

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1 Gambaran Umum Objek Penelitian**

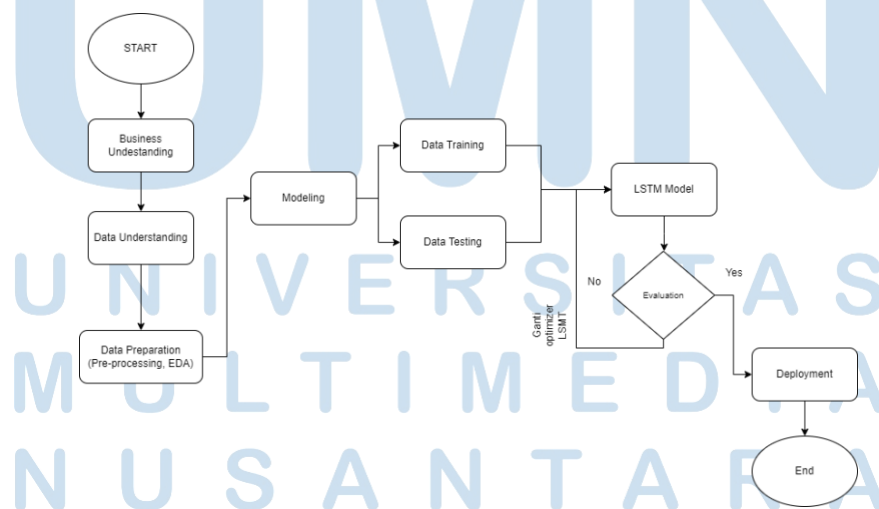
Penelitian ini dilakukan dengan objek penelitian yaitu data *night-time light* di Jakarta yang berasal dari bangunan-bangunan yang dipilih secara random. Untuk pemilihan kota yang dilakukan yaitu hanya kota DKI Jakarta saja, karena DKI Jakarta merupakan salah satu kota terbesar di Indonesia dan juga DKI Jakarta sendiri merupakan ibukota negara. Data *night-time light* sendiri diambil melalui platform google earth engine yaitu sebuah platform yang dimiliki oleh Google yang dimana platform ini merupakan *computing platform* yang dapat mengizinkan pengguna untuk melakukan *geospatial* analisis dalam infrastruktur yang dimiliki oleh Google. *Dataset* yang digunakan dalam penelitian ini adalah *VIIRS Stray Light Corrected Nighttime Day/Night Band Composites Version 1*. Data yang didapat dari *dataset* ini adalah data gambar yang kemudian diubah menjadi numerik untuk bisa diolah dengan menggunakan nilai median dari setiap bulan yang ada. Data yang tersedia di dalam *dataset* itu dari tahun 2012 bulan April hingga 2023 bulan Januari, dikarenakan penelitian ini ingin melihat pada saat terjadinya pandemic covid-19 saja maka data yang diambil mulai dari Januari 2019 hingga Juni 2022. Data yang digunakan berisikan 382 baris data dengan 51 atribut secara keseluruhan yaitu *ID, Name, Address, District, Type, Latitude, Longitude*, 2019-01-01 hingga 2022-06-01.

#### **3.2 Metode Penelitian**

##### **3.2.1 Alur Penelitian**

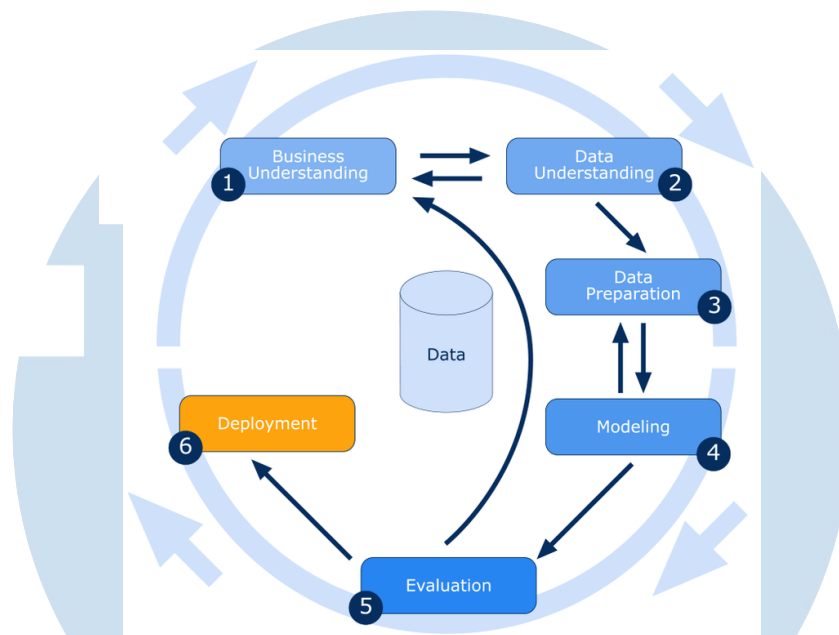
Alur Penelitian ini, setelah menemukan topik yang ingin diteliti maka proses selanjutnya dalam melakukan penelitian ini yaitu melakukan pengumpulan data yang dimana data tersebut didapatkan dari Google Earth Engine. Setelah mendapatkan data tersebut dari Google Earth Engine, data yang didapat merupakan data gambar yang kemudian di proses untuk menjadi data numerik sehingga mendapatkan nilai dari intensitas cahaya dari setiap bangunan yang di pilih. Pada

saat data sudah menjadi numerik, maka proses selanjutnya yaitu melakukan *Pre-Processing Data*, tujuannya yaitu ingin menemukan apakah data tersebut terdapat *missing value*, kemudian memilih mana saja dari atribut yang akan digunakan dalam proses selanjutnya yaitu *Exploratory Data analysis* (EDA). EDA ini sendiri bertujuan untuk melihat pola dari data yang akan diteliti, kemudian menemukan juga *anomaly* pada kumpulan data yang ada. Setelah dilakukan EDA, maka data akan langsung dibagi menjadi 2 yaitu *data training* dan *data testing*. Pembagian data tersebut akan digunakan untuk membuat LSTM model. Dalam membuat LSTM model ini sendiri juga diperlukan mengatur estimator yang diinginkan agar dapat menjadi model tersebut. Pada saat estimator pada model LSTM sudah di atur maka akan menghasilkan prediksi dengan hasil grafik, kemudian setelah itu akan dinilai atau di evaluasi dengan cara melihat nilai RMSE dari model yang sudah dibuat. Ketika nilai RMSE besar maka hasil prediksi dari model tersebut kurang akurat, tetapi apabila model tersebut memiliki nilai RMSE nya kecil maka hasil akurasi dari model tersebut bagus. Ketika, nilai RMSE hasilnya tidak bagus maka akan dilakukan penggantian optimizer dari model LSTM tersebut untuk mendapatkan hasil yang terbaik dan dapat mengetahui optimizer mana yang paling bagus. Setelah mendapatkan hasil yang bagus maka model tersebut akan dibuatkan visualisasi yaitu plot untuk melihat hasil prediksi tersebut agar mudah untuk dilihat. Seluruh alur penelitian ini dapat dilihat secara ringkas pada Gambar 3.1 dibawah ini.



**Gambar 3. 1 Alur Penelitian**  
**Sumber: mandiri**

### 3.2.2 Metode Data Mining



Gambar 3. 2 Tahapan CRISP-DM  
Sumber: Google Image

Dalam melakukan analisis, metode yang di gunakan yaitu CRISP-DM. CRISP-DM (*Cross Industry Standard Process for Data Mining*) sendiri adalah metode yang paling umum yang menggunakan model pemrosesan untuk pengembangan data dan juga yang paling banyak digunakan oleh para ahli untuk memecahkan suatu masalah. Gambar 3.2 diatas menjelaskan bahwa CRISP-DM sendiri memiliki 6 tahapan yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Pemilihan CRISP-DM sendiri dikarenakan metode ini merupakan metode yang paling umum digunakan dalam tahapan *data science*[43]. Berikut ini merupakan tahapan dari CRISP-DM yang diimplementasikan pada penelitian ini, antara lain:

#### 1. *Business Understanding*

*Business Understanding* adalah tahapan paling pertama dalam metode CRISP-DM yang bertujuan untuk mengetahui terlebih dahulu tujuan dilakukannya penelitian berdasarkan kebutuhan. Tujuan dilakukannya penelitian ini sendiri yaitu untuk mengetahui nilai intensitas cahaya dan juga memprediksi nilai tersebut di provinsi DKI Jakarta dan

juga akan mengetahui *optimizer* dari tensorflow manakah yang paling baik untuk data dalam penelitian ini.

## 2. *Data Understanding*

Tahapan kedua, setelah mengetahui tujuan dari dilakukannya penelitian ini yaitu tahapan *data understanding*. Tahapan ini merupakan tahapan yang terdiri dari pengumpulan data, melakukan deskripsi pada data yang akan digunakan, mengevaluasi data, dan pemilihan atribut yang akan digunakan[44]. Penelitian ini menggunakan data intensitas cahaya pada kota DKI Jakarta dengan memilih sekitar 382 bangunan di wilayah yang tersebar di DKI Jakarta yang diambil melalui dataset *VIIRS Stray Light Corrected Nighttime Day/Night Band Composites Version 1* pada Google Earth Engine. Dataset itu sendiri menyediakan data dari 1 April 2012 hingga terakhir yaitu 1 Januari 2023, akan tetapi yang digunakan dalam penelitian ini yaitu dari 1 Januari 2019 hingga 1 Juni 2022

## 3. *Data Preparation*

Tahapan ketiga yaitu *data preparation*. *Data preparation* sendiri adalah tahapan dimana akan dilakukannya perbaikan terhadap data yang akan di bentuk atau dijadikan model. Tahapan ini mempunyai peranan yang penting dalam penelitian ini, karena hasil dari model yang akan di hasilkan akan dipengaruhi oleh data yang akan digunakan[45]. Terdapat beberapa proses yang dilakukan dalam penelitian ini yaitu *formatting data*, penambahan kolom data, hingga melakukan *feature selection*. *Formatting data* sendiri dilakukan guna untuk mengecek setiap baris data yang akan digunakan. Penambahan kolom data sendiri dilakukan untuk memberikan informasi tambahan dalam analisa yang dilakukan. Terakhir yaitu *feature selection* dilakukan guna untuk melakukan pemilihan fitur mana yang akan digunakan dan membuang variable yang tidak digunakan dalam penelitian ini.

#### 4. *Modeling*

Tahapan yang keempat yaitu *modelling*. Modeling sendiri merupakan tahapan pemilihan dan pembentukan model yang sesuai dengan data yang dimiliki agar menghasilkan hasil model yang optimal[46]. Pada penelitian ini dilakukan pembentukan model *forecasting* dengan menggunakan data yang sudah melewati tahapan *data preprocessing* dengan memanfaatkan python sebagai bahasa pemrograman untuk melakukan hal tersebut. Model *forecasting* yang diambil dan digunakan untuk melakukan penelitian ini adalah *Long Short-Term Memory* (LSTM). Dalam *Long Short-Term Memory* ini sendiri menggunakan optimizer yang disediakan oleh Tensorflow, penelitian ini juga akan menggunakan berbagai macam optimizer yang akan di gunakan.

#### 5. *Evaluation*

Tahapan yang kelima yaitu tahap *Evaluation*. Tahapan ini adalah tahapan dimana akan dilakukannya pengukuran terhadap performa dari model yang sudah dibuat. Tahapan ini sendiri akan menentukan apakah model tersebut akan dilakukan *deployment* atau tidak. Apabila, dalam tahapan ini model memiliki hasil yang kurang baik maka akan dilakukan tahapan *modelling* kembali agar dapat mendapatkan hasil yang terbaik.

Dalam penelitian ini, pemilihan indikator yang digunakan dalam Evaluation ini adalah nilai RMSE. RMSE sendiri adalah akar kuadrat dari rata-rata kesalahan kuadrat, dan ini juga lebih sensitive terhadap outliers daripada metrik akurasi lainnya. Dalam RMSE, nilai yang lebih rendah maka menunjukkan model yang lebih akurat.

#### 6. *Deployment*

Tahapan ini adalah tahapan keenam dan yang terakhir dari metode CRISP-DM ini sendiri. Tahapan ini merupakan tahapan dimana kita melakukan penyebaran informasi yang didapatkan ke *end-user* mengenai analisa yang sudah dilakukan dalam penelitian. Penelitian ini akan

memanfaatkan visualisasi dari data yang sudah diolah dan diterapkan model yang sudah pilih.

### **3.3 Variabel Penelitian**

Variabel yang digunakan dalam penelitian ini adalah variable dependen dan variable independen.

#### **1. Variabel Dependen**

Variabel dependen adalah variabel hasil yang perubahannya itu sendiri dipengaruhi dan terikat oleh variable independen[47]. Variable dependen dalam penelitian ini adalah nilai Median dari setiap lokasi yang ditentukan. Penggunaan nilai median ini dikarenakan nilai-nilai yang ada dalam setiap kolom dalam *dataset* tidak menentu yang berarti terdapat nilai yang terlalu besar dan nilai yang terlalu kecil, dan ketika terdapat situasi tersebut maka penggunaan nilai tengah merupakan hal yang tepat sehingga tidak terdapat kesimpangan data akibat dari hal tersebut[12].

#### **2. Variabel Independen**

Variabel Independen adalah variabel penyebab yang menjadi penyebab terjadinya perubahan pada variabel dependen[47]. Variabel independen dari data penelitian ini adalah nilai setiap kolom dari setiap bulan yang digunakan. Kolom yang dimaksud yaitu kolom tanggal 01-01-2019, kolom tanggal 01-02-2019, kolom tanggal 01-03-2019, kolom tanggal 01-04-2019, kolom tanggal 01-05-2019, kolom tanggal 01-06-2019, kolom tanggal 01-07-2019, kolom tanggal 01-08-2019, kolom tanggal 01-09-2019, kolom tanggal 01-10-2019, kolom tanggal 01-11-2019, kolom tanggal 01-12-2019, kolom tanggal 01-01-2020, kolom tanggal 01-02-2020, kolom tanggal 01-03-2020, kolom tanggal 01-04-2020, kolom tanggal 01-05-2020, kolom tanggal 01-06-2020, kolom tanggal 01-07-2020, kolom tanggal 01-08-2020, kolom tanggal 01-09-2020, kolom tanggal 01-10-2020, kolom tanggal 01-11-2020, kolom tanggal 01-12-2020, kolom tanggal 01-01-2021, kolom tanggal 01-02-2021, kolom tanggal 01-03-2021, kolom tanggal 01-04-2021, kolom tanggal 01-05-2021, kolom tanggal 01-06-2021, kolom tanggal 01-07-2021, kolom tanggal 01-08-2021, kolom tanggal 01-09-2021, kolom



tanggal 01-10-2021, kolom tanggal 01-11-2021, kolom tanggal 01-12-2021, kolom tanggal 01-01-2022, kolom tanggal 01-02-2022, kolom tanggal 01-03-2022, kolom tanggal 01-04-2022, kolom tanggal 01-05-2022, dan kolom tanggal 01-06-2022

### **3.4 Teknik Pengumpulan Data**

Penelitian ini akan menggunakan metode penelitian kuantitatif, dimana pengumpulan data akan dilakukan dari Google Earth Engine dengan menggunakan dataset *VIIRS Stray Light Corrected Nighttime Day/Night Band Composites Version 1*. Metode kuantitatif sendiri akan diterapkan dengan cara mengambil data dari dataset tersebut untuk mendapatkan data gambar yang kemudian akan di ubah menjadi data numerik. Pengambilan bangunan yang dipilih pun secara random dan itu tersebar di seluruh daerah di Ibukota Indonesia yaitu DKI Jakarta.

### **3.5 Teknik Analisis Data**

Berdasarkan metodologi penelitian yang telah di tentukan dan akan digunakan yaitu dengan framework CRISP-DM, algoritma LSTM dan *forecasting*, maka akan di perlukan *tools* untuk melakukan *forecasting* dengan menggunakan metodologi tersebut. Tools yang akan digunakan yaitu python dengan menggunakan Jupyter Notebook, mulai dari mengkonversi data gambar menjadi numerik hingga ke menghasilkan plot visualisasi untuk prediksi 1 tahun ke depan. Penggunaan LSTM nya itu sendiri juga akan dibagi menjadi seluruh Jakarta kemudian akan dipecah menjadi perbagian Jakarta, lalu yang terakhir ada menjadi per tipe bangunan. Setelah mendapatkan itu semua maka akan dicoba 5 *optimizer* yang akan dipilih secara random dan hasilnya akan di bandingkan antar *optimizer* 1 dengan yang lainnya untuk melihat nilai RMSE mana pada *optimizer* apa yang akan memiliki nilai yang paling bagus.