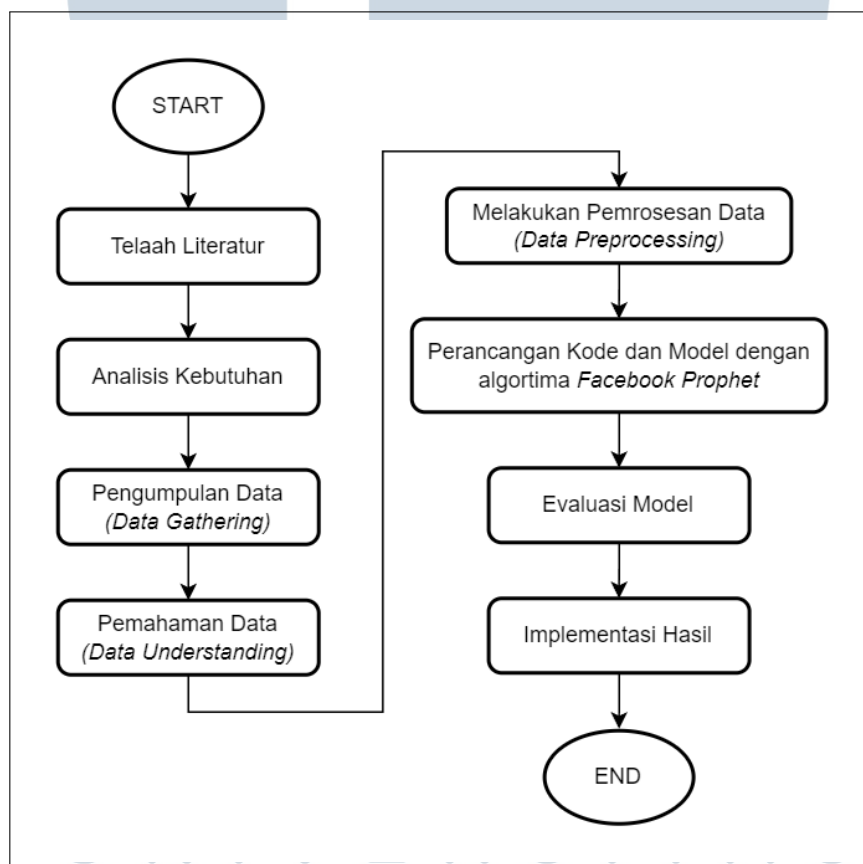


BAB 3 METODOLOGI PENELITIAN

3.1 Metodologi Penelitian

Metodologi penelitian adalah sebuah aturan ilmiah yang umumnya digunakan sebagai upaya dalam menemukan atau mendapatkan data untuk suatu tujuan tertentu, yang dalam prosesnya dapat membantu dalam menjabarkan secara ilmiah mengenai langkah demi langkah proses yang dilakukan dalam penyusunan serta pengerjaan dari sebuah penelitian.



Gambar 3.1. *Flowchart* Metodologi Penelitian.

Gambar 3.1 merupakan *flowchart* dari tahap-tahap dan metodologi yang dilakukan dalam pembuatan model prediksi penggunaan listrik harian dengan menggunakan algoritma *Facebook Prophet model*.

3.1.1 Telaah Literatur

Pada tahap Telaah Literatur, dilakukan proses analisis yang mendalam mengenai teori-teori serta tinjauan literatur terhadap hasil dari penelitian terdahulu yang memiliki topik maupun objek penelitian yang berkaitan dengan topik penelitian, seperti teknik prediksi *time series*, metode serta implementasi algoritma *FB Prophet*, dan juga faktor-faktor yang dapat mempengaruhi penggunaan daya listrik harian. Adapun sumber yang kemudian dijadikan referensi merupakan data penelitian dengan rentang waktu 5 tahun terakhir.

3.2 Analisis Kebutuhan

Dalam tahap analisis kebutuhan, dilakukan analisis mengenai kebutuhan yang diperlukan untuk dapat melakukan penelitian ini. Berdasarkan tinjauan literatur yang telah dilakukan, kemudian hasil dari analisis kebutuhan akan ditekankan pada tujuan penelitian, yaitu pada manfaat hasil akhir atau *output* dari penelitian.

3.2.1 Pengumpulan Data (*Data Gathering*)

Pada tahap pengumpulan data, dilakukan pencarian serta pengumpulan data yang sesuai dengan topik penelitian. Sehingga untuk melakukan penelitian ini dibutuhkan kumpulan data penggunaan listrik yang akan digunakan untuk kebutuhan analisis dan prediksi.

Dataset yang akan digunakan merupakan data sekunder diperoleh melalui sumber terpercaya seperti Kaggle. Adapun data yang digunakan merupakan data penggunaan listrik harian di kota Tetouan, Morocco pada tahun 2017, serta data faktor-faktor yang memengaruhi penggunaan listrik harian seperti *temperature*, *general diffuse flows*, *diffuse flows*, *humidity* dan *wind speed*.

3.2.2 Pemahaman Data (*Data Understanding*)

Dataset yang digunakan untuk penelitian ini berasal dari *dataset Electric Power Consumption* yang diperoleh dari laman Kaggle, yang memiliki kategori *Energy* dan *Electricity*. Dataset ini terdiri dari sebuah *file* atau berkas yang disimpan dengan nama *file* "*powerdata*" dengan format *file* ".csv" yang berukuran 4,33 MB.

Adapun potongan isi dari *dataset* yang digunakan dapat dilihat pada Tabel 3.1 dan Tabel 3.2 berikut.

Tabel 3.1. Data Penggunaan Listrik kota Tetouan, Maroko tahun 2017 - Bagian 1.

Datetime	T	H	W	GDF	DF
1/1/2017 0:00	6.559	73.8	0.083	0.051	0.119
1/1/2017 0:10	6.414	74.5	0.083	0.07	0.085
1/1/2017 0:20	6.313	74.5	0.08	0.062	0.1
...
12/30/2017 23:30	6.9	72.8	0.086	0.084	0.074
12/30/2017 23:40	6.758	73	0.08	0.066	0.089
12/30/2017 23:50	6.58	74.1	0.081	0.062	0.111

Tabel 3.2. Data Penggunaan Listrik kota Tetouan, Maroko tahun 2017 - Bagian 2.

Datetime	PC1	PC2	PC3
1/1/2017 0:00	34055.6962	16128.87538	20240.96386
1/1/2017 0:10	29814.68354	19375.07599	20131.08434
1/1/2017 0:20	29128.10127	19006.68693	19668.43373
...
12/30/2017 23:30	29590.87452	25277.69254	13806.48259
12/30/2017 23:40	28958.1749	24692.23688	13512.60504
12/30/2017 23:50	28349.80989	24055.23167	13345.4982

Tabel tersebut merupakan potongan isi dari *dataset* yang digunakan. Adapun masing-masing keterangan dari setiap kolom adalah sebagai berikut.

- **Datetime**: Mencatat waktu dan tanggal pengambilan data dengan format "DD/MM/YYYY HH:MM" untuk analisis tren dan pola. (Satuan: tanggal dan waktu)
- **T (Temperature)**: Ukuran suhu udara yang berpengaruh pada penggunaan energi, khususnya dalam sistem pemanasan dan pendinginan. (Satuan: °C)
- **H (Humidity)**: Tingkat kelembaban udara yang mempengaruhi kenyamanan termal dan penggunaan energi alat pengatur kelembaban. (Satuan: %)
- **WS (Wind Speed)**: Kecepatan angin yang mempengaruhi permintaan energi pada bangunan, terutama dalam sistem ventilasi. (Satuan: km/h)
- **GeneralDiffuseFlows**: Total aliran difusi radiasi surya yang diterima, penting untuk analisis generasi energi solar. (Satuan: kW/m²)

- ***DiffuseFlows***: Aliran difusi radiasi surya yang lebih terfokus, penting untuk pemahaman distribusi radiasi yang diterima. (Satuan: kW/m²)
- ***PC1 (Power Consumption Zone 1)***: Konsumsi daya listrik di stasiun Quads, Tétouan, digunakan untuk analisis kebutuhan energi lokal. (Satuan: kW)
- ***PC2 (Power Consumption Zone 2)***: Konsumsi daya listrik di stasiun Smir, Tétouan, penting untuk pemahaman distribusi dan penggunaan energi di area tersebut. (Satuan: kW)
- ***PC3 (Power Consumption Zone 3)***: Konsumsi daya listrik di stasiun Boussafou, Tétouan, memberikan data untuk pemantauan efisiensi energi. (Satuan: kW)

Data tersebut menunjukkan jumlah energi yang dikonsumsi pada tiga wilayah stasiun berbeda yakni, Quads, Smir, dan Boussafou, yang termasuk dalam jaringan distribusi listrik kota Tetouan. Ketiga stasiun ini merupakan stasiun pemasok listrik utama pada setiap zona untuk mengirimkan listrik ke berbagai bagian kota dan berperan dalam infrastruktur energi kota.

3.3 Praproses Data (*Data Preprocessing*)

Pada tahap ini, dijalankan tahap pra-pemrosesan data, yang mencakup kegiatan pembersihan data, penanganan nilai yang tidak lengkap, dan transformasi data agar dapat digunakan secara efektif dalam proses pemodelan. Tujuan dari data *preprocessing* adalah untuk membersihkan, mengubah, dan mempersiapkan data agar dapat digunakan dalam pemodelan. Berikut merupakan penjelasan mengenai tahapan data *preprocessing* yang dilakukan.

3.3.1 Eksplorasi Data (*Data Exploration*)

Pada tahap eksplorasi data, dilakukan analisis terhadap *dataset* di mana kita berusaha untuk memahami karakteristik pada *dataset*. Tahap ini dilakukan untuk dapat mengidentifikasi pola, yang mencakup beberapa kegiatan seperti, pemahaman data, normalisasi data, dan visualisasi data. Berikut merupakan tahap-tahap yang dijalankan pada proses eksplorasi data.

A *Data Understanding*

Proses *data understanding*, merupakan langkah awal yang dilakukan dalam eksplorasi data di mana dilakukan pemahaman mengenai struktur dan isi dalam *dataframe*. Pada proses ini dilakukan ekstraksi terhadap kolom *datetime* yang digunakan untuk mengelompokkan data deret waktu sesuai dengan hari, minggu, bulan dan musim sehingga mudah untuk memvisualisasikan *dataframe* sesuai dengan kategori tertentu.

Proses selanjutnya adalah melihat deskripsi variabel untuk memahami atribut-atribut yang ada, tipe datanya, serta jumlah data yang tersedia untuk setiap atribut. Sedangkan deskripsi statistik dilakukan untuk memahami kolom-kolom numerik dalam dataset. Deskripsi ini mencakup informasi seperti jumlah data, rata-rata, standar deviasi, nilai minimum dan maksimum, serta quartil.

B *Data Normalization*

Data normalization adalah proses mengubah data ke dalam skala yang konsisten untuk memastikan bahwa semua fitur berkontribusi secara seimbang dalam analisis data dan algoritma machine learning. Pada proses ini normalisasi data dilakukan terhadap 5 komponen variabel regresor yang memiliki rentang jarak antara min dan max cukup besar, proses normalisasi dilakukan dengan menggunakan *standard scaler*.

Standard scaler merupakan salah satu teknik normalisasi data yang digunakan dalam prapemrosesan data. Teknik ini bertujuan untuk mengubah data sedemikian rupa sehingga memiliki distribusi dengan nilai rata-rata nol dan deviasi standar satu. *Standard scaler* digunakan untuk memastikan bahwa setiap fitur dalam dataset memiliki skala yang sama. Ini membantu dalam memperbaiki kinerja model pembelajaran mesin dengan mengurangi bias yang disebabkan oleh skala yang berbeda dari fitur yang berbeda.

Adapun rumus dari *standard scaler* adalah sebagai berikut.

$$z = \frac{x - \mu}{\sigma} \quad (3.1)$$

Keterangan:

- z : nilai yang ditransformasikan dari fitur,
- x : nilai asli dari fitur,

- μ : nilai rata-rata dari fitur,
- σ : standar deviasi dari fitur.

Berikut merupakan rumus perhitungan rata-rata (*mean*) beserta keterangan dari masing-masing komponen dapat dilihat pada persamaan 3.2.

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.2)$$

Keterangan:

- μ : simbol yang digunakan untuk rata-rata dari dataset.
- n : jumlah total elemen atau observasi dalam dataset.
- x_i : elemen individu dalam dataset, di mana i adalah indeks yang berjalan dari 1 sampai n .

Sedangkan rumus perhitungan standar deviasi dan juga keterangan untuk masing-masing komponen dapat dilihat pada persamaan 3.3.

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \quad (3.3)$$

Keterangan:

- σ (sigma) : simbol yang digunakan untuk standar deviasi dari *dataset*.
- n : jumlah total elemen atau observasi dalam *dataset*.
- x_i : elemen individu dalam *dataset*, di mana i adalah indeks yang berjalan dari 1 sampai n .
- μ : rata-rata dari semua elemen x_i dalam *dataset*.

C *Data Visualization*

Tahap visualisasi data merupakan sebuah proses untuk mengubah data yang terkandung dalam sebuah dataframe menjadi grafik atau plot. Proses ini sangat penting untuk menampilkan data secara grafis, yang memudahkan identifikasi pola, tren, dan hubungan yang mungkin tidak langsung terlihat melalui analisis statistik saja. Visualisasi data umumnya digunakan untuk mempresentasikan data dalam

bentuk visual seperti grafik, diagram, dan peta. Metode ini membantu dalam memberikan pemahaman dan mempermudah menangkap pola, tren, dan data yang menyimpang, yang membantu dalam membuat interpretasi yang lebih efektif [41].

3.4 Perancangan Kode dan Model

Pada tahapan perancangan kode dan model, melibatkan beberapa langkah, seperti pemilihan fitur, penyesuaian parameter model, dan validasi model. Pemilihan fitur dilakukan untuk memilih variabel yang paling relevan dalam pemodelan.

Penyesuaian parameter model dilakukan untuk menentukan parameter yang optimal dalam model prediksi. Sedangkan validasi terhadap model dilakukan untuk dapat memastikan bahwa model yang dirancang dapat memberikan hasil yang akurat dan dapat diandalkan.

3.5 Evaluasi Model

Tahap ini dilakukan dengan melakukan uji coba terhadap model untuk melakukan prediksi terhadap data *test* serta melibatkan evaluasi kinerja model prediksi menggunakan data penggunaan listrik harian. Dilakukan uji coba untuk memastikan bahwa model mampu memberikan prediksi yang akurat dan dapat diandalkan. Proses evaluasi kinerja model *forecasting Facebook Prophet*, menggunakan metrik keakuratan seperti *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, dan *Mean Absolute Percentage Error (MAPE)*. Analisis ini bertujuan untuk dapat membantu dalam mengidentifikasi kelebihan serta kekurangan model dalam memprediksi data masa depan, serta memberikan rekomendasi untuk peningkatan.

Keempat metrik ini (MAE, MSE, RMSE, dan MAPE) merupakan bagian pada tahap evaluasi yang penting untuk dilakukan dalam *time series forecasting*, yang membantu dalam melihat nilai akurasi dari model yang digunakan dengan hasil yang didapatkan dari perhitungannya. Sehingga dapat memberikan pandangan yang komprehensif tentang performa model dari berbagai sudut pandang, serta membantu untuk mengidentifikasi kekurangan dan perbaikan dari model yang diimplementasikan.

3.6 Implementasi Hasil

Langkah terakhir dalam penelitian ini adalah melakukan implementasi hasil terhadap *dataset dummy* yang memiliki informasi data berupa *timeseries* dan juga informasi dari masing-masing *regressors*. Implementasi ini dilakukan dengan tujuan untuk memprediksi penggunaan listrik dengan interval 1 jam untuk dua bulan ke depan. Adapun informasi tabel lengkap mengenai *dataset dummy* yang digunakan dapat dilihat pada Tabel 3.3 berikut.

Tabel 3.3. Data *Dummy Timeseries* dan Regressor tahun 2018.

Datetime	T	H	WS	GDF	DF
1/1/2018 0:00	7.559	73.8	83	51	119
1/1/2018 1:00	6.714	74.5	83	0.07	85
1/1/2018 2:00	6.513	74.5	0.08	62	0.1
...
3/1/2018 21:00	10.69	86.3	85	77	174
3/1/2018 22:00	10.33	86.5	83	73	108
3/1/2018 23:00	10.74	86.5	88	66	96

Apabila dilihat pada tabel tersebut, *dataset dummy* berisikan kolom *datetime* dan kelima *regressor* yang akan digunakan oleh model untuk melakukan prediksi terhadap penggunaan listrik selama beberapa hari kedepan dengan mempertimbangkan data-data tersebut.