

Model Prediksi Kasus Covid-19 di Indonesia Menggunakan Metode Linear Regresi dan Polynomial Regresi

1st Amid Rakhman
Universitas Buana Perjuangan
Karawang, Indonesia
if18.amidrakhman@mhs.ubpkarawang.ac.id

2nd Yana Cahyana
Universitas Buana Perjuangan
Karawang, Indonesia
yana.cahyana@ubpkarawang.ac.id

3rd Rahmat
Universitas Buana Perjuangan
Karawang, Indonesia
rahmat@ubpkarawang.ac.id

Abstract—

Organisasi Kesehatan Dunia pada 11 Maret 2020 telah mengumumkan bahwa *Coronavirus Disease 2019* (Covid-19) sebagai sebuah pandemi. Covid-19 ialah sebuah penyakit yang disebabkan oleh jenis *coronavirus* baru yaitu *Sars-CoV-2* yang mengganggu sistem pernafasan. Hingga saat ini kasus terkonfirmasi positif Covid-19 di Indonesia masih terjadi setiap harinya. Penelitian ini bertujuan untuk melakukan prediksi penambahan kasus Covid-19 di Indonesia. Data yang digunakan bersumber dari *API* publik laman covid19.go.id berupa penambahan jumlah kasus Covid-19 di Indonesia sebanyak 122 baris data. Prediksi dilakukan dengan menggunakan metode linear regresi dan polynomial regresi sebagai pembanding. Evaluasi pada metode linear regresi mendapatkan nilai $R^2 = 0,57$, sedangkan metode polynomial regresi mendapatkan nilai $R^2 = 0,84$. Berdasarkan evaluasi tersebut, metode polynomial regresi mendapatkan hasil yang lebih baik dibandingkan dengan metode linear regresi. Prediksi kasus Covid-19 di Indonesia pada bulan Januari sampai bulan Maret 2022 menggunakan metode polynomial regresi diprediksi penambahan kasus Covid-19 akan naik kembali.

Kata kunci — Covid-19, Prediksi, Linear Regresi, Polynomial Regresi

I. PENDAHULUAN

Pada 11 Maret 2020 Organisasi Kesehatan Dunia telah mengumumkan bahwa *Coronavirus Disease 2019* (Covid-19) sebagai sebuah pandemi. Covid-19 ialah sebuah penyakit yang diakibatkan oleh jenis *coronavirus* baru yaitu *Sars-CoV-2* yang mengganggu sistem pernafasan sehingga mengakibatkan radang pada paru-paru [1]. Penyakit ini pertama kali muncul pada awal bulan Desember 2019 di Kota Wuhan, Tiongkok. Sementara di Indonesia, Covid-19 pertama kali terdeteksi pada 2 Maret 2020 [2]. Semenjak itu kasus terkonfirmasi positif Covid-19 terus bertambah setiap harinya seiring dengan semakin banyaknya interaksi di masyarakat serta pelanggaran kebijakan yang dibuat oleh pemerintah terkait aktivitas saat pandemi Covid-19.

Data mengenai kondisi Covid-19 di Indonesia tentunya dapat dijadikan sebagai bahan pertimbangan dalam menentukan kebijakan-kebijakan yang dilakukan oleh pemerintah, karena dalam perkembangan teknologi saat ini, data merupakan bagian yang sangat penting dan tidak bisa dipisahkan dalam pengambilan sebuah informasi. Informasi terkait kasus Covid-19 pada masa yang akan datang dapat diperoleh dengan proses data mining berdasarkan data yang telah terjadi sebelumnya. Data mining merupakan aktivitas untuk menganalisis data dari sudut pandang yang berbeda dan menyimpulkannya menjadi sebuah informasi atau pengetahuan penting [3]. Salah satu penerapan data mining yaitu untuk memprediksi suatu kejadian pada masa yang akan datang berdasarkan data yang telah terjadi dengan metode tertentu. Beberapa metode yang dapat digunakan untuk melakukan sebuah peramalan atau prediksi, yaitu Linear Regresi dan Polynomial Regresi. Regresi Linear adalah analisis statistika yang memodelkan hubungan beberapa variabel menurut bentuk hubungan persamaan linier eksplisit [4]. Sedangkan Regresi Polynomial merupakan model regresi linear yang dibentuk dengan menjumlahkan pengaruh masing-masing variabel prediktor (X) yang dipangkatkan meningkat sampai orde ke-k [5].

Berdasarkan penelitian sebelumnya yang telah dilakukan, penggunaan metode linear mendapatkan hasil dengan kriteria sangat baik dalam memprediksi penjualan smartphone [6]. Pada penelitian lain dengan metode yang sama juga mendapatkan hasil dengan kategori sangat baik dalam memprediksi penjualan properti pada PT XYZ [7]. Sementara, penelitian menggunakan metode polynomial regresi dapat berjalan dengan baik dalam memprediksi kasus harian Covid-19 di DKI Jakarta [8]. Penggunaan metode polynomial regresi dalam memprediksi produksi daging sapi nasional juga mendapatkan nilai yang baik berdasarkan dataset produksi daging sapi nasional tahun 1984 s.d. 2019 [9].

Berdasarkan permasalahan yang telah dipaparkan dan penelitian yang telah membuktikan bahwa metode linear regresi dan polynomial regresi mendapatkan hasil dengan kategori yang baik dalam melakukan sebuah peramalan atau prediksi. Sehingga pada penelitian ini dilakukan prediksi kasus Covid-19 di Indonesia menggunakan metode linear regresi dan polynomial regresi.

II. DATA DAN METODE

A. Covid-19

Covid-19 merupakan suatu penyakit radang paru-paru yang terjadi karena seseorang terinfeksi oleh virus korona. Virus korona sendiri terdapat banyak macamnya dan yang dapat menyebabkan seseorang terkena penyakit Covid-19 yaitu jenis *SARS Corona Virus-2*. Virus ini memiliki ukuran yang sangat kecil, berkisar antara 50 s.d. 200 nanometer [10].

B. Prediksi

Prediksi atau peramalan adalah suatu kegiatan yang dilakukan untuk meramalkan atau memperkirakan kejadian pada waktu yang akan datang dengan menggunakan metode atau pendekatan ilmu tertentu. Data yang digunakan untuk melakukan suatu prediksi atau peramalan biasanya merupakan data kuantitatif [11].

Evaluasi pada sebuah model prediksi bisa dilakukan dengan cara menghitung nilai dari Koefisien Determinasi (R^2). Koefisien determinasi merupakan ukuran statistik yang mengindikasikan pengaruh yang diberikan variabel independen atau bebas (X) terhadap variabel dependen atau terikat (Y). Koefisien determinasi bernilai antara 0 sampai dengan 1. Apabila nilai R^2 semakin besar atau mendekati nilai 1 maka semakin kuat pengaruh variabel bebas (X) terhadap variabel terikat (Y). Menurut Eka dkk [9], Bentuk umum dari koefisien determinasi dapat dilihat pada persamaan 1 berikut.

$$R^2 = \frac{SSR}{SST} = \frac{\sum(y' - \bar{y})^2}{\sum(y - \bar{y})^2} \quad (1)$$

dengan SSR merupakan jumlah kuadrat dari selisih nilai prediksi (y') dengan nilai aktual rata-rata (\bar{y}), sementara SST merupakan jumlah kuadrat dari selisih nilai aktual (y) dengan nilai aktual rata-rata (\bar{y}).

C. Linear Regresi

Linear Regresi adalah analisis statistika yang dilakukan untuk melakukan peramalan atau prediksi dengan memodelkan hubungan antara beberapa variabel, yaitu variabel dependen (Y) dengan variabel independen (X). Menurut Ayuni dan Fitriana [7] rumus untuk regresi linear sederhana dijelaskan pada persamaan berikut.

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \quad (2)$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (3)$$

$$y = a + bx \quad (4)$$

dengan y adalah kuantiti kasus, x adalah periode waktu, n adalah banyaknya data, a adalah intersep dan b adalah slope.

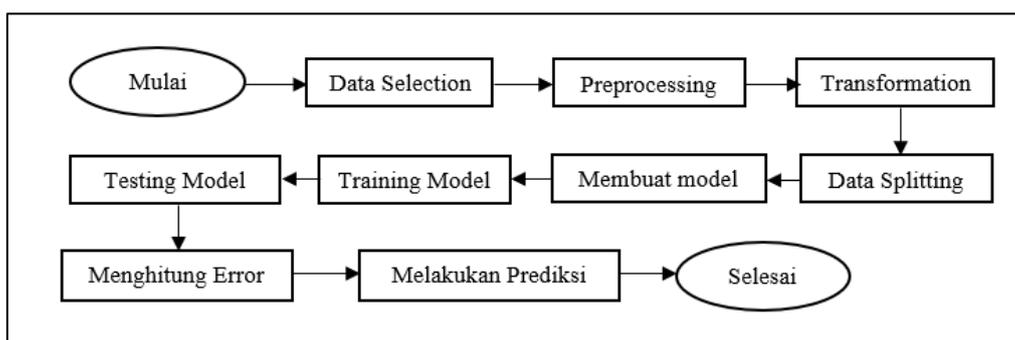
D. Polynomial Regresi

Regresi polynomial merupakan suatu jenis regresi khusus yang terdapat pada hubungan lengkung (*curvilinear*) antara nilai dependen (Y) dan nilai independen (X). Regresi Polynomial ialah model Regresi Linier yang dimodelkan dengan menjumlahkan pengaruh dari setiap variabel prediktor (X) yang dipangkatkan sampai orde ke-n. Secara umum, menurut Eka dkk [9] model Regresi Polynomial ditulis dalam persamaan berikut.

$$Y = b_0 + b_1X + b_2X^2 + \dots + b_nX^n \quad (5)$$

dimana Y merupakan variabel yang diprediksi, b_0 merupakan intersep, b_1, b_2, \dots, b_n merupakan slope atau koefisien-koefisien regresi, X merupakan variabel bebas, n merupakan orde atau derajat polynomial.

E. Gambaran Umum Penelitian



Gambar 1. Gambaran Umum Penelitian

Penelitian dimulai dengan melakukan seleksi data yang diakses melalui API publik dari laman covid19.go.id. Data yang diseleksi sebanyak 122 baris data dengan 2 atribut, yaitu atribut periode waktu (X) dan jumlah kasus Covid-19 (Y). Kemudian melakukan *preprocessing* pada data dan melakukan transformasi pada atribut periode waktu (variabel X). Selanjutnya dilakukan pembuatan model dengan metode Linear Regresi dan Polynomial Regresi pada data. Setelah melakukan pembuatan model, kemudian dilakukan *testing* data dan menghitung nilai *error* dengan menggunakan koefisien determinasi (R^2) pada kedua model. Model yang lebih baik kemudian digunakan untuk memprediksi kasus penambahan Covid-19 pada bulan Januari s.d. Maret 2022.

III. HASIL DAN PEMBAHASAN

A. Hasil Penelitian

Pada tahap data *selection* dilakukan seleksi data yang 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': {'id': 1,
'jumlah_odp': 2992,
'jumlah_pdp': 0,
'total_spesimen': 102088338,
'total_spesimen_negatif': 88673512},
'update': {'harian': [{'doc_count': 1,
'jumlah_dirawat': {'value': 2},
'jumlah_dirawat_kum': {'value': 2},
'jumlah_meninggal': {'value': 0},
'jumlah_meninggal_kum': {'value': 0},
'jumlah_positif': {'value': 2},
'jumlah_positif_kum': {'value': 2},
'jumlah_sembuh': {'value': 0},
'jumlah_sembuh_kum': {'value': 0},
'key': 1583107200000,
'key_as_string': '2020-03-02T00:00:00.000Z'},
{'doc_count': 1,
'jumlah_dirawat': {'value': 0},
'jumlah_dirawat_kum': {'value': 2},
'jumlah_meninggal': {'value': 0},
'jumlah_meninggal_kum': {'value': 0},
'jumlah_positif': {'value': 0},
'jumlah_positif_kum': {'value': 2},
'jumlah_sembuh': {'value': 0},
'jumlah_sembuh_kum': {'value': 0},
'key': 1583193600000,
'key_as_string': '2020-03-03T00:00:00.000Z'},
```

Gambar 2. Data Asli dari *API* publik laman covid19.go.id

Data yang telah diseleksi berjumlah 122 baris data dari tanggal 1 September 2021 s.d. 31 Desember 2021. Data terdiri dari dua atribut yaitu periode waktu (X) dan jumlah positif (Y)

	waktu	jumlah positif
0	2021-09-01	10337
1	2021-09-02	8955
2	2021-09-03	7797
3	2021-09-04	6727
4	2021-09-05	5403
...
117	2021-12-27	120
118	2021-12-28	278
119	2021-12-29	194
120	2021-12-30	189
121	2021-12-31	180

Gambar 3. Data Hasil Seleksi

Transformasi data pada atribut waktu menjadi index waktu berdasarkan nilai atribut waktu itu sendiri dikurangi dengan nilai minimum dari atribut waktu tersebut agar data dapat diproses saat melakukan pembuatan model.

	waktu	index waktu	jumlah positif
0	2021-09-01	0	10337
1	2021-09-02	1	8955
2	2021-09-03	2	7797
3	2021-09-04	3	6727
4	2021-09-05	4	5403

Gambar 4. Transformasi Data

Pembuatan Model dilakukan dengan metode Linear Regresi dan Polynomial Regresi dengan menggunakan python. Atribut index waktu sebagai variabel X dan jumlah positif sebagai variabel Y.

```
linear = LinearRegression()
linear = linear.fit(x, y)
```

Gambar 5. Pembuatan Model Linear Regresi

```
print('Nilai intercept (a): ', linear.intercept_)
print('Nilai slope (b): ', linear.coef_)

Nilai intercept (a): 3968.209516193522
Nilai slope (b): [-42.1626556]
```

Gambar 6. Nilai Intercept (a) dan Nilai Slope (b) Model Linear Regresi

Perhitungan manual model Linear Regresi:

Tabel 1. Perhitungan Manual Linear Regresi

Waktu	x	y	xy	x ²
01/09/2021	0	10337	0	0
02/09/2021	1	8955	8955	1
03/09/2021	2	7797	15594	4
04/09/2021	3	6727	20181	9
...
...
29/12/2021	119	194	23086	14161
30/12/2021	120	189	22680	14400
31/12/2021	121	180	21780	14641
Jumlah	7381	172919	4081947	597861

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$= \frac{(172919 \cdot 597861) - (7381 \cdot 4081947)}{(122 \cdot 597861) - (7381)^2}$$

$$= 3968,21$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$= \frac{(122 \cdot 4081947) - (7381 \cdot 172919)}{(122 \cdot 597861) - (7381)^2}$$

$$= -42,16$$

Sehingga, rumus untuk mencari nilai prediksi pada model Linear Regresi sebagai berikut.

$$y = 3968,21 - 42,16x$$

```
poly = PolynomialFeatures()
x_polynomial = poly.fit_transform(x)

polynomial = LinearRegression()
polynomial.fit(x_polynomial, y)

LinearRegression()
```

Gambar 7. Pembuatan Model Polynomial Regresi

```
print('Nilai intercept: ', polynomial.intercept_)
print('Nilai slope: ', polynomial.coef_)

Nilai intercept: 6196.332170357626
Nilai slope: [ 0.          -153.56878831   0.92071184]
```

Gambar 8. Nilai Intercept dan Nilai Slope Model Polynomial Regresi

Perhitungan manual pada Polynomial Regresi:

Tabel 2. Perhitungan Manual Polynomial Regresi

waktu	x	y	xy	x ²	x ³	x ⁴	x ² y
01/09/2021	0	10337	0	0	0	0	0
02/09/2021	1	8955	8955	1	1	1	8955
03/09/2021	2	7797	15594	4	8	16	31188
04/09/2021	3	6727	20181	9	27	81	60543
...
...
29/12/2021	119	194	23086	14161	1685159	200533921	2747234
30/12/2021	120	189	22680	14400	1728000	207360000	2721600
31/12/2021	121	180	21780	14641	1771561	214358881	2635380
Jumlah	7381	172919	4081947	597861	54479161	5295254877	213650469

$$\begin{bmatrix} n & \sum x & \sum x^2 \\ \sum x & \sum x^2 & \sum x^3 \\ \sum x^2 & \sum x^3 & \sum x^4 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum x^0 y \\ \sum x^1 y \\ \sum x^2 y \end{bmatrix}$$

Penyelesaian menggunakan metode Eliminasi Gauss:

$$\begin{bmatrix} 122 & 7381 & 597861 \\ 7381 & 597861 & 54479161 \\ 597861 & 54479161 & 5295254877 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 172919 \\ 4081947 \\ 213650469 \end{bmatrix}$$

$$\left| \begin{array}{ccc|c} 1 & 0 & 0 & 6196,33217 \\ 0 & 1 & 0 & -153,56879 \\ 0 & 0 & 1 & 0,92071184 \end{array} \right|$$

b₀ = 6196,33217

b₁ = -153,56879

b₂ = 0,92071184

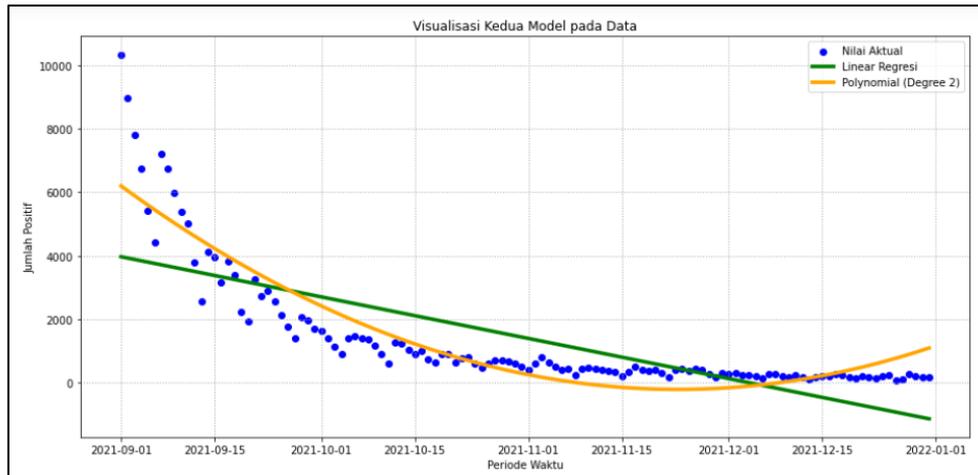
Sehingga, rumus untuk mencari nilai prediksi pada model Polynomial Regresi sebagai berikut.

$$y = 6196,33 - 153,57x + 0,92x^2$$

Testing Model dilakukan pada data dengan melakukan prediksi menggunakan persamaan pada Model Linear Regresi dan Polynomial Regresi yang telah diketahui.

	waktu	index waktu	jumlah positif	prediksi linear regresi	prediksi polynomial regresi
0	2021-09-01	0	10337	3968.209516	6196.332170
1	2021-09-02	1	8955	3926.046861	6043.684094
2	2021-09-03	2	7797	3883.884205	5892.877441
3	2021-09-04	3	6727	3841.721549	5743.912212
4	2021-09-05	4	5403	3799.558894	5596.788407

Gambar 9. Hasil Testing Model Pada Linear Regresi dan Polynomial Regresi



Gambar 10. Visualisasi Model Linear Regresi dan Polynomial Regresi

Evaluasi Model Linear Regresi dan Polynomial Regresi dengan menghitung nilai koefisien determinasi (R^2).

```
r2_linear = r2_score(y, y_pred_linear)
r2_polynomial = r2_score(y, y_pred_polynomial)

print('R2 Score Model Linear Regresi :', r2_linear, '\n')
print('R2 Score Model Polynomial Regresi :', r2_polynomial)

R2 Score Model Linear Regresi : 0.5688755069698165
R2 Score Model Polynomial Regresi : 0.8379796911399235
```

Gambar 11. Evaluasi Model Linear Regresi dan Polynomial Regresi

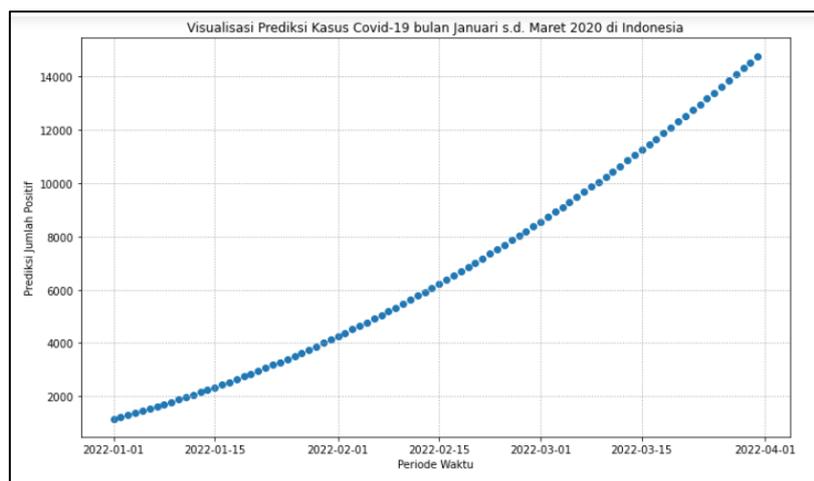
Berdasarkan nilai koefisien determinasi (R^2) pada kedua model tersebut, model dengan menggunakan metode Linear Regresi mendapatkan nilai $R^2 = 0,57$, sementara model dengan menggunakan metode Polynomial Regresi mendapatkan nilai $R^2 = 0,84$. Sehingga dapat dinyatakan bahwa model dengan menggunakan metode Polynomial Regresi dapat memprediksi lebih baik dibandingkan dengan metode Linear Regresi. Kemudian dilakukan prediksi untuk kasus penambahan Covid-19 di Indonesia pada bulan Januari s.d. Maret 2022 dengan menggunakan Polynomial Regresi.

```
x_prediksi = data_prediksi['index waktu'].values.reshape(-1,1)
poly = PolynomialFeatures()
x_prediksi = poly.fit_transform(x_prediksi)
y_prediksi = polynomial.predict(x_prediksi)

data_prediksi['prediksi jumlah positif'] = np.array(y_prediksi)
data_prediksi.head()
```

	date	index waktu	prediksi jumlah positif
0	2022-01-01	122	1164.815032
1	2022-01-02	123	1236.820644
2	2022-01-03	124	1310.667681
3	2022-01-04	125	1386.356141
4	2022-01-05	126	1463.886024

Gambar 12. Hasil Prediksi Bulan Januari s.d. Maret 2022



Gambar 13. Visualisasi Hasil Prediksi Bulan Januari s.d. Maret 2022

B. Pembahasan Penelitian

Evaluasi pada Model Prediksi kasus Covid-19 di Indonesia menggunakan metode Linear Regresi mendapatkan nilai $R^2 = 0,57$, sementara metode Polynomial Regresi mendapatkan nilai $R^2 = 0,84$. Sehingga, metode Polynomial Regresi mendapatkan nilai yang lebih baik dalam memprediksi kasus Covid-19 di Indonesia dibandingkan dengan metode Linear Regresi. Metode Polynomial Regresi digunakan untuk memprediksi penambahan kasus Covid-19 di Indonesia pada bulan Januari 2022.

IV. KESIMPULAN DAN SARAN

A. KESIMPULAN

Berdasarkan hasil dan pembahasan dari penelitian, dapat disimpulkan bahwa metode Polynomial Regresi mendapatkan nilai yang lebih baik dibandingkan dengan metode Linear Regresi. Evaluasi pada metode Regresi mendapatkan nilai $R^2 = 0,57$, sedangkan pada metode Polynomial Regresi mendapatkan nilai $R^2 = 0,84$. Hasil prediksi kasus Covid-19 di Indonesia pada bulan Januari s.d. Maret 2022 menggunakan Polynomial Regresi diprediksi penambahan kasus Covid-19 akan naik kembali.

B. SARAN

Saran bagi penelitian selanjutnya diharapkan menggunakan metode yang lain dan data yang lebih banyak agar menghasilkan model yang lebih baik. Bagi pemerintah diharapkan bisa dijadikan sebagai bahan pertimbangan dalam menentukan kebijakan-kebijakan selanjutnya mengenai penanganan Covid-19 di Indonesia. Bagi masyarakat diharapkan meningkatkan kewaspadaan, karena kasus Covid-19 di Indonesia masih terus terjadi hingga saat ini.

PENGAKUAN

Naskah ilmiah ini adalah penelitian tugas akhir milik Amid Rakhman dengan judul Model Prediksi Kasus Covid-19 di Indonesia Menggunakan Metode Linear Regresi dan Polynomial Regresi yang dibimbing oleh Yana Cahyana dan Rahmat.

DAFTAR PUSTAKA

- [1] Wahyudin dan Heri Purwanto. (2021). *Prediksi Kasus Covid-19 Di Indonesia Menggunakan Metode Backpropagation Dan Regresi Linear*. Journal of Information System, Applied, Management, Accounting and Research, 5(2), 331. <https://doi.org/10.52362/jisamar.v5i2.420>.
- [2] Widyarsi, L., dkk. (2021). *Prediksi Kasus Covid-19 Melalui Analisis Data Google Trend Di Indonesia : Pendekatan Metode Long Short Term Memory (LSTM)*. JURNAL SAINTIKA UNPAM : Jurnal Sains Dan Matematika Unpam, 3(2), 161–177.
- [3] Mabror, Angga Ginanjar dan Riani Lubis. (2012). *Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit*. Jurnal Komputer Dan Informatika (KOMPUTA), 1(1), 53–57.
- [4] Nafi'iyah, N. (2016). *Perbandingan Regresi Linear , Backpropagation Dan Fuzzy Mamdani Dalam Prediksi Harga Emas*. Seminar Nasional Inovasi Dan Aplikasi Teknologi Di Industri, 291–296.
- [5] Malensang, J. S., H Komalig., & D. Hatidja. (2013). *Pengembangan Model Regresi Polinomial Berganda Pada Kasus Data Pemasaran*. Jurnal Ilmiah Sains, 12(2), 149. <https://doi.org/10.35799/jis.12.2.2012.740>.
- [6] Indarwati, T., T. Irawati., dan E. Rimawati. (2019). *Penggunaan Metode Linear Regression Untuk Prediksi Penjualan Smartphone*. Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN), 6(2), 2–7. <https://doi.org/10.30646/tikomsin.v6i2.369>.
- [7] Ayuni, Ghebyla Najla dan Devi Fitrihanah. (2019). *Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ*. Jurnal Telematika, 14(2), 79–86. <https://journal.ithb.ac.id/telematika/article/view/321>
- [8] Rory dan Rita Diana. (2020). *Pemodelan Data Covid-19 Menggunakan Regresi Polinomial Lokal*. Seminar Nasional Official Statistics 2020, 2, 91–98.
- [9] Putra, Andika Eka dan Asep Juarna. (2021). *Prediksi Produksi Daging Sapi Nasional dengan Metode Regresi Linier dan Regresi Polinomial*. Jurnal Ilmiah KOMPUTASI, 20, 209–215.

- [10] Sabrina, D. S., dkk. (2020). *Buku Praktis Penyakit Virus Corona 19 (COVID-19)*. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9).
- [11] Nugroho, K. (2016). *Model Analisis Prediksi Menggunakan Metode Fuzzy Time Series*. *Infokam*, 12(1), 46–50.