

BAB III

METODOLOGI PENELITIAN

3.1 Objek Penelitian

Penelitian ini bertujuan untuk menganalisis pengaruh deforestasi terhadap transformasi hutan di Kalimantan, dengan fokus pada hubungan antara data *Global Forest Change* (GFC) pada perubahan lahan hutan dan *Night-Time Light* (NTL) pada intensitas cahaya malam hari. Data dari kedua model ini diperoleh melalui Google Earth Engine menggunakan dataset “*Hansen Global Forest Change v1.11*” untuk GFC dan “*VIIRS Nighttime Day/Night Band Composites Version 1*” untuk NTL, dengan periode studi dari tahun 2001 hingga 2023 yang mencakup data *time-series* secara tahunan untuk GFC dan bulanan dari 2013 hingga 2023 untuk NTL. Periode studi dari tahun 2001 dan 2013 hingga 2023 dipilih karena ketersediaan data GFC dan NTL yang mempertemukan kedua ketersediaan data pada periode tersebut.

Wilayah administratif yang diteliti meliputi provinsi Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, dan Kalimantan Utara. Pemilihan wilayah administratif ini dilakukan agar skala relevansi data bisa dicapai seakurat mungkin, mengingat luasnya wilayah dan variasi kondisi hutan di Kalimantan. Proses pengambilan *dataset* dilakukan secara 2 macam, antara lain secara per kota dan kabupaten, serta secara provinsi. Data dengan metode pengambilan secara per kota dan kabupaten mencatat gabungan kota dan kabupaten di Pulau Kalimantan bertotal 56 dari 5 provinsi berbeda. Data dengan metode pengambilan secara per provinsi menghasilkan data dengan nilai untuk 5 provinsi berbeda di Kalimantan. Metode ini dipilih karena memudahkan pengambilan data yang lebih terperinci dan mudah untuk diidentifikasi berdasarkan wilayah perbatasannya.

Masing-masing metode pengambilan tersebut dilakukan dengan *dataset* GFC dan NTL sehingga menghasilkan total 4 *dataset* inti, terdiri dari 2 *dataset* (GFC dan

NTL) per kota dan kabupaten, dan 2 *dataset* (GFC dan NTL) per provinsi. Berikut *detail* dari masing-masing *dataset* yang dihasilkan:

- 1) *Dataset* GFC per kota dan kabupaten (1,288 baris, 7 kolom)
Kolom: nama, wilayah, provinsi, *longitude*, *latitude*, *year*, *area*
- 2) *Dataset* NTL per kota dan kabupaten (7,336 baris, 7 kolom)
Kolom: nama, wilayah, provinsi, *longitude*, *latitude*, *date*, *mean*
- 3) *Dataset* GFC per provinsi (115 baris, 5 kolom)
Kolom: provinsi, *longitude*, *latitude*, *year*, *loss_year*
- 4) *Dataset* NTL per provinsi (660 baris, 5 kolom)
Kolom: provinsi, *longitude*, *latitude*, *date*, *mean_radiance*

Hasil analisis ditampilkan melalui data *visualization dashboard* pada Tableau yang menampilkan persebaran data dari GFC dan NTL per tahun, beserta prediksinya untuk tahun 2024. Tujuan visualisasi tersebut membantu dalam memahami dinamika transformasi di Pulau Kalimantan dan efeknya terhadap intensitas cahaya malam hari.

3.2 Metode Penelitian

Metode yang digunakan pada penelitian ini berdasarkan kesimpulan hasil dari penelitian terdahulu, yakni *Time-Series Analysis* dengan memanfaatkan teknik *data mining* sebagai kerangka kerjanya. Analisis berbasis *time-series* ANN dipilih karena kemampuannya yang kuat dalam pemetaan nonlinear, toleransi terhadap kesalahan, dan kemampuan belajar mandiri [50]. Hal ini juga didukung karena ANN adalah pendekatan terbaik untuk melakukan prediksi dengan dasar nilai *time-series* yang telah dikonversikan ke dalam bentuk numerik secara efektif [51]. Metode model yang digunakan pada penelitian ini adalah LSTM. LSTM ini dipakai untuk memprediksi nilai numerikal dari *dataset* yang sudah dilakukan tahap *preprocessing*. Metode ini masih satu turunan dari algoritma metode ANN dan LSTM merupakan salah satu penyempurnaan metode dari RNN yang digunakan untuk analisis numerikal secara prediktif.

Sebelum menentukan teknik *data mining* yang paling sesuai, dilakukan perbandingan antar beberapa kerangka kerja *data mining*, yakni CRISP-DM (*Cross-Industry Standard Process for Data Mining*), KDD (*Knowledge Discovery in Databases*), dan SEMMA (*Sample, Explore, Modify, Model, Assess*).

3.2.1 Perbandingan Metode

Metode penelitian yang umum digunakan untuk melakukan pengolahan data adalah CRISP-DM, KDD, dan SEMMA.

Tabel 3.1 Tabel Perbandingan Metode Penelitian *Data Mining* [23]

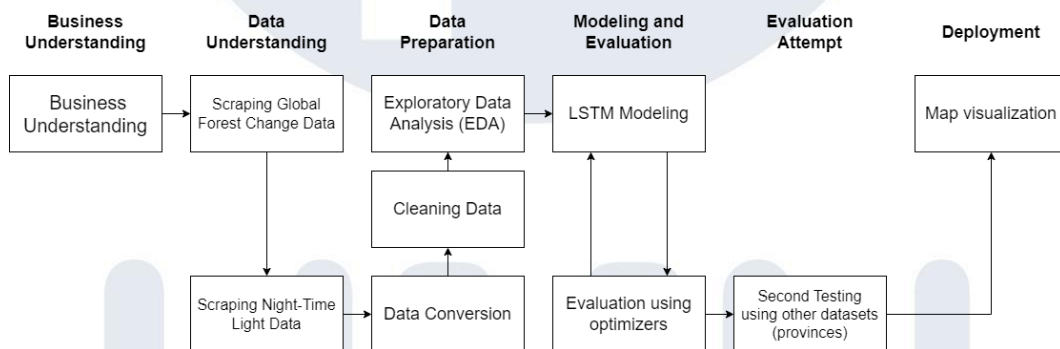
Metode	CRISP-DM	KDD	SEMMA
Tahap	<ol style="list-style-type: none"> 1. <i>Business Understanding</i> 2. <i>Data Understanding</i> 3. <i>Data Preparation</i> 4. <i>Modeling</i> 5. <i>Evaluation</i> 6. <i>Deployment</i> 	<ol style="list-style-type: none"> 1. <i>Data Selection</i> 2. <i>Pre-processing</i> 3. <i>Transformation</i> 4. <i>Data mining</i> 5. <i>Interpretation / Evaluation</i> 	<ol style="list-style-type: none"> 1. <i>Sample</i> 2. <i>Explore</i> 3. <i>Modify</i> 4. <i>Model</i> 5. <i>Assessment</i>

Metodologi yang digunakan untuk memahami data penelitian ini, sebagaimana tercantum dalam tabel 3.1, adalah CRISP-DM (*Cross-Industry Standard Process for Data Mining*). CRISP-DM dipilih karena dianggap sebagai standar yang dapat menghasilkan keputusan berbasis data yang berkualitas tinggi dan lebih komprehensif dibandingkan SEMMA [23]. CRISP-DM dipilih karena popularitasnya yang tinggi dan sebagai model proses standar terbuka yang menggambarkan pendekatan umum dalam *data mining*, serta menyediakan proses yang andal dan berulang [37].

3.2.2 Alur Penelitian

Penelitian ini mengikuti metode CRISP-DM, dimulai dengan pemilihan topik dan pemahaman bisnis, diikuti oleh pengumpulan dan *understanding dataset* GFC dan NTL. Data dikumpulkan empat kali dengan dua metode: per kota dan kabupaten, serta per provinsi. Data awal berupa gambar diolah menjadi data numerik untuk mendapatkan nilai rata-rata tahunan intensitas cahaya dan transformasi hutan di Kalimantan. Data GFC diperoleh bulanan (2001-2023) dan data NTL tahunan (2013-2023).

Tahap berikutnya adalah *data preparation* dengan *pre-processing* untuk mendeteksi *missing values* dan pemilihan atribut untuk *Exploratory Data Analysis* (EDA). EDA bertujuan memahami pola dan mencari anomali dalam data *time-series*. Data dibagi menjadi data *training* dan testing untuk pemodelan *Long Short-Term Memory* (LSTM). Proses pemodelan menghasilkan prediksi yang dievaluasi dengan *Root Mean Square Error* (RMSE). RMSE kecil menunjukkan akurasi yang baik, sedangkan RMSE besar memerlukan optimasi *hyperparameter*, pengecekan kualitas data, dan perbandingan dengan model *time-series* lain untuk hasil optimal. Pengujian kedua dilakukan dengan *dataset* berbeda (data per kota dan kabupaten atau per provinsi) untuk mencocokkan korelasi dan kevalidan data. Hasil akhir divisualisasikan menggunakan Tableau untuk memudahkan interpretasi dalam bentuk peta visual. Seluruh alur penelitian dijelaskan dalam gambar 3.1.



Gambar 3.1 Alur Penelitian

3.2.2.1 Business Understanding

Business Understanding adalah tahapan awal dalam metodologi analisis menggunakan CRISP-DM. *Business Understanding* membantu dalam memahami tujuan penelitian berdasarkan kebutuhan yang ada. Tujuan dari penelitian ini adalah melakukan komparasi model dan mencari korelasi terkait model prediktif dari data kehutanan *Global Forest Change* (GFC) dan data aktivitas *Night-Time Light* (NTL) yang memiliki pengaruh dari perubahan tutupan hutan di Kalimantan akibat deforestasi.

3.2.2.2 Data Understanding: Pengumpulan Data Global Forest Change

Data *Global Forest Change* merupakan informasi yang didapatkan melalui satelit Google Earth, tetapi proses pengambilan gambarnya dilakukan pada siang hari, berbeda dengan *Night-Time Light* yang fokus pada gambar malam. *Dataset* ini mencakup beragam informasi terkait luas hutan dan perubahan yang terjadi, baik pengurangan maupun penambahan luas hutan. Untuk penelitian ini, *dataset* yang dipilih adalah *Hansen Global Forest Change v1.11* dengan relevansi data mencapai tahun 2023 sehingga penelitian ini menggunakan rentang penelitian dari tahun 2001 hingga 2023 yang memerlukan observasi lebih banyak melalui pengolahan data tahunan. Keunggulan *dataset* ini adalah mencakup wilayah Asia, termasuk Pulau Kalimantan yang menjadi fokus penelitian. Pemilihan *dataset* ini strategis karena memberikan gambaran komprehensif terkait perubahan hutan di wilayah yang diteliti. *Dataset* ini mempergunakan variabel metrik “*loss_year*”, yaitu metrik yang menghitung kehilangan area hutan secara luas hektar per tahunan.

3.2.2.3 Data Understanding: Pengumpulan Data Night-Time Light

Data *Night-Time Light* adalah hasil tangkapan satelit pada Google Earth pada waktu malam yang mencerminkan intensitas cahaya saat itu. Intensitas cahaya ini memberikan gambaran visual mengenai kegiatan manusia dan perkembangan infrastruktur, khususnya di wilayah perkotaan. Kota-kota besar akan tampak lebih terang karena banyaknya infrastruktur yang menggunakan lampu terang, sekaligus mencerminkan kepadatan populasi di wilayah tersebut. Dalam penelitian ini, digunakan *dataset VIIRS Nighttime Day/Night Band Composites Version 1*, dengan rentang waktu dari tahun 2013 bulan Januari hingga tahun 2023 bulan Desember dengan fokus di wilayah administratif Pulau Kalimantan, Indonesia.

3.2.2.4 Data Preparation: Konversi Gambar Satelit

Data yang terhimpun, mencakup *Global Forest Change* secara tahunan dan *Night-Time Light* secara bulanan yang ditransformasi dari format gambar sebelum akhirnya dikonversi ke bentuk numerik. Proses konversi menjadi bentuk numerik ini bersifat strategi, mengingat bahwa proses pemodelan *Long Short-Term Memory* (LSTM) membutuhkan data yang bersifat numerik. Langkah ini memastikan bahwa data yang akan digunakan dalam proses analisis memiliki representasi numerik yang sesuai, mempermudah interpretasi, dan optimasi model. Pendekatan ini meliputi *formatting data*, pengaturan *data frame*, dan *feature selection*. Data diformat untuk diperiksa setiap baris yang akan digunakan, menambah kolom untuk memberikan informasi tambahan, dan memilih variabel fitur apa saja dalam *data frame* yang akan digunakan.

3.2.2.5 Data Preparation: Splitting Data

Data split berperan sebagai acuan untuk bahan evaluasi dari ketepatan dan keberhasilan setelah penerapan suatu model *machine learning*. *Data split* mempergunakan *sampling*, yakni teknik dalam mengambil jumlah objek dari suatu populasi untuk kebutuhan pengukuran. Pengambilan sampel dalam teknik *sampling* dapat menggunakan perhitungan probabilitas sederhana atau *simple random sampling*, yaitu pengambilan jumlah beberapa sampel secara acak yang ditentukan berdasarkan peluang sampel. Metode *sampling* dalam penerapan model membagi *dataset* ke dalam dua kategori menjadi *data training* dan *data testing* yang berisi bobot terbagi ke dalam rasio yang sesuai kebutuhan, yakni 75% pada *data training* dan 25% pada *data testing* untuk data GFC dan NTL.

3.2.2.6 Modeling

Proses pemodelan ini melibatkan penggunaan *Long Short-Term Memory* (LSTM) berbasis *time-series*, yang didasarkan pada data *Global Forest Change* dan *Night-Time Light*. Fokus utama dari pemodelan ini adalah tidak hanya memahami perkembangan transformasi tutupan hutan, termasuk lokasi tempat tinggal penduduk pada periode sebelumnya di Pulau Kalimantan, tetapi juga memiliki kemampuan untuk melakukan analisis prediktif terhadap perkembangan tersebut di masa depan. Model NTL mempergunakan parameternya yang telah disesuaikan dan model GFC yang sudah disesuaikan juga sehingga tidak memiliki perbandingan terhadap acuan parameter model LSTM. Dengan menerapkan teknik *predictive analytics*, pemodelan ini bertujuan untuk memberikan wawasan yang lebih mendalam tentang pola-pola yang mungkin terjadi dalam transformasi pengalihan lahan hutan di Pulau Kalimantan.

3.2.2.7 Evaluasi Model

Setelah proses *training* model selesai, dilakukan evaluasi untuk menilai sejauh mana kinerja model tersebut. Evaluasi dilakukan menggunakan berbagai metrik utama RMSE. Model membandingkan RMSE dari masing-masing *dataset* yang menggunakan *optimizer* tertentu. Terdapat 5 *optimizer* yang dibandingkan, terdiri dari Adam (sebagai acuan utama/*default*), RMSprop, SGD, Adamax, dan Adadelta. Dari perbandingan RMSE dengan *optimizer* yang berbeda tersebut, akan ditentukan *optimizer* apa dengan hasil optimasi RMSE terbaik dari total model yang menggunakan *optimizer* tersebut. Jika hasil evaluasi menunjukkan performa model yang masih belum optimal, langkah selanjutnya melibatkan kembali tahapan pemodelan dengan melakukan optimalisasi. Proses ini melibatkan perubahan *hyperparameter* atau *optimizer*, pengecekan kualitas data ulang, dan perbandingan dengan model *time-series* lainnya. Tujuannya adalah untuk meningkatkan kualitas dan ketepatan model prediksi agar dapat

memberikan hasil yang lebih baik dalam memprediksi pola dan persebaran intensitas cahaya yang akurat.

3.2.2.8 Pembuatan *Dashboard* dan *Deployment*

Langkah selanjutnya setelah mendapatkan hasil analisis adalah merancang dan membuat *dashboard* visualisasi peta yang dapat mempresentasikan temuan dengan jelas. Proses ini melibatkan pemilihan alat visualisasi data, diikuti oleh ekstraksi dan *preprocessing* data agar sesuai dengan kebutuhan visualisasi. Setelah *dashboard* dibuat, langkah berikutnya adalah melakukan *deploy* kepada khalayak luas sebagai pemaparan informasi baru. Pemeliharaan dan pembaruan secara berkala diperlukan untuk memastikan bahwa *dashboard* tetap responsif dan dapat memberikan informasi terkini kepada pengguna.

3.3 Variabel Penelitian

Variabel dalam penelitian terbagi menjadi dua macam, yaitu variabel dependen dan variabel independen. Variabel dependen sebagai variabel yang dipengaruhi oleh variabel lainnya. Sebaliknya, variabel independen tidak perlu pengaruh dari variabel lain, melainkan mandiri.

Pada penelitian ini, variabel dependen (terikat) adalah variabel *measure* berupa nilai numerik dari area kehilangan hutan dan persebaran intensitas cahaya yang diambil secara rata-rata (numerik), seperti “mean”, “area”, “*mean_radiance*”, dan “*loss_year*”. Dari nilai rata-rata tersebut, akan dikelompokkan secara provinsi dengan menggunakan nilai median karena *dataset* memiliki nilai yang tidak konsisten, termasuk nilai dengan jarak yang signifikan dengan nilai lainnya. Penggunaan median dapat mengatasi ketidakseimbangan data yang disebabkan oleh kesenjangan nilai tersebut [14]. Variabel independen (tidak terikat) penelitian ini adalah variabel *dimension* yang mendukung angka numerik variabel *measure*. Pada penjelasan penggunaan *dataset* pada objek penelitian, contoh variabel

independen antara lain “nama”, “wilayah”, “provinsi”, “*longitude*”, “*latitude*”, “*date*”, dan “*year*”.

3.4 Teknik Pengumpulan Data

Penelitian ini bermetode penelitian kuantitatif dan pengumpulan data menggunakan *dataset Hansen Global Forest Change v1.11* dari tahun 2001 hingga 2023 untuk GFC dan *VIIRS Nighttime Day/Night Band Composites Version 1* untuk NTL dari tahun 2013 hingga tahun 2023. Pengumpulan data mempergunakan fungsi dari Python untuk mengolah data di Google Colab dan Google Earth Engine dengan mengekstrak data secara mandiri yang terdiri dari 4 macam hasil pengumpulan data:

- 1) *Dataset* GFC per kota dan kabupaten, 56 titik (data utama)
- 2) *Dataset* NTL per kota dan kabupaten, 56 titik (data utama)
- 3) *Dataset* GFC per provinsi, 5 provinsi (data pengujian kedua)
- 4) *Dataset* NTL per provinsi, 5 provinsi (data pengujian kedua)

Metode kuantitatif akan diterapkan dengan mengambil data gambar dari *dataset* tersebut dan mengonversinya menjadi data numerik. Objek tempat dan lokasi adalah keseluruhan wilayah administratif Indonesia di Pulau Kalimantan yang terbagi per provinsi, yakni Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Kalimantan Selatan, dan Kalimantan Barat.

3.4.1 Periode Pengambilan Data

Periode pengambilan data dilakukan pada rentang tahun 2001 hingga 2023, dengan deskripsi GFC 2001 hingga 2023 dan NTL 2013 hingga 2023. Hal ini merujuk pada titik tengah dari ketersediaan data *Global Forest Change* (tersedia dari 2000-2023) dan *Night-Time Light* (tersedia dari 2012-2023) yang digunakan agar bisa diintegrasikan. Pada implementasi pengambilan datanya, data *time-series Global Forest Change* dari Hansen hanya bisa dijabarkan secara tahunan (2001, 2002, 2003, ..., 2023), sedangkan data *time-series* pada *Night-Time Light* memiliki penjabaran secara per bulan dari tahun 2013 hingga 2023 (2013 bulan 1, 2013 bulan 2,

2013 bulan 3, ..., 2023 bulan 12). Batasan masalah yang dirumuskan dari kedua kondisi yang berbeda tersebut menyebabkan data *Night-Time Light* yang harus disesuaikan untuk hasil *output* prediksi yang dijabarkan secara per tahun mengikuti data *Global Forest Change*.

3.5 Teknik Analisis Data

Teknik analisis geospasial data *time-series* dalam penelitian ini menggunakan metode CRISP-DM, tepatnya dengan algoritma *Long Short-Term Memory* (LSTM), yakni berkombinasi penggunaan data berbasis *time-series* yang dimodelkan dengan teknik penjarangan dengan kebutuhan prediksi transformasi di masa yang akan datang. Analisis data dilakukan dengan menggunakan bahasa pemrograman Python dengan Jupyter Notebook. Diawali dari proses mengonversi data gambar menjadi data numerik hingga pembuatan *plot* visualisasi untuk prediksi, dan pemodelan menggunakan *Long Short-Term Memory* (LSTM). Setelah mendapatkan hasil dari seluruh tahapan tersebut, dilakukan pengujian. Hasilnya kemudian dibandingkan antara satu *optimizer* dengan *optimizer* lainnya untuk menilai nilai evaluasi mana yang paling optimal. Setelah pemodelan, diperlukan *deployment* ke visualisasi peta untuk publikasi hasil visualisasi data dengan *dashboard*.

