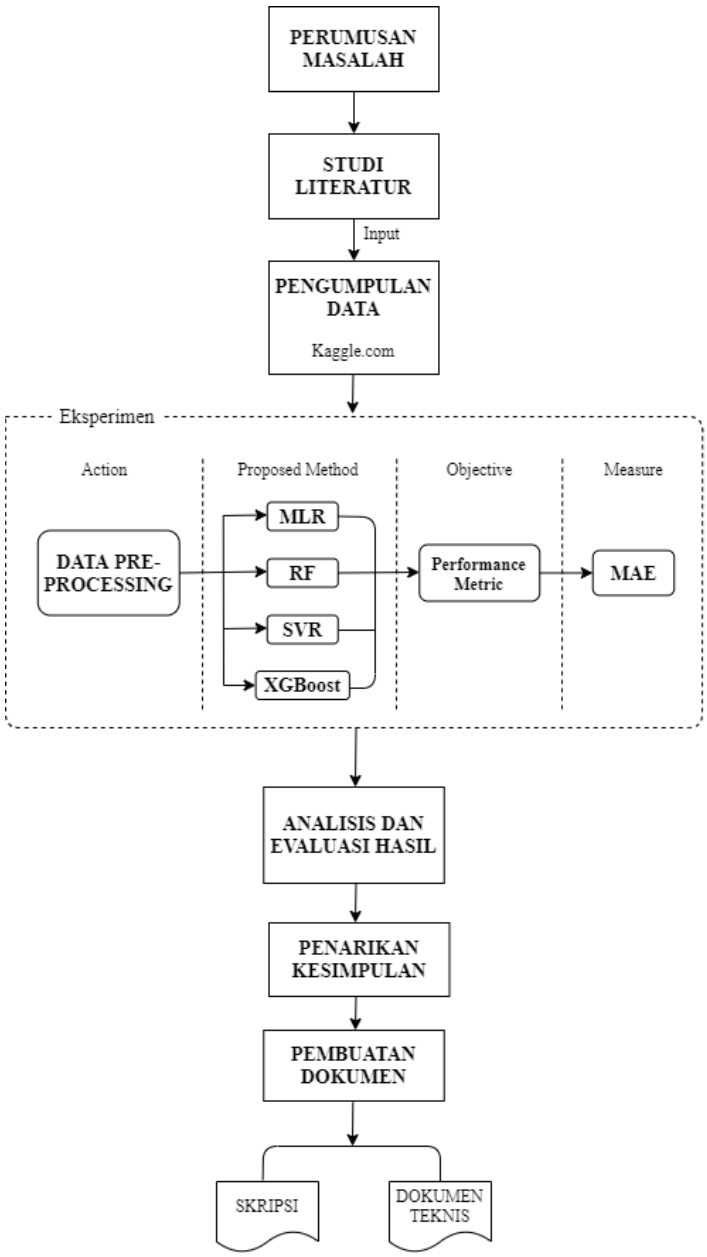
BAB III
METODOLOGI PENELITIAN

Bab ini merupakan penjabaran mengenai teori-teori dasar yang mendukung serta berhubungan dengan penelitian ini.

3.1 Desain Penelitian

Pada penelitian ini dilakukan beberapa tahapan penelitian sebagaimana ditunjukkan pada Gambar 3.1. berikut ini:



Gambar 3.1 Desain Penelitian

Gina Fauziah, 2021
ANALISIS IMPLEMENTASI MACHINE LEARNING UNTUK PREDIKSI JUMLAH KASUS DEMAM BERDARAH DENGUE (DBD)
Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

3.1.1 Perumusan Masalah

Langkah pertama dalam alur penelitian ini adalah perumusan masalah penelitian tentang Analisis Implementasi *Machine Learning* untuk Prediksi Jumlah Kasus Demam Berdarah Dengue (DBD) yaitu menentukan latar belakang yang merupakan dasar pemikiran penelitian, rumusan masalah dan batasan masalah agar memperjelas penelitian, serta tujuan penelitian untuk memfokuskan penelitian agar lebih terarah.

3.1.2 Studi Literatur

Studi literatur dilakukan dengan cara mencari dan mempelajari referensi yang berkaitan dengan penelitian ini meliputi DBD, *forecasting*, *multiple linear regression*, *random forest regressor*, *support vector regression*, *xgboost*, *gridsearch-cv*, metriks evaluasi *Mean Absolute Error*, *Mean Squared Error dan Root Mean Squared Error* serta jurnal-jurnal atau artikel ilmiah tentang penelitian sebelumnya yang berkaitan dengan penelitian ini.

3.1.3 Pengumpulan Data

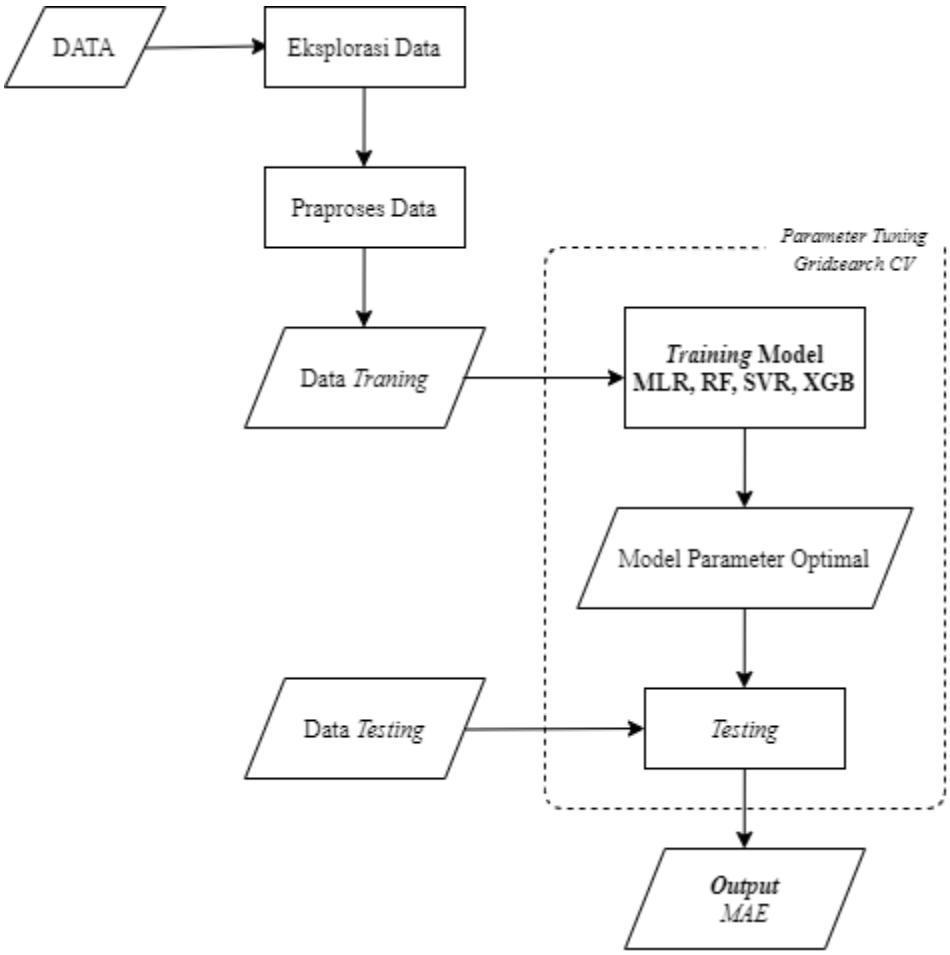
Penulis menggunakan set data publik jumlah kasus DBD dan data perubahan iklim di Kota San Juan (Puerto Rico) dan Kota Iquitos (Peru) yang merupakan set data yang digunakan untuk kompetisi pada situs DRIVENDATA yang bernama "*DengAI: Predicting Disease Spread*". Dapat diunduh pada tautan https://dengueforecasting.noaa.gov/.

Pada situs DRIVENDATA dijelaskan bahwa Data-data tersebut berasal dari berbagai sumber yang bertujuan untuk mendukung "*Predict the Next Pandemic Initiative*". Data pengawasan DBD disediakan oleh *Centers for Disease Control and prevention* Amerika Serikat, serta *Department of Defense's Naval Medical Research Unit 6* dan *The Armed Forces Health Surveillance Center*, bekerja sama dengan Pemerintah Peru dan Universitas Amerika Serikat. Data lingkungan dan iklim disediakan oleh *National Oceanic and Atmospheric Administration* (NOAA), sebuah agensi dari *Department of Commerce* Amerika Serikat. Variabel iklim yang digunakan diantaranya variabel pengukuran suhu (suhu maksimum, suhu minimum dan suhu rata-rata), pengukuran kelembaban, pengukuran curah hujan, dan

pengukuran indeks vegetasi (ndvi) dari citra satelit yang menunjukan tingkat kehijauan vegetasi atau konsentrasi klorofil daun.

3.1.4 Eksperimen

Pada tahapan ini akan dilakukan eksperimen arsitektur sesuai dengan desain eksperimen pada Gambar 3.2.



Gambar 3.2 Desain Eksperimen

1) Eksplorasi Data

Dimulai dengan data yang ada akan dieksplorasi dan dianalisis yaitu mengecek jumlah null setiap kolom, menganalisis bagaimana dengan korelasi datanya terhadap jumlah kasus, dan mengecek *trend* jumlah kasus setiap minggu untuk memudahkan membuat grafik *line* yang berdasarkan waktu karena nilainya kontinyu.

2) Praproses Data

Selanjutnya pada praproses data diawali dengan pengisian null value pada kolom yang null dengan nilai setelahnya (*backwardfill*), lalu melakukan split data yaitu membagi data menjadi dua bagian berdasakan nama kota karena akan dibuat dua model, yaitu San Juan dan Iquitos. Selanjutnya melakukan *training* dan *testing* dengan *spliting* data sebanyak 80% data untuk *training model* dan 20% untuk *testing model*. Hal ini bertujuan agar data tidak akan terlalu *overfitting* terhadap data *testing*.

3) *Parameter Tuning*

Kemudian data *training* akan dilatih pada *training model* dengan metode *Multiple Linear Regression, Random Forest, Support Vector Regression, dan XGBoost*. Lalu menggunakan *GridSearch-CrossValidation* untuk mencari *hyperparameter tuning* atau mencari kombinasi terbaik dari parameter sehingga menghasilkan parameter optimal untuk setiap modelnya.

4) Testing Model

Setelah melakukan *training* 80% data dan mendapatkan model dengan parameter optimal, langkah selanjutnya yaitu melakukan *testing model* dengan 20% data latih pada model yang telah ada. Output dari testing berupa evaluasi metrik MAE (*Mean Absolute Error*).

3.1.5 Analisis dan Evaluasi Hasil Eksperimen

Tahap proses berikutnya akan dilakukan analisis dan evaluasi terhadap hasil eksperimen yang telah didapatkan. Aspek yang dievaluasi adalah *MAE* (*Mean Absolute Error*). Kemudian hasil evaluasi akan disajikan dalam bentuk tabel dan grafik untuk dibandingkan unjuk kerja terbaik dari setiap metode.

3.1.6 Penarikan Kesimpulan

Pada tahap ini hasil analisis pada tahapan sebelumnya akan menjadi dasar pada penarikan kesimpulan. Tahap ini akan menentukan kesimpulan akhir dari penelitian yang telah dilakukan. Selanjutnya dapat melakukan dokumentasi skripsi dan dokumen teknis.

3.2 Lingkungan Komputasi

Lingkungan komputasi pada penelitian ini adalah seperangkat komputer dan beberapa perangkat lunak pendukung. Pada penelitian ini, spesifikasi perangkat keras komputer sebagai berikut:

1) RAM 8 GB
2) HDD 1 TB
3) *Processor Intel Core i5-8250U*
4) *Mouse* dan *Keyboard*
5) VGA Nvidia GeForce MX130

Adapun perangkat lunak pada komputer tersebut sebagai berikut:

1) *Microsoft Windows 10*
2) *Anaconda Prompt*
3) *Text Editor*: *Jupyter Notebook for Python 3*
4) *Google Colab*