10 menit di dalam sebuah rumah, Data tersebut dicatat dari musim dingin ke musim semi dikarenakan di Belgia terdapat 4 musim. *Dataset* ditemukan di Kaggle dan merupakan data yang dipublikasi pada tahun 2017.

### 3.3 Pengambilan Data

Data yang diambil merupakan data dari Kaggle. Data yang diambil memiliki kurang lebih 19.000 baris data. Data yang sudah diambil kemudian diberikan label baru sehingga data yang digunakan lebih familiar dan mudah untuk dibaca. Tahap ini membuat list dari setiap data dapat dibaca dengan lebih mudah. Label dibuat dengan mengubah kepala kolom menjadi *kitchen\_temp*, *kitchen\_hum*, *office\_temp*, *office\_hum*, *bedroom\_temp*, *bedroom\_hum*, dan lain-lain. Hal ini dilakukan karena pada *dataset* tidak tertera ruangan yang dimaksud adalah ruangan apa sehingga sulit untuk mengartikan maksud dari *dataset* yang sudah dibuat.

### 3.4 Pembersihan Data

Dalam *dataset* yang sudah digunakan terdapat beberapa data yang tidak relevan sehingga perlu dibersihkan. Contohnya data *appliances\_energy* data ini merupakan salah satu data yang akan diprediksi sehingga data ini harus dibuang dari *dataset* sehingga model dapat bekerja dengan lebih optimal. Data yang dibuang selanjutnya adalah data *rv1* dan *rv2* data ini merupakan data yang tidak relevan dengan penelitian ini sehingga data tersebut dihapus agar tidak ikut diprediksi dalam pembuatan model. Pembersihan data ini dilakukan beberapa kali sampai data yang akan digunakan sesuai.

### 3.5 Menghapus Outliers

Dalam *dataset* yang sudah diambil terdapat nilai-nilai yang tidak masuk ke dalam model sebagai nilai yang digunakan dalam prediksi. *Outliers* tersebut dihapus sehingga data yang tersisa merupakan data yang relevan dan berdampak dalam prediksi nilai yang akan dijalankan oleh model yang dibuat. Penghapusan *outliers* dilakukan dengan menghitung terlebih dahulu data yang dibuat. Kemudian dicari nilai yang tidak berada di bawah dan di atas nilai yang akan diprediksi. Data tersebut selanjutnya akan dihapus sehingga nilai yang akan diprediksi berkurang jumlahnya. Penghapusan *outliers* ini dilakukan untuk menghasilkan hasil yang lebih akurat.

### 3.6 Menghapus Data yang Kosong

Pada *dataset* yang telah diolah, terdapat beberapa baris data yang kosong. Baris data yang kosong ini akan dihapus dan tidak dipergunakan untuk proses pelatihan model yang akan dibuat. Hal ini bertujuan untuk mengurangi baris data yang akan diprediksi oleh model sehingga model dapat memprediksi lebih sedikit baris yang membuat model dapat bekerja dengan lebih baik.

### 3.7 Melakukan Reduksi Dimensi

Selanjutnya, dimensi dari data akan direduksi sehingga data yang berkategori serupa dapat menjadi satu baris data. Hal ini dilakukan sehingga data yang dicek atau diprediksi dapat berkurang sehingga dapat lebih mudah di prediksi dan hasil dari pengukuran metrik dapat lebih baik. Reduksi dilakukan menjadi 10 kategori besar dari 34 kolom yang ada dalam dataset dikarenakan banyak kolom yang dinilai memiliki kemiripan di dalam isi datanya sehingga kemudian dijadikan ke dalam satu kategori atau kolom yang sama.

### 3.8 Pembagian Data

Data yang sudah ada dibagi dengan perbandingan 80:20. Data yang digunakan untuk pelatihan model ialah sebesar 80 persen dan data yang digunakan untuk pengujian model sebesar 20 persen data. Pembagian ini dilakukan agar model dapat berlatih dengan waktu yang cukup dan juga data yang menjadi *testing* dapat mewakili keseluruhan data dengan baik.

### 3.9 Proses Data

Data yang sudah dibagi ke dalam perbandingan 80:20 akan diproses menggunakan model yang sudah dibuat. Model yang sudah dibuat menggunakan algoritma *polynomial regression* akan memulai *training* dengan 80 persen data dan *testing* dengan 20 persen data. Proses ini akan berlangsung terus menerus hingga hasil yang dikeluarkan dinilai maksimal.

### 3.10 Evaluasi Model

Setelah melakukan proses prediksi, model akan dievaluasi menggunakan metrik Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R-squared, dan Mean Absolute Percentage Error (MAPE). Proses evaluasi ini dilakukan hingga hasil dari proses yang dilakukan maksimal dan sesuai yang diinginkan oleh penulis. Evaluasi model dilakukan agar dapat mengetahui nilai *loss* dari model. Pada metrik RMSE dan MAE, semakin mendekati 0 nilainya, maka model memiliki akurasi yang semakin baik. Pada metrik R-Squared jika nilai R-Squared semakin mendekati 1 artinya model semakin baik. Sementara pada MAPE semakin kecil nilai presentase dari metrik, maka model semakin baik.

