

BAB 3 METODOLOGI PENELITIAN

Penelitian ini menggunakan algoritma Linear Regression. Metode Linear Regression adalah metode statistika yang menghubungkan satu atau lebih variabel bebas X dengan sebuah variabel respon Y metode ini digunakan untuk menghasilkan hasil prediksi tingkat kematian COVID-19 di Indonesia [18]. Untuk evaluasi dari prediksi yang telah dibuat dalam penelitian ini, akan menggunakan perhitungan *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE) dan *R Squared* (R^2). Berikut merupakan tahap-tahapan yang akan dilakukan dalam penelitian.

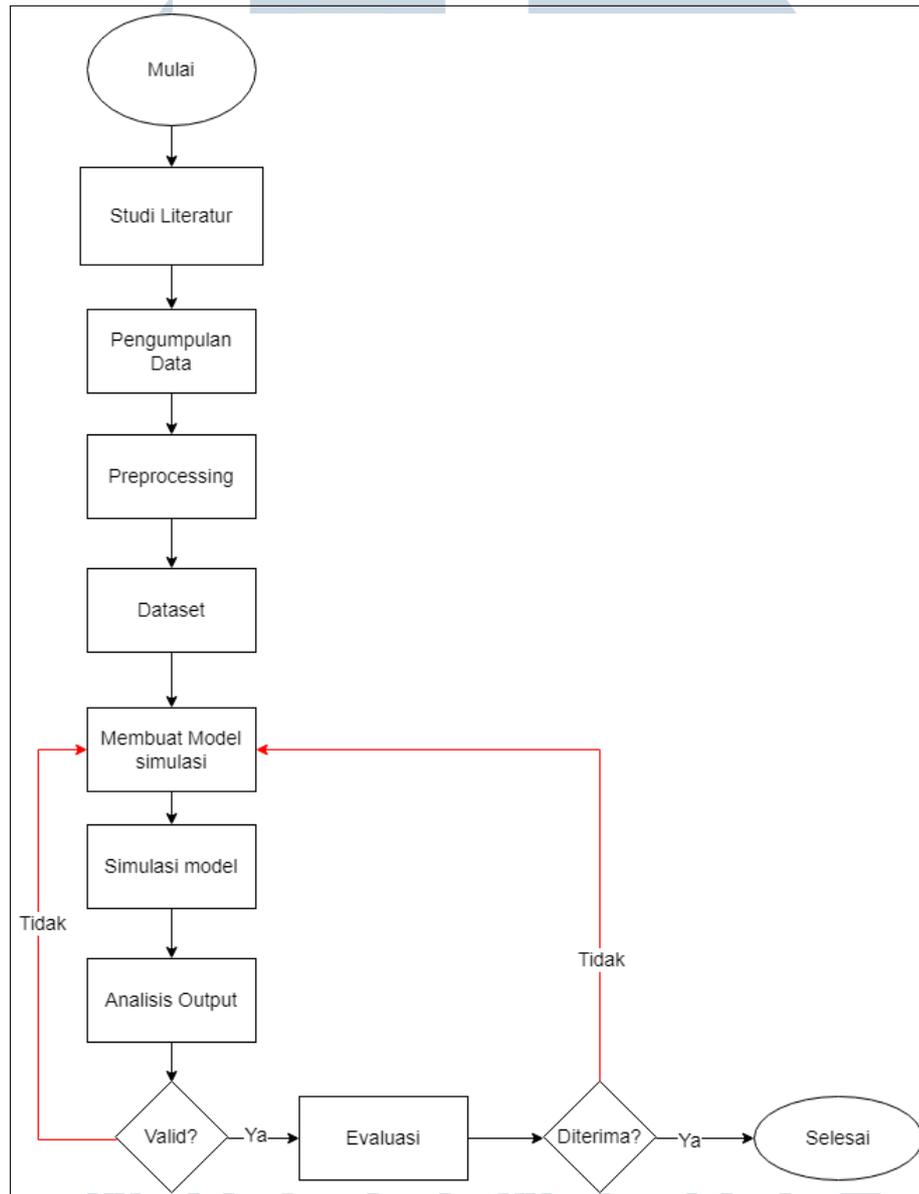
3.1 Analisis Kebutuhan

Pada sistem prediksi peningkatan kematian COVID-19 di Indonesia yang akan dibangun, terdapat beberapa kebutuhan yang dibutuhkan untuk menjadi tujuan dari sistem yang akan dibuat. Kebutuhan sistem prediksi yang akan dibuat berhubungan dengan algoritma yang digunakan yaitu *Linear Regression* pemilihan algoritma ini dipilih berdasarkan tingkat akurasi pada penelitian yang serupa sebelumnya, dan algoritma ini akan dipadukan dengan *Polynomial Features* dikarenakan *Polynomial Features* memiliki kecocokan terhadap pola nonlinear sehingga cukup efektif penggunaannya terhadap algoritma *Linear Regression*.

Setelah memahami algoritma yang digunakan akan dilanjutkan dengan tahapan pengambilan data yang diperlukan dalam penelitian. Data yang digunakan merupakan data yang telah disediakan oleh Kawalcovid.com, sehingga penelitian dapat dilanjutkan ke tahapan selanjutnya yaitu tahapan pembersihan data. Pada tahapan ini pembersihan data dilakukan dengan metode mengisi data yang kosong dengan data yang ditentukan. Data dinilai layak digunakan pada tahap pelatihan apabila data yang digunakan telah bersih dari data-data kosong, sehingga pada tahap pelatihan algoritma yang digunakan dapat menghasilkan hasil prediksi yang cukup akurat. Setelah dilakukan pelatihan data maka hasil prediksi akan dievaluasi kembali keakuratannya dengan menggunakan perhitungan *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE) dan *R Squared* (R^2).

3.2 Flowchart

Dalam melakukan prediksi tingkat kematian Covid-19 di Indonesia terdapat beberapa tahap yang harus dilakukan yaitu sebagai berikut yang telah digambarkan dalam flowchart :



Gambar 3.1. Flowchart implementasi Algoritma terhadap model

3.2.1 Pemahaman Study Literatur

Melakukan proses pemahaman berbagai macam literatur yang telah dipilih tentang konsep-konsep yang berkaitan dengan rumusan masalah, mempelajari setiap identifikasi, tujuan dan batasan masalah yang ada pada penelitian. Pemahaman studi literatur ini dimulai dari pemahaman tentang algoritma yang digunakan yaitu *Linear Regression* dan dalam tahap pembuatannya algoritma ini akan dipadukan kembali dengan fitur *Polynomial Features*. Pemilihan *Polynomial Features* dipilih dikarenakan kecocokan data dengan fitur ini dan hal tersebut juga didukung dengan beberapa referensi. Dan akan diakhiri dengan tahapan evaluasi yang akan digunakan pada penelitian ini.

3.2.2 Pengumpulan Data

Data dari penelitian ini diambil dari data excel yang telah disiapkan oleh Kawalcovid.com pada bulan desember 2021. Data yang digunakan terdiri dari : (1)Jumlah Kasus Meninggal harian, (2)Jumlah Kasus Sembuh, (3)Jumlah Kasus Aktif dan (4)Kasus Meninggal Harian [10] yang berjumlah 656 data untuk setiap variabel. Data yang digunakan dimulai dari tanggal 2 Maret 2020 hingga 19 Desember 2021. Data yang diambil dari kawalcovid.com telah dikonfirmasi keakuratan datanya dengan data yang ada pada website Covid19.go.id dengan cara pengecekan secara manual dengan bantuan dari Microsoft Excel untuk setiap data yang digunakan.

Covid19.go.id adalah website resmi dari pemerintah Indonesia yang berisikan data penyebaran COVID-19 di Indonesia dan berita-berita terbaru dari perkembangan virus COVID-19. Berikut adalah 5 data teratas yang digunakan pada penelitian ini :

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

data.head(5)

	Day	Tanggal	Kasus harian	Kasus Impor	Kasus Lokal	Total kasus	Kasus aktif	% kasus aktif	Sembuh\n(baru)	Sembuh	...	Unnamed: 77	Pos Rate DKI	Pos Rate DKI - 7 harian
0	1	2020-03-02	2	0.0	2.0	2	2	1.0	0	0	...	NaN	NaN	NaN
1	2	2020-03-03	0	0.0	0.0	2	2	1.0	0	0	...	NaN	NaN	NaN
2	3	2020-03-04	0	0.0	0.0	2	2	1.0	0	0	...	NaN	NaN	NaN
3	4	2020-03-05	0	0.0	0.0	2	2	1.0	0	0	...	NaN	NaN	NaN
4	5	2020-03-06	2	0.0	2.0	4	4	1.0	0	0	...	NaN	NaN	NaN

5 rows x 87 columns

Gambar 3.2. Dataset

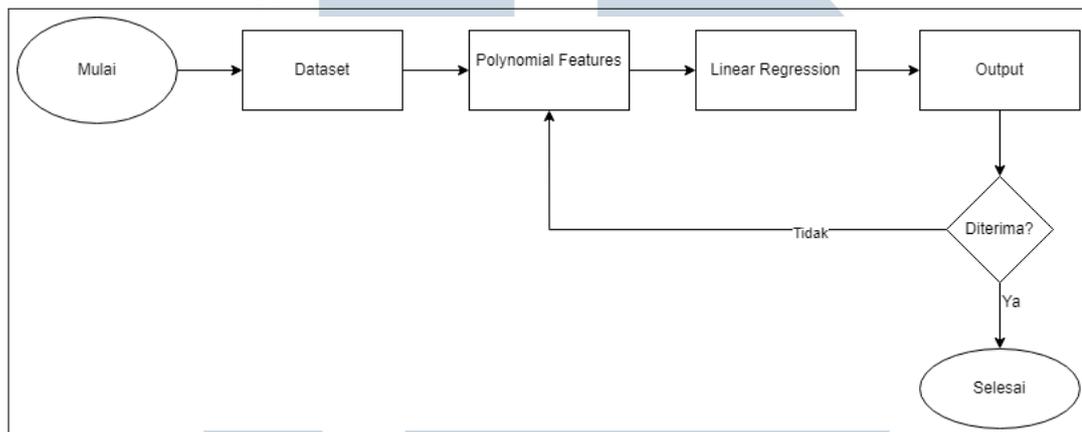
3.2.3 Pembersihan Data

Sebelum melakukan proses pelatihan data dilakukan *preprocessing* data yang bertujuan untuk meningkatkan akurasi dari proses pelatihan data yang akan dilakukan. Dalam melakukan proses *preprocessing* menggunakan metode *fillna*[19]. Metode ini berfungsi untuk mengisi data yang kosong dengan data yang ditentukan dengan mengembalikan objek *DataFrame* baru. Metode *fillna* dipilih dikarenakan data yang digunakan merupakan data *actual* sehingga tidak dapat dimanipulasi datanya sehingga pada proses ini setiap data *NULL* akan diisi dengan nilai 0, dan setiap data *NULL* yang diisi dengan nilai 0 akan dicek kembali secara manual terhadap setiap data yang diisi dengan nilai 0. Hal ini dilakukan untuk meningkatkan kualitas dan keakuratan data berdasarkan data dari website resmi pemerintah yaitu covid.go.id.

3.2.4 Pemodelan dan Pelatihan Data

Tahap pemodelan dan pelatihan data merupakan inti dari proses yang dilakukan pada *machine learning*. Pada penelitian ini, proses pelatihan dilakukan dengan menggunakan beberapa kali kombinasi parameter proses percobaan. Dataset yang telah disiapkan sebelumnya dan telah dipilih *variabel target* serta *variabel feature*. Dalam proses pelatihan data terhadap algoritma *Linear Regression* menggunakan parameter *Normalize* dengan nilai *True*, *fit_intercept* dengan nilai *True* dan *Degree* yang digunakan pada *Polynomial Feature* adalah bernilai 3, tujuan penggunaan parameter pada *polynomial feature* untuk menghasilkan proses dengan hasil yang maksimum sehingga dapat menghasilkan prediksi dengan tingkat akurasi

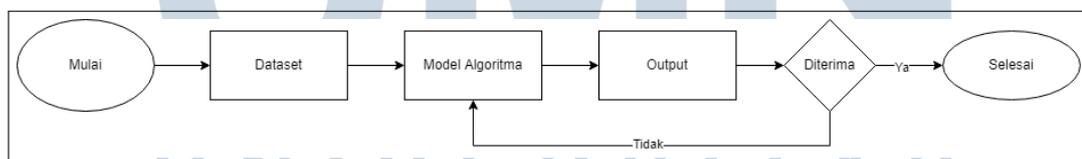
yang tinggi. Dalam proses pelatihan, pemilihan penggunaan *Polynomial Features* dikarenakan *Polynomial Features* memiliki kecocokan algoritma perhitungan data terhadap dataset yang digunakan dalam penelitian. Berikut merupakan flowchart pelatihan data terhadap model.



Gambar 3.3. Flowchart Pemodelan dan Pelatihan Data

3.2.5 Implementasi

Setelah dilakukan proses pelatihan data dengan menggunakan model algoritma *Linear Regression* dan *Polynomial Feature* bentuk data yang dilatih menghasilkan bentuk baru dikarenakan degree yang digunakan pada *polynomial feature*. Sehingga dari bentuk baru tersebut diharapkan dapat meningkatkan tingkat akurasi dari hasil prediksi yang dihasilkan. Berikut adalah flowchart pelatihan model terhadap dataset.



Gambar 3.4. Flowchart Model Latih