

BAB III METODE PENELITIAN

3.1 Bahan Penelitian

Bahan penelitian yang akan digunakan oleh peneliti kali ini merupakan data rekapitulasi kasus Covid-19 di Indonesia yang bersumber dari Kementerian Kesehatan Republik Indonesia yang didapatkan dengan cara mengakses *API* publik dari laman covid19.go.id dengan alamat tautan <https://data.covid19.go.id/public/api/update.json>. Data yang digunakan pada penelitian ini merupakan akumulasi kasus Covid-19 di Indonesia dari tanggal 1 September 2021 s.d. 31 Desember 2021.

Tabel 3. 1. Data Kasus Harian Covid-19 di Indonesia



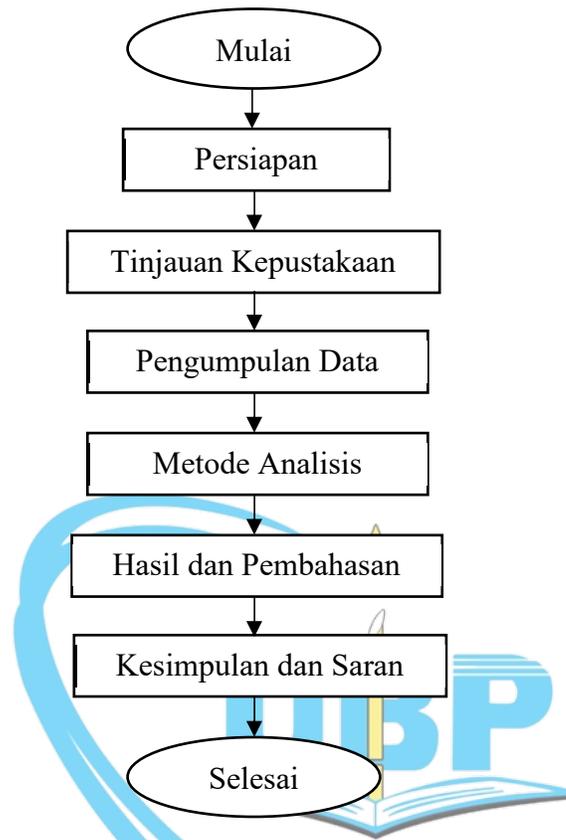
Tanggal	Jumlah Kasus Harian
01/09/2021	10329
02/09/2021	8355
03/09/2021	7797
04/09/2021	8727
05/09/2021	5458
...	...
27/12/2021	1131
28/12/2021	2128
29/12/2021	1654
30/12/2021	1528
31/12/2021	1420



3.2 Peralatan Penelitian

Penelitian yang akan dilakukan membutuhkan peralatan untuk mendukung proses penelitian. Peralatan-peralatan tersebut berupa kebutuhan perangkat lunak (*software*) dan perangkat keras (*hardware*).

3.4 Prosedur Penelitian



Gambar 3. 1 Diagram Prosedur Penelitian

Gambaran mengenai keseluruhan dari prosedur penelitian yang akan dilakukan dalam penelitian ini dituangkan dalam gambar 3.1. Penjelasan mengenai setiap tahapan yaitu:

1. Persiapan

Tahap persiapan merupakan kegiatan yang akan dilakukan sebelum melakukan penelitian, yaitu dengan menyiapkan alat serta bahan yang akan digunakan dalam penelitian, seperti persiapan kebutuhan piranti *hardware* maupun *software*.

2. Tinjauan Kepustakaan

Tinjauan Kepustakaan dilakukan dengan mencari referensi baik buku ataupun jurnal yang berkaitan dengan objek penelitian dan metode yang digunakan dalam penelitian kali ini.

3. Pengumpulan Data

Pengumpulan data dilakukan dengan mengakses *API* publik pada laman Covid19.go.id

4. Metode Analisis

Metode analisis dilakukan dengan mengikuti kaidah dari proses data *mining* atau metode KDD (*Knowledge Discovery in Database*)

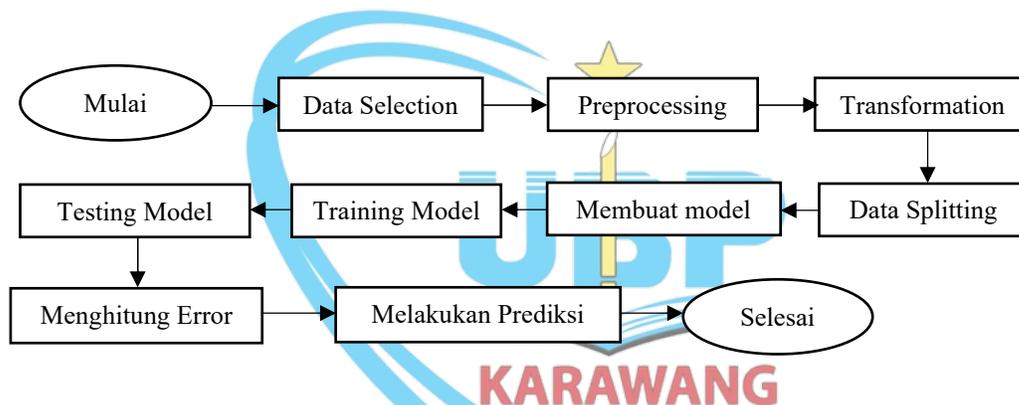
5. Hasil dan Pembahasan

Pada tahap ini menjelaskan hasil dari proses data *mining* sesuai dengan rumusan masalah yang dituangkan dalam penelitian ini.

6. Kesimpulan dan Saran

Penulis membuat kesimpulan dari hasil penelitian dan memberikan saran untuk penelitian selanjutnya.

3.5 Metode Analisis



Gambar 3. 2 Metode Analisis

Metode Analisis pada penelitian ini mengikuti tahapan dari proses data *mining* atau metode KDD (*Knowledge Discovery in Database*), tahapannya sebagai berikut.

3.5.1 Data Selection

Pemilihan (seleksi) data merupakan langkah awal sebelum tahap penggalian informasi. Pada tahap data *selection* dilakukan pengumpulan data. Data yang digunakan pada penelitian ini bersumber dari API publik dari laman covid19.go.id yang berbentuk file json, dengan alamat tautan <https://data.covid19.go.id/public/api/update.json>, data tersebut merupakan akumulasi data Covid-19 di Indonesia. Data yang digunakan yaitu dari tanggal 1 September 2021 s.d. 31 Desember 2021.

```
... : "key_as_string": "2020-03-21T09:09:49.000Z",
... : "key": "202003210909",
... : "doc_count": 1,
... : "jumlah_meninggal": {
... :   "value": 0
... : },
... : "jumlah_sembuh": {
... :   "value": 0
... : },
... : "jumlah_positif": {
... :   "value": 2
... : },
... : "jumlah_dirawat": {
... :   "value": 2
... : },
... : "jumlah_positif_kum": {
... :   "value": 2
... : },
... : "jumlah_sembuh_kum": {
... :   "value": 0
... : }
```

Gambar 3. 3 Data Covid-19 dari Kementerian Kesehatan Indonesia

Berdasarkan Gambar 3.3, data memiliki beberapa atribut diantaranya *key_as_string*, *key*, *doc_count*, *jumlah_meninggal*, *jumlah_sembuh*, *jumlah_positif*, *jumlah_dirawat*, *jumlah_positif_kum*, *jumlah_sembuh_kum*, *jumlah_meninggal_kum* dan *jumlah_dirawat_kum*. Dari banyak atribut tersebut akan diseleksi dan hanya dua atribut yang akan digunakan untuk proses data *mining*, atribut tersebut yaitu sebagai berikut.

1. *Key_as_string* merupakan atribut berisi informasi waktu penambahan kasus Covid-19
2. *Jumlah_positif* merupakan atribut yang berisi informasi jumlah penambahan kasus positif Covid-19

3.5.2 Preprocessing

Preprocessing merupakan tahap untuk memperbaiki kualitas data sebelum dilakukan proses *mining*. Pada tahap ini dilakukan pembersihan data pada sebuah data yang akan digunakan agar data menjadi berkualitas, seperti membuang data yang tidak diperlukan dan memperbaiki kesalahan pada data. Pada umumnya, data yang diperoleh dari sumbernya memiliki isian-isian yang tidak sempurna seperti data yang hilang, *missing value* atau hanya sekedar salah ketik. Data-data yang tidak relevan itu juga lebih baik dibuang karena keberadaannya bisa mengurangi akurasi dari hasil data *mining* nantinya. Pembersihan data juga akan mempengaruhi performansi dari sistem data *mining* karena data yang akan ditangani akan berkurang jumlah dan kompleksitasnya. Untuk mengetahui ada atau tidaknya data yang salah, maka dilakukan visualisasi data.

3.5.3 Transformation

Pada tahap *transformation* dilakukan perubahan tipe data yang digunakan. Perubahan tipe data dilakukan agar data dapat diproses saat proses *mining*. Proses transformasi merupakan proses yang sangat tergantung pada jenis atau pola informasi yang akan dicari dalam data. Pada tahap ini transformasi dilakukan pada atribut prediktor (x), sehingga data siap digunakan dalam proses selanjutnya.

3.5.4 Data Splitting

Pada tahap Data Splitting dilakukan pemisahan data yang digunakan untuk memisahkan dari *independent* variabel dan *dependent* variabel.

3.5.5 Pembuatan Model

Pada tahap ini dilakukan pembuatan model untuk melakukan sebuah prediksi. Metode yang digunakan yaitu Linear Regresi dan Polynomial Regresi. Regresi Linear adalah analisis statistika yang memodelkan hubungan beberapa variabel menurut bentuk hubungan persamaan linier eksplisit (Nafi'iyah, 2016). Sedangkan Regresi Polynomial merupakan model regresi linear yang dibentuk dengan menjumlahkan pengaruh masing-masing variabel prediktor (X) yang dipangkatkan meningkat sampai orde ke-k (Malensang., dkk, 2013).

3.5.6 Training Model

Pada tahap ini dilakukan proses *training* terhadap model yang sudah dibuat terhadap data train yang telah ditentukan saat data splitting.

3.5.7 Testing Model

Setelah model melakukan *training*, selanjutnya model digunakan untuk melakukan sebuah prediksi. Data hasil prediksi nantinya akan dibandingkan dengan data testing yang telah ditentukan saat proses data *splitting*

3.5.8 Menghitung Nilai Error

Hasil prediksi pada tahap sebelumnya, akan dibandingkan dengan data *testing* untuk mencari nilai error dari setiap prediksi yang telah dilakukan. Perhitungan dilakukan dengan cara menghitung nilai dari Koefisien Determinasi (R^2). Dari hasil ini, maka nantinya dapat kita ketahui metode yang lebih baik dalam melakukan sebuah prediksi.

3.5.9 Melakukan Prediksi

Setelah menemukan metode yang lebih baik diantara linear regresi atau polynomial regresi, tahap selanjutnya yaitu melakukan prediksi terhadap kasus harian Covid-19 pada bulan Januari s.d. Maret 2022 di Indonesia. Prediksi dapat dilakukan dengan menentukan atribut prediktor (x) untuk dimasukkan kedalam model yang telah dibuat pada pembuatan model.

