

BAB 3

METODE PENELITIAN

Penelitian ini adalah analisis prediksi kasus COVID-19 di Indonesia dengan algoritma Linear Regression. Penelitian ini membutuhkan dataset yang berisi kasus terkonfirmasi positif, kasus kematian, dan kesembuhan penderita di Indonesia, selanjutnya dilakukan pengolahan data berupa *preprocessing* untuk melakukan pembersihan dan transformasi data. Data tersebut nantinya digunakan untuk melatih model yang akan digunakan untuk memprediksi kasus COVID-19.

3.1 BAHAN DAN ALAT PENELITIAN

Dalam penelitian ini bahan yang dibutuhkan antara lain dataset. Dataset ini digunakan sebagai sumber data yang dipakai untuk melakukan analisis prediksi kasus COVID-19 di Indonesia. Dataset yang digunakan dalam penelitian ini berasal dari website <https://covid19.go.id/> yang merupakan media edukasi dan informasi serta situasi perkembangan COVID-19 di Indonesia berbentuk *Application Programming Interface* (API).

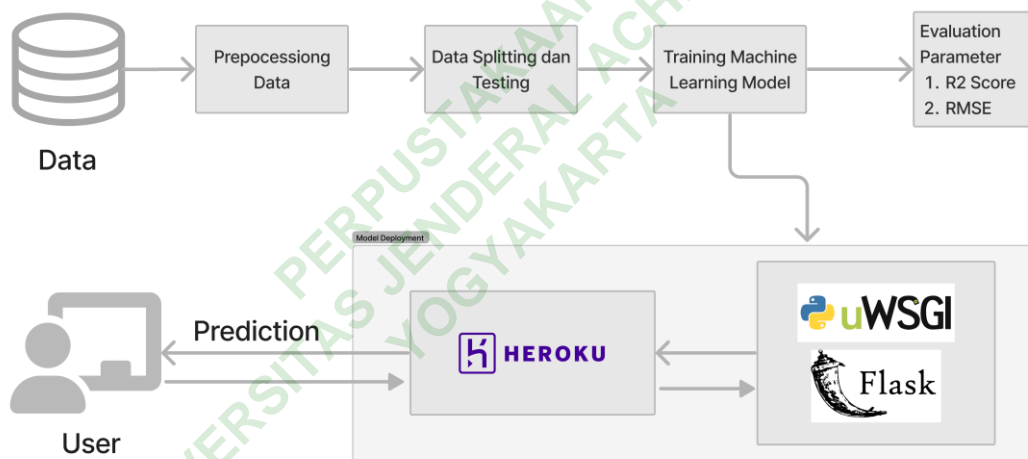
Alat yang digunakan dalam penelitian ini adalah komputer dengan spesifikasi dan konektivitas internet yang memadai untuk menjalankan sistem operasi dan pengembangan perangkat lunak. Sistem operasi dan program aplikasi yang digunakan untuk mengembangkan aplikasi ini adalah:

1. Sistem Operasi: Windows 11.
2. Aplikasi Jupyter Notebook: versi 6.3.0 atau versi terbaru
3. Aplikasi Visual Studio Code: versi 1.70.1 atau versi terbaru
4. Aplikasi Microsoft Excel: versi 2019 atau versi terbaru
5. *Library* Python, *library* python yang akan digunakan antara lain:
 - a. NumPy versi 1.21.4 atau versi terbaru
 - b. Pandas versi 1.3.4 atau versi terbaru
 - c. Matplotlib versi 3.3.2 atau versi terbaru
 - d. Sklearn versi 0.22.1 atau versi terbaru
 - e. Datetime versi 4.4 atau versi terbaru

- f. Flask versi 2.0.2 atau versi terbaru
- g. Joblib versi 1.1.0 atau versi terbaru
- h. Heroku versi 7.59.2 atau versi terbaru

3.2 JALAN PENELITIAN

Selanjutnya pada Bab ini perlu dituliskan metode-metode yang akan digunakan dalam proses penelitian, sesuai dengan topik / bidang penelitian. Pada penelitian ini, metode yang digunakan dibagi menjadi 6 tahap, yaitu tahap *Data Preprocessing* (pembersihan data), *Data Splitting* (pemisahan data), *Testing Set*, *Trained Model* (melatih model), *Evaluation Parameter* (evaluasi model), dan *Visualization* (visualisasi). Adapun alur tahap tersebut ditunjukkan pada gambar 3.3-1.



Gambar 3.2-1 Alur Proses Penelitian

3.2.1 Pengambilan Data

Tahap pertama dimulai dengan melakukan pengambilan dataset COVID-19. Tujuan dari penelitian ini adalah prediksi kasus COVID-19 dengan fokus pada jumlah kasus positif, jumlah kematian, dan jumlah kesembuhan. Dataset yang digunakan dalam penelitian ini diperoleh dari website <https://covid19.go.id/> yang merupakan media edukasi dan informasi serta situasi perkembangan COVID-19 di Indonesia berbentuk *Application Programming Interface* (API). Dalam API tersebut berisi data *JSON time series* harian, termasuk jumlah kasus yang

terkonfirmasi, kematian, dan kesembuhan secara *real-time* yang di update setiap hari. Semua data berasal dari laporan kasus harian dan frekuensi pembaruan data adalah satu hari.

3.2.2 Exploratory Data Analysis (EDA)

Dalam tahap *Exploratory Data Analysis* (EDA, Indonesia : Analisis Data Eksplorasi) melakukan pengamatan data sebelum membuat asumsi. Ini dapat membantu mengidentifikasi kesalahan yang jelas, serta lebih memahami pola dalam data, mendeteksi *outlier* atau anomali, menemukan hubungan yang menarik di antara variabel.

3.2.3 Preprocessing Data

Kemudian dilanjutkan dengan melakukan *preprocessing* data untuk menemukan statistik dari jumlah kasus terinfeksi yang dikonfirmasi setiap hari, jumlah kasus sembuh, dan jumlah kasus kematian. Hasil *time series* yang dihasilkan kemudian diekstraksi untuk dilakukan ke tahap berikutnya.

3.2.4 Data Splitting dan Testing Set

Setelah langkah *preprocessing* data awal, dataset dibagi menjadi dua subset, yaitu: *training set* (set pelatihan) untuk melatih model dan *testing set* (set pengujian) untuk menguji model sesuai data latih.

3.2.5 Training Model

Untuk pelatihan model yang digunakan dalam penelitian ini adalah menggunakan algoritma Linear regression. Model ini kemudian dilatih pada hari dan kasus yang baru dikonfirmasi, kasus sembuh, dan kasus kematian. Dalam teknik pemodelan ini, algoritma mengambil kumpulan data dengan *instance input* yang sesuai untuk melatih model. Kemudian model yang sudah dilatih dilakukan pengujian dengan membandingkan hasil aktual kasus COVID-19 dengan hasil prediksi model.

3.2.6 Evaluation Model

Model pembelajaran kemudian dievaluasi berdasarkan metrik penting seperti R^2 -score, dan RMSE kemudian dilaporkan hasilnya performansi model untuk memprediksi kasus.

3.2.7 Visualisasi

Dalam tahap ini, hasil prediksi yang telah diperoleh kemudian divisualisasikan dalam bentuk grafik dan dianalisis agar bisa menemukan *insights* bagaimana performa prediksi kasus COVID-19 di Indonesia dengan menggunakan algoritma *Linear Regression*.

3.2.8 Deployment Model

Pada tahap ini model yang telah dilatih kemudian disimpan dalam format *.pkl* menggunakan library *joblib*.

Setelah model telah selesai, maka proses selanjutnya adalah menyimpan model. Pada tahap ini membuat serial objek sehingga dapat disimpan ke file, dan dimuat dalam program dengan menggunakan *module* *Pickle*. *Pickle* digunakan untuk membuat *serializing* dan *de-serialized* struktur objek Python, proses ini juga disebut *marshaling* atau *flattening*. *Serializing* mengacu pada konversi objek dalam memori *byte stream* yang dapat disimpan pada *disk*. Kemudian, karakter *stream* ini dapat diambil dan *de-serialized* kembali ke objek Python (*Pickle — Python Object Serialization — Python 3.10.6 Documentation, n.d.*).

Setelah model tersimpan dalam *file* format *.pkl* yang telah disimpan dan menerapkan kepada user sebagai sistem *web service* layanan REST API yang ditulis menggunakan *framework* *Flask*. *Flask* adalah kerangka kerja web Python yang kecil dan ringan. *Flask* menyediakan alat dan fitur berguna yang membuat pembuatan aplikasi web dengan Python lebih mudah. *Flask* juga dapat diperluas dan tidak memaksa struktur direktori tertentu atau memerlukan kode yang rumit. Sistem tersebut akan di *deploy* dalam bentuk pada server aplikasi yang sesuai dengan protokol WSGI dengan dengan *cloud Platform as a Service* (PaaS) Heroku sebagai *web server*nya.

Heroku adalah salah satu layanan *cloud* berbasis *container-based* sebagai *Platform as a Service* (PaaS). Banyak Developer yang menggunakan Heroku untuk *deploy*, *manage*, dan menskalakan aplikasi modern. Heroku pertama kali digunakan pada tahun 2007, dan pada tahun 2008 dirilis sebagai *open-source*. Dalam Tugas Akhir ini, men-deploy Flask API menggunakan Heroku Command Line Interface (Heroku CLI).

PERPUSTAKAAN
UNIVERSITAS JENDERAL ACHMAD YANI
YOGYAKARTA